

**UNIVERSIDADE FEDERAL DE VIÇOSA**

**Predição da vazão e da qualidade da água subterrânea em região semiárida**

Maria Fernanda Antunes Collares  
*Doctor Scientiae*

**VIÇOSA - MINAS GERAIS  
2025**

**MARIA FERNANDA ANTUNES COLLARES**

**Predição da vazão e da qualidade da água subterrânea em região semiárida**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Engenharia Agrícola, para obtenção do título de *Doctor Scientiae*.

Orientador: Andre Pereira Rosa

**VIÇOSA - MINAS GERAIS  
2025**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade  
Federal de Viçosa - Campus Viçosa**

T

C697p  
2025 Collares, Maria Fernanda Antunes, 1995-  
Predição da vazão e da qualidade da água subterrânea em  
região semiárida / Maria Fernanda Antunes Collares. – Viçosa,  
MG, 2025.

1 dissertação eletrônica (141 f.): il. (algumas color.).

Orientador: André Pereira Rosa.

Tese (doutorado) - Universidade Federal de Viçosa,  
Departamento de Engenharia Agrícola, 2025.

Inclui bibliografia.

DOI: <https://doi.org/10.47328/ufvbbt.2026.114>

1. Águas subterrâneas - Qualidade. 2. Regiões áridas -  
Minas Gerais. 3. Aprendizado do computador. 4. Salinidade.  
I. Rosa, André Pereira, 1983-. II. Universidade Federal de  
Viçosa. Departamento de Engenharia Agrícola. Programa de  
Pós-Graduação em Engenharia Agrícola. III. Título.

CDD 22. ed. 628.114

**MARIA FERNANDA ANTUNES COLLARES**

**Predição da vazão e da qualidade da água subterrânea em região semiárida**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Engenharia Agrícola, para obtenção do título de *Doctor Scientiae*.

APROVADA: 26 de agosto de 2025.

Assentimento:

---

Maria Fernanda Antunes Collares  
Autora

---

Andre Pereira Rosa  
Orientador

Essa tese foi assinada digitalmente pela autora em 20/05/2026 às 12:08:37 e pelo orientador em 21/05/2026 às 09:21:50. As assinaturas têm validade legal, conforme o disposto na Medida Provisória 2.200-2/2001 e na Resolução nº 37/2012 do CONARQ. Para conferir a autenticidade, acesse <https://siadoc.ufv.br/validar-documento>. No campo 'Código de registro', informe o código **IN9Z.JN4N.11UF** e clique no botão 'Validar documento'.

A quem me amou, acreditou e esperou junto comigo.

## AGRADECIMENTOS

Ao concluir esta etapa tão significativa da minha trajetória, agradeço primeiramente a Deus, por sustentar meus passos com amor, fé e esperança, mesmo nos dias mais difíceis.

Ao meu orientador, professor André Pereira Rosa, meu sincero agradecimento pela orientação segura, pelas discussões sempre enriquecedoras, pela paciência, incentivo e confiança depositada em mim ao longo de todo o desenvolvimento deste trabalho. Obrigada por acreditar que seria possível!

À minha família, meu alicerce e motivação diária. À minha mãe Tânia, por todo amor incondicional, força e exemplo de dedicação. À minha irmã Isabela, pela cumplicidade, apoio constante e pela leveza com que sempre soube estar presente, assim como à sua família, que me acolheu com tanto carinho. Ao meu pai Magno, pelo apoio em todos esses anos.

Ao meu esposo Kêuron, por ser meu porto seguro em todas as fases desta jornada. Obrigada por acreditar em mim, por compartilhar os desafios, dividir responsabilidades e me lembrar diariamente do que realmente importa. Obrigada por acreditar em mim, mesmo quando nem eu mesma acreditava. Você nunca me deixou desistir. Ao nosso filho Bernardo, que chegou no meio do doutorado e me ensinou o verdadeiro sentido de persistir com doçura e coragem, você é minha maior conquista.

Aos membros do Laboratório da Qualidade Ambiental (LQA), agradeço por toda a ajuda. Um agradecimento especial à Tayane, que mesmo de longe, sempre esteve presente na minha caminhada.

Ao Gustavo, agradeço pelo auxílio nas etapas de processamento de dados e por todo conhecimento compartilhado.

À Universidade Federal de Viçosa (UFV), ao Departamento de Engenharia Agrícola (DEA) e ao Programa de Pós-Graduação em Engenharia Agrícola, pela formação acadêmica de excelência e pela infraestrutura que possibilitou o desenvolvimento desta pesquisa.

Ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), pelo fomento à pesquisa por meio da concessão da bolsa de doutorado.

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e ao IGAM, agradeço pelo apoio institucional e financiamento do projeto que subsidiou parte desta pesquisa.

Agradeço também aos membros da banca examinadora, pela leitura atenta e pelas valiosas contribuições para o aprimoramento deste trabalho.

Por fim, a todos os colegas, familiares, amigos e profissionais que, de forma direta ou indireta, colaboraram para a realização desta tese: meu

mais sincero obrigada.

Este trabalho foi realizado com o apoio das seguintes agências de pesquisa brasileiras: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

## RESUMO

COLLARES, Maria Fernanda Antunes, D.Sc., Universidade Federal de Viçosa, agosto de 2025. **Predição da vazão e da qualidade da água subterrânea em região semiárida.** Orientador: Andre Pereira Rosa.

As águas subterrâneas desempenham papel estratégico para a segurança hídrica global, especialmente em regiões semiáridas, onde a disponibilidade de recursos superficiais é limitada e irregular. No norte de Minas Gerais, essa dependência é intensificada pelo crescimento das demandas urbanas e agrícolas, exigindo estratégias integradas para avaliação da quantidade e qualidade desse recurso. A área de estudo abrange a região norte do estado de Minas Gerais, marcada por clima semiárido, solos heterogêneos, relevo variado e forte dependência das águas subterrâneas. Essas características tornam o território particularmente sensível a pressões antrópicas e ambientais, destacando a importância de investigações técnico-científicas voltadas à gestão sustentável dos aquíferos locais. A metodologia adotada combinou análise bibliométrica, modelagem espacial da vazão de poços tubulares e avaliação preditiva da qualidade da água subterrânea para fins de irrigação. A análise bibliométrica baseou-se em publicações indexadas na base Scopus entre 2000 e 2024, evidenciando as principais tendências, colaborações e abordagens científicas globais sobre qualidade das águas subterrâneas. Para a modelagem quantitativa, foram utilizados 3.700 registros de vazão georreferenciados obtidos no Sistema de Informações de Águas Subterrâneas (SIAGAS), integrados a camadas ambientais representando fatores condicionantes (clima, geologia, topografia, uso do solo). Na avaliação qualitativa, foram analisadas 672 observações obtidas pelo Instituto Mineiro de Gestão das Águas (IGAM) entre 2017 e 2023, distribuídas em períodos secos e chuvosos, com base em 95 pontos de coleta. A qualidade da água foi avaliada por meio de índices como Condutividade Elétrica (CE), Carbonato de sódio residual (RSC), Razão de adsorção de sódio (RAS), Porcentagem de sódio (%Na), Proporção Kelly (KR) e Índice de Qualidade da água subterrânea para fins de irrigação (IQAi). Também foram aplicadas análises multivariadas (Componentes Principais (PCA) e a Análise de Agrupamento Hierárquico (HCA)), interpolação espacial (IDW) e algoritmos de aprendizado de máquina (MLR, RNA, RF, KNN) para modelar os padrões de salinidade. Os resultados revelaram um crescimento expressivo na produção científica internacional a partir de 2020, com destaque para o uso crescente de abordagens integrativas que combinam geoprocessamento e inteligência artificial. A modelagem espacial da vazão demonstrou elevado

desempenho do algoritmo KNN, que obteve  $R^2 = 0,751$ ,  $RMSE = 0,150$ ,  $MAE = 0,094$  e  $LCCC = 0,862$ . As variáveis climáticas, sobretudo precipitação e radiação solar em meses chuvosos, mostraram-se determinantes para a recarga e produtividade dos poços. A análise espacial indicou maior potencial aquífero em domínios carbonáticos e psamíticos, enquanto litologias fraturadas máficas e félsicas apresentaram menor vazão. Por outro lado, a análise estatística não indicou correlação significativa entre os índices de qualidade da água subterrânea e os volumes de precipitação, sugerindo que os padrões geoquímicos observados são majoritariamente controlados por características locais do aquífero e não por variações sazonais de chuva. Em relação à qualidade da água, a maioria das amostras foi classificada como adequada para irrigação, embora valores extremos de Cl, Na e  $SO^2$  indiquem a necessidade de monitoramento contínuo. O modelo Random Forest destacou-se na predição de índices derivados de múltiplas interações iônicas, como %Na e KR.

De forma geral, os resultados reforçam o potencial das geotecnologias e do aprendizado de máquina para compreender, mapear e prever padrões de disponibilidade e qualidade da água subterrânea em regiões semiáridas. A abordagem integrada adotada nesta tese oferece subsídios técnicos relevantes para o planejamento hídrico, a definição de políticas públicas e a adoção de práticas sustentáveis no uso dos recursos subterrâneos.

Palavras-chave: água subterrânea; regiões semiáridas; aprendizado de máquina; modelagem espacial; salinidade; bibliometria.

## ABSTRACT

COLLARES, Maria Fernanda Antunes, D.Sc., Universidade Federal de Viçosa, August, 2025. **Prediction of Groundwater Flow and Quality in a Semi-Arid Region.** Adviser: Andre Pereira Rosa.

Groundwater plays a strategic role in global water security, especially in semi-arid regions where the availability of surface water resources is limited and irregular. In Northern Minas Gerais, this dependency is intensified by the growing urban and agricultural demands, requiring integrated strategies for assessing both the quantity and quality of this resource. The study area encompasses the northern region of the state of Minas Gerais, characterized by a semi-arid climate, heterogeneous soils, varied relief, and significant reliance on groundwater. These characteristics make the territory particularly sensitive to anthropogenic and environmental pressures, highlighting the importance of technical-scientific investigations focused on the sustainable management of local aquifers.

The methodology combined bibliometric analysis, spatial modeling of tubular well discharge, and predictive assessment of groundwater quality for irrigation purposes. The bibliometric analysis was based on publications indexed in the Scopus database from 2000 to 2024, highlighting global scientific trends, collaborations, and approaches related to groundwater quality. For the quantitative modeling, 3,700 georeferenced well discharge records were obtained from the Groundwater Information System (SIAGAS) and integrated with environmental layers representing conditioning factors (climate, geology, topography, land use). For the qualitative assessment, 672 observations were analyzed from the database of the Minas Gerais Institute of Water Management (IGAM) collected between 2017 and 2023, distributed across dry and wet seasons, based on 95 sampling points.

Water quality was evaluated using indices such as Electrical Conductivity (EC), Residual Sodium Carbonate (RSC), Sodium Adsorption Ratio (SAR), Sodium Percentage (%Na), Kelly's Ratio (KR), and the Groundwater Quality Index for Irrigation (GWQI). Additionally, multivariate analyses (Principal Component Analysis – PCA, and Hierarchical Cluster Analysis – HCA), spatial interpolation (IDW), and machine learning algorithms (MLR, ANN, RF) were applied to model salinity patterns. The results revealed a significant increase in international scientific production after 2020, with a notable rise in integrative approaches combining geoprocessing and artificial intelligence. Spatial modeling of groundwater discharge demonstrated high performance of the KNN algorithm, which achieved  $R^2 = 0.751$ ,  $RMSE = 0.150$ ,  $MAE = 0.094$ , and

LCCC = 0.862. Climatic variables, especially precipitation and solar radiation during rainy months, proved to be key drivers of recharge and well productivity. Spatial analysis indicated higher aquifer potential in carbonate and psammitic domains, while fractured mafic and felsic lithologies showed lower yields.

On the other hand, statistical analysis did not indicate a significant correlation between groundwater quality indices and precipitation volumes, suggesting that the observed geochemical patterns are primarily controlled by local aquifer characteristics rather than seasonal rainfall variability. Regarding water quality, most samples were classified as suitable for irrigation, although extreme values of Cl, Na, and  $\text{SO}_4^{2-}$  highlight the need for continuous monitoring. The Random Forest model showed the best predictive performance for indices derived from complex ionic interactions, such as %Na and KR.

Overall, the results reinforce the potential of geotechnologies and machine learning to understand, map, and predict patterns of groundwater availability and quality in semi-arid regions. The integrated approach adopted in this thesis provides relevant technical support for water planning, public policy formulation, and the implementation of sustainable groundwater management practices.

Keywords: groundwater; semi-arid regions; machine learning; spatial modeling; salinity; bibliometrics.

## SUMÁRIO

1. INTRODUÇÃO GERAL .....	11
2. HIPÓTESES DE INVESTIGAÇÃO .....	14
3. OBJETIVO GERAL .....	15
<b>3.1 Objetivos Específicos</b> .....	15
4. ESTRUTURAÇÃO DA TESE .....	16
CAPÍTULO 1 .....	18
<b>1.1 Introduction</b> .....	18
<b>1.2 Materials and Methods</b> .....	19
<b>1.2.1 Database and Techniques for Article Selection</b> .....	20
<b>1.3 Results and discussion</b> .....	21
<b>1.3.1 Overview of Groundwater Publications</b> .....	21
<b>1.3.2 Dynamics of Publications in Scientific Journals and Authorship</b> .....	23
<b>1.4 Conclusion</b> .....	36
<b>1.5 References</b> .....	37
CAPÍTULO 2 .....	43
<b>2.1 Introdução</b> .....	43
<b>2.2 Materiais e Métodos</b> .....	45
<b>2.2.1 Área de estudo</b> .....	47
<b>2.2.2 Base de Dados</b> .....	49
<b>2.2.3 Covariáveis Ambientais</b> .....	52
<b>2.2.4 Classificação Litológica Baseada na Composição Química e Mineralógica</b> .....	55
<b>2.2.4.1 Procedimentos de Reclassificação</b> .....	56
<b>2.2.5 Seleção e Processamento das Covariáveis</b> .....	57
<b>2.2.6 Modelagem da Vazão com Técnicas de Aprendizado de Máquina</b> .....	58
<b>2.2.7 Geração do Mapa de Potencial Aquífero</b> .....	60
<b>2.3 Resultados e Discussão</b> .....	60
<b>2.3.1 Monitoramento quantitativo da vazão dos poços</b> .....	60
<b>2.3.2 Desempenho preditivo dos Modelos de <i>Machine Learning</i></b> .....	62
<b>2.3.2.1 Comparação com estudos anteriores e contribuições do modelo</b> .....	68
<b>2.3.4 Importância relativa das covariáveis ambientais</b> .....	71
<b>2.3.5 Mapeamento do potencial aquífero regional</b> .....	73
<b>2.3.6 Mapa de incerteza das predições (coeficiente de variação)</b> .....	82
<b>2.4 Conclusão</b> .....	84

2.5 Referências .....	85
CAPÍTULO 3 .....	94
3.1 Introdução .....	94
3.2 Material e Métodos .....	96
3.2.1 Área de estudo.....	97
3.2.2 Obtenção e tratamento dos dados .....	97
3.2.3 Análise comparativa sazonal da pluviosidade .....	97
3.2.4 Determinação dos índices de qualidade de água subterrânea para fins de irrigação.....	98
3.2.5 Análise estatística multivariada .....	101
3.2.6 Análise espaço-temporal dos índices de qualidade da água para fins de irrigação .....	102
3.2.7 Previsão da qualidade da água de irrigação .....	102
3.3 Resultados e Discussão .....	104
3.3.1 Análise comparativa sazonal da pluviosidade .....	104
3.3.2 Características físico-químicas das águas subterrâneas .....	105
3.3.3 Características dos índices de adequação da água de irrigação para diferentes estações do ano .....	110
3.3.4 Avaliação da água de irrigação usando a abordagem multiparamétrica.....	112
3.3.5 Análise estatística multivariada .....	115
3.3.6 Previsão de índices de qualidade da água para fins de irrigação.....	122
3.3.7 Recomendações e perspectivas do estudo .....	128
3.4 Conclusões .....	129
3.5 Referências .....	130
5 CONSIDERAÇÕES GERAIS.....	135
6 SUGESTÃO PARA TRABALHO FUTUROS.....	138
7 REFERÊNCIAS GERAIS.....	139

## 1. INTRODUÇÃO GERAL

Nos últimos anos, o uso das águas subterrâneas tem ganhado destaque como alternativa estratégica frente à crescente escassez de recursos hídricos superficiais. O aumento da demanda por água, impulsionado pelo crescimento populacional, urbanização acelerada e expansão da agricultura irrigada, tem pressionado os sistemas aquíferos, especialmente em regiões onde a distribuição espacial e temporal da precipitação é irregular. Estima-se que as águas subterrâneas respondam por aproximadamente 26% da retirada global de água doce, suprimindo quase metade do consumo humano e cerca de 43% da irrigação global (ROBLES e MONJARDIN, 2025). Em regiões áridas e semiáridas, as águas subterrâneas muitas vezes representam a única fonte confiável de abastecimento.

No contexto do semiárido mineiro, essa dependência se intensifica. A irregularidade pluviométrica, aliada à elevada evapotranspiração e à escassez de rios perenes, limita severamente a disponibilidade de água superficial, tornando os aquíferos a principal fonte de abastecimento humano, uso doméstico e irrigação agrícola. A produtividade dessas reservas subterrâneas, no entanto, é altamente variável no espaço, em função de condicionantes como litologia, estrutura geológica, presença de fraturas e características de recarga (KOZAR *et al.*, 2025). A compreensão dessa variabilidade, expressa com frequência pela vazão dos poços, é fundamental para a gestão eficiente dos recursos hídricos subterrâneos.

Além da quantidade, a qualidade da água subterrânea é igualmente preocupante, sobretudo no que se refere à sua adequação para fins agrícolas. A salinidade representa um dos principais entraves à irrigação sustentável, podendo comprometer tanto o rendimento das culturas quanto a saúde dos solos. Parâmetros como sódio ( $\text{Na}^+$ ), cloreto ( $\text{Cl}^-$ ) e sulfato ( $\text{SO}_4^{2-}$ ), comumente associados à salinidade, devem ser rigorosamente monitorados, uma vez que seu acúmulo pode desencadear processos de salinização e sodificação dos solos.

Para avaliar essa aptidão, indicadores como a condutividade elétrica (CE), razão de adsorção de sódio (RAS), carbonato de sódio residual (CSR) e o índice de qualidade da água para irrigação (IQAi) são amplamente utilizados (AYERS e WESTCOT, 1994; SANTOS *et al.*, 2020). Esses índices refletem não apenas os aspectos geoquímicos dos aquíferos, mas também a influência de processos naturais, como intemperismo e evapoconcentração, e antrópicos, a exemplo do uso excessivo de fertilizantes e práticas inadequadas de irrigação (ALMEIDA *et al.*, 2025). Em ambientes semiáridos, onde a renovação da água subterrânea é limitada, esses fatores contribuem para o agravamento da salinidade, comprometendo a sustentabilidade dos sistemas agrícolas e ambientais.

De acordo com Sarkar *et al.* (2024), a junção de fatores ambientais como geologia, clima, relevo, uso e cobertura do solo pode auxiliar na identificação de zonas críticas de vulnerabilidade, áreas com maior potencial de recarga e regiões propensas à salinização. Essa abordagem integrada possibilita o mapeamento de áreas prioritárias para monitoramento e intervenção, além de fornecer subsídios para políticas públicas voltadas à gestão sustentável dos recursos hídricos subterrâneos. No entanto, a ausência de redes de monitoramento hidrogeológico contínuas e a carência de dados geoespaciais em escala regional dificultam a análise integrada da qualidade e disponibilidade hídrica subterrânea, especialmente em territórios ambientalmente frágeis, como o norte de Minas Gerais (RIBEIRO, 2024).

A variabilidade espacial da qualidade e da quantidade da água subterrânea resulta, portanto, da interação entre múltiplas variáveis ambientais que atuam de forma interdependente e complexa. Estudos apontam que a composição geológica, a profundidade dos poços, o tipo de aquífero (freático ou confinado) e o uso e cobertura do solo têm papel fundamental na definição desses padrões (NÉGREL *et al.*, 2021; GULGUNDI e SHETTY, 2018). Segundo Ghoto *et al.* (2025), a integração de dados sobre geologia, clima, relevo e uso do solo pode auxiliar na delimitação de zonas críticas de vulnerabilidade, áreas com maior potencial de recarga e regiões propensas à contaminação ou salinização das águas subterrâneas. Esses fatores, quando analisados de forma conjunta, possibilitam uma compreensão mais aprofundada dos processos hidrológicos que regulam tanto a disponibilidade quanto a qualidade da água em ambientes semiáridos, contribuindo para estratégias mais eficazes de manejo e preservação dos recursos hídricos subterrâneos.

Nesse cenário, o uso de técnicas avançadas de análise espacial e modelagem preditiva, especialmente aquelas baseadas em algoritmos de aprendizado de máquina (machine learning – ML), tem se mostrado uma alternativa promissora para superar as limitações dos métodos tradicionais. Esses algoritmos têm a capacidade de integrar grandes volumes de dados ambientais, hidroquímicos e climáticos, detectar padrões complexos e realizar previsões com elevado grau de acurácia (RAHMATI *et al.*, 2019; FARZIN *et al.*, 2021). Modelos como *Random Forest* (RF), *Support Vector Machines* (SVM) e *k-Nearest Neighbors* (KNN) têm sido amplamente utilizados na estimativa da salinidade, na previsão da vazão específica e na identificação de áreas com maior potencial de recarga (SANZ PÉREZ *et al.*, 2025).

Além disso, a combinação desses algoritmos com ferramentas de sistemas de informação geográfica (SIG) amplia sua aplicabilidade, permitindo análises espaciais detalhadas e suporte à gestão territorial. A integração de métodos estatísticos multivariados (como PCA e HCA), interpoladores espaciais (como IDW) e algoritmos de ML constitui uma

estratégia robusta para compreender os sistemas aquíferos em sua totalidade, identificar zonas prioritárias para monitoramento e subsidiar a tomada de decisão. Em regiões semiáridas, marcadas por fragilidade hídrica e pressão antrópica crescente, esses instrumentos ganham ainda mais relevância.

Os resultados oriundos dessas análises podem subsidiar políticas públicas e orientar intervenções técnicas voltadas à conservação dos recursos hídricos subterrâneos, especialmente em contextos de alta vulnerabilidade climática. O presente trabalho foi estruturado em três capítulos interdependentes. No primeiro capítulo é apresentada uma análise bibliométrica da produção científica internacional sobre qualidade da água subterrânea, destacando as principais abordagens, tendências e lacunas temáticas. No segundo capítulo é abordada a modelagem espacial da disponibilidade hídrica no semiárido mineiro, estimando a vazão de poços tubulares a partir de variáveis ambientais e hidrogeológicas. No terceiro capítulo, é realizada a avaliação da qualidade da água subterrânea para irrigação, com ênfase na sazonalidade e nos processos de salinização, utilizando análises estatísticas e geoespaciais. Além disso, neste capítulo aspectos qualitativos da água subterrânea são integrados por meio de modelos de predição baseados em machine learning, com o objetivo de estimar índices de salinidade e identificar os principais fatores ambientais associados aos padrões observados.

## 2. HIPÓTESES DE INVESTIGAÇÃO

### Hipótese 1

A produção científica global com a temática de qualidade das águas subterrâneas apresenta crescimento expressivo e evolução de temas emergentes e ferramentas metodológicas nas últimas décadas, apesar dos estudos se concentrarem em países restritos.

### Hipótese 2

A variabilidade espacial da disponibilidade hídrica subterrânea (vazão de poços) em regiões semiáridas mineiras é influenciada por fatores ambientais e hidrogeológicos, os quais podem explicar padrões espaciais observados.

### Hipótese 3

É possível estimar a vazão de águas subterrâneas no semiárido mineiro por meio de algoritmos de aprendizado de máquina, desde que sejam incorporadas variáveis ambientais e hidrogeológicas relevantes.

### Hipótese 4

A sazonalidade pluviométrica exerce influência sobre parâmetros de qualidade da água subterrânea para irrigação, especialmente os relacionados à salinidade.

### Hipótese 5

Modelos de aprendizado de máquina podem estimar com precisão índices de qualidade da água subterrânea relacionados à salinidade, permitindo identificar áreas críticas para uso agrícola.

### **3. OBJETIVO GERAL**

Realizar uma análise integrada dos aspectos qualitativos e quantitativos das águas subterrâneas, por meio de revisão bibliométrica e da avaliação de fatores ambientais e climáticos, visando compreender padrões, tendências e variabilidade em regiões semiáridas mineiras.

#### **3.1 Objetivos Específicos**

1. Investigar as tendências, lacunas e áreas emergentes na pesquisa global sobre qualidade das águas subterrâneas, bem como identificar os principais atores e padrões de colaboração científica.
2. Avaliar a influência de fatores ambientais e hidrogeológicos sobre a disponibilidade hídrica subterrânea em região semiárida.
3. Estimar a vazão de águas subterrâneas na região do semiárido mineiro a partir uso de algoritmos de aprendizado de máquina
4. Investigar a influência da sazonalidade pluviométrica sobre os parâmetros de qualidade da água subterrânea para fins de irrigação.
5. Estimar índices de qualidade de água subterrânea focados em salinidade para o uso da água para fins de irrigação.

#### 4. ESTRUTURAÇÃO DA TESE

O Quadro 1 apresenta a estruturação da tese por meio das hipóteses e objetivos estabelecidos.

Quadro 1 – Estruturação da tese

Hipótese	Objetivo Específico	Aspectos Avaliados	Resultados Esperados	Capítulo
<p><b>Hipótese 1</b> A produção científica global sobre qualidade das águas subterrâneas apresenta crescimento expressivo e evolução temática nas últimas décadas, revelando lacunas geográficas e metodológicas, bem como áreas emergentes de pesquisa.</p>	<p><b>Objetivo Específico 1</b> Investigar as tendências, lacunas e áreas emergentes na pesquisa global sobre qualidade das águas subterrâneas, bem como identificar os principais atores e padrões de colaboração científica.</p>	<ul style="list-style-type: none"> <li>- Revisão sistemática e análise bibliométrica de publicações (2000–2025).</li> <li>- Identificação de tendências, áreas emergentes e redes de colaboração.</li> </ul>	<ul style="list-style-type: none"> <li>- Evidência de crescimento contínuo da produção científica.</li> <li>- Identificação de lacunas e concentração de estudos em determinadas regiões e temas.</li> </ul>	<p><b>Capítulo 1</b></p>
<p><b>Hipótese 2</b> A variabilidade espacial da disponibilidade hídrica subterrânea em regiões semiáridas é significativamente influenciada por fatores ambientais e hidrogeológicos, os quais podem explicar padrões espaciais observados.</p>	<p><b>Objetivo Específico 2</b> Avaliar a influência de fatores ambientais e hidrogeológicos sobre a disponibilidade hídrica subterrânea em região semiárida.</p>	<ul style="list-style-type: none"> <li>- Integração de dados ambientais e hidrogeológicos.</li> <li>- Análise da relação entre as variáveis e a vazão.</li> <li>- Construção de mapas que indicarão a influência da geologia na vazão</li> </ul>	<ul style="list-style-type: none"> <li>- As variáveis relacionadas aos aspectos hidrogeológicos estão entre as mais importantes para predição de vazão</li> </ul>	<p><b>Capítulo 2</b></p>

<p><b>Hipótese 3</b> É possível estimar a vazão de águas subterrâneas no semiárido mineiro por meio de algoritmos de aprendizado de máquina, desde que sejam incorporadas variáveis ambientais e hidrogeológicas relevantes.</p>	<p><b>Objetivo Específico 3</b> Estimar a vazão de águas subterrâneas na região do semiárido mineiro a partir uso de algoritmos de aprendizado de máquina.</p>	<ul style="list-style-type: none"> <li>- Treinamento e validação de modelos de aprendizado de máquina (LM, RF e KKNN).</li> <li>- Avaliação do desempenho por métricas estatísticas.</li> </ul>	<ul style="list-style-type: none"> <li>- Desenvolvimento de um modelo para predição para estimativa da vazão de águas subterrâneas a partir do conhecimento de variáveis ambientais e hidrogeológicas.</li> <li>- Geração de mapas preditivos do potencial hídrico subterrâneo.</li> </ul>	<p><b>Capítulo 2</b></p>
<p><b>Hipótese 4</b> A sazonalidade pluviométrica exerce influência significativa sobre parâmetros de qualidade da água subterrânea para irrigação, especialmente os relacionados à salinidade.</p>	<p><b>Objetivo Específico 4</b> Investigar a influência da sazonalidade pluviométrica sobre os parâmetros de qualidade da água subterrânea para fins de irrigação.</p>	<ul style="list-style-type: none"> <li>- Comparação estatística da pluviosidade ao longo do tempo.</li> <li>- Avaliação os índices separados por período seco e chuvoso</li> </ul>	<ul style="list-style-type: none"> <li>- A pluviosidade exerce influência sobre a qualidade da água subterrânea para fins de irrigação;</li> <li>- Identificação de parâmetros significativamente afetados pela sazonalidade.</li> </ul>	<p><b>Capítulo 3</b></p>
<p><b>Hipótese 5</b> Modelos de aprendizado de máquina podem estimar com precisão índices de qualidade da água subterrânea relacionados à salinidade, permitindo identificar áreas críticas para uso agrícola.</p>	<p><b>Objetivo Específico 5</b> Estimar índices de qualidade de água subterrânea focados em salinidade para o uso da água para fins de irrigação.</p>	<ul style="list-style-type: none"> <li>- Desenvolvimento de modelos preditivos com múltiplos parâmetros físico-químicos.</li> <li>- Validação de modelos com dados observados.</li> </ul>	<ul style="list-style-type: none"> <li>- Algoritmo capaz de estimar os índices de qualidade.</li> <li>- Mapas de adequabilidade da água para irrigação.</li> </ul>	<p><b>Capítulo 3</b></p>

## CAPÍTULO 1

### **The state of the art on groundwater quality studies: a bibliometric analysis of the topic at a global level<sup>1</sup>**

<sup>1</sup>Artigo publicado na Revista Acque Sotteranee - Italian Journal of Groundwater.

#### **1.1 Introduction**

Water is an essential resource for sustaining life and socioeconomic development, found in various natural reservoirs such as rivers, lakes, aquifers, and glaciers. However, surface water quality can vary significantly depending on its source and exposure to environmental and anthropogenic factors, due to its direct exposure to human activities, including effluent discharge, the use of agrochemicals, and accidental chemical spills (VLIET *et al.*, 2021). In contrast, groundwater typically exhibits greater chemical and biological stability over time due to its limited exposure to external contaminants. Despite this relative protection, aquifers can still be impacted by anthropogenic activities, particularly in areas with intensive fertilizer use, improper disposal of industrial waste, and unregulated land occupation. Studies indicate that groundwater contamination, although slower, can have persistent and difficult-to-remediate effects, impacting large areas and compromising water quality for extended periods (CHEGBELEH *et al.*, 2020).

Given the growing dependence on groundwater as a water supply source, both in rural and urban areas, it is essential to understand contamination patterns and the factors influencing its quality. In addition to its importance as a potable water source, aquifers play a crucial role in maintaining ecological flows and recharging surface water systems (BOSE *et al.*, 2023; OIRO *et al.*, 2020). Groundwater exploitation depends on locational aspects and geological formations, requiring geophysical investigations, which require the use of equipment and complex studies integrating various factors (geological, geomorphological, pedological, topographical, among others). Despite the practical limitations in determining groundwater quality, efforts must be made to promote monitoring of this resource, particularly to prevent contamination from anthropogenic sources (OIRO *et al.*, 2020).

The growing concern about groundwater quality has driven the use of innovative analytical approaches. Recent studies have applied techniques such as multivariate statistical analysis, including Principal Component Analysis (PCA) (RIBOLI e LINDINO, 2023),

Hierarchical Cluster Analysis (HCA) (CHOTPANTARAT *et al.*, 2020; WISITTHAMMASRI *et al.*, 2020), and spatiotemporal analyses based on Geographic Information Systems (GIS) and remote sensing (SINHA *et al.*, 2021). These approaches enable the identification of contamination patterns and the assessment of groundwater contamination risk, by integrating the intrinsic vulnerability of aquifers with the spatial distribution and intensity of anthropogenic pressure.

Despite advancements in groundwater quality research, there is a limited number of systematic review and bibliometric articles in the literature compared to studies on surface waters (ADENIYI e GIWA, 2021; BARRETT *et al.*, 1999; DEVIc *et al.*, 2014; FOSTER *et al.*, 2013; GRAHAM e POLIZZOTTO, 2013; KATZ *et al.*, 2011; LOFTIS, 1996). Furthermore, recent studies, such as those by Kannazarova *et al.* (2024), indicate an increase in research on the use of groundwater for irrigation. However, these analyses are still limited to regional contexts, such as the countries of the Commonwealth of Independent States (CIS). This gap highlights the need for studies that systematically compile and organize information on groundwater quality at a global scale, considering both the evolution of research and the methodologies employed (LI *et al.*, 2019; ZHANG *et al.*, 2017; ZHOU *et al.*, 2020; ZYOUD e FUCHS-HANUSCH, 2017).

In this context, bibliometric analysis emerges as a promising tool to identify cognitive structures and intellectual relationships in the field of groundwater quality. This approach allows for the examination of publication patterns, collaboration between researchers and institutions, and emerging themes in the literature (DONTHU *et al.*, 2021). Thus, this study aimed to conduct a systematic review of the literature on groundwater quality at a global level, with specific objectives: (i) to map and examine the most prevalent and emerging themes in groundwater quality publications; (ii) to detect knowledge gaps and suggest directions for future research that can contribute to a more comprehensive understanding of the topic; (iii) to identify the leading authors, institutions, and countries involved in groundwater quality research, analyzing co-authorship networks and patterns of international collaboration.

By integrating these elements, this study provides a comprehensive overview of the evolution of groundwater quality research, contributing to the development of sustainable management and protection strategies for this essential resource.

## 1.2 Materials and Methods

### 1.2.1 Database and Techniques for Article Selection

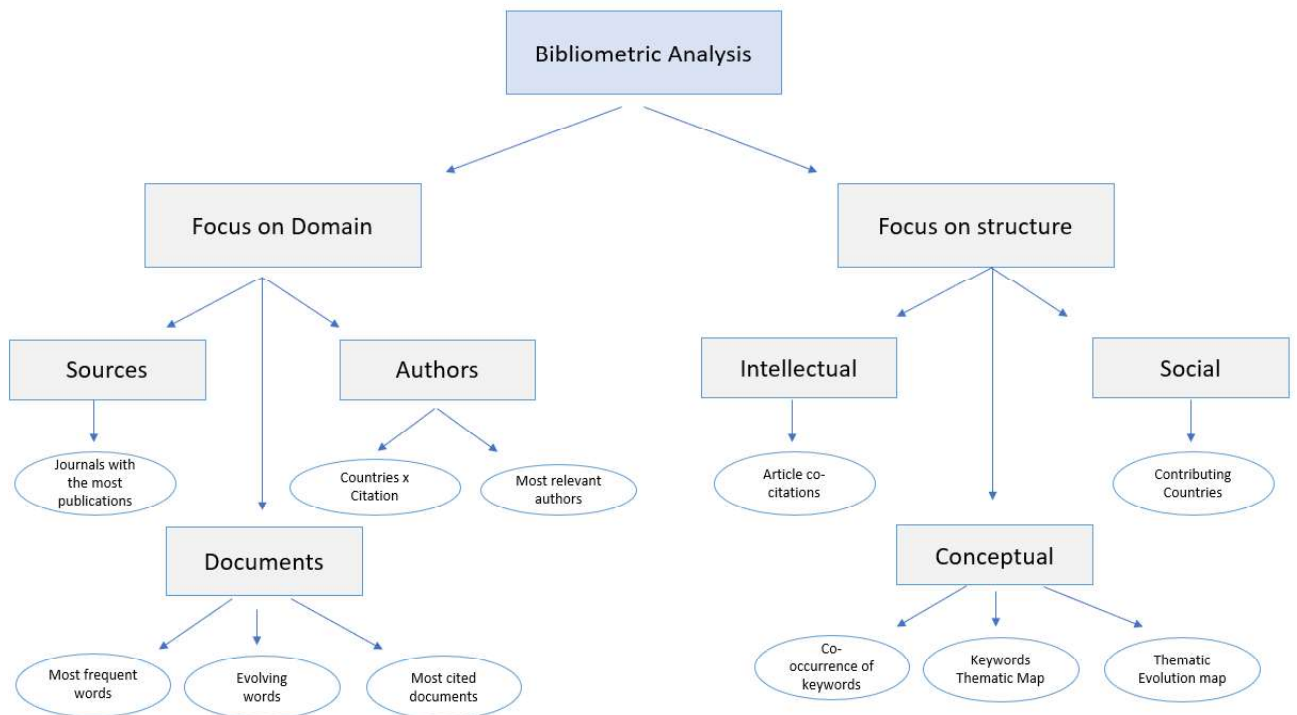
The search for scientific articles for the bibliometric analysis on groundwater quality was conducted in the Scopus database (<https://www-scopus.ez35.periodicos.capes.gov.br/home.uri>). Scopus is widely recognized as the largest database of abstracts and citations of peer-reviewed literature, covering a broad range of publications, such as scientific journals, books, and conference proceedings (BAAS *et al.*, 2020; KANNAZAROVA *et al.*, 2024).

The search was conducted using the following keywords and term combinations: "groundwater" AND "machine learning"; "groundwater quality" AND "machine learning"; "groundwater quality" AND "prediction"; "groundwater quality" AND "classification"; "groundwater quality index" AND "machine learning"; "underground water quality" AND "remote sensing"; "groundwater quality" AND "landsat"; "groundwater quality" AND "sentinel". The search query was constructed as follows: (("groundwater" AND "machine learning") OR ("groundwater quality" AND "machine learning") OR ("groundwater quality" AND "prediction") OR ("groundwater quality" AND "classification") OR ("groundwater quality index" AND "machine learning") OR ("underground water quality" AND "remote sensing") OR ("groundwater quality" AND "landsat") OR ("groundwater quality" AND "sentinel")). The search considered publications from the years 2000 to 2024. Initially, the search resulted in 2,524 potentially relevant publications, covering various studies on groundwater quality and its estimates under different conditions, including the use of machine learning techniques and remote sensing. To refine the results, additional filters were applied, considering only articles published in English, between 2000 and 2024, and classified as "articles." After this filtering, 1,686 articles were selected for a more detailed analysis.

The bibliometric analysis was conducted based on the methodology adapted from Bose *et al.* (2023), which structures the evaluation of the articles into two main axes: (i) domain and (ii) structure. In the domain axis, the articles were examined for: (i) source types; (ii) document aspects; and (iii) origins and relevance of the authors. In the structure axis, the analysis covered: (i) intellectual aspects; (ii) conceptual aspects; and (iii) social aspects, with an emphasis on the countries of origin of the research on groundwater quality (Figure 1.1).

The analyses were conducted using RStudio software and the Bibliometrix package, which allowed for the extraction of information on sources, authors, and documents, providing insights into the domain and structure of the articles. For the keyword co-occurrence analysis

(conceptual aspect), VOSviewer software was used, enabling the visualization of relationships between the main terms found in the analyzed literature.



**Figure 1.1.** Diagram reporting the aspects considered in the bibliometric analysis.

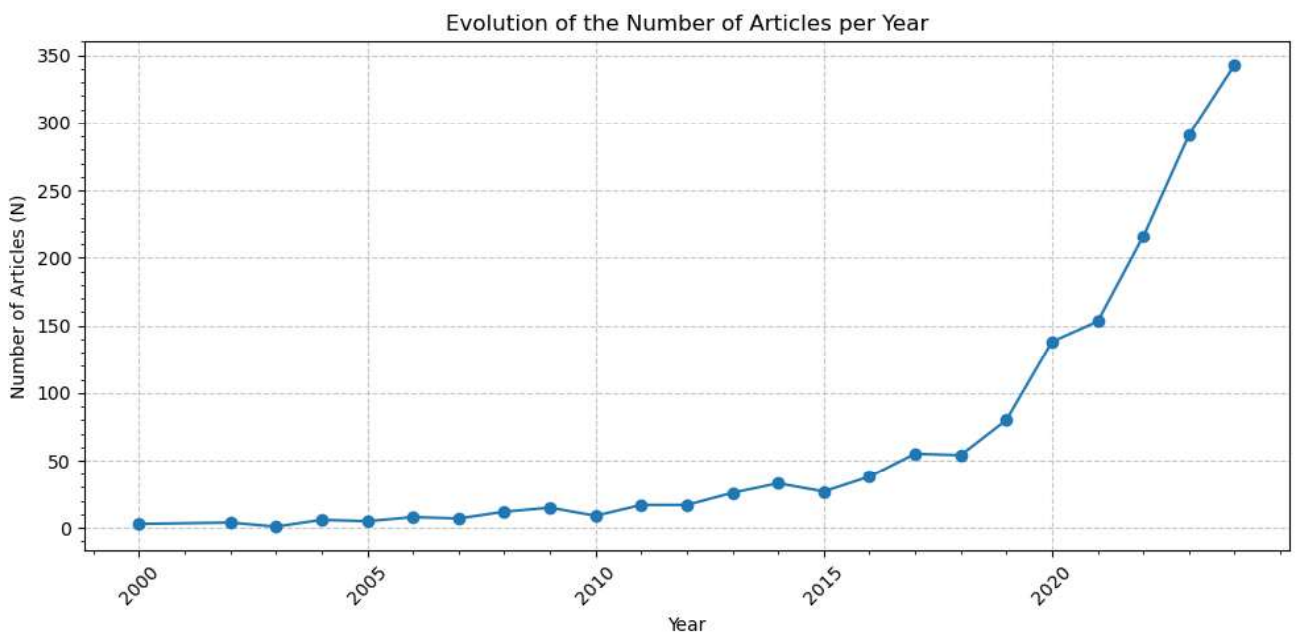
## 1.3 Results and discussion

### 1.3.1 Overview of Groundwater Publications

The bibliometric analysis revealed that, over the past 24 years, a total of 1,686 articles on groundwater have been published. Of these, 26 were written by individual authors, while the others resulted from collaborations among multiple researchers. On average, the published articles had more than four authors and received at least 22 citations each.

It is observed that the scientific interest in groundwater quality began to intensify in the late 20th century, coinciding with the increase in the global population and irrigation. This growth was largely driven by concerns about the deterioration of groundwater quality, as recently pointed out by Singh *et al.* (2024). However, since the 2000s, research on groundwater has increasingly integrated with other areas of knowledge, such as climate sciences (RAHAMAN *et al.*, 2024), agricultural sciences (SAKTHIPRIYA *et al.*, 2024), environmental chemistry (ABBA *et al.*, 2024), computer sciences (REN *et al.*, 2024), and geosciences/remote sensing (INAM ULLAH *et al.*, 2024), expanding the perspectives and research approaches.

Figure 1.2 illustrates the number of articles published on groundwater quality over the analyzed period. A clear trend of continuous growth in scientific production is evident, with a highlight in 2020, which saw the highest annual growth rate (60.92%), totaling 140 published articles. This accelerated growth can be interpreted as a response to the increasing demand for solutions to the challenges of groundwater management and preservation, especially in the face of climate change and the expansion of agricultural activities. It is worth noting that, until 2018, the number of publications on the topic did not exceed 60 articles per year, reflecting a substantial increase in interest and relevance of the topic in recent decades.

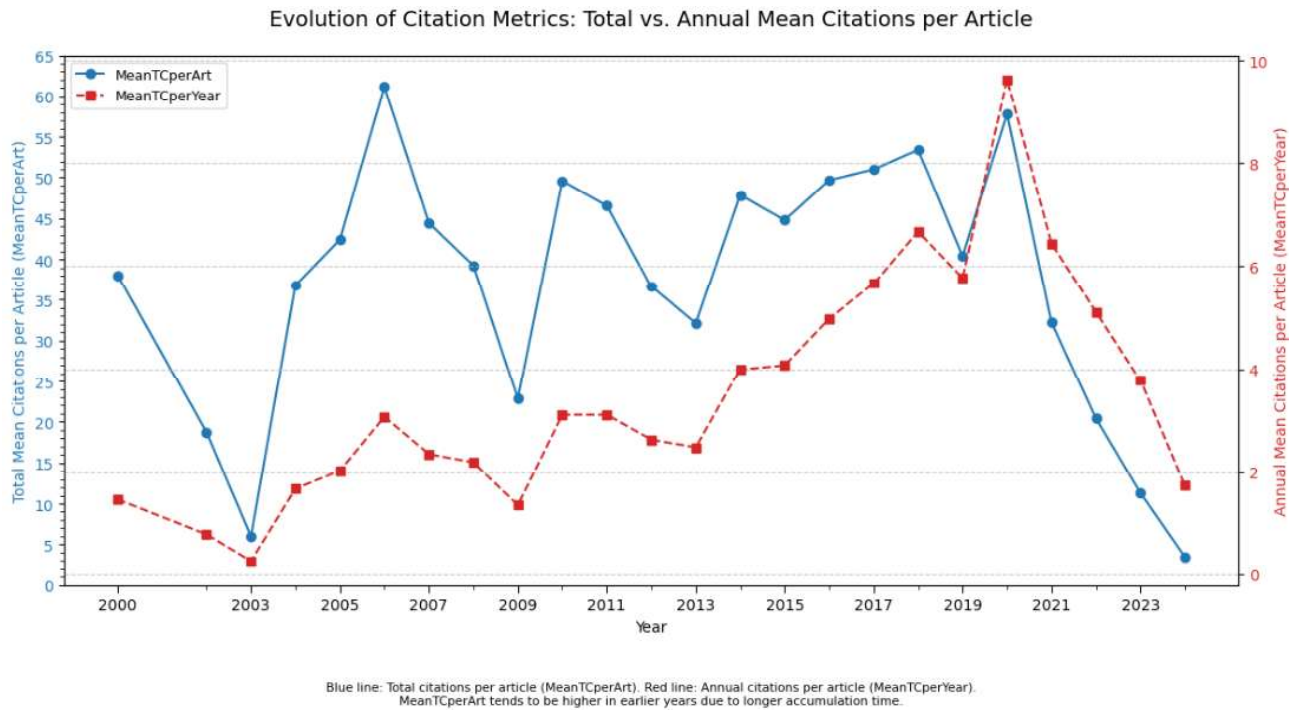


**Figure 1.2.** Number of articles published related to groundwater quality during the study period (2000–2025).

In addition to the increase in the number of publications over the years, an important aspect to evaluate the impact of research on groundwater quality is the average number of citations per published article (MeanTCperArt), as shown in Figure 1.3. It is observed that older articles generally have a higher average number of citations, such as those from 2006, which recorded an average of 61.12 citations per article. This behavior was expected, as articles published earlier have more opportunities to be cited. However, when analyzing the average citations per year (MeanTCperYear), it is observed that the impact of more recent scientific production has been significant, with a peak in 2020 (9.63 citations per year), followed by a downward trend in the subsequent years.

This decline can be explained by the maturation time required for newly published articles to consolidate in the literature and become more widely cited. Starting from 2021, for

example, the average citations per article gradually decreased, reaching 1.73 in 2024, reflecting the short time window available for citation accumulation. Therefore, the interpretation of these indicators should consider both the total number of citations accumulated over the years and the annual growth rate of citations.



**Figure 1.3.** Evolution of the average number of citations per published article and per year in publications related to groundwater quality.

Another relevant aspect is the relationship between interdisciplinarity and scientific impact. The increasing integration of groundwater quality research with areas such as climate change, water resource management, and artificial intelligence has broadened the scope of these studies and potentially accelerated their recognition within the scientific community. This phenomenon may explain the peak in citations per year observed in 2020, possibly driven by the growing global interest in water security and the application of new technologies in the field.

### 1.3.2 Dynamics of Publications in Scientific Journals and Authorship

The analysis of the publication sources of groundwater quality studies allows for the identification of the main scientific outlets disseminating research on this topic. Identifying the most relevant journals not only indicates where the scientific community has focused its efforts but also reveals editorial trends and areas of greatest impact within the field.

Table 1.1 presents a list of the scientific journals that have published most of the articles on the topic of groundwater quality. The results show that *Science of the Total Environment* stands out as the leading journal in terms of the number of publications, with a total of 80 articles, followed by *Water (Switzerland)* with 70, and *Journal of Hydrology* with 63 publications. In general, these journals have a broad scope covering environmental studies, water management, and applied hydrology, which explains their relevance for the dissemination of research on groundwater quality.

**Table 1.1.** Journals with most of the publications on groundwater quality between 2000 and 2024.

<b>Position</b>	<b>Journal (ISSN)</b>	<b>Number of Published Articles</b>
1	Science of the Total Environment (0048-9697)	80
2	Water (Switzerland) (2073-4441)	70
3	Journal of Hydrology (0022-1694)	63
4	Environmental Earth Sciences (1866-6299)	53
5	Environmental Science and Pollution Research (0944-1344)	46
6	Environmental Monitoring and Assessment (1573-2959)	43
7	Groundwater for Sustainable Development (2352-801X)	39
8	Remote Sensing (2072-4292)	26
9	Water Resources Management (0920-4741)	26
10	Arabian Journal of Geosciences (1866-7511)	24

It is important to note that the high number of publications in these journals does not necessarily imply a specific editorial commitment to the topic, but rather a convergence between the demands of the scientific community and the editorial focus of these publications. The recurrent presence of these journals reinforces that groundwater quality studies have found a priority space in journals dedicated to environmental science and hydrology. Furthermore, the analysis of the distribution of publications across journals may indicate the interdisciplinary

nature of the topic, as groundwater quality studies are not confined to strictly hydrological journals but are also published in journals focused on geosciences, environmental sciences, and engineering. This diversity highlights the need for integrated approaches to understanding the challenges and solutions related to the sustainable management of groundwater resources.

Table 1.2 presents the group of authors who have made the most significant contributions to groundwater research, based on data obtained from the bibliometric analysis. The ranking was initially determined by the h-index, followed by Total Citations (TC).

Among the most influential researchers, Lee S. (South Korea) stands out for having the highest h-index (10), a metric that quantifies a researcher's productivity and impact, based on their most cited articles. Since 2019, his research has explored various innovative approaches in groundwater management and quality, with a particular focus on the use of artificial intelligence (AI) and machine learning to predict nitrate contamination susceptibility and model pollutant adsorption processes such as fluoride. His research also includes studies on the impacts of road transportation systems on groundwater quality, employing Explainable Artificial Intelligence (XAI), as well as modeling the suitability of groundwater for human consumption through neural networks and geospatial analysis (LEE *et al.*, 2022). The author's research has been crucial in developing strategies and policies for the sustainable management of groundwater resources through the development and application of validated models to prevent environmental issues (MALLICK *et al.*, 2022), which were further enhanced by Jayasingle *et al.* (2021) to map suitable locations for groundwater extraction.

The author with the most cited works is Wu J. (China), who is the principal author of 9 articles published during the period, resulting in 967 citations, and has authored 16 articles. Among the highlighted researches, the author is a pioneer in publishing articles on groundwater, with his first publication in 2008, establishing himself as one of the most influential researchers in the field. Wu's contributions range from analyzing groundwater quality in urban and rural areas to modeling changes in water storage in large river basins, such as the Ordos Basin in China (WU *et al.*, 2024). Furthermore, Wu J. has been prominent in applying advanced techniques for modeling and predicting groundwater quality using machine learning (WU *et al.*, 2022) and investigating the relationship between water quality and land use patterns (WU *et al.*, 2020).

**Table 1.2.** Identification of the most relevant researchers in publications related to groundwater quality.

<b>Autores</b>	<b>h_index</b>	<b>g_index</b>	<b>m_index</b>	<b>TC</b>	<b>NP</b>	<b>PY_start</b>
LEE S	10	13	1.429	700	13	2019
WU J	9	16	0.500	967	16	2008
PHAM BT	9	10	1.125	629	10	2018
MOSAVI A	9	9	1.500	574	9	2020
ISLAM ARMT	9	12	1.000	413	12	2017
LU W	9	13	1.000	179	13	2017
LI P	8	12	0.667	773	12	2014
PRADHAN B	8	9	1.000	653	9	2018
PRAKASH I	8	8	1.000	600	8	2018
PAL SC	8	12	1.333	290	12	2020

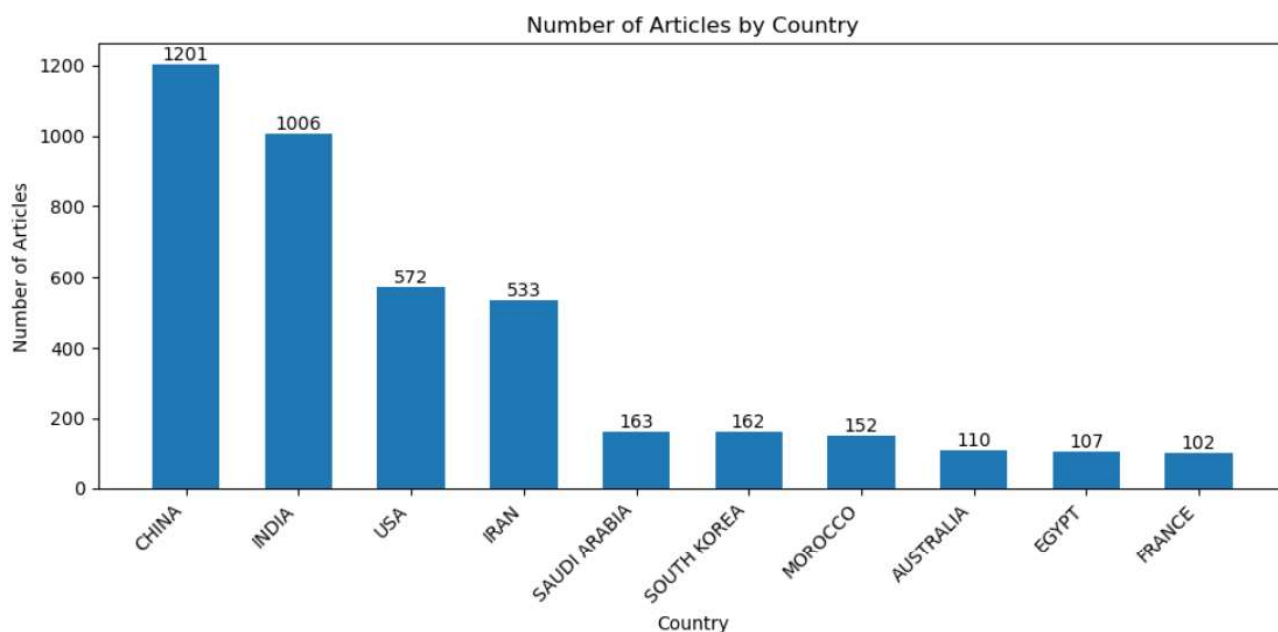
Legend: **h\_index**: Hirsch index, which measures the productivity and impact of an author's publications based on the number of citations received; **g\_index**: Egghe index, which gives more weight to highly cited publications, reflecting the global impact of the researcher; **m\_index**: Index that adjusts the h-index for the researcher's career length, allowing for fairer comparisons between researchers with different career durations; **TC (Total Citations)**: Total number of citations received by the author's publications; **NP (Number of Publications)**: Total number of the author's publications; **PY\_start (Publication Year Start)**: Year of the first publication registered in the analysis.

Another prominent researcher is Islam ARMT, who has published 12 articles since 2017 and received 413 citations. His research focuses on analyzing the variation in groundwater quality in response to climate change and land use, integrating statistical modeling and machine learning to predict the presence of pollutants. Lu W., also with 13 publications and an h-index of 9, has made significant contributions to the hydrochemical characterization of groundwater and the identification of contamination sources, employing geostatistical models and optimization algorithms for the sustainable management of aquifers. Among the researchers who have advanced the application of artificial intelligence in hydrogeology, Mosavi A. and Pham BT deserve special mention. Mosavi A., with 9 publications since 2020 and an h-index of 9, has led studies on the prediction of groundwater levels and aquifer recharge analysis, considering the impacts of climate change. Pham BT, with an h-index of 9 and 10 published articles since 2018, has specialized in the use of Explainable Artificial Intelligence (XAI) to enhance the interpretation of groundwater quality prediction models, contributing to more transparent and effective decision-making.

Other researchers, such as Li P. (h-index 8, 773 citations), Pal SC. (h-index 8, 290 citations), and Pradhan B. (h-index 8, 653 citations), also rank among the most influential, with significant contributions to groundwater quality modeling and the development of innovative methodologies for the management of groundwater resources. The increasing relevance of these

researchers reflects the evolution of the field of hydrogeology and groundwater quality, with advances driven by the use of artificial intelligence, machine learning, and geostatistical modeling. Interdisciplinary collaboration and the development of new approaches have been key to enhancing monitoring, prediction, and mitigation of environmental risks, ensuring the sustainability of groundwater resources for future generations.

Regarding scientific production on groundwater, there is a predominance of a limited number of countries (Figure 1.4), which may be related to the high demand for water resources in these nations, as well as the availability of funding for research on this topic. According to the data on scientific production by country, China (1,201 articles) and India (1,006 articles) lead the number of publications from 2000 to 2024, followed by the United States (572), Iran (533), and Saudi Arabia (163). Along with the scientific prominence of China and India in publications, it is important to note that these are the most populous countries in the world, and consequently, there is a high demand for water resources, particularly groundwater sources. As highlighted by the National Ground Water Association of China, the intense exploitation of these resources presents challenges that drive scientific research in search of solutions for the sustainable management of groundwater.



**Figure 1.4.** Scientific production by country in publications related to groundwater quality

In addition to the number of publications, it is important to analyze the distribution of corresponding authors, which indicates scientific leadership and international collaboration. In

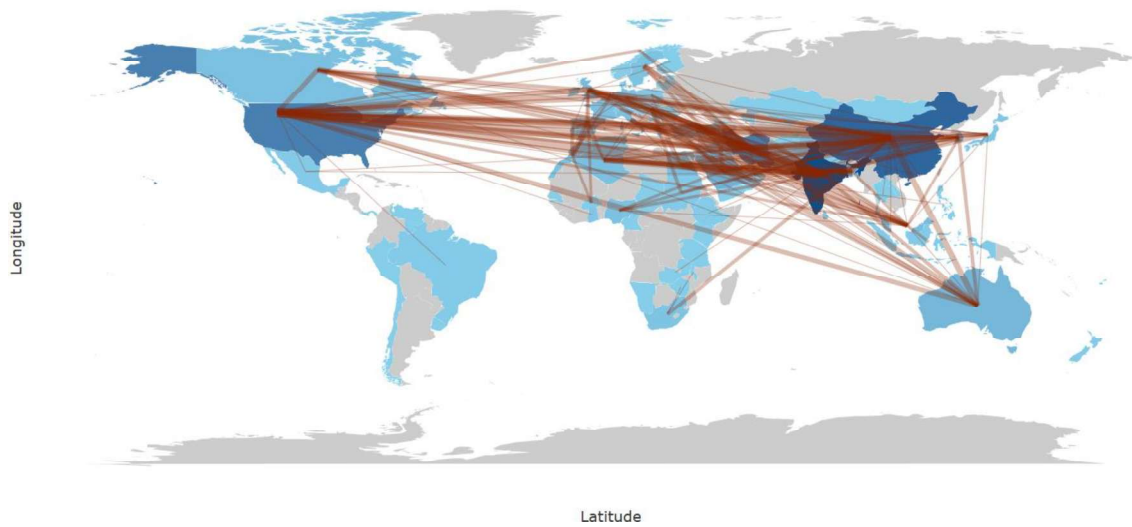
this regard, China (356 articles), India (277), and the United States (230) also occupy the top positions (Table 1.3). However, while China and India have a higher number of publications with national collaboration, Iran stands out for its high rate of international collaboration (MCP\_Ratio = 0.42), suggesting that its research often involves cooperation with foreign institutions. This metric represents the proportion of a country's publications involving international collaboration, calculated by dividing the number of multi-country publications (MCP) by the total number of the country's publications (PETERMANN *et al.*, 2024). In the case of Iran, 42% of publications were produced in international collaboration, while China had a rate of 25.8%, indicating a greater focus on national scientific output. According to FAO data (AQUASTAT), Iran is among the countries with the highest consumption of groundwater, reinforcing the need for effective management strategies, thereby promoting international collaborations for the development of scientific based solutions.

**Table 1.3.** Distribution of scientific publications and international collaboration by country in publications related to groundwater quality.

Country	Articles	SCP	MCP	Freq	MCP Ra
China	356	264	92	0.174	0.258
India	277	224	53	0.136	0.191
Usa	230	185	45	0.113	0.196
Iran	181	105	76	0.089	0.420
Australia	50	26	24	0.024	0.480
Korea	45	31	14	0.022	0.311
Germany	40	28	12	0.020	0.300
Canada	34	19	15	0.017	0.441
Italy	33	26	7	0.016	0.212

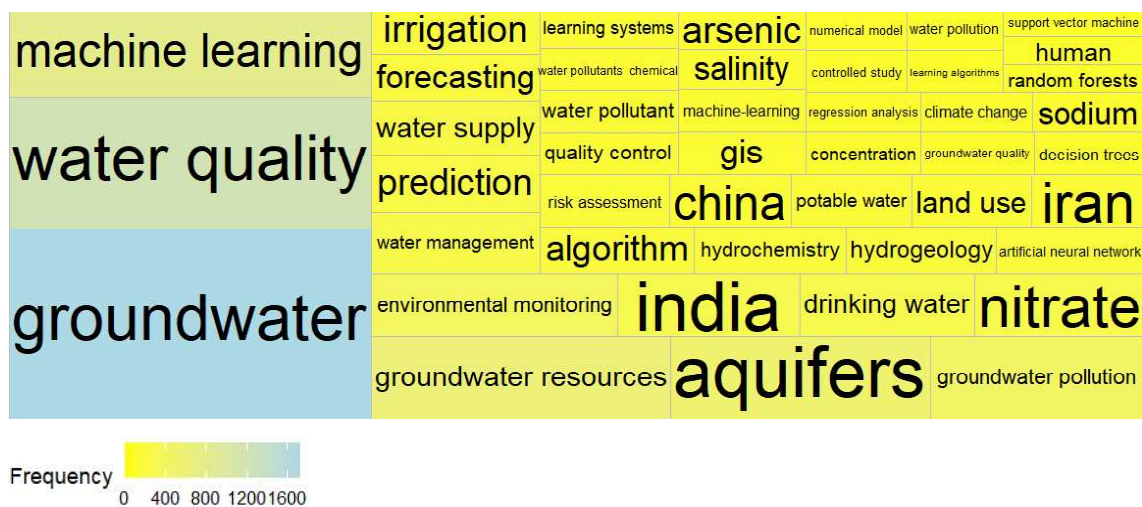
Legend: **SCP:** single country publication; **MCP:** multiple countries publication; **Freq:** represents the relative frequency of scientific publications from each country in relation to the total number of articles in the sample analyzed.

The bibliometric analysis also reveals a social structure within the research field, highlighting the contribution and collaboration among researchers and their institutions (Figure 1.5). Among the collaborations established through co-authorships, notable joint publications include those between Iranian and American researchers (23 publications), Chinese and American researchers (22 publications), Indian and Iranian researchers (20 publications), and researchers from India, Bangladesh, and the United States, with 18 publications. These data highlight how collaboration among researchers influences the quality of research and contributes to the global advancement of scientific knowledge.



**Figure 1.5.** World map illustrating the collaboration of authors and institutions between countries in publications related to groundwater quality.

Based on the survey of articles (1,342), the number of keywords counted in publications on groundwater quality exceeded 2,000. Figure 1.6 highlights the 30 more frequently used keywords; the most common term was "groundwater" (1,344 occurrences), followed by "water quality" (1,188 occurrences), "machine learning" (805 occurrences), and "groundwater pollution" (434 occurrences). Notably, there is a significant relationship between keywords referencing methodological tools associated with statistical analysis aimed at understanding the impact of chemical and geological parameters on groundwater quality. Within this context, the terms "prediction" (237 occurrences) and "algorithm" (183 occurrences) stand out.

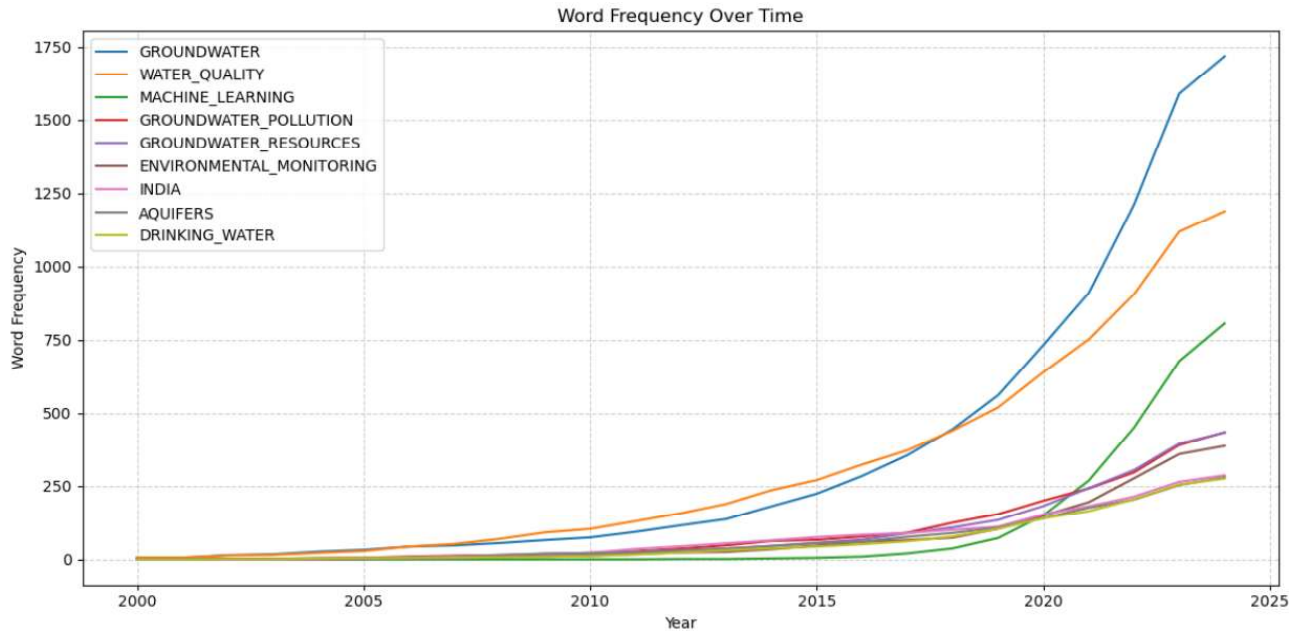


**Figure 1.6.** Most frequency keywords used in articles related to groundwater quality.

The selection of keywords reflects the focus of the research on aspects related to contamination, the presence of nitrates and salinity, land use changes, as well as the identification of hazardous substances. The frequency of the keywords also reveals the emphasis on statistical methods and other computational tools used in the research, such as machine learning (805), multivariate statistics (130), Geographic Information Systems (GIS) (225), and geostatistics (kriging) (64). In recent years, hydrogeological and hydrochemical aspects have been studied in conjunction with modern computational techniques, indicating a research trend within this field (MOHAMMED *et al.*, 2024), as well as the use of advanced machine learning techniques and data mining (VAROUCHAKIS *et al.*, 2023; TIAN *et al.*, 2024).

Complementarily, Figure 1.7 presents the evolution of the accumulated occurrence of the ten most frequently used keywords in scientific publications for the period from 2000 to 2024, based on the bibliometric analysis. From 2000 to 2012, the number of scientific articles was limited, with only 108 articles published (Figure 1.2), as well as the frequency of the keywords. However, after 2014, there was a significant increase in the number of publications, accompanied by a rapid shift in the use of specific keywords. In particular, there was a notable growth in the keywords "machine learning", "groundwater," and "water quality," which highlights the increasing use of predictive tools for simulating groundwater quality.

This trend aligns with the advancement of machine learning (ML) as an approach widely used in science, driven by increased computational capacity and the availability of large volumes of data starting in the 21st century (FU *et al.*, 2022). Although traditional statistical techniques had already been applied in groundwater quality modeling, the use of ML began to gain greater relevance once the scientific community recognized its potential to handle complex and nonlinear patterns in environmental data (HAO *et al.*, 2024, LI *et al.*, 2023, ZHAN *et al.*, 2025). Thus, the adoption of machine learning in the field of hydrogeology was not immediate, but rather occurred as the technology became more established in other fields and more accessible. The progressive use of ML, particularly in the prediction of groundwater quality, supports decision-making for more sustainable resource management.



**Figure 1.7.** Evolution in the frequency of use of key keywords in articles on groundwater quality between 2000–2024.

Table 1.4 presents the 10 most cited articles on groundwater quality for the analysis period. Among the most notable studies related to groundwater quality is the research of Podgorski e Berg (2020), who assess the impacts of arsenic contamination.

**Table 1.4.** List of the 10 most cited articles on the topic of groundwater for the evaluated period (2000–2024).

Position	Title	Number of citations	Citation
1	Global threat of arsenic in groundwater	694	Podgorski e Berg (2020)

2	GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran	490	Naghibi <i>et al.</i> (2017)
3	Predictive modeling of groundwater nitrate pollution using Random Forest and multisource variables related to intrinsic and specific vulnerability: A case study in an agricultural setting (Southern Spain)	316	Rodriguez-Galiano <i>et al.</i> (2014)
4	Application of Support Vector Machine, Random Forest, and Genetic Algorithm Optimized Random Forest Models in Groundwater Potential Mapping	313	Naghibi <i>et al.</i> (2017)
5	Seasonal variation of groundwater quality in a part of Guntur District, Andhra Pradesh, India	250	Subba (2006)
6	A novel machine learning-based approach for the risk assessment of nitrate groundwater contamination	248	Sajedi-Hosseini <i>et al.</i> (2018)
7	Machine learning algorithms for modeling groundwater level changes in agricultural regions of the United States	243	Sahoo <i>et al.</i> (2017)
8	GIS-based groundwater potential analysis using novel ensemble weights-of-evidence with logistic regression and functional tree models	235	Chen <i>et al.</i> (2018)
9	Origin and assessment of groundwater pollution and associated health risk: A case study in an industrial park, northwest China	223	Li <i>et al.</i> (2014)
10	Hydrogeochemistry for the assessment of groundwater quality in Varanasi: a fast-urbanizing center in Uttar Pradesh, India	217	Janardhana <i>et al.</i> (2011)

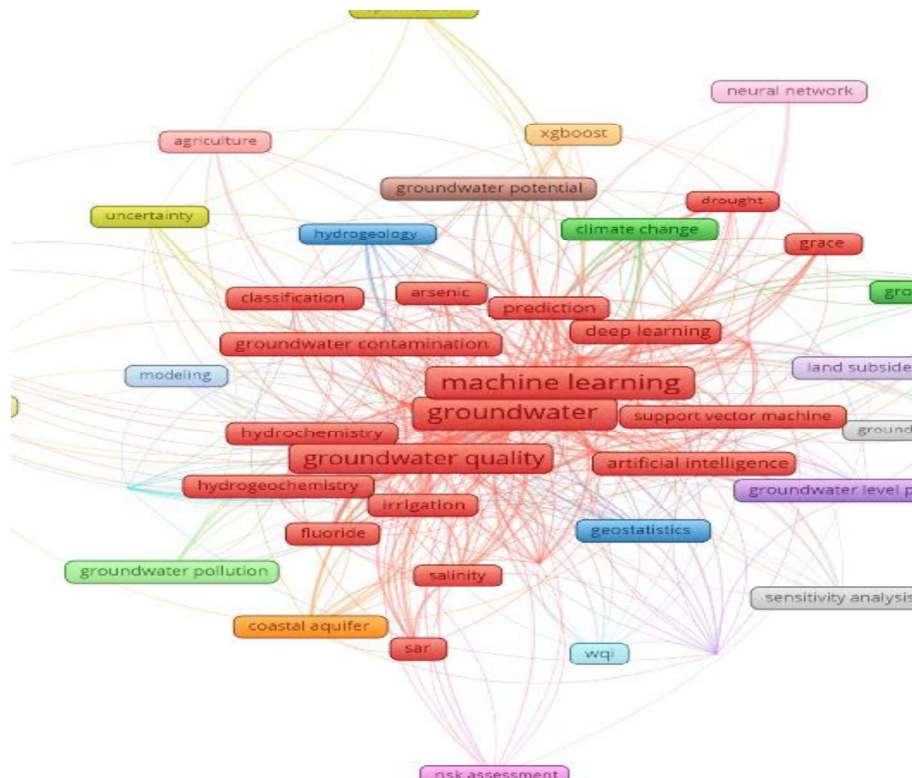
They employ advanced modeling techniques to map the potential risk in various regions of the world. This study provided an understanding of the risks associated with arsenic presence, particularly in areas facing water security challenges. Additionally, Naghibi *et al.* (2016) explored the mapping of groundwater potential in the Chaharmahal-e-Bakhtiari province of Iran, using machine learning models to produce accurate spring maps, which provided valuable insights for the sustainable management of these resources. Another significant study was conducted by Rodriguez-Galiano *et al.* (2014), who investigated the use of the Random Forest method in predictive modeling of nitrate pollution in groundwater. The authors demonstrated the effectiveness of this technique in identifying pollution sources and creating vulnerability maps, highlighting the potential of these approaches in water resource management.

In addition, Naghibi *et al.* (2017) applied machine learning methods, such as Support Vector Machine (SVM), Random Forest (RF), and Genetic Algorithm-optimized Random Forest (RFGA), to map groundwater sources near springs. The study, conducted at 842 locations, indicated that RF and RFGA were more effective than SVM. The most influential variables in this process were altitude, the Topographic Wetness Index (TWI), and slope angle. This methodology proved to be promising for application in other regions with water scarcity. Subba (2006) also made a significant contribution by studying the seasonal variation in groundwater quality in an area of the Guntur district, Andhra Pradesh, India. The author analyzed water samples collected from 40 wells before and after the rainy season over three years, observing that most samples did not meet quality standards for either human consumption or irrigation, particularly in the post-rainy period, which was attributed to factors such as the semi-arid climate, water-rock interaction, and human activities. Finally, Sajedi-Hosseini *et al.* (2018) developed a new method using machine learning to assess the risk of nitrate contamination in groundwater in an arid region of Iran. Samples from 102 water wells were analyzed, and the results pointed to a high risk of nitrate pollution in the central region of the plain, emphasizing the importance of such studies for risk management in semi-arid areas.

These studies highlight the relevance of using advanced modeling techniques and machine learning in the management and preservation of groundwater, particularly in regions vulnerable to contamination and water scarcity. The adoption of these approaches is crucial for the development of sustainable strategies that ensure the quality and availability of the resource.

Figure 1.8 presents a network diagram, obtained using the VOSviewer software, which is a tool used to construct and visualize bibliometric networks, such as term co-occurrence networks, citation networks, or co-authorship networks. In bibliometric analyses, a term refers to a word or expression that frequently appears in the analyzed literature, representing key

concepts and thematic trends in the field. The co-occurrence of terms indicates how often they appear together in the same documents, helping to identify research patterns and emerging topics. At the center of the diagram, the terms "machine learning" and "groundwater" are prominently used, indicating that these are the most frequently mentioned concepts that come together in the literature. This suggests the emphasis of research on the use of machine learning techniques for the study and management of groundwater.



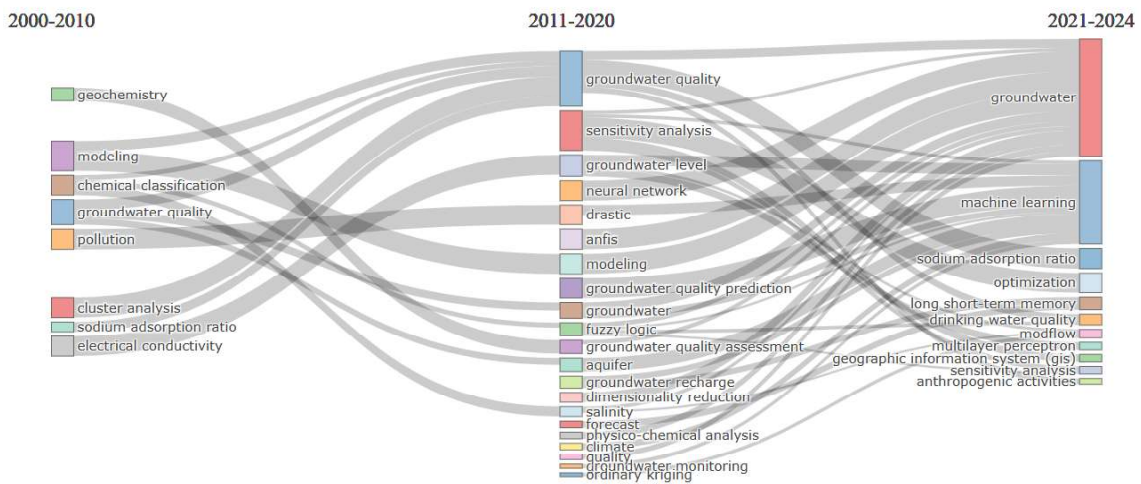
**Figure 1.8.** Network diagram showing a bibliometric analysis of terms related to groundwater research.

The different clusters are represented by distinct colors and indicate groups of terms that frequently appear together, reflecting specific research areas within the broader field of groundwater and machine learning. In this bibliometric analysis, the terms "groundwater" and "machine learning" are highlighted, reflecting the growing trend of applying machine learning techniques in the study and management of groundwater. Both terms are closely related to water quality, salinity, irrigation, and quantification methods such as artificial intelligence, prediction, and classification of chemical and geological parameters. Issues related to health, arsenic contamination, and salinity have emerged as consequences associated with the domain. This cluster illustrates the complexity and interconnectivity of groundwater research, highlighting

the growing importance of machine learning techniques and artificial intelligence in the analysis, management, and optimization of groundwater resources. Each cluster reveals different subareas of focus, ranging from water quality and contamination to advanced modeling and mapping techniques.

The thematic evolution of keywords in articles published on groundwater quality from 2000 to 2024 (Figure 1.9) is represented by a Sankey diagram, where the nodes correspond to the themes identified in different periods, and the arrows indicate the transition of these themes over time. The thickness of the arrows is proportional to the inclusion index, meaning that the greater the inclusion, the higher the similarity between the themes of different periods. Additionally, the area of the boxes represents the abundance of the theme in the literature, indicating its relevance within the field of study. To construct this analysis, a bibliometric approach was applied to track how specific research topics evolved over time. The co-occurrence of keywords in published studies was examined across different periods, revealing shifts in scientific focus. Keywords that frequently appeared together in older studies were compared to those in more recent periods, allowing the identification of emerging trends and methodological advances. This enables the visualization of how initial research topics, such as physicochemical characterization, gradually integrated new approaches like statistical modeling and, more recently, artificial intelligence techniques.

During the period from 2000 to 2010, topics such as "chemical classification," "electrical conductivity," "sodium absorption," and "pollutant contamination" were prominent, reflecting the initial concern with the physicochemical characterization of groundwater and its environmental impacts. In the following decade, there was a shift toward interdisciplinary methods, with an increased use of combined approaches involving geoprocessing, hydrological modeling, and statistical analyses. This period marks a maturation of research, with greater integration of different areas of knowledge.



**Figure 1.9.** Thematic evolution of keywords in groundwater quality papers from 2000 to 2024.

In the last four years (2020–2024), the thematic evolution indicates a convergence toward the application of machine learning techniques and complex statistical tools in the assessment of groundwater quality. This trend suggests an enhancement of computationally advanced analyses, allowing greater accuracy in predicting contamination patterns, as well as in identifying critical factors for the sustainable management of water resources.

#### 1.4 Conclusion

Since 2020, a notable increase has been observed in scientific production on groundwater quality, driven by global challenges such as population growth, climate change, and agricultural expansion, which exacerbate groundwater contamination and degradation issues. This trend reflects an intensification of interest in understanding and mitigating the threats that compromise the quality of these waters, particularly in countries with large populations or significant investments in science and technology, such as India, China, Iran, and the United States.

The bibliometric analysis reveals an important shift in research areas, with a growing focus on interdisciplinary approaches, such as the use of machine learning, geoprocessing, and hydrological modeling, which have deepened the understanding and resolution of problems related to groundwater quality. These technological and methodological advancements demonstrate a clear shift towards more integrated studies, utilizing innovative tools for water quality monitoring and management, going beyond traditional physicochemical characterization.

The rise of predictive methods and advanced tools, such as machine learning, not only improves the ability to predict changes in water quality but also indicates the maturation of research, which now embraces more complex and interdisciplinary solutions. Finally, the state of the art in groundwater research highlights the growing importance of international collaboration and the central role of new technologies in overcoming challenges related to the quality and management of this resource. The evolution of research reflects a robust response to contemporary challenges and a continued commitment to the sustainability and protection of groundwater, essential for human and environmental well-being.

### 1.5 References

Adeniyi, A.; Giwa, O. (2021). Accumulation and health effects of metals in selected urban groundwater. **Physical Sciences Reviews**. <https://doi.org/10.1515/psr-2020-0089>.

Barrett, M. H., Hiscock, K. M., Pedley, S., Lerner, D. N., Tellam, J. H., e French, M. J. (1999). Marker species for identifying urban groundwater recharge sources: A review and case study in Nottingham, UK. **Water Research**, 33, 3083-3097. [https://doi.org/10.1016/S0043-1354\(99\)00021-4](https://doi.org/10.1016/S0043-1354(99)00021-4).

Baas, J., Schotten, M., Plume, A., Côté, G., e Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. **Quantitative Science Studies**, 1(1), 377–386. [https://doi.org/10.1162/qss\\_a\\_00019](https://doi.org/10.1162/qss_a_00019).

Bose, S., Mazumdar, A., e Basu, S. (2023). Evolution of groundwater quality assessment on urban area,a bibliometric analysis. **Groundwater for Sustainable Development**, 20, 100894. <https://doi.org/10.1016/j.gsd.2022.100894>.

Chen, W., Li, H., Hou, E., Wang, S., Wang, G., Panahi, M., Li, T., Peng, T., Guo, C., Niu, C., Xiao, L., Wang, J., Xie, X., Ahmad, B. B. (2018). GIS-based groundwater potential analysis using novel ensemble weights-of-evidence with logistic regression and functional tree models. *Science of The Total Environment*, 634, 853-867. <https://doi.org/10.1016/j.scitotenv.2018.04.055>.

Chegbeleh, L. P., Akurugu, B. A., e Yidana, S. M. (2020). Assessment of groundwater quality in the Talensi District, Northern Ghana. **Scientific World Journal**, 2020, 8450860. <https://doi.org/10.1155/2020/8450860>.

Chotpantararat, S., Parkchai, T., e Wisitthammasri, W. (2020). Multivariate statistical analysis of hydrochemical data and stable isotopes of groundwater contaminated with nitrate at Huay Sai Royal Development Study Center and adjacent areas in Phetchaburi Province, Thailand. **Water**, 12, 1127. <https://doi.org/10.3390/W12041127>.

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., e Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. **Journal of Business Research**, 133, 285–296. <https://doi.org/10.1016/j.jbusres.2021.04.070>.

Devic, G., Djordjevic, D., e Sakan, S. (2014). Natural and anthropogenic factors affecting the groundwater quality in Serbia. **Science of the Total Environment**, 468–469, 933-942. <https://doi.org/10.1016/j.scitotenv.2013.09.011>.

Foster, S., Chilton, J., Nijsten, G.J., e Richts, A. (2013). Groundwater, a global focus on the ‘local resource’. **Current Opinion in Environmental Sustainability**, 5(6), 685-695. <https://doi.org/10.1016/j.cosust.2013.10.010>.

Fu, Y., Cao, W., Pan, D., e Ren, Y. (2022). Changes of groundwater arsenic risk in different seasons in Hetao basin based on machine learning model. **Science of the Total Environment**, 817, 153058. <https://doi.org/10.1016/j.scitotenv.2022.153058>.

Graham, J. P., e Polizzotto, M. L. (2013). Pit latrines and their impacts on groundwater quality: a systematic review. **Environmental Health Perspectives**, 121(5), 521-530. <https://doi.org/10.1289/ehp.1206028>.

Hao, H., Li, P., Li, K., Shan, Y., Liu, F., Hu, N., Zhang, B., Li, M., Sang, X., Xu, X., Lv, Y., Chen, W., Jiao, W. (2024). A novel prediction approach driven by graph representation learning for heavy metal concentrations. **Science of the Total Environment**, 947, 174713. <https://doi.org/10.1016/j.scitotenv.2024.174713>.

Li, H., Dong, Q., Zhang, M., Gong, T., Zan, R., Wang, W. (2023). Transport behavior difference and transport model of long- and short-chain per- and polyfluoroalkyl substances in underground environmental media: a review. **Environmental Pollution**, 327, 121579. <https://doi.org/10.1016/j.envpol.2023.121579>.

Inam Ullah Ehsan, Ahmad, S., Khokhar, M. F., Khayyam, U., Azmat, M., Arshad, M., e Kaiser, F. U. R. (2024). Analyzing land use land cover (LULC) changes induced by the run-of-river project and respondent survey: a case of Ghazi Barotha Hydropower Project on Indus River, Pakistan. **Environmental Research Communications**, 6(3), 035002. <https://doi.org/10.1088/2515-7620/ad2bb5>.

Janardhana Raju, N.; Shukla, U. K.; Ram, P. (2011). Hydrogeochemistry for the assessment of groundwater quality in Varanasi: a fast-urbanizing center in Uttar Pradesh, India. **Environmental Monitoring and Assessment**, v. 173, p. 279–300, 2011. <https://doi.org/10.1007/s10661-010-1387-6>.

Kannazarova, Z., Juliev, M., Abuduwaili, J., Muratov, A., e Bekchanov, F. (2024). Drainage in irrigated agriculture: Bibliometric analysis for the period of 2017–2021. **Agricultural Water Management**, 305, 109118. <https://doi.org/10.1016/j.agwat.2024.109118>.

Katz, B. G., Eberts, S. M., e Kauffman, L. J. (2011). Using Cl/Br ratios and other indicators to assess potential impacts on groundwater quality from septic systems: a review and examples from principal aquifers in the United States. **Journal of Hydrology**, 397(3-4), 151-166.

Lee, S., Kaown, D., Koh, E. H., Lee, H. L., Ko, K. S., e Lee, K. K. (2022). Advanced utilization of multi-learning algorithm: ensemble super learner to map groundwater potential for potable mineral water. **Geocarto International**, v. 37, n. 25, p. 9897–9916. <https://doi.org/10.1080/10106049.2022.2025921>.

Li, P., Wu, J., Qian, H., Lyu, X., e Liu, H. (2014). Origin and assessment of groundwater pollution and associated health risk: a case study in an industrial park, northwest China. **Environmental Geochemistry and Health**, v. 36, p. 693–712, 2014. <https://doi.org/10.1007/s10653-013-9590-3>.

Li, C., Xie, L., e Xiong, Y. (2019). Bioelectrochemical systems for groundwater remediation: the development trend and research front revealed by bibliometric analysis. **Water**, 11(8), 1532. <https://doi.org/10.3390/w11081532>.

Li, H., Dong, Q., Zhang, M., Gong, T., Zan, R., e Wang, W. (2023). Transport behavior difference and transport model of long- and short-chain per- and polyfluoroalkyl substances in underground environmental media: a review. **Environmental Pollution**, 327, 121579. <https://doi.org/10.1016/j.envpol.2023.121579>.

Loftis, J. C. (1996). Trends in groundwater quality. **Hydrological Processes**, 10(2), 335-355. [https://doi.org/10.1002/\(SICI\)1099-1085\(199602\)10:2<335::AID-HYP359>3.0.CO;2-T](https://doi.org/10.1002/(SICI)1099-1085(199602)10:2<335::AID-HYP359>3.0.CO;2-T).

Mallick, J., Naikoo, M. W., Talukdar, S., Ahmed, I. A., Rahman, A., Islam, A. R. M. T., Pal, S., Ghose, B., e Shashtri, S. (2021). Developing groundwater potentiality models by coupling ensemble machine learning algorithms and statistical techniques for sustainable groundwater management. **Geocarto International**, 37(25), 7927–7953. <https://doi.org/10.1080/10106049.2021.1987535>.

Naghibi, S. A., Pourghasemi, H. R., e Dixon, B. (2016). GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. **Environmental Monitoring and Assessment**, 188(1), 1-27.

Naghibi, S. A., Ahmadi, K., e Daneshi, A. (2017). Application of support vector machine, random forest, and genetic algorithm optimized random forest models in groundwater potential mapping. **Water Resources Management**, 31(2), 1-15.

Oiro, S., Comte, J.-C., Soulsby, C., MacDonald, A., e Mwakamba, C. (2020). Depletion of groundwater resources under rapid urbanisation in Africa: recent and future trends in the Nairobi Aquifer System, Kenya. **Hydrogeology Journal**, 28(8), 2635-2656. <https://doi.org/10.1007/s10040-020-02236-5>.

Podgorski, J., e Berg, M. (2020). Global threat of arsenic in groundwater. *Science*, 368(6493), 845-850. <https://doi.org/10.1126/science.aba1510>.

Rahaman, E. I. U., Ahmad, S., Khokhar, M. F., Khayyam, U., Azmat, M., Arshad, M., e Qaiser, F. U. R. (2024). Analyzing land use land cover (LULC) changes induced by the run-of-river project and respondent survey: A case of Ghazi Barotha Hydropower Project on Indus River, Pakistan. **Environmental Research Communications**, 6(3), 035002. <https://doi.org/10.1088/2515-7620/ad2bb5>.

Rodriguez-Galiano, V., Mendes, M. P., Garcia-Soldado, M. J., Chica-Olmo, M., e Ribeiro, L. (2014). Predictive modeling of groundwater nitrate pollution using Random Forest and multisource variables related to intrinsic and specific vulnerability: A case study in an agricultural setting (Southern Spain). **Science of The Total Environment**, 476–477, 189–206. <https://doi.org/10.1016/j.scitotenv.2014.01.001>.

Sajedi-Hosseini, F., Malekian, A., Choubin, B., Rahmati, O., Cipullo, S., Coulon, F., e Pradhan, B. (2018). A novel machine learning-based approach for the risk assessment of nitrate groundwater contamination. **Science of The Total Environment**, v. 644, p. 954-962, 2018. <https://doi.org/10.1016/j.scitotenv.2018.07.054>.

Sahoo S., Russo T. A., Elliott J. (2017). Machine learning algorithms for modeling groundwater level changes in agricultural regions of the U.S. **Water Resources Research**, v. 53, p. 3878–3895, 2017. Disponível em: <https://doi.org/10.1002/2016WR019933>.

Sakthipriya, E., e Chandrakumar, T. (2024). Weather based paddy yield prediction using machine learning regression algorithms. *Journal of Agrometeorology*, 26(3), 344–348. <https://doi.org/10.54386/jam.v26i3.2598>. Subba Rao, N. (2006). Seasonal variation of groundwater quality in a part of Guntur District, Andhra Pradesh, India. *Environmental Geology*, 49, 413–429. <https://doi.org/10.1007/s00254-005-0089-9>.

Riboli, S. A. e Lindino, C. (2023). Análise de componentes principais (PCA) na discriminação de fontes de água potável. **Revista Tecnia**, 5(1). <https://periodicos.ifg.edu.br/tecnica/article/view/577>.

Singha, S., Pasupuleti, S., Singha, S. S., Singh, R., e Kumar, S. (2021). Prediction of groundwater quality using efficient machine learning technique. **Chemosphere**, 276, 130265. <https://doi.org/10.1016/j.chemosphere.2021.130265>.

Vliet, M. T. H. Van, Jones, E. R., Flörke, M., Franssen, W. H. P., e Hanasaki, N. (2021). Global water scarcity including surface water quality and expansions of clean water technologies. **Global Water Scarcity**.

Wisitthammasri, W., Chotpantarat, S., e Thitimakorn, T. (2020). Multivariate statistical analysis of the hydrochemical characteristics of a volcano sedimentary aquifer in Saraburi Province, Thailand. **Journal of Hydrology: Regional Studies**, 32, 100745. <https://doi.org/10.1016/j.ejrh.2020.100745>.

Wu, Z., Lu, C., Sun, Q., Lu, W., He, X., Qin, T., Yan, L., e Wu, C. (2023). Predicting groundwater level based on machine learning: a case study of the Hebei Plain. **Water**, 15, 823. <https://doi.org/10.3390/w15040823>.

Wu, J., Cao, Y., Islam, M. S., e Emch, M. (2025). Application of machine learning to identify influential factors for fecal contamination of shallow groundwater. **Water**, 17, 160. <https://doi.org/10.3390/w17020160>.

Wu, J., Li, P., Wang, D., Ren, X., e Wei, M. (2020). Statistical and multivariate statistical techniques to trace the sources and affecting factors of groundwater pollution in a rapidly growing city on the Chinese Loess Plateau. **Human and Ecological Risk Assessment: An International Journal**, 26(6), 1603–1621. <https://doi.org/10.1080/10807039.2019.1594156>.

Fu, Y., Cao, W., Pan, D., Y. Ren. (2022) Changes of groundwater arsenic risk in different seasons in Hetao basin based on machine learning model. **Sci. Total Environ.**, 817 (2022), 10.1016/j.scitotenv.2022.153058.

Zhan, X., Liu, W., Chen, S., Yao, R., Yang, C., Zhang, X., Li, J., Wang, Y., e Zhang, Y. (2025). Machine learning approaches to identify hydrochemical processes and predict drinking water quality for groundwater environment in a metropolis. **Journal of Hydrology: Regional Studies**, 58, 102227. <https://doi.org/10.1016/j.jhydrol.2025.102227>.

Zhang, S., Mao, G., Crittenden, J., Liu, X., e Du, H. (2017). Groundwater remediation from the past to the future: a bibliometric analysis. **Water Research**, 119, 114-125. <https://doi.org/10.1016/j.watres.2017.01.029>.

Zhou, Y., Yang, F., Wu, X., Jia, C., Liu, S., e Gao, Y. (2020). Bibliometric analysis of research progress on karst groundwater pollution. IOP Conference Series: **Earth and Environmental Science**, 568, 012040. <https://doi.org/10.1088/1755-1315/568/1/012040>.

Zyoud, S. H., e Fuchs-Hanusch, D. (2017). Estimates of Arab world research productivity associated with groundwater: a bibliometric analysis. **Applied Water Science**, 7(3), 1255-1272. <https://doi.org/10.1007/s13201-016-0520-2>.

## CAPÍTULO 2

### Potencial hídrico subterrâneo: modelagem espacial a partir de machine learning

#### 2.1 Introdução

Nas últimas décadas, a demanda por águas subterrâneas vem aumentando, impulsionada por fatores como crescimento populacional, urbanização acelerada e expansão da agricultura irrigada (DECHASA DIRIBA, 2024). Em diversas regiões, esse recurso desempenha papel estratégico tanto no abastecimento urbano quanto rural, além de subsidiar atividades agropecuárias essenciais (BEKELE, 2021; BELAY, 2022). Contudo, garantir seu uso sustentável exige compreender não apenas a disponibilidade, mas também a dinâmica de fluxo e as condições de recarga dos aquíferos.

Em escala global, a distribuição dos recursos hídricos é desigual. Países como Canadá e Brasil possuem bacias hidrográficas extensas e abundantes em água (LI e WU, 2024), embora a sua distribuição heterogênea promova grave escassez em muitas regiões (MARENGO e BERNASCONI, 2015; CARVALHO *et al.*, 2020).

Dentro deste contexto, regiões com clima semiárido, caracterizadas por distribuição irregular de precipitação, longos períodos de estiagem e elevadas taxas de evapotranspiração, enfrentam restrições hídricas severas, uma situação cuja tendência é de agravamento diante das mudanças climáticas (KAHIL, DINAR; ALBIAC, 2015). No Brasil, apesar da grande reserva hídrica, a distribuição espacial é altamente desigual; em regiões como o norte de Minas Gerais há historicamente a escassez de água e forte dependência de aquíferos para abastecimento humano e atividades produtivas (MARENGO e BERNASCONI, 2015; CARVALHO *et al.*, 2020).

A complexidade hidrogeológica dessas regiões está associada a aquíferos fraturados e, por vezes, cársticos, cuja recarga é fortemente condicionada por eventos pluviométricos intensos, cobertura vegetal e heterogeneidade litológica (HARTMANN *et al.*, 2017; SMAIL *et al.*, 2019; TURKELTAUB e BEL, 2024). Esses fatores, combinados com mudanças no uso e cobertura da terra, desmatamento e práticas agrícolas inadequadas, elevam a vulnerabilidade ecológica e podem acelerar processos de desertificação (SILVA *et al.*, 2020; CASTRO *et al.*, 2019). Assim, compreender o potencial aquífero nessas condições demanda métodos capazes de integrar múltiplas variáveis ambientais e hidrogeológicas.

Diversos estudos têm utilizado variáveis como geologia, geomorfologia, declividade, uso e cobertura da terra, características do solo, densidade de drenagem e precipitação para delimitar zonas potenciais de água subterrânea (*Groundwater Potential Zones – GWPZs*) (ABDOWUDAD, 2021; AGUMASE e KINDIE, 2019; YIFRU *et al.*, 2020; SEIFU *et al.*, 2023). Tradicionalmente, técnicas geoestatísticas como krigagem ou interpolação inversa da distância têm sido empregadas para estimar a distribuição espacial de parâmetros como nível estático ou vazão (GOOVAERTS, 1997; CHITSAZAN *et al.*, 2015). No entanto, esses métodos assumem continuidade espacial baseada na proximidade entre pontos amostrados, o que pode limitar sua precisão em contextos com heterogeneidade litológica acentuada, baixa densidade amostral ou transições abruptas nas características do meio físico.

Nesse cenário, o aprendizado de máquina (*machine learning – ML*) surge como alternativa promissora, capaz de integrar dados ambientais multiescalares e capturar relações não lineares entre variáveis explicativas e a vazão dos poços (RAHMATI *et al.*, 2016). Algoritmos como *Support Vector Machines* (SVM), *Random Forest* (RF), *k-Nearest Neighbors* (KKNN) e regressão linear múltipla (LM) conseguem capturar padrões não lineares e interações complexas entre variáveis, muitas vezes superando modelos puramente baseados em processos físicos (KARIMI *et al.*, 2012; WEI *et al.*, 2021; TAO *et al.*, 2022). Além disso, sua integração com dados de sensoriamento remoto e sistemas de informações geográficas (SIG) amplia a capacidade de análise espacial permitindo gerar mapas preditivos robustos e de maior aplicabilidade para o planejamento hídrico (TAMIRU e WAGARI, 2021; TAMESGEN, ATLABACHEW e JOTHIMANI 2023). Apesar de sua relevância, os modelos baseados em processos físicos na hidrologia de águas subterrâneas apresentam limitações, como a dependência de parâmetros de entrada de difícil mensuração e a alta complexidade para representar sistemas hidrogeológicos heterogêneos. Essas restrições podem comprometer a precisão das estimativas, especialmente em regiões com dados escassos (BORZI, 2025).

Diante destes desafios, reitera-se a necessidade da integração de métodos de aprendizado de máquina, a fim de complementar modelos tradicionais, com vistas ao aprimoramento dos modelos preditivos. Modelos de aprendizado de máquina usam dados históricos para identificar padrões e tendências e não exigem uma compreensão de processos hidrológicos físicos complexos (KARIMI *et al.*, 2012). Notavelmente, modelos de aprendizado de máquina podem superar modelos baseados em processos físicos na previsão de níveis de águas subterrâneas (WEI *et al.*, 2021), os tornando de interesse para a previsão de níveis de águas subterrâneas (TAO *et al.*, 2022). Para esta finalidade, estudos reportam o emprego de redes neurais artificiais (LALLAHEM *et al.*, 2005), modelos fuzzy (NADIRI *et al.*,

2019), modelos baseados em árvore (Wang *et al.*, 2018), programação genética (KASIVISWANATHAN *et al.*, 2016) e técnicas emergentes como Gradient Boosting (HUSSEIN *et al.*, 2020), XGBoost (OSMAN *et al.*, 2021) e Extreme Learning Machine (ALIZAMIR *et al.*, 2018).

Entretanto, grande parte dos estudos de modelagem de potencial aquífero realizam a identificação qualitativa de áreas favoráveis à ocorrência de água subterrânea, não realizando estudos quantitativos em termos da vazão dos poços. Em linhas gerais, em regiões semiáridas e hidrogeologicamente complexas, a vazão representa um indicador direto da produtividade do aquífero e é essencial para o planejamento hídrico. Adicionalmente, poucos trabalhos exploram comparativamente o desempenho de diferentes algoritmos de ML para a estimativa da vazão de poços.

Diante desse cenário, este estudo tem como objetivo modelar espacialmente o potencial aquífero em uma região semiárida no norte de Minas Gerais (Brasil), estimando a vazão específica dos poços tubulares a partir de um conjunto abrangente de covariáveis ambientais e hidrogeológicas. Os objetivos específicos do estudo correspondem: (i) identificar as variáveis ambientais mais relevantes para a predição da vazão; (ii) comparar o desempenho dos algoritmos de machine learning para determinar a vazão da água subterrânea; e (iii) gerar mapas preditivos que subsidiem a gestão eficiente e o planejamento sustentável dos recursos hídricos subterrâneos. A originalidade desta pesquisa reside na aplicação integrada de algoritmos supervisionados para estimativa de vazão em ambiente semiárido, oferecendo uma abordagem inovadora para apoiar a tomada de decisão em áreas vulneráveis à escassez hídrica.

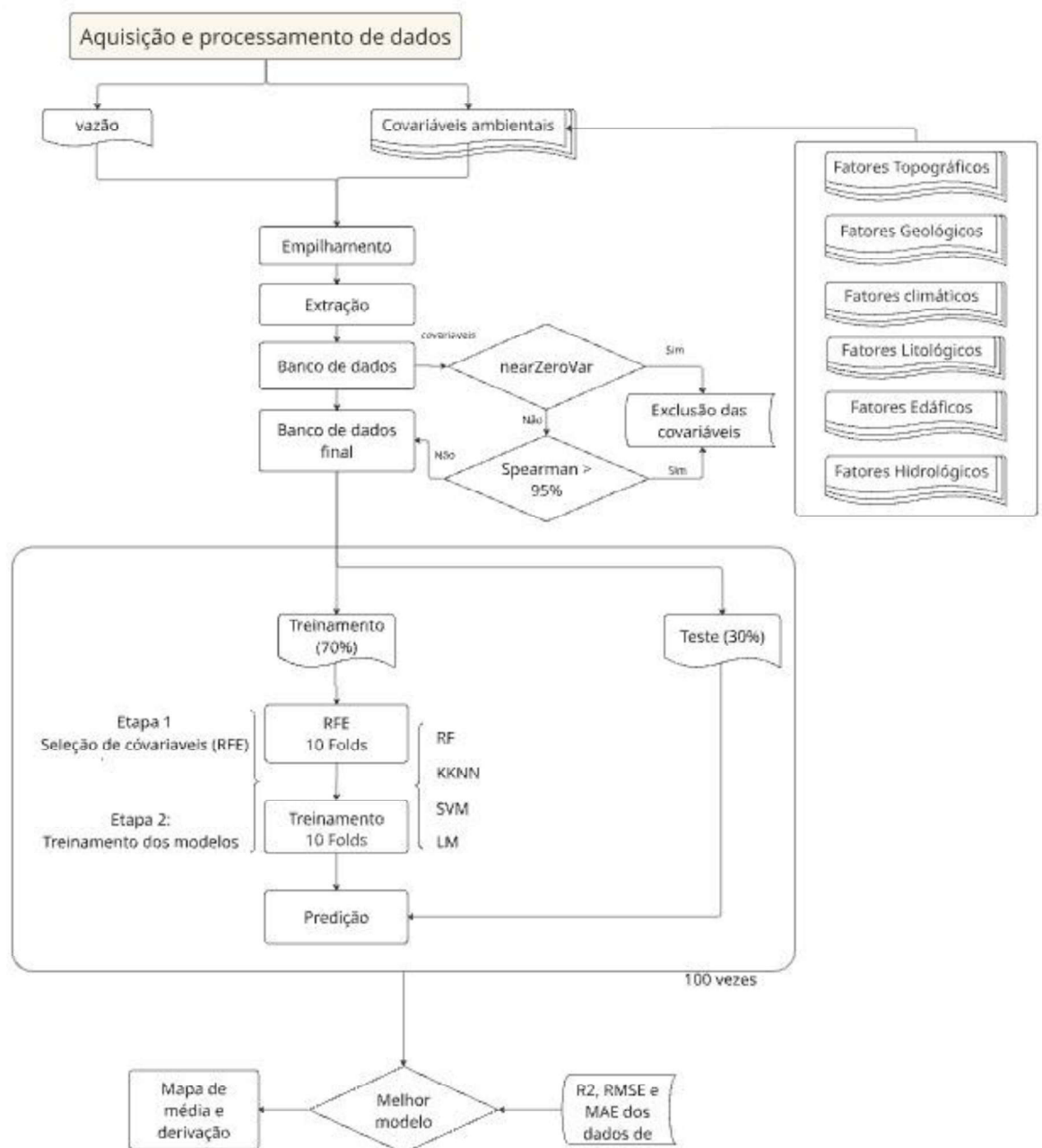
## 2.2 Materiais e Métodos

A metodologia adotada neste estudo envolveu a aplicação de técnicas de aprendizado de máquina para modelar o potencial hídrico subterrâneo com base em covariáveis ambientais multiescalares. O processo metodológico foi estruturado em quatro etapas principais:

- (i) aquisição e pré-processamento de dados;
- (ii) seleção e preparação das variáveis explicativas;
- (iii) treinamento e validação dos modelos; e
- (iv) geração e análise dos mapas de predição.

Na Figura 2.1 apresenta-se o fluxograma geral da metodologia empregada, desde a integração dos dados de vazão com as covariáveis ambientais até a seleção do modelo com

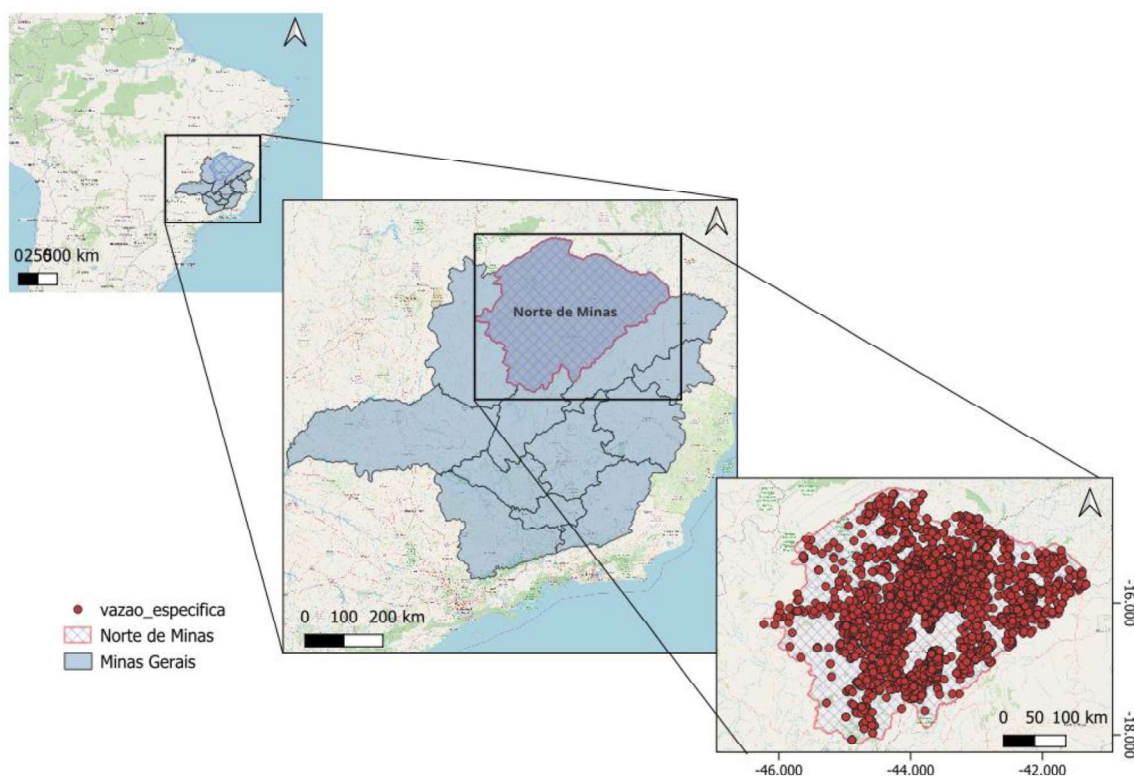
melhor desempenho e a elaboração dos mapas finais. A etapa de seleção de variáveis envolveu a aplicação dos critérios *near zero variance* e correlação de *Spearman* superior a 95% para reduzir a multicolinearidade. Em seguida, os dados foram divididos em subconjuntos de treino (70%) e teste (30%) para avaliar o desempenho dos algoritmos *Random Forest* (RF), *Weighted k-Nearest Neighbors* (KKNN), *Support Vector Machine* (SVM) e Regressão Linear Múltipla (LM). Os modelos foram comparados com base nas métricas  $R^2$ , RMSE e MAE, sendo o modelo com melhor performance utilizado para gerar os mapas contínuos de potencial aquífero.



**Figura 2.1.** Fluxograma do processo metodológico de aquisição, seleção de covariáveis, modelagem e geração de mapas preditivos do potencial hídrico subterrâneo.

### 2.2.1 Área de estudo

A área de estudo corresponde à região norte do estado de Minas Gerais, situada na porção setentrional da região Sudeste do Brasil (Figura 2.2). Essa região apresenta clima semiárido, solos heterogêneos, relevo variado e elevada dependência dos recursos hídricos subterrâneos para o abastecimento humano e atividades agrícolas. Geograficamente, está compreendida entre as coordenadas aproximadas de  $41^{\circ}33'$  a  $46^{\circ}28'$  de longitude oeste e  $14^{\circ}24'$  a  $18^{\circ}33'$  de latitude sul.



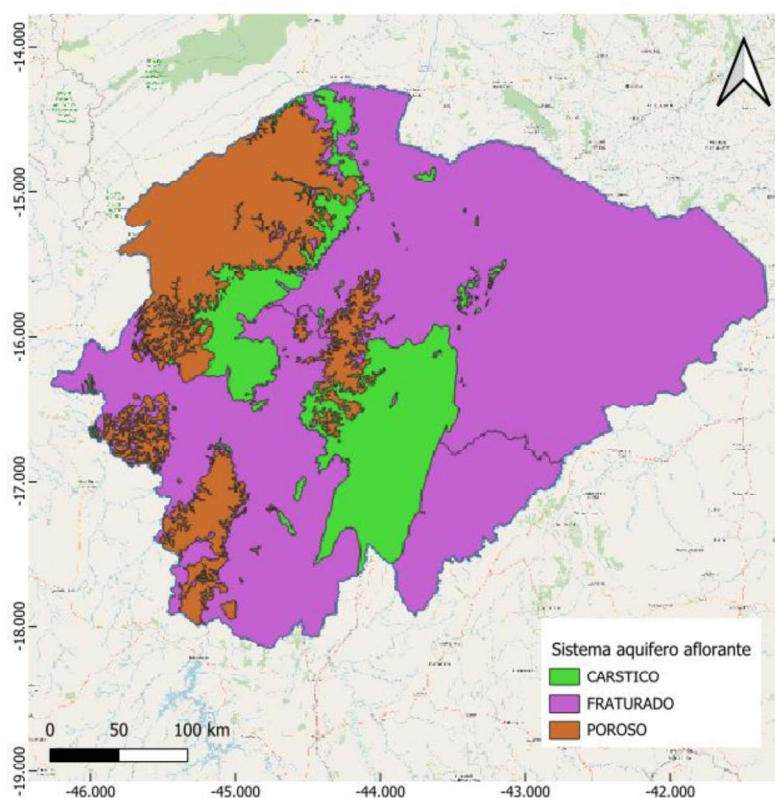
**Figura 2.2.** Localização da área de estudo no norte de Minas Gerais, Brasil, com distribuição espacial dos poços tubulares utilizados na modelagem.

Segundo a Associação dos Municípios da Área Mineira da Sudene (AMAMS, 2023), o norte de Minas Gerais abrange 89 municípios distribuídos em uma área estimada de 128.454 km<sup>2</sup>, com população de aproximadamente 1.779.701 habitantes. Montes Claros destaca-se

como o principal centro urbano e polo econômico da região, concentrando elevadas demandas de abastecimento hídrico.

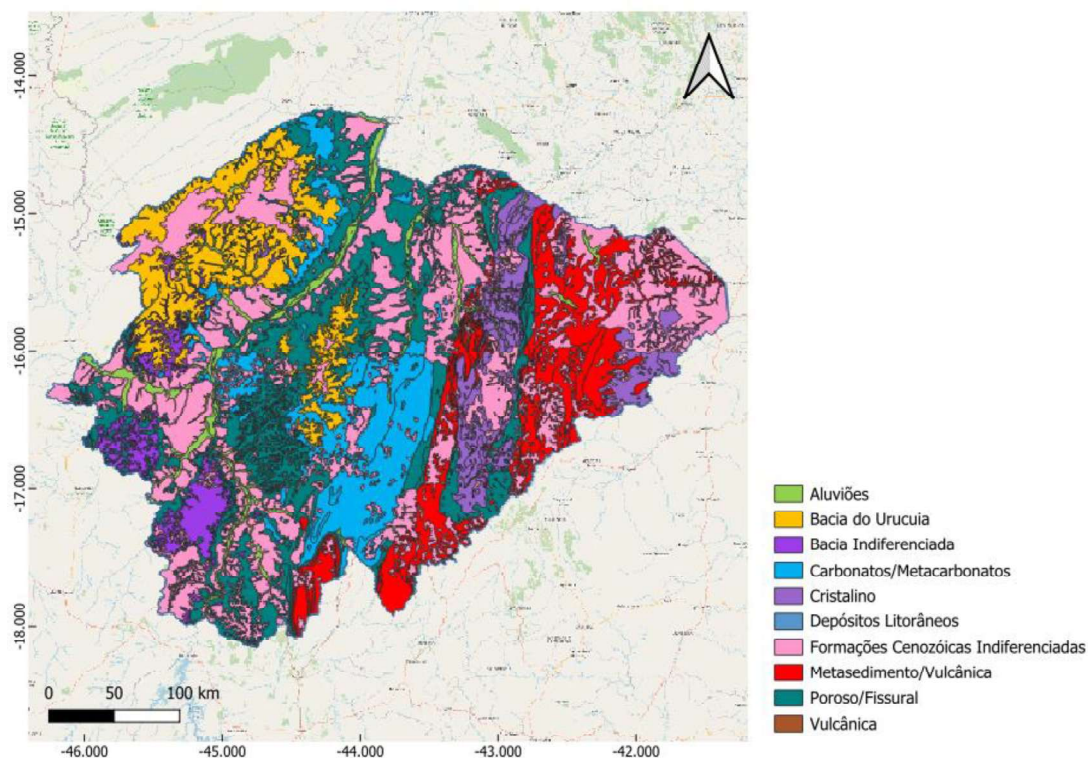
Em regiões semiáridas como o norte de Minas Gerais, a escassez de recursos hídricos superficiais e a elevada variabilidade climática limitam o desenvolvimento socioeconômico (CÂNDIDO *et al.*, 2019). O regime pluviométrico é fortemente concentrado em poucos meses do ano, com um período seco de aproximadamente oito meses e uma estação chuvosa de cerca de quatro meses. A precipitação anual varia entre 600 e 1.200 mm, com gradiente decrescente do sudoeste para o nordeste, e a evapotranspiração potencial elevada contribui para uma balança hídrica frequentemente negativa (MONTENEGRO E RAGAB, 2010; JARDIM *et al.*, 2023).

Geologicamente, a região apresenta alta diversidade litológica e topográfica, com ocorrência de formações sedimentares, rochas carbonáticas, sistemas cársticos e compartimentos cristalinos fraturados (SILVA *et al.*, 2020; CASTRO *et al.*, 2019). A Figura 2.3 apresenta o sistema aquífero aflorante na área de estudo, destacando a distribuição espacial das principais unidades aquíferas.



**Figura 2.3.** Sistema aquífero aflorante, indicando a distribuição espacial das unidades aquíferas presentes na área de estudo.

Além disso, para caracterizar o arcabouço geológico regional, foi elaborado o mapa de domínios hidrogeológicos (Figura 2.4), mantendo a taxonomia do mapeamento de referência (e.g., cristalino, carbonatos/metacarbonatos, bacias sedimentares e aluviões). Este produto tem caráter descritivo e não foi utilizado diretamente como variável de entrada nos modelos.



**Figura 2.4** – Domínios hidrogeológicos, mantendo a taxonomia do mapeamento de referência

A rede hidrográfica da região é formada principalmente pelas bacias dos rios São Francisco, Pardo e Jequitinhonha. O clima predominante é classificado como Aw, segundo Köppen, com forte sazonalidade pluviométrica.

Para fins deste estudo, aspectos geológicos, geomorfológicos, pedológicos e climáticos da área foram avaliados para subsidiar a modelagem espacial do potencial aquífero. Os dados espaciais dessas variáveis foram integrados a informações de poços tubulares existentes, cuja distribuição está apresentada na Figura 2.2.

### 2.2.2 Base de Dados

A variável resposta utilizada neste estudo foi a vazão de poços tubulares ( $m^3/h$ ), obtida a partir do Sistema de Informações de Águas Subterrâneas (SIAGAS) (2025). A base original

é composta da aproximadamente 3.700 registros georreferenciados, abrangendo toda a região de estudo. Cada registro inclui informações sobre a localização geográfica dos poços, características construtivas e medições de vazão realizadas durante testes de bombeamento ou levantamento cadastral.

Além da variável resposta, foram integradas ao banco de dados camadas ambientais auxiliares que representam potenciais fatores explicativos da variação de vazão, tais como litologia, geomorfologia, topografia, uso e cobertura da terra, tipos de solo e variáveis climáticas. Essas informações foram obtidas de bases públicas e institucionais, compatíveis com a escala regional e em formato georreferenciado, possibilitando a extração posterior dos valores correspondentes a cada poço amostrado.

Além da variável resposta, foram incorporadas ao banco de dados camadas ambientais consideradas potenciais preditoras da variação de vazão, a saber:

- Litologia: derivada do mapeamento geológico oficial (CPRM, 2023), posteriormente reclassificada segundo critérios químicos e mineralógicos.
- Geomorfologia: obtida do mapeamento geomorfológico do IBGE (2019).
- Topografia: extraída de modelo digital de elevação SRTM (NASA, 2022).
- Uso e cobertura da terra: proveniente do MapBiomias (Coleção 8, 2023).
- Tipos de solo: obtidos do Sistema Brasileiro de Classificação de Solos – SiBCS (EMBRAPA, 2022).
- Variáveis climáticas: incluindo precipitação média anual e evapotranspiração potencial, extraídas do banco *WorldClim* v2.1 (FICK e HIJMANS, 2017).

As coordenadas originais dos poços foram avaliadas e padronizadas para o sistema de referência SIRGAS 2000 (EPSG:4674), garantindo consistência espacial nas análises. Registros duplicados foram identificados e removidos com base no código único do poço e em sua localização geográfica. Também foram eliminados valores inconsistentes, como vazões negativas ou manifestamente incompatíveis com a realidade hidrogeológica regional.

Para evitar distorções estatísticas causadas por valores extremos e melhorar a robustez da modelagem, aplicou-se a regra do intervalo interquartil (IQR), conforme proposta por Tukey (1977), para detecção e remoção de outliers. Esse procedimento resultou na exclusão de observações com vazões que se encontravam fora do limite calculado por:

$$\text{IQR} = Q3 - Q1$$

$$\text{Limite inferior} = Q1 - 1.5 \times \text{IQR}$$

$$\text{Limite superior} = Q3 + 1.5 \times \text{IQR}$$

Manter valores extremos na modelagem espacial de vazão pode gerar mapas superestimados, induzindo interpretações incorretas sobre o potencial aquífero. Em áreas onde não existem as mesmas condições estruturais ou hidrogeológicas que favorecem altas vazões, o modelo poderia atribuir produtividades irreais, levando a decisões de perfuração equivocadas e investimentos improdutivos. Além disso, tais valores influenciam desproporcionalmente a ponderação das covariáveis nos algoritmos, reduzindo a capacidade de identificar padrões representativos da maioria dos poços. Alguns estudos recomendam a remoção ou tratamento desses valores para minimizar vieses e aumentar a robustez preditiva (RAHMATI *et al.*, 2016; ARABAMERI *et al.*, 2020; BAI *et al.*, 2022). Assim, a exclusão dos outliers neste trabalho não apenas melhora a qualidade estatística dos dados, mas também assegura maior confiabilidade na aplicação prática dos resultados para o planejamento e gestão dos recursos hídricos subterrâneos.

A filtragem permitiu obter uma distribuição mais representativa das condições hidrogeológicas regionais, reduzindo o viés causado por poços atípicos com vazões excepcionalmente elevadas.

Após a consolidação e limpeza da base, os dados foram espacializados em ambiente de Sistema de Informação Geográfica (SIG), possibilitando a análise de sua distribuição geográfica e a posterior interseção com as camadas ambientais selecionadas. Essa etapa preparatória gerou o conjunto de covariáveis que será utilizado na modelagem preditiva, conforme descrito nas próximas seções da metodologia.

Esse procedimento permitiu a obtenção de uma distribuição mais representativa da realidade hidrogeológica regional, reduzindo o viés causado por poços atípicos com vazões excepcionalmente elevadas. Todas as análises foram conduzidas no ambiente R versão 4.4.3 (R Core Team, 2023).

Os dados de vazão dos poços foram inicialmente padronizados, sendo realizada a remoção de registros duplicados, com base na identificação única dos poços e nas coordenadas geográficas, além da exclusão de valores inconsistentes ou considerados outliers, como vazões negativas ou extremamente elevadas que destoavam do comportamento geral da amostra.

Nesta fase, procedeu-se à análise espacial dos poços, com o objetivo de verificar a distribuição espacial das vazões e avaliar possíveis padrões geográficos. Também foi realizada a interseção espacial dos pontos com diferentes camadas ambientais auxiliares, como litologia, topografia, uso e cobertura da terra, e variáveis climáticas, como o intuito de relacionar a base

de dados com informações ambientais, fornecendo o conjunto de covariáveis necessário para os modelos preditivos.

### **2.2.3 Covariáveis Ambientais**

Para compreender os fatores que influenciam a produtividade da vazão atual dos poços, foram utilizadas como covariáveis diversas informações ambientais e físico-geográficas. A seleção das variáveis ambientais e físico-geográficas (Tabela 2.1) foi baseada em sua reconhecida influência sobre a ocorrência e a produtividade de aquíferos, conforme evidenciado em estudos prévios de modelagem de potencial hídrico subterrâneo.

**Tabela 2.1** Covariáveis ambientais utilizadas na modelagem preditiva da vazão de poços tubulares no norte de Minas Gerais, com respectivas classes, tipo de dado, fontes, resoluções/escala e processos hidrogeológicos influenciados.

<b>Categoria</b>	<b>Variáveis</b>	<b>Fonte</b>	<b>Resolução</b>	<b>Processo hidrogeológico influenciado</b>	<b>Referência</b>
Climáticas	Precipitação mensal e anual (1985–2022), temperatura média, máxima e mínima (1985–2022), radiação solar Variáveis bioclimáticas (Bio1–Bio19)	WorldClim	1 km	Determina o volume potencial de recarga dos aquíferos, especialmente em eventos concentrados no semiárido; Afeta a evapotranspiração potencial, controlando o balanço hídrico e a infiltração efetiva	Fick e Hijmans (2017) Kumar <i>et al.</i> (2020) Rahmati <i>et al.</i> (2016)
Topográficas	Declividade, orientação, índice topográfico de umidade (TWI), MDE	SRTM	30 m	Regula o escoamento superficial e a capacidade de infiltração; declives suaves favorecem recarga; Influencia a incidência solar, evaporação e retenção de umidade no solo; Representa a tendência de acúmulo de água em função da forma do relevo	Farr <i>et al.</i> (2007) Adiat <i>et al.</i> (2012) Machiwal <i>et al.</i> (2011);
Geológicas	Litologia reclassificada por composição química/mineralógica, geomorfologia, domínio hidrogeológico, sistema aquífero aflorante	CPRM	1:100.000	Controla a porosidade e a permeabilidade do aquífero; influencia diretamente a transmissividade	Silva <i>et al.</i> (2020) Todd e Mays (2005);
Edáficas	Classes de solos	Levantamento de Solos da UFV	1:100.000	Afetam a infiltração, armazenamento de água e percolação para o aquífero	UFV <i>et al.</i> (2010) Jesiya e Gopinath (2020)

Uso e cobertura da terra	NDVI, SAVI, classes de uso do MapBiomas	Landsat 7 / MapBiomas	30 m	Cobertura vegetal aumenta infiltração e reduz escoamento; áreas agrícolas ou urbanas tendem a reduzir recarga e aumentar contaminação	USGS (2023) Machiwal <i>et al.</i> (2011); Aslam <i>et al.</i> (2023)
Geofísicas	K, Th, U (gamaespectrometria)	CPRM	500 m	Associados à mineralogia superficial e à alteração de rochas, indicadores indiretos de condições de permeabilidade	Pinto <i>et al.</i> (2003) Carvalho <i>et al.</i> (2014)
Hidrológicas	Densidade de drenagem, distância a rios, erosividade	ANA / IBGE	1:100.000	Altos valores indicam baixa permeabilidade e maior escoamento superficial; Áreas próximas a rios podem ter maior recarga induzida; zonas distantes podem indicar aquíferos isolados	Cândido <i>et al.</i> (2019) Nag e Ghosh (2013)

A seleção das variáveis ambientais e físico-geográficas (Tabela 2.1) foi fundamentada em sua reconhecida influência sobre processos de recarga, armazenamento e transmissividade dos aquíferos, que controlam diretamente a disponibilidade e a vazão dos poços (RAHMATI *et al.*, 2016). Estudos recentes destacam que a integração de múltiplos fatores condicionantes, incluindo aspectos climáticos, topográficos, geológicos, edáficos, de uso e cobertura da terra, geofísicos e hidrológicos, aumenta a acurácia e a robustez de modelos preditivos de potencial aquífero (ARABAMERI *et al.*, 2020; BAI *et al.*, 2022; MADANI e NIYAZI, 2023).

Os dados após essa análise foram projetados em sistemas de coordenadas *South America Lambert Conformal Conic* (ESRI:102015), juntamente com os pontos de vazão observados para modelagem dos dados.

#### **2.2.4 Classificação Litológica Baseada na Composição Química e Mineralógica**

A variável litológica, integrante da categoria Geológica das covariáveis ambientais (Tabela 2.1), foi reclassificada para uso como variável categórica no modelo preditivo. Essa reclassificação foi baseada na composição química e mineralógica predominante das rochas, ao invés de critérios puramente granulométricos ou texturais, e contou com o apoio de geólogos especialistas na geologia regional. O objetivo foi categorizar os litotipos de forma a refletir suas características geoquímicas dominantes, permitindo uma análise mais aprofundada das relações entre litologia, mineralogia e processos geológicos na região estudada.

O objetivo foi agrupar os litotipos de forma a refletir suas características geoquímicas dominantes, permitindo uma representação mais fiel de seu comportamento hidrogeológico e favorecendo a análise das relações entre litologia, mineralogia e processos de recarga e transmissividade dos aquíferos.

As unidades geológicas originais foram agregadas nas seguintes classes:

- Rochas psefíticas: Incluem sedimentos com granulometria predominantemente grossa, como conglomerados e cascalhos.
- Formações ferríferas: Caracterizam-se por altos teores de ferro, incluindo formações ferríferas bandadas e aglomerados ferruginosos.
- Rochas félsicas: Ricas em elementos como potássio e sódio, incluindo granitoides e sienitos.
- Rochas carbonáticas: Predominantemente compostas por minerais carbonáticos, como calcita e dolomita.

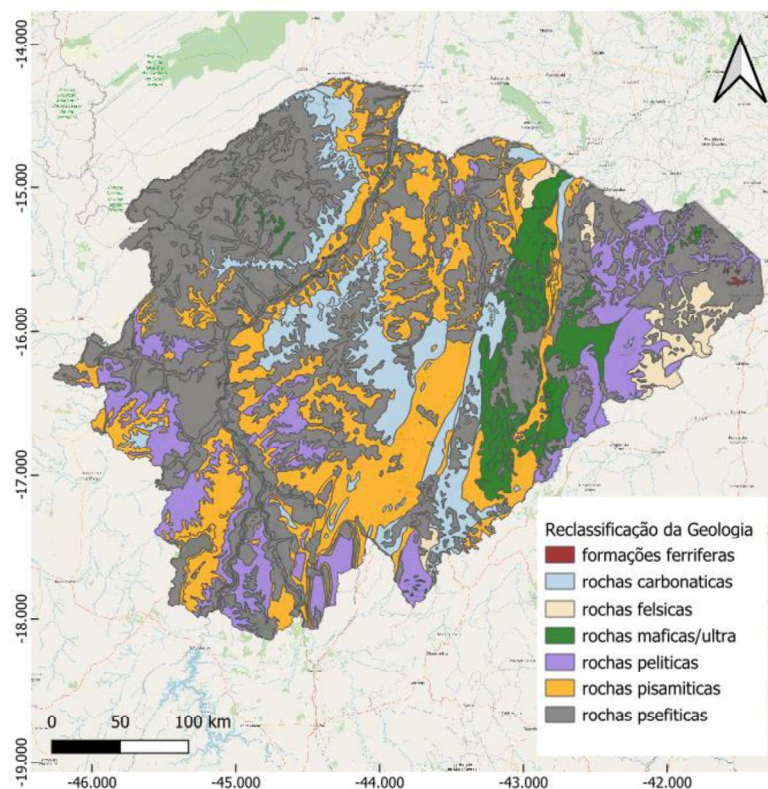
- Rochas pisamíticas: Agrupam rochas sedimentares de granulação grossa, como arcósios e quartzitos.
- Rochas máficas e ultramáficas: Compostas por minerais ricos em cálcio, magnésio e ferro, incluindo basaltos, anfíbolitos e peridotitos.
- Rochas pelíticas: De origem marinha, ricas em alumínio e potássio, incluindo filitos, xistos e siltitos, podendo conter elementos-traço como boro.

#### *2.2.4.1 Procedimentos de Reclassificação*

Os dados geológicos foram inicialmente obtidos a partir de mapas geológicos da região do norte de Minas Gerais. Posteriormente, as litologias foram reclassificadas considerando sua composição química e mineralógica predominante. Essa abordagem permitiu uma categorização mais alinhada aos processos geoquímicos regionais e às características dos depósitos sedimentares e ígneos presentes.

Essa metodologia busca oferecer uma nova perspectiva para a análise das unidades geológicas, permitindo correlacionar melhor a distribuição das rochas com aspectos geoquímicos e metalogenéticos da região.

As informações geológicas foram reclassificadas com base na composição química e mineralógica predominante das formações, agrupando as unidades em classes como rochas máficas, félsicas, pelíticas e carbonáticas. Essa reclassificação permitiu representar com maior fidelidade o comportamento hidrogeológico associado a cada litotipo. Na Figura 2.5 apresenta-se a distribuição espacial das classes litológicas reclassificadas, as quais foram utilizadas como variáveis categóricas nos modelos de predição da vazão dos poços.



**Figura 2.5.** Mapa das classes litológicas reclassificadas com base na composição química e mineralógica das rochas.

### 2.2.5 Seleção e Processamento das Covariáveis

Após a obtenção e organização das variáveis ambientais e físico-geográficas (Tabela 2.2), foi realizado um processo de filtragem e seleção com o objetivo de definir quais covariáveis seriam efetivamente utilizadas na modelagem. Essa etapa buscou reduzir redundâncias, minimizar problemas de multicolinearidade e priorizar atributos com maior relevância para a predição da vazão, conforme recomendado por Rahmati *et al.* (2016), Gomes *et al.* (2019) e Sena *et al.* (2020).

Para alcançar maior eficiência preditiva, a seleção das covariáveis seguiu três etapas principais:

1. Filtragem por variância: covariáveis de baixa variabilidade (próxima a zero) foram eliminadas por meio da função *nearZeroVar* (KUHN, 2008), garantindo que apenas atributos informativos fossem mantidos (HUJOEL *et al.*, 2018; KERN *et al.*, 2017), visto que elas não melhoram a qualidade da predição somente elevando o peso computacional do treinamento;

2. Remoção de variáveis altamente correlacionadas: utilizou-se a correlação de *Spearman* para identificar e remover redundâncias, seguindo metodologia adotada por Coelho *et al.* (2019), Gomes *et al.* (2019) e Souza *et al.* (2022), sendo utilizado para remover covariáveis de alta correlação que pode dominar a predição dos modelos;

3. Seleção baseada na importância preditiva: cada modelo avaliou a relevância das variáveis na amostra de treinamento para a previsão da vazão dos poços tubulares, garantindo um conjunto de covariáveis otimizado e estatisticamente significativo para predição dessa amostra para algoritmo avaliado. Esse processo é realizado partir da *recursive feature elimination* (RFE) do pacote “*caret*” (citação). O RFE trabalha calculado a importância do conjunto completo de covariáveis, posteriormente com essa informação da importância ele calcular os valores de performance dos algoritmos avaliados com grupos “*subset*” de covariáveis definidos pelo operador. Neste trabalho foram utilizados, utilizados, além do conjunto total de covariáveis que terminaram a seleção por correlação valores de 5 a 20, 25,30,40 e 50.

A abordagem adotada buscou assegurar a robustez da modelagem, minimizar o risco de *overfitting* e promover um equilíbrio entre precisão preditiva e eficiência computacional, possibilitando a construção de mapas confiáveis para estimar o potencial aquífero em diferentes regiões.

### 2.2.6 Modelagem da Vazão com Técnicas de Aprendizado de Máquina

A predição da vazão de poços tubulares em ambientes hidrogeologicamente complexos exige métodos capazes de lidar com relações não lineares, interações multivariadas e transições abruptas no espaço (RAHMATI *et al.*, 2016; JAAFARZADEH *et al.*, 2021). Ao contrário de abordagens convencionais, como a geoestatística, que assumem continuidade espacial e dependência baseada unicamente na distância, técnicas de aprendizado de máquina têm demonstrado maior desempenho em regiões de forte heterogeneidade geológica, como o norte de Minas Gerais, onde interfaces entre aquíferos fraturados e compactos resultam em variações bruscas de produtividade (GHEISARI *et al.*, 2023).

Com base na literatura e na diversidade de famílias algorítmicas, foram selecionados quatro modelos supervisionados:

- Regressão Linear Múltipla (LM) – amplamente utilizada como modelo de referência (baseline) em estudos de potencial aquífero, permitindo avaliar o ganho obtido com modelos não lineares (DARGAHI *et al.*, 2022);

- Máquina de Vetores de Suporte (SVM) – eficaz para capturar relações não lineares e trabalhar com conjuntos de dados multivariados e de alta dimensionalidade (ELKIRAN, NOURANI e ABBA 2019; RIAZ *et al.*, 2024);
- Kernel k-Nearest Neighbors (KKNN) – abordagem baseada em instâncias, sensível à proximidade espacial e capaz de modelar padrões locais de variação, com ponderação por função kernel (SCHLIEP, 2025);
- Random Forest (RF) – algoritmo baseado em árvores de decisão, robusto a multicolinearidade e capaz de estimar a importância das variáveis (ARABAMERI *et al.*, 2020; BAI *et al.*, 2022).

Cada algoritmo foi treinado com 75% das amostras (conjunto de treinamento) e validado com os 25% restantes (conjunto de teste) por *hold-out testing*. A seleção do subconjunto ótimo de covariáveis foi realizada por meio de validação cruzada (10 folds, 1 repetição), utilizando como métrica de escolha o Coeficiente de Correlação de Concordância de Lin (LCCC).

Após a seleção, os modelos finais foram otimizados via validação cruzada repetida (10 folds, 3 repetições), ajustando hiperparâmetros específicos de cada algoritmo. Esse procedimento foi repetido 100 vezes para cada modelo, com diferentes divisões treino/teste, permitindo avaliar a robustez das previsões.

#### 2.2.6.1 Métricas de desempenho

A avaliação comparativa entre modelos foi baseada nas seguintes métricas:

- Coeficiente de Correlação de Concordância de Lin (LCCC) – avalia simultaneamente precisão e exatidão, sendo considerado alto desempenho quando  $LCCC > 0,90$  (LIN, 1989);
- Coeficiente de Determinação ( $R^2$ ) – indica a proporção da variabilidade explicada pelo modelo; valores acima de 0,80 são geralmente interpretados como muito bons em estudos hidrogeológicos (MORIASI *et al.*, 2007);
- Erro Quadrático Médio (RMSE) – mede a magnitude média dos erros, com valores menores indicando melhor ajuste;
- Erro Médio Absoluto (MAE) – semelhante ao RMSE, porém menos sensível a valores extremos.

Essas métricas foram aplicadas a todos os modelos e comparadas para identificar o algoritmo mais eficiente na previsão da vazão dos poços tubulares na região de estudo.

### 2.2.7 Geração do Mapa de Potencial Aquífero

Com base no algoritmo de melhor desempenho preditivo, foi gerado um mapa contínuo de potencial aquífero na área de estudo, representando a distribuição espacial da vazão estimada. Para isso, os valores preditos foram extrapolados para toda a área de estudo, considerando a combinação espacial das covariáveis selecionadas.

Para representar a robustez das predições, os resultados das 100 rodadas do modelo foram sintetizados em dois produtos cartográficos:

- Mapa da Média da Vazão Preditiva: representa o valor médio da vazão estimada em cada pixel, refletindo o potencial aquífero regional conforme modelado pelo algoritmo.
- Mapa do Coeficiente de Variação (CV): expressa a variabilidade relativa entre as 100 simulações em cada pixel, evidenciando as áreas com maior incerteza nas predições. Valores elevados de CV indicam regiões onde o modelo apresentou maior instabilidade, o que pode estar associado a lacunas nos dados, variabilidade ambiental ou menor densidade de amostras.

Essa abordagem permite não apenas visualizar o potencial aquífero estimado, mas também avaliar a confiabilidade espacial das predições realizadas, contribuindo para uma análise mais crítica da aplicabilidade do modelo em diferentes zonas da área de estudo.

## 2.3 Resultados e Discussão

### 2.3.1 Monitoramento quantitativo da vazão dos poços

A Tabela 2.2 apresenta a estatística descritiva da vazão dos poços tubulares antes e após a remoção de outliers. A análise foi realizada com base em 3.370 registros válidos, resultantes da exclusão de valores extremos (8,9% de dados excluídos) segundo o critério do intervalo interquartil (IQR).

Após o processamento, a distribuição das vazões revelou comportamento assimétrico, porém menos acentuado. A mediana foi de 6,98 m<sup>3</sup>/h, enquanto os quartis 10% e 90% situaram-se em 1,15 m<sup>3</sup>/h e 18,72 m<sup>3</sup>/h, respectivamente, indicando que a maior parte dos poços apresenta vazões relativamente baixas, com alguns valores ainda representando condições mais favoráveis.

A média da amostra foi de 8,93 m<sup>3</sup>/h, com desvio padrão de 7,10 m<sup>3</sup>/h, evidenciando uma redução na dispersão em comparação com o conjunto original. A assimetria (skewness =

1,12) e a curtose (0,72) indicam uma distribuição moderadamente enviesada e próxima da normalidade, com menor influência de valores extremos.

É importante destacar que a exclusão dos outliers não implica desconsiderar a existência de poços com alta produtividade na região. Esses valores podem estar associados a condições hidrogeológicas excepcionais, como a presença de aquíferos cársticos ou poços com características construtivas diferenciadas. No entanto, do ponto de vista estatístico e preditivo, a permanência desses extremos compromete a calibragem dos modelos, resultando em superestimações em áreas onde tais condições não ocorrem.

A remoção dos outliers, portanto, foi adotada como medida de controle para garantir maior robustez estatística e representatividade dos modelos de aprendizado de máquina, evitando que casos pontuais dominassem a inferência espacial da vazão.

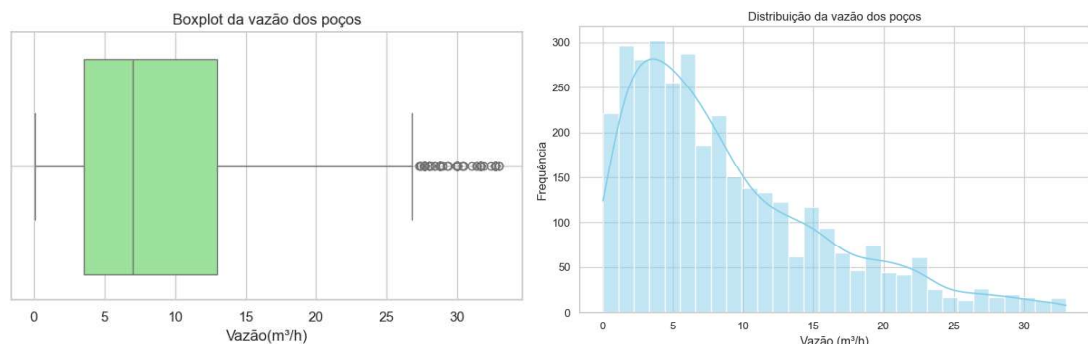
**Tabela 2.2.** Estatística descritivas da vazão antes e após remoção de outliers

<b>Estatísticas</b>	<b>Antes da remoção de outliers</b>	<b>Após a remoção de outliers</b>
n°	3700	3370
Média	15,11	8,93
Std	26,79	7,10
Mínimo	0,04	0,04
25%	3,82	3,53
50%	7,92	6,98
75%	15,55	13,00
máximo	329,04	32,97
skewness	5,64	1,12
kurtosis	41,82	0,72

A distribuição das vazões observadas nos poços tubulares (mediana  $\approx 6,98$  m<sup>3</sup>/h; quartis 10% = 1,15 m<sup>3</sup>/h e 90% = 18,72 m<sup>3</sup>/h) revela padrão típico de aquíferos fraturados em regiões semiáridas, marcados pela predominância de poços com baixa produtividade e forte assimetria na distribuição dos dados. Essa configuração é coerente com resultados reportados por Bianchi *et al.* (2023), que analisaram aquíferos cristalinos na África e observaram medianas de aproximadamente 0,5 L/s ( $\approx 1,8$  m<sup>3</sup>/h), com a maioria dos poços registrando rendimentos entre 2 e 20 L/s ( $\approx 7,2 \approx 72$  m<sup>3</sup>/h). Valores similares também são encontrados em aquíferos sedimentares costeiros, nos quais zonas de baixa produtividade (“*poor yield*”) apresentam

vazões inferiores a 6 L/s, enquanto zonas médias (“*medium*”) situam-se entre 6,1 e 10 L/s, e zonas mais produtivas (“*good*”) ultrapassam os 11 L/s (NJOCK *et al.*, 2025). Tais evidências corroboram a adequação da distribuição observada neste estudo, situando a amostra dentro de um intervalo esperado para formações hidrogeológicas de baixa a média produtividade.

A Figura 2.6a (boxplot) evidencia a acentuada assimetria positiva, destacando a alta concentração de poços com vazões reduzidas e a presença de uma cauda superior longa, indicativa de valores extremos. Já o histograma com curva de densidade (Figura 2.6b) reforça essa tendência, mostrando que a maior parte dos poços apresenta vazão inferior a 10 m<sup>3</sup>/h, um comportamento frequentemente associado a aquíferos fraturados ou com baixa permeabilidade, como os encontrados em diversas áreas do semiárido do norte de Minas Gerais.



**Figura 2.6.** Distribuição das vazões de poços tubulares no norte de Minas Gerais

(a) Boxplot da vazão específica dos poços; (b) Histograma com curva de densidade das vazões específicas.

Essa distribuição altamente concentrada em vazões reduzidas sugere um cenário de baixa disponibilidade hídrica subterrânea, coerente com a complexidade litológica e com a natureza descontínua dos sistemas aquíferos regionais (MACDONALD *et al.*, 2005; CHILTON e FOSTER, 1995). A heterogeneidade estrutural das rochas, combinada com a baixa capacidade de armazenamento e transmissão, limita a produtividade dos poços e gera grande variabilidade entre locais relativamente próximos (WRIGHT, 1992; BIANCHI *et al.*, 2023).

### 2.3.2 Desempenho preditivo dos Modelos de *Machine Learning*

A Tabela 2.3 apresenta os resultados obtidos pelos quatro algoritmos avaliados, Regressão Linear Múltipla (LM), Máquina de Vetores de Suporte (SVM), *Random Forest* (RF) e *Kernel k-Nearest Neighbors* (KKNN), na predição da vazão específica de poços tubulares. O modelo de Regressão Linear Múltipla (LM) apresentou o menor desempenho entre os

algoritmos avaliados. No conjunto de teste, obteve  $R^2 = 0,462$ ,  $RMSE = 0,247$ ,  $MAE = 0,169$  e  $LCCC = 0,681$ , indicando limitação em capturar as complexas relações não-lineares entre as variáveis ambientais e a vazão dos poços. O modelo SVM mostrou desempenho superior ao LM, com  $R^2 = 0,598$  e  $LCCC = 0,754$ , destacando-se pelo menor desvio padrão do  $R^2$  (0,019) e  $LCCC$  (0,017) entre os modelos baseados em margens, o que indica relativa robustez à variabilidade dos dados.

O modelo *Random Forest* (RF) apresentou resultados satisfatórios, com  $R^2 = 0,639$ ,  $RMSE = 0,205$ ,  $MAE = 0,141$  e  $LCCC = 0,792$ . A boa capacidade de modelagem não linear e de interação entre variáveis o posicionou como uma alternativa competitiva, embora não tenha alcançado o melhor desempenho. Já o modelo KKNN destacou-se como o mais eficaz em todas as métricas analisadas. No conjunto de teste, obteve o maior coeficiente de determinação ( $R^2 = 0,751$ ), o menor  $RMSE$  (0,150) e o menor  $MAE$  (0,094), além de apresentar  $LCCC = 0,862$ , evidenciando alta concordância entre valores observados e preditos. O modelo também manteve baixo coeficiente de variação nas 100 rodadas, indicando elevada estabilidade preditiva.

**Tabela 2.3.** Desempenho de modelos para prever a distribuição espacial dos valores de vazão de poços tubulares. As métricas referem-se à média após 100 previsões.

a- treino										
	Modelo	$R^2$	$R^2$	$R^2$	RMSE	RMSE	RMSE	MAE	MAE	MAE
			SD	CV						
			(%)		(%)		(%)		(%)	
Treino	LM	0,481	0,020	4,2	0,244	0,009	3,7	0,166	0,007	4,2
	SVM	0,614	0,015	2,4	0,211	0,008	3,8	0,142	0,006	4,2
	RF	0,652	0,014	2,1	0,199	0,007	3,5	0,138	0,006	4,3
	KKNN	<b>0,769</b>	<b>0,010</b>	<b>1,3</b>	<b>0,142</b>	<b>0,006</b>	<b>4,2</b>	<b>0,091</b>	<b>0,004</b>	<b>4,4</b>
b- teste										
Teste	LM	0,462	0,025	5,4	0,247	0,013	5,3	0,169	0,010	5,9
	SVM	0,598	0,019	3,2	0,216	0,010	4,6	0,145	0,008	5,5
	RF	0,639	0,021	3,3	0,205	0,009	4,4	0,141	0,007	5,0
	KKNN	<b>0,751</b>	<b>0,009</b>	<b>1,2</b>	<b>0,150</b>	<b>0,007</b>	<b>4,7</b>	<b>0,094</b>	<b>0,005</b>	<b>5,3</b>

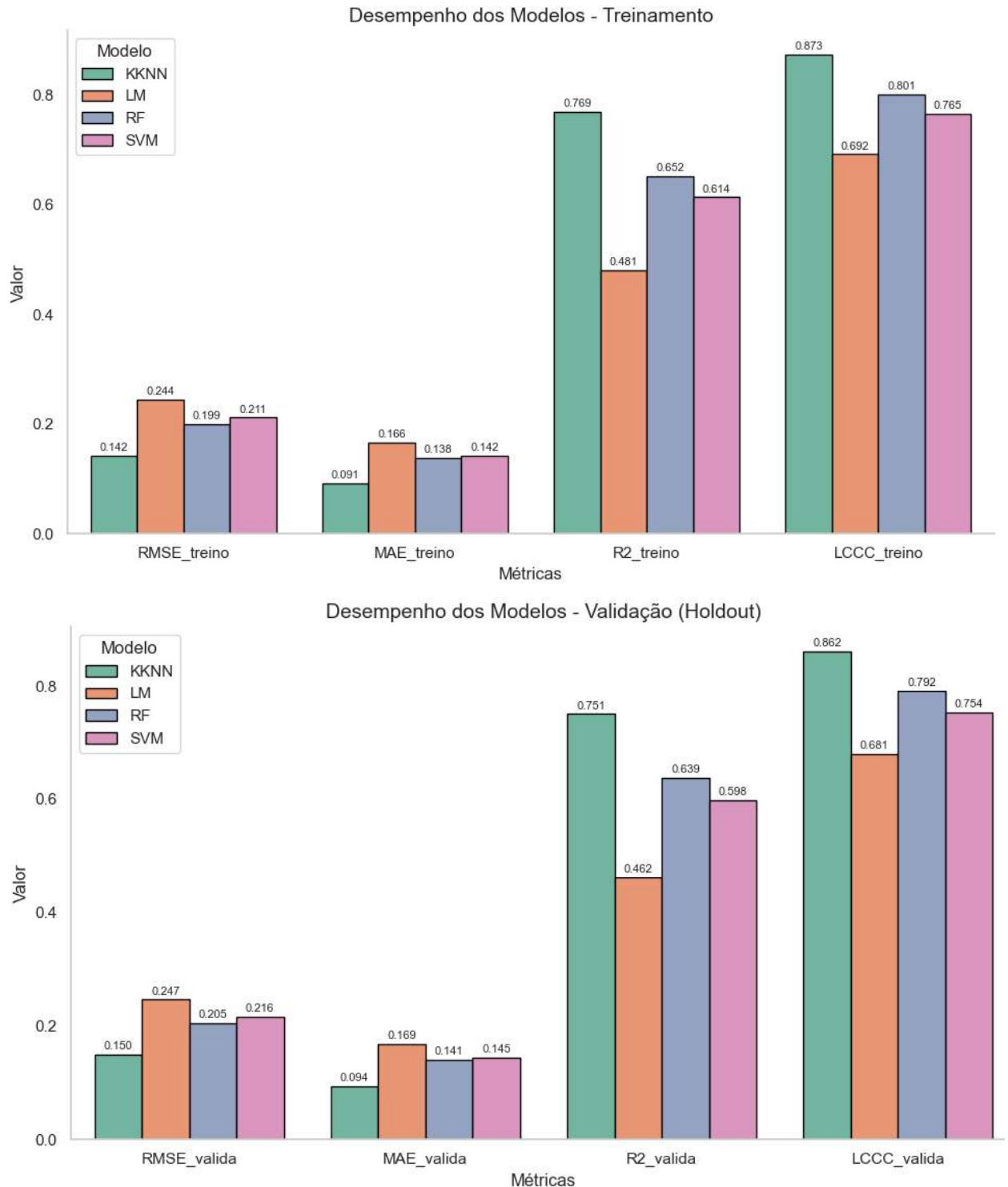
Modelos: Modelo de Regressão Linear Múltipla (LM), Máquina de Vetores de Suporte (SVM), *Kernel k-Nearest Neighbors* (KKNN) e *Random Forest* (RF). Métricas: Métrica R-quadrado ( $R^2$ ), erro quadrático médio RMSE, LCCC. SD – desvio padrão, CV – coeficiente de variação.

Ao comparar os modelos, observa-se que o KKNN é superior aos demais algoritmos, especialmente em termos de precisão preditiva e capacidade de generalização. A diferença entre o MAE do KKNN e do RF (0,094 vs. 0,141) demonstra maior capacidade do KKNN em capturar nuances locais da variabilidade da vazão, o que é coerente com a natureza espacialmente heterogênea do sistema aquífero estudado.

Embora o RF apresente boa performance, seu desempenho foi inferior ao do KKNN, sugerindo que abordagens baseadas em instâncias (como o KKNN), por considerarem a vizinhança local, são mais sensíveis a variações espaciais e padrões regionais complexos de controle da vazão.

A superioridade do KKNN também se evidencia na validação cruzada, com  $R^2 = 0,7685$  e  $RMSE = 0,1421$ , confirmando sua eficiência e consistência na modelagem da distribuição espacial da vazão em cenários hidrogeológicos heterogêneos.

A Figura 2.7 apresenta a comparação gráfica entre os modelos de *machine learning* quanto ao seu desempenho preditivo nas métricas  $R^2$ , RMSE, MAE e LCCC, calculadas com base na média das 100 rodadas de validação. Observa-se que o modelo KKNN se destaca visualmente com os melhores resultados em todas as métricas, evidenciando seu equilíbrio entre precisão e robustez. O gráfico também permite identificar com clareza a diferença de desempenho entre os algoritmos, facilitando a interpretação das vantagens relativas de cada abordagem e reforçando a superioridade do KKNN na modelagem da vazão específica dos poços tubulares.



**Figura 2.7.** Desempenho dos modelos de aprendizado de máquina – treinamento e validação

Embora o KKNN ainda seja pouco explorado em estudos hidrogeológicos, a literatura tem apontado o potencial de modelos baseados em KNN, especialmente quando combinados com outros algoritmos, na predição de variáveis hidrológicas e hidrogeológicas. Por exemplo, Kombo *et al.* (2020) aplicaram um modelo híbrido KNN-RF (*K-Nearest Neighbors* com *Random Forest*) para prever variações sazonais do nível freático em aquíferos fraturados com

dados escassos na África Oriental. Os resultados indicaram melhor capacidade de generalização e maior precisão preditiva, especialmente quando integrados com dados hidrometeorológicos.

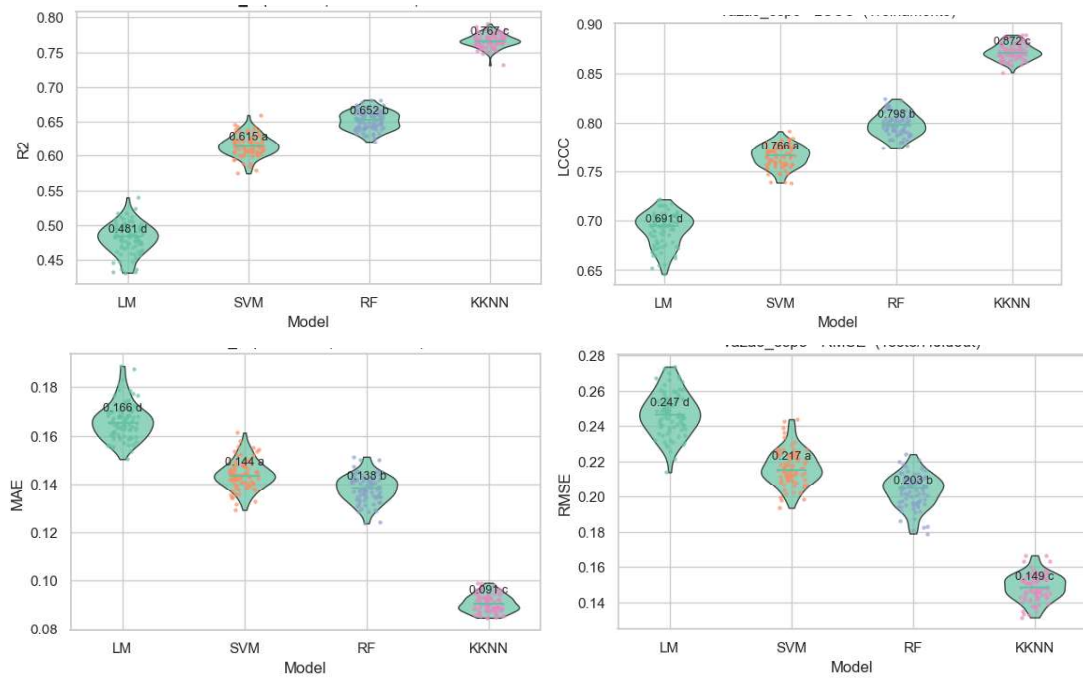
Recentemente, Toure *et al.* (2024) utilizaram o mesmo algoritmo KNN-RF para modelar níveis freáticos sazonais em aquíferos fraturados na África do Sul, novamente destacando robustez e acurácia em cenários hidrogeológicos complexos.

Além disso, diversos estudos em sondagens e mapeamentos de vazão utilizam o *K-Nearest Neighbors* (KNN) como uma das abordagens comparativas, embora não com kernel ponderado, e enfatizam sua simplicidade e adaptabilidade, mesmo que modelos como *Random Forest* ou *Gradient Boosting* frequentemente apresentem melhor desempenho (MASKOONI *et al.*, 2020).

Esses exemplos indicam que, embora o KNN tenha sido explorado, a aplicação do KKNN na predição da vazão específica de poços surge como uma abordagem inovadora. O uso de pesos via kernel confere maior sensibilidade à estrutura espacial e às relações não-lineares entre as covariáveis, refletindo-se nos excelentes resultados observados ( $R^2 \approx 0,75$ ; RMSE  $\approx 0,15$ ; MAE  $\approx 0,09$ ; LCCC  $\approx 0,87$ ) em seu estudo, resultados que superam os de modelos tradicionais como RF, SVM e LM.

A análise estatística por meio do teste de *Kruskal-Wallis* reforçou essa constatação, indicando diferenças significativas ( $p < 0,05$ ) entre os modelos para todas as métricas avaliadas. As letras distintas atribuídas nos gráficos tipo *violin plot* (Figura 2.8) indicam agrupamentos estatisticamente distintos, evidenciando a superioridade do KKNN frente aos demais algoritmos. O coeficiente de concordância de Lin (LCCC), por sua vez, reforça a maior aderência das previsões do KKNN em relação aos valores observados, à frente dos modelos RF (0,801), SVM (0,765) e LM (0,692).

Esses achados estão em consonância com Coelho *et al.* (2019) e Souza *et al.* (2022), que destacam a eficácia de algoritmos não paramétricos na modelagem de sistemas hidrogeológicos complexos. Contudo, destaca-se que a aplicação do *Kernel k-Nearest Neighbors* (KKNN) para predição de vazão específica ainda é incipiente na literatura. Embora estudos anteriores tenham explorado o KNN clássico em combinação com outros métodos, como o modelo híbrido KNN-RF utilizado por Kombo *et al.* (2020) e Toure *et al.* (2024) para prever níveis freáticos, o uso do KKNN com ponderação por kernel representa uma abordagem inovadora. Esse tipo de modelagem incorpora informações espaciais com maior sensibilidade, resultando em ganhos expressivos de acurácia e robustez em cenários marcados por escassez de dados e elevada variabilidade ambiental.



**Figura 2.8.** Comparação do desempenho dos modelos de aprendizado de máquina. Violin plots das métricas de desempenho com teste de Kruskal-Wallis e agrupamentos por diferença estatística ( $p < 0,05$ )

A superioridade do modelo KKNN observado nos resultados deste estudo reflete uma contribuição inovadora na modelagem preditiva de vazões de poços tubulares em ambientes hidrogeológicos complexos. Embora o uso de variantes ponderadas do KNN, como o KKNN, ainda seja raro na literatura de águas subterrâneas, sua eficácia já foi demonstrada em outras áreas de sensoriamento ambiental. Guevara *et al.* (2019), por exemplo, aplicaram o KKNN para fazer *downscaling* de umidade do solo a partir de satélite, obtendo previsões mais precisas e estáveis em comparação ao KNN clássico. Esse desempenho superior pode ser atribuído à capacidade do KKNN de aplicar pesos diferenciados aos vizinhos mais próximos por meio de funções kernel, o que o torna particularmente sensível a padrões espaciais complexos, uma vantagem crucial em sistemas aquíferos descontínuos como os do semiárido mineiro.

Embora o uso de variantes kernelizadas como o KKNN ainda seja raro em estudos hidrogeológicos, o algoritmo KNN clássico tem ganhado espaço em diferentes aplicações ligadas às águas subterrâneas. Por exemplo, Kombo *et al.* (2020) aplicaram uma abordagem híbrida KNN-RF para prever variações sazonais do nível freático em ambientes semiáridos de Ruanda, reportando baixo RMSE e elevado coeficiente de determinação. Isso reforça a robustez do KNN em contextos com dados escassos e alta variabilidade climática, características que também definem o semiárido brasileiro. De forma semelhante, Sahour *et al.* (2023) utilizaram o KNN, entre outros algoritmos, para mapear a qualidade da água subterrânea no norte do Irã.

Os autores observaram desempenho competitivo do KNN frente ao RF e ao SVM, com destaque para métricas como precisão e F1-score. Além disso, estudos que avaliaram a salinidade e potabilidade de aquíferos em regiões iranianas identificaram que o KNN apresentou maior estabilidade e desempenho em áreas marcadas por heterogeneidade geológica, superando modelos paramétricos como regressão linear e SVM. Essas evidências reforçam o potencial de abordagens baseadas em instâncias, especialmente em contextos com elevada complexidade litológica e escassez de dados, características recorrentes em regiões semiáridas (ZHANG *et al.*, 2018; HE *et al.*, 2021). Assim, a adoção do KKNN neste estudo amplia os avanços obtidos com o KNN tradicional, ao incorporar uma sensibilidade aprimorada à estrutura espacial dos dados por meio de funções kernel. Os bons resultados obtidos, com destaque para o alto  $R^2$  e baixo erro médio, confirmam sua robustez e apontam o KKNN como uma alternativa promissora para a modelagem do potencial aquífero em cenários hidrogeológicos desafiadores.

#### *2.3.2.1 Comparação com estudos anteriores e contribuições do modelo*

Os resultados obtidos neste estudo demonstram avanços significativos em relação a trabalhos anteriores que buscaram estimar a produtividade de poços ou modelar o potencial aquífero em regiões semiáridas. Os modelos baseados em aprendizado de máquina, como o k-Nearest Neighbors ponderado (KKNN) e o Random Forest (RF), apresentaram desempenho superior às abordagens tradicionais comumente aplicadas, como a regressão linear múltipla (LM) e técnicas geoestatísticas clássicas.

Estudos prévios realizados em regiões com características hidrogeológicas semelhantes, como no norte de Minas Gerais, interior do Nordeste ou em partes semiáridas da Índia e Irã, frequentemente utilizaram interpoladores determinísticos (IDW, Krigagem) ou regressões lineares para modelagem do potencial hídrico subterrâneo. Nessas abordagens, os valores de  $R^2$  raramente ultrapassam 0,20 e os erros absolutos médios (MAE) costumam exceder 0,5 L/s/m, o que limita a capacidade de representar adequadamente a heterogeneidade espacial dos aquíferos.

Em comparação, o presente estudo alcançou valores de  $R^2$  de até 0,76 e MAE inferior a 0,09 com o modelo KKNN, além de coeficientes de correlação concordantes (LCCC) superiores a 0,87, o que representa um ganho expressivo de desempenho. Tais resultados se devem, principalmente:

- à incorporação de múltiplas covariáveis ambientais (geologia, geomorfologia, clima, relevo, solos),

- ao uso de modelos não lineares capazes de capturar interações complexas, e
- à validação robusta com 100 partições independentes, o que garante confiabilidade na generalização.

Além disso, enquanto métodos geoestatísticos exigem alta densidade de dados para gerar mapas confiáveis, os modelos de machine learning aplicados aqui se mostraram eficazes mesmo em cenários com dados esparsos, desde que alimentados com variáveis preditoras ambientais de boa resolução espacial.

A Tabela 2.4 compara os resultados obtidos neste estudo com os de pesquisas anteriores realizadas em contextos hidroclimáticos semelhantes. Observa-se que o modelo KKNN apresentou um dos menores valores de erro absoluto médio, com desempenho competitivo em termos de  $R^2$  (0,76), superando inclusive abordagens amplamente difundidas como regressão linear e interpolação. Essa superioridade está associada à capacidade do algoritmo de lidar com dados complexos e não lineares, além do uso integrado de múltiplas covariáveis ambientais.

**Tabela 2.4.** Comparação com estudos anteriores

<b>Estudo</b>	<b>Região</b>	<b>Método</b>	<b>R<sup>2</sup></b>	<b>MAE</b>	<b>Observações</b>
Arabameri <i>et al.</i> (2021)	Irã (semiárido)	Regressão linear múltipla	0.12	0.48	Uso de variáveis topográficas e geológicas
Razavi-Termeh <i>et al.</i> (2019)	Irã	Random Forest	0.27	0.42	Considerou 15 variáveis ambientais
Singha <i>et al.</i> (2021)	Índia (semiárido)	SVM	0.18	0.44	Incluiu precipitação, NDVI e declividade
Nogueira <i>et al.</i> (2019)	norte de Minas (Brasil)	IDW + , , Krigagem			Foco em interpolação de dados observados
Este estudo	norte de Minas	KKNN	0.75	0.094	Menor MAE e maior R <sup>2</sup> , com robustez em 100 rodadas de validação
Este estudo	norte de Minas	RF	0.64	0.141	Bom desempenho e alta capacidade de generalização

Todos os estudos apresentados na Tabela 2.4 têm como objetivo central a modelagem ou estimativa da produtividade aquífera, a partir de variáveis ambientais ou geoespaciais. Embora apresentem abordagens distintas, esses trabalhos compartilham contextos climáticos e hidrogeológicos semelhantes ao do presente estudo, majoritariamente regiões semiáridas, com elevada dependência da água subterrânea. Arabameri *et al.* (2021), Razavi-Termeh *et al.* (2019) e Singha *et al.* (2021) empregaram algoritmos preditivos como regressão linear múltipla, Random Forest e SVM para estimar a vazão de poços tubulares, com base em variáveis como declividade, geologia, precipitação e NDVI. Já o estudo de Nogueira *et al.* (2019), realizado também no norte de Minas Gerais, aplicou métodos geoestatísticos (IDW e Krigagem) para interpolar valores observados de vazão, sem uso de aprendizado de máquina.

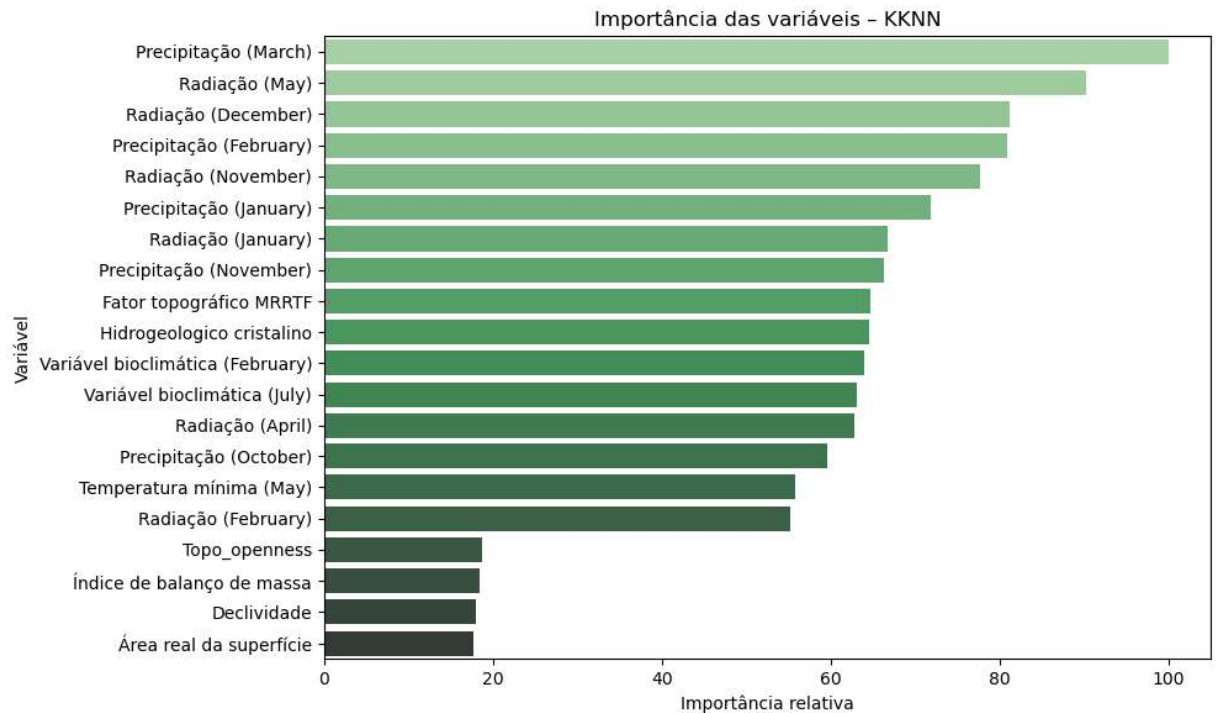
Dessa forma, embora as metodologias e variáveis utilizadas variem entre os trabalhos, todos compartilham o foco na estimativa da vazão de poços em contextos semiáridos, sendo, portanto, comparáveis em termos de objetivo e escala. O presente estudo se diferencia ao empregar modelos de machine learning com validação robusta (100 rodadas) e um conjunto

abrangente de covariáveis, o que resultou em desempenho preditivo superior, especialmente no modelo Kernel KNN (KKNN), que apresentou os menores erros (MAE) e maior coeficiente de determinação ( $R^2$ ) entre os métodos analisados.

Esses resultados evidenciam o potencial dos algoritmos baseados em aprendizado de máquina para superar limitações de métodos tradicionais amplamente utilizados em estudos hidrogeológicos, sobretudo em regiões semiáridas com escassez de dados. O uso do modelo KKNN, aliado à incorporação de variáveis ambientais multiescalares e à validação robusta, não apenas permitiu mapear o potencial hídrico com maior precisão, como também gerou produtos aplicáveis à gestão territorial e à tomada de decisão em contextos de vulnerabilidade hídrica. Assim, este estudo avança a fronteira do conhecimento ao propor uma abordagem confiável, replicável e adaptável a outras regiões com desafios similares, fortalecendo a integração entre ciência de dados e gestão sustentável dos recursos subterrâneos.

#### **2.3.4 Importância relativa das covariáveis ambientais**

A análise da importância das variáveis no modelo KKNN permitiu identificar padrões consistentes e coerentes com o comportamento hidrogeológico da região, reforçando sua capacidade de representar a complexidade física da dinâmica de fluxo subterrâneo no norte de Minas Gerais (Figura 2.9). As covariáveis utilizadas no treinamento incluíram atributos relacionados à litologia, classes de solos, geomorfologia, variáveis topográficas derivadas do MDE, além de fatores climáticos e ambientais.



**Figura 2.9.** Importância das variáveis para o modelo KKNN

A análise das variáveis mais influentes no modelo KKNN revela que os fatores climáticos sazonais, como a precipitação de março e a radiação solar em diversos meses chuvosos, exercem papel central na modelagem da vazão específica dos poços. Este padrão reflete a dinâmica típica de recarga em regiões semiáridas, onde eventos pluviométricos intensos e concentrados determinam a maior parte do aporte hídrico subterrâneo (BRIAN *et al.*, 2016).

A influência da radiação solar e da evapotranspiração também se destaca, pois afetam diretamente a disponibilidade hídrica efetiva, reforçando que pequenas mudanças climáticas podem provocar grandes impactos no balanço hídrico regional (FATICHI *et al.*, 2021). A relevância da precipitação de março, em particular, está associada ao fato de este ser um dos últimos meses do período chuvoso, quando o solo já se encontra saturado, favorecendo uma recarga mais efetiva dos aquíferos antes do início do período seco.

Estudos realizados em diferentes regiões de Minas Gerais corroboram essa interpretação. Por exemplo, Silva *et al.* (2012), ao estimar a recarga direta no Alto Rio Grande, observaram que os maiores valores ocorrem no final da estação úmida, especialmente em áreas com cobertura florestal e solos com maior capacidade de infiltração. De forma semelhante, Souza *et al.* (2019), ao analisarem a bacia do rio Doce, identificaram que aproximadamente 32% da precipitação anual contribui para a recarga subterrânea, com forte dependência da distribuição sazonal das chuvas. No campus da UFMG em Belo Horizonte, Ribeiro *et al.* (2014)

destacaram a rápida resposta dos níveis freáticos aos eventos de chuva, confirmando a relação direta entre precipitação e recarga em aquíferos granular e fissural.

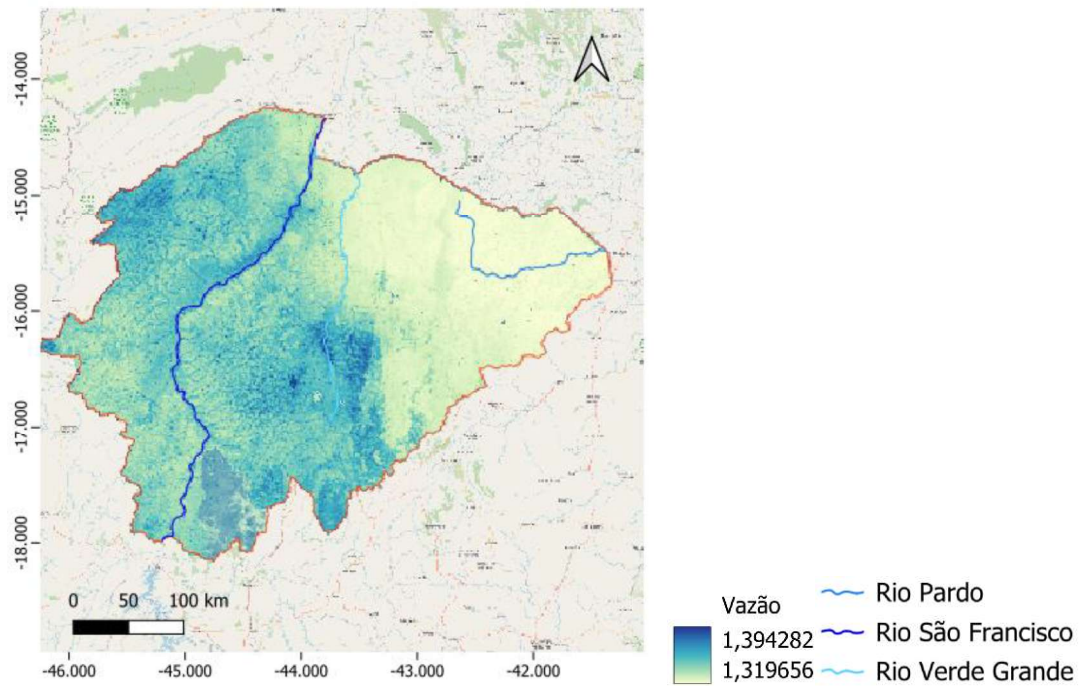
A influência da radiação solar e da evapotranspiração também se destaca entre as variáveis do modelo, pois afetam diretamente a disponibilidade hídrica efetiva, reforçando que pequenas mudanças climáticas podem provocar grandes impactos no balanço hídrico regional. A presença recorrente de variáveis de radiação (nos meses de maio, dezembro, novembro e janeiro), combinadas com índices bioclimáticos (como os de fevereiro e julho), indica que o KKNN foi capaz de captar não apenas os efeitos diretos da chuva, mas também os efeitos indiretos relacionados à energia disponível para evapotranspiração e estresse térmico.

Além disso, variáveis topográficas e geológicas, como o fator topográfico MRRTF, a estrutura cristalina hidrogeológica e a declividade, destacam a importância da morfologia e da litologia para garantir caminhos e armazenamento eficientes de água subterrânea. Isso corrobora estudos como o de Tenenwurcel (2022), na bacia do rio Bicudo, que demonstram que áreas com relevo suave, vegetação preservada e solos estruturados favorecem zonas com maior potencial de recarga.

Outros atributos, como temperatura mínima (maio), índice de balanço de massa, área real da superfície e curvatura topográfica (*topo\_openness*), completam a lista das variáveis mais relevantes, reforçando o caráter multifatorial do controle da vazão. Esses resultados evidenciam que o KKNN é capaz de integrar múltiplas dimensões ambientais, com ênfase na sazonalidade climática, na estrutura do relevo e na composição geológica, para modelar com precisão a complexa dinâmica hidrogeológica do semiárido mineiro.

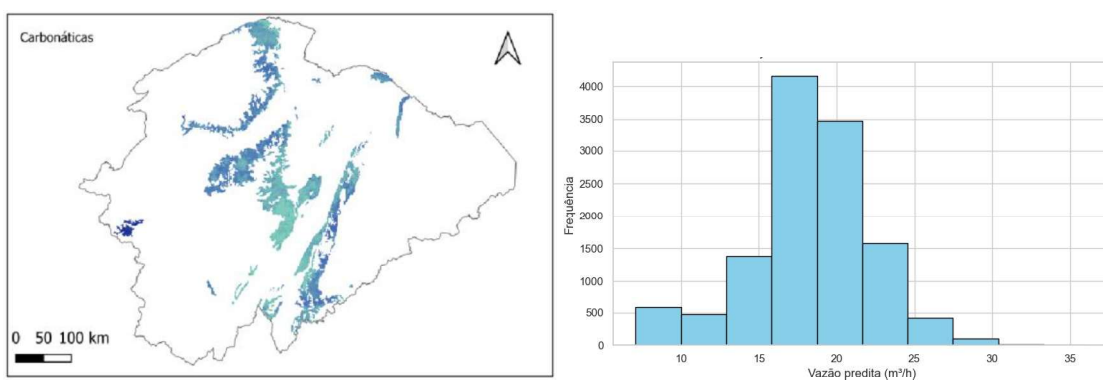
### **2.3.5 Mapeamento do potencial aquífero regional**

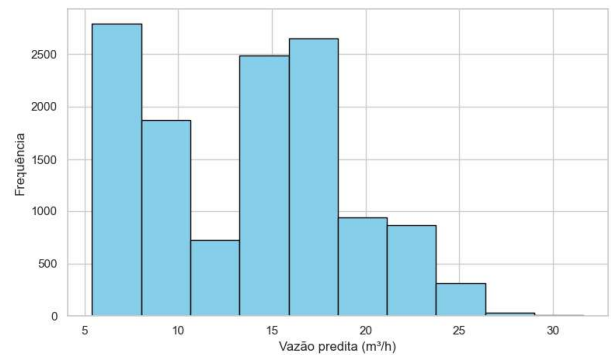
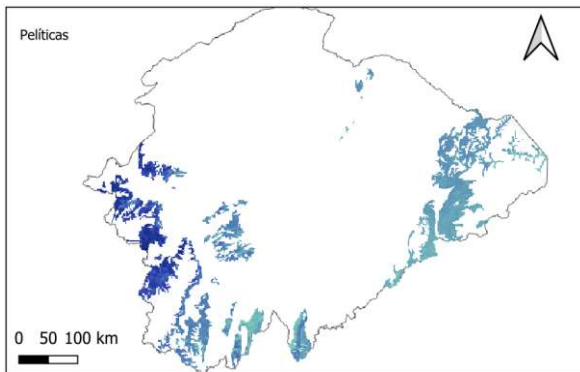
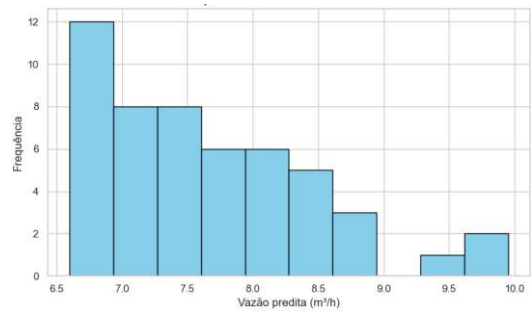
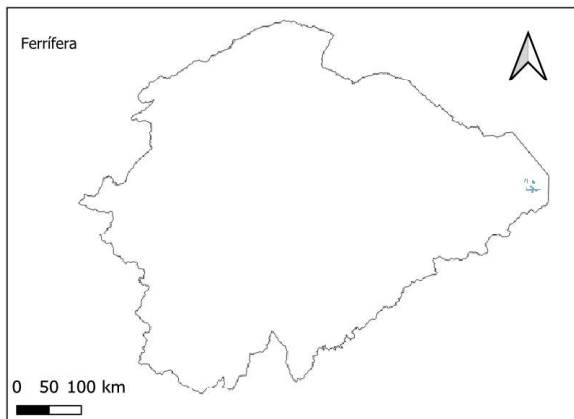
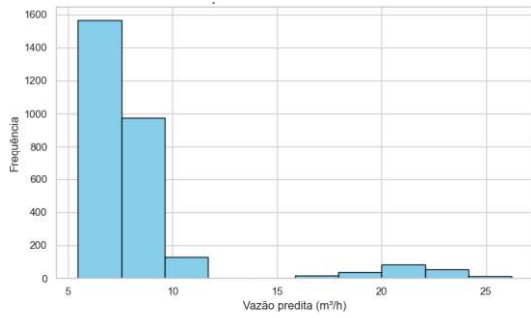
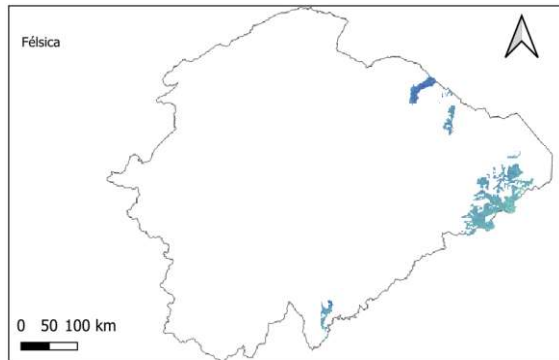
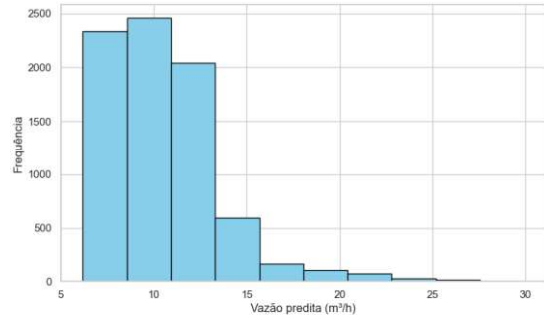
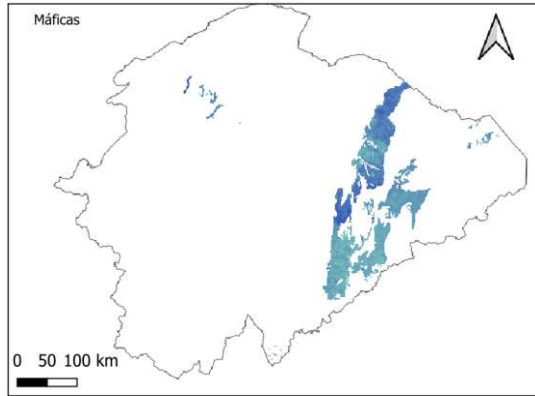
Com base nos modelos treinados, foi gerado um mapa contínuo do potencial aquífero, representando a média da vazão predita para cada pixel da área de estudo (Figura 2.10). Essa representação espacial permite visualizar, com maior detalhamento, os gradientes de produtividade hídrica ao longo do território, destacando zonas prioritárias para captação, preservação e monitoramento. Além disso, os padrões espaciais observados mostram-se coerentes com a distribuição geológica e morfométrica da região, reforçando a aplicabilidade do modelo como ferramenta de apoio ao planejamento estratégico da gestão hídrica subterrânea.

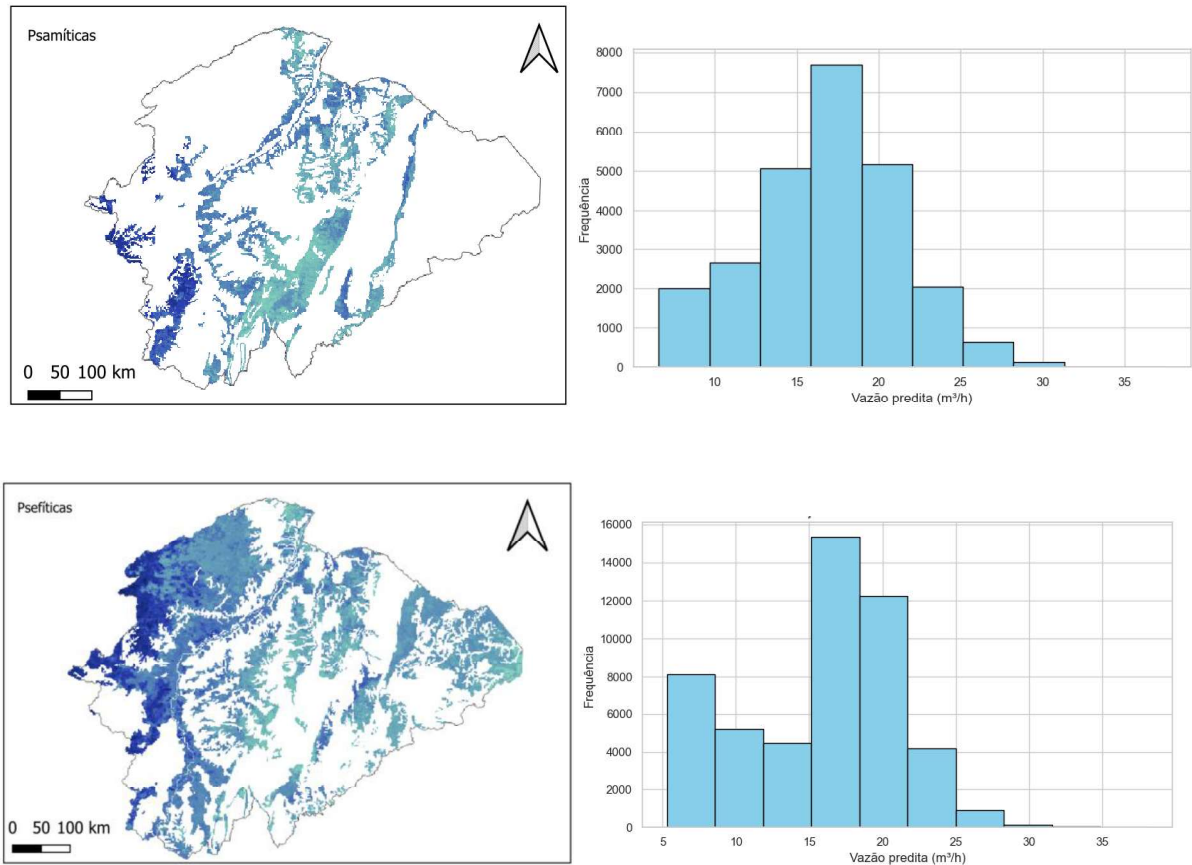


**Figura 2.10.** Potencial de água subterrânea obtido pelo modelo KKNN

A Figura 2.11 apresenta a distribuição dos valores de potencial hídrico subterrâneo por domínio hidrogeológico, evidenciada por histogramas. Os resultados permitem comparar a produtividade da vazão estimada entre diferentes litologias, destacando a variabilidade entre os sistemas aquíferos fraturados, cársticos e porosos presentes na área de estudo.





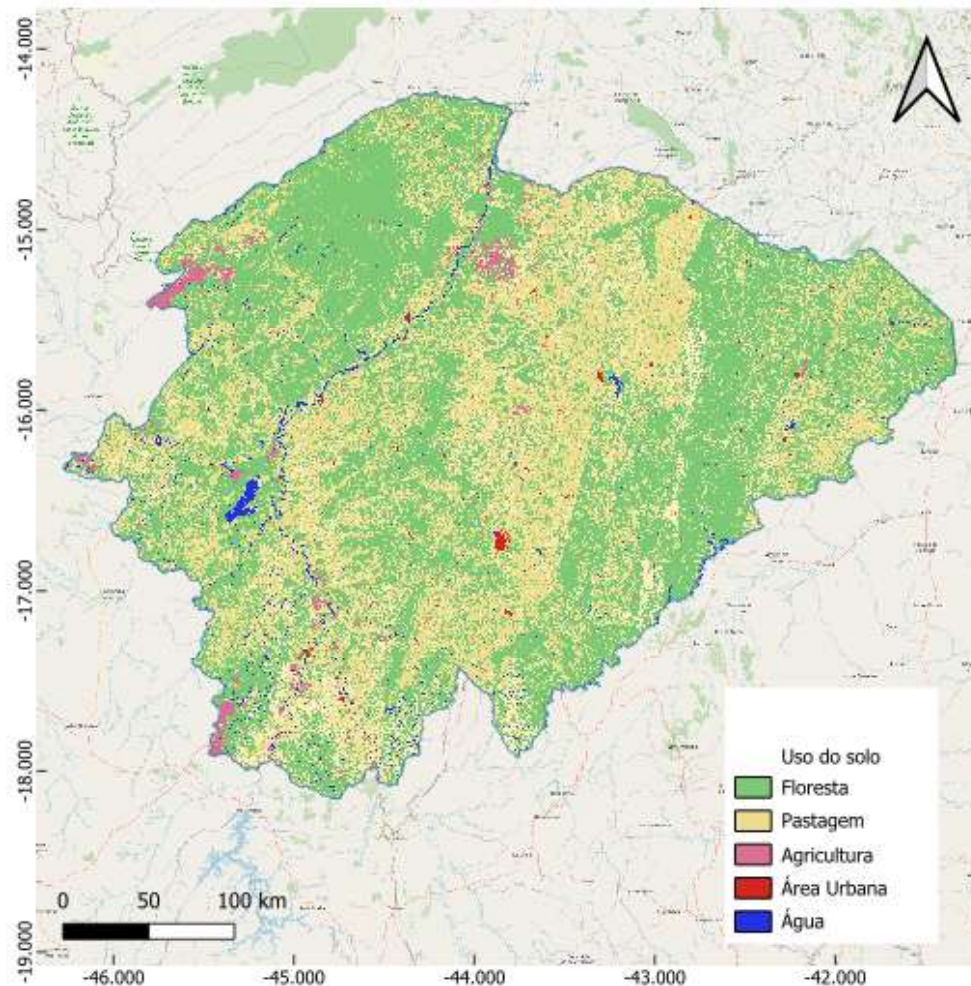


**Figura 2.11.** Distribuição dos valores do potencial hídrico subterrâneo nos domínios hidrogeológicos com a representação do histograma de valores

A análise da distribuição da vazão predita revelou influência direta da litologia e do uso do solo sobre o potencial hídrico subterrâneo na região de estudo. As formações litológicas carbonáticas se destacaram por concentrarem os maiores valores de vazão, com mais de 11 mil pixels classificados nesse intervalo. Esse desempenho pode ser atribuído às características típicas dos sistemas cárstico-fissurais, nos quais a água subterrânea circula por fraturas, falhas e condutos de dissolução, favorecendo a recarga e a condução de grandes volumes de água, como já evidenciado por Sanz Pérez *et al.* (2025). Tais sistemas, comumente associados a rochas carbonáticas, apresentam elevada conectividade hidráulica e potencial aquífero, como também registrado por Atman *et al.* (2011) e Pessoa *et al.* (2020) na mesma região, e por Banks *et al.* (1996) em estudos clássicos sobre aquíferos fraturados.

Além da litologia, o uso do solo (Figura 2.12) nas áreas de domínio carbonático também contribui para a eficiência do processo de recarga. Observou-se uma predominância de uso para pastagem (7,9 milhões de pixels) e cobertura florestal (7,5 milhões de pixels), ambos associados a maior infiltração e proteção do solo. A vegetação natural desempenha papel fundamental na manutenção do ciclo hidrológico e da disponibilidade de água subterrânea, conforme já

discutido por Bosquilia *et al.* (2019) e Silva *et al.* (2020), sobretudo em regiões com clima semiárido e solos rasos.



**Figura 2.12.** Uso e ocupação do solo da área de estudo

As litologias psamíticas e psefíticas também apresentaram elevada produtividade hídrica. As psamíticas, compostas majoritariamente por arenitos, exibiram mais de 19.930 pixels na faixa de ação superiores a  $13,5 \text{ m}^3/\text{h}$ , enquanto as psefíticas, compostas por conglomerados e sedimentos grossos, registraram mais de 35.783 pixels nessa mesma faixa. Esses materiais sedimentares, por sua natureza porosa e permeável, favorecem a infiltração da água da chuva e a formação de aquíferos intersticiais. As áreas associadas a essas litologias também apresentaram usos do solo voltados à pastagem e à cobertura florestal, o que pode contribuir para a manutenção da recarga, especialmente em regiões de relevo mais plano e menor escoamento superficial.

Em contraste, as litologias félsicas e ferríferas, associadas a rochas ígneas e metamórficas compactas, mostraram baixa produtividade hídrica. As felsicas apresentaram apenas 204 pixels com vazão acima de 13,5 m<sup>3</sup>/h, enquanto as ferríferas praticamente não apresentaram pixels nas faixas superiores, reforçando sua baixa capacidade de armazenamento e transmissão, típica de rochas com porosidade primária desprezível. faixa. Esses resultados são coerentes com a baixa capacidade de armazenamento e transmissão de água dessas rochas, cujas porosidades primárias são geralmente negligenciáveis, sendo a ocorrência de água subterrânea restrita a zonas fraturadas. Esse comportamento foi amplamente discutido por Inocêncio *et al.* (2021), que relataram incertezas quanto à produtividade hídrica em poços perfurados em rochas cristalinas, sobretudo na ausência de critérios técnicos apropriados.

As rochas máficas, embora compactas, totalizaram 893 pixels com vazão acima de 13,5 m<sup>3</sup>/h, indicando potencial moderado, o que pode ser atribuído à presença de fraturas ou zonas de intemperismo que aumentam a capacidade de armazenamento e transmissão de água. Esses resultados sugerem que, embora a litologia seja um fator condicionante fundamental, aspectos estruturais e geomorfológicos locais, como relevo, presença de fraturas e tipo de cobertura do solo, podem modificar substancialmente o potencial aquífero. As pelíticas, por sua vez, apresentaram distribuição mais equilibrada entre as faixas de vazão: 7.219 pixels superaram 13,5 m<sup>3</sup>/h, sugerindo uma permeabilidade moderada, possivelmente relacionada a zonas de alteração secundária ou fraturamento. A Tabela 2.5 apresenta a síntese das interações entre litologia, tipo de aquífero, faixa de vazão e uso do solo, acompanhadas de uma interpretação hidroambiental

**Tabela 2.5.** Relação entre classes litológicas, tipo de aquífero, faixa de vazão predominante, uso do solo associado e interpretação ambiental

<b>Litologia</b>	<b>Tipo de Aquífero Predominante</b>	<b>Faixa de vazão predominante (m<sup>3</sup>/h)</b>	<b>Uso do solo predominante</b>	<b>Interpretação ambiental e hídrica</b>
Carbonáticas	Cárstico Fraturado	- 13,5 – 21,7	Pasto > Floresta	Alta produtividade associada à presença de aquífero cárstico; uso intensivo pode pressionar recarga.
Félsicas	Fraturado	5,3 – 13,5	Pasto ≈ Floresta	Predomínio de aquíferos fraturados com uso misto; recarga moderada e sensível a mudanças no uso.
Ferríferas	Fraturado	5,3 – 13,5	Floresta	Baixa densidade de uso antrópico; potencial aquífero reduzido por estrutura da rocha.
Máficas	Fraturado	5,3 – 13,5	Pasto ≈ Floresta	Vazões moderadas; ocupação equilibrada entre natural e agropecuária; atenção à intensificação do uso.
Pelíticas	Fraturado	13,5 – 21,7	Pasto > Floresta	Alta frequência de uso agropecuário; áreas com potencial moderado sensíveis à pressão antrópica.
Psamíticas	Poroso	13,5 – 21,7	Pasto > Floresta	Grande cobertura de pastagens; aquíferos porosos com boa recarga, mas sujeitos à superexploração.
Psefiticas	Poroso Fraturado	- 13,5 – 21,7	Floresta > Pasto	Cobertura natural predominante com bom

De maneira geral, os dados analisados corroboram os achados de estudos prévios na região norte de Minas Gerais, como o de Costa (2021), que apontam maior potencial hídrico nas porções central e oeste do território, onde predominam rochas carbonáticas e sedimentos do tipo Urucuia. Essas áreas apresentam características favoráveis à recarga, como maior precipitação, solos mais profundos e cobertura vegetal que favorece a infiltração. Em contrapartida, as áreas orientais, dominadas por rochas cristalinas, relevos escarpados e menor índice pluviométrico, apresentam restrições naturais à recarga, sendo associadas a menores valores de vazão.

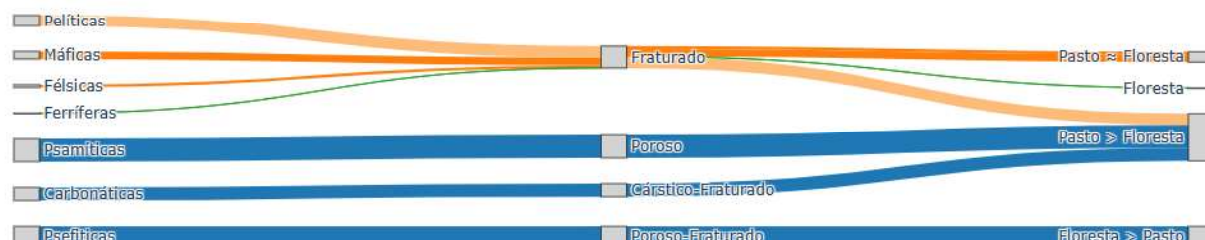
Portanto, os resultados obtidos evidenciam a relevância da integração entre litologia e uso do solo para compreender a distribuição espacial do potencial hídrico subterrâneo. A adoção de estratégias de manejo adequadas, especialmente em áreas de alta vulnerabilidade geológica e climática, é essencial para a conservação dos recursos hídricos na região.

Além dos condicionantes geológicos e de uso do solo, os resultados obtidos também refletem pressões antrópicas relevantes sobre os recursos hídricos subterrâneos, sobretudo nas regiões com maior densidade populacional e atividade agrícola intensiva. Áreas como a do entorno de Montes Claros, importante polo urbano e agroindustrial do norte de Minas Gerais, concentram demandas elevadas por água subterrânea, o que pode acentuar os efeitos de superexploração e reduzir a disponibilidade hídrica local. Estudos anteriores, como o de Durães *et al.* (2022), já indicam que o aumento das captações para abastecimento público, irrigação e uso industrial tem alterado o balanço hídrico em diversas porções da bacia do rio São Francisco.

A Figura 2.13 apresenta um diagrama Sankey que sintetiza a interação entre as classes litológicas predominantes, os tipos de aquíferos associados e os principais usos do solo na área de estudo. Essa abordagem permite identificar zonas prioritárias para conservação, monitoramento ou uso estratégico dos recursos hídricos subterrâneos. As formações carbonáticas, por exemplo, são fortemente associadas a aquíferos cárstico-fraturados com alta vazão (29,9–38,1 m<sup>3</sup>/h) e uso predominante de pastagens, o que indica elevada produtividade natural, mas também potencial pressão antrópica sobre a recarga. Já as litologias psamíticas e psefiticas, ligadas a aquíferos porosos ou poroso-fraturados, apresentam ampla distribuição espacial e ocorrência de vazões elevadas, principalmente em áreas de pasto e vegetação natural, sendo relevantes para ações de conservação hídrica. A espessura dos fluxos é proporcional ao número de pixels por combinação, enquanto a coloração representa a faixa de vazão

predominante ( $\text{m}^3/\text{h}$ ): azul-escuro (29,9–38,1), azul-claro (21,7–29,9), verde (13,5–21,7) e amarelo (5,2–13,5). A visualização permite compreender os padrões de produtividade hídrica e suas condicionantes geológicas e antrópicas, subsidiando o planejamento do uso sustentável da água subterrânea.

Fluxo: Litologia → Tipo de Aquífero → Uso do Solo (colorido por vazão)



**Figura 2.13.** Diagrama Sankey ilustrando a interação entre classes litológicas, tipos de aquíferos e principais categorias de uso do solo na área de estudo.

Nas litologias máficas, pelíticas e félsicas, associadas a aquíferos fraturados, predominam vazões moderadas, com usos do solo variados, incluindo pastagem, silvicultura e mosaicos agrícolas. Esse padrão evidencia que áreas com produtividade intermediária requerem atenção quanto à intensificação do uso do solo. Por outro lado, as formações ferríferas, menos frequentes, concentram-se em zonas com vegetação natural e vazões mais baixas, refletindo menor produtividade associada à baixa permeabilidade das rochas.

Essa análise integrada reforça a importância de considerar as características litológicas, hidrogeológicas e de uso do solo para orientar políticas públicas, definir zonas de exploração sustentável e implementar medidas de proteção em áreas com maior fragilidade hidroambiental. Além disso, destaca-se o potencial do uso de modelos preditivos e visualizações integradas para embasar estratégias de planejamento hídrico em regiões semiáridas, onde a gestão eficiente dos recursos subterrâneos é essencial para a segurança hídrica regional.

O uso do solo também exerce influência direta sobre a recarga dos aquíferos. O avanço de áreas agrícolas em substituição a formações florestais, notadamente nas regiões de litologia sedimentar, tende a reduzir a capacidade de infiltração e aumentar o escoamento superficial, prejudicando o armazenamento de água no subsolo. Culturas de ciclo longo e elevado consumo hídrico, como eucalipto, tem sido amplamente implantado em planaltos da região, impactando negativamente o balanço hídrico, especialmente sob regimes de baixa precipitação. Silva *et al.* (2020) e Ouyang *et al.* (2021) observaram que monoculturas de eucalipto em áreas semiáridas

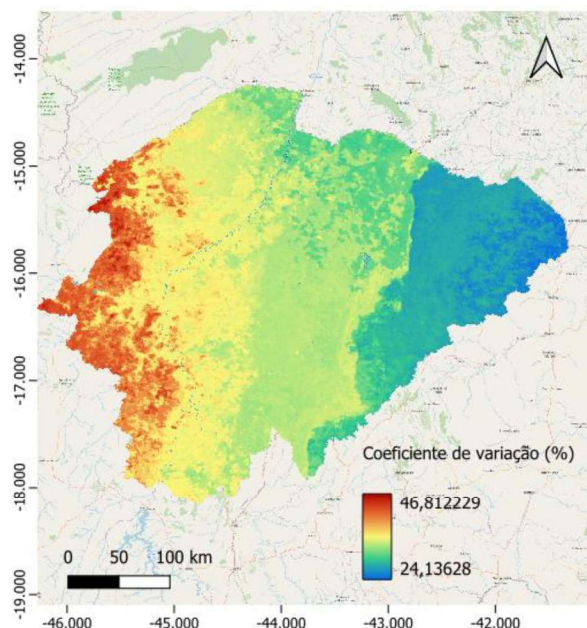
podem apresentar taxas de evapotranspiração comparáveis a ambientes úmidos, o que contribui para a redução do fluxo de recarga e afeta a sustentabilidade do uso da água subterrânea.

Além disso, a expansão urbana desordenada, com aumento de áreas impermeabilizadas, e a escassez de políticas públicas voltadas à proteção das zonas de recarga acentuam a vulnerabilidade dos aquíferos. A ocupação de áreas planas com elevada capacidade de infiltração, como as associadas ao sistema Urucuia, compromete não apenas a disponibilidade hídrica local, mas também a capacidade de manutenção dos fluxos base de rios e córregos intermitentes. A ausência de infraestrutura adequada de saneamento e a disposição inadequada de resíduos sólidos em áreas rurais e urbanas também podem levar à contaminação dos aquíferos, agravando os problemas de quantidade e qualidade da água subterrânea (LUCAS *et al.*, 2021).

Nesse contexto, os resultados deste estudo reforçam a necessidade de considerar o uso e ocupação do solo nas estratégias de gestão integrada dos recursos hídricos. Identificar áreas de alta produtividade aquífera deve ser acompanhada de ações de planejamento territorial, controle do uso agrícola intensivo e proteção das áreas de recarga. Medidas como reflorestamento com espécies nativas, incentivo ao manejo sustentável e delimitação de zonas de proteção de aquíferos (ZPA) são essenciais para garantir a conservação e o uso racional da água subterrânea em regiões semiáridas.

### **2.3.6 Mapa de incerteza das predições (coeficiente de variação)**

A construção de um mapa de incerteza, baseado no coeficiente de variação (CV) das 100 rodadas do modelo KKN, fornece informações cruciais sobre a confiança nos valores preditos. O CV é uma métrica estatística amplamente utilizada para expressar a variabilidade relativa dos dados em relação à média, sendo especialmente útil na análise da robustez de modelos preditivos aplicados a sistemas ambientais complexos (GOOVAERTS, 2001). Regiões com alta variabilidade indicam áreas de maior incerteza, podendo estar associadas à baixa densidade de dados amostrais, elevada complexidade geológica ou variáveis explicativas menos representativas no treinamento do modelo. Essa abordagem permite avaliar não apenas “quanto” foi predito, mas também “quão confiável” é essa predição em cada local do território (HUANG *et al.*, 202). Na Figura 2.14, observa-se que os maiores coeficientes de variação concentram-se na porção oeste e sudoeste da área de estudo, onde predominam domínios litológicos mais heterogêneos, como os sedimentos psamíticos e psefíticos.



**Figura 2.14.** Mapa do coeficiente de variação (%) da vazão predita pelo modelo KKNN, calculado a partir de 100 rodadas de simulação

O gradiente de cores representa a variabilidade relativa dos valores de vazão média, indicando o grau de incerteza associado à predição em cada pixel. Valores mais elevados (vermelhos) indicam maior incerteza, enquanto valores mais baixos (azul-esverdeados) apontam maior estabilidade nas estimativas do modelo.

Essa maior variabilidade pode refletir a influência de aquíferos porosos de alta variabilidade local, complexidade estrutural ou descontinuidade de atributos ambientais, além da escassez de poços de referência em certas regiões. Em contrapartida, a região leste apresenta menor coeficiente de variação, sugerindo maior estabilidade nas predições, possivelmente em função de maior homogeneidade geológica e cobertura de dados amostrais mais representativa. Tais padrões reforçam a importância de integrar a análise de incerteza à etapa de interpretação dos mapas de potencial aquífero, uma vez que áreas com alta variabilidade requerem maior cautela na tomada de decisão (FILDES, 2022; HRACHOWITZ *et al.*, 2014).

Além disso, o mapeamento da incerteza pode ser utilizado como critério para orientar futuras campanhas de monitoramento e amostragem, priorizando regiões com baixa confiabilidade preditiva. Estudos anteriores demonstraram que incorporar mapas de incerteza em processos de gestão adaptativa contribui significativamente para uma alocação mais eficiente de recursos em programas de monitoramento hidrogeológico (CONDON *et al.*, 2021). Assim, a análise espacial do CV se consolida como uma ferramenta complementar e estratégica para avaliação do desempenho de modelos baseados em aprendizado de máquina,

principalmente quando aplicados a ambientes semiáridos com alta variabilidade natural e limitada disponibilidade de dados.

## 2.4 Conclusão

Dentre os modelos testados, o algoritmo *Kernel k-Nearest Neighbors* (KKNN) apresentou o melhor desempenho preditivo, com  $R^2$  de 0,75 e MAE de 0,094 superando os algoritmos RF e SVM. A seleção de covariáveis ambientais foi um aspecto central da metodologia, com destaque para atributos litológicos, pedológicos, geomorfológicos, topográficos e climáticos mensais. A análise de importância das covariáveis no modelo KKNN evidenciou que fatores como precipitação nos meses chuvosos, radiação solar e morfologia do terreno foram fundamentais para a predição da vazão específica, atendendo ao primeiro objetivo do estudo. A abordagem adotada permitiu captar a variabilidade espacial da produtividade dos poços, mesmo em áreas com baixa densidade de dados, gerando mapas contínuos do potencial aquífero. Essas representações espaciais, associadas à avaliação da incerteza, contribuem para o planejamento sustentável e a priorização de áreas críticas de monitoramento.

Adicionalmente, verificou-se que áreas associadas a domínios geológicos carbonáticos e psamíticos concentraram os maiores valores de vazão predita, enquanto regiões com litologias máficas e félsicas apresentaram menor potencial aquífero. O uso intensivo do solo, especialmente para agricultura e silvicultura, também esteve associado às zonas de maior exploração dos recursos subterrâneos, indicando a influência de fatores antrópicos na dinâmica de captação. Apesar dos resultados promissores, limitações como a utilização de dados secundários e a ausência de séries temporais contínuas indicam a necessidade de avanços futuros. Recomenda-se ampliar a base de dados com medições em campo, incorporar variáveis hidrogeofísicas e explorar modelos híbridos que integrem machine learning e modelagem física. Tais estratégias podem aprimorar a capacidade preditiva dos modelos e fortalecer o suporte à gestão dos recursos hídricos subterrâneos em regiões semiáridas.

Dessa forma, a aplicação do modelo KKNN, integrada a dados ambientais multivariados, representa um avanço metodológico na estimativa da produtividade de poços tubulares em contextos de elevada complexidade hidrogeológica, oferecendo suporte técnico ao planejamento hídrico em áreas com monitoramento limitado.

## 2.5 Referências

ABDOWUDAD, D. B. Delineation of groundwater potential zones using GIS and remote sensing techniques: case of Korahe Zone, Somali Regional State, Ethiopia. *International Journal of Scientific and Research Publications*, v. 11, n. 8, p. 412–432, 2021. Disponível em: <https://doi.org/10.29322/IJSRP.11.08.2021.p11653>. Acesso em: 18 ago. 2025.

ADIAT, K. A. N.; NAWAWI, M. N. M.; ABDULLAH, K. Assessing the accuracy of GIS-based elementary multi-criteria decision analysis as a spatial prediction tool – a case of predicting potential zones of sustainable groundwater resources. *Journal of Hydrology*, v. 440–441, p. 75–89, 2012.

AGUMASE, T.; KINDIE, T. E. Spatial analysis of groundwater potential using GIS-based multi-criteria decision analysis method in Lake Tana Basin, Ethiopia. In: **Groundwater of Ethiopia**. Springer, Nature Switzerland, 2019. p. 439–460. Disponível em: [https://doi.org/10.1007/978-3-030-15357-1\\_37](https://doi.org/10.1007/978-3-030-15357-1_37).

ALIZAMIR, M.; KISI, Ö.; ZOUNEMAT-KERMANI, M. Modelling long-term groundwater fluctuations by extreme learning machine using hydro-climatic data. *Hydrological Sciences Journal*, v. 63, n. 1, p. 63–73, 2018. Disponível em: <https://doi.org/10.1080/02626667.2017.1410891>.

AMAMS – ASSOCIAÇÃO DOS MUNICÍPIOS DA ÁREA MINEIRA DA SUDENE. Perfil socioeconômico da região do Norte de Minas Gerais. 2023. Disponível em: <http://www.amams.org.br>

ASLAM, R. A.; SHRESTHA, S.; PANDEY, V. P. Groundwater vulnerability to climate change: a review of the assessment methodology. *Science of the Total Environment*, v. 612, p. 853–875, 2018. Disponível em: <https://doi.org/10.1016/j.scitotenv.2017.08.237>.

ATMAN, D.; VELÁSQUEZ, L. N. M.; FANTINEL, L. M. Controle estrutural na circulação e composição das águas no sistema aquífero cárstico-fissural do Grupo Bambuí, Norte de Minas Gerais. *Águas Subterrâneas*, v. 25, n. 1, 2011. Disponível em: <https://doi.org/10.14295/ras.v25i1.21023>.

ARABAMERI, A. et al. Novel ensemble of MCDM–Artificial Intelligence techniques for groundwater-potential mapping in arid and semi-arid regions (Iran). *Remote Sensing*, v. 12, n. 3, 490, 2020. Disponível em: <https://doi.org/10.3390/rs12030490>. Acesso em: 18 ago. 2025.

BAI, Z. et al. Groundwater potential mapping in Hubei region of China using machine learning, ensemble learning, deep learning and AutoML methods. *Natural Resources Research*, v. 31, p. 2549–2569, 2022. Disponível em: <https://doi.org/10.1007/s11053-021-09931-5>. Acesso em: 18 ago. 2025.

BANKS, D. et al. Permeability and stress in crystalline rocks. *Terra Nova*, v. 8, n. 3, p. 223–235, 1996. Disponível em: <https://doi.org/10.1111/j.1365-3121.1996.tb00751.x>.

BEKELE, D. Modeling the impacts of land use and land cover dynamics on groundwater recharge processes in the Keleta watershed, Awash River Basin. *Environmental Monitoring*

**and Assessment**, v. 193, 2021. Disponível em: <https://doi.org/10.1007/s10661-021-19476-2>. Acesso em: 18 ago. 2025.

BELAY, S. A. Evaluating land suitability and water availability for surface irrigation in the Abbay Basin of Ethiopia. **Royal Society Open Science**, v. 9, n. 12, 2022. Disponível em: <https://doi.org/10.1098/rsos.220674>.

BORZÌ, Iolanda. Modeling Groundwater Resources in Data-Scarce Regions for Sustainable Management: Methodologies and Limits. **Hydrology**, v. 12, n. 1, art. 11, 2025. Disponível em: <https://doi.org/10.3390/hydrology12010011>. Acesso em: 18 ago. 2025.

BOSQUILIA, R. W. D. et al. Evaluation of evapotranspiration variations according to soil type using multivariate statistical analysis. **Geoderma**, v. 355, 113906, 2019. Disponível em: <https://doi.org/10.1016/j.geoderma.2019.113906>.

CÂNDIDO, G. A.; MATOS, A. T.; SALGADO, A. A.; MOREIRA, M. C. Estimativa da recarga potencial de aquíferos em região semiárida utilizando SIG. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 23, n. 3, p. 157–163, 2019. Disponível em: <https://doi.org/10.1590/1807-1929/agriambi.v23n3p157-163>.

CÂNDIDO, M. O. et al. Projeto Águas do Norte de Minas, PANM: estudo da disponibilidade hídrica subterrânea do Norte de Minas Gerais: relatório de integração. **CPRM**, 2019

CÂNDIDO, G. A.; MATOS, A. T.; SALGADO, A. A.; MOREIRA, M. C. Estimativa da recarga potencial de aquíferos em região semiárida utilizando SIG. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 23, n. 3, p. 157–163, 2019. Disponível em: <https://doi.org/10.1590/1807-1929/agriambi.v23n3p157-163>.

CARVALHO, T. S. et al. Impacto das mudanças de uso da terra e do clima sobre a disponibilidade hídrica na Bacia do Rio Grande. RBRH - **Revista Brasileira de Recursos Hídricos**, v. 25, e14, 2020. Disponível em: <https://doi.org/10.1590/2318-0331.252020190119>.

CASTRO, M. A. et al. Modelagem integrada da recarga de aquíferos no semiárido brasileiro. **Revista Brasileira de Recursos Hídricos**, v. 24, e2, 2019.

CPRM – SERVIÇO GEOLÓGICO DO BRASIL. **Sistema de Informações de Águas Subterrâneas (SIAGAS)**. Disponível em: <https://siagasweb.sgb.gov.br/layout/>.

CHILTON, P. J.; FOSTER, S. S. D. Hydrogeological characterisation of aquifers for groundwater resource assessment: hydrogeological characterisation for groundwater resources evaluation. **British Geological Survey Technical Report**, 1995.

CHITSAZAN, M. et al. Groundwater potential mapping using analytical hierarchy process and GIS in Mehran Region, Iran. **Arabian Journal of Geosciences**, v. 8, p. 733–746, 2015. Disponível em: <https://doi.org/10.1007/s12517-013-1161-9>.

COELHO, M. T. P.; DINIZ-FILHO, J. A.; RANGEL, T. F. A parsimonious view of the parsimony principle in ecology and evolution. **Ecography**, v. 42, n. 5, p. 968–976, 2019. Disponível em: <https://doi.org/10.1111/ecog.04228>

CONDON, L. E. *et al.* Global groundwater modeling and monitoring: opportunities and challenges. **Water Resources Research**, v. 57, e2020WR029500, 2021. Disponível em: <https://doi.org/10.1029/2020WR029500>. Acesso em: 18 ago. 2025.

COSTA, L. R. F. Considerações sobre as macrounidades geomorfológicas do estado de Minas Gerais – Brasil. *William Morris Davis – Revista de Geomorfologia*, v. 2, n. 1, p. 11–18, 2021. Disponível em: <https://doi.org/10.48025/ISSN2675-6900.v2n1.2021.109>.

CPRM – SERVIÇO GEOLÓGICO DO BRASIL. Mapas geológicos do Brasil em escala 1:250.000. 2023. Disponível em: <http://geosgb.cprm.gov.br>.

DARGAHI, P. *et al.* Prediction models for groundwater quality parameters: application of multiple linear regression techniques in Kermanshah province, western Iran. **Environmental Monitoring and Assessment**, v. 194, art. 182, 2022. Disponível em: <https://doi.org/10.1007/s10661-022-10020-4>.

DECHASA DIRIBA, S. *et al.* Delineation of groundwater potential zonation using geoinformatics and AHP techniques with remote sensing data. **Heliyon**, v. 10, n. 3, e25532, 2024. Disponível em: <https://doi.org/10.1016/j.heliyon.2024.e25532>

DURÃES, R. C. F. *et al.* Lowering of the water table in the urban spaces of the city of Montes Claros-MG. **Research, Society and Development**, v. 11, n. 7, e20211730018, 2022. Disponível em: <https://doi.org/10.33448/rsd-v11i7.30018>.

ELKIRAN, G.; NOURANI, V.; ABBA, S. I. Multi-step ahead modelling of river water quality parameters using ensemble artificial intelligence-based approach. **Journal of Hydrology**, v. 577, 123962, 2019. Disponível em: <https://doi.org/10.1016/j.jhydrol.2019.123962>

FARR, T. G. *et al.* The Shuttle Radar Topography Mission. **Reviews of Geophysics**, v. 45, RG2004, 2007. Disponível em: <https://doi.org/10.1029/2005RG000183>. Acesso em: 18 ago. 2025.

FATICHI, S. *et al.* An ecohydrological journey of 4500 years reveals a stable but threatened precipitation–groundwater recharge relation around Jerusalem. **Science Advances**, v. 7, n. 37, eabe6303, 2021. Disponível em: <https://doi.org/10.1126/sciadv.abe6303>.

FICK, S. E.; HIJMANS, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. **International Journal of Climatology**, v. 37, p. 4302–4315, 2017. Disponível em: <https://doi.org/10.1002/joc.5086>.

FILDES, S. G. *et al.* Integrating spatially explicit sensitivity and uncertainty analysis in a multi-criteria decision analysis-based groundwater potential zone model. **Journal of Hydrology**, v. 610, 127837, 2022. Disponível em: <https://doi.org/10.1016/j.jhydrol.2022.127837>

GHEISARI, N. *et al.* Incorporation of three different creep models into large-strain consolidation analysis of a clayey tailings deposit. **Computers and Geotechnics**, v. 161, 105533, 2023. Disponível em: <https://doi.org/10.1016/j.compgeo.2023.105533>.

GOMES, L. C. *et al.* Modelling and mapping soil organic carbon stocks in Brazil. **Geoderma**, v. 340, p. 337–350, 2019. Disponível em: <https://doi.org/10.1016/j.geoderma.2019.01.007>.

GOOVAERTS, P. *Geostatistics for Natural Resources Evaluation*. New York: **Oxford University Press**, 1997.

GUEVARA, M.; VARGAS, R. Downscaling satellite soil moisture using geomorphometry and machine learning. **PLoS ONE**, v. 14, n. 9, e0219639, 2019. Disponível em: <https://doi.org/10.1371/journal.pone.0219639>.

HARTMANN, A. et al. A large-scale simulation model to assess karstic groundwater recharge over Europe and the Mediterranean. **Geoscientific Model Development**, v. 10, p. 1109–1131, 2017. Disponível em: <https://doi.org/10.5194/gmd-10-1109-2017>.

HE, Z. et al. Variations of groundwater dynamics in alluvial aquifers with reclaimed water restoring the overlying river, Beijing, China. **Water**, v. 13, 806, 2021. Disponível em: <https://doi.org/10.3390/w13060806>.

HRACHOWITZ, M. et al. Process consistency in models: the importance of system signatures, expert knowledge, and process complexity. **Water Resources Research**, v. 50, p. 7445–7469, 2014. Disponível em: <https://doi.org/10.1002/2014WR015484>

HUANG, S. et al. Assessing geological structure uncertainties in groundwater models using coefficient of variation from stochastic simulations. **Journal of Hydrology**, 2025.

HUSSEIN, A. et al. Predicting groundwater level using gradient boosting machine learning algorithm in Ardabil plain, Iran. **Environmental Earth Sciences**, v. 79, 179, 2020. Disponível em: <https://doi.org/10.1007/s12665-020-08971-9>.

HUJOEL, I. A. et al. Machine learning in detection of undiagnosed celiac disease. **Clinical Gastroenterology and Hepatology**, v. 16, n. 8, p. 1354, 2018.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Mapeamento geomorfológico do Brasil: escala 1:250.000. 2019. Disponível em: <https://www.ibge.gov.br>.

INOCÊNCIO, T. D. M. et al. Linking drought propagation with episodes of climate-induced water insecurity in Pernambuco State – Northeast Brazil. **Journal of Arid Environments**, v. 193, 104593, 2021. Disponível em: <https://doi.org/10.1016/j.jaridenv.2021.104593>.

JARDIM, E. F. F. et al. Impactos da variabilidade climática sobre os recursos hídricos no semiárido mineiro. **Revista Brasileira de Recursos Hídricos**, v. 28, e47, 2023. Disponível em: <https://doi.org/10.1590/2318-0331.2820230051>

JAAFARZADEH, M. S.; TAHMASEBIPOOR, N.; HAGHIZADEH, A. et al. Groundwater recharge potential zonation using an ensemble of machine learning and bivariate statistical models. **Scientific Reports**, v. 11, Art. 5587, 2021.

JESIYA, N. P.; GOPINATH, G. A fuzzy-based MCDM–GIS framework to evaluate groundwater potential index for sustainable groundwater management – A case study in an urban-peri-urban ensemble, southern India. **Groundwater for Sustainable Development**, v. 11, art. 100466, 2020. Disponível em: <https://doi.org/10.1016/j.gsd.2020.100466>

KAHIL, M. T.; DINAR, A.; ALBIAC, J. Modeling water scarcity and droughts for policy adaptation to climate change in arid and semiarid regions. **Journal of Hydrology**, v. 522, p. 95–109, 2015.

KARIMI, S. et al. A hybrid model of neural networks and GIS for groundwater potential mapping in Iran. **Hydrogeology Journal**, v. 20, p. 1651–1662, 2012. Disponível em: <https://doi.org/10.1007/s10040-012-0900-3>.

KASIVISWANATHAN, K. S. et al. Groundwater level forecasting using empirical mode decomposition and artificial neural networks. **Journal of Hydrology**, v. 538, p. 162–173, 2016. Disponível em: <https://doi.org/10.1016/j.jhydrol.2016.04.027>.

KERN, A. N. et al. Machine learning based predictive modeling of debris flow probability following wildfire in the Intermountain Western United States. **Mathematical Geosciences**, v. 49, n. 6, p. 717–735, 2017.

KOMBO, O. H. et al. Long-term groundwater level prediction model based on hybrid KNN RF technique. **Hydrology**, v. 7, n. 3, p. 59, 2020. Disponível em: <https://doi.org/10.3390/hydrology7030059>

KUHN, M. Building predictive models in R using the caret package. **Journal of Statistical Software**, v. 28, n. 5, p. 1–26, 2008. Disponível em: <https://doi.org/10.18637/jss.v028.i05>.

KUMAR, V. et al. Advanced machine learning techniques to improve hydrological prediction: a comparative analysis of streamflow prediction models. **Water**, v. 15, p. 2572, 2023. Disponível em: <https://doi.org/10.3390/w15202572>.

LALLAHEM, S. et al. On the use of neural networks to evaluate groundwater levels in fractured media. **Journal of Hydrology**, v. 307, p. 92–111, 2005. Disponível em: <https://doi.org/10.1016/j.jhydrol.2004.10.013>.

LI, P.; WU, J. Water resources and sustainable development. **Water**, v. 16, p. 134, 2024.

LIN, L. I.-K. A concordance correlation coefficient to evaluate reproducibility. **Biometrics**, v. 45, n. 1, p. 255–268, mar. 1989.

LUCAS, M. C. et al. Significant baseflow reduction in the São Francisco River Basin. **Water**, v. 13, n. 1, 2, 2021. Disponível em: <https://doi.org/10.3390/w13010002>.

MACDONALD, A. M.; CALOW, R. C.; NICOL, A. L.; ROBINS, N. S. Developing groundwater: a guide for rural water supply. **British Geological Survey Technical Report WC/00/33**, 2005.

MACHIWAL, D.; JHA, M. K.; MAL, B. C. Assessment of groundwater potential in a semi-arid region of India using remote sensing, GIS and MCDM techniques. **Water Resources Management**, v. 25, n. 3, p. 1359–1386, 2011.

MADANI, A.; NIYAZI, B. Groundwater potential mapping using remote sensing and Random Forest machine learning model: a case study from the lower part of Wadi Yalamlam, western

Saudi Arabia. **Sustainability**, v. 15, n. 3, 2772, 2023. Disponível em: <https://doi.org/10.3390/su15032772>.

MASKOONI, E. K. et al. Application of advanced machine learning algorithms to assess groundwater potential using remote sensing derived data. **Remote Sensing**, v. 12, n. 17, 2742, 2020. Disponível em: <https://doi.org/10.3390/rs12172742>.

MARENGO, J. A.; BERNASCONI, M. Mudanças climáticas e os recursos hídricos no Brasil. In: SANTOS, M.; MARENGO, J. A.; VALVERDE, M. C. (eds.). **Águas do Brasil: análise estratégica**. Brasília: ANA, 2015. p. 215–232.

MONTENEGRO, S.; RAGAB, R. Impact of possible climate and land use changes in the semi-arid regions: a case study from North Eastern Brazil. **Journal of Hydrology**, v. 383, n. 3–4, p. 232–241, 2010. Disponível em: <https://doi.org/10.1016/j.jhydrol.2009.12.035>.

MORIASI, D. N. et al. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. **Transactions of the ASABE**, v. 50, n. 3, p. 885–900, 2007. Disponível em: <https://doi.org/10.13031/2013.23153>.

NADIRI, A. A. et al. Groundwater quality prediction using hybrid soft computing models: a case study in Iran. **Environmental Earth Sciences**, v. 78, 279, 2019. Disponível em: <https://doi.org/10.1007/s12665-019-8254-5>.

NAG, S. K.; GHOSH, P. Variation in groundwater levels and water quality in Chhatna Block, Bankura district, West Bengal, a GIS approach. **Journal of the Geological Society of India**, v. 81, p. 261–280, 2013. Disponível em: <https://doi.org/10.1007/s12594-013-0029-3>.

NASA – NATIONAL AERONAUTICS AND SPACE ADMINISTRATION. Shuttle Radar Topography Mission (SRTM) Digital Elevation Data. 2022. Disponível em: <https://www2.jpl.nasa.gov/srtm/>.

NOGUEIRA MARTINS, R. et al. Comparing IDW and ordinary kriging to map soil attributes in a farm area, Northern Minas Gerais, Brazil. **Communications in Soil Science and Plant Analysis**, v. 50, n. 20, p. 2570–2578, 2019. Disponível em: <https://doi.org/10.1080/00103624.2019.1670836>

OUYANG, L. et al. Consumption of precipitation by evapotranspiration indicates potential drought for broadleaved and coniferous plantations in hilly lands of South China. **Agricultural Water Management**, v. 252, 106927, 2021. Disponível em: <https://doi.org/10.1016/j.agwat.2021.106927>.

PESSOA, P.; ATMAN, D.; KIMURA, G. Environmental problems in the Lagoa Santa karst. In: AULER, A.; PESSOA, P. (Ed.). Lagoa Santa Karst: Brazil's iconic karst region. Cham: **Springer International Publishing**, 2020. p. 283–303. Disponível em: [https://doi.org/10.1007/978-3-030-35940-9\\_13](https://doi.org/10.1007/978-3-030-35940-9_13).

RAHMATI, O. et al. Groundwater potential mapping using AHP and GIS techniques: a case study from the Central Iran. **Hydrogeology Journal**, v. 24, p. 425–446, 2016. Disponível em: <https://doi.org/10.1007/s10040-015-1348-8>. Acesso em: 18 ago. 2025.

RAZAVI TERMEH, S. V.; SADEGHI NIARAKI, A.; CHOI, S.-M. Groundwater potential mapping using an integrated ensemble of three bivariate statistical models with random forest and logistic model tree models. **Water**, v. 11, n. 8, 1596, 2019. Disponível em: <https://doi.org/10.3390/w11081596>.

R CORE TEAM. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2023. Disponível em: <https://www.r-project.org>.

RIAZ, M. T. et al. An integrated approach of support vector machine (SVM) and weight of evidence (WOE) techniques to map groundwater potential and assess water quality. **Scientific Reports**, v. 14, art. 26186, 2024. Disponível em: <https://doi.org/10.1038/s41598-024-76607-3>.

RIBEIRO, L. O. A. Avaliação do sistema de monitoramento de águas subterrâneas do Estado de Minas Gerais: contribuições para a governança dos recursos hídricos. 2024. 144 f. **Dissertação** (Mestrado em Geografia) – Universidade Federal de Minas Gerais, Belo Horizonte, 2024. Disponível em: <https://repositorio.ufmg.br/handle/1843/77418>.

RIBEIRO, J. P. M. et al. Análise da recarga no sistema aquífero granular e fissural na área do campus Pampulha da Universidade Federal de Minas Gerais. **Geonomos**, v. 22, n. 2, p. 28–43, 2014.

SAHOUR, H.; GHOLAMI, V.; VAZIFEDAN, M. A. A comparative analysis of statistical and machine learning techniques for mapping the spatial distribution of groundwater salinity in a coastal aquifer. **Journal of Hydrology**, v. 591, 125321, 2020. Disponível em: <https://doi.org/10.1016/j.jhydrol.2020.125321>.

SANZ PÉREZ, E. et al. A procedure to estimate global natural recharge in karst aquifers. **Water**, v. 17, n. 12, 1779, 2025. Disponível em: <https://doi.org/10.3390/w17121779>.

SCHLIEP, K. kkn: Weighted k-Nearest Neighbors. R package version 1.3.2, 2025. Disponível em: <https://CRAN.R-project.org/package=kkn>

SEIFU, T. K. et al. Groundwater potential mapping using GIS and remote sensing with multi-criteria decision-making in Shinile sub-basin, eastern Ethiopia. **Songklanakarin Journal of Science & Technology**, v. 45, n. 1, p. 122–132, 2023.

SENA, N. C. et al. Analysis of terrain attributes in different spatial resolutions for digital soil mapping application in southeastern Brazil. **Geoderma Regional**, v. e00268, 2020.

SILVA, R. M. et al. Impacto do uso e cobertura da terra sobre a recarga de aquíferos no semiárido brasileiro. **Revista Brasileira de Geografia Física**, v. 13, n. 5, p. 2020–2034, 2020.

SILVA, J. L. et al. As secas no Jequitinhonha: demandas, técnicas e custos do abastecimento no semiárido de Minas Gerais. **Revista Brasileira de Estudos Urbanos e Regionais**, v. 22, 2020. Disponível em: <https://doi.org/10.22296/2317-1529.rbeur.202013>.

SILVA, L. A. et al. Estimating water loss in an environmental protection area – Minas Gerais, southeast Brazil. **Caderno de Geografia**, v. 30, n. 62, p. 76–88, 2020. Disponível em: <https://doi.org/10.5752/P.2318-2962.2020v30n62p76>.

SINGHA, S.; PASUPULETI, S. Delineation of groundwater prospect zones in Arang Block, Raipur District, Chhattisgarh, Central India, using Analytical Network Process. **Journal of the Geological Society of India**, 2021. Disponível em: <https://doi.org/10.1007/s12594-020-1487-z>.

SMAIL, F. et al. Assessing the impact of climate change on groundwater recharge in semi-arid regions: a case study in Algeria. **Groundwater for Sustainable Development**, v. 8, p. 554–566, 2019. Disponível em: <https://doi.org/10.1016/j.gsd.2019.01.002>. Acesso em: 18 ago. 2025.

SOUZA, C. M. P. et al. Spatiotemporal prediction of rainfall erosivity by machine learning, Minas Gerais State, Brazil. **Geocarto International**, v. 37, n. 2, p. 1–19, 2022.

TAMESGEN, Y.; ATLABACHEW, A.; MURALITHARAN, J. Groundwater potential assessment in the Blue Nile River catchment, Ethiopia, using geospatial and multi-criteria decision making techniques. **Heliyon**, v. 9, n. 6, e17616, 2023. Disponível em: <https://doi.org/10.1016/j.heliyon.2023.e17616>.

TAMIRU, H.; WAGARI, M. Evaluation of data-driven model and GIS technique performance for identification of groundwater potential zones: a case of Fincha catchment, Abay Basin, Ethiopia. **Journal of Hydrology: Regional Studies**, 2021. Disponível em: <https://doi.org/10.1016/j.ejrh.2021.100902>

TENENWURCEL, M. A. et al. An improved model for the evaluation of groundwater recharge based on the concept of conservative use potential: a study in the River Pandeiros watershed, Minas Gerais, Brazil. **Water**, v. 12, n. 4, 1001, 2020. Disponível em: <https://doi.org/10.3390/w12041001>.

TAO, T. et al. Predicting groundwater recharge under different climate change scenarios using machine learning. **Journal of Hydrology**, v. 604, 127232, 2022. Disponível em: <https://doi.org/10.1016/j.jhydrol.2021.127232>

TODD, D. K.; MAYS, L. W. Groundwater hydrology. 3. ed. **Hoboken**: John Wiley and Sons, 2005. 652 p.

TOURÉ, H. KNN RF model is well-suited for forecasting seasonal variations in groundwater depths. **Environmental and Earth Sciences**, 2024.

TURKELTAUB, T.; BEL, G. Changes in mean evapotranspiration dominate groundwater recharge in semiarid regions. **Hydrology and Earth System Sciences**, v. 28, n. 18, p. 4263–4274, 2024. Disponível em: <https://doi.org/10.5194/hess-28-4263-2024>.

TUKEY, J. W. Exploratory data analysis. Reading, MA: Addison-Wesley, 1977.

UNIVERSIDADE FEDERAL DE VIÇOSA. **Mapa de solos do estado de Minas Gerais (1:650.000)**. Disponível em: <https://dps.ufv.br/software/>.

WEI, W. et al. Identifying the impacts of land use and topography on groundwater recharge in semi-arid regions using multiple linear regression. **Hydrology Research**, v. 52, n. 2, p. 489–505, 2021. Disponível em: <https://doi.org/10.2166/nh.2021.110>. Acesso em: 18 ago. 2025.

WRIGHT, E. P. The hydrogeology of crystalline basement aquifers in Africa. **Hydrogeology Journal**, v. 1, n. 2, p. 65–76, 1992. Disponível em: <https://doi.org/10.1007/BF00868670>.

YIFRU, B. A. et al. Groundwater potential mapping using SWAT and GIS-based multi-criteria decision analysis. **KSCE Journal of Civil Engineering**, v. 24, p. 2546–2559, 2020. Disponível em: <https://doi.org/10.1007/s12205-020-0168-1>.

ZHANG, H. et al. LSTM time series model for predicting water table depth. **Frontiers in Environmental Science**, 2018. Disponível em: <https://doi.org/10.3389/fenvs.2024.1291327>

## CAPÍTULO 3

### **Qualidade da água subterrânea para fins de irrigação em região semiárida: diagnóstico multiparamétrico e predição por aprendizado de máquina**

#### **3.1 Introdução**

A intensificação da demanda por água subterrânea em regiões semiáridas é reflexo de suas condições climáticas desafiadoras, como baixos índices pluviométricos, distribuição irregular das chuvas ao longo do ano e alta evapotranspiração, resultando em um déficit hídrico (HUANG *et al.*, 2021; LAPIS, 2024). Segundo a classificação do Programa das Nações Unidas para o Meio Ambiente (UNEP), essas regiões apresentam um índice de aridez entre 0,2 e 0,5 e abrangem porções da superfície de países como Brasil, Índia, Irã, Etiópia, México e Austrália (UNEP, 1992; HUANG *et al.*, 2021). No semiárido brasileiro, embora as precipitações anuais variem entre 200 e 800 mm, valores superiores aos de outras regiões semiáridas do mundo (80 e 250 mm.ano<sup>-1</sup>), esse volume não é suficiente para suprir as elevadas perdas por evaporação que podem atingir até 3.000 mm anuais (LAPIS, 2024). Além disso, a baixa profundidade dos solos cristalinos e a rápida drenagem superficial comprometem a recarga dos aquíferos, reduzindo significativamente a disponibilidade de água subterrânea (SILVA *et al.*, 2020).

Esse conjunto de fatores torna a água subterrânea uma importante fonte para o abastecimento urbano, rural e para a agricultura irrigada. Estima-se que cerca de 60% da água utilizada na irrigação em regiões semiáridas de países em desenvolvimento tenha origem em aquíferos (AMWELE *et al.* 2021), refletindo sua importância socioeconômica e ambiental. A agricultura, principal atividade econômica dessas regiões, desempenha papel relevante na economia local e na segurança alimentar, mas o uso contínuo e intensivo deste recurso requer atenção à sua qualidade e disponibilidade (KHOSRAVI *et al.*, 2018).

Entre os desafios associados ao uso da água subterrânea destaca-se a salinidade, que pode ocorrer por fatores naturais, associada à dissolução de minerais evaporíticos, intemperismo de rochas, troca iônica e mobilização de sais acumulados nos solos (ABDULRAHMAN *et al.*, 2023). Entretanto, práticas agrícolas intensivas, como o uso excessivo de fertilizantes, irrigação sem controle apropriado, e a própria extração de grandes volumes de água subterrânea, que pode promover a ascensão de zonas salinas, agravam a qualidade da água utilizada na irrigação (OMRANI *et al.*, 2021; WANG *et al.*, 2022). Assim, a aplicação de águas subterrâneas salinas na agricultura afeta não apenas a produtividade das

culturas, mas também acelera a salinização dos solos, prejudicando a estrutura física do solo, a disponibilidade de nutrientes e, conseqüentemente, o crescimento vegetal (ZHAO *et al.*, 2023). A salinidade das águas subterrâneas representa riscos à saúde humana e ao crescimento das culturas (LIU *et al.*, 2018; WEN *et al.*, 2019) podendo afetar a fertilidade do solo e características da vegetação, favorecendo a salinização das zonas radiculares e a redução do crescimento das culturas e de sua produtividade (ZHANG *et al.*, 2024; WEN *et al.*, 2019).

Dentro deste contexto, faz-se necessária a busca de métodos para o diagnóstico da salinidade de águas subterrâneas. Embora a análise laboratorial convencional seja mais precisa e confiável, a mesma é onerosa, e de elevada complexidade em função da logística e do tempo para obtenção dos resultados, implicando em limitações operacionais para o monitoramento *in loco* em extensas áreas (SHAH *et al.*, 2023). Estudos como os de Singh *et al.* (2021) reportam a influência da pluviosidade sobre a qualidade da água subterrânea em regiões semiáridas, destacando mudanças na composição iônica após a estação chuvosa. No entanto, apesar dessas evidências, ainda são escassos os estudos que avaliam de forma sistemática os impactos da sazonalidade sobre índices multiparamétricos relacionados à salinidade, como CE, RAS, RSC e %Na, indicadores essenciais para o diagnóstico da aptidão da água para irrigação.

De forma complementar, alguns estudos têm avançado no desenvolvimento de métodos de previsão da qualidade da água baseados em modelos de aprendizado de máquina, os quais são caracterizados pela aplicação das relações não lineares entre variáveis independentes e variáveis dependentes, assim como pela melhoria na previsão da qualidade de água subterrâneas a partir dos modelos matemáticos desenvolvidos (BOUDIBI *et al.*, 2024; SARKAR *et al.*, 2024; TAGHIZADEH-MEHRJARDI *et al.*, 2021). De acordo com Omrani *et al.* (2021), algoritmos de aprendizado de máquina têm sido amplamente aplicados para estimar parâmetros de salinidade e concentração iônica de forma rápida e eficiente em diferentes regiões do mundo. No entanto, conforme apontado por Shaw e Sharma (2024), ainda há lacunas no que se refere à escolha das variáveis de entrada mais adequadas que proporcionem maior acurácia nas previsões para os modelos preditivos, assim como na identificação das combinações que garantam maior acurácia e aplicabilidade prática nas previsões. Além disso, a influência sazonal das chuvas nos parâmetros de salinidade, como condutividade elétrica (CE), Relação de Absorção de Sódio (RAS), RSC e %Na, ainda não é amplamente abordada na literatura. A maioria dos estudos existentes foca na qualidade da água subterrânea de forma estática, sem considerar como as variações climáticas e a precipitação afetam esses índices ao longo do ano (PINGHENG *et al.* 2025). Esse aspecto é fundamental, pois a interação entre as chuvas e os parâmetros de qualidade da água pode influenciar diretamente a viabilidade do uso

dessa água para irrigação.

Grande parte das abordagens atuais ainda se limita ao uso de parâmetros físico-químicos, como condutividade elétrica (CE) ou total de sólidos dissolvidos (SDT), desconsiderando índices complementares importantes, a exemplo, da Relação de Absorção de Sódio (RAS) e do Percentual de Sódio e do RSC (BOUDIBI *et al.*, 2024; SARKAR *et al.*, 2024). Essa visão fragmentada pode comprometer a avaliação global da qualidade da água para irrigação, uma vez que o uso de diferentes indicadores reflete aspectos distintos da interação entre salinidade, solo e cultura. Apesar da importância desses índices, estudos recentes têm mostrado que muitos modelos preditivos de salinidade ainda focam na estimativa de parâmetros físicos isolados, sem considerar a complexidade das interações entre os diversos indicadores e sua variação espaço-temporal, especialmente em ambientes semiáridos, onde essas variáveis são altamente heterogêneas (SHAW e SHARMA, 2024).

Nesse cenário, métodos baseados em aprendizado de máquina têm ganhado destaque por sua capacidade de identificar relações não lineares e realizar previsões robustas a partir de grandes volumes de dados (TAGHIZADEH-MEHRJARDI *et al.*, 2021). Tais modelos permitem o desenvolvimento de sistemas inteligentes para monitoramento e gestão da qualidade da água, reduzindo a necessidade de análises laboratoriais onerosas e de difícil logística em regiões remotas (OMRANI *et al.*, 2021; SARKAR *et al.*, 2024).

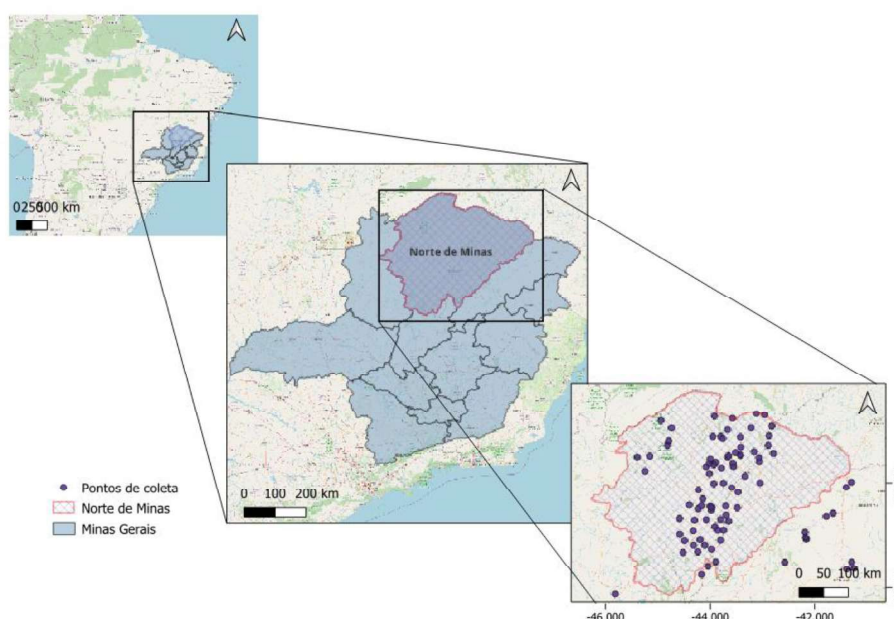
Com base nessas limitações, este estudo teve por objetivos específicos: i) Avaliar a influência da pluviosidade na qualidade de água subterrânea, no período seco e chuvoso, em uma região semiárida; ii) Avaliar a qualidade das águas subterrâneas para fins de irrigação em uma região semiárida; e (iii) Desenvolver modelos preditivos com algoritmos de aprendizado de máquina capazes de estimar índices de qualidade de água focados em salinidade para o uso da água subterrânea para fins de irrigação.

O presente estudo busca contribuir com a construção de modelos de aprendizado de máquina que integrem múltiplos indicadores de salinidade e explorem sua variação em escala espaço-temporal. A novidade desta pesquisa reside na aplicação conjunta de ferramentas computacionais avançadas e na análise multivariada de parâmetros de qualidade de água relacionados à salinidade, permitindo um diagnóstico mais assertivo do uso deste recurso para fins agrícolas, a fim de subsidiar a tomada de decisão na gestão hídrica de regiões semiáridas.

### **3.2 Material e Métodos**

### 3.2.1 Área de estudo

A área de estudo está localizada na região do semiárido mineiro que abrange a região norte do estado de Minas Gerais (Figura 3.1), tendo como características predominantes baixos índices pluviométricos (400 – 900 mm), altas temperaturas (24°C - 27°C), elevada intensidade de luz solar e baixa umidade relativa (SENA *et al.*, 2019).



**Figura 3.1.** Localização da área de estudo

### 3.2.2 Obtenção e tratamento dos dados

A obtenção dos dados do monitoramento da qualidade da água subterrânea na região foi realizada a partir da consulta ao banco de dados do Instituto Mineiro de Gestão das Águas (IGAM) no período de 2017 a 2023. O banco de dados é composto por 95 pontos de coletas (Figura 3.1) com 672 observações, as quais foram realizadas nos meses de abril, maio, julho, agosto, novembro e dezembro no período analisado, os quais foram organizados em períodos seco (abril. – setembro) e chuvoso (outubro – março).

### 3.2.3 Análise comparativa sazonal da pluviosidade

Como forma de embasar a avaliação dos efeitos sazonais sobre a qualidade da água subterrânea na região do semiárido mineiro, foi realizada uma análise comparativa da pluviosidade entre os períodos seco (abril a setembro) e chuvoso (outubro a março). Para isso, aplicou-se o teste t de Student para amostras independentes, a fim de verificar se havia diferença estatisticamente significativa entre as médias de precipitação nos dois períodos.

Os dados de precipitação mensal foram obtidos no site do Instituto Nacional de Meteorologia (INMET), abrangendo o período de 2017 a 2023, para as principais estações meteorológicas localizadas na área de estudo. A análise foi conduzida utilizando a linguagem de programação R, com nível de significância estabelecido em 5% ( $\alpha < 0,05$ ).

Para as variáveis que atenderam à suposição de normalidade, aplicou-se o teste t de Student. Para aquelas que não atenderam a esse pressuposto, optou-se por uma comparação não paramétrica adequada às distribuições observadas.

### **3.2.4 Determinação dos índices de qualidade de água subterrânea para fins de irrigação**

No estudo os parâmetros de qualidade da água subterrânea considerados foram pH, sulfato ( $\text{SO}_4^{2-}$ ), condutividade elétrica (CE), sólidos dissolvidos totais (SDT), sódio ( $\text{Na}^+$ ), magnésio ( $\text{Mg}^{+2}$ ), cálcio ( $\text{Ca}^{+2}$ ), potássio ( $\text{K}^+$ ), carbonato ( $\text{CO}_3^{-2}$ ), cloreto ( $\text{Cl}^-$ ), e bicarbonato ( $\text{HCO}_3^-$ ), os quais foram utilizados para: i) determinação de índices de salinidade associados à água subterrânea para fins de irrigação (Tabela 3.1) e ii) identificação da aptidão deste recurso para fins de irrigação agrícola. A tratamento dos dados e a realização dos cálculos foram realizados utilizando a linguagem de programação R. Na Tabela 3.1 são reportadas as classificações de cada índice em termos qualitativos.

**Tabela 3.1.** Índices de adequação das águas subterrâneas para fins de irrigação.

Índice	Faixa	Classificação	Equação	Referência
Condutividade Elétrica (CE)	<250	Excelente	Obtido no banco de dados	Wilcon (1955)
	250 -750	Bom		
	750-2000	Permitida		
	2000-3000	Duvidoso		
Carbonato de sódio residual (RSC)	>3000	Inadequado	$RSC = (HCO_3^- + CO_3^{2-}) - (Ca^{2+} + Mg^{2+})$	Enaton (1950)
	>1,25	Bom		
	1,25-2,5	Duvidoso		
Razão de adsorção de sódio (RAS)	>2,5	Inadequado	$RAS = \frac{Na^+}{\sqrt{\frac{(Ca^{2+} + Mg^{2+})}{2}}}$	Ayers and Westcot (1985); Richards (1954)
	<10	Excelente		
	10-18	Bom		
	18-26	Duvidoso		
Porcentagem de sódio (%Na)	>26	Inadequado	$Na\% = \frac{(Na^+ + K^+)}{Ca^{2+} + Mg^{2+} + Na^+ + K^+}$	Wilcox (1955)
	<20	Excelente		
	20-40	Bom		
	40-60	Permitida		
Proporção Kelly (KR)	60-80	Duvidoso	$KR = \frac{Na^+}{Ca^{2+} + Mg^{2+}}$	Kelly (1940)
	>80	Inadequado		
Índice de Qualidade da água subterrânea para fins de irrigação (IQAi)	<1	Adequado	$IQAi = \sum \frac{W_i * Q_i}{W_i}$	Singh <i>et al.</i> (2024); Gantait <i>et al.</i> (2022)
	>1	Inadequado		
	<150	Excelente		
	150-300	Bom		
	300-450	Moderado		
	>450	Inadequado		

A determinação do Índice de Qualidade da água subterrânea para fins de irrigação (IQAi) fez uso da metodologia adaptada de Singh *et al.* (2024), na qual parâmetros físico-químicos e índices apresentados na Tabela 2 foram considerados (TEŞAN, 2023).

O Índice de Qualidade da Água para Irrigação (IQAi) foi calculado a partir das equações (1), (2), (3) e (4), com base na metodologia proposta pela Organização das Nações Unidas para Alimentação e Agricultura – FAO (1985). Esse índice integra múltiplos parâmetros físicos e químicos da água com pesos relativos definidos conforme os critérios estabelecidos pela FAO, permitindo classificar a água em categorias como “excelente”, “boa”, “moderada” e “inadequada” para irrigação (vide classificação apresentada na Tabela 3.1). Para o cálculo do IQAi, foram considerados os seguintes parâmetros: pH, SDT, CE, Ca<sup>2+</sup>, K<sup>+</sup>, Na<sup>+</sup>, Mg<sup>2+</sup>, HCO<sub>3</sub><sup>-</sup>, CO<sub>3</sub><sup>2-</sup>, Cl<sup>-</sup>, SO<sub>4</sub><sup>2-</sup>, RAS, RSC, Na% e KR.

A aplicação desse índice tem sido adotada por diversos autores em diferentes contextos agroambientais, como Gantait *et al.* (2022), Bahrami e Zarei (2023) e Batarseh *et al.* (2021), os quais utilizaram o padrão da FAO para avaliar a aptidão da água para irrigação.

$$IQAi = \sum \frac{W_i * Q_i}{W_i} \quad (1)$$

Q<sub>i</sub> é a classificação de qualidade para o i-ésimo parâmetro de amostra para uma determinada amostra de água.

W<sub>i</sub> é o peso unitário para o i-ésimo parâmetro

$$W_i = \frac{V_a - V_i}{V_s - V_i} \quad (A) \quad W_i = \frac{K}{S_i} \quad (B) \quad K = \frac{1}{\sum_{i=1}^n \frac{1}{S_i}} \quad (C) \quad (2,3,4)$$

V<sub>a</sub> = Valor real do i-ésimo parâmetro de uma determinada amostra de água.

V<sub>i</sub> = Valor ideal para o i-ésimo parâmetro (é 0 para todos os parâmetros exceto pH para o qual é 7,0)

K = Constante relativa.

S<sub>i</sub> = Valor padrão do i-ésimo parâmetro conforme FAO (1985).

**Tabela 3.2.** Padrão de qualidade da água de irrigação pela FAO e peso unitário dos parâmetros para IQAi.

Parâmetros	Unidade	FAO (1985) (Si)	Peso unitário (Wi)
CO <sub>3</sub> <sup>2-</sup>	meqL <sup>-1</sup>	60	0,0062
HCO <sub>3</sub> <sup>3-</sup>	meqL <sup>-1</sup>	610	0,0006
Cl <sup>-</sup>	meqL <sup>-1</sup>	1065	0,0003
CE	μS/m	1000	0,0004
SDT	meqL <sup>-1</sup>	1000	0,0004
Ca <sup>2+</sup>	meqL <sup>-1</sup>	400	0,0009
pH	-	8,5	0,0437
Mg <sup>2+</sup>	meqL <sup>-1</sup>	60	0,0062
Na <sup>+</sup>	meqL <sup>-1</sup>	920	0,0004
K <sup>+</sup>	meqL <sup>-1</sup>	2	0,1856
SO <sub>4</sub> <sup>2-</sup>	meqL <sup>-1</sup>	1920	0,0002
Dureza total (como CaCO <sub>3</sub> )	meqL <sup>-1</sup>	712	0,0005
RSC	meqL <sup>-1</sup>	1,5	0,2474
RAS	-	3	0,1237
Na %	%	30	0,0124
KR	-	1	0,3711
ΣWi			1,0000

### 3.2.5 Análise estatística multivariada

Com o intuito de compreender os padrões latentes e as inter-relações entre os parâmetros físico-químicos da água subterrânea, aplicaram-se duas técnicas estatísticas multivariadas: a Análise de Componentes Principais (PCA) e a Análise de Agrupamento Hierárquico (HCA). Essas abordagens foram empregadas para reduzir a dimensionalidade dos dados, identificar grupos homogêneos de amostras e revelar possíveis origens comuns de salinização e variabilidade geoquímica. A PCA foi realizada com rotação ortogonal do tipo Varimax, buscando maximizar a interpretabilidade dos componentes extraídos. Essa técnica permitiu isolar os fatores dominantes responsáveis pela variabilidade da qualidade da água subterrânea, com foco especial nos parâmetros de salinidade utilizados para fins de irrigação. Apenas os componentes com autovalores superiores a 1 foram considerados significativos, conforme o critério de Kaiser. A análise de agrupamento hierárquico (HCA), por sua vez, foi

aplicada no *Rstudio*, considerando os parâmetros físico-químicos como variáveis de agrupamento.

Para a formação dos grupos, utilizou-se o método de ligação de Ward, reconhecido por sua eficiência na minimização da variância intra-grupo. Previamente à análise, os dados foram normalizados por meio da padronização do escore-z, de modo a eliminar o efeito de diferentes escalas e unidades de medida entre as variáveis. O dendrograma resultante auxiliou na visualização das similaridades entre os parâmetros, contribuindo para a identificação de conjuntos de indicadores de qualidade da água com comportamento semelhante ao longo do espaço e do tempo ( Rao *et al.*, 2019 ; UNIGWE *et al.*, 2022 ).

### **3.2.6 Análise espaço-temporal dos índices de qualidade da água para fins de irrigação**

Para delinear a distribuição espacial dos parâmetros da água subterrânea, utilizou-se a técnica de interpolação *Inverse Distance Weighted (IDW)* (TAYYAB *et al.*, 2023), disponível no módulo de análise espacial do software QGIS. Inicialmente, as coordenadas das estações de amostragem foram compiladas em uma planilha Excel, salvas em formato de valores separados por vírgula (.csv), e importadas no QGIS como uma camada de pontos. Cada ponto de amostragem foi identificado com um código único, armazenado na tabela de atributos, que também contém os valores dos parâmetros estudados. A interpolação IDW foi aplicada para gerar mapas de distribuição espacial desses parâmetros de qualidade da água. Os métodos de interpolação espacial são frequentemente aplicados em diversos campos para estimar valores de uma variável em locais sem medições diretas (BOUDIBI *et al.*, 2024). A geodatabase criada no QGIS possibilitou a geração de mapas detalhados de distribuição dos parâmetros, auxiliando na visualização e análise da área estudada.

### **3.2.7 Previsão da qualidade da água de irrigação**

Com o objetivo de prever índices de qualidade da água subterrânea relacionados à salinidade (Tabela 3.1) e, assim, avaliar sua aptidão para irrigação agrícola, foram aplicados modelos de aprendizado de máquina (ML). A escolha dos algoritmos considerou tanto sua capacidade de modelar relações complexas entre as variáveis quanto seu bom desempenho relatado em estudos semelhantes voltados à predição de parâmetros associados à salinização do solo (GAUTAM *et al.*, 2023).

Os modelos testados foram de Regressão Linear Múltipla (MLR), que é particularmente útil para examinar relacionamentos entre múltiplas variáveis independentes (NISHAT *et al.*, 2025); Rede Neural Artificial (RNA) que possui capacidade de modelar relações não linearidades complexas (NISHAT *et al.*, 2025); e Random Forest (RF), que é uma técnica de aprendizado conjunto que constrói múltiplas árvores de decisão para melhorar a precisão da previsão (KAKAEI *et al.*, 2024). RF é amplamente aplicado pelos pesquisadores devido à sua alta precisão, eficiência, rapidez e apresenta menor suscetibilidade ao overfitting (WANG *et al.* 2024; LI *et al.* 2021).

Para tanto, as variáveis de entrada dos modelos foram: pH, CE,  $Na^+$ ,  $K^+$ ,  $Cl^-$ , e  $HCO_3^-$ . Já os parâmetros-alvo (variáveis de saída) incluíram os índices relacionados à salinidade da água de irrigação (RSC, RAS, %Na, KR, IQAi). A seleção dessas variáveis de entrada baseou-se em sua reconhecida relevância nos processos que controlam a salinidade da água subterrânea utilizada para irrigação. A CE é amplamente utilizada como indicador indireto da concentração total de sais dissolvidos na água, sendo fortemente correlacionada com a salinidade (LI *et al.*, 2021; GANTAIT *et al.*, 2022). Os íons  $Na^+$ ,  $K^+$  e  $Cl^-$  são os principais constituintes dos sais dissolvidos e estão frequentemente associados à salinização dos solos irrigados (AYERS e WESTCOT, 1985; WANG *et al.*, 2024). Já o íon  $HCO_3^-$ , embora não seja um sal em si, contribui para a alcalinidade da água e influencia diretamente índices como o RSC e a precipitação de carbonatos, afetando a disponibilidade de cálcio e magnésio no solo (BAHRAMI e ZAREI, 2023). Por fim, o pH é um parâmetro fundamental que modula as reações químicas e a estabilidade dos íons em solução, influenciando a solubilidade dos sais e a formação de precipitados. Assim, a escolha desses parâmetros reflete tanto a base conceitual do processo de salinização quanto o uso consolidado na literatura científica voltada à predição da qualidade da água para irrigação. Os modelos foram, então, validados e calibrados a partir do uso do conjunto de dados na proporção 70% para treinamento e 30% para teste (GORGAN-MOHAMMADIE *et al.*, 2023). O desempenho dos modelos foi avaliado a partir de duas métricas de desempenho: o coeficiente de determinação ( $R^2$ ) e a raiz do erro quadrático médio (RMSE).

O valor de  $R^2$  quantifica a porcentagem de variabilidade em uma variável considerada pelo modelo, sendo que o valor varia de 0 a 1, onde o valor 0 indica que o modelo tem um desempenho ruim na previsão do resultado, e o valor de 1 revela que o modelo prevê o resultado com precisão. Enquanto que o RMSE fornece medidas de precisão de previsão ao quantificar a diferença entre valores observados e previstos (MAHDIAN *et al.*, 2023). As Equações 5, 6 e 7 descrevem o cálculo dos coeficientes utilizados. Chicco, Warrens e Jurman (2021) reportam

que RMSE, uma métrica popular de avaliação de regressão, calcula a raiz quadrada da média das diferenças quadradas entre os valores previstos e verdadeiros fornecendo uma medida do erro típico nas previsões do modelo. Já o coeficiente de determinação ( $R^2$ ) indica a proporção de variância na variável dependente (alvo) que pode ser explicada pelas variáveis independentes (características) no modelo (CHICCO, WARRENS e JURMAN, 2021).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5) \quad MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (6) \quad RMSE = \sqrt{MSE}$$

(7)

Em que:

$y_i$  são os valores reais (monitorados).

$\hat{y}_i$  são os valores previstos pelo modelo.

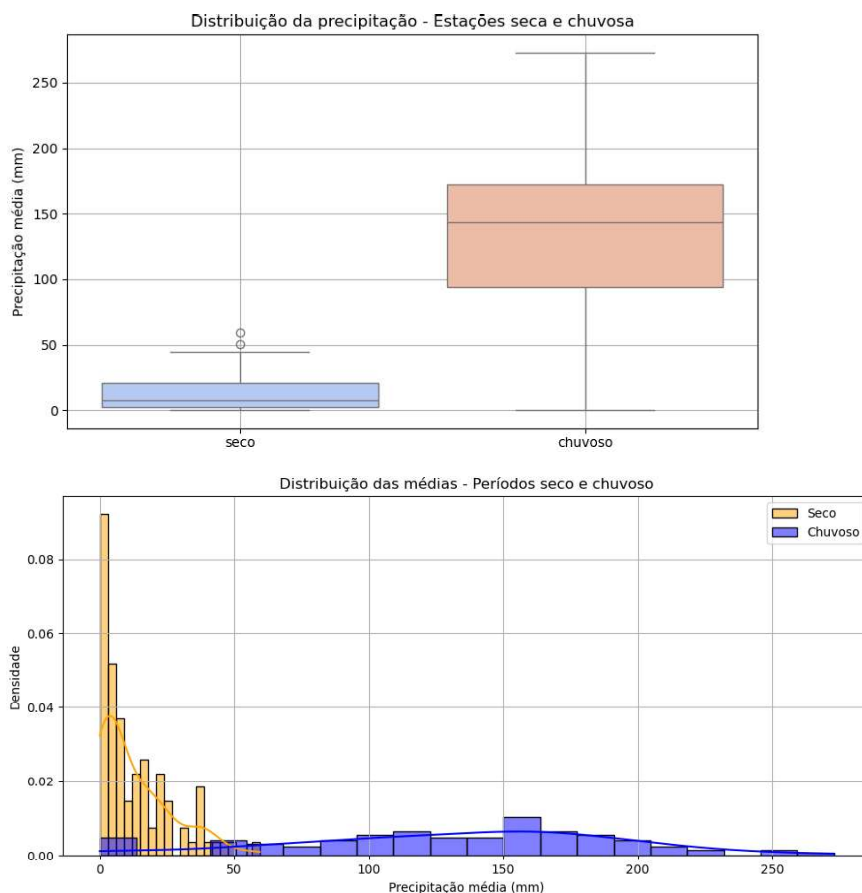
$\bar{y}$  é a média dos valores reais.

$n$  é o número de amostras

### 3.3 Resultados e Discussão

#### 3.3.1 Análise comparativa sazonal da pluviosidade

Na Figura 3.2 apresenta-se a distribuição da precipitação para os períodos seco e chuvoso, evidenciando maior dispersão de valores no período chuvoso (*boxplot*, a) e padrões de distribuição distintos entre os períodos (histograma de densidade, b). A comparação entre os períodos seco e chuvoso revelou uma diferença estatisticamente significativa nos valores de precipitação ( $p < 0,001$ ), conforme demonstrado pelo teste t de *Student*. A média de precipitação no período seco foi de 12,95 mm, enquanto no período chuvoso atingiu 132,18 mm, com desvios padrão de 13,72 mm e 60,90 mm, respectivamente. O valor elevado de t (-18,22) e o p-valor extremamente baixo ( $2,08 \times 10^{-33}$ ) evidenciam a presença de uma sazonalidade pluviométrica fortemente marcada, característica típica do semiárido mineiro (dos SANTOS *et al.*, 2024). Essa distinção também é visível nos gráficos de distribuição.



**Figura 3.2.** Distribuição da precipitação nos períodos seco e chuvoso no norte de Minas Gerais: (a) boxplot mostrando a variação e dispersão dos dados; (b) histograma de densidade evidenciando diferenças no padrão de distribuição entre os períodos

Esses resultados justificam a segmentação temporal das análises da qualidade da água subterrânea em função dos períodos do ano, considerando que a pluviosidade pode influenciar diretamente os processos de recarga, diluição de sais e mobilização geoquímica no aquífero. Estudos anteriores já apontam que em regiões semiáridas, a qualidade das águas subterrâneas depende significativamente da precipitação, sendo comum a ocorrência de alterações na composição iônica após a estação chuvosa (SINGH *et al.*, 2021).

### 3.3.2 Características físico-químicas das águas subterrâneas

Na Tabela 3.3 apresenta-se a estatística descritiva dos resultados dos parâmetros físico-químicos da água subterrânea na área de estudo, organizados em período seco e chuvoso. Optou-se por manter todos os valores observados uma vez que estes refletem condições naturais e pontuais do sistema aquífero e, portanto, são relevantes para a caracterização hidroquímica da região. Na área de estudo, o pH médio das águas subterrâneas foi levemente alcalino em

ambos os períodos, com valores de 7,29 e 7,26, para os períodos seco e chuvoso, respectivamente. Valores próximos da neutralidade e levemente alcalinos são comuns em ambientes onde há contato prolongado da água com minerais carbonáticos (LEI *et al.*, 2023; ZHANG *et al.*, 2019). Pequenos aumentos no pH durante períodos pós-precipitação também podem estar associados à maior interação entre a água infiltrada e a matriz do solo (SUBRAMANIAN e SAXENA, 1983).

A CE média das águas subterrâneas foi de 855,63  $\mu\text{S}/\text{cm}$  no período seco e 813,53  $\mu\text{S}/\text{cm}$  no período chuvoso, o que pode indicar diluição sazonal, ainda que a dispersão dos dados ( $\text{DP} > 500 \mu\text{S}/\text{cm}$ ) revele uma ampla variabilidade espacial. A CE é um indicador da concentração total de íons dissolvidos na água e é influenciada por fatores como evapotranspiração, recarga hídrica e atividades antrópicas locais (SOLANGI *et al.*, 2019), no presente estudo, os resultados foram compatíveis com os observados por BOUNAB *et al.* (2022) (463 e 6844  $\mu\text{S}/\text{cm}$ ) onde a água subterrânea é utilizada para agrícola intensiva.

**Tabela 3.3.** Resultados dos parâmetros físico-químicos da água subterrânea na área de estudo.

Estatística		CE	HCO <sub>3</sub>	Mg	Ca	K	Na	Cl	Dureza	SDT	pH	SO <sub>4</sub>	Pluviosidade
		μS/cm	meq/L	meq/L	meq/L	meq/L	meq/L	meq/L	meq/L	meq/L	mg/L	-	meq/L
n	seco	314	314	314	314	314	314	314	314	314	314	314	91
	chuvoso	358	358	358	358	358	358	358	358	358	358	358	91
Média	seco	855,6	4,29	1,26	5,26	0,08	1,77	59,74	337,12	492,88	7,29	48,31	12,95
	chuvoso	813,5	4,27	1,27	4,89	0,09	1,71	61,99	328,96	484,73	7,26	51,24	132,18
DP	seco	585,4	1,54	1,18	3,33	0,11	2,31	92,18	210,24	365,27	0,46	89,71	13,72
	chuvoso	534,06	1,49	1,23	3,16	0,12	2,14	102,96	212,28	351,68	0,51	97,25	60,90
Percentil 10	seco	302,0	2,26	0,18	1,36	0,02	0,07	0,5	100,2	172,2	6,8	5	0
	chuvoso	28707	2,34	0,17	1,33	0,02	0,17	0,5	96,41	168,7	6,74	5	51,73
Percentil 90	Seco	1565,2	6,05	3,01	8,94	0,18	4,52	172,32	588,86	921,6	7,8	124,36	36,48
	chuvoso	1431,1	6,08	3,03	8,46	0,18	4,25	183,74	597,78	908,2	8	149,6	203,50

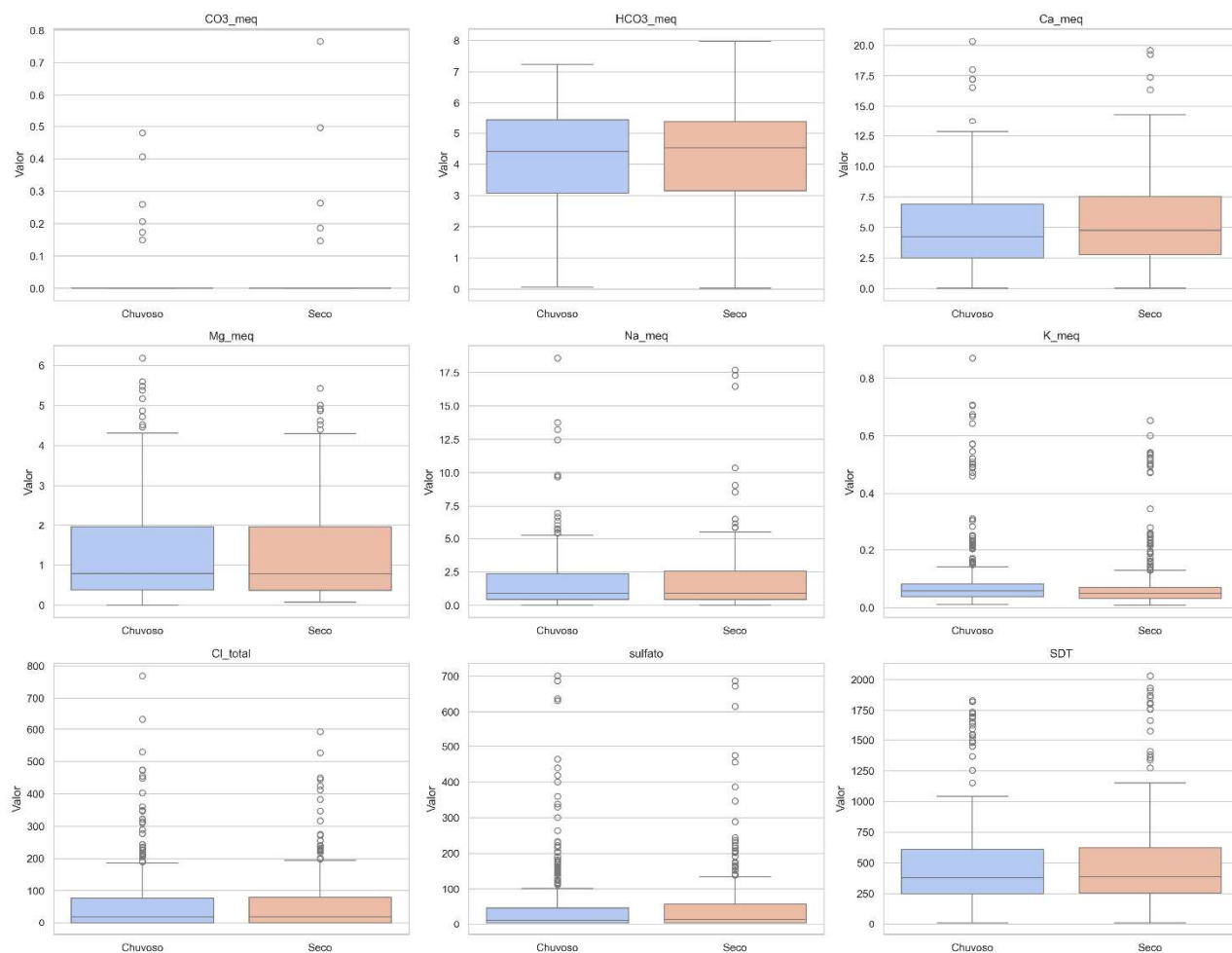
CE: Condutividade Elétrica (μS/cm); HCO<sub>3</sub>: Bicarbonato (meq/L); Mg: Magnésio (meq/L); Ca: Cálcio (meq/L); K: Potássio (meq/L); Na: Sódio (meq/L); Cl: Cloreto (meq/L); Dureza: Dureza Total (meq/L); SDT: Sólidos Dissolvidos Totais (mg/L); pH: Potencial Hidrogeniônico (sem unidade); SO<sub>4</sub>: Sulfato (meq/L); n (número de amostras), DP (Desvio Padrão).

O teor de sólidos dissolvidos totais (SDT) foi semelhante entre os períodos, com valores médios de 492,88 mg/L (seco) e 484,73 mg/L (chuvoso). A concentração de SDT é influenciada pela mineralização natural da água e pelos processos de lixiviação promovidos pelas chuvas, que podem favorecer o transporte de sais para o lençol freático (SHARMA *et al.*, 2017; SMAIL e DIŞLI, 2023).

Quanto à dureza total, observou-se leve redução no valor médio entre os períodos (337,12 meq/L no seco e 328,96 meq/L no chuvoso). A dureza está diretamente associada à presença de íons cálcio e magnésio, cuja origem está ligada à dissolução de minerais presentes na rocha matriz. Segundo Sharma *et al.* (2017), a variação na dureza pode refletir o tempo de residência da água no aquífero e a maior solubilização de carbonatos após períodos chuvosos.

A concentração de sódio ( $\text{Na}^+$ ) variou de forma sutil entre os períodos (1,77 meq/L (seco) e 1,71 meq/L (chuvoso), com distribuições semelhantes conforme observado na Figura 3.3. Essa estabilidade sugere que os aportes de  $\text{Na}^+$  podem estar mais relacionados a fontes localizadas ou ao rebaixamento do nível freático durante períodos de extração intensa, o que favorece a concentração de sais residuais (SINGH *et al.*, 2021; DEY *et al.*, 2023).

Comparação entre Chuvoso e Seco - Parâmetros



**Figura 3.3.** *Boxplots* comparativos dos parâmetros físico-químicos da água subterrânea entre os períodos chuvoso e seco no norte de Minas Gerais.

O cálcio ( $\text{Ca}^{2+}$ ), apesar da média maior no período seco, apresentou distribuição semelhante entre os períodos, conforme evidenciado na Figura 3. A dissolução de carbonatos, intensificada pela formação de ácido carbônico durante as chuvas, pode explicar o aumento pontual de  $\text{Ca}^{2+}$  após eventos chuvosos (SAHA e RAY, 2019). O magnésio ( $\text{Mg}^{2+}$ ) manteve-se praticamente estável (1,26 meq/L no seco e 1,27 meq/L no chuvoso), indicando controle geoquímico mais consistente e menor sensibilidade à sazonalidade.

O bicarbonato ( $\text{HCO}_3^-$ ) também se mostrou estável entre os períodos, com médias de 4,29 e 4,27 meq/L, respectivamente. Essa consistência está relacionada à atuação do intemperismo de rochas carbonáticas e à atividade microbológica no solo, fontes dominantes desse ânion (PATEL *et al.*, 2020; SINGH *et al.*, 2022). O cloreto ( $\text{Cl}^-$ ), por outro lado, apresentou um leve aumento médio no período chuvoso (de 59,74 para 61,99 meq/L),

acompanhado de ampla variação e presença de outliers, como evidenciado na Figura 3.2. Esse comportamento pode ser atribuído à lixiviação superficial e à percolação de sais após eventos de precipitação mais intensos (SUBBA RAO *et al.*, 2022; MANIKANDAN *et al.*, 2020). A concentração de sulfato ( $\text{SO}_4^{2-}$ ) também aumentou ligeiramente no período chuvoso (48,31 para 51,24 meq/L), o que pode estar associado à entrada de águas superficiais carregadas de partículas e ao uso de fertilizantes contendo enxofre, os quais têm sido identificados como fontes relevantes desse ânion em aquíferos agrícolas (YU *et al.*, 2024).

A comparação gráfica entre os períodos seco e chuvoso (Figura 3.3) indica estabilidade nas medianas para a maioria dos parâmetros físico-químicos, com diferenças visuais pouco expressivas. Essa relativa homogeneidade sazonal é coerente com estudos realizados em outros ambientes semiáridos, como os de Subba Rao *et al.* (2022) e Manikandan *et al.* (2020), que também não identificaram influência marcante da pluviosidade sobre as concentrações iônicas da água subterrânea. Esses autores atribuíram tal estabilidade à recarga rápida e localizada, à predominância de fontes geoquímicas estáveis, como intemperismo de minerais carbonáticos e halita, e à influência antrópica contínua, fatores que mantêm a composição química relativamente constante ao longo do ano.

Com a inclusão dos dados de pluviosidade, observa-se que, embora a sazonalidade hídrica seja marcante no semiárido, as diferenças nas médias dos parâmetros físico-químicos entre os períodos seco e chuvoso foram pouco expressivas. O aumento da precipitação no período chuvoso (confirmado pelo teste t com  $p < 0,001$ ) não se traduziu em mudanças significativas na composição iônica, possivelmente devido ao predomínio de controles geoquímicos estáveis e de fontes antrópicas persistentes. Resultados semelhantes foram relatados por Subba Rao *et al.* (2022) e Manikandan *et al.* (2020), indicando que, em muitos aquíferos de regiões semiáridas, a qualidade da água subterrânea mantém-se relativamente constante ao longo do ano. Ainda assim, a sazonalidade climática deve ser registrada e monitorada, pois pode influenciar processos pontuais de recarga e transporte de sais, especialmente em anos atípicos de precipitação.

### **3.3.3 Características dos índices de adequação da água de irrigação para diferentes estações do ano**

A Tabela 3.4 apresenta a estatística descritiva dos índices de qualidade da água subterrânea de interesse para fins de irrigação na região semiárida de Minas Gerais. De modo geral, os valores médios dos índices RSC, RAS, %Na, KR e IQAi apresentaram pequenas

variações entre os períodos seco e chuvoso.

O RSC apresentou médias negativas em ambos os períodos, com valores de -2,22 no seco e -1,88 no chuvoso, indicando que a água apresenta baixo potencial de acúmulo excessivo de sódio no solo.

No entanto, o ligeiro aumento no valor médio durante o período chuvoso, pode indicar maior mobilização de bicarbonatos e carbonatos na zona vadosa devido à recarga hídrica, como relatado em investigações anteriores em regiões semiáridas (GUGULOTHU *et al.*, 2022; RAGHUNATH, 1987). O índice RAS manteve-se relativamente estável entre os dois períodos, com médias de 1,18 no seco e 1,21 no chuvoso. Esses valores situam-se dentro da faixa considerada ideal para irrigação, demonstrando baixo risco de impacto na estrutura do solo, especialmente em sistemas com solos de textura média a fina, conforme discutido por Ayers e Westcot (1985) e corroborado por Sarkar *et al.* (2022).

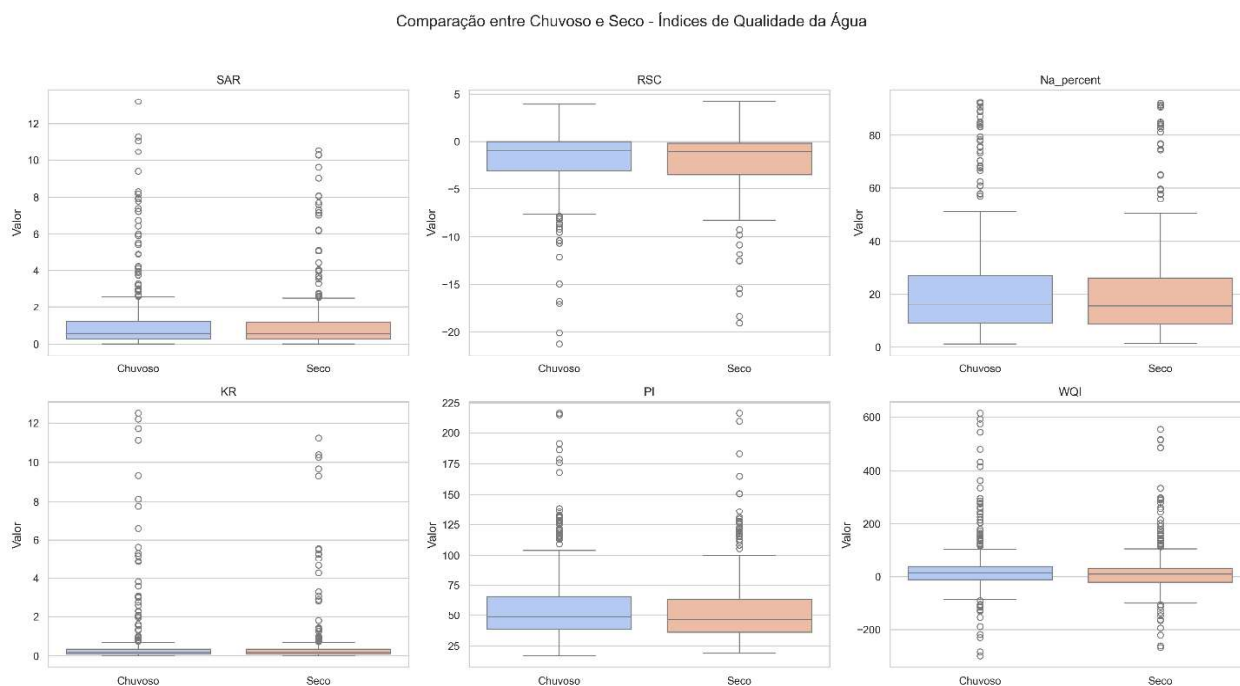
**Tabela 3.4.** Estatísticas dos índices de águas subterrâneas para irrigação

Estatística		RS	RAS	%Na	KR	IQAi
n	seco	31	314	314	314	314
	chuvoso	35	358	358	358	358
Média	seco	1,18	1,18	21,16	0,56	21,53
	chuvoso	1,21	1,21	22,13	0,62	32,78
DP	seco	3,3	1,83	19,91	1,51	99,84
	chuvoso	3,3	1,90	20,58	1,66	106,40
Percentil 10	seco	0,1	0,1	4,51	0,04	49,24
	chuvoso	0,15	0,15	4,98	0,04	40,99
Percentil 90	Seco	2,56	2,56	46,75	0,85	113,43
	chuvoso	2,69	2,69	49,16	0,92	144,06

KR: Proporção de Kelly; RAS: Razão de Adsorção de Sódio; %Na: Porcentagem de Sódio (%); RSC: Carbonato de Sódio Residual (meq/L); IQAi: Índice de Qualidade da Água para fins de irrigação.

Em relação à porcentagem de sódio (%Na), observou-se um pequeno aumento entre os períodos, com a média variando de 21,16% no seco para 22,13% no chuvoso (Figura 3.4). O aumento do índice pode estar relacionado à maior lixiviação de sais durante as chuvas, que pode elevar temporariamente a concentração de sódio na água subterrânea, efeito já documentado

em outras regiões semiáridas (SANTACRUZ-De LEÓN *et al.*, 2022).



**Figura 3.4.** *Boxplot* comparativo Índices de adequação das águas subterrâneas para fins de irrigação para o período seco e chuvoso

A Proporção de Kelly (KR) apresentou média inferior a 1 nos dois períodos, indicando adequação da água para irrigação segundo os critérios estabelecidos por Kelly (1940). Contudo, a elevação do valor médio do índice de 0,56 para 0,62 sugere que, em parte das amostras, o sódio já ultrapassa os teores de cálcio e magnésio, o que pode favorecer a sodificação progressiva, especialmente em ambientes com baixa taxa de infiltração. Esse padrão já foi observado em estudos conduzidos na Índia e na Namíbia, os quais alertam para os efeitos cumulativos da irrigação com águas levemente alcalinas em solos suscetíveis (AMWELE *et al.*, 2021; BHAT *et al.*, 2018).

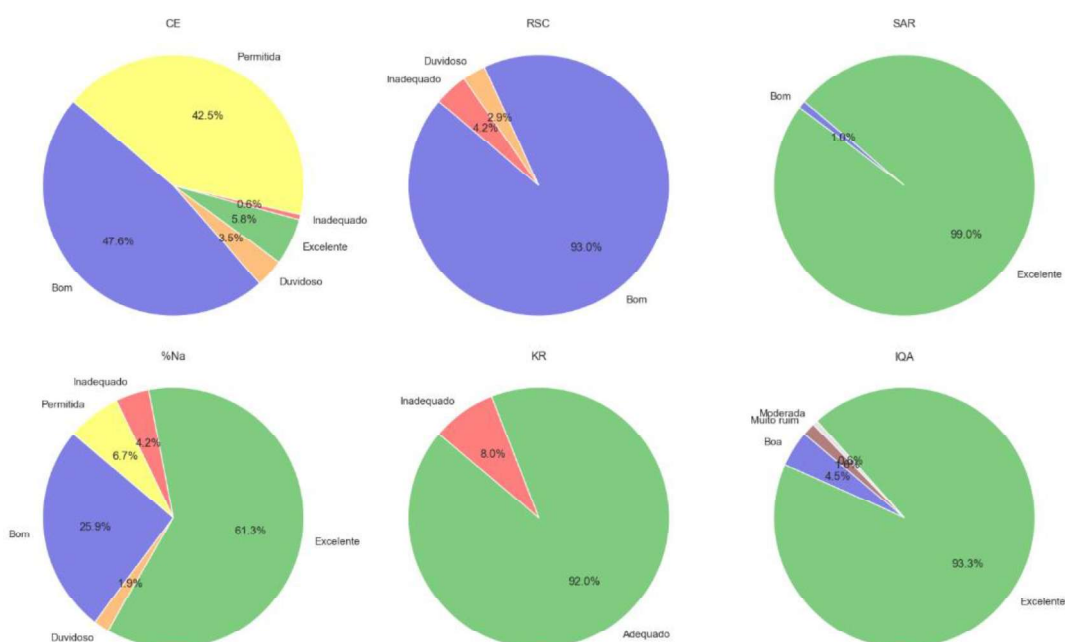
Por fim, o Índice de Qualidade da Água Subterrânea para fins de irrigação (IQAi), que resume a adequação hídrica com base em múltiplos parâmetros, apresentou variação mais expressiva entre os períodos. O IQAi teve um acréscimo de 21,53 (período seco) para 32,78 (período chuvoso), indicando uma melhoria na qualidade da água subterrânea para fins de irrigação. Conclusão semelhante foi reportado por Dey *et al.* (2023) e Gao *et al.* (2023), os quais apontaram heterogeneidade espacial e temporal da qualidade da água subterrânea em regiões semiáridas, especialmente após eventos de recarga.

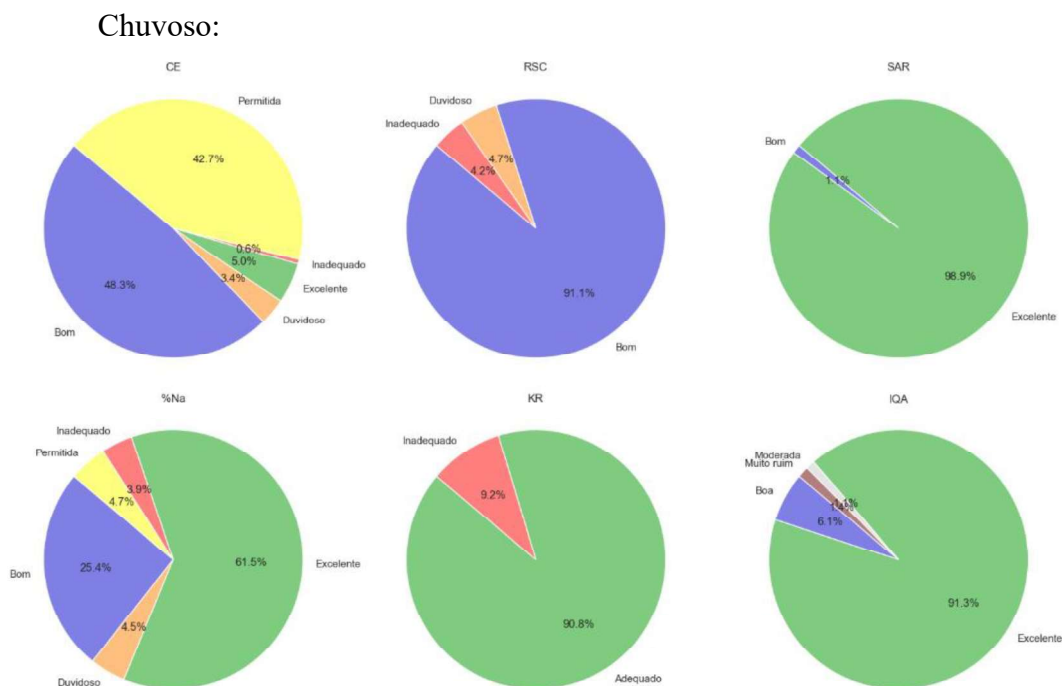
### 3.3.4 Avaliação da água de irrigação usando a abordagem multiparamétrica

A classificação da qualidade das águas subterrâneas para fins de irrigação utilizando diferentes índices é apresentada na Figura 3.5, revelando padrões relativamente estáveis entre os períodos seco e chuvoso, mas com variações que indicam possíveis efeitos sazonais.

A condutividade elétrica, um indicador geral da salinidade, apresentou predominância da classe “Bom” em ambos os períodos, com uma leve elevação de 47,60% para 48,32% no chuvoso. A classe “Permitida” manteve proporção semelhante, também com discreto aumento, enquanto as categorias “Duvidoso” e “Inadequado” representaram menos de 4% do total das amostras, sugerindo baixa ocorrência de salinidade severa. Esses resultados indicam, em geral, uma condição favorável ao uso da água para irrigação quanto à salinidade, mesmo após o período de recarga hídrica, o que pode estar relacionado à diluição sazonal ou à baixa mobilização de sais nas áreas de recarga, como observado em estudos anteriores conduzidos em regiões semiáridas com características semelhantes (BOUNAB *et al.*, 2022; SOLANGI *et al.*, 2019).

Seco:





**Figura 3.5.** Adequação de águas subterrâneas para irrigação utilizando diferentes índices.

Legenda: Condutividade elétrica (CE), Carbonato de sódio residual (RSC), Razão de adsorção de sódio (RAS), Porcentagem de sódio (%Na), Proporção Kelly (KR), Índice de Qualidade da água na irrigação (IQAI).

No que se refere ao Carbonato de Sódio Residual (RSC), mais de 90% das amostras foram classificadas como “Bom” em ambos os períodos, o que demonstra baixa propensão à sodificação dos solos pela presença de carbonatos e bicarbonatos. Houve, no entanto, um ligeiro aumento nas classes “Duvidoso” e “Inadequado” no período chuvoso, o que pode refletir uma maior mobilização desses íons durante o escoamento vertical da água da chuva, como já discutido por Gugulothu *et al.* (2022) e Sharma *et al.* (2017).

A Razão de Adsorção de Sódio (RAS) reforça o cenário positivo, com mais de 98% das amostras classificadas como “Excelente” nos dois períodos. Esse resultado indica que a maioria das amostras de água analisadas apresentam baixa relação entre sódio e íons cálcio e magnésio, minimizando os riscos de dispersão da estrutura do solo quando a água é utilizada para irrigação (AYERS e WESTCOT, 1985; SARKAR *et al.*, 2022).

A distribuição das classes da porcentagem de sódio (%Na), embora centrada na categoria “Excelente” (mais de 61%), apresentou aumento nas classes “Duvidoso” e “Inadequado” no período chuvoso. Esse padrão sugere que, embora a maioria das amostras mantenha boas condições de uso, a recarga hídrica pode mobilizar sais de sódio em algumas áreas específicas, elevando pontualmente a salinidade relativa desse íon. Esse comportamento

heterogêneo é típico de sistemas aquíferos em regiões semiáridas com uso agrícola intensivo (SINGH *et al.*, 2021; SANTACRUZ-De LEÓN *et al.*, 2022).

A Proporção de Kelly (KR), outro indicador de sodificação, indicou que mais de 90% das amostras são adequadas para irrigação em ambos os períodos, com pequeno aumento na proporção de amostras classificadas como “Inadequadas” no chuvoso. Essa elevação, embora modesta, pode refletir um acúmulo temporário de sódio em detrimento do cálcio e magnésio, especialmente em zonas com recarga localizada e limitada drenagem (AMWELE *et al.*, 2021; BHAT *et al.*, 2018).

O IQAi, que sintetiza múltiplos parâmetros, também apresentou estabilidade entre os períodos, com predominância da classe “Excelente” (acima de 91%). No entanto, verificou-se ligeiro aumento da classe “Boa” de 4,47% para 6,15% no chuvoso, além da duplicação da frequência de amostras “Moderadas”, o que aponta para uma maior dispersão dos resultados e reforça a necessidade de abordagens locais na gestão da irrigação. Essa variabilidade pode estar associada à influência combinada da recarga sazonal, do uso agrícola e das características hidrogeológicas locais, conforme já observado por Gao *et al.* (2023) e Dey *et al.* (2023).

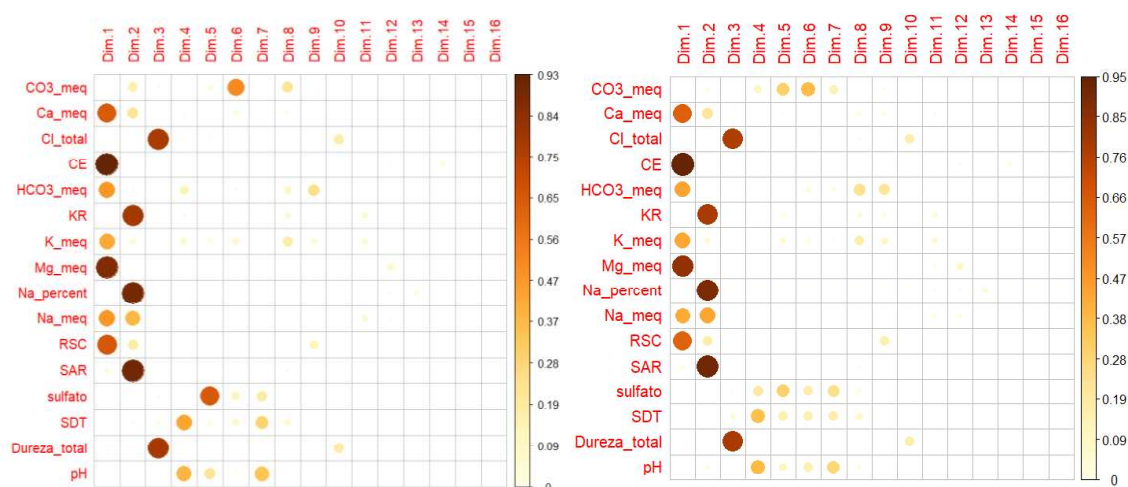
### **3.3.5 Análise estatística multivariada**

A análise de componentes principais (PCA) resultou na extração de cinco componentes principais para o período seco e cinco para o período chuvoso, todos com autovalores superiores a 1, uma abordagem amplamente utilizada para seleção de componentes significativos (Kaiser, 1960). Essa regra assume que componentes com autovalores superiores a 1 explicam mais variância do que uma variável original padronizada e, portanto, justificam ser retidos para análise posterior (JOLLIFFE e CADIMA, 2016). No período seco, as cinco primeiras dimensões explicaram aproximadamente 75% da variância total, enquanto no chuvoso explicaram aproximadamente 73% da variância total dos dados (Tabela 3.5).

**Tabela 3.5.** Autovalor e porcentagem da variância acumulada da PCA

Dimensão	Período	Autovalor	Variância	%variância acumulada
Dim 1	seco	4,59	28,73	28,73
	chuv	4,43	27,70	27,70
	seco	3,63	22,70	51,43
Dim 2	chuv	3,59	24,26	50,13
	seco	1,66	10,35	61,78
Dim 3	chuv	1,68	10,48	60,62
Dim 4	seco	1,15	7,22	68,99
	chuv	1,07	6,69	67,31
Dim 5	Seco	1,06	6,62	75,62
	chuv	1,03	6,42	73,73

A Dimensão 1 (PC1) é a mais expressiva em ambos os períodos, com 28,73% e 27,70% da variância, respectivamente, e apresentou composição muito semelhante entre os períodos seco e chuvoso, com cargas elevadas para CE, Mg, Ca, Na e K, sugerindo que os principais fatores relacionados à salinidade e dureza da água subterrânea mantiveram-se constantes ao longo das estações conforme ilustrado na Figura 3.6. Isso reforça a hipótese de que tais características são controladas por fatores geológicos e hidrogeológicos locais, mais do que pela sazonalidade da pluviosidade (GUO *et al.*, 2021; SUNDARAY *et al.*, 2009).



**Figura 3.6.** Correlograma da qualidade da representação ( $\cos^2$ ) das variáveis nas duas primeiras componentes principais para os períodos seco e chuvoso

Este agrupamento de cargas elevadas negativas sugere uma forte associação entre condutividade elétrica e os íons principais de dureza e salinidade, indicando um possível controle geoquímico conjunto, como processos de intemperismo de rochas e dissolução de sais, além da influência de atividades antrópicas, como irrigação e uso de fertilizantes (SUNDARAY *et al.*, 2009).

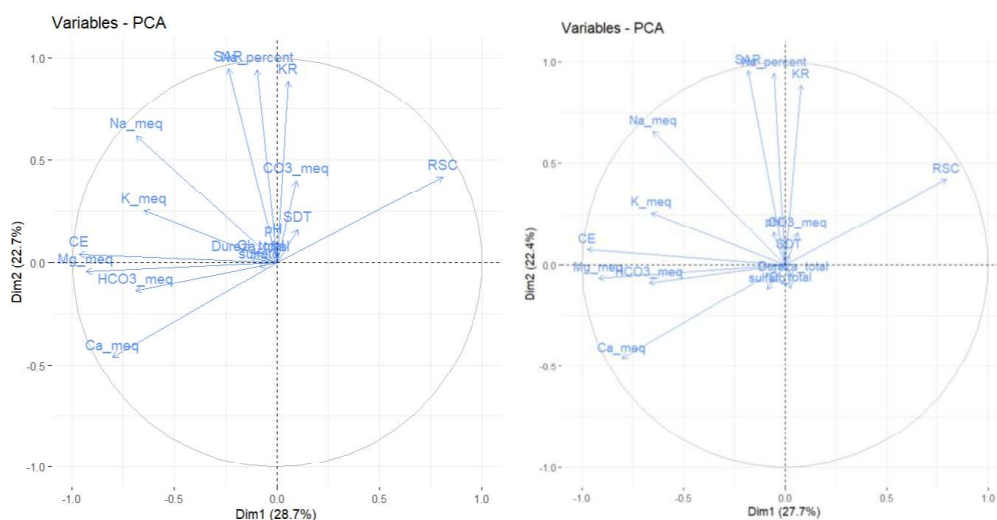
A Dimensão 2 (PC2) do período chuvoso explicou aproximadamente 24% da variância no período seco e 22,7% no chuvoso, sendo fortemente influenciada por %Na ( $\approx 0,94$ ), RAS ( $\approx 0,95$ ) e KR ( $\approx 0,89$ ) em ambos os casos. Essa combinação de indicadores é amplamente empregada na avaliação da aptidão da água para irrigação e na avaliação do risco de desequilíbrio da estrutura do solo, altas concentrações relativas de sódio podem levar à dispersão de partículas argilosas, reduzindo a infiltração de água e a aeração do solo (AYERS e WESTCOT, 1999; RICHARDS, 1954). Os elevados valores desses índices na PC2 podendo refletir a presença persistente de sódio na água subterrânea, mesmo durante o período chuvoso, possivelmente advindo de mobilização superficial ou de práticas agrícolas.

Essa combinação de indicadores é comumente utilizada para avaliar a aptidão da água para irrigação e possíveis riscos à estrutura do solo, uma vez que concentrações elevadas de sódio em relação a cálcio e magnésio podem comprometer a infiltração e a aeração do solo devido à dispersão de partículas argilosas (AYERS e WESTCOT, 1999; RICHARDS, 1954). Assim, os altos valores desses índices na PC2 podem sugerir uma influência antrópica ou geológica que eleva o teor relativo de sódio na água durante o período chuvoso.

No período seco, os coeficientes de PC2 para esses índices foram muito semelhantes aos do chuvoso, indicando que a sazonalidade da pluviosidade não alterou o perfil relativo desses indicadores da razão de adsorção de sódio na água subterrânea, como RAS, %Na e KR. Isso sugere que, nesta região do norte de Minas, a composição iônica da água subterrânea, especialmente os parâmetros de sódio, é controlada mais pelo tipo de solo (solos cristalinos de baixa capacidade tampão) e pela dinâmica de recarga limitada do aquífero do que pela variação sazonal da chuva, fenômeno também observado em outros estudos de aquíferos fraturados em zonas semiáridas (GUO *et al.*, 2021).

A Figura 3.7 apresenta o biplot das variáveis para o período seco e chuvoso, respectivamente. O comprimento das setas representa a qualidade de representação das variáveis no plano das duas primeiras componentes, quanto maior a seta, maior a contribuição daquela variável para a estrutura dos dados e mais bem representada ela está no espaço bidimensional. A orientação das setas também é relevante: variáveis com vetores na mesma

direção possuem correlação positiva, enquanto vetores opostos indicam correlação negativa. Nesse sentido, CE, Mg, Ca, Na e HCO apresentam as maiores setas, indicando forte influência na PC1. Já %Na, RAS e KR apresentam vetores direcionados para o quadrante superior, com ângulos relativamente próximos entre si, refletindo correlação positiva e associação predominante com a PC2. Embora os vetores de %Na, RAS e KR não estejam perfeitamente alinhados, sua direção geral e magnitude permitem afirmar que essas variáveis compartilham estrutura comum, provavelmente relacionada ao balanço de sódio na água subterrânea. A CE está fortemente alinhada a Mg, sugerindo que a condutividade elétrica está mais associada à presença de íons salinos em geral, sobretudo aqueles resultantes de processos de intemperismo ou dissolução de minerais. Em contraste, CE e RSC apontam para direções opostas, indicando correlação negativa entre essas variáveis nesse período.



**Figura 3.7.** Biplot das variáveis da PCA para o período seco e chuvoso respectivamente.

No período seco, observa-se um pequeno aumento no comprimento das setas dessas variáveis, principalmente para o  $\text{CO}_3$ , que se desloca para o quadrante superior direito, evidenciando uma contribuição positiva tanto em PC1 quanto em PC2. Já no período chuvoso, o SDT e o  $\text{CO}_3$  possuem vetores curtos e orientados de forma discreta para o centro, indicando fraca influência sobre as duas primeiras dimensões. Essa mudança sugere uma maior influência relativa de carbonatos e sólidos dissolvidos totais durante o período seco, possivelmente associada à redução na recarga hídrica e menor diluição da água subterrânea nesse intervalo. Apesar da estrutura global da variância explicada pelas componentes ter se mantido semelhante entre os períodos, o aumento da representação dessas variáveis pode indicar condições geoquímicas pontuais mais concentradas, como maior dissolução de minerais ou menor

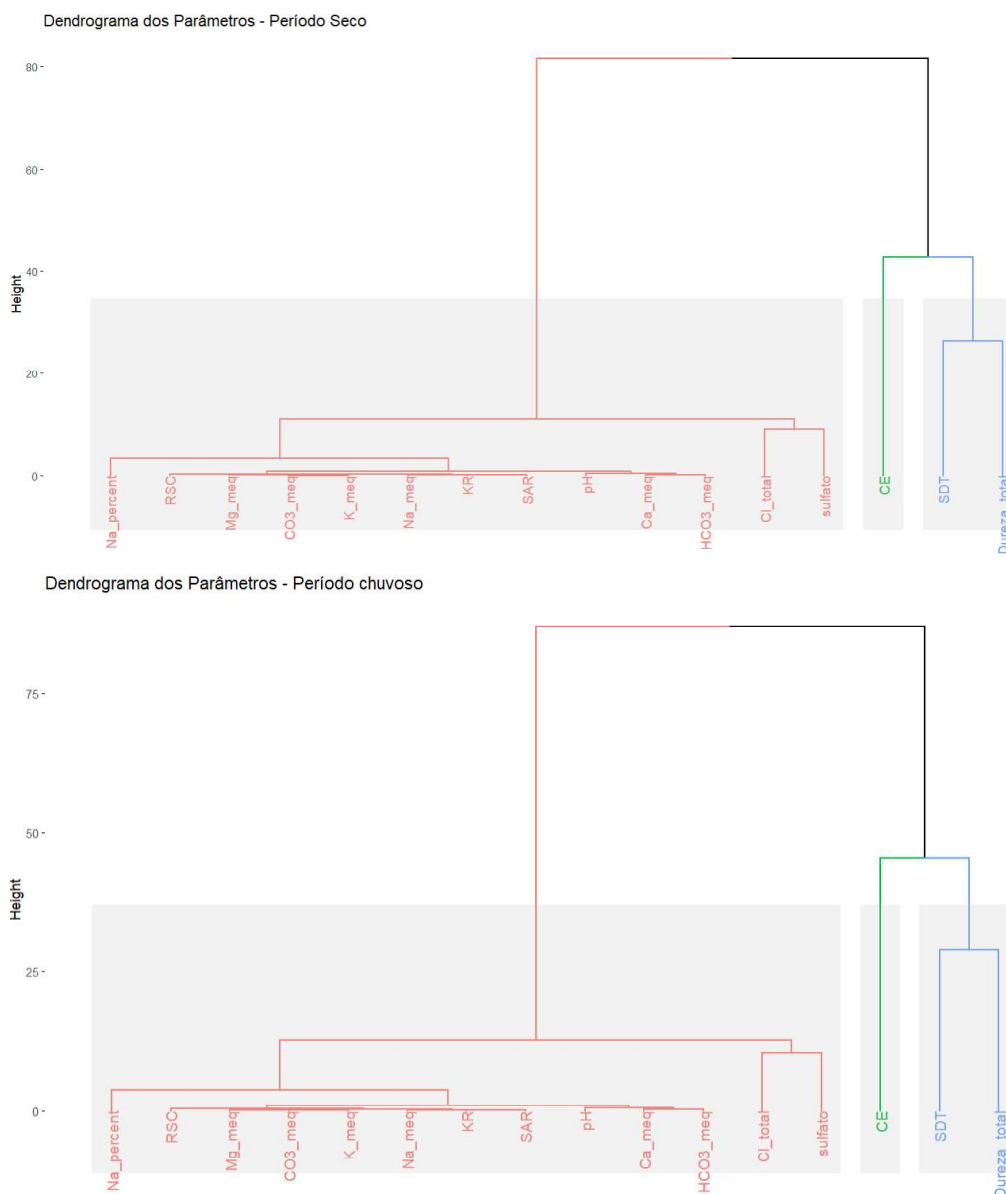
renovação da água nos aquíferos rasos durante a estação seca, padrão também observado em outras regiões semiáridas com aquíferos fraturados (GUO *et al.*, 2021).

A Análise de Agrupamento, representada pelo dendrograma (Figura 3.8), organizou as variáveis em três grupos principais bem definidos em ambos os períodos avaliados (seco e chuvoso), com estrutura semelhante, com base na similaridade de seus comportamentos estatísticos. Essa segmentação foi obtida utilizando o método de ligação de Ward com padronização prévia (escore z), o que permitiu minimizar a variância intra-grupo.

O primeiro cluster (em vermelho), o mais numeroso nos dois períodos, agrupa variáveis diretamente associadas aos processos de salinização e sodificação da água subterrânea, incluindo  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Mg}^{2+}$ ,  $\text{Ca}^{2+}$ ,  $\text{Cl}^-$ ,  $\text{SO}_4^{2-}$ , RAS, RSC, KR, %Na e pH. Essa associação indica uma dinâmica geoquímica típica de ambientes semiáridos, marcada pelo acúmulo de sais e pelo risco de degradação estrutural dos solos irrigados. A proximidade entre os índices RAS e KR reforça que ambos capturam aspectos complementares do risco de sodificação, padrão também destacado por Singh *et al.* (2024). O agrupamento conjunto dessas variáveis foi consistente com os resultados da PCA, na qual apresentaram cargas fatoriais elevadas na Dimensão 1, refletindo seu papel dominante na variabilidade química da água.

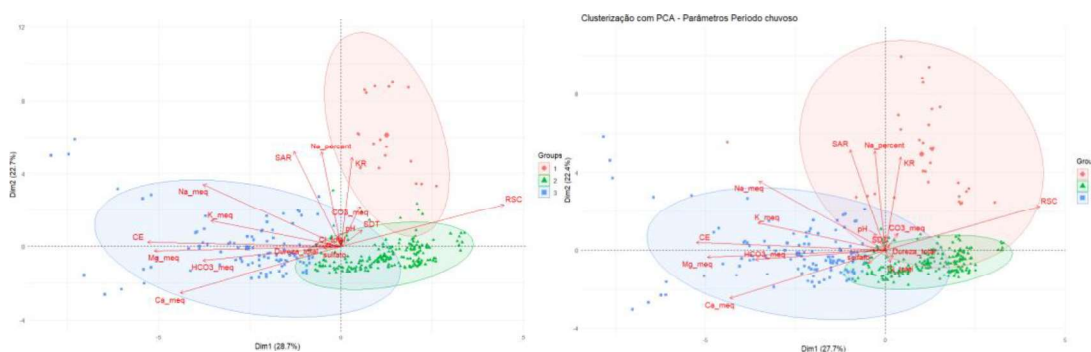
O segundo cluster (em verde) é representado exclusivamente pela condutividade elétrica (CE), o que denota um comportamento isolado da variável em relação aos demais parâmetros. Essa singularidade pode ser atribuída à natureza integradora da CE, que expressa o somatório da concentração iônica total, mas não distingue os íons individualmente. A elevada contribuição da CE na primeira dimensão da PCA reforça sua importância como indicador global da salinidade, ainda que sem especificidade quanto à origem ou composição dos sais.

O terceiro cluster (em azul) agrupa SDT e dureza total, variáveis que refletem principalmente a influência da geologia local, intemperismo e dissolução de minerais carbonáticos. Esses parâmetros apresentaram maiores contribuições nas Dimensões 3 e 4 da PCA, o que indica uma dinâmica hidrogeoquímica distinta dos processos de salinização, mais relacionada à matriz litológica do aquífero. A correlação positiva entre dureza total e íons como  $\text{Ca}^{2+}$  e  $\text{HCO}_3^-$  sugere a presença de rochas carbonáticas como fonte principal desses elementos, especialmente em áreas associadas ao Grupo Bambuí, caracterizado por formações calcárias e dolomíticas (MMA, 2006; SANTOS *et al.*, 2020). A Formação Bambuí, de idade neoproterozoica, é composta por carbonatos e pelitos, que formam importantes aquíferos fraturados e cársticos na região do médio e alto São Francisco, incluindo o norte de Minas Gerais (CPRM, 2011).



**Figura 3.8.** Análise de cluster para o período seco e chuvoso

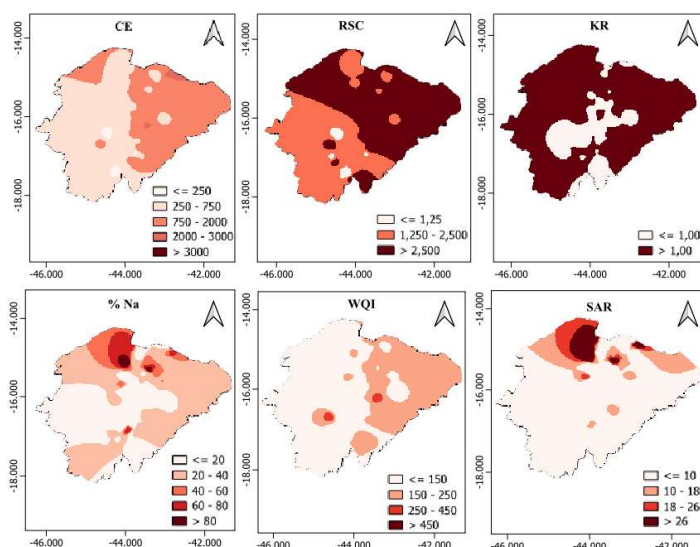
Apesar de a análise de agrupamento hierárquico com base nos dendrogramas não ter evidenciado mudanças significativas na estrutura dos clusters entre os períodos seco e chuvoso, uma vez que o número de grupos e a associação geral entre os parâmetros permaneceram semelhantes, a análise de agrupamento realizada via PCA (Figura 3.9) revelou diferenças um pouco mais expressivas na distribuição espacial dos grupos.



**Figura 3.9.** Gráfico de clusterização com PCA para o período seco e chuvoso

No gráfico de clusterização com PCA (Figura 3.9), observa-se que, embora os três grupos se mantenham nos dois períodos, houve pequenas variações na dispersão e na sobreposição entre eles. No período seco, os grupos apresentaram separação, especialmente o Grupo 1 (vermelho), que concentrou amostras associadas a altos teores de sódio, RSC e RAS, indicadores amplamente utilizados para avaliação da aptidão da água para irrigação (Ay#ers e WESTCOT, 1999; RICHARDS, 1954). Já no período chuvoso, observou-se maior sobreposição entre os grupos, sugerindo um efeito de homogeneização das características químicas da água subterrânea decorrente da recarga hídrica superficial. Esse tipo de abordagem combinando análise sazonal, índices de irrigação e técnicas multivariadas tem sido aplicada de forma fragmentada na literatura (ZHANG *et al.*, 2020; YIDANA *et al.*, 2012), mas são escassos os estudos que integram essas dimensões no contexto hidrogeológico semiárido do Brasil. A comparação sistemática entre os períodos seco e chuvoso com base na estrutura dos dados e nos agrupamentos químicos pode fornecer subsídios valiosos para o gerenciamento racional da água subterrânea, especialmente em regiões suscetíveis à salinização e degradação da estrutura do solo por irrigação.

Na Figura 3.10 apresenta-se os mapas de distribuição espacial dos principais parâmetros de qualidade da água subterrânea utilizados na avaliação da aptidão para irrigação. Observa-se que as áreas com maiores valores de condutividade elétrica (CE), RSC, KR, RAS e porcentagem de sódio (%Na) concentram-se predominantemente na porção centro-norte da região estudada, indicando zonas críticas quanto ao risco de salinização e sodificação.



**Figura 3.10.** Distribuição espacial com os resultados dos índices reportados para o estudo.

A CE apresentou valores superiores a 3000  $\mu\text{S}/\text{cm}$  em algumas localidades, sugerindo elevada carga de sais dissolvidos, resultado de processos como intemperismo, dissolução de minerais solúveis e possível influência antrópica, como irrigação excessiva e fertilização (AYERS e WESTCOT, 1999). O índice RSC revelou valores superiores a 2,5 meq/L em boa parte da área, sugerindo desequilíbrio entre carbonatos/bicarbonatos e os cátions cálcio e magnésio, o que pode favorecer a precipitação desses íons e acentuar o efeito do sódio trocável no solo. Os elevados valores de KR ( $>1$ ) e RAS ( $>26$ ) reforçam esse cenário, evidenciando um potencial risco de dispersão de partículas argilosas e degradação da estrutura física do solo. Esses resultados são consistentes com as altas porcentagens de sódio observadas no noroeste da área, indicando possível contribuição tanto de fontes naturais (litologia local) quanto de práticas agrícolas intensivas.

### 3.3.6 Previsão de índices de qualidade da água para fins de irrigação

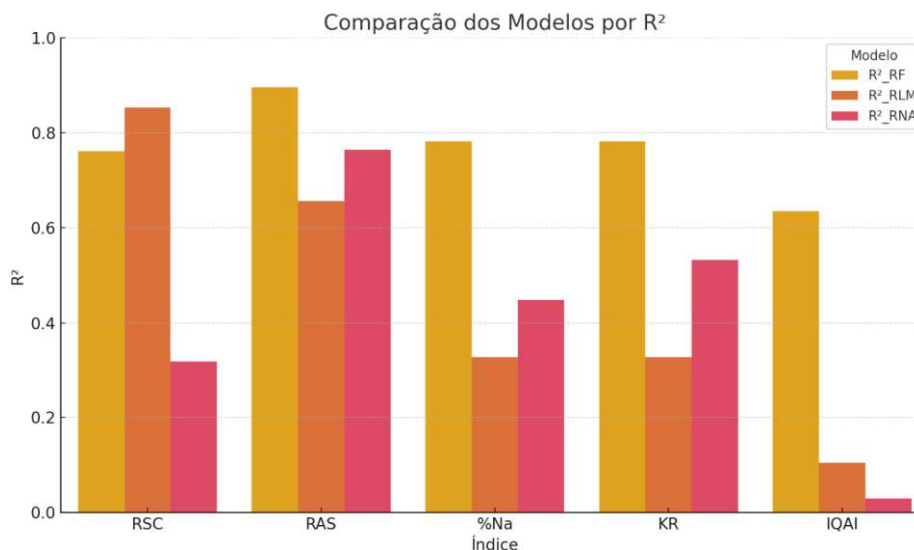
Embora tenham sido observadas variações pontuais entre os períodos seco e chuvoso nas análises multivariadas, a estrutura geral da qualidade da água subterrânea permaneceu relativamente estável, com os mesmos grupos de variáveis influentes e padrões espaciais semelhantes. Dessa forma, optou-se por modelar os índices de qualidade de forma agregada, incorporando o conjunto completo de dados para maximizar o poder preditivo dos algoritmos e garantir representatividade estatística. Essa abordagem é comum em estudos onde a variabilidade temporal é pequena ou onde os modelos devem ser aplicáveis em cenários gerais

de manejo agrícola (RASHEED *et al.*, 2024; ABU *et al.*, 2024). Os resultados das métricas estatísticas utilizadas para comparação dos modelos de aprendizado de máquina (RF, RLM e RNA) da análise dos índices de qualidade de água subterrânea para fins de uso na irrigação são apresentados na Tabela 3.7, sendo destacados os melhores resultados para os índices. De forma geral, valores de  $R^2$  superiores a 0,70 são reportados como adequados para indicar a robustez de modelos (XIONG *et al.*, 2023; ABU *et al.*, 2024), condição também recomendada no desenvolvimento de modelos de qualidade de água a partir de dados monitorados ambiental (RASHEED *et al.*, 2024).

**Tabela 3.7.** Comparação dos modelos de aprendizado de máquina para os índices de qualidade de água para fins de irrigação.

Índice	RF		RLM		RNA	
	$R^2$	RMSE	$R^2$	RMSE	$R^2$	RMSE
RSC	0,7608	<b>3,95</b>	<b>0,8523</b>	6,7582	0,3167	5,754
RAS	<b>0,8952</b>	<b>3,43</b>	0,6559	6,6520	0,7639	5,2846
%Na	<b>0,7819</b>	4,85	0,3255	<b>2,3860</b>	0,4483	8,3046
KR	<b>0,7819</b>	<b>2,09</b>	0,3255	4,0477	0,5325	3,3701
IQA <sub>I</sub>	<b>0,6352</b>	<b>5,07</b>	0,1043	7,0676	0,0296	7,9545

Neste estudo, observou-se que o modelo Random Forest (RF) apresentou desempenho superior na predição dos índices KR, RAS e %Na, com coeficientes de determinação ( $R^2$ ) superiores a 0,78 e valores de erro quadrático médio (RMSE) mais baixos em comparação com a Regressão Linear Múltipla (RLM) e Redes Neurais Artificiais (RNA) (FIGURA 3.11).



**Figura 3.11.** Comparação dos modelos por R<sup>2</sup>

No caso do IQAI, o valor de R<sup>2</sup> para RF foi elevado (0,63), indicando a eficácia do modelo de aprendizado de máquina para predição da qualidade da água para irrigação, o que favorece o seu emprego em práticas de gestão agrícola. Kim *et al.* (2024) e Sidek *et al.* (2024) também reportam o melhor ajuste do modelo RF para a determinação da qualidade de água de águas superficiais. Embora o *Random Forest* não forneça equações explícitas, o modelo é extremamente eficiente na predição de índices de qualidade da água, especialmente em ambientes com muitas variáveis correlacionadas. Além disso, substitui a necessidade de fórmulas com uma abordagem baseada em aprendizado empírico a partir dos dados. De maneira geral, os resultados confirmam que modelos não-lineares, como o RF, são mais eficazes para prever índices complexos de qualidade da água em áreas com características hidrogeológicas heterogêneas, como o norte de Minas Gerais. Além disso, o RF se mostrou consistente com os achados das análises multivariadas (PCA e HCA), uma vez que as variáveis mais relevantes na predição são também aquelas mais associadas às variabilidades nas amostras. Isso fortalece o uso do RF como ferramenta prática para auxiliar a gestão da água subterrânea para fins agrícolas, permitindo prever com precisão riscos de salinização e orientar medidas corretivas.

O modelo de RF é amplamente reconhecido por sua capacidade de minimizar o risco de *overfitting*, uma vez que cada árvore é treinada em subconjuntos aleatórios dos dados e variáveis, o que melhora significativamente a generalização do modelo (CONOSCENTI *et al.*, 2021; EMAMGHOLIZADEH *et al.*, 2022). A estrutura do modelo permite capturar interações complexas entre variáveis sem comprometer a precisão, o que o torna adequado para prever a qualidade da água em cenários de alta variabilidade dos dados. De forma semelhante ao

observado por Sahoo *et al.* (2021), o RF apresentou desempenho superior aos modelos lineares na predição de variáveis como CE, nitrato e SDT em aquíferos do sul da Índia, alcançando valores de  $R^2$  acima de 0,85. Liu *et al.* (2023) aplicaram RF e XGBoost para modelar o risco de salinização de aquíferos costeiros na China, obtendo excelente acurácia ( $R^2 > 0,90$ ) na estimativa de RAS e RSC, mesmo sob elevada heterogeneidade espacial. Esses achados reforçam a eficácia do RF na predição de índices de qualidade da água subterrânea para fins de irrigação em regiões semiáridas, como é o caso deste estudo.

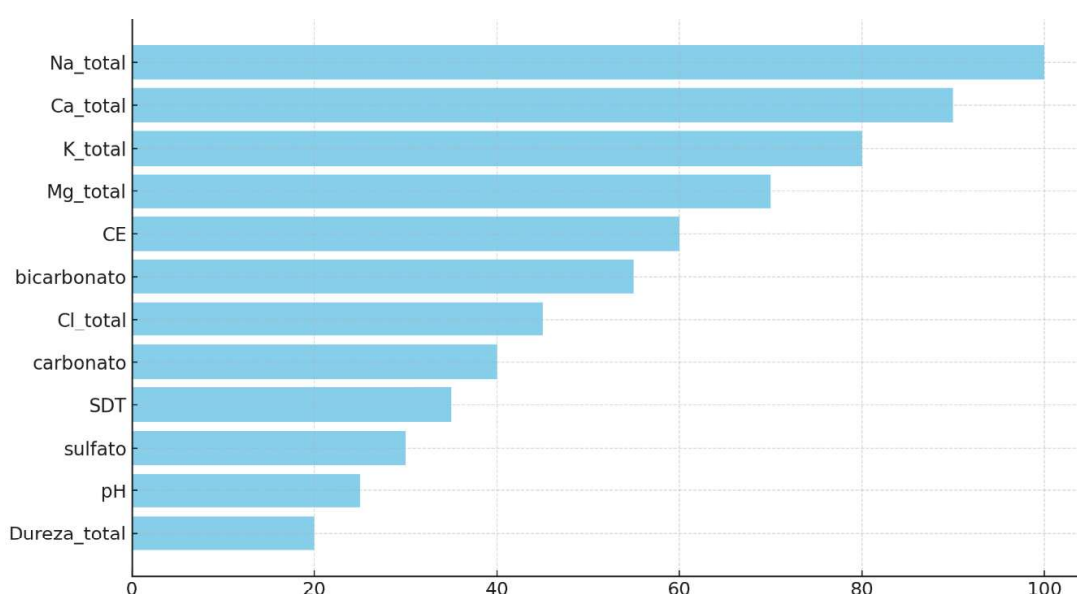
Complementando essa evidência, Hussein *et al.* (2024), em um estudo com 166 amostras de águas subterrâneas, relataram elevados valores de  $R^2$  para predição de índices como RAS, RSC e %Na utilizando RF ( $R^2 \approx 0,82$ ), XGBoost ( $\approx 0,98$ ) e SVR ( $\approx 0,99$ ), o que confirma a capacidade desses algoritmos em capturar correlações complexas entre múltiplas variáveis físico-químicas. Tais achados reforçam a robustez do RF observada no presente trabalho, sobretudo na predição dos índices que incorporam múltiplas interações iônicas, como o KR e o %Na. Como a análise estatística não indicou influência significativa da sazonalidade sobre os parâmetros avaliados, optou-se por utilizar um conjunto de dados único, reunindo observações dos períodos seco e chuvoso. Essa abordagem possibilitou ampliar a variabilidade amostral e aprimorar a capacidade do modelo em representar diferentes condições hidroquímicas típicas da região semiárida.

De forma semelhante, estudo de Al-Adhaileh, (2022) também apontou a efetividade de algoritmos de aprendizado de máquina na estimativa de índices como RAS, RSC e KR, ressaltando que a seleção adequada das variáveis de entrada tem impacto direto na acurácia do modelo. Essa observação encontra respaldo na presente análise, que apontou Na\_total, Ca\_total, K\_total, Mg\_total e CE como as variáveis mais influentes nos modelos RF para diferentes índices, refletindo seu papel determinante nos processos de salinização e sodificação dos solos irrigados. Estudos realizados em diferentes regiões semiáridas do mundo, como Argélia e Índia, também identificaram esses mesmos parâmetros como determinantes na avaliação da qualidade da água subterrânea para irrigação (MOKHTAR *et al.*, 2022; HUSSEIN *et al.*, 2024).

Ainda dentro desse contexto, Dewi *et al.* (2024) reforçam a superioridade do RF frente às redes neurais em estudos de qualidade de água em sistemas fluviais, atribuindo seu melhor desempenho à capacidade de lidar com não-linearidades sem sofrer overfitting, além de apresentar menores valores de RMSE e MAE. Tal comportamento foi observado no presente trabalho, onde as RNAs apresentaram  $R^2$  inferiores em praticamente todos os índices avaliados, sobretudo no IQAI ( $R^2 = 0,03$ ), enquanto o RF alcançou  $R^2 = 0,635$  para o mesmo índice.

Além do desempenho preditivo, a análise de importância das variáveis no RF (Figura 12) evidenciou que os principais contribuintes para a predição dos índices foram as concentrações de  $\text{Na}^+$ ,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{K}^+$  e a Condutividade Elétrica (CE), todos diretamente envolvidos nos processos de salinização e sodificação. Esse resultado reforça a relevância dessas variáveis para a compreensão integrada da qualidade da água subterrânea, como também demonstrado por Deshinta *et al.* (2024) em estudos fluviais e agrícolas no sudeste asiático.

A Figura 3.12 apresenta as variáveis mais relevantes para o modelo de Random Forest, conforme a métrica de aumento percentual no erro quadrático médio (%IncMSE).



**Figura 3.12.** Análise de importância das variáveis no RF

As variáveis com maior influência na predição dos índices de qualidade da água para irrigação foram: Na\_total (Sódio total), Ca\_total (Cálcio total), e K\_total (Potássio total), seguidas por Mg\_total (Magnésio total) e CE (Condutividade Elétrica), demonstrando sua relevância na composição iônica e nos efeitos sobre a salinidade e sodificação do solo. Em adição, os parâmetros bicarbonato, carbonato, SDT (Sólidos Dissolvidos Totais) e Cl\_total se destacam, os quais estão associados ao RSC e à precipitação de cátions como cálcio e magnésio, afetando a estrutura do solo e disponibilidade de água para as plantas.

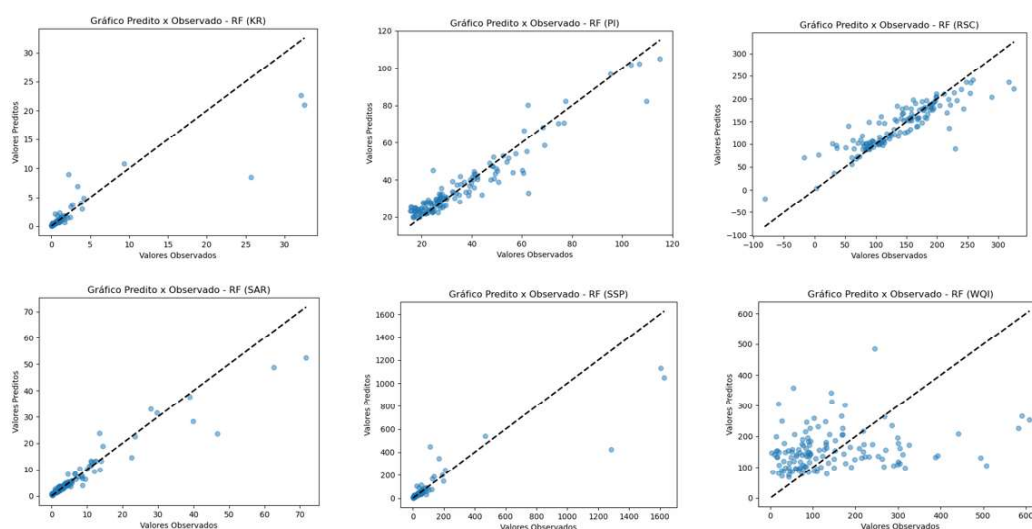
Tasnim *et al.* (2023) reportam que a aplicação combinada de técnicas poderia ser uma alternativa válida com vistas a uma maior precisão de modelos de qualidade de água, em especial para águas subterrâneas.

A Figura 3.13 apresenta os gráficos de dispersão entre os valores observados e os valores preditos pelo modelo Random Forest (RF) para os diferentes índices de qualidade da água

subterrânea utilizados na avaliação da aptidão para irrigação. A linha tracejada representa a identidade ( $y = x$ ), indicando o ponto ideal onde os valores preditos coincidem exatamente com os observados.

Observa-se que os modelos apresentaram alto grau de aderência à linha de identidade para os índices RAS, KR, RSC e PI, o que indica bom desempenho preditivo e baixa dispersão dos resíduos, em conformidade com os elevados valores de  $R^2$  reportados anteriormente. O modelo RF para RAS, por exemplo, exibiu concentração expressiva de pontos ao longo da linha, sugerindo excelente capacidade de generalização.

Por outro lado, os modelos voltados à predição do IQAi e SSP demonstraram maior dispersão dos pontos, indicando maior dificuldade na modelagem desses índices, possivelmente devido à sua composição multivariada e maior sensibilidade a ruídos ou variáveis não representadas adequadamente no conjunto de entrada.



**Figura 3.13.** Dados reais versus os previstos para os dados de validação modelo RF

Esses resultados reforçam a robustez do modelo RF, especialmente para índices cuja formulação depende diretamente da concentração de íons monovalentes e bivalentes, como  $\text{Na}^+$ ,  $\text{Ca}^{2+}$  e  $\text{Mg}^{2+}$ , enquanto apontam a necessidade de ajustes ou seleção de variáveis mais específicas para índices compostos como o IQAi. De modo geral, os resultados desta etapa demonstram a eficácia do modelo Random Forest na predição de índices de qualidade da água subterrânea para irrigação, especialmente aqueles fortemente associados à concentração de íons dissolvidos, como RAS, KR, RSC e %Na.

### 3.3.7 Recomendações e perspectivas do estudo

Com base nos resultados obtidos quanto à qualidade da água subterrânea na região semiárida do norte de Minas Gerais, associada ao risco de salinização dos solos agrícolas, torna-se evidente a necessidade de práticas de manejo voltadas à sustentabilidade do uso desse recurso para fins de irrigação. Dentre as estratégias recomendadas, destaca-se a importância de assegurar um adequado nivelamento do solo e garantir boas condições de drenagem, principalmente em áreas propensas ao acúmulo de sais, além da escolha criteriosa de culturas mais tolerantes à salinidade, conforme sugerido por Singh *et al.* (2022).

Técnicas de manejo irrigatório também desempenham papel fundamental. O uso de sistemas como a irrigação por sulcos, que distribui a água entre fileiras de plantas de forma mais controlada, aliado ao cultivo em fileiras elevadas, contribui para reduzir o contato direto das raízes com a água salinizada (MOMTAZI *et al.*, 2024; SARKER *et al.*, 2024). Além disso, práticas conservacionistas como a cobertura do solo com palha ou resíduos vegetais, o plantio direto e a rotação de culturas são recomendadas por favorecerem a manutenção da umidade e da estrutura do solo, reduzindo os efeitos adversos da salinidade (SOLANGI *et al.*, 2019).

Em áreas onde há indícios de sodicidade, o uso de aditivos como gesso agrícola ( $\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$ ) e fontes orgânicas, como esterco de curral, pode ser eficaz para melhorar a estrutura do solo e reduzir a concentração de sódio trocável (SINGH *et al.*, 2021). Adicionalmente, a implementação de sistemas de recarga de águas pluviais, o fortalecimento das redes de drenagem e o uso racional de fertilizantes e defensivos agrícolas são medidas que contribuem para o equilíbrio hídrico e químico do solo (GUGULOTHU *et al.*, 2022). Além das estratégias de manejo recomendadas, destaca-se o potencial do uso de algoritmos de aprendizado de máquina como ferramentas complementares à gestão ambiental da qualidade da água subterrânea. A adoção de modelos preditivos, como o Random Forest, pode oferecer aos órgãos gestores uma abordagem eficiente, de baixo custo e escalável para antecipar riscos de degradação da qualidade hídrica em áreas irrigadas. Esses modelos permitem a detecção de padrões espaciais e temporais relacionados à salinização, à recarga hídrica e à dinâmica iônica, viabilizando respostas mais ágeis e fundamentadas na tomada de decisão. Ademais, a integração de dados de monitoramento com plataformas geoespaciais e sistemas de alerta automatizados pode fortalecer ações de fiscalização, priorização de áreas críticas e planejamento de políticas públicas voltadas à segurança hídrica em regiões semiáridas. Nesse contexto, as ferramentas de ML se apresentam como aliadas estratégicas à gestão integrada dos recursos hídricos subterrâneos, promovendo inovação, eficiência e sustentabilidade no monitoramento ambiental.

Por fim, destaca-se a necessidade de ações contínuas de capacitação e conscientização ambiental entre produtores e técnicos locais, com vistas à adoção de práticas sustentáveis e ao fortalecimento da gestão integrada dos recursos hídricos subterrâneos. Como salientado por Subba Rao *et al.* (2022), a percepção ambiental e o conhecimento técnico são elementos-chave para garantir o uso eficiente e responsável da água em regiões vulneráveis à degradação hidroambiental.

### 3.4 Conclusão

O presente estudo avaliou a qualidade da água subterrânea para fins de irrigação no semiárido do norte de Minas Gerais, integrando análises físico-químicas, índices de adequação e técnicas de machine learning, considerando os períodos seco e chuvoso. Os parâmetros analisados apresentaram variações médias pequenas entre as estações e ausência de influência significativa da pluviosidade sobre a maioria das variáveis, o que permitiu unir os dados em um conjunto único para a etapa de modelagem. Essa estabilidade, também observada em outros aquíferos fraturados de regiões semiáridas, reflete a predominância de controles geoquímicos locais e de fontes pontuais de sais sobre a dinâmica hidroquímica.

Entre os índices de qualidade avaliados, RAS, KR e %Na permaneceram majoritariamente dentro de faixas adequadas para irrigação, embora a ocorrência de valores extremos para alguns íons ( $\text{Cl}^-$ ,  $\text{Na}^+$  e  $\text{SO}_4^{2-}$ ) indique heterogeneidade espacial e a necessidade de monitoramento direcionado. A PCA revelou padrões consistentes entre as estações, com aumento da contribuição relativa de SDT e  $\text{CO}_3$  no período seco, possivelmente associado à menor diluição hídrica.

Na modelagem preditiva, o Random Forest destacou-se pelo melhor desempenho entre os algoritmos testados, obtendo altos valores de  $R^2$  e baixos erros, especialmente para índices derivados de múltiplas interações iônicas, como KR e %Na. Esses resultados demonstram o potencial do RF para lidar com a variabilidade espacial e não linearidade das relações hidroquímicas em ambientes semiáridos.

Assim, embora não tenham sido detectadas alterações sazonais marcantes na qualidade da água, a aplicação integrada de análises estatísticas e machine learning mostrou-se eficiente para compreender padrões e prever indicadores críticos, contribuindo para o planejamento do uso agrícola e a gestão sustentável dos recursos hídricos na região.

### 3.5 Referências

ARABAMERI, A. *et al.* Modeling Groundwater Potential Using Novel Gis-Based Machine-Learning Ensemble Techniques. **Journal Of Hydrology: Regional Studies**, v. 36, 2021.

AMWELE, H. R.; KGABI, N. A.; KANDJIBI, L. I. Sustainability of groundwater for irrigation purposes in semi-arid parts of Namibia. **Frontiers in Water**, v. 3, 2021

AL-ADHAILEH, M. H. Groundwater Quality: The Application of Artificial Intelligence. **Wiley Online Library**, 2022. DOI: 10.1155/2022/8425798.

BATARSEH, M. *et al.* Assessment of groundwater quality for irrigation in the arid regions using irrigation water quality index (IWQI) and GIS-Zoning maps: case study from Abu Dhabi Emirate, UAE. **Groundwater Sustain Dev**, v. 14, 2021, Artigo 100611.

BAHRAMI, M.; ZAREI, A.R. Assessment And Modeling Of Groundwater Quality For Drinking, Irrigation, And Industrial Purposes Using Water Quality Indices And Gis Technique In Fasarud Aquifer (Iran). **Modeling Earth Systems and Environment**, p. 1–15, 2023.

BOUNAB, N.; *et al.* *Assessment of groundwater quality in the Ain Larbi area: geostatistical modeling and PCA.* 2022.

BHAT, M. A.; WANI, A. S.; VIJAY, K.; JYOTIRMAYA, S.; DINESH, T.; *et al.* An Overview of the Assessment of Groundwater Quality for Irrigation. **Journal of Agricultural Science and Food Research**, v. 9, no. 1, p. 209, 2018.

BOUDIBI, S.; *et al.* Groundwater quality assessment and irrigation water quality index prediction using machine learning algorithms. **Water**, v. 16, 2024. DOI: 10.3390/w16020264.

CONOSCENTI, C.; MARTINELLO, C.; ALFONSO-TORREÑO, A.; GÓMEZ-GUTIÉRREZ, Á. Predicting sediment deposition rate in check-dams using machine learning techniques and high-resolution DEMs. **Environmental Earth Sciences**, v. 80, n. 10, p. 1–19, 2021. DOI: 10.1007/s12665-021-09695-3.

DEY, Sangita; RAJU, Nandimandalam; GOSSEL, Wolfgang; MALL, Rajesh. Hydrogeochemical characterization and geochemical modeling for the evaluation of groundwater quality and health risk assessment in the Varuna River basin, India. **Environmental Geochemistry and Health**, v. 45, 2023. DOI: <https://doi.org/10.1007/s10653-023-01521-y>.

DEWI, Deshinta Arrova; SAM WEI, Aik; LIN, Leong Chi; HENG, Chang Ding. Water Quality Prediction using Random Forest Algorithm and Optimization. **Journal of Applied Data Sciences**, v. 5, n. 3, p. 1354–1362, set. 2024.

dos SANTOS, A.L.M.; Gonçalves, W.A.; Andrade, L.d.M.B.; Rodrigues, D.T.; Batista, F.F.; Lima, G.C.; e Silva, C.M.S. Space–Time Characterization of Extreme Precipitation Indices for the Semiarid Region of Brazil. **Climate** **2024**, *12*, 43. <https://doi.org/10.3390/cli12030043>

EMAMGHOLIZADEH, S.; BAZOOBANDI, A.; MOHAMMADI, B.; GHORBANI, H. Prediction of soil cation exchange capacity using enhanced machine learning approaches in the

southern region of the Caspian Sea. **Ain Shams Engineering Journal**, v. 13 (2022), p. –, 2022. DOI: 10.1016/j.asej.2022.101876

ESTÉVEZ, V. *et al.* Machine learning techniques for acid sulfate soil mapping in southeastern Finland. **Geoderma**, v. 406, 2022, Artigo 115446.

GANTAIT, A. *et al.* Hydrogeochemical evolution and quality assessment of groundwater of Ajmer district, Rajasthan, India. **J. Earth Syst. Sci.**, v. 131, n. 4, 2022, p. 236.

Gao, M., K. Knobelspiesse, B.A. Franz, P.-W. Zhai, B. Cairns, X. Xu, and J.V. Martins, 2023: The impact and estimation of uncertainty correlation for multi-angle polarimetric remote sensing of aerosols and ocean color. **Atmos. Meas. Tech.**, **16**, no. 8, 2067-2087, doi:10.5194/amt-16-2067-2023.

GAUTAM, Vinay Kumar; KOTHARI, Mahesh; AL-RAMADAN, Baqer M.; SINGH, P. K.; UPADHYAY, Harsh; PANDE, Chaitanya B.; et al. Groundwater quality characterization using an integrated water quality index and multivariate statistical techniques. **PLoS ONE**, v. 19, n. 2, e0294533, 23 fev. 2024. DOI: <https://doi.org/10.1371/journal.pone.0294533>.

GORGAN-MOHAMMADI, F.; RAJAEI, T.; ZOUNEMAT-KERMANI, M. Investigating machine learning models in predicting lake water quality parameters as a 3-year moving average. **Environmental Science and Pollution Research**, v. 30, p. 63839–63863, 2023. DOI: 10.1007/s11356-023-26830-8

GUGULOTHU, Sakram; SUBBARAO, N.; DAS, Rashmirekha; DHAKATE, Ratnakar. Geochemical evaluation of groundwater and suitability of groundwater quality for irrigation purposes in an agricultural region of South India. **Applied Water Science**, v. 12, art. 142, 2022. DOI: 10.1007/s13201-022-01583-w.

GUERRA, J.V.S. *et al.* Agronomic Performance Of Irrigated Crop Rotations Under Conventional And No-Tillage Systems In The Semiarid Region Of Minas Gerais, Brazil. **Rev. Caatinga, Mossoró**, v. 35, n. 1, p. 33–43, jan.–mar., 2022.

LIU, X. *et al.* Effects of sodium adsorption ratio and electrolyte concentration on soil saturated hydraulic conductivity. **Geoderma**, v. 414, 2022, Artigo 115772.

KHOSRAVI, K.; et al. Spatial prediction of groundwater spring potential mapping based on an adaptive neuro-fuzzy inference system and metaheuristic optimization. **Hydrology and Earth System Sciences**, v. 22, p. 4771–4792, 2018.

ELUMALAI, Manikandan; NATARAJAN, Rajmohan; ANBAZHAGAN, Siddan. Monsoon impact on groundwater chemistry and geochemical processes in the shallow hard rock aquifer. **Catena**, v. 195, p. 104766, dez. 2020. DOI: 10.1016/j.catena.2020.104766.

MOKHTAR, A., ELBELTAGI, A., GYASI-AGYEI, Y. *et al.* Prediction of irrigation water quality indices based on machine learning and regression models. **Appl Water Sci** **12**, 76 (2022). <https://doi.org/10.1007/s13201-022-01590-x>

NISHAT, M. H.; et al. Comparative analysis of machine learning models for predicting water quality index in Dhaka's rivers of Bangladesh. **Environmental Sciences Europe**, v. 37, art. 31, 2025.

PATEL, Praharsh; SAHA, Dipankar; SHAH, Tushaar Natwarlal. Sustainability of groundwater through community-driven distributed recharge: An analysis of arguments for water scarce regions of semi-arid India. **Journal of Hydrology: Regional Studies**, v. 29, p. 100680, jun. 2020. DOI: 10.1016/j.ejrh.2020.100680.

LI, PINGHENG et al. "Groundwater Quality Assessment for Rangeland Dynamic: Integration of Multicriteria Decision Analysis with Remote Sensing Data." **Rangeland Ecology & Management** (2025): n. pag.

NABIOLLAHI, K. *et al.* Assessing agricultural salt-affected land using digital soil mapping and hybridized random forests. **Geoderma**, v. 385, 2021, Artigo 114858.

OMRANI, M. *et al.* Application of remote sensing indices to digital soil salt composition and ionic strength mapping in the east shore of Urmia Lake, Iran. *Remote Sens. Appl.: Soc. Environ.*, v. 22, 2021, Artigo 100498.

RAGHUNATH, H. M. *Ground Water*. 2. ed. Wiley, 1987

SAHOO, Alok; SINGH, Satyapriya; NATH, Anshuman; SUNANI, Sunil. Impact of agricultural pollutants on water resources and their management. 2021. p. 17–30. ISBN 978-93-90471-69-0.

SOLANGI GS, Siyal AA, Babar MM, Siyal P (2019) Groundwater quality evaluation using the water quality index (WQI), the synthetic pollution index (SPI), and geospatial tools: a case study of Sujawal district, Pakistan. *Hum Ecol Risk Assess Int J*. <https://doi.org/10.1080/10807039.2019.1588099>

SUBBA Rao N, Chaudhary M (2019) Hydrogeochemical processes regulating the spatial distribution of groundwater contamination, using pollution index of groundwater (PIG) and hierarchical cluster analysis (HCA): a case study. **Groundw Sustain Dev** 9:100238. <https://doi.org/10.1016/j.gsd.2019.100238>

RICHARDS, A., Allison, L. E., Bernstein, C., Bower, L. A., Brown, J. W., Fireman, M., et al. (1954). United States salinity laboratory staff. **Agriculture Handbook**: United States Departement of Agriculture, 60.

SAHA, D.; RAY, R. K. Groundwater resources of India: Potential, challenges and management. In: SIKDAR, P. (ed.). **Groundwater development and management**. Cham: Springer International Publishing, 2019, p. 19–42.

SANTACRUZ-DE LEÓN, G.; MORAN-RAMÍREZ, J.; RAMOS-LEAL, J. A. Impact of Drought and Groundwater Quality on Agriculture in a Semi-Arid Zone of Mexico. **Agriculture**, v. 12, n. 9, art. 1379, set. 2022. DOI: <https://doi.org/10.3390/agriculture12091379>

SARKAR, Manasree; PAL, Subodh Chandra; ISLAM, Abu Reza Md Towfiqul. Groundwater quality assessment for safe drinking water and irrigation purposes in Malda district, Eastern India. **Environmental Earth Sciences**, v. 81, n. 2, art. 52, jan. 2022. DOI: 10.1007/s12665-022-10188-0.

SARKAR, S.K., Rudra, R.R., Talukdar, S. *et al.* Future groundwater potential mapping using machine learning algorithms and climate change scenarios in Bangladesh. **Sci Rep** **14**, 10328 (2024). <https://doi.org/10.1038/s41598-024-60560-2>

SENA, F.H.S. *et al.* Levantamento Fitossociológico De Plantas Daninhas Em Pomares De Mangueira No Semiárido Mineiro. **Nativa**, Sinop, v. 7, n. 5, p. 500-505, 2019.

SINGH, G. *et al.* Assessment of Groundwater Suitability For Sustainable Irrigation: A Comprehensive Study Using Indexical, Statistical, and Machine Learning Approaches. **Groundwater For Sustainable Development**, v. 24, 2024.

TAGHIZADEH-MEHRJARDI, R. *et al.* Enhancing the accuracy of machine learning models using the super learner technique in digital soil mapping. **Geoderma**, v. 399, 2021, Artigo 115108.

TEŞAN, S. Estimation of groundwater quality using an integration of water quality index, artificial intelligence methods and GIS: case study, Central Mediterranean Region of Turkey. **Appl. Water Sci.**, v. 13, n. 1, 2023, p. 15.

TIWARI, T.N.; MISHRA, M.A. A preliminary assignment of water quality index of major Indian rivers. **Indian J. Environ. Protect.**, v. 5, n. 4, 1985, pp. 276-279.

UNIGWE, Chinanu; EGBUERI, Johnbosco. Drinking water quality assessment based on statistical analysis and three water quality indices (MWQI, IWQI and EWQI): a case study. **Environment, Development and Sustainability**, v. 23, p. 6592–6617, 2022. DOI: <https://doi.org/10.1007/s10668-020-00868-2>.

TAGHIZADEH-MEHRJARDI, Ruhollah; HAMZEHPUR, Nikou; HASSANZADEH, Maryam; *et al.* Enhancing the accuracy of machine learning models using the super learner technique in digital soil mapping. **Geoderma**, [S.l.], v. 399, p. 115108, abr. 2021. DOI: 10.1016/j.geoderma.2021.115108.

TAYYAB, Muhammad; ASLAM, Rana Ammar; FAROOQ, Umar; *et al.* Comparative study of geospatial techniques for interpolating groundwater quality data in agricultural areas of Punjab, Pakistan. **Water**, [S.l.], v. 16, n. 1, p. 139, dez. 2023. DOI: 10.3390/w16010139.

YIDANA, Sandow Mark; BANOENG-YAKUBO, Bruce; ALIOU, Abdul Samed; AKABZAA, Thomas M. Groundwater quality in some Voltaian and Birimian aquifers in northern Ghana , application of multivariate statistical methods and geographic information systems. **Hydrological Sciences Journal**, v. 57, n. 6, p. 1168–1183, ago. 2012. DOI: 10.1080/02626667.2012.693612.

UNIGWE, Chinanu; EGBUERI, Johnbosco. Drinking water quality assessment based on statistical analysis and three water quality indices (MWQI, IWQI and EWQI): a case study.

*Environment, Development and Sustainability*, v. 23, p. 6592–6617, 2022. DOI: <https://doi.org/10.1007/s10668-020-00868-2>.

WANG, H. *et al.* Quantitatively estimating main soil water-soluble salt ions content based on Visible-near infrared wavelength selected using GC. **SR and VIP. PeerJ**, v. 7, 2019, p. e6310.

WANG, F. *et al.* Characterizing soil salinity at multiple depth using electromagnetic induction and remote sensing data with random forests: A case study in Tarim River Basin of southern Xinjiang, China. **Sci. Total Environ.**, v. 754, 2021, Artigo 142030.

WILCOX, L.V. Classification and Use of Irrigation Waters. **USDA Circular** No. 969, Washington, DC, 1955, 19 p.

WU, Y. *et al.* A Novel Kernel Extreme Learning Machine Model Coupled With K-Means Clustering And Firefly Algorithm For Estimating Monthly Reference Evapotranspiration In Parallel Computation. **Agric. Water Manag.**, v. 245, 2021, Artigo 106624

YUJUAN Lei, Yunde Liu, Ziyong Sun, Changjian Zou, Rui Ma, Lihe Yin, Huanying Pan, Influences of paleoclimatic environment and hydrogeochemical evolution on groundwater salinity in an arid inland plain in northwestern China, **Applied Geochemistry**, Volume 154, 2023, 105688, ISSN 0883-2927, <https://doi.org/10.1016/j.apgeochem.2023.105688>.

YIDANA, Sandow Mark; BANOENG-YAKUBO, Bruce; ALIOU, Abdul Samed; AKABZAA, Thomas M. Groundwater quality in some Voltaian and Birimian aquifers in northern Ghana , application of multivariate statistical methods and geographic information systems. **Hydrological Sciences Journal**, v. 57, n. 6, p. 1168–1183, ago. 2012. DOI: 10.1080/02626667.2012.693612

ZHAO, D., Arshad, M., Li, N., Triantafilis, J., 2023. Predicting soil physical and chemical properties using vis-NIR in Australian cotton areas. **Catena** 196, 104938.

ZHANG, H., Yin, S., Chen, Y., Shao, S., Wu, J., Fan, M., Gao, C., 2020b. Machine learning-based source identification and spatial prediction of heavy metals in soil in a rapid urbanization area, eastern China. **J. Cleaner Prod.** 273, 122858.

ZHANG, Qiyong; QIAN, Hui; REN, Wenhao; XU, Panpan; LI, Weiqing; YANG, Qiaoyang; SHANG, Jiatao. Salinization of shallow groundwater in the Jiaokou Irrigation District and associated secondary environmental challenges. **Science of The Total Environment**, v. 908, p. 168445, 15 jan. 2024. DOI: <https://doi.org/10.1016/j.scitotenv.2023.168445>.

CHICCO D, Warrens MJ, Jurman G (2021) The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. **PeerJ Comput Sci** 7:e623. <https://doi.org/10.7717/peerj-cs.623>

PINGHENG Li, Iram Naz, Rana Waqar Aslam, Muhammad Azeem Liaquat, Yahia Said, Groundwater Quality Assessment for Rangeland Dynamic: Integration of Multicriteria Decision Analysis with Remote Sensing Data, **Rangeland Ecology e Management**, Volume 102, 2025, Pages 110-127, ISSN 1550-7424, <https://doi.org/10.1016/j.rama.2025.06.010>.

MOMTAZI *et al.*, 2024; Farahnaz Momtazi, Hamidreza Miri, Barmak Jafari Haghighi, Abdolreza Jafari, Hamidreza Ebrahimi; Agrophysiological responses of bread wheat to raised-bed planting and irrigation level. **Water Supply** 1 November 2024; 24 (11): 3697–3707. doi: <https://doi.org/10.2166/ws.2024.228>

SARKER *et al.*, 2024. Anower, Skip Furrow Irrigation Maintains Yield and Salinity, Saves Water and Increases Water Productivity of Sunflower in Saline Soils. Available at SSRN: <https://ssrn.com/abstract=5318373> or <http://dx.doi.org/10.2139/ssrn.5318373>

SUBRAMANIAN, V.; SAXENA, K. K. Hydro-geochemistry of groundwater in the Delhi region of

ZHAO, D., ARSHAD, M., Li, N., TRIANTAFILIS, J., 2023. Predicting soil physical and chemical properties using vis-NIR in Australian cotton areas. **Catena** 196, 104938.

ZHANG, H., *et al.* Machine learning- based source identification and spatial prediction of heavy metals in soil in a rapid urbanization area, eastern China. **J. Cleaner Prod.** 273, 122858., 2020.

ZHANG, Qiyang; QIAN, Hui; REN, Wenhao; XU, Panpan; LI, Weiqing; YANG, Qiaoyang; SHANG, Jiatao. Salinization of shallow groundwater in the Jiaokou Irrigation District and associated secondary environmental challenges. **Science of The Total Environment**, v. 908, p. 168445, 15 jan. 2024. DOI: <https://doi.org/10.1016/j.scitotenv.2023.168445>.

## 5 CONSIDERAÇÕES GERAIS

Esta tese abordou de forma integrada os aspectos qualitativos e quantitativos das águas subterrâneas em regiões semiáridas, combinando análise bibliométrica global, modelagem espacial do potencial hídrico e avaliação preditiva da qualidade para irrigação. A articulação

desse três eixos de investigação permitiu compreender não apenas o estado da arte científico, mas também explorar metodologias inovadoras e aplicadas para subsidiar a gestão sustentável desse recurso estratégico.

O levantamento bibliométrico revelou um crescimento expressivo da produção científica sobre qualidade da água subterrânea a partir de 2020, impulsionado por desafios globais como o crescimento populacional, as mudanças climáticas e a expansão agrícola. Observou-se uma transição de abordagens tradicionais, centradas na caracterização físico-química, para estudos interdisciplinares que integram *machine learning*, geoprocessamento e modelagem hidrológica. Essa evolução metodológica reflete a maturação da pesquisa e o reconhecimento da importância de ferramentas preditivas e colaborativas na proteção e no manejo dos aquíferos.

No âmbito quantitativo, a modelagem espacial do potencial hídrico no norte de Minas Gerais demonstrou a eficácia de algoritmos de aprendizado de máquina, destacando-se o *Weighted k-Nearest Neighbors* (KKNN) como o método mais robusto na estimativa da vazão específica em contexto de elevada heterogeneidade geológica e escassez de dados. A incorporação de múltiplas covariáveis ambientais e a validação cruzada rigorosa possibilitaram representar com maior fidelidade a variabilidade espacial da produtividade dos poços. Os mapas contínuos de potencial e incerteza gerados constituem ferramentas estratégicas para o planejamento hídrico regional, permitindo priorizar áreas para exploração e conservação.

Quanto ao componente qualitativo, a análise da água subterrânea para irrigação indicou estabilidade nos parâmetros físico-químicos entre as estações seca e chuvosa, evidenciando a predominância de controles geoquímicos locais sobre a sazonalidade pluviométrica. Embora a maioria dos índices de adequação tenha permanecido em faixas favoráveis, a detecção de valores extremos para íons como  $\text{Cl}^-$ ,  $\text{Na}^+$  e  $\text{SO}_4^{2-}$  reforça a necessidade de monitoramento direcionado. Na modelagem preditiva, o algoritmo Random Forest apresentou desempenho superior, especialmente para índices derivados de múltiplas interações iônicas, confirmando seu potencial para lidar com padrões não lineares e variabilidade espacial da qualidade da água.

De forma geral, os resultados desta tese demonstram que a integração de análises bibliométricas, modelagem preditiva e diagnóstico hidroquímico oferece uma abordagem abrangente e eficaz para a gestão de recursos hídricos subterrâneos em regiões semiáridas. Além de avançar o conhecimento científico, os produtos gerados, como mapas de potencial aquífero, modelos preditivos de qualidade e identificação de lacunas de pesquisa, fornecem subsídios práticos para o planejamento agrícola, a formulação de políticas públicas e a implementação de estratégias de conservação.

Desse modo, acredita-se que os resultados obtidos neste estudo possam fornecer informações valiosas para a tomada de decisão no planejamento e na gestão integrada dos recursos hídricos subterrâneos, especialmente em áreas de elevada vulnerabilidade climática. As ferramentas e metodologias aplicadas demonstram potencial para serem adaptadas e replicadas em diferentes contextos hidrogeológicos, fortalecendo a capacidade de gestores e formuladores de políticas públicas em promover o uso sustentável e seguro da água subterrânea.

Vale ressaltar que o trabalho desenvolvido também apresenta limitações que precisam ser consideradas. Entre elas, destacam-se a disponibilidade e a qualidade dos dados, especialmente em regiões com baixa densidade de monitoramento, e a dependência de informações secundárias que podem conter incertezas associadas à coleta e ao processamento. Ademais, a escolha e parametrização dos algoritmos de *machine learning* influenciam diretamente os resultados, sendo recomendável que futuras pesquisas explorem diferentes combinações de modelos e técnicas de seleção de variáveis, bem como incorporem séries temporais mais longas e dados de monitoramento contínuo para aprimorar a capacidade preditiva e a aplicabilidade prática dos modelos.

## 6 SUGESTÃO PARA TRABALHO FUTUROS

Este estudo contribuiu significativamente para o entendimento da disponibilidade e da qualidade das águas subterrâneas em regiões semiáridas, combinando abordagens estatísticas, geoespaciais e algoritmos de aprendizado de máquina. Contudo, os resultados também indicam caminhos relevantes para investigações futuras, que podem aprimorar e expandir as análises realizadas:

1. **Ampliação e qualificação das bases de dados**  
Uma limitação recorrente foi a escassez de dados primários com alta resolução espacial e temporal. Futuras pesquisas podem focar na realização de campanhas sistemáticas de campo, com medições padronizadas de vazão e qualidade da água em diferentes estações do ano, além da instalação de redes de monitoramento contínuo.

2. **Integração de variáveis hidrogeofísicas**  
A inclusão de variáveis como resistividade elétrica do solo, espessura do aquífero, transmissividade e dados geofísicos de sondagens pode aumentar a robustez dos modelos de predição da vazão e da salinidade, especialmente em áreas com geologia complexa.

3. **Modelos híbridos e baseados em processos físicos**  
Estudos futuros podem explorar a integração entre algoritmos de ML e modelos hidrológicos baseados em processos, como MODFLOW ou SWAT, permitindo representar com maior fidelidade os mecanismos de recarga, fluxo e transporte de solutos nos aquíferos.

4. **Análise multitemporal e cenários futuros**  
A incorporação de análises multitemporais e modelagem preditiva de cenários climáticos pode ajudar a avaliar o impacto das mudanças climáticas na disponibilidade e salinização da água subterrânea. O uso de dados climáticos projetados por modelos como o CMIP6 pode ser integrado às análises existentes.

5. **Avaliação de outros parâmetros de qualidade da água**  
Além da salinidade, futuras investigações podem incluir indicadores como presença de nitrato, metais pesados, pesticidas ou parâmetros microbiológicos, a fim de ampliar a compreensão sobre os riscos à saúde humana e ao meio ambiente.

## 7 REFERÊNCIAS GERAIS

ALMEIDA, M. B. *et al.* Variabilidade espaço-temporal da qualidade da água subterrânea para irrigação: um estudo de caso no Vale Aluvial do Mimoso, região semiárida do Brasil. **ResearchGate**, 2025. Disponível em: <https://www.researchgate.net/publication/388669644>.

AYERS, R. S.; WESTCOT, D. W. Water quality for agriculture. **Rome: FAO**, 1985. (FAO Irrigation and Drainage Paper, 29 Rev.1).

FARZIN, M. *et al.* Assessment of ensemble models for groundwater potential modeling and prediction in a karst watershed. **Water**, v. 13, n. 18, p. 2540, 2021. Disponível em: <https://doi.org/10.3390/w13182540>.

GHOTO, S. M. *et al.* Mapping and assessing impacts of land use land cover and climate conditions on groundwater quality using RS & GIS. **Applied Water Science**, v. 15, art. 53, 2025. Disponível em: <https://doi.org/10.1007/s13201-024-02351-8>

GULGUNDI, M. S.; SHETTY, A. Groundwater quality assessment of urban Bengaluru using multivariate statistical techniques. **Applied Water Science**, v. 8, art. 43, 2018. Disponível em: <https://doi.org/10.1007/s13201-018-0684-z>.

KOZAR, M. D. *et al.* Factors affecting the distribution of water bearing fractures in bedrock aquifers: topographic, stratigraphic, lithologic, depth, and physiographic controls. **U.S. Geological Survey Scientific Investigations Report**, 2025. Disponível em: <https://pubs.usgs.gov/publication/sir20255044/full>.

NÉGRIEL, P. *et al.* Groundwater inputs towards surface water: quantification and impact on the river water quality using chemical and isotope fingerprints (e.g.,  $^{87}\text{Sr}/^{86}\text{Sr}$ ), example of the Loire River (France). In: **EGU General Assembly**, 2021. Disponível em: <https://meetingorganizer.copernicus.org/EGU21>.

RAHMATI, O. *et al.* Groundwater potential mapping using logistic regression, weights of evidence and evidential belief function models. **Hydrogeology Journal**, v. 24, n. 2, p. 425–446, 2016. Disponível em: <https://doi.org/10.1007/s10040-015-1348-8>.

RIBEIRO, L. O. A. Avaliação do sistema de monitoramento de águas subterrâneas do Estado de Minas Gerais: contribuições para a governança dos recursos hídricos. 2024. 144 f. Dissertação (Mestrado em Geografia) – Universidade Federal de Minas Gerais, Belo Horizonte, 2024. Disponível em: <https://repositorio.ufmg.br/handle/1843/77418>.

ROBLES, K. P. V.; MONJARDIN, C. E. F. Assessment and monitoring of groundwater contaminants in heavily urbanized areas: a review of methods and applications for Philippines. **Water**, v. 17, art. 1903, 2025. Disponível em: <https://doi.org/10.3390/w17131903>. Acesso em: 17 ago. 2025.

SANZ PÉREZ, E. *et al.* A procedure to estimate global natural recharge in karst aquifers. **Water**, v. 17, n. 12, art. 1779, 2025. Disponível em: <https://doi.org/10.3390/w17121779>. Acesso em: 18 ago. 2025.

SARKAR, S. K. *et al.* Future groundwater potential mapping using machine learning algorithms and climate change scenarios in Bangladesh. **Scientific Reports**, v. 14, art. 10328, 2024. Disponível em: <https://doi.org/10.1038/s41598-024-60560-2>. Acesso em: 18 ago. 2025.