

UNIVERSIDADE DE SÃO PAULO

ESCOLA POLITÉCNICA

KARINA OLIVEIRA FERRAZ SILVA

**Avaliação da potencial contaminação da água subterrânea na região hidrográfica  
do Alto do Tietê utilizando dados públicos de monitoramento ambiental**

Versão Original

São Paulo  
2024

KARINA OLIVEIRA FERRAZ SILVA

**Avaliação da potencial contaminação da água subterrânea na região hidrográfica  
do Alto do Tietê utilizando dados públicos de monitoramento ambiental**

Monografia apresentada à Escola Politécnica da Universidade de São Paulo como parte dos requisitos para a obtenção do título de Especialista em Gestão de Áreas Contaminadas, Desenvolvimento Urbano Sustentável e Revitalização de Brownfields.

Orientadora: Natália de Souza Pelinson

São Paulo

2024

## RESUMO

**SILVA, K. O. F. Avaliação da potencial contaminação da água subterrânea na região hidrográfica do Alto do Tietê utilizando dados públicos de monitoramento ambiental.** Monografia (MBA em Gestão de Áreas Contaminadas, Desenvolvimento Urbano Sustentável e Revitalização de Brownfields) – Escola Politécnica, Universidade de São Paulo, São Paulo, 2024.

A complexidade da gestão dos recursos hídricos na Região Metropolitana de São Paulo (RMSP) é acentuada pela densidade populacional que ultrapassa os 22 milhões de habitantes. Em face dos desafios impostos pela urbanização acelerada e as mudanças climáticas, a água subterrânea surge como uma alternativa para o abastecimento, destacando-se em meio à crescente escassez hídrica superficial. O monitoramento periódico da qualidade da água subterrânea é imperativo para identificar e controlar contaminações, especialmente em locais suscetíveis como postos de combustíveis, onde os componentes do petróleo podem apresentar potenciais riscos à saúde pública. A sistemática coleta de amostras, com ênfase em poços gerenciados pela Companhia Ambiental do Estado de São Paulo (CETESB) torna-se uma ferramenta valiosa na preservação da qualidade da água subterrânea, expandindo a perspectiva de gestão para além das fontes superficiais. Neste projeto, propõe-se a implementação de análises avançadas, incluindo métodos de aprendizado de máquina, para avaliação da qualidade da água na RMSP, Unidade de Gerenciamento de Recursos Hídricos do Alto do Tietê (UGRHI-6), apesar dos desafios decorrentes de lacunas nos dados e frequência de amostragem semestral. A integração de análises estatísticas quantitativas representa uma oportunidade para aprimorar a segurança hídrica, fornecendo subsídios para estratégias específicas visando o controle e prevenção de contaminações em postos de combustíveis e áreas circunvizinhas. Essa abordagem não apenas propicia os usos múltiplos da água subterrânea, mas também consolida a gestão responsável de áreas potencialmente contaminadas com base em dados específicos de qualidade do meio físico, com perspectiva de replicabilidade em diversas regiões do Brasil.

**Palavras-chave:** Água subterrânea; contaminação ambiental; métodos quantitativos estatísticos; Aprendizado de Máquina aplicado à Gestão de Áreas Contaminadas.

## ABSTRACT

SILVA, K. O. F. **Assessment of potential groundwater contamination in the Alto do Tietê watershed based on public environmental monitoring data.** Monograph (MBA in Contaminated Areas Management, Sustainable Urban Development, and Brownfield Revitalization) – Polytechnic School, University of São Paulo, São Paulo, 2024.

The complexity of water resources management in the Metropolitan Region of São Paulo (MRSP) is accentuated by a population density exceeding 22 million inhabitants. Confronted with challenges imposed by urbanization and climate change, groundwater emerges as an alternative for supply, standing out amidst the escalating scarcity of surface water. Regular monitoring of groundwater quality is imperative to identify and control contaminations, particularly in vulnerable locations such as fuel stations, where petroleum components may pose potential risks to public health. Systematic sample collection, with an emphasis on wells managed by the Environmental Company of the State of São Paulo (CETESB), becomes a valuable tool in preserving groundwater quality, broadening the management perspective beyond surface sources. In this research, the implementation of advanced analyses, including machine learning methods, is proposed for assessing water quality in the MRSP, specifically in the Alto do Tiete Water Resources Management Unit (UGRHI-6), despite challenges arising from data gaps and a semi-annual sampling frequency. The data analysis aims to strengthen historical understanding, providing insights for specific strategies targeting the control and prevention of contaminants in fuel stations and surrounding areas. The integration of quantitative statistical analyses presents an opportunity to enhance water security, allowing swift responses to losses in water quality through real-time environmental monitoring and control. This approach not only facilitates the multiple uses of groundwater but also solidifies sustainable water resources management and effective handling of potentially contaminated areas based on São Paulo-specific data, with prospects for replicability in various regions across Brazil.

**Keywords:** groundwater; environmental contamination; quantitative statistical methods; Machine Learning applied to management of contaminated sites.

## SUMÁRIO

<b>1 INTRODUÇÃO.....</b>	<b>5</b>
<b>2 OBJETIVOS.....</b>	<b>7</b>
2.1 Objetivo principal.....	7
2.2 Objetivos específicos.....	7
<b>3 REVISÃO BIBLIOGRÁFICA.....</b>	<b>8</b>
3.1 Contaminação de aquíferos.....	8
3.2 Legislação de contaminação de águas subterrâneas.....	9
3.3 Métodos de estatística de aprendizado de máquina aplicados para Águas Subterrâneas....	10
<b>4 MATERIAL E MÉTODOS.....</b>	<b>15</b>
4.1 Área de estudo.....	15
4.2 Coleta de dados públicos para análise de dados.....	15
4.3 Análise de dados: aplicação de métodos de aprendizado de máquina.....	16
<b>5 RESULTADOS E DISCUSSÃO.....</b>	<b>17</b>
<b>6 CONSIDERAÇÕES FINAIS.....</b>	<b>28</b>
<b>7 REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>29</b>

## 1 INTRODUÇÃO

A gestão sustentável dos recursos hídricos é uma preocupação atual em regiões altamente urbanizadas, como o Estado de São Paulo e, em particular, a Região Metropolitana de São Paulo (RMSP). Assim, dentre as 22 Unidades de Gerenciamento de Recursos Hídricos (UGRHI) em São Paulo, destaca-se a região do Alto do Tietê (UGRHI-6), como a mais populosa do Brasil, com mais de 22 milhões de habitantes.

A RMSP enfrenta desafios substanciais associados à escassez de água, sendo influenciada pela urbanização, mudanças climáticas e variabilidade hidrológica (Miyazaki; Venturi, 2022; Faro *et al.*, 2019). Nesse contexto, a água subterrânea, armazenada em aquíferos, surge como uma alternativa para complementar o abastecimento de água, considerando os usos múltiplos de abastecimento público, industrial e agrícola, especialmente, mas não somente, durante períodos de seca e escassez de água superficial.

O monitoramento sistemático da qualidade da água subterrânea, portanto, torna-se essencial para identificar precocemente possíveis contaminações e implementar medidas preventivas e corretivas para a manutenção da qualidade da água. Esta abordagem tem sido enfatizada em várias partes do mundo, com estudos como o realizado na Europa por Balderacchi *et al.* (2013), nos Estados Unidos por Rosecrans *et al.* (2021), e na Nigéria por Ighalo e Adeniyi (2020), ressaltando a importância do acompanhamento contínuo da qualidade das águas subterrâneas.

Isso requer a coleta regular de amostras em diferentes pontos estratégicos para a análise de parâmetros físico-químicos e microbiológicos. A amostragem dos poços de monitoramento da água subterrânea realizada pela CETESB, e disponibilizada anualmente em formato de relatórios, permite a identificação de possíveis contaminações e na preservação da qualidade da água subterrânea, ampliando assim a visão da gestão para além das fontes superficiais. Sendo assim, integrar o monitoramento desses poços ao contexto mais amplo da gestão de recursos hídricos é essencial para garantir a sustentabilidade a longo prazo.

O monitoramento contínuo da qualidade da água subterrânea, especialmente em áreas com presença significativa de postos de combustíveis, é essencial para detectar precocemente a presença de contaminantes derivados do petróleo (Gomes *et al.*, 2023). Os componentes do petróleo, como hidrocarbonetos aromáticos e compostos orgânicos voláteis presentes nos combustíveis, podem representar sérios riscos à qualidade da água subterrânea (Gomes *et al.*, 2023).

Vazamentos de tanques de armazenamento subterrâneo e práticas inadequadas de descarte de resíduos desses estabelecimentos podem resultar na infiltração desses poluentes, comprometendo a integridade do aquífero. A implementação de medidas preventivas e corretivas poderia evitar danos irreversíveis à qualidade da água, protegendo não apenas o abastecimento público, mas também a saúde humana e a preservação do meio ambiente.

O presente estudo busca analisar os dados de qualidade da água subterrânea da UGRHI-6 (Alto do Tietê) por meio de métodos de aprendizado de máquina supervisionados (regressões lineares) e não supervisionados (análise de componentes principais, métodos de agrupamentos e análises geoespaciais). Vale destacar que os dados de qualidade da água podem apresentar lacunas de amostragem e frequências não ideais para estudos com maior resolução e detalhamento de informações. De forma análoga, o banco disponibilizado pelo sistema InfoAguas (CETESB) foi analisado quanto aos aspectos de qualidade dos dados e às limitações práticas de análise das condições históricas de qualidade da água subterrânea.

Na região em análise, foram conduzidos estudos que destacam a importância da avaliação da qualidade da água subterrânea, conforme evidenciado por Bertolo *et al.* (2015) e Conicelli *et al.* (2021).

Ao considerar a gestão hídrica na região do Alto do Tietê, por ser imperativo incluir estratégias específicas para o controle e prevenção da contaminação por postos de combustíveis, análises utilizando métodos estatísticos quantitativos podem aumentar a segurança hídrica a longo prazo. Nessa área crítica do Estado de São Paulo, propostas como esta podem subsidiar respostas mais rápidas futuras com complementação de monitoramento ambiental em tempo real, com interpretação dos dados coletados diretamente em painéis de controle (*dashboards*).

O monitoramento da qualidade da água subterrânea e respostas eficientes à possíveis contaminações não apenas proporcionam os usos múltiplos deste recurso natural, mas também contribui para que uma base sólida para a tomada de decisões na gestão sustentável dos recursos hídricos em São Paulo, que pode futuramente ser aplicada a outras regiões brasileiras.

## 2 OBJETIVOS

### 2.1 Objetivo principal

O principal objetivo desta monografia é analisar a qualidade da água subterrânea na Unidade de Gerenciamento de Recursos Hídricos do Alto do Tietê (UGRHI-6), com enfoque na presença de componentes do petróleo, nitrato e metais verificando a viabilidade de métodos estatísticos (não-supervisionados).

### 2.2 Objetivos específicos

Dentre os objetivos específicos deste trabalho, podemos destacar:

- A partir de análise de dados em áreas potencialmente contaminadas, propor uma análise da qualidade da água subterrânea utilizando dados secundários públicos;
- Investigar a aplicação de análises estatísticas quantitativas para aprimorar a segurança hídrica, possibilitando o estabelecimento de novas estratégias para respostas rápidas diante de perdas na qualidade da água;
- Aprimorar a tomada de decisão em relação à gestão de recursos hídricos, permitindo respostas eficientes e garantindo a comunicação de riscos diante de alterações na qualidade da água subterrânea.

### **3 REVISÃO BIBLIOGRÁFICA**

#### **3.1 Contaminação de aquíferos**

Com o aumento populacional, houve a necessidade de uma maior exploração de recursos hídricos e, neste contexto, tornou-se comum o uso das águas subterrâneas para abastecimento humano, sobretudo por ser uma opção economicamente melhor. A água subterrânea é muito importante para o equilíbrio da dinâmica da infiltração e escoamento da água, bem como, há muitos anos vem servindo de reservatórios para futuras demandas e implicações por uma iminente falta de água (Franco, 2018).

Segundo a Associação Brasileira de Água Subterrâneas (ABAS), os aquíferos são formações geológicas do subsolo, constituídas por rochas permeáveis, que armazenam água em seus poros ou fraturas. Além dessa definição, a ABAS também conceitua os aquíferos como sendo materiais geológicos capazes de servirem de repositórios e transmissores da água ali armazenadas.

De acordo com João Manoel Filho, em sua participação no livro “Geologia - Conceitos e Aplicações, 3<sup>a</sup> Edição de 2008, a contaminação das águas subterrâneas podem ter origens diversas, sendo as mais comuns aquelas com relação direta com atividades industriais, domésticas e agrícolas.

As principais fontes de contaminação da água subterrânea, segundo Fetter (1993), ocorrem por meio de fossas sépticas, poços em geral, aplicações de fertilizantes no solo, aterros, vazamentos de óleos e combustíveis etc. Os principais tipos de contaminantes da água subterrânea, podem ser classificados em três tipos, segundo João Manoel Filho:

- Contaminantes Inorgânicos: nitrogênio agrícola, produtos radioativos, nitritos, nitratos e metais pesados;
- Contaminantes Orgânicos: pesticidas, aditivos de gasolina e produtos de petróleo (benzeno, tolueno, etilbenzeno e xileno);
- Contaminantes Biológicos: bactérias e vírus.

Segundo Varnier e Hirata (2002), o nitrato é o poluente de ocorrência mais frequente na água subterrânea. Devido a sua alta mobilidade, pode atingir áreas extensas.

Conforme mencionado por Lilian Caramel, na sua publicação sobre contaminação de aquíferos para o portal do Instituto de Geociências da Universidade de São Paulo, em 2023, segundo a OMS (Organização Mundial da Saúde), o nitrato quando apresenta concentrações superiores ao padrão de potabilidade, pode vir a gerar riscos à saúde humana, como a

metemoglobinemia (síndrome do bebê azul), alguns tipos de câncer e doenças no sistema reprodutivo, se ingerido.

Além dessas consequências, segundo documento publicado em 2002 por Álvaro Resende, pela EMBRAPA (Empresa Brasileira de Pesquisa Agropecuária), o excesso de nitrato na água pode causar impacto ao meio ambiente, como a eutrofização de mananciais, pois quando em alta concentrações, pode favorecer a proliferação de algas e plantas aquáticas, provocando redução da disponibilidade de oxigênio, e gerando a mortandade de peixes e organismos.

Para Oliveira e Loureiro (1998), a contaminação da água subterrânea por derivados de petróleo tem se revelado um problema ambiental grave, principalmente em países mais industrializados.

### **3.2 Legislação de contaminação de águas subterrâneas**

Em esfera nacional, a Resolução CONAMA nº 420/2009 é uma das principais normativas que estabelece critérios e diretrizes para o gerenciamento de áreas contaminadas por substâncias químicas e resíduos perigosos. A Tabela 1, apresenta alguns dos principais compostos químicos contaminantes, juntamente com seus valores orientadores para água subterrânea no Brasil.

Tabela 1 - Contaminantes e valores orientadores para água subterrânea no Brasil

<b>Composto</b>	<b>Fontes de Origem</b>	<b>Valor Orientador (<math>\mu\text{g/L}</math>)</b>
Nitratos	Agricultura	10
Nitritos	Agricultura	1,0
Atrazina	Uso de Pesticidas	2,0
Aldicarbe	Uso de Pesticidas	0,1
Clorpirifós	Uso de Pesticidas	2,0
Chumbo	Atividades Industriais, Mineração	10,0
Mercúrio	Atividades Industriais, Mineração	1,0
Cádmio	Atividades Industriais, Mineração	5,0
Cromo	Atividades Industriais, Mineração	50,0
Hidrocarbonetos	Vazamentos de tanques subterrâneos, atividades industriais	Variável

Além da legislação brasileira, a Agência de Proteção Ambiental dos Estados Unidos também possui diversas regulamentações e diretrizes relacionadas à contaminação da água subterrânea. Algumas das principais regulamentações e diretrizes da EPA incluem:

- Regulamento de Água Subterrânea: Esta regulamentação estabelece padrões de qualidade da água potável para sistemas de água pública que utilizam água subterrânea, visando proteger a saúde pública ao controlar a contaminação microbiológica na água potável proveniente de fontes subterrâneas.
- Regulamentação de Águas Subterrâneas Substancial: Este conjunto de regulamentações aborda a injeção subterrânea de fluidos para proteger as águas subterrâneas de contaminação proveniente de atividades como a injeção de resíduos industriais ou a recuperação de petróleo.
- Programa de Gerenciamento de Águas Subterrâneas: A EPA promove programas de gerenciamento de águas subterrâneas em colaboração com estados para monitorar, avaliar e proteger as águas subterrâneas de contaminação e esgotamento excessivo.
- Guia para Avaliação de Riscos de Contaminação de Águas Subterrâneas: A EPA fornece diretrizes e protocolos para avaliar e mitigar a contaminação da água subterrânea em locais contaminados, incluindo métodos de amostragem, avaliação de riscos e desenvolvimento de planos de remediação.

### **3.3 Métodos de estatística de aprendizado de máquina aplicados para Águas Subterrâneas**

O monitoramento ambiental apresenta um papel importante na avaliação e gestão da qualidade dos recursos naturais, em especial, em processos relacionados ao Gerenciamento de Áreas contaminadas, que demandam de investigação de meio físico para analisar se há ou não contaminação em uma área de interesse.

Nos últimos anos, a utilização de técnicas de aprendizado de máquina (do inglês, *Machine Learning* - ML) tem sido significativamente aprimorada, refletindo em análises de dados ambientais integradas, otimizando, entre outras aplicações, a análise da qualidade de águas subterrâneas (Haggerty *et al.*, 2023). De uma forma simplificada, a aplicação de métodos de *machine learning* para avaliação da qualidade da água subterrânea pode ser utilizada para identificar possíveis fontes de contaminação e garantir a gestão sustentável dos recursos hídricos (Xu *et al.*, 2024). Tal avaliação é um domínio complexo que demanda

técnicas analíticas avançadas para elucidar relações multifacetadas entre diversos parâmetros físico-químicos que podem ser considerados na qualidade natural de águas, desde os mais básicos como pH, condutividade elétrica e potencial de oxirredução, aos mais complexos, como por exemplo, a partir de isótopos estáveis.

Assim, podemos destacar que a versatilidade dos métodos de ML permite que pesquisadores e profissionais extraiam interpretações de conjuntos de dados obtidos em diferentes tipos de monitoramento ambiental, incluindo a possibilidade de análise de grandes volumes de dados. As técnicas estatísticas de *Machine Learning* abrangem duas metodologias principais: métodos de ML supervisionados e não supervisionados, sendo que a principal distinção consiste na natureza dos dados e nos objetivos gerais da análise a ser realizada.

No aprendizado de máquina supervisionado, o algoritmo desenvolvido é testado com dados rotulados, em situações que cada entrada está associada a uma saída correspondente e tem como objetivo “aprender uma função de mapeamento que possa prever com precisão a saída de dados novos e não observados” (Sarker, 2021). A premissa básica de haver variáveis dependentes e variáveis respostas faz com que as técnicas mais comuns para os métodos supervisionados sejam classificação e regressão, sendo a regressão linear, as árvores de decisão e as máquinas de vetores de suporte métodos consolidados internacionalmente (Taherdoost, 2023).

A regressão linear é uma técnica amplamente consolidada na modelagem preditiva e opera dentro do paradigma de aprendizagem supervisionada tem uma ampla aplicabilidade na previsão de resultados contínuos (Fávero; Belfiore; Souza, 2023). De forma complementar, as Máquinas de Vetores de Suporte (SVMs) e redes neurais representam ferramentas sofisticadas de ML supervisionado (Taherdoost, 2023). Taherdoost (2023) complementa ainda que as SVMs demonstram habilidade em tarefas de classificação, enquanto as redes neurais, particularmente, as arquiteturas de aprendizagem profunda (do inglês, *Deep Learning*) revelam capacidades incomparáveis no reconhecimento de padrões complexos, como evidenciado na classificação de imagens e no processamento de linguagem natural.

Por outro lado, a aprendizagem de máquina não supervisionada, em geral, é aplicada para dados não rotulados, com o objetivo de descobrir padrões, estruturas ou relacionamentos inerentes ao conjunto de dados (Fávero; Belfiore; Souza, 2023). Os métodos de ML não-supervisionados tornou-se comumente adotados para tarefas como agrupamento (*clustering*), redução de dimensionalidade (análise multivariada) e detecção de anomalias. Assim, embora métodos de ML supervisionados sejam excelentes para análise com dados rotulados e objetivos bem definidos, a aprendizagem não supervisionada é uma ferramenta

valiosa para explorarmos padrões inerentes em conjuntos de dados não rotulados, sem uma variável de resposta (Gallatin; Albon, 2023).

Na avaliação da qualidade da água subterrânea, a presença de variáveis interconectadas demanda abordagens analíticas que vão além das técnicas tradicionais isoladas. Neste contexto, as análises multivariadas constituem uma classe de métodos estatísticos que exploram simultaneamente as relações entre múltiplas variáveis, oferecendo uma visão integrada das complexas interações (Haggerty *et al.*, 2023), comumente presentes nos dados ambientais. O emprego de análises multivariadas nesse contexto pode oferecer uma abordagem robusta e abrangente para identificar padrões e correlações inerentes de parâmetros da qualidade dos meios físicos monitorados.

Dentre as técnicas de análises multivariadas, podemos destacar a Análise de Componentes Principais (PCA), que é uma técnica estatística amplamente empregada para a avaliação de conjuntos de dados fornecendo a identificação de padrões e as relações entre diferentes variáveis (Taherdoost, 2023). Vale destacar que a PCA é uma técnica clássica de aprendizado de máquina não supervisionado focada na redução de dimensionalidade, que transforma variáveis originais em componentes principais não correlacionados, simplificando conjuntos de dados complexos ao reter informações essenciais em um número reduzido de componentes (Patnaik; Tudu; Bagal, 2024).

Os componentes principais resultantes são ordenados em termos de sua contribuição para a variância total, destacando as principais fontes de variabilidade nos dados de água subterrânea (Gallatin; Albon, 2023). A interpretação desses componentes permite a identificação de padrões dominantes e a compreensão das contribuições individuais de cada variável para a variabilidade observada.

Antes de aplicar a PCA, realiza-se uma análise de correlação para compreender as relações entre diferentes parâmetros de qualidade da água. Essa etapa identifica possíveis redundâncias e avalia a “força” das conexões entre variáveis (Fávero; Belfiore; Souza, 2023). A PCA pode ser aprimorada incorporando informações espaciais. Gráficos Scree-plots auxiliam na determinação do número de componentes principais a serem retidos, indicando onde componentes adicionais contribuem menos significativamente para a variância geral.

Os autovalores (eigenvalues) representam a quantidade de variância explicada por cada componente principal, e os autovetores indicam a direção e a força da correlação entre as variáveis originais e o componente principal. A PCA busca maximizar a variância explicada pelos primeiros componentes principais. A variância cumulativa explicada por todos os componentes fornece insights sobre a informação geral retida do conjunto de dados original

(Gallatin; Albon, 2023). Os componentes principais são ortogonais e não correlacionados, assegurando que cada componente capture um aspecto único da variabilidade do conjunto de dados. Os fatores são ordenados pela quantidade de variância que explicam, sendo o primeiro componente aquele que explica a maior variância. No contexto do monitoramento da qualidade da água subterrânea, vários parâmetros, como pH, oxigênio dissolvido, nitratos, condutividade, cloretos e concentrações de metais pesados, são considerados. Esses parâmetros definem coletivamente a natureza multidimensional dos conjuntos de dados de qualidade da água.

Gráficos de carga são utilizados para interpretar as contribuições de cada parâmetro de qualidade da água para os componentes principais (Garbas; Idris; Gambo, 2023). Valores de carga elevados indicam uma forte influência em um componente específico, auxiliando na identificação de variáveis-chave que moldam a variabilidade observada.

Em análises de água subterrânea, a distribuição geográfica das localizações dos poços de monitoramento pode ser analisada junto com os escores dos componentes principais, fornecendo um contexto espacial aos padrões identificados. A PCA auxilia na identificação de padrões dominantes e possíveis fontes de contaminação em conjuntos de dados de qualidade da água subterrânea (Haggerty *et al.*, 2023). Escores elevados em componentes principais específicos podem indicar fontes de contaminação ou poluição.

Desta forma, ao reduzir a dimensionalidade dos dados e revelar padrões subjacentes, a PCA contribui com interpretações valiosas para um monitoramento eficaz da qualidade da água subterrânea e identificação de fontes de contaminação (Deng; Ye; Du, 2023; Garba, Idris; Gambo, 2023; Patnaik; Tudu; Bagal, 2024). PCAs podem revelar agrupamentos de poços de monitoramento com perfis de contaminação semelhantes ou destacar as principais fontes de variabilidade nos parâmetros de qualidade da água subterrânea (Patnaik; Tudu; Bagal, 2024).

A abrangência da aprendizagem não supervisionada é, geralmente, também representada por algoritmos de agrupamento, com destaque agrupamento K-Means (Fávero; Belfiore; Souza, 2023). O *clustering* K-Means é um algoritmo não supervisionado representativo que apresenta os dados em clusters 'k' com base na similaridade entre os atributos dos dados. Cada agrupamento (ou *cluster*) pode ser representado por seu centróide, por exemplo, e o algoritmo atribui iterativamente pontos de dados ao centróide mais próximo, atualizando os centróides até a convergência.

O processo de integração envolve a vinculação dos dados dos poços de monitoramento com as localizações geográficas de locais contaminados conhecidos. A análise espacial

usando sistemas de informações geográficas (SIG) facilita a identificação de similaridades entre pontos de monitoramento (McHaffie; Hwang; Follet, 2023); em especial, isso pode ser aplicado a poços de monitoramento com baixa qualidade da água e as áreas contaminadas listadas. Além disso, a incorporação de informações espaciais enriquece a análise multivariada, especialmente na identificação de padrões geográficos, possibilitando a visualização espacial dos resultados da PCA e de outros métodos, tais como a clusterização de amostragens.

Apesar dos benefícios, a aplicação de análises multivariadas enfrenta desafios, incluindo a qualidade dos dados, a interpretabilidade dos resultados e a necessidade de considerar a variabilidade temporal. Entretanto, os modelos de aprendizado de máquina podem ser treinados para prever a probabilidade de contaminação em poços não monitorados com base em padrões observados no conjunto de dados analisado (Deng; Ye; Du, 2023; Patnaik; Tudu; Bagal, 2024).

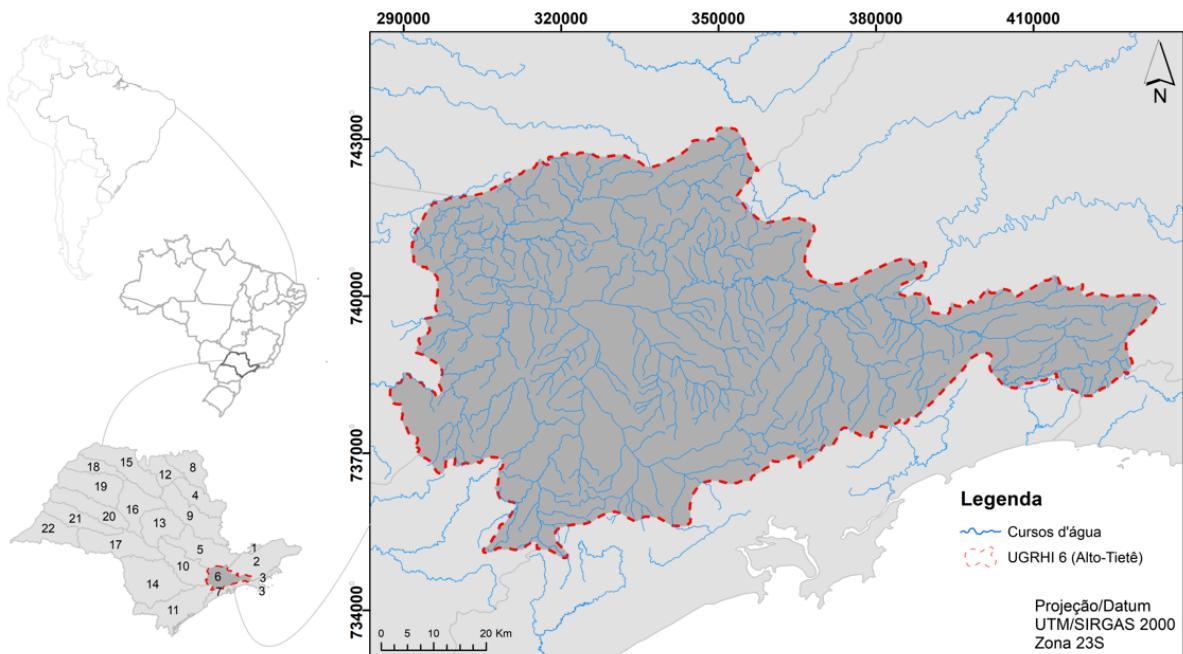
Desta forma, podemos observar que o aprendizado de máquina aplicado à água subterrânea tem se desenvolvido grandemente nos últimos anos (Haggerty *et al.*, 2023). Os avanços recentes testemunham a convergência de técnicas supervisionadas e não supervisionadas, além do desenvolvimento de modelos híbridos. Vale evidenciar que há uma aplicação ML comum para a avaliação de águas naturais por diferentes métodos visando a elaboração de índices de qualidade (Jayaraman *et al.*, 2024), tanto para águas superficiais (Costa *et al.*, 2020) quanto para águas subterrâneas (Abdelaziz *et al.*, 2020). Por fim, cabe ressaltar que pesquisas, como a de Jayaraman *et al.* (2024), fortalecem a utilização de ML em conjunto com outras metodologias, desde a aplicação da Internet das Coisas (IoT) para automatização do monitoramento da qualidade de águas, até análises de cenários relacionados às mudanças climáticas.

## 4 MATERIAL E MÉTODOS

### 4.1 Área de estudo

O escopo deste projeto engloba a análise da Unidade de Gerenciamento de Recursos Hídricos Alto do Tietê (UGRHI-06), situada no estado de São Paulo, Brasil (Figura 1). A UGRHI-06, composta por 34 municípios paulistas, incluindo grandes cidades como São Paulo, Guarulhos e Santo André, enfrenta desafios significativos relacionados à contaminação das águas subterrâneas. A complexidade dessa área de estudo é acentuada pela ocupação não-organizada do território e pelo despejo de águas residuárias tratadas ou não, especialmente provenientes de áreas residenciais humanas, atividades industriais e postos de combustíveis.

**Figura 1** - Área de estudo localizada no estado de São Paulo, Brasil.



### 4.2 Coleta de dados públicos para análise de dados

Os dados de qualidade da água subterrânea coletados na plataforma InfoAguas, criada e gerenciada pela Companhia Ambiental do Estado de São Paulo (CETESB) para disponibilização de dados de monitoramento semestral que ocorre de forma contínua. As variáveis de qualidade coletadas foram desde parâmetros básicos de qualidade da água, tais como pH, condutividade elétrica e turbidez, até analitos mais específicos tais como benzeno, tolueno, cloretos e nitratos.

### **4.3 Análise de dados: aplicação de métodos de aprendizado de máquina**

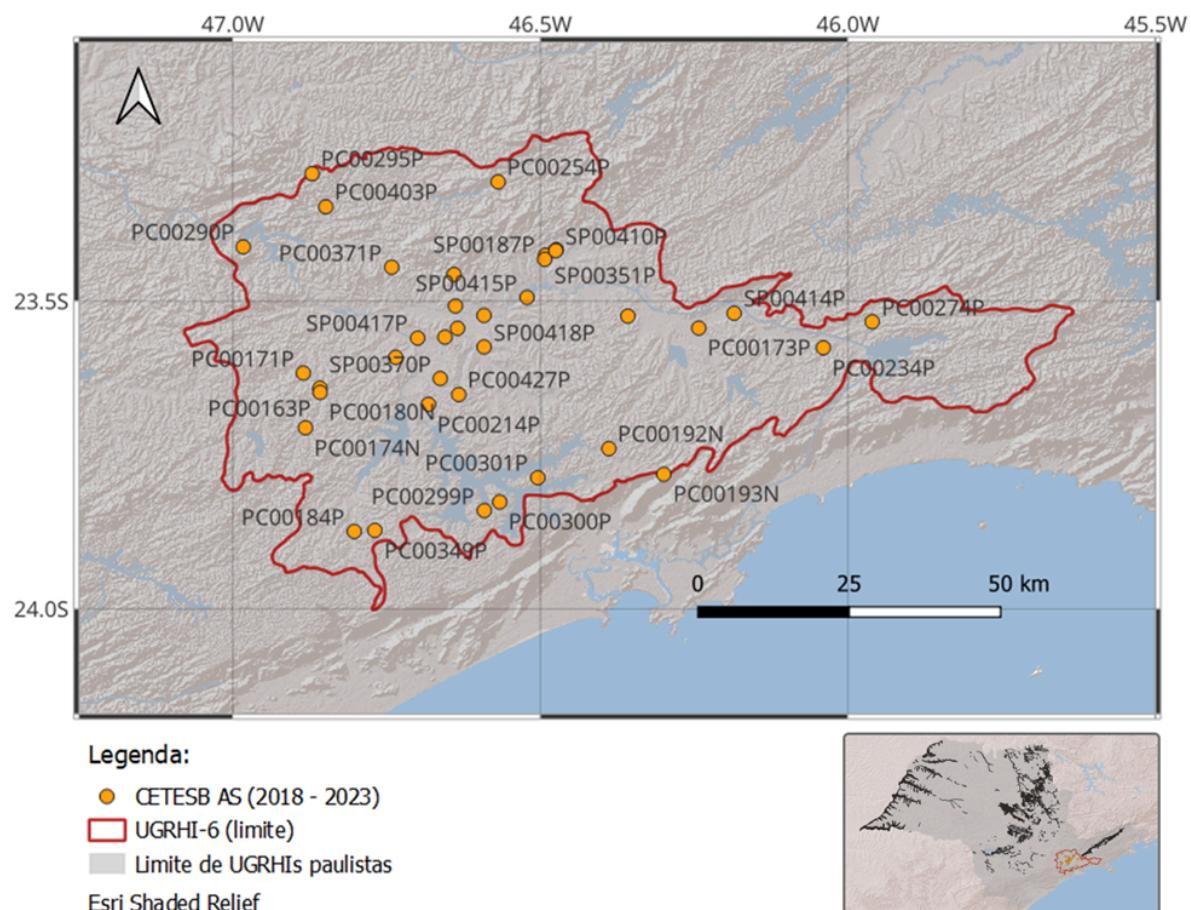
A aplicação da ciência de dados revela-se uma ferramenta útil em pesquisas ambientais, especialmente na avaliação da qualidade de águas subterrâneas, com foco nas possíveis contaminações provenientes de postos de combustíveis. A metodologia adotada neste estudo compreende duas etapas distintas. Na primeira fase, empregou-se a análise geoespacial para mapear a variabilidade regional das águas subterrâneas, considerando dados do Alto do Tietê gerados pelo monitoramento realizado pela CETESB. A partir da elaboração de mapas temáticos, foi possível analisar o uso e cobertura da terra, com dados disponíveis pela Secretaria do Meio Ambiente do Estado de São Paulo (Secretaria de Meio Ambiente, Infraestrutura e Logística do Estado de SP (Semil).

Enquanto, na segunda etapa, houve a aplicação de uma análise de componentes principais (ACP) e análise de correlação das variáveis, ambos utilizando a linguagem R, aplicando para parâmetros de qualidade da água subterrânea disponibilizadas publicamente, com a finalidade de se avaliar as variáveis significantes a esse monitoramento ambiental específico.

## 5 RESULTADOS E DISCUSSÃO

A Unidade de Gerenciamento de Recursos Hídricos (UGRHI-06) destaca-se por sua posição geográfica, abrangendo a Região Metropolitana de São Paulo (RMSP), abarcando a problemática de potenciais contaminações industriais e por postos de combustíveis que possivelmente contribuem para a perda da qualidade das águas subterrâneas nesta região altamente populosa. A Companhia Ambiental do Estado de São Paulo (CETESB) realiza o monitoramento de 37 poços quanto à qualidade da água subterrânea na UGRHI-6, cujos pontos podem ser observados na Figura 2.

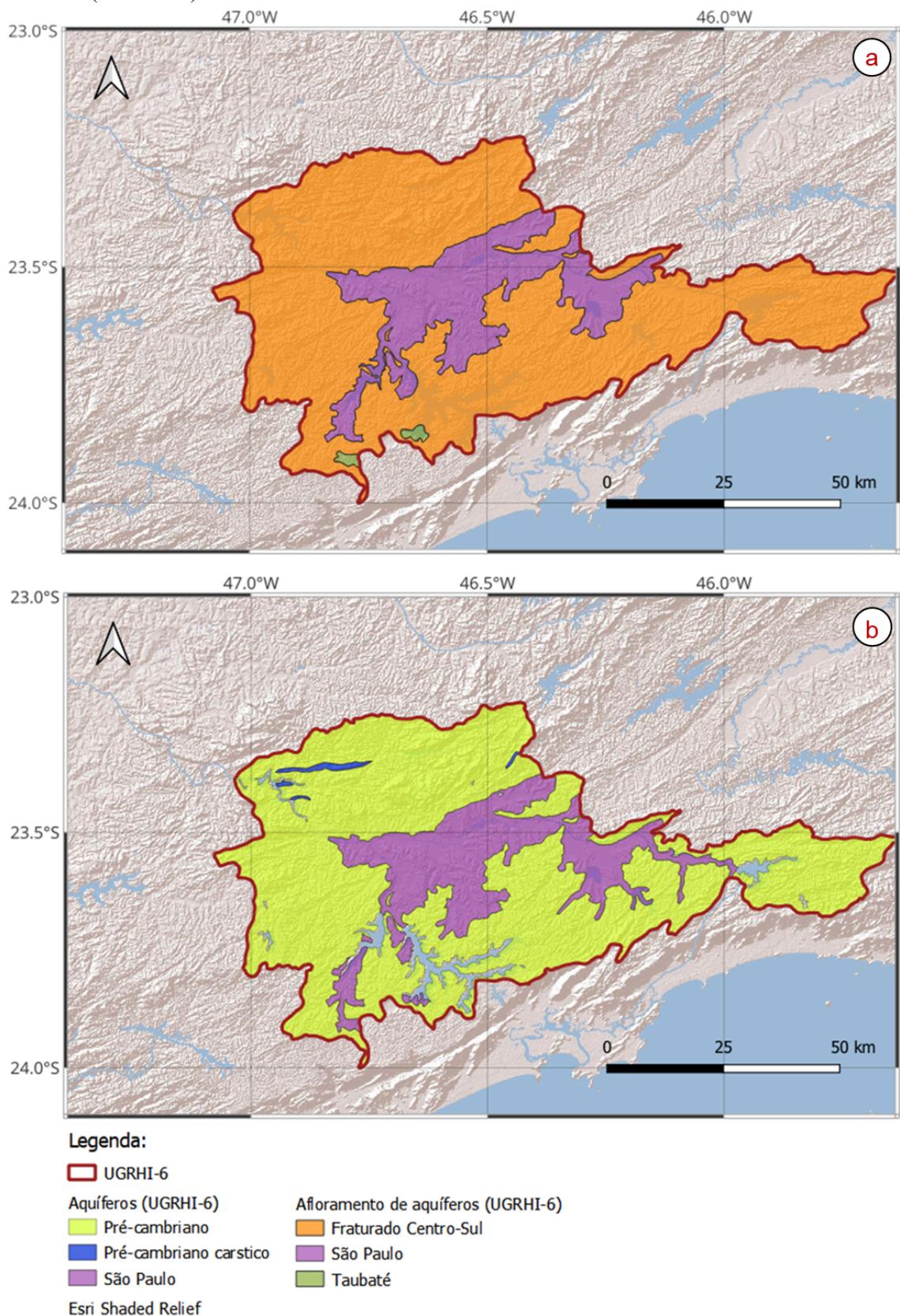
**Figura 2** - Mapa de localização de poços de monitoramento de águas subterrâneas disponibilizados pela CETESB na plataforma do InfoAguas com detalhe (no canto inferior direito) do mapa de vulnerabilidade de aquíferos.



**Fonte:** Autoria própria (Dados do InfoAguas).

A litologia diversificada da Região Metropolitana de São Paulo (RMSP) é marcada pela predominância de estratos argilosos e entremeados por depósitos de areia que se distribuem de maneira irregular na parte central da UGRHI-6 (Rocha, 2005), coincidente com a porção de Aquífero São Paulo (Figura 3).

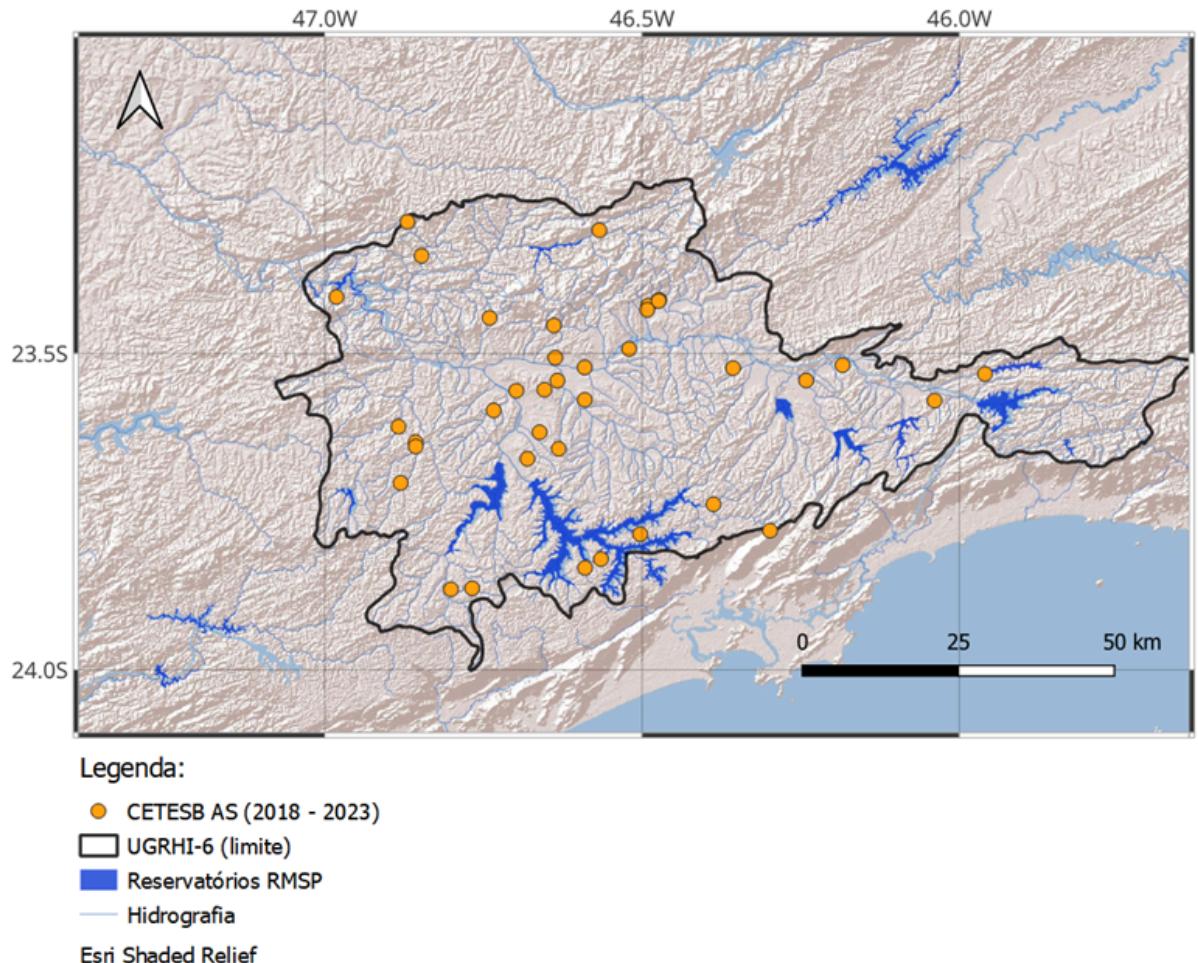
**Figura 3** - Mapa dos tipos de aquíferos da Unidade de Gerenciamento de Recursos Hídricos do Alto Tietê - SP (UGRHI-6).



**Fonte:** Autoria própria (Dados disponíveis: a) Instituto Geológico & Secretaria de Meio Ambiente, Infraestrutura e Logística do Estado de SP - Semil; e b) LEBAC - Laboratório de Estudos de Bacias - LEBAC-UNESP/Rio Claro-SP.

Em complementação, no mapa da Figura 4, pode ser observada a localização dos reservatórios superficiais que podem ser relevantes para análises futuras sobre a interação rio-aquífero.

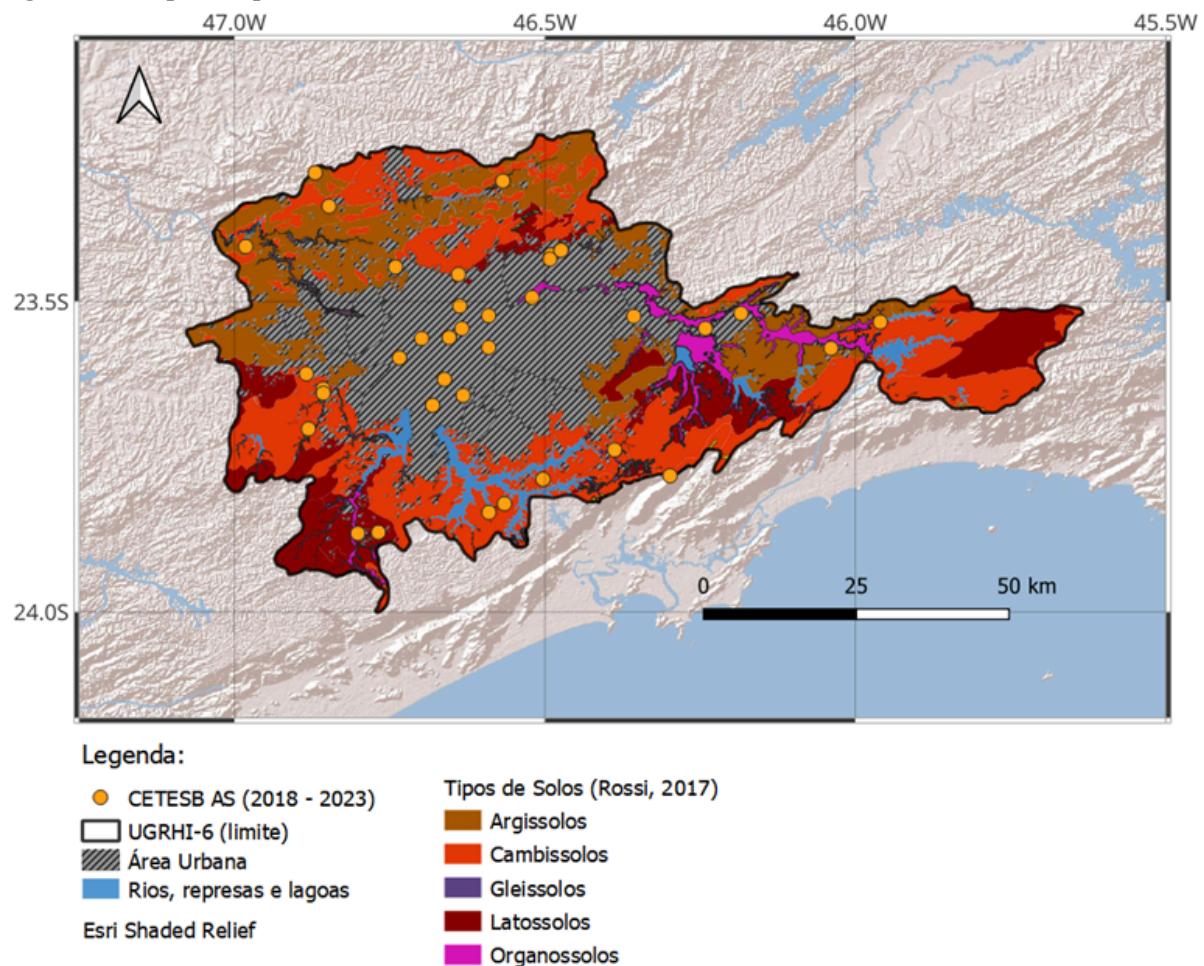
**Figura 4** - Localização dos principais reservatórios de água superficial na UGRHI-6.



**Fonte:** Autoria própria (Acervo pessoal Natália de Souza Pelinson).

Os tipos de solo podem ser classificados principalmente como argissolo, latossolo e cambissolo (Figura 5). É importante observar que a porção central da UGRHI-6 é altamente urbanizada e não possui seu solo classificado por Rossi (2017), indicando uma possível impermeabilização do solo na área.

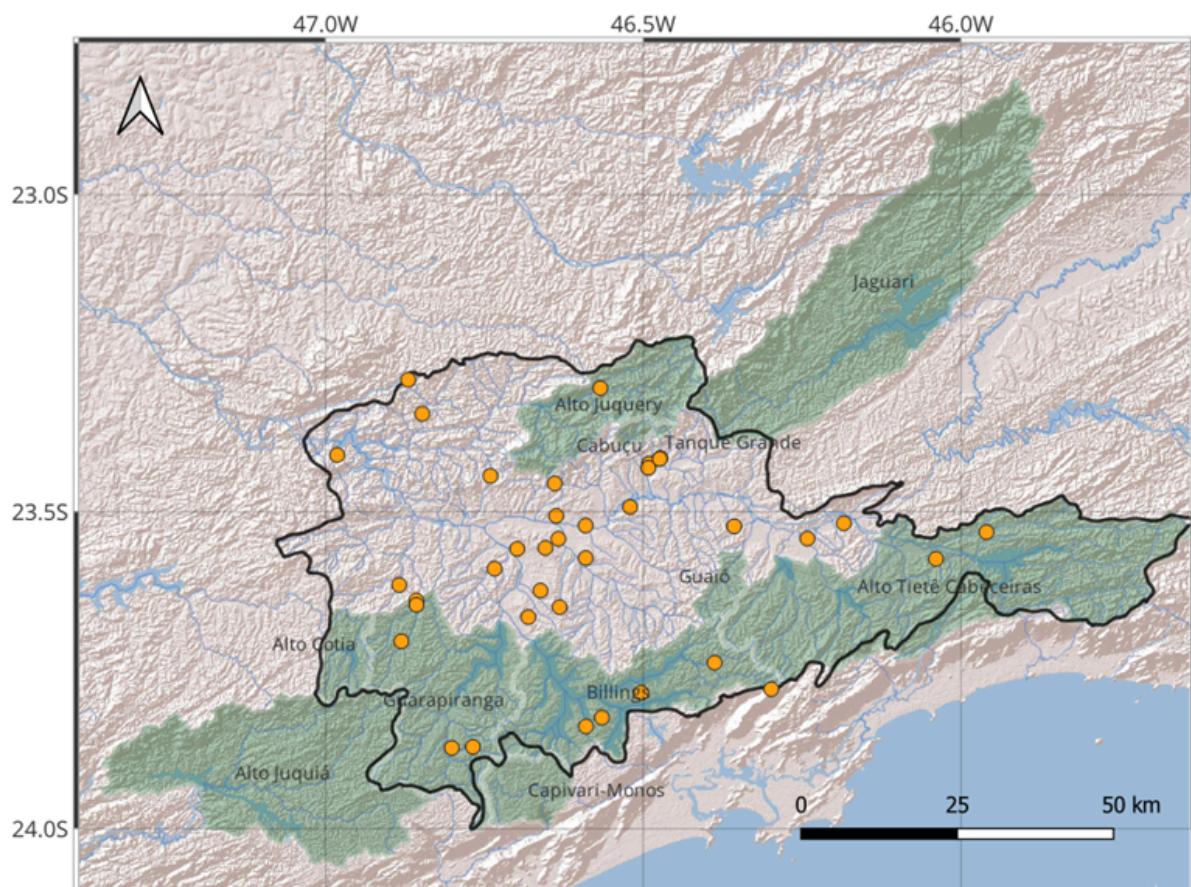
**Figura 5** - Mapa de tipos de solo evidenciando a área urbanizada da UGRHI-6.



**Fonte:** Adaptado de Rossi (2017).

A localização das Áreas de Proteção e Recuperação dos Mananciais (APRM) é apresentada na Figura 6. A área central do mapa, onde se encontra a maioria dos pontos monitorados, está ausente de APRM, o que demonstra que a proteção e recuperação dessas áreas não está sendo priorizada na região central.

**Figura 6 - Localização da APRM na UGRHI-6.**



**Legenda:**

- CETESB AS (2018 - 2023)
- ◻ UGRHI-6 (limite)
- Hidrografia
- APRM (UGRHI-6)

Esri Shaded Relief

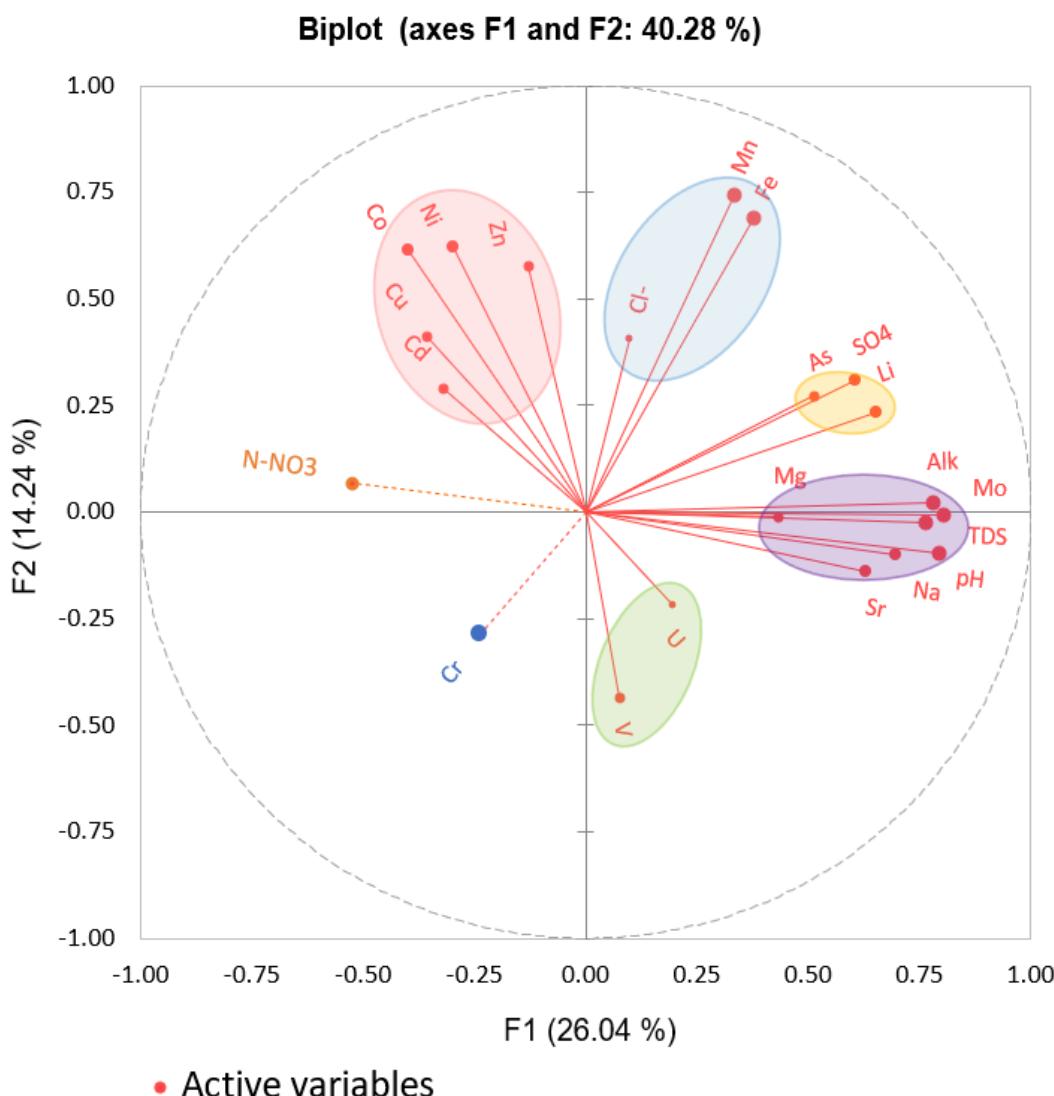
**Fonte:** Comitê de Bacia do Alto do Tietê (*on-line*).

Os principais dados usados neste trabalho foram retirados de relatórios de monitoramento de água da Companhia Ambiental do Estado de São Paulo (CETESB), acessados por meio da plataforma InfoAguas (CETESB).

Lembrando que a PCA é um método não supervisionado amplamente adotado que se destaca por preservar informações essenciais e reduzir o número de variáveis (Fávero; Belfiore; Souza, 2023). A sua aplicação na análise de dados ambientais ilustra a sua eficácia na identificação de padrões e relações dominantes. Ao transformar as variáveis originais em um conjunto de componentes principais não correlacionados, a PCA simplifica a interpretação de conjuntos de dados complexos. No contexto da análise da qualidade da água subterrânea, a PCA pode ajudar a identificar padrões dominantes e relações entre variáveis. Além disso, os

componentes principais identificados pela PCA podem servir como indicadores de contaminantes específicos ou fontes de poluição. Isso não apenas auxilia na compreensão do estado atual da qualidade da água subterrânea, mas também fornece estratégias açãoáveis para práticas mais eficazes de gestão ambiental.

**Figura 7** - Análise de Componentes Principais (PCA) utilizando dados de águas subterrâneas da UGRHI-6.



**Fonte:** Autoria própria (Acervo pessoal Natália de Souza Pelinson).

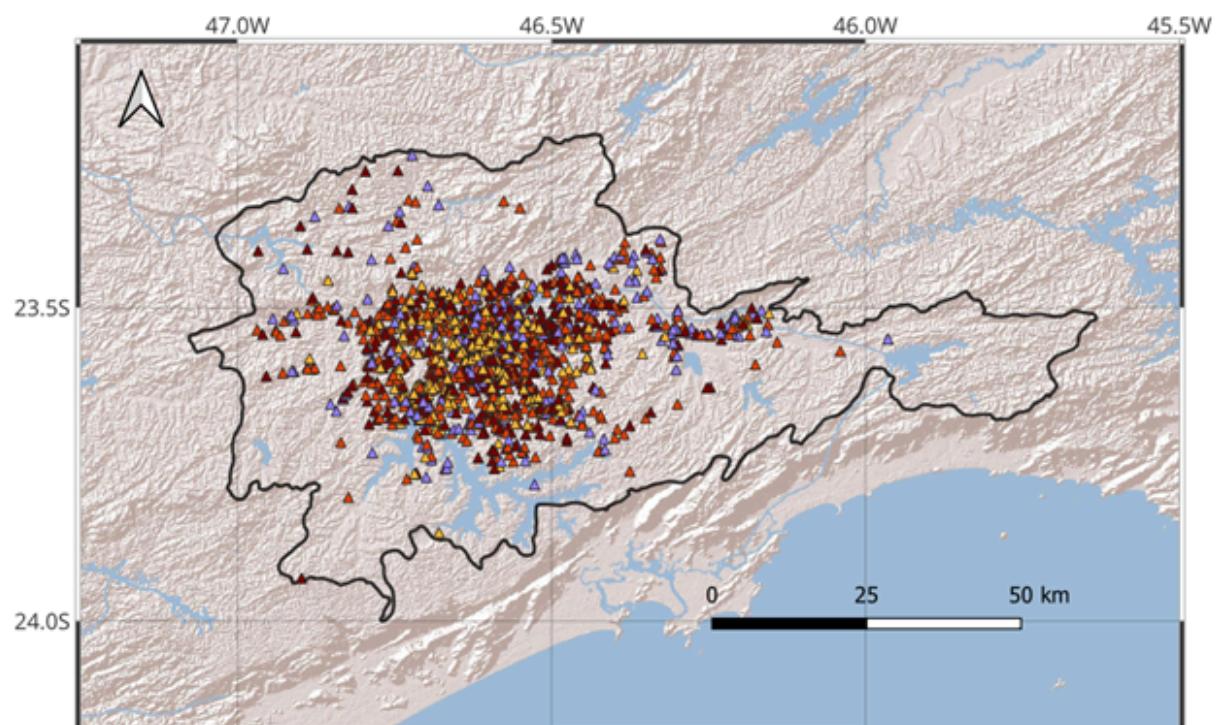
Note que a PCA tem suas bases em álgebra linear e conceitos estatísticos, visando transformar um conjunto de variáveis correlacionadas em um novo conjunto de variáveis não correlacionadas, conhecidas como componentes principais. O objetivo principal é reduzir a dimensionalidade do conjunto de dados, mantendo o máximo possível da variabilidade

original. Isso é alcançado por meio da extração de componentes principais, que são combinações lineares das variáveis originais.

Os valores do cosseno quadrado servem como indicadores da qualidade de representação das variáveis do estudo, enfatizando a contribuição de um componente através da distância quadrática da observação à origem (ABDI; WILLIAMS, 2010). Consequentemente, o cosseno quadrado do ângulo obtido pode revelar os componentes com valores elevados de  $\cos^2$  que dão contribuições mais substanciais à análise.

Apesar de, pela análise da PCA, não ficar caracterizada uma região crítica de preocupação a partir do monitoramento de águas subterrâneas, é possível observar, na Figura, 8 que há inúmeras áreas já catalogadas pela Lista de Áreas Contaminadas da CETESB (2020) como sendo contaminadas, com risco confirmado e até mesmo em processo de reutilização.

**Figura 8 -** Lista de Áreas Contaminadas da CETESB (2020).



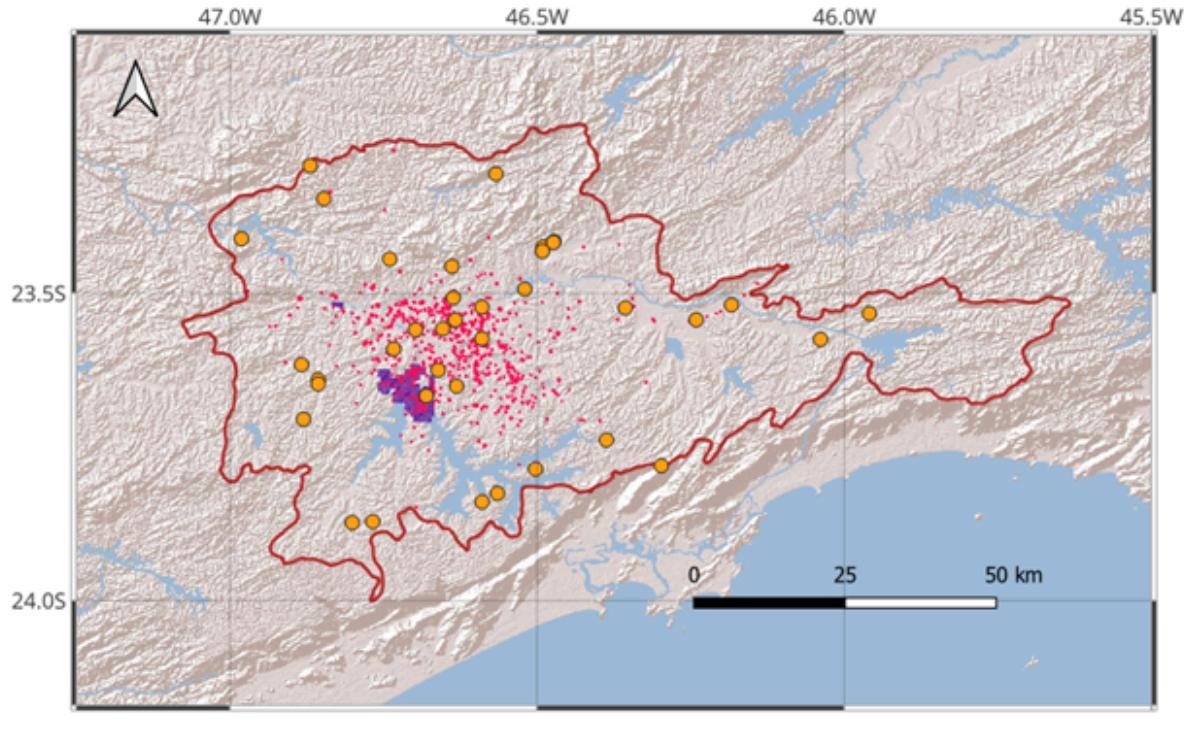
#### Legenda:

- Classificação das áreas contaminadas
- ▲ Área Contaminada com Risco Confirmado (ACRi)
  - ▲ Área Contaminada em Processo de Remediação (ACRe)
  - ▲ Área Contaminada em Processo de Reutilização (ACRu)
  - ▲ Área Contaminada sob Investigação (ACI)
- UGRHI-6 (limite)
- Esri Shaded Relief

**Fonte:** Lista de Áreas Contaminadas (2020).

Áreas em que a Companhia Ambiental do Estado de São Paulo e o Departamento de Águas e Energia Elétrica - DAEE indicaram como uso restrito para água subterrânea no ano de 2021.

**Figura 9** - Áreas de uso restrito de águas subterrâneas na região da UGRHI 6.



**Legenda:**

- CETESB AS (2018 - 2023)
  - Áreas de Restrição (CETESB)
  - Áreas de Restrição (DAAE)
  - UGRHI-6 (limite)
- Esri Shaded Relief

De forma a complementar à PCA obtida, podemos ainda analisar a correlação entre as variáveis aplicando a Correlação de Spearman (Tabela 2). As correlações positivas ficam evidenciadas em tons de azul, enquanto as correlações negativas estão marcadas em tons de vermelho.

**Tabela 2** - Análise de Correlação de Spearman dos parâmetros físico-químicos mensurados na UGRHI Alto do Tietê.

Variáveis	Alk	As	Cd	Cl-	Co	Cu	Cr	Sr	Fe	Li	Mg	Mn	Mo	Ni	N-NO <sub>3</sub>	pH	Na	TDS	SO4	U	V	Zn
Alk	<b>1.00</b>	<b>0.36</b>	<b>-0.16</b>	0.07	<b>-0.22</b>	<b>-0.17</b>	-0.02	<b>0.45</b>	<b>0.32</b>	<b>0.44</b>	<b>0.38</b>	<b>0.23</b>	<b>0.62</b>	-0.07	<b>-0.30</b>	<b>0.77</b>	<b>0.39</b>	<b>0.75</b>	<b>0.28</b>	<b>0.20</b>	<b>0.21</b>	-0.03
As	<b>0.36</b>	<b>1.00</b>	-0.09	0.10	<b>-0.15</b>	-0.02	<b>-0.15</b>	0.06	<b>0.35</b>	<b>0.46</b>	<b>0.15</b>	<b>0.32</b>	<b>0.40</b>	-0.04	<b>-0.32</b>	<b>0.39</b>	0.22	<b>0.31</b>	<b>0.29</b>	-0.04	-0.09	0.02
Cd	<b>-0.16</b>	-0.09	<b>1.00</b>	<b>0.18</b>	<b>0.39</b>	<b>0.16</b>	0.12	<b>-0.26</b>	0.02	<b>-0.22</b>	<b>-0.19</b>	0.02	-0.11	<b>0.20</b>	0.13	<b>-0.16</b>	<b>-0.27</b>	<b>-0.21</b>	-0.06	-0.01	-0.07	<b>0.15</b>
Cl-	0.07	0.10	<b>0.18</b>	<b>1.00</b>	<b>0.19</b>	<b>0.17</b>	<b>-0.14</b>	0.03	0.13	<b>0.17</b>	0.12	<b>0.25</b>	0.08	<b>0.20</b>	<b>0.33</b>	-0.12	0.03	<b>0.17</b>	<b>0.46</b>	<b>0.19</b>	-0.05	-0.05
Co	<b>-0.22</b>	<b>-0.15</b>	<b>0.39</b>	<b>0.19</b>	<b>1.00</b>	<b>0.25</b>	-0.01	<b>-0.31</b>	<b>0.23</b>	<b>-0.14</b>	<b>-0.23</b>	<b>0.32</b>	<b>-0.24</b>	<b>0.52</b>	0.27	<b>-0.32</b>	<b>-0.33</b>	<b>-0.24</b>	-0.07	0.02	-0.10	<b>0.38</b>
Cu	<b>-0.17</b>	-0.02	<b>0.16</b>	<b>0.17</b>	<b>0.25</b>	<b>1.00</b>	0.09	-0.10	0.05	-0.06	0.06	-0.08	<b>-0.37</b>	<b>0.55</b>	<b>0.28</b>	<b>-0.28</b>	<b>-0.21</b>	<b>-0.26</b>	-0.11	-0.09	-0.04	<b>0.43</b>
Cr	-0.02	<b>-0.15</b>	0.12	<b>-0.14</b>	-0.01	0.09	<b>1.00</b>	-0.06	-0.13	<b>-0.23</b>	<b>0.15</b>	<b>-0.26</b>	<b>-0.20</b>	0.03	<b>0.16</b>	<b>-0.13</b>	-0.11	-0.10	<b>-0.37</b>	0.04	<b>0.47</b>	0.05
Sr	<b>0.45</b>	0.06	<b>-0.26</b>	0.03	<b>-0.31</b>	-0.10	-0.06	<b>1.00</b>	0.09	<b>0.32</b>	<b>0.76</b>	0.05	<b>0.30</b>	<b>-0.15</b>	<b>-0.19</b>	<b>0.35</b>	<b>0.82</b>	<b>0.39</b>	<b>0.33</b>	0.01	0.10	-0.06
Fe	<b>0.32</b>	<b>0.35</b>	0.02	0.13	<b>0.23</b>	0.05	-0.13	0.09	<b>1.00</b>	<b>0.30</b>	<b>0.16</b>	<b>0.75</b>	<b>0.26</b>	<b>0.17</b>	<b>-0.26</b>	<b>0.24</b>	0.17	<b>0.25</b>	<b>0.29</b>	<b>-0.28</b>	<b>-0.20</b>	<b>0.39</b>
Li	<b>0.44</b>	<b>0.46</b>	<b>-0.22</b>	<b>0.17</b>	<b>-0.14</b>	-0.06	<b>-0.23</b>	<b>0.32</b>	<b>0.30</b>	<b>1.00</b>	<b>0.24</b>	<b>0.22</b>	<b>0.45</b>	0.07	<b>-0.35</b>	<b>0.43</b>	<b>0.29</b>	<b>0.42</b>	<b>0.61</b>	<b>0.32</b>	0.09	0.12
Mg	<b>0.38</b>	<b>0.15</b>	<b>-0.19</b>	0.12	<b>-0.23</b>	0.06	<b>0.15</b>	<b>0.76</b>	<b>0.16</b>	<b>0.24</b>	<b>1.00</b>	0.12	0.07	-0.04	0.12	<b>0.16</b>	<b>0.59</b>	<b>0.29</b>	0.12	<b>-0.16</b>	0.05	-0.01
Mn	<b>0.23</b>	<b>0.32</b>	0.02	<b>0.25</b>	<b>0.32</b>	-0.08	<b>-0.26</b>	0.05	<b>0.75</b>	<b>0.22</b>	0.12	<b>1.00</b>	<b>0.28</b>	<b>0.18</b>	-0.12	<b>0.15</b>	<b>0.19</b>	<b>0.23</b>	<b>0.34</b>	<b>-0.26</b>	<b>-0.46</b>	<b>0.29</b>
Mo	<b>0.62</b>	<b>0.40</b>	-0.11	0.08	<b>-0.24</b>	<b>-0.37</b>	<b>-0.20</b>	<b>0.30</b>	<b>0.26</b>	<b>0.45</b>	0.07	0.28	<b>1.00</b>	<b>-0.24</b>	<b>-0.45</b>	<b>0.74</b>	<b>0.49</b>	<b>0.63</b>	<b>0.47</b>	<b>0.35</b>	0.05	-0.13
Ni	-0.07	-0.04	<b>0.20</b>	<b>0.20</b>	<b>0.52</b>	<b>0.55</b>	0.03	<b>-0.15</b>	<b>0.17</b>	0.07	-0.04	<b>0.18</b>	<b>-0.24</b>	<b>1.00</b>	<b>0.27</b>	<b>-0.25</b>	<b>-0.25</b>	<b>-0.18</b>	-0.02	0.00	-0.07	<b>0.47</b>
N-NO <sub>3</sub>	<b>-0.30</b>	<b>-0.32</b>	0.13	<b>0.33</b>	<b>0.27</b>	<b>0.28</b>	<b>0.16</b>	<b>-0.19</b>	<b>-0.26</b>	<b>-0.35</b>	0.12	-0.12	<b>-0.45</b>	<b>0.27</b>	<b>1.00</b>	<b>-0.50</b>	<b>-0.26</b>	<b>-0.25</b>	<b>-0.31</b>	0.05	0.02	0.05
pH	<b>0.77</b>	<b>0.39</b>	<b>-0.16</b>	-0.12	<b>-0.32</b>	<b>-0.28</b>	<b>-0.13</b>	<b>0.35</b>	<b>0.24</b>	<b>0.43</b>	<b>0.16</b>	<b>0.15</b>	<b>0.74</b>	-0.25	<b>-0.50</b>	<b>1.00</b>	0.42	<b>0.62</b>	<b>0.32</b>	<b>0.22</b>	0.13	-0.07
Na	<b>0.39</b>	<b>0.22</b>	<b>-0.27</b>	0.03	<b>-0.33</b>	<b>-0.21</b>	-0.11	<b>0.82</b>	<b>0.17</b>	<b>0.29</b>	<b>0.59</b>	<b>0.19</b>	<b>0.49</b>	-0.25	<b>-0.26</b>	<b>0.42</b>	<b>1.00</b>	<b>0.38</b>	<b>0.36</b>	0.07	0.04	-0.04
TDS	<b>0.75</b>	<b>0.31</b>	<b>-0.21</b>	<b>0.17</b>	<b>-0.24</b>	<b>-0.26</b>	-0.10	<b>0.39</b>	<b>0.25</b>	<b>0.42</b>	<b>0.29</b>	<b>0.23</b>	<b>0.63</b>	<b>-0.18</b>	<b>-0.25</b>	<b>0.62</b>	<b>0.38</b>	<b>1.00</b>	<b>0.39</b>	<b>0.25</b>	<b>0.19</b>	<b>-0.13</b>
SO4	<b>0.28</b>	<b>0.29</b>	-0.06	<b>0.46</b>	-0.07	-0.11	<b>-0.37</b>	<b>0.33</b>	<b>0.29</b>	<b>0.61</b>	0.12	<b>0.34</b>	<b>0.47</b>	-0.02	<b>-0.31</b>	<b>0.32</b>	<b>0.36</b>	<b>0.39</b>	<b>1.00</b>	<b>0.18</b>	-0.10	-0.08
U	<b>0.20</b>	-0.04	-0.01	<b>0.19</b>	0.02	-0.09	0.04	0.01	<b>-0.28</b>	<b>0.32</b>	<b>-0.16</b>	<b>-0.26</b>	<b>0.35</b>	0.00	0.05	<b>0.22</b>	0.07	<b>0.25</b>	<b>0.18</b>	<b>1.00</b>	<b>0.44</b>	-0.03
V	<b>0.21</b>	-0.09	-0.07	-0.05	-0.10	-0.04	<b>0.47</b>	0.10	<b>-0.20</b>	0.09	0.05	<b>-0.46</b>	0.05	-0.07	0.02	0.13	0.04	<b>0.19</b>	-0.10	<b>0.44</b>	<b>1.00</b>	0.01
Zn	-0.03	0.02	<b>0.15</b>	-0.05	<b>0.38</b>	<b>0.43</b>	0.05	-0.06	<b>0.39</b>	0.12	-0.01	<b>0.29</b>	-0.13	<b>0.47</b>	0.05	-0.07	-0.04	<b>-0.13</b>	-0.08	-0.03	0.01	<b>1.00</b>

Fonte: Autoria própria (Acervo pessoal Natália de Souza Pelinson).

É preciso analisar se, tecnicamente, a atual liberação de dados de monitoramento pelo InfoAguas e pelos Mapa Interativo da Lista de Áreas Contaminadas apresenta a possibilidade dos pontos destacados no Decreto 53.263/2012 ainda não estarem sendo atendidos, mesmo após mais de 10 anos de regulamentação da Lei 13.577/2009 (São Paulo, 2009). Questionamentos acerca da transparência à informação e sobre as melhores práticas acerca da comunicação de riscos podem reforçar que o interesse privado não deveria sobrepor o interesse coletivo. A importância em se divulgar as informações de monitoramento de água superficial e/ou subterrânea realizada por processos de GAC é reforçada pelo fato de que os compartimentos ambientais são interconectados e a água é um bem público em que o abastecimento humano (e dessedentação de animais) deve ter prioridade frente a outros usos múltiplos da água (Brasil, 1997).

Apesar de ser um assunto complexo, uma vez que garantir a disseminação transparente de informações sobre a qualidade das águas subterrâneas e, ao mesmo tempo, proteger dados sensíveis requer uma abordagem cuidadosa e estratégica. Porém, deveríamos considerar a divulgação de informações de forma responsável como uma prioridade, de forma a integrar as informações tornando-as públicas, realmente priorizando a proteção da água como preconizado pela Lei de Águas (Brasil, 1997).

Complementarmente, ainda que não haja ainda um banco de dados nacional ainda alimentado pelos estados brasileiros, a proteção da qualidade do solo e da água subterrânea também é definida pela Resolução 420/2009 (Brasil, 2009) e outras legislações vigentes. Vale destacar que não há ainda uma plataforma que dados da qualidade do solo sejam liberados, nem à população, nem a profissionais, pesquisadores ou outros interessados, dificultando o acesso à informação pública, ainda que haja interesse coletivo em diversas situações.

Como sugestão para viabilizar a liberação de dados de empresas juntamente a monitoramentos públicos da qualidade dos meios físicos, poderíamos destacar as seguintes recomendações:

- No momento de disponibilização de dados, anonimizar os dados, removendo qualquer informação pessoalmente identificável do conjunto de dados, como nomes, endereços ou quaisquer outros detalhes que possam vincular diretamente os poços de monitoramento a indivíduos ou entidades.
- Podemos adotar Zonas Geográficas de interesse para liberação de dados, apresentando os dados com base em zonas em vez de localizações exatas. Podemos, por exemplo, agrupar dados de poços de monitoramento em regiões ou áreas específicas para fornecer uma visão geral da qualidade das águas subterrâneas, sem a apresentação de

dados em pontos. Assim, ao invés de uma rede de poços, poderíamos ter uma malha pré-definida de coordenadas a serem “abastecidas nas plataformas”.

- Quanto às plataformas, podem ser implementados controles de acesso com usuário cadastrado, adotando uma plataforma online segura com controles de acesso para restringir o acesso às informações (similar ao InfoAguas). Isto poderia garantir que apenas indivíduos autorizados, como autoridades ambientais ou investigadores, possam visualizar dados detalhados completos (pontuais, enquanto outros apenas malhas simplificadas). Assim, as informações voltadas ao público podem ser mais generalizadas, apenas mudando o papel do usuário no sistema.
- Poderiam ser atribuídos identificadores exclusivos, substituindo informações identificáveis por identificadores para cada poço de monitoramento. Isso garante que as informações ainda possam ser rastreadas em seu sistema, mas não comprometem a privacidade.
- Ainda que os dados agregados ou resumidos, em vez de fornecer dados brutos para cada poço, ofereçam médias, tendências ou comparações ao longo do tempo, protegem detalhes específicos ao mesmo tempo que transmitem informações importantes sobre a qualidade das águas subterrâneas.
- Priorizar a Conscientização Pública, envolvendo-se com as partes interessadas relevantes, incluindo organizações ambientais, representantes da comunidade local e meios de comunicação. Incentivar discussões abertas sobre a importância da qualidade das águas subterrâneas e os riscos potenciais associados à contaminação.
- Defender a transparência na gestão ambiental, enfatizando a importância de fornecer informações claras e comprehensíveis ao público. A transparência gera confiança e incentiva a responsabilização, promovendo um comportamento responsável entre empresas e órgãos reguladores. Podemos utilizar as regulamentações ambientais existentes para promover a transparência.
- No Brasil, onde a água é priorizada para o abastecimento humano (Brasil, 1997), as regulamentações podem apoiar o direito do público de acessar informações relacionadas à qualidade da água. Manter uma comunicação com as autoridades reguladoras dos comitês de bacias hidrográficas poderia propiciar que as políticas públicas estejam alinhadas com o interesse público.

Sendo assim, é válido observarmos que é de suma relevância que haja engajamento diretamente com a comunidade local, organizando reuniões públicas, workshops ou sessões

informativas para discutir a qualidade das águas subterrâneas, os riscos potenciais e as medidas que estão a ser tomadas para lidar com possíveis focos de contaminação. As comunidades informadas podem participar ativamente na defesa dos seus direitos e da proteção ambiental, para que não apenas novas normas e leis sejam criadas ou adequadas, mas para que haja cobrança no cumprimento da legislação vigente.

## 6 CONSIDERAÇÕES FINAIS

A integração de análises estatísticas quantitativas representa uma oportunidade para aprimorar a segurança hídrica, fornecendo subsídios para estratégias específicas visando o controle e prevenção de contaminações em postos de combustíveis e áreas circunvizinhas. Essa abordagem não apenas propicia os usos múltiplos da água subterrânea, mas também consolida a gestão responsável de áreas potencialmente contaminadas com base em dados específicos de qualidade do meio físico, com perspectiva de replicabilidade em diversas regiões do Brasil.

## 7 REFERÊNCIAS BIBLIOGRÁFICAS

ABAS - ÁGUAS SUBTERRÂNEAS. Aquíferos. Disponível em:  
<https://www.abas.org/aguas-subterraneas-o-que-sao/>. Acesso em: 12 fev.24

ABDELAZIZ, S.; GAD, M. I.; El Tahan, A. H. M. H. Groundwater quality index based on PCA: Wadi El-Natrun, Egypt, **Journal of African Earth Sciences**, Volume 172, 2020, 103964, ISSN 1464-343X. Disponível em: <https://doi.org/10.1016/j.jafrearsci.2020.103964>. Acesso em: 23 fev. 24.

ABNT – Associação Brasileira de Normas Técnicas. **NBR 15515-1:2007 - Passivo Ambiental**, 2007. Disponível em : [https://saturno.crea-rs.org.br/pop/profissional/ABNT\\_NBR\\_15515\\_1\\_2021.pdf](https://saturno.crea-rs.org.br/pop/profissional/ABNT_NBR_15515_1_2021.pdf). Acesso em: 13fev.24

BALDERACCHI, M.; BENOIT, P.; CAMBIER, P.; EKLO, O. M.; GARGINI, A., GEMITZI, A.; GUREL, M.; KLOVE, B.; NAKIC, Z.; PREDA, E.; RUZICIC, S. WACHNIEW, P.; TREVISAN, M. (2013). Groundwater pollution and quality monitoring approaches at the European level. **Critical reviews in environmental science and technology**, 43(4), 323-408. Disponível em: <https://doi.org/10.1080/10643389.2011.604259>. Acesso em: 23 fev. 24.

BERTOLO, R. A.; HIRATA, R.; CONICELLI, B., SIMONATO, M., PINHATTI, A., & FERNANDES, A. (2015). Água subterrânea para abastecimento público na Região Metropolitana de São Paulo: é possível utilizá-la em larga escala?. **Revista DAE**, 63(199), 6-17. Disponível em: <http://dx.doi.org/10.4322/dae.2014.148>. Acesso em: 23 fev. 24.

CETESB – Companhia Ambiental do Estado de São Paulo. **Avaliação Potencial de Contaminação em Imóveis**. São Paulo, 2003. Disponível em:  
[https://cetesb.sp.gov.br/wp-content/uploads/sites/17/2015/06/guia\\_aval\\_pot\\_con\\_imoveis.pdf](https://cetesb.sp.gov.br/wp-content/uploads/sites/17/2015/06/guia_aval_pot_con_imoveis.pdf). Acesso em: 15fev.24

CETESB - Companhia Ambiental do Estado de São Paulo. **Decisão de Diretoria Nº 038/2017/C**, de 07 de fevereiro de 2017. Dispõe sobre a aprovação do “Procedimento para a Proteção da Qualidade do Solo e das Águas Subterrâneas”, da revisão do “Procedimento para o Gerenciamento de Áreas Contaminadas” e estabelece “Diretrizes para Gerenciamento de Áreas Contaminadas no Âmbito do Licenciamento Ambiental. Disponível em:  
<https://cetesb.sp.gov.br/wp-content/uploads/2014/12/DD-038-2017-C.pdf>. Acesso em: 22 fev. 24.

CONAMA – CONSELHO NACIONAL DE MEIO AMBIENTE. **Resolução nº 420, de 28 de dezembro de 2009**. Valores Orientadores de qualidade de solo e água subterrânea do Brasil. 2009. Disponível em:  
<https://cetesb.sp.gov.br/areas-contaminadas/wp-content/uploads/sites/17/2017/09/resolucao-conama-420-2009-gerenciamento-de-acr.pdf>. Acesso em: 10 jan.24

CONICELLI, B.; HIRATA, R.; GALVÃO, P.; BERNARDINO, M.; SIMONATO, M., ABREU, M. C.; ARANDA, N.; TERADA, R. (2021). Determining groundwater availability

and aquifer recharge using GIS in a highly urbanized watershed. **Journal of South American Earth Sciences**, Volume 106, 2021, 103093, ISSN 0895-9811. Disponível em: <https://doi.org/10.1016/j.jsames.2020.103093>. Acesso em: 23 fev. 24.

COSTA, D. A.; AZEVEDO, J. P. S.; SANTOS, M. A.; ASSUMPÇÃO, R. S. F. V. Water quality assessment based on multivariate statistics and water quality index of a strategic river in the Brazilian Atlantic Forest. **Sci Rep** 10, 22038 (2020). Disponível em: <https://doi.org/10.1038/s41598-020-78563-0>. Acesso em 23/02/2024.

DENG, Y.; YE, X.; Du, X. Predictive modeling and analysis of key drivers of groundwater nitrate pollution based on machine learning. **Journal of Hydrology**, Volume 624, 2023, 129934, ISSN 0022-1694. Disponível em: <https://doi.org/10.1016/j.jhydrol.2023.129934>. Acesso em 02 Mar. 24.

EMBRAPA - EMPRESA BRASILEIRA DE CULTURA E AGROPECUÁRIA. Agricultura e Qualidade da Água: Contaminação da Água por Nitrato. Disponível em: [https://ainfo.cnptia.embrapa.br/digital/bitstream/CPAC-2009/24718/1/doc\\_57.pdf](https://ainfo.cnptia.embrapa.br/digital/bitstream/CPAC-2009/24718/1/doc_57.pdf). Acesso em: 05 fev.24

FARO, G. T. C.; GARCIA, J. I. B.; OLIVEIRA, C. DE P. M.; RAMOS, M. R. S. (2019). Application of indices for water resource systems stress assessment. **RBRH**, 24, e7. Disponível em: <https://doi.org/10.1590/2318-0331.241920180106>. Acesso em: 23 fev. 24.

FÁVERO, LUIZ PAULO LOPES; BELFIORE, P.; SOUZA, R. F. **Data Science, Analytics and Machine Learning with R**. 1. ed. Cambridge: Academic Press, 2023. 660 p. <https://doi.org/10.1016/C2020-0-01649-X>.

FORMIGA-JOHNSSON, R. M.; KEMPER, K. (2005). **Institutional and Policy analysis of river basin management**: the Alto-Tietê River Basin, São Paulo, Brazil. São Paulo, Brazil (June 2005). Disponível em: <https://documents1.worldbank.org/curated/en/403111468229455654/pdf/wps3650.pdf> Acesso em: 23 fev. 24

FRANCO, A. O; ARCOS, F. O.; PEREIRA, J. S. Uso do solo e a qualidade da água subterrânea: estudo de caso do aquífero Rio Branco, Acre, Brasil. **Águas Subterrâneas**, v. 32, n. 3 (2018) - Seção Estudos de Caso e Notas Técnicas, 2018. Disponível em: <https://aguassubterraneas.abas.org/asubterraneas/article/view/29178> . Acesso em 23 fev.2024

GALLATIN, K.; ALBON, C. **Machine Learning with Python Cookbook**. 2. Ed. O'Reilly Media, Inc: Sebastopol, CA (EUA). ISBN: 978-1-098-13572-0, 2023.

GARBA, A.; IDRIS, A.M.; GAMBO, J. (2023). Groundwater Quality Assessment Using Principal Component and Cluster Analysis. In: Sherif, M., Singh, V.P., Sefelnasr, A., Abrar, M. (eds) **Water Resources Management and Sustainability**. Water Science and Technology Library, V. 121. Springer, Cham. Disponível em: [https://doi.org/10.1007/978-3-031-24506-0\\_22](https://doi.org/10.1007/978-3-031-24506-0_22). Acesso em: 02 Mar. 2024.

GOMES, K. J. M.; OLIVA, P. A. C., ROCHA, H. O.; ALCANTARA MENDES, R.; COSTA, A. C. G.; SANTOS MIRANDA, C.; ALMEIDA, N. O. (2023). Evaluation of the contamination of the subsurface and groundwater by monoaromatic hydrocarbons in an

eastern Amazonian town in northern Brazil. **Environmental Earth Sciences**, 82(1), 23. Disponível em: <https://doi.org/10.1007/s12665-022-10680-7>. Acesso em 23 fev. 24.

HAGGERTY, R.; SUN, J.; YU, H.; LI, Y. Application of machine learning in groundwater quality modeling - A comprehensive review, **Water Research**, Volume 233, 2023, 119745, ISSN 0043-1354. Disponível em: <https://doi.org/10.1016/j.watres.2023.119745>. Acesso em: 23 fev. 24.

IGHALO, J. O.; ADENIYI, A. G. (2020). A comprehensive review of water quality monitoring and assessment in Nigeria. **Chemosphere**, 260, 127569. Disponível em: <https://doi.org/10.1016/j.chemosphere.2020.127569>. Acesso em 23 fev. 24.

JAYARAMAN, P.; NAGARAJAN, K. K.; PARTHEEBAN, P.; KRISHNAMURTHY, V. Critical review on water quality analysis using IoT and machine learning models, **International Journal of Information Management Data Insights**, Volume 4, Issue 1, 2024, 100210, ISSN 2667-0968. Disponível em: <https://doi.org/10.1016/j.jjimei.2023.100210>. Acesso em: 23 fev. 24.

MCHAFFIE, P.; HWANG, S.; FOLLET, C. **GIS - An Introduction to Mapping Technologies**. CRC Press (Taylor & Francis Group, LLC): Boca Raton, FL (EUA), 2023. ISBN: 978-1-003-30718-1 . Disponível em: <https://doi.org/10.1201/9781003307181>. Acesso em: 02 Mar. 24.

MIYAZAKI, L.; VENTURI, L. A. B. “Water Resources in the São Paulo Metropolitan Area (RMSP): The Underuse of Billings Reservoir”. **International Journal of Water Management and Diplomacy**, vol. 1, no. 5, 2022, pp. 21-54. Disponível em: <https://dergipark.org.tr/en/download/article-file/2473743>. Acesso em: 23 fev. 24.

PATNAIK, M.; TUDU, C.; BAGAL, D.K. Monitoring groundwater quality using principal component analysis. **Appl Geomat** (2024). Disponível em: <https://doi.org/10.1007/s12518-024-00552-z>. Acesso em: 02 Mar. 24.

ROCHA, G. **Mapa de águas subterrâneas do Estado de São Paulo**: escala 1:1.000.000: Nota explicativa. São Paulo - SP: DAEE - Departamento de Águas e Energia Elétrica; IG-Instituto Geológico; IPT - Instituto de Pesquisas Tecnológicas do Estado de São Paulo; CPRM - Serviço Geológico do Brasil, 2005. Disponível em: [www.infraestruturaeambiente.sp.gov.br/wp-content/uploads/sites/233/2012/03/Nota%20Explanativa%20Mapa%20Aguas%20Subterraneas.pdf](http://www.infraestruturaeambiente.sp.gov.br/wp-content/uploads/sites/233/2012/03/Nota%20Explanativa%20Mapa%20Aguas%20Subterraneas.pdf). Acesso em: 02 Mar. 24.

ROSECRANS, C. Z.; LANDON, M. K.; MCMAHON, P. B.; GILLESPIE, J. M.; KULONGOSKI, J. T., STEPHENS, M. J.; HUNT, A. G.; SHIMABURUO, D. H.; DAVIS, T. A. (2021). Groundwater quality of aquifers overlying the oxnard oil field, Ventura County, California. **Science of The Total Environment**, 771, 144822. Disponível em: <https://doi.org/10.1016/j.scitotenv.2020.144822>. Acesso em: 23 fev. 24.

SANCHES, L. E. A. Desativação de empreendimentos industriais: um estudo sobre o passivo ambiental. Tese (Livre Docência) – Escola Politécnica, Universidade de São Paulo, 1998. Disponível em: <https://repositorio.usp.br/item/000987290>. Acesso em: 23 fev. 24.

SARKER, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. **Sn Comput. Sci.** 2, 160 (2021). Disponível em: <https://doi.org/10.1007/s42979-021-00592-x>. Acesso em: 23 fev. 24.

SÃO PAULO (Estado). **Decreto nº 59.263**, de 5 de junho de 2013. Regulamenta a Lei nº 13.577, de 8 de julho de 2009, que dispõe sobre diretrizes e procedimentos para a proteção da qualidade do solo e gerenciamento de áreas contaminadas, e dá providências correlatas. Assembleia Legislativa do Estado de São Paulo. Disponível em: <https://www.al.sp.gov.br/repositorio/legislacao/decreto/2013/decreto-59263-05.06.2013.html>. Acesso em: 22 fev. 24.

OLIVEIRA, L.; LOUREIRO, C. Contaminação de aquíferos por combustíveis orgânicos em Belo Horizonte: Avaliação Preliminar (1998). Disponível em: <https://aguassubterraneas.abas.org/asubterraneas/article/view/22287/14630>. Acesso em: 26 fev. 24

TAHERDOOST, H. Deep Learning and Neural Networks: Decision-Making Implications. **Symmetry**. 2023; 15(9):1723. Disponível em: <https://doi.org/10.3390/sym15091723>. Acesso em: 23 fev. 24.

VARNIER, C.; HIRATA, R. Contaminação por nitrato no parque ecológico do Tietê - São Paulo, Brasil (2002). **Águas subterrâneas**, 16 (1). Disponível em: <https://aguassubterraneas.abas.org/asubterraneas/article/view/1303/1081>. Acesso em: 20 fev. 24.