

## **Efeito das características físicas e da cobertura da terra na vazão e turbidez de nascentes sob diferentes abordagens estatísticas**

Vinicius Barros Rodrigues

Universidade Federal de Viçosa - UFV

[viniciusbrbio@gmail.com](mailto:viniciusbrbio@gmail.com)

Thales Braga Capetine

Universidade Federal de Viçosa - UFV

[thales.b.capetine@hotmail.com](mailto:thales.b.capetine@hotmail.com)

Fillipe Tamiozzo Pereira Torres

Universidade Federal de Viçosa - UFV

[tamiozzo@ufv.br](mailto:tamiozzo@ufv.br)

### **Resumo**

A falta de planejamento adequado e o uso desordenado do solo afeta significativamente os recursos naturais, especialmente os hídricos. O aumento do volume de dados para enfrentar desafios complexos, como a proteção e aumento da disponibilidade de recursos hídricos, introduz potenciais problemas como sobreajuste e multicolinearidade durante as análises. Assim, o objetivo deste estudo foi utilizar diferentes métodos estatísticos para mitigar essas questões, analisando os efeitos e as influências dos fatores ambientais e das características físicas do solo na vazão e na turbidez de 82 nascentes. Foram utilizadas diferentes abordagens: *Random Forest* e Critério de Informação de Akaike, e ambos se mostraram consistentes em relação às variáveis analisadas. A cobertura arbórea apresentou maior correlação com o aumento da vazão, enquanto que a cobertura arbustiva favoreceu o aumento da turbidez. A cobertura arbustiva pode favorecer o escoamento superficial, o que também aumenta a turbidez do fluxo. A cobertura vegetal (arbórea e arbustiva) pode ter favorecido os processos de infiltração da água, aumentando a vazão. Os resultados fornecidos possuem o potencial de desempenhar um importante papel ao orientar as melhores práticas na análise de dados, bem como nas estratégias de recuperação e proteção de nascentes.

**Palavras-chave:** Random Forest, Machine Learning, Recuperação de Áreas Degradadas, AIC

### **Abstract**

The lack of proper planning and disorderly land use significantly impacts natural resources, especially water. The increase in data volume to address complex challenges, such as protecting and enhancing water resource availability, introduces potential issues like overfitting and multicollinearity during analyses. Therefore, this study aimed to employ different statistical methods to mitigate these issues, examining the effects and influences of environmental factors and soil physical characteristics on the flow and turbidity of 82 springs. Various approaches were utilized, including Random Forest and Akaike Information Criterion, both proving consistent concerning the analyzed variables. Arboreal coverage showed a stronger correlation with increased flow, while shrub coverage favored higher turbidity. Shrub coverage can facilitate surface runoff, contributing to increased flow turbidity. The vegetation coverage (arboreal and shrub) may have favored water infiltration processes, enhancing flow. The provided results have the potential to play a crucial role in guiding best practices in data analysis and strategies for the recovery and protection of springs.

**Keywords:** Random Forest, Machine Learning, Recovery of Degraded Areas, AIC

**Introdução**

Os ecossistemas naturais estão sob grave ameaça em todo o mundo, principalmente devido às alterações drásticas do uso da terra e às mudanças climáticas, causando um declínio na biodiversidade e nos serviços ecossistêmicos. Estudos em diversas regiões apontam para a necessidade urgente de conservação dos ecossistemas naturais remanescentes, uso sustentável dos recursos ambientais e restauração de ecossistemas degradados para evitar danos irreversíveis à própria existência humana (Mello, 2022). Como parte importante desse aumento na exploração dos recursos naturais, responder a essa escala de transformação e alcançar o desenvolvimento sustentável é atualmente um dos princípios da conservação ecológica (Bos; Gupta, 2019; Hong et al., 2019). Assim, embora exista um consenso geral sobre a necessidade de um planejamento adequado, ainda há um intenso debate sobre a maneira mais eficiente de desenvolver as ações voltadas para a conservação e restauração desses recursos naturais, bem como para a manutenção de seus serviços ecossistêmicos.

A exploração das terras e o crescimento econômico de grande parte das regiões brasileiras têm sido constantemente prejudicados pela falta de um planejamento adequado. A ocupação e o uso inadequado do solo têm um impacto significativo nos recursos naturais, especialmente nos recursos hídricos (Oliveira et al., 2017). Neste contexto, um dos principais desafios para administradores públicos, pesquisadores e a população é assegurar a quantidade e a qualidade da água disponível para uso e consumo (Costa et al., 2015).

Neste contexto, muitas ações e metodologias têm sido desenvolvidas e testadas para a proteção e aumento da disponibilidade dos recursos hídricos. Essas abordagens não são triviais, pois também devem considerar muitos fatores além dos ambientais, como custos de implantação, análises do solo e da água e da dependência socioeconômica do recurso (Van Camp; Walraevens, 2009). Para que esses problemas complexos sejam devidamente abordados e modelados, são necessários grandes volumes de dados, o que resulta em um aumento significativo no número de observações e variáveis coletadas e armazenadas. Embora essa quantidade de dados possibilite novas oportunidades de análises, também causam sérios problemas durante a análise de dados, como o sobreajuste (*overfitting*) e a multicolinearidade (Chan et al., 2022). A presença de algum desses problemas pode apresentar resultados enviesados. O sobreajuste pode ocorrer quando há a incorporação de um número excessivo de variáveis independentes no modelo. Já a multicolinearidade ocorre quando as variáveis

independentes são fortemente intercorrelacionadas e podem ser representadas por um número menor de variáveis (Xiong et al., 2019).

Atualmente, novas metodologias têm surgido e vêm sendo testadas buscando mitigar estes problemas, de forma a otimizar os processos envolvidos nas decisões, nas análises e nas interpretações dos dados. As ferramentas de Inteligência Artificial (IA), como o *Machine Learning* (ML), e seleção de modelos pelo Critério de Informação de Akaike (AIC) abriram novas perspectivas neste cenário, principalmente para monitoramento e avaliação ambiental. Ambas as ferramentas, ML e AIC, são alternativas promissoras para solucionar os problemas de sobreajuste e multicolinearidade (Xiong et al., 2019; Chan et al., 2022). Essas metodologias possuem baixas premissas estatísticas e são adequadas para modelar interações complexas e não-lineares (Yamashita et al., 2007; Viana et al., 2021).

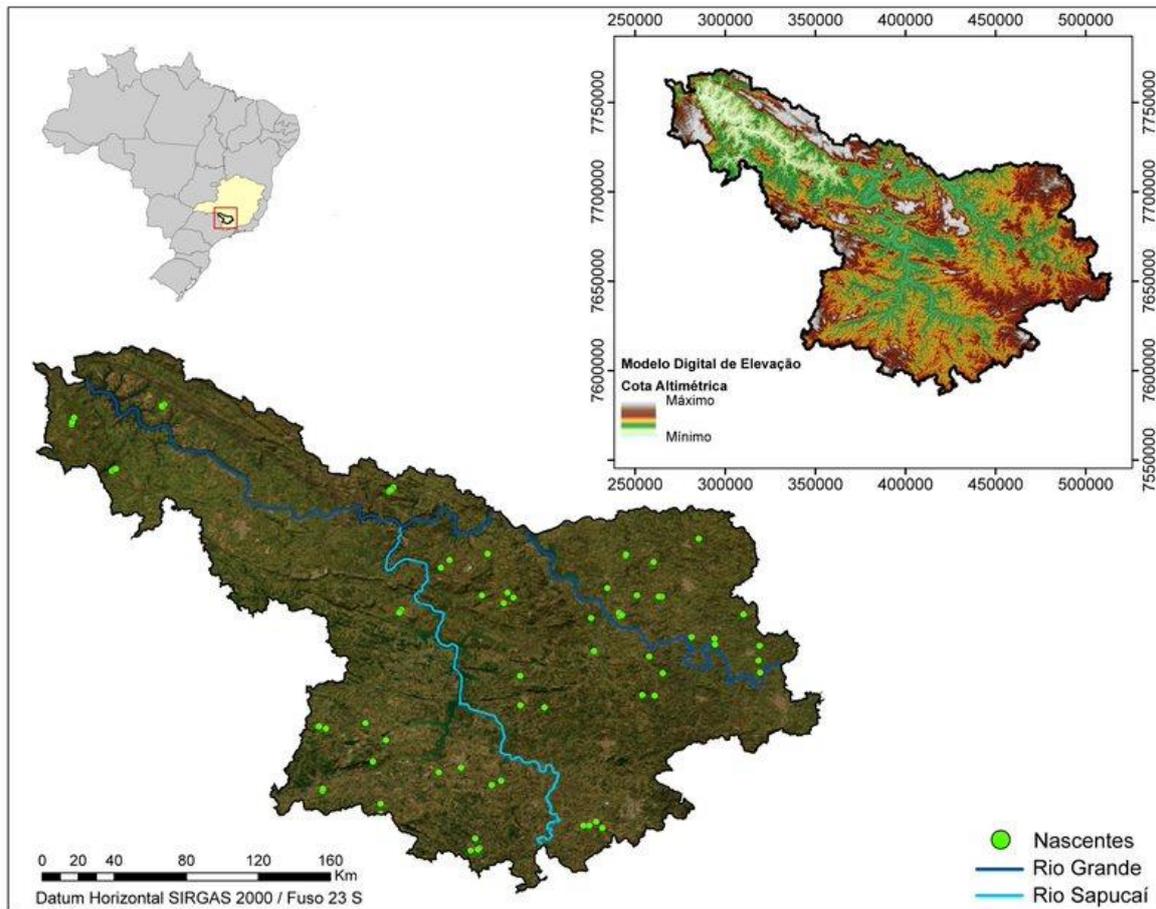
Desta forma, o objetivo deste estudo foi combinar métodos estatísticos com diferentes paradigmas, ML e AIC, para avaliar os efeitos e as relações dos fatores ambientais e das características físicas do solo na vazão e na turbidez de 82 nascentes. Além disso, buscamos entender melhor as relações entre as principais variáveis e sugerir quais ações podem apresentar maior eficiência na recuperação das áreas de nascentes.

## **Material e métodos**

### *Área de estudo*

O trabalho foi conduzido na sub-bacia hidrográfica do Rio Grande, abrangendo vinte e cinco municípios na porção sul do estado de Minas Gerais. Foram selecionadas 82 parcelas representativas das zonas de recarga de aquíferos contribuintes dos reservatórios das Usinas Hidroelétricas de Furnas e Peixoto. A seleção das parcelas (Figura 1) ocorreu de acordo com a disponibilidade e interesse dos proprietários em participar do projeto e foram distribuídas em 25 municípios: Aguanil (5 parcelas), Alfenas (2 parcelas), Areado (6 parcelas), Boa Esperança (2 parcelas), Cabo Verde (5 parcelas), Campo Belo (4 parcelas), Campos Gerais (3 parcelas), Cana Verde (4 parcelas), Candeias (2 parcelas), Capitólio (4 parcelas), Carmo do Rio Branco (2 parcelas), Cássia (6 parcelas), Conceição da Aparecida (2 parcelas), Coqueiral (4 parcelas), Cristais (3 parcelas), Delfinópolis (1 área), Divisa Nova (4 parcelas), Eloi Mendes (2 parcelas), Fama (1 parcela), Guapé (3 parcelas), Ibiraci (1 parcelas), Ilícinea (4 parcelas), Lavras (1 parcela), Machado (4 parcelas) e Monte Belo (7 parcelas).

**Figura 01.** Localização dos pontos de monitoramento na sub-bacia hidrográfica do Rio Grande.



Fonte: Os autores, 2023.

## Coleta de dados

### *Características físicas do solo*

Foram coletadas amostras de solo para análise granulométrica (porcentagem de argila, silte e areia) conforme recomendado por Teixeira et al. (2017). Em cada parcela foram selecionadas as áreas mais homogêneas possíveis. Uma vez selecionados os pontos de amostragens, as superfícies dos solos foram limpas, removendo eventuais folhas e gravetos e, com o auxílio de uma pá, amostras representativas do solo foram coletadas nos primeiros 10 cm do perfil, acondicionadas em saco plástico e enviadas para análises. As análises granulométricas foram realizadas no Departamento de Solos da Universidade Federal de Viçosa, Viçosa-MG, seguindo os protocolos da Embrapa (1997).

### *Diagnóstico da cobertura do solo*

Para análise do uso e ocupação do solo, o procedimento foi executado conforme a metodologia preconizada no trabalho de Gama et al. (2021). As fotografias aéreas foram realizadas nos anos de 2020 e 2021, utilizando drones DJI Phantom 4 Pro e DJI Mavic Pro com altura de voo entre 100 e 200 metros (<5 cm/pixel). A Classificação das imagens foi realizada pelo classificador de máxima verossimilhança (MaxVer) que por meio de parâmetros estatísticos considera a ponderação das distâncias entre as médias dos valores dos pixels das classes (De Brito et al., 2021). De acordo com Instituto Nacional de Pesquisas Espaciais (2008), esse método considera que as bandas têm distribuição normal para calcular a probabilidade de um dado pixel pertencer a uma classe específica, por tanto expressa confiabilidade na classificação. Por meio da ferramenta ImageClassification no Arcgis, realizou-se a representação das classes de ocupação do solo na área de estudo. O sistema de coordenadas adotado foi o SIRGAS 2000 UTM Zona 23S, que utiliza a projeção Universal Transversa de Mercator, considerada ideal para o cálculo de área em grande e média escala.

O procedimento seguinte foi a geração do arquivo de feições pontuais, que representa as nascentes da Sub-bacia Hidrográfica do Rio Grande. Para tanto foi utilizado o arquivo *shapefile* referente aos cursos d'água de cada área estudada. A camada foi extraída da plataforma Infraestrutura de Dados Espaciais (IDE-SISEMA).

Dessa forma, além do tamanho da área (hectare - ha), também foi calculado a cobertura arbórea (ha), a cobertura arbustiva (ha), cobertura herbácea (ha) e solo exposto (ha).

### *Vazão e turbidez*

A vazão foi mensurada utilizando o medidor portátil FlowTracker2, que opera com o método Doppler Acústico de Velocidade. Para a mensuração do nível de turbidez das nascentes, foi utilizado o turbidímetro portátil Akso TU-430, que possui faixa de medição de 0 a 1000 NTU e curva de calibração de turbidez linear de 4 escalas. As mensurações também foram realizadas no mês de julho de 2021 e não houve mudanças no regime de chuvas no período.

*Análises estatísticas*

Como objetivo de analisar os efeitos das variáveis explicativas ambientais nas variáveis respostas das nascentes: (i) vazão e (ii) turbidez, foram utilizadas duas metodologias para a seleção das variáveis do modelo: Critério de Informação de Akaike (AIC) e *Random Forest* (RF).

Inicialmente, como o recomendado (Lantz, 2015), todos os dados passaram por um pré-processamento, onde foram normalizados e os *outliers* removidos utilizando *boxplots*, com o pacote “*rstatix*” (Kassambara, 2021). Após isso, foram removidas do modelo as variáveis com altos valores de correlação com o pacote “*PerformanceAnalytics*”.

O AIC é um método que permite comparar e classificar os melhores modelos dentre os concorrentes (Symonds; Moussalli, 2011). Então, para a seleção do melhor modelo foi utilizado o pacote “*bootStepAIC*” (Rizopoulos; Rizopoulos, 2009), que utiliza o método de *stepwise bootstrap*, juntamente com o AIC, para a seleção das variáveis que melhor contribuem com o modelo.

Em seguida utilizou-se modelos de RF, que é um dos algoritmos mais utilizados em *Machine Learning* e foi desenvolvido por Breiman (2001). É um método que utiliza múltiplas árvores de decisão durante o treinamento e a predição. Para eliminar as variáveis redundantes e identificar aquelas diretamente relacionadas a cada variável resposta, foi utilizado o pacote “*VSURF*” (Genuer et al., 2015), como feito por Klock et al. (2022). As variáveis mais importantes do modelo final foram acessadas com as funções disponíveis no pacote “*randomForestSRC*”.

Ao final, foi feita uma análise de correlação de Pearson com as variáveis selecionadas pelos dois métodos utilizando o pacote “*GGally*”. Para os gráficos utilizou-se o pacote “*ggplot2*”. Todas as análises foram realizadas no R (R Core Team, 2021).

**Resultados e discussão**

Os valores descritivos das variáveis coletadas no trabalho podem ser observados na Tabela 1.

**Tabela 1.** Valores descritivos das variáveis coletadas, com seus valores mínimos, médios e máximos das parcelas experimentais analisadas

Variável	Valor mínimo	Valor médio	Valor máximo
Tamanho da área (ha)	0,06	1,30	6,99
Cobertura arbórea (ha)	0	0,38	6,76
Cobertura arbustiva (ha)	0	0,27	3,16
Cobertura herbácea (ha)	0	0,37	2,63
Solo exposto (ha)	0	0,01	0,41
Turbidez (NTU)	0	262,4	953,0
Vazão (m <sup>3</sup> /h)	0	0,002	0,023
Argila (%)	2,0	33,4	64,0
Silte (%)	7,0	20,9	50,0
Areia (%)	10,0	46,2	80,0

Fonte: Os autores, 2023.

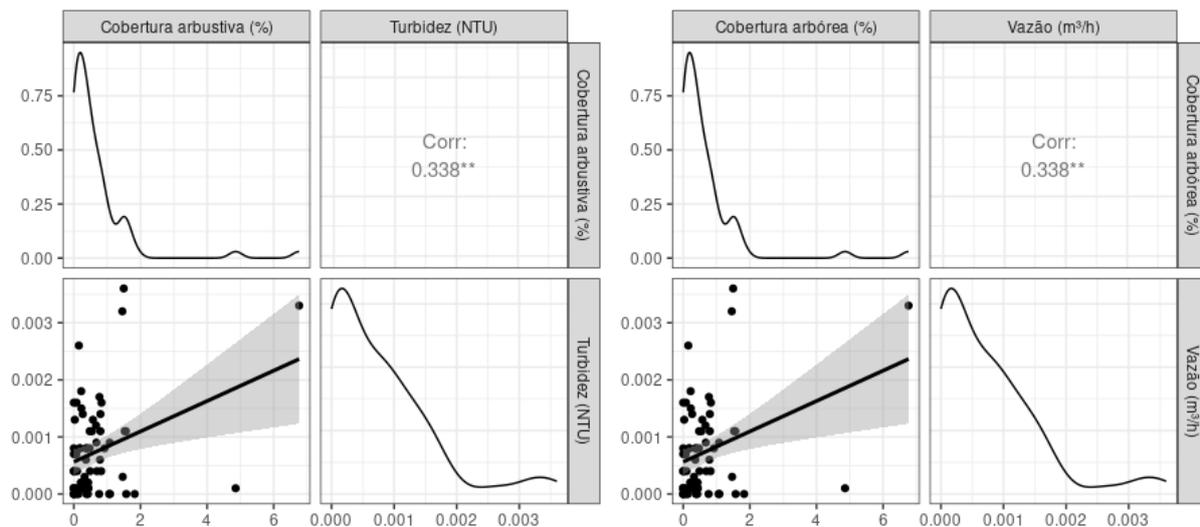
Em relação à variável resposta “turbidez”, o modelo final selecionado com o AIC continha apenas a variável explicativa “cobertura arbustiva”, com um peso de 78% no modelo. As demais variáveis apresentaram valores abaixo de 36%. De acordo com alguns autores, 60% é um valor ótimo de corte (Austin; TU, 2004). Quando o modelo foi analisado utilizando o algoritmo RF, as variáveis explicativas foram reduzidas para cinco variáveis, sendo a variável “cobertura arbórea” como a mais importante, com valor normalizado de 60,7%. As demais variáveis foram “cobertura arbustiva”, “argila”, “silte” e “areia”, com valores 15%, 16,43%, 0,01% e 7,86% respectivamente.

Por outro lado, o modelo final da variável resposta “vazão” continha apenas a variável explicativa “cobertura arbórea”, com peso de 83% no modelo AIC. O *Random Forest* reduziu para três variáveis explicativas, também com a variável “cobertura arbórea” como a mais importante no modelo, com valor normalizado de 92,83%, seguido das variáveis “argila” e “areia”, com valores respectivos de 7,15% e 0,02%.

Por fim, foram calculadas as correlações entre as principais variáveis selecionadas ao final de cada modelo: “cobertura arbórea” e “cobertura arbustiva”, com as variáveis “vazão” e “turbidez”. Os maiores valores de correlação foram entre “vazão” e “cobertura arbórea” foram

de 0,33 ( $p > 0,05$ ) e a entre a “turbidez” e a “cobertura arbustiva” foi de 0,21 ( $p > 0,005$ ). Os gráficos da relação entre as variáveis estão disponíveis na Fig. 02.

**Figura 02.** Relação entre as variáveis Turbidez (NTU) e a Cobertura Arbustiva (%); Vazão ( $m^3/h$ ) e Cobertura Arbórea (%). Cada ponto representa uma parcela experimental.



Fonte: Os autores, 2023.

O estudo utilizou a combinação de técnicas estatísticas e medidas de informação (IA e AIC) para avaliar os efeitos das características ambientais e físicas do solo na vazão e turbidez de nascentes, com o objetivo de fornecer novas percepções para futuras análises ambientais. Nós verificamos que a cobertura arbórea e arbustiva das parcelas estudadas apresenta impactos nas características de vazão e turbidez das nascentes.

Nossos resultados corroboram com outros trabalhos como Pinto et al. (2004), Fernandes et al. (2014) e Pessi et al. (2018), nos quais observaram que o grau de cobertura do solo possui impactos positivos na preservação dos corpos hídricos. Isso ocorre, provavelmente, devido à maior infiltração das águas das chuvas no solo nessas áreas.

De acordo com Felipe (2013), a perenidade dos cursos d'água superficiais é assegurada pela recarga das águas subterrâneas, que se dá pela infiltração da água no solo. Formas de facilitar e incrementar a água no solo inclui a infiltração subsuperficial ao longo das raízes e canais formados direta ou indiretamente pelo sistema subterrâneo da vegetação. Ademais, pelo aumento dos coeficientes de infiltração superficiais decorrentes do recobrimento vegetativo e da matéria orgânica decaída.

As nascentes perdem a capacidade quantitativa e qualitativa da água, quando é alterado por ações antrópicas em ambientes de contribuição natural de infiltração em seu entorno e na área de recarga do lençol freático, comprometendo seu reabastecimento e sua produção de água (Sophocleous, 2002). A dinâmica das nascentes e do lençol freático pode ser modificada conforme diferentes formas de uso e cobertura do solo, tais como: remoção das espécies vegetais com sistema radicular profundo substituindo por espécies com raízes superficiais, dificultando assim a liberação da água nas camadas do âmbito do sistema radicular para fornecimento das nascentes (Marmontel; Rodrigues, 2015)

Assim, os fluxos subterrâneos convergem para zonas de menor potencial hidrométrico, promovendo uma diferença de energia que resulta na exfiltração da água subterrânea para a superfície. Essa exfiltração é a principal responsável pela formação de nascentes que irão alimentar os diversos corpos d'água, o que explica a variável cobertura arbórea ser a mais correlacionada com a variável resposta vazão. Corroborando, estudo realizado por Gomes et al. (2021), na bacia hidrográfica do rio Pontal, aponta que o desmatamento ocorrido entre os anos de 1992 e 2020, contribuiu para redução de água no balanço hídrico, impactando na disponibilidade de água no solo.

Por sua vez, a cobertura arbustiva no local de estudo está relacionada à degradação do solo, visto se enquadrar em área original de Mata Atlântica. Desta forma, a cobertura arbustiva apresenta um efeito diferente na infiltração da água quando comparado com a cobertura arbórea, favorecendo o escoamento superficial, que por sua vez tende a carregar material em suspensão para os afloramentos, aumentando sua turbidez. Marmontiel e Rodrigues (2021) apontam que o desmatamento da mata ciliar é cada vez mais frequente, e isso tem contribuído para o aumento da turbidez, elevação da temperatura da água, diminuição da porosidade, e diversidade de habitats/microhabitats, além da eutrofização artificial. Resultados semelhantes também encontraram Rocha *et al.* (2016) mostram que as coberturas florestais atenuam o impacto das fontes de poluição por meio do efeito “filtro”, destacando a retenção de sedimentos finos, matéria orgânica e produtos tóxicos. Isso reduz o assoreamento dos corpos d'água, previne a erosão e estabiliza os solos marginais, atuando como uma barreira física eficaz.

A degradação das nascentes em zonas rurais está relacionada a fatores como o acesso de bovinos às localidades e utilização direta da água para uso doméstico (Diel et al.

2000; Resende et al. 2009; Faria et al. 2012; Pinto et al. 2012). Entretanto, nas áreas amostradas não foram constatadas tais práticas.

É importante ressaltar que outros fatores podem influenciar as características das nascentes, como a infiltração da água da chuva no solo, diminuição da evapotranspiração pelo sombreamento, dentre outros efeitos hidrológicos que favorecem a quantidade e qualidade da água superficial (Bacellar, 2005). Portanto, para entendermos de forma mais precisa os impactos e as dinâmicas, o acompanhamento e futuras análises serão necessárias.

Em relação aos métodos estatísticos utilizados, os modelos reduziram, em ambos os casos, o número total de variáveis estudadas para apenas uma variável explicativa para cada variável resposta. Assim posto, apesar dos valores relativamente baixos de correlação entre as variáveis, a resposta foi consistente entre os modelos, o que evidencia a relativa importância destas variáveis no contexto. O coeficiente de correlação nos mostra caminhos para futuras investigações e, como qualquer valor estatístico, possui pouca importância se não for apropriadamente interpretado (Taylor, 1990). Por fim, novas variáveis serão coletadas no futuro e inseridas nos modelos com o intuito de verificar a coerência e consistência dos resultados encontrados.

### **Considerações finais**

Os resultados desta pesquisa evidenciam a influência da cobertura arbórea na vazão e turbidez das nascentes, enquanto a cobertura arbustiva emerge como um fator favorável ao aumento da turbidez. Devido à grande quantidade de variáveis e os eventuais problemas de multicolinearidade e sobreajuste, nós utilizamos diferentes paradigmas estatísticos, como o *Random Forest* e o AIC, e ambos se mostraram consistentes em relação às variáveis analisadas. O diagnóstico fornecido tem o potencial de desempenhar um importante papel ao orientar as melhores práticas na análise de dados, bem como nas estratégias de recuperação e proteção de nascentes.

### **Agradecimentos**

O estudo em questão faz parte como produto do Projeto pertencente ao Programa de P&D regulado pela ANEEL, Nº 00394-2103/2021, com o título: “Utilização de Inteligência Artificial no Desenvolvimento de Metodologias Inovadoras de Recuperação e Proteção de Nascentes e Áreas Degradadas em Zonas de Recargas de Aquíferos Contribuintes dos

Reservatórios das UHE's Furnas e Peixoto". Esse Projeto de P&D foi desenvolvido por Furnas Centrais Elétricas S/A juntamente com a empresa Ingá Engenharia e Consultoria Ltda.

### Referências

AUSTIN, P. C.; TU, J. V. Bootstrap methods for developing predictive models. *The American Statistician*, v. 58, n. 2, p. 131–137, 2004.

BOS, K; GUPTA, J. Stranded assets and stranded resources: Implications for climate change mitigation and global sustainable development. *Energy Research & Social Science*, New York, v. 56, p. 101215, 2019.

CHAN, J.Y.-L.; LEOW, S.M.H.; BEA, K.T.; CHENG, W.K.; PHOONG, S.W.; HONG, Z.-W.; CHEN, Y.-L. Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review. *Mathematics*, n.10, v.8, p.1283, 2022.

COSTA, W. C.; PIGA, F. G.; MORAES, M. C. P.; DORICI, M.; SANGUINETTO, E. de C.; de LOLLO, J. A.; MOSCHINI, L. E.; LORANDI, R.; OLIVEIRA, L. J. Fragilidade ambiental e escassez hídrica em bacias hidrográficas: Manancial do Rio das Araras – Araras, SP. *Revista Brasileira de Recursos Hídricos*, Porto Alegre, v. 20, n. 4, p.946-958, 2015.

DE BRITO, A. P. M.; DE MELO, I. M. S; SILVA, R. M. Variabilidade espaço-temporal da estrutura da paisagem e fragmentação florestal na APA da Serra de Baturité no Ceará. *Ambiente: Gestão e Desenvolvimento*, Boa Vista, v. 1, n. 1, p. 96-104, 2021.

FERNANDES, J. de F.; DE SOUZA, A. L. T.; TANAKA, M. O. Can the structure of a riparian forest remnant influence stream water quality? A tropical case study. *Hydrobiologia*, New York, v. 724, n. 1, p. 175-185, 2014.

DIEL, P. S.; DA COSTA, P. F.; DE OLIVEIRA, P. S. R.; DA SILVA, N. L. S.; ROSSET, J. S. Capacidade de uso de solo das propriedades rurais na microbacia do córrego Guará no Município de Marechal Cândido Rondon/PR. *Scientia Agraria Paranaensis*, Marechal Cândido Rondon, v. 12, p. 400–410, 2000.

GAMA, L. H. O. M.; PAIVA, P. F. P. R.; SILVA, O. M. J; RUIVO, M. L. P. Modelagem ambiental e uso de inteligência artificial para prognóstico de desmatamento: o caso da Rebio do Gurupi-MA. *Research, Society and Development*, Vargem Grande Paulista, v. 10, n. 2, p. e13810211609, 2021.

GENUER, R.; POGGI, J.; TULEAU-MALOT, C.. VSURF: an R package for variable selection using random forests. *The R Journal*, Vienna, v. 7, n. 2, p. 19-33, 2015.

GOMES, V. P.; ARAÚJO, M.S. B.; GALVÍNCIO, J. D. Mudanças espaço-temporais no uso e cobertura da terra na bacia hidrográfica do rio Pontal a partir de dados referenciais do Google Earth Pro. *Revista Brasileira de Geografia Física*, Recife, v. 14, n. 7, p.4148-4160, 2021.

HONG, C.; ZHANG, Q.; ZHANG, Y.; DAVIS, S. J.; TONG, D.; ZHENG, Y. Impacts of climate change on future air quality and human health in China. *Proceedings of the National Academy of Sciences*, Washington, v. 116, n. 35, p. 17193-17200, 2019.

KASSAMBARA, A. rstatix: Pipe-friendly framework for basic statistical tests. 2021. *R package version 0.6.0*. Available in: < Available in: <https://cran.r-project.org/web/packages/rstatix/index.html>>. Acessado em: 17 Abril 2023.

KLOCK, A. M.; VOGT, K. A; VOGT, D. J.; GORDON, J. G.; SCULLION, J. J.; SUNTANA, A. S.; MAFUNE, K. K.; POLYAKOV, A. Y.; GMUR, S. J.; DE LA ROSA, C. G. See the forest not the trees! Ecosystem-based assessment of response, resilience, and scope for growth of global forests, *Ecological Indicators*, New York, v. 140, p. 108973, 2022.

LANTZ, B. *Machine Learning with R*, 2nd ed., Packt Publishing, 2015, 396 p.

MELLO, K. D.; BRITES, A; BORGES, C. M.; TAVARES, P. A.; METZGER, J. P.; RODRIGUES, R. R.; SANTOS, L. S.; JOLY, C.A.; SPAROVEK, G. Science and environmental policy establishment: the case of the Forest Act in the State of São Paulo, Brazil. *Biota Neotropica*, São Paulo, v. 22, e20211373, 2022.

OLIVEIRA, R. C. M; LIMA, P. V. P. S; SOUSA, R. P. Gestão ambiental e gestão dos recursos hídricos no contexto do uso e ocupação do solo nos municípios. *Gestão & Regionalidade*, São Caetano do Sul, v. 33, n. 97, p. 48-64, 2017.

MARMONTEL, C. V. F.; RODRIGUES, V. A. Avaliação qualitativa de nascentes com diferentes coberturas do solo e conservação da vegetação em seu entorno no Córrego Pimenta. *Scientia Agraria Paranaensis*, Curitiba, v. 14, n. 1, p.53-59, 2015.

PESSI, D. D.; EREIO, P. K. B; ALVES, G. B. M; MARTARELLO, A. P; LOVERDE-OLIVEIRA, S. M. Qualidade da cobertura vegetal em áreas de preservação permanente de nascentes. *Anuário do Instituto de Geociências*, Rio de Janeiro, v. 41, n. 3, p. 270-280, 2019.

PINTO, L. V. A.; BOTELHO, S. A.; DAVIDE, A. C.; FERREIRA, E. Estudo das nascentes da bacia hidrográfica do Ribeirão Santa Cruz, Lavras, MG. *Scientia Florestalis*, Piracicaba, n. 65, p. 197-206, 2004.

PINTO, L. V. A.; ROMA, T. N.; BALIEIRO, K. R. C. Avaliação qualitativa da água de nascentes com diferentes usos do solo em seu entorno. *Cerne*, Lavras, v. 18, p. 495-505, 2012.

R CORE TEAM. *R: A language and environment for statistical computing*. *R Foundation for Statistical Computing*. Disponível em: <<https://www.r-project.org>> Acesso em: 28/11/2022.

RESENDE, H. C.; MENDES, D. R.; MENDES, J. E. D. G; BERNARDES, W. A. Diagnóstico e ações de conservação e recuperação para as nascentes do córrego-feio, Patrocínio, MG. *Bioscience Journal*, Uberlândia, v. 25, n.5, p. 112-119, 2009.

RIZOPOULOS, D.; RIZOPOULOS, M. D. (2009). Package ‘bootStepAIC’. *R Package Version 1.3-0*, <<https://cran.r-project.org/web/packages/bootStepAIC/index.html>> Acesso em: 28/11/2022.

ROCHA, C. H. BARRA.; PEREIRA, B. H. C.; SILVA, A. F. R. da; OLIVEIRA, M. de; CASQUIN, A. P.; FIGUEIREDO, M. R. de. Impactos do uso do solo nos recursos hídricos da bacia de contribuição do Lago dos Manacás, Minas Gerais, Brasil. *Revista Ambiente e Água*, Taubaté, v. 11, n. 4, p.929-942, 2016.

SOPHOCLEOUS, M. Interactions between groundwater and surface water: the state of the science. *Hydrogeology Journal*, n. 10, p. 52–67, 2002.

SYMONDS, M. R.; MOUSSALLI, A. A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology*, New York, v. 65, n. 1, p. 13-21, 2011.

TAYLOR, R. Interpretation of the correlation coefficient: a basic review. *Journal of Diagnostic Medical Sonography*, Thousand Oaks, v. 6, n. 1, p. 35-39, 1990.

TEIXEIRA, P. C.; DONAGEMMA, G. K.; FONTANA, A. I.; TEIXEIRA, W. G. *Manual de métodos de análise de solo*. Brasília: EMBRAPA Solos, 3 ed. 2017, 574 p.

VAN CAMP, M.; WALRAEVENS, K. Recovery scenarios for deep over-exploited aquifers with limited recharge: methodology and application to an aquifer in Belgium. *Environmental Geology*, v. 56, p. 1505-1516, 2009.

VIANA, D. S.; KEIL, P.; JELIAZKOV, A. Disentangling spatial and environmental effects: flexible methods for community ecology and macroecology. *Ecosphere*, Washington, v. 13, n. 4, p. e4028, 2022.

XIONG, B.; NEWTON, S.; LI, V.; SKITMORE, M.; XIA, B Hybrid approach to reducing estimating overfitting and collinearity. *Engineering, Construction and Architectural Management*, n. 26, v. 10, p. 2170-2185, 2019.

YAMASHITA, T.; YAMASHITA, K.; KAMIMURA, R. A stepwise AIC method for variable selection in linear regression. *Communications in Statistics - Theory and Methods*, London, v. 36, n. 13, p. 2395-2403, 2007.