

Gamaespectrometria para modelagem e mapeamento de argila do solo com uso de aprendizado de máquina

Gammaspectrometry for soil clay modeling and mapping using machine learning

MOQUEDACE, Cássio Marques¹; BALDI, Clara Glória Oliveira¹; SIQUEIRA, Rafael Gomes¹, PEREIRA, Luís Flávio¹; ROSOLEM, Gabriel Phelipe Nascimento²; FERNANDES-FILHO, Elpídio Inácio¹

¹Laboratório de Geoprocessamento e Pedometria (LabGeo) da Universidade Federal de Viçosa - UFV, cassiomoquedace@gmail.com, clara.gloria.oliveira@gmail.com, rafael.geo.siqueira@gmail.com, luis.flavio@ufv.br, elpidio@ufv.br; ²Universidade Federal de Santa Catarina, gabriel.rosolem@posgrad.ufsc.br

RESUMO EXPANDIDO TÉCNICO CIENTÍFICO

Eixo Temático: Crise ecológica e mudança climática: resistências e impactos na agricultura, nas águas e nos bens

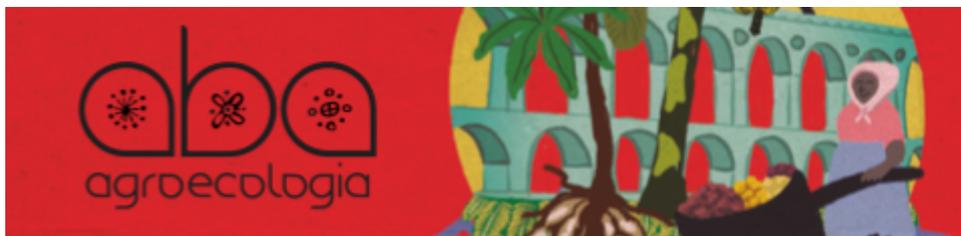
Resumo: Os atributos do solo variam espacial e temporalmente em diferentes escalas. Utilizar a gamaespectrometria como proxies ambientais podem melhorar a capacidade dos modelos preditivos e auxiliar na compreensão dos processos naturais. Objetivou-se modelar os teores de argila dos solos de Minas Gerais, Brasil e avaliar os principais preditores que controlam a sua distribuição. A área de estudo foi o estado de Minas Gerais, Brasil. Utilizou-se o banco de dados com 685 amostras para o ajuste do modelo. O ajuste do quantile random forest – qrf se deu na separação dos dados em treinamento e teste. Todo o processamento dos dados foi realizado no software R. As métricas de performance denotaram um bom ajuste para o qrf. O equivalente de tório (eTh) se destacou em importância para o modelo, sugerindo forte relação do eTh com a argila do solo. As maiores incertezas foram associadas às áreas com pouca argila. Os mapas de argila podem auxiliar na compreensão da dinâmica dos ecossistemas.

Palavras-chave: textura do solo; sensoriamento remoto; tório; raios gama; quantile random forest.

Introdução

Os atributos do solo variam espacialmente e temporalmente nas mais diferentes escalas. Estas variações têm implicações não somente nos ecossistemas manejados, mas também nos ecossistemas naturais. No contexto do mapeamento digital de solos (MDS), um dos campos do conhecimento que mais crescem na ciência do solo, melhorar a compreensão destas variações e relações tem sido objeto de estudo, especialmente mapeando de forma contínua atributos do solo (GOMES et al., 2019; SIQUEIRA et al., 2023).

Métodos computacionais robustos como o aprendizado de máquina permitem modelar e mapear com precisão conhecida estes atributos em escala de propriedades rurais até escala global. Estes modelos estatísticos detectam relações entre amostras medidas em campo com preditores ambientais. O gargalo no uso do aprendizado de máquina no MDS está em encontrar preditores ambientais que se relacionem de forma direta com o atributo, uma vez que a maioria dos sensores



remotos capturam a reflectância da faixa superficial do solo (se o solo estiver exposto) ou da vegetação.

Nesse contexto, a espectroscopia de raios gama se apresenta como ferramenta promissora. Com esta técnica é possível detectar emissões radioativas naturais de raios gama no decaimento de potássio (K40), urânio (eU) e tório (eTh) do solo. Os sensores que captam raios gama penetram até 40 cm em profundidade e podem ser utilizados remotamente com o uso de aeronaves ou de forma proximal em campo.

Utilizar estes dados como proxies explicativos de atributos do solo pode melhorar a capacidade dos modelos em produzir dados e informações espacialmente explícitas com maior precisão. No contexto da agroecologia, tais mapas podem auxiliar nas previsões das dinâmicas dos processos de ecossistemas naturais e manejados subsidiando as tomadas de decisões para os formuladores de políticas públicas tangentes ao uso sustentável dos bens naturais e mitigação das mudanças climáticas. Objetivou-se modelar os teores de argila dos solos de Minas Gerais, Brasil e avaliar os principais preditores que controlam a sua distribuição.

Metodologia

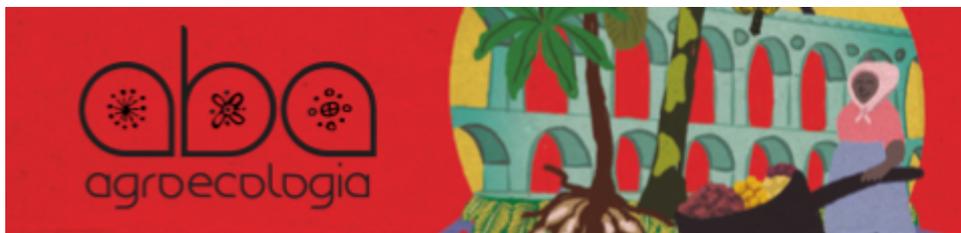
A área de estudo foi o estado de Minas Gerais, região Sudeste do Brasil. O estado contém 588.358,8 km² e o clima varia entre As, Aw, Cfa, Cfb, Cwa e Cwb (ALVARES et al., 2013). Há dominância dos biomas Cerrado e Mata Atlântica com manchas de Caatinga no norte do estado. Os solos são caracterizados por intenso intemperismo e grandes profundidades, com exceção das regiões serranas onde a erosão é mais pronunciada.

Como base de dados para as concentrações de argila do solo utilizou-se o banco de dados de Guevara et al. (2018) e Souza et al. (2015) que contém 685 amostras de solo distribuídas pelo estado 20 cm (Figura 1). A determinação dos teores de argila foi realizada pelo método da pipetagem associada aos princípios da Lei de Stokes. Como preditores ambientais utilizou-se informações derivadas dos fatores de formação do solo como clima, organismos, relevo e solo (Tabela 1). Os preditores ambientais foram harmonizados na resolução espacial de 30m.

No processo de modelagem utilizou-se o algoritmo baseado em árvores de decisão quantile random forest (qrf). O ajuste do modelo se deu na separação dos dados em 80% para treinamento e validação cruzada com 10 folds e 20% para teste (avaliação de performance). Sob essa estrutura o modelo foi ajustado 100 vezes, variando os subconjuntos de treinamento e teste, garantindo assim uma avaliação de performance mais plausível com a realidade.

Como avaliação da performance utilizou-se o erro médio absoluto (MAE), a raiz quadrada do erro médio (RMSE) e o coeficiente de correlação de concordância (ρ_c).

Além disso, como avaliação de viabilidade adicional do modelo foi calculado um modelo nulo baseado na média. Foi avaliado a importância média de cada variável para a previsão da argila do solo ao longo das 100 rodadas. Todo o processamento



dos dados e análises gráficas foram realizados no software R (R CORE TEAM, 2023).

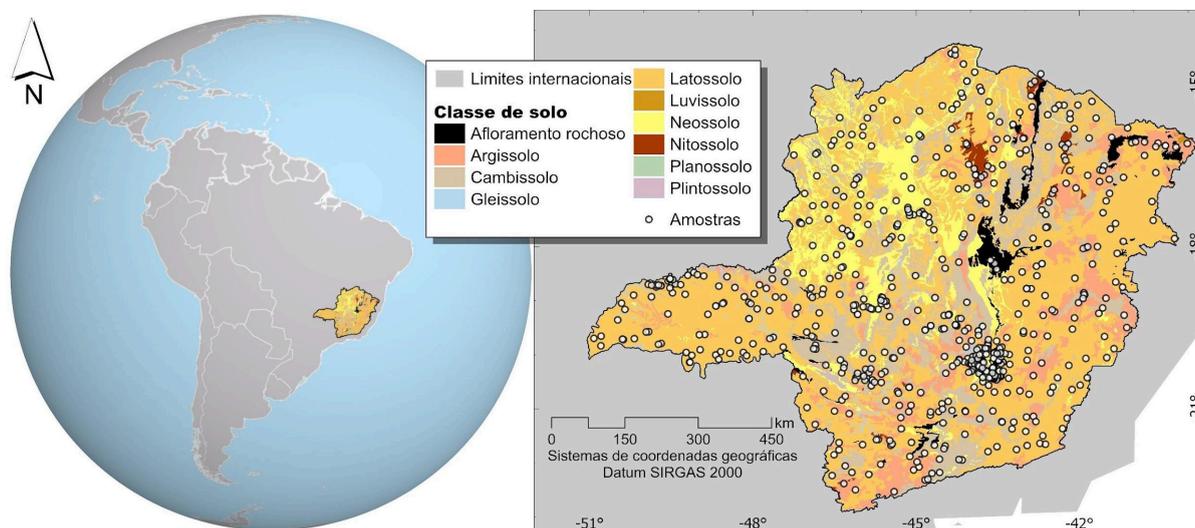


Figura 1. Mapa de localização das amostras em cada ordem de solos no estado de Minas Gerais, Brasil.

Tabela 1. Preditores ambientais utilizados para modelagem dos teores de argila do solo no estado de Minas Gerais.

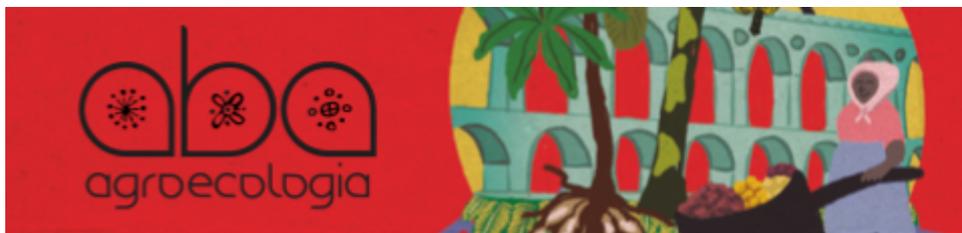
Clima	NDVI	Solo	K40
TMA	NPP	CT	DZ1
PMA	Relevo	eTh	IMT
Organismos	MDE	eU	Umidade do solo

Em que: TMA = temperatura média anual; PMA = precipitação média anual; NDVI = índice de vegetação por diferença normalizada; NPP = produtividade primária líquida; MDE = modelo digital de terreno; CT = contagem total da radiação gama; eTh = equivalente de tório; eU = equivalente de urânio; K40 = equivalente de potássio; DZ1 = 1ª derivada vertical da anomalia magnetométrica e; IMT = magnometria – intensidade magnética total.

Resultados e Discussão

As métricas de performance denotaram um ajuste razoável do modelo qrf (Figura 2). Ao comparar os erros (MAE e RMSE) do modelo nulo com o qrf verifica-se que os erros do último modelo são consideravelmente menores que os do modelo nulo. O coeficiente de correlação de concordância alcançou valores médios maiores que 0,5 no conjunto de teste, evidenciando que mesmo com a remoção do bias (proporcionado pelo cálculo da métrica) o modelo é capaz de generalizar as concentrações de argila do solo de Minas Gerais com relativa acurácia.

As diferentes classes de preditores ambientais se distribuíram de forma variada no ranking de importância nas 100 execuções do modelo (Figura 3). Três das variáveis relacionadas ao solo (eTh, CT e Umidade do solo) foram as mais importantes, seguidas pelo clima (TMA) e organismos (NPP). Destaca-se a importância do eTh para o modelo, pois de todos os preditores foi o que expressou maior importância e



a menor amplitude nas 100 execuções, o que sugere a forte relação do radionuclídeo eTh com a argila do solo em Minas Gerais.

Isso pode ser explicado pela maior intensidade de raios gama do eTh emitidos pelo solo, pois aumenta a capacidade do equipamento captar o sinal em maiores profundidades. Além disso, a mineralogia do solo é fortemente influenciada por estas assinaturas espectrais de raios gama (TAYLOR et al., 2002), o que explica as variáveis de gamaespectrometria apresentarem elevada relação com a argila do solo.

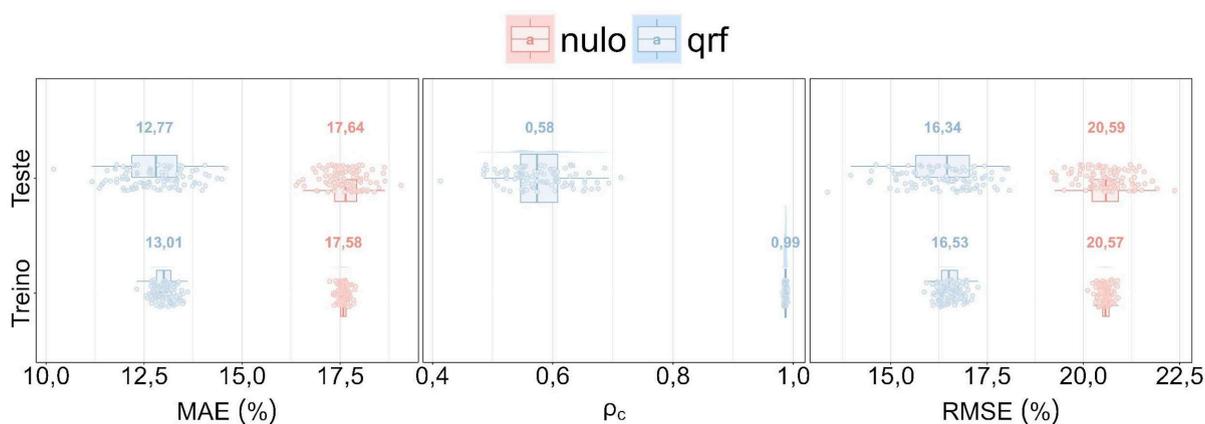


Figura 2. Erro médio absoluto (MAE) e raiz quadrada do erro médio (RMSE) para o quantile random forest (qrf) e modelo nulo e coeficiente de correlação de concordância (ρ_c) para o qrf ajustado para o teor de argila do solo de Minas Gerais, Brasil. Os valores acima dos boxplots indicam valores médios.

A importância da umidade do solo pode ser explicada devido à sua relação direta com a argila, pois solos argilosos tendem a armazenar mais água que solos arenosos. Além disso, o ERA5 (COPERNICUS, 2019) utiliza textura do solo em seus modelos se correlacionado de forma direta com os teores de argila, o que auxilia na explicação das fortes relações encontradas neste estudo.

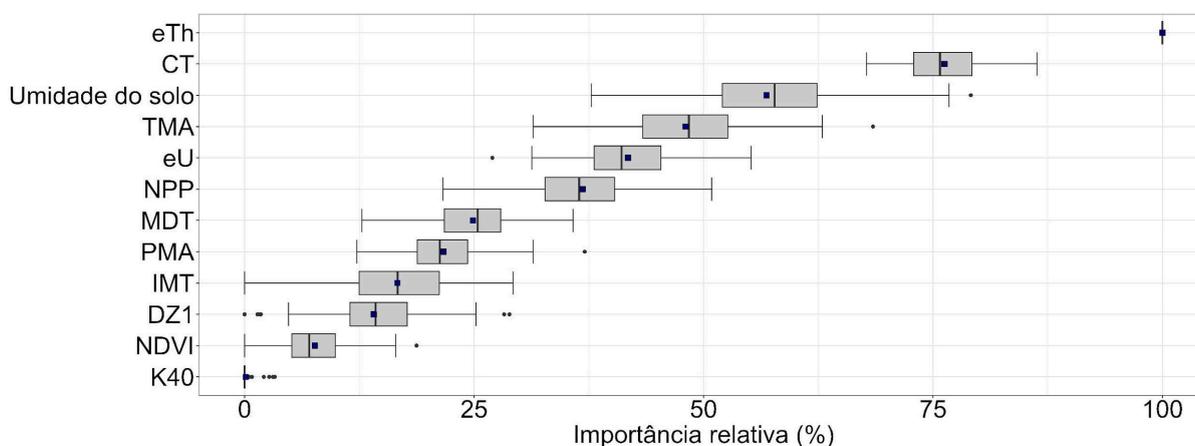
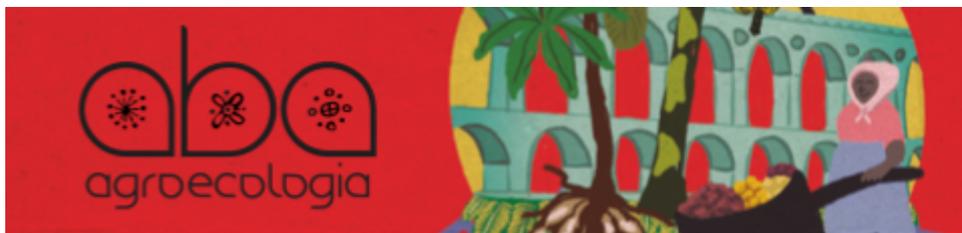


Figura 3. Importância relativa dos preditores ambientais utilizados para modelagem dos teores de argila dos solos de Minas Gerais, Brasil. Quadrados azuis indicam valores médios.



A distribuição da concentração de argila expressou ampla variação no estado (Figura 4). Concentrações menores foram estimadas na região norte e oeste do estado, enquanto na porção central e na Zona da Mata mineira foram observadas maiores concentrações. A ampla variabilidade é fundamentada na literatura, uma vez que as áreas com baixas concentrações de argila são dominadas por solos resultantes do intemperismo de rochas ígneas com granulometrias mais grosseiras e predominância de quartzo e o arenito (sedimentar) presente na bacia sedimentar São Francisco.

Isto consequentemente reduz os teores de argila. Já na porção central, encontram-se solos argilosos derivados de rochas sedimentares mais finas (CPRM, 2004) e do clima úmido predominante no estado. As maiores incertezas foram associadas às áreas com baixos teores de argila. Isso relaciona-se com tanto com a baixa representatividade amostral de solos arenosos, quanto com a dificuldade do qrf em realizar predições com valores extremos, o que consequentemente dificulta a detecção de padrões do modelo para esta faixa de dados, aumentando as incertezas.

Os mapas produzidos podem auxiliar os agricultores familiares no acesso a crédito, por exemplo. Grande parte dos programas de acesso a crédito destinado aos agricultores familiares dependem de garantias fornecidas pelos camponeses acerca do potencial produtivo de suas áreas para o banco. Uma das características muito utilizadas é a textura do solo, inclusive solicitada de forma obrigatória por instituições financeiras. Os mapas de argila em 1km de resolução espacial podem subsidiar informações e apontar perspectivas da capacidade produtiva do solo, sobretudo na capacidade de armazenamento de água.

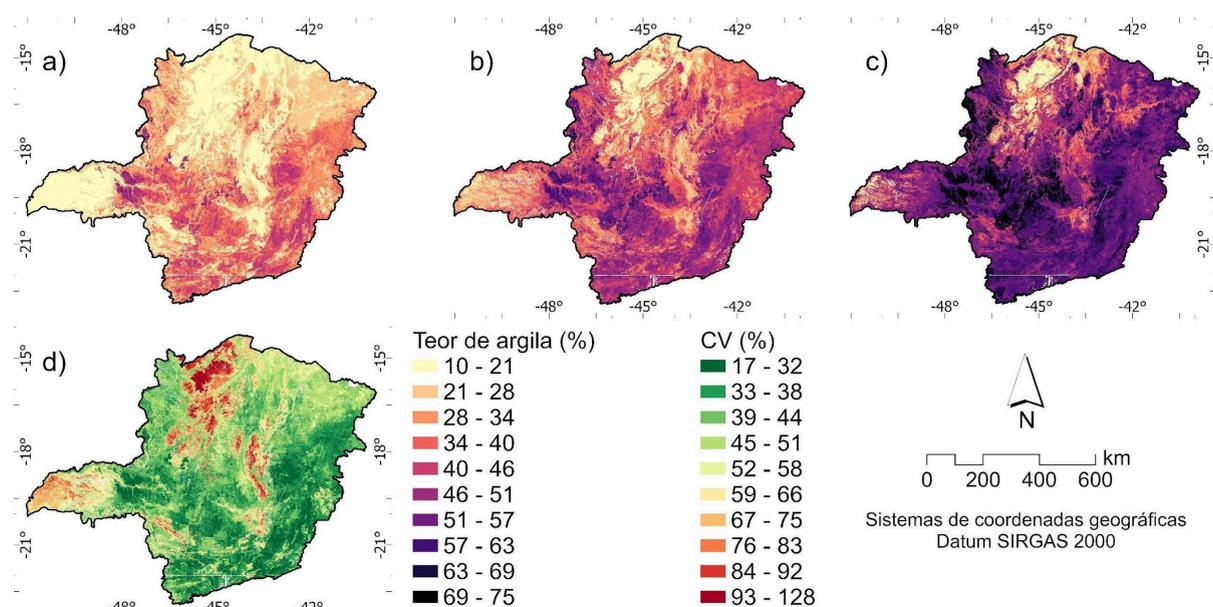
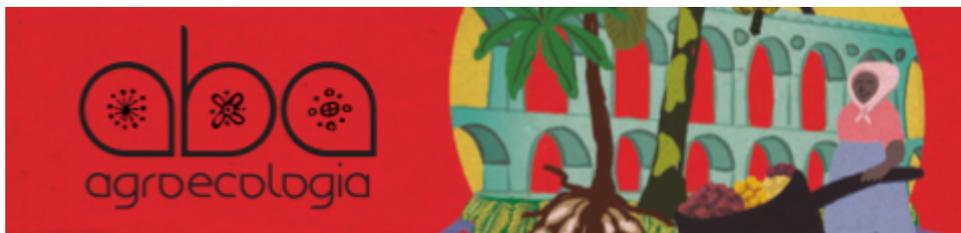


Figura 4. Mapas do quantil 25% (a), médio (b), quantil 75%(c) e coeficiente de variação (d) dos teores de argila preditos com o quantile random forest para Minas Gerais, Brasil.



Nesse contexto, os produtos gerados dialogam tanto com a agricultura familiar quanto com a compreensão da dinâmica dos processos ambientais naturais e antrópicos. Isso oferece perspectivas para orientar os formuladores de políticas públicas sobre estratégias de planejamento do uso e ocupação da terra, o que é consequentemente capaz de mitigar a crise ecológica.

Conclusões

O modelo ajustado alcançou performance satisfatória para predição dos mapas. O solo e o clima são os principais fatores que controlam a distribuição de argila no estado de Minas Gerais.

Referências bibliográficas

ALVARES, Clayton. A. et al. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, n. 6, p. 711–728, 2013. DOI: 10.1127/0941-2948/2013/0507. Disponível em: <https://doi.org/10.1127/0941-2948/2013/0507>.

COPERNICUS. **ERA5-Land hourly data from 1981 to present. Climate Data Store**, ECMWF, 2019. DOI: 10.24381/CDS.E2161BAC. Disponível em: <https://doi.org/10.24381/cds.e2161bac>.

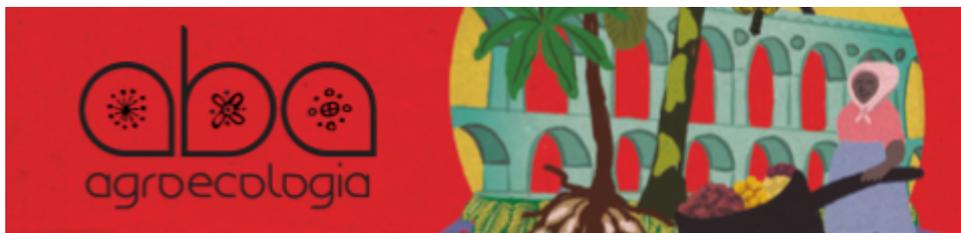
CPRM. **Carta Geológica do Brasil ao Milionésimo: sistema de informações geográficas–SIG [Geological Map of Brazil 1: 1.000. 000 scale: Geographic Information System–GIS]**. Brasília, 2004.

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

GUEVARA, Yang. Z. C. et al. Reference values of soil quality for the Rio Doce Basin. **Revista Brasileira de Ciência do Solo**, v. 42, 2018. DOI: 10.1590/18069657rbc20170231. Disponível em: <https://doi.org/10.1590/18069657rbc20170231>.

R CORE TEAM. **R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing**. Vienna, Austria, 2023. Disponível em: <http://www.r-project.org>.

SIQUEIRA, Rafael. G. et al. Machine learning applied for Antarctic soil mapping: Spatial prediction of soil texture for Maritime Antarctica and Northern Antarctic Peninsula. **Geoderma**, v. 432, p. 116405, 2023. DOI: 10.1016/j.geoderma.2023.116405. Disponível em: <https://doi.org/10.1016/j.geoderma.2023.116405>.



SOUZA, José. J. L. L. et al. Geochemistry and spatial variability of metal(loid) concentrations in soils of the state of Minas Gerais, Brazil. **Science of the Total Environment**, v. 505, p. 338–349, 2015. DOI: 10.1016/j.scitotenv.2014.09.098. Disponível em: <https://doi.org/10.1016/j.scitotenv.2014.09.098>.

TAYLOR, Miranda. J. et al. Relationships between soil properties and high-resolution radiometrics, central eastern Wheatbelt, Western Australia. **Exploration Geophysics**, v. 33, n. 2, p. 95–102, 2002. DOI: 10.1071/EG02095. Disponível em: <https://doi.org/10.1071/EG02095>.