

Mapeamento Digital dos Teores de Fe_2O_3 , MnO , Nb e TiO_2 em Morro dos Seis Lagos – AM, Brasil

Digital Mapping of Fe_2O_3 , MnO , Nb and TiO_2 Contents in Morro Seis Lagos – AM, Brazil

Niriele Bruno Rodriguesⁱ

Universidade Federal Rural do Rio de Janeiro
Seropédica, Brasil

Júlio Cesar Lopes da Silvaⁱⁱ

Universidade Federal do Rio de Janeiro
Rio de Janeiro, Brasil

Renan Pereira Marinatti da Silvaⁱⁱⁱ

Universidade Federal Rural do Rio de Janeiro
Seropédica, Brasil

Helena Saraiva Koenow Pinheiro^{iv}

Universidade Federal Rural do Rio de Janeiro
Seropédica, Brasil

Waldir de Carvalho Júnior^v

Embrapa Solos
Jardim Botânico, Brasil

Resumo: Conceitos em pedometria e técnicas de *Machine Learning* são cada vez mais utilizados na execução de levantamentos de solos, empregando procedimentos de mapeamento digital de solos. O objetivo do estudo foi avaliar o desempenho de modelos *Multivariate Adaptive Regression Spline* (MARS), *Radial Support Vector Machine* (SVMRadial) e *Random Forest* (RF), para predição espacial de Fe_2O_3 , MnO , Nb e TiO_2 , em Morro dos Seis Lagos-AM, Brasil. A metodologia consistiu em: Revisão bibliográfica; Compilação dos dados geoquímicos; Tratamento e análise dos dados (*input data*); Seleção de

ⁱ Doutoranda pelo Programa de Pós-graduação em Agronomia-Ciência do Solo, Departamento de Solos, Instituto de Agronomia. nirielebr@ufrj.br. <https://orcid.org/0000-0001-6230-0058>

ⁱⁱ Doutorando pelo Programa de Pós-graduação em Geologia, Instituto de Geociências. jlopes@geologia.ufrj.br. <https://orcid.org/0000-0001-6230-0058>

ⁱⁱⁱ Graduando em Geologia, Departamento de Geociências. marinatti.renan@gmail.com. <https://orcid.org/0009-0006-8015-9071>

^{iv} Profa. Adjunta no Departamento de Solos, Instituto de Agronomia. koenow@ufrj.br. <https://orcid.org/0000-0001-5742-7556>.

^v Pesquisador na Embrapa Solos. waldir.carvalho@embrapa.br. <https://orcid.org/0000-0002-8619-0989>.

covariáveis; Aplicação de algoritmos para predição de elementos; Obtenção dos mapas, análise dos resultados e interpretações. Os resultados demonstraram maior acurácia para a predição de teores de óxido de ferro (Fe_2O_3), manganês (MnO) e nióbio (Nb) com o modelo RF, já para titânio (TiO_2), melhor desempenho foi observado com o modelo SVMRadial. As covariáveis morfométricas foram mais relevantes do que covariáveis derivadas de índices espectrais.

Palavras-chave: Pedometria; *Machine-Learning*; Áreas de Difícil Acesso.

Abstract: Concepts in pedometrics and Machine Learning techniques are increasingly used in the execution of soil surveys, employing digital soil mapping procedures. The objective of the study was to evaluate the performance of Multivariate Adaptive Regression Spline (MARS), Radial Support Vector Machine (SVMRadial) and Random Forest (RF) models, for spatial prediction of Fe_2O_3 , MnO , Nb and TiO_2 , in the Morro dos Seis Lagos-AM, Brazil. The methodology consisted of: Thematic review; Compilation of geochemical data; Treatment and analysis of data (input data); Selection of covariates; Application of algorithms for prediction of elements; obtaining the maps, analysis of results and interpretations. The results showed greater accuracy for the prediction of iron oxide (Fe_2O_3), manganese (MnO) and niobium (Nb) contents, with the RF model, while for titanium (TiO_2), better performance was observed with the SVMRadial model. Morphometric covariates were more relevant than covariates derived from spectral indices.

Keywords: Pedometrics; Machine Learning; Poorly-Accessible Areas.

Introdução

A demanda por acesso à informação de recursos naturais teve aumento significativo nas últimas décadas, em especial sobre o recurso solo. Nesse contexto, o mapeamento digital de solos (MDS) desempenha um papel fundamental ao lidar com o vasto volume de dados e gerar informações sobre os solos e suas propriedades. Nesse sentido, os acervos existentes de dados de levantamento de solo legados e patrimoniais em todo o mundo estão sendo resgatados, compilados e processados em um conjunto de dados, consistente e geograficamente contíguo de propriedades relevantes do solo que cobrem a superfície terrestre do planeta (ARROUAYS *et al.*, 2017).

O uso do mapeamento digital para predizer a distribuição espacial de propriedades e classes dos solos requer a organização de conjuntos de dados espaciais, incluindo dados de solos e variáveis ambientais relacionadas. Segundo Fernandes Filho *et al.* (2023), as novas técnicas disponíveis e avanços tecnológicos no mapeamento digital de solos podem ser aplicados para conduzir avaliações de solos mais rápidas, e com custo-benefício vantajoso, permitindo a continuidade dos levantamentos de solos no Brasil. Ressalta-se que os levantamentos pedológicos não se limitam apenas a mapas cartográficos, mas também proporcionam uma gama de informações acerca do recurso para suporte a tomada de decisões.

Atualmente, a prospecção mineral tem se baseado em sistema de informação geográfica (SIG), com a finalidade de suporte a exploração mineral (HRONSKY; KREUZER, 2019). Nesse sentido, o aprendizado de máquina tem sido empregado em diversas áreas de conhecimento, associado à disponibilidade de conjuntos de dados e à evolução dos algoritmos, contribuindo para o sucesso das previsões na obtenção de informações mineralógicas de qualidade e com confiabilidade conhecida. Conforme Prado *et al.* (2020), os métodos de mapeamento para a prospecção mineral fornecem ferramentas objetivas para a integração de grandes quantidades de dados geocientíficos. Dessa forma, esses métodos podem ser aplicados em áreas com pouco ou nenhum local mineralizado conhecido (PRADO *et al.*, 2020).

Estudos recentes, como os dos autores Cracknell e Reading (2014), Costa *et al.* (2019), Pinheiro *et al.* (2021) e Wang *et al.* (2020), utilizam as técnicas de *Machine-Learning* para a previsão de dados geoquímicos e, principalmente quando integrado com dados de sensoriamento remoto, têm contribuído para o mapeamento de teores de elementos e compostos, que superam as técnicas convencionalmente utilizadas. Segundo Demattê *et al.* (2018), o uso do sensoriamento remoto (SR) tem se tornado cada vez mais importante na modelagem espacial e no mapeamento digital de solos. Isso se deve à existência de uma extensa relação entre o solo e a interação com a energia eletromagnética na forma de distintos padrões espectrais, o que reforça a relevância dessa técnica.

A área de estudo, denominada Complexo Carbonatítico Morro dos Seis Lagos, onde o intemperismo da siderita carbonatito originou espessa crosta laterítica (espessura maior que 200 metros), é composta predominantemente de *goethita/hematita* (CPRM, 2019; BENTO *et al.*, 2022). Frisa-se que tais carbonatitos e lateritas hospedam minerais traços com elementos que têm potencial econômico (como Nb e TiO_2 , por exemplo). Esta ocorrência está localizada no bioma amazônico sob densa floresta em áreas de proteção ambiental e reserva indígena, de difícil acesso.

Sob essa única e complexa conjuntura de hospedagem de elementos com potencial econômico, bem como área de estudo de difícil acesso, o objetivo dessa pesquisa foi avaliar o desempenho dos modelos *Multivariate Adaptive Regression Spline* (MARS), *Radial Support Vector Machine* (SVMRadial) e *Random Forest* (RF), para previsão espacial dos conteúdos de Fe_2O_3 , MnO , Nb e TiO_2 , em Morro dos Seis Lagos, São Gabriel da Cachoeira, Amazônia, Brasil.

Metodologia

Os procedimentos metodológicos dividem-se em 6 etapas (Figura 1): 1) Revisão bibliográfica; 2) Compilação dos dados geoquímicos; 3) Tratamento e análise da base de dados (*input data*); 4) Seleção de covariáveis; 5) Aplicação de algoritmos de mapeamento digital para previsão de elementos, e posterior validação; e 6) Integração análise dos resultados e interpretações.

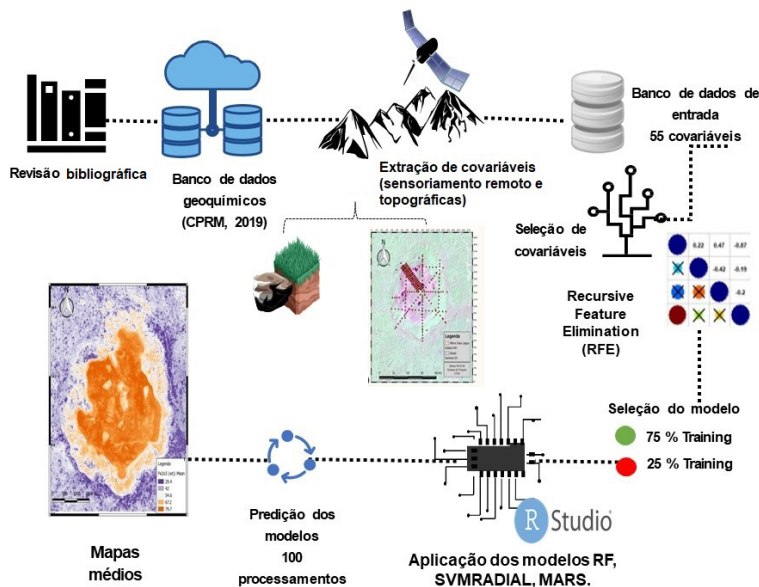


Figura 1 – Etapas dos procedimentos metodológicos.

A região de Morro dos Seis Lagos (Figura 2) está localizada na porção noroeste da bacia Amazônica, inserida nas coordenadas 756825.32 m E, 35078.26 m N, com área total de 35,96 km². Essa área é limítrofe (< 150km) com a Venezuela e a Colômbia. Está inserida no contexto de três unidades de conservação (UCs) (BRASIL, 2000) com sobreposição: a Terra Indígena Balaio (Decreto nº 21 de 21/12/2009), Parque Nacional do Pico da Neblina (Decreto nº 83.550 de 05/06/1979) e a Reserva Biológica do Morro dos Seis Lagos (Decreto nº 12.836 de 09/03/1990).

Os dados geoquímicos de solo, sedimentos e materiais rochosos utilizados foram disponibilizados pelo Serviço Geológico do Brasil (CPRM) através da plataforma GeoSBG (disponível em: <https://geoportal.cprm.gov.br/>), e encontram-se sumarizados no Projeto Avaliação do Potencial de Terras Raras no Brasil – Área Morro dos Seis Lagos, Noroeste do Amazonas (CPRM, 2009). A malha amostral superficial é de 341 pontos, onde foram realizadas determinações geoquímicas pelo Laboratório SGS Geosol (Belo Horizonte), para os 56 elementos foram aplicados os seguintes métodos (CPRM, 2009):

- Método SGS XRF79C, com identificação dos minerais feita através da comparação do difratograma do banco de dados de ICSS-PDF (*International Center for Diffraction Data-Powder Diffraction file*), para os óxidos principais;
- A Espectrometria de Massa com Plasma Indutivamente Acoplado (ICP-MS) com Espectrometria de Massa com Plasma Indutivamente Acoplado (ICP-OES) para análise de 14 Elementos de terras raras (ETR), Nb, Sn, W, Y e outros 11 elementos; e
- ISE03A e B para os elementos F e Cl.

A análise estatística exploratória dos parâmetros foi executada através do *software* RStudio (v.3.4).

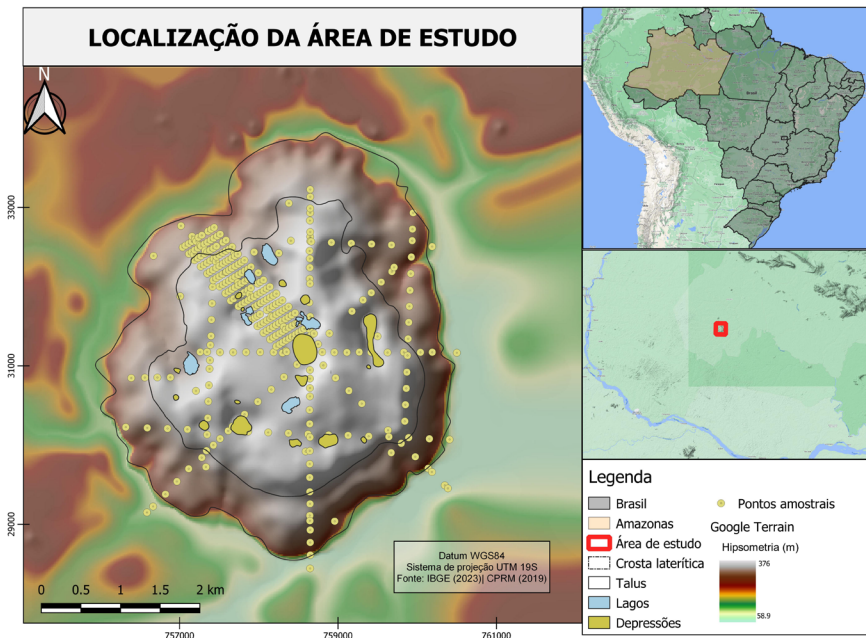


Figura 2 – Localização da área de estudo.

Covariáveis Ambientais de Entrada

Os atributos do terreno foram derivados de um Modelo Digital de Elevação Hidrologicamente Consistente (MDE-HC), com resolução espacial de 20 m, obtido através de interpolador (*Topo to Raster*) no ArcGIS Desktop v. 10.6 (ESRI), a partir de dados primários vetoriais de curvas de nível com equidistância de 10 metros, pontos cotados, hidrografia, extraído da base cartográfica digital – escala 1:25.000, produzida pela Fundação Instituto Brasileiro de Geografia e Estatística (IBGE), em parceria com a Secretaria do Estado do Ambiente (SEA). Os atributos topográficos derivados do MDE-HC foram obtidos através do programa Sistema de Análise Geocientífica Automatizadas – SAGA-GIS v.2.1.2 (CONRAD, 2007), sendo eles: *Aspect*, *Real surface area*, *Convergence Index*, *Curvature Flow line*, *Curvature General*, *Curvature maxima*, *Curvature minimo*, *Curvature Plan*, *Curvature prolife*, *Curvature Tangencial*, *Curvature Total*, *Elevation*, *Mid Slope Position*, *Multiresolution Index of Valley Bottom Flatness (MRVBF)*, *Multiresolution Index of The Ridge Top Flatness (MRRTF)*, *SAGA Weteness Index*, *Vector Terrain Ruggedness (VRM)*, *Standardized Height*, *Slope Height*, *Terrain Surface Convexity*, *Terrain Surface Texture* e *Valley Depth*.

As covariáveis proveniente de dados de sensoriamento remoto foram obtidas pelo sensor *Multispectral Sensor Instrument* (MSI) do satélite Sentinel 2A de 07/01/2020 referente a órbita/ponto 4677, e do sensor *Advanced Spaceborne Thermal Emission and Reflection Radiometer* (ASTER) do satélite Terra de 06/02/2020 referentes a órbita/ponto 0030107, na resolução espacial de 20 m, sendo os índices gerados através da combinação de bandas espectrais descritos com detalhes na Tabela 1.

Tabela 1 – Índices provenientes da combinação de bandas espectrais dos sensores MSI-Sentinel 2A e ASTER.

Índices	Equação	Referência
<i>Ferrous Silicates</i>	Banda 12/ banda 11	Van Der Meer <i>et al.</i> (2014)
Zonas alteradas	Banda 8/ banda 11	Sora <i>et al.</i> (2018)
Rochas alteradas hidrotermicamente	8-1-2-3	Sora <i>et al.</i> (2018)
<i>Laterite</i>	Banda 11/ banda 12	Van der Meer <i>et al.</i> (2014)
<i>Clay Minerals</i>	Clay Minerals = banda 6/banda 7	Sabins (1997)
<i>Iron Oxide</i>	Iron Oxide = banda 4/banda 2	Sabins (1997)
<i>Normalized Difference Vegetation Index</i> (NDVI)	NDVI = RNIR – Rred/ RNIR + Rred	Rouse <i>et al.</i> (1973)
<i>Ferroginous regolith</i>	(Banda 8A/ banda3)	Rowan <i>et al.</i> (2003)
	ASTER	
Deteccção da alteração argilítica avançada	Banda 4 /banda 6	Sora <i>et al.</i> (2018)
<i>Ferric iron, Fe3+</i>	Banda 2/ banda 1	Kalinowski e Oliver (2004)
<i>Ferrous iron, Fe2+</i>	Banda 5/banda 3 + banda 1/banda 2	Kalinowski e Oliver (2004)
Gossan	Banda 4/ banda 2	Kalinowski e Oliver (2004)

A seleção de covariáveis ambientais desempenha um papel crucial na construção de modelos, buscando capturar os padrões e relações ambientais e criar modelos mais simples e eficientes. Essa etapa aborda aspectos fundamentais, como a redução da dimensionalidade, o aprimoramento da precisão do modelo, a interpretabilidade dos modelos e a eficiência computacional. Nesse contexto, é adotado o princípio geral de simplicidade e parcimônia, enfatizando a interpretabilidade dos modelos (HASTIE *et al.*, 2009).

O procedimento de seleção de covariáveis para construção dos modelos preditivos envolveu três etapas, sendo elas: 1) Remoção de covariáveis com variância zero ou quase zero; 2) Remoção de covariáveis altamente correlacionadas com limiar de 0.95; e 3) Implementação do algoritmo *Recursive Feature Elimination* (JEONG *et al.*, 2017), para seleção das covariáveis preditores potenciais.

O processo de remoção de variáveis com variância zero ou quase zero, também conhecido como *NearZeroVar*, foi considerado na etapa de pré-processamento, buscando a identificação daquelas que possuem uma variância abaixo de um limiar pré-definido. As covariáveis foram removidas a partir do programa RStudio (3.6.1) utilizando o pacote *Caret*. Nesta primeira etapa, foram removidas as seguintes covariáveis: *Closed Depressions*, *Hill*, *Hill index*, *Slope index*, *Surface specific points*, *Valley* e *Valley index*, sendo consideradas não informativas para o conjunto de dados de *input*.

Posteriormente, foi realizada uma nova análise entre as covariáveis para exclusão por alta correlação utilizando o pacote *Caret*. Nesse contexto, considera-se que a remoção de covariáveis altamente correlacionadas favorece melhor precisão e a interpretabilidade do modelo. Desta forma, foi estabelecido um critério de limiar de 0,95 para identificar as covariáveis com correlação significativa. Com base nisso, as seguintes variáveis foram eliminadas: Banda 7, Banda 8, Banda 8a, Banda 11, Banda 5, *Mass Balance Index*, *Ferrous silic*, *Terrain Ruggedness Index*, *Clay Minerals*, *Curvature cross-sectional*, *Curvature longitudinal* e *Gossan* (ASTER).

A terceira etapa para a seleção das covariáveis preditoras, consistiu na utilização do algoritmo *Recursive Feature Elimination* – RFE (JEONG *et al.*, 2017). O método visa a eliminação recursiva por interação entre as covariáveis e a variável predita, selecionando aquelas com maior importância e retirando aquelas que apresentam colinearidade (multicolineridade), e que podem prejudicar o desempenho dos modelos (SVETNIK *et al.*, 2004).

O algoritmo RFE (*Recursive Feature Elimination*) foi implementado em cada modelo usando o pacote *Caret* (KUHNS, 2008). Deste modo, o hiperparâmetro de *subset* do RFE foi ajustado para gerar subconjuntos com diferentes números de covariáveis, com base nas covariáveis remanescentes das etapas anteriores de seleção. Foram testados cinco conjuntos específicos, incluindo as covariáveis (2:25, 30, 35, 40, 45). Assim, a seleção dos subconjuntos foi realizada por meio de validação cruzada com cinco dobras, avaliando a métrica de precisão R-quadrado para cada subconjunto.

Em termos de modelos preditivos foram testados os modelos *Multivariate Adaptive Regression Splines* (MARS) via pacote *Earth* (MILBORROW *et al.*, 2017), *Random Forest* (RF) através de pacote *randomForest* (BREIMAN, 2001), *Support Vector Machine* (SVMRadial) via pacote *Caret* (KUHNS, 2008), em função de suas características dessemelhantes, a fim de avaliar a predição mais precisa dos elementos (Fe_2O_3 , MnO , Nb e TiO_2).

Nesse sentido, os modelos diferem em especial pela gestão dessemelhante do conjunto de dados de treinamento e teste, contribuindo assim para escolhas de covariáveis distintas, avaliando a importância de cada uma. Assim, com fins de análise comparativa entre os modelos e prezando pela seleção de covariáveis altamente relacionadas aos elementos preditos, foi utilizado o algoritmo *Recursive Feature Elimination* (RFE), implementado em cada modelo como RFE- RF, RFE- SVMRadial e RFE- MARS.

Para avaliar o desempenho e selecionar o modelo com melhor desempenho para a predição de cada elemento, foi utilizado um conjunto ideal de covariáveis pré-selecionadas via RFE para cada modelo. Assim, os dados foram separados em treinamento (75%) e teste (25%) para validação. O desempenho do modelo foi avaliado com base nas métricas: Raiz quadrada do Erro médio (*Root Mean Squared Error* – RMSE) (eq. (1)), Erro

Absoluto Médio (*Mean Absolute Error* – MAE) (eq. (2)) e coeficiente de determinação (R^2), (eq. (3)).

$$RMSE = \sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (1)$$

$$MAE = \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2)$$

$$R^2 = \frac{SQR}{SQT} = \frac{\sum_{i=1}^n |\hat{y}_i - \bar{y}|}{\sum_{i=1}^n |y_i - \bar{y}|} \quad (3)$$

O processamento de cada modelo foi realizado com 100 repetições para cada elemento, possibilitando desta forma o cálculo do mapa médio e de amplitude (mínimos e máximos preditos) para cada elemento. Essa abordagem segue o procedimento adotado por Gomes *et al.* (2019) e Reis *et al.* (2021), que consiste em repetir o processamento dos conjuntos de dados para gerar o cálculo da variabilidade, demonstrando assim maior consistência na predição. De acordo com Kuhn e Johnson (2013), a repetição múltipla é importante para determinar a variabilidade da predição, uma vez que diferentes grupos de conjuntos de dados de treinamento e validação podem resultar em modelos distintos, apresentando valores de precisão diferentes.

Resultados e Discussão

Estatística Descritiva

Os resultados dos modelos preditivos foram, inicialmente, submetidos à análise estatística descritiva (Tabela 2), obtendo-se média, mediana, valor máximo, valor mínimo, coeficiente de variação e coeficiente de curtose. O óxido de ferro (Fe_2O_3) apresentou coeficiente de variação de 508,01 em relação à média de 68,99. Esse fator pode estar correlacionado com o contexto geológico da área de estudo, apresentando diferentes fontes de materiais, alteração e/ou translocação de elementos.

Para Titânio (TiO₂) exibiu desvio padrão de 4,09, com conteúdo médio de 2,93 e coeficiente de variação de 16,69. E para Nb obteve-se média de 0,74, com valores de mínimo a máximo de 0–4,02, contribuindo para coeficiente de variação baixo de 0,76.

Tabela 2 – Estatística descritiva de elementos.

Elementos	Média	Med.	Min.	Máx.	DP	CV (%)	Cs	Ck
Fe ₂ O ₃	68,99	77,4	0,12	95,1	22,59	508,01	1,66	1,88
TiO ₂	2,93	1,42	0	29,9	4,09	16,69	2,79	9,86
Nb	0,74	0,39	0	4,02	0,88	0,76	1,43	1,50
MnO	5,09	0,20	0,00	64,10	14,02	2,75	10,93	3,02

Med = mediana; Min = mínimo; Máx = máximo; DP = desvio padrão; CV = coeficiente de variação; CS = coeficiente de assimetria; e Ck = Curtose.

Os valores máximos e mínimos refletem a composição do material geológico da área de estudo, pois conforme Paye *et al.* (2010), a distribuição de metais no solo, sob condições naturais é heterogênea, e influenciada por fatores como a composição do material de origem, fatores de formação do solo e características físicas e químicas do solo.

A partir de uma análise das métricas de desempenho dos modelos (Tabela 3), no que se refere ao erro absoluto médio (MAE) e a raiz do erro quadrático médio (RMSE), dos modelos com melhor desempenho no R², verificaram-se valores significativamente mais baixos de MAE e RMSE representando assim uma a maior precisão (WILLMOTT, MATSUURA, 2005).

Constatou-se que para o Nb o RF apresentou R² 0,23, no desempenho, e RMSE e MAE inferior em relação aos demais modelos. O SVMRadial apresentou menor desempenho com R² de 0,15. Bento *et al.* (2022), através do método de interpolação *Inverse Distance Weighting* (IDW), estimaram o teor médio para Nb equivalente a 1,17 (% em peso) em Morro dos Seis Lagos. Ainda segundo o autor destaca-se que o depósito de Nb em Morro dos Seis Lagos é classificado como supergênico, sendo semelhante a outros exemplos brasileiros, como Araxá (Minas Gerais) e Catalão I e II (Goiás) que possuem a variação de teores de 1,08 e 2,48 (% em peso) Nb₂O₅.

Para o Titânio (TiO₂) o SVMRadial apresentou melhor performance com R² de 0,14, em comparação aos demais modelos. Adel *et al.* (2021), na região do Iran, constataram, para o elemento TiO₂ o R² 0,51, a partir do método de agrupamento *K-means* e rede neural artificial. Destacou-se que ambos os métodos produzem os resultados desejados, mas o método da rede neural artificial tem resultados mais precisos. Para Ferro (Fe₂O₃) o modelo mais adequado foi o RF com R² 0,23. O elevado teor de ferro na região é produto do intemperismo de um carbonatito siderítico combinado com sua exposição a um clima tropical úmido, resultando na formação de uma crosta laterítica extremamente espessa (BENTO *et al.*, 2022).

No que tange à modelagem do conteúdo de manganês (MnO) o modelo *Random Forest* apresentou melhor desempenho com R² equivalente a 0,20, e os demais modelos

são semelhantes (Tabela 3). Recentemente, Bento *et al.* (2022) destacaram que as crostas manganíferas, possuem elevados teores de elementos de terras raras (ETR) referente a furos de sondagem. A formação de óxidos de Mn deriva da lixiviação das partes próximas à superfície do perfil de intemperismo, formando complexos coloidais de Mn que precipitam devido ao aumento do pH e supersaturação da solução (ULRICH *et al.*, 2019).

Tabela 3 – Performance dos modelos *Random Forest* (RF), *Multivariate Adaptive Regression Splines* (MARS) e *Radial Support Vector Machine* (SVMRadial).

Elementos (% em peso)	Modelos	RMSE (% em peso)	MAE (% em peso)	R ² (adimensional)
Nb	RF	19,92	14,24	0,23
	MARS	20,60	15,29	0,16
	SVMRadial	21,25	14,10	0,15
TiO₂	RF	4,00	2,59	0,12
	MARS	4,00	2,61	0,11
	SVMRadial	4,03	2,32	0,14
Fe₂O₃	RF	19,77	14,12	0,23
	MARS	20,60	15,09	0,17
	SVMRadial	20,73	13,76	0,17
MnO	RF	13,13	6,57	0,20
	MARS	13,02	6,88	0,17
	SVMRadial	13,93	5,73	0,17

RF: *Random Forest*; MARS: *Multivariate Adaptive Regression Splines*; e SVM-R: *Support Vector Machine Radial Sigma*.

De modo geral, as covariáveis predictoras para os modelos apresentaram uma característica heterogênea na composição de seus conjuntos (Figura 3). De acordo com o grau de importância as covariáveis morfométricas apresentaram maior importância quando comparadas com os índices espectrais. Nesse sentido, entre as mais frequentes nos rankings destacam-se: *Elevation*, *Multiresolution Index of Valley Bottom Flatness* (MRVBF), *Saga Wetness Index*, *Standardized Height* e *Vector Ruggedness*, as quais exerceram influência significativa na explicação da variabilidade dos elementos.

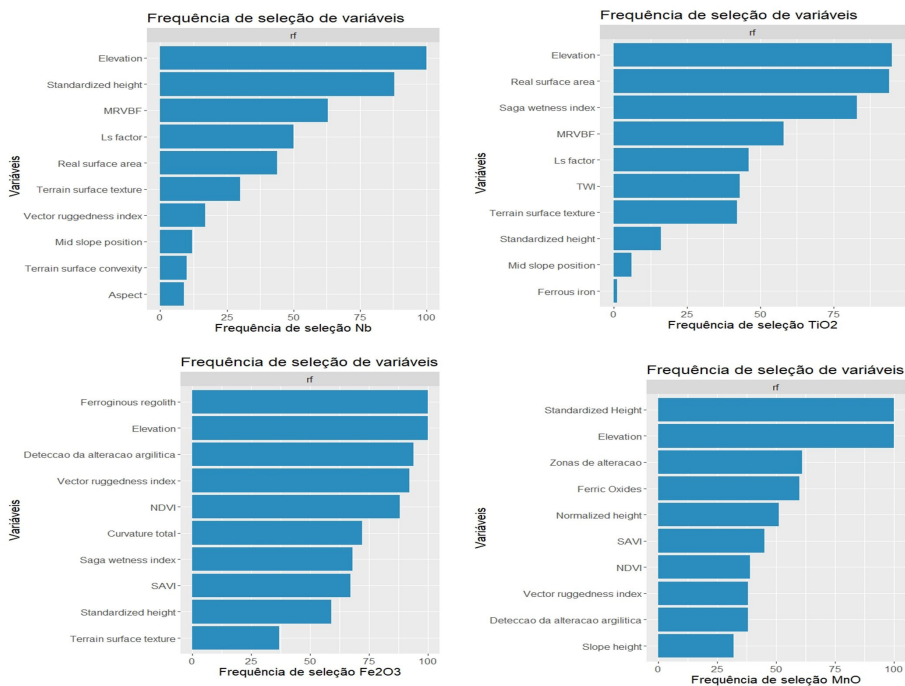


Figura 3 – Frequência de seleção das covariáveis, para a predição espacial dos elementos, a partir do modelo RFE para os modelos.

Constatou-se que as covariáveis de índices espectrais (*Ferric Iron*, *Ferrous Iron*, *Deteccão da alteração argilítica avançada*) obtiveram maior destaque nos modelos, uma vez que estão relacionadas diretamente com alterações mineralógicas, principalmente relacionadas a precipitação de óxidos de ferro. Esses resultados são relevantes para a compreensão dos processos de intemperismo das rochas carbonáticas e gênese das crostas lateríticas e, como consequência, o acúmulo de elementos como Nb e TiO_2 .

Análise da Variabilidade Espacial dos Elementos

Os mapas de média e amplitude apresentados nas Figuras 4, 5, 6 e 7 exibem a distribuição espacial dos teores dos elementos em superfície.

Na variabilidade do Nióbio (Figura 4), verifica-se que no mapa médio obteve-se uma variação do teor de 0,24–0,45 (% em peso), considerada baixa, sob o material de talus. Os maiores teores de Nb ocorreram entre 0,89–1,11 (% em peso), sob material de crosta laterítica. No mapa de desvio padrão, os maiores teores foram de 0,37–0,46 (% em peso) em áreas de depressões e em lagos.

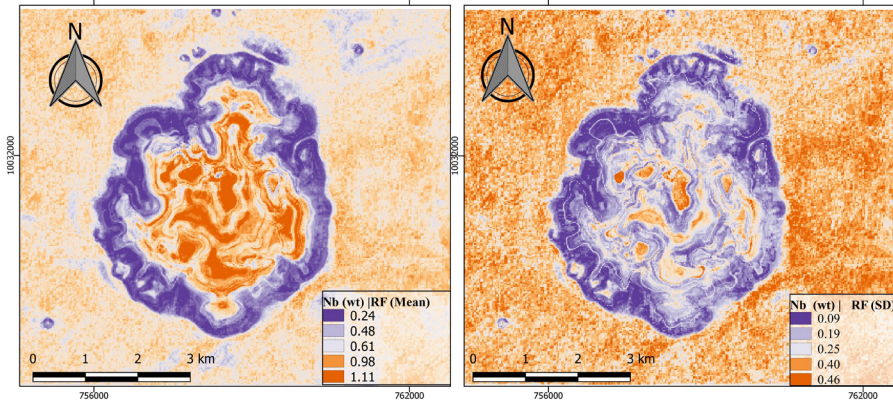


Figura 4 – Mapeamento dos teores de Nb em superfície por MARS.

As covariáveis morfométricas *Elevation*, *Aspect* e *Multiresolution Index of Valley Bottom Flatness* (MRVBF), e *LS-Factor* (Figura 3) indicaram maior importância na explicação da variabilidade do elemento. Frisa-se que estas covariáveis estão correlacionadas com a heterogeneidade topográfica e as principais morfoestruturas no entorno do Morro de Seis Lagos. Conforme Rossoni *et al.* (2016), as direções das estruturas regionais são refletidas nas estruturas rúpteis na superfície do Morro dos Seis Lagos, influenciando a carstificação, colapsos na crosta laterítica e favorecendo a concentração de minerais por processos supergênicos.

Nesse sentido, os índices espectrais *Ferrous Iron*, detecção da alteração argilítica avançada, e *ferroginous regolith*, presentes na seleção do modelo RFE-RF, podem ser correlacionados com o material. De acordo com Mitchell (2015), os carbonatos da rocha primária são lixiviados e os silicatos se alteram para argilominerais, predominando na laterita os óxidos (hematita) e hidróxidos (*goethita*) de ferro (Fe).

O elemento Titânio (TiO_2) no mapa médio (Figura 5) apresentou maior variação de teores na área de crosta laterítica com variação de 2,17 – 2,72 (% em peso). Ao oposto com menores teores de 0,52 – 1,08 (% em peso) na área de depósitos de tálus. Quanto ao mapa de desvio padrão, os maiores valores foram de 0,8 – 1,02 (% em peso), em lagos e depressões, correspondendo às áreas sem alocação de amostras. Como covariáveis preditoras importantes apresentou *Elevation*, *LS-Factor*, *Saga Wetness Index* e *Multiresolution Index of Valley Bottom Flatness* (MRVBF).

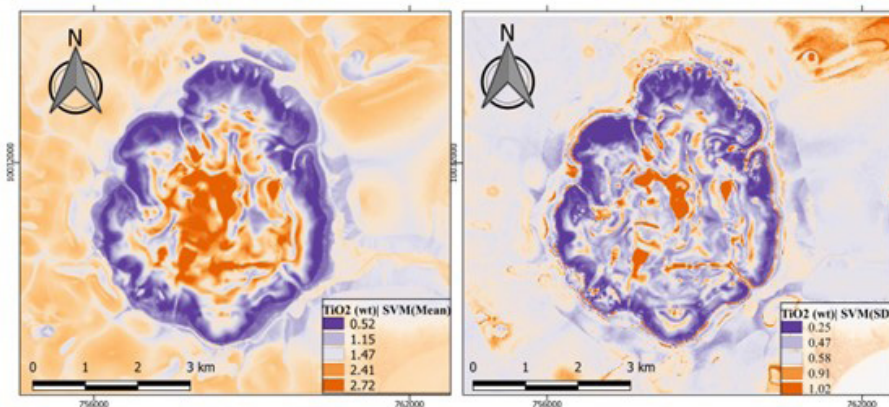


Figura 5 – Mapeamento dos teores de TiO_2 em superfície por SVM Radial.

Para Fe_2O_3 (Figura 6), os maiores teores entre 67,3–79,8 (% em peso), estão relacionados a presença da crosta laterítica, e no depósito de tálus se localizam menores teores do que na crosta laterítica. No mapa de desvio padrão ressaltam-se os valores entre 1,78–6,82 (% em peso) tanto em depressões e lagos quanto na área de tálus. As covariáveis *Elevation*, *Ferroginous regolith*, Detecção da alteração argilítica avançada, *Vector Ruggnedness index*, *NDVI*, *Saga Wetness index*, foram consideradas as mais importantes na predição da distribuição espacial dos teores de Fe_2O_3 (Figura 3).

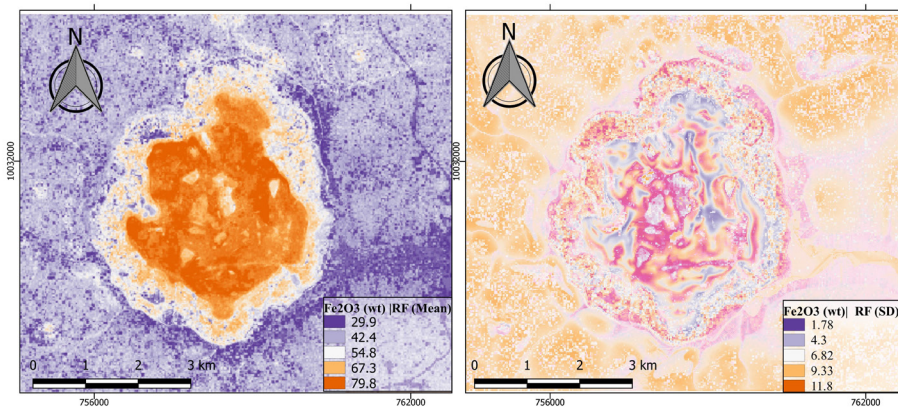


Figura 6 – Mapeamento dos teores de Fe_2O_3 em superfície por *Random Forest*.

O mapa que representa os teores médios preditos para o elemento Manganês (MnO) (Figura 7) apresentou menor teor na crosta laterítica, com valores entre 0,11 – 11,0 (% em peso), e maior teor, variando entre 32,9 – 43,9 (% em peso) na área de depósitos de tálus. No mapa de desvio padrão, os maiores valores foram de 13,6 – 18,1 (% em peso), em lagos e depressões (áreas sem alocação de amostras). As covariáveis importantes fo-

ram *Elevation*, *Ferric oxides*, Zonas de alteração, *SAVI* e Detecção da alteração argilítica avançada.

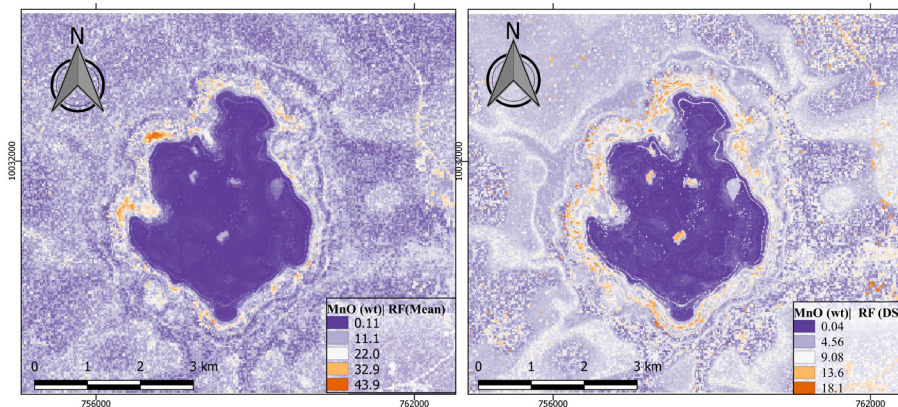


Figura 7 – Mapeamento dos teores de MnO em Morro em superfície por *Random Forest*.

Nesse sentido, o modelo *Random Forest* obteve melhor desempenho para os elementos Fe_2O_3 , Nb e MnO, indicando como as mais importantes covariáveis: *Elevation*, *SAGA Wetness Index* e Detecção argilítica, que foram pré-selecionados através do método de eliminação recursiva (RFE-RF), resultando em uma otimização do conjunto de dados de entrada.

Para o TiO_2 a melhor performance foi no modelo SVMRadial, em comparação com outros modelos, e indicando como covariáveis significativas: *Real surface area*, *Topographic Wetness index*, *Ls-Factor* e *Mid Slope Position*. De uma forma geral, em termos de índices espectrais, os mais importantes selecionados pelos modelos para a predição dos elementos foram: Detecção da alteração argilítica avançada, *Ferroginous regolith*, Zonas de alteração.

As covariáveis dos índices espectrais demonstraram um desempenho abaixo do esperado neste contexto, podendo estar sujeitas a limitações referentes a resolução espacial. El Desoky *et al.* (2022) obtiveram boa performance com a utilização do Sentinel-2A e ASTER no mapeamento geológico de média resolução.

Conclusão

Constatou-se, a partir do modelo *Random Forest* associado ao método do RFE, a obtenção de melhor desempenho na predição dos elementos Fe_2O_3 , MnO e Nb, em comparação aos modelos SVMRadial e MARS. Verificou-se que os maiores teores dos elementos Nb, TiO_2 , Fe_2O_3 estão localizados na crosta laterítica, e, por conseguinte, os teores baixos correspondendo ao material presente no tálus.

A partir do método de eliminação recursiva de feições (RFE-RF) o modelo indicou no ranking de importância as seguintes covariáveis na predição: *Elevation*, *Ls-Factor*, *Multiresolu-*

tion Index of Valley Bottom Flatness (MRVBF), SAGA Wetness Index e Detecção da alteração argilítica avançada, resultando em uma otimização do conjunto de dados de entrada.

As covariáveis morfométricas desempenharam um papel mais significativo na explicação da variabilidade espacial dos elementos químicos, em comparação com as covariáveis derivadas de índices espectrais, as quais demandam, para trabalhos futuros, análises que considerem uma resolução com maior detalhamento. Dessa forma, ressalta-se o controle topográfico do relevo associado à drenagem existente na área como reguladores da abundância dos elementos/compostos químicos analisados.

Como perspectivas futuras frisa-se que os modelos podem ser melhorados utilizando um conjunto de dados de entrada com nova malha amostral, que pode ser coletada utilizando como referência os mapas de desvio padrão, onde estes apresentam valores mais elevados. Além disso, pode-se utilizar outras covariáveis, como dados geofísicos, por exemplo, explorando diferentes modelos preditivos e algoritmos de predição, considerando o nível de detalhamento requerido na análise.

Referências Bibliográficas

ADEL, S.; MANSOUR, Z.; ARDESHIR, H. Geochemical behavior investigation based on k-means and artificial neural network prediction for titanium and zinc, Kivi region, Iran. *Bulletin of the Tomsk Polytechnic University. Geo Assets Engineering*, v. 332, n. 3, p. 113-125, 2021.

ARROUAYS, D.; LEENAARS, J. G.; RICHER DE FORGES, A. C.; ADHIKARI, K.; BALLABIO, C.; GREVE, M.; RODRIGUEZ, D. Soil legacy data rescue via GlobalSoilMap and other international and national initiatives. *GeoResJ*, v. 14, p. 1-19, 2017.

BENTO, J. P. P.; PORTO, C. G.; TAKEHARA, L., SILVA, F. J. D.; BASTOS NETO, A. C.; MACHADO, M. L.; DUARTE, A. C. Mineral potential re-evaluation of the Seis Lagos Carbonatite Complex, Amazon, Brazil. *Brazilian Journal of Geology*, v. 52, p. 1-19, 2022.

BRASIL. *Decreto-Lei nº 6.001, de 21 de dezembro de 2009*. Homologa a demarcação administrativa da Terra Indígena Balaio, localizada no Município de São Gabriel da Cachoeira, no Estado do Amazonas. Disponível em: http://www.planalto.gov.br/ccivil_03/ato2007-2010/2009/dnn/dnn12364.htm. Acesso em: 24 ago. 2021.

_____. *Decreto-Lei nº 83.550, de 05 de junho de 1979*. Cria, no Estado do Amazonas, o Parque Nacional do Pico da Neblina, com os limites que especifica e dá outras providências. Disponível em: http://www.planalto.gov.br/ccivil_03/decreto/1970-1979/d83550.htm. Acesso em: 24 ago. 2021.

_____. *Decreto-Lei nº 9.985, de 18 de julho de 2000*. Regulamenta o art. 225, § 1º, incisos I, II, III e VII da Constituição Federal, institui o Sistema Nacional de Unidades de Conservação da Natureza e dá outras providências. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/19985.htm. Acesso em: 24 ago. 2021.

Niriele Bruno Rodrigues, Júlio Cesar Lopes da Silva, Renan Pereira Marinatti da Silva, Helena Saraiva Koenow Pinheiro e Waldir de Carvalho Júnior

_____. *Lei nº 6.634, de 02 de maio de 1979*. Dispõe sobre a Faixa de Fronteira, altera o Decreto-Lei nº 1.135, de 3 de dezembro de 1970, e dá outras providências. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l6634.htm. Acesso em: 24 ago. 2021.

BREIMAN, L. Random forests. *Machine Learning Journal*. Hingham, v. 45, p. 5-32, 2001.

CAMPBELL, P. M. D. M.; FRANCELINO, M. R.; FERNANDES, E. I.; ROCHA, P. D. A.; AZEVEDO, B. C. D. Digital mapping of soil attributes using machine learning. *Revista Ciência Agronômica*, v. 50, p. 519-528, 2019.

CONRAD, O. SAGA–System für Automatisierte Geowissenschaftliche Analysen. *Presentation held at FOSSGIS*, 2007.

COSTA, I. S. L.; TAVARES, F. M.; DE OLIVEIRA, J. K. M. Predictive lithologic mapping using machine learning methods: a case study in the Grey Lineament, Carajas Province, Brazil. *Journal of the Geological Survey of Brazil*, v. 2, n. 1, p. 26-36, 2019.

CPRM. *Project Assessment of the Rare Earth Potential in Brazil: Morro dos Seis Lagos area, Northwest Amazonas* / Organized by Lucy Takehara. 4. ed. Brasília: CPRM, 2009. Disponível em: <https://geoportal.cprm.gov.br/>

CRACKNELL, M. J.; READING, A. M. Geological mapping using remote sensing data: a comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Computers & Geosciences*, v. 63, p. 22-33, 2014.

DEMATTÊ, J. A. M.; FONGARO, C. T.; RIZZO, R.; SAFANELLI, J. L. Geospatial Soil Sensing System (GEOS3): a powerful data mining procedure to retrieve soil spectral reflectance from satellite images. *Remote Sensing of Environment*, v. 212, p. 161-175, 2018.

EL-DESOKY, H. M.; SHEBL, A.; ABDEL-RAHMAN, A. M.; FAHMY, W.; EL-AWNY, H.; EL-SHERIF, A. M.; CSÁMER, A. Multiscale mineralogical investigations for mineral potentiality mapping of Ras El-Kharit-Wadi Khashir district, Southern Eastern Desert, Egypt. *The Egyptian Journal of Remote Sensing and Space Science*, v. 25, n. 4, p. 941-960, 2022.

FERNANDES FILHO, E. I.; MENDONÇA-SANTOS, M. L.; SCHAEFER, C. E.; DALMOLIN, R. S.; FRANCELINO, M. R.; CHAGAS, C. S.; CARVALHO JÚNIOR, W.; DEMATTÊ, J. A. M.; GOMES, L. C. The future of brazilian pedology: pedometrics and advanced methods for soil survey. In: KUHN, M.; JOHNSON, K. *The Soils of Brazil*, p. 423-433. Cham: Springer International Publishing 2023.

GOMES, L. C.; FARIA, R. M.; DE SOUZA, E.; VELOSO, G. V.; SCHAEFER, C. E. G.; FERNANDES FILHO, E. I. Modelling and mapping soil organic carbon stocks in Brazil. *Geoderma*, v. 340, p. 337-350, 2019.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The elements of statistical learning: Data Mining, Inference, and Prediction* (Títulos da Série – Springer Series in Statistics). Nova York: Springer, 2009.

HRONSKY, J. M.; KREUZER, O. P. Applying spatial prospectivity mapping to exploration targeting: Fundamental practical issues and suggested solutions for the future. *Ore Geology Reviews*, v. 107, p. 647-653, 2019.

JEONG, G.; OEVERDIECK, H.; PARK, S. J.; HUWE, B.; LIEB, M. Spatial soil nutrients prediction using three supervised learning methods for assessment of land potentials in complex terrain. *Catena*, v. 154, p. 73-84, 2017.

KALINOWSKI, A.; OLIVER, S. ASTER mineral index processing manual. *Remote Sensing Applications, Geoscience Australia*, v. 37, p. 36, 2004.

KUHN, M. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, v. 28, n. 5, p. 1-26, 2008. Disponível em: <https://cran.r-project.org/package=caret>. Acesso em: 20 nov. 2022.

_____; JOHNSON, K. *Applied predictive modeling*. Nova York: Springer, 2013.

MILBORROW, S.; HASTIE, T.; TIBSHIRANI, R.; MILLER, A.; LUMLEY, T. Earth: multivariate adaptive regression splines. *Package version*, v. 5, n. 2, p. 1-52, 2017.

MITCHELL, R. H. Primary and secondary niobium mineral deposits associated with carbonatites. *Ore Geology Reviews*, v. 64, p. 626-641, 2015.

PAYE, H. S.; MELLO, J. W. V.; ABRAHÃO, W. A. P.; FERNANDES FILHO, E. I.; DIAS, L. C. P.; CASTRO, M. L. O.; MELO, S. B.; FRANÇA, M. M. Valores de referência de qualidade para metais pesados em solos no Estado do Espírito Santo. *Revista Brasileira de Ciência do Solo*. v. 34, p. 2041-2051, 2010.

PINHEIRO, H. S. K.; LOPES, J. C.; DE MELLO, F. M.; BILAL, E. Digital mapping of P-LI-BE contents: a proposal to recognize LCT type pegmatites. In: 50º CONGRESSO BRASILEIRO DE GEOLOGIA, L, *Anais*. Brasília – DF, 2021. Disponível em: <https://hal-emse.ccsd.cnrs.fr/emse-03457498/>. Acesso em: 05 maio 2022.

PRADO, E. M. G.; SOUZA FILHO, C. R.; CARRANZA, E. J. M.; MOTTA, J. G. Modeling of Cu-Au prospectivity in the Carajás mineral province (Brazil) through machine learning: Dealing with imbalanced training data. *Ore Geology Reviews*, v. 124, p. 103-611, 2020.

REIS, G. B.; DA SILVA, D. D.; FERNANDES FILHO, E. I.; MOREIRA, M. C.; VELOSO, G. V.; DE SOUZA FRAGA, M.; PINHEIRO, S. A. R. Effect of environmental covariable selection in the hydrological modeling using machine learning models to predict daily streamflow. *Journal of Environmental Management*, v. 290, p. 112625, 2021.

Niriele Bruno Rodrigues, Júlio Cesar Lopes da Silva, Renan Pereira Marinatti da Silva, Helena Saraiva Koenow Pinheiro e Waldir de Carvalho Júnior

ROUSE, J. W.; HAAS, R. H.; SCHELL, J. A.; DEERING, D. W. Monitoring Vegetation Systems in the Great Plains with ERTS. *Third ERTS Symposium*, NASA special publication, NASA, v. 351, p. 309- 317, 1973.

ROSSONI, M. B.; NETO, A. C. B.; SALDANHA, D. L.; SOUZA, V. S.; GIOVANNINI, A. L.; PORTO, C. G. Aplicação de técnicas de sensoriamento remoto na investigação do controle do posicionamento do Complexo Carbonatítico Seis Lagos e no estudo do depósito (Nb) laterítico associado (Amazonas, Brasil). *Pesquisas em Geociências*, v. 43, n. 2, p. 111-125, 2016.

ROWAN, L. C.; MARS, J. C. Lithologic mapping in the Mountain Pass, California area using Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) data. *Journal of Remote Sensing of Environment*, v. 84, p. 350-366, 2003.

SABINS, F. F. *Remote sensing: principles and interpretation*. 3rd ed. Nova York: W. H. Freeman and Company, 1997.

SORA, A. M.; SIMBE, M.; DIAS, J.; UACANE, M. S. Integração de imagens de satélites para identificação de alterações na Suíte Carbonatítica de Xiluvo-Nhamatanda. *Educa-mazônia-Educação, Sociedade e Meio Ambiente*, v. 21, n. 2, p. 251-263, 2018.

SVETNIK, V., LIAW, A., TONG, C., & WANG, T. Application of Breiman's random forest to modeling structure-activity relationships of pharmaceutical molecules. *International Workshop on Multiple Classifier Systems*, p. 334-343. 2004.

ULRICH, M.; CATHELINÉAU, M.; MUÑOZ, M.; BOIRON, M. C.; TEITLER, Y.; KARPOFF, A. M. The relative distribution of critical (Sc, REE) and transition metals (Ni, Co, Cr, Mn, V) in some Ni-laterite deposits of New Caledonia. *Journal of Geochemical Exploration*, v. 197, p. 93-113, 2019.

VAN DER MEER, F. D.; VAN DER WERFF, H. M. A.; VAN RUITENBEEK, F. J. A. Potential of ESA's Sentinel-2 for geological applications. *Remote sensing of environment*, v. 148, p. 124-133, 2014.

WANG, C.; PAN, Y.; CHEN, J.; OUYANG, Y.; RAO, J.; JIANG, Q. Indicator element selection and geochemical anomaly mapping using recursive feature elimination and random forest methods in the Jingdezhen region of Jiangxi Province, South China. *Applied Geochemistry*, v. 122, p. 104-760, 2020.

WILLMOTT, C. J.; MATSUURA, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, v. 30, n. 1, p. 79-82, 2005.

Recebido em: 09/08/2023. Aceito em: 27/10/2023.