

UFRRJ
INSTITUTO DE AGRONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM
AGRONOMIA - CIÊNCIA DO SOLO

TESE

Mapeamento Digital em Formações Petroféricas
Mineralizadas: Estudo de Caso - Morro de Seis
Lagos na Amazônia brasileira

Niriele Bruno Rodrigues

2025



**UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
INSTITUTO DE AGRONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA
CIÊNCIA DO SOLO**

**MAPEAMENTO DIGITAL EM FORMAÇÕES PETROFÉRRICAS
MINERALIZADAS: ESTUDO DE CASO - MORRO DE SEIS LAGOS NA
AMAZÔNIA BRASILEIRA**

NIRIELE BRUNO RODRIGUES

Sob a Orientação do Professora
Helena Saraiva Koenow Pinheiro

e Coorientação dos Pesquisadores
Waldir de Carvalho Junior
Zamir Libohova

Tese submetida como requisito parcial
para obtenção do grau de **Doutora**, no
Programa de Pós-Graduação em
Agronomia – Ciência do Solo, Área de
concentração em Pedologia e Física do
Solo.

Seropédica - RJ
Julho de 2025

Universidade Federal Rural do Rio de Janeiro
Biblioteca Central/Seção de Processamento Técnico

Ficha catalográfica elaborada
Com os dados fornecidos pelo(a) autor(a)

R696m

Rodrigues, Niriele Bruno, 1994-
Mapeamento Digital em Formações Petroféricas Mineralizadas: Estudo de caso - Morro de Seis Lagos na Amazônia Brasileira/Niriele Bruno Rodrigues. – Seropédica, 2025.
136 f.: il.

Orientadora: Helena Saraiva Koenow Pinheiro.

Tese (Doutorado). – – Universidade Federal Rural do Rio de Janeiro, Programa de Pós-Graduação em Agronomia Ciência do Solo, 2025.

1. Pedometria. 2. Aprendizagem de máquina. 3. Áreas de Difícil Acesso. 4. Processos pedogenéticos. 5. Ferrocronatitos. I. Pinheiro, Helena Saraiva Koenow, orient. II. Carvalho Junior, Waldir de, coorient. III. Universidade Federal Rural do Rio de Janeiro. Programa de Pós-Graduação em Agronomia - Ciência do Solo. IV. Título.

É permitida a cópia parcial ou total desta tese, desde que seja citada a fonte.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.



HOMOLOGAÇÃO DE TESE DE DOUTORADO Nº 14 / 2025 - CPGACS (12.28.01.00.00.00.27)

Nº do Protocolo: 23083.042224/2025-88

Seropédica-RJ, 01 de agosto de 2025.

UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO

INSTITUTO DE AGRONOMIA

PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA - CIÊNCIA DO SOLO

NIRIELE BRUNO RODRIGUES

Tese submetida como requisito parcial para obtenção do grau de **Doutora**, no Programa de Pós-Graduação em Agronomia – Ciência do Solo, Área de Concentração em Pedologia e Física do Solo.

TESE APROVADA EM 31/07/2025.

Helena Saraiva Koenow Pinheiro. Dra. UFRRJ
(Orientadora)

Lúcia Helena Cunha dos Anjos. Ph.D. UFRRJ

Francisco Jose da Silva. Dr. UFRRJ

Carlos Ernesto Gonçalves Reynaud Schaefer. Dr. UFV

Gustavo de Mattos Vasques. Dr. Embrapa Solos

(Assinado digitalmente em 02/08/2025 06:23)

FRANCISCO JOSE DA SILVA
PROFESSOR DO MAGISTERIO SUPERIOR
DPETRO (11.39.40)
Matrícula: 1580057

(Assinado digitalmente em 05/08/2025 08:54)

HELENA SARAIVA KOENOW PINHEIRO
PROFESSOR DO MAGISTERIO SUPERIOR
DeptS (12.28.01.00.00.00.33)
Matrícula: 2223668

(Assinado digitalmente em 01/08/2025 14:05)

LUCIA HELENA CUNHA DOS ANJOS
PROFESSOR DO MAGISTERIO SUPERIOR
DeptS (12.28.01.00.00.00.33)
Matrícula: 387335

(Assinado digitalmente em 01/08/2025 15:05)

GUSTAVO DE MATTOS VASQUES
ASSINANTE EXTERNO
CPF: 084.272.437-07

(Assinado digitalmente em 12/08/2025 10:15)

CARLOS ERNESTO GONÇALVES REYNAUD SCHAEFER
ASSINANTE EXTERNO
CPF: 540.091.466-72

DEDICATÓRIA

Dedico este trabalho aos meus pais, Nádia Bruno e Dilson Francisco, pela vida e suporte, aos meus irmãos, Pablo e Thauan, e aos meus sobrinhos; a todos os meus amigos, pelo incentivo ao longo da caminhada.

AGRADECIMENTOS

Primeiramente, agradeço a Deus por ter me sustentado, e dado sabedoria e força para que chegasse até aqui.

Gostaria de agradecer aos meus pais, Nádia dos Reis Bruno e Dilson Francisco Rodrigues e irmãos Pablo e Thauan, e a minha avó Dulce Francisco e avô João Mendes (*in memoriam*), pelo apoio no decorrer dessa trajetória, e por acreditarem nas minhas realizações. Aos meus pets e 'assistentes' nesta jornada: Simba, Pretinha (*in memoriam*), Katarina e Florzinha.

Aos amigos pela força na caminhada, parceria e momentos de trocas de conhecimentos. Ao grupo de retiro de escrita científica da profa. Aretha Guimarães. À professora Laura Mendes, por me apresentar e incentivar no caminho da Ciência do Solo durante a minha trajetória na Geografia, e ao grupo de pesquisa GP- EVOL, onde iniciei a jornada acadêmica, sou grata pelo suporte e boas energias.

A professora Helena Pinheiro, pela contribuição no desenvolvimento deste trabalho, paciência e orientação. E o pesquisador Waldir de Carvalho Júnior, pela contribuição no processo de aprendizagem das técnicas de Mapeamento Digital de Solos. À professora Lúcia Anjos, expresse meu sincero agradecimento pelas importantes contribuições a esta pesquisa e pelo incentivo à minha participação no edital PSDE/CAPES. Agradeço, em especial, pela conversa que foi um impulso para a minha experiência internacional.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo privilégio da concessão bolsa de doutorado e doutorado sanduíche, possibilitando dedicação a essa fase de construção de conhecimento. A *Crop, Soil, and Environmental Sciences - University of Arkansas* pela concessão bolsa de doutorado sanduíche, possibilitando a extensão da minha permanência.

Ao Dr. Phillip Owens, Dr. Zamir Libohova, Dr. David Miller e equipe que possibilitaram a minha estadia na University of Arkansas em parceria com o U.S. *Department of Agriculture* (USDA) e a *Ozark Literacy Council*, gratidão por me apresentarem um novo mundo de diversidade cultural.

Aos colegas João Victor Bueno, Theresa Rocco, Marcelo Mancini e Quentim D. Read pelo acompanhamento nas etapas de processamento dos dados.

Meus sinceros agradecimentos à banca examinadora pelas pertinentes contribuições, que foram essenciais para a qualidade final deste estudo.

Aos funcionários do Departamento de Solos, em especial aos secretários Marquinhos, Wagner e Michelle pelo acompanhamento nessa jornada.

Ao Programa de Pós-Graduação em Agronomia – Ciência do Solo.

Agradeço a todas as pessoas que direta ou indiretamente, contribuíram para que este trabalho fosse desenvolvido e realizado com sucesso.

Muito Obrigada!

“O rio atinge seus objetivos porque aprendeu a contornar obstáculos”.
Lao-Tse

BIOGRAFIA

Niriele Bruno Rodrigues, filha de Nádia dos Reis Bruno e Dilson Francisco Rodrigues, nasceu em 12 de junho de 1994, na cidade do Rio de Janeiro (RJ). Mulher negra, cientista e periférica, cursou o ensino fundamental e médio em escolas públicas no subúrbio do Rio de Janeiro.

Em 2017, graduou-se em Geografia (Licenciatura) pelo Instituto Multidisciplinar da Universidade Federal Rural do Rio de Janeiro (UFRRJ), tornando-se a primeira pessoa da sua família a conquistar um diploma em uma universidade pública federal.

Durante a graduação, foi bolsista de iniciação científica em projetos do Grupo de Pesquisa de Evolução e Dinâmica da Paisagem (GP EVOL), vinculado ao Laboratório de Geociências (IM-UFRRJ), em parceria com o Laboratório de Meio Ambiente do Instituto Federal do Rio de Janeiro. Atuou no projeto *"Águas e Solos Urbanos: Uma análise na bacia hidrográfica do rio Botas, na Baixada Fluminense (RJ), para identificar geoindicadores de mudanças ambientais antropogênicas."*

Em 2018, ingressou no Programa de Pós-Graduação em Agronomia – Ciência do Solo (UFRRJ), concluindo o mestrado em novembro de 2021. Doutoranda no mesmo programa, realizando um período de estágio de doutorado sanduíche na *University of Arkansas* (EUA) e no *United States Department of Agriculture* (USDA), através do edital Capes PDSE.

Nasceu no bairro de Guadalupe, na zona norte do Rio de Janeiro, e desde cedo teve interesse na área científica, através das suas visitas de veraneio à região da Costa Verde (RJ). Seu desejo inicial era ser investigadora/pesquisadora, e o interesse pela ciência do solo surgiu a partir de reportagens sobre movimentos de massa na região da Costa Verde, lugar que frequentava na infância com sua família. Esse interesse se aprofundou durante a graduação, em uma disciplina chamada "Tópicos especiais em Geografia Física", onde teve o desafio de apresentar um seminário sobre o recurso natural solo.

Desde então, vem se dedicando ao estudo dos solos, integrando grupos de pesquisa na Geografia e na Agronomia, com o objetivo de contribuir para o avanço do conhecimento nessa área.

RESUMO

RODRIGUES, Niriele Bruno. **Mapeamento digital em formações petroféricas mineralizadas: Estudo de caso - Morro dos Seis Lagos na Amazônia Brasileira**. 2025. 136f. Tese (Doutorado em Agronomia, Ciência do Solo) – Instituto de Agronomia, Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, 2025.

A combinação de dados de detecção remota e topográficos associados a modelos de aprendizado de máquina (*Machine Learning* - ML), especialmente para a cartografia digital geológica e pedológica, tem contribuído para a identificação de áreas com potencial econômico para a prospecção mineral. No entanto, os aspectos pedológicos dessas formações são ainda pouco explorados. Os carbonatitos, em especial, são raros e a Amazônia brasileira abriga uma mineralização particular de siderita carbonatito, representando um depósito mineral de potencial classe mundial de nióbio (Nb_2O_5). Essa formação contém uma espessa crosta laterítica (>200 m), onde os processos de intemperismo atuando sobre a siderita contribuíram para produzir uma crosta de goethita/hematita. Os objetivos da presente pesquisa foram: investigar o processo de laterização em carbonatito siderítico raro (ferrocarbonatitos) avaliando: a variabilidade espacial dos elementos químicos; o efeito do desenho da malha amostral sobre os elementos Al_2O_3 , Fe_2O_3 , MnO , Nb_2O_5 e TiO_2 ; a distribuição dos materiais geológicos (subclasses); e os índices de intemperismo (*Index of Lateritisation* - IOL, *Chemical Index of Alteration* - CIA, e o *Chemical Index of Weathering* - CIW). Para isso, utilizou-se aprendizado de máquina combinando covariáveis morfométricas com imagens de sensoriamento remoto. O conjunto de dados de entrada reúne dados geoquímicos de 341 amostras (solo, sedimentos e materiais rochosos), combinados com covariáveis morfométricas e índices espectrais, obtidos a partir de dados de detecção remota. Essas informações foram extraídas de bandas dos sensores Sentinel-1A SAR, Sentinel-2A e *Advanced Spaceborne Thermal Emission and Reflection Radiometer* (ASTER), além de atributos topográficos derivados de um modelo digital de elevação hidrologicamente condicionado (MDE-HC), com resolução de 20 metros. Os dados coletados foram modelados utilizando algoritmos de *machine learning*. As covariáveis mais importantes para cada composto mineral e para cada modelo foram selecionadas por meio do algoritmo *Recursive Feature Elimination* (RFE). Os modelos que obtiveram melhor desempenho preditivo foram o *Random Forest* (RF), seguido pelo *K-Nearest Neighbors* (KNN). O ranking de importância fornecido pelo RFE indicou que, inicialmente, os atributos do terreno exercem uma influência mais significativa na variabilidade dos elementos contidos nas lateritas do que os índices espectrais de sensoriamento remoto. Nesse contexto, verificou-se que as características do relevo local desempenham um papel fundamental na compreensão da variação espacial dos compostos minerais, dada a maior influência das covariáveis morfométricas e SAR na predição dos diferentes elementos e compostos.

Palavras-chave: Pedometria. Aprendizagem de máquina. Áreas de Difícil Acesso. Processos pedogenéticos. Ferrocarbonatitos.

ABSTRACT

RODRIGUES, Niriele Bruno. **Digital mapping in mineralized petroferic formations: Case study - Morro dos Seis Lagos in the Brazilian Amazon.** 2025. 136p. Thesis (Doctorate in Agronomy, Soil Science). Instituto de Agronomia, Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, 2025.

The combination of remote sensing and topographic data associated with Machine Learning (ML) models, especially for mapping digital geological and pedological formations, has contributed to the identification of areas with economic potential for mineral prospecting. However, the pedological aspects of these formations are still little explored. The carbonatite formations, in particular, are rare and the Brazilian Amazon is home to a particular mineralization of siderite carbonatite, representing the largest known deposit of niobium (Nb_2O_5). This formation contains a thick lateritic crust (>200 m), where the carbonation processes of the siderite have produced a goethite/hematite crust. The objectives of this research were: to investigate the lateritization process in rare siderite carbonatite (ferrocarbonatites) by evaluating: spatial variability; the effect of the design of the sample mesh on the elements Al_2O_3 , Fe_2O_3 , MnO , Nb_2O_5 and TiO_2 ; the distribution of geological materials (subclasses); and weathering indices (Index of Lateritisation - IOL, Chemical Index of Alteration - CIA, and the Chemical Index of Weathering - CIW). Machine learning was used to combine morphometric covariates with remote sensing images. The input data set includes geochemical data from 341 samples (soil, sediment and rock materials), combined with morphometric covariates and spectral indices obtained from remote sensing data. This information was extracted from bands of the Sentinel-1A SAR, Sentinel-2A and Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) sensors, as well as topographic attributes derived from a hydrologically conditioned digital elevation model (MDE-HC), with a resolution of 20 meters. The data collected was modeled using machine learning algorithms. The most important covariates for each mineral compound and for each model were selected using the Recursive Feature Elimination (RFE) algorithm. The models with the best predictive performance were Random Forest (RF), followed by K-Nearest Neighbors (KNN). The ranking of importance provided by the RFE indicated that, initially, terrain attributes exert a more significant influence on the variability of the elements contained in the laterites than remote sensing spectral indices. In this context, it was found that the characteristics of the local relief play a fundamental role in understanding the spatial variation of mineral compounds, given the greater influence of morphometric and SAR covariates in predicting the different elements and compounds.

Key words: Pedometry. Machine learning; Poorly-Accessible Areas. Pedogenetic processes. Ferrocarbonatites.

LISTA DE FIGURAS

Figura 1. Distribuição de ferrocarbonatitos no mundo. Fonte: adaptado de Giovannini et al. (2020).	4
Figura 2. Inversão do relevo. Fonte: Ollier e Sheth (2008) adaptado de Pain e Ollier (1995).	10
Figura 3. Lateritas de Morro de Seis Lagos. Fonte: Giovannini (2013).	10
Figura 4. Grau de evolutivo (agração e degradação) de fácies de duricrust ferruginoso. Fonte: Campos (2023), adaptado de Tardy (1993).	11
Figura 5. Perda de Si no processo de ferratização. Fonte: Desenvolvido pela Autora.	12
Figura 6. Intensidade do intemperismo. Fonte: Adaptado de Teixeira et al. (2000).	13
Figura 7. A interdisciplinaridade da Pedometria. Fonte: Adaptado de Hengl (2003).	18
Figura 8. Diagrama mostrando o avanço das técnicas de Mapeamento Digital de Solos. Fonte: Minasny e McBratney (2016).	19
Figura 9. Técnicas pedométricas. Fonte: McBratney et al. (2000).	21
Figura 10. Etapas dos procedimentos metodológicos. Fonte: Elaborado pela autora (2023).	28
Figura 11. Localização da área de estudo. Fonte: Elaborado pela autora (2023).	29
Figura 12. Frequência de seleção das covariáveis, para a predição espacial dos elementos, a partir do modelo RFE para os modelos. Fonte: Elaborado pela autora (2023).	36
Figura 13. Mapeamento superficial de Nb ₂ O ₅ em Morro de Seis Lagos por Random Forest. Fonte: Elaborado pela autora (2023).	36
Figura 14. Mapeamento superficial de TiO ₂ em Morro de Seis Lagos por SVMRadial. Fonte: Elaborado pela autora (2023).	37
Figura 15. Mapeamento superficial de Fe ₂ O ₃ em Morro de Seis Lagos por Random Forest. Fonte: Elaborado pela autora (2023).	38
Figura 16. Mapeamento superficial de MnO em Morro de Seis Lagos por Random Forest. Fonte: Elaborado pela autora (2023).	38
Figure 17. The methodological steps included: (1) Compilation of geochemical data from CPRM [22]; (2) Extraction of remote sensing covariates (ASTER, SAR, Sentinel-2A) and topographic data (DEM); (3) Harmonization of geochemical data and covariates into a single database; (4) Selection of covariates using Spearman's correlation (threshold 0.95) and Recursive Feature Elimination (RFE), with partitioning of the database into test and training and use of (Leave-One-Out Cross-Validation); (5) Definition of sampling scenarios; (6) Processing models (RF, SVM, NNET, KNN, and GLMNET); (7) Model performance evaluation based on precision, accuracy, and sampling design; and (8) Calculation of the average spatial predictions of elements. Source: Prepared by the author (2025).	47
Figure 18. Location of the study area and the sampling point locations. Source: Prepared by the author (2025).	48
Figure 19. Different scenarios for a combination of sampling designs as training and validation sets for modeling. Source: Prepared by the author (2025).	52
Figure 20. Predictive performance of the SVMRadial, RF, KNN, GLMNET, and NNET models by sampling scenario (Ia V) for the chemical elements: A) Al ₂ O ₃ ; B) Fe ₂ O ₃ ; C) MnO; D) Nb ₂ O ₅ ; E) TiO ₂ ; F) SiO ₂ , evaluated by the RMSE, MAE, and R ² metrics. Groups with the same letters indicate statistically equivalent performance, according to the Kruskal-Wallis test followed by	

Dunn's post-test ($p \leq 0.05$). ND indicates absence of detection of the chemical element. Source: Prepared by the author (2025).	57
Figure 21. Coefficient of determination (R^2) (A) among scenarios for all models and elements combined; (B) among models for all scenarios and elements combined; (C) among elements for all models and scenarios combined. Source: Prepared by the author (2025).	60
Figure 22. RMSE for each element by scenario and model (A) and for each sampling scenario by model and element (B). Source: Prepared by the author (2025).	63
Figure 23. Distribution of the expected values for the sampling scenarios. Source: Prepared by the author (2025).	67
Figure 24. Distribution of observed and predicted values for Fe_2O_3 , Nb_2O_5 , and TiO_2 based on the KNN, GLMNET, and RF models (Appendix D), respectively, and location of sampling points underlined by two main formations (laterite crust and slope). Source: Prepared by the author (2025).	68
Figure 25. Spatial distribution of element content. Source: Prepared by the author (2025).	71
Figure 26. Visualization of soil data distribution by sampling design. Source: Prepared by the author (2025).	72
Figure 27. The methodological procedures were divided into three stages. Stage I - Preparation of the database, with extraction of terrain covariates and remote sensing indices; Stage II - Selection of covariates using Spearman's correlation and the Recursive feature elimination (RFE) method. The Random Forest and Cubist models were then processed with 20 repetitions, and the statistics of the performance metrics were calculated; Stage III - Selection of the model with the best performance based on the metrics, followed by generation of the mean prediction and amplitude maps. Source: Prepared by the author (2025).	79
Figure 28. Location of the study area in the Six Lakes Project, in São Gabriel da Cachoeira, Brazilian Amazon, Amazonas State, highlighting the surface sampling collection points ($n = 341$) and the four boreholes. Source: Prepared by the author (2025).	80
Figure 29. Spatial distribution of the IOL index - mean map (A), standard deviation map (B) and Variable Importance Ranking (C), using the Random Forest (RF) model in the Morro dos Seis Lagos area, São Gabriel da Cachoeira, Amazonas State, Brazil. Source: Prepared by the author (2025).	88
Figure 30. Spatial distribution of the CIA index - mean map (A), standard deviation map (B), and Variable Importance Ranking (C), using the Random Forest (RF) model in the Morro dos Seis Lagos area, São Gabriel da Cachoeira, Amazonas State, Brazil. Source: Prepared by the author (2025).	90
Figure 31. Spatial distribution of the CIW index - mean map (A), standard deviation map (B), and Variable Importance Ranking (C), using the Random Forest (RF) model in the Morro dos Seis Lagos area, São Gabriel da Cachoeira, Amazonas State, Brazil. Source: Prepared by the author (2025).	92
Figure 32. Geochemical bivariate maps (A) IOL versus Nb_2O_5 ; (B) CIA versus Nb_2O_5 ; and (C) CIW versus Nb_2O_5 in the Morro dos Seis Lagos area, São Gabriel da Cachoeira, Amazonas State, Brazil. Source: Prepared by the author (2025).	94
Figure 33. Geochemical bivariate maps (A) IOL versus TiO_2 ; (B) CIA versus TiO_2 ; and (C) CIW versus TiO_2 in the Morro dos Seis Lagos area, São Gabriel da Cachoeira, Amazonas State, Brazil. Source: Prepared by the author (2025).	97
Figure 34. Geochemical bivariate maps (A) IOL versus P_2O_3 ; (B) CIA versus P_2O_3 ; and (C) CIW	

versus P_2O_3 in the Morro dos Seis Lagos area, São Gabriel da Cachoeira, Amazonas State, Brazil. Source: Prepared by the author (2025).	99
Figura 35 - Distribuição estatística, densidade e intervalos de confiança para os teores de Al_2O_3 , Fe_2O_3 , MnO, Nb_2O_5 , TiO_2 , P_2O_5 e SiO_2	120

LISTA DE TABELAS

Tabela 1. Índices provenientes da combinação de bandas espectrais dos sensores MSI- Sentinel 2A e ASTER.	30
Tabela 2. Estatística descritiva de elementos.	34
Tabela 3. Performance dos modelos Random Forest (RF), Multivariate Adaptive Regression Splines (MARS) e Radial Support Vector Machine (SVMRadial).	35
Table 4. Summary statistics of elements based on geologic formation.	49
Table 6. Mean of performance parameters among sampling scenarios for all models combined.	58
Table 7. Mean of performance parameters among models for all scenarios combined.	61
Table 8. Geochemical indicators of weathering.	81
Table 9. Terrain morphometrics covariates, categorical covariates, and spectral indices from Morro dos Seis Lagos area, São Gabriel da Cachoeira, Amazonas State, Brazil.	82
Table 10. Summary statistics of weathering indexes and mean comparison between Laterite and Talus.	86
Table 11. Performance metrics for the accuracy predictions of surface weathering indices (IOL - Lateralization Index; CIA - Chemical Index of Alteration; and CIW - Chemical Index of Weathering for Random Forest and Cubist Models in the Morro dos Seis Lagos area, São Gabriel da Cachoeira, Amazonas State, Brazil.	87

LISTA DE ABREVIACÕES E SÍMBOLOS

Sigla	Descrição
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
CB	Cubist
CPRM	Companhia de Pesquisa de Recursos Minerais
CV	Cross Validation
GLMNET	Generalized Linear Models with Elastic Net Regularization
IBGE	Instituto Brasileiro de Geografia e Estatística
K-KNN	Nearest Neighbors
LOOCV	Leave One Out Cross-Validation
MAE	Mean Absolute Error
MARS	Multivariate Adaptive Regression Spline
MDE-HC	Modelo Hidrologicamente Consistido
MDS	Mapeamento Digital de Solos
ML	Machine learning
PC	Concordance Correlation Coefficient
R^2	Coeficiente de determinação
RFE	Recursive Feature Elimination
RF	Random Forest
RMSE	Root Mean Squared Error
RNA	Redes Neurais
SVMR	Radial Support Vector Machine
UFRRJ	Universidade Federal Rural do Rio de Janeiro

SUMÁRIO

1. INTRODUÇÃO GERAL.....	1
2. REVISÃO DE LITERATURA GERAL	4
2.1. Carbonatitos: Gênese, Ocorrência e Classificação	4
2.2. Lateritas: Gênese, Mineralogia, Fatores e Processos Associados.....	5
2.3. Pedosfera: Sistema Solo e Paisagem	7
2.4.1 Lateritas em ambientes tropicais.....	8
2.4.2 Processos: ferralitização (latossolização)	12
2.4.3 Plintitização (laterização)	14
2.5. Mapeamento Digital de Solos (MDS)	16
2.5.1. Avanços das técnicas de MDS e pedometria.....	17
2.5.2 Modelos preditivos e fatores de formação do solo	19
2.5.3 Algoritmos de aprendizagem de máquina.....	21
2.5.4 Dados legados	22
3. CAPÍTULO I: MAPEAMENTO DIGITAL DOS TEORES DE Fe_2O_3 , MnO , Nb_2O_5 E TiO_2 EM MORRO DOS SEIS LAGOS-AM, BRASIL.....	25
3.1 RESUMO.....	25
3.2 ABSTRACT.....	26
3.3 INTRODUÇÃO	27
3.4 MATERIAL E MÉTODOS	28
3.4.1 Covariáveis ambientais de entrada.....	29
3.5 RESULTADOS E DISCUSSÃO.....	34
3.5.1 Estatística descritiva.....	34
3.5.2 Análise da variabilidade espacial dos elementos	36
3.6 CONCLUSÕES	40
4. CAPÍTULO II: SAMPLING DESIGN AND MODEL SELECTION INFLUENCE ON PREDICTIONS OF CHEMICAL COMPOUNDS IN PETROFERRIC FORMATIONS IN THE BRAZILIAN AMAZON.....	42
4.1 RESUMO.....	42
4.2 ABSTRACT.....	43
4.3 INTRODUCTION	44
4.4 MATERIAL AND METHODS	47
4.4.1 Study area.....	47
4.4.2 Environmental covariates.....	49
4.4.3 Selection of environmental covariates	50
4.4.4 Modelling	51
4.4.5 Sampling scenarios.....	51
4.4.6 Model performance evaluation	53
4.4.7 Comparison criteria and statistical analysis	54

4.5 RESULTS	55
4.5.1 Factors influencing performance.....	55
4.5.2. Accuracy prediction comparisons among scenarios	58
4.5.3 Accuracy predictions among models	60
4.5.4 Interaction of models and sampling scenarios on accuracy predictions	62
4.6 DISCUSSION	64
4.6.1. Sampling scenarios	64
4.6.2. Spatial variability and geological context.....	67
4.6.3. Geological formation of the study area	69
4.6.4. Predictive models	69
4.6.5. Sampling designs and properties.....	70
4.7 CONCLUSIONS	73
5. CAPÍTULO III: WEATHERING PATTERN RECOGNITION IN PETROFERRIC FORMATIONS IN BRAZILIAN AMAZON.....	75
5.1 RESUMO.....	75
5.2 ABSTRACT.....	76
5.3 INTRODUCTION	77
5.4 MATERIAL AND METHODS	79
5.4.1 Study area	79
5.4.2 Laboratory analysis.....	80
5.4.3 Weathering indexes- surface.....	81
5.4.4 Environmental covariates	81
5.4.5 Spatial modeling of weathering indices.....	83
5.4.5.1 Selection of covariates	83
5.4.5.2 Models	83
5.4.6 Model performance evaluation	84
5.5 RESULTS AND DISCUSSION	86
5.5.1 Weathering indices difference between geologic/geomorphic units	86
5.5.2 Spatial predictions of surface weathering indices.....	86
5.5.3 Bivariate spatial distribution of weathering rates versus Nb ₂ O ₅ , P ₂ O ₃ , and TiO ₂	93
5.6 CONCLUSIONS	100
6. CONCLUSÕES GERAIS	101
7. CONSIDERAÇÕES FINAIS	103
8. REFERÊNCIAS BIBLIOGRÁFICAS	104
9. APÊNDICES	119

1. INTRODUÇÃO GERAL

Nas últimas décadas, modelos de aprendizagem de máquinas têm sido amplamente utilizados em análises de mapeamento digital, pedológico, geológico, fitofisionômicas, entre outros. Estudos de minerais em rochas, por exemplo, apresentam vantagens no sensoriamento remoto devido à sua origem controlada por processos geológicos e à precisão das assinaturas espectrais. O espectro de um mineral representa a assinatura espectral mais precisa dos materiais, apresentando um espectro distinto para cada mineral (SHIRMARD et al., 2022).

Destaca-se, atualmente, a combinação de detecção remota e a análise de dados associada a aprendizagem de máquina (ROSSITER et al., 2018; PADARIAN et al., 2020; WADOUX; ODEH; MCBRATNEY, 2021; REDDY; DAS BHABANI, 2023), em especial para o mapeamento geológico (unidades litológicas, tipos de alteração, estruturas e minerais indicadores) de áreas de difícil acesso, visto que uma das etapas fundamentais na exploração mineral é identificar as feições geológicas que estão relacionadas à mineralização alvo fornecendo e investigando mapas geológicos (COSTA et al., 2019; SHIRMARD et al., 2022).

Nesse contexto geopolítico global, o Brasil detém 94% das reservas mundiais de nióbio (842 Mt) e a segunda maior reserva de terras raras (21 Mt; CPRM/SGB, 2025), posicionando o país na disputa por minerais críticos da transição energética e da demanda de tecnologias. Nesse contexto, torna-se imperativo caracterizar depósitos tropicais como Morro dos Seis Lagos (AM), o exemplar de ferrocarbonatitos, onde processos supergênicos redistribuem Nb_2O_5 , TiO_2 e elementos terras raras (ETR). Nesse sentido, essa pesquisa preenche lacuna com abordagem pedogeoquímica focando em formações petroféricas resultantes de intensos processos de intemperismo e pedogênese no manto amazônico. O uso das técnicas de pedometria na predição de teores de elementos associados a índices geoquímicos posiciona essa pesquisa na fronteira disciplinar entre Ciência do Solo e Geoquímica, ressaltando o solo como corpo natural hospedeiro de recursos naturais estratégicos.

Uma parte muito importante na busca por minerais de interesse econômico é identificar as características do solo e das rochas que estão ligadas à mineralização que queremos encontrar. Para isso, usamos e estudamos mapas geológicos, como explicam Costa et al. (2019) e Shirmard et al. (2022). Sendo assim, essas técnicas podem apoiar a prospecção mineral de elementos com potencial econômico, para atender à demanda crescente para fins industriais e energéticos. E, atualmente, a disponibilidade de múltiplos dados geográficos tem sido de grande valia para o mapeamento e prospecção mineral, o que contribui para previsões de alta qualidade e baixo custo. Estudos recentes como de Cracknell e Reading (2014), Costa, Samuel-Rosa e Anjos (2018), Pinheiro et al. (2021) e Wang et al. (2020) utilizaram técnicas de aprendizagem de máquina na análise geológica.

O Morro dos Seis Lagos, localidade situada na parte noroeste do bioma da Amazônia brasileira, se destaca pela condição rara entre os dezenove principais exemplos de depósitos de nióbio - Nb associado a carbonatitos, sendo considerado o maior depósito de Nb do mundo (CHAKHMOURADIAN et al., 2015). O siderita carbonatito de Morro dos Seis Lagos está relacionado ao estágio magmático tardio ao hidrotérmico, e a área contém uma crosta espessa de laterita (>200 m), onde os processos de carbonatação da siderita produziram uma crosta de

goethita/hematita. À medida que ocorre o intemperismo, o pirocloro primário sofre alterações físico-químicas, resultando na formação do pirocloro secundário. Nesse processo ocorre a lixiviação de cátions (Ca^{2+} e Na^+) e o enriquecimento residual de Nb e Fe, sendo o pirocloro secundário de cor variando de marrom escura a preta, e ocorrendo em horizontes lateríticos (PORTO et al., 2025). A área possui elevação de, aproximadamente, 400 metros de altitude, posicionada em um terreno plano com uma elevação média de aproximadamente 90 metros de altitude. Os seis lagos ocorrem em duricrusts de Fe em zonas de tálus, em posições de altitudes intermediárias, sendo a forma circular do corpo do Morro dos Seis Lagos delineada pela base do depósito de tálus (PORTO et al., 2025). Ressalta-se que os carbonatitos e lateritas hospedam metais traços, elementos que têm potencial econômico. O complexo é o único exemplo de depósito de Nb no qual o rutilo, rico em Nb, é o principal mineral de minério (GIOVANNINI et al., 2017).

Diante dos aspectos físicos e químicos da área de estudo, tem-se a hipótese de que o processo de intenso intemperismo em ambiente tropical ao longo do tempo geológico atua na morfologia e gênese das lateritas associadas aos ferrocarbonatitos. Desta forma, esses materiais delateritas são testemunhos dos processos pedogenéticos sobre a rocha matriz rica em ferro, condicionando o enriquecimento em elementos no solo. Nesse sentido, acredita-se que há influência em diferentes teores de elementos associados às lateritas que afloram na área, diante da heterogeneidade dos fatores e processos de formação atuantes, permitindo o mapeamento da variabilidade espacial dos elementos e materiais lateríticos a partir de técnicas de aprendizagem de máquina aliadas a dados analíticos e características da paisagem.

Nesse sentido, o objetivo da pesquisa consiste no mapeamento de elementos/compostos químicos a partir da associação de dados analíticos e variáveis ambientais derivadas de sensoriamento remoto e modelos numéricos de superfície do terreno, através de técnicas de aprendizagem de máquina (*machine-learning*), bem como na elucidação do processo de laterização associado à siderita carbonatito de Morro dos Seis Lagos (AM).

De forma a atender ao objetivo geral proposto, os objetivos específicos englobam: a. Revisão temática sobre a área de estudo; a gênese e composição de siderita carbonatitos, e fatores e processos associados à laterização; b. Aquisição e compilação de banco de dados, organização/compatibilização e tratamento de dados; c. Preparação e análise descritiva dos dados de entrada (*input data*); d. Seleção de covariáveis ambientais para predição dos elementos/compostos de interesse. Aplicação e análise de algoritmos de *machine-learning* para mapeamento de elementos/compostos; e f. Interpretação dos resultados e discussão a respeito dos processos associados à laterização e ao enriquecimento de Nb_2O_5 e outros elementos/compostos químicos em Morro dos Seis Lagos.

Assim, com o intuito de comprovar as hipóteses, a tese é dividida em três capítulos, descritos a seguir:

O Capítulo I apresenta a predição da variabilidade espacial de elementos no Morro Seis Lagos, Amazônia. Este capítulo objetiva a predição da distribuição espacial de elementos como nióbio (Nb_2O_5), titânio (TiO_2), manganês (MnO) e ferro (Fe_2O_3), tal como a análise comparativa dos modelos *Multivariate Adaptive Regression Spline* (MARS), *Radial Support Vector Machine* (SVMRadial) e *Random Forest* (RF).

O Capítulo II explora a modelagem a partir de diferentes desenhos amostrais na predição dos elementos em formações petroféricas na Amazônia brasileira, com objetivo de investigar

a influência de várias combinações de amostragem (cenários) de cinco planos de amostragem nas previsões espaciais de compostos químicos (Al_2O_3 , Fe_2O_3 , MnO , Nb_2O_5 , TiO_2 e SiO_2), utilizando múltiplos algoritmos de aprendizagem de máquina, tais como: *Generalized Linear Models with Elastic Net Regularization* (GLMNET), *Nearest Neighbors* (KNN), *Neural Network* (NNET), *Random Forest* (RF) e *Support Vector Machine* (SVMRadial) combinados com imagens de sensoriamento remoto dos sensores ASTER, Sentinel- 1A e Sentinel- 1B (SAR).

O Capítulo III aborda a modelagem de índices geoquímicos de intemperismo como *Index of Lateritisation* (IOL), *Chemical Index of Alteration* (CIA) e *Chemical Index of Weathering* (CIW), em formações petroféricas na Amazônia brasileira. Assim, objetiva-se quantificar a intensidade do intemperismo (laterização) em diferentes unidades geomórficas na camada superficial.

Os três capítulos constituem uma abordagem metodológica de modelagem tridimensional (X, Y e Z), na qual o Capítulo I buscou entender o desempenho de melhores modelos de aprendizado de máquina (*Multivariate Adaptive Regression Spline* (MARS), *Radial Support Vector Machine* (SVMRadial) e *Random Forest* (RF)) na predição de elementos químicos (Nióbio (Nb_2O_5), Titânio (TiO_2), Manganês (MnO) e Ferro (Fe_2O_3)), visando também compreender seus padrões espaciais. O Capítulo II avaliou o desempenho de cinco desenhos amostrais na predição dos elementos (Al_2O_3 , Fe_2O_3 , MnO , Nb_2O_5 , TiO_2 e SiO_2), utilizando os algoritmos *Generalized Linear Models with Elastic Net Regularization* (GLMNET), *K- Nearest Neighbors* (KNN), *Neural Network* (NNET), *Random Forest* (RF) e *Support Vector Machine with Radial Basis Function Kernel* (SVMRadial), com o objetivo de compreender a melhor estratégia de delineamento amostral que auxilie na representação dos padrões espaciais dos elementos no contexto a paisagem. O Capítulo III integra os resultados dos capítulos anteriores por meio da aplicação da modelagem de índices geoquímicos de intemperismo, como o *Index of Lateritisation* (IOL), *Chemical Index of Alteration* (CIA) e *Chemical Index of Weathering* (CIW), com o objetivo de quantificar e compreender a intensidade do intemperismo (laterização), bem como a gênese do material local em diferentes unidades geomórficas da camada superficial.

2. REVISÃO DE LITERATURA GERAL

2.1. Carbonatitos: Gênese, Ocorrência e Classificação

De acordo com a classificação do sistema da *International Union of Geological Sciences* (IUGS), os carbonatitos são rochas ígneas contendo em sua composição mais de 50% de minerais carbonáticos primários (mágmatícos) e que possuem menos de 20% de SiO_2 (LE MAITRE et al., 2005). Conforme a classificação baseada em concentrações modais dos minerais, carbonatitos podem ser divididos em: i) Calcita carbonatito (rocha com granulação grossa denominada sovito, se possuir granulação fina é alvikito); ii) Dolomita carbonatito (beforsito); iii) Ferrocronatito (essencialmente composto de carbonatos de ferro); e iv) Natrocarbonatito (LE MAITRE et al., 2005).

Os carbonatitos são considerados raros, sua distribuição na crosta (Figura 1) possui menos de 1% de área de superfície emersa (GIOVANNINI et al., 2017; TAKEHARA, 2019). Segundo Gomes et al. (2019), os constituintes dos carbonatitos se distinguem por conter minerais pouco comuns, acessórios raros, primários ou secundários, de composição complexa trazendo estrôncio (Sr), nióbio (Nb), zircônio (Zr), elementos terras-raras (ETRs), entre outros, cuja presença é resultado das condições especiais de formação dessas rochas. Nesse sentido, são consideradas rochas raras e depósitos de elementos com potencial estratégico na economia.

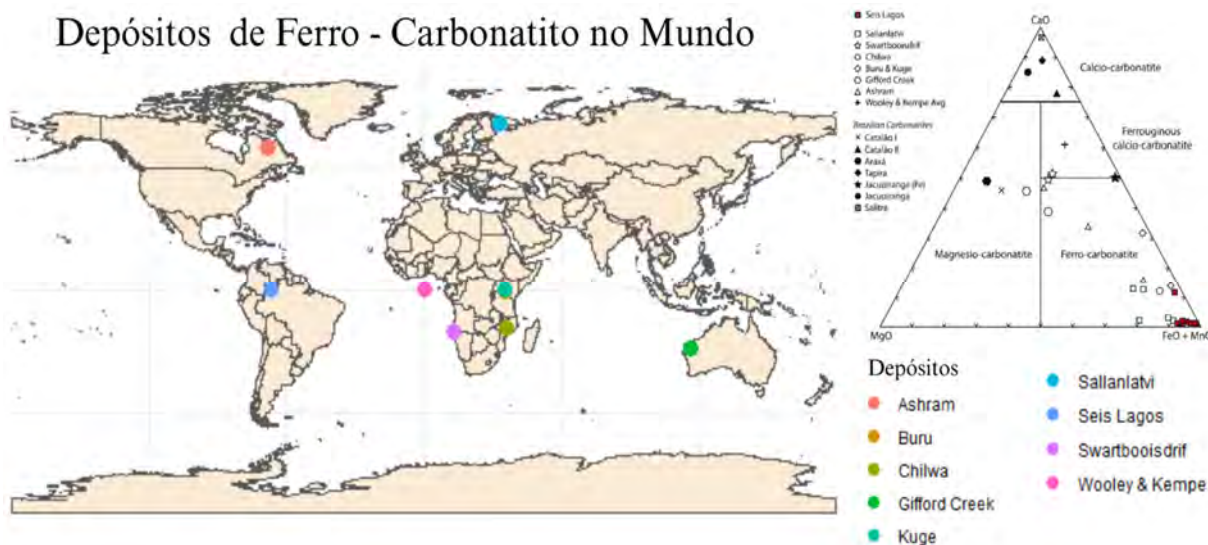


Figura 1. Distribuição de ferrocronatitos no mundo. Fonte: adaptado de Giovannini et al. (2020).

Esse material é produto de manifestações magmáticas intrusivas e extrusivas, associado a processos endógenos desde o Pré-cambiano até atualmente, sobretudo em ambiente tectônico extensivo e intraplacas (GIOVANNINI et al., 2017). Conforme os mesmos autores, esse material é formado a partir da fusão do manto, cristalização fracionada e imiscibilidade de líquidos magmáticos ricos no íon carbonato.

Segundo Giovannini et al. (2020), o Complexo Carbonatítico do Morro dos Seis Lagos representa a parte superior de um sistema magmático carbonatítico e o carbonatito siderítico está relacionado a processos magmáticos tardios a posteriores. A composição do siderita carbonatito

na região central é mais rica em Fe_2O_3 , ETRs e P que a borda da intrusão (GIOVANNINI et al., 2020).

O fator de enriquecimento de Nb no carbonatito está correlacionado ao processo de substituição de $3\text{Ti}^{4+} = \text{Fe}^{2+} + 2\text{Nb}^{5+}$ em brookita rica em Nb, indicando um ambiente redutor (GIOVANNINI et al., 2020). Segundo estudos petrológicos de Giovannini et al. (2020), o rutilo é um mineral secundário proveniente do intemperismo da brookita rica em Nb, sendo o Morro dos Seis Lagos um exemplar para a formação de rutilo rico em Nb.

Em Giovannini et al. (2020), em Seis lagos são encontrados três tipos de siderita carbonatitos: (1) porção central, composto por siderita (até 95% em volume), hematita, pirocloro, Nb-brookita, Ti-maghemita e thorbastnasita; (2) porção central rica em ETRs e P, com siderita (até 95% em volume), hematita, pirocloro, monazita e bastnasita; e (3) porção da borda com siderita (~70% vol.), barita, gorceixita, rabdofano e pirocloro.

Acima dos siderita carbonatito do Morro dos Seis Lagos existe uma crosta ferruginosa (canga laterítica) com mais de 250 metros de espessura, cuja elevação se destaca no relevo (TAKEHARA, 2019). A área tem potencial mineral significativo, além de ser importante na prospecção geoquímica, no mapeamento geológico e pedológico e no estudo da evolução geológica e geomorfológica da região (COSTA, 1991).

2.2. Lateritas: Gênese, Mineralogia, Fatores e Processos Associados

A denominação “laterita” tem sido empregada para designar depósitos residuais endurecidos oriundos do intemperismo de rochas e materiais superficiais, situados em posições variadas do relevo (ESPINDOLA; DANIEL, 2008). Conforme Ridley (2013), o conceito de laterita se refere ao subsolo com abundância em óxido de ferro e alumina, de cor vermelha, que é oxidante e lixiviado de bases em um intemperismo químico típico de zonas climáticas tropicais. O processo de formação da crosta ferruginosa das lateritas está associado a tempos geológicos e condições ambientais pretéritas. Nesse sentido, esse material é testemunho de eventos paleoclimáticos, com formação de paleosolos, dada sua estabilidade garantida na paisagem pela ação protetora das lateritas (ESPINDOLA; DANIEL, 2008), bem como pelo seu posicionamento no interior do Cráton Amazônico.

A laterização ocorre a partir da alteração dos minerais primários (micas, anfibólios, piroxênios, feldspatos, carbonatos) que são hidrolisados. Íons de ferro, alumínio, magnésio, sódio e cálcio são lixiviados (elementos móveis) para a solução e para as águas subterrâneas. Ferro e alumínio têm migração lateral e acúmulo em locais preferenciais (BOURMAN, 1993). Desse modo, a gênese das lateritas propicia concentração de elementos de forma heterogênea e irregular. Assim, os perfis lateríticos são tipicamente camadas/horizontes com diferentes componentes minerais e texturas, que refletem o aumento da intensidade de oxidação, hidratação e lixiviação de matérias solúveis em direção à superfície (EL-DESOKY *et al.*, 2020).

De acordo com a Pedologia, ramo que estuda a origem, a estrutura e a classificação das lateritas (SANTOS et al., 2018), esse material está associado às classes e propriedades diagnósticas da ordem dos Latossolos (solos profundos e bem desenvolvidos) e dos Plintossolos (formação de material concrecionário/ferruginoso). Nos Latossolos observamos o predomínio do processo pedogenético de latossolização (Ferratilização e Dessilicação) e os processos múltiplos de perdas e transformação de materiais no solo. A partir da perda de bases, ou seja,

liberação e remoção parcial a total dos cátions, sílica e/ou carbonato no solo (KER et al., 2012), obtém-se o aumento relativo das concentrações de óxidos de ferro e alumínio.

No processo de laterização e plintitização, predominam os processos múltiplos de transformação e translocação no solo. Esse processo resulta no acúmulo de plintita em áreas com variação de umidade sazonal correlacionada com a presença de baixos teores de colóides orgânicos. O processo pedogenético envolve a segregação de ferro (redução e oxidação); em ambiente redutor, ocorre a solubilização do Fe^{3+} , e, por conseguinte, atranslocação de Fe na forma reduzida (maior mobilidade), logo após a sua precipitação por oxidação, produzindo plintita e concreções localizadas de óxidos de ferro (cor vermelha ou brunada), em profundidade no perfil (KER et al., 2012).

A plintita é constituída por um conjunto de minerais de argila ricos em ferro e é pobre em carbono (OLIVEIRA, 2011). Ainda segundo o autor, esse material, quando submetido à alternância de ciclos de umidificação e secagem, está sujeito a eventual endurecimento irreversível (petroplintita) ou canga laterítica (horizonte concrecionário). Conforme Anjos et al. (1995), no Brasil ocorrem grandes extensões de solos com plintita e concreções ferruginosas. As camadas cimentadas são conhecidas como lateritas ou ferricretos (KER et al., 2012).

Os Plintossolos são solos constituídos por material mineral, apresentando horizonte plíntico, litoplíntico ou concrecionário em uma das seguintes condições: iniciando dentro de 40 cm da superfície; ou iniciando dentro de 200 cm da superfície quando precedidos de horizonte glei ou imediatamente abaixo do horizonte A, E ou horizonte que apresente cores pálidas, variegadas ou com mosqueados em quantidade abundante (SANTOS et al., 2018). Ainda segundo o Sistema Brasileiro de Classificação de Solos (SiBCS), esses solos podem ser classificados quanto ao segundo categórico (subordens) em: Plintossolos Pétricos quando possuir horizonte concrecionário ou horizonte litoplíntico; Plintossolos Argilúvicos- quando atender ao critério de horizonte plíntico e caráter argilúvico; e Plintossolos Háplicos- quando não se enquadrar nas classes de solos anteriores.

Na Amazônia, a crosta ferruginosa, por se encontrar próxima à superfície, está atualmente em processo de intemperismo e assim recoberta também por Latossolos areno-argilosos autóctones e alóctones (COSTA, 1991). A ocorrência de lateritas na área é observada em platôs ou morros, cujos topos representam superfícies lateríticas reliquias. Os platôs isolados atingem até 40 km² de área, formando-se sobre lateritas bauxíticas (COSTA, 1991).

A gênese das lateritas está associada à formação do pirocloro secundário, junto com cériopirocloro-(Ce), goethita rica em Nb e cerianita-(Ce). Com o aumento da profundidade, o volume desses minerais de 0,12% em peso, no tipo pisolítico, cresce para 3,5% em peso no tipo de laterita classificada como “laterita marrom”, segundo Giovannini et al. (2017). Os mesmos autores elucidam que as lateritas do Morro dos Seis Lagos (AM) estão associadas ao processo de intemperismo do siderita carbonatito.

O perfil laterítico da região possui elevado teor de ferro, sendo classificado em seis tipos pela textura e composição mineralógica, sendo (do topo para a base): laterita pisolítica, laterita fragmentada, laterita mosqueada, laterita roxa, laterita manganífera e laterita marrom (GIOVANNINI et al., 2017). Assim, na laterita marrom e roxa, observa-se a maior predominância de hematita. Já na laterita fragmentada tem-se o enriquecimento da goethita em relação à hematita, elucidando que essas foram hidratadas. A laterita manganífera tem maior predominância do óxido de Mn (hollandita e pirolusita) e minerais como rutilo rico em Nb,

brookita rica em Nb, cério-florencita-(Ce) e ankerita.

Segundo Giovannini et al. (2017), as lateritas apresentam, na sua parte superior, uma média de 2,91% de Nb₂O₅ e 5% de TiO₂ em peso. É importante notar que o rutilo, que contém entre 11,26% e 22,23% de Nb₂O₅ em peso, está presente em todas as lateritas. Ainda segundo os mesmos autores, há predominância de óxi-hidróxidos de Fe nas lateritas, no entanto, ocorrem altas concentrações de goethita nas lateritas inferiores e em lateritas pisolíticas e fragmentadas. Quanto à hematita, possivelmente produto do processo de desidratação da goethita, é predominante na laterita roxa superior.

Conforme Giovannini et al. (2020), as lateritas estão submetidas ao forte intemperismo tropical que, por sua vez, próximo à superfície, contribui para a hidrodinâmica onde o Fe é frequentemente mobilizado e precipitado como goethita, na presença de água no sistema, sendo capaz de incorporar maiores quantidades de Al em comparação com a hematita.

2.3. Pedosfera: Sistema Solo e Paisagem

A importância do solo como corpo natural deriva, em grande parte, do seu papel de interface entre o mundo das rochas (a litosfera), o ar (a atmosfera), a água (a hidrosfera) e os seres vivos (a biosfera). Os ambientes onde todos estes quatro mundos interagem são frequentemente os mais complexos e produtivos da Terra. O solo, ou pedosfera, é um exemplo deste tipo de ambiente (BRADY; WEIL, 2013).

Sabemos que o solo é produto das interações entre fatores e processos de formação, constituindo um sistema dinâmico que interage com o ambiente externo. Segundo Dijkerman (1974), o solo é um sistema aberto porque perde e recebe material e energia.

Nesse sentido, destacam-se as características que atribuem o solo como um sistema aberto, tais como: intercâmbio de materiais (recebimento de materiais externos); interações biológicas; interações hidrológicas e com a atmosfera. Strahler e Strahler (1973), destacam as seguintes caracterizações que envolvem a estrutura de um sistema aberto, tais como:

- i) Os sistemas possuem entradas e saídas de energia e matéria que atravessam esses limites;
- ii) Sistemas possuem caminhos de transporte de energia e transformação associados à matéria dentro deles;
- iii) A matéria pode ser transportada dentro dos sistemas ou ser submetida a transformações físicas ou químicas;
- iv) Sistemas abertos tendem a buscar um equilíbrio dinâmico, com taxas de entrada e saída de energia e matéria iguais, enquanto o armazenamento permanece constante;
- v) Quando as taxas de entrada ou saída de um sistema aberto mudam, o sistema busca um novo equilíbrio; e
- vi) O armazenamento de energia e matéria aumenta ou diminui de acordo com as variações na taxa de energia e fluxo de matéria através do sistema.

Destaca-se que esse sistema possui um papel importante nos aspectos de serviços ecossistêmicos, mudanças climáticas, segurança alimentar e saúde do solo. Não obstante, os processos de pedogênese ocorrem em um espaço de tempo longo, e o recurso está sendo degradado vastamente nas últimas décadas decorrente do manuseio inadequado.

2.4 Processos Pedogenéticos Específicos de Ferralitização Versus Plintitização

Os processos pedogenéticos específicos de Ferralitização e Plintitização ocorrem em ambientes distintos e antecedem o processo de dessilicação no solo, e, por conseguinte, a concentração e/ou redistribuição de óxidos de Fe e Al (KÄMPF; CURI, 2012).

As semelhanças entre os processos de Ferralitização e Plintitização residem na dinâmica do Fe (transformação), onde ocorre oxidação e perdas com a remoção dos produtos mais solúveis para o processo de Ferralitização, e redução e oxidação do Fe que é transferido de posição dentro do perfil (translocação). Quando o sistema é exposto ao ar, o Fe no seu estado reduzido (Fe^{2+}) desloca-se das regiões de maior concentração para as de concentração mais baixa, precipitando-se no processo de desenvolvimento de plintitas (Plintitização) (KÄMPF; CURI, 2012). Ressalta-se que esses processos pedogenéticos imprimem na morfologia do solo camadas ou estruturas distintas.

As diferenças significativas estão nos processos pedogenéticos múltiplos (perdas e translocação) condicionados pelo relevo (condições de drenagem) e conhecimento das fontes de Fe. A ferralitização tem ocorrência em maior variedade de ambientes associada a solos bem drenados, com horizontes enriquecidos de minerais de Fe (oxidação e precipitação de ferro (Fe^{2+} a Fe^{3+})), com cores vermelha e amarela. Quanto à Plintitização é um processo físico e químico que envolve condições específicas de alternância nos ciclos de umedecimento e secagem. Nesse sentido, tende a ocorrer em climas mais secos com uma longa estação seca e endurecimento irreversível da crosta ferruginosa (petroplintita) ou lençol freático mais elevado (plintita) (SCHAEFER et al., 2023).

Os principais fatores de formação incluem o clima e o relevo. O relevo proporciona boa drenagem possibilitando a remoção de sílica e cátions básicos (produtos mais solúveis), e, por conseguinte, a formação de óxi- hidróxidos de ferro (hematita e goethita).

2.4.1 Lateritas em ambientes tropicais

O termo laterita pode ser definido como acumulação relativa (autóctone) e absoluta (alóctone) de Fe e Al (D' HOORE, 1964), causando novas discordâncias na comunidade científica. As lateritas atualmente abrangem os termos de ferricretes, *duricrusts* ferruginosos e aluminosos, crostas ferruginosas, bauxitas, Plintossolos, couraças, carapaças, horizontes mosqueados, e até solos ferruginosos tropicais (TARDY et al., 1994). De acordo com IUSS *Working Group WRB* (2014) e *Soil Survey Staff*, algumas terminologias para materiais resultantes da segregação de Fe são utilizadas: *Laterite*, *Plinthite*, *Petroplinthite*, *Pisolith*, *Ferricrete* e *Duricrusts*.

A gênese das lateritas ainda não está bem definida, embora existam várias teorias que formulem sua evolução. Frisa-se que, a associação da formação de lateritas a climas tropicais quentes, caracterizados por ciclos de estações secas e úmidas, em regiões tectonicamente estáveis (NAHON; TARDY, 1992). Eggleton e Taylor (1998), consideram três critérios associados à gênese de perfis lateríticos, sendo: clima tropical sazonal, migrações laterais e fundos de vales.

Os principais processos pedológicos específicos associados à sua formação são: dessilicação, ferralitização, alitização, laterização, bauxitização e plintitização (KÄMPF; CURI, 2012). O processo de laterização, é uma junção entre o processo de dessilicação

(lixiviação intensa de sílica), remoção total de elementos álcalis e alcalinos-terrosos, e concentração de Fe (Ferratilização), que culmina na precipitação de hematita, goethita e caulinita na forma de mosqueados, nódulos e revestimentos, e posteriormente no endurecimento irreversível do material associado às oscilações de umidade do meio (KÄMPF; CURI, 2012).

Paisagens lateríticas são caracterizadas por extensos platôs parcialmente preservados, algumas vezes em diferentes altitudes formando escalonamento do relevo, e circundados por planícies e sedimentos detríticos classificados como pedimentos (KING, 1976). Esse material é associado a superfícies cratônicas antigas do cretáceo ao Cenozóico, localizados em regiões de clima tropical úmido como, por exemplo, América do Sul, África, Austrália e Índia (TARDY, 1992).

Para a formação do perfil de laterita, é necessário um ambiente de clima tropical com estações bem definidas e regimes cíclicos. Tardy e Roquin (1998) destacam as características: precipitação pluviométrica média de 1450 mm ano⁻¹; temperatura média anual de 28 °C; umidade relativa do ar média 70%; e estação seca de 5 meses; ano⁻¹.

De acordo com Campos (2023), considera-se dois modelos (ou hipóteses) básicos para explicar a formação da laterita (ou couraça ferrugionosa, ou ferricrete), sendo ambos relacionados com a evolução da paisagem: (i) Acumulação relativa de Fe decorrente da dessilicação parcial (caulinitização) e remoção de outros constituintes mais solúveis; e (ii) Acumulação absoluta por meio do aporte de Fe²⁺ em solução, sua oxidação, precipitação e acumulação.

O primeiro modelo leva em consideração a formação de laterita *in situ* com a incorporação do Fe liberado no sistema via a intemperização do material de origem e da sucessiva evolução em sequência vertical da base para o topo com: horizontes caulíníticos com mosqueados e nódulos- carapaça -couraça (TARDY et al., 1994). Os horizontes anteriores são progressivamente integrados na couraça à medida que a superfície da paisagem e a frente de intemperização sofrem um rebaixamento simultâneo, causado pela perda dos minerais dissolvidos (CAMPOS, 2023).

Nesse sentido, a cimentação da laterita, que precede à saturação da caulinita remanescente por óxidos de ferro, constitui uma etapa importante na formação da couraça ferrugionosa. Esta, por sua vez, pode evoluir por dessilicação (caulinita-gibsitita), dando origem à bauxita. As bauxitas são reconhecidas como produtos secundários predominantemente ricos em óxidos e hidróxidos de alumínio e têm sua origem relacionada à alteração hidrolítica de diferentes tipos de rochas (OLIVEIRA, 2011).

O segundo modelo (Figura 2) parte do pressuposto de que o ferro liberado pelo intemperismo em áreas mais elevadas da paisagem seja transportado na forma reduzida (íons Fe²⁺), através de fluxos laterais subsuperficiais para áreas mais baixas do relevo, onde ocorre a oxidação e subsequente precipitação na forma de óxidos de ferro (TARDY et al., 1994).

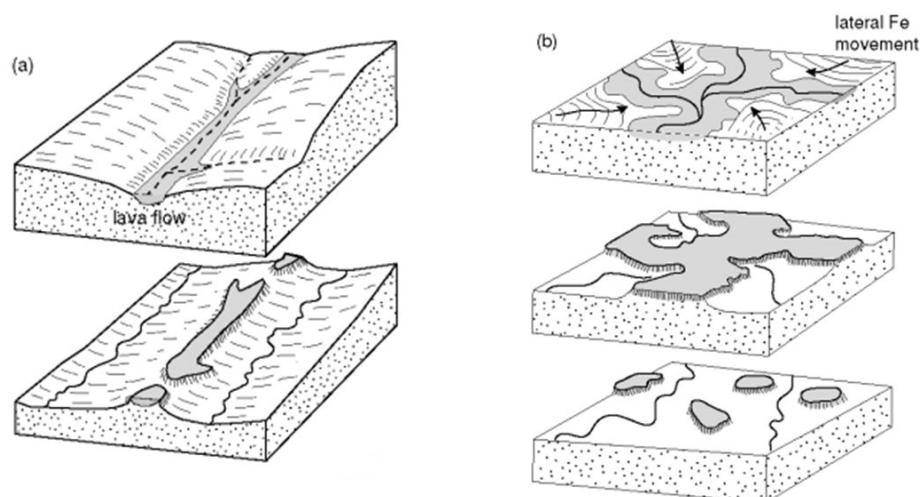


Figura 2. Inversão do relevo. Fonte: Ollier e Sheth (2008) adaptado de Pain e Ollier (1995).

De acordo com Pain e Ollier (1995), na seção (a) um fluxo de lava preenche um vale; os cursos de água originais estão representados a tracejado. Novos cursos de águas laterais formam-se em cada lado do fluxo de lava, e o alargamento do vale por eles provocam uma mesa de lava; e na seção (b) O fundo e as encostas inferiores de um sistema de vales fluviais estão parcialmente cobertos por aluviões e coluviões que são cimentados para formar a laterita. A erosão posterior ataca rocha intemperizada vizinha e a laterita resistente forma agora forma as mesas. No cenário de geocronologia, pesquisas como a de Vasconcelos e Carmo (2018), utilizam a datação de minerais supérgeos precipitados nos perfis de intemperismo.

De acordo com Retallack (2010), os perfis lateríticos podem ser bem espessos, atingindo algumas dezenas de metros de profundidade, sendo que, quando se desenvolvem sobre formações ferríferas bandadas (BIFs) ou sobre rochas ultramáficas podem se tornar economicamente viáveis para exploração de ferro e níquel, respectivamente.

Em Morro de Seis Lagos, Giovannini et al. (2020), destaca que o intenso intemperismo de ferrocarnatitos, pode originar lateritas (Figura 3) com Ti, Nb e outros metais. Frisa-se que minerais de Nb (pirocloro) do carbonatito podem concentrar residualmente, ou por decomposição parcial ou completa, gerando rutilo com Nb e brookita com Nb.

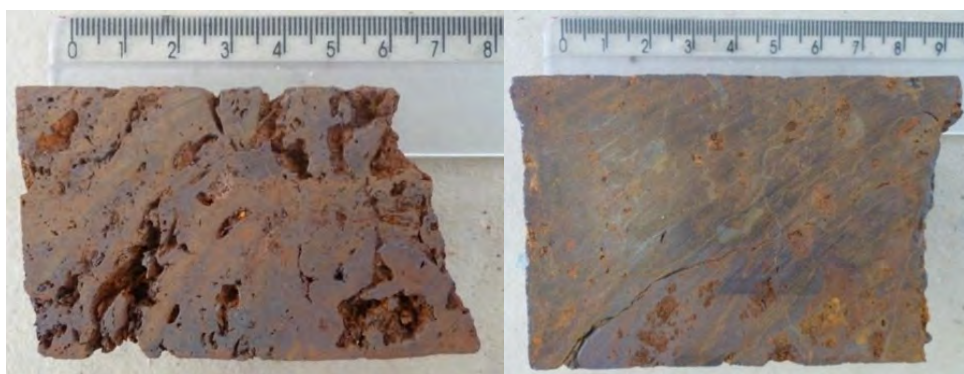


Figura 3. Lateritas de Morro de Seis Lagos. Fonte: Giovannini (2013).

As fácies podem ser caracterizadas segundo Tardy et al. (1994), pelo grau de agradação/degradação (Figura 4). No que tange à agradação temos os estágios: 1) fácies nodular- é denominada como o início da formação de um *duricrust*, caracterizada por um fundo matricial pouco ferruginoso e pela formação de nódulos hematíticos; 2) fácies meta nodular- com o aumento do ferro no sistema, os nódulos aumentam de tamanho e começam a fundir entre si; e 3) fácies maciça- os nódulos já em estágio avançado de mistura, fundem com o fundo matricial, resultando no grau evolutivo de *duricrust* ferruginoso. Sob o aspecto de degradação, temos os estágios: 4 e 5) fácies protopisolítica e pisolítica.

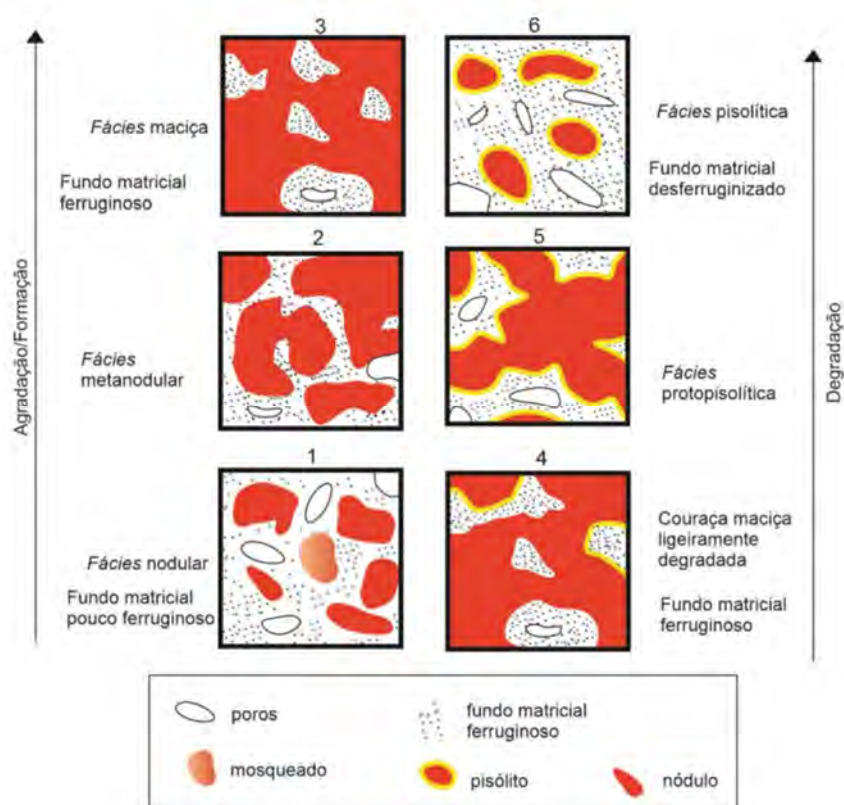


Figura 4. Grau de evolutivo (agração e degradação) de fácies de duricrust ferruginoso. Fonte: Campos (2023), adaptado de Tardy (1993).

O aumento na umidade proporciona a degradação com desferruginização do fundo matricial e formação de córtex de goethita, gerando fácies protopisolítica e pisolítica (Campos, 2023), onde se ressalta a característica complexa da morfologia em função da sobreposição de eventos ambientais que influenciam a formação.

A ocorrência da distribuição geográfica das lateritas está mais associada às paleoclimas do que ao clima atual, o que, em parte, justifica a falta de consenso quanto à origem e à definição dessas formações, de acordo com diversos autores.

O desenvolvimento das lateritas está associado a ambientes geológicos estáveis, conforme a teoria da inversão do relevo. Porém, trabalhos recentes apontam para atividades de neotectônica posterior à formação de lateritas, o que modifica a sua distribuição espacial na paisagem.

A formação da plintita como a da laterita, são processos de acumulação de óxido de Fe

e de Al. A plintita pode ter se formado recentemente ou até mesmo estar se formando atualmente, enquanto a laterita exige períodos muito mais longos de tempo, variando de milhões a dezenas de milhões de anos para se desenvolver (Campos, 2023).

Constatou-se que os processos pedogenéticos específicos de Ferratização e Plintitização ocorrem em ambientes distintos, e compartilham os fatores de formação clima e relevo, os quais influenciam na dinâmica de óxidos de Fe e Al, resultando em registros no solo dessemelhantes quanto aos aspectos físicos e químicos.

2.4.2 Processos: ferralitização (latossolização)

O processo de ferralitização, também conhecido por latossolização, é caracterizado pela perda intensa de bases e sílica coloidal (dessilicação), com aumento relativo das concentrações dos óxidos de ferro e alumínio. A dessilicação (liberação e lixiviação de Si) (Figura 5) é caracterizada pela remoção de silício, iniciando-se com a intemperização dos minerais primários no material de origem, tendo continuidade como processo pedogenético na evolução do solo atuando sobre minerais primários e secundários, e ocorrendo associado a vários outros processos pedogenéticos específicos, tais como a fersialitização, ferratização e outros (KÄMPF; CURI, 2012).

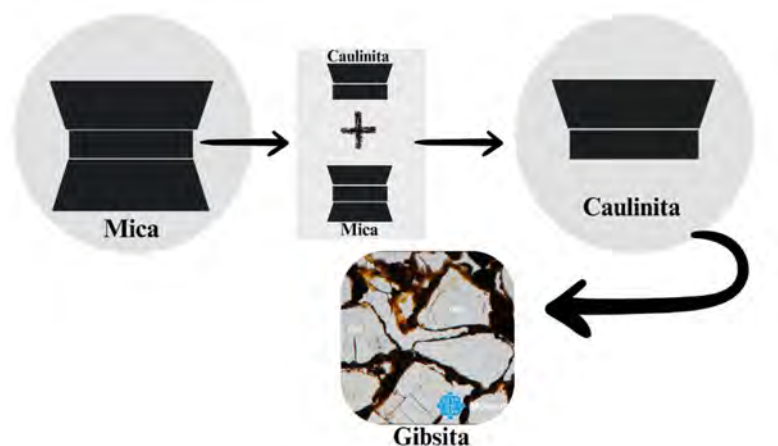


Figura 5. Perda de Si no processo de ferratização. Fonte: Desenvolvido pela Autora.

Frise-se que a dessilicação apresenta vários estágios e/ou graus de remoção do Si, assim refletindo na composição mineralógica dos solos. Na fase inicial do processo ocorre a dessilicação parcial (Fersialitização), onde os argilominerais 2:1 (esmectitas e vermiculitas) apresentam intercalação com alumínio, resultando em esmectitas com hidróxi- entre camadas (EHes) e vermiculitas com hidróxi- entre camadas (VHEs), ou podem desaparecer por completo (KÄMPF; CURI, 2012).

Em estágio posterior ocorre a dessilicação em estágio avançado (Ferralitização), onde os aluminossilicatos primários e os argilominerais 2:1 são transformados em argilominerais 1:1, tais como a caulinita (KÄMPF; CURI, 2012). Ressalta-se que, em processo de lixiviação extrema (Ferralitização), ocorre a dessilicação intensa, levando à formação de gibbsite.

Em regiões tropicais úmidas com altos índices pluviométricos (Figura 6), temperatura elevada e drenagem livre, a lixiviação remove cátions básicos (Ca, Mg, K, Na) e parte ou todo o silício mineral (dessilicação), em função dos óxidos de Fe e Al em condições de pH >5 ter

menor solubilidade, diferente da sílica que é removida do sistema, resultando no acúmulo de Al e Fe. Assim, o Fe^{2+} é oxidado e precipita como óxidos (goethita, hematita). O Al é liberado e se associa com o silício remanescente (monossilicização: formação de argilominerais 1:1), formando caulinita ou, em ocorrência de dessilicação intensa (elitização), o Al precipita como gibbsita ($\text{Al}(\text{OH})_3$). Esses processos em conjunto promovem a dessilicação e acumulação residual de óxidos de Fe e Al, além de caulinita, são conhecidos como ferratização (KÄMPF; CURI, 2012).

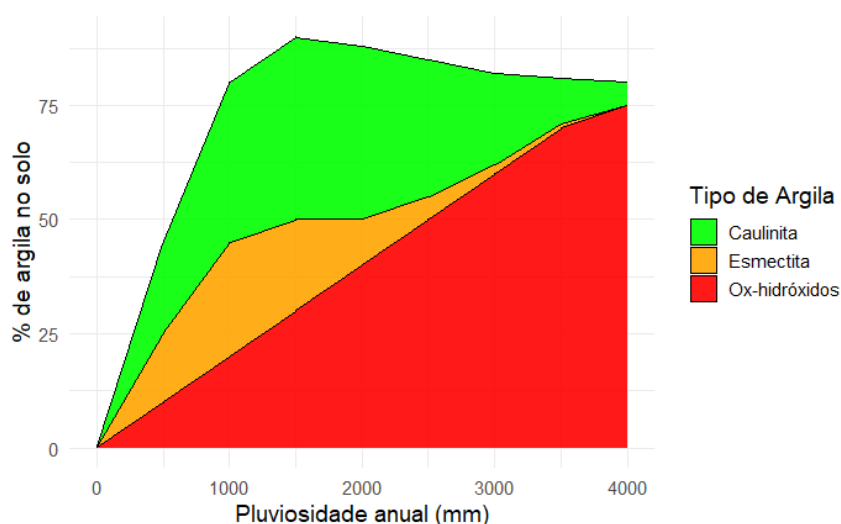


Figura 6. Intensidade do intemperismo. Fonte: Adaptado de Teixeira et al. (2000).

Durante o processo de Latossolização, quando a litologia principal é rica em ferro, os elementos solúveis são removidos e os elementos restantes precipitam em forma de óxidos e oxihidróxidos de ferro, perto da superfície (PEREIRA, 2016). Em estágio superior de evolução os depósitos (nódulos, concreções ou pisólitos) evoluem para uma cimentação, caracterizando-se como "canga laterítica".

O processo de Ferralitização está expresso na classe de Latossolos (SANTOS et al., 2018), que possuem uma extensão superior a 30% do território nacional, constituindo, assim, a categoria de maior relevância. Os Latossolos têm correspondência aos *Oxisols* e *Ferralsols* (horizonte B ferrálico) na classificação WRB (IUSS Working Group WRB, 2014). De acordo com Pereira et al. (2019), os Latossolos encontram-se amplamente distribuídos por todas as regiões do país, especialmente em ambientes mais úmidos que favorecem a intensa intemperização.

Os Latossolos desenvolvem-se sobre diversos materiais de origem em climas tropicais úmidos ou equatoriais em regiões cobertas por floresta tropical, e particularmente nas encostas bem drenadas das paisagens (TARDY et al., 1994). Apresentam tipicamente uma estrutura micronodular constituída por um conjunto de caulinita e quartzo cimentado por óxidos de ferro.

Os fatores de formação dos Latossolos, englobam principalmente o relevo, clima e tempo (PEREIRA et al., 2019). As condições de paisagem com elevada topografia do relevo condicionam boa drenagem (infiltração de água no sistema). Esse tipo de processo é caracterizado pelos processos pedogenéticos múltiplos de transformação e perdas, relacionado à hidrólise (intemperismo químico). Sendo associado ao fator clima, fundamental para o fornecimento de água no sistema durante as reações químicas, pode ser tanto o clima presente

quanto registros de condições climáticas do passado (paleoclimas) e o tempo.

Destacam-se as características morfológicas de textura (homogeneidade/sem gradiente textural), cor uniforme, transições graduais a difusas entre os sub-horizontes, e os teores de Fe. Estas feições indicam uma ação intensa e uniforme de intemperização química no regolito, atuando por longo tempo (milhões de anos) (KÄMPF; CURI, 2012).

Ainda segundo os autores, os Latossolos possuem a presença significativa de Fe e Al e têm um efeito estabilizador sobre os argilominerais (caulinita), impedindo sua dispersão e eluviação. Na macroestrutura, originam-se microagregados fortes, tornando a consistência do solo muito friável (quando úmido). A forte agregação implica baixos valores nas análises de argila dispersa em água, bem como uma frequente subestimativa do teor de argila na análise granulométrica diante da dificuldade na dispersão do solo em laboratório.

Quanto às características mineralógicas destaca-se a razão $\text{SiO}_2/(\text{Al}_2\text{O}_3 + \text{Fe}_2\text{O}_3)$ reduzida, capacidade de troca catiônica (CTC) baixa, argila com baixa atividade e teores reduzidos de minerais primários suscetíveis à intemperização. Frisa-se que estas feições são requisitos do horizonte diagnóstico B latossólico (Bw) (SANTOS et al., 2018). Muitos desses solos são, provavelmente, formados por material alóctone pré-intemperizado e redepositado; outros são originados pelo dismantelamento de couraças lateríticas (KÄMPF; CURI, 2012). Ressalta-se que, os processos de dessilicação e ferratização também estão expressos em outras classes de solos, como os Nitossolos e em grau menor em Argissolos (KÄMPF; CURI, 2012).

Solos resultantes do processo de Ferralitização são geralmente considerados, poligenéticos. Estudos, como o de Calegari (2008) e Pereira (2016), demonstraram a caracterização de caulinitas para entender o cenário de evolução em Latossolos por meio de técnicas de datação, e essas evidências elucidam que as características do material não são representativas da intemperização atual.

2.4.3 Plintitização (laterização)

O processo de plintitização é caracterizado pelo acúmulo de plintita, por redução e oxidação, Fe^{2+} - Fe^{3+} , originando um aglomerado de óxido de ferro, nas condições de variação de umidade sazonal (clima tropical e subtropical), que auxiliam na alteração química dos minerais no solo, e associado a baixos teores de coloides orgânicos, assim formando a plintita (cor vermelha ou brunada) e subsequente a petroplintita por meio de ciclos alternados de secamentos e umedecimento.

De acordo com Kämpf e Curi (2012), as acumulações localizadas de óxidos de Fe (hematita e goethita), na forma de mosqueados e nódulos macios, geralmente de cor vermelha, capazes de endurecer irreversivelmente por meio de ciclos de umedecimento e secagem, são identificadas pelo termo plintita. Ainda segundo os autores, já em estágio superior, presume que, durante a alternância de ciclos, haja novos aportes de Fe que impregnam nos nódulos, bem como a cristalização dos óxidos, produzindo a sua cimentação; quando endurecido esse material passa a ser chamado de petroplintita (KÄMPF; CURI, 2012).

A ocorrência dos Plintossolos no Brasil corresponde a cerca de 6% da área do território nacional (maiores extensões se encontram na região Amazônica). Os solos com plintita ocorrem em áreas bem drenadas (Acre, Roraima, Amazônia, Pantanal, semiárido nordestino), enquanto

aqueles com petroplintita (concreções ferruginosas e laterita endurecida) são comuns nos topos e bordas dos planaltos de altitude (SCHAEFER et al., 2023).

Os Plintossolos são caracterizados pela presença do horizonte plíntico, litoplíntico ou concrecionário iniciando dentro de 40 cm a partir da superfície ou dentro dos 200 cm desde que seja precedido de horizonte glei ou horizontes com cores pálidas, variegadas ou com mosqueado em quantidade abundante. A presença de plintita em teor $\geq 15\%$ caracterizam o horizonte diagnóstico plíntico (Bf, Cf), que identifica os Plintossolos (SANTOS et al., 2018). Essa classe é correspondente aos *Plinthosols* no sistema *World Reference Base* (IUSS Working Group WRB, 2014) e no *Soil Taxonomy* ao subgrupo *Plinthic* (várias classes de *Oxisols*, *Ultisols*, *Alfisols*, *Entisols*, *Inceptisols*).

De acordo com WRB (IUSS Working Group WRB, 2014), os *Plinthosols* foram classificados nos subgrupos *Plinthic Plinthosols* e *Petroplinthic Plinthosols*; o horizonte petroplíntico é semelhante às lateritas ou “*ironstone*”. Os Plintossolos Pétricos possuem horizonte concrecionário ou litoplíntico, ou seja, são Plintossolos que têm como principal característica a ocorrência de petroplintita no perfil. A petroplintita, quando em forma contínua e com espessura superior a 10 cm, caracteriza o horizonte litoplíntico, que por sua vez, é uma característica diagnóstica do Plintossolo Pétrico Litoplíntico. Deste modo, se esses solos apresentam uma matriz solta, com ocorrência descontínua de nódulos ou concreções, ocupando mais de 50% em volume (massa do solo), caracteriza-se o horizonte concrecionário, que é uma característica diagnóstica dos Plintossolos pétricos concrecionários.

Destaca-se como fatores de formação o clima e o relevo, associados a uma fonte de ferro. Neste contexto, o ferro está sendo removido das partes mais altas e se acumulando nas partes mais baixas. E como processos pedogenéticos múltiplos a transformação, onde ocorre alternância de oxidação e redução do ferro (ambiente redutor), propiciando a elevada mobilidade e, por conseguinte, a translocação no perfil do solo.

Logo, a plintita se forma devido à precipitação de ferro na região de variação do nível do lençol freático, assim, em estágio inicial com a expressão de manchas irregulares, em estágios médio e avançado evoluem em nódulos e concreções. Assim, esse material, uma vez cimentado e reconhecido como uma camada extensiva, é denominado lateritas ou ferricretos (seção lateritas).

Anjos et al. (1995), em estudo no estado do Maranhão, relatou 13,5% de ocorrência de Plintossolos, sendo comum no Nordeste solos com plintita e/ou arenito (petroplintita). Ressalta-se que, as concreções ocorrem a diferentes profundidades nos perfis do solo, de acordo com a posição da paisagem. Segundo Anjos et al. (1995), a maioria dos solos plínticos no Brasil possui propriedades associadas a argilas de baixa atividade, mas foi relatado que esses solos possuem propriedades associadas a argilas de alta atividade.

Anjos et al. (2007) destacam que, em Plintossolos do Maranhão, identificaram plintita com argilominerais de alta atividade e índices Ki superiores a 2,2. A acidez acentuada em profundidade, atribuída ao processo de ferrólise, é decorrente da alternância redox (ANJOS et al., 2007). Anjos et al. (2007), observaram que a cristalização dos óxidos de ferro (relação $Feo/Fed > 0,05$) nestes Plintossolos indica um processo de laterização ativo, sustentado por reservas de ferros e minerais primários como a biotita; isso ressalta que tais formações não são apenas relictos geológicos inertes.

No grupo de rochas de formações ferríferas, lateritas ferruginosas (província estrutural

do cráton amazônico), as rochas ferruginosas tendem a ocorrer em associação com quartzitos na paisagem geral da encosta da montanha (SCHAEFER et al., 2023). No entanto, os solos são comumente petroplínticos ou concrecionários, e estas concreções são estáveis, ajudando a proteger o solo da erosão a longo prazo, em climas secos ou sazonais. Destaca-se a ocorrência comum nas áreas de Carajás (PA) e Quadrilátero Ferrífero (MG).

Na região de depressão central do Rio Grande do Sul, Miguel et al. (2013) constataram um mecanismo de elevados teores de ferro de baixa cristalinidade (Feo) foram identificados na parte interna das petroplintitas, atingindo valores de 1,7 a 1,8 g kg⁻¹ nos horizontes C BC do perfil 1, e de 1,6 g kg⁻¹ no horizonte Ec do perfil 3. A presença de Feo no interior de nódulos endurecido indica a ocorrência de ciclos de dissolução e reprecipitação dos óxidos cristalinos na estrutura concrecionária. Nesse sentido, destacam-se estudos de formações petroféricas como sistemas biogeoquímicos dinâmicos em constante evolução, em resposta ao ambiente pedogenético atual.

2.5. Mapeamento Digital de Solos (MDS)

O Mapeamento Digital do Solo (MDS) baseia-se em técnicas estatísticas e geoestatísticas para a predição espacial de atributos e classes de solo e integra o ramo da pedologia quantitativa, atualmente conhecida como Pedometria. Deste modo, a pedometria tem buscado quantificar e mensurar a variabilidade espacial de atributos e classes do solo, por meio da combinação de métodos de mapeamento e de ferramentas de estatística espacial, com base em algoritmos determinísticos, estocásticos ou híbridos (MCBRATNEY et al., 2000).

Estudos acerca de atributos e classes de solos no Brasil, como exemplo, Pinheiro et al. (2021), Samuel-Rosa et al. (2022) e Carvalho Junior et al. (2024), apontam para a importância da obtenção de informações acerca da variabilidade de atributos químicos, físicos e biológicos, para fins de manejo e planejamento de uso do solo, em escalas de maior detalhe. Recentemente, como política nacional, o Brasil instituiu o Programa Nacional de Solos do Brasil (POLIDORO et al., 2021), com a retomada de levantamentos pedológicos sistemáticos como suporte do planejamento das atividades de uso e conservação da terra.

No contexto histórico, os levantamentos de solos no Brasil tiveram início na década de 1950, impulsionados pela industrialização e pela necessidade de expansão agrícola, como o Programa de Levantamento e Reconhecimento dos Solos Brasileiros, aprovado em 1953. Conforme ressalta Carvalho et al. (2013), esse programa teve por objetivo atender aos critérios de reconhecimento e de planejamento em âmbito nacional, adotando-se mapeamentos generalizados, com nível de detalhe insuficiente para o planejamento de uso do solo.

Segundo Polidoro et al. (2016), atualmente uma pequena parte do País (cerca de 5%) conta com mapas de solos em escala 1:100.000 ou menor, o que contrasta de forma acentuada com os EUA, país integralmente coberto por mapas de solos em escalas entre 1:20.000 e 1:40.000. O quadro atual demonstra que, apesar do significativo progresso alcançado no conhecimento dos solos nacionais, há deficiência de informações detalhadas para orientação das atividades no meio rural, que podem ser sanadas com a retomada do levantamento sistemático nacional com base em técnicas modernas de processamento de dados. Ainda segundo Carvalho et al. (2013), destacam que o Mapeamento Digital de Solos (MDS) e a pedometria são ferramentas essenciais que utilizam geotecnologias e modelos matemáticos para prever classes e atributos do solo, otimizando recursos e permitindo a atualização das

informações.

2.5.1. Avanços das técnicas de MDS e pedometria

De acordo com Carvalho Junior et al. (2020), entre os anos de 2015 e 2019 nos periódicos *Geoderma*, *Regional Geoderma*, *Catena* e *Pedosphere*, a busca pelas palavras "*digital soil mapping*", constatou que 22% dos artigos publicados se referiam a estudos de mapeamento digital de classes de solos ou unidade de mapeamento. Nesse sentido frisa-se a lacuna no desenvolvimento e nas aplicações de MDS para classes de solos ou unidade de mapeamento para a contribuição na produção de trabalhos com escala de maior detalhe (1:100.000, 1:50.000 e 1:25.000), para o melhor conhecimento dos recursos pedológicos e planejamento das atividades de uso e conservação da terra.

A abordagem de mineração de dados para o mapeamento geológico, associada à disponibilidade de conjuntos de dados, e a evolução dos algoritmos contribuem para o sucesso das predições na obtenção de informações mineralógicas e litológicas de alta qualidade e baixo custo. Estudos recentes como os dos autores Cracknell e Reading (2014), Costa, Tavares e Oliveira (2019), Pinheiro et al. (2021), Wang et al. (2020) têm utilizado as técnicas de machine-learning, para a predição de litologias ou variabilidade de elementos/compostos geoquímicos, e que quando integradas às técnicas de sensoriamento hiperespectral multiescalar, têm contribuído para otimização de levantamentos geológicos com custo-benefício favorável.

Com o intuito de agregar à abordagem tradicional, nos últimos 30 anos, em resposta às limitações da mesma, tem havido uma adoção crescente da abordagem quantitativa para descrever e reconhecer os padrões e a distribuição do solo na paisagem (MCBRATNEY et al., 2000). O mapeamento digital de solos (MDS) é baseado em modelos numéricos, técnicas de sensoriamento remoto, variáveis ambientais e planos de informação temáticos/categóricos (Geologia, geomorfologia, uso e cobertura da terra), e/ou contínuos aplicados a um banco de dados geográficos com a finalidade de reconhecer os padrões da variabilidade espacial. Em McBratney et al. (2000) e McBratney, Santos e Minasny (2003), essa técnica é denominada pedometria.

Conforme Carvalho Junior et al. (2014), o mapeamento digital de solos é baseado essencialmente em banco de dados que espacializa as propriedades e classes do solo obtidas por meio de amostragem e modelagem estatístico-matemática da paisagem por métodos determinísticos e/ou estocásticos.

Nesse sentido, as geotecnologias (SIG e sensoriamento remoto) são uma fonte dos principais meios de obtenção de planos de informação (covariáveis) no processo de predição de classes e atributos do solo. Conforme Meneses, Almeida e Baptista (2019), os sensores remotos, em seus níveis de operação orbital, aéreo ou terrestre, podem captar (ou emitir e captar) a energia eletromagnética de fonte natural ou artificial, que possibilita as estimativas de propriedades e constituintes dos alvos. Ressalta-se que, essas técnicas têm obtido suporte de processamento em nuvem, atribuindo eficiência em levantamento de recursos, observação de tendências ou recuperação de cenários diversos (NOVAIS, 2021).

Nesse sentido, a pedometria tem buscado quantificar/mensurar a variabilidade espacial dos atributos e classes do solo, por meio da combinação de métodos de mapeamento de forma interdisciplinar entre as áreas de conhecimento da ciência do solo, a estatística e a

geoinformação (Figura 7).



Figura 7. A interdisciplinaridade da Pedometria. Fonte: Adaptado de Hengl (2003).

Minasny e McBratney (2016) frisam que os sistemas de posicionamento geográfico (GPS) tornaram acessível o referenciamento de amostras de campo associado ao registro de covariáveis ambientais no SIG, possibilitando a construção de modelos espaciais de solo (classes e/ou propriedades), a partir da correlação ambiental.

No âmbito de correlação ambiental Padarian, McBratney e Minasny (2020), enfatiza que o uso de métodos estatísticos e de aprendizado de máquina na modelagem de classes de solo ou propriedades com fatores ambientais é um componente central do mapeamento digital de solo. Visto que a correlação ambiental permite através de modelos estatísticos a predição de locais não observados, por meio das variáveis ambientais. Segundo Minasny e McBratney (2016), alguns autores argumentam que a correlação ambiental é simplesmente a evolução natural e quantitativa do clorpt, este modelo espacial do solo é puramente baseado em uma abordagem estatística com incerteza de previsão conhecida.

Nesse contexto, a evolução das técnicas de mapeamento digital de solos ganhou repercussão em 1980, com suporte do avanço na área de tecnologia da informação e sensoriamento remoto (MINASNY; MCBRATNEY, 2016). Em 1988 é instituída a comissão 1.6 *Pedometrics* (MINASNY; MCBRATNEY, 2016), na *International Society of Soil Sciences* (ISSS) na *International Union of Soil Sciences* (IUSS). Posteriormente em 2009 é criada a Comissão Brasileira de Pedometria (OLIVEIRA; MEDEIROS; FARIAS, 2015).

Nesse sentido, os avanços das técnicas de mapeamento digital de solos (Figura 8) são o somatório do advento e progresso do SIG, poder computacional, disponibilidade de dados espaciais e nova geração de cientistas. Assim, a cartografia digital do solo visa criar e preencher sistemas de informação espacial do solo (SSINFOS) que possam ajudar os usuários a tomar decisões sobre os territórios para lidar com problemas agrícolas, agroambientais, entre outros (LAGACHERIE, 2008).

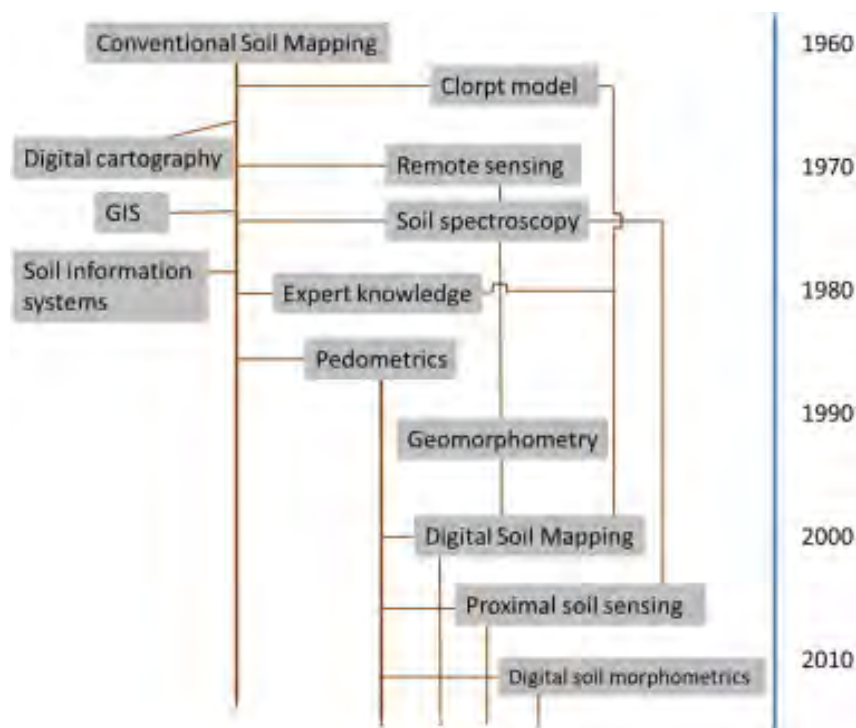


Figura 8. Diagrama mostrando o avanço das técnicas de Mapeamento Digital de Solos. Fonte: Minasny e McBratney (2016).

Ressalta-se que, o mapeamento digital não está ocorrendo só na ciência do solo, estudos têm usado mapeamento na busca por elucidar a distribuição espacial de fenômenos através da modelagem ambiental por meio de técnicas de mineração de dados (MAXWELL et al., 2019; CAMPBELL et al., 2020; FARHADI; NAJAFZADEH, 2021). Acredita-se que, o mapeamento digital a partir do uso de covariáveis ambientais pode impulsionar essa abordagem como uma fronteira do conhecimento.

Estudos recentes Arrouays et al. (2017), Demattê et al. (2018), Carvalho Junior et al. (2020), Adel, Mansour e Ardeshtir (2021) e Chen et al. (2022), em diversas partes do mundo têm mostrado o ganho em otimização de tempo e o suporte de técnicas a partir de SRTM; SAR; LiDAR (*Light Detection And Ranging*), espectroscopia (VIS-NIR-MIR). Esse fato demonstra o ganho do mapeamento digital de solos em termo de produção de informação com melhor precisão, quando comparado ao mapeamento convencional.

2.5.2. Modelos preditivos e fatores de formação do solo

O solo é um sistema natural, aberto e complexo, desse modo, pedólogos utilizam modelos matemáticos para elucidar as dinâmicas e mecanismo desse recurso. O sistema natural complexo é explicado por um modelo mais simples ou mais abstrato, que pode ser mais facilmente entendido (DIJKERMAN, 1974).

Nesse sentido, pedólogos fazem uso de modelos como, por exemplo, os clássicos de Jenny (1994) e McBratney, Santos e Minasny (2003), apesar de possuírem finalidades distintas, possuem correlação com a abordagem da influência dos fatores de formação e gênese dos solos, contribuindo com a compreensão da ocorrência de classes de solo e seus atributos, na paisagem.

A partir da compreensão das paisagens e dos fatores de formação, adotando uma abordagem clássica, Jenny (1994) em “*Factors of Soil Formation: a system of quantitative*

pedology", destaca os cinco principais fatores que exercem influência na formação do solo: i) clima (temperatura, umidade e pressão atmosférica); ii) organismos vivos (animais e vegetais); iii) relevo; iv) material de origem; e v) tempo (tempo geológico). Desse modo, esses fatores interagem de forma dinâmica ao longo do tempo e do espaço na unidade do solo, resultando nas características específicas de uma paisagem e na variação e distribuição dos solos nessa paisagem.

Destaca-se que o modelo fatorial de formação de solo de Jenny (1994) utilizado conceitualmente para o mapeamento convencional, o "CLORPT" (Eq. 1):

$$S = f(\text{cl, o, r, p, t, ...})$$

Onde os fatores do solo (S): Clima (Cl), organismos (o), relevo (r), material de origem (p), tempo (t), e fatores desconhecidos (...). Esse método qualitativo propôs um modelo preditivo de classes pedológicas.

Considerando a formação do solo sob o viés espacial, McBratney, Santos e Minasny (2003), desenvolveram o modelo estatístico chamado de "*Digital Soil Mapping*", onde faz uso de informações espaciais para prever os atributos e classes do solo em uma determinada área. Nesse sentido, para explicar as relações quantitativas ou qualitativas entre solo e outros fatores ambientais, adotou-se o modelo "SCORPAN" (MCBRATNEY, SANTOS E MINASNY, 2003), (Eq. 2):

$$S = f(s, c, o, r, p, a, n)$$

Onde as classes ou atributos do solo são expressos como funções das informações do solo (S) em um determinado ponto; condições climáticas (C) (ponto específico do ambiente); organismos (O) (vegetação, fauna e atividades humanas); relevo (R) e variáveis morfométricas; material de origem (P); da idade (A); e da posição (N) no espaço.

Destaca-se que o diferencial desse modelo é os fatores "s" indicando que a predição de propriedades/classes do solo pode ser predita com prévias informações disponíveis sobre esse recurso. E "n" aponta para a relação espacial entre as propriedades do solo e as características do ambiente, determinando que as propriedades do solo podem ser preditas a partir de uma posição geográfica específica (coordenadas X e Y) (SAMUEL-ROSA, 2017). Assim, estudos podem definir as variáveis com maior influência/importância a partir da representação dos fatores de formação.

Samuel-Rosa (2017) ressalta que um valor acrescido à abordagem desse modelo com cunho espacial é o uso de variáveis ambientais, como dados topográficos, imagens de satélite, características geológicas e outros dados auxiliares, para realizar a predição de atributos e classes dos solos. Além disso, frisam-se os produtos do modelo scorpan que incluem os conceitos de incerteza e acurácia associados às FPESe, os quais aparentemente não estariam disponíveis quando do uso do modelo clorpt.

Os modelos de técnicas pedométricas podem ser divididos em determinísticos e estocásticos (Figura 9) de acordo com McBratney et al. (2000).

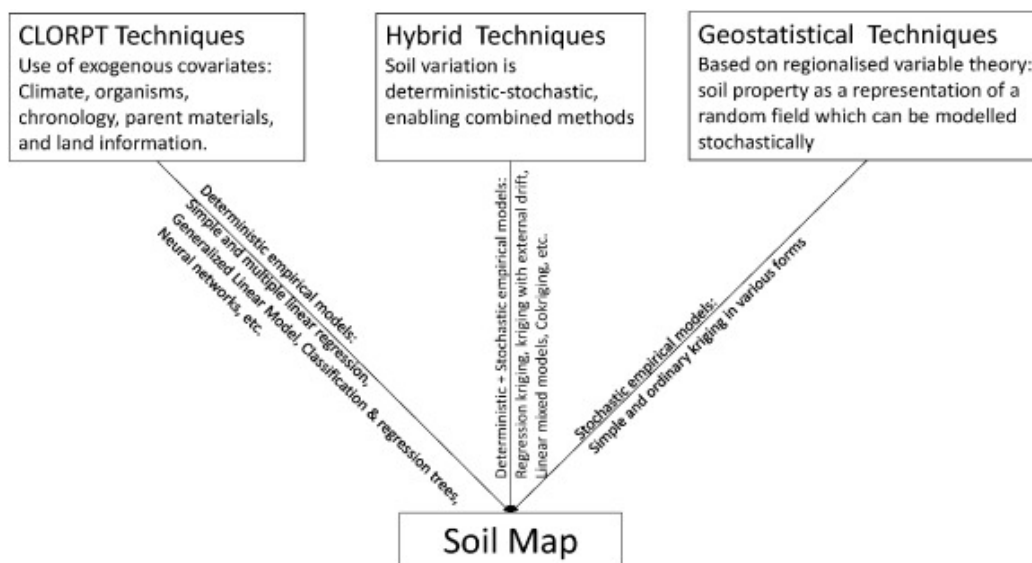


Figura 9. Técnicas pedométricas. Fonte: McBratney *et al.* (2000).

Os modelos determinísticos, possuem uma relação de causa e efeito. Utilizam variáveis exógenas, tais como: clima, organismos, tempo e uso da terra. Os modelos comumente usados são regressões lineares e polinomiais, modelos baseados em árvores, modelos aditivos generalizados (GAM), redes neurais (NN) entre outros.

Modelos estocásticos utilizam-se da teoria de variáveis regionalizadas (técnicas geoestatísticas), tendo como referência as propriedades do solo com uma representação de um campo aleatório que pode ser modelado estocasticamente. Já os modelos determinísticos-estocásticos, conhecidos como técnicas híbridas, fazem uso de métodos combinados como, por exemplo, regressão/krigagem, krigagem com incerteza, co-krigagem, entre outros. Chagas (2006) ressalta que a aplicação de cada uma das técnicas pedométricas depende do objetivo, resolução e organização do levantamento, visto que o uso final da informação do levantamento do solo determina a exatidão necessária.

2.5.3 Algoritmos de aprendizagem de máquina

A aplicação de algoritmos de aprendizagem de máquinas na predição de classes e atributos de solos e materiais geológicos vem sendo cada vez mais utilizada em levantamentos no campo das Geociências. Como exemplo destes algoritmos destacam-se as Redes Neurais Artificiais (RNA), Cubist (CB), *Random Forest* (RF) e *Support Vector Machine* (SVM) entre outros.

Entre os modelos estatísticos preditivos, destaca-se o *Random Forest* (RF), desenvolvido por Breiman (2001), como uma extensão do *Classification and Regression Trees* (CART). É considerado uma técnica não paramétrica que tem o desempenho de melhorar a predição deste modelo. Segundo Breiman (2001), consiste numa combinação de árvores preditoras (floresta), na qual é gerada a partir de um vetor aleatório, amostrado de forma independente, que obedece à mesma distribuição para as demais árvores. A aplicação da RF, em caso de regressão, tem resultado final que equivale à média dos resultados de todas as árvores.

No modelo RF, Breiman (2001) destaca vantagens como a capacidade de modelar

relações dimensionais não-lineares; uso de variáveis categóricas e contínuas; resistência a "overfitting"; robustez em relação à presença de ruído nos dados; estabelecimento de uma medida imparcial da taxa de erro; capacidade de determinar a relevância das variáveis utilizadas; e exigência de poucos parâmetros para implementação. Assim, Heung, Bulmer e Schmidt (2014) também ressaltam as características do modelo considerado como um algoritmo robusto, pouco sensível a ruídos nos dados e ao *overfitting* (super ajuste), tendo como característica a capacidade de fornecer medidas de importância das covariáveis na predição.

O algoritmo *Multivariate Adaptive Regression Splines* (MARS) foi desenvolvido por Friedman (1991), sendo um método não paramétrico. O método adota como estratégia a divisão e conquista, sendo os conjuntos de dados de treinamento divididos em segmentos lineares separados em partes (slips) e gradientes diferentes (inclinação) (ZHANG et al., 2022). De acordo com Hosseinzadeh et al. (2020), consideram-se os finais de segmentos dos pontos como nós, sendo um demarcador de região de dados, dando início a outra região.

As vantagens do MARS são: seleção automática das variáveis mais importantes para o modelo e produção de um conjunto de equações, permitindo reprodutibilidade por outros usuários. De acordo com Abreu (2016), adequa-se a problemas multidimensionais, e embora este método tenda a ser mais rápido do que as redes neurais feed-forward usando back-propagation, problemas mais complexos com múltiplas aproximações tendem a tornar este método bastante lento.

O algoritmo *Support Vector Machines Radial Sigma* (SVM-R) constitui-se em regressões não lineares normalmente obtidas pela aplicação do chamado truque do kernel (DIAS et al., 2021). Ainda segundo o autor, a característica distintiva do SVM é a sua capacidade de definir funções de decisão complexas que separam de forma otimizada duas classes de amostras de dados.

O *Cubist* (CB) é um algoritmo baseado em regras usado como uma extensão da árvore de regressão não paramétrica que aborda a relação linear entre a variável alvo e as variáveis preditoras (BASSER, 1992; QUINLAN, 1995). De acordo com Nguyen et al. (2019), o processamento do CB consiste em quatro etapas: i) Dividir os dados para criar uma árvore completa; ii) Criar um modelo de regressão em cada nó para preparar a poda e a previsão; iii) podar a árvore para evitar o problema de sobreajuste; e iv) Suavizar a árvore para ocultar os limites de descontinuidade causados pela divisão.

2.5.4 Dados legados

Os dados conhecidos como legados consiste em informações cartográficas e relatórios obtidos previamente, disponíveis em várias fontes. Esses bancos de dados são de suma importância para a Pedometria, pois podem ser usados tanto para criar mapas digitais de solos quanto para validar os produtos finais. De acordo com Medeiros et al. (2024), o uso desses dados traz alguns desafios, como a distribuição irregular das amostras, resolução limitada dos mapas, dificuldade em representar toda a complexidade do solo, além de questões relacionadas à inclusão de mapas antigos de solos nos processos de mapeamento digital.

Nesse contexto, a colaboração global é importante para melhorar nossa compreensão dos recursos limitados do solo no mundo, para que possam ser usados de forma sustentável (FAO, 2015). O mapeamento colaborativo pode ser uma das estratégias a serem adotadas nos

programas de levantamento de solos para atualização de dados. De acordo com McBratney, Santos e Minasny (2003), o mapeamento colaborativo é a agregação de mapas on-line e de conteúdo gerado pelo usuário, de um grupo de indivíduos ou entidades, e pode tomar muitas formas distintas.

Em relação aos primeiros esforços para reutilizar dados históricos de solos, vale destacar as estruturas genéricas para mapeamento de solos propostas por McBratney (2003), Chagas (2004) e Sanchez et al. (2009), as quais possibilitaram que diversas plataformas digitais fossem desenvolvidas para tornar esses dados globalmente acessíveis, como o GlobalSoilMap (Arrouays et al., 2017c), Grades de Solo (Hengl et al., 2014, 2017), Biblioteca Espectral de Solos do Brasil (Demattê et al., 2019) e Dados do Solo MapBiomias (2023). Isso só se tornou viável graças à maior eficiência e aos avanços tecnológicos, aliados à crescente popularidade resultante do acesso mais fácil às ferramentas necessárias.

No Brasil têm se propagado iniciativas de atualização e resgate de dados, as quais buscam contribuir com os avanços das técnicas de mapeamento digital de solos, visando disseminar dados, que costumavam frequentemente ficar velados dentro de laboratórios e atualmente são acessíveis em plataformas digitais. Samuel-Rosa et al. (2019) destacam que o Brasil tem uma vasta quantidade de dados de solos provenientes de levantamentos de solos e projetos de pesquisa sobre os diversos aspectos da ciência do solo, que em grande parte são utilizados apenas uma vez. No caso de dados legados, por muitas das vezes esses dados estão incompletos ou incorretos, conforme ressaltam os estudos de Arrouays et al. (2017) e Samuel-Rosa et al. (2019).

Assim, para verificar o potencial de reutilização e dos legados, é necessária a adoção de metodologias que considerem aspectos como a análise da escala cartográfica, a idade, a finalidade dos mapas, além da quantidade e da precisão das informações, com a finalidade de avaliar o potencial de reutilização dos dados sobre solos. Samuel Rosa et al. (2020), apresentam essa abordagem de estudo com a plataforma *SoilData* (disponível em <https://www.pedometria.org/soildata/>); a plataforma funciona como um repositório central e interface de acesso ao Repositório Brasileiro Livre para Dados Abertos do Solo (FEBR), que, por sua vez, visa promover os princípios da ciência aberta, a recuperação e a qualidade dos dados geoespaciais, bem como a disponibilização de dados legados sobre os solos do Brasil.

Estudos bibliométricos recentes de Medeiros et al. (2024) demonstraram que a utilização de dados legados evoluiu crescentemente entre 1979 e 2022, consolidando-se, assim, como um pilar estratégico para a pedometria. Nesse contexto, países como Austrália, Estados Unidos e Brasil têm utilizado esses acervos para modelos preditivos. Contudo, essa temática traz o desafio da dicotomia entre a ampla disponibilidade de dados existentes e a real usabilidade desse material. Frisa-se que o uso de dados legados, associado às técnicas de modelagem preditiva, não depende apenas do regaste de acervos, mas sobretudo, da harmonização para mitigar as incertezas inerentes a dados passados.

3. CAPÍTULO I

MAPEAMENTO DIGITAL DOS TEORES DE Fe_2O_3 , MnO , Nb_2O_5 E TiO_2 EM MORRO DOS SEIS LAGOS-AM, BRASIL

RODRIGUES, Niriele Bruno; SILVA, Júlio Cesar Lopes da; MARINATTI DA SILVA, Renan Pereira; KOENOW PINHEIRO, Helena Saraiva; JÚNIOR, Waldir de Carvalho. Mapeamento Digital dos Teores de Fe_2O_3 , MnO , Nb_2O_5 e TiO_2 em Morro dos Seis Lagos – AM, Brasil. Espaço Aberto, Rio de Janeiro, Brasil, v. 14, n. 1, p. 157–174, 2024. DOI: 10.36403/espacoaberto.2024.60234. Disponível em: <https://revistas.ufrj.br/index.php/EspacoAberto/article/view/60234>.

3.1 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_2O_5 e TiO_2 , em Morro dos Seis Lagos-AM, Brasil. A metodologia consistiu em: Revisão temática; Compilação dos dados geoquímicos; Tratamento e análise dos dados (*input data*); Seleção de 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_2O_5), 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. Aprendizagem de máquina. Área de difícil acesso.

3.2 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_2O_5 , 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_2O_5) 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.

Key words: Pedometrics. Machine learning. Poorly-accessible areas.

3.3 INTRODUÇÃO

A demanda por acesso à informação de recursos naturais aumentou muito nas últimas décadas, em especial sobre o recurso solo. Nesse contexto, o mapeamento digital de solos (MDS) desempenha papel fundamental ao lidar com vasto volume de dados e gerar informações sobre os solos e suas propriedades. Portanto, os acervos 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 de MDS para prever 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 à tomada de decisões.

Atualmente, a prospecção mineral tem se baseado em sistema de informação geográfica (SIG), com a finalidade de suporte à 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 poucos ou nenhum local mineralizado conhecido (PRADO et al., 2020).

Estudos dos autores Cracknell e Reading (2014), Costa, Tavares e De Oliveira (2019), Pinheiro et al. (2021) e Wang et al. (2020) utilizam as técnicas de *machine-learning*, para a previsão de dados geoquímicos e, principalmente, se integradas com dados de sensoriamento remoto (SR), contribuem para o mapeamento de teores de elementos e compostos, que superam as técnicas convencionais. Segundo Demattê et al. (2018), o uso de SR é cada vez mais importante na modelagem espacial e no MDS. Isso se deve a uma relação entre o solo e a interação com a energia na forma de distintos padrões espectrais.

Na área de estudo, o Complexo Carbonatítico Morro dos Seis Lagos, intemperismo do siderita carbonatito originou espessa crosta laterítica (maior que 200 metros), composta predominantemente de *goethita/ hematita* (TAKEHARA, 2019; BENTO et al., 2022). Frisa-se que, tais carbonatitos e lateritas hospedam minerais traço com elementos com potencial econômico (Nb_2O_5 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 do estudo 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_2O_5 e TiO_2 , em Morro dos Seis Lagos, São Gabriel da Cachoeira (AM).

3.4 MATERIAL E MÉTODOS

Os procedimentos metodológicos dividem-se em 6 etapas (Figura 10): i) Revisão temática; ii) Compilação dos dados geoquímicos; iii) Tratamento e análise da base de dados (*input data*); iv) Seleção de covariáveis; v) Aplicação de algoritmos de mapeamento digital para predição de elementos e posterior validação; e vi) Integração e análise dos resultados e interpretações.

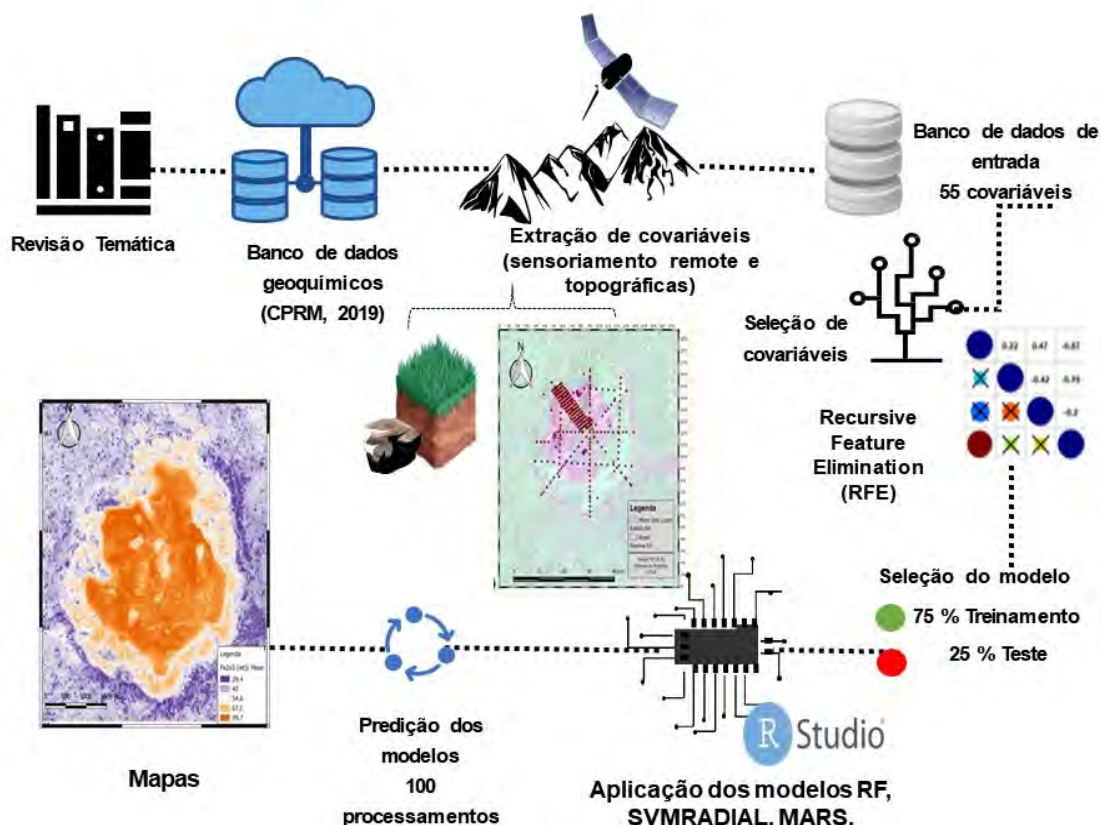


Figura 10. Etapas dos procedimentos metodológicos. Fonte: Elaborado pela autora (2023).

A região de Morro dos Seis Lagos (Figura 11) 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 (BRASIL, 2009), Parque Nacional do Pico da Neblina (BRASIL, 1979) e a Reserva Biológica do Morro dos Seis Lagos (BRASIL, 2009).

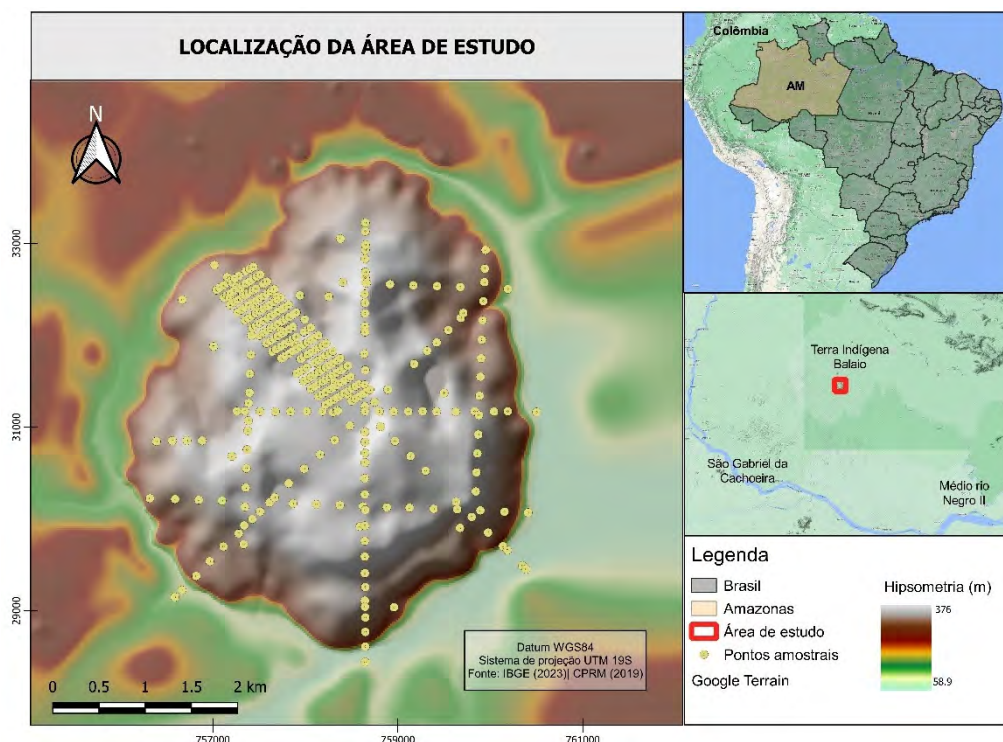


Figura 11. Localização da área de estudo. Fonte: Elaborado pela autora (2023).

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 (TAKEHARA, 2019). 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 (TAKEHARA, 2019):

- Método SGS XRF79C, com identificação dos minerais feita através da comparação do difratograma do banco de dados de ICSS-PDF (*International Center dor Diffraction Data-Powder Diffracton file*), para os óxidos principais;
- 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), Nb₂O₅, Sn, W, Y e outros 11 elementos; e
- ISE03A e B para os elementos F e Cl.

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

3.4.1 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 do Instituto Brasileiro de Geografia e Estatística (IBGE)- escala 1:25.000. 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 provenientes de dados de sensoriamento remoto foram obtidas pelo sensor *Multispectral Sensor Instrument* (MSI) do satélite Sentinel 2A e pelo sensor *Advanced Spaceborne Thermal Emission and Reflection Radiometer* (ASTER) do satélite Terra, na resolução espacial de 20 m, sendo os índices gerados através da combinação de bandas espectrais descritos com detalhes na Tabela 1, a seguir.

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, Van der Werff e Van Ruitenbeek (2014)
Zonas alteradas	Banda 8/ banda 11	Sora et al. (2018)
Rochas alteradas hidrotermicamente	8-1-2-3	Sora et al. (2018)
<i>Laterite</i>	Banda 11/ banda 12	Van der Meer, Van der Werff e Van Ruitenbeek (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 (NDVI)</i>	$NDVI = \frac{RNIR - Rred}{RNIR + Rred}$	Rouse et al. (1974)
<i>Ferroginous regolith</i>	(Banda 8A/ banda3)	Rowan & Mars (2003)
ASTER		
Deteção da alteração argilítica avançada	Banda 4 /banda 6	Sora et al. (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)
<i>Gossan</i>	Banda 4/ banda 2	Kalinowski e Oliver (2004)

Fonte: Elaborado pela autora (2023).

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; TIBSHIRANI; FRIEDMAN, 2009).

O procedimento de seleção de covariáveis para construção dos modelos preditivos envolveu três etapas, sendo elas: i) Remoção de covariáveis com variância zero ou quase zero; ii) Remoção de covariáveis altamente correlacionadas com limiar de 0.95; e iii) 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. Por conseguinte, foi estabelecido um critério de limiar de 0,95 para identificar as covariáveis com correlação significativa. Portanto, 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 (multicolinearidade), 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* (KUHN, 2008). Em vista disso, utilizou-se o método de validação usada repetida (*repeated cross-validation*) com 10 *folds* e 2 repetições. O algoritmo baseou-se em três etapas recursivas: (i) Ajuste do modelo e aprendizado de máquinas junto às covariáveis disponíveis; (ii) Cálculo e importância de cada variável com base na métrica do modelo; e (iii) Eliminação das variáveis menos importantes e do hiperparâmetro. O espaço de busca para o número ideal de covariáveis foi a partir do *subset size* do RFE, ajustado para gerar subconjuntos com diferentes números de covariáveis, com base nas covariáveis remanescentes das etapas anteriores de seleção. Foram testados cinco conjuntos específicos, variando de 2 a 25, além de 30, 40 e 45 covariáveis. Assim, a seleção dos subconjuntos foi realizada por meio de validação cruzada com cinco dobras, que avaliou a métrica de precisão R^2 para cada subconjunto. Essa abordagem permitiu identificar o número mínimo de covariáveis para otimizar predição e reduzir a redundância e colinearidade do conjunto de dados.

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* (KUHNN, 2008), em função de suas características dessemelhantes, a fim de avaliar a predição mais precisa dos elementos (Fe₂O₃, MnO, Nb e TiO₂).

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 e avaliar a importância de cada uma. Portanto, para 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. A otimização dos hiperparâmetros internos de cada modelo foi implementada via pacote *caret*. Adotou-se a estratégia de busca padrão (*tuneLength=1*), baseada na métrica de minimização do Erro Absoluto Médio (MAE), ajustando-se os parâmetros: *mtry* para *Random Forest*; C e Sigma para SVMRadial; e *nprune* e *degree* para o 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 via as 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 (R²) (eq. (3)).

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

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

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

Em que:

- n*: número total de observações (amostras) do conjunto de dados;
- y_i*: valor observado (real a variável alvo para a *i*-ésima amostra;
- ŷ_i*: valor predito (estimado) pelo modelo para *i*-ésima amostra;
- ȳ*: média aritmética os valores observados;
- i*: índice da amostra, variando de 1 a *n*.

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. 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.

3.5 RESULTADOS E DISCUSSÃO

3.5.1 Estatística descritiva

Os elementos foram, inicialmente, submetidos à análise estatística descritiva (Tabela 2 e Apêndice A), 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 (TiO_2) exibiu desvio padrão de 4,09, com conteúdo médio de 2,93 e coeficiente de variação de 16,69. E para Nb_2O_5 obteve-se média de 0,74, com valores de mínimo a máximo de 0 a 4,02, contribuindo para coeficiente de variação baixo de 0,76.

Tabela 2. Estatística descritiva de elementos.

Elementos (% em peso)	Média	Med.	Min.	Máx.	DP	CV(%)	Cs	Ck
Fe_2O_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_2O_5	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

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

Fonte: Elaborado pela autora (2023).

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 2), no que se refere ao erro absoluto médio (MAE) e à raiz do erro quadrático médio (RMSE), dos modelos com melhor desempenho no R^2 , verificaram-se valores significativamente mais baixos de MAE e RMSE representando assim uma maior precisão (WILLMOTT; MATSUURA, 2005).

Constatou-se que para o Nb_2O_5 o RF apresentou R^2 0,23, no desempenho, e RMSE e MAE inferiores em relação aos demais modelos. O SVMRadial apresentou menor desempenho com R^2 de 0,15. Bento et al. (2022) através do método de interpolação *Inverse Distance Weighting* (IDW) estimaram o teor médio para Nb_2O_5 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) de Nb_2O_5 .

Para o Titânio (TiO_2) o SVMRadial apresentou melhor performance com R^2 de 0,14, em comparação aos demais modelos. Adel, Mansour e Ardeshtir (2021) na região do Iran constataram para o elemento TiO_2 o R^2 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 (Fe_2O_3) o modelo mais adequado foi o RF com $R^2 = 0,23$. 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 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 R^2 equivalente a 0,20, e os demais modelos não apresentaram diferença estatística entre si (Tabela 3). Recentemente Bento et al. (2022) destacaram que nas crostas manganesíferas, há elevados teores de elementos de terras raras (ETR) referentes 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 à 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_2O_5	RF	19,92	14,24	0,23
	MARS	20,60	15,29	0,16
	SVMRadial	21,25	14,10	0,15
TiO_2	RF	4,00	2,59	0,12
	MARS	4,00	2,61	0,11
	SVMRadial	4,03	2,32	0,14
Fe_2O_3	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

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

Fonte: Elaborado pela autora (2023).

De modo geral, as covariáveis preditoras para os modelos apresentaram uma característica heterogênea na composição de seus conjuntos (Figura 12). 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*, *MRVBF*, *Saga Wetness Index*, *Standardized Height* e *Vector Ruggedness*, as quais exerceram influência significativa na explicação da variabilidade dos elementos.

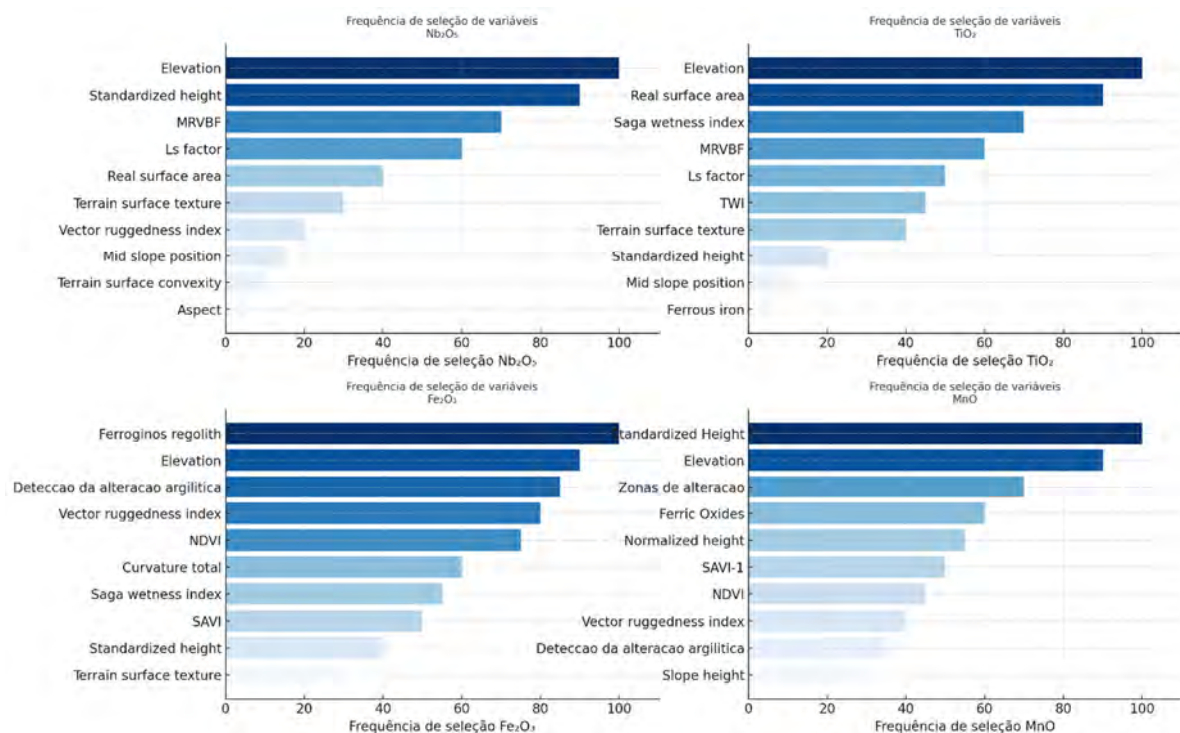


Figura 12. Frequência de seleção das covariáveis, para a predição espacial dos elementos, a partir do modelo RFE para os modelos. Fonte: Elaborado pela autora (2023).

Constatou-se que as covariáveis de índices espectrais (*Ferric Iron*, *Ferrous Iron*, detecção da alteração argilítica avançada) obtiveram destaque nos modelos, uma vez que estão relacionadas diretamente com alterações mineralógicas, principalmente as relacionadas à precipitação de óxidos de ferro. Esses resultados são relevantes para a compreensão dos processos de intemperismo das rochas carbonatíticas e da gênese das crostas lateríticas e, como consequência, do acúmulo de elementos como Nb₂O₅ e TiO₂.

3.5.2 Análise da variabilidade espacial dos elementos

Os mapas de média e amplitude apresentados nas Figuras 13, 14, 15 e 16 exibem a distribuição espacial dos teores dos elementos em superfície.

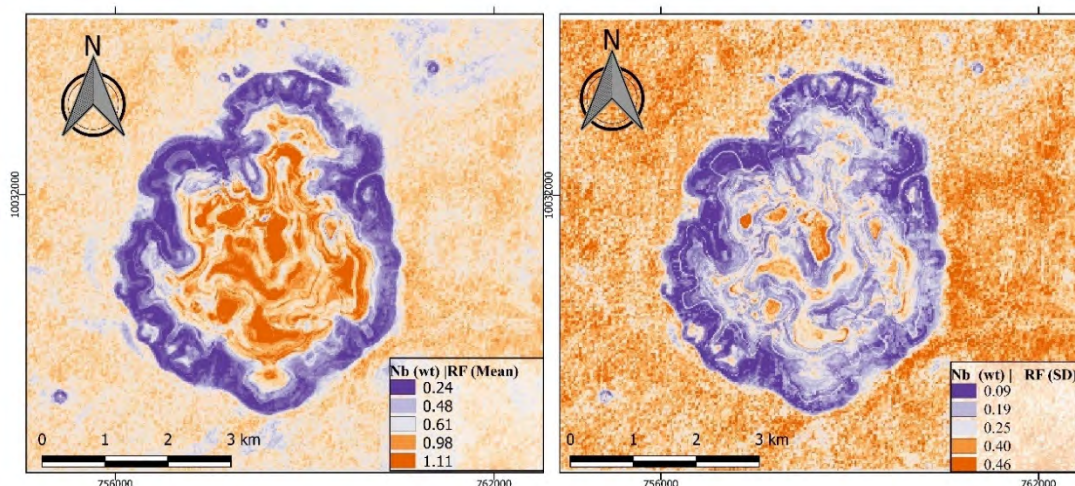


Figura 13. Mapeamento superficial de Nb₂O₅ em Morro de Seis Lagos por *Random Forest*. Fonte: Elaborado pela autora (2023).

Na variabilidade do Nióbio (Nb_2O_5) (Figura 13), 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 e 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.

As covariáveis morfométricas *Elevation*, *Aspect* e *Multiresolution Index of Valley Bottom Flatness* (MRVBF) e *LS-factor* (Figura 14) foram as que indicaram maior importância na explicação da variabilidade do elemento. Frisa-se que 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 o 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).

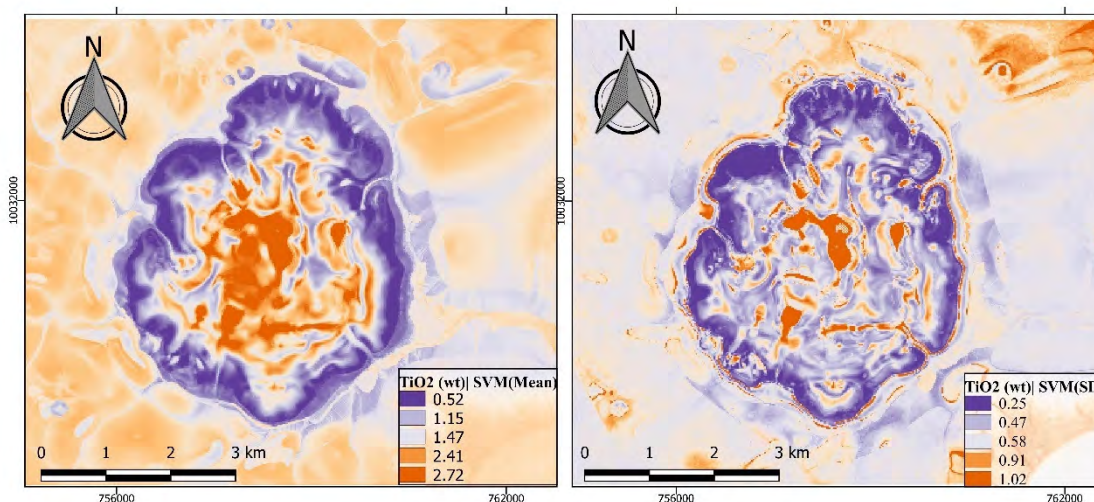


Figura 14. Mapeamento superficial de TiO_2 em Morro de Seis Lagos por SVMRadial. Fonte: Elaborado pela autora (2023).

O elemento Titânio (TiO_2) no mapa médio (Figura 14) 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 talus. 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 apresentaram *Elevation*, *LS-factor*, *Saga Wetness Index* e *Multiresolution Index of Valley Bottom Flatness* (MRVBF).

Para Fe_2O_3 (Figura 15) os maiores teores entre 67,3 - 79,8 (% em peso), estão relacionados à presença da crosta laterítica, e no depósito de talus se localiza uma menor variação de 54,8-67,3 (% em peso). 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 talus. As covariáveis *Elevation*, *Ferroginous regolith*, Detecção da alteração argilítica avançada, *Vector*

Ruggedness index, *NDVI*, *Saga Wetness index* foram consideradas as mais importantes na predição da distribuição espacial dos teores de Fe (Figura 12).

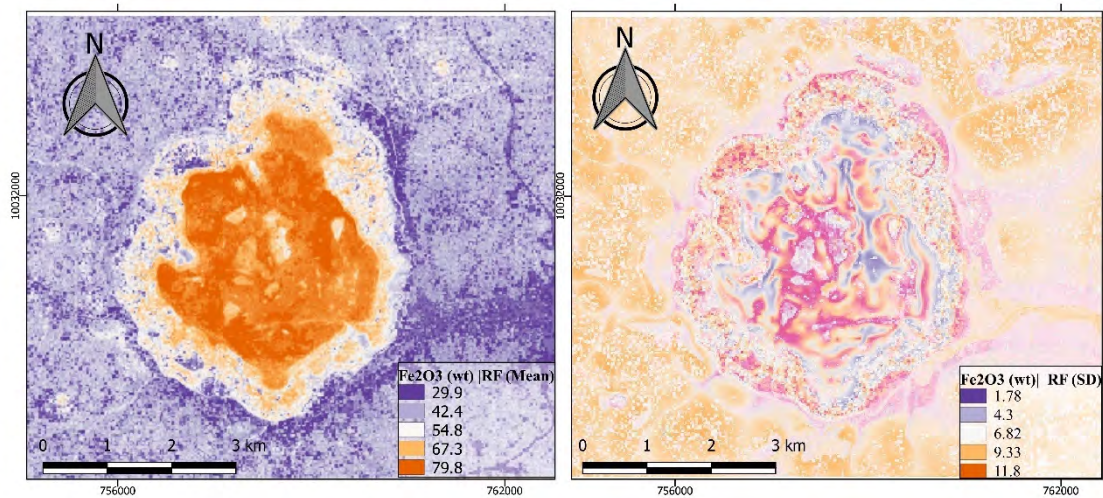


Figura 15. Mapeamento superficial de Fe_2O_3 em Morro de Seis Lagos por *Random Forest*.
Fonte: Elaborado pela autora (2023).

O mapa que representa os teores médios preditos para o elemento Manganês (MnO) (Figura 16) apresentou menor variabilidade na crosta laterítica, com valores entre 0,11 – 11,0 (% em peso), e maiores teores, variando entre 32,9 – 43,9 (% em peso) na área de depósitos de talus. 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). O padrão de enriquecimento de MnO no talus indica uma mobilidade geoquímica distinta, com lixiviação no topo do perfil e reprecipitação secundária ou acumulação nas zonas de deposição de encostas. Destaca-se que esse padrão é diverso quando se observa a dinâmica dos elementos Fe_2O_3 , TiO_2 e Nb_2O_5 na área de estudo, pois tendem a concentrar-se residualmente na crosta laterítica devido à lixiviação dos componentes móveis. As covariáveis importantes foram *Elevation*, *Ferric Oxides*, *Zonas de alteração*, *SAVI* e *Deteccção da alteração argilítica avançada*.

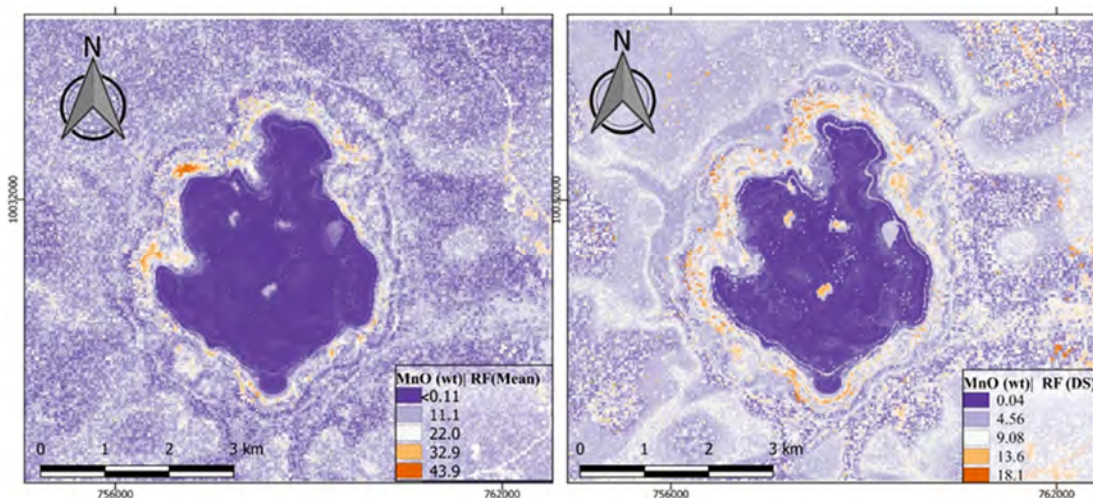


Figura 16. Mapeamento superficial de MnO em Morro de Seis Lagos por *Random Forest*.
Fonte: Elaborado pela autora (2023).

Nesse sentido, o modelo *Random Forest* obteve melhor desempenho para os elementos Fe_2O_3 , Nb_2O_5 e MnO , indicando as mais importantes covariáveis: Elevation, SAGA Wetness Index e Detecção argilítica, que foram pré-selecionadas 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, 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 à resolução espacial. Em El-Desoky et al. (2022), com a utilização do Sentinel-2A e ASTER demonstrou boa resposta espectral entre as bandas para o mapeamento geológico com média resolução.

3.6 CONCLUSÕES

Ao considerar que o objetivo do trabalho foi avaliar o desempenho dos modelos MARS, SVMRadial e RF, para predição espacial dos conteúdos de Fe_2O_3 , MnO , Nb_2O_5 e TiO_2 , em Morro dos Seis Lagos, São Gabriel da Cachoeira, Amazônia, conclui-se que o mesmo foi alcançado. 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_2O_5 , em comparação aos modelos SVMRadial e MARS. Verificou-se que os maiores teores dos elementos Nb_2O_5 , 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*, *Multiresolution Index of Valley Bottom Flatness* (MRVBF), *SAGA Wetness Index* e Detecção da alteração argilítica avançada, resultando em otimização do conjunto de dados de entrada.

As covariáveis morfométricas desempenharam papel 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 regulador da abundância dos elementos / compostos químicos analisados.

Como perspectivas futuras frisa-se que os modelos podem ser aperfeiçoados 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.

4. CAPÍTULO II

SAMPLING DESIGN AND MODEL SELECTION INFLUENCE ON PREDICTIONS OF CHEMICAL COMPOUNDS IN PETROFERRIC FORMATIONS IN THE BRAZILIAN AMAZON

Rodrigues, N. B., Barbosa, T. R., Pinheiro, H. S. K., Mancini, M., Read, Q. D., Blackstock, J., Winzeler, E. H., Miller, D., Owens, P. R., & Libohova, Z. (2025). Influences of Sampling Design and Model Selection on Predictions of Chemical Compounds in Petroferric Formations in the Brazilian Amazon. **Remote Sensing**, 17(9), 1644. <https://doi.org/10.3390/rs17091644>

4.1 RESUMO

Morro de Seis Lagos é uma região na Amazônia brasileira, abriga uma rara formação (menos de 1%) de carbonatitos de siderita, considerada uma das maiores reservas mundiais de nióbio. Essa ocorrência geológica e pedológica altamente intemperizada torna o local ideal para o estudo do processo pedogenético de lateralização e da variabilidade espacial dos elementos químicos. O objetivo do presente estudo foi investigar a influência de várias combinações de amostragem (cenários) derivadas de três projetos de amostragem nas previsões espaciais de compostos químicos (Al_2O_3 , Fe_2O_3 , MnO , Nb_2O_5 , TiO_2 e SiO_2), obtidos por algoritmos de aprendizado de máquina combinados com imagens de sensoriamento remoto, bem como comparar o desempenho desses métodos por meio do teste de Kruskal-Wallis e posthoc. O conjunto de dados compreendeu 341 amostras do Serviço Geológico do Brasil (CPRM). As covariáveis incluíram dados de sensoriamento remoto coletados do Sentinel-2 MSI, Sentinel-1A e Radiômetro Espacial Avançado de Emissão e Reflexão Térmica (*Advanced Spaceborne Thermal Emission and Reflection Radiometer* - ASTER), e atributos topográficos calculados a partir de um modelo digital de elevação de 20 m derivado de dados hidrológicos (*Digital Elevation Model Derived from Hydrologic* - HC-DEM). Algoritmos de aprendizado de máquina (Modelos Lineares Generalizados com Regularização Elástica (*Generalized Linear Models with Elastic Net Regularization* - GLMNET), *Nearest Neighbors* (KNN), *Neural Network* (NNET), *Random Forest* (RF) e *Support Vector Machine* (SVMRadial) foram usados em combinação com covariáveis e elementos medidos em locais pontuais para mapear espacialmente a concentração desses elementos químicos. As covariáveis ideais para modelagem foram selecionadas usando *Recursive Feature Elimination* (RFE), processando 10 execuções para cada elemento químico. Os modelos RF, SVMRadial e KNN tiveram melhor desempenho, seguidos pelos modelos do grupo de redes neurais (NNET). Os diferentes cenários de amostragem não foram significativamente diferentes com base no erro quadrático médio; no entanto, diferenças significativas foram observadas para o coeficiente de determinação em todos os modelos. Os atributos do terreno foram significativamente mais influentes nas previsões espaciais dos elementos contidos nas lateritas do que os índices espectrais de sensoriamento remoto, provavelmente devido à estrutura espacial subjacente das duas formações (laterita e talude) que ocorrem em diferentes elevações.

Palavras-chave: Pedometria. Aprendizado de máquina. Áreas de difícil acesso. Laterita. Otimização de amostragem.

4.2 ABSTRACT

Morro de Seis Lagos, a region in the Brazilian Amazon, hosts a rare (less than 1%) formation of siderite carbonatites, considered one of the world's largest Niobium reserves. This highly weathered geological and pedological occurrence makes the site ideal for studying pedogenetic process of lateralization and the spatial variability of the chemical elements. The objective of this study was to investigate the influence of various sampling combinations (scenarios) derived from three sampling designs on the spatial predictions of chemical compounds (Al_2O_3 , Fe_2O_3 , MnO , Nb_2O_5 , TiO_2 , and SiO_2), obtained by machine learning algorithms combined with remote sensing images, as well as to compare the performance of these methods using the Kruskal-Wallis and post hoc tests. The dataset comprised 341 samples from the Geological Survey of Brazil (CPRM). Covariates included remotely sensed data collected from Sentinel-2 MSI, Sentinel-1A, and Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), and topographic attributes calculated from a 20-m digital elevation model derived from hydrologic data (HC-DEM). Machine learning algorithms (Generalized Linear Models with Elastic Net Regularization (GLMNET), Nearest Neighbors (KNN), Neural Network (NNET), Random Forest (RF), and Support Vector Machine (SVMRadial)) were used in combination with covariates and measured elements at point locations to spatially map the concentration of these chemical elements. The optimal covariates for modeling were selected using Recursive Feature Elimination (RFE), processing 10 runs for each chemical element. RF, SVMRadial, and KNN models performed better, followed by the models from the neural network group (NNET). The different sampling scenarios were not significantly different based on root mean square error; however, significant differences were observed for the coefficient of determination across all models. The terrain attributes were significantly more influential on the spatial predictions of the elements contained in laterites than the remote sensing spectral indexes, likely due to the underlying spatial structure of the two formations (laterite and talus) occurring at different elevations.

Keywords: Pedometrics. Machine learning. Poorly accessible areas. Laterite. sampling optimization.

4.3 INTRODUCTION

Soil is the product of various forming factors (JENNY 1994; MCBRATNEY; SANTOS; MINASNY, 2003) that influence pedogenetic processes through weathering (chemical, physical and biological), and translocation of weathered materials over and through soil (RUNGE, 1973; SIMONSON, 1959). Pedogenetic processes as an expression of past and current soil environment are reflected by current observed features and measured soil properties (KÄMPF; CURI, 2012). These pedogenetic processes that leave records in the form of morphological, physical and chemical properties are used to differentiate various soil types. Further, the interactions of pedogenetic processes with lithological conditions define regions and pedogeomorphological units, that influence drainage patterns and hydrological functions (PINHEIRO JUNIOR et al., 2020).

Morro dos Seis Lagos is an example of a combination of a rare formation of siderite carbonatite with laterization/paleo-pedogenesis weathering processes. The formation is unique from the perspective of pedological, geomorphological and geological characteristics associated with the tropical climate, that leads to intense weathering, thus reshaping the landscape and controlling the mechanisms of soil formation processes in a tropical environment. In addition, this formation contains certain minerals that contains chemical elements of economic interest such as niobium (Nb), titanium (Ti), and tungsten (W), (GIOVANNINI et al., 2020). Due to intense weathering these minerals can concentrate in shallow surfaces which makes them attractive for mineral extraction (GIOVANNINI *et al.*, 2017; 2020). Niobium is known for properties such as resistance against chemical corrosion and high temperatures with many applications in construction material for ocean exploitation, space power equipment, pharmaceutical and chemical equipment (ZARDIACKAS; KRAAY; FREESE, 2006; BABAEI; FATTAH-ALHOSSEINI; CHAHARMAHALI, 2020). Niobium has a limited biological role in the human body; however, niobates and niobium chloride that are soluble in water are toxic and can induce damage to human DNA, causing immune cell mortality (CHEN; THOUAS, 2015).

Morro dos Seis Lagos is a Nb-rich rutile formation from the weathering of Nb-rich brookite associated with the formation of secondary pyrochlore, along with ceriopyrochlore (Ce), Nb-rich goethite and cerianite (Ce). With increasing depth, the volume of these minerals increases from 0.12% by weight in the pisolitic type to 3.5% by weight in the laterite type classified as "brown laterite", according to Giovannini et al. (2017). This unique formation allows for studying the laterization process under intense weathering conditions and the resulting spatial distribution and concentration of elements/minerals with economic potential value.

However, in many instances these rich formations occur in isolated areas with limited access. Remote sensing combined with field data and machine learning has been used successfully for high-quality predictions for digital soil mapping applications especially for hard-to-reach areas (FERREIRA et al., 2023; GELSLEICHTER et al., 2023). According to Wadoux, Brus e Heuvelink (2019), map accuracy depends in part on the number and location of sampling sites used for calibration and validation of machine learning models. The effect of sampling designs on property prediction for digital soil mapping is still in its infancy (WADOUX; BRUS; HEUVELINK, 2019). Designing proper sampling schemes is challenging

and depends on the study objectives whether be soil fertility campaigns, soil survey or other mapping efforts. Some of the major sampling designs in soil science are systematic grids for detailed soil fertility mapping; transects and/or purposive sampling for traditional soil survey campaigns, random stratified and/or complete randomized for ecological studies, spatial coverage like Conditioned Latin Hypercube (cLHC) (MINASNY; MCBRATNEY, 2006; ROUDIER, 2011) for DSM applications, among many other designs.

The *Morro dos Seis Lagos* sampling project is a combination of the projects that includes: *Projeto Seis Lagos* (VIEGAS FILHO; BONOW, 1976); *Projeto Uaupés* (JUSTO, 1983); and *Projeto Terras Raras* (TAKEHARA, 2019) summarized in the *Projeto Avaliação do Potencial de Terras Raras no Brasil - Área Morro dos Seis Lagos, Noroeste do Amazonas* (TAKEHARA, 2019). Each project has implemented unique sampling designs and density, thus offering a possibility to test the performance of sampling designs and models.

Morro dos Seis Lagos represents a unique situation for assessing different sampling designs on spatial accuracy predictions of minerals in the siderite carbonatite formation using the DSM approach. Under the DSM approach we focus on some of the most commonly used machine learning techniques or models such as Generalized Linear Models with Elastic Net Regularization (GLMNET) (FRIEDMAN; HASTIE; TIBSHIRANI, 2010); Neural Network (NNET) (RIPLEY, 2019); Nearest Neighbors (KNN) (HECHENBICHLER; SCHLIEP, 2004); and Random Forest (RF) (BREIMAN, 2001). Briefly, GLMNET was proposed by Friedman, Hastie e Tibshirani (2010) and fits a generalized linear model via likelihood, combining generalized linear models with elastic net models.

The Feed-forward Neural Network (NNET) is a neural network that uses a multilayer feed-forward perceptron architecture with a sigmoidal transfer function and logarithmic learning for error backpropagation and resilient propagation variation (RPROP). In this algorithm, the connection weights are adjusted by feedback until convergence is reached, which determines the backpropagation error.

K-Nearest Neighbor (KNN) is a non-parametric regression technique, based on the theory of weighting observations based on the values of their nearest neighbors. The algorithm consists of determining a classification or average value from among the k known objects (k nearest neighbors) that are closest to the unknown samples (MCROBERTS et al., 2007).

Random Forest (RF) is a technique based on trees generated from a random vector sampled independently, which obeys the same distribution as the other trees (BREIMAN, 2001). The best tree resulting from the average of the results of all the trees is selected. The model is considered robust as it is not very sensitive to noise in the data and overfitting, it provides measures of the importance of covariates in the prediction (HEUNG; BULMER; SCHMIDT, 2014) and can handle continuous and categorical data.

The Support Vector Machines Radial Sigma (SVM-R) algorithm consists of non-linear regressions usually obtained by applying the kernel trick (DIAS et al., 2021). The SVM-R model operates as a binary classifier, separating into two categories of samples in order to minimize empirical risk while maximizing confidence, its hyperparameters include the “kernel” and “C” (penalty) (DUAN et al., 2024). According to the author, SVM's distinctive feature is its ability to define complex decision functions that optimally separate two classes of data samples.

The objectives of this study were to evaluate the performance of 5 different machine

learning models, the influence of 5 sampling scenarios derived from 3 sampling designs, and their interaction on the accuracy of predictions of the main 6 elements/minerals (Al_2O_3 , Fe_2O_3 , MnO , Nb_2O_5 , TiO_2 and SiO_2) in the Morro dos Seis Lagos petroferric/lateritic formation. The following alternative hypotheses will be tested:

- i) The prediction of different chemical elements is influenced by the sampling design;
- ii) The performance of all the models will depend on the sampling design;
- iii) The model performance will differ for different chemical elements; and
- iv) The prediction of different chemical elements is influenced by the sampling design and the model (interaction).

4.4 MATERIAL AND METHODS

The methodological procedures are divided into three stages (Fig. 17). Stage one includes compilation of model inputs (geochemical data and covariates from topography and remote sensing). Stage two includes the selection of covariates through spearman correlation and recursive feature elimination, and modeling based on different sampling scenarios derived from the initial sampling designs. Stage three includes evaluation of model performance and spatial predictions of the selected elements.

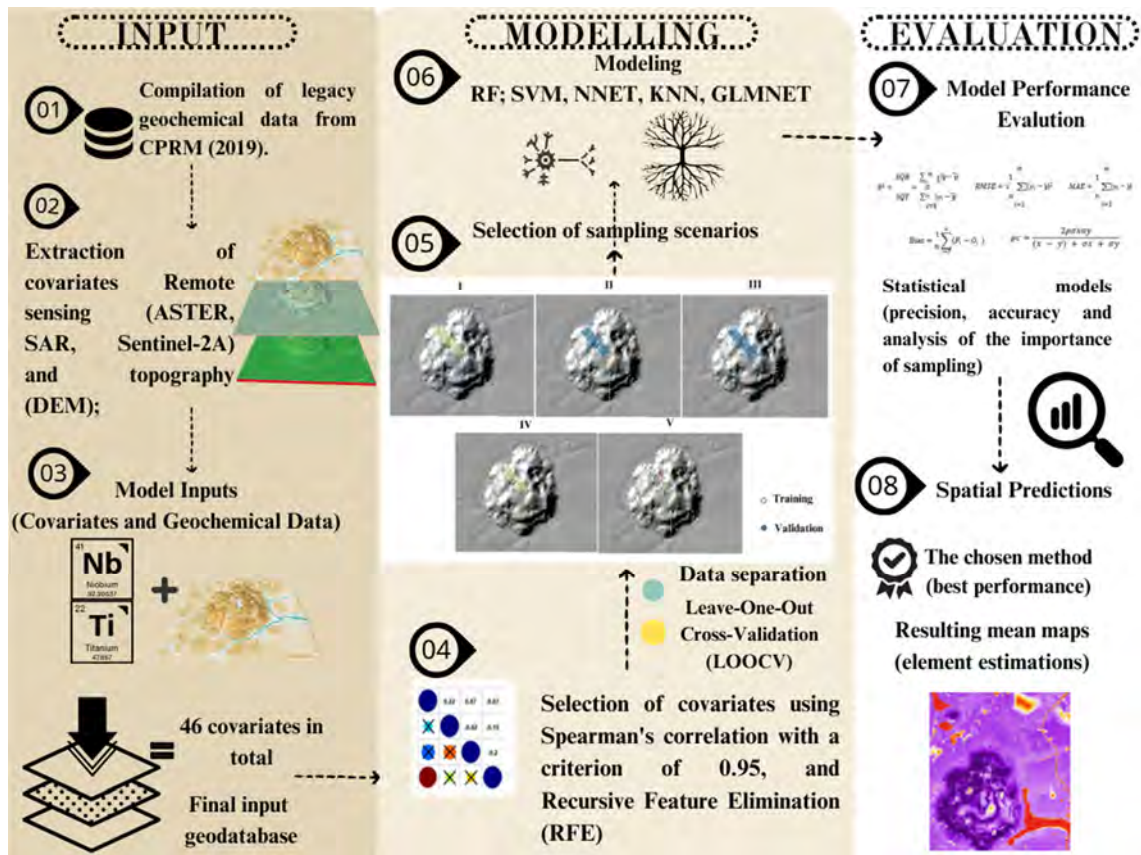


Figure 17. The methodological steps included: (1) Compilation of geochemical data from CPRM [22]; (2) Extraction of remote sensing covariates (ASTER, SAR, Sentinel-2A) and topographic data (DEM); (3) Harmonization of geochemical data and covariates into a single database; (4) Selection of covariates using Spearman's correlation (threshold 0.95) and Recursive Feature Elimination (RFE), with partitioning of the database into test and training and use of (Leave-One-Out Cross-Validation); (5) Definition of sampling scenarios; (6) Processing models (RF, SVM, NNET, KNN, and GLMNET); (7) Model performance evaluation based on precision, accuracy, and sampling design; and (8) Calculation of the average spatial predictions of elements. Source: Prepared by the author (2025).

4.4.1 Study area

The *Morro dos Seis Lagos* region (Fig. 18) is located in the northwestern part of the Amazon basin, (43° 12' 34.56" W, 22° 58' 12.34" S), with a total area of 35.96 km² bordered (<150 km) by Venezuela and Colombia. The area is within three conservation units (UCs)

(BRASIL, 2000) which overlap: the Balaio Indigenous Land (BRASIL, 2009), the Pico da Neblina National Park (BRASIL,1979) and the Morro dos Seis Lagos Biological Reserve (BRASIL, 2009).

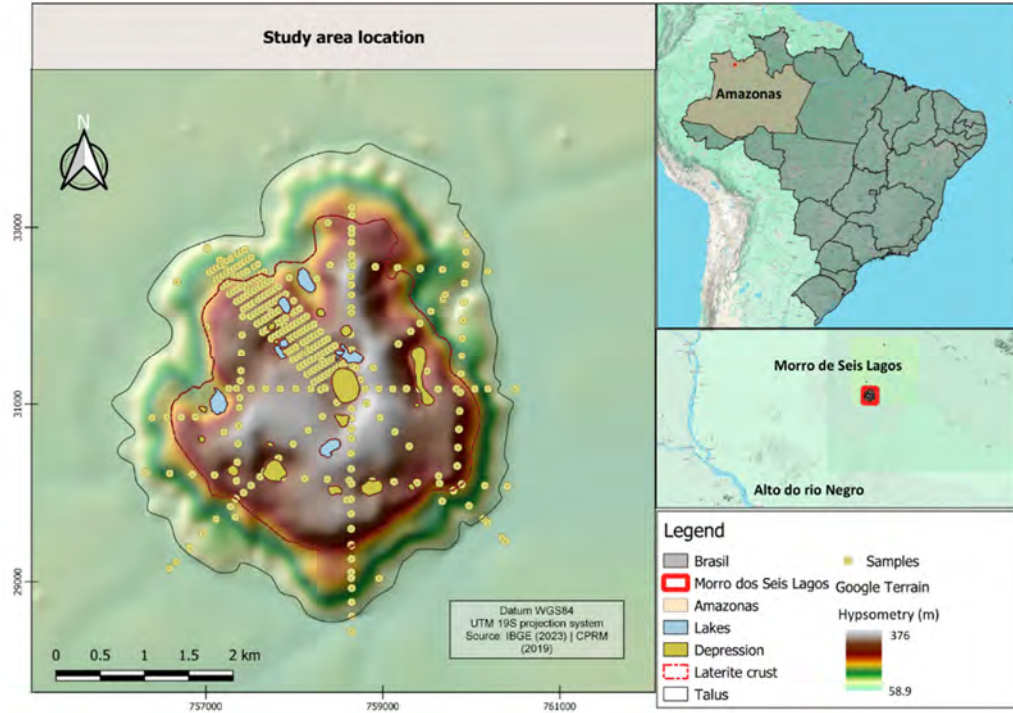


Figure 18. Location of the study area and the sampling point locations. Source: Prepared by the author (2025).

The soil geochemical data, sediments, and rock materials used in this study were made available by the Geological Survey of Brazil (CPRM) through the GeoSBG platform (available at: <https://geoportal.cprm.gov.br/>) and are summarized in the Rare Earth Elements Potential Assessment in Brazil - Morro dos Seis Lagos Area, Northwest Amazonas (TAKEHARA, 2009). The surface sampling (0-40 cm) grid consists of 341 points, where geochemical analysis was carried out by the SGS Geosol Laboratory (Belo Horizonte) for 56 elements. The samples were analyzed based on the following methods (TAKEHARA, 2019): SGS XRF79C for identification of the minerals by comparing the diffractogram of the ICSS-PDF database (International Center for Diffraction Data - Powder Diffraction file) for the principal oxides (provide the abbreviated formulas); Inductively Coupled Plasma Mass Spectrometry (ICP-MS) with Inductively Coupled Plasma Mass Spectrometry (ICP_OES) for determining 4 rare earth elements (REE) plus Nb, Sn, W, Y and 11 other elements; and ISE03A and B for the elements F and Cl. The summary statistics of the earth elements were carried out using R software (v. 4.1.3) and Rstudio (v. 2022.02.3 Build 492) (Table 4).

Table 4. Summary statistics of elements based on geologic formation.

Elements (%)	Geology	Mean	Median	Min	Max	SD	CV (%)	Cs	Ck
Al ₂ O ₃	Combined	3.28	1.67	0.19	58	6.4	195.9	5.3	31.4
	Laterite/Depressions	4.73	1.76	0.75	36.7	8.9	189.4	3.0	7.8
	Laterite	2.07	1.42	0.19	32.5	3.1	148.2	7.4	65.4
	Talus	6.31	2.71	0.26	58.0	10.5	165.6	3.1	9.6
Fe ₂ O ₃	Combined	67.0	77.4	0.12	95.1	22.6	508.0	1.7	1.9
	Laterite/Depressions	68.6	73.9	7.10	87.5	19.7	28.7	19.7	3.5
	Laterite	75.8	78.8	6.96	95.1	13.5	17.9	-2.1	6.3
	Talus	51.3	61.0	0.12	92.7	31	60.3	-0.4	-1.5
TiO ₂	Combined	2.93	1.42	bdl	29.9	4.1	16.7	2.8	9.9
	Laterite/Depressions	4.82	2.36	0.73	17.1	4.9	100.9	1.3	0.3
	Laterite	3.46	1.86	0.04	29.9	4.4	126.6	2.5	8.2
	Talus	1.25	0.50	bdl	18.3	2.3	182.1	5.0	33.1
Nb ₂ O ₅	Combined	0.74	0.39	0.00	4.02	0.9	0.8	1.4	1.5
	Laterite/Depressions	1.53	1.12	0.53	3.81	1.0	66.9	0.9	-0.4
	Laterite	0.81	0.49	bdl	4.02	0.9	111.8	1.2	0.8
	Talus	0.34	0.22	bdl	2.38	0.5	149.0	2.2	4.8
MnO	Combined	5.09	0.20	bdl	64.1	14	2.8	10.9	3.0
	Laterite/Depressions	0.41	0.06	bdl	4.51	1.1	272.6	3.1	8.1
	Laterite	1.62	0.17	bdl	63.8	6.8	416.5	6.5	47.0
	Talus	14.67	0.66	bdl	64.1	22.0	150.1	1.1	-0.4
SiO ₂	Combined	2.74	0.56	bdl	97.8	11.6	422.6	7.0	50.7
	Laterite/Depressions	2.95	0.56	0.12	36.5	9.3	314.3	3.1	8.5
	Laterite	0.84	0.46	bdl	26.7	2.2	259.7	9.3	95.9
	Talus	7.86	1.45	0.14	97.8	21.5	273.7	3.5	10.9

Med=median; Min=minimum; Max=maximum; SD=standard deviation; CV=coefficient of variation; CS=coefficient of asymmetry; and Ck = Kurtosis. bdl = below detection limit.

Source: Prepared by the author (2025).

4.4.2 Environmental covariates

The terrain attributes were derived from a Hydrologically Consistent Digital Elevation Model (HC-DEM), with a spatial resolution of 20 meters, obtained using an interpolator (Topo to Raster) in ArcGIS Desktop v. 10.6 (ESRI). The DEM was based on primary vector data of contour lines with an equidistance of 10 meters, elevation points, hydrography, extracted from the cartographic base of the Brazilian Institute of Geography and Statistics (IBGE) - scale 1:25.000. The following terrain attributes derived from the HC-DEM were obtained using the program System for Automated Geoscientific Analyses- SAGA-GIS v.2.1.2 (CONRAD, 2007): Aspect, Real Surface Area, Convergence Index, Curvature Flow line, Curvature features

(General, Maximum, Minimum , Planar , Prolife, Tangencial, and Total), Elevation, Mid Slope Position, Multiresolution Index of Valley Bottom Flatness (MRVBF), Multiresolution Index of The Ridge Top Flatness (MRRTF), SAGA Wetness Index, Vector Terrain Ruggedness (VRM), Standardized Height, Slope Height, Terrain Surface Convexity, Terrain Surface Texture, Valley Depth, Closed Depressions, Hill, Hill Index, Slope Index, Surface Specific Points, Valley and Valley Index.

The covariates from remote sensing data were obtained by the Multispectral Sensor Instrument (MSI) sensor of the Sentinel-1, 2A satellite and the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) sensor of the Terra satellite at a spatial resolution of 20 meters. Several indices were generated by combining spectral bands from Sentinel 2A to identify: Ferrous silicates, Altered zones (Band 8/ band 11) Sora et al. (2018), Hydrothermally altered rocks, Lateritic clay minerals, Iron oxide, Normalized difference vegetation index (NDVI), and Ferrous regolith. For the Sentinel 1A (SAR), the indices were Vertical transmit-vertical receive (VV) and Vertical transmit-horizontal receive (VH), while for the ASTER sensor advanced clay alteration (Band 4 /Band 6) Sora et al. (2018), Ferric iron, Fe^{3+} , Ferrous iron, Fe^{2+} , and Gossan were identified.

4.4.3 Selection of environmental covariates

The terrain attributes were derived from a Hydrologically Consistent Digital Elevation Model (HC-DEM), with a spatial resolution of 20 meters, obtained using an interpolator (Topo to Raster) in ArcGIS Desktop v. 10.6 (ESRI). The DEM was based on primary vector data of contour lines with an equidistance of 10 meters, elevation points, hydrography, extracted from the cartographic base of the Brazilian Institute of Geography and Statistics (IBGE) - scale 1:25.000. The following terrain attributes derived from the HC-DEM were obtained using the program System for Automated Geoscientific Analyses- SAGA-GIS v.2.1.2 (CONRAD, 2007): Aspect, Real Surface Area, Convergence Index, Curvature Flow line, Curvature features (General, Maximum, Minimum , Planar , Prolife, Tangencial, and Total), Elevation, Mid Slope Position, Multiresolution Index of Valley Bottom Flatness (MRVBF), Multiresolution Index of The Ridge Top Flatness (MRRTF), SAGA Wetness Index, Vector Terrain Ruggedness (VRM), Standardized Height, Slope Height, Terrain Surface Convexity, Terrain Surface Texture, Valley Depth, Closed Depressions, Hill, Hill Index, Slope Index, Surface Specific Points, Valley and Valley Index.

The covariates from remote sensing data were obtained by the Multispectral Sensor Instrument (MSI) sensor of the Sentinel-1, 2A satellite and the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) sensor of the Terra satellite at a spatial resolution of 20 meters. Several indices were generated by combining spectral bands from Sentinel 2A to identify: Ferrous silicates, Altered zones (Band 8/ band 11) Sora et al. (2018), Hydrothermally altered rocks, Lateritic clay minerals, Iron oxide, Normalized difference vegetation index (NDVI), and Ferrous regolith. For the Sentinel 1A (SAR), the indices were Vertical transmit-vertical receive (VV) and Vertical transmit-horizontal receive (VH), while for the ASTER sensor advanced clay alteration (Band 4 /Band 6) Sora et al. (2018), Ferric iron, Fe^{3+} , Ferrous iron, Fe^{2+} , and Gossan were identified.

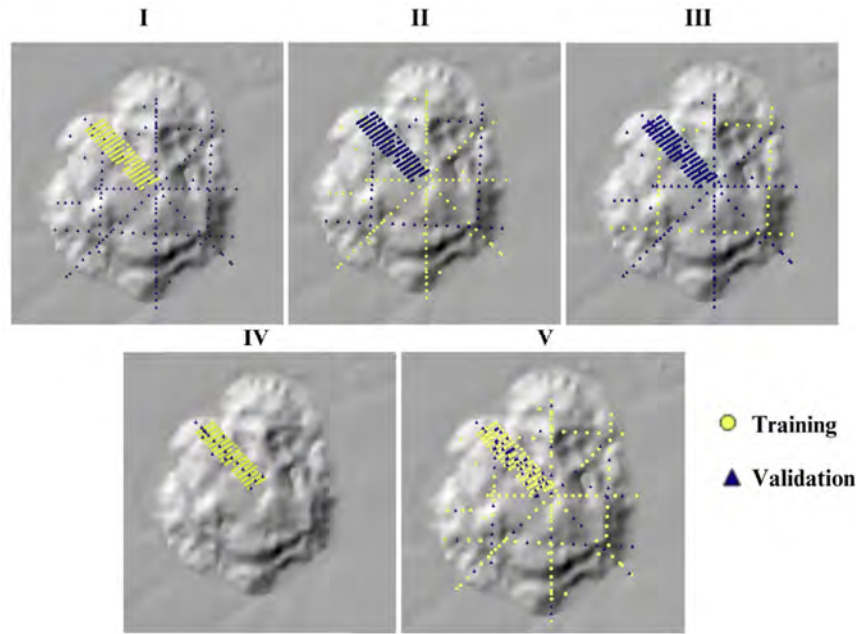
4.4.4 Modelling

Predictive Models differ due to the different management of the training and validation data sets, and selection of different covariates based on their importance. The following models for selected elements (Al_2O_3 , Fe_2O_3 , MnO , Nb_2O_5 , TiO_2 and SiO_2) were tested: Generalized Linear Models with Elastic Net Regularization (GLMNET); Neural Network (NNET); Nearest Neighbors (KNN); Random Forest (RF); and Support Vector Machine (SVMRadial).

The GLMNET algorithm was implemented via the `glmnet` package (FRIEDMAN; HASTIE; TIBSHIRANI, 2010), with hyperparameters α (elastic net mixing parameter: 0.1, 1, 10) optimized via grid search with 10-fold repeated cross-validation (3 repeats) using the `caret` package (KUHN, 2008). The Feed-forward neural network (NNET) was implemented via the `nnet` package (RIPLEY, 2019), with hyperparameters size (hidden units: 5, 10, 15) and decay (weight decay: 0.01, 0.001, 0.0001) optimized via the same `caret` procedure. The model was carried out in two steps: "one step forward" (propagation) and "one step back" (backpropagation) (SILVEIRA et al., 2013). The K-Nearest Neighbor (KNN) was implemented via the `caret` package (KUHN, 2008), where the hyperparameter k (neighbors 3,5,7) was optimized. Random Forest (RF) was implemented via the `caret` with the `randomforest` package (LIAW; WIENER, 2002), where three hyperparameters: i) `ntree` (number of trees: 500), ii) `nodesize` (minimum node size: 1), iii) `mtry` (number of variables sampled: 2,4,6). The Support Vector Machines Radial Sigma (SVM-R) algorithm was implemented via the `caret` with the `kernlab` package (KARATZOGLOU; SMOLA; HORNIK, 2016), with hyperparameters: `sigma` (fixed at 0.1, 1, 10) and `C` (cost: 1, 10, 100) adjusted.

4.4.5 Sampling scenarios

The surficial samples were collected based on three transect designs. Two designs consisted of points collected across the site in diagonal and perpendicular fashion (D&P) and around the site in box fashion (B). The third design consisted of transects with high point density (D) collected in a small area. To evaluate the influence of sampling design several sampling scenarios consisting of combinations of D&P, B and D were tested where one design serves as a training set while the other as a validation set. We tested five scenarios (Fig. 19) thus, the data was separated according to scenarios 1, 2, 3, 4, and 5 into training and validation. The LOOCV method was applied to the training set in order to validate the model.



	Sampling Designs		
	D&P	B	D
Scenario 1	Validation	Validation	Training
Scenario 2	Validation	Training	Training
Scenario 3	Validation	Training	Validation
Scenario 4*	Validation	Validation	Training
Scenario 5	Validation/Training	Validation/Training	Validation/Training

* Validation data from D&P and B falling within D are used only

Figure 19. Different scenarios for a combination of sampling designs as training and validation sets for modeling. Source: Prepared by the author (2025).

In scenario I (Fig.19), the models were trained using only the dense sampling grid (D) and validated with diagonal and perpendicular transect plans (D&P) plus the box-shaped transect (B). In scenario II, training combined the dense grid (D) and the box-shaped transect (B), while validation was performed exclusively with the diagonal and perpendicular transects (D&P). In scenario III, the opposite configuration was used: training was based on the box-shaped transect (B), while validation used both the dense grid (D) and the D&P transects. Scenario IV used the dense grid (D) for training, but validation was restricted to the D&P and B points located within the spatial extent of D, ensuring that training and validation were confined to the same geographic area. And scenario V integrated all sampling plans (D, D&P, and B), randomly dividing them into 70% for training and 30% for validation, using the Leave-One-Out cross-validation (LOOCV) approach.

4.4.6 Model performance evaluation

The model's performance was evaluated using the following metrics: Root Mean Squared Error (RMSE) (eq. (1)), Mean Absolute Error (MAE) (eq. (2)), Coefficient of determination (R^2) (eq. (3)), Concordance correlation coefficient (ρ_c) (eq. (4)), and Bias (eq. (5)).

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

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

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

$$\rho_c = \frac{2\rho_{xy}}{(x - y) + \sigma_x + \sigma_y} \quad (4)$$

$$Bias = \frac{1}{n} \sum_{i=0}^n (y_i - \hat{y}_i) \quad (5)$$

Each model was run 10 times (nrep=10) for each element, using fixed train/test partitions derived from three distinct sampling designs (Designer 1 to Designer 4), with no random resampling of training/ validation samples. From each repetition, Recursive Feature Elimination (RFE) was first performed using 10-fold repeated cross-validation (5 repeats) to select the optimal number of predictors (subsets: 2-25, 30, 35, 40, 46 variables), followed by hyperparameter tuning via grid search with 10-fold repeated cross-validation (3 repeats) using caret package (KUHNS, 2008). This approach generates an average map and the range repetitions for each element (predicted minimum and maximums), following the procedure adopted by Reis et al. (2021) and Moquedace et al. (2024) to quantify prediction variability. All steps were

implemented in R 4.4.1 using parallel processing across 15 CPU cores.

4.4.7 Comparison criteria and statistical analysis

The scenarios (I to V), chemical elements, and performance metrics (RMSE, MAE, R^2) obtained by the Random Forest, KNN, SVMRadial, MARS, and NNET models were compared using the nonparametric Kruskal-Wallis test, considering the results of 10 repetitions performed during the processing of the predictive models.

When significant differences were detected ($p < 0.05$), multiple comparisons were performed using Dunn's test, with Holm correction for type I error control. The results were expressed in Compact Letter Display (CLD) format, in which identical letters indicate statistically similar performance ($p > 0.05$), while different letters represent significant differences between scenarios or models ($p \leq 0.05$), allowing statistical performance to be ranked.

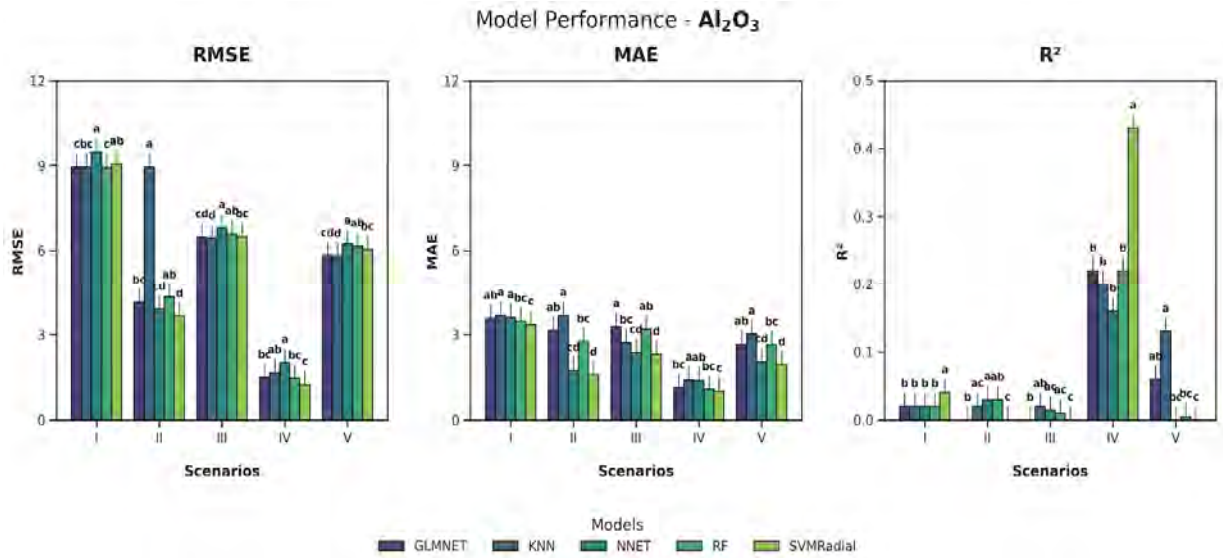
All statistical analyses were conducted in the RStudio environment (R Core Team, 2023), using the `dunn.test`, `rstatix`, and `multcompView` packages.

4.5 RESULTS

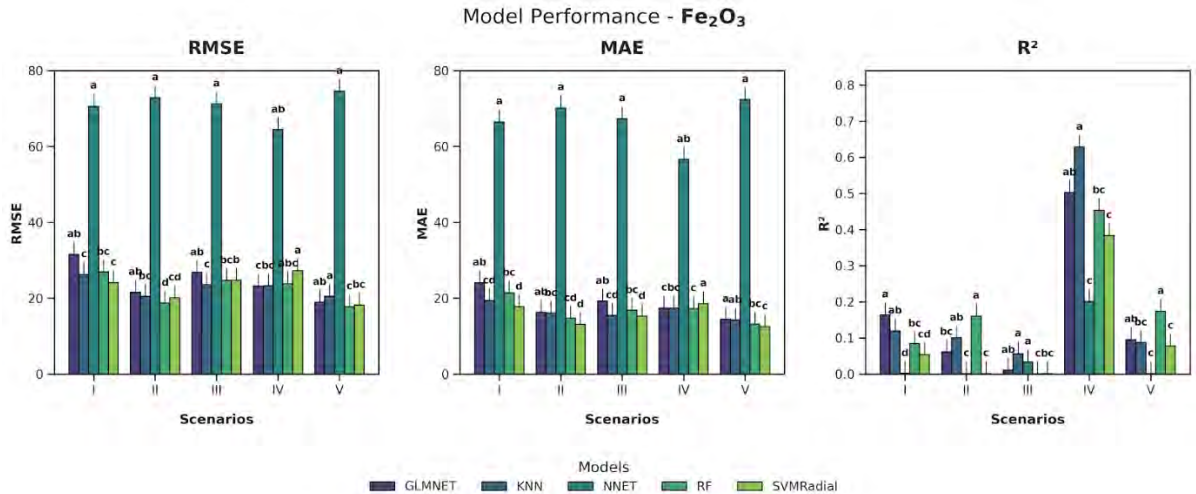
4.5.1 Factors influencing performance

Figure 20 (Appendix C) details the performance of predictive models by chemical element and scenario, considering the medians of the RMSE, MAE, and R^2 metrics, in addition to the statistical groupings obtained by the Kruskal-Wallis and Dunn tests. Although the Random Forest, SVMRadial, and KNN models show the best results most frequently, they vary according to the scenario and spatial behavior of the elements.

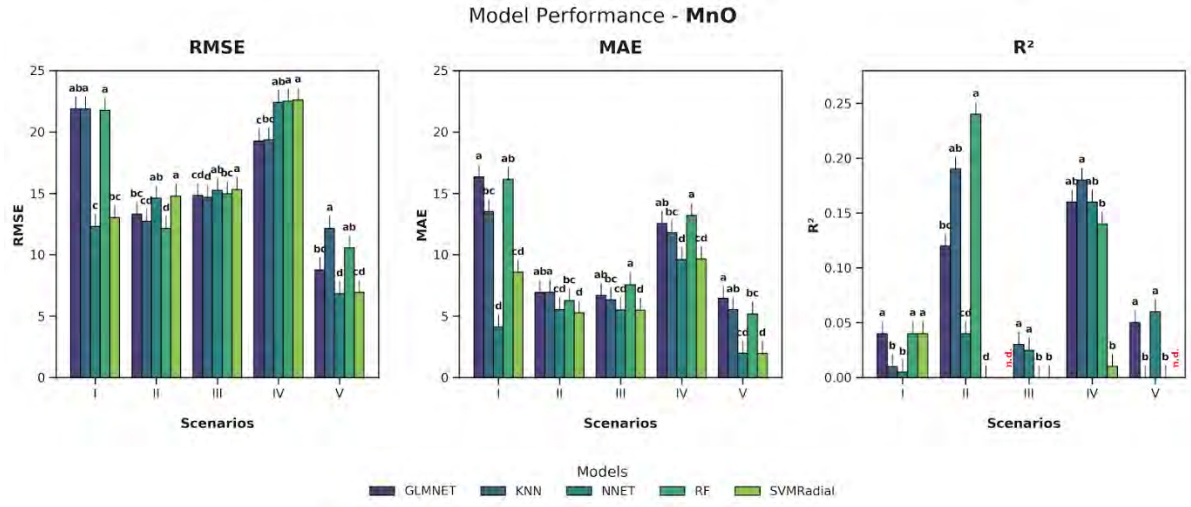
A



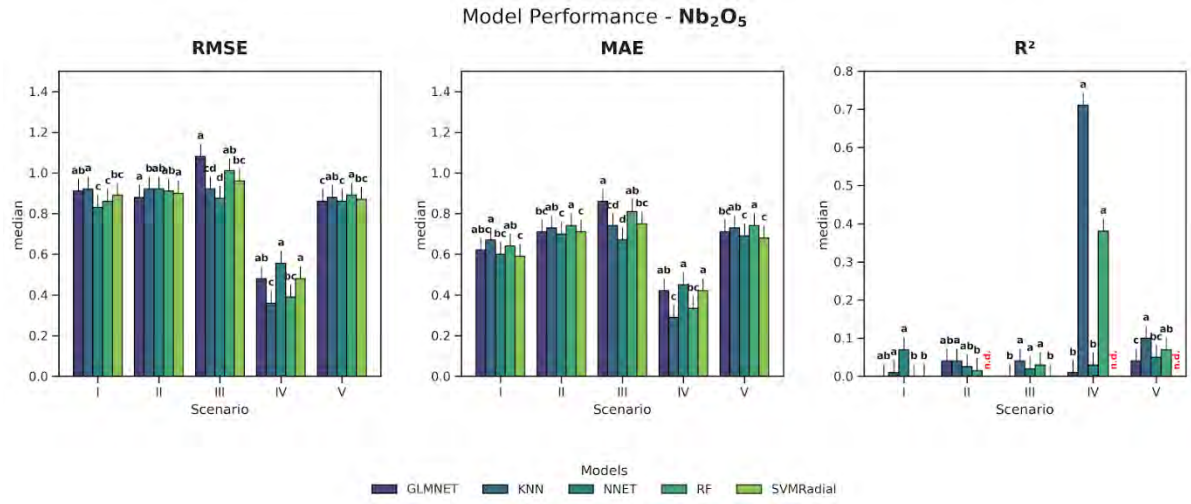
B



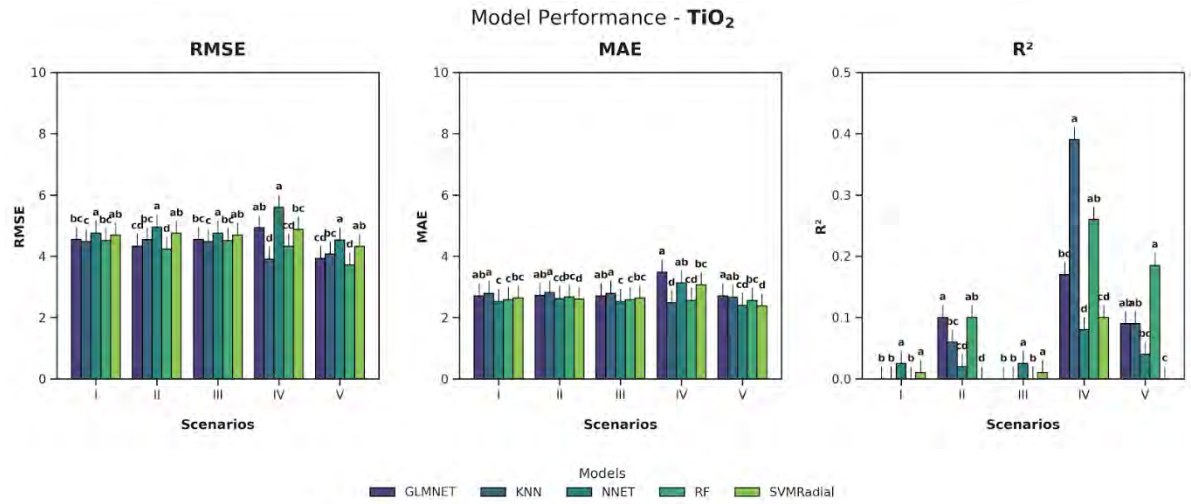
C



D



E



F

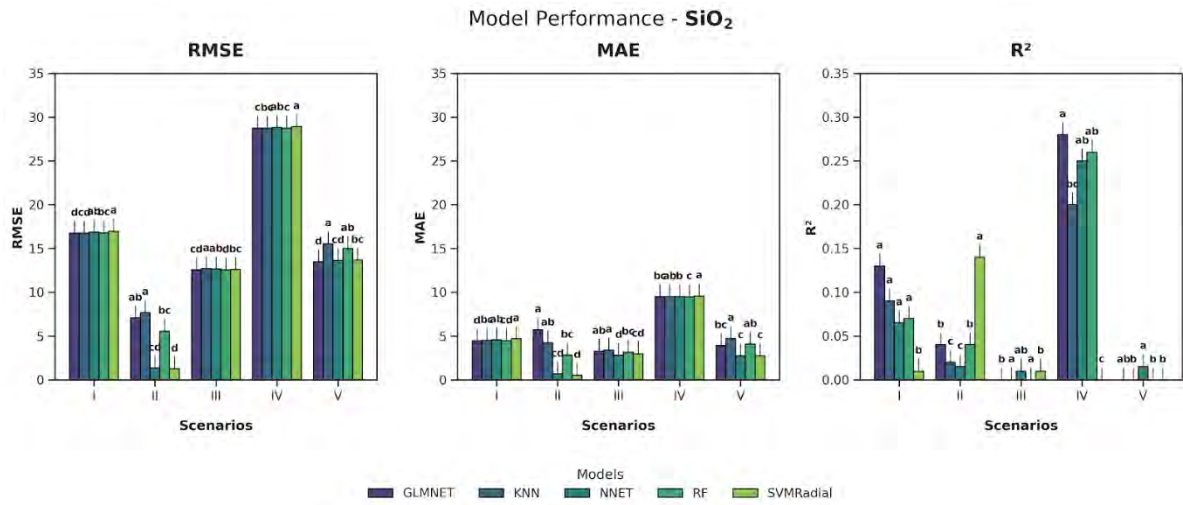


Figure 20. Predictive performance of the SVMRadial, RF, KNN, GLMNET, and NNET models by sampling scenario (Ia V) for the chemical elements: A) Al₂O₃; B) Fe₂O₃; C) MnO; D) Nb₂O₅; E) TiO₂; F) SiO₂, evaluated by the RMSE, MAE, and R² metrics. Groups with the same letters indicate statistically equivalent performance, according to the Kruskal-Wallis test followed by Dunn's post-test ($p \leq 0.05$). ND indicates absence of detection of the chemical element. Source: Prepared by the author (2025).

Overall, scenario IV (“c”) had the best performance with sample optimization for geochemical modeling, with superior performance for most elements (Al₂O₃, Fe₂O₃, SiO₂, Nb₂O₅, and TiO₂), combining lower MAE and RMSE values with higher R² values (Fig. 20A, B, D, E, and F). In this scenario, SVMRadial showed the best fit for Fe₂O₃ (e.g., MAE= 18.5a; RMSE= 27.2a) (Fig. 20B), followed by the Random Forest model (MAE= 9.4a SiO₂; Fig. 20F) and KNN (MAE= 17.4a Fe₂O₃; Fig. 20F), indicating the good ability of these models to capture nonlinear patterns associated with the spatial variability of oxides.

Scenario V (“b”) was also relevant in sample optimization, especially for MnO, with the lowest absolute errors (MAE = 6.2a SVMRadial; RMSE = 7.8a SVMRadial and 8.4b (RF) (Fig. 20C), surpassing scenarios I to IV). The KNN model stands out mainly for Fe₂O₃ (scenario IV -c), achieving MAE= 16.0a, RMSE= 19.5a, and R²= 0.45b, in addition to good performance with the elements Al₂O₃ and TiO₂ (group “a/b”), demonstrating superiority for oxides.

Scenarios I (“a”) and III (“d”) showed the worst overall performance in predicting the elements, with significantly high MAE and RMSE medians ($p < 0.05$; Kruskal-Wallis and Dunn) for Fe₂O₃ (MAE = 24.1ab GLMNET - Scenario I; RMSE = 31.6ab Scenario I); SiO₂ (MAE = 4.5d GLMNET- Scenario I) and Nb₂O₅ (MAE = 0.67a KNN- Scenario I; R² = ≤ 0.07) in all models. These scenarios did not perform well in capturing spatial variability, probably due to the low point density and irregular distribution (Fig. 19), resulting in unfavorable statistical groupings (group c/d- Fe₂O₃, SiO₂, Nb₂O₅) and R² < 0.20 for most elements.

The subsampling of scenarios I and III hindered the capture of patterns in the transition zone in lateritic crust and talus, both of high Fe content. This corroborated with an increase in error in metrics such as MAE (24.1ab GLMNET in Scenario I versus 17.4c KNN in Scenario IV for Fe₂O₃) and RMSE (31.6ab GLMNET Scenario I versus 23.3 bc KNN Scenario IV for Fe₂O₃) (Figure 20B).

4.5.2. Accuracy prediction comparisons among scenarios

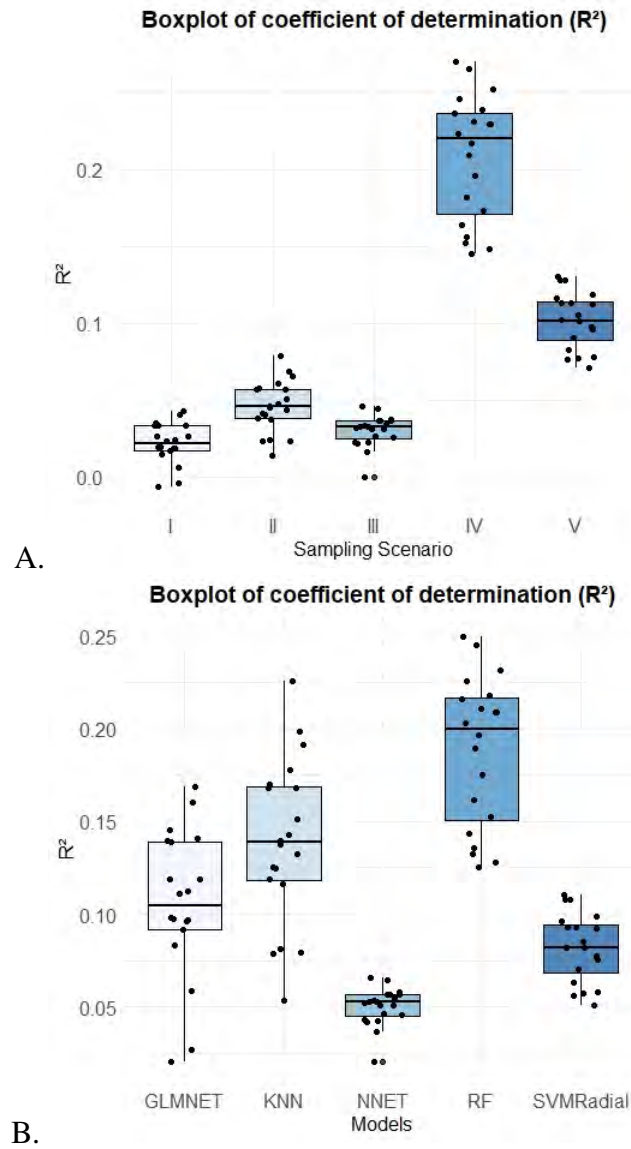
Overall, there were differences in model performance parameters among scenarios for all elements (Table 6). The average coefficient of determination (R^2) varied significantly, as did the MAE. All models slightly overpredicted elements concentrations regardless of scenarios.

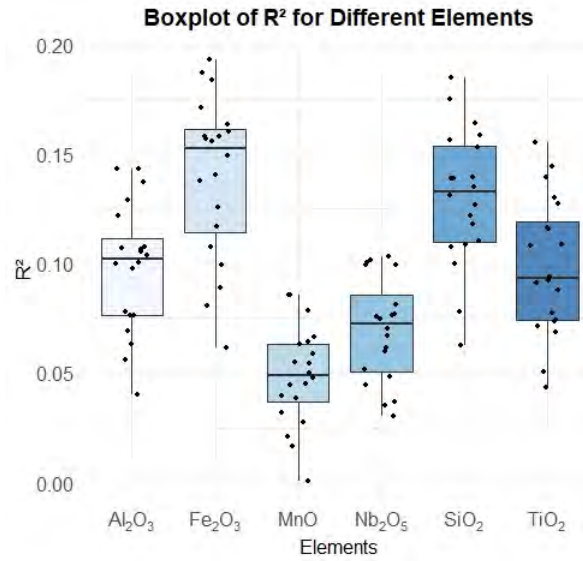
Table 5. Mean of performance parameters among sampling scenarios for all models combined.

Element (%)	Scenario	RMSE	MAE	R^2	RMSE	MAE	R^2	Pc	Bias
Training		Validation							
Al ₂ O ₃	I	1.99	1.20	0.33	9.08	3.54	0.02	0.11	-2.33
	II	9.58	4.56	0.46	4.22	2.49	0.02	0.07	0.24
	III	6.63	3.04	0.48	6.57	2.79	0.01	0.04	-0.35
	IV	2.02	1.22	0.31	1.59	1.21	0.24	0.32	-0.02
	V	6.63	2.84	0.43	6.02	2.47	0.04	0.15	-0.25
	Mean	5.37	2.57	0.40	5.50	2.50	0.07	0.14	-0.54
Fe ₂ O ₃	I	27.82	23.27	0.44	35.80	29.72	0.09	0.24	16.85
	II	33.92	27.19	0.23	30.79	26.12	0.07	0.22	15.87
	III	25.21	22.17	0.39	34.20	26.86	0.02	0.95	-9.96
	IV	28.50	23.74	0.39	32.40	25.46	0.44	0.53	-4.39
	V	30.43	24.84	0.30	30.01	25.33	0.09	0.25	17.29
	Mean	29.18	24.24	0.35	32.64	26.70	0.14	0.44	12.87
MnO	I	12.13	5.84	0.50	18.21	11.71	0.03	0.16	6.52
	II	13.64	5.87	0.42	13.51	6.17	0.12	0.27	-2.18
	III	7.03	3.19	0.32	15.00	6.31	0.02	0.11	-3.38
	IV	12.73	5.90	0.58	21.09	11.46	0.13	0.32	-2.46
	V	13.51	6.38	0.59	9.05	4.24	0.03	-0.21	1.88
	Mean	11.81	5.44	0.48	15.37	7.98	0.07	0.13	0.07
Nb ₂ O ₅	I	0.86	0.65	0.34	0.89	0.63	0.02	0.06	-0.12
	II	0.70	0.54	0.47	0.91	0.72	0.03	0.16	0.05
	III	0.84	0.59	0.38	0.96	0.77	0.02	0.02	0.10
	IV	0.86	0.66	0.47	0.46	0.38	0.26	0.29	-0.01
	V	0.83	0.62	0.49	0.87	0.71	0.07	0.25	0.09
	Mean	0.82	0.61	0.43	0.82	0.64	0.08	0.16	0.02
TiO ₂	I	4.59	2.81	0.32	3.76	2.69	0.03	0.13	0.07
	II	3.28	2.05	0.28	4.57	2.69	0.05	0.20	-0.95
	III	2.60	1.73	0.30	4.61	2.66	0.01	-0.26	-0.88
	IV	4.49	2.73	0.31	4.72	2.94	0.21	0.43	-1.48
	V	4.04	2.45	0.46	4.12	2.54	0.08	0.25	-0.66
	Mean	3.80	2.36	0.33	4.36	2.71	0.07	0.15	-0.78
SiO ₂	I	3.64	2.26	0.34	4.47	2.71	0.08	0.15	-0.95
	II	16.30	7.39	0.27	4.59	2.80	0.05	0.68	1.82
	III	2.33	1.21	0.29	12.61	3.12	0.01	0.01	-1.44
	IV	0.70	0.38	0.36	28.79	9.51	0.20	0.39	-9.26
	V	10.21	2.77	0.51	14.26	3.65	0.00	0.04	-0.98
	Mean	6.64	2.80	0.35	12.94	4.36	0.07	0.25	-2.16

Source: Prepared by the author (2025).

The performance parameters decreased substantially between training and validation for all elements and scenarios. However, the decrease was less for Scenario IV. Scenario IV performed consistently better across all elements with R^2 values varying from 0.13 (MnO) to 0.44 (Fe_2O_3) followed by Scenario III for Fe_2O_3 ($R^2 = 0.02$) and MnO ($R^2 = 0.02$) (Fig. 21A). Scenario II performed worst across all elements with R^2 values varying from 0.00 to 0.12.





C.

Figure 21. Coefficient of determination (R^2) (A) among scenarios for all models and elements combined; (B) among models for all scenarios and elements combined; (C) among elements for all models and scenarios combined. Source: Prepared by the author (2025).

Similar trends were observed for the other performance parameters for Scenario IV, however with some noticeable exceptions. For example, for Scenario IV the RMSE (21.1%) and bias (-2.46) for MnO were the highest though it had the largest R^2 (0.13). Similarly, the SiO_2 with the largest R^2 (0.20) had also the largest RMSE (28.8%), Pc (-9.3) and bias (-9.26). However, there was a greater variation among scenarios for each element with values ranging from 0.00 (Al and Fe_2O_3 , Scenario II) and (SiO_2 , Scenario V) to 0.44 (Fe_2O_3 , Scenario IV). The performance parameters varied also widely among models and elements within each scenario (Tables 5 and 6).

4.5.3 Accuracy predictions among models

Overall, all models performed poorly with R^2 values ranging from 0.00 (SVMRadial model; SiO_2) to 0.17 (GLMNET model; Fe_2O_3) (Table 4). The mean R^2 for Fe_2O_3 for all the models combined was 0.14 followed by other elements ranging from 0.07 (SiO_2) to 0.16 (TiO_2).

Table 6. Mean of performance parameters among models for all scenarios combined.

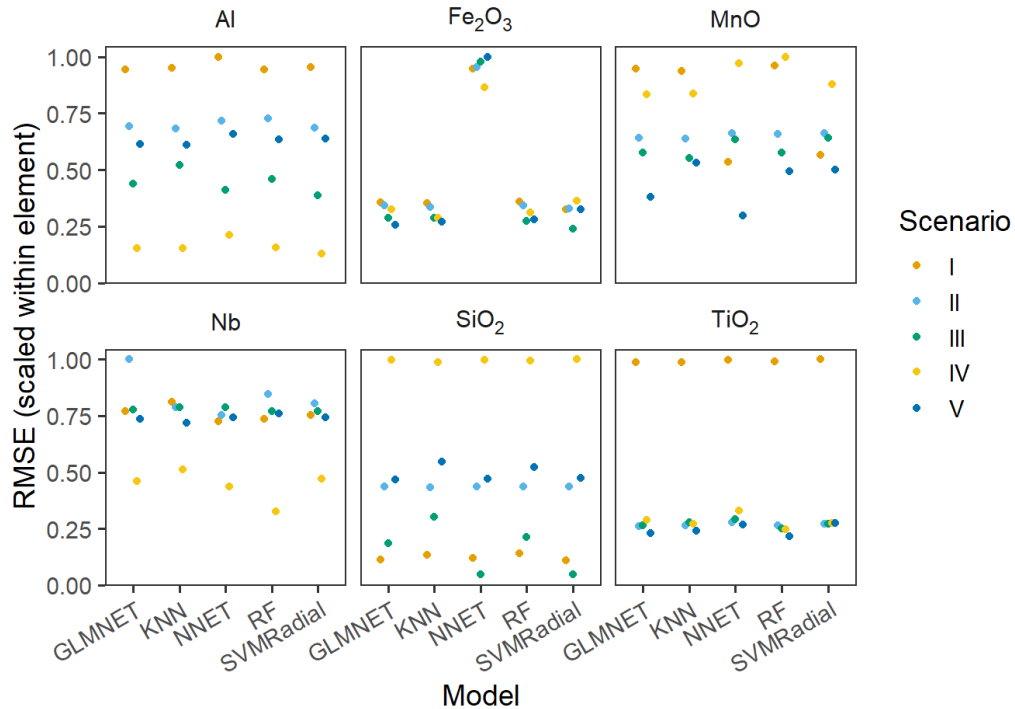
Element (%)	Model	RMSE	MAE	R ²	RMSE	MAE	R ²	Pc	Bias
Training				Validation					
Al ₂ O ₃	RF	5.16	2.62	0.12	5.51	2.64	0.06	0.19	0.20
	SVMRadial	5.30	2.21	0.64	5.30	2.05	0.09	0.17	-1.03
	NNET	5.82	2.46	0.17	5.69	2.24	0.05	-0.10	-2.06
	GLMNET	5.27	2.77	1.00	5.38	2.77	0.09	0.26	-0.10
	KNN	5.29	2.80	0.08	5.58	2.80	0.08	0.24	0.22
	Mean	5.37	2.57	0.40	5.50	2.50	0.07	0.15	-0.56
Fe ₂ O ₃	RF	17.36	12.47	0.34	22.37	16.57	0.18	-1.54	0.35
	SVMRadial	19.16	13.06	0.39	22.82	15.51	0.11	4.22	0.26
	NNET	71.85	68.53	0.04	70.78	66.61	0.05	-66.58	0.01
	GLMNET	18.42	13.63	0.76	24.41	18.26	0.17	1.00	0.31
	KNN	19.08	13.50	0.21	22.83	16.53	0.20	-0.57	0.40
	Mean	29.18	24.24	0.35	32.64	26.70	0.14	-12.69	0.27
MnO	RF	10.65	5.25	0.31	14.47	7.55	0.08	1.21	0.14
	SVMRadial	12.48	4.98	0.78	14.37	6.18	0.02	-3.08	0.09
	NNET	13.82	5.11	0.04	14.29	5.36	0.06	-4.44	0.13
	GLMNET	11.11	6.58	1.00	15.59	9.78	0.09	1.88	0.29
	KNN	11.37	5.48	0.26	16.15	8.83	0.08	1.96	0.22
	Mean	11.88	5.48	0.48	14.97	7.54	0.07	-0.50	0.18
Nb ₂ O ₅	RF	0.38	0.70	0.47	0.44	0.75	0.37	0.18	0.08
	SVMRadial	0.70	0.80	0.68	0.22	0.78	0.61	-0.01	-0.03
	NNET	0.26	0.78	0.50	0.22	0.78	0.50	0.10	-0.01
	GLMNET	0.97	1.29	1.02	0.21	1.50	0.93	0.11	-0.10
	KNN	0.27	0.78	0.51	0.34	0.77	0.51	0.33	0.07
	Mean	0.52	0.87	0.64	0.28	0.92	0.59	0.14	0.00
TiO ₂	RF	3.48	2.28	0.19	4.19	2.72	0.12	0.25	-0.11
	SVMRadial	3.80	2.32	0.32	4.38	2.63	0.02	0.07	-1.01
	NNET	4.33	2.34	0.04	4.66	2.52	0.04	0.18	-2.12
	GLMNET	3.20	2.55	1.19	3.88	3.33	0.51	-0.03	-0.62
	KNN	3.74	2.40	0.12	4.08	2.61	0.12	0.22	-0.17
	Mean	3.71	2.38	0.37	4.24	2.76	0.16	0.14	-0.81
SiO ₂	RF	6.11	2.69	0.19	15.73	4.82	0.07	0.69	-2.95
	SVMRadial	6.06	1.97	0.35	14.68	4.10	0.03	0.11	-3.58
	NNET	6.15	2.04	0.13	14.67	4.06	0.08	0.18	-3.55
	GLMNET	6.15	3.00	0.86	15.72	5.37	0.09	0.23	-1.94
	KNN	5.77	2.42	0.22	16.28	5.26	0.06	0.17	-2.13
	Mean	6.05	2.43	0.35	15.41	4.72	0.07	0.27	-2.83

Source: Prepared by the author (2025).

However, the performance parameters for Fe_2O_3 between the models varied more compared to the other elements. For example, the R^2 for Fe_2O_3 varied from 0.18 (RF) to 0.05 (NNET) compared to the other elements, with R^2 values varying from 0.00 (SVMRadial) to 0.11 (KNN). In general, for all scenarios and elements combined, the RF, GLMET and KNN models tied in terms of performance, followed by SVMRadial and NNET with the worst performance (Fig. 21B). However, no model performed consistently better for each element.

4.5.4 Interaction of models and sampling scenarios on accuracy predictions

Model performance varied by element and scenario (Fig. 22). For aluminum (Al_2O_3) performance was responsive to scenarios but insensitive to models. For iron (Fe_2O_3) model performance was somewhat sensitive to scenarios, however, the differences seem to be obscured by the much greater variability among models (NNET performs very poorly, and all other models perform about the same). For manganese (MnO) there was a similar performance between models, but a strong dependence on scenarios. However, there is some interaction where some models perform relatively better in some scenarios and worse in others. Niobium (Nb_2O_5) has a stronger dependence on scenarios than on models, with much better performance in scenario IV. Silicon (SiO_2) is also more dependent on scenarios, but in an opposite as the performance is very poor in Scenario IV but good for I and III. Lastly, titanium (TiO_2) is highly dependent on scenarios but only because Scenario I performed very poorly compared to all other scenarios.



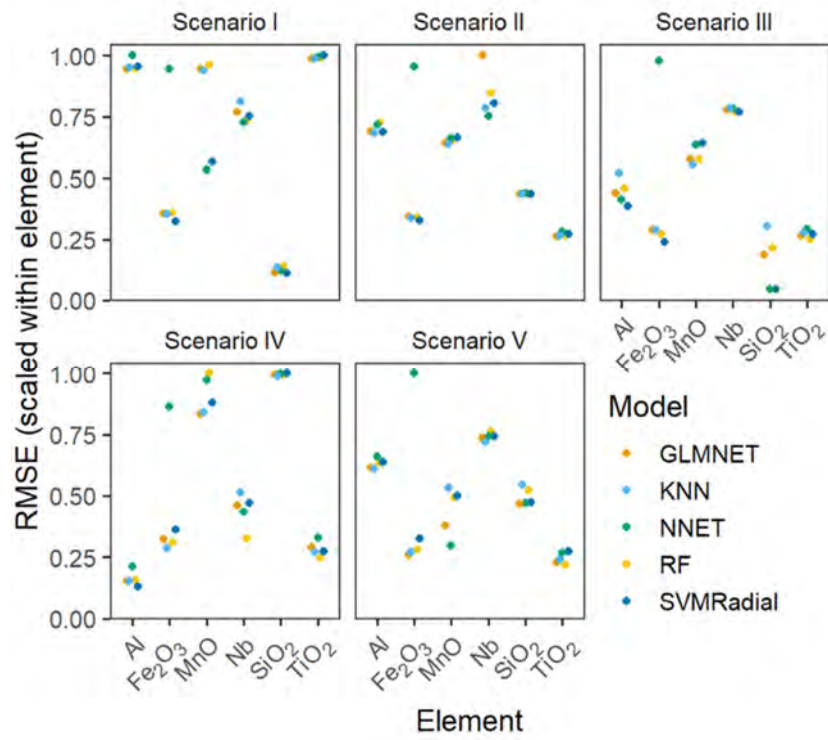


Figure 22. RMSE for each element by scenario and model (A) and for each sampling scenario by model and element (B). Source: Prepared by the author (2025).

4.6 DISCUSSION

At the beginning of our study, we formulated several hypotheses regarding the influence of sampling designs (5) and models (5) on the spatial predictions for selected elements (6). We postulated that the predictions of different chemical elements are not influenced by the sampling design (Hypothesis 1) and that the performance of all the models will be similar, regardless of the sampling design (Hypothesis 2). We further stated that the model performance will be similar for different chemical elements (Hypothesis 3) and that the predictions of different chemical elements are not influenced by the sampling design and the model (Hypothesis 4). We examine each hypothesis in light of the results from our analysis.

4.6.1. Sampling scenarios

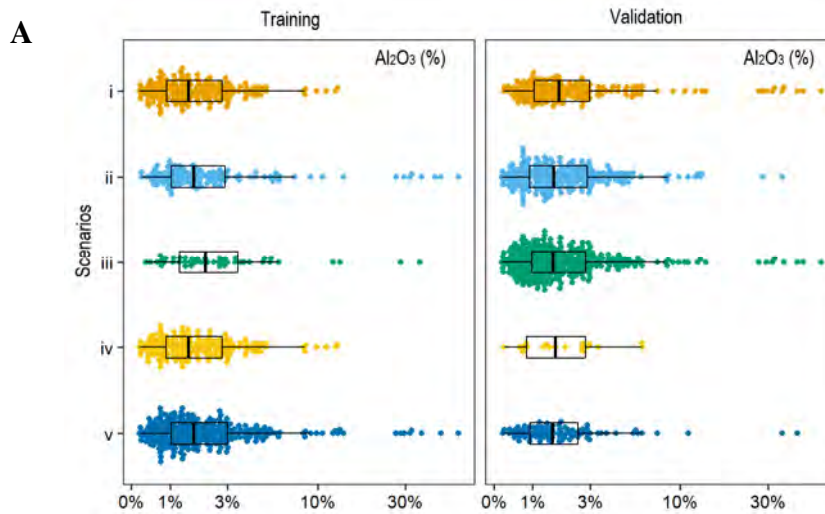
The sampling scenarios tested in this study sought to evaluate the representativeness of existing samples to capture the variability of the elements for spatial modeling based on the relationships between the soil and its environment. The location and distribution of existing sampling points for each set were different as they were designed to meet certain objectives of three different projects (TAKEHARA, 2019).

Scenarios I, II, and III did not perform well across all models as the combination of different sampling designs for training and validation did not equally capture the variability for any of the measured elements. For example, for Scenario I, models were trained on the high point density grid (D) but validated from the combination of diagonally and perpendicular (D&P) transects through the study site and surrounding it in a box square fashion (B) (Fig. 19). The difference in the distribution and density between the sample sets used for training (D) and validation (D&P and B) and their ability to capture the variability likely contributed to the poor performance of the models for Scenario I. A similar situation can also explain the poor performance of the models for Scenario II and III as the model training was based on B and D (Scenario II) and B (Scenario III) while validation was based on D&P for Scenario II and a combination of D&P and D for Scenario III. In other words, when training was based on the dense grid (D) the validation was based on transects (D&P and B) and vice versa. This resulted in uneven distribution of samples between training and validation sets. For example, for Scenario I there were 163 samples used for training and 173 for validation, while for Scenario II, 118 for training and 223 for validation. Scenario III was the most extreme case with 47 points used for training and 294 for validation, which explains among other things its worst performance. Scenario IV on the other hand was almost the reverse of Scenario III with 175 samples used for training and only 19 for validation and all confined to a small area, followed by Scenario V with 256 points used for training and 85 for validation but that were spread throughout the study area, contrary to Scenario IV.

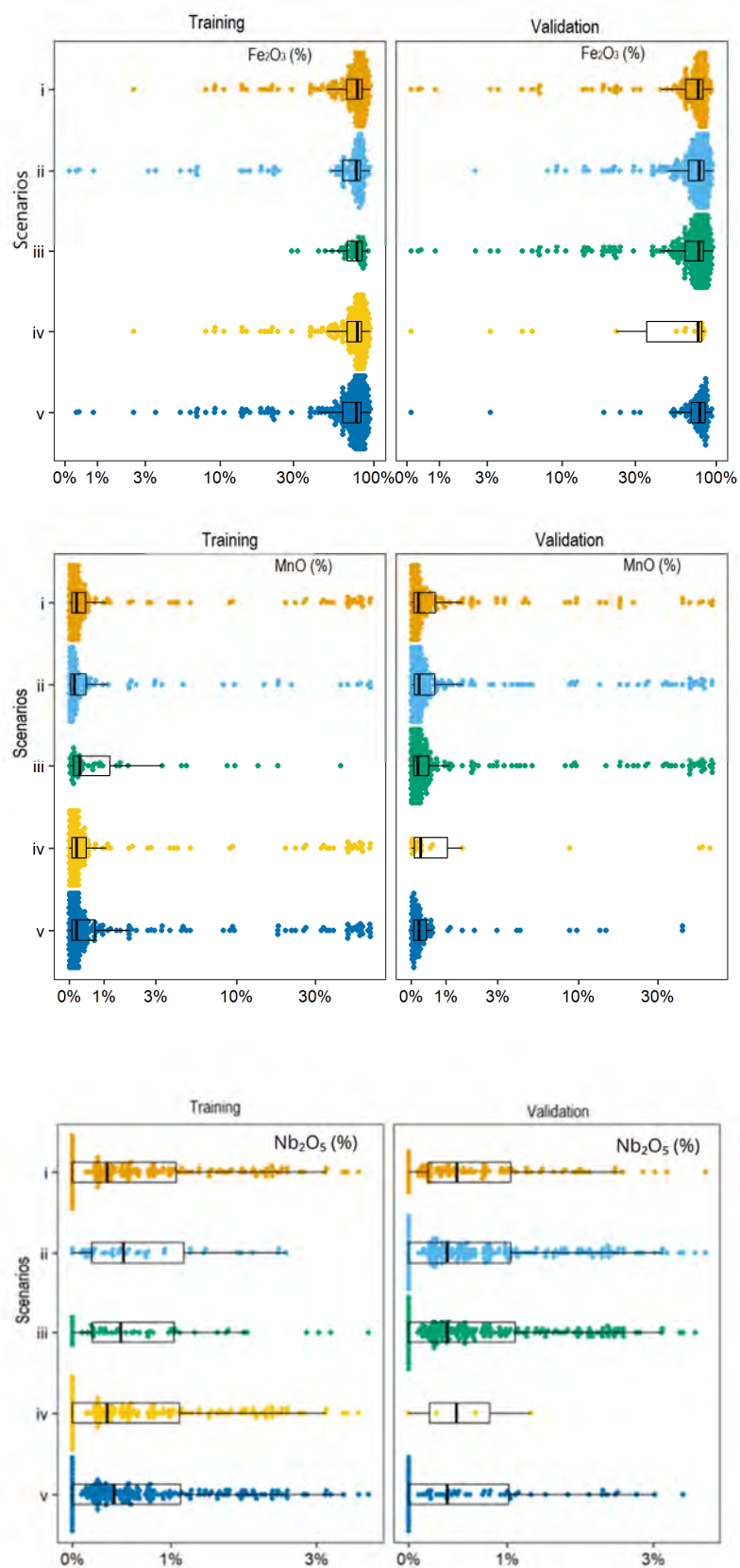
Transect sampling designs have been used successfully in many studies to explore the best spacing for capturing patterns in the landscape. According to the study by Rodrigues et al. (2020), accurate soil ECa maps were obtained by increasing transect spacing simulations by up to 150 m. For example, Grimm et al. (2008) used a random stratified sampling design with topography and geology as the stratifying variables to capture soil organic carbon (SOC) variations along a gradient of environmental covariates for spatial predictions using Random Forest and environmental covariates and obtained low R^2 values with an RMSE varying from

0.33 to 1.71%. The SOC spatial patterns reflected the catena sequence showing decreasing trends (toeslopes > ridges > midslope) proving to be effective for landscapes where environmental factors, especially topography, drive the distribution of SOC.

In our *Morro dos Seis Lagos* study area, the measured elements at different point locations whether based on dense grids (D) or two types of transects (D&P and B) did not show any correlations with the topography as represented by terrain attribute covariates and demonstrated by studies conducted in other areas (Kirkwood et al., 2016). In addition, previous studies have shown that the study area does not present a great heterogeneity of geological materials (BENTO et al., 2022). The dominant geologic formations were laterite crust in the center and talus in the periphery surrounding the laterite crust (Fig. 18) both with distinct concentrations of elements (Table 4). Also, the points from the dense grid (D) and the diagonally and perpendicular (D&P) transects fell almost entirely within laterite crust. Not surprisingly, the distribution of predicted values showed two major clusters of values (Fig. 22) indicating the underlying structure from the distribution of sampling points between two geologic materials (laterite crust and talus) that was also reflected by sampling scenarios (Fig 23, A, B and C). The dispersion of the observed and predicted values shows that the models tended to underestimate the high levels of the elements and display a somewhat symmetrical patterns of the predicted values. The predicted values tend to cluster in two distinct groups with high and low concentrations.



B



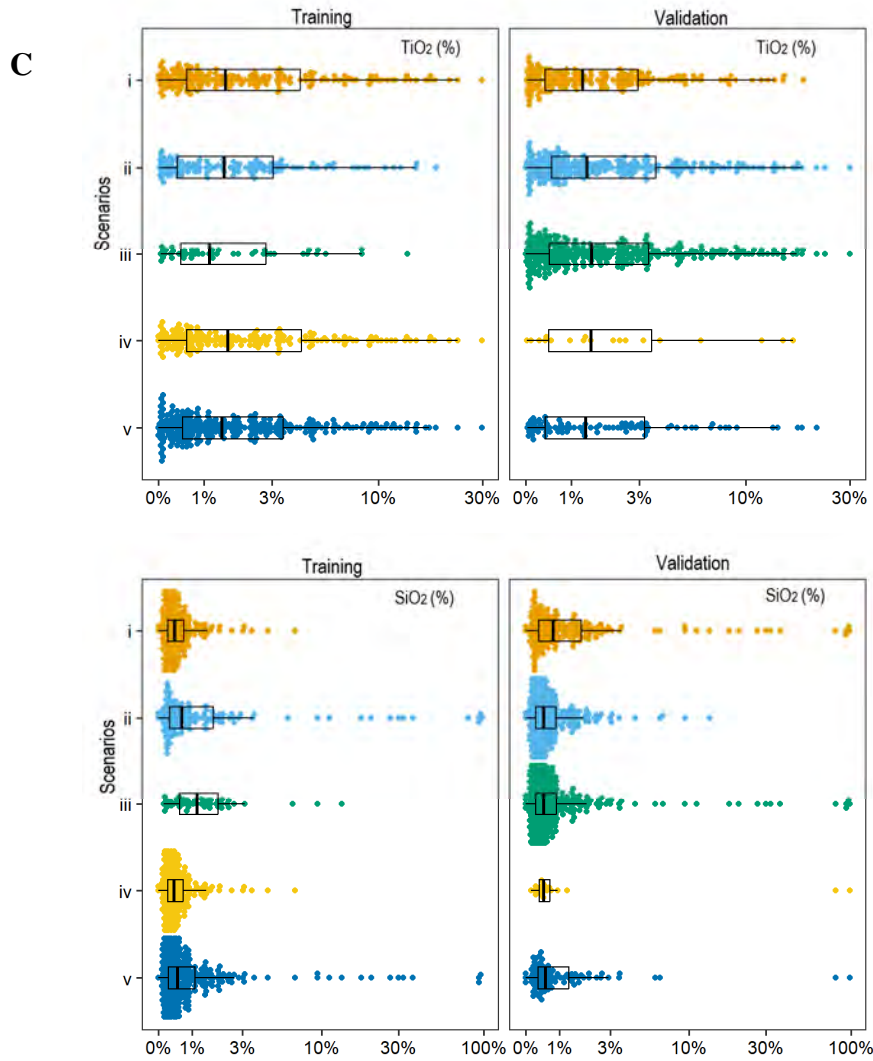


Figure 23. Distribution of the expected values for the sampling scenarios. Source: Prepared by the author (2025).

The models performed relatively better for the other scenarios, especially Scenario IV (Fig. 20A). Different from the first three scenarios, both training and validation points were geo co-located within the same spatial area despite of the fact that for training the dense grid (D) was used while for validation the D&P transect. For Scenario V all sampling designs were used for both training and validation which resulted in slightly better performance compared to the first three sampling scenarios, but still worse compared to Scenario IV.

The training versus validation analysis (Fig. 23 A, B and C), showed a greater range of element contents for the training data compared to the validation data. The range discrepancies were consistent but varied among elements which explains also the differences in model performance.

4.6.2. Spatial variability and geological context

The bimodal distribution of measured values, due to underlying geology, was present in all sampling scenarios (Fig. 23) and influenced the predictions that showed also a bimodal distribution for many of the elements (Fig. 24).

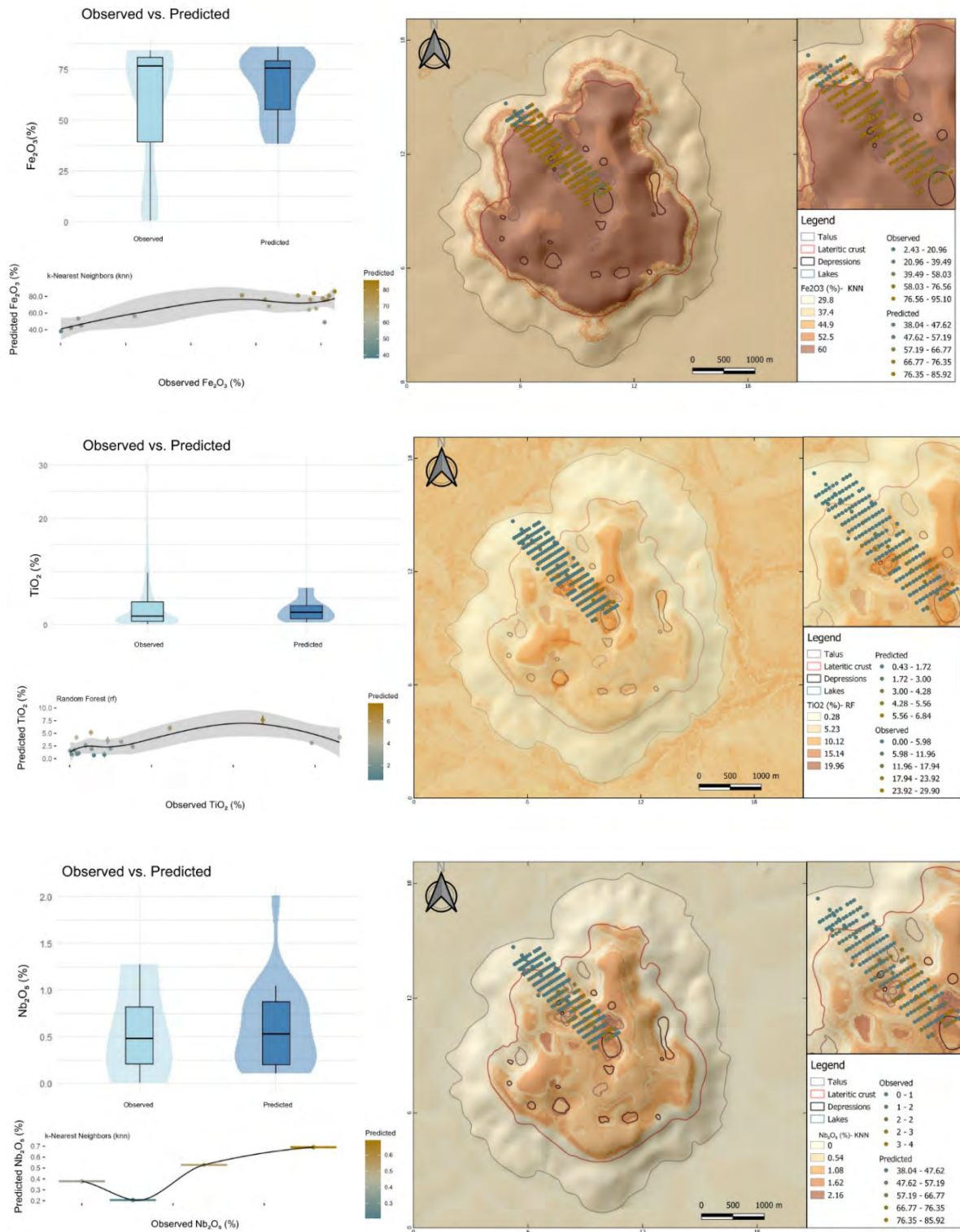


Figure 24. Distribution of observed and predicted values for Fe₂O₃, Nb₂O₅, and TiO₂ based on the KNN, GLMNET, and RF models (Appendix D), respectively, and location of sampling points underlined by two main formations (laterite crust and slope). Source: Prepared by the author (2025).

Other studies have demonstrated the importance of understanding the spatial context of the data (KIRKWOOD et al., 2016) and the usefulness of the auxiliary variables (geophysics,

remote sensing and morphometric) to capture it for improved model predictions (ODEH; MCBRATNEY; CHITTLEBOROUGH, 1995; MCKENZIE; RYAN, 1999; MCBRATNEY; SANTOS; MINASNY, 2003; HENGL; HEUVELINK; STEIN, 2004). Minasny and McBratney (2016), have found that for DSM applications transect based sampling may not always capture the patterns of environmental covariates at small scales. In our study site transects covered the two major geologic formations. However, they likely either failed to capture the topographic features or the features were not well expressed especially at the sampling locations. The same but to a lesser degree was the case for the dense grid, perhaps due to the ability to capture spatial variability over short distances as found by Stumpf et al. (2016) in another study area. Although there were differences in model performance between sampling scenarios either due to uneven distribution of samples between training and validation or their spatial representation of the environmental covariates or underlying geology, overall performance of all models was poor regardless of sampling scenarios. The evidence did not support our initial hypothesis that predictions of chemical elements would not be influenced by the sampling design.

4.6.3. Geological formation of the study area

The spatial distribution of the elements was more driven by geology than topography contrary to some soil properties such as SOC. Each unit (Talus, Laterite Crust and Lacustrine Sediments) has a unique composition, that resulted from the different weathering rate processes. In the Crust laterite unit, the characteristics of a high concentration of Fe_2O_3 , Nb_2O_5 and TiO_2 stand out (Fig. 24). Kasay (2018) showed that the weathering process, especially intense in tropical environments, causes laterites to have high concentrations of minerals. The laterites are highly enriched in oxidized forms of iron (Fe_2O_3) ranging from 39.0 to 70.0% and manganese (MnO) ranging from 8 to 39%. According to Giovannini et al. (2017), high concentrations of these elements occur in the upper and lower purple laterites and upper portions of the brown type. The talus on the other hand, has a morphology that is independent of the underlying rock. It is highly weathered with lower Fe_2O_3 , Nb_2O_5 and TiO_2 contents as shown by the asymmetrical relationship between the predicted and observed values. The Fe_2O_3 showed similar spatial variability with Nb_2O_5 and TiO_2 . The study by Oliveira et al. (2012) points to very high values of Fe and Ti in soils, characteristics inherited from the rock, which is naturally enriched in these elements. The authors emphasized that both Fe and Ti have similar weathering rates, so that the $\text{Fe}_2\text{O}_3/\text{TiO}_2$ ratio is approximately the same as that of the rock. A study by Rodrigues et al. (2022) shows that the spatial variability of Fe_2O_3 is highly correlated with Nb. Understanding the spatial structure of factors that determine the spatial distribution of properties as shown in this study could be more important, especially at the beginning for an efficient sampling design and accurate predictions.

4.6.4. Predictive models

Overall, all models performed poorly with the highest coefficient of determination ($R^2 = 0.44$ - Fe_2O_3 , S-IV). However, there were performance differences among models as shown by the wide range of performance parameters. The evidence did not support our initial

hypothesis that all models would perform similarly across sampling scenarios. The results showed that the performance of the models not only differed by scenario but was also dependent on the element. Thus, overall models performed better for Fe₂O₃ followed by TiO₂ and Nb₂O₅ (Fig. 21C). Across and within each element not a single model was clearly identified as superior, except for RF and KNN and only to a certain extent (Fig. 22B). While for Fe₂O₃ KNN was the best performer ($R^2 = 0.19$) for Nb, the GLMNET performed best ($R^2 = 0.12$). This suggests that some of the models were perhaps not only more sensitive to the location, spacing and imbalances between the size of training and validations data sets, but the degree of sensitivity also varied by element indicating an interaction between models, sampling scenarios and elements, albeit weak. In a sampling study, Safaee et al. (2024) highlighted the low prediction accuracy for the properties (SOM and CEC), with the interactions between the sites, the model and the sample density being mainly driven by site differences.

Studies by Rodrigues et al. (2020) and Safaee et al (2024) indicate good performance for models with nearest neighbor assumptions. For the Morro dos seis Lagos area, we obtained good performance for the KNN model for predicting the elements. The effect of sampling density on prediction accuracy assumes a spatial dependence but that is sensitive to noise with a high coefficient of variation (MCROBERTS et al., 2007).

The RF model performed better for the almost all elements, probably because it considers the weight of the covariates influencing the prediction and manages their variation through each tree, utilizing it for improving predictions (BREIMAN, 2001). This was also evidenced by Safaee et al. (2024) when comparing other models with RF.

4.6.5. Sampling designs and properties

The accuracy predictions as influenced by models, sampling scenarios and element point to a complex picture for improving predictions. Ensemble modeling has been used successfully to improve the accuracy predictions (MALONE et al., 2014; ZHANG et al., 2022), when no one model is superior. However, the poor performance of all models did not warrant such an approach. Our analysis conducted for 6 elements showed also spatial autocorrelation between elements that could be leveraged to decrease the number of interactions. For example, Fe₂O₃, SiO₂ and TiO₂ elements were spatially autocorrelated positively or negatively indicating the geochemical affinity among them (Fig. 25). For example, while there was a greater predominance of high Fe₂O₃ contents in the laterite crust unit, higher predominance for Nb and TiO₂ occurred in Talus.

It was found that the spatial distribution of the elements follows a trend according to the geological classes (Talus, Laterite) present in the study area. (Fig. 24).

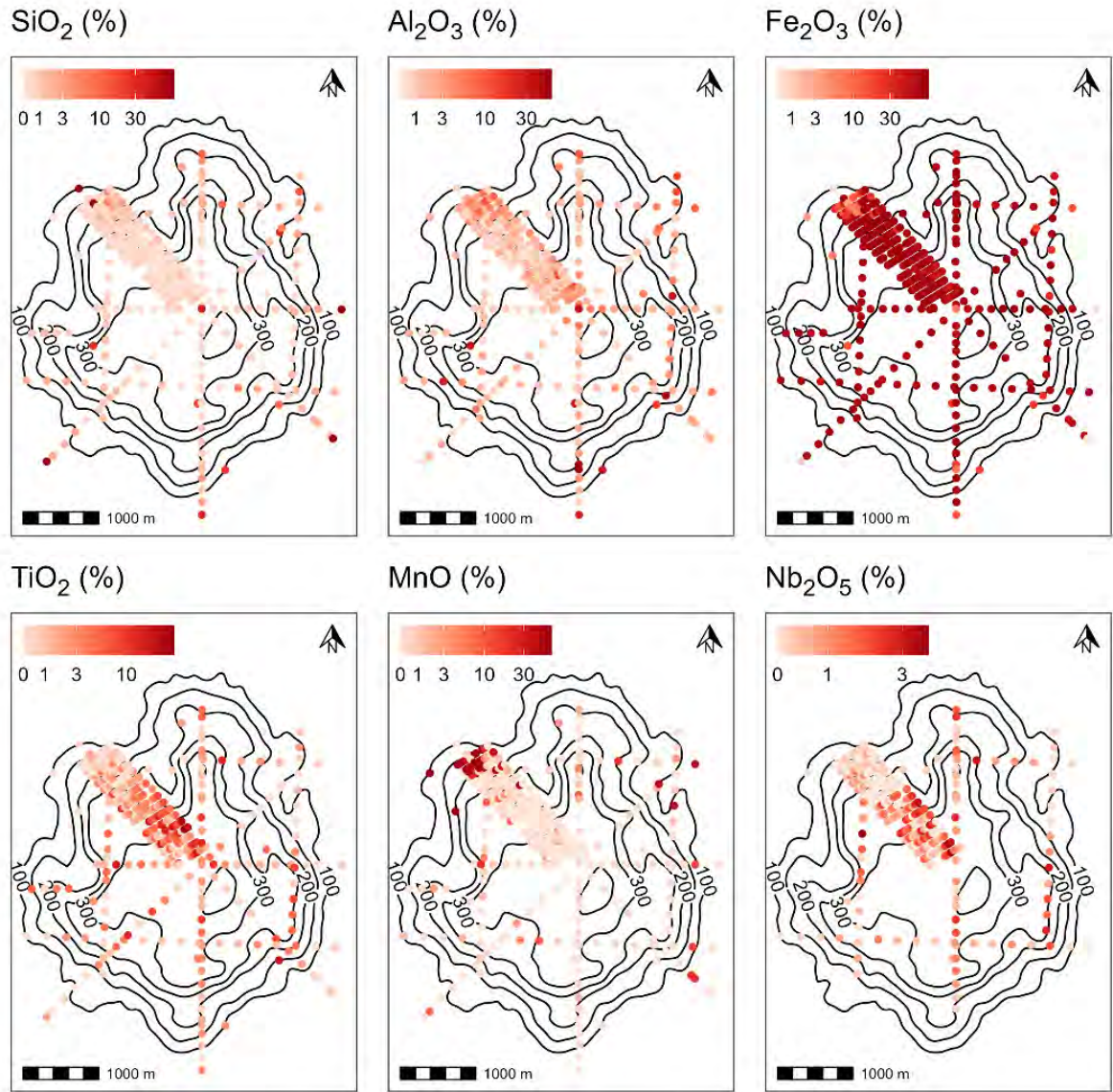


Figure 25. Spatial distribution of element content. Source: Prepared by the author (2025).

The Laterite class showed low values of SiO₂ (<10%) and Al₂O₃ (<12%) and the highest values of TiO₂ and Nb₂O₅, despite the internal compositional gradient. Another geochemical characteristic of the laterite class is the high average concentration of Fe₂O₃ and MnO elements. The Talus class showed low levels of Nb₂O₅ and TiO₂ (<3%). The strong positive correlation between TiO₂ and Nb₂O₅ stands out.

The concentrations of Fe₂O₃ content were quite high, and the sampling design (D&P) showed a better distribution of the data and amplitude of the minimum and maximum values, with a greater number of outliers, indicating a greater variation in the data. The sampling design (P) showed relatively less variability with concentration values varying between 30 and 100 % and mostly closer to 100% (Fig. 26).

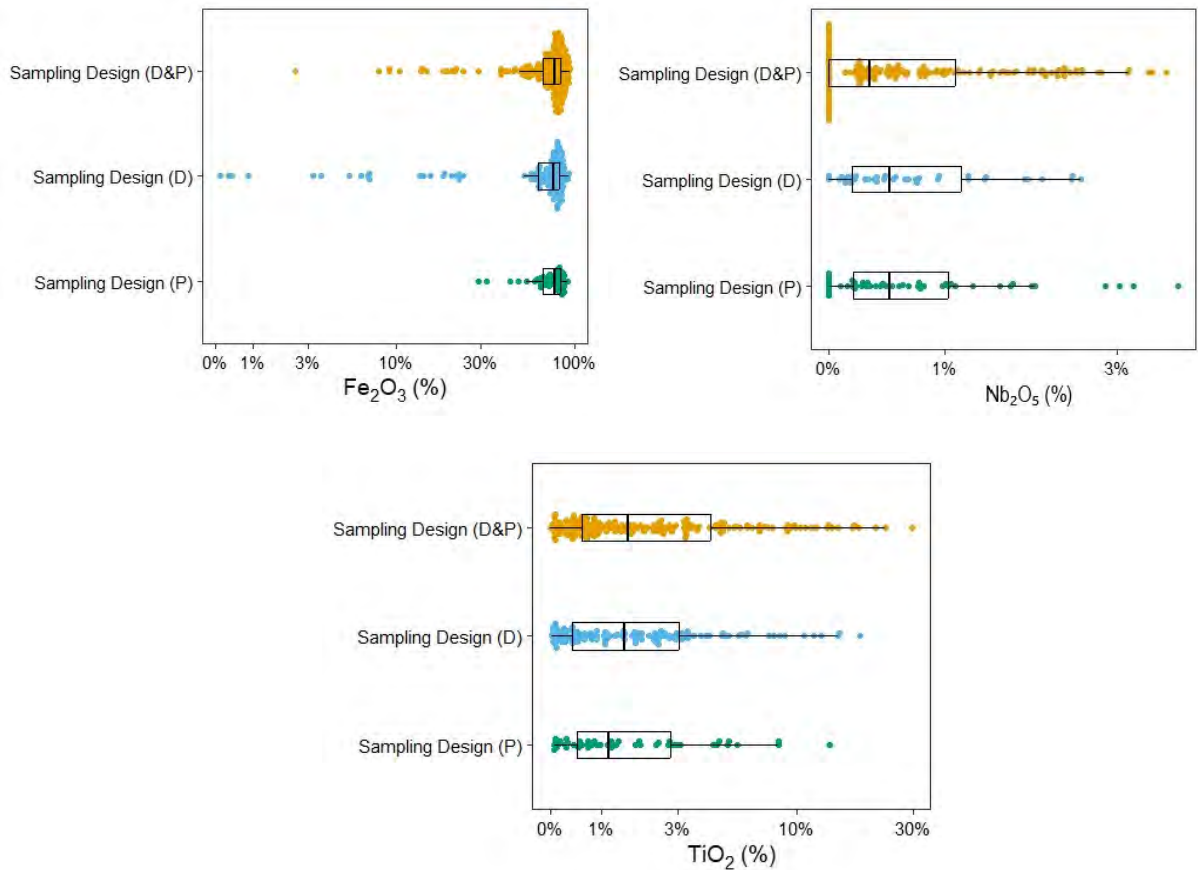


Figure 26. Visualization of soil data distribution by sampling design. Source: Prepared by the author (2025).

The Sampling Design (D&P) showed greater variation in the data for Nb_2O_5 and contained more outliers than the other sampling designs. There was a significant amount of zero values and/or values concentrated close to zero, a factor that may have influenced the model performance metrics. With regard to Fe_2O_3 , the (D) and (P) sampling designs showed less variation in the data, which was not as suitable for efficient representation.

For the TiO_2 , the sampling design (D&P) also showed greater variation in the data, with a significant number of outliers. Again, the (D) and (P) sampling designs showed less variation. The Sampling Design (D&P) has a better representation of the distribution of elements across the geologic units, and likely better captures the spatial patterns.

4.7 CONCLUSIONS

We investigated the influence of sampling designs and models on the prediction accuracy of selected chemical compounds (Al_2O_3 , Fe_2O_3 , MnO , Nb_2O_5 , TiO_2 , and SiO_2). Model performance was influenced by the sampling design, and the degree of influence varied by element. The five sampling scenarios created by alternating sampling designs between training and validation led to an unbalanced number of samples, which affected the performance of the models. Thus, the performance parameters decreased substantially from training to validation for all elements.

The Kruskal-Wallis test with Dunn's post-hoc analysis confirmed that Scenario IV ('c') established the optimal sampling design, showing significantly superior performance for most elements (Al_2O_3 , Fe_2O_3 , TiO_2), using SVMRadial and Random Forest models. Scenario V ('b') demonstrated complementary relevance for MnO and Nb_2O_5 optimization. The KNN model excelled for oxides (Al_2O_3 , Fe_2O_3 , SiO_2), while NNET performed poorly (SiO_2 , scenario I), evidencing its inability to capture spatial heteroscedasticity in lateritic materials.

However, these results demonstrate that there is no single universally superior model, algorithm selection depends on the chemical element, data spatial structure, and sampling design. In general, model performance remained moderate to low across all sampling scenarios and chemical elements, mainly due to weak correlation between element concentrations and predictors, compounded by underlying geological complexity. This reflects the need for geological stratification and new mineralogical covariates. Future analyses should explore hybrid modeling techniques to enhance predictive accuracy in these complex pedogenic environments.

Thus, future analyses should incorporate new covariates (geological and mineralogical indices) and hybrid modeling techniques to enhance predictive capacity, given the inherent complexity of highly weathered soil environments such as *Morro de Seis Lagos*. Understanding the signature of geology on the concentration and spatial distribution of chemical elements in conditions of high weathering, such as those observed on the *Morro de Seis Lagos*, is as important a factor in characterizing the pedogenic process as the characteristics of the terrain and landscapes.

5. CAPÍTULO III

WEATHERING PATTERN RECOGNITION IN PETROFERRIC FORMATIONS IN BRAZILIAN AMAZON

5.1 RESUMO

Índices de intemperismo, como o *Index of Lateritisation* (IOL), *Chemical Index of Alteration* (CIA), e o *Chemical Index of Weathering* (CIW), são uma janela para a compreensão dos processos geoquímicos da dinâmica solo-relevo. Esses índices podem ser utilizados para prever e monitorar tendências pedogenéticas espaço-temporais em climas tropicais úmidos, como na Bacia Amazônica, onde esses estudos são raros. O objetivo desta pesquisa foi quantificar a intensidade do intemperismo (laterização), utilizando os índices CIA, CIW e IOL, em diferentes unidades geomorfológicas no Morro de Seis Lagos, Amazônia, empregando modelos de aprendizado de máquina *Random Forest* (RF) e *Cubist* (CB). Os procedimentos metodológicos tiveram como base: i) Revisão temática; ii) Estatística descritiva; iii) Filtragem por *Recursive Feat Elimination* (RFE); e iv) implementação de algoritmos preditivos e elaboração de mapas bivariais. O delineamento utilizou 341 pontos (solos, sedimentos e materiais rochosos) da camada superficial, provenientes de dados geoquímicos legados pelo Serviço Geológico do Brasil (CPRM). As covariáveis morfométricas foram geradas a partir do Modelo Digital de Elevação (MDE-HC) e, para o conjunto de covariáveis de sensoriamento remoto, foram utilizados o sensor Sentinel-2A e o Radiômetro de Emissão Térmica e Reflexão de Satélites Avançados (ASTER) para derivar índices espectrais mineralógicos. Os resultados indicaram o melhor desempenho dos índices IOL, CIA e CIW para o modelo *Random Forest*, seguido pelo modelo *Cubist*. A variação entre unidades geológicas e topográficas foi atribuída ao uso de covariáveis via RFE-RF, com ambos os modelos destacando a importância de covariáveis obtidas das bandas VV e VH do SAR - Sentinel-1A, associadas a covariáveis morfométricas de relevo e índices espectrais do sensor ASTER, como aqueles relacionados ao índice de Goethita e materiais ferruginosos, mostrando padrões nos processos de intemperismo e na evolução de perfis lateríticos.

Palavras-chave: Geoquímica. Lateritas. Mapeamento digital. Pedogeomorfologia. Processos pedogenéticos.

5.2 ABSTRACT

Weathering indexes such as the Index of Lateritisation (IOL), Chemical Index of Alteration (CIA), and the Chemical Index of Weathering (CIW) are a window to understanding the geochemical processes of soil-landscape dynamics. These indexes can be used to predict and monitor spatial-temporal pedogenetic trends in humid tropical climates, such as the Amazon basin, where such studies are rare. The aim of this research was to quantify the intensity of weathering (laterization), using the CIA, CIW, and IOL indexes, in different geomorphic units in Morro de Seis Lagos, Amazon, using Random Forest (RF) and Cubist machine learning models. The methodological procedures were based on: i) Thematic review; ii) Descriptive statistics; iii) Filtering via Recursive Feature Elimination (RFE); and iv) Implementation of forecasting algorithms and making of bivariate maps. The design used 341 points (soils, sediments, and rock materials) from the surface layer, sourced from legacy geochemical data from the Geological Survey of Brazil (CPRM). The morphometric covariates were generated from the Digital Elevation Model (MDE-HC), and for the set of remote sensing covariates, the Sentinel -2A sensor and Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) were used for mineralogical spectral indexes. The results obtained showed the best performance for the IOL, CIA, and CIW indices for the Random Forest model, followed by the Cubist model. The variation in geological and topographical units was found to be due to the use of covariates via RFE -RF, with both models showing the importance of covariates from the VV and VH bands of SAR - Sentinel-1A, associated with terrain morphometric covariates and spectral indexes from the ASTER sensor, such as those related to Goethite and ferruginous materials, showing patterns in weathering processes and the evolution of laterite profiles.

Keywords: Geochemistry. Laterites. Digital mapping. Pedogeomorphology. Pedogenetic processes.

5.3 INTRODUCTION

Soil is the product of chemical, physical, and biological weathering (RETALLACK, 1997) through pedogenesis acting on the lithology/parent material, governed by climate, weather, organisms, and terrain (JENNY, 1994). According to Tardy (1986), weathering precedes erosion, causing the loss of elements through the leaching process and resulting in the relative accumulation of less soluble elements in the weathering profile. In terms of mineralogical composition, weathering removes the primary silicate phases and replaces them with poor crystalline aluminosilicates. Over time, weathering can produce iron oxide and oxyhydroxide phases, which evolve into more crystalline varieties, as demonstrated by several studies on lateritic and ferruginous crusts (COSTA, 1991; DE FIGUEIREDO FILHO; DE OLIVEIRA; BARROS, 2019; MARTINS et al., 2024). Quantifying weathering rates is important for understanding landscape evolution processes from a geological and geomorphological perspective (BIGARELLA, 1994), modeling natural processes (MELLO et al., 2023; MEI et al., 2021; MOHANTY; GUPTA; DAS, 2016), and supporting environmental planning.

In this context, geochemical indices such as the Chemical Index of Alteration (CIA), the Chemical Index of Weathering (CIW), and the Laterization Index (IOL) are potential tools for assessing and understanding the dynamics of chemical alteration and geological materials (rocks, sediments, and soils). As highlighted by the studies of Santos et al. (2024), who apply the IOL index to regolith profiles in the region and Caparaó Mountains (MG/ES) to quantify stages of laterization, and Goydaragh et al. (2021), who use CIA CIW via FTIR to predict initial leaching in soils in the Zona da Mata Mineira (MG), these indices allow for the intensity of leaching and the mobility of elements, being applicable to regions with a humid tropical climate.

Nesbitt and Young (1982) proposed the Chemical Index of Alteration (CIA) as an indicator of the degree of chemical weathering in soils and sediments, with implications for paleoclimatic reconstruction. Thus, the removal of mobile cations during chemical weathering, facilitated by warmer and wetter climatic conditions, produces higher CIA values. The CIA is calculated by the equation: $CIA = [Al_2O_3 / (Al_2O_3 + CaO + Na_2O + K_2O)] \times 100$. Where CaO represents the calcium contained in the fraction and in the silicates. The McLennan (1993) correction procedure was adopted to exclude the influence of carbonates and phosphates, ensuring the accuracy of the index only for silicate alteration.

The Chemical Index of Weathering (CIW) was developed by Harnois (1988) and is a geochemical indicator used to quantify the intensity of chemical alteration based on the composition of major oxides in rocks and soils. It is used to assess the removal of mobile cations in environments of intense weathering where intense leaching occurs. The CIW is represented by the expression: $CIW = [Al_2O_3 / (Al_2O_3 + CaO + Na_2O)] \times 100$. According to Harnois (1988), in this index, alumina (Al_2O_3) represents the component that is relatively immobilized during weathering, while CaO and Na_2O represent more mobile cations, whose progressive removal reflects the advance of chemical alteration.

The Laterization Index (IOL) was proposed by Schellmann (1981) and is a geochemical indicator developed specifically to quantify the degree of laterization in tropical alteration profiles based on the composition of major oxides in rocks, saprolites, and soils. The IOL is represented by the expression: $IOL = [(Al_2O_3 + Fe_2O_3) / (SiO_2 + Al_2O_3 + Fe_2O_3)] \times 100$.

According to Schellmann (1981), in this index, sesquioxides (Al_2O_3) and iron (Fe_2O_3) represent residual components immobilized or enriched during laterization in environments of intense weathering, where there is strong leaching of SiO_2 and enrichment of Al and Fe (BOURMAN; OLLIER, 2002).

The CIA, CIW, and IOL indices represent universal geochemical indicators applicable to complete weathering profiles (rock, saprolite, soil, sediments), allowing the quantification of initial leaching (CIA, CIW) to advanced laterization (IOL), as proposed by Nesbitt and Young (1982), Harnois (1988), and Schellmann (1981), respectively. Unlike these, the Ki and Kr indices (OLIVEIRA et al., 1983) are strictly pedological in nature and are defined in the Brazilian Soil Classification System (SiBCS) to characterize highly weathered Latosols (Oxisols).

In this context, the application of geochemical indices is a potential tool for elucidating the dynamics and degree of weathering intensity in humid tropical climate regions, a scenario exemplified by the *Morro dos Seis Lagos* area, with records of petroferic formations with an advanced degree of pedogenesis. Local geological, geomorphological, climatic and hydrological characteristics lead to the formation of various mineral endmembers that document pedogenetic processes recorded through morphological, physical and chemical properties. The *Morro dos Seis Lagos* area in the municipality of São Gabriel da Cachoeira (AM), on the upper Negro River of the Amazon Basin is home to the rare formation of siderite carbonatite which presents an opportunity for studying petroferic formations from the pedological and/or pedogenic perspectives. The formation is rare covering less than 1% of world surface area (GIOVANNINI, 2013; TAKEHARA, 2019) and considered one the largest known Nb deposits. It occurs on highly weathered geomorphic surfaces, containing a thick crust of laterite (>200 m) (TAKEHARA, 2019), where the carbonation processes of the siderite have produced a crust of goethite/hematite.

In *Morro dos Seis Lagos*, the occurrence of minerals with economic potential is associated with the intense weathering process of the carbonatite. The Nb is in the form of pyrochlore ($\text{A}_{2-x}\text{B}_2(\text{O}, \text{OH})_6(\text{OH}, \text{F}, \text{H}_2\text{O})_{1-}$) and is considered an accessory mineral in siderite carbonatites. In tropical humid environments, intense chemical weathering promotes the formation of thick weathering profiles on carbonatites, with leaching of mobile elements such as calcium and silica, and consequent residual enrichment of immobilized elements such as Nb (PORTO; BASTOS NETO; BENTO, 2025). Thus, secondary enrichment of Nb occurs in laterite zones and residual soils (Ferralsols).

The study area is considered a paleoenvironmental archive for soils in Neotropics and offers an opportunity to quantify soil weathering rates or intensity resulting from the interactions between vegetation, relief, time, and chemical and physical processes. The objectives of this research are focused on studying weathering intensity of rare siderite/petroferic carbonatite formation as captured by CIA, CIW, and IOL geochemical indexes and factors controlling their spatial distribution using machine learning combined with remote sensing.

The following alternative hypotheses will be tested: i) Terrain characteristics influence the spatial distribution of Petroferic formation; and ii) Geomorphic and/or geological units represent different weathering intensities.

5.4 MATERIAL AND METHODS

The methodological procedures are divided into three stages (Fig. 27). The first stage includes the compilation of model inputs (geochemical data, weathering indexes, and terrain characteristics or covariates). The second phase consists of applying the Random Forest and Cubist models to spatially predict CIA, CIW, and IOL indexes based on covariates, using the Caret package (KUHNN, 2008). The third stage consists of modeling CIA, CIW, and IOL indexes.

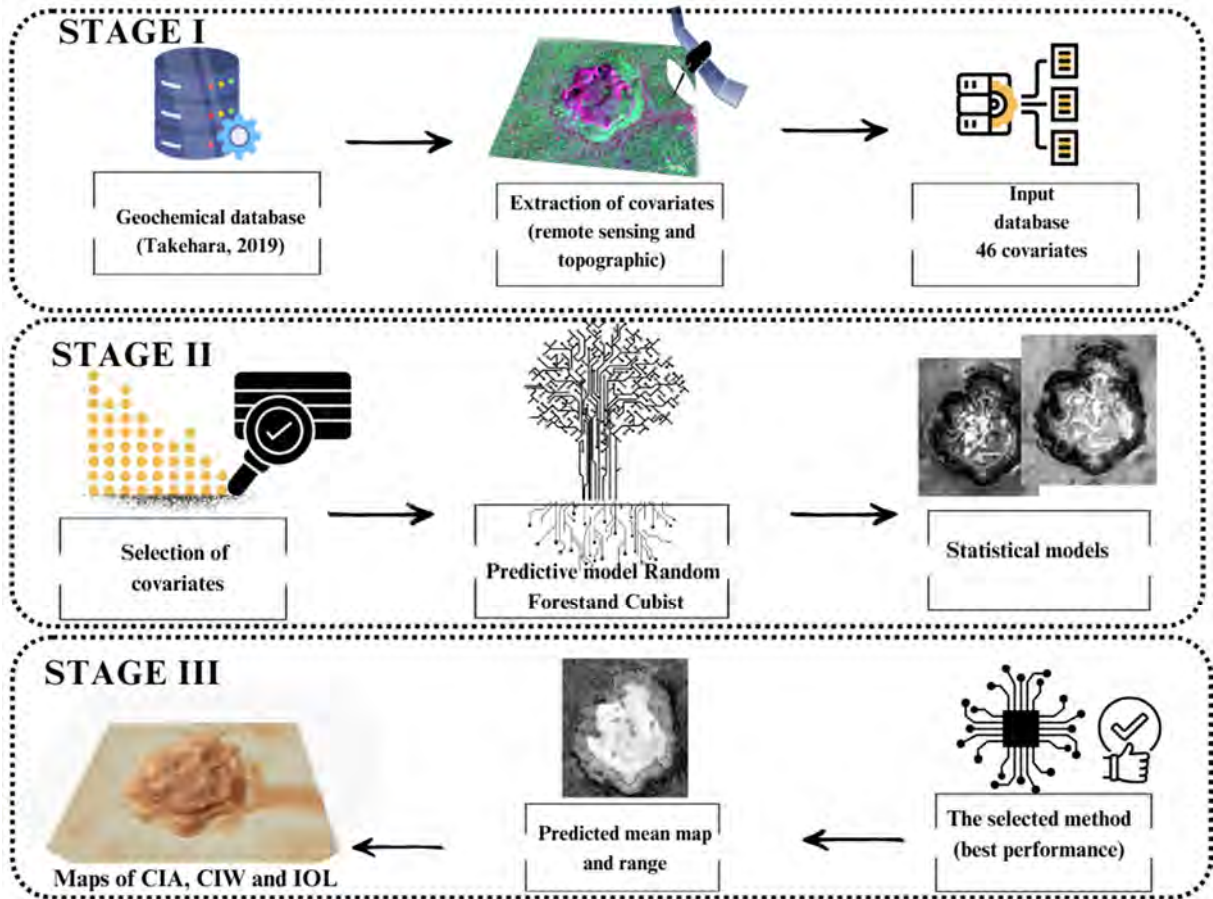


Figure 27. The methodological procedures were divided into three stages. Stage I - Preparation of the database, with extraction of terrain covariates and remote sensing indices; Stage II - Selection of covariates using Spearman's correlation and the Recursive feature elimination (RFE) method. The Random Forest and Cubist models were then processed with 20 repetitions, and the statistics of the performance metrics were calculated; Stage III - Selection of the model with the best performance based on the metrics, followed by generation of the mean prediction and amplitude maps. Source: Prepared by the author (2025).

5.4.1 Study area

The *Morro dos Seis Lagos* region (Fig. 28) is located in the northwestern part of the Amazon basin, (43° 12' 34.56" W, 22° 58' 12.34" S), with a total area of 35.96 km² bordered (<150 km) by Venezuela and Colombia. The area is within three conservation units (UCs)

(BRASIL, 2000) which overlap: the Balaio Indigenous Land, the Pico da Neblina National Park and the Morro dos Seis Lagos Biological Reserve.

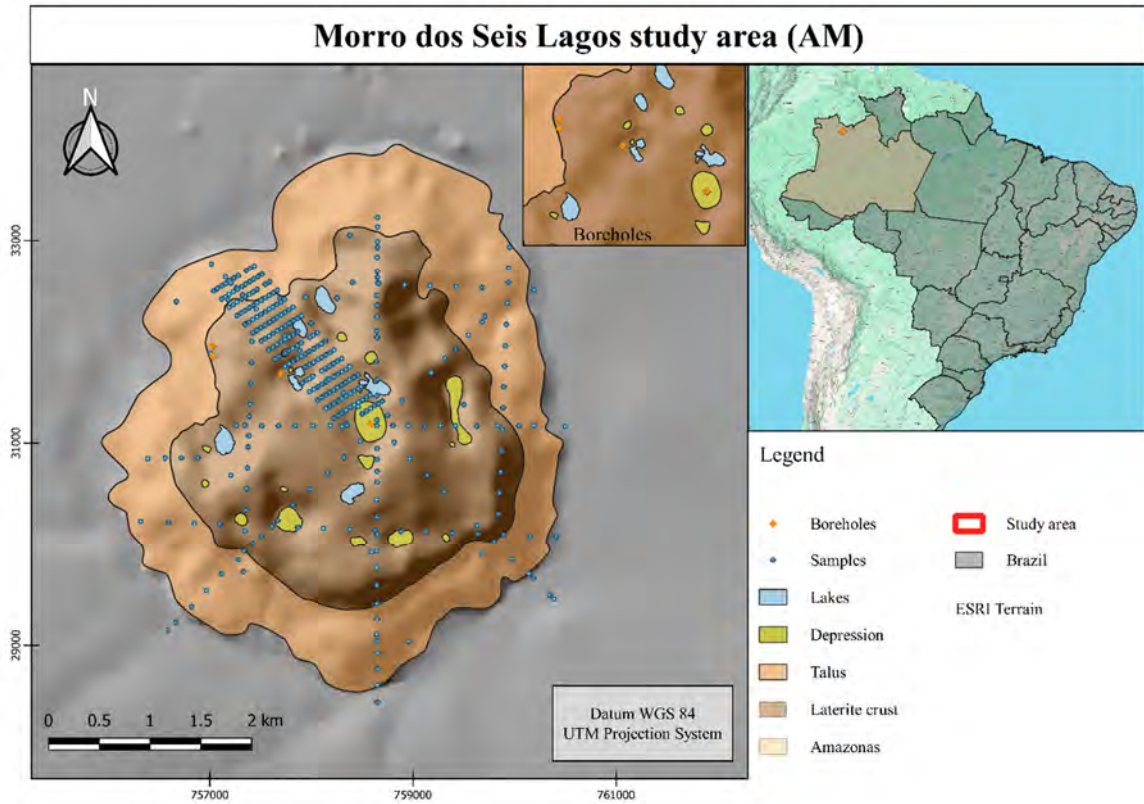


Figure 28. Location of the study area in the Six Lakes Project, in São Gabriel da Cachoeira, Brazilian Amazon, Amazonas State, highlighting the surface sampling collection points ($n = 341$) and the four boreholes. Source: Prepared by the author (2025).

The region is composed of a granite-gneiss substrate with the local geology mostly of the Imeri domain (TAKEHARA, 2019). The region is under lateritic crust, with a thickness of more than 250 meters (TAKEHARA, 2019). According to Justo (1983), *Morro dos Seis Lagos* is one of three isolated volcanic chimneys with an alkaline-carbonate constitution. The local geology of MSL (Fig. 28) is made up of talus, laterite crust and depressions. According to Bento et al. (2022), Morro dos Seis Lagos has a circular shape and is supported by ferriferous Canga formed from the original igneous intrusive rock, which has been eroded considerably.

The Northwest Amazon region receives annually between 2,000- and 3,490-mm rainfall with an even distribution throughout the year, while the average annual temperature vary between 25 and 26°C (NIEMER, 1979). The hot and humid climate is classified by Koppen (1998) as Af (Hot and Humid Equatorial Climate).

5.4.2 Laboratory analysis

A total of 810 samples and pulps from the Project Assessment of Rare Earth Potential in Brazil in Morro dos Seis Lagos Area, Northwest Amazonas (TAKEHARA, 2019) were reanalyzed by the SGS Geosol Laboratory (Belo Horizonte, Brazil). Three analytical methods were used: (i) X-ray fluorescence (XRF) for 10 main oxides ($\text{Al}_2\text{O}_3\%$, $\text{CaO}\%$, $\text{Fe}_2\text{O}_3\%$, $\text{K}_2\text{O}\%$,

MgO%, MnO%, Na₂O%, P₂O₅%, SiO₂%, TiO₂%); (ii) ICP-MS combined with ICP-OES after metaborate fusion followed by acid dissolution and analysis for 14 rare earth elements (La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu) plus Nb, Sn, W, Y and 11 other elements; and (iii) ICP-MS combined with ICP-OES after aqua regia digestion and analysis for Sc, Zn and 23 other elements.

5.4.3 Weathering indexes- surface

In order to understand the conditions of physical and chemical weathering, as well as the soil's pedogenetic processes, the following geochemical proxies were used: Index of Lateritization (IOL), Chemical Index of Alteration (CIA) and Chemical Index of Weathering (CIW) (Table 8).

Table 7. Geochemical indicators of weathering.

INDEXES	FORMULA
Index of Lateritisation (IOL)	$IOL = [(Al_2O_3 + Fe_2O_3) / (SiO_2 + Al_2O_3 + Fe_2O_3)] \times 100$ Schellmann (1981)
Chemical Index of Alteration (CIA)	$CIA = [Al_2O_3 / (Al_2O_3 + CaO + Na_2O + K_2O)] \times 100$ Nesbitt e Young (1982) e Bahlburg e Dobrzinski, N. (2011)
Chemical Index of Weathering (CIW)	$CIW = [Al_2O_3 / (Al_2O_3 + CaO + Na_2O)] \times 100$ Harnois (1988)

Source: Prepared by the author (2025).

5.4.4 Environmental covariates

The terrain attributes were derived from a Hydrologically Consistent Digital Elevation Model (MDE-HC), with a spatial resolution of 20 meters, using Topo to Raster interpolator in ArcGIS Desktop v. 10.6 (ESRI, 2016) (Table 8). The DEM was based on primary vector data of contour lines with an equidistance of 10 meters, elevation points and hydrography, extracted from the cartographic base of the Brazilian Institute of Geography and Statistics (IBGE) - scale 1:25,000 (IBGE, 2025). The following terrain attributes were derived from the MDE-HC using the System for Automated Geoscientific Analyses (SAGA-GIS) v.2.1.2 (CONRAD, 2007). The covariates from remote sensing data were obtained using the Multispectral Sensor Instrument (MSI) of the Sentinel-1 and 2A satellites, and the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) of the Terra satellite, with a spatial resolution of 20 meters. The categorical covariate of Geology was also added (BENTO et al., 2022) (Table 9).

Table 8. Terrain morphometrics covariates, categorical covariates, and spectral indices from Morro dos Seis Lagos area, São Gabriel da Cachoeira, Amazonas State, Brazil.

Category	Variable / Index	Equation (if applicable)	Reference
Geomorphological Indices & Factors	SAGA Wetness Index (WI)	-	Conrad (2001)
	Slope Height	-	Thompson <i>et al.</i> (2001); Wilson & Gallant (2000)
	Aspect	-	Conrad (2013)
	Elevation	-	Hutchinson & Gallant (2000); Moore <i>et al.</i> (1991)
Curvature and Morphometry	Vector Ruggedness Measure (VRM)	-	Wilson <i>et al.</i> (2007)
	Real Surface Area	-	Olaya (2004)
	Convergence Index (CI)	-	Conrad (2001)
	Curvature Flow Line	-	Conrad (2001)
	General Curvature (GC)	-	Conrad (2001)
	Maximum Curvature	-	Conrad (2001)
	Minimum Curvature	-	Conrad (2001)
	Planar Curvature (PlanC)	-	Conrad (2001)
	Profile Curvature (ProfC)	-	Conrad (2001)
	Tangential Curvature (TanC)	-	Conrad (2001)
	Total Curvature (TotalC)	-	Conrad (2001)
	Mid-Slope Position	-	Conrad <i>et al.</i> (2008)
	MRRTF	-	Gallant & Dowling (2003)
	MRVBF	-	Gallant & Dowling (2003)
	Slope Index	-	Conrad (2010)
	Standardized Height	-	Conrad <i>et al.</i> (2008)
	Terrain Surface Convexity	-	Conrad (2012)
	Terrain Surface Texture	-	Conrad (2012)
	Valley Depth	-	Conrad (2012)
	Closed Depressions	-	Conrad (2012)
	Hill	-	Olaya (2004)
	Hill Index	-	Olaya (2004)
	Valley	-	Conrad (2015)
	Valley Index	-	Conrad (2015)
	Surface Specific Points	-	Conrad (2005)
Spectral Indices	NDVI	$(\text{RNIR} - \text{Rred}) / (\text{RNIR} + \text{Rred})$	Rouse <i>et al.</i> (1974)
	SAVI	$(\text{RNIR} - \text{Rred}) / (\text{RNIR} + \text{Rred} + \text{L})$	Huete (1988)
	Ferrous Silicates	Band 12 / Band 11	Van der Meer, Van der Werff e Van Ruitenbeek (2014)
	Clay Minerals Ferric Iron, Fe ³⁺	Band 6 / Band 7 Band 2 / Band 1	Sabins (1997) Kalinowski & Oliver (2004)
	Laterite	Band 11 / Band 12	Van der Meer, Van der Werff e Van Ruitenbeek (2014)

	Ferrous Iron, Fe2+	Band 5 / Band 3 + Band 1 / Band 2	Kalinowski & Oliver (2004)
	Ferroginous Regolith Ferrous Iron	Band 8A / Band 3 (Band 5 + Band 1) / (Band 3 + Band 2)	Rowan & Mars (2003) Rowan & Mars (2003)
	Altered Zones	Band 8 / Band 11	Sora et al. (2018)
	Advanced Clay	Band 4 / Band 6	Sora et al. (2018)
	Alteration Detection		
	Hydrothermally Altered Rocks	Band 8 – Band 1 – Band 2 – Band 3	Sora et al. (2018)
	Gossan Index	Band 4 / Band 2	Sora et al. (2018)
	VV Polarization (Sentinel-1A SAR)	-	Sentinel (2021)
	VH Polarization (Sentinel-1A SAR)	-	Sentinel (2021)
Categorical variable	Geology		Bento et al. (2022)

Source: Prepared by the author (2025).

5.4.5 Spatial modeling of weathering indices

5.4.5.1 Selection of covariates

Prior to spatial modeling by RF and Cubist the Recursive Feature Elimination (RFE) (JEONG et al., 2017) algorithm was used to select the predictor covariates. The method carries out a recursive elimination by assessing the collinearity (multicollinearity) between covariates and the predicted variables (SVETNIK et al., 2004). The subset of covariates (2:30, 35, 40, 45) was processed in the RFE subset hyperparameter. The subset was selected based on 5-fold cross-validation using R-squared performance metric for each subset.

5.4.5.2 Models

a) Random forest (RF) model

The Random Forest (RF), developed by Breiman (2001) as an extension of the Classification and Regression Trees (CART), is considered a non-parametric technique. According to Breiman (2001), it consists of a combination of predictor trees (forest), which are generated from an independently sampled random vector that follows the same distribution as the other trees. The application of RF, in the case of regression, leads to a result that is equivalent to the average of all the trees. Heung, Bulmer e Schmidt (2014) considers the model to be robust, not very sensitive to noise in the data and overfitting, with the ability to provide measures of the importance of the covariates in the prediction.

The RF model (caret package) adjustment used grid search for hyperparameter mtry (2,4,6), selecting the best configuration based on the lowest MAE under 10-fold cross-validation repeated three times.

b) Cubist (CB) model

Cubist (CB) is an advanced rule-based algorithm used as an extension of the non-parametric regression tree that addresses the linear relationship between the target variable and predictor variables (BASSER, 1992; QUINLAN, 1995). According to Nguyen *et al.* (2019)

CB processing consists of four steps: i) splitting the data to grow a complete tree; ii) creating a regression model at each node to prepare for pruning and prediction; iii) pruning the tree to avoid the overfitting; and iv) smoothing the tree to hide discontinuity boundaries caused by splitting.

The Cubist model was implemented via the caret package, with hyperparameter optimization via grid search considering the parameters committees (1, 10, 50) and neighbors (0, 1, 3, 5). The selection was based on the MAE metric, applied in 10-fold cross-validation with three repetitions.

5.4.6 Model performance evaluation

The model's performance was evaluated using the following metrics: Root Mean Squared Error (RMSE) (eq. 1), Mean Absolute Error (MAE) (eq. 2), Coefficient of determination (R^2) (eq. 3), Concordance correlation coefficient (ρ_c) (eq. 4), and Bias (eq. 5).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (obs_i - pred_i)^2}{n}} \quad (1)$$

Where:

obs_i : observed value of the response variable; (2)

$pred_i$: model predicted value;

n : total number of observations.

$$MAE = \left| \frac{\sum_{i=1}^n obs_i - pred_i}{n} \right|$$

Where:

obs_i : observed value;

$pred_i$: predicted value;

n : total number of observations.

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

Where:

y_i : observed value of the response variable;

$\hat{y}_i$: value predicted by the model;

$\bar{y}$: mean of the observed values;

n : total number of observations

$$\rho_c = \frac{2\rho\sigma_{obs}\rho\sigma_{pred}}{\sigma_{pred}^2 + \sigma_{obs}^2 + (\mu_{pred} + \mu_{obs})^2} \quad (4)$$

Where:

ρ : Pearson correlation between observed and predicted values;

σ_{obs} : Standard deviation of observed values;

σ_{pred} : Standard deviation of predictions;

μ_{obs} : Average of observed values;

μ_{pred} : Average of predicted values.

$$\text{Bias} = \frac{1}{n} \sum_{i=0}^n (pred_i - obs_i) \quad (5)$$

Where:

P_i represents the predicted values;

O_i represents the observed values.

Where “obs” and “pred” are observed and predicted values for each element.

Each model was run 20 times (repetitions) for each weathering index, and the mean prediction, standard deviation maps for each index were generated. For each run, the data was randomly split 70% for training and 30% for validation, following the procedure by Reis et al. (2021) and Rodrigues et al. (2024). The bivariate maps were processed using the bi-scale package (PRENER et al., 2022) in the Rstudio program (RStudio Team, 2020), while the average maps of the Nb₂O₅ and TiO₂ elements, using the Random Forest model, were taken from the study by Rodrigues et al. (2024).

5.5 RESULTS AND DISCUSSION

5.5.1 Weathering Indices difference between geologic/geomorphic units

The data set (Table 10) was initially subjected to descriptive statistical analysis (Table 10), obtaining the mean, coefficient of variation and interquartile range.

Table 9. Summary statistics of weathering indexes and mean comparison between Laterite and Talus.

Geologic Units	Mean	Median	Min.	Max.	StDev	CV (%)	Cs	Ck	Sig.(p<0.05)
Weathering Indexes									
IOL									
Laterite	98.7	99.4	54.5	100	4.11	4.17	-8.66	80.0	non-normal
Talus	90.6	97.6	0.7	99.7	22.5	24.8	-3.33	9.96	non-normal
CIA									
Laterite	84.1	86.3	38.6	97.6	10.3	12.2	-1.73	4.03	non-normal
Talus	89.2	92.0	51.0	98.9	9.71	10.9	-1.94	4.30	non-normal
CIW									
Laterite	84.5	87.0	40	97.7	10.5	12.4	-1.60	3.13	non-normal
Talus	90.3	92.8	51.0	99.3	8.99	9.96	-2.08	5.17	non-normal

IOL = Lateralization Index; **CIA** = Chemical Index of Alteration; **CIW** = Chemical Index of Weathering; **StDev** = standard deviation; **CV** = coefficient of variation; **Cs** = coefficient of asymmetry; **Ck** = Kurtosis; **Sig.** = significant at p = 0.05.

Source: Prepared by the author (2025).

The IOL index for the laterite geological unit showed a low standard deviation of 4.11 and a coefficient of variation of 4.17, with an average of 98.7, indicating homogeneity of the dataset and conditions for a material with a high degree of weathering. The talus unit showed a standard deviation of 22.5 and a coefficient of variation of 24.8 %, indicating a more dispersed distribution, with negative asymmetry and kurtosis, which could also be indicative of heavily weathered material.

The CIA index averaged > 80 for both the Laterite and Talus units, indicating advanced weathering (HUANG et al., 2024). The CIW index showed similar behavior to the CIA, with a mean > 80 for both geological units, with a non-normal distribution. All the data sets are non-normal (p < 0.05), so it was necessary to adopt non-parametric methods.

5.5.2 Spatial predictions of surface weathering indices

The average maps, standard deviation and ranking of importance of the covariates are presented in Figures 29, 30 and 31, and show the spatial distribution of the Index of Lateritization (IOL), Chemical Index of Alteration (CIA) and Chemical Index of Weathering (CIW), considering the intensity of weathering and the residual mineralization of oxides such as Nb₂O₅ and TiO₂.

The performance evaluations of the Random Forest and Cubist models (Table 11) were based on the coefficient of determination index (R²), which obtained a significant explanatory variation of 0.03-0.28 (%), and on the mean absolute error (MAE).

Table 10. Performance metrics for the accuracy predictions of surface weathering indices (IOL - Lateralization Index; CIA - Chemical Index of Alteration; and CIW - Chemical Index of Weathering for Random Forest and Cubist Models in the *Morro dos Seis Lagos* area, São Gabriel da Cachoeira, Amazonas State, Brazil.

.Weathering Index (%)	RMSE		MAE		R ²		BIAS	CCC
	Random Forest							
	Training	Validation	Training	Validation	Training	Validation		
	n	n	n	n	n	n		
IOL	9.51	13.31	3.91	4.24	0.19	0.28	0.65	0.53
CIA	9.87	10.05	7.49	6.83	0.15	0.03	-0.5	0.18
CIW	10.36	8.89	7.55	6.59	0.15	0.12	-0.84	0.34
	Cubist							
IOL	0.08	0.15	0.03	0.05	0.32	0.12	-0.02	0.35
CIA	9.78	10.58	7.60	7.58	0.18	0.02	0.18	0.14
CIW	10.57	8.99	7.93	6.62	0.15	0.03	-0.01	0.29

RMSE: Root Mean Square Error; **MAE:** Mean Absolute Error; **R²:** Coefficient of Determination; **BIAS:** Mean Bias Error; **CCC:** Concordance Correlation Coefficient.

Source: Prepared by the author (2025).

The RF algorithm showed $R^2 = 0.28$, $MAE = 4.24$ for IOL; CIA with $R^2 = 0.03$, $MAE = 6.83$; and CIW with $R^2 = 0.12$, $MAE = 6.59$ (Table 11). When compared to Cubist, the Random Forest model performed better in predicting weathering indices. As also shown in studies such as Goydaragh et al. (2021), where there is the presence of more robust performance than other models for environmental attributes and indices, there is a better sensitivity to the dataset. Thus, the values showed moderate performance accuracy in correctly predicting weathering intensity.

The IOL index (Figure 29), calculated by the RF model, showed high values according to the studies by Sahoo et al. (2017). In relation to the lateritic crust zone, the value was $> 90.5\%$, while the talus presented an average between 87.9 and 95.6 (%), both indicating an advanced stage and lateritization. According to Pereira et al. (2024), an IOL index between 85-95% indicates intermediate laterites, and $< 90\%$ indicates mature lateritic crusts with almost total SiO_2 leaching. In this area, a significant standard deviation of 0.04-17.2% was also obtained, showing greater homogeneity in the location of the lateritic crust, which contrasts with slopes, evidencing topographic control.

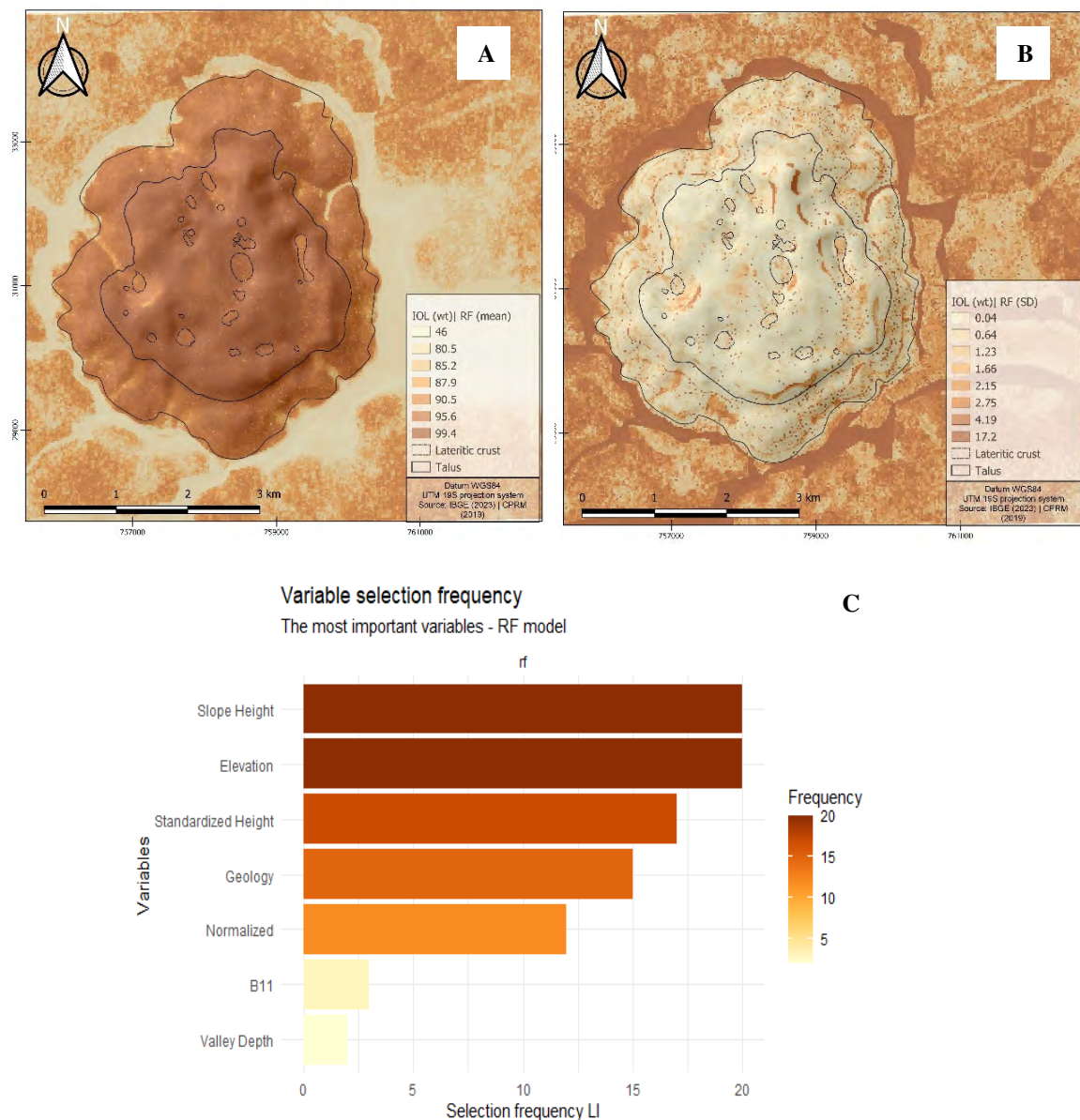


Figure 29. Spatial distribution of the IOL index - mean map (A), standard deviation map (B) and Variable Importance Ranking (C), using the Random Forest (RF) model in the Morro dos Seis Lagos area, São Gabriel da Cachoeira, Amazonas State, Brazil. Source: Prepared by the author (2025).

The covariates with the highest frequency of importance (Figure 29 C) in predicting the IOL index were Slope Height, Elevation Standardized Height, Geology, B11 (Sentinel 2A), and Valley Depth. It should be noted that the petroferic formations are intrinsically related to the morphometric covariates of the terrain that provide local geological and geomorphological conditions and will therefore reflect on the intensity of weathering and the stage of pedogenesis. The study by Mello et al. (2023) also highlights the importance of the morphometric covariates Slope Height, Digital Elevation Model, Terrain Surface Convexity in the prediction of weathering intensity by the RF model, contributing more than 75% to the decrease in average accuracy, while the digital elevation model (DEM) contributed 50%.

The degree of importance for the Goethite, Ferroginous and Clay Detection (ASTER Sensor) indices - can be associated with low K_i , the presence of iron oxides (hematite and

goethite) typical of soils with low Ki (SANTOS et al., 2018), being a spectral signature to identify zones with a high degree of lateritization.

The average IOL values ranged from 46 to 99.4 (%) (Figure 29), being considered a moderately to highly weathered material. According to Sahoo et al. (2017), catchment laterites fall within strong lateritization. Still, weathered sediments fall between the strong to weak field divisions, and some are in the kaolinized field, indicating chemical reworking of sediments. Thus, high IOL values are associated with Fe concentrations, reinforcing it as an indicator of extremely weathered and reworked areas.

It is worth highlighting the consideration of climatic influence from a paleoenvironmental perspective, based on the dynamics of ancient and poorly eroded surfaces in the laterite crust, such as the importance of the Standardized Height and Slope Height covariates. These covariates favor the process of silica leaching and contribute to the supergene process of mineral enrichment in the study area. As pointed out by Giovannini et al. (2020), the local geology controls the availability of minerals such as Fe and Al, conditioning the secondary mineralogy, forming a lithotype of ferricrete horizons, iron-rich material, and low Al contents. Porto, Bastos Neto, and Bento (2025) point out that supergene processes promote the enrichment of niobium (Nb) through the intense chemical weathering that affects carbonatites exposed at the surface, especially in tropical regions. Thus, the continuous action of rainwater and percolation promotes the leaching of mobile elements (Ca, Na, K, and Si), leaving behind more resistant minerals such as Nb oxides. In contrast, the Valley Depth covariate stands out mainly in areas of depression and lakes in the laterite crust, creating environments conducive to the deposition of highly weathered materials from the higher parts of the landscape (retransported laterite crust).

For the CIA index (Figure 30), there was a range of average contents between 63.4 and 94.5, indicating a moderate to advanced stage of weathering that predominates in environments with hot and humid climate conditions, as pointed out by Huang et al. 2024. From a paleoclimate perspective, values closer to 100 (%) indicate older soils or sediments that have undergone intense leaching. In the Chaohu basin (HUANG et al., 2024), values > 80 (%) of CIA were found in sediments, which are considered a reflection of paleoclimatic characteristics and the chemical composition of the bedrock.

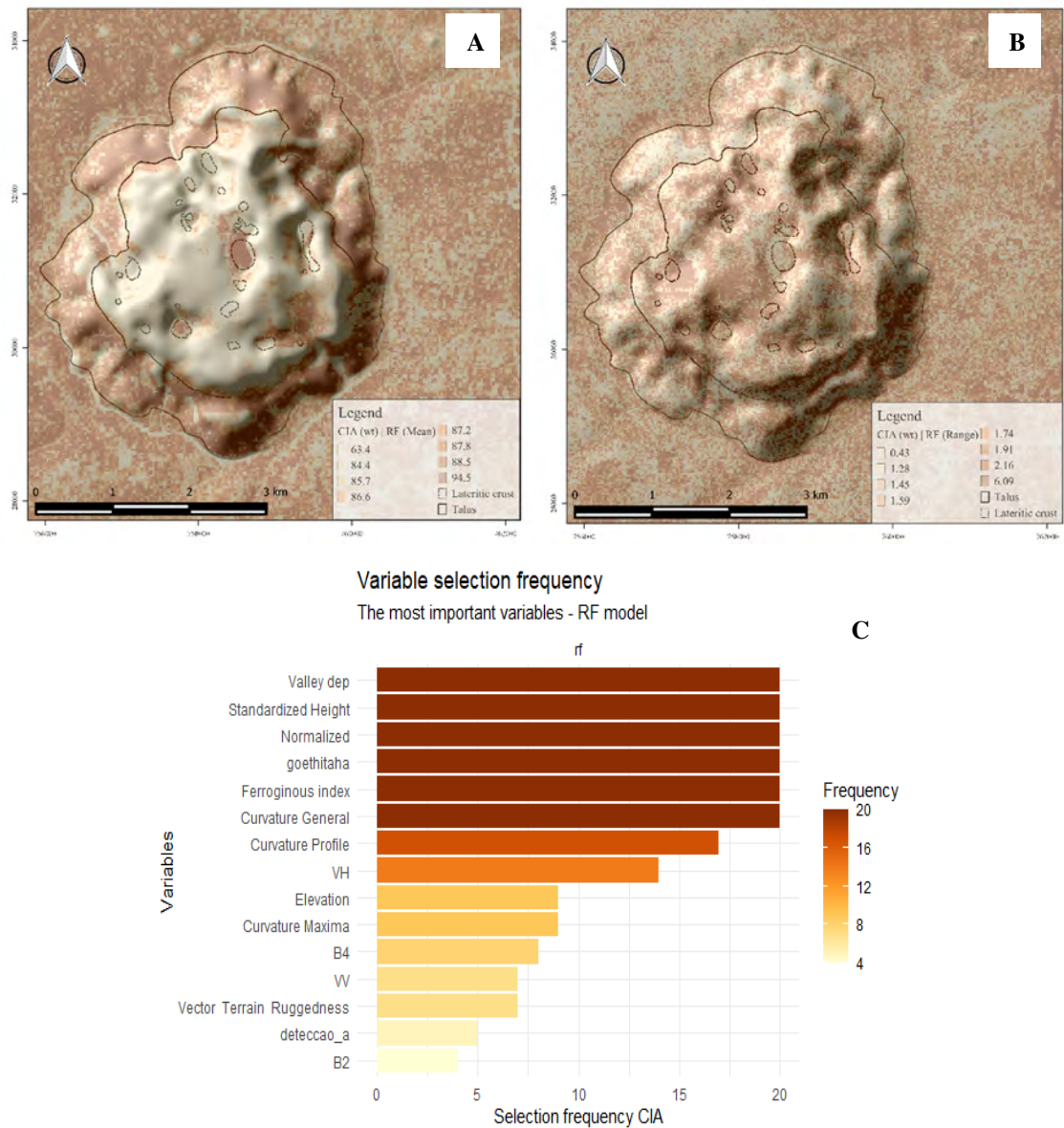


Figure 30. Spatial distribution of the CIA index - mean map (A), standard deviation map (B), and Variable Importance Ranking (C), using the Random Forest (RF) model in the Morro dos Seis Lagos area, São Gabriel da Cachoeira, Amazonas State, Brazil. Source: Prepared by the author (2025).

In the Talus area, depressions, and lakes, the CIA (87.2 - 94.5%) indicates moderate to advanced chemical weathering, characterized by intense leaching of mobile cations and under the influence of restrictive hydrological conditions and redox potential that favor the occurrence of autochthonous geochemical processes, such as precipitation of secondary carbonates and phosphates. Sahoo et al. (2017) points out that the chemical composition of lake bottom sediments is affected by several factors, such as hydraulic classification, weathering rates, erosion, transport pathways, source rock composition and the tectonic configuration of the basin, and post-depositional diagenetic changes, including the precipitation of Fe-oxyhydroxides (e.g. goethite, jarosite and magnetite), sulfides, carbonates, can also control the

redistribution of trace elements in lake sediments. Given that the geochemical signature of this material is a tool with the potential to reconstruct weathering histories and distribution patterns (TAYLOR; MCLENNAN, 1986).

On the other hand, in the lateritic crust, 63.4-86.6 was obtained, indicative of areas of moderate to intense weathering, containing mature lateritic soils in the condition of intense silica leaching and base removal. Studies by Adama et al. (2025) on lateritic soils in the district of Koubou, East Cameroon, also showed that the lateritic soil samples studied showed CIA values ≥ 90 , indicating a hot and humid climate influenced by intense chemical weathering. It should be noted that mineral deposits in laterite profile areas are formed from the decomposition of primary minerals through the removal of bases via leaching and are determined based on the chemistry of the parent rock, climatic environment, and mineralogical characteristics. During this process, the mobile elements contained in the parent rocks are removed, leaving behind immobile elements that accumulate as stable eluvial/residual minerals (SANTORO et al., 2022).

For CIA, the ranking of importance highlighted the Goethite, Ferrous indices, B4 and B2, indicating the spectral signature of the presence of Fe oxyhydroxides, when correlated with CIA > 85. And for SAR, the VV and VH bands helped to detect rough surfaces, probably due to high backscattering in conditions of greater moisture retention and compaction due to the presence of exposed laterite crusts. Studies by Stanyer et al. (2025) found that the spectral response of C-band SAR data depends on soil texture properties and strongly suggest that C-band VV backscatter depends on soil texture, with average gradients ranging from a high sensitivity of 1.69% VWC SM per dB S1 response for a sandy soil to a low sensitivity of 4.81% VWC SM per dB S1 response for a clay soil.

In the case of the CIW index, which obtained an average variation of 70.06- 95.65 in the total area (Figure 31), it was found that the high values in the talus area reflect the interaction between geomorphological and pedogenetic processes.

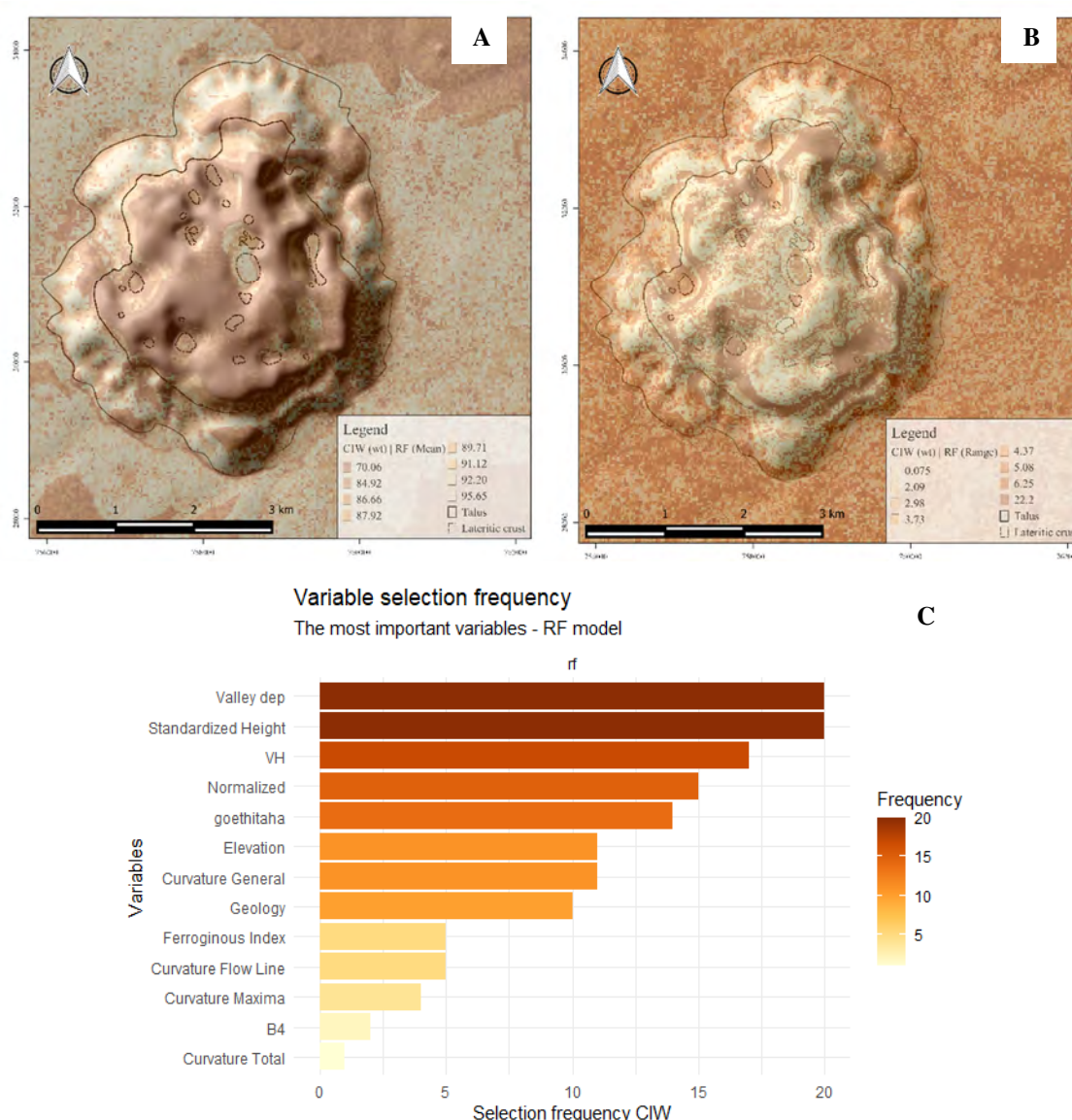


Figure 31. Spatial distribution of the CIW index - mean map (A), standard deviation map (B), and Variable Importance Ranking (C), using the Random Forest (RF) model in the Morro dos Seis Lagos area, São Gabriel da Cachoeira, Amazonas State, Brazil. Source: Prepared by the author (2025).

The Talus region presented CIW values > 87.92 (%), inferring the influence of pedogenic restructuring of this colluvial material, which already indicates significant loss of bases and accumulation of Al_2O_3 , giving a higher value to the CIW index. Like the CIA, the CIW is close to 100%, indicating conditions of extreme weathering. Studies conducted by Goydaragh et al. (2021) found that the CIA and CIW were the most suitable for assessing the intensity of weathering. The RF models confirmed the alteration of potassium minerals and the relative increase in smectites, characteristic of tropical soils.

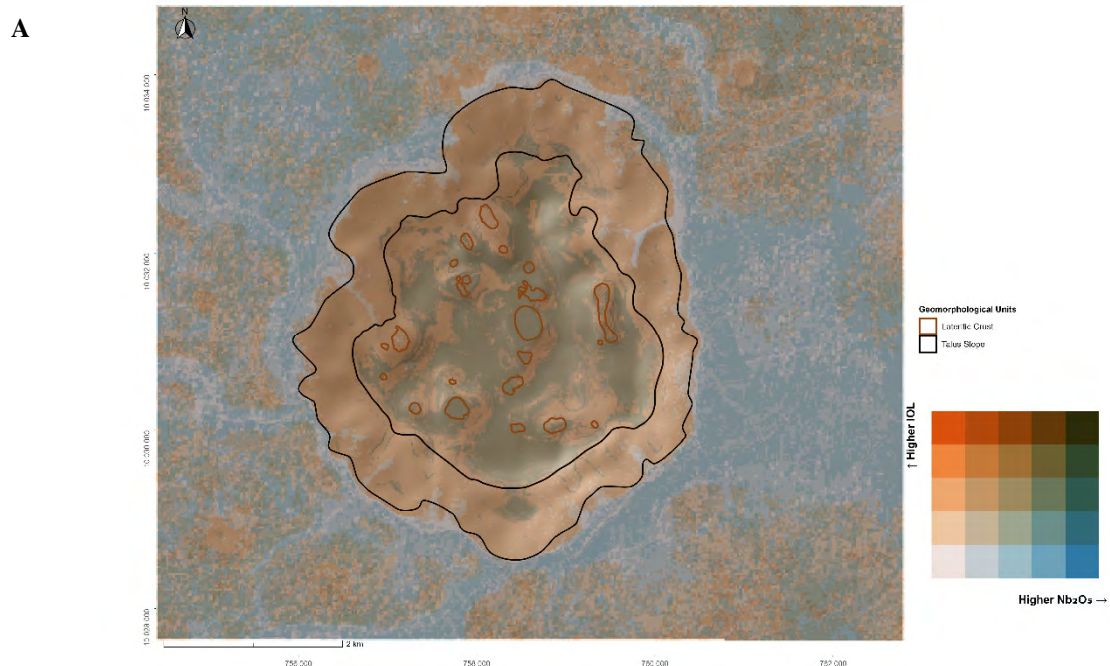
In the lateritic crust, the CIW ranged from 70 to 86 (%), which is relatively low considering the advanced stage of alteration. This occurrence correlates with geochemical stability due to the high residual concentration of Nb_2O_5 and TiO_2 , elements that are stable under conditions of intense weathering. Laterization controls the relative enrichment and distribution of major and trace elements in petroferic formations (EL-LEIL; EL-DESOKY;

SHAFEA, 2017), resulting in the depletion of mobile elements (Na, K, Mg, Ca, and Si) and the relative accumulation of Fe and Al. It is important to note that Fe enrichment is a factor not considered in this index.

5.5.3 Bivariate spatial distribution of weathering rates versus Nb₂O₅, P₂O₃, and TiO₂

Due to differences in local pedogeomorphological conditions, the influence of pedogenetic processes on the spatial distribution of chemical element concentrations was confirmed (Figure 32). It was identified that the sites with the most significant enrichment are in the area of the lateritic crust for both Nb₂O₅ (A, B and C), TiO₂ (A, B and C) and P₂O₅ (A, B, and C), when crossed with the weathering indices IOL, CIA, and CIW, which is strong evidence of the intense laterization process. A study by Rodrigues et al. (2025) in Morro dos Seis Lagos highlights the concentration/mineralization of elements such as Nb₂O₅, TiO₂, and MnO in the laterite crust, and consequently, in the Talus area, shows high levels of Al₂O₃ and SiO₂ associated with the local physiographic aspects.

In this context, Iza et al. (2018), in the southwest of the Brazilian Amazon, also identified a strong spatial correlation between geochemical indices such as the Lateritic index and the concentrations of elements such as I, Ce, Co, Cr, Cu, and Fe, observed in geochemical associations (soil and stream sediments), using lower concentrations of alkali as a geochemical criterion for determining highly weathered areas and identifying lateritic duricrust.



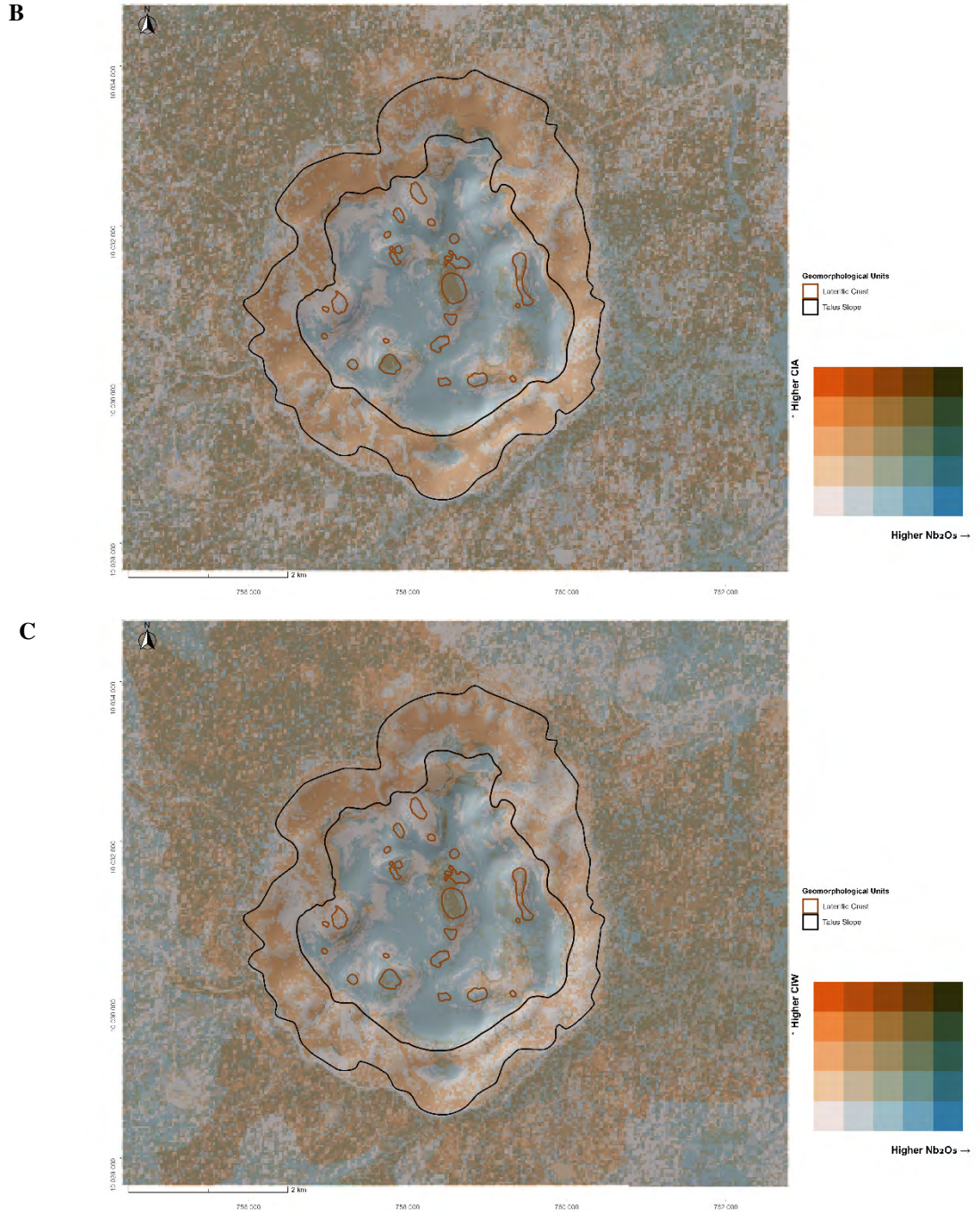


Figure 32. Geochemical bivariate maps (A) IOL versus Nb_2O_5 ; (B) CIA versus Nb_2O_5 ; and (C) CIW versus Nb_2O_5 in the Morro dos Seis Lagos area, São Gabriel da Cachoeira, Amazonas State, Brazil. Source: Prepared by the author (2025).

The CIA and CIW indices (Figure 32) showed conditions of areas in the lateritic crust unit and transition areas with highly weathered thallus and mature laterization with areas of $\text{CIA} > 60$ (%), being classified as low to extremely weathered according to Nesbitt and Young (1982), and also corroborated by the relationship between CIA and CIW versus Nb_2O_5 (Figure 32 B and C), with the strong occurrence of Nb_2O_5 and TiO_2 in the lateritic crust and areas of depressions and lakes. Osinuga et al. (2025) also observed similar patterns with the CIA, MIA, and CIW indices in all soil profiles, classified as lightly weathered at intermediate stages of development, indicating the complete conversion of primary minerals into their respective

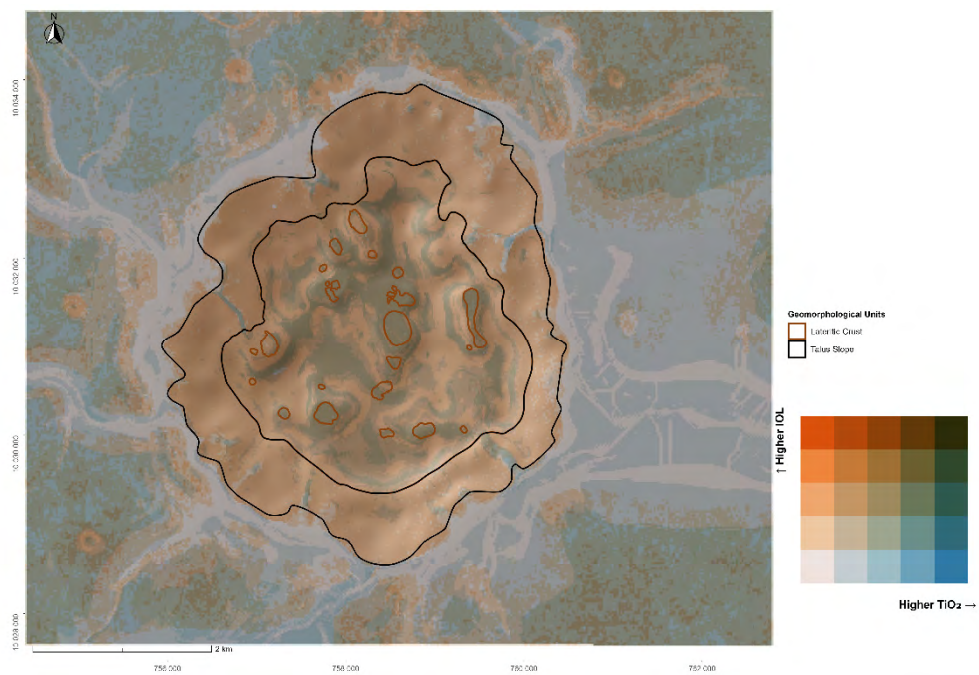
weathering products.

The spatial pattern of TiO_2 (Figure 33) follows the geomorphological unit of Morro dos Seis Lagos, with higher concentrations in the central lateritic crust, while relatively lower values occur in the lower areas of talus and depressions. In the central areas of the lateritic crust, it was found that the increase in TiO_2 follows the dynamics of areas of intense weathering (high IOL, CIA, and CIW), indicating a residual enrichment of the element with control of lateritic formation, as also pointed out in the studies by Bhaskar et al. (2017) in their analysis of lateritic soils in the Western Ghats found TiO_2 increasing from 9.75% (surface) to 15.75% (basal horizon), indicating the occurrence of progressive leaching of silica and bases with residual Ti concentration as a proxy for indicating hard laterization conditions.

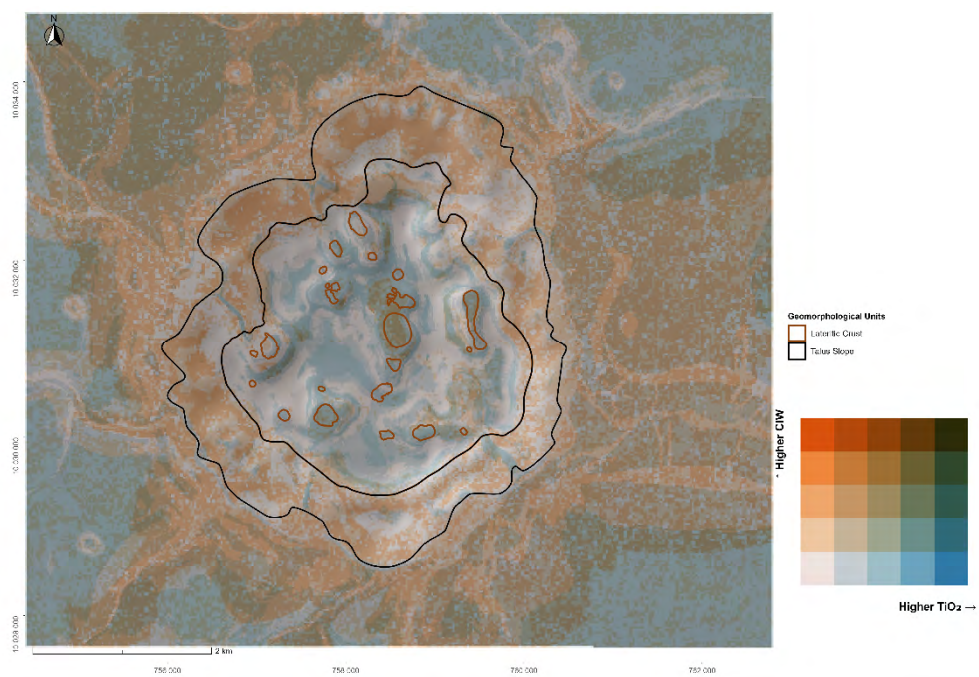
Similar to what was observed in the spatial pattern of Nb_2O_5 , the high TiO_2 contents associated with more stable geomorphological compartments evidences the action of selective supergene processes. Intense leaching of silica and bases favors the residual concentration of Ti in secondary phases of high stability, typical of ancient and thick horizons (Santos et al. 2016; Silveira et al. 2022). In contrast, in areas and talus, moderate TiO_2 levels reflect colluvial input, mantle rejuvenation, and the influence of less evolved materials (Ito et al., 2021).

The unique condition of occurrence of Nb and Ti mineralization with most of the reported trace elements, including REEs, in areas of lateritic crust in the studies by Giovannini et al. (2017), Bento et al. (2022), Rodrigues et al. (2024) in Morro dos Seis Lagos, are indicative of stages of intense weathering of carbonatites, and consequently of the laterization process controlling the enrichment and distribution of minerals in the area. The origin of the Nb found in the area is identified as rutile (11-27% by weight of Nb_2O_5), brookite (10-16% by weight of Nb_2O_5), goethite (up to 3% by weight of Nb_2O_5), and Cerianite (up to 0.75% by weight of Nb_2O_5) (GIOVANNINI et al., 2017). Thus, the intense weathering of carbonatites corroborates the enrichment of Nb that host minerals of the pyrochlore supergroup ($\text{A}_{2-x}\text{B}_2(\text{O}, \text{OH})_6(\text{OH}, \text{F}, \text{H}_2\text{O})_{1-y}$) lead to the enrichment of Nb about the primary carbonatite (MITCHELL, 2015). And through the speciation process, where the A site of the pyrochlore can be occupied by various cations, including Na^+ , Ca^{2+} , Mn^{2+} , Sr^{2+} , Ba^{2+} , Pb^{2+} , or rare earth elements (REE^{3+}). The B site is occupied by cations coordinated six times, such as Nb^{5+} , Ta^{5+} , Ti^{4+} , Fe^{3+} , Al^{3+} , Si^{4+} (ATENCIO et al., 2010).

A



B



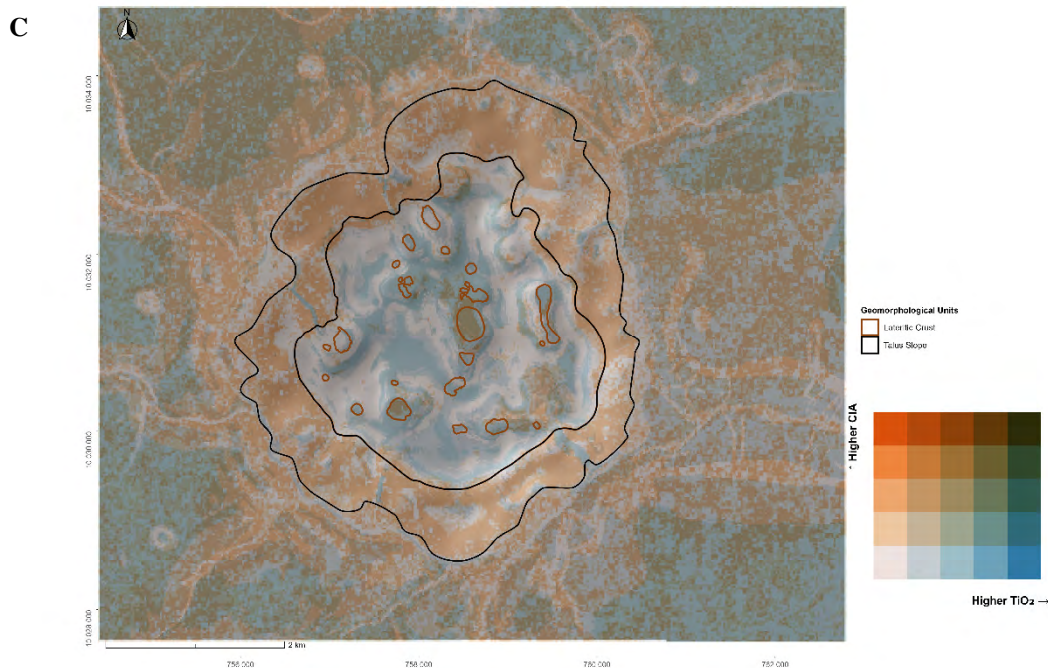
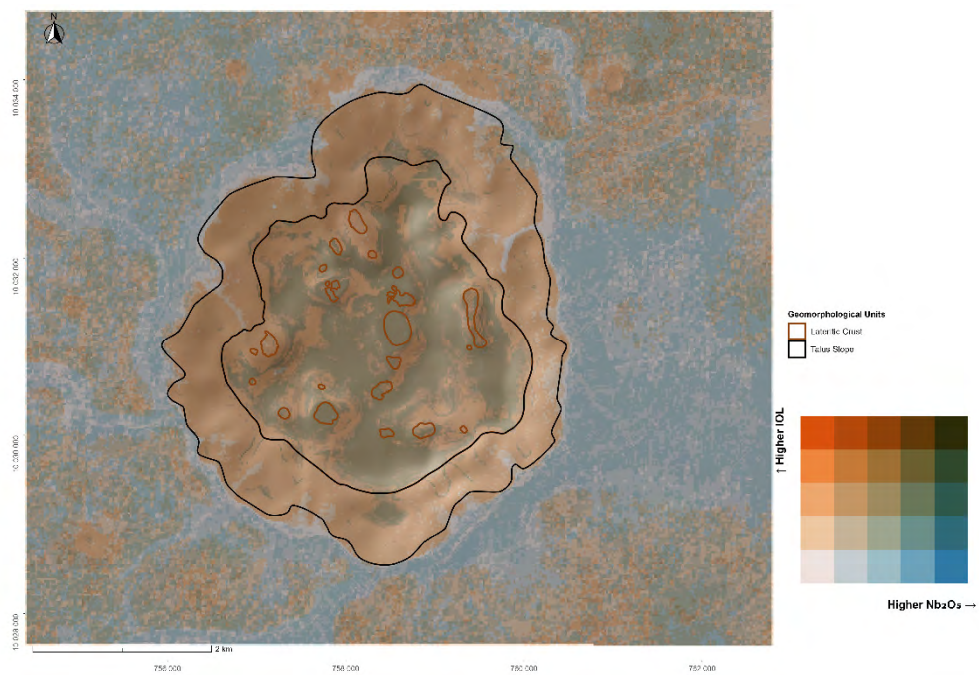


Figure 33. Geochemical bivariate maps (A) IOL versus TiO_2 ; (B) CIA versus TiO_2 ; and (C) CIW versus TiO_2 in the Morro dos Seis Lagos area, São Gabriel da Cachoeira, Amazonas State, Brazil. Source: Prepared by the author (2025).

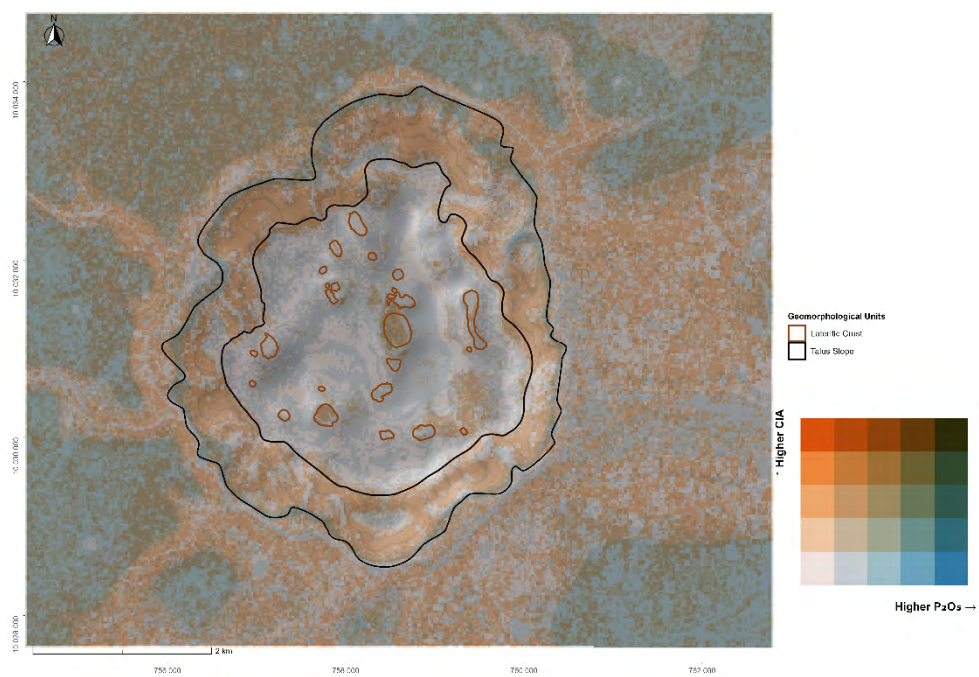
The spatial behavior of P_2O_3 (Figure 34) corroborates the geochemical indices and reflects the topographic and lithogeochemical control of the soil landscape in the Amazon, where the strong negative correlation between P and sesquioxides is associated with the degree of weathering and the mineralogy of the clay fraction. Therefore, marked depletion is observed in lateritic crusts (IOL - high), where phosphorus has been intensely leached and/or adsorbed by Fe and Al oxides, being inversely proportional to the talus area, with also high contents (2.89%) in weathered colluvial slopes. According to Campos et al. (2018), high point anomalies in the crust reflect the lithogeochemical heterogeneities of the carbonate parent rock, without altering the general trend of progressive P loss with lateritization.

In highly weathered tropical soils, such as Oxisols, the dynamics of phosphorus are typical of a highly mobile element, being rapidly leached under acidic conditions ($\text{pH} < 5.5$), which are predominant in the region, and strongly adsorbed on surfaces with variable loads of Fe oxides (goethite, hematite), while it is preserved on slopes due to shorter residence time and colluvial transport (Novais et al., 2007).

A



B



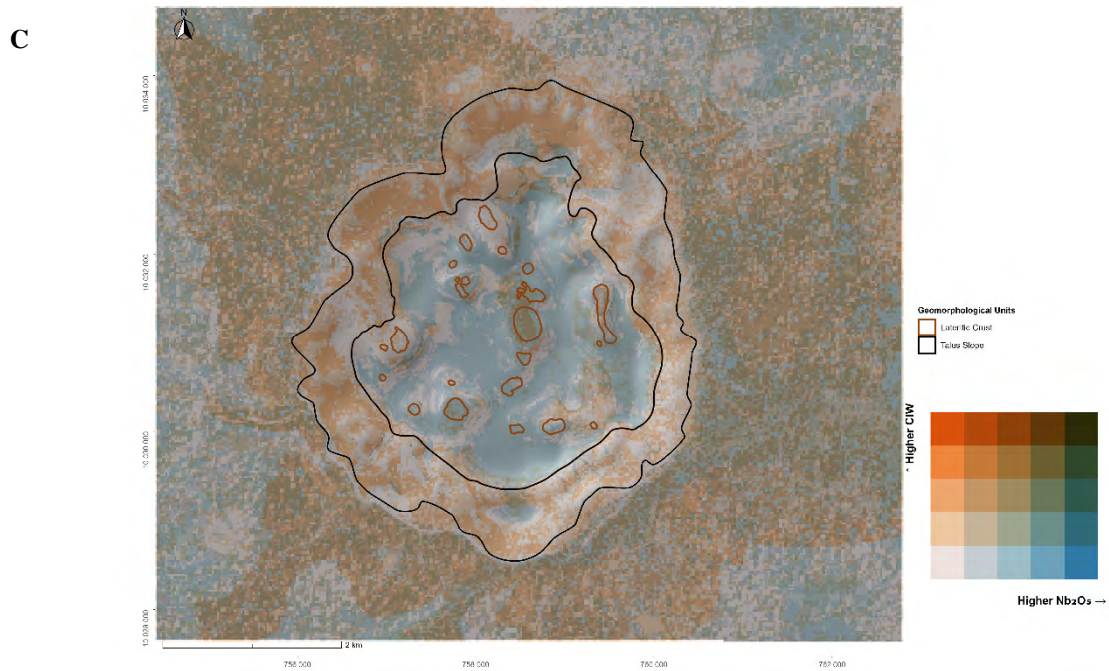


Figure 34. Geochemical bivariate maps (A) IOL versus P_2O_3 ; (B) CIA versus P_2O_3 ; and (C) CIW versus P_2O_3 in the Morro dos Seis Lagos area, São Gabriel da Cachoeira, Amazonas State, Brazil. Source: Prepared by the author (2025).

5.6 CONCLUSIONS

The research fills a critical gap by demonstrating that the IOL index obtained maximum values in lateritic crusts (lateritization). At the same time, CIA and CIW exhibited patterns of early stages predominantly in talus. We examined the influence of geochemical indices (IOL, CIA, and CIW) on the geological units of lateritic crust and talus, combined with the elements Nb_2O_5 and TiO_2 . The most intense weathering occurred in the lateritic crust region for the IOL, CIW, and CIA indexes, while the lowest intensities were observed in the talus area.

The Random Forest model showed better performance in predicting the IOL and CIW weathering indexes when compared to Cubist, in terms of accuracy, sensitivity, and sensitivity. The most important covariates for prediction were the combination of morphometric variables (Elevation, Slope Height, and Valley Depth), the categorical Geology and Ferruginous index, and the VV and VH bands, which capture information about roughness, terrain moisture, and mineralogical composition, contributing to the degree of weathering.

It was found that niobiferous carbonatites, under intense lateritization in the high relief areas, generate secondary residual enrichment of Nb_2O_5 and TiO_2 in lateritic crusts, highlighting how the source material controls the spatial patterns of mineralization in this formation. This soil-landscape dynamic, associated with the intense pedogenic processes of the carbonatitic facies, directly reflects the spatial distribution of the contents of these elements, correlated with the local lithology and the properties of the derived soils.

This study makes an unprecedented contribution to the field of digital mapping and soils by integrating geochemical analyses and pedogenetic processes in the context of *Morros dos Seis Lagos*, in Amazon State (BR), a difficult-to-access area with complex and particular geology, which is still little explored in national and international literature. The research fills a gap in the understanding of weathering patterns in petroferic formations under the influence of mineralization associated with elements of economic interest.

6. CONCLUSÕES GERAIS

Considerando que o Brasil possui reservas mundiais de nióbio e a segunda maior reserva de elementos de terras raras (ETR), esse fator posiciona o país como um protagonista geopolítico na disputa por minerais críticos essenciais à transição energética e à soberania tecnológica.

Esta tese estabelece fronteira de conhecimento entre Ciência do Solo e Geoquímica, ao caracterizar e integrar processos pedogenéticos supergênicos no Morro dos Seis Lagos (AM) com o exemplar de ferrocarbonatitos, por meio da predição espacial de teores de Nb_2O_5 , TiO_2 , P_2O_5 e outros elementos, bem como de índices geoquímicos de intemperismo (CIA, CIW e IOL). Essa abordagem incorpora os vetores lateríticos que controlam a redistribuição de elementos economicamente estratégicos em solos tropicais. Geociências integradas subsidiam o planejamento ambiental fronteiriço e políticas de apoio à soberania mineral nacional.

Considerando a base teórica e experimental do conjunto dos três capítulos, há evidências de que o Morro dos Seis Lagos, em São Gabriel da Cachoeira, localizado no estado do Amazonas, está diretamente associado à geoquímica do material de origem (relevo residual), que controla as características morfométricas e os processos pedogenéticos que influenciam a mobilização e a imobilização química. Esse contexto influencia diretamente a mineralização e a redistribuição dos elementos, em especial Al_2O_3 , Fe_2O_3 , MnO , Nb_2O_5 e TiO_2 , cujo comportamento mostrou tendência condicionada pelos compartimentos geomorfológicos e pedológicos. A integração entre os capítulos evidencia a interação entre geoquímica e o controle topográfico (capítulo I), associado à otimização de delineamento amostral mais representativo quanto à captura de padrões na predição de elementos (capítulo II) e à confirmação da condição de pedogenética e intemperismo extremo na crosta laterítica versus talus (capítulo III), com enriquecimento residual e Nb_2O_5 e TiO_2 condicionado pela dinâmica de supergência.

Os resultados apresentados no capítulo I constataam maior influência de controle topográfico e influência da drenagem na predição da mineralização dos elementos da área de estudo através do modelo *Recursive Feature Elimination* (REF), onde o ranking de importância indicou as seguintes covariáveis: *Elevation*, *Ls-Factor*, *Multiresolution Index of Valley Bottom Flatness* (MRVBF), *SAGA Wetness Index* e detecção da alteração argilítica avançada. Como perspectivas futuras frisa-se que os modelos podem ser aperfeiçoados a partir da junção de novos dados na malha amostral e com o uso de outros algoritmos de predição.

No capítulo II, foi constatada a influência de modelos e delineamentos de amostragem na precisão de predição de compostos químicos selecionados (Al_2O_3 , Fe_2O_3 , MnO , Nb_2O_5 , TiO_2 e SiO_2). Sendo a predição influenciada pelo cenário de amostragem, porém apenas ligeiramente, e o grau de influência variou por elemento. Obtivemos destaque com o desenho amostral do Cenário IV, no qual o modelo RF obteve o melhor desempenho, provavelmente em função da característica das amostras de treinamento ser maior do que as amostras de validação, se ressalta que as amostras de treinamento e validação estavam contidas em uma pequena área representativa com malha adensada.

É importante destacar que a heterogeneidade espacial das formações geológicas pode ter contribuído para o baixo desempenho observado em todos os modelos, bem como para a

falta de interação com os cenários de amostragem. Ressalta-se que o delineamento amostral adotado no presente estudo, em formato de "estrela", é proveniente de estágios anteriores da Pedometria, ignora os pressupostos da pedometria (regionalização variável, continuidade e estrutura espacial). A densidade, insuficiente continuidade e dependência espacial, ao contrário do estudo. Os demais delineamentos também apresentaram um desafio adicional ao mapeamento de elementos, devido à densidade inadequada (amostras muito distanciadas entre si), dificultando a modelagem. Assim, ressalta-se que compreender a assinatura da geologia nas concentrações e distribuições espaciais de elementos químicos, encontrados nas condições de alto intemperismo observadas no Morro de Seis Lagos, é um fator tão importante na caracterização dos processos pedogenéticos quanto as características do terreno e das paisagens.

No capítulo III, se investigou a influência dos índices geoquímicos (IOL, CIA e CIW), nas unidades geológicas crosta laterítica e tálus, combinadas com os elementos Nb_2O_5 e TiO_2 . A maior intensidade de intemperismo ocorreu na região da crosta laterítica, para os índices IOL, CIW e CIA, enquanto as menores intensidades ocorreram na área do tálus. Assim, como nos estudos anteriores de predição de elementos químicos, o modelo RF obteve melhor desempenho na previsão dos índices de intemperismo IOL e CIW em comparação com o Cubist, em termos de precisão e sensibilidade. Enfatiza-se a contribuição na predição dos índices resultantes da combinação das covariáveis morfométricas, em conjunto com índices de sensoriamento remoto, em especial o índice Ferruginoso e as bandas VV e VH, que auxiliam na identificação de informações sobre rugosidade, umidade do terreno e composição mineralógica, contribuem com as informações sobre o grau de intemperismo. Nesse sentido, foi possível evidenciar que as áreas com intenso intemperismo da fácies do carbonatito e lateritização estão associadas à composição química do material parental e contribuíram, assim, para o enriquecimento de óxidos secundários (Nb_2O_5 e TiO_2) nas partes mais altas do relevo, na unidade da crosta laterítica.

7. CONSIDERAÇÕES FINAIS

Recomenda-se que investigações futuras priorizem a definição dos tipos de lateritas e materiais aflorantes, com análise dos padrões composicionais (assinaturas geoquímicas e padrões de ocorrência), para a obtenção das subclasses litogeoquímicas (*Laterite*, *Nb Laterite*, *Marginal lake* e *Talus*) da área de estudo. Para fins de mapeamento de subclasses, as perspectivas incluem o uso de outras metodologias, como a estimativa de probabilidade para cada pixel, ajudando a identificar onde as subclasses podem ocorrer e também avaliando a incerteza para aprimorar os modelos.

Além disso, é importante explorar outros algoritmos de previsão, como as redes neurais artificiais mais robustas (*Deep Neural Networks* (DNNs), *Convolutional Neural Networks* (CNNs), *Deep Belief Networks* (DBN) entre outros), por exemplo. Paralelamente, propõe-se a modelagem regional com o conjunto de bancos de dados dos morros adjacentes, como Morro do Meio e Morro do Norte, em São Gabriel da Cachoeira.

8. REFERÊNCIAS BIBLIOGRÁFICAS

- ABREU, J. M. F. **Modelos de Data Mining para suporte a avaliações sensoriais do vinho**. 2016. 15f. Dissertação (Mestre em Engenharia Informática) - Instituto Superior de Engenharia do Porto, Portugal. 2016.
- 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.
- ALBUQUERQUE, M. F. D. S.; HORBE, A. M. C. Mineralogia, geoquímica e evolução da lateritização em Apuí, sudeste do Amazonas. *Brazilian Journal of Geology*, v. 45, n. 4, p. 569–590, 2015.
- ANJOS, L. H. C. dos; PEREIRA, M. G.; PÉREZ, D. V.; RAMOS, D. P. Caracterização e classificação de Plintossolos no município de Pinheiro-MA. *Revista Brasileira de Ciência do Solo, Viçosa*, v. 31, n. 5, p. 1035-1044, set./out. 2007.
- ARROUAYS, D.; LEENAARS, J. G. B.; RICHER-DE-FORGES, A. C.; ADHIKARI, K.; BALLABIO, C.; GREVE, M.; GRUNDY, M.; GUERRERO, E.; HEMPEL, J.; HENGL, T.; HEUVELINK, G.; BATJES, N.; CARVALHO, E.; HARTEMINK, A.; HEWITT, A.; HONG, S.; KRASILNIKOV, P.; LAGACHERIE, P.; LELYK, G.; LIBOHOVA, Z.; LILLY, A.; MCBRATNEY, A.; MCKENZIE, N.; VASQUEZ, G. M.; MULDER, V. L.; MINASNY, B.; MONTANARELLA, L.; ODEH, I.; PADARIAN, J.; POGGIO, L.; ROUDIER, P.; SABY, N.; SAVIN, I.; SEARLE, R.; SOLBOVOY, V.; THOMPSON, J.; SMITH, S.; SULAEMAN, Y.; VINTILA, R.; ROSSEL, R. V.; WILSON, P.; ZHANG, G.; SWERTS, M.; OORTS, K.; KARKLINS, A.; FENG, L.; NAVARRO, A. R. I.; LEVIN, A.; LAKTIONOVA, T.; DELL'ACQUA, M.; SUVANNANG, N.; RUAM, W.; PRASAD, J.; PATIL, N.; HUSNJAK, S.; PÁSZTOR, L.; OKX, J.; HALLETT, S.; KEAY, C.; FAREWELL, T.; LILJA, H.; JUILLERET, J.; MARX, S.; TAKATA, Y.; KAZUYUKI, Y.; MANSUY, N.; PANAGOS, P.; LIEDEKERKE, M. V.; SKALSKY, R.; SOBOCKA, J.; KOBZA, J.; EFTEKHARI, K.; ALAVIPANAH, S. K.; MOUSSADEK, R.; BADRAOUI, M.; SILVA, M. da.; PATERSON, G.; GONÇALVES, M. da C.; THEOCHAROPOULOS, S.; YEMEFACK, M.; TEDOU, S.; VRSCAJ, B.; GROB, U.; KOZÁK, J.; BORUVKA, L.; DOBOS, E.; TABOADA, M.; MORETTI, L.; RODRIGUEZ, D. Soil legacy data rescue via GlobalSoilMap and other international and national initiatives. *GeoResJ*, v. 14, p. 1-19, 2017.
- ATENCIO, D.; ANDRADE, M. B.; CHRISTY, A. G.; GIERÉ, R.; KARTASHOV, P. M. The pyrochlore supergroup of minerals: nomenclature. *The Canadian Mineralogist*, v. 48, n. 3, p. 673-698, 2010.
- BABAEI, K.; FATTAH-ALHOSSEINI, A.; CHAHARMAHALI, R. A review on plasma electrolytic oxidation (PEO) of niobium: Mechanism, properties and applications. *Surfaces and Interfaces*, v. 21, p. 100719, 2020.
- BAHLBURG, H.; DOBRZINSKI, N. A review of the Chemical Index of Alteration (CIA) and its application to the study of Neoproterozoic glacial deposits and climate transitions. In: ARNAUD, E.; HALVERSON, G. P.; SHIELDS-ZHOU, G. (eds) **The Geological Record of Neoproterozoic Glaciations**. Geological Society, London, Memoirs, 36, 81–92. 2011.
- BASSER, J. R. Q. **Learning with continuous classes**. Sydney, Australia, 1992. 343–348 p.
- 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. e20210031, 2022.

- BIGARELLA, J. J.; BECKER, R. D.; SANTOS, G. D.; PASSOS, E.; SUGUIO, K. **Estrutura e origem das paisagens tropicais e subtropicais**. Florianópolis: Ufsc. 1994.
- BOURMAN, R. P. Perennial problems in the study of laterite: a review. **Australian Journal of Earth Sciences**, v. 40, n. 4, p. 387-401, 1993.
- BOURMAN, R. P.; OLLIER, C. D. A critique of the Schellmann definition and classification of 'laterite'. **Catena**, v. 47, n. 2, p. 117-131, 2002.
- BHASKAR, B. P.; TIWARI, G. Geochemical interpretations of laterite associated soils of East Coast Andhra Pradesh. **Chemical Science Review and Letters**, v. 6, n. 21, p. 187-197, 2017.
- BRADY, N. C.; WEIL, R. R. **Elementos da natureza e propriedades dos solos**. Bookman Editora, 2009.
- BRASIL. **Constituição da República Federativa do Brasil**. Título VIII: Da Ordem Social. Capítulo VI: Do Meio Ambiente. Brasília, DF: Presidência da República, 5 out. 1988. Disponível em: http://www.planalto.gov.br/ccivil_03/Constituicao/Constituicao.htm. Acesso em: 02 ago. 2023.
- 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.
- BRASIL. **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.
- BRASIL. **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.
- BRASIL. **Lei nº 6.634, de 02 de maio de 1979a**. 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, Jan. 2001.
- CALEGARI, M. R. **Ocorrência e significado paleoambiental do horizonte A húmico em Latossolos**. 2008. 259f. Tese (Doutorado em Agronomia) - Universidade de São Paulo, Piracicaba, 259p., 2008.
- CAMPBELL, A. M.; RACAULT, M. F.; GOULT, S.; LAURENSEN, A. Cholera risk: A machine learning approach applied to essential climate variables. **International Journal of Environmental Research and Public Health**, v. 17, n. 24, p. 9378, 2020.
- CAMPOS, D. S. de. **Duricrusts ferruginosos e aluminosos e relações pedologia geomorfologia-geologia no Alto Jequitinhonha (MG)**. 2023. 230f. Tese (Doutorado em Ciências) - Universidade de São Paulo, Piracicaba, 23p., 2023.
- CAMPOS, M. de; ANTONANGELO, J. A.; VAN DER ZEE, S. E. A. T. M.; ALLEONI, L. R. F. Degree of phosphate saturation in highly weathered tropical soils. **Agricultural Water Management**, Amsterdam, v. 206, p. 135-146, 2018.
- CARVALHO JUNIOR, W. de.; PEREIRA, N. R.; BHERING, S. B.; CALDERANO FILHO, B.; CHAGAS, C. da S.; PINHEIRO, H. S. K.; PEREIRA, J. R.; LOPES, C. H. L.; LEAL, R. B.

Exploratory Analysis from Harmonized Legacy Soil Data to Support Digital Soil Mapping in Brazilian Midwest. In: **Pedometrics in Brazil**. Cham: Springer Nature Switzerland, 2024. p. 115-127.

CARVALHO JUNIOR, W. de; LAGACHERIE, P.; CHAGAS, C. da S.; CALDERANO FILHO, B.; BHERING, S. B. A regional-scale assessment of digital mapping of soil attributes in a tropical hillslope environment. **Geoderma**, v. 232–234, p. 479–486, 2014.

CARVALHO JUNIOR, W.; PEREIRA, N.R.; FERNANDES FILHO E.I.; CALDERANO FILHO, B.; PINHEIRO, H.S.K.; CHAGAS, C.S.; BHERING, S.B.; PEREIRA, V.R.; LAWAL, S. Sample design effects on soil unit prediction with machine: randomness, uncertainty, and majority map. **Revista Brasileira de Ciência do Solo**, v. 44, p. e0190120, 2020.

CARVALHO, C. C. N. de; NUNES, F. C.; ANTUNES, M. A. H. Histórico do levantamento de solos no Brasil: da industrialização brasileira à era da informação. **Revista Brasileira de Cartografia**, Rio de Janeiro, n. 65/5, p. 997-1013, set./out. 2013.

CHAGAS, C. D. S. **Mapeamento digital de solos por correlação ambiental e redes neurais em uma bacia hidrográfica no domínio de mar de morros**. 2006. 286 f., Tese (Doutorado em Solos e Nutrição de Plantas) - Universidade Federal de Viçosa, Viçosa. 2006. 286p.

CHAKHMOURADIAN, A. R.; MITCHELL, R. H.; YANG, P. Carbonatite-hosted niobium deposit at Aley, northern British Columbia, Canada. **Ore Geology Reviews**, Amsterdam, v. 64, p. 642–666, 2015. DOI: 10.1016/j.oregeorev.2014.07.004.

CHEN, Q.; THOUAS, G. A. Metallic implant biomaterials. **Materials Science and Engineering: R: Reports**, v. 87, p. 1-57, 2015.

CHEN, S.; ARROUAYS, D.; MULDER, V. L.; POGGIO, L.; MINASNY, B.; ROUDIER, P.; LIBOHOVA, Z.; LAGACHERIE, P.; SHI, Z.; HANNAM, J.; MEERSMANS, J.; RICHER-DE-FORGES, A. C.; WALTER, C. Digital mapping of GlobalSoilMap soil properties at a broad scale: A review. **Geoderma**, v. 409, p. 115567, 2022.

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

CONRAD, O.; BECHTEL, B.; BOCK, M.; DIETRICH, H.; FISCHER, E.; GERLITZ, L.; WEHBERG, J.; WICHMANN, V.; BÖHNER, J. System for automated geoscientific analyses (SAGA) v. 2.1. 4. **Geoscientific model development**, v. 8, n. 7, p. 1991-2007, 2015.

COOPER, M.; MENDES, L. M. S.; SILVA, W. L. C.; SPAROVEK, G. A national soil profile database for Brazil available to international scientists. **Soil Science Society of America Journal**, v. 69, n. 3, p. 649-652, 2005.

COSTA, E. M.; SAMUEL-ROSA, A.; ANJOS, L. H. C. dos. Digital elevation model quality on digital soil mapping prediction accuracy. **Ciência e Agrotecnologia**, v. 42, n. 6, p. 608-622, 2018.

COSTA, I. S. L.; TAVARES, F. M.; DE OLIVEIRA, J. K. M. Predictive lithological mapping through machine learning methods: a case study in the Cinzento Lineament, Carajás Province, Brazil. **Journal of the Geological Survey of Brazil**, v. 2, n. 1, p. 26-36, 2019. Disponível em: <https://rigeo.sgb.gov.br/handle/doc/25024>. Acesso em: 10 dez. 2024.

COSTA, M. L. Aspectos geológicos dos lateritos da Amazônia. **Brazilian Journal of Geology**, v. 21, p. 146–160, 1991.

CPRM. **Geologia e Recursos Minerais do Estado do Amazonas**. Manaus: CPRM – Serviço Geológico do Brasil, 2006. 125 p.

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.

D' HOORE, J. **Soil map of Africa at a scale of 1/5,000,000**. Soil map of Africa at a scale of 1/5,000,000 (revised definitions). 1964.

DE FIGUEIREDO FILHO, W. F. A.; DE OLIVEIRA, F. S.; BARROS, L. F. de P.; MACHADO, H. A.; MAGALHÃES JUNIOR, A. P. O papel da dinâmica do relevo na evolução de couraças no Vale do Rio Conceição, Quadrilátero Ferrífero-MG. **Revista Brasileira de Geomorfologia**, v. 20, n. 1. 2019.

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.

DIAS, R. L.S.; SILVA, D. D. da; FERNANDES-FILHO, E. I.; AMARAL, C.H. DO; SANTOS, E. P. dos; MARQUES, J. F.; VELOSO, G.V. Machine learning models applied to TSS estimation in a reservoir using multispectral sensor onboard to RPA. **Ecological Informatics**, v. 65, p. 101414, 2021.

DIJKERMAN, J. C. Pedology as a science: The role of data, models and theories in the study of natural soil systems. **Geoderma**, v. 11, n. 2, p. 73-93, 1974.

DOS ANJOS, L. H. C.; FRANZMEIER, D. P.; SCHULZE, D. G. Formation of soils with plinthite on a toposequence in Maranhão State, Brazil. **Geoderma**, v. 64, n. 3-4, p. 257-279, 1995.

DUAN, M.; SONG, X.; LI, Z.; ZHANG, X.; DING, X.; CUI, D. Identifying soil groups and selecting a high-accuracy classification method based on multi-textural features with optimal window sizes using remote sensing images. **Ecological Informatics**, v. 81, p. 102563, 2024.

EGGLETON, R. A.; TAYLOR, G. Selected thoughts on laterite. In: TAYLOR, G.; PAIN, C. **New Approaches to an Old Continent**: Proceedings of the 3rd Australian Regolith Conference. Dordrecht: Springer, 1998. p. 209-226.

EL-DESOKY, H. M.; SHEBL, A.; ABDEL-RAHMAN, A. M.; FAHMY, W.; EL-AWNY, H.; EL-SHERIF, A. M.; EL-RAHMANY, M. M.; CSÁMER, Á. 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.

EL-DESOKY, M. M.; MORAD, I.; WASFY, M. H.; MANSOUR, A. F. Synthesis, structural and electrical properties of PVA/TiO₂ nanocomposite films with different TiO₂ phases prepared by sol-gel technique. **Journal of Materials Science: Materials in Electronics**, v. 31, n. 20, p. 17574-17584, 2020.

EL-LEIL, I. A.; EL-DESOKY, H. M.; SHAFEA, E. A. Geochemistry, diagenesis and mineral associations of laterite deposits along the contact zone of Precambrian basement rocks and Nubia sandstone between Wadi Dungash and Wadi Shait, South Eastern Desert. **Egyptian Journal of American Science**, v. 13, n. 8, 2017.

ESPINDOLA, C. R.; DANIEL, L. A. Laterita e solos lateríticos no Brasil. **Boletim Técnico da FATECSP**, v. 24, p. 21, 2008.

FAO. **Revised World Soil Charter**. Roma, Italy, p. 4-5, june 2015. Disponível em: <https://www.fao.org/documents/card/ar/c/e60df30b-0269-4247-a15f-db564161fee0/>. Acesso em: 20 ago. 2023.

- FARHADI, H.; NAJAFZADEH, M. Flood risk mapping by remote sensing data and random forest technique. **Water**, v. 13, n. 21, p. 3115, 2021.
- 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**. Cham: Springer International Publishing. p. 423-433, 2023.
- FERREIRA, A. C. S.; PINHEIRO, É. F. M.; COSTA, E. M.; CEDDIA, M. B. Predicting soil carbon stock in remote areas of the Central Amazon region using machine learning techniques. **Geoderma Regional**, v. 32, p. e00614, 2023.
- FRIEDMAN, J. H. Multivariate adaptive regression splines. **The annals of statistics**, v. 19, n. 1, p. 1-67, 1991.
- FRIEDMAN, J.; HASTIE, T.; TIBSHIRANI, R. Regularization paths for generalized linear models via coordinate descent. **Journal of statistical software**, v. 33, p. 1-22, 2010.
- GELSLEICHTER, Y. A.; COSTA, E. M.; DOS ANJOS, L. H. C.; MARCONDES, R. A. T. Enhancing Soil Mapping with Hyperspectral Subsurface Images generated from soil lab Vis-SWIR spectra tested in southern Brazil. **Geoderma Regional**, v. 33, p. e00641, 2023.
- GIOVANNINI, A. L. **Contribuição à geologia e geoquímica do carbonatito e da jazida (Nb, ETR) de Seis Lagos** (Amazonas). 2013. 112 f. Dissertação (Mestrado) — Instituto de Geociências, Universidade Federal do Rio Grande do Sul, Porto Alegre. 2013.
- GIOVANNINI, A. L.; NETO, A. C. B.; PORTO, C. G.; PEREIRA, V. P.; TAKEHARA, L.; BARBANSO, L.; BASTOS, P. H. S. Mineralogy and Geochemistry of Laterites from the Morro Dos Seis Lagos Nb (Ti, REE) Deposit (Amazonas, Brazil). **Ore Geology Reviews**, v. 88, p. 461-480, 2017.
- GIOVANNINI, A. L.; NETO, A. C. B.; PORTO, C. G.; TAKEHARA, L.; PEREIRA, V. P.; BIDONE, M. H. REE mineralization (primary, supergene and sedimentary) associated to the Morro dos Seis Lagos Nb (REE, Ti) deposit (Amazonas, Brazil). **Ore Geology Reviews**, v. 137, p. 104308, 2021.
- GIOVANNINI, A.L.; MITCHELL, R.H.; NETO, A.C.B.; MOURA, C.A.V.; PEREIRA, V.P.; PORTO, C.G. Mineralogy and geochemistry of the Morro dos Seis Lagos siderite carbonatite, Amazonas, Brazil. **Lithos**, v. 360, p. 105433, 2020.
- 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.
- GOYDARAGH, M. G.; TAGHIZADEH-MEHRJARDI, R.; GOLCHIN, A.; JAFARZADEH, A. A.; LADO, M. Predicting weathering indices in soils using FTIR spectra and random forest models. **Catena**, v. 204, p. 105437, 2021.
- GRIMM, R.; BEHRENS, T.; MÄRKER, M.; ELSENBEER, H. Soil organic carbon concentrations and stocks on Barro Colorado Island—Digital soil mapping using Random Forests analysis. **Geoderma**, v. 146, n. 1-2, p. 102-113, 2008.
- HARNOIS, L. The CIW index: a new chemical index of weathering. **Sedimentary geology**, v. 55, n. 3, p. 319-322, 1988.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The elements of statistical learning: Data Mining, Inference, and Prediction**. 2nd edn New York: Springer Series in Statistics, 2009.

- HECHENBICHLER, K.; SCHLIEP, K. Weighted K-Nearest-Neighbor Techniques and Ordinal Classification. **Collaborative Research Center 386**, Discussion Paper 399. 2004.
- HENGL, T. **Pedometric Mapping**: bridging the gaps between conventional and pedometric approaches. 2003. 24f. Thesis (Ph.D.) – Wageningen University and ITC with summary in Dutch and Croatian. 2003. Disponível em: <https://www.proquest.com/openview/fe221a82dd13570a16b0acb55c605fe3/1?pq-origsite=gscholar&cbl=2026366&diss=y>. Acesso em: 25 nov. 2023.
- HENGL, T.; HEUVELINK, G. B. M.; STEIN, A. A generic framework for spatial prediction of soil variables based on regression-kriging. **Geoderma**, v. 120, n. 1-2, p. 75-93, 2004.
- HEUNG, B.; BULMER, C. E.; SCHMIDT, M. G. Predictive soil parent material mapping at a regional-scale: A Random Forest approach. **Geoderma**, v. 214, p. 141-154, 2014.
- HOSSEINZADEH, A.; ZHOU, J. L.; ALTAEE, A.; BAZIAR, M.; LI, D. Effective modelling of hydrogen and energy recovery in microbial electrolysis cell by artificial neural network and adaptive network-based fuzzy inference system. **Bioresource Technology**, v. 316, p. 123967, 2020.
- HORBE, A. M. C.; ALBUQUERQUE, M. F. D. S.; DANTAS, E. L. Nd and Sr isotopes and REE investigation in tropical weathering profiles of Amazon region. **Frontiers in Earth Science**, 10, 845224, 2022.
- 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.
- HUANG, M.; LIU, Y.; CHEN, F.; TAO, Y.; SUN, Y. Correlations of chemical weathering indicators with major chemical constituents in sediments to obtain palaeoclimate information from Chaohu Lake, China. **Scientific Reports**, v. 14, n. 1, p. 17258, 2024.
- HUETE, Alfredo R. A soil-adjusted vegetation index (SAVI). **Remote sensing of environment**, v. 25, n. 3, p. 295-309, 1988.
- IBGE. **Banco de Dados de Informações Ambientais**. Relatório pedológico do município de São Gabriel da Cachoeira – AM. 2025. Disponível em: <https://bdia.ibge.gov.br>. Acesso em: 12 maio 2025.
- IBGE. **Manual Técnico de Pedologia**. 2 ed. Rio de Janeiro: IBGE - Coordenação de Recursos Naturais e Estudos Ambientais. 323p. 2007. Disponível em: <https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=295017>. Acesso em: 10 abr. 2025.
- ITO, M.; OTAKE, Y.; OHBA, T.; SATO, H.; YONEDA, S.; MATSUMOTO, R. Geochemical constraints on the mobilization of Ni and critical metals in laterite deposits, Sulawesi, Indonesia: A mass-balance approach. **Resource Geology**, v. 71, n. 4, p. 437-456, 2021. DOI: 10.1111/rge.12266.
- IZA, E. R. H. de F.; HORBE, A. M. C.; CASTRO, C. C.; HERRERA, I. L. I. E. Integration of geochemical and geophysical data to characterize and map lateritic regolith: an example in the Brazilian Amazon. **Geochemistry, Geophysics, Geosystems**, v. 19, n. 9, p. 3254–3271, 2018.
- JENNY, H. **Factors of soil formation**: a system of quantitative pedology. Reimpressão. New York: Courier Corporation. 281p. 1994.
- 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.

- JUSTO, L. J. E. C. **Projeto Uaupés**. v. 1. Manaus: Companhia de Pesquisa de Recursos Minerais – CPRM [Relatório Final de Pesquisa], 1983. 214 p.
- KALINOWSKI, A.; OLIVER, S. ASTER mineral index processing manual. **Remote Sensing Applications, Geoscience Australia**, v. 37, p. 36, 2004.
- KÄMPF, N.; CURI, N. Formação e evolução do solo (Pedogênese). In: KER, J. C. Curi, N. Schaefer, C. E. G. R., Vidal-Torrado, P (Eds.). **Pedologia Fundamentos**. 1. ed. Viçosa, MG: SBCS, p. 207-302, 2012.
- KARATZOGLOU, A.; SMOLA, A.; HORNIK, K. Kernlab: Kernel-based machine learning lab. **R package version 0.9**, v. 25, 2016.
- KASAY, G. M. **Geochemistry and Economic Potential of the Bingo Carbonatite and Its Associated Laterites in Beni Territory, North Kivu, Democratic Republic of Congo (DRC)**. 2018. 96f. Dissertation (Master of Science in Geology), University of Nairobi, Kenya. 2018.
- Disponível em: https://www.researchgate.net/publication/337223787_Geology_Geochemistry_and_Economic_Potential_of_the_Bingo_Carbonatite_and_its_associated_laterites_in_Beni_North_Kivu_DRC. Acesso em: 20 jun. 2024.
- KER, J. C. Latossolos do Brasil: uma revisão. **Pesquisas Agropecuárias Brasileiras**, Brasília, v. 33, n. 5, p. 511-523, maio 1998.
- KER, J. C.; CURI, N.; SCHAEFER, C. E. G. R.; VIDAL-TORRADO, P. (Eds.). **Pedologia: fundamentos**. Viçosa, MG: Sociedade Brasileira de Ciência do Solo, 2012. 343 p.
- KING, P. M. **The secondary minerals of the tertiary lavas of northern and central Skye-zeolite zonation patterns their origin and formation**. University of Aberdeen (United Kingdom), 1976.
- KIRKWOOD, C.; CAVE, M.; BEAMISH, D.; GREBBY, S.; FERREIRA, A. A machine learning approach to geochemical mapping. **Journal of Geochemical Exploration**, v. 167, p. 49-61, 2016.
- KÖPPEN, W. **Climatologia: con un studio de los climas de la Tierra**. México: Fondo de Cultura Económica, 1998.
- 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.
- KUHN, M.; JOHNSON, K. **Applied predictive modeling**. New York: Springer, v. 16, p. 13. 2013.
- LAGACHERIE, P. Digital soil mapping: a state of the art. **Digital soil mapping with limited data**, p. 3-14, 2008.
- LE MAITRE, R. W.; STRECKEISEN, A.; ZANETTIN, B.; LE BAS, M. J.; BONIN, B.; BATEMAN, P. **Igneous rocks: a classification and glossary of terms: recommendations of the International Union of Geological Sciences Subcommission on the Systematics of Igneous Rocks**. Cambridge University Press, 2005.
- LIAW, A.; WIENER, M. Classification and regression by randomForest. **R news**, v. 2, n. 3, p. 18-22, 2002.
- LIMA, L. A. de S.; NEUMANN, M. R. B.; REATTO, A.; ROIG, H. L. **Mapeamento de Solos: do tradicional ao digital**. Documentos / Embrapa Cerrados. (INFOTECA) Planaltina: 2013. Disponível em: <https://www.infoteca.cnptia.embrapa.br/bitstream/doc/1005933/1/doc316.pdf>. Acesso em: 05 ago. 2023.

- MALONE, B. P.; MINASNY, B.; ODGERS, N. P.; MCBRATNEY, A. B. Using model averaging to combine soil property rasters from legacy soil maps and from point data. **Geoderma**, v. 232, p. 34-44, 2014.
- MARTINS, R. M.; VARAJÃO, A. F. D. C.; DE OLIVEIRA, F. S.; ALLARD, T.; VARAJÃO, C. A. C.; SOUZA, C. D. Degradation systems of ferruginous duricrust in the Barreiras Formation, Southeastern Brazil. **Catena**, v. 246, p. 108463, 2024.
- MARTINS, T. F.; SEOANE, J. C. S.; TAVARES, F. M. Cu–Au exploration target generation in the eastern Carajás Mineral Province using random forest and multi-class index overlay mapping. **Journal of South American Earth Sciences**, v. 116, p. 103790, 2022.
- MAXWELL, A. E.; STRAGER, M. P.; WARNER, T. A.; RAMEZAN, C. A.; MORGAN, A. N.; PAULEY, C. E. Large-area, high spatial resolution land cover mapping using random forests, GEOBIA, and NAIP orthophotography: Findings and recommendations. **Remote Sensing**, v. 11, n. 12, p. 1409, 2019.
- MCBRATNEY, A. B.; ODEH, I. O.; BISHOP, T. F.; DUNBAR, M. S.; SHATAR, T. M. An overview of pedometric techniques for use in soil survey. **Geoderma**, v. 97, n. 3-4, p. 293-327, 2000.
- MCBRATNEY, A. B.; SANTOS, M. L. M.; MINASNY, B. On digital soil mapping. **Geoderma**, v. 117, n. 1-2, p. 3-52, 2003.
- MCKENZIE, N. J.; RYAN, P. J. Spatial prediction of soil properties using environmental correlation. **Geoderma**, v. 89, n. 1-2, p. 67-94, 1999.
- MCROBERTS, R. E.; TOMPPA, E. O.; FINLEY, A. O.; HEIKKINEN, J. Estimating areal means and variances of forest attributes using the k-Nearest Neighbors technique and satellite imagery. **Remote sensing of environment**, v. 111, n. 4, p. 466-480, 2007.
- MEDEIROS, B. M.; SEQUINATTO-ROSSI, L.; TEN CATEN, A.; PEREIRA, G. E.; SILVA, E. B. D.; DABOIT, K. T. U. Soil legacy data: An opportunity for digital soil mapping. **Revista Brasileira de Ciência do Solo**, v. 48, p. e0230130, 2024.
- MEI, H.; JIAN, X.; ZHANG, W.; FU, H.; ZHANG, S. Behavioral differences between weathering and pedogenesis in a subtropical humid granitic terrain: Implications for chemical weathering intensity evaluation. **Catena**, v. 203, p. 105368, 2021.
- MELLO, D. C. de; FERREIRA, T. O.; VELOSO, G. V.; DE LANA, M. G.; MELLO, F. A. de O.; DI RAIMO, L. A. de L.; DEMATTÊ, J. A. M. Digital mapping of soil weathering using field geophysical sensor data coupled with covariates and machine learning. **Journal of South American Earth Sciences**, v. 128, p. 104449, 2023.
- MENDONÇA-SANTOS, M. L.; SANTOS, H. The state of the art of Brazilian soil mapping and prospects for digital soil mapping. In: LAGACHERIE, P.; MCBRATNEY, A.B.; VOLTZ, M. [Eds.]. **Developments in Soil Science**. New York: Elsevier, v.31, p.39-54, 2007.
- MENESES, P. R.; DE ALMEIDA, T.; BAPTISTA, G. M. de M. **Refletância dos materiais terrestres**. São Paulo: Oficina de textos, 2019.
- MILBORROW, S.; HASTIE, T.; TIBSHIRANI, R.; MILLER, A.; LUMLEY, T. earth: Multivariate adaptive regression splines. **R package version**, v. 5, n. 2, 2017.
- MIGUEL, P.; DALMOLIN, R. S. D.; PEDRON, F. de A.; FINK, J. R.; MOURA-BUENO, J. M. Caracterização de plintitas e petroplintitas em solos da Depressão Central do Rio Grande do Sul. **Ciência Rural**, Santa Maria, v. 43, n. 6, p. 999-1005, jun. 2013.
- MINASNY, B.; MCBRATNEY, A. B. A conditioned Latin hypercube method for sampling in the presence of ancillary information. **Computers & geosciences**, v. 32, n. 9, p. 1378-1388,

2006.

MINASNY, B.; MCBRATNEY, A. B. Digital soil mapping: A brief history and some lessons. **Geoderma**, v. 264, p. 301–311, 2016.

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

MOHANTY, B.; GUPTA, A.; DAS, B. S. Estimation of weathering indices using spectral reflectance over visible to mid-infrared region. **Geoderma**, v. 266, p. 111–119, 2016.

MOQUEDACE, C. M.; BALDI, C. G. O.; SIQUEIRA, R. G.; CARDOSO, I. M.; SOUZA, E. F. M. de; FONTES, R. L. F.; FRANCELINO, M. R.; GOMES, L. C.; FERNANDES-FILHO, E. I. High-resolution mapping of soil carbon stocks in the western Amazon. **Geoderma Regional**, v. 36, p. e00773, 2024.

NAHON, D.; TARDY, Y. The ferruginous laterite. In: GOVETT, G. J. S. (Ed.). **Handbook of exploration geochemistry**. London: Elsevier, v. 4, p. 41–78, 1992.

NESBITT, H. W.; YOUNG, G. M. Early Proterozoic climates and plate motions inferred from major element chemistry of lutites. **Nature**, v. 299, n. 5885, p. 715–717, 1982.

NGUYEN, H.; BUI, X.-N.; TRAN, Q.-H.; MAI, N.-L. A new soft computing model for estimating and controlling blast-produced ground vibration based on Hierarchical K-means clustering and Cubist algorithms. **Applied Soft Computing**, v. 77, p. 376–386, 2019.

NOVAIS, R. F.; SMYTH, T. J.; LEITE, N. A. Fósforo em solos tropicais altamente intemperizados: mecanismos de retenção e disponibilidade. In: *Fertilidade do solo*. Viçosa: **Sociedade Brasileira de Ciência do Solo**, p. 147–190, 2007.

NOVAIS, J. de J. **Modelagem de Dados de Sensoriamento Remoto para o Mapeamento Digital de Solos e Índice Relativo de Umidade dos Solos do Distrito Federal**. 2021. 251 f., il. Tese (Doutorado em Agronomia) — Universidade de Brasília, Brasília, 2021.

ODEH, I. O. A.; MCBRATNEY, A. B.; CHITTLEBOROUGH, D. J. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. **Geoderma**, v. 67, n. 3–4, p. 215–226, 1995.

OLIVEIRA, A. J. de; SCHULTZ, G. E.; CURI, N.; OLIVEIRA, I. F. de. Relações químicas e mineralogia de solos altamente intemperizados da região sudeste do Brasil. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 7, n. 2, p. 187–193, 1983.

OLIVEIRA, B. L.; MEDEIROS, L. R.; FARIAS, G. S. de. **Sociedade Brasileira de Ciência do Solo – Um olhar sobre sua história**. – 3. ed. rev. e amp. - Viçosa, MG: SBCS, p. 177. 2015. Disponível em: <https://www.sbcs.org.br/wp-content/uploads/2014/01/ebook.pdf>. Acesso em: 10 maio. 2023.

OLIVEIRA, F. S. de. **A bauxita de Barro Alto (GO): gênese e evolução mineralógica, micromorfológica e geoquímica**. 2011. 149 f. Tese (Doutorado) - Universidade Federal de Ouro Preto, Ouro Preto, 149p. 2011.

OLIVEIRA, J. R. S. D.; PRUSKI, F. F.; SILVA, J. M. A. D.; SILVA, D. P. D. Comparative analysis of the performance of mixed terraces and level and graded terraces. **Acta Scientiarum. Agronomy**, v. 34, p. 351–357, 2012.

OLLIER, C. D.; SHETH, H. C. The High Deccan duricrusts of India and their significance for the ‘laterite’ issue. **Journal of Earth System Science**, v. 117, p. 537–551, 2008.

ONU. **Relatório Anual das Nações Unidas no Brasil 2021**. 2021. Disponível em: <https://brasil.un.org/pt-br/177803-relat%C3%B3rio-anual-das-na%C3%A7%C3%B5es-unidas-no-brasil-2021>. Acesso em: 14 set. 2023.

OSINUGA, O. A.; OLUTIMI, M. S.; TOBORE, A. O.; SENJOBI, B. A.; OYEGOKE, C. O.; ADEOFUN, C. O. Elemental Oxides and Weathering Extent of Wetland Soils Under Different Geological Formations of Ogun State, Nigeria. **Communications in Soil Science and Plant Analysis**, v. 56, n. 4, p. 617-641, 2025.

PADARIAN, J.; MCBRATNEY, A. B.; MINASNY, B. Game theory interpretation of digital soil mapping convolutional neural networks. **Soil**, v. 6, n. 2, p. 389-397, 2020.

PADARIAN, J.; MINASNY, B.; MCBRATNEY, A. B. Machine learning and soil sciences: A review aided by machine learning tools. **Soil**, v. 6, n. 1, p. 35-52, 2020.

PAIN, C. F.; OILIER, C. D. Inversion of relief—a component of landscape evolution. **Geomorphology**, v. 12, n. 2, p. 151-165, 1995.

PALMIERI, M.; BROD, J. A.; CORDEIRO, P.; GASPAS, J. C.; BARBOSA, P. A. R.; ASSIS, L. C. de; JUNQUEIRA-BROD, T. C.; SILVA, S. E.; MILANEZI, B. P.; MACHADO, S. A.; JÁCOMO, M. H. The carbonatite-related Morro do Padre niobium deposit, Catalão II Complex, central Brazil. **Economic Geology**, v. 117, n. 7, p. 1497-1520, 2022.

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.

PEREIRA, L. F. **Caracterização e Reconstituição Paleoambiental da Formação Içá na Bacia do Rio Negro - AM**. 2016. 155f. Tese (Doutor em Geografia) - Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte, 155p. 2016.

PEREIRA, M. G.; DOS ANJOS, L. H. C.; PINHEIRO JUNIOR, C. R.; PINTO, L. A. da S. R.; SILVA NETO, E. C. da; FONTANA, A. Formação e caracterização de solos. In: TULLIO, L. (Org.). **Formação, classificação e cartografia dos solos**. p. 1-20. Ponta Grossa: Atena Editora, 2019.

PINHEIRO JUNIOR, C. R.; PEREIRA, M. G.; SILVA NETO, E. C.; ANJOS, L. H. C. Solos do Brasil: gênese, classificação e limitações ao uso. In: RIBEIRO, J. C. (org.). **Ciências exatas e da terra: conhecimentos estratégicos para o desenvolvimento do país**. Ponta Grossa: Atena, 2020. cap. 15, p. 183-199.

PINHEIRO, H. S. K. **Métodos de mapeamento digital aplicados na predição de classes e atributos dos solos da Bacia Hidrográfica do Rio Guapi Macacu, RJ**. 2015. 178 p. Tese (Doutorado) — Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2015.

PINHEIRO, H. S. K.; CHAGAS, C. D. S.; CARVALHO JÚNIOR, W. D.; ANJOS, L. H. C. D. Ferramentas de pedometria para caracterização da composição granulométrica de perfis de solos hidromórficos. **Pesquisa Agropecuária Brasileira**, v. 51, p. 1326-1338, 2016.

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**. Sociedade Brasileira de Geologia (SBGEO), 2021. p. 335 à 345. Disponível em: <https://hal-emse.ccsd.cnrs.fr/emse-03457498/>. Acesso em: 05 maio de 2022.

POLIDORO, J. C.; MENDONÇA-SANTOS, M. L.; LUMBRERAS, J. F.; COELHO, M. R.; CARVALHO FILHO, A.; MOTTA, P. E. F.; CARVALHO JUNIOR, W.; ARAÚJO FILHO, J. C.; CURCIO, G. R.; CORREIA, J. R.; MARTINS, É. S.; SPERA, S. T.; OLIVEIRA, S. R. M.; BOLFE, E. L.; MANZATTO, C. V.; TÔSTO, S. G.; VENTURIERI, A.; SÁ, I. B.;

- OLIVEIRA, V. Á.; SHINZATO, E.; ANJOS, L. H. C.; VALLADARES, G. S.; RIBEIRO, J. L.; MEDEIROS, P. S. C.; MOREIRA, F. M. S.; SILVA, L. S. L.; SEQUINATTO, L.; AGLIO, M. L. D.; DART, R.O. **Programa Nacional de Solos do Brasil (PronaSolos)**. Embrapa Solos, Rio de Janeiro, RJ, Brazil. Embrapa Solos-Documents (INFOTECa-E), 2016.
- POLIDORO, J.; COELHO, M.; CARVALHO FILHO, A. D.; LUMBRERAS, J.; de OLIVEIRA, A. P.; VASQUES, G. D. M.; MACARIO, C. G. do N; VICTORIA, D. de C; BHERING, S. B; FREITAS, P. L. de; QUARTAROLI, C. F; MENDONÇA-SANTOS, M. de L. **Programa Nacional de Levantamento e Interpretação de Solos do Brasil (PronaSolos): diretrizes para implementação**. Embrapa Solos-Documents (INFOTECa-E), 2021. Disponível em: <https://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1135056>. Acesso em: 05 set. 2023.
- PORTO, C. G.; BASTOS NETO, A. C.; BENTO, J. P. P. Regolith Hosted Niobium Deposits in Carbonatite Complexes in Brazil. *In: Geology, Geochemistry and Formation of Supergene Mineral Deposits in Deeply Weathered Terrain*. Springer, Cham, 2025. p. 207-225.
- PRADO, E. M. G.; DE 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.
- PORTO, C. G., BASTOS NETO, A. C., & PROENÇA BENTO, J. P. Regolith Hosted Niobium Deposits in Carbonatite Complexes in Brazil. *In: Geology, Geochemistry and Formation of Supergene Mineral Deposits in Deeply Weathered Terrain* (pp. 207-225), 2025. Springer, Cham. DOI: 10.1007/978-3-031-75733-4_8.
- PRENER, C.; GROSSENCACHER, T.; ZEHR, A.; STEVENS, J. biscale: Tools and Palettes for Bivariate Thematic Mapping, Versão 1.0.0. cran, 2022.
- QUINLAN, J. R. Learning with continuous classes. *In: Proceedings of the 5th Australian joint Conference on Artificial Intelligence*, Singapore: World Scientific:343 – 348, 1995.
- R CORE TEAM, R. R: **A language and environment for statistical computing**. R Foundation for Statistical Computing: Vienna, Austria, 2021.
- REDDY, N. N.; DAS BHABANI, S. Digital soil mapping of key secondary soil properties using pedotransfer functions and Indian legacy soil data. **Geoderma**, v. 429, p. 116265, 2023.
- 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.
- RETALLACK, G. J. **A colour guide to paleosols**. Chichester: Wiley, 1997. 175 p.
- RETALLACK, G. J. Lateritization and bauxitization events. **Economic Geology**, v. 105, n. 3, p. 655-667, 2010.
- RStudio Team. **RStudio: Integrated Development Environment for R**; RStudio, PBC: Boston, MA, USA, 2020.
- RIDLEY, J. Ore deposit geology [First published]. Colorado State University. 2013.
- RIPLEY, B.D.; VENABLES, W. nnet: Feed-Forward Neural Networks and Multinomial Log-Linear Models. Versão 7.3-20.2025. DOI:10.32614/CRAN. package.nnet
- RODRIGUES, H. M.; VASQUES, G. M.; OLIVEIRA, R. P.; TAVARES, S. R.; CEDDIA, M. B.; HERNANI, L. C. Finding suitable transect spacing and sampling designs for accurate soil ECa mapping from EM38-MK2. **Soil Systems**, v. 4, n. 3, p. 56, 2020.

- RODRIGUES, N. B.; BARBOSA, T. R.; PINHEIRO, H. S. K.; MANCINI, M.; READ, Q. D.; BLACKSTOCK, J.; LIBOHOVA, Z. Influences of Sampling Design and Model Selection on Predictions of Chemical Compounds in Petroferric Formations in the Brazilian Amazon. **Remote Sensing**, v. 17, n. 9, p. 1644, 2025.
- RODRIGUES, N. B.; SILVA, J. C. L. da; SILVA, R. P. M. da; PINHEIRO, H. S. K.; CARVALHO JÚNIOR, W. de. Mapeamento Digital dos Teores de Fe_2O_3 , MnO , Nb e TiO_2 em Morro dos Seis Lagos – AM, Brasil. **Espaço Aberto**, v. 14, n. 1, p. 157–174, 2024.
- RODRIGUES, N.; LOPES, J. C. da S.; SILVA, R. P. M. da, KOENOW PINHEIRO, H. R.; CARVALHO JUNIOR, W. Mapping of Fe_2O_3 , Nb and TiO_2 , AS a Support to Classify Outcropping Materials In" Morro DOS Seis Lagos" Carbonatite Complex, Brazilian Amazon. *In: EGU General Assembly Conference Abstracts*. p. EGU22-12326. 2022.
- ROSSITER, D. G. Past, present & future of information technology in pedometrics. **Geoderma**, v. 324, p. 131-137, 2018.
- 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.
- ROUDIER, P. A R Package for Conditioned Latin Hypercube Sampling: ‘clhs’R Pacote 'CLHS'. 2011. Disponível em: <https://github.com/pierrerroudier/clhs/CLH>. Acesso em: 29 out. 2023.
- ROUSE, J. W.; HAAS, R. H.; SCHELL, J. A.; DEERING, D. W. **Monitoring vegetation systems in the great plains with ERTS**. [s. l.], v. 1, p. 309–317, 1974. Disponível em: <https://repository.exst.jaxa.jp/dspace/handle/a-is/570457>. Acesso em: 10 jan. 2024.
- 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.
- RUNGE, E. C. Soil development sequences and energy models. **Soil Science**, v. 115, n. 3, p. 183-193, 1973.
- SABINS, F. F. **Remote sensing: Principles and Interpretation**. 3ª ed. New York: W. H. Freeman and Company. p. 432, 1997.
- SAFAEE, S.; LIBOHOVA, Z.; KLADIVKO, E.J.; BROWN, A.; WINZELER, E.; READ, Q.; RAHMANI, S.; ADHIKARI, K. Influence of sample size, model selection, and land use on prediction accuracy of soil properties. **Geoderma Regional**, v. 36, p. e00766, 2024.
- SAHOO, P. K.; GUIMARÃES, J. T. F.; SOUZA-FILHO, P. W. M.; SILVA, M. S. da; NUNES JÚNIOR, W.; POWELL, M. A.; COSTA, V. E. Geochemical characterization of the largest upland lake of the Brazilian Amazonia: Impact of provenance and processes. **Journal of South American Earth Sciences**, v. 80, p. 541–558, 2017.
- SAMUEL-ROSA, A. **Pedometria Feita Simples: Parte I - Fundamentação histórica e conceitual**. (ed.). Rio Grande do Sul: Universidade Federal de Santa Maria. 2017. Disponível em: <https://samuel-rosa.github.io/pedometria-feita-simples/index.html>. Acesso em: 10 jun. 2023.
- SAMUEL-ROSA, A., DALMOLIN, R. S. D.; MOURA-BUENO, J. M.; TEIXEIRA, W. G.; ALBA, J. M. F. Open legacy soil survey data in Brazil: geospatial data quality and how to improve it. **Scientia Agricola**, v. 77, n. 1, p. e20170430, 2019.

- SAMUEL-ROSA, A.; DALMOLIN, R. S. D.; MOURA-BUENO, J. M.; TEIXEIRA, W. G.; ALBA, J. M. F. Open legacy soil survey data in Brazil: geospatial data quality and how to improve it. **Scientia Agricola**, v. 77, n. 1, p. e20170430, 2020.
- SAMUEL-ROSA, A.; DO AMARAL, L. R.; HORST, T. Z.; PUSCH, M. Novidades e desafios do mapeamento digital do solo de alta resolução no Brasil. In: **Open Science Research I**. Editora Científica Digital, 2022. p. 315-330.
- SANTORO, L.; PUTZOLU, F.; MONDILLO, N.; BONI, M.; HERRINGTON, R. Trace element geochemistry of iron-(oxy)-hydroxides in Ni(Co)-laterites: review, new data and implications for ore forming processes. **Ore Geology Reviews**, v. 140, p. 104501, 2022.
- SANTOS, M. L.; SANTOS, R. S.; ROCHA, V.; SOUZA, L. F. P.; RIBEIRO, R. C.; SILVA, M. A.; COSTA, M. L.; VARAJÃO, A. F. D. C. Caparaó Mountains, southeastern Brazil: Unraveling geochemical weathering signatures and provenance insights from a tropical regolith profile. **Catena**, v. 238, 107849, maio 2024.
- SANTOS, H. D. dos; JACOMINE, P. T.; ANJOS, L. H. C. dos; OLIVEIRA, V. Á. de; LUMBRERAS, J. F.; COELHO, M. R.; OLIVEIRA, J. B. de. Sistema Brasileiro de Classificação de Solos. 5. Ed [ver. e ampl.]. Brasília: Embrapa, 356 p. 2018.
- SANTOS, P. H. C.; VASCONCELOS, P. M.; ANDRADE, F. R. D.; MENDONÇA, I. M.; FARIA, K. C.; SOUZA, A. C. The Piriá aluminous lateritic profile: mineralogy, geochemistry and parent rock. **Brazilian Journal of Geology**, v. 46, n. 4, p. 503-516, 2016. DOI: 10.1590/2317-4889201620160039.
- SILVEIRA, C. S.; VIANA, J. H. M.; VIEIRA, L. S.; RIBEIRO, F. G. R.; FERREIRA, R. S.; SOUZA, A. P.; BASTOS, R. M. Geochemistry of a spheroidal weathering profile in a tropical mountainous landscape, Rio de Janeiro, Brazil. **Anais da Academia Brasileira de Ciências**, v. 94, supl. 4, e20211091, 2022. DOI: 10.1590/0001-3765202220211091.
- SCHAEFER, C. E.; CORRÊA, G. R.; DE OLIVEIRA, F. S.; FILHO, E. I. F.; FRANCELINO, M. R.; GOMES, L. C. The Making of Brazilian Soilscapes: A Geosystemic Vista on Neotropical Pedology. In: SCHAEFER, C. E. **The Soils of Brazil**. Cham: Springer International Publishing, 2023. p. 25-70.
- SERVIÇO GEOLÓGICO DO BRASIL. Serviço Geológico do Brasil esclarece dúvidas sobre potencial do país para terras raras e minerais estratégicos. Brasília, DF, 5 ago. 2025. Disponível em: <https://www.sgb.gov.br/w/servico-geologico-do-brasil-esclarece-duvidas-sobre-potencial-do-pais-para-terras-raras-e-minerais-estrategicos>. Acesso em: 23 Dezembro. 2025.
- SENTINEL, E. S. A. **SAR User Guide** [em linha]. 1. ESA: Paris, France, 2021. Disponível em: <https://sentinel.esa.int>. Acesso em: 10 jan. 2024.
- SHIRMARD, H.; FARAHBAKHSH, E.; MÜLLER, R. D.; CHANDRA, R. A review of machine learning in processing remote sensing data for mineral exploration. **Remote Sensing of Environment**, v. 268, p. 112750, 2022.
- SILVEIRA, C. T.; OKA-FIORI, C.; SANTOS, L. J. C.; SIRTOLI, A. E.; SILVA, C. R.; BOTELHO, M. F. Soil prediction using artificial neural networks and topographic attributes. **Geoderma**, v. 195, p. 165-172, 2013.
- SIMONSON, R. W. Outline of a generalized theory of soil genesis. **Soil Science Society of America Journal**, v. 23, n. 2, p. 152-156, 1959.
- 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. **Educamazônia-Educação, Sociedade e Meio Ambiente**, v. 21, n. 2, p. 251-263, 2018.

- SOUZA, C. M. P. D.; THOMAZINI, A.; SCHAEFER, C. E. G. R.; VELOSO, G. V.; MOREIRA, G. M.; FERNANDES FILHO, E. I. Multivariate analysis and machine learning in properties of Ultisols (Argissolos) of Brazilian Amazon. **Revista Brasileira de Ciência do Solo**, v. 42, p. e0170419, 2018.
- STANYER, C.; SECO-RIZO, I.; ATZBERGER, C.; MARTI-CARDONA, B. Soil Texture, Soil Moisture, and Sentinel-1 Backscattering: Towards the Retrieval of Field-Scale Soil Hydrological Properties. **Remote Sensing**, v. 17, n. 3, p. 542, 2025.
- STRAHLER, A. N.; STRAHLER, A. H. Environmental geoscience: interaction between natural systems and man. (**No Title**), 1973.
- STUMPF, F.; SCHMIDT, K.; BEHRENS, T.; SCHÖNBRODT-STITT, S.; BUZZO, G.; DUMPERT, C.; WADOUX, A.; XIANG, W.; SCHOLTEN, T. Incorporating limited field operability and legacy soil samples in a hypercube sampling design for digital soil mapping. **Journal of Plant Nutrition and Soil Science**, v. 179, n. 4, p. 499-509, 2016.
- SVETNIK, V.; LIAW, A.; TONG, C.; WANG, T. Application of Breiman's random forest to modeling structure-activity relationships of pharmaceutical molecules. In: ROLI, F.; KITTLER, J.; WINDEATT, T. (eds) **International workshop on multiple Classifier systems**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004. p. 334-343.
- TAKEHARA, L. Projeto Avaliação do Potencial de Terras Raras no Brasil: área Morro dos Seis Lagos, Noroeste do Amazonas. [Informe de Recursos Minerais, Série Minerais Estratégicos, n. 4]. Brasília: CPRM, 2019. 112p. Disponível em: <https://geoportal.cprm.gov.br>. Acesso em: 14 jun. 2025.
- TARDY, Y. Diversity and terminology of lateritic profiles. In: MARTINI, J. P.; CHESWORTH, W. **Developments in earth surface processes**. Elsevier, 1992. p. 379-405.
- TARDY, Y. **Le cycle de l'eau: climats, paléoclimats et géochimie globale**. Paris: Masson, 1986.
- TARDY, Y.; MORTATTI, J.; RIBEIRO, A.; VICTORIA, R.; PROBST, J. L. Fluctuations de la pluviosité, de l'écoulement et de la température sur le bassin de l'Amazonie et oscillations du climat global au cours du siècle écoulé. **Comptes Rendus de L'Académie des Sciences. Série 2 Sciences de la Terre et des Plantes**, v. 318, p. 955-60, 1994.
- TARDY, Y.; ROQUIN, C. **Dérive des continents paléoclimats et altérations tropicales**. éd. BRGM, 1998.
- TAYLOR, S. R.; MCLENNAN, S. M. **The continental crust: its composition and evolution**. Oxford: Blackwell Scientific Publications, 1986.
- ULRICH, M.; CATHELINEAU, 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.
- VASCONCELOS, P. M.; CARMO, I. de O. Calibrating denudation chronology through ⁴⁰Ar/³⁹Ar weathering geochronology. **Earth-Science Reviews**, v. 179, p. 411-435, 2018.
- VENABLES, W. N.; RIPLEY, B. D. **Modern applied statistics with S**. New York: Springer Science & Business Media, 2002.
- VIEGAS FILHO, J. de R.; BONOW, C. W. de. Projeto Seis Lagos: relatório final. 2v. Manaus:

CPRM, 1976. Disponível em: <https://rigeo.sgb.gov.br/handle/doc/7601?mode=full>. Acesso em: 10 nov. 2023.

WADOUX, A. M. J. C.; ODEH, I. O. A.; MCBRATNEY, A. B. **Overview of Pedometrics**. Reference Module in Earth Systems and Environmental Sciences [em linha]. 2021.

WADOUX, A. MJ-C.; BRUS, D. J.; HEUVELINK, G. B. M. Sampling design optimization for soil mapping with random forest. **Geoderma**, v. 355, p. 113913, 2019.

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.

WANG, W.; XUE, C.; ZHAO, J.; YUAN, C.; TANG, J. Machine learning-based field geological mapping: A new exploration of geological survey data acquisition strategy. **Ore Geology Reviews**, v. 166, p. 105959, 2024.

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.

WRB. **A framework for international classification, correlation and communication**. Food and Agriculture Organization of the United Nations. Rome: IUSS/ISRIC/FAO (World Soil Resources Reports, 103), 2014. 191 p.

ZARDIACKAS, L. D.; KRAAY, M. J.; FREESE, H. L. **Titanium, Niobium, Zirconium, and Tantalum for Medical and Surgical Applications**. ASTM: International, n. 1471, 2006.

ZHANG, J.; SCHMIDT, M.G.; HEUNG, B.; BULMER, C.E.; KNUDBY, A. Using an ensemble learning approach in digital soil mapping of soil pH for the Thompson-Okanagan region of British Columbia. **Canadian Journal of Soil Science**, v. 102, n. 03, p. 579-596, 2022.

9. APÊNDICES

9.1 Apêndice A – Estatística descritiva dos elementos Al_2O_3 , Fe_2O_3 , MnO , Nb_2O_5 , TiO_2 , P_2O_5 e SiO_2 .

Elementos (%)	Média	Mediana	Mínimo	Máximo	Desvio Padrão	Coefficiente de variação	Assimetria	Curtose	Amplitude
Al_2O_3	3.28	1.67	0.19	58.00	6.43	195.97	5.34	32.06	57.81
Fe_2O_3	69.00	77.40	0.12	95.10	22.59	32.74	-1.67	1.89	94.98
MnO	5.09	0.20	0.00	64.10	14.01	275.09	3.04	8.07	64.10
Nb_2O_5	0.74	0.39	0.00	4.02	0.88	118.69	1.44	1.50	4.02
TiO_2	2.94	1.42	0.00	29.90	4.09	139.42	2.79	9.86	29.90
P_2O_5	732.74	0.84	0.03	10673.00	1189.51	162.34	2.70	14.38	10672.98
SiO_2	2.75	0.56	0.00	97.80	11.61	422.66	7.07	51.77	97.80

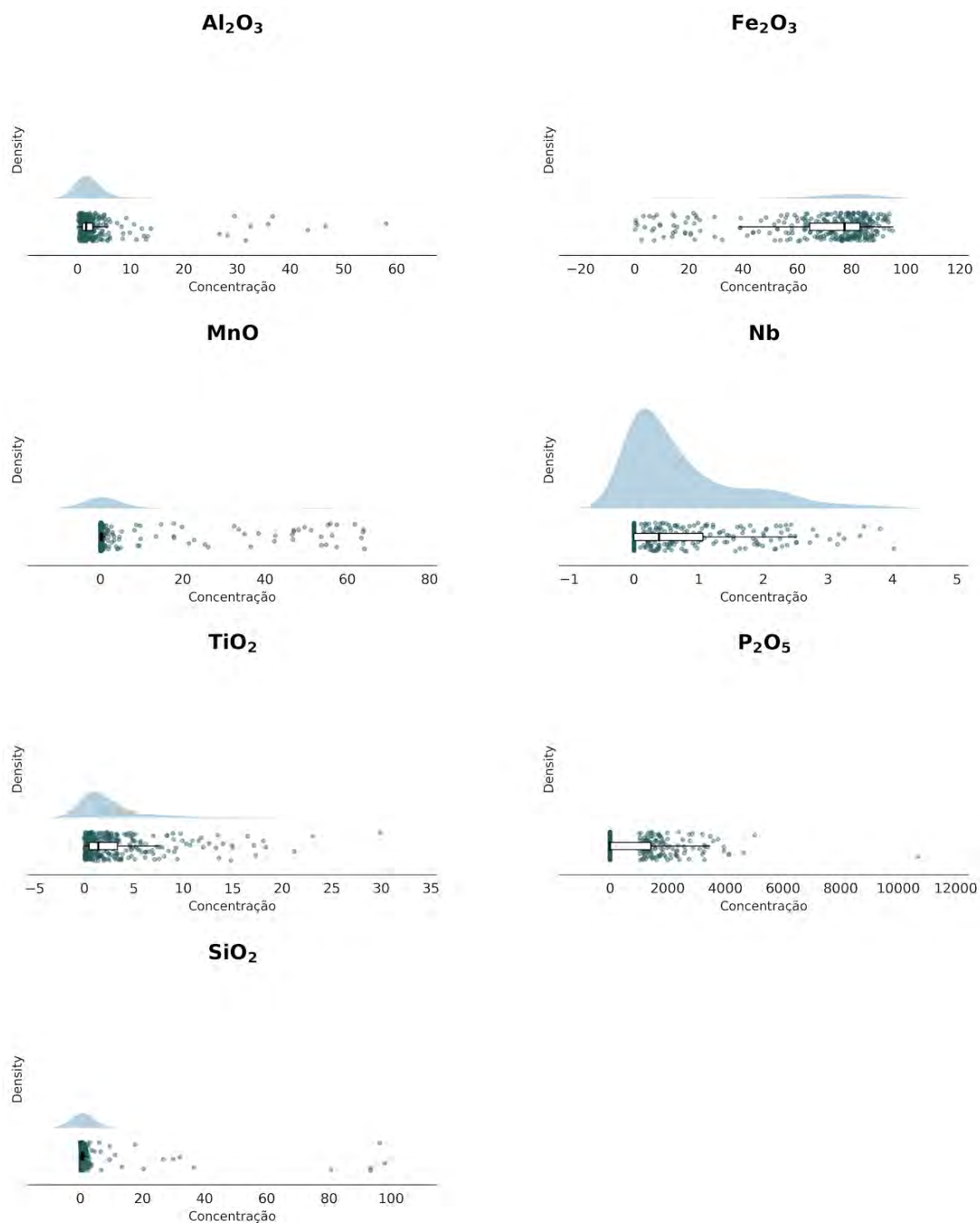


Figura 35 - Distribuição estatística, densidade e intervalos de confiança para os teores de Al_2O_3 , Fe_2O_3 , MnO , Nb_2O_5 , TiO_2 , P_2O_5 e SiO_2

9.2 Apêndice B – Tabela de Covariáveis

Category	Variable/Index	Equation (if applicable)	Reference
Geomorphological Indices & Factors	SAGA Wetness Index (WI)	-	Conrad (2001)
	Slope Height	-	Thompson <i>et al.</i> (2001); Wilson & Gallant (2000)
	Aspect	-	Conrad (2013)
	Elevation	-	Hutchinson & Gallant (2000); Moore <i>et al.</i> (1991)
	Vector Ruggedness Measure (VRM)	-	Wilson <i>et al.</i> (2007)
Curvature and Morphometry	Real Surface Area	-	Olaya (2004)
	Convergence Index (CI)	-	Conrad (2001)
	Curvature Flow Line	-	Conrad (2001)
	General Curvature (GC)	-	Conrad (2001)
	Maximum Curvature	-	Conrad (2001)
	Minimum Curvature	-	Conrad (2001)
	Planar Curvature (PlanC)	-	Conrad (2001)
	Profile Curvature (ProfC)	-	Conrad (2001)
	Tangential Curvature (TanC)	-	Conrad (2001)
	Total Curvature (TotalC)	-	Conrad (2001)
	Mid-Slope Position	-	Conrad <i>et al.</i> (2008)
	Multi-Resolution Ridge Top Flatness (MRRTF)	-	Gallant & Dowling (2003)
	Multi-Resolution Valley Bottom Flatness (MRVBF)	-	Gallant & Dowling (2003)
	Slope index	-	Conrad (2010)
	Standardized height	-	Conrad <i>et al.</i> (2008)
	Terrain surface convexity	-	Conrad (2012)
	Terrain surface texture	-	Conrad (2012)
	Valley Depth	-	Conrad (2012)
	Closed depressions	-	Conrad (2012)
	Hill	-	Olaya (2004)
	Hill index	-	Olaya (2004)
	Valley	-	Conrad (2015)
	Valley index	-	Conrad (2015)
	Surface specific points	-	Conrad (2005)
Spectral Indices	NDVI (Normalized Difference Vegetation Index)	$(RNIR - Rred) / (RNIR + Rred)$	Rouse <i>et al.</i> (1973)
	SAVI (Soil-Adjusted Vegetation Index)	$(RNIR - Rred) / (RNIR + Rred + L)$	Huete (1988)
	Ferrous Silicates	Band 12/ band 11	Van der Meer <i>et al.</i> (2014)
	Clay Minerals	band 6/ band 7	Sabins (1997)
	Ferric iron, Fe3+	Band 2/ band 1	Kalinowski e Oliver (2004)
	Laterite	Band 11/ band 12	Van der Meer <i>et al.</i> (2014)
	Ferrous iron, Fe2+	Band 5/ band 3 + band 1/ band 2	Kalinowski e Oliver (2004)
	Ferroginous regolith	Band 8A/ band 3	Rowan <i>et al.</i> (2003)

Ferrous Iron	Band 5 + Band 1/ Band 3+ Band 2	Rowan <i>et al.</i> (2003)
Altered zones	Band 8/ band 11	Sora <i>et al.</i> (2018)
Detection of advanced clay alteration	Band 4/ band 6	Sora <i>et al.</i> (2018)
Hydrothermally altered rocks	Band 8-Band 1- Band 2- Band 3	Sora <i>et al.</i> (2018)
Gossan Index	Band 4/ band 2	Sora <i>et al.</i> (2018)
VV Polarization (Sentinel-1A SAR)		ESA (2021)
VH Polarization (Sentinel-1A SAR)		ESA (2021)
VH Polarization (Sentinel-1A SAR)		ESA (2021)

References

- Conrad, O. Topographic Position Index (TPI). SAGA GIS Documentation 2011.
- Conrad, O. (2001). *SAGA - System for Automated Geoscientific Analyses*. Geoscientific Model Development, 8, 1991–2007. <https://doi.org/10.5194/gmd-8-1991-2015>
- Conrad, O. (2008). *Mid-Slope Position Analysis in SAGA GIS*. Geoscientific Model Development.
- Conrad, O. (2010). *Slope Index Calculation in SAGA GIS*. SAGA GIS Documentation.
- Conrad, O. System for Automated Geoscientific Analyses (SAGA) – version 2.1.0. 2013.
- European Space Agency (ESA). Sentinel-1 SAR User Guide; ESA: Paris, France, 2021. Available online: <https://sentinel.esa.int> (accessed on 10 January 2024).
- Gallant, J.C.; Dowling, T.I. A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resources Research* 2003, 39(12).
- Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25(3), 295–309. [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X)
- Hutchinson, M.F.; Gallant, J.C. Digital elevation models and representation of terrain shape. *Mathematics and Computers in Simulation* 2000, 42(4), 135-150.
- Kalinowski, A.; Oliver, S. ASTER mineral index processing manual; Geoscience Australia Technical Report 2004.
- Moore, I.D.; Burch, G.J. Physical basis of the length-slope factor in the Universal Soil Loss Equation. *Soil Science Society of America Journal* 1986, 50(5), 1294-1298.
- Olaya, V. (2004). SAGA GIS - Morphometry Tools. Available at: https://saga-gis.sourceforge.io/saga_tool_doc/2.2.3/ta_morphometry_6.html
- Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring vegetation systems in the Great Plains NASA Special Publication 1973, 351, 309-317.
- Rowan, L.C.; Mars, J.C.; Simpson, C.J. Hyperspectral analysis of the ultramafic complex of the Gossan Lead, Virginia, USA. *Remote Sensing of Environment* 2003, 88(2), 123-139.
- Sabins, F.F. *Remote Sensing Principles and Interpretation*, 3rd ed.; W.H. Freeman and Company: New 1997.
- Sora, S.; Van der Meer, F.; Hecker, C. Spectral indices for detecting advanced argillic alteration in active hydrothermal systems. *International Journal of Applied Earth Observation and Geoinformation* 2018, 69, 134-147.
- Thompson, J.A.; Bell, J.C.; Butler, C.A. Digital elevation model resolution: Effects on terrain attribute calculation and quantitative soil-landscape modeling. *Geoderma* 2001, 100(1-2), 67-89.
- Van der Meer, F.D.; Van der Werff, H.M.A.; Van Ruitenbeek, F.J.A.; *et al.* Multi- and hyperspectral geologic remote sensing: A review. *International Journal of Applied Earth Observation and Geoinformation* 2014, 33, 161-175
- Wilson, J. P., Gallant, J. C., & Walsh, S. J. (2007). *Topographic Mapping and Analysis in GIS*. In R. H. Ramsay & R. E. McCleary (Eds.), *Geographic Information Systems and Science* (pp. 87-112).

9.3 Apêndice C – Tabela - Desempenho comparativo dos modelos preditivos por cenários (I - V) de amostragem e elementos (Al₂O₃, Fe₂O₃, TiO₂, Nb₂O₅, MnO, SiO₂): medianas (MAE, RMSE, R²) com agrupamento estatístico (Kruskal-Wallis + Dunn, CLD, $\alpha = 0,05$).

Element	Scenario	Metric	Model	Median	Group	Kruskal_pvalue
Al ₂ O ₃	I	MAE	GLMNET	3.58	ab	1.46e-08
Al ₂ O ₃	I	MAE	KNN	3.68	a	1.46e-08
Al ₂ O ₃	I	MAE	NNET	3.61	a	1.46e-08
Al ₂ O ₃	I	MAE	RF	3.485	bc	1.46e-08
Al ₂ O ₃	I	MAE	SVMRadial	3.36	c	1.46e-08
Al ₂ O ₃	II	MAE	GLMNET	3.16	ab	7.97e-10
Al ₂ O ₃	II	MAE	KNN	3.68	a	7.97e-10
Al ₂ O ₃	II	MAE	NNET	1.75	cd	7.97e-10
Al ₂ O ₃	II	MAE	RF	2.765	bc	7.97e-10
Al ₂ O ₃	II	MAE	SVMRadial	1.6	d	7.97e-10
Al ₂ O ₃	III	MAE	GLMNET	3.3	a	1.08e-09
Al ₂ O ₃	III	MAE	KNN	2.73	bc	1.08e-09
Al ₂ O ₃	III	MAE	NNET	2.38	cd	1.08e-09
Al ₂ O ₃	III	MAE	RF	3.19	ab	1.08e-09
Al ₂ O ₃	III	MAE	SVMRadial	2.32	d	1.08e-09
Al ₂ O ₃	IV	MAE	GLMNET	1.14	bc	2.55e-09
Al ₂ O ₃	IV	MAE	KNN	1.41	a	2.55e-09
Al ₂ O ₃	IV	MAE	NNET	1.39	ab	2.55e-09
Al ₂ O ₃	IV	MAE	RF	1.08	bc	2.55e-09
Al ₂ O ₃	IV	MAE	SVMRadial	1.01	c	2.55e-09
Al ₂ O ₃	V	MAE	GLMNET	2.67	ab	6.88e-10
Al ₂ O ₃	V	MAE	KNN	3.04	a	6.88e-10
Al ₂ O ₃	V	MAE	NNET	2.05	cd	6.88e-10
Al ₂ O ₃	V	MAE	RF	2.655	bc	6.88e-10
Al ₂ O ₃	V	MAE	SVMRadial	1.95	d	6.88e-10

Element	Scenario	Metric	Model	Median	Group	Kruskal_pvalue
Fe ₂ O ₃	I	MAE	GLMNET	24.08614	ab	5.51e-09
Fe ₂ O ₃	I	MAE	KNN	19.41545	cd	5.51e-09
Fe ₂ O ₃	I	MAE	NNET	66.48357	a	5.51e-09
Fe ₂ O ₃	I	MAE	RF	21.44187	bc	5.51e-09
Fe ₂ O ₃	I	MAE	SVMRadial	17.75558	d	5.51e-09
Fe ₂ O ₃	II	MAE	GLMNET	16.31359	ab	7.05e-09
Fe ₂ O ₃	II	MAE	KNN	16.05018	bc	7.05e-09
Fe ₂ O ₃	II	MAE	NNET	70.21677	a	7.05e-09
Fe ₂ O ₃	II	MAE	RF	14.76439	cd	7.05e-09
Fe ₂ O ₃	II	MAE	SVMRadial	13.10989	d	7.05e-09
Fe ₂ O ₃	III	MAE	GLMNET	19.26883	ab	1.16e-08
Fe ₂ O ₃	III	MAE	KNN	15.51267	cd	1.16e-08
Fe ₂ O ₃	III	MAE	NNET	67.32949	a	1.16e-08
Fe ₂ O ₃	III	MAE	RF	16.8656	bc	1.16e-08
Fe ₂ O ₃	III	MAE	SVMRadial	15.29639	d	1.16e-08
Fe ₂ O ₃	IV	MAE	GLMNET	17.40119	c	5.7e-06
Fe ₂ O ₃	IV	MAE	KNN	17.41684	bc	5.7e-06
Fe ₂ O ₃	IV	MAE	NNET	56.65842	ab	5.7e-06
Fe ₂ O ₃	IV	MAE	RF	17.26957	c	5.7e-06
Fe ₂ O ₃	IV	MAE	SVMRadial	18.52801	a	5.7e-06
Fe ₂ O ₃	V	MAE	GLMNET	14.42367	a	3.99e-08
Fe ₂ O ₃	V	MAE	KNN	14.23776	ab	3.99e-08
Fe ₂ O ₃	V	MAE	NNET	72.37012	a	3.99e-08
Fe ₂ O ₃	V	MAE	RF	13.16489	bc	3.99e-08
Fe ₂ O ₃	V	MAE	SVMRadial	12.4872	c	3.99e-08

Element	Scenario	Metric	Model	Median	Group	Kruskal_pvalue
MnO	I	MAE	GLMNET	16.33	a	2.57e-09
MnO	I	MAE	KNN	13.51	bc	2.57e-09
MnO	I	MAE	NNET	4.13	d	2.57e-09
MnO	I	MAE	RF	16.13	ab	2.57e-09
MnO	I	MAE	SVMRadial	8.57	cd	2.57e-09
MnO	II	MAE	GLMNET	6.93	ab	1.06e-09
MnO	II	MAE	KNN	6.95	a	1.06e-09
MnO	II	MAE	NNET	5.53	cd	1.06e-09
MnO	II	MAE	RF	6.25	bc	1.06e-09
MnO	II	MAE	SVMRadial	5.26	d	1.06e-09
MnO	III	MAE	GLMNET	6.68	ab	1.98e-09
MnO	III	MAE	KNN	6.32	bc	1.98e-09
MnO	III	MAE	NNET	5.52	cd	1.98e-09
MnO	III	MAE	RF	7.57	a	1.98e-09
MnO	III	MAE	SVMRadial	5.49	d	1.98e-09
MnO	IV	MAE	GLMNET	12.55	ab	3.27e-09
MnO	IV	MAE	KNN	11.8	bc	3.27e-09
MnO	IV	MAE	NNET	9.61	d	3.27e-09
MnO	IV	MAE	RF	13.21	a	3.27e-09
MnO	IV	MAE	SVMRadial	9.66	cd	3.27e-09
MnO	V	MAE	GLMNET	6.43	a	3.08e-09
MnO	V	MAE	KNN	5.54	ab	3.08e-09
MnO	V	MAE	NNET	1.99	cd	3.08e-09
MnO	V	MAE	RF	5.18	bc	3.08e-09
MnO	V	MAE	SVMRadial	1.97	d	3.08e-09
MnO	I	R ²	GLMNET	0.04	a	6.07e-06
MnO	I	R ²	KNN	0.01	b	6.07e-06
MnO	I	R ²	NNET	0.005	b	6.07e-06
MnO	I	R ²	RF	0.04	a	6.07e-06
MnO	I	R ²	SVMRadial	0.04	a	6.07e-06
MnO	II	R ²	GLMNET	0.12	bc	7.41e-10
MnO	II	R ²	KNN	0.19	ab	7.41e-10
MnO	II	R ²	NNET	0.04	cd	7.41e-10
MnO	II	R ²	RF	0.24	a	7.41e-10
MnO	II	R ²	SVMRadial	0.0	d	7.41e-10
MnO	III	R ²	KNN	0.03	a	8.42e-07
MnO	III	R ²	NNET	0.025	a	8.42e-07
MnO	III	R ²	RF	0.0	b	8.42e-07
MnO	III	R ²	SVMRadial	0.0	b	8.42e-07
MnO	IV	R ²	GLMNET	0.16	ab	1.7e-05
MnO	IV	R ²	KNN	0.18	a	1.7e-05
MnO	IV	R ²	NNET	0.16	ab	1.7e-05
MnO	IV	R ²	RF	0.14	b	1.7e-05
MnO	IV	R ²	SVMRadial	0.01	b	1.7e-05
MnO	V	R ²	GLMNET	0.05	a	3.44e-06
MnO	V	R ²	KNN	0.0	b	3.44e-06

MnO	V	R ²	NNET	0.06	a	3.44e-06
MnO	V	R ²	RF	0.0	b	3.44e-06
MnO	I	RMSE	GLMNET	21.86	ab	2.33e-08
MnO	I	RMSE	KNN	21.87	a	2.33e-08
MnO	I	RMSE	NNET	12.3	c	2.33e-08
MnO	I	RMSE	RF	21.765	a	2.33e-08
MnO	I	RMSE	SVMRadial	13.03	bc	2.33e-08
MnO	II	RMSE	GLMNET	13.29	bc	6.77e-10
MnO	II	RMSE	KNN	12.71	cd	6.77e-10
MnO	II	RMSE	NNET	14.61	ab	6.77e-10
MnO	II	RMSE	RF	12.15	d	6.77e-10
MnO	II	RMSE	SVMRadial	14.77	a	6.77e-10
MnO	III	RMSE	GLMNET	14.82	cd	1.71e-09
MnO	III	RMSE	KNN	14.67	d	1.71e-09
MnO	III	RMSE	NNET	15.235	ab	1.71e-09
MnO	III	RMSE	RF	14.965	bc	1.71e-09
MnO	III	RMSE	SVMRadial	15.29	a	1.71e-09
MnO	IV	RMSE	GLMNET	19.25	c	2.53e-06
MnO	IV	RMSE	KNN	19.35	bc	2.53e-06
MnO	IV	RMSE	NNET	22.405	ab	2.53e-06
MnO	IV	RMSE	RF	22.51	a	2.53e-06
MnO	IV	RMSE	SVMRadial	22.59	a	2.53e-06
MnO	V	RMSE	GLMNET	8.78	bc	1.28e-09
MnO	V	RMSE	KNN	12.15	a	1.28e-09
MnO	V	RMSE	NNET	6.84	d	1.28e-09
MnO	V	RMSE	RF	10.55	ab	1.28e-09
MnO	V	RMSE	SVMRadial	6.92	cd	1.28e-09

Element	Scenario	Metric	Model	Median	Group	Kruskal_pvalue
Nb ₂ O ₅	I	MAE	GLMNET	0.62	abc	3.92e-06
Nb ₂ O ₅	I	MAE	KNN	0.67	a	3.92e-06
Nb ₂ O ₅	I	MAE	NNET	0.6	bc	3.92e-06
Nb ₂ O ₅	I	MAE	RF	0.64	ab	3.92e-06
Nb ₂ O ₅	I	MAE	SVMRadial	0.59	c	3.92e-06
Nb ₂ O ₅	II	MAE	GLMNET	0.71	bc	2.03e-06
Nb ₂ O ₅	II	MAE	KNN	0.73	ab	2.03e-06
Nb ₂ O ₅	II	MAE	NNET	0.7	c	2.03e-06
Nb ₂ O ₅	II	MAE	RF	0.74	a	2.03e-06
Nb ₂ O ₅	II	MAE	SVMRadial	0.71	c	2.03e-06
Nb ₂ O ₅	III	MAE	GLMNET	0.86	a	1.35e-08
Nb ₂ O ₅	III	MAE	KNN	0.74	cd	1.35e-08
Nb ₂ O ₅	III	MAE	NNET	0.67	d	1.35e-08
Nb ₂ O ₅	III	MAE	RF	0.81	ab	1.35e-08
Nb ₂ O ₅	III	MAE	SVMRadial	0.75	bc	1.35e-08
Nb ₂ O ₅	IV	MAE	GLMNET	0.42	ab	2.06e-08
Nb ₂ O ₅	IV	MAE	KNN	0.29	c	2.06e-08
Nb ₂ O ₅	IV	MAE	NNET	0.45	a	2.06e-08
Nb ₂ O ₅	IV	MAE	RF	0.335	bc	2.06e-08
Nb ₂ O ₅	IV	MAE	SVMRadial	0.42	a	2.06e-08
Nb ₂ O ₅	V	MAE	GLMNET	0.71	bc	8e-09
Nb ₂ O ₅	V	MAE	KNN	0.73	ab	8e-09
Nb ₂ O ₅	V	MAE	NNET	0.69	c	8e-09
Nb ₂ O ₅	V	MAE	RF	0.74	a	8e-09
Nb ₂ O ₅	V	MAE	SVMRadial	0.68	c	8e-09
Nb ₂ O ₅	I	R ²	GLMNET	0.0	ab	0.000625424
Nb ₂ O ₅	I	R ²	KNN	0.01	a	0.000625424
Nb ₂ O ₅	I	R ²	NNET	0.07	a	0.000625424
Nb ₂ O ₅	I	R ²	RF	0.0	b	0.000625424
Nb ₂ O ₅	I	R ²	SVMRadial	0.0	b	0.000625424
Nb ₂ O ₅	II	R ²	GLMNET	0.04	ab	0.004565541
Nb ₂ O ₅	II	R ²	KNN	0.04	a	0.004565541
Nb ₂ O ₅	II	R ²	NNET	0.025	ab	0.004565541
Nb ₂ O ₅	II	R ²	RF	0.015	b	0.004565541
Nb ₂ O ₅	III	R ²	GLMNET	0.0	b	3.06e-07
Nb ₂ O ₅	III	R ²	KNN	0.04	a	3.06e-07
Nb ₂ O ₅	III	R ²	NNET	0.02	a	3.06e-07
Nb ₂ O ₅	III	R ²	RF	0.03	a	3.06e-07
Nb ₂ O ₅	III	R ²	SVMRadial	0.0	b	3.06e-07
Nb ₂ O ₅	IV	R ²	GLMNET	0.01	b	8.53e-07
Nb ₂ O ₅	IV	R ²	KNN	0.71	a	8.53e-07
Nb ₂ O ₅	IV	R ²	NNET	0.03	b	8.53e-07
Nb ₂ O ₅	IV	R ²	RF	0.38	a	8.53e-07
Nb ₂ O ₅	V	R ²	GLMNET	0.04	c	1.88e-06
Nb ₂ O ₅	V	R ²	KNN	0.1	a	1.88e-06
Nb ₂ O ₅	V	R ²	NNET	0.05	bc	1.88e-06
Nb ₂ O ₅	V	R ²	RF	0.07	ab	1.88e-06
Nb ₂ O ₅	I	RMSE	GLMNET	0.91	ab	2.65e-06
Nb ₂ O ₅	I	RMSE	KNN	0.92	a	2.65e-06

Nb ₂ O ₅	I	RMSE	NNET	0.83	c	2.65e-06
Nb ₂ O ₅	I	RMSE	RF	0.86	c	2.65e-06
Nb ₂ O ₅	I	RMSE	SVMRadial	0.89	bc	2.65e-06
Nb ₂ O ₅	II	RMSE	GLMNET	0.88	b	0.001545531
Nb ₂ O ₅	II	RMSE	KNN	0.92	a	0.001545531
Nb ₂ O ₅	II	RMSE	NNET	0.92	ab	0.001545531
Nb ₂ O ₅	II	RMSE	RF	0.91	ab	0.001545531
Nb ₂ O ₅	II	RMSE	SVMRadial	0.9	b	0.001545531
Nb ₂ O ₅	III	RMSE	GLMNET	1.08	a	3.13e-09
Nb ₂ O ₅	III	RMSE	KNN	0.92	cd	3.13e-09
Nb ₂ O ₅	III	RMSE	NNET	0.875	d	3.13e-09
Nb ₂ O ₅	III	RMSE	RF	1.01	ab	3.13e-09
Nb ₂ O ₅	III	RMSE	SVMRadial	0.96	bc	3.13e-09
Nb ₂ O ₅	IV	RMSE	GLMNET	0.48	ab	2.17e-09
Nb ₂ O ₅	IV	RMSE	KNN	0.36	c	2.17e-09
Nb ₂ O ₅	IV	RMSE	NNET	0.555	a	2.17e-09
Nb ₂ O ₅	IV	RMSE	RF	0.39	bc	2.17e-09
Nb ₂ O ₅	IV	RMSE	SVMRadial	0.48	a	2.17e-09
Nb ₂ O ₅	V	RMSE	GLMNET	0.86	c	2.54e-07
Nb ₂ O ₅	V	RMSE	KNN	0.88	ab	2.54e-07
Nb ₂ O ₅	V	RMSE	NNET	0.86	c	2.54e-07
Nb ₂ O ₅	V	RMSE	RF	0.89	a	2.54e-07
Nb ₂ O ₅	V	RMSE	SVMRadial	0.87	bc	2.54e-07

Element	Scenario	Metric	Model	Median	Group	Kruskal_pvalue
TiO ₂	I	MAE	GLMNET	2.71	ab	5.05E-09
TiO ₂	I	MAE	KNN	2.79	a	5.05E-09
TiO ₂	I	MAE	NNET	2.525	c	5.05E-09
TiO ₂	I	MAE	RF	2.585	c	5.05E-09
TiO ₂	I	MAE	SVMRadial	2.64	bc	5.05E-09
TiO ₂	II	MAE	GLMNET	2.72	ab	1.19E-09
TiO ₂	II	MAE	KNN	2.81	a	1.19E-09
TiO ₂	II	MAE	NNET	2.62	cd	1.19E-09
TiO ₂	II	MAE	RF	2.67	bc	1.19E-09
TiO ₂	II	MAE	SVMRadial	2.6	d	1.19E-09
TiO ₂	III	MAE	GLMNET	2.71	ab	5.05E-09
TiO ₂	III	MAE	KNN	2.79	a	5.05E-09
TiO ₂	III	MAE	NNET	2.525	c	5.05E-09
TiO ₂	III	MAE	RF	2.585	c	5.05E-09
TiO ₂	III	MAE	SVMRadial	2.64	bc	5.05E-09
TiO ₂	IV	MAE	GLMNET	3.48	a	2.05E-09
TiO ₂	IV	MAE	KNN	2.48	d	2.05E-09
TiO ₂	IV	MAE	NNET	3.13	ab	2.05E-09
TiO ₂	IV	MAE	RF	2.56	cd	2.05E-09
TiO ₂	IV	MAE	SVMRadial	3.07	bc	2.05E-09
TiO ₂	V	MAE	GLMNET	2.71	a	6.93E-10
TiO ₂	V	MAE	KNN	2.66	ab	6.93E-10
TiO ₂	V	MAE	NNET	2.4	cd	6.93E-10
TiO ₂	V	MAE	RF	2.56	bc	6.93E-10
TiO ₂	V	MAE	SVMRadial	2.38	d	6.93E-10
TiO ₂	I	R ²	GLMNET	0	b	8.66E-09
TiO ₂	I	R ²	KNN	0	b	8.66E-09
TiO ₂	I	R ²	NNET	0.025	a	8.66E-09
TiO ₂	I	R ²	RF	0	b	8.66E-09
TiO ₂	I	R ²	SVMRadial	0.01	a	8.66E-09
TiO ₂	II	R ²	GLMNET	0.1	a	2.24E-09
TiO ₂	II	R ²	KNN	0.06	bc	2.24E-09
TiO ₂	II	R ²	NNET	0.02	cd	2.24E-09
TiO ₂	II	R ²	RF	0.1	ab	2.24E-09
TiO ₂	II	R ²	SVMRadial	0	d	2.24E-09
TiO ₂	III	R ²	GLMNET	0	b	8.66E-09
TiO ₂	III	R ²	KNN	0	b	8.66E-09
TiO ₂	III	R ²	NNET	0.025	a	8.66E-09
TiO ₂	III	R ²	RF	0	b	8.66E-09
TiO ₂	III	R ²	SVMRadial	0.01	a	8.66E-09
TiO ₂	IV	R ²	GLMNET	0.17	bc	4.06E-09
TiO ₂	IV	R ²	KNN	0.39	a	4.06E-09
TiO ₂	IV	R ²	NNET	0.08	d	4.06E-09
TiO ₂	IV	R ²	RF	0.26	ab	4.06E-09
TiO ₂	IV	R ²	SVMRadial	0.1	cd	4.06E-09
TiO ₂	V	R ²	GLMNET	0.09	ab	8.35E-10
TiO ₂	V	R ²	KNN	0.09	ab	8.35E-10
TiO ₂	V	R ²	NNET	0.04	bc	8.35E-10
TiO ₂	V	R ²	RF	0.185	a	8.35E-10
TiO ₂	V	R ²	SVMRadial	0	c	8.35E-10

Element	Scenario	Metric	Model	Median	Group	Kruskal_pvalue
SiO ₂	I	MAE	GLMNET	4.45	d	1.49E-09
SiO ₂	I	MAE	KNN	4.49	bc	1.49E-09
SiO ₂	I	MAE	NNET	4.56	ab	1.49E-09
SiO ₂	I	MAE	RF	4.48	cd	1.49E-09
SiO ₂	I	MAE	SVMRadial	4.68	a	1.49E-09
SiO ₂	II	MAE	GLMNET	5.71	a	8.88E-10
SiO ₂	II	MAE	KNN	4.22	ab	8.88E-10
SiO ₂	II	MAE	NNET	0.67	cd	8.88E-10
SiO ₂	II	MAE	RF	2.835	bc	8.88E-10
SiO ₂	II	MAE	SVMRadial	0.53	d	8.88E-10
SiO ₂	III	MAE	GLMNET	3.26	ab	8.42E-10
SiO ₂	III	MAE	KNN	3.39	a	8.42E-10
SiO ₂	III	MAE	NNET	2.8	d	8.42E-10
SiO ₂	III	MAE	RF	3.16	bc	8.42E-10
SiO ₂	III	MAE	SVMRadial	2.97	cd	8.42E-10
SiO ₂	IV	MAE	GLMNET	9.5	bc	7.07E-08
SiO ₂	IV	MAE	KNN	9.51	ab	7.07E-08
SiO ₂	IV	MAE	NNET	9.51	b	7.07E-08
SiO ₂	IV	MAE	RF	9.47	c	7.07E-08
SiO ₂	IV	MAE	SVMRadial	9.57	a	7.07E-08
SiO ₂	V	MAE	GLMNET	3.91	bc	2.78E-09
SiO ₂	V	MAE	KNN	4.71	a	2.78E-09
SiO ₂	V	MAE	NNET	2.74	c	2.78E-09
SiO ₂	V	MAE	RF	4.07	ab	2.78E-09
SiO ₂	V	MAE	SVMRadial	2.75	c	2.78E-09
SiO ₂	I	R ²	GLMNET	0.13	a	2.03E-05
SiO ₂	I	R ²	KNN	0.09	a	2.03E-05
SiO ₂	I	R ²	NNET	0.065	a	2.03E-05
SiO ₂	I	R ²	RF	0.07	a	2.03E-05
SiO ₂	I	R ²	SVMRadial	0.01	b	2.03E-05
SiO ₂	II	R ²	GLMNET	0.04	b	6.06E-08
SiO ₂	II	R ²	KNN	0.02	c	6.06E-08
SiO ₂	II	R ²	NNET	0.015	c	6.06E-08
SiO ₂	II	R ²	RF	0.04	b	6.06E-08
SiO ₂	II	R ²	SVMRadial	0.14	a	6.06E-08
SiO ₂	III	R ²	GLMNET	0	a	3.93E-05
SiO ₂	III	R ²	KNN	0	b	3.93E-05
SiO ₂	III	R ²	NNET	0.01	ab	3.93E-05
SiO ₂	III	R ²	RF	0	b	3.93E-05
SiO ₂	III	R ²	SVMRadial	0.01	a	3.93E-05
SiO ₂	IV	R ²	GLMNET	0.28	a	7.46E-07
SiO ₂	IV	R ²	KNN	0.2	bc	7.46E-07
SiO ₂	IV	R ²	NNET	0.25	ab	7.46E-07
SiO ₂	IV	R ²	RF	0.26	ab	7.46E-07
SiO ₂	IV	R ²	SVMRadial	0	c	7.46E-07

SiO ₂	V	R ²	GLMNET	0	ab	0.00145
SiO ₂	V	R ²	KNN	0	b	0.00145
SiO ₂	V	R ²	NNET	0.015	a	0.00145
SiO ₂	V	R ²	RF	0	b	0.00145
SiO ₂	V	R ²	SVMRadial	0	b	0.00145
SiO ₂	I	RMSE	GLMNET	16.75	d	2.87E-09
SiO ₂	I	RMSE	KNN	16.76	cd	2.87E-09
SiO ₂	I	RMSE	NNET	16.87	ab	2.87E-09
SiO ₂	I	RMSE	RF	16.8	bc	2.87E-09
SiO ₂	I	RMSE	SVMRadial	16.96	a	2.87E-09
SiO ₂	II	RMSE	GLMNET	7.07	ab	8.32E-10
SiO ₂	II	RMSE	KNN	7.69	a	8.32E-10
SiO ₂	II	RMSE	NNET	1.345	cd	8.32E-10
SiO ₂	II	RMSE	RF	5.55	bc	8.32E-10
SiO ₂	II	RMSE	SVMRadial	1.27	d	8.32E-10
SiO ₂	III	RMSE	GLMNET	12.57	cd	2.86E-09
SiO ₂	III	RMSE	KNN	12.67	a	2.86E-09
SiO ₂	III	RMSE	NNET	12.65	ab	2.86E-09
SiO ₂	III	RMSE	RF	12.56	d	2.86E-09
SiO ₂	III	RMSE	SVMRadial	12.58	bc	2.86E-09
SiO ₂	IV	RMSE	GLMNET	28.73	c	5.90E-09
SiO ₂	IV	RMSE	KNN	28.74	bc	5.90E-09
SiO ₂	IV	RMSE	NNET	28.82	ab	5.90E-09
SiO ₂	IV	RMSE	RF	28.74	c	5.90E-09
SiO ₂	IV	RMSE	SVMRadial	28.91	a	5.90E-09
SiO ₂	V	RMSE	GLMNET	13.48	d	9.80E-10
SiO ₂	V	RMSE	KNN	15.52	a	9.80E-10
SiO ₂	V	RMSE	NNET	13.64	cd	9.80E-10
SiO ₂	V	RMSE	RF	15.01	ab	9.80E-10
SiO ₂	V	RMSE	SVMRadial	13.67	bc	9.80E-10

9.4 Apêndice D – Valores observados vs. preditos de Nb_2O_5 , TiO_2 e Fe_2O_3 nos cenários de amostragem I-V para os modelos de aprendizagem automática avaliados.

