

UNIVERSIDADE FEDERAL DE VIÇOSA

**Modelagem e mapeamento dos teores de ferro em solos do Quadrilátero
Ferrífero: aplicações de mapeamento digital de solos**

Marcelo Batista Krause
Magister Scientiae

**VIÇOSA - MINAS GERAIS
2025**

MARCELO BATISTA KRAUSE

**Modelagem e mapeamento dos teores de ferro em solos do Quadrilátero
Ferrífero: aplicações de mapeamento digital de solos**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Solos e Nutrição de Plantas, para obtenção do título de *Magister Scientiae*.

Orientador: Marcio Rocha Francelino

Coorientador: Elpidio I. Fernandes Filho

**VIÇOSA - MINAS GERAIS
2025**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

K91m
2025 Krause, Marcelo Batista, 1998-
Modelagem e mapeamento dos teores de ferro em solos do
Quadrilátero Ferrífero: aplicações de mapeamento digital de
solos / Marcelo Batista Krause. – Viçosa, MG, 2025.
1 dissertação eletrônica (89 f.): il. (algumas color.).

Orientador: Márcio Rocha Francelino.
Dissertação (mestrado) - Universidade Federal de Viçosa,
Departamento de Solos, 2025.

Inclui bibliografia.

DOI: <https://doi.org/10.47328/ufvbbt.2025.267>

Modo de acesso: World Wide Web.

1. Solos - Teor de ferro - Quadrilátero Ferrífero, Região do
(MG). 2. Aprendizado do computador. 3. Sustentabilidade.
4. Solos tropicais. 5. Mapeamento digital. I. Francelino, Márcio
Rocha, 1966-. II. Universidade Federal de Viçosa. Departamento
de Solos. Programa de Pós-Graduação em Solos e Nutrição de
Plantas. III. Título.

CDD 22. ed. 631.4

MARCELO BATISTA KRAUSE

**Modelagem e mapeamento dos teores de ferro em solos do Quadrilátero
Ferrífero: aplicações de mapeamento digital de solos**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Solos e Nutrição de Plantas, para obtenção do título de *Magister Scientiae*.

APROVADA: 25 de fevereiro de 2025.

Assentimento:

Marcelo Batista Krause
Autor

Marcio Rocha Francelino
Orientador

Essa dissertação foi assinada digitalmente pelo autor em 13/05/2025 às 20:37:01 e pelo orientador em 25/05/2025 às 11:56:53. As assinaturas têm validade legal, conforme o disposto na Medida Provisória 2.200-2/2001 e na Resolução nº 37/2012 do CONARQ. Para conferir a autenticidade, acesse <https://siadoc.ufv.br/validar-documento>. No campo 'Código de registro', informe o código **Y82M.7V9K.9DS6** e clique no botão 'Validar documento'.

Aos meus pais e pessoas queridas.

AGRADECIMENTOS

Antes de tudo, gostaria de agradecer aos meus pais, Maria Célia e João Krause, por todo o suporte ao longo da vida. Foram eles que me permitiram finalizar mais essa etapa. Aquele calouro de Geografia, completamente perdido, que achava legal o título de mestre e sempre sonhou em ser um, hoje defende sua dissertação. Ao longo do mestrado e da vida acadêmica, muitas pessoas, de certa forma, colaboraram com minha trajetória. É até difícil citar o nome de todas, mas gostaria de agradecer, em especial, às pessoas que fazem parte do Departamento de Solos (DPS) e à equipe do Laboratório de Pedometria e Geoprocessamento (LabGeo). Ao professor Márcio, que, desde a iniciação científica, sempre me deu total liberdade para desenvolver o que eu quisesse. Ao professor Elpídio, pelos conselhos, puxões de orelha e ideias mirabolantes que sempre surgiam e que, como ele sempre disse: “o mestrado só serve para fazer doutorado” — e quem disse que ele estava errado? Aos amigos e companheiros de LabGeo, em especial ao Luís, Rafa e Cássio. Aos dois primeiros, agradeço por todo o suporte ao longo do mestrado, desde as discussões em campo para definir pontos de coleta até as conversas e ideias que surgiram ao longo da pesquisa. Deixo aqui registrado todo o meu agradecimento por aceitarem encarar um campo no Quadrilátero de Voyage. Ao Cássio, pelo suporte desde a graduação e pelas ajudas nas modelagens. Sem vocês três, essa dissertação não teria nascido. Gostaria de agradecer ao Waldir, que, quando a situação ficou complicada, surgiu como uma luz no fim do túnel. Sem ele, nenhuma das análises teria ocorrido. Aos meus queridos Merecedores, obrigado por todos os momentos. Nada mais rejuvenescedor do que uma cerveja na Rita. Em muitos momentos, estar com vocês foi a forma de fugir dos problemas e conseguir deixar a cabeça livre. Não poderia deixar de agradecer à pessoa mais importante de todas. Mari! Obrigado por todos os momentos e por todo o suporte! Mesmo à distância, você sempre foi meu porto seguro! Todas as risadas, raivas, ligações e estresses diários valeram a pena. Ressalto o agradecimento ao Banco de Solos de Minas Gerais (FEAM/UFV) e ao coordenador do banco, Professor José João Leal de Souza, pela liberação e acesso aos resultados analíticos. Para acabar, obrigado UFV! Obrigado DPS! Obrigado LabGeo! Este trabalho foi realizado com o apoio das seguintes agências de pesquisa brasileiras: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

*“É necessário sempre acreditar que o sonho é possível!
Que o céu é o limite e você, truta, é imbatível!
Que o tempo ruim vai passar e é só uma fase!”.*

Racionais MC's – A vida é desafio.

Aivaldo Pereira Alves.

Cristian De Souza Augusto.

RESUMO

KRAUSE, Marcelo Batista, M.Sc., Universidade Federal de Viçosa, fevereiro de 2025. **Modelagem e mapeamento dos teores de ferro em solos do Quadrilátero Ferrífero: aplicações de mapeamento digital de solos.** Orientador: Marcio Rocha Francelino. Coorientador: Elpidio Inacio Fernandes Filho.

O Quadrilátero Ferrífero (QF) é a principal província mineralógica do sudeste brasileiro e possui uma riqueza ambiental ímpar. Apesar de ser historicamente marcado pelo extrativismo mineral e contar com diversos estudos ambientais detalhados, o conhecimento sobre seus solos ainda é incipiente. Nesse contexto, o Mapeamento Digital de Solos surge como uma alternativa acessível para a geração de informações pedológicas em escala adequada. Os métodos quantitativos proporcionaram uma nova perspectiva para o mapeamento de solos e possibilitam a melhor representação espacial de atributos do solo, de forma rápida, precisa e acurada. Assim, objetivou-se com esta pesquisa construir modelos preditivos, definidos como funções de pedotransferência (FPT) para a predição dos teores de ferro em solos do QF e espacializar por meio de mapas os teores de ferro. Utilizou-se dados de amostras do Banco de Solos de Minas Gerais, complementados por coletadas realizadas na profundidade de 0 a 20 cm, totalizando 165 amostras. Para a seleção dos subconjuntos das covariáveis mais importantes e reduzir a redundância, empregou-se a seleção por correlação ($|r| < 0,95$) e a seleção pela importância, por meio do Recursive Feature Elimination. Testou-se seis algoritmos de machine learning e executou-se os processos de seleção, treinamento, validação, teste e predição por 100 vezes. Observou-se que o Random Forest (RF) e o Gradient Boosting Machine (GBM) apresentaram-se como os modelos mais consistentes durante a construção das FPT. No que diz respeito à espacialização e construção dos mapas, os modelos RF e K-Nearest Neighbors (KNN) se destacaram. Em ambas as análises, as covariáveis bioclimáticas, morfométricas e aquelas relacionadas aos materiais de origem, como os dados de gamaespectrometria, emergiram como as mais relevantes. As amostras coletadas representaram bem a ampla diversidade litoestratigráfica do QF. Além disso, observou-se que as maiores concentrações de ferro em solos do QF ocorrem em regiões predominantemente compostas por materiais ferruginosos, que geralmente apresentam maior resistência ao intemperismo. Por outro lado, os teores mais baixos encontrados, foram nas regiões de menores altitudes, onde o material originário possui menor concentração de ferro. De forma geral, os modelos desenvolvidos demonstraram alta capacidade de identificação dos padrões presentes nos dados, seja para a construção

de FPT ou para a geração dos mapas.

Palavras-chave: Aprendizado de máquina; Sustentabilidade; Pedometria; Solos tropicais

ABSTRACT

KRAUSE, Marcelo Batista, M.Sc., Universidade Federal de Viçosa, February, 2025. **Modeling and mapping of iron contents in soils of the Quadrilátero Ferrífero: applications of digital soil mapping.** Adviser: Marcio Rocha Francelino. Co-adviser: Elpidio Inacio Fernandes Filho.

The Quadrilátero Ferrífero (QF) is the main mineralogical province in southeastern Brazil and possesses unique environmental richness. Despite being historically marked by mineral extraction and having numerous detailed environmental studies, knowledge about its soils remains limited. In this context, Digital Soil Mapping emerges as an accessible alternative for generating pedological information at an appropriate scale. Quantitative methods have provided a new perspective for soil mapping, enabling a more accurate, precise, and rapid spatial representation of soil attributes. Thus, this research aimed to develop predictive models, defined as pedotransfer functions (PTF), to predict iron contents concentrations in QF soils and to spatialize these concentrations through maps. Data from soil samples in the Minas Gerais Soil Bank were used, complemented by additional collections at a depth of 0 to 20 cm, totaling 165 samples. To select the most important covariate subsets and reduce redundancy, correlation selection ($|r| < 0.95$) and importance-based selection using Recursive Feature Elimination (RFE) were applied. Six machine learning algorithms were tested, and the selection, training, validation, testing, and prediction processes were executed 100 times. The results showed that the Random Forest (RF) and Gradient Boosting Machine (GBM) models were the most consistent during PTF development. Regarding spatialization and map construction, the RF and K-Nearest Neighbors (KNN) models stood out. In both analyses, bioclimatic and morphometric covariates, as well as those related to parent materials—such as gamma spectrometry data—emerged as the most relevant. The collected samples effectively represented the broad lithostratigraphic diversity of the QF. Furthermore, it was observed that the highest iron concentrations in QF soils occur in regions predominantly composed of ferruginous materials, which are generally more resistant to weathering. Conversely, the lowest iron concentrations were found in lower-altitude regions, where the parent material has lower iron content. Overall, the developed models demonstrated a high capacity for identifying patterns within the data, both for PTF construction and for generating spatial maps.

Keywords: Machine learning; Sustainability; Pedometrics; Tropical soils

LISTA DE FIGURAS

CAPÍTULO 1

Figura 1 – Mapa de localização do QF e dos locais de coleta das amostras.	22
Figura 2 – Mapa dos locais de coleta na área central do QF.	24
Figura 3A – Susceptibilímetro magnético Terraplus KT10.	25
Figura 3B – Realização da leitura, para evitar contaminação entre o sensor e amostra.	25
Figura 4 – Mapa de agrupamento das litoestruturas para fins pedológicos do QF.	29
Figura 5 – Mapa de agrupamento em “pedorochas” para fins pedológicos do QF.	30
Figura 6 – Fluxograma da metodologia desenvolvida.	31
Figura 7 – Valores preditos x valores observados para os modelos RF e GBM.	35
Figura 8 – Amplitude de valores determinados de ferro separados de acordo com a litoestratigrafia e a classificação de “pedorochas”.	36
Figura 9 – Performance dos modelos utilizados. “sem SM” são as métricas sem a susceptibilidade magnética e “com SM” são com a presença da susceptibilidade.	37
Figura 10 – LCCC dos modelos durante o processo do RFE. “sem SM” são as métricas sem a susceptibilidade magnética e “com SM” são com a presença da susceptibilidade.	39
Figura 11 – MAE dos modelos durante o processo do RFE. “sem SM” são as métricas sem a susceptibilidade magnética e “com SM” são com a presença da susceptibilidade.	40
Figura 12 – RMSE dos modelos durante o processo do RFE. “sem SM” são as métricas sem a susceptibilidade magnética e “com SM” são com a presença da susceptibilidade.	41
Figura 13 – Importância das variáveis escolhidas de acordo com os modelos. Os valores indicam a frequência de repetição. “sem SM” são as métricas sem a susceptibilidade magnética e “com SM” são com a presença da susceptibilidade.	42
Figura 14 – Resultado da correlação de Spearman entre as variáveis pedológicas.	43

CAPÍTULO 2

Figura 1 – Mapa de localização da área de estudo.	60
Figura 2 – Agrupamento litoestratigráfico para fins pedológicos, considerando os “litotipos”.	63
Figura 3 – Agrupamento litoestratigráfico para fins pedológicos, considerando as “pedorochas”.	64
Figura 4 – Fluxograma da metodologia realizada.	67
Figura 5 – Desempenho dos modelos de acordo com as diferentes métricas utilizadas para avaliação. Perceptível as diferenças estatísticas com exceção ao R^2	71
Figura 6 – Importância das variáveis escolhidas de acordo com os modelos.	74
Figura 7 – Mapas das predições médias dos teores de ferro em solos do QF.	76
Figura 8 – Mapas das predições medianas dos teores de ferro em solos do QF.	77
Figura 9 – Mapas dos coeficientes de variação das predições dos teores de ferro em solos do QF.	78
Figura 10 – Mapas dos desvios padrões das predições dos teores de ferro em solos do QF. ..	79
Figura 11 – Mapas do quartil de 5% dos valores preditos dos teores de ferro em solos do QF.	80

Figura 12 – Mapas do quartil de 95% dos valores preditos dos teores de ferro em solos do QF.	81
Figura 13 – Distribuição dos teores de ferro em solos do QF agrupados para fins pedológicos.	83

LISTA DE TABELAS

CAPÍTULO 1

Tabela 1 – Lista de covariáveis utilizadas para a geração dos locais de coleta via cLHS.	23
Tabela 2 – Covariáveis representativas do modelo SCORPAN selecionadas para a modelagem dos teores de Ferro nos solos do QF.	26
Tabela 3 – Agrupamentos de rochas para fins pedológicos.	28
Tabela 4 – Equações das métricas utilizadas para avaliar o desempenho dos modelos.	33
Tabela 5 – Ganhos de performance dos modelos comparados aos modelos nulos.	38
Tabela 6 – Avaliação dos melhores e piores desempenhos dos modelos RF e GBM, considerando o número de covariáveis utilizadas para alcançar tal valor.	38

CAPÍTULO 2

Tabela 1 – Covariáveis bioclimáticas escolhidas para a predição espacial dos teores de ferro nos solos do QF.....	62
Tabela 2 – Covariáveis morfométricas escolhidas para a predição espacial dos teores de ferro nos solos do QF.	65
Tabela 3 – Agrupamentos de rochas para fins pedológicos.	65
Tabela 4 – Covariáveis geofísicas escolhidas para a predição espacial dos teores de ferro nos solos do QF.	65
Tabela 5 – Covariáveis representativas de atividades biológicas, usos do solo e índice de solo escolhidas para a predição espacial dos teores de ferro nos solos do QF.	66
Tabela 6 – Equações das métricas utilizadas para avaliação dos modelos.	69
Tabela 7 – Avaliação das métricas médias dos modelos após 100 rodadas.	72
Tabela 8 – Comparações entre valores preditos e observados entre os dois modelos preditores.	81

SUMÁRIO

INTRODUÇÃO GERAL	13
REFERÊNCIAS	14
CAPÍTULO 1	17
1. INTRODUÇÃO	19
2. MATERIAIS E MÉTODOS.....	20
2.1 Caracterização da área de estudo.....	20
2.2 Bases de dados e amostragem	22
2.3 Análises laboratoriais	23
2.4 Susceptibilidade Magnética (SM)	24
2.5 Covariáveis explicativas e preditoras	25
2.6 Seleção e importância de covariáveis preditoras.....	29
2.7 Modelos de machine learning	31
2.8 Treino, teste e modelagem	31
2.9 Análises de correlação.....	32
2.10 Remoção de Susceptibilidade.....	33
3. RESULTADOS.....	33
3.1 Amostragem e representatividade	33
3.2 Performance dos Modelos	35
3.3 Seleção de modelos	37
3.4 Importância dos preditores	40
4. DISCUSSÕES.....	42
4.1 Escolha dos modelos para construção da FPT	42
4.2 Subconjuntos escolhidos pelo RFE para modelagem dos teores de ferro.....	44
4.3 Importância das variáveis e suas relações com os teores de ferro	45
4.4 Por que utilizar uma FPT?.....	47
5. CONCLUSÕES	48
REFERÊNCIAS	48
CAPÍTULO 2	55
1. INTRODUÇÃO	57
2. MATERIAIS E MÉTODOS.....	58
2.1 Caracterização da área de estudo.....	58
2.2 Covariáveis ambientais e dados de solos	60
2.3 Seleções de covariáveis e importâncias dos preditores	65
2.4 Modelos de Machine Learning.....	67
2.5 Treinos, testes, previsões, especializações, otimizações e incertezas	68
3. RESULTADOS E DISCUSSÕES	69
3.1 Avaliações das performances dos modelos	69
3.2 Importância e frequência das covariáveis e suas relações com os teores de ferro .	72
3.3 Distribuição espacial dos teores de ferro, previsões, incertezas e usos	74
4. CONCLUSÕES	82
REFERÊNCIAS.....	83
CONSIDERAÇÕES FINAIS	89

INTRODUÇÃO GERAL

O Quadrilátero Ferrífero (QF) é o berço da mineração organizada no Brasil e o coração da região das “minas” do período colonial (Madeira et al., 2015; Tarso et al., 2020). Historicamente, o QF desempenhou um papel fundamental na economia nacional e na construção da identidade brasileira, atraindo o interesse de técnicos, exploradores e acadêmicos de diversas áreas do conhecimento (Tarso et al., 2020; Salgado & Carmo, 2015). Além de sua relevância econômica e histórica, o QF destaca-se por suas características singulares que resultam em paisagens marcantes fornecedoras de serviços ambientais essenciais à sociedade e estratégicas para a conservação da biodiversidade (Carmo, 2010; Carmo & Kamino, 2015).

Descrita por Saadi (1991) como um “quebra-cabeças geológico-geomorfológico”, o QF combina fortes contrastes altimétricos e alta complexidade tectono-estrutural (Barbosa & Oliveira, 2020; Lana, 2015; Salgado & Carmo, 2015), o que o torna objeto de estudos voltados tanto à exploração de recursos minerais quanto à preservação de seus recursos naturais (Beato et al., 2005; Davis et al., 2005; Medina et al., 2005; Oliveira et al., 2005; Shinzato & Carvalho-Filho, 2005; Silva et al., 2005; Madeira et al., 2015; Schaefer et al., 2015; Assis et al., 2021). No entanto, entre os recursos naturais do QF, os solos permanecem como um dos elementos menos estudados (Carvalho-Filho et al., 2010; Schaefer et al., 2015; Coelho et al., 2017).

À predominância histórica da mineração na região relegou a segundo plano o mapeamento pedológico, tradicionalmente associado à avaliação da aptidão agrícola em outras áreas (Costa, 2003; Carvalho-Filho, 2008; Carvalho-Filho et al., 2010; 2011; Schaefer et al., 2015; Coelho et al., 2017). Apesar disso, o estudo dos solos é crucial para o planejamento ambiental e o uso sustentável das terras (Demattê et al., 2019; Reddy & Das, 2023), especialmente em regiões marcadas por intensa atividade antrópica. Segundo Shinzato & Carvalho-Filho (2005), o levantamento e o mapeamento de solos são ferramentas indispensáveis ao planejamento do uso das terras por serem os mapas de solos, importantes estratificadores de ambientes (Resende et al., 2023).

Nesse contexto, o Mapeamento Digital de Solos (MDS) tem se mostrado uma abordagem viável e inovadora (McBratney et al., 2003; Wadoux et al., 2020; Araújo et al., 2024; Bastos et al., 2024), especialmente em áreas de difícil acesso. O avanço de tecnologias como softwares, hardwares e sistemas automatizados têm permitido maior precisão e agilidade no estudo dos solos, além de reduzir a subjetividade associada aos modelos mentais convencionais dos pedólogos (ten Caten et al., 2012; Ma et al., 2019; Wadoux et al., 2020). Ademais, o uso de dados legados, provenientes de levantamentos pedológicos convencionais, têm contribuído para otimizar recursos humanos e financeiros (ten Caten et al., 2012; Carvalho-Júnior et al., 2013;

Mondejar & Tongco, 2019; Carvalho-Júnior et al., 2024), enquanto amplia a base de informações para o MDS.

A partir da combinação de dados legados e com covariáveis ambientais, representativas dos fatores de formação dos solos, através de relações matematicamente mensuradas como fatores de formação do solo (Wadoux et al., 2020), o MDS possibilita a produção rápida e acessível de informações pedológicas (McBratney et al., 2003; Carvalho-Júnior et al., 2024; Ceddia et al., 2024). Isso é particularmente relevante para o QF devido ao conhecimento limitado sobre os seus solos. Portanto, esta pesquisa busca utilizar técnicas de MDS, como diferentes modelos de Machine Learning, para aprofundar o conhecimento sobre os solos do QF, com ênfase nos teores de ferro, por meio de dois capítulos (artigos).

O primeiro propõe a construção de modelos preditivos, descritos como “funções de pedotransferência” para estimar os teores de ferro em solos do QF, com o objetivo de reduzir o uso de reagentes químicos potencialmente tóxicos, que são convencionalmente utilizados em análises de laboratório. Já o segundo, apresenta mapas de distribuição espacial dos teores de ferro construídos por diferentes modelos, que podem constituir-se como uma ferramenta essenciais para a construção de dados e informações imprescindíveis ao planejamento, preservação, conservação e sustentabilidade do uso das áreas ferríferas do QF.

REFERÊNCIAS

- Araújo, C. F., Oliveira-Júnior, R. C., & Beldini, T. P. (2024). Multiscalar Geomorphometric Generalization to Delineate Soil Textural Patterns on Amazon Watersheds Landscapes. In *Pedometrics in Brazil* (pp. 15–29). https://doi.org/10.1007/978-3-031-64579-2_2
- Assis, L. M., Francelino, M. R., Daher, M., Fernandes-Filho, E. I., Veloso, G. V., Gomes, L. C., & Schaefer, C. E. G. R. (2021). Modeling regolith thickness in iron formations using machine learning techniques. *Catena*, 207. <https://doi.org/10.1016/j.catena.2021.105629>
- Barbosa, M. S. C., & Oliveira, L. A. S. (2020). Geofísica do Quadrilátero Ferrífero). In P. A. C. Tarso, I. Endo, & A. L. Gandini (Eds.), *Quadrilátero Ferrífero: Avanços do conhecimento nos últimos 50 anos*. (pp. 114–141).
- Bastos, B. P., Pinheiro, H. S. K., Carvalho-Júnior, W., & Anjos, L. H. C. (2024). Aerogeophysical Data to Modeling Soil Properties: A Study Case in Bom Jardim—RJ. In *Pedometrics in Brazil* (pp. 159–170). https://doi.org/10.1007/978-3-031-64579-2_11
- Beato, D. A. C.; Monsore, A. L. M.; Bertachinni, A. C. Projeto APA Sul RMBH: Hidrogeologia e mapas hidrogeológicos, Belo Horizonte: SEMAD/CPRM, 2005.
- Carmo, F. F. do. Importância Ambiental e Estado de Conservação dos Ecossistemas de Cangas no Quadrilátero Ferrífero e Proposta de Áreas-Alvo para Investigação e Proteção da Biodiversidade em Minas Gerais. Dissertação (Mestrado). Universidade Federal de Minas Gerais, Belo Horizonte, 2010
- Carmo, F. F., & Kamino, L. H. Y. (2015). Introdução de Geossistemas Ferruginosos no Brasil. In *Geossistemas ferruginosos no Brasil*. 3i editora.

- Carvalho-Filho, A. Solos e ambientes do Quadrilátero Ferrífero (MG) e aptidão silvicultural dos Tabuleiros Costeiros. Tese (Doutorado). Universidade Federal de Lavras, Lavras, 2008.
- Carvalho-Filho, A., Curi, N., Marques, J. J. G. S. M., Shinzato, E., de Jesus, E. A., Freitas, D. A. F., & Massahud, R. T. L. R. (2011). Óxidos de manganês em solos do quadrilátero ferrífero (MG). *Revista Brasileira de Ciência Do Solo*, 35, 793–804. <https://doi.org/10.1590/S0100-06832011000300015>
- Carvalho-Filho, A., Curi, N., & Shinzato, E. (2010). Relações solo-paisagem no Quadrilátero Ferrífero em Minas Gerais. In *Pesq. agropec. bras* (Issue 8). <https://doi.org/10.1590/S0100-204X2010000800017>
- Carvalho-Júnior, W., Chagas, C. da S., Calderano-Filho, B., & Bhering, S. B. (2013). *Funções de pedotransferência para estimativa da densidade dos solos de áreas tropicais montanhosas*. https://www.researchgate.net/publication/279854424_Funcoes_de_pedotransferencia_para_estimativa_da_densidade_dos_solos_de_areas_tropicais_montanhosas
- Carvalho-Júnior, W., Pereira, N. R., Bhering, S. B., Calderano-Filho, B., Chagas, C. S., Pinheiro, H. S. K., Pereira, J. R., Lopes, C. H. L., & Leal, R. B. (2024). Exploratory Analysis from Harmonized Legacy Soil Data to Support Digital Soil Mapping in Brazilian Midwest. In *Pedometrics in Brazil* (pp. 115–127). https://doi.org/10.1007/978-3-031-64579-2_9
- Coelho, M. R.; Vasques, G. de M.; Tassinari, D.; Souza, Z. R. de.; Oliveira, A. P. de.; Moreira, F. M. de. S. Solos do Quadrilátero Ferrífero sob diferentes coberturas vegetais e materiais de origem. Boletim de pesquisa e desenvolvimento, Embrapa Solos, Rio de Janeiro, 2017.
- Costa, S. A. D. da. Caracterização química, física, mineralógica e classificação de solos ricos em ferro do Quadrilátero Ferrífero. Tese (Doutorado). Universidade Federal de Viçosa, Viçosa, 2003.
- Ceddia, M. B., Rodrigues, H. M., Souza-Ferreira, A. C., Costa, E. M., Pinheiro, É. F. M., & Fernandes, D. A. C. (2024). Soil Organic Carbon Stock Estimation Using Legacy Data: A Case Study of North Fluminense Region—BR. In *Pedometrics in Brazil* (pp. 129–158). https://doi.org/10.1007/978-3-031-64579-2_10
- Davis, E. G.; Pinto, E. J. de. A.; Pinto, M. C. F. Projeto APA Sul RMBH: Estudos do meio físico, Hidrologia, Belo Horizonte: SEMAD/CPRM, 2005.
- Demattê, J. A. M., Dotto, A. C., Paiva, A. F. S., Sato, M. V., Dalmolin, R. S. D., Araújo, M. S. B., Silva, E. B., Nanni, M. R., ten Caten, A., Noronha, N. C., Lacerda, M. P. C., Araújo-Filho, J. C., Rizzo, R., Bellinaso, H., Francelino, M. R., Schaefer, C. E. G. R., Vicente, L. E., Santos, U. J., Sampaio, E. V. S. B., ... Couto, H. T. Z. (2019). The Brazilian Soil Spectral Library (BSSL): A general view, application and challenges. *Geoderma*, 354. <https://doi.org/10.1016/j.geoderma.2019.05.043>
- Lana, C. E. (2015). A Geomorfologia do Quadrilátero Ferrífero. In P. A. C. Tarso, I. Endo, & A. L. Gandini (Eds.), *Quadrilátero Ferrífero: Avanços do conhecimento nos últimos 50 anos*. (pp 194 – 215)
- Ma, Y., Minasny, B., Malone, B. P., & Mcbratney, A. B. (2019). Pedology and digital soil mapping (DSM). *European Journal of Soil Science*, 70(2), 216–235. <https://doi.org/10.1111/ejss.12790>
- Madeira, J. A., Martins, F. D., Ribeiro, K. T., & Carvalho, A. S. (2015). Geossistemas ferruginosos e áreas protegidas. In F. F. Carmo & L. H. Y. Kamino (Eds.), *Geossistemas ferruginosos no Brasil* (pp 521 – 546)
- McBratney, A. B., Mendonça-Santos, M. L., & Minasny, B. (2003). On digital soil mapping. *Geoderma*, 117(1–2), 3–52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)

- Medina, A. I. de. M.; Dantas, M. E.; Saadi, A. Projeto APA Sul RMBH: Geomorfologia, e mapas geomorfológicos, Belo Horizonte: SEMAD/CPRM, 2005.
- Mondejar, J. P., & Tongco, A. F. (2019). Estimating topsoil texture fractions by digital soil mapping - A response to the long outdated soil map in the Philippines. *Sustainable Environment Research*, 1(1). <https://doi.org/10.1186/s42834-019-0032-5>
- Oliveira, G. da. S. R.; Jacques, P. D.; Shinzato, E. Projeto APA Sul RMBH: Uso e cobertura da terra, mapas de cobertura e uso da terra, Belo Horizonte: SEMAD/CPRM, 2005.
- Reddy, N. N., & Das, B. S. (2023). Digital soil mapping of key secondary soil properties using pedotransfer functions and Indian legacy soil data. *Geoderma*, 429. <https://doi.org/10.1016/j.geoderma.2022.116265>
- Resende, M., Ker, J. C., de Rezende, S. B., Poggere, G., Zimmer, J. B., Mancini, M., & Curi, N. (2023). *Roteiro pedológico de Minas Gerais: Conhecendo e interpretando os principais solos de diferentes ambientes*. https://www.researchgate.net/publication/370124938_ROTEIRO_PEDOLOGICO_DE_MINAS_GERAIS_Conhecendo_e_interpretando_os_principais_solos_de_diferentes_a_mmbientes
- Saadi, A. Ensaio sobre a morfotectônica de Minas Gerais. Instituto de Geociências, UFMG Belo Horizonte, Tese de Livre Docência, 285 p. 1991.
- Salgado, A. A. R., & Carmo, F. F. (2015). ‘Quadrilátero Ferrífero’: A Beautiful and Neglected Landscape Between the Gold and Iron Ore Reservoirs. In *Landscapes and Landforms of Brazil* (pp. 319–330). https://doi.org/10.1007/978-94-017-8023-0_29
- Schaefer, C. E. G. R., Cândido, H. C., Corrêa, G. R., Pereira, A., Nunes, J. A., Souza, O. F., Marins, A., Fernandes-Filho, E. I., & Ker, J. C. (2015). Solos desenvolvidos sobre canga ferruginosa no Brasil: Uma revisão crítica e papel ecológico de termiteiros. In F. F. Carmo & L. H. Y. Kamino (Eds.), *Geossistemas ferruginosos no Brasil* (pp. 77–102).
- Shinzato, E.; Carvalho-Filho, A. Projeto APA Sul RMBH: Estudos do meio físico, Pedologia, Belo Horizonte: CPRM/EMBRAPA/SEMAD, 2005.
- Silva, S. L. da.; Monteiro, E. A.; Baltazar, O. F.; Zucchetti, M. Projeto APA Sul RMBH: Geologia e mapas geológicos, Belo Horizonte: SEMAD/CPRM, 2005.
- Tarso, P. A. C., Endo, I., & Gandini, A. L. (2020). In P. A. C. Tarso, I. Endo, & A. L. Gandini (Eds.), *Quadrilátero Ferrífero: Avanços do conhecimento nos últimos 50 anos*. (pp 9 – 15)
- ten Caten, A., Dalmolin, R. S. D., Mendonça-Santos, M. L., & Giasson, E. (2012). *Mapeamento digital de classes de solos: características da abordagem brasileira*. <https://doi.org/10.1590/S0103-84782012001100013>
- Wadoux, A. M. J. C., Minasny, B., & McBratney, A. B. (2020). Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Science Reviews*, 210, 103359. <https://doi.org/10.1016/j.earscirev.2020.103359>
- Wadoux, A. M. J. C., Samuel-Rosa, A., Poggio, L., & Mulder, V. L. (2020). A note on knowledge discovery and machine learning in digital soil mapping. *European Journal of Soil Science*, 71(2), 133–136. <https://doi.org/10.1111/ejss.12909>

CAPÍTULO 1

Machine learning como ferramenta para a construção de modelos preditivos (“funções de pedotransferência”) e suas potencialidades para solos ferríferos.

KRAUSE, M. B.¹; PEREIRA, L. F.¹; SIQUEIRA, R. G.¹; MOQUEDACE, C. M. M. d. S.¹; FRANCELINO, M. R.¹; CARVALHO-JÚNIOR, W.²; FERNANDES-FILHO, E. I.¹.

1 – Departamento de Solos, Universidade Federal de Viçosa, Viçosa – MG.

2 – Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA - Solos), Rio de Janeiro – RJ.

RESUMO

O uso de ferramentas de machine learning (ML) na ciência do solo cresceu nos últimos anos e tem impulsionado cada vez mais a Pedometria. Aliando o avanço tecnológico ao conhecimento pedológico convencional, tornou-se comum a criação de funções de pedotransferência (FPT) para a determinação de diferentes atributos ou propriedades dos solos, que em campo, são difíceis de mensurar continuamente. Nesse sentido, objetivou-se com esse artigo a criação de modelos preditivos, descritos como FPT, para estimar os teores de ferro em solos do Quadrilátero Ferrífero (QF). Para isso, utilizou-se amostras do Banco de Solos de Minas Gerais e, para completar a base de dados, realizaram-se coletas na profundidade de 0 a 20 m. No total, foram analisadas 165 amostras. Para a modelagem, empregaram-se seis algoritmos de ML. A seleção das covariáveis predictoras ocorreu por meio da eliminação por correlação ($|\rho| < 0,95$) e da importância das covariáveis, utilizando o método *Recursive Feature Elimination*. Todo o processo de seleção, treinamento, validação, teste e predição foi repetido 100 vezes. Destaca-se que uma diferenciação metodológica impactou diretamente a performance dos modelos. As FPT foram construídas de duas formas: a primeira utilizando informações obtidas a partir de um sensor proximal que mediu a susceptibilidade magnética (SM) das amostras; e a segunda sem a SM, devido à alta correlação dessa variável com os teores de ferro. Os modelos Random Forest (RF) e Gradient Boosting Machine (GBM) foram os mais consistentes. O RF apresentou o melhor desempenho quando a SM foi incluída, enquanto o GBM foi superior na ausência dessa variável. Com a SM, os valores de MAE, LCCC e RMSE para o RF foram, respectivamente 23,85; 0,78 e 33,4. Sem a SM, os valores para o GBM foram 37,08; 0,45; 52,28. A presença da SM melhorou a performance dos modelos em aproximadamente 30%, além de aumentar sua capacidade de detecção de padrões nos dados. Quanto às variáveis frequentemente selecionadas, observou-se uma predominância de covariáveis pedológicas, como a SM e a textura do solo. Destacaram-se as covariáveis bioclimáticas, como a precipitação e temperatura, além de variáveis morfométricas relacionadas à altitude e as covariáveis gamaespectrométricas. De modo geral, pode-se afirmar que os modelos conseguiram captar as relações existentes entre os dados e que as FPT construídas podem contribuir para ampliar o conhecimento sobre os solos do QF.

Palavras-Chave: Dados legados; Atributos pedológicos; Random Forest; Gradient Boosting Machine.

ABSTRACT

The use of machine learning (ML) tools in soil science has grown in recent years and has increasingly driven Pedometrics. Combining technological advancements with conventional pedological knowledge, the creation of pedotransfer functions (PTFs) has become common for determining various soil attributes or properties that are difficult to measure continuously in the field. In this context, this study aimed to develop predictive models, described as PTFs, to estimate iron contents in the soils of the Quadrilátero Ferrífero (QF). To achieve this, soil samples from the Minas Gerais Soil Bank were used, and additional samples were collected at depths ranging from 0 to 20 m to complement the dataset. In total, 165 samples were analyzed. For modeling, six ML algorithms were employed. The selection of predictor covariates was performed through correlation-based elimination ($|\rho| < 0.95$ | $|\rho| < 0.95$) and covariate importance using the Recursive Feature Elimination method. The entire process of selection, training, validation, testing, and prediction was repeated 100 times. It is noteworthy that a methodological differentiation directly impacted the models' performance. The PTFs were developed in two ways: the first using information obtained from a proximal sensor that measured the magnetic susceptibility (MS) of the samples, and the second without MS, due to its high correlation with iron contents. The *Random Forest* (RF) and *Gradient Boosting Machine* (GBM) models were the most consistent. RF showed the best performance when MS was included, while GBM performed better in the absence of this variable. With MS, the MAE, LCCC, and RMSE values for RF were 23.85, 0.78, and 33.4, respectively. Without MS, the values for GBM were 37.08, 0.45, and 52.28. The presence of MS improved model performance by approximately 30%, in addition to enhancing their ability to detect patterns in the data. Regarding the frequently selected variables, a predominance of pedological covariates, such as MS and soil texture, was observed. Bioclimatic covariates, such as precipitation and temperature, also stood out, along with morphometric variables related to altitude and gamma spectrometry-derived covariates. Overall, it can be stated that the models successfully captured the relationships within the data and that the developed PTFs can contribute to expanding knowledge about the soils of the QF.

Keywords: Legacy data; Pedological attributes, Random Forest, Gradient Boosting Machine.

1. INTRODUÇÃO

O Quadrilátero Ferrífero (QF) é a principal província mineral do sudeste brasileiro (Santos et al., 2004; Costa et al., 2014; Diniz et al., 2014) e a mais antiga área de mineração do Brasil (Carvalho-Filho et al., 2010; Souza et al., 2015; Schaefer et al., 2016; Barbosa & Oliveira, 2020). Sua diversidade de condições ambientais favoreceu o desenvolvimento de estudos no ramo das geociências, embora pesquisas dedicadas aos solos ainda sejam escassas (Carvalho-Filho et al., 2010; Assis et al., 2021; Fernandes-Filho et al., 2022).

Os solos são recursos essenciais e não renováveis, que desempenham papel central na provisão e preservação de serviços ambientais (Pereira et al., 2024). Segundo Schaefer et al. (2015), os estudos pedológicos podem contribuir para o desenvolvimento de técnicas de recuperação de ambientes degradados, especialmente em áreas mineradas.

Nos últimos anos, técnicas de Machine Learning (ML) e o Mapeamento Digital de Solos (MDS) têm impulsionado avanços na ciência do solo (Demattê & Terra, 2014; Padarian et al., 2020; Mello et al., 2022a; 2022b). Essas abordagens permitem compreender propriedades do solo em áreas com pouca ou nenhuma informação pedológica, como o QF (van Leeuwen et al., 2024). Minasny & Hartemink (2011) destacam que o MDS utiliza informações espaciais e covariáveis ambientais para modelar características dos solos por meio de relações matemáticas lineares e não-lineares.

A predição de atributos do solo tem sido aprimorada pelo uso de Funções de Pedotransferência (FPT), que são modelos preditivos usados para estimar propriedades complexas dos solos tropicais (da Silva et al., 2023; dos Santos et al., 2024). As FPT permitem a substituição de métodos laboratoriais tradicionais, muitas vezes poluentes e geradores de resíduos não sustentáveis, além de possibilitar análises mais rápidas e acessíveis (Minasny & Hartemink, 2011; Carvalho-Júnior et al., 2013; da Silva et al., 2023).

Em ambientes tropicais, as FPT, são amplamente aplicadas para estimar propriedades hidrofísicas e estoques de carbono, ainda que os métodos estatísticos convencionais tenham predominado historicamente sobre os modelos de ML (da Silva et al., 2023). No entanto, Witten & Frank (2002) ressaltam que as técnicas de ML são especialmente adequadas para modelar funções não-lineares, como as que caracterizaram a interação entre atributos do solo e fatores de formação (da Silva et al. 2023).

Os solos tropicais, em geral, são caracterizados por grande profundidade, intemperismo avançado, predominância de mineralogia oxídica e caulínica (Ferreira et al., 1999; Resende et al., 2023). Nesse contexto, os óxidos de ferro desempenham um papel crucial na fração argila e influenciam na estabilidade dos agregados, na retenção de água, na adsorção de fósforo e de

metais-traços e na formação de complexos com a matéria orgânica (Donagemma et al., 2003; Costa et al., 2014; Camêlo et al., 2018; Almeida et al., 2020).

Além das propriedades estruturais, os óxidos de ferro afetam atributos como a cor do solo e a Susceptibilidade Magnética (SM) e por isso, Camêlo et al. (2018) enfatizam a necessidade de entender melhor os solos oxídicos com altos teores de ferro. Barbosa et al. (2021) definem a SM como uma medida quantitativa do grau de magnetização sob aplicação de um campo magnético. Segundo Mello et al. (2020; 2023b), a SM é influenciada pelo material de origem, pela mineralogia e pela geoquímica de minerais ferromagnéticos do solo.

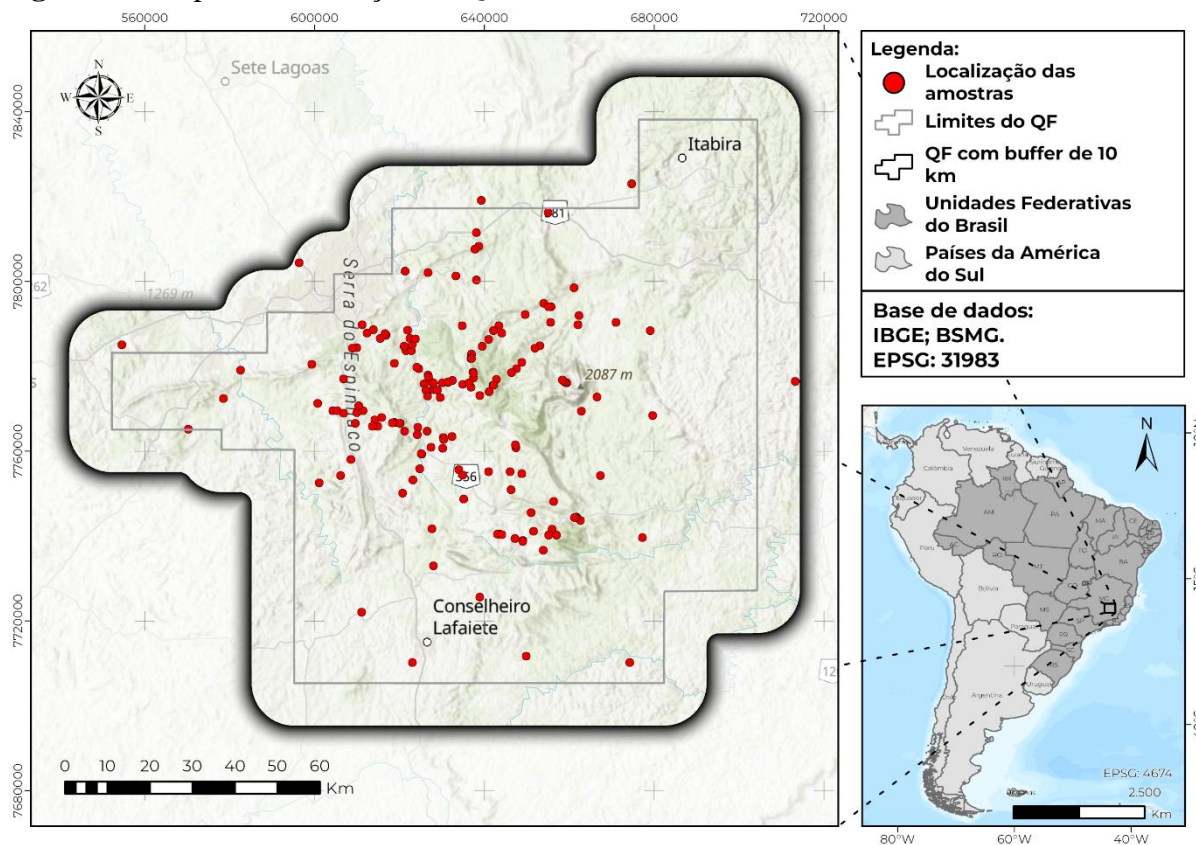
Dessa forma, a integração de modelos de ML com dados legados, observações de campo e covariáveis ambientais pode fornecer informações estratégicas que ampliam o entendimento sobre os solos do QF (Assis et al., 2021). O presente estudo tem como objetivo desenvolver modelos preditivos definidos como FPT para estimar os teores de Ferro em solos do QF. Adicionalmente, busca-se avaliar diferentes modelos de ML e selecionar aqueles que melhores compreendem as relações do Ferro com as covariáveis e analisar as principais variáveis que influenciam a presença e determinação dos teores de Ferro. Essa abordagem visa substituir métodos laboratoriais convencionais poluentes ao utilizar dados gratuitos e públicos.

2. MATERIAIS E MÉTODOS

2.1 Caracterização da área de estudo

O QF é uma região de destaque entre as terras altas tropicais no cenário geológico global, devido à sua diversidade de rochas arqueanas e proterozoicas, que abrigam grandes reservas de ouro, ferro, manganês, alumínio e rochas industriais em uma estrutura geológica aparentemente estável (Dorr, 1969; Souza et al., 2015; Varajão et al., 2020; Fernandes-Filho et al., 2022). Com altitudes que variam de 650 até 2000 metros e cobrem uma área de aproximadamente 7000 km², localiza-se na região centro-sudeste do estado de Minas Gerais, sudeste brasileiro (Santos & Varajão, 2003; Santos et al., 2004; Carvalho-Filho et al., 2011) (Figura 1).

Figura 1 – Mapa de localização do QF e dos locais de coleta das amostras.



Fonte: Elaborado pelo autor.

Os contornos do QF correspondem ao alinhamento litoestrutural e a processos de erosão diferencial. As feições pedogeomorfológicas, estão intimamente ligadas ao controle estrutural (Carvalho-Filho, 2008), as quais apresenta três principais unidades principais: as encostas elevadas, que estão presentes em altitudes acima de 1500 metros; as encostas das colinas, correspondentes às unidades intermediárias que variam entre 800 e 1400 metros e; por fim, as depressões ou áreas de baixo declive, que variam entre 1200 e 1400 metros (Santos & Varajão, 2003).

O clima da região é caracterizado por chuvas intensas concentradas entre os meses de outubro a março, seguida por uma estação seca bem definida de setembro a abril. A precipitação média anual varia entre 950 e 1550 mm, enquanto a temperatura média anual oscila entre 18°C e 22°C. O clima, de acordo com a classificação de Köppen é o Cwa, marcado por inverno seco e verão chuvoso (Diniz et al., 2014; Carvalho-Filho et al., 2015).

A vegetação do QF reflete as variações de condições ambientais em cada compartimento da paisagem. Sendo assim, é típica de uma zona ecotonal com transições entre Cerrado, Floresta Atlântica, Campos Rupestres Ferruginosos e/ou Quartzíticos (Jacobi et al., 2007; Schaefer et al., 2016; Coelho et al., 2017).

2.2 Bases de dados e amostragem

Utilizou-se como base de dados inicial, 99 amostras legadas do Banco de Solos de Minas Gerais (BSMG) (Souza et al., 2015). Os critérios para escolha seguiram os preceitos: amostras georreferenciadas; localizadas em um raio de dez (10) quilômetros de distância dos limites do QF; com existência de análises de textura, do pH determinado em água e a presença da determinação dos teores de Ferro. (figura 1).

Ao espacializar os locais de coletas do BSMG, identificou-se uma lacuna representativa na parte central do QF, onde haviam poucas amostras. Para mitigar essa deficiência, foi utilizado um método de design inteligente de amostragem, com o objetivo de reduzir a subjetividade na escolha dos locais representativos e garantir uma cobertura adequada em quantidade e qualidade, permitindo a generalização do modelo (Moquedace et al., 2024). O método escolhido foi o Conditioned Latin Hypercube Sampling (cLHS) (Sena et al., 2021).

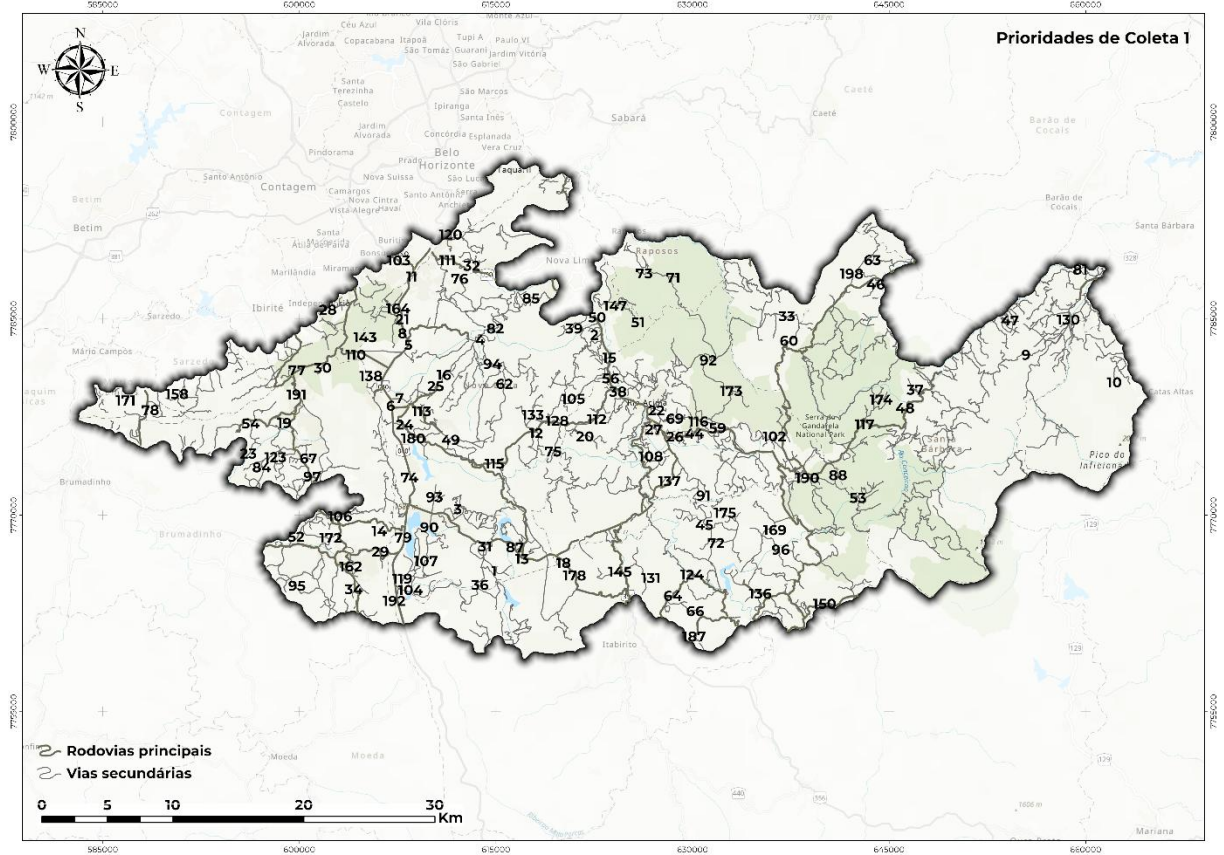
Esse método realiza uma separação aleatória estratificada, gerando uma malha amostral capaz de representar a variabilidade de covariáveis ambientais utilizadas (Tabela 1) em modelos de predição (McBratney et al., 2003; Minasny & McBratney, 2006; Vasconcelos, 2016). A escolha das variáveis ocorreu por meio de uma análise de componentes principais e selecionou-se as variáveis de maiores importância. Assim, definiu-se 200 pontos principais de coleta (Figura 2), com mais 5 pontos extras de reserva, para o caso de alguma intercorrência que não possibilitaria o acesso ao ponto principal, conforme preconizado por Sena et al. (2021).

Tabela 1 – Lista de covariáveis utilizadas para a geração dos locais de coleta via cLHS.

Covariáveis representativas
Mapeamento da litoestratigrafia
Calor anisotrópico diário
Contagem Total de Gamaespectroscopia
Contagem de Tório
Contagem de Urânio
Contagem de Potássio
Índice de óxidos de ferro
Magnetometria
Rodovias (“custo”)

Fonte: Elaborado pelo autor.

Figura 2 – Mapa dos locais de coleta na área central do QF.



Fonte: Elaborado pelo autor.

Introduziu-se a malha rodoviária como custo à análise e os locais muito distantes às rodovias penalizados e consequentemente descartados. Ressalta-se que todos os pontos extras eram representativos dos pontos principais. No total foram gerados 1200 pontos, com 1.000.000 de iterações. Em campo, coletou-se 66 amostras de 0 a 20 cm definidas pelo cLHS. As 134 amostras restantes não foram coletadas devido às restrições logísticas de acesso. Ao final, a base de dados utilizada foi de 165 amostras de solos com a litoestratigrafia do QF bem representada.

2.3 Análises laboratoriais

Para equiparar as análises realizadas nas amostras coletadas as dos dados legados do BSMG, seguiu-se a mesma metodologia analítica de determinação, conforme as diretrizes de Teixeira et al. (2017). Todas as amostras foram secas ao ar e passadas por uma peneira de aço inoxidável de 2,0 milímetros, determinando assim a Terra Fina Seca ao Ar (TFSA).

Em conformidade com o objetivo da pesquisa, para não utilizar covariáveis determinadas por meio de reagentes químicos, optou-se por mensurar apenas o pH do solo medido em água. Em relação à textura do solo, apesar da utilização de 0,1 M de NaOH para a dispersão das

frações do solo, preferiu-se determinar as concentrações devido a elevada agregação de argilas oxídicas e o método de determinação usado foi o da peneira-pipeta.

Para determinar a concentração dos teores de Ferro, triturou-se a TFSA e peneirou-se através de uma peneira de malha de 200 mesh. A metodologia seguiu o protocolo EPA 3051a (EPA et al, 2007) e as orientações descritas por Souza et al. (2015) para digestão ácida das amostras. Determinou-se a quantificação das concentrações a partir do espectrômetro de emissão óptica por plasma induzido acoplado (ICP-OES).

2.4 Susceptibilidade Magnética (SM)

Por meio do sensor proximal Terraplug KT10 (Figura 3^a), mensurou-se a SM de baixa frequência nas 165 amostras. O equipamento pode medir até 2 centímetros de profundidade da superfície do solo, com a precisão do equipamento de 10^{-6} no sistema internacional de unidades (“SI units”) que corresponde a expressão em $\text{m}^3 \text{kg}^{-1}$ de solo (Terraplug inc, 2018; Mello et al., 2022b).

Figura 3. A – Susceptibilímetro magnético Terraplug KT10. B – Realização da leitura, para evitar contaminação entre o sensor e amostra. Fonte: **3. A** – GEOSensor. **3. B** – Autores.

A**B**

A partir de testes preliminares, percebeu-se que a massa e o volume das amostras impactavam no valor de SM. Diante disso, adaptou-se o método de leitura que consistiu no uso de um filme plástico entre o sensor e a amostra. Desta forma, a medição seria coerente e principalmente evitaria a contaminação entre as amostras, conforme apresentado na figura 3B.

Seguindo as recomendações do fabricante (Terraplus inc, 2018), realizou-se a calibração do equipamento conforme Mello et al. (2020; 2022b; 2023b; 2023d), anteriormente a todas as análises. Para isso, determina-se inicialmente a oscilação da frequência externa e posteriormente é determinada a SM da amostra, quando o sensor é colocado em contato com a amostra. Esse processo é necessário porque o princípio da leitura ocorre com a diferença da SM do material em análise e do ambiente externo.

As leituras foram realizadas no modo “scanner” devido à melhor correlação com os valores de SM, além de providenciar uma quantificação rápida e precisa (Mello, et al., 2023a; 2023c). Durante a leitura, colocou-se o sensor em contato com a amostra durante alguns segundos, até o momento em que o valor da SM era estabilizado. Após estabilização, registrou-se o valor e realizou-se cinco leituras para cada amostra. Ao final, utilizou-se o valor médio das cinco leituras.

2.5 Covariáveis explicativas e preditoras

A seleção de covariáveis para a construção das FPT visou atender ao modelo SCORPAN (McBratney et al., 2003). Cada letra do acrônimo representa um atributo, propriedade ou fator de formação do solo (Jenny, 1941). Para o fator “S”, considerou-se os dados legados do BSMG e aqueles determinados em laboratório após as coletas de campo (Tabela 2). Além do mais, considerou-se 5 índices espectrais que se correlacionam com a presença de óxidos de Ferro no solo.

Tabela 2 – Covariáveis representativas do modelo SCORPAN selecionadas para a modelagem dos teores de Ferro nos solos do QF.

Covariáveis (descrição dos gráficos – referente a qual fator da equação SCORPAN)		
Susceptibilidade magnética (S)	Areia grossa (S)	Silte (S - S)
pH determinado em água (S)	Areia fina (S)	Argila (S)
Índ de minerais de argila (S)	Índ de ferro ferroso (S)	Índ de minerais ferrosos (S)
Índ de óxidos de ferro (S)	Índ de silicatos de ferro (S)	Temp média anual do ar (TMAA – C)
Temp média no quarto do ano mais quente (TMQAMQ – C)	Precip média anual (PMA – C)	Temp média no quarto do ano mais frio (TMQAMF– C)
Precip no mês mais úmido (PMMU – C)	Precip no mês mais seco (PMMS – C)	Sazonalidade da precipitação (C)

Covariáveis (descrição dos gráficos – referente a qual fator da equação SCORPAN)		
Precip no quarto do ano mais úmido (PQAMU – C)	Precip no quarto do ano mais seco (PQAMS – C)	Precip no quarto do ano mais quente (PQAMQ – C)
Precip no quarto do ano mais frio (PQAMF – C)	Variação média da temp diária (VMTD – C)	Razão entre a variação diária da temp (RVDT – C)
Temp média no quarto do ano mais seco (TMQAMS – C)	Temp máxima no mês mais quente (TMMMQ– C)	Temp mínima no mês mais frio (TMMMFMF – C)
Variação anual da temp (VAT – C)	Sazonalidade da temperatura (C)	Temp média no quarto do no mais úmido (TMQAMU – C)
Índice de clorofila verde (GCI - O)	Índice de diferença normalizada de vegetação verde (GNDVI - O)	Índice de diferença normalizada de umidade (NDMI - O)
Índice de diferença normalizada de vegetação (NDVI - O)	Índice de diferença normalizada de Água (NDWI - O)	Composição RGB (O)
Índice de vegetação ajustado ao solo (SAVI - O)	Índice resistente à atmosfera na região visível (VARI - O)	Área de captação (R)
Áreas de captações declividade (R)	Aspecto (R)	Índice de convergência (R)
Seção transversal de curvatura (R)	Linha de fluxo da curvatura (R)	Curvatura geral (R)
Curvatura mínima (R)	Curvatura longitudinal (R)	Curvatura máxima (R)
Tangência de curvatura (R)	Perfil de curvatura (R)	Altura efetiva do fluxo de ar (R)
Curvatura total (R)	Calor anisotrópico diário (R)	Índice de balanço de massa (R)
Altitude (R)	Sombreamento (R)	Planicidade do vale (MRVBF - R)
Posição de meia encosta (R)	Área superficial real (R)	Nível de cume (R)
Planicidade do topo (MRRTF - R)	Declividade graus (R)	Altura de encosta (R)
Altura normalizada (R)	Potencial de insolação (R)	Proporção solar (R)
Fator de visão do céu (R)	Pontos específicos de superfície (R)	Radiação total (R)
Insolação solar (R)	Textura superficial do terreno (R)	Índice de rugosidade do terreno (R)
Tempo de nascer solar (R)	Tempo de pôr do sol (R)	Convexidade da superfície do terreno (R)
Altura padronizada (R)	Vales (R)	Abertura dos topos (R)
Índice de posição topográfica (R)	Índice laterítico 1 (P)	Índice laterítico 2 (P)
Índice máfico 1 (P)	Índice máfico 2 (P)	Índice máfico 3 (P)
Índice máfico 4 (P)	Parâmetro F (P)	Razão eU/eTh (P)
Razão eTh/ ⁴⁰ K (P)	Razão eU/ ⁴⁰ K (P)	Mapa da litoestratigrafia (P)
Mapa das “pedorochas” (P)		

Fonte: Elaborado pelo autor.

O fator “C” refere-se ao clima ou às propriedades climáticas e utilizou-se as 19 variáveis bioclimáticas provenientes do WorldClim (Hijmans et al., 2005). De acordo com Siqueira et al. (2023), as variáveis bioclimáticas fornecem informações biológicas significativas, representando tendências anuais, sazonalidade e fatores ambientais limitantes.

Para representar o fator “O”, que é a atividade dos organismos, vegetação ou atividades humanas, derivou-se 7 índices espectrais de imagens Landsat-09 obtidas em março de 2023. Optou-se por usar as imagens Landsat-09 devido à resolução ser próxima a do MDE e a escolha da data ocorreu pela pouca presença de nuvens nas imagens. Para o fator “R”, que se trata do relevo e atributos da paisagem, derivou-se 58 covariáveis morfométricas do modelo digital de elevação FABDEM (Hawker et al., 2022), utilizando-se o pacote RSAGA (Brenning, 2008), presente no software R (R Core Team, 2024).

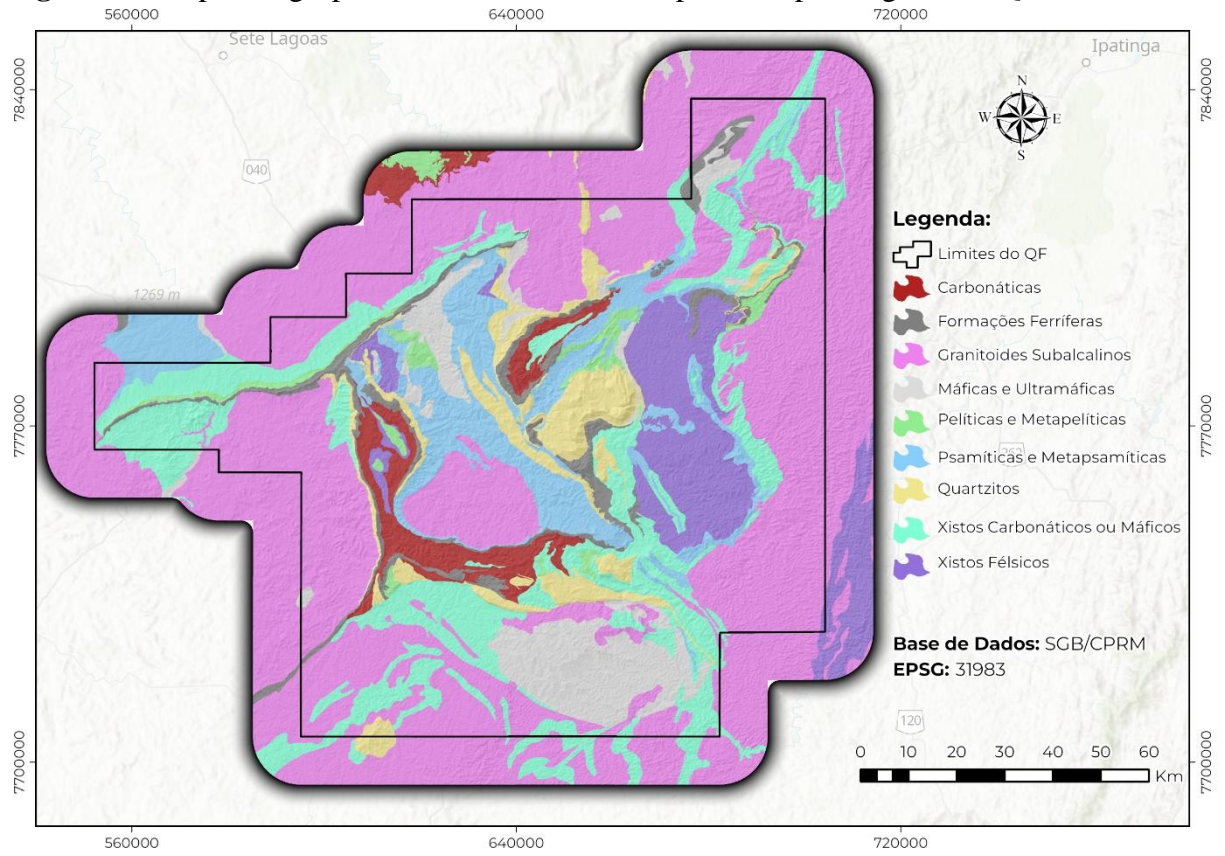
O fator “P” é associado ao material de origem ou à litologia. Para isso, realizou-se um agrupamento litológico em um mapa produzido por Brasil (2010) e com base em Resende et al. (2023) que separaram grupos de rochas para fins pedológicos (tabela 3 e figuras 4 e 5). Além disso, utilizou-se dados referentes à magnetometria e gamaespectroscopia e derivou-se índices geofísicos e razões entre Urânio (eU), Tório (eTh) e Potássio (^{40}K) para auxílio na separação litológica (tabela 6) (Gomes, 2021).

Tabela 3 – Agrupamentos de rochas para fins pedológicos.

Agrupamentos por litotipos (Brasil, 2010)	Agrupamentos por fins pedológicos (Resende et al., 2023)
Carbonáticas	Pedocalcárias
Formações Ferríferas	Pedoferruginosas
Granitoides Subalcalinos a Alcalinos	Pedogranitoides
Máficas e Ultramáficas	Pedomáficas
Pelíticas e Metapelíticas	Pedopelíticas
Psamíticas e Metapsamíticas com Componentes Finos	Pedopsefíticas
Quartzitos	Pedopsamíticas
Xistos com Componentes Carbonáticos ou Máficos	Pedopelíticas
Xistos com Componentes Félsicos	

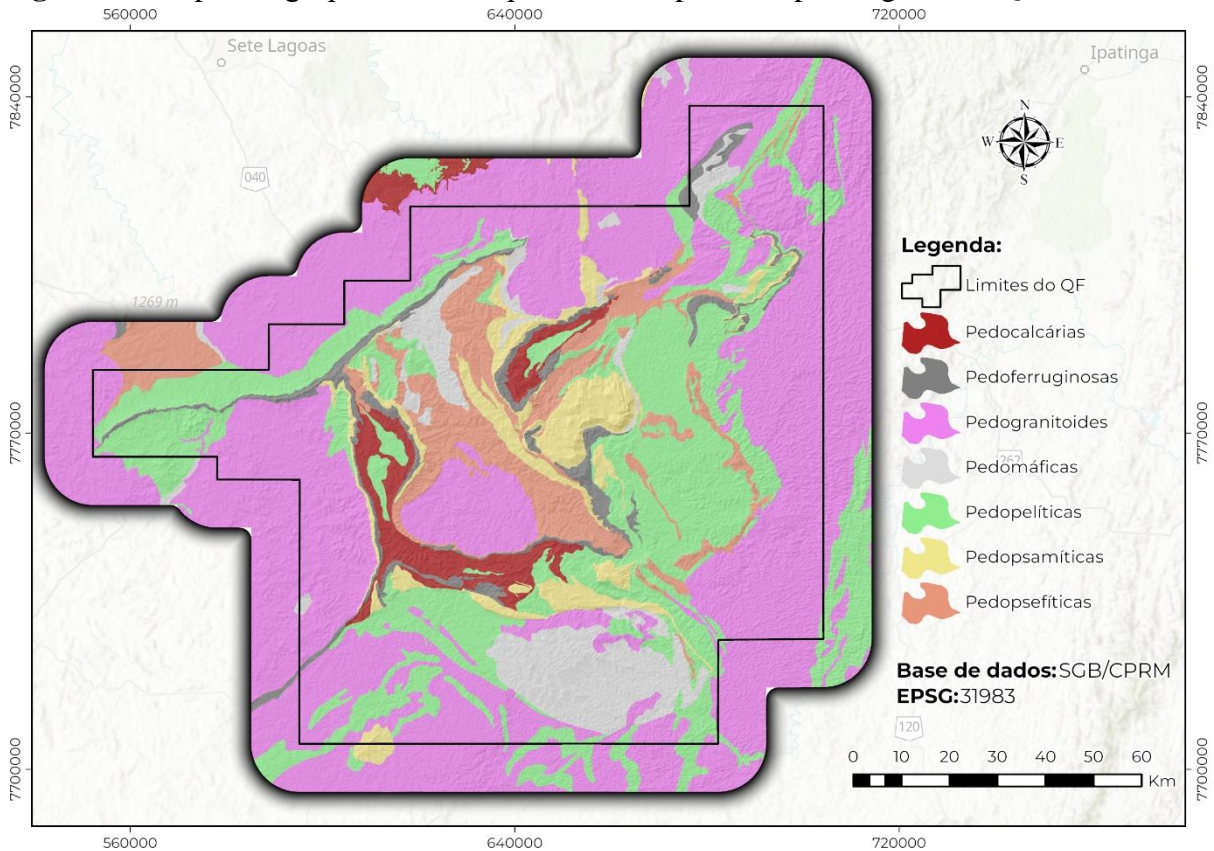
Fonte: Elaborado pelo autor.

Figura 4 – Mapa de agrupamento das litoestruturas para fins pedológicos do QF.



Fonte: Elaborado pelo autor.

Figura 5 – Mapa de agrupamento em “pedorochas” para fins pedológicos do QF.



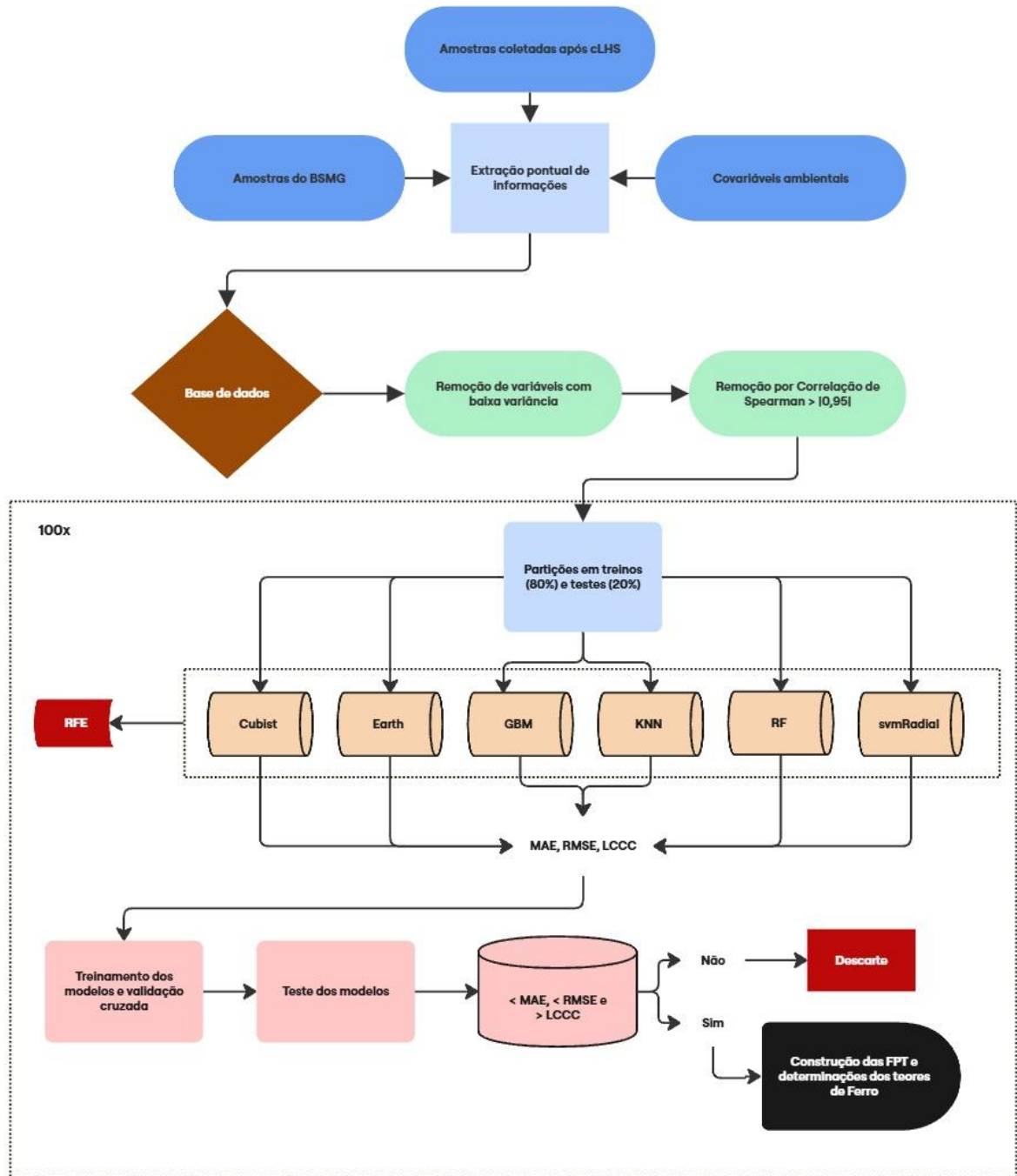
Fonte: Elaborado pelo autor.

O fator “A” é referente ao fator tempo (Jenny, 1941) e é possível que o modelo SCORPAN não apresente correlações significativas com solos policíclicos (McBratney et al., 2003). Por fim, o fator “N”, é a posição espacial de todas as covariáveis. Todos os processos de modelagens e predições foram realizados no software R (R Core Team, 2024), utilizando principalmente o pacote caret (Kuhn, 2020) e outros pacotes associados.

2.6 Seleção e importância de covariáveis preditoras

Com base nas localizações das coletas realizadas, extraiu-se, pontualmente, as informações de todas as covariáveis e considerou-se as covariáveis categóricas como fatores. A base de dados final consistiu em 117 covariáveis explicativas e 165 amostras. Com base nas metodologias descritas por Fernandes-Filho et al. (2024) e Mello et al. (2025) para a escolha das covariáveis mais importantes, redução de possíveis redundâncias, facilitar a interpretabilidade dos modelos e reduzir os custos computacionais, aplicou-se uma seleção sequencial seguindo três etapas. A sequência metodológica está descrita na figura 6.

Figura 6 – Fluxograma da metodologia desenvolvida.



Fonte: Elaborado pelo autor.

A primeira consistiu na remoção de covariáveis com baixa variância aplicando a função *nearZeroVar*, presente no pacote *caret* (Kuhn, 2020). A segunda ocorreu pela identificação de pares de covariáveis que apresentam valores de correlação de Spearman (ρ) maiores que 95% e uma das duas covariáveis é removida de acordo com a maior correlação média absoluta, esse processo é feito por meio da função *findcorrelation* do pacote *caret* (Kuhn, 2020). Por fim, a

terceira etapa consiste na aplicação do *Recursive Feature Elimination* (RFE) sobre o conjunto de covariáveis selecionadas após a primeira e segunda etapas.

O RFE realiza uma seleção do melhor subconjunto de preditores com base em sua importância para a variável alvo, por isso, ele é aplicado sobre o treinamento dos modelos (Pereira et al., 2024). Dessa forma, os modelos serão otimizados por subconjuntos de covariáveis ótimas, com ganho ou no mínimo sem perda significativa da capacidade para predição de dados (Gomes et al., 2019; Siqueira et al., 2023; Moquedace et al., 2024). Trabalhos como o de Gomes et al. (2019) e o de Mello et al. (2025) destacaram que o RFE considera diferentes algoritmos base e possui funções específicas para cada ajuste. Desta forma, utilizou-se para cada modelo de ML a sua “*Funcs*” específica.

2.7 Modelos de machine learning

Foram selecionados os seguintes algoritmos de ML: Random Forest (RF) (Wadoux et al., 2020), Cubist (Zeraatpisheh et al., 2019), Adaptive multivariate regression (Earth) (Friedman, 1991), Stochastic Gradient Boosting (GBM) (Friedman, 2001), Support Vector Machine with Radial Basis Linear (svmRadial) (Cortes & Vapnik, 1995) e o K-Nearest Neighbors (KNN) (Hechenbichler & Schliep, 2004).

Algoritmos de diferentes “famílias” têm características específicas para o seu processamento e sua otimização a partir de diferentes estruturas e formas distintas de tratamentos de dados (Khaledian & Miller, 2020). Por exemplo, RF, Cubist e GBM são algoritmos de árvores de regressão, o Earth é um modelo linear baseado em spline, já o svmRadial é um modelo baseado em aprendizagem estatística com princípios de margem máxima e ajustes de hiperplanos e, por fim, o KNN é baseado na similaridade dos dados (Siqueira et al., 2023; Fernandes-Filho et al., 2024; Moquedace et al., 2024; Pereira et al., 2024; Mello et al., 2025).

2.8 Treino, teste e modelagem

Para a otimização dos modelos, separou-se 80% dos dados para treinamento e 20% para teste e avaliação da performance dos modelos e utilizou-se a validação cruzada com 10 folds. A separação foi realizada pela função do pacote caret *createDataPartition* que executa uma amostragem aleatória em percentis para garantir que ocorra uma representação de toda a base de dados (Kuhn, 2020).

Para avaliação da performance, descrição da acurácia e seleção dos melhores modelos utilizou-se as métricas: o erro médio absoluto (MAE), a raiz quadrada do erro médio (RMSE)

e o Coeficiente de Concordância de Correlação de Lins (LCCC). A escolha do LCCC e não o R^2 é devido a Khaledian & Miller (2020) afirmarem que o LCCC deve ser considerado mais apropriado para esse tipo de avaliação, pois consegue captar melhor o “*bias*” em modelos de predição e possibilita a avaliação da precisão e acurácia das estimativas. As equações das métricas estão apresentadas na tabela 4.

Tabela 4 – Equações das métricas utilizadas para avaliar o desempenho dos modelos.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - x_i| \quad RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2}$$

$$LCCC = \frac{2\rho\sigma_{x_i}\sigma_{y_i}}{\sigma^2_{x_i} + \sigma^2_{y_i} + (\bar{x} - \bar{y})^2}$$

Fonte: Elaborado pelo autor.

Além disso, como uma validação adicional, aplicou-se o “MAE Null” e o “RMSE Null”, calculando as métricas para um modelo nulo com as observações médias como preditores (Siqueira et al., 2023; Fernandes-Filho et al., 2024; Moquedace et al., 2024; Pereira et al., 2024; Mello et al., 2025). O uso de modelos “Null” é uma estratégia eficiente para estabelecer limites e avaliar os modelos, já que permite a comparação dos modelos testados com um modelo simples.

Realizou-se todo o processo de seleção de covariáveis, separações em treinos e testes, construções dos modelos, predições e validações por 100 vezes, por meio de diferentes subconjuntos de treinos e testes. Por fim, para a comparação entre os modelos, considerou-se como base os valores médios e medianos das métricas de avaliação. Além disso, realizou-se o teste estatístico ANOVA, com $p < 0,05$, para verificar se haveria diferença significativa entre as médias dos valores das métricas dos modelos.

2.9 Análises de correlação

Adicionalmente às métricas de avaliação e para a interpretação das escolhas das variáveis pelos modelos e sua relação com a variável alvo, realizou-se análises estatísticas de correlações. Devido à ausência da distribuição normal dos dados, efetuou-se a análise de ρ nas variáveis mais importantes. Ademais, desenvolveu-se análises somente nas variáveis escolhidas pelo RFE. Todas as análises foram realizadas no software R.

2.10 Remoção de Susceptibilidade

Para evitar tendências nos modelos e considerando que as propriedades magnéticas dos solos refletem predominantemente solos com minerais ferromagnéticos como a Magnetita (Mt) e a Maghemita (Mh) (Valaee et al., 2016; Mello et al., 2020), optou-se pela realização de uma nova modelagem, onde removeu-se a variável SM. Realizou-se isso para buscar e compreender melhor a participação e a importância dessa variável explicativa na construção dos modelos quando esteve presente e entender melhor a importância de outras covariáveis.

Todo o processo posterior seguiu os mesmos passos destacados inicialmente. Realizou-se a remoção das covariáveis de baixa variância e posteriormente, houve a identificação dos pares de covariáveis com $\rho > 95\%$ e remoção de uma covariável do par que apresentou a maior correlação média absoluta. As covariáveis selecionadas na primeira e segunda etapa, após a fase de separação para treinamento dos modelos foram submetidas ao RFE.

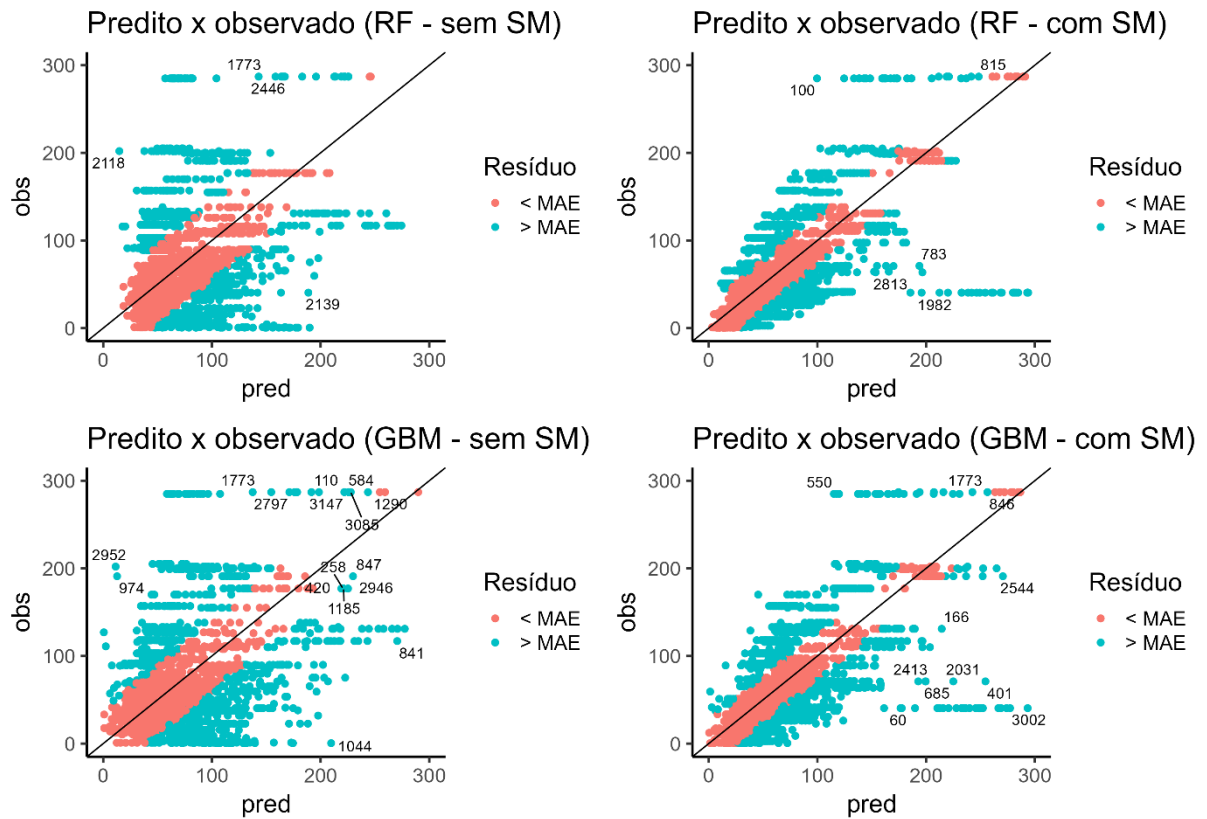
A separação manteve os 80% dos dados para treinamento e 20% para teste e avaliação da performance dos modelos e a validação cruzada ocorreu com 10 folds. O processo foi repetido por 100 vezes e por fim, avaliou-se os valores médios e medianos das métricas MAE, RMSE e LCCC, com validação pelos modelos nulos “MAE Null” e “RMSE Null”.

3. RESULTADOS

3.1 Amostragem e representatividade

As 165 amostras utilizadas permitiram a modelagem, predição e geração de FPT para determinação dos teores de ferro no solo de forma adequada. Os valores preditos apresentaram uma ampla variedade em relação aos valores de observados (figura 7). Porém, para nenhum dos modelos, houve uma extrapolação dos valores máximos e mínimos observados.

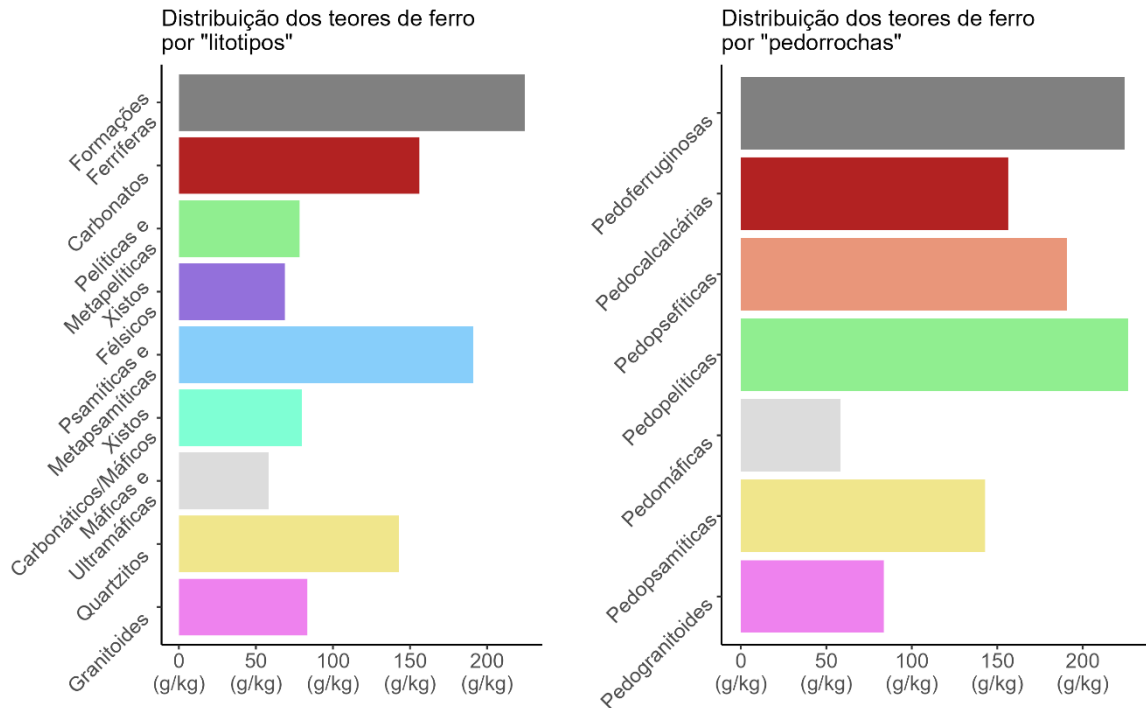
Figura 7 – Valores preditos x valores observados para os modelos RF e GBM.



Fonte: Elaborado pelo autor.

A boa capacidade de predição, pode estar relacionada as coletas realizadas. Houve uma ótima representação da variabilidade dos teores de ferro, de acordo com os agrupamentos para fins pedológicos (figura 8).

Figura 8 – Amplitude de valores determinados de ferro separados de acordo com a litoestratigrafia e a classificação de “pedorochas”.



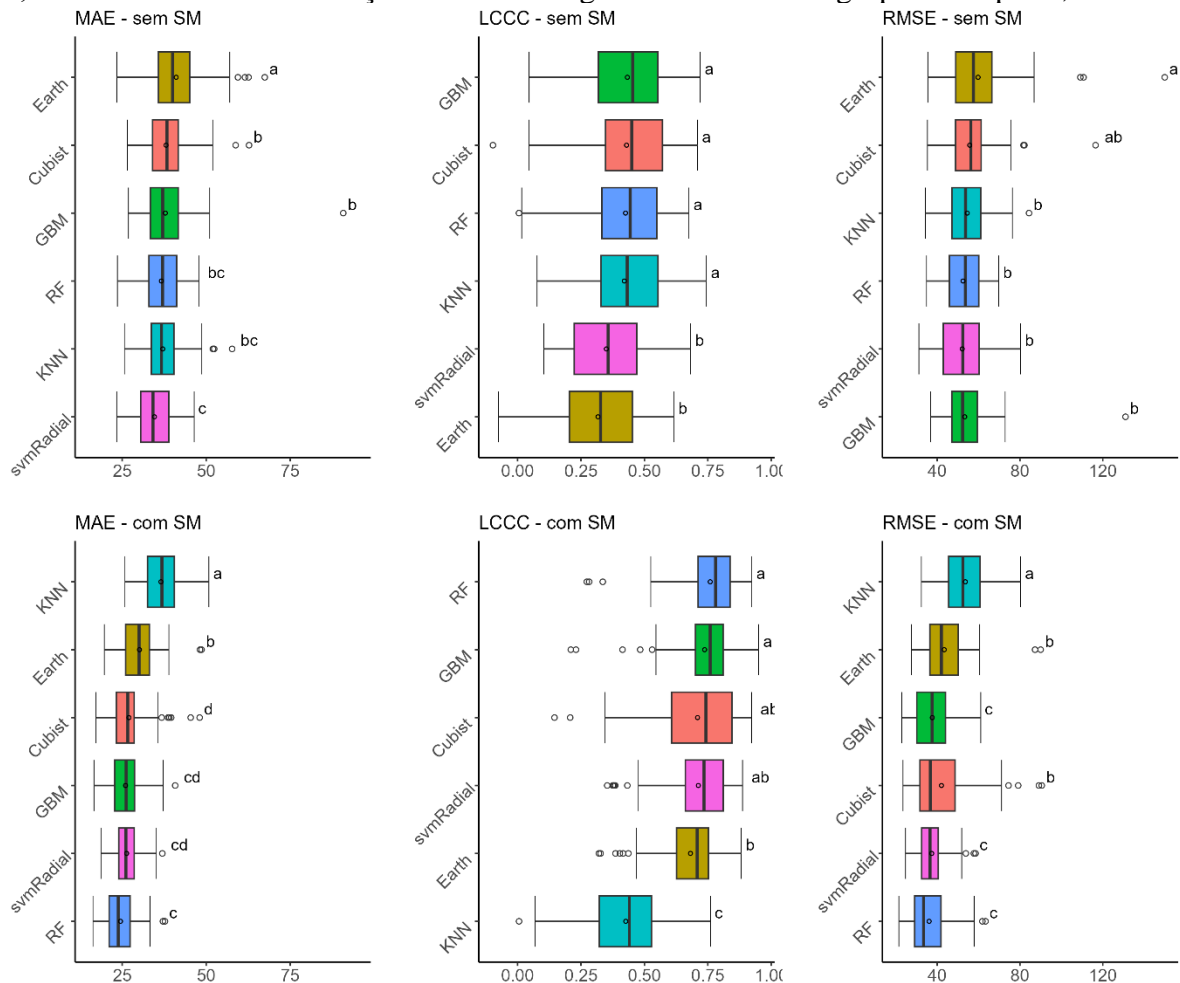
Fonte: Elaborado pelo autor.

De forma geral, observou-se que os maiores erros encontrados durante as predições, para todos os modelos, foram em dados legados, cujas as amostras tinham elevados teores de ferro e as SM eram baixas. Além disso, o comportamento inverso também causou dificuldade à detecção de padrões por parte dos modelos, onde a SM era elevada, mas os teores de ferro eram baixos. A explicação ocorre por dois motivos: pelo tamanho da partícula, consequentemente pela sua área de superfície específica e por características mineralógicas e ferromagnéticas do material amostrado. Além disso, entende-se que os teores de ferro podem estar subestimados nos dados legados.

3.2 Performance dos Modelos

Os algoritmos de ML utilizados variaram sua performance, principalmente pela presença ou ausência da SM (figura 9). De forma geral, todos os modelos apresentaram melhores performances quando a SM foi utilizada e o RF foi o melhor modelo, enquanto o KNN, o pior. Em relação às métricas, os valores de MAE variaram de 23,85 a 36,82, o LCCC de 0,78 a 0,44 e o RMSE de 33,40 a 52,41.

Figura 9 – Performance dos modelos utilizados. “sem SM” refere-se as métricas sem a susceptibilidade magnética e “com SM” são com a presença da susceptibilidade. As letras “a”, “b”, “c” e “d” indicam diferenças estatísticas significativas entre os grupos com $p < 0,05$.



Fonte: Elaborado pelo autor.

Com a ausência da SM, o svmRadial obteve o valor mais baixo de MAE com 34,18. Os modelos GBM e Cubist obtiveram os maiores valores de LCCC com 0,45 e o GBM obteve o menor RMSE de 52,28. O Earth foi o pior modelo com os maiores valores de MAE e RMSE de, respectivamente, 39,94 e 57,50. Ressalta que sem a SM, os modelos apresentaram menores diferenças estatísticas.

De forma geral, todos os modelos performaram melhor que os modelos nulos. Com exceção ao KNN, a presença da SM significou um ganho de aproximadamente 30% de performance (Tabela 5). Quando ausente a SM, os modelos tiveram performance pouco superiores a 10%. Destaca-se que o modelo Earth performou muito próximo ao modelo nulo, mostrando sua dificuldade para detectar padrões e avaliar os dados.

Tabela 5 – Ganhos de performance dos modelos comparados aos modelos nulos.

Modelos	Métricas			
	MAE		RMSE	
	com SM	sem SM	com SM	sem SM
Cubist	36,01%	7,92%	38,02%	4,92%
Earth	27,85%	4,05%	29,02%	2,82%
GBM	37,14%	10,92%	36,60%	11,64%
KNN	11,62%	11,81%	11,42%	9,28%
RF	42,75%	11,04%	43,55%	9,38%
svmRadial	37,33%	17,89%	38,28%	11,49%

Fonte: Elaborado pelo autor.

3.3 Seleção de modelos

Ao avaliar as métricas, MAE, LCCC e RMSE, percebeu-se que o RF e o GBM foram os melhores modelos com e sem a SM, respectivamente e que estatisticamente, performaram parecidos. Dessa forma, selecionou-se os dois algoritmos para os ajustes das FPT e observou-se que o tamanho dos subconjuntos selecionados pelo RFE, para cada modelo, variou largamente ao longo das 100 rodadas.

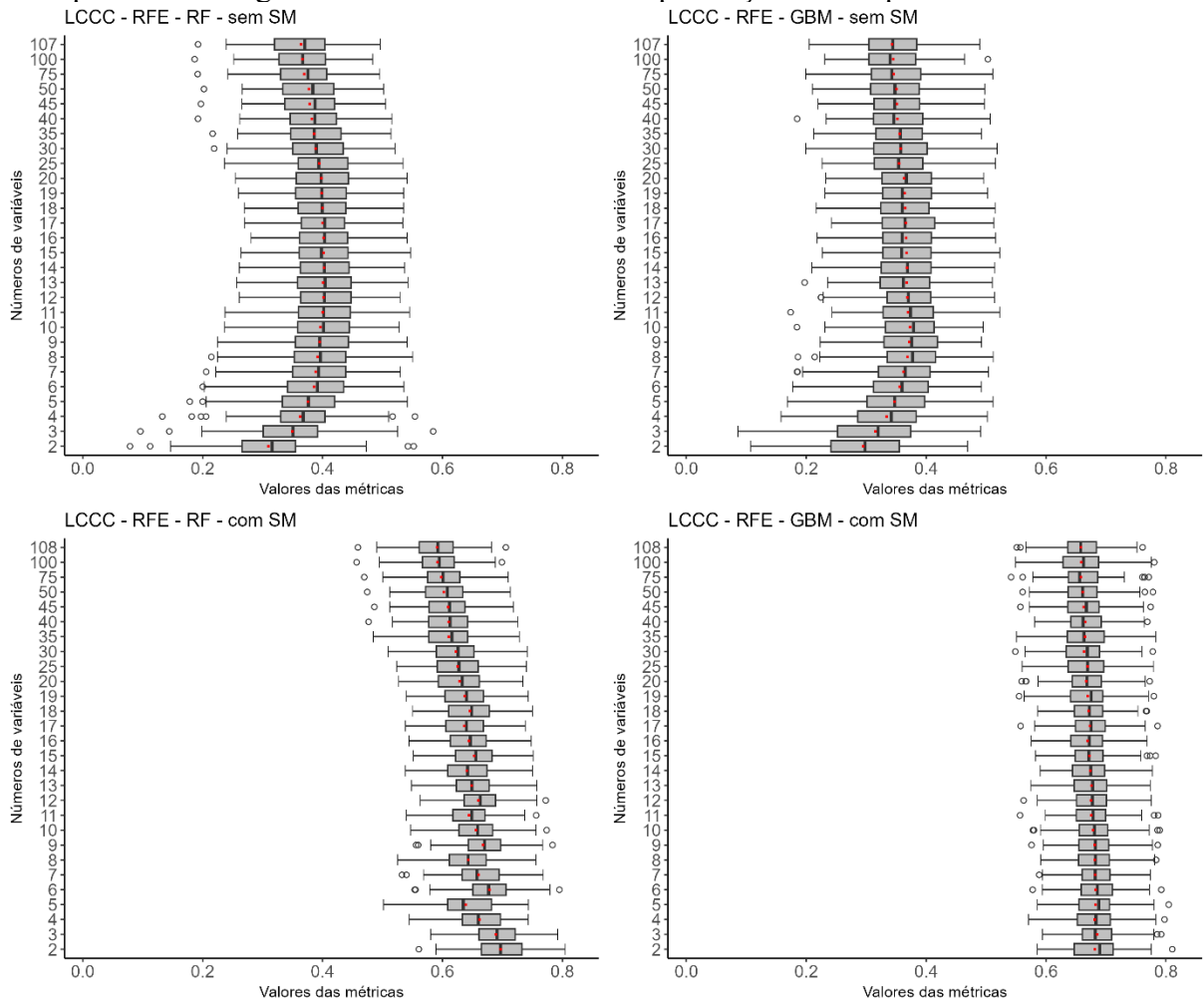
Com a seleção de variáveis, verificou-se que os modelos com a presença da SM apresentaram uma tendência de melhor performance, à medida que o número de covariáveis reduziu-se (Tabela 6 e Figuras 10, 11 e 12). Assim, os modelos captaram melhor as relações entre os dados. Por outro lado, sem a SM, os dois modelos apresentaram melhores performances com o acréscimo de variáveis e isso ocorreu porque os modelos selecionam mais variáveis que sejam explicativas do fenômeno em análise. Entre os modelos selecionados, o RF foi o melhor modelo.

Tabela 6 – Avaliação dos melhores e piores desempenhos dos modelos RF e GBM, considerando o número de covariáveis utilizadas para alcançar tal valor.

Modelos	Métricas após RFE					
	MAE		LCCC		RMSE	
	(números de variáveis)		(números de variáveis)		(números de variáveis)	
	Com SM	Sem SM	Com SM	Sem SM	Com SM	Sem SM
RF	26,01 (3)	32,72 (25)	0,61 (117)	0,30 (2)	36,45 (3)	46,73 (35)
	28,07 (117)	39,06 (2)	0,69 (2)	0,48 (30)	39,58 (117)	55,87 (2)
GBM	26,94 (14)	33,58 (40)	0,66 (117)	0,26 (2)	38,48 (13)	48,37 (35)
	28,09 (2)	39,42 (2)	0,68 (11)	0,44 (15)	39,94 (2)	56,22 (2)

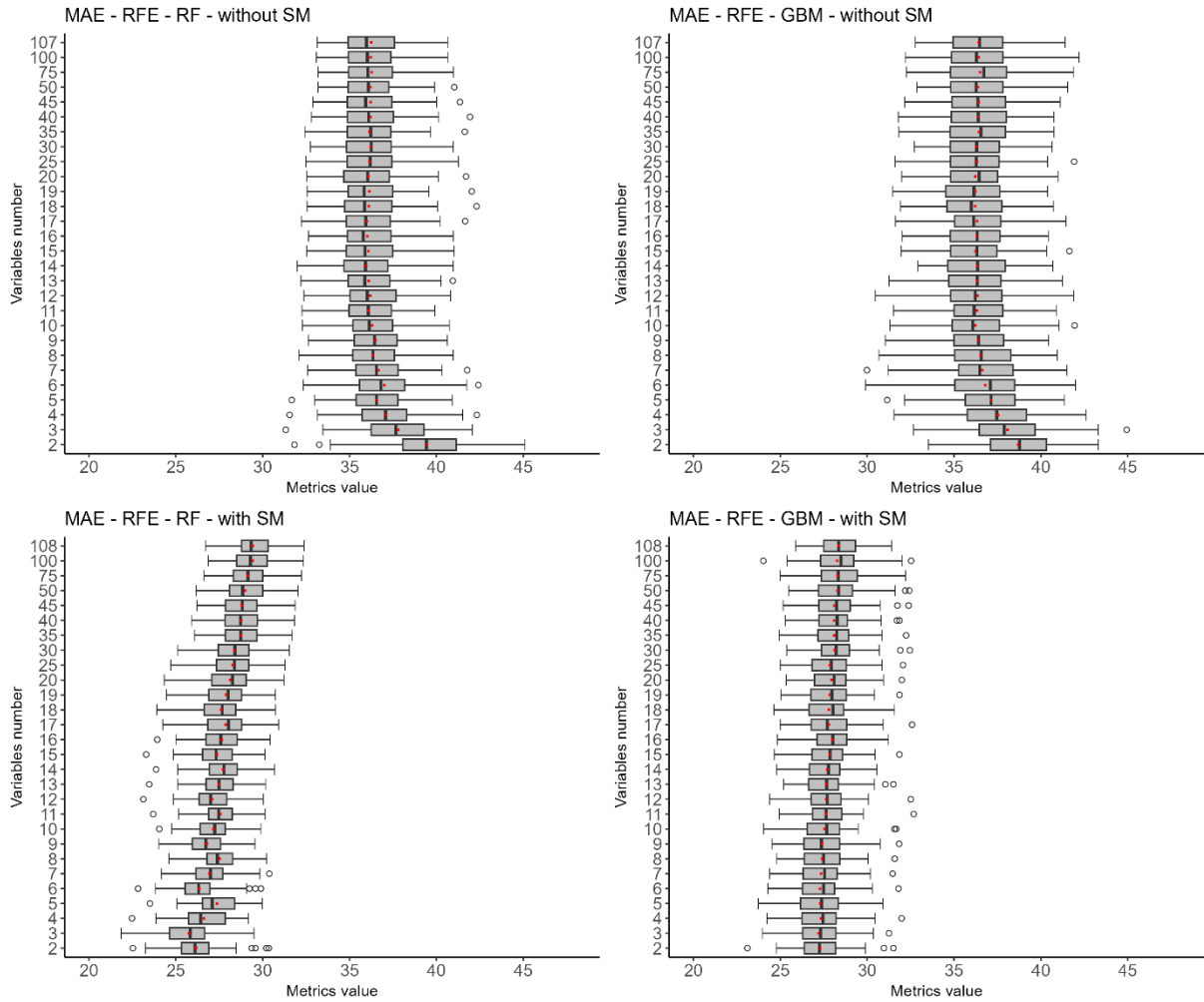
Fonte: Elaborado pelo autor.

Figura 10 – LCCC dos modelos durante o processo do RFE. “sem SM” são as métricas sem a susceptibilidade magnética e “com SM” são com a presença da susceptibilidade.



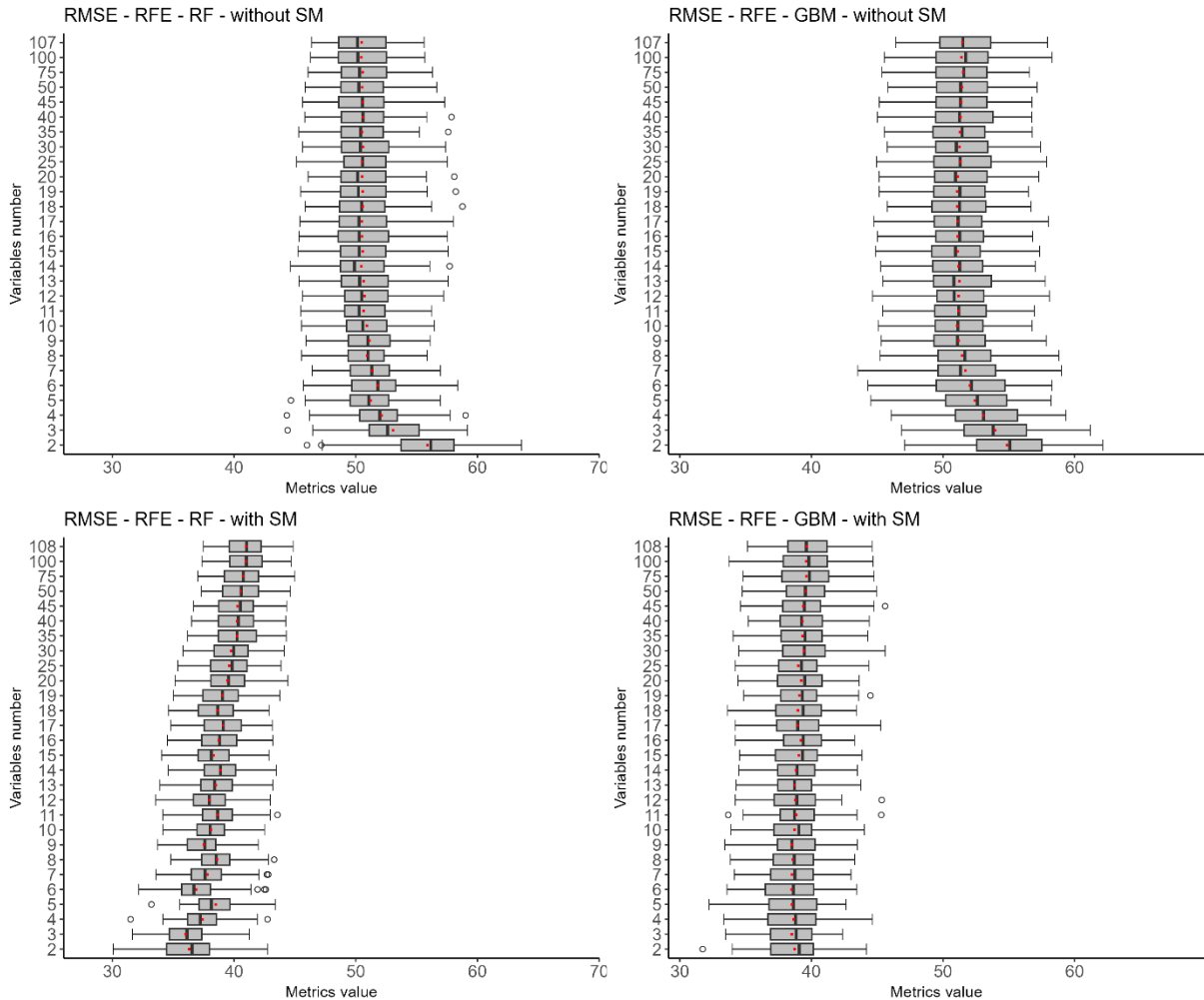
Fonte: Elaborado pelo autor.

Figura 11 – MAE dos modelos durante o processo do RFE. “sem SM” são as métricas sem a susceptibilidade magnética e “com SM” são com a presença da susceptibilidade.



Fonte: Elaborado pelo autor.

Figura 12 – RMSE dos modelos durante o processo do RFE. “sem SM” são as métricas sem a susceptibilidade magnética e “com SM” são com a presença da susceptibilidade.



Fonte: Elaborado pelo autor.

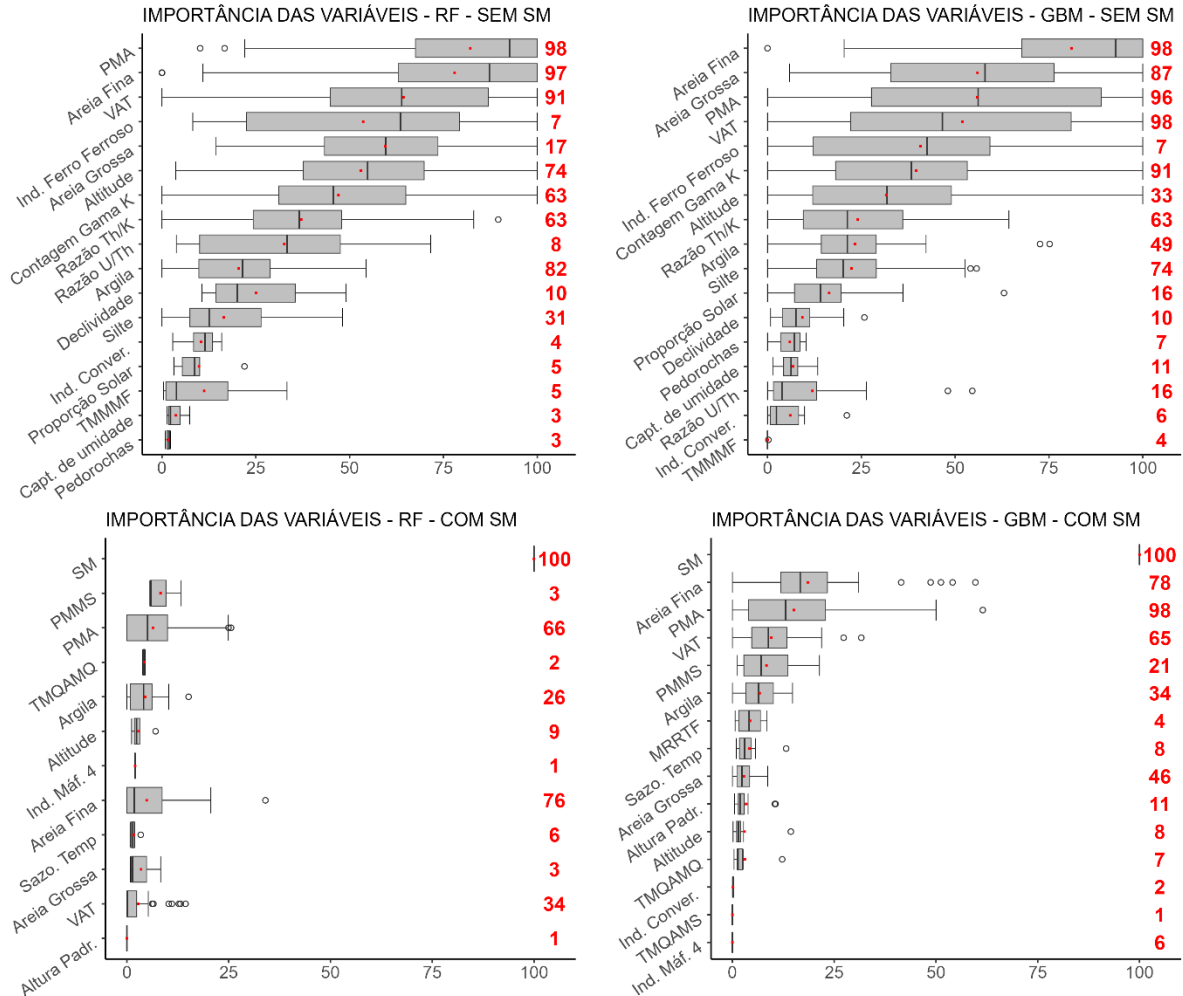
De forma geral, os modelos obtiveram as piores métricas com menos variáveis, com exceção à SM que é uma variável altamente correlacionada ($\rho = 0,76$) com os teores de ferro. Os subconjuntos construídos pelo RFE indicaram que a correlação entre as covariáveis explicativas com a variável resposta é fundamental para a melhoria das métricas dos modelos, o que implica na maior capacidade de extração de padrão entre os dados.

3.4 Importância dos preditores

Durante o processo de seleção das variáveis, observou-se o predomínio da SM, para ambos os modelos (figura 13), como a variável mais vezes selecionada e de maior importância. Além disso, ressalta-se que os subconjuntos selecionados pelo RFE, quando a SM esteve presente, foram menores. A menor frequência de seleção e as menores importâncias das outras variáveis mostram que os modelos se estabilizaram melhor com a SM, o que indica que o

acréscimo de outras variáveis, com a presença da SM, seria apenas para ajustes de ruídos dos modelos

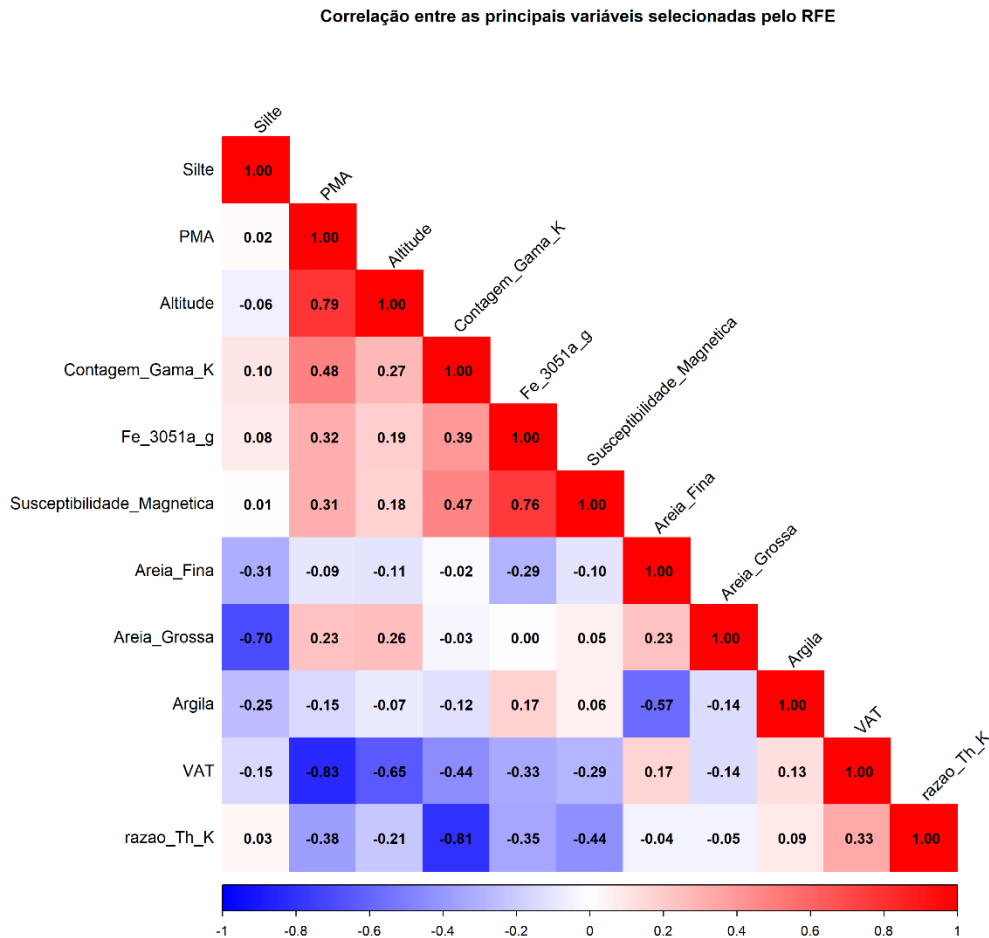
Figura 13 – Importância das variáveis escolhidas de acordo com os modelos. Os valores indicam a frequência de repetição. “sem SM” são as métricas sem a susceptibilidade magnética e “com SM” são com a presença da susceptibilidade.



Fonte: Elaborado pelo autor.

Com a ausência da SM, para o RF, observou-se que a PMA, a areia fina (AF) e a VAT foram as variáveis de maiores importâncias selecionadas em mais de 90% das rodadas. A PMA teve correlação de 0,32, a AF de -0,29 e a VAT de -0,33 com os teores de ferro (Figura 14). Para o GBM, observou-se que a AF, a PMA, a VAT e a Contagem gama de ^{40}K (Kperc) foram selecionadas em mais de 90% das rodadas. Destaca-se que a areia grossa (AG) foi outra variável de elevada importância. Em relação as correlações com os teores de ferro, a Kperc obteve 0,39 e a AG não teve correlação significativa.

Figura 14 – Resultado da correlação de Spearman entre as principais variáveis selecionadas pelo RFE.



Fonte: Elaborado pelo autor.

Observou-se que a PMA, AF e a VAT foram as variáveis que estiveram presentes, frequentemente, após a seleção pelo RFE, com e sem SM, e possuem elevadas importâncias. De certa forma, houve a predominância da seleção de covariáveis referentes às propriedades dos solos e as covariáveis bioclimáticas. Em relação ao fator “P”, destaca-se a importância das covariáveis Kperc e a razão $eTh/^{40}K$. A altitude, representativa do fator “R”, esteve presente com elevada importância apenas nas variáveis selecionadas pelo RF sem a SM e a sua correlação com os teores de ferro foi de 0,19.

4. DISCUSSÕES

4.1 Escolha dos modelos para construção da FPT

Como afirmado por Moquedace et al. (2024), a variação da performance mostrada pelas métricas demonstra a importância da repetição durante o processo da construção dos modelos

de ML sobre diferentes subconjuntos, o que evita resultados otimistas ou pessimistas que não correspondem à realidade do fenômeno em análise. Durante a construção das FPT que estimaram os teores de ferro no solo, a diferença entre os modelos pode ser explicada pela escolha das covariáveis e como eles são estruturados para analisar os dados. De forma geral, observou-se que as covariáveis escolhidas nos subconjuntos construídos pelo RFE, ocorreu, principalmente, pela relação existente entre a variável os teores de ferro.

Dentre todos os modelos escolhidos, o RF apresentou a melhor performance, quando a SM esteve presente. Alguns estudos prévios destacam o RF consistentemente como o mais usado e o melhor modelo de ML para predição de atributos pedológicos (Hengl et al., 2015; 2018; Siqueira et al., 2024). Por ser um modelo que se beneficia de um conjunto de árvores de decisão aleatórias treinadas independentemente, o RF consegue melhorar a sua capacidade de generalização e determinar uma predição mais precisa e estável, ou seja, com baixa variância (Siqueira et al., 2023). Dessa forma, ele foi o modelo mais consistente para captar relações hierárquicas e não-lineares nos dados.

No Brasil, para fins de classificação de solos, é comum a determinação dos teores dos óxidos por meio de ataque sulfúrico na fração argila. Além disso, alguns trabalhos procuram espacializar esses dados como os de Silva et al. (2020) e Mendes et al. (2022), que, respectivamente, utilizaram a geoestatística e o ML. Os trabalhos destacados encontraram valores de RMSE melhores do que os nossos. Contudo, os objetivos desses trabalhos foram realizar mapeamentos de solos e conseqüentemente, utilizaram-se mapas de solos, o que facilitou para os modelos a detecção de padrões. Para este trabalho, além do objetivo distinto, não houve a inserção de uma covariável de classes de solos, devido a sua inexistência.

Silva et al. (2020) também utilizaram o ataque sulfúrico para determinação dos teores de ferro na fração argila. Porém, para validar seus dados, aplicaram uma regressão linear simples para predição, o que pode explicar os seus baixos valores de RMSE que também estão em desacordo com os nossos dados. Mello et al. (2022) destacaram diversas novas tecnologias e diferentes modelos de ML que podem auxiliar no entendimento dos variados atributos de solo e encontraram valores de RMSE e MAE superiores aos nossos. Ademais, os teores de ferro foram determinados por ataque sulfúrico e o RF não foi o melhor modelo. Essa diferenciação dos resultados pode ser explicada pelos diferentes métodos analíticos empregados ou pelas diferentes covariáveis ambientais empregadas.

Para os nossos modelos, a presença de uma covariável dependente dos teores de ferro, como a SM, mostrou que a detecção de padrões entre os dados foi facilitada pela habilidade dos modelos em identificar relações que podem ser não-lineares e até mesmo hierárquicas entre a

variável alvo e os preditores ambientais (Moquedace et al., 2024). Sem a SM, o GBM se mostrou mais eficiente e Fernandes-Filho et al. (2024) destacaram que a utilização de diferentes modelos ajuda a reduzir as incertezas das predições, selecionando assim, o modelo mais adequado para cada questão. Ressalta-se que com a ausência da SM, em algumas rodadas, ambos os modelos escolheram um número maior de covariáveis, incluindo algumas com baixa correlação significativa com os teores de ferro. Entende-se que isso ocorreu, como uma forma de redução dos ruídos nos modelos, ajudando-os a melhorar sua capacidade preditiva.

Diferentes modelos conseguem analisar diversas bases de dados podendo acumular ganhos ou perdas de performances (Khaledian & Miller, 2020). Por sua vez, a comparação dos modelos analisados com os modelos nulos expôs suas usabilidades (Fernandes-Filho et al., 2024). O modelo KNN, quando a SM esteve presente e o modelo Earth, com a ausência da SM, não conseguiram extrair, de forma satisfatória, as relações existentes entre os dados.

O KNN, por ser um modelo mais simples, com a presença de dados muito dimensionais, não consegue captar as relações existentes e somente memoriza as informações ao perder a noção do “vizinho mais próximo”. Dessa forma, suas previsões se tornam instáveis e ruidosas. A baixa capacidade de detecção de padrões pelo Earth, se deu por ser um modelo linear. Em relação os algoritmos de árvores de decisão, como o RF e o GBM, Siqueira et al. (2023) afirmam que além do poder preditivo, eles oferecem uma grande eficiência contra ruídos e “*overfitting*”, bom desempenho em pequenos conjuntos de dados, boa interpretabilidade, facilidade de uso e o principal, otimização de relações não aditivas e não lineares.

4.2 Subconjuntos escolhidos pelo RFE para modelagem dos teores de ferro

Como destacado por Gomes et al. (2019), o RFE tem como objetivo simplificar os modelos e é amplamente utilizado para atender aos princípios da parcimônia, que pode ser interpretado como: o melhor modelo deve explicar o mesmo fenômeno com menos variáveis sem a perda de significativa de performance. Em acordo com essa afirmação, observou-se que com a ausência da SM, os modelos perderam em média aproximadamente 40% da performance absoluta. Embora, acompanhou-se o aumento do tamanho dos subconjuntos com covariáveis ótimas selecionadas.

Para os modelos sem a SM, o acréscimo de covariáveis significou um ganho de performance até determinado ponto, sendo que isso está em acordo com Gomes et al. (2019) e não se observou um ganho de capacidade preditiva. Quando a SM esteve presente, os modelos performaram melhor e as melhores métricas foram obtidas com um número menor de covariáveis ótimas. Com a presença da SM, a remoção de covariáveis resultou em ganho de

performance para ambos os modelos até determinando ponto, resultados esses que estão em acordo com Moquedace et al. (2024).

A perda de performance com o aumento expressivo de covariáveis escolhidas durante o RFE, quando a SM esteve presente, pode ser explicada porque ela é a covariável de maior importância. Além disso, escolheu-se a SM em todas as rodadas com a importância máxima. O acréscimo de covariáveis ocorreu apenas para a correção de ruídos, por parte dos modelos. Apesar das métricas mostrarem que o GBM foi o melhor modelo que performou sem a SM, percebeu-se que com a seleção de covariáveis executada durante o RFE, o RF foi o modelo que apresentou o melhor ajuste. Principalmente com a presença da SM, o RF construiu modelos com fortes capacidades de generalização e de predição ao utilizar diferentes subconjuntos mais otimizados (Gomes et al., 2019; Siqueira et al., 2023; Moquedace et al., 2024), sendo o RF o melhor modelo.

4.3 Importância das variáveis e suas relações com os teores de ferro

A SM foi a variável de maior importância e selecionada em todas as execuções. Mello et al. (2020) destacaram que a composição do material de origem e conseqüentemente a mineralogia das rochas e dos sedimentos que formam os solos, são os principais parâmetros que influenciam propriedades magnéticas. Os autores afirmam que a SM está associada à quantidade, forma, tipo e estrutura dos óxidos de ferro do solo e por isso é um importante indicador do grau de intemperismo e desenvolvimento pedogenético de solos tropicais.

As maiores magnetizações são observadas em solos com maiores conteúdos de óxidos como a Mt, comumente presente na fração areia e a Mh, presente na fração argila (Ker, 1997). De acordo com Barbosa & Oliveira (2020), a Mt é o mineral ferromagnético mais encontrado na natureza. Dessa forma, há uma relação intrínseca da SM com minerais de óxidos de ferro e o que explica sua predominância em todas as execuções.

Ao modelarem a SM de baixa frequência em solos do Brasil na fração argila, Barbosa et al. (2021) encontraram que os óxidos de ferro foram a principal covariável em nível de importância e demonstraram a relação direta e dependente entre as duas variáveis em acordo com os nossos resultados (Figura 12). Para Camargo et al. (2014) o comportamento magnético é mais evidente na fração argila. Os solos desenvolvidos de materiais ferruginosos do QF possuem valores elevados de ferro (Souza et al., 2015) e isso influencia diretamente na variação espacial da cristalinidade dos óxidos de ferro e da manifestação da SM.

A frequente importância e a presença constante das covariáveis bioclimáticas como a PMA e VAT demonstram as relações dessas variáveis com o processo de intemperização das

rochas e liberação de ferro no solo. Schaefer et al. (2015) destacaram como fatores comuns aos locais com alta presença de ferro nos solos do QF as amplitudes térmicas ao longo dos dias e alta exposição solar a ventos constantes.

Schaefer et al. (2016) ao descreverem o processo de formação e degradação das cangas lateríticas, que são formações ferruginosas, demonstraram a importância da precipitação ao longo de diferentes ciclos climáticos que influenciam o movimento vertical e lateral de produtos do intemperismo. Aliada a atividade biológica, os solventes e solutos derivados do intemperismo vão alterando o manto saprolítico, que neste caso são itabiríticos, que ajudam no processo de dissolução e reprecipitação dos óxidos de ferro (Schaefer et al., 2016).

Para Ferrari et al. (2016) a diferença de temperatura anual é explicada pelo albedo, onde as áreas ferruginosas, são mais capazes de absorver maior energia, independente da época do ano. Portanto, a variação da maior temperatura do ar no mês mais quente para a menor temperatura do ar no mês mais frio, é menor. A afirmação é confirmada pela correlação negativa entre os teores de ferro e a VAT. De certa forma, em relação ao clima, pode-se afirmar que a precipitação anual é o principal fator para acumulação de ferro no solo, entendendo que a água é o principal agente regulador de ciclos e reações biogeoquímicas.

No que diz respeito as covariáveis morfométricas, pode-se destacar a maior relevância da covariável altitude, com uma maior presença de ferro em porções mais altas da paisagem, onde predominam as formações ferríferas. Para Fernandes-Filho et al. (2022) as áreas em que os solos possuem mais ferro acumulado no QF, se encontram principalmente entre 900 e 2000 metros de altitude. Para Schaefer et al. (2023) as formações geológicas ferríferas exercem um forte controle estrutural sobre o relevo regional devido a sua resistência à erosão, sendo os solos um importante marcador da evolução da paisagem nesses locais.

Em relação às variáveis pedológicas, a predominância da AF e a aparição de outros componentes da textura do solo entre as variáveis selecionadas pelos modelos, mostram a relação direta com a mineralogia e a textura do solo. A correlação negativa entre a AF e os teores de ferro pode ser explicado pela maior presença de ferro na fração argila. Além disso, devido às Formações Quartzosas, muito comuns no QF, o Quartzo (Qz), presente na fração areia, possui baixos teores de ferro. Destaca-se que em solos ferríferos do QF, é habitual a presença de pseudocomponentes, onde agregados oxídicos de alta estabilidade, ocasionalmente não são totalmente quebrados pela dispersão química e mecânica (Donagemma et al., 2003), e por isso, a fração areia não é somente composta por Qz.

A Hematita (Hm) primária é o principal mineral encontrado na fração areia dos solos do QF desenvolvidos sobre as formações ferríferas (BIFs), porém a Mt é presente (Fabris et al.,

1997). De acordo com Carvalho-Filho et al. (2015) na fração argila, é possível encontrar a Hm pedogenética, Goethita (Gt) e a Mh. A Mt e Mh são os minerais que apresentam características ferromagnéticas, enquanto que Hm e Gt são minerais conhecidos como antiferromagnéticos, são marcados por baixos valores de SM e são os óxidos de ferro mais comuns em solos tropicais (Fabris et al., 1997; 1998; Santos et al., 2023).

Schaefer et al. (2016) ao definirem os solos das áreas ferríferas afirmam que elas estão em um estado nutricional muito pobre, com capacidade de troca de cátions (CTC) extremamente baixa, refletindo a natureza oxidica. Além disso, os valores de pH são levemente ácidos, como resultado do alto grau de intemperismo, o que deixou a CTC esgotada, que depende quase inteiramente de uma matéria orgânica muito pobre e ácida (Schaefer et al., 2016).

Em relação as informações gamaespectrométricas, segundo Gomes (2021), geralmente esses dados apresentam baixos valores em BIFs do QF, o que é justificada a presença da covariável K_{perc} entre as mais importantes. Além disso, a Razão $eTh/^{40}K$ também aparece entre as principais covariáveis, o que pode ser explicado por dois fatores: a baixa concentração de ^{40}K nas BIFs, já que é um elemento facilmente lixiviado e a alta capacidade de adsorção do eTh por argilas oxidicas, o que eleva a sua concentração residual (Mello et al., 2023a 2023b).

4.4 Por que utilizar uma FPT?

A grande vantagem de se utilizar FPT são as estimativas de atributos ou de propriedades do solo que são custosas de serem mensuradas em campo ou o preenchimento de dados não existentes em informações legadas (Arbor et al., 2023; van Leeuwen et al., 2024). Além disso, as FPT se mostram eficazes, pois ajudam na redução da utilização de ácidos e outros produtos possivelmente tóxicos à saúde humana, que são comumente utilizados na determinação dos elementos nos solos, como é o caso dos teores de ferro.

Neste trabalho, os modelos preditivos construídos apresentaram boas performances, principalmente com a presença da SM. Por sua vez, a utilização de um sensor ativo não destrutivo, como o susceptibilímetro magnético, colaborou na construção de modelos mais assertivos para a predição dos teores de ferro nos solos do QF. De acordo com Reddy & Das (2023), os solos são responsáveis por importantes serviços ecossistêmicos e o desenvolvimento de estratégias limpas, como as FPT, colaboram de forma sustentável e ajudam a construir informações decisivas de forma rápida, barata e precisa.

Para Miti et al. (2023) os dados legados são fundamentais para o bom desenvolvimento de FPT, já que eles podem acrescentar informações ao objeto em análise. Em consenso com a

afirmação de Arbor et al. (2024), a principal defasagem das FPT são o desenvolvimento e as aplicações locais e regionais. Contudo, os resultados obtidos foram satisfatórios e a qualidade das FPT segundo Choudhury et al. (2023) é dependente das covariáveis utilizadas. Sendo assim, em uma região com poucas informações pedológicas, as FPT construídas conseguiram explicar a relações entre os teores de ferro e as covariáveis preditoras a medida em que se mostram uma alternativa viável, econômica e sustentável.

5. CONCLUSÕES

Neste estudo, criou-se modelos preditivos definidos como FPT utilizando técnicas de ML para a determinação dos teores de ferro em solos do QF. Dentre os modelos avaliados, o RF apresentou as melhores métricas de desempenho e foi considerado o modelo mais adequado para este estudo. As principais covariáveis preditoras selecionadas foram: atributos de solo, como o SM e a granulometria; as covariáveis bioclimáticas, como a precipitação e a temperatura; além das covariáveis relacionadas à altitude e a gamaespectroscopia.

Observou-se que os modelos de ML apresentaram melhor desempenho quando a SM esteve presente. Isso mostra que em um ambiente tão diverso, a presença de uma variável independente que tenha relação direta com a variável alvo ajudou os modelos a detectarem melhor os padrões existentes entre os dados. A construção de modelos de predições robustos, como para FPT, demonstrou-se eficaz, já que não é possível mensurar atributos ou propriedades dos solos em cada localização de forma direta.

Este estudo contribuiu para ampliar o conhecimento sobre os solos do QF, que, devido à sua inadequação ao cultivo e baixa presença de elementos básicos, ainda são pouco estudados. Destaca-se que a utilização de dados gratuitos disponíveis na internet, como as covariáveis utilizadas como preditoras, se mostraram uma ferramenta poderosa para o desenvolvimento das FPT. Por outro lado, ressalta-se a importância do avanço de análises e do conhecimento sobre os solos do QF que podem ser uma ferramenta fundamental e subsidiar o desenvolvimento de técnicas de recuperação e restauração ambiental nas mais diversas áreas degradadas pela atividade minerária.

REFERÊNCIAS

- Adhikari, K., & Hartemink, A. E. (2016). Linking soils to ecosystem services - A global review. In *Geoderma* (Vol. 262, pp. 101–111). Elsevier. <https://doi.org/10.1016/j.geoderma.2015.08.009>

- Almeida, C. C., Fontes, M. P. F., Dias, A. C., Pereira, T. T. C., & Ker, J. C. (2020). Adsorption and desorption of arsenic and its immobilization in soils. *Scientia Agricola*, 78(3). <https://doi.org/10.1590/1678-992x-2018-0368>
- Arbor, A., Schmidt, M., Saurette, D., Zhang, J., Bulmer, C., Filatow, D., Kasraei, B., Smukler, S., & Heung, B. (2023). A framework for recalibrating pedotransfer functions using nonlinear least squares and estimating uncertainty using quantile regression. *Geoderma*, 439. <https://doi.org/10.1016/j.geoderma.2023.116674>
- Arbor, A., Schmidt, M., Zhang, J., Bulmer, C., Filatow, D., Kasraei, B., Smukler, S., & Heung, B. (2024). Filling the gaps in soil data: A multi-model framework for addressing data gaps using pedotransfer functions and machine-learning with uncertainty estimates to estimate bulk density. *Catena*, 245. <https://doi.org/10.1016/j.catena.2024.108310>
- Assis, L. M., Francelino, M. R., Daher, M., Fernandes-Filho, E. I., Veloso, G. V., Gomes, L. C., & Schaefer, C. E. G. R. (2021). Modeling regolith thickness in iron formations using machine learning techniques. *Catena*, 207. <https://doi.org/10.1016/j.catena.2021.105629>
- Barbosa, J. Z., Poggere, G. C., Silva, S. H. G., Mancini, M., Motta, A. C. V., Marques, J. J. G. S. M., & Curi, N. (2021). National-scale spatial variations of soil magnetic susceptibility in Brazil. *Journal of South American Earth Sciences*, 108. <https://doi.org/10.1016/j.jsames.2021.103191>
- Barbosa, M. S. C., & Oliveira, L. A. S. (2020). Geofísica do Quadrilátero Ferrífero. In P. T. A. Castro, I. Endo, & A. L. Gandini (Eds.), *Quadrilátero Ferrífero: Avanços do conhecimento nos últimos 50 anos* (pp. 114–141).
- Brasil. (2010). *Mapa de Geodiversidade do estado de Minas Gerais* (Vol. 12). <https://rigeo.sgb.gov.br/handle/doc/14704>
- Brenning, A. (2008). Statistical geocomputing combining R and SAGA: The example of landslide susceptibility analysis with generalized additive models. In J. Boehner, T. Blaschke, & L. Montanarella (Eds.), *SAGA - Seconds Out (= Hamburger Beitrage zur Physischen Geographie und Landschaftsoekologie)* (Vol. 19, pp. 23–32).
- Camargo, L. A., Marques-Júnior, J., Pereira, G. T., & Bahia, A. R. de S. (2014). Clay mineralogy and magnetic susceptibility of Oxisols. *Scientia Agricola*. <https://doi.org/10.1590/S0103-90162014000300010>
- Camêlo, D. de L., Ker, J. C., Fontes, M. P. F., da Costa, A. C. S., Corrêa, M. M., & Leopold, M. (2018). Mineralogy, magnetic susceptibility and geochemistry of Fe-rich oxisols developed from several parent materials. *Scientia Agricola*, 75(5), 410–419. <https://doi.org/10.1590/1678-992x-2017-0087>
- Carvalho-Filho, A. Solos e ambientes do Quadrilátero Ferrífero (MG) e aptidão silvicultural dos Tabuleiros Costeiros. Tese (Doutorado). Universidade Federal de Lavras, Lavras, 2008.
- Carvalho-Filho, A., Curi, N., & Shinzato, E. (2010). Relações solo-paisagem no Quadrilátero Ferrífero em Minas Gerais. In *Pesq. agropec. bras* (Issue 8). <https://doi.org/10.1590/S0100-204X2010000800017>
- Carvalho-Filho, A., Curi, N., Marques, J. J. G. S. M., Shinzato, E., de Jesus, E. A., Freitas, D. A. F., & Massahud, R. T. L. R. (2011). Óxidos de manganês em solos do quadrilátero ferrífero (MG). *Revista Brasileira de Ciência Do Solo*, 35, 793–804. <https://doi.org/10.1590/S0100-06832011000300015>
- Carvalho-Filho, A., Inda, A. V., Fink, J. R., & Curi, N. (2015). Iron oxides in soils of different lithological origins in Ferriferous Quadrilateral (Minas Gerais, Brazil). *Applied Clay Science*, 118, 1–7. <https://doi.org/10.1016/j.clay.2015.08.037>
- Carvalho-Júnior, W., Chagas, C. da S., Calderano-Filho, B., & Bhering, S. B. (2013). *Funções de pedotransferência para estimativa da densidade dos solos de áreas tropicais montanhosas*.

https://www.researchgate.net/publication/279854424_Funcoes_de_pedotransferencia_para_estimativa_da_densidade_dos_solos_de_areas_tropicais_montanhasas

- Choudhury, B. U., Santra, P., Singh, N., & Chakraborty, P. (2023). Development of land-use-specific pedotransfer functions for predicting bulk density of acidic topsoil in eastern Himalayas (India). *Geoderma Regional*, 34. <https://doi.org/10.1016/j.geodrs.2023.e00671>
- Coelho, M. R.; Vasques, G. de M.; Tassinari, D.; Souza, Z. R. de.; Oliveira, A. P. de.; Moreira, F. M. de. S. Solos do Quadrilátero Ferrífero sob diferentes coberturas vegetais e materiais de origem. Boletim de pesquisa e desenvolvimento, Embrapa Solos, Rio de Janeiro, 2017.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>
- Costa, S. A. D., Ker, J. C., Simões, D. F. de F., Fontes, M. P. F., & Fabris, J. D. (2014). Pedogênese e classificação de latossolos desenvolvidos de itabiritos no Quadrilátero Ferrífero, MG. *Revista Brasileira de Ciência Do Solo*. <https://doi.org/10.1590/S0100-06832014000200001>
- Da Silva, L. de C. M., Amorim, R. S. S., Fernandes-Filho, E. I., Bocuti, E. D., & Da Silva, D. D. (2023). Pedotransfer functions and machine learning: Advancements and challenges in tropical soils. In *Geoderma Regional* (Vol. 35). Elsevier B.V. <https://doi.org/10.1016/j.geodrs.2023.e00720>
- Demattê, J. A. M., & Terra, F. da S. (2014). Spectral pedology: A new perspective on evaluation of soils along pedogenetic alterations. *Geoderma*, 217–218, 190–200. <https://doi.org/10.1016/j.geoderma.2013.11.012>
- Diniz, J. M. F. de S., Dos Reis, A. A., Acerbi-Júnior, F. W., & Gomide, L. R. (2014). Detecção da expansão da área minerada no Quadrilátero Ferrífero, Minas Gerais, no período de 1985 a 2011 através de técnicas de sensoriamento remoto. *Boletim de Ciências Geodesicas*, 20(3), 683–700. <https://doi.org/10.1590/S1982-21702014000300039>
- Donagemma, G. K., Ruiz, H. A., Fontes, M. P. F., Ker, J. C., & Schaefer, C. E. G. R. (2003). Dispersão de Latossolos em resposta à utilização de pré-tratamentos na análise textural. *Revista Brasileira de Ciência Do Solo*. <https://doi.org/10.1590/S0100-06832003000400021>
- Dorr, J. V. N. (1969). *Physiographic, Stratigraphic and Structural Development of the Quadrilatero Ferrifero*.
- Dos Santos, L. A. C., Araújo, F. P., Souza-Junior, V. S., Silva, L. S., Campos, M. C. C., & Araujo, J. K. S. (2023). Magnetic and spectral signature of anthropogenic and non-anthropogenic soils to identify pedogenetic processes in Amazonia, Brazil. *Geoderma Regional*, 32. <https://doi.org/10.1016/j.geodrs.2023.e00622>
- Dos Santos, P. A., Koenow, H. S. P., Carvalho-Júnior, W., Pereira, N. R., Bhering, S. B., & da Silva, I. L. (2024). Predicting Soil Physical-Hydric Attributes Based on Pedotransfer Functions and Algorithms for Quantitative Pedology. In *Pedometrics in Brazil* (pp. 47–62). https://doi.org/10.1007/978-3-031-64579-2_4
- EPA, U., & of Resource Conservation, O. (2007). *Method 3051A: Microwave Assisted Acid Digestion of Sediments, Sludges, Soils, and Oils, part of Test Methods for Evaluating Solid Waste, Physical/Chemical Methods*.
- Fabris, J. D., Coey, J. M. D., Da, W., & Mussel, N. (1998). Magnetic soils from mafic lithodomains in Brazil. In *Hyperfine Interactions* (Vol. 113).
- Fabris, J. D., De, M. F., Filho, Jesus., Coey, J. M. D., Da, W., Mussel, N., & Goulart, A. T. (1997). Iron-rich spinels from Brazilian soils. In *Hyperfine Interactions* (Vol. 110).
- Fernandes-Filho, E. I., Moquedace, C. M., Pereira, L. F., Veloso, G. V., & Carvalho-Júnior, W. (2024). Methods and Challenges in Digital Soil Mapping: Applied Modelling with R

- Examples. In *Pedometrics in Brazil* (pp. 263–283). https://doi.org/10.1007/978-3-031-64579-2_19
- Fernandes-Filho, E. I., Schaefer, C. E. G. R., Faria, R. M., Lopes, A. O., Francelino, M. R., & Gomes, L. C. (2022). The unique and endangered Campo Rupestre vegetation and protected areas in the Iron Quadrangle, Minas Gerais, Brazil. *Journal for Nature Conservation*, 66. <https://doi.org/10.1016/j.jnc.2022.126131>
- Ferrari, L. T., Schaefer, C. E. G. R., Fernandes, R. B. A., Mendonça, B. A. F., Gjorup, D. F., Corrêa, G. R., & Senra, E. O. (2016). Thermic and Hydric Dynamics of Ironstone (Canga) and Quartzite Rupestrian Grasslands in the Quadrilátero Ferrífero: The Ecological Importance of Water. In *Ecology and Conservation of Mountaintop grasslands in Brazil* (pp. 71–85). Springer International Publishing. https://doi.org/10.1007/978-3-319-29808-5_4
- Ferreira, M. M., Fernandes, B., & Curi, N. (1999). Influência da mineralogia da fração argila nas propriedades físicas de latossolos da região sudeste do Brasil. *Revista Brasileira de Ciência Do Solo*, 23, 515–524. <https://doi.org/10.1590/S0100-06831999000300004>
- Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, 19(1). <https://doi.org/10.1214/aos/1176347963>
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5). <https://doi.org/10.1214/aos/1013203451>
- Gomes, D. G. da. C. (2021). *Atlas aerogeofísico do estado de Minas Gerais*. www.cprm.gov.br
- Gomes, L. C., Faria, R. M., de Souza, E., Veloso, G. V., Schaefer, C. E. G. R., & Fernandes-Filho, E. I. (2019). Modelling and mapping soil organic carbon stocks in Brazil. *Geoderma*, 340, 337–350. <https://doi.org/10.1016/j.geoderma.2019.01.007>
- Hawker, L., Uhe, P., Paulo, L., Sosa, J., Savage, J., Sampson, C., & Neal, J. (2022). A 30 m global map of elevation with forests and buildings removed. *Environmental Research Letters*, 17(2). <https://doi.org/10.1088/1748-9326/ac4d4f>
- Hechenbichler, K., & Schliep, K. (2004). *Weighted k-Nearest-Neighbor techniques and ordinal classification*.
- Hengl, T., Heuvelink, G. B. M., Kempen, B., Leenaars, J. G. B., Walsh, M. G., Shepherd, K. D., Sila, A., MacMillan, R. A., de Jesus, J. M., Tamene, L., Tondoh, J. E., 2015. Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. 10, 1–26. <https://doi.org/10.1371/journal.pone.0125814>.
- Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. M., Graler, B., 2018. Random forest as a generic framework for predictive modelling of spatial and spatio-temporal variables. <https://doi.org/10.7717/peerj.5518>.
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. r G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25(15), 1965–1978. <https://doi.org/10.1002/joc.1276>
- Jacobi, C. M., Do Carmo, F. F., Vincent, R. C., & Stehmann, J. R. (2007). Plant communities on ironstone outcrops: A diverse and endangered Brazilian ecosystem. *Biodiversity and Conservation*, 16(7), 2185–2200. <https://doi.org/10.1007/s10531-007-9156-8>
- Jenny, H. (1941) *Factors of Soil Formation: A System of Quantitative Pedology*. Dover Publications, New York, 281 p.
- Khaledian, Y., & Miller, B. A. (2020). Selecting appropriate machine learning methods for digital soil mapping. *Applied Mathematical Modelling*, 81, 401–418. <https://doi.org/10.1016/j.apm.2019.12.016>
- Kuhn, M., 2020. caret: Classification and Regression Training.
- McBratney, A. B., Mendonça-Santos, M. L., & Minasny, B. (2003). On digital soil mapping. *Geoderma*, 117(1–2), 3–52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)

- Mello, D. C., Demattê, J. A. M., Silvero, N. E. Q., Di Raimo, L. A. D. L., Poppiel, R. R., Mello, F. A. O., Souza, A. B., Safanelli, J. L., Resende, M. E. B., & Rizzo, R. (2020). Soil magnetic susceptibility and its relationship with naturally occurring processes and soil attributes in pedosphere, in a tropical environment. *Geoderma*, 372. <https://doi.org/10.1016/j.geoderma.2020.114364>
- Mello, D. C., Ferreira, T. O., Veloso, G. V., Lana, M. G., Mello, F. A. O., Di Raimo, L. A. D. L., Cabrero, D. R. O., Souza, J. J. L. L., Fernandes-Filho, E. I., Francelino, M. R., & Demattê, J. A. M. (2023a). Digital mapping of soil weathering using field geophysical sensor data coupled with covariates and machine learning. *Journal of South American Earth Sciences*, 128. <https://doi.org/10.1016/j.jsames.2023.104449>
- Mello, D. C., Ferreira, T. O., Veloso, G. V., Lana, M. G., Mello, F. A. O., Di Raimo, L. A. D. L., Cabrero, D. R. O., Souza, J. J. L. L., Fernandes-Filho, E. I., Francelino, M. R., & Demattê, J. A. M. (2023b). Digital mapping of soil weathering using field geophysical sensor data coupled with covariates and machine learning. *Journal of South American Earth Sciences*, 128. <https://doi.org/10.1016/j.jsames.2023.104449>
- Mello, D. C., Osório, T. F., Veloso, G. V., Lana, M. G., Mello, F. A. O., Di Raimo, L. A. D. L., Schaefer, C. E. G. R., Francelino, M. R., Fernandes-Filho, E. I., & Demattê, J. A. M. (2022). Pedogenetic processes operating at different intensities inferred by geophysical sensors and machine learning algorithms. *Catena*, 216. <https://doi.org/10.1016/j.catena.2022.106370>
- Mello, D. C., Veloso, G. V., Lana, M. G., Mello, F. A. O., Poppiel, R. R., Cabrero, D. R. O., Di Raimo, L. A. D. L., Schaefer, C. E. G. R., Fernandes-Filho, E. I., Leite, E. P., & Demattê, J. A. M. (2022). A new methodological framework for geophysical sensor combinations associated with machine learning algorithms to understand soil attributes. *Geoscientific Model Development*, 15(3), 1219–1246. <https://doi.org/10.5194/gmd-15-1219-2022>
- Mello, D. C., Veloso, G. V., Mello, M. F., Lana, M. G., de Angeli, I. O., Mello, F. A. O., Siqueira, R. G., Gomes, L. C., Fernandes-Filho, E. I., Schaefer, C. E. G. R., Francelino, M. R., Leite, E. P., Ferreira, T. O., & Demattê, J. A. M. (2025). Integrating proximal geophysical sensing and machine learning for digital soil mapping: Spatial prediction and model evaluation using a small dataset. *Soil Advances*, 3, 100024. <https://doi.org/10.1016/j.soilad.2024.100024>
- Mello, D. C., Veloso, G. V., Moquedace, C. M., de Angeli, I. O., Francelino, M. R., Oliveira, F. S., Souza, J. J. L. L., Gomes, L. C., Schaefer, C. E. G. R., Medeiros-Júnior, E., Fernandes-Filho, E. I., & Demattê, J. A. M. (2023). Chemical weathering detection in the periglacial landscapes of Maritime Antarctica: New approach using geophysical sensors, topographic variables and machine learning algorithms. *Geoderma*, 438. <https://doi.org/10.1016/j.geoderma.2023.116615>
- Mello, D. C., Veloso, G. V., Moquedace, C. M., de Angeli, I. O., Oliveira, F. S., Gomes, L. C., Souza, J. J. L. L., Francelino, M. R., Fernandes-Filho, E. I., Schaefer, C. E. G. R., & Demattê, J. A. M. (2023). Radiometric and magnetic susceptibility characterization of soil profiles: Geophysical data and their relationship with Antarctic periglacial processes, pedogenesis, and lithology. *Catena*, 232. <https://doi.org/10.1016/j.catena.2023.107427>
- Mendes, W. de S., Demattê, J. A. M., Minasny, B., Silvero, N. E. Q., Bonfatti, B. R., Safanelli, J. L., Rizzo, R., & Costa, A. C. S. da. (2022). Free iron oxide content in tropical soils predicted by integrative digital mapping. *Soil and Tillage Research*, 219. <https://doi.org/10.1016/j.still.2022.105346>
- Minasny, B., & Hartemink, A. E. (2011). Predicting soil properties in the tropics. In *Earth-Science Reviews* (Vol. 106, Issues 1–2, pp. 52–62). <https://doi.org/10.1016/j.earscirev.2011.01.005>

- Minasny, B., & McBratney, A. B. (2006). A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers and Geosciences*, 32(9), 1378–1388. <https://doi.org/10.1016/j.cageo.2005.12.009>
- Miti, C., Mbanye, V., Mtangadura, T., Magwero, N., Namaona, W., Njira, K., Sandram, I., Lubinga, P. N., Chisanga, C. B., Nalivata, P. C., Chimungu, J. G., Nezomba, H., Phiri, E., & Lark, R. M. (2023). The appraisal of pedotransfer functions with legacy data; an example from southern Africa. *Geoderma*, 439. <https://doi.org/10.1016/j.geoderma.2023.116661>
- 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. (2024). High-resolution mapping of soil carbon stocks in the western Amazon. *Geoderma Regional*, 36. <https://doi.org/10.1016/j.geodrs.2024.e00773>
- Padarian, J., Minasny, B., & McBratney, A. B. (2020). Machine learning and soil sciences: A review aided by machine learning tools. In *SOIL* (Vol. 6, Issue 1, pp. 35–52). Copernicus GmbH. <https://doi.org/10.5194/soil-6-35-2020>
- Pereira, L. F., Moquedace, C. M., Rosolem, G. P. N., de Sousa, M. da C., Francelino, M. R., & Fernandes-Filho, E. I. (2024). Applying Machine Learning Techniques to Model and Map Soil Surface Texture Using Limited Legacy Data. In *Pedometrics in Brazil* (pp. 31–45). https://doi.org/10.1007/978-3-031-64579-2_3
- R Core Team. (2024). *A language and environment for statistical computing*. <https://www.R-project.org/>
- Reddy, N. N., & Das, B. S. (2023). Digital soil mapping of key secondary soil properties using pedotransfer functions and Indian legacy soil data. *Geoderma*, 429. <https://doi.org/10.1016/j.geoderma.2022.116265>
- Resende, M., Ker, J. C., de Rezende, S. B., Poggere, G., Zimmer, J. B., Mancini, M., & Curi, N. (2023). *Roteiro pedológico de Minas Gerais: Conhecendo e interpretando os principais solos de diferentes ambientes*. https://www.researchgate.net/publication/370124938_ROTUIRO_PEDOLOGICO_DE_MINAS_GERAIS_Conhecendo_e_interpretando_os_principais_solos_de_diferentes_ambientes
- Santos, M. do C., & Varajão, A. F. D. C. (2003). Sedimentation and pedogenic features in a clay deposit in Quadrilátero Ferrífero, Minas Gerais, Brazil. *Anais Da Academia Brasileira de Ciências*, 76(1), 147–159. <https://doi.org/10.1590/S0001-37652004000100013>
- Santos, M. do C., Varajão, A. F. D. C., & Yvon, J. (2004). Genesis of clayey bodies in Quadrilátero Ferrífero, Minas Gerais, Brazil. *Catena*, 55(3), 277–291. [https://doi.org/10.1016/S0341-8162\(03\)00106-1](https://doi.org/10.1016/S0341-8162(03)00106-1)
- Schaefer, C. E. G. R., Cândido, H. C., Corrêa, G. R., Pereira, A., Nunes, J. A., Souza, O. F., Marins, A., Fernandes-Filho, E. I., & Ker, J. C. (2015). Solos desenvolvidos sobre canga ferruginosa no Brasil: Uma revisão crítica e papel ecológico de termiteiros. In F. F. Carmo & L. H. Y. Kamino (Eds.), *Geossistemas ferruginosos no Brasil* (pp. 77–102).
- Schaefer, C. E. G. R., Candido, H. G., Corrêa, G. R., Campos, P. V., Senra, E. O., Gjorup, D. F., & Fernandes-Filho, E. I. (2023). Soils of Campos Rupestres (Rupestrian Grasslands) of the Old Brazilian Mountain Ranges. In *Ecology and Conservation of Mountaintop grasslands in Brazil* (pp. 343–362). https://doi.org/10.1007/978-3-031-19949-3_12
- Schaefer, C. E. G. R., Corrêa, G. R., Candido, H. G., Arruda, D. M., Nunes, J. A., Araujo, R. W., Rodrigues, P. M. S., Fernandes-Filho, E. I., Pereira, A. F. S., Brandão, P. C., & Neri, A. V. (2016). The Physical Environment of Rupestrian Grasslands (Campos Rupestres) in Brazil: Geological, Geomorphological and Pedological Characteristics, and Interplays.

- In *Ecology and Conservation of Mountaintop grasslands in Brazil* (pp. 15–53). Springer International Publishing. https://doi.org/10.1007/978-3-319-29808-5_2
- Sena, N. C., Veloso, G. V., Lopes, A. O., Francelino, M. R., Fernandes-Filho, E. I., Senra, E. O., Silva Filho, L. A. da, Condé, V. F., Silva, D. L. de A., & Araújo, R. W. de. (2021). Soil sampling strategy in areas of difficult access using the cLHS method. *Geoderma Regional*, 24. <https://doi.org/10.1016/j.geodrs.2020.e00354>
- Silva, L. S., Marques-Júnior, J., Barrón, V., Gomes, R. P., Teixeira, D. D. B., Siqueira, D. S., & Vasconcelos, V. (2020). Spatial variability of iron oxides in soils from Brazilian sandstone and basalt. *Catena*, 185. <https://doi.org/10.1016/j.catena.2019.104258>
- Silva, S. H. G., Silva, E. A., Poggere, G. C., Junior, A. L. P., Gonçalves, M. G. M., Guilherme, L. R. G., & Curi, N. (2020). Modeling and prediction of sulfuric acid digestion analyses data from pXRF spectrometry. *Scientia Agricola*, 77(4). <https://doi.org/10.1590/1678-992x-2018-0132>
- Siqueira, R. G., Moquedace, C. M., Fernandes-Filho, E. I., Schaefer, C. E. G. R., Francelino, M. R., Sacramento, I. F., & Michel, R. F. M. (2024). Modelling and prediction of major soil chemical properties with Random Forest: Machine learning as tool to understand soil-environment relationships in Antarctica. *Catena*, 235. <https://doi.org/10.1016/j.catena.2023.107677>
- Siqueira, R. G., Moquedace, C. M., Francelino, M. R., Schaefer, C. E. G. R., & Fernandes-Filho, E. I. (2023). Machine learning applied for Antarctic soil mapping: Spatial prediction of soil texture for Maritime Antarctica and Northern Antarctic Peninsula. *Geoderma*, 432. <https://doi.org/10.1016/j.geoderma.2023.116405>
- Souza, J. J. L. L., Abrahão, W. A. P., de Mello, J. W. V., da Silva, J., da Costa, L. M., & de Oliveira, T. S. (2015). Geochemistry and spatial variability of metal(loid) concentrations in soils of the state of Minas Gerais, Brazil. *Science of the Total Environment*, 505, 338–349. <https://doi.org/10.1016/j.scitotenv.2014.09.098>
- Teixeira, P. C., Donagemma, G. K., Fontana, A., & Teixeira, W. G. (2017). *Manual de métodos de análise de solo - Embrapa*.
- Valaee, M., Ayoubi, S., Khormali, F., Lu, S. G., & Karimzadeh, H. R. (2016). Using magnetic susceptibility to discriminate between soil moisture regimes in selected loess and loess-like soils in northern Iran. *Journal of Applied Geophysics*, 127, 23–30. <https://doi.org/10.1016/j.jappgeo.2016.02.006>
- van Leeuwen, C. C. E., Mulder, V. L., Batjes, N. H., & Heuvelink, G. B. M. (2024). Effect of measurement error in wet chemistry soil data on the calibration and model performance of pedotransfer functions. *Geoderma*, 442. <https://doi.org/10.1016/j.geoderma.2023.116762>
- Varajão, A. F. D. C., Mateus, A. C. C., Santos, M. do C., Varajão, C. A. C., Oliveira, F. S., & Yvon, J. (2020). The Cenozoic deposits of the ancient landscapes of Quadrilátero Ferrífero highlands, Southeastern Brazil: Sedimentation, pedogenesis and landscape evolution. *Catena*, 195. <https://doi.org/10.1016/j.catena.2020.104813>
- Vasconcelos, B. N. F. (2016). *Mapeamento digital de solos em diferentes escalas: Abordagem metodológica*.
- Wadoux, A. M. J. C., Minasny, B., & McBratney, A. B. (2020). Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Science Reviews*, 210, 103359. <https://doi.org/10.1016/j.earscirev.2020.103359>
- Witten, I. H., & Frank, E. (2002). Data mining. *ACM SIGMOD Record*, 31(1), 76–77. <https://doi.org/10.1145/507338.507355>
- Zeraatpisheh, M., Ayoubi, S., Jafari, A., Tajik, S., & Finke, P. (2019). Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. *Geoderma*, 338, 445–452. <https://doi.org/10.1016/j.geoderma.2018.09.006>

CAPÍTULO 2

Predições dos teores de ferro em solos do Quadrilátero Ferrífero: O uso do Mapeamento Digital de Solos para o aprofundamento do conhecimento pedológico na região.

KRAUSE, M. B.¹; MOQUEDACE, C. M. M. d. S.¹; PEREIRA, L. F.¹; SIQUEIRA, R. G.¹; FRANCELINO, M. R.¹; CARVALHO-JÚNIOR, W.²; FERNANDES-FILHO, E. I.¹.

1 – Departamento de Solos, Universidade Federal de Viçosa, Viçosa – MG.

2 – Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA - Solos), Rio de Janeiro – RJ.

RESUMO

Os solos são recursos reguladores dos principais ciclos biogeoquímicos que ocorrem no planeta, e seus estudos são essenciais para a manutenção da biodiversidade. Nesse sentido, torna-se necessário aprofundar o conhecimento sobre os solos em áreas com poucas pesquisas voltadas para esse objetivo, como o Quadrilátero Ferrífero (QF) que é a principal província mineralógica do sudeste brasileiro, embora ainda haja poucos estudos dedicados à compreensão das dinâmicas pedológicas. O avanço das técnicas de Mapeamento Digital de Solos (MDS) tem contribuído para o aprimoramento das informações pedológicas, especialmente em áreas com escassez de dados detalhados. Com base nessas técnicas, este artigo tem como objetivo modelar e mapear a distribuição espacial dos teores de ferro nos solos do QF, utilizando diferentes algoritmos de machine learning (ML). Para isso, foram utilizadas 165 amostras, algumas delas provenientes do Banco de Solos de Minas Gerais. A modelagem foi realizada com seis algoritmos de ML e a seleção das covariáveis preditoras seguiu dois critérios principais: a eliminação por correlação ($|\rho| < 0,95$) e a seleção baseada na importância das variáveis pelo método *Recursive Feature Elimination*. Todo o processo de seleção, treinamento, validação, teste e predição foi repetido 100 vezes. A avaliação dos mapas considerou valores medianos, médios, coeficiente de variação, desvio padrão e os quartis de 5% e 95%. Os modelos Random Forest e K-Nearest Neighbors apresentaram os melhores desempenhos, com MAE de 39,06 e 39,22, respectivamente; LCCC de 0,36 e 0,34; R^2 de 0,20 para ambos; e RMSE de 56,66 e 57,52. Em relação às covariáveis, observou-se a predominância das bioclimáticas, com destaque para a precipitação e a temperatura. A altitude também se mostrou uma covariável relevante e as razões de gamaespectrometria emergiram como importantes preditores. Embora os materiais originários não tenham aparecido entre as covariáveis mais importantes, eles desempenham um papel fundamental na distribuição espacial dos teores de ferro nos solos do QF. A análise dos mapas revelou que as maiores concentrações de ferro ocorrem em áreas ferríferas, que são berços da biodiversidade no QF e estão altamente ameaçadas pela atividade minerária. Nossos resultados fornecem informações relevantes para a gestão e o planejamento ambiental, auxiliando a tomada de decisão em políticas de conservação e uso sustentável dos solos do QF. Além disso, reforça-se a importância das técnicas de MDS no detalhamento de padrões espaciais, contribuindo para estratégias sustentáveis voltadas à recuperação, proteção e conservação ambiental.

Palavras-Chave: Biodiversidade; Conservação; Preservação; Machine learning.

ABSTRACT

Soils are regulatory resources of the main biogeochemical cycles occurring on the planet, and their study is essential for maintaining biodiversity. In this context, it is necessary to deepen the knowledge of soils in areas with limited research on this subject, such as the Quadrilátero Ferrífero (QF), which is the main mineral province in southeastern Brazil, although there are still few studies dedicated to understanding pedological dynamics. The advancement of Digital Soil Mapping (DSM) techniques has contributed to the improvement of pedological information, especially in areas with a scarcity of detailed data. Based on these techniques, this study aims to model and map the spatial distribution of iron content in QF soils using different machine learning (ML) algorithms. For this purpose, 165 samples were used, some of which were obtained from the Minas Gerais Soil Bank. The modeling was performed using six ML algorithms, and the selection of predictor covariates followed two main criteria: elimination by correlation ($|\rho| < 0.95$) and selection based on variable importance using the Recursive Feature Elimination method. The entire process of selection, training, validation, testing, and prediction was repeated 100 times. The map evaluation considered median values, mean values, coefficient of variation, standard deviation, and the 5% and 95% quartiles. The Random Forest and K-Nearest Neighbors models showed the best performances, with MAE of 39.06 and 39.22, respectively; LCCC of 0.36 and 0.34; R^2 of 0.20 for both; and RMSE of 56.66 and 57.52. Regarding the covariates, bioclimatic variables were predominant, with precipitation and temperature standing out. Altitude was also identified as a relevant covariate, and gamma spectrometry ratios emerged as important predictors. Although parent materials did not appear among the most important covariates, they play a fundamental role in the spatial distribution of iron content in QF soils. The analysis of the maps revealed that the highest iron concentrations occur in ferriferous areas, which are biodiversity hotspots in the QF and are highly threatened by mining activities. Our results provide relevant information for environmental management and planning, assisting decision-making in conservation policies and the sustainable use of QF soils. Furthermore, the importance of DSM techniques is reinforced in detailing spatial patterns, contributing to sustainable strategies aimed at environmental recovery, protection, and conservation.

Keywords: Biodiversity; Conservation; Preservation; Machine learning.

1. INTRODUÇÃO

Em vista da sustentabilidade, o solo é um recurso vital para o planeta, sustentando diversas funções ecossistêmicas essenciais (Teixeira et al., 2017), o que faz que o seu mapeamento e classificação são ações de grande relevância ambiental em diferentes escalas. Definidos como corpos naturais resultantes da interação entre o clima, organismos, relevo e material de origem que agem juntos em diferentes intensidades ao longo do tempo (Jenny, 1941; Mendonça-Santos & dos Santos, 2006), os solos são fundamentais para processos biogeoquímicos que sustentam a vida no planeta e sua saúde é essencial à segurança ambiental (Polidoro et al., 2016).

Aliado ao desenvolvimento de novas tecnologias, a inserção de modelos de Machine Learning (ML) na ciência do solo contribuiu para o avanço e desenvolvimento do Mapeamento Digital de Solos (MDS) (McBratney et al., 2003), alterando a forma como cientistas do solo elaboram seus mapas (Wadoux et al., 2020) e ampliando as potencialidades de uso desses produtos. Para Fernandes-Filho et al. (2024), os diferentes modelos de ML utilizados no MDS se constituem como uma poderosa ferramenta para a interpretação dos complexos pedoambientes.

Os grandes desafios à modelagem de distribuição espacial de propriedades, características e classes de solos, ao qual são muito comuns para o MDS, são as escolhas e as representatividades das covariáveis a serem utilizadas (Wadoux et al., 2020). Muitas vezes, dados legados podem não ser padronizados e aqueles derivados de sensores remotos ou proximais podem garantir apenas informações da cobertura do solo. Dessa forma, os modelos encontrarão apenas padrões indiretos que levam a produção de dados e informações com mais incertezas e erros (Fernandes-Filho et al., 2024).

O Brasil, devido ao seu tamanho continental, ainda necessita aprofundar o conhecimento sobre seus solos e principalmente produzir mapas com escalas mais refinadas (Polidoro et al., 2016). O desenvolvimento de novos trabalhos de MDS torna-se necessário, já que de acordo com Mendonça-Santos & dos Santos (2006) eles fornecem suporte à avaliação, planejamento e gestão do solo seja como recurso econômico ou provedor de serviços ambientais (Polidoro et al., 2016).

Dessa forma, destaca-se o Quadrilátero Ferrífero (QF), que se constitui como a principal província mineralógica do sudeste brasileiro. Apesar de sua diversidade de condições ambientais, historicamente, poucas pesquisas exploraram os solos do QF, concentrando-se em tópicos específicos ou áreas limitadas (Carvalho-Filho et al., 2010). O aprofundamento do conhecimento dos solos do QF, torna-se importante, já que esse recurso se relaciona com os

diferentes materiais de origem da área, além de suas relações com componentes bióticos e abióticos que contribuem para a diversidade e sustentabilidade regional (Carvalho-Filho et al., 2015).

Devido aos materiais originários, é comum encontrar solos de constituição ferrífera no QF (Shinzato & Carvalho-filho, 2005). Os óxidos de ferro têm papel importante nas funções do solo, sendo que refletem condições de pH, potencial redox, umidade, temperatura, tem forte relação com as cores e mineralogia, além de desempenhar papel estrutural e ajudar na estabilidade (Viscarra-Rossel et al., 2010). Por causa da presença de cargas positivas na superfície dos minerais, os óxidos de ferro ainda ajudam na adsorção de elementos potencialmente tóxicos e metais pesados (Schwertmann & Taylor, 1989; Viscarra-Rossel et al., 2010).

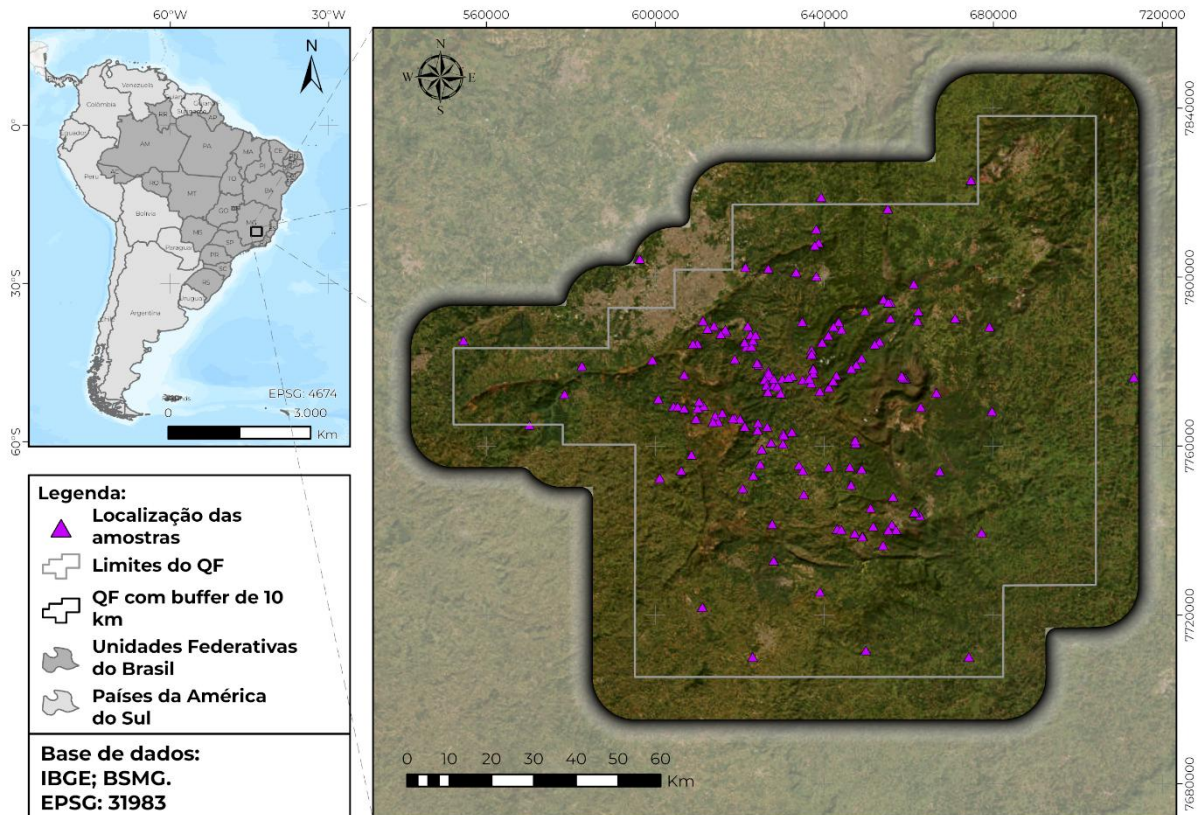
Assim, entende-se que aprofundar o conhecimento sobre os solos do QF, por meio de técnicas de MDS, torna-se fundamental para o auxílio à manutenção da sustentabilidade em uma área historicamente degradada (Schaefer et al., 2015). Dessa forma, este artigo tem como objetivo avaliar e mapear a distribuição espacial dos teores de ferro em solos do QF de 0 a 20 cm por meio de técnicas de MDS e modelos de ML, com o intuito de identificar padrões espaciais e compreender melhor a relação dos teores de ferro com a paisagem do QF. Além disso, entende-se que aprofundar os conhecimentos sobre a distribuição espacial dos teores de ferro pode contribuir com informações centrais ao processo de preservação ambiental.

2. MATERIAIS E MÉTODOS

2.1 Caracterização da área de estudo

O QF ocupa uma área de aproximadamente 7.200 km² na porção centro-sudeste do estado de Minas Gerais (MG) (Figura 1). A região é caracterizada por uma topografia acidentada decorrente de processos tectônicos que moldaram a paisagem ao longo de milhões de anos (Resende et al., 2023). Formado ao longo de tempos juro-cretáceos, o QF, possui raízes em um passado geológico em que terras da América do Sul, África, Oceania e Antártica ainda compunham o supercontinente Gondwana (Carvalho-Filho et al., 2010). Esse contexto histórico-geológico é determinante para a formação dos seus depósitos minerais. (Dorr, 1969; Carvalho-Filho et al., 2011).

Figura 1 – Mapa de localização da área de estudo.



Fonte: Elaborado pelo autor.

Historicamente, o QF foi fundamental para a mineração de ferro e ouro, com papel central no desenvolvimento econômico e cultural do Brasil (Schaefer et al., 2015; Souza et al., 2015). Além disso, é a região de maior concentração urbana do estado de MG (Fernandes-Filho et al., 2022). A riqueza ambiental do QF é evidenciada por sua vegetação diversificada, que reflete uma transição ecotonal entre os domínios dos Cerrados, da Floresta Atlântica e dos Campos Rupestres Ferruginosos ou Quartzíticos (Schaefer et al., 2016; Fernandes-Filho et al., 2022).

Com altitudes que variam entre 650 a mais de 2000 metros, o QF apresenta clima predominante Cwa, conforme a classificação de Köppen, caracterizado por verões chuvosos e invernos secos (Santos et al., 2004; Diniz et al., 2014; Carvalho-Filho et al., 2015). Apesar dessas condições definirem características importantes, os estudos sobre os solos da região ainda são limitados, sendo que o mapeamento realizado por Shinzato & Carvalho-Filho (2005) na região central do QF constitui o estudo mais aprofundado sobre os solos da região.

De forma geral, esses autores classificaram os solos como minerais, pouco evoluídos, de baixa fertilidade natural, com textura variando de média a argilosa e presença comum de cascalhos e fragmentos de rocha. O material de origem exerce grande influência na pedogênese, sendo comum, nas áreas ferríferas, a mineralogia oxidada, com altos teores de ferro, cores

avermelhadas, concentrações ferruginosas e presença de magnetismo. Por outro lado, é comum encontrar solos com baixos teores de ferro, o que demonstra o contraste na região (Shinzato & Carvalho-Filho, 2005; Carvalho-Filho et al., 2010; 2015; Coelho et al., 2017).

2.2 Covariáveis ambientais e dados de solos

Para modelar a variabilidade espacial da concentração dos teores de ferro nos solos do QF, utilizou-se um conjunto de preditores ambientais que se relacionam ao modelo SCORPAN (McBratney et al., 2003). O modelo relaciona cada letra a algum atributo, propriedade ou fator de formação do solo (Jenny, 1941). Sendo que o fator “S” se relaciona as informações pedológicas, o “C” aos dados climáticos, o “O” as atividades dos organismos, o “R” as variáveis de relevo, o “P” as informações dos materiais originários, o “A” ao fator tempo e o “N” às informações espaciais.

Os dados pedológicos incluíram concentrações dos teores de ferro obtidos em amostras armazenadas no Banco de Solos de MG (BSMG) (Souza et al., 2015). Complementarmente, foram realizadas algumas coletas de amostras de solos na profundidade de 0 a 20 cm na região central do QF, o que totalizou 165 amostras representativas da variável litoestratigrafia do QF. Para padronizar os dados referentes às amostras coletadas e aos dados legados, as análises laboratoriais seguiram as mesmas diretrizes.

Realizou-se a determinação da concentração dos teores de ferro na terra fina seca ao ar, previamente separada conforme descrito por Teixeira et al. (2017) e peneirou-se através de uma peneira de malha de 200 *mesh*. A determinação da concentração seguiu o protocolo EPA 3051a (EPA & of Resource Conservation, 2007) e a digestão ácida das amostras ocorreu conforme as orientações descritas por Souza et al. (2015). Por fim, a quantificação das concentrações foi realizada por meio de um espectrômetro de emissão óptica por plasma induzido acoplado (ICP-OES).

A representação das covariáveis climáticas ocorreu com o grupo das 19 variáveis bioclimáticas provenientes do *WorldClim* (Hijmans et al., 2005) e essas consistem em dados de precipitação e temperatura média em uma série de 30 anos (1970 – 2000) (tabela 1). Modelou-se o relevo por meio de 58 covariáveis morfométricas (tabela 2) derivadas do modelo digital de elevação FABDEM (Hawker et al., 2022), construídas pelo pacote RSAGA (Brenning, 2008) do software R (R Core Team, 2024).

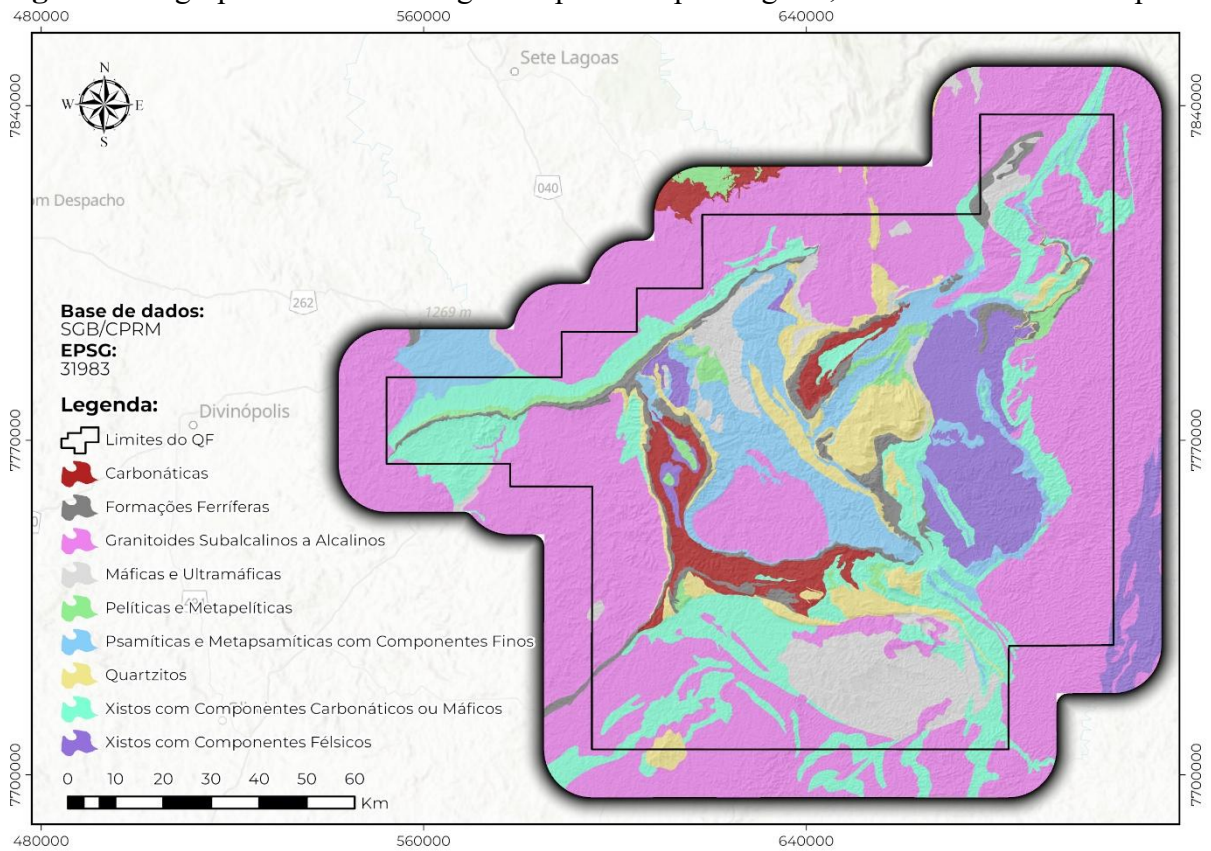
Tabela 1 – Covariáveis bioclimáticas escolhidas para a predição espacial dos teores de ferro nos solos do QF.

Variáveis bioclimáticas		
Temp média anual do ar (TMAA)	Temp média no quarto do ano mais úmido (TMQAMU)	Temp média no quarto do ano mais quente (TMQAMQ)
Precip média anual (PMA)	Precip no mês mais seco (PMMS)	Temp média no quarto do ano mais frio (TMQAMF)
Precip no mês mais úmido (PMMU)	Precip no quarto do ano mais seco (PQAMS)	Varição média da temp diária (VMTD)
Precip no quarto do ano mais frio (PQAMF)	Precip no quarto do ano mais úmido (PQAMU)	Precip no quarto do ano mais quente (PQAMQ)
Varição anual da temp (VAT)	Temp máxima no mês mais quente (TMMMQ)	Temp média no quarto do ano mais seco (TMQAMS)
Sazonalidade da precip	Razão entre a variação diária da temp (RVDT)	Temp mínima no mês mais frio (TMMMMF)
Sazonalidade da temp		

Fonte: Elaborado pelo autor.

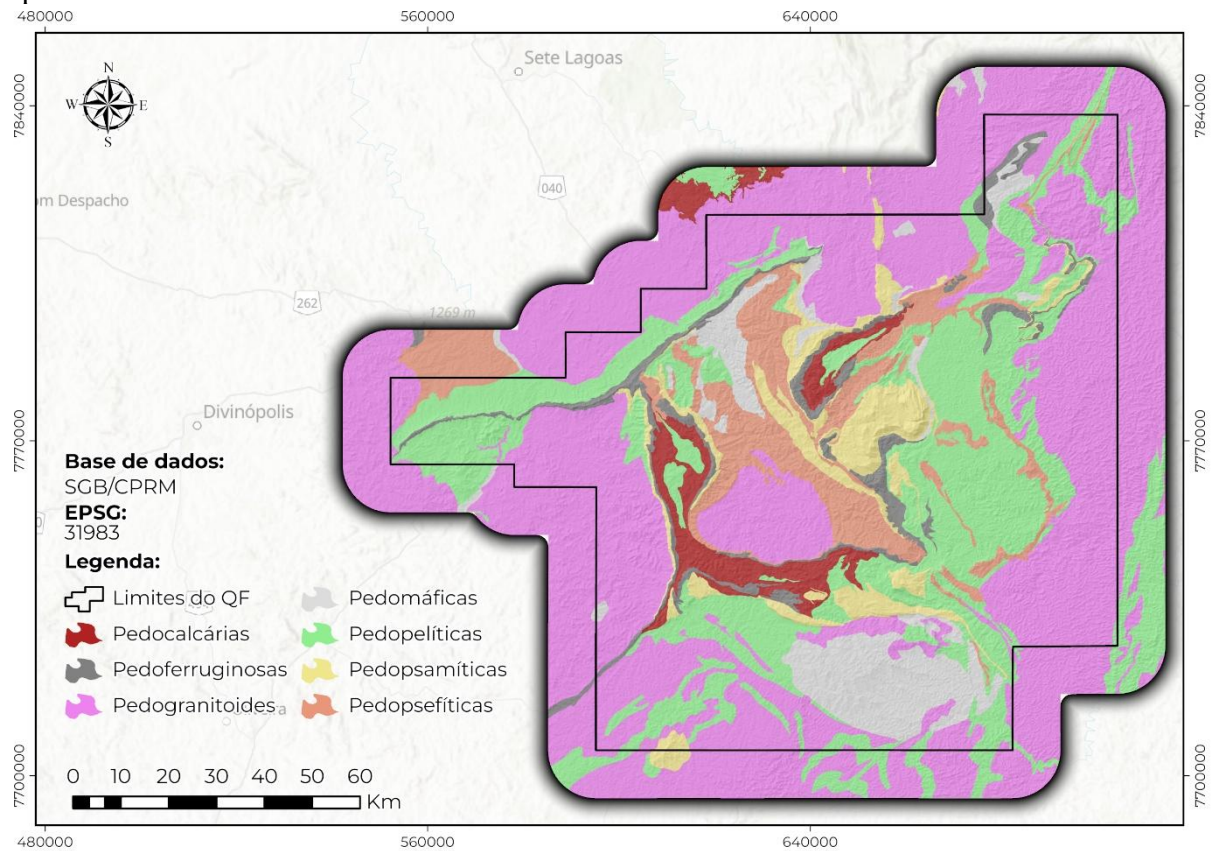
Dados de gamaespectroscopia, magnetometria, índices geofísicos e mapas de agrupamentos litoestratigráficos para fins pedológicos (tabela 3 e figuras 2 e 3) (BRASIL, 2010, Gomes, 2021; Resende et al., 2023) representaram o material de origem (tabela 4). Os índices geofísicos foram determinados conforme diretrizes descritas em Gomes (2021) e os agrupamentos litoestratigráficos seguiram as descrições de Resende et al. (2023).

Figura 2 – Agrupamento litoestratigráfico para fins pedológicos, considerando os “litotipos”.



Fonte: Elaborado pelo autor.

Figura 3 – Agrupamento litoestratigráfico para fins pedológicos, considerando as “pedorochas”.



Fonte: Elaborado pelo autor.

Além disso, utilizou-se 11 índices espectrais para retratar as atividades biológicas e usos do solo derivados de imagens Landsat-09 obtidas em março de 2023 (tabela 5). A data escolhida foi pela baixa presença de nuvens. Por fim, harmonizou-se todos os preditores para 30 metros de resolução e os projetou no sistema SIRGAS 2000 UTM Zona 23K (EPSG: 31983).

Tabela 2 – Covariáveis morfométricas escolhidas para a predição espacial dos teores de ferro nos solos do QF.

Covariáveis morfométricas derivadas do FABDEM (Hawker et al., 2022)		
Área de captação	Áreas de captações de umidade	Aspecto
Plano de curvatura	Linha de fluxo da curvatura	Índice de convergência
Curvatura mínima	Curvatura longitudinal	Curvatura geral
Tangência de curvatura	Seção transversal de curvatura	Curvatura máxima
Altitude	Pontos específicos de superfície	Perfil de curvatura
Índice do balanço de massa	Posição de meia encosta	Calor anisotrópico diário
Planicidade de fundo de vale (MRVBF)	Altura normalizada	Sombreamento
Nível de cume	Fator de visão do céu	Planicidade de topos (MRRTF)
Altura de encosta	Convexidade da superfície do terreno	Área superficial real
Radiação direta	Duração solar	Declividade graus
Proporção solar	Tempo de nascer solar	Difusão de radiação
Radiação total	Altura padronizada	Potencial de insolação
Abertura dos topos	Índice de rugosidade do terreno	Tempo de pôr do sol
Vales	Índice de posição topográfica	Curvatura total
Declividade	Textura superficial do terreno	

Fonte: Elaborado pelo autor.

Tabela 3 – Agrupamentos de rochas para fins pedológicos.

Agrupamentos por litotipos (Brasil, 2010)	Agrupamentos por fins pedológicos (Resende et al., 2023)
Carbonáticas	Pedocalcárias
Formações Ferríferas	Pedoferruginosas
Granitoides Subalcalinos a Alcalinos	Pedogranitoides
Máficas e Ultramáficas	Pedomáficas
Pelíticas e Metapelíticas	Pedopelíticas
Psamíticas e Metapsamíticas com Componentes Finos	Pedopsefíticas
Quartzitos	Pedopsamíticas
Xistos com Componentes Carbonáticos ou Máficos	Pedopelíticas
Xistos com Componentes Félsicos	

Fonte: Elaborado pelo autor.

Tabela 4 – Covariáveis geofísicas escolhidas para a predição espacial dos teores de ferro nos solos do QF.

Índices gamaespectrométricos (Gomes, 2021)		
Índice laterítico 1	Índice laterítico 2	Índice máfico 1
Índice máfico 2	Índice máfico 3	Índice máfico 4
Parâmetro F	Razão eTh/ ⁴⁰ K	Razão eU/ ⁴⁰ K
Razão eU/eTh	Contagem de ⁴⁰ K	Contagem de eU
Contagem de eTh	Contagem gama total	Magnetometria

Fonte: Elaborado pelo autor.

Tabela 5 – Covariáveis representativas de atividades biológicas, usos do solo e índice de solo escolhidas para a predição espacial dos teores de ferro nos solos do QF.

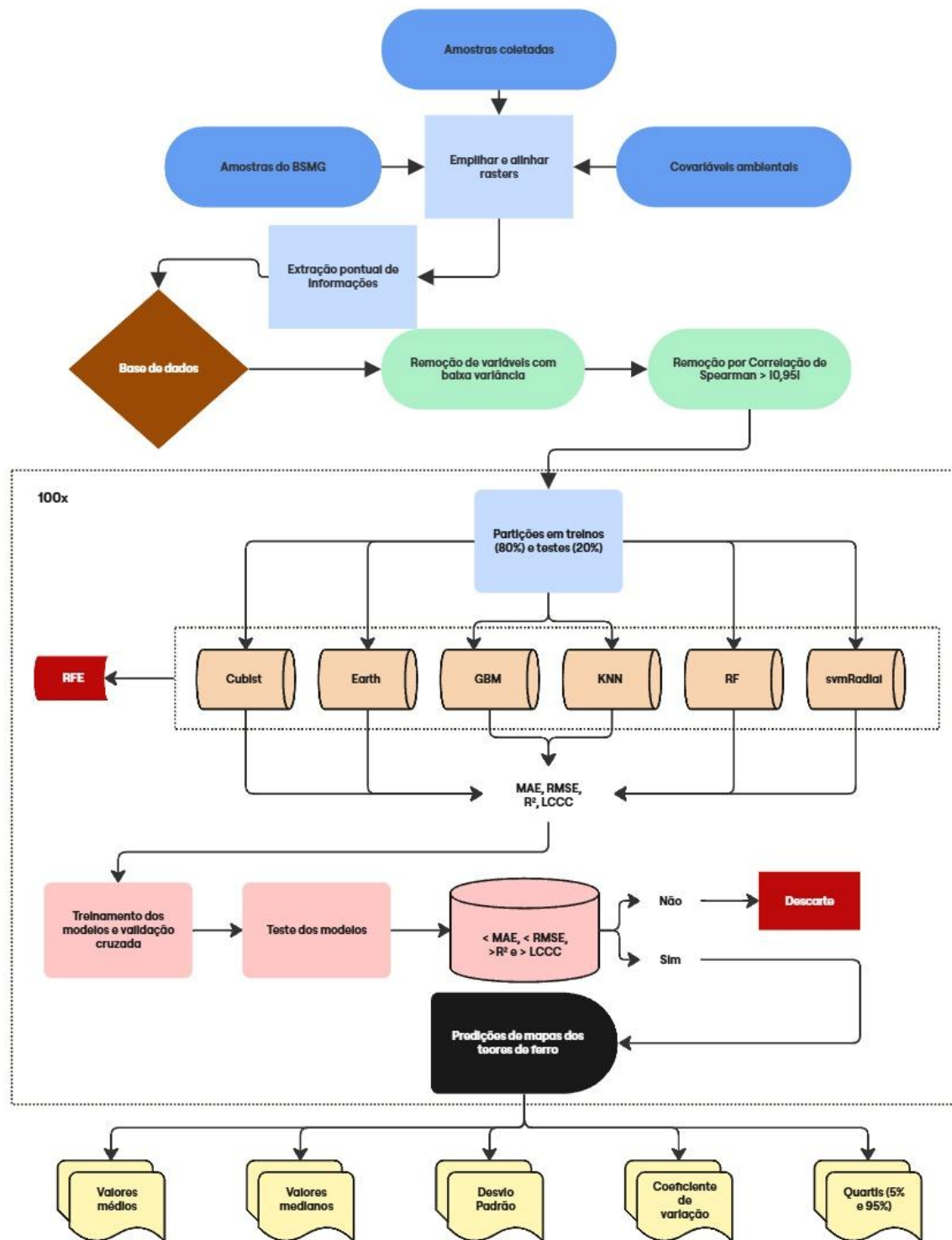
Índices espectrais		
Índice de minerais de argila	Índice de ferro ferroso	Índice de minerais ferrosos
Índice de óxidos de ferro	Índice de silicatos de ferro	Índice de clorofila verde (GCI)
Índice de diferença normalizada de vegetação verde (GNDVI)	Índice de diferença normalizada de Água (NDWI)	Índice de diferença normalizada de umidade (NDMI)
Índice de diferença normalizada de vegetação (NDVI)	Índice resistente à atmosfera na região visível (VARI)	Composição RGB
	Índice de vegetação ajustado ao solo (SAVI)	

Fonte: Elaborado pelo autor.

2.3 Seleções de covariáveis e importâncias dos preditores

Para selecionar os preditores mais importantes, reduzir possíveis redundâncias, facilitar a interpretabilidade dos modelos e reduzir os custos computacionais, baseou-se nas metodologias descritas por Pereira et al., (2024), Moquedace et al. (2024) e Siqueira et al. (2024) e aplicou-se uma seleção sequencial descrita em três passos. O primeiro foi a remoção das covariáveis de baixa variância pela função *nearZeroVar* do pacote *caret* (Kuhn, 2020). Em seguida, aplicou-se a função *findcorrelation* do pacote *caret* (Kuhn, 2020) para identificar pares de covariáveis com coeficientes de correlação de Spearman (ρ) superiores a 0,95 e eliminar umas das covariáveis do par com a maior correlação média absoluta. A sequência metodológica está descrita na figura 4.

Figura 4 – Fluxograma da metodologia realizada.



Fonte: Elaborado pelo autor.

No terceiro passo, após a separação de treinamento dos modelos e utilizando as covariáveis pré-selecionadas na primeira e segunda etapa, aplicou-se o método *Recursive Feature Elimination* (RFE) que identificou os melhores subconjuntos de covariáveis dentre as

selecionadas nas etapas iniciais. O RFE baseia-se na importância das covariáveis para a predição da variável-alvo e assim os modelos são otimizados por utilizar subconjuntos de covariáveis ótimas. (Gomes et al., 2019; Pereira et al., 2024). Ressalta que o RFE considera diferentes algoritmos base e por isso, cada modelo é ajustado individualmente.

Embora Fernandes-Filho et al. (2024) tenham destacado que o aumento no número de covariáveis pode melhorar o desempenho dos modelos, também os tornam mais complexos e difíceis de interpretar. É por isso que o RFE otimiza o modelo, sem perda significativa do poder preditivo (Moquedace et al., 2024). Para determinar a importância das covariáveis dos subconjuntos, aplicou-se a função *varImp* do pacote *caret* (Kuhn, 2020) e avaliou-se em conjunto a frequência de seleção das covariáveis nos subconjuntos construídos pelo RFE.

2.4 Modelos de Machine Learning

Os modelos de ML são capazes de identificar relações não lineares existentes entre preditores ambientais (Moquedace et al., 2024) e na ciência do solo tem crescido a sua utilização de forma a auxiliar as compreensões pedoespaciais (ten Caten et al., 2012). Estudar quantitativamente como o solo varia ao longo das paisagens ajuda a reduzir a subjetividade dos modelos mentais (McBratney et al., 2003; ten Caten et al., 2012; Carvalho-Júnior et al., 2013; Pereira et al., 2024).

Neste contexto, avaliou-se a performance de seis modelos de ML: *Adaptive Multivariate Regression* (Earth); *Cubist*; *K-Nearest Neighbors* (KNN); *Random Forest* (RF); *Stochastic Gradient Boosting* (GBM) e o *Support Vector Machine with Radial Basis Linear* (svmRadial). O Earth é um modelo linear flexível de regressão, que ao identificar relacionamentos entre os dados constrói funções de base que aproximam segmentos similares (Friedman, 1991).

O Cubist é um algoritmo que cria regras “se, depois e então”. Cada regra é associada a um modelo linear multivariado que calcula o valor predito quando a regra é atendida (Zeraatpisheh et al., 2019). O KNN é um algoritmo não paramétrico que considera a técnica do vizinho mais próximo, em geral, ponderando as observações mais próximas com pesos maiores e medindo a distância (Hechenbichler & Schliep, 2004; Schliep & Hechenbichler, 2016). O RF é um algoritmo baseado em um conjunto de árvores de decisão. O valor previsto é a média das previsões de todas as árvores e o algoritmo é treinado para reduzir a redundância e a possibilidade de *overfitting* do modelo (Liam & Wiener, 2002; Wadoux et al., 2020).

O GBM é um algoritmo que combina *boosting* e *bagging* para construir árvores sequenciais. A partir da primeira árvore, a subsequente é construída considerando os erros da árvore anterior. O *boosting* é implementado selecionando uma amostra aleatória para cada

etapa, sem reposição (Friedman, 2001; Rahman *et al.*, 2020). Por fim, o *svmRadial* é um algoritmo baseado no aprendizado estatístico e no princípio da margem máxima, maximizando a separação entre os vetores de suporte a partir da aplicação de uma função de *kernel* radial (Cortes & Vapnik, 1995). Todos os modelos foram ajustados por seus pacotes específicos presentes no software R (R Core Team, 2024). Todos os hiperparâmetros dos foram ajustados conforme especificações particulares pelo pacote *caret* (Kuhn, 2020).

2.5 Treinos, testes, predições, espacializações, otimizações e incertezas

Em todos os modelos de ML, aplicou-se uma validação cruzada repetida com 10 *folds*, com a divisão dos dados em 80% para treino e 20% para teste. A separação ocorreu por meio da função *createDataPartition* do pacote *caret* (Kuhn, 2020), que executa uma amostragem aleatória em percentis e garante a representação de toda a base de dados. O processo de modelagem incluiu as etapas de separação em subconjuntos de treinos e testes, seleções das covariáveis, predições e espacializações, repetidas 100 vezes.

O desempenho dos modelos foi avaliado pelas métricas de acurácia, incluindo o erro médio absoluto (MAE), a raiz quadrada do erro médio (RMSE), o Coeficiente de Concordância de Correlação de Lins (LCCC) e o coeficiente de determinação (R^2), as equações estão apresentadas na tabela 6. Apesar de Khaledian e Miller (2020) afirmarem que o R^2 não é uma métrica apropriada para a avaliação de modelos tão robustos, ele foi incluído devido à sua ampla utilização na literatura, permitindo assim, comparações e discussões dos resultados.

Tabela 6 – Equações das métricas utilizadas para avaliação dos modelos

$MAE = \frac{1}{n} \sum_{i=1}^n y_i - x_i $	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2}$
$LCCC = \frac{2\rho\sigma_{x_i}\sigma_{y_i}}{\sigma_{x_i}^2 + \sigma_{y_i}^2 + (\bar{x} - \bar{y})^2}$	$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - x_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$

Fonte: Elaborado pelo autor.

Como validação adicional, alguns autores destacam a relevância de modelos nulos como estratégia para estabelecer limites de avaliação (Siqueira *et al.*, 2023; Fernandes-Filho *et al.*, 2024; Moquedace *et al.*, 2024; Pereira *et al.*, 2024; Mello *et al.*, 2025). Assim, foram calculadas as métricas “*MAE Null*” e “*RMSE Null*”, em que os modelos são simples e os preditores são nulos, dessa forma, permite-se a comparação e interpretação dos resultados produzidos e buscase inferir a qualidade dos modelos. Além disso, realizou-se o teste estatístico de análise de

variância (ANOVA), com $p < 0,05$, para verificar se haveria diferença estatística significativa entre os desempenhos dos modelos.

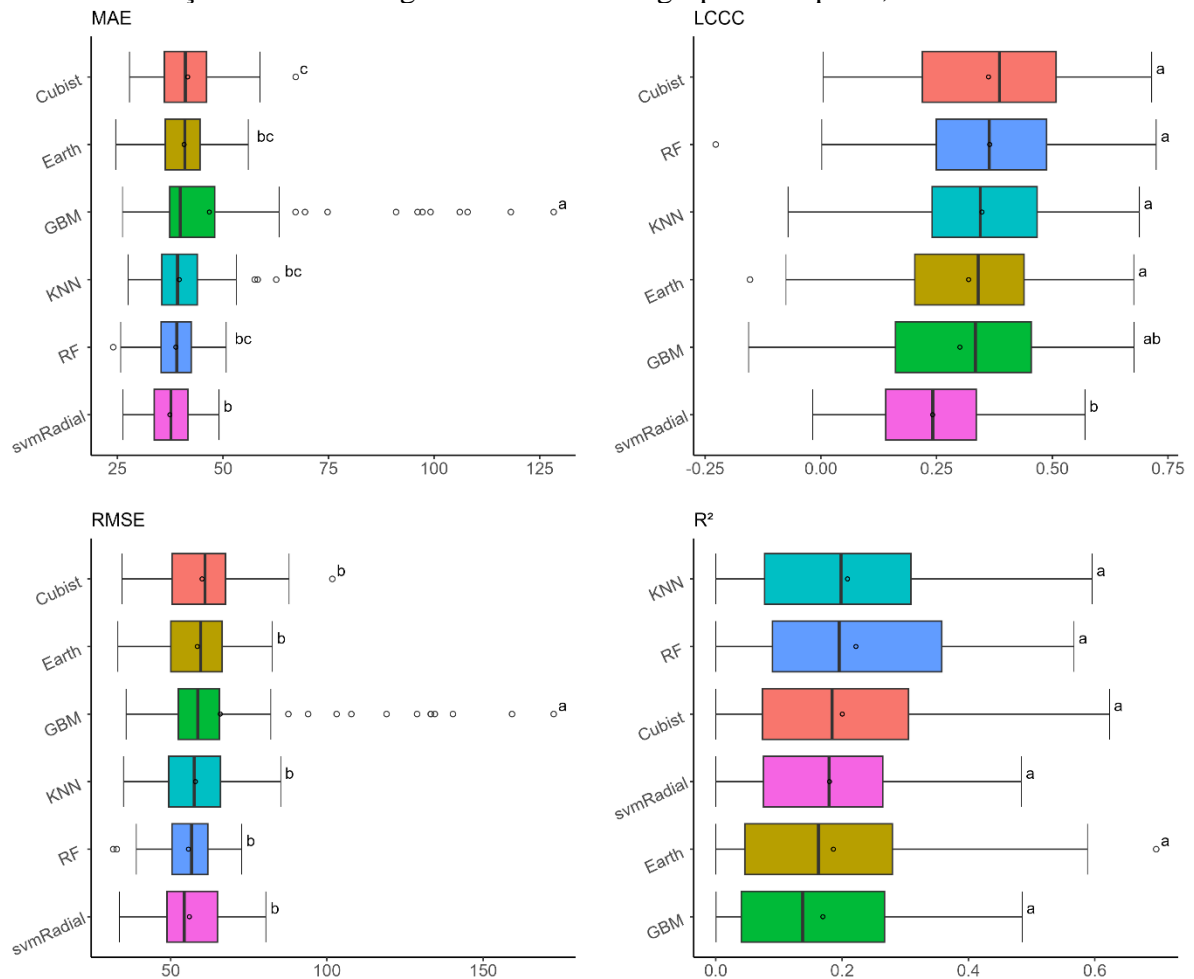
Adicionalmente, ao final de cada rodada, gerou-se um mapa da concentração dos teores de ferro no solo. Assim, para análise e discussão dos dados produziram-se mapas relacionados à média, à mediana, ao desvio padrão, ao coeficiente de variação e aos percentis mais baixo (5%) e mais alto (95%) dos valores dos teores de ferro nos solos do QF. Toda a modelagem foi realizada em software R (R Core Team, 2024).

3. RESULTADOS E DISCUSSÕES

3.1 Avaliações das performances dos modelos

A variação da performance das métricas (Figura 5) demonstrou a importância da repetição durante o processo de construção dos modelos de ML com diferentes subconjuntos, evitando resultados enviesados que não representam a realidade (Moquedace et al., 2024). Ao considerar o número de amostras para o treinamento e testes dos modelos, as métricas obtidas apresentaram desempenho satisfatório e acredita-se que com mais amostras, as métricas poderiam ter sido melhores. Os modelos também apresentaram diferença estatística significativa.

Figura 5 – Desempenho dos modelos de acordo com as diferentes métricas utilizadas para avaliação. Perceptível as diferenças estatísticas com exceção ao R². As letras “a”, “b” e “c” indicam diferenças estatísticas significantes entre os grupos com $p < 0,05$.



Fonte: Elaborado pelo autor.

O Cubist obteve os valores mais altos de MAE e RMSE (tabela 7). Esses resultados estão alinhados com os encontrados por Mello et al. (2022) e Mendes et al. (2022), ao predizerem os teores de ferro em solos de região com diversidade geológica. Apesar disso, o Cubist apresentou o melhor índice de concordância entre os valores preditos e observados. Esses valores corroboram estudos que exploram diferentes modelos de ML para predizer propriedades de solos em diversas regiões do mundo (Yang et al., 2016; Mulder et al., 2016; Taghizadeh-Mehrjardi et al., 2016; Wadoux, 2019; Dharumarajan et al., 2017; Siqueira et al., 2023; Moquedace et al., 2024).

Tabela 7 – Avaliação das métricas médias dos modelos após 100 rodadas.

Modelos	MAE	MAE NULL	LCCC	R ²	RMSE	RMSE NULL
Cubist	41,07	42,96	0,39	0,18	60,97	59,35
Earth	40,96	42,45	0,34	0,16	59,57	59,38
GBM	39,88	42,54	0,33	0,14	58,65	59,45
KNN	39,22	40,99	0,34	0,20	57,52	59,44
RF	39,06	43,07	0,36	0,20	56,66	59,45
svmRadial	37,68	38,92	0,24	0,18	54,34	59,71

Fonte: Elaborado pelo autor.

Para Khaledian e Miller (2020) as diferenças entre as métricas decorrem, principalmente, da quantidade de amostras no conjunto de dados, que influencia diretamente os cálculos dos erros e as suas variâncias. Assim, os valores de MAE e RMSE do Cubist indicam que seus erros tiveram magnitudes elevadas. Contudo, o LCCC por descrever a acurácia do modelo, indica que o modelo conseguiu capturar bem os padrões existentes nos dados. Esses resultados indicam que o Cubist subestima, constantemente, os valores preditos, todavia, os dados mantêm um padrão linear.

O svmRadial, por sua vez, apresentou comportamento inverso ao Cubist, registrando os valores mais baixos de MAE, RMSE e LCCC. Gomes et al. (2019) explicam que o svmRadial, ao buscar minimizar os erros, ignora os pontos corretamente classificados que estão distantes do limite de decisão, desta forma, os erros são reduzidos sem necessariamente otimizar a acurácia. Logo, o modelo pode ter um bom ajuste global, no entanto haverá pouca concordância ao ser avaliado. Ressalta-se que os resultados encontrados para o svmRadial, são indicativos de que o modelo não conseguiu capturar a real variação existente entre os dados.

Os modelos RF e KNN obtiveram os melhores valores de R², além de apresentarem bons valores de LCCC. Mello et al. (2025) ao utilizar o RF para modelar atributos geofísicos de solos tropicais, reportaram valores de R² que variaram entre 0,49 a 0,017. De maneira semelhante, Mendes et al. (2021) e Rosin et al. (2023) encontraram valores de R² variando de 0,62 a 0,17 ao modelar os óxidos de ferro em solos brasileiros. Todos os autores citados utilizaram diversas outras variáveis pedológicas em suas modelagens, o que pode explicar os melhores desempenhos em seus modelos. Apesar desse contraste, os nossos resultados estão dentro de uma variação aceitável.

De acordo com Mello et al. (2025), a boa performance dos modelos está relacionada à sua capacidade de generalização e de lidar com relações não-lineares entre as covariáveis. Em relação ao KNN, entende-se que sua boa performance é explicada por ser um método não-

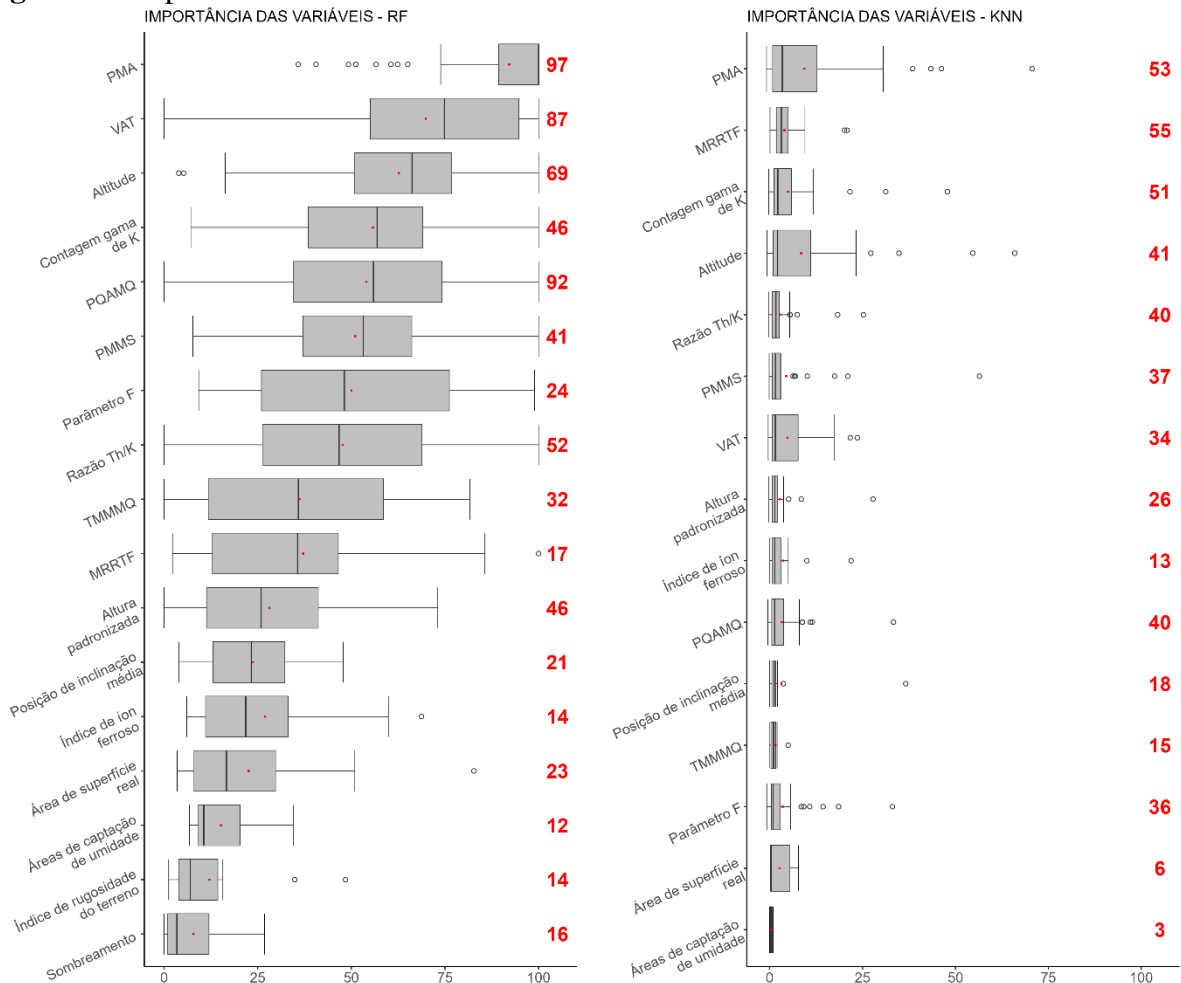
paramétrico que conseguiu representar muito bem as relações entre as variáveis. Além disso, pode-se afirmar que a litoestratigrafia foi bem representada, já que o KNN é muito sensível às desigualdades presentes nas classes dos objetos (Dragovic e Onjia, 2007).

Em relação aos modelos nulos, para a métrica MAE, todos os algoritmos foram superiores. Contudo, os modelos Cubist e Earth apresentaram valores de RMSE piores do que os modelos nulos. Mello et al. (2025) apontaram que modelos lineares, como o Earth, possuem limitações ao lidar com dados não-lineares. No caso do Cubist, seus resultados inferiores reforçam a necessidade de critérios mais rigorosos durante a construção do modelo.

Como destacado por Mello et al. (2022) e Fernandes-Filho et al. (2024), para que um modelo seja útil em MDS, ele deve apresentar desempenho melhor que modelo nulo, caso contrário, o seu custo computacional o torna inviável. Sendo assim, optou-se pelo RF e KNN para as previsões espaciais, por apresentarem boas métricas e por obterem desempenhos satisfatórios em relação aos modelos nulos.

3.2 Importância e frequência das covariáveis e suas relações com os teores de ferro

De maneira geral, os resultados indicaram que as covariáveis bioclimáticas foram mais frequentemente selecionadas pelo RFE, com destaque para a PMA, que se mostrou a covariável mais importante em ambos os modelos (Figura 6). Para o RF, além da PMA, outras covariáveis também apresentaram relevância significativa, como a VAT, a PMMS, a altitude, o MRRTF, a Contagem de ^{40}K e a razão $e\text{Th}/^{40}\text{K}$

Figura 6 – Importância das variáveis escolhidas de acordo com os modelos.

Fonte: Elaborado pelo autor.

Por outro lado, no modelo KNN, observou-se que a frequência das covariáveis selecionadas e suas importâncias relativas foram menores. Apesar disso, houve uma concordância entre as covariáveis selecionadas por ambos os modelos, reforçando o papel fundamental das covariáveis bioclimáticas, como a precipitação e temperatura, como agentes reguladores do intemperismo em áreas com a litoestratigrafia ferrífera.

De acordo com Schaefer et al. (2008), a concentração dos teores de ferro está diretamente relacionada à intensidade do intemperismo e às condições de drenagem, sendo que em regiões com sistemas de drenagem livre e temperaturas elevadas, a água atua como o principal agente intempérico, promovendo a remoção de sílica e a concentração de elementos residuais. Fernandes-Filho et al. (2022) destacaram que a amplitude térmica diária e anual são características em áreas ferríferas e Ker (1997) ressaltou que, apesar das elevadas precipitações e umidade relativa no QF, a mineralogia hematítica, herdada do material de origem, é o principal fator para a elevada concentração dos teores de ferro nesses solos.

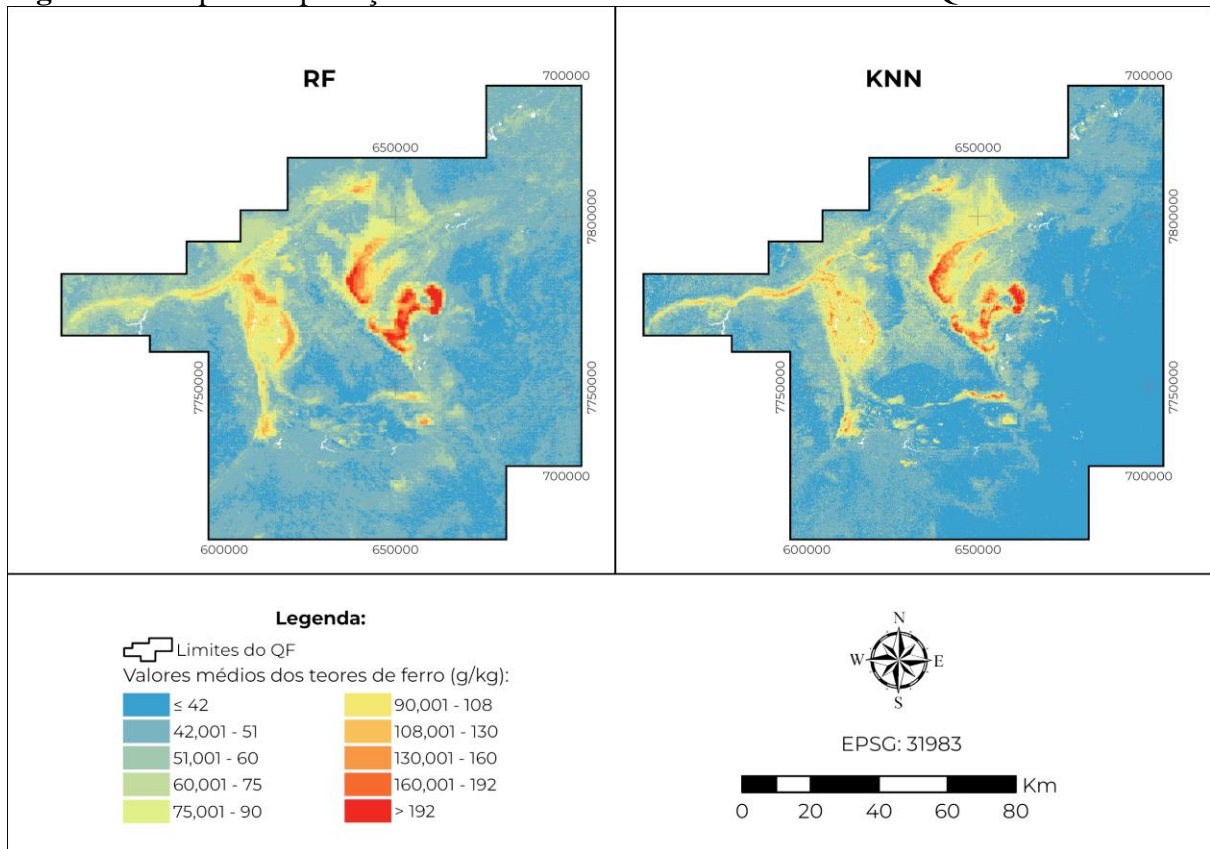
A frequente presença de covariáveis morfométricas, como a altitude e o MRRTF pode ser explicada pelas “ilhas de ferro”, áreas distribuídas nos topos das encostas, que consistem em platôs resistentes ao intemperismo situados acima de 900 metros de altitude (Jacobi e Carmo, 2008). Essas áreas, são conhecidas como Campos Rupestres Ferruginosos (CRF) e são caracterizados por mosaicos de alta diversidade edáfica que controlam comunidades vegetacionais associadas a afloramentos antigos de formações ferríferas bandadas (BIFs) (Schaefer et al., 2023; Figueira et al., 2016). Carvalho-Filho et al. (2010) complementaram que a resistência das BIFs ao intemperismo é fundamental para preservar as antigas estruturas geológicas em níveis topográficos mais suaves e elevados.

No que diz respeito às variáveis gamaespectrométricas, Gomes (2021) afirma que, geralmente, essas apresentam valores baixos nas BIFs do QF. Contudo, a elevada razão $eTh/^{40}K$ nas formações ferríferas encontrada em nossos resultados, pode ser atribuído aos seguintes fatores. Apesar da composição geoquímica das rochas não ser enriquecida nesses elementos (Mello et al., 2023a), o ^{40}K é um cátion altamente solúvel em condições de superfície, sendo facilmente lixiviado das partes mais altas da paisagem, principalmente em solos quimicamente intemperizados, com pH ácido e drenagem livre (Mello et al., 2021; 2022). Além disso, o eTh apresenta baixa mobilidade no solo, sendo facilmente adsorvido por argilas oxídicas (Mello et al., 2023a; Mello et al., 2023b).

3.3 Distribuição espacial dos teores de ferro, predições, incertezas e usos

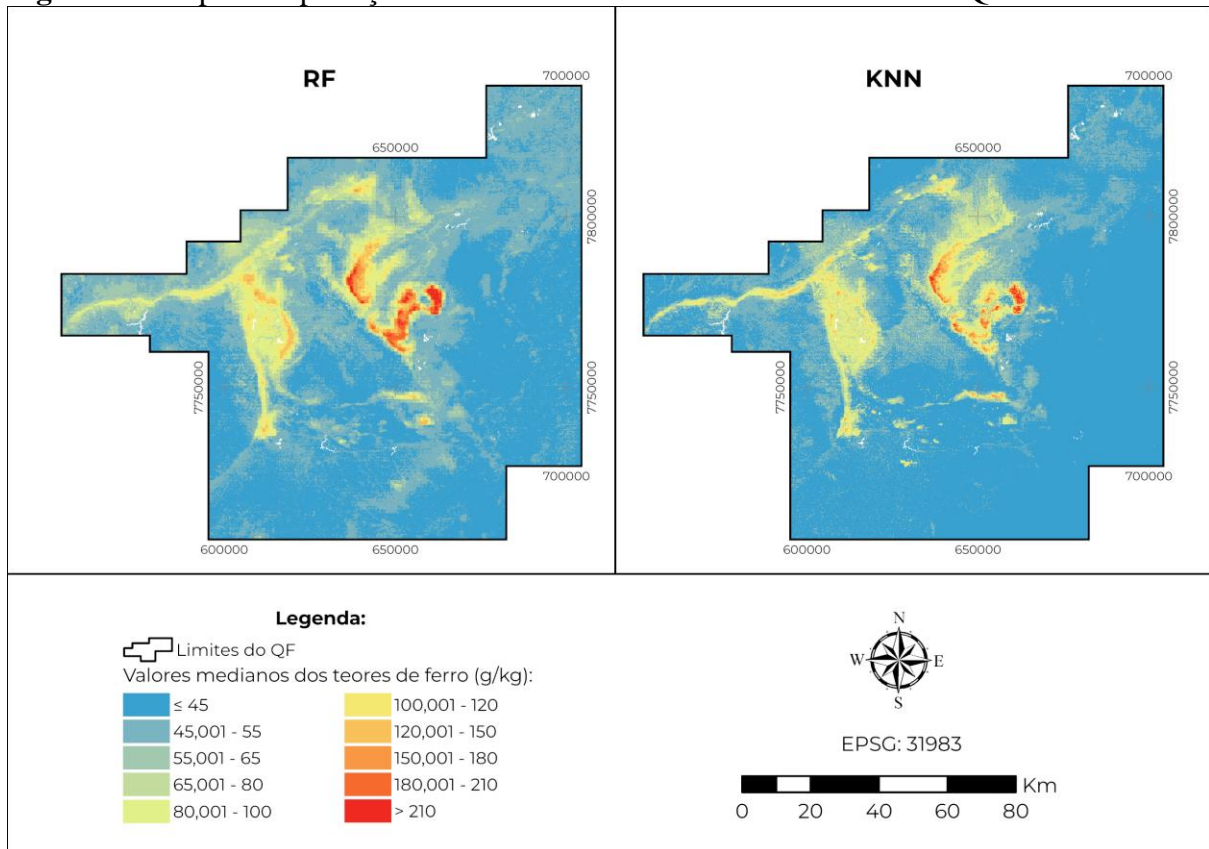
As figuras 7 e 8 destacam os mapas referentes as médias e medianas das 100 predições espaciais dos teores de ferro em solos do QF com os modelos selecionados.

Figura 7 – Mapas das predições médias dos teores de ferro em solos do QF.



Fonte: Elaborado pelo autor.

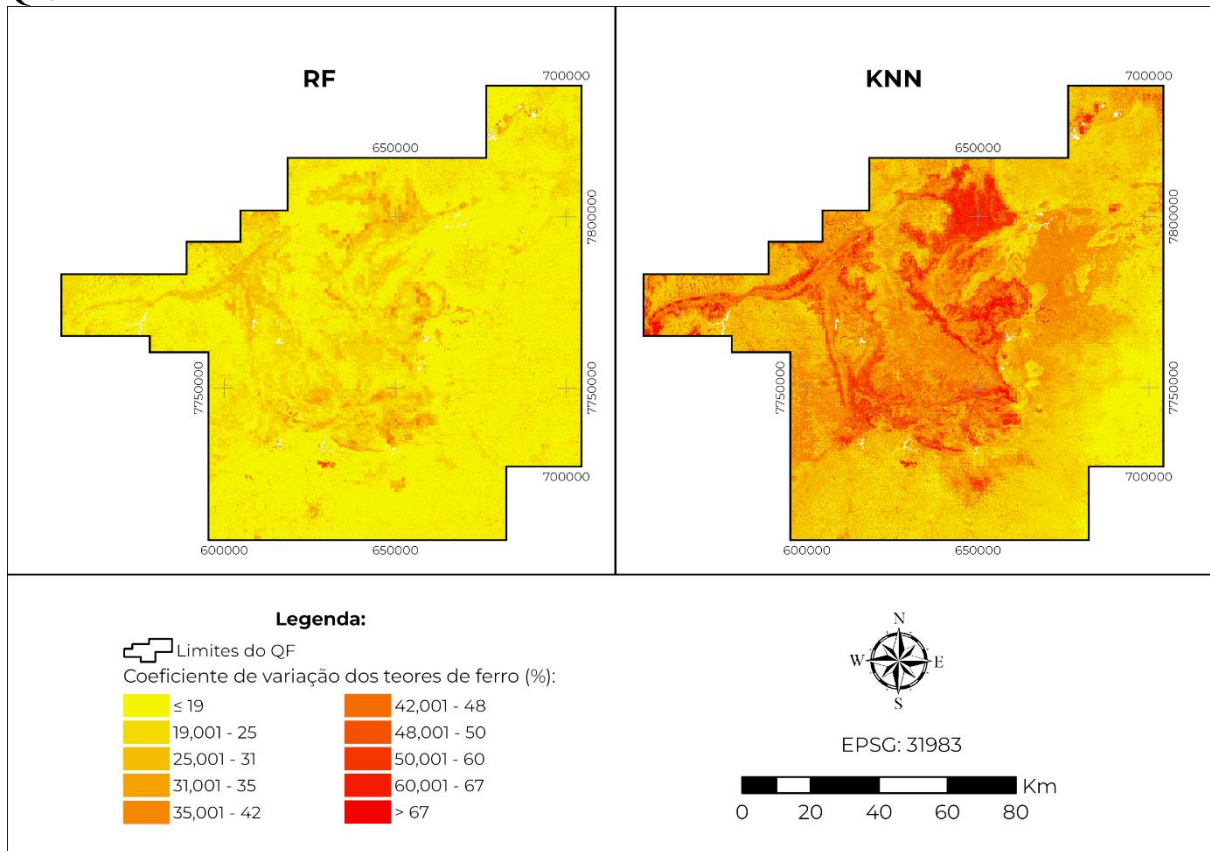
Figura 8 – Mapas das predições medianas dos teores de ferro em solos do QF.



Fonte: Elaborado pelo autor.

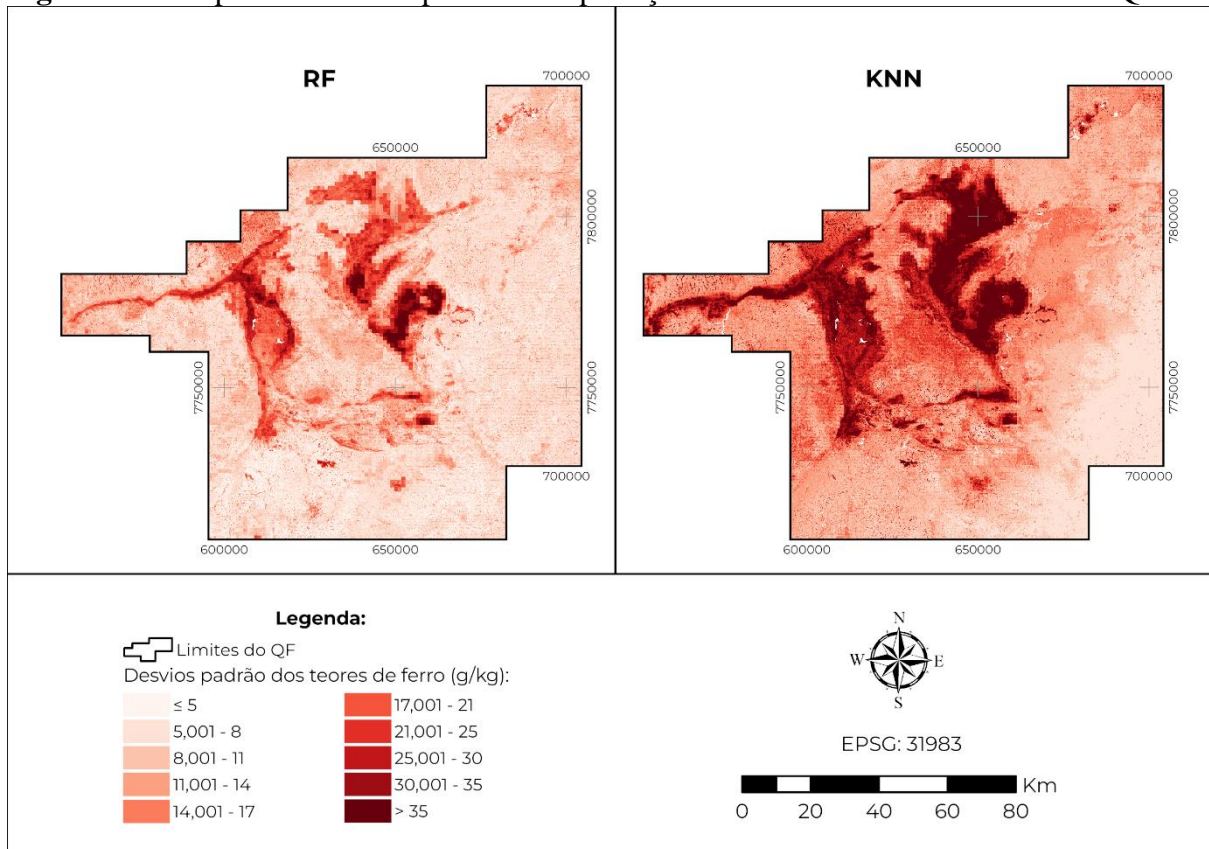
As figuras 9, 10, 11 e 12 destacam, respectivamente, os coeficientes de variação, desvios padrões e os quartis de 5% e 95% dos teores de ferro em solos do QF. De acordo com os resultados apresentados na tabela 8, o RF mostrou valores mais próximos à referência, indicando um desempenho superior na predição das concentrações dos teores de ferro. Além disso, nenhum dos modelos apresentou subestimação ou superestimação dos valores mínimos e máximos dos teores de ferro, o que indica que as predições se mantiveram próximas às médias e medianas observadas. Os desvios padrões e coeficientes de variação também foram semelhantes entre os modelos, sugerindo desempenho similar.

Figura 9 – Mapas dos coeficientes de variação das predições dos teores de ferro em solos do QF.



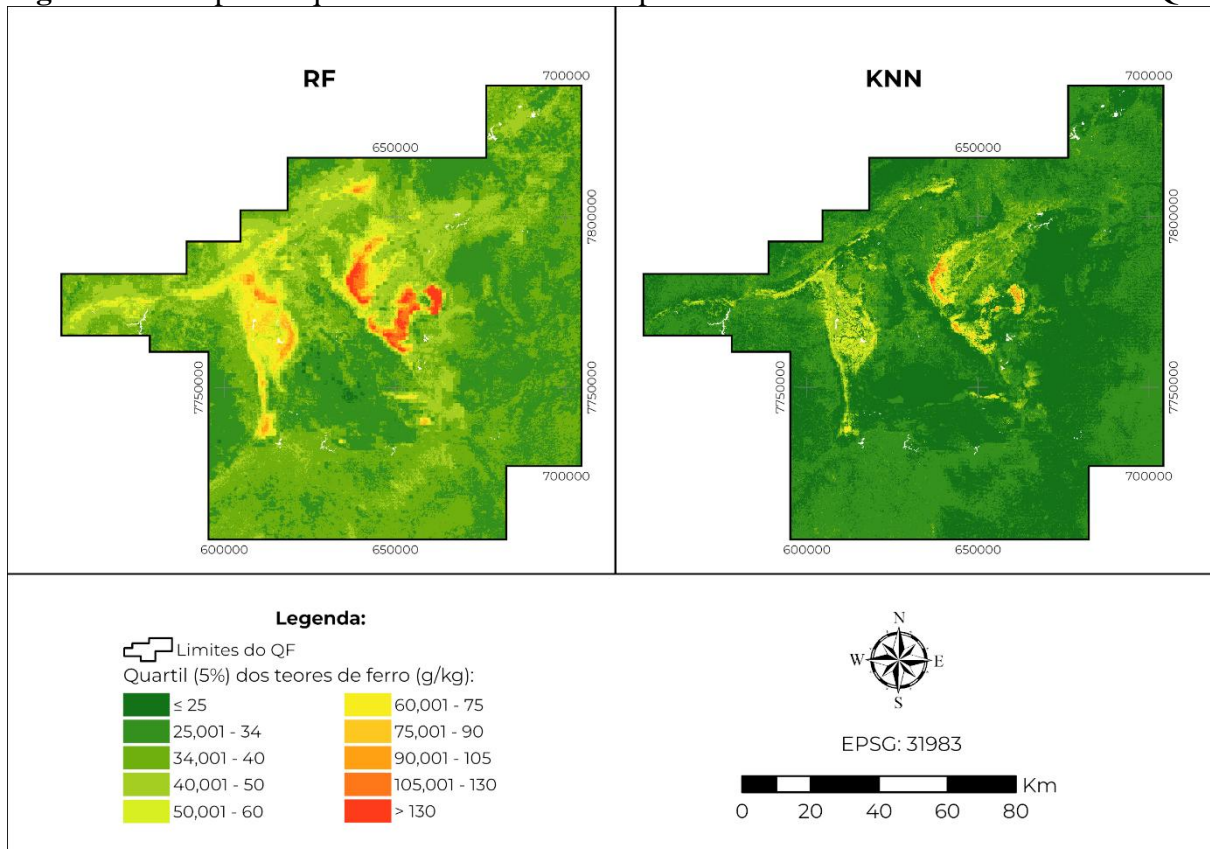
Fonte: Elaborado pelo autor.

Figura 10 – Mapas dos desvios padrões das predições dos teores de ferro em solos do QF.

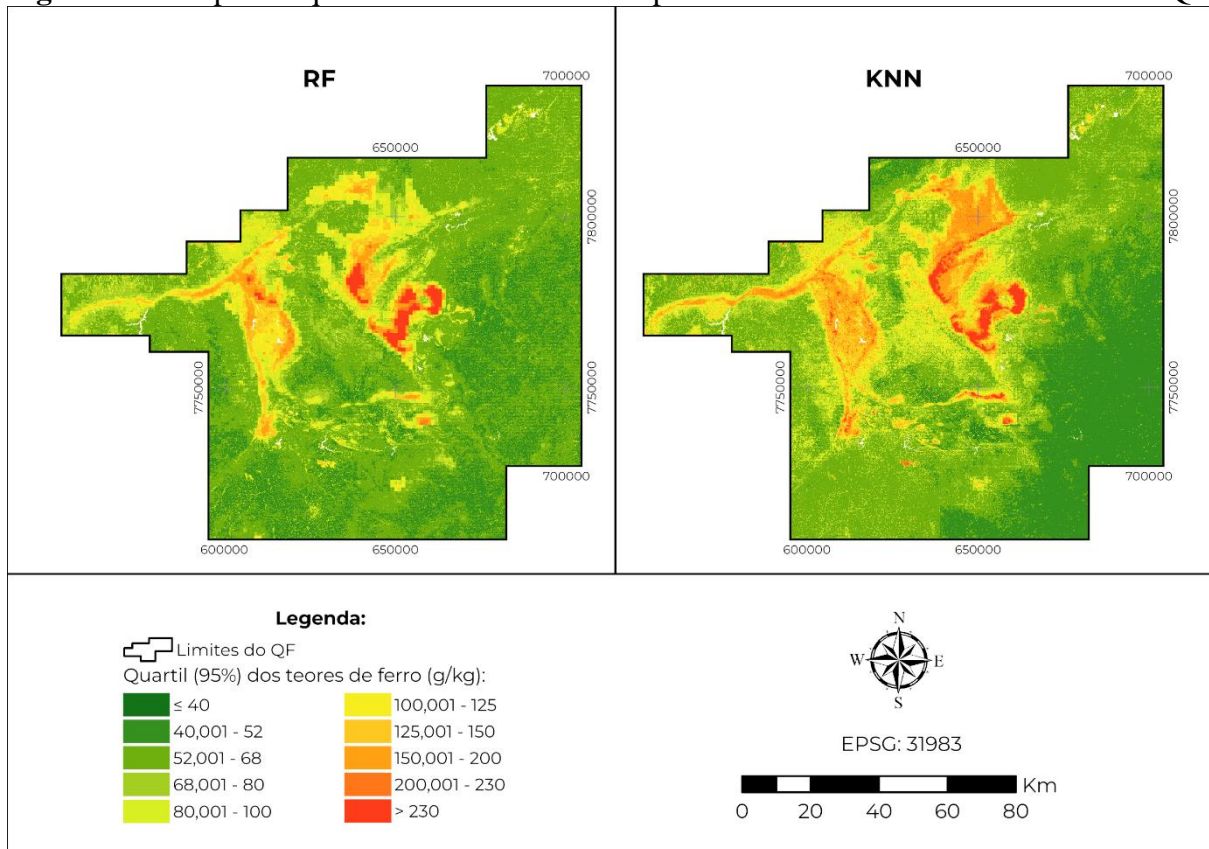


Fonte: Elaborado pelo autor.

Figura 11 – Mapas do quartil de 5% dos valores preditos dos teores de ferro em solos do QF.



Fonte: Elaborado pelo autor.

Figura 12 – Mapas do quartil de 95% dos valores preditos dos teores de ferro em solos do QF.

Fonte: Elaborado pelo autor.

Tabela 8 – Comparações entre valores preditos e observados entre os dois modelos preditores.

Valores	Fe 3051a (g)	Mapa médio		Mapa mediano	
	Observados	KNN	RF	KNN	RF
Mínimos	0,50	25,55	29,95	24,62	29,14
Medianos	46,34	43,13	46,99	42,20	46,06
Médios	65,62	49,37	53,47	47,67	52,36
Máximos	329,00	238,50	241,23	244,67	249,52
SD	60,75	20,50	21,56	19,18	21,26
CV	0,92	41,57	40,35	40,26	40,65
Q – 5%	9,11	33,39	38,91	32,99	38,17
Q – 95%	197,40	92,97	93,99	88,94	92,21

Fonte: Elaborado pelo autor.

Os mapas produzidos e os resultados apresentados corroboram com estudos anteriores, como o Carvalho-Filho et al. (2010), que identificaram que os maiores teores de ferro em solos do QF estão associados às formações ferríferas Moeda e Gandarela, presentes, na região centro-leste da área de estudo. Além disso, verificou-se que as áreas das bordas, sustentadas pelas BIFs e que delimitam o polígono do QF (Dorr, 1969, Gomes, 2021), apresentam altas concentrações de ferro. Essas regiões são responsáveis pela preservação dos platôs ferruginosos em altitudes

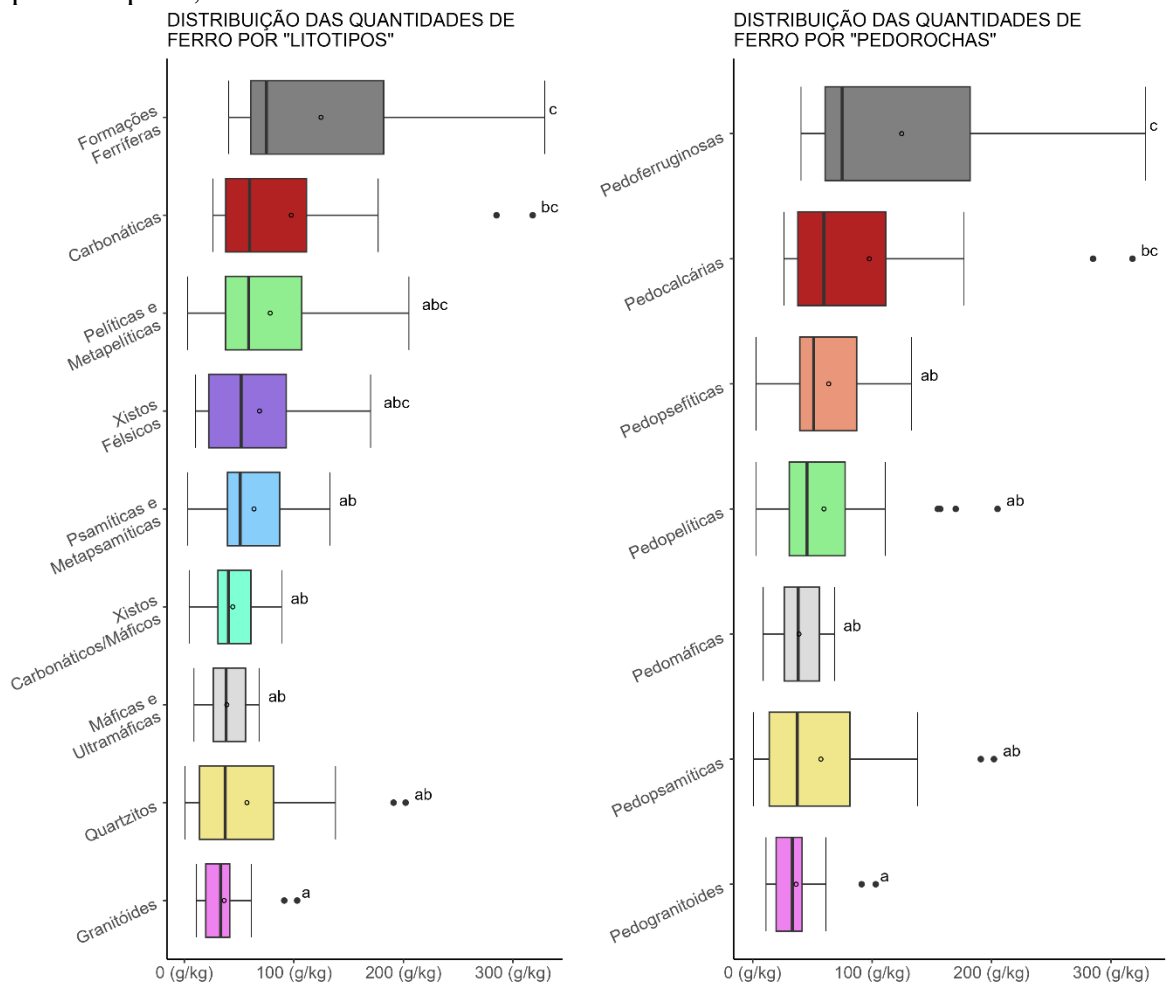
elevadas, uma vez que apresentam elevada resistência aos processos intempérico e se mantêm como superfícies residuais.

Essas regiões mais resistentes correspondentes aos CRF, fundamentais para a conservação da biodiversidade do QF e historicamente, ameaçadas pela exploração mineral (Jacobi e Carmo, 2008; Schaefer et al., 2015). Salienta-se que as áreas ferruginosas do QF são responsáveis pela manutenção e reabastecimento de importantes aquíferos. Portanto, a preservação desses ecossistemas é crucial não apenas para a conservação ambiental, mas também para a sustentabilidade ambiental de toda a região e suas áreas dependentes.

As menores concentrações dos teores de ferro foram observadas em áreas dominadas por terrenos graníticos-gnáissicos (“pedogranitoides”), típicos de rochas félsicas com menores influências de ambientes ferruginosos. Dessa forma, os modelos demonstraram boa capacidade de capturar os padrões de distribuição espacial dos teores de ferro em solos do QF, refletindo as condições ambientais, pedológicas e litoestratigráficas da região.

Embora a litoestratigrafia não tenha emergido como uma covariável de grande importância nos modelos analisados, ela exerce uma influência indireta sobre as concentrações dos teores de ferro, como evidenciado pelos padrões espaciais observados. Essa relação está associada à distribuição espacial dos grupos litoestratigráficos que determinam a disponibilidade de ferro nos solos (Figura 13).

Figura 13 – Distribuição dos teores de ferro em solos do QF agrupados para fins pedológicos. As letras “a”, “b”, “c” indicam diferenças estatísticas significantes entre os grupos entre os grupos com $p < 0,05$.



Fonte: Elaborado pelo autor.

A variação estatística significativa entre os agrupamentos para fins pedológicos, é indicativo de como os teores de ferro são bastante variáveis nas diferentes litologias do QF e como isso afeta diretamente a gênese e a sua presença nos solos. Para Rosin et al. (2025), a litoestratigrafia desempenha um papel essencial na distribuição dos teores de ferro em solos tropicais, devido à complexa interação entre os materiais de origem, as taxas de intemperismo e os processos de redistribuição.

4. CONCLUSÕES

O presente estudo avaliou a distribuição espacial dos teores de ferro em solos do QF por meio de diferentes modelos de ML. Dentre os algoritmos testados, o RF apresentou o melhor desempenho com destaque para as covariáveis bioclimáticas, morfométricas e de gamaespectroscopia como as mais relevantes na predição da distribuição dos teores de ferro.

Embora os modelos tenham atribuído baixa relevância a litoestratigrafia, os mapas e os gráficos produzidos sugerem que ela desempenha um papel significativo na distribuição dos teores de ferro em solos do QF e evidencia uma discrepância entre os resultados dos modelos e o comportamento observado. Isso reforça a necessidade de estudos futuros que integrem de forma mais detalhada as variáveis litoestratigráficas e pedológicas para aprimorar os resultados e explorar essa relação com maior profundidade.

Historicamente pressionado pela exploração mineral, o QF possui áreas ferríferas de elevada biodiversidade que são prioritárias à conservação. Os resultados deste estudo não apenas fornecem dados e informações essenciais para o planejamento e as tomadas de decisão, mas também evidenciam a importância de estratégias de uso sustentável, proteção e recuperação ambiental para toda a região.

Por fim, a aplicação de técnicas de MDS mostrou-se indispensável ao detalhamento de padrões espaciais dos solos do QF. Além de contribuir para uma gestão sustentável do território, a metodologia proposta neste trabalho pode ser aplicada a outras regiões com relevância socioambiental e econômica e promover uma contribuição mais ampla do entendimento da relação entre os solos e a paisagem.

REFERÊNCIAS

- Brasil. (2010). *Mapa de Geodiversidade do estado de Minas Gerais* (Vol. 12). <https://rigeo.sgb.gov.br/handle/doc/14704>
- Brenning, A. (2008). Statistical geocomputing combining R and SAGA: The example of landslide susceptibility analysis with generalized additive models. In J. Boehner, T. Blaschke, & L. Montanarella (Eds.), *SAGA - Seconds Out (= Hamburger Beitrage zur Physischen Geographie und Landschaftsoekologie)* (Vol. 19, pp. 23–32).
- Carvalho-Filho, A., Curi, N., Marques, J. J. G. S. M., Shinzato, E., de Jesus, E. A., Freitas, D. A. F., & Massahud, R. T. L. R. (2011). Óxidos de manganês em solos do quadrilátero ferrífero (MG). *Revista Brasileira de Ciência Do Solo*, 35, 793–804. <https://doi.org/10.1590/S0100-06832011000300015>
- Carvalho-Filho, A., Curi, N., & Shinzato, E. (2010). Relações solo-paisagem no Quadrilátero Ferrífero em Minas Gerais. In *Pesq. agropec. bras* (Issue 8). <https://doi.org/10.1590/S0100-204X2010000800017>
- Carvalho-Filho, A., Inda, A. V., Fink, J. R., & Curi, N. (2015). Iron oxides in soils of different lithological origins in Ferriferous Quadrilateral (Minas Gerais, Brazil). *Applied Clay Science*, 118, 1–7. <https://doi.org/10.1016/j.clay.2015.08.037>
- Carvalho-Júnior, W., Chagas, C. da S., Calderano-Filho, B., & Bhering, S. B. (2013). *Funções de pedotransferência para estimativa da densidade dos solos de áreas tropicais montanhosas*. https://www.researchgate.net/publication/279854424_Funcoes_de_pedotransferencia_para_estimativa_da_densidade_dos_solos_de_areas_tropicais_montanhasas

- Coelho, M. R.; Vasques, G. de M.; Tassinari, D.; Souza, Z. R. de.; Oliveira, A. P. de.; Moreira, F. M. de S. Solos do Quadrilátero Ferrífero sob diferentes coberturas vegetais e materiais de origem. *Boletim de pesquisa e desenvolvimento*, Embrapa Solos, Rio de Janeiro, 2017.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>
- Dharumarajan, S., Hegde, R., & Singh, S. K. (2017). Spatial prediction of major soil properties using Random Forest techniques - A case study in semi-arid tropics of South India. *Geoderma Regional*, 10, 154–162. <https://doi.org/10.1016/j.geodrs.2017.07.005>
- Diniz, J. M. F. de S., Dos Reis, A. A., Acerbi-Júnior, F. W., & Gomide, L. R. (2014). Detecção da expansão da área minerada no Quadrilátero Ferrífero, Minas Gerais, no período de 1985 a 2011 através de técnicas de sensoriamento remoto. *Boletim de Ciências Geodesicas*, 20(3), 683–700. <https://doi.org/10.1590/S1982-21702014000300039>
- Dorr, J. V. N. (1969). *Physiographic, Stratigraphic and Structural Development of the Quadrilátero Ferrífero*.
- Dragovic, S., & Onjia, A. (2007). Classification of soil samples according to geographic origin using gamma-ray spectrometry and pattern recognition methods. *Applied Radiation and Isotopes*, 65(2), 218–224. <https://doi.org/10.1016/j.apradiso.2006.07.005>
- EPA, U., & of Resource Conservation, O. (2007). *Method 3051A: Microwave Assisted Acid Digestion of Sediments, Sludges, Soils, and Oils, part of Test Methods for Evaluating Solid Waste, Physical/Chemical Methods*.
- Fernandes-Filho, E. I., Moquedace, C. M., Pereira, L. F., Veloso, G. V., & Carvalho-Júnior, W. (2024). Methods and Challenges in Digital Soil Mapping: Applied Modelling with R Examples. In *Pedometrics in Brazil* (pp. 263–283). https://doi.org/10.1007/978-3-031-64579-2_19
- Fernandes-Filho, E. I., Schaefer, C. E. G. R., Faria, R. M., Lopes, A. O., Francelino, M. R., & Gomes, L. C. (2022). The unique and endangered Campo Rupestre vegetation and protected areas in the Iron Quadrangle, Minas Gerais, Brazil. *Journal for Nature Conservation*, 66. <https://doi.org/10.1016/j.jnc.2022.126131>
- Figueira, J. E. C., Ribeiro, K. T., Ribeiro, M. C., Jacobi, C. M., França, H., Neves, A. C. O., Conceição, A. A., Mourão, F. A., Souza, J. M., & Miranda, C. A. K. (2016). Fire in Rupestrian Grasslands: Plant Response and Management. In *Ecology and Conservation of Mountaintop grasslands in Brazil* (pp. 415–448). Springer International Publishing. https://doi.org/10.1007/978-3-319-29808-5_18
- Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, 19(1). <https://doi.org/10.1214/aos/1176347963>
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5). <https://doi.org/10.1214/aos/1013203451>
- Gomes, D. G. da C. (2021). *Atlas aerogeofísico do estado de Minas Gerais*.
- Gomes, L. C., Faria, R. M., de Souza, E., Veloso, G. V., Schaefer, C. E. G. R., & Fernandes-Filho, E. I. (2019). Modelling and mapping soil organic carbon stocks in Brazil. *Geoderma*, 340, 337–350. <https://doi.org/10.1016/j.geoderma.2019.01.007>
- Hawker, L., Uhe, P., Paulo, L., Sosa, J., Savage, J., Sampson, C., & Neal, J. (2022). A 30 m global map of elevation with forests and buildings removed. *Environmental Research Letters*, 17(2). <https://doi.org/10.1088/1748-9326/ac4d4f>
- Hechenbichler, K. & Schliep, K. (2004). Weighted k-Nearest-Neighbor Techniques and Ordinal Classification. *Mol. Ecol.* 399, 17. <https://doi.org/10.5282/ubm/epub.1769>
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. r G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25(15), 1965–1978. <https://doi.org/10.1002/joc.1276>

- Jacobi, C. M., & Carmo, F. F. (2008). *Diversidade dos campos rupestres ferruginosos no Quadrilátero Ferrífero, MG*. Megadiversidade, vol 4. N 1-2.
- Jenny, H. (1941) Factors of Soil Formation: A System of Quantitative Pedology. Dover Publications, New York, 281 p.
- Ker, J. C. (1997). Latossolos do Brasil: Uma revisão. In *GEONOMOS* (Vol. 5, Issue 1). <https://doi.org/10.18285/geonomos.v5i1.187>
- Khaledian, Y., & Miller, B. A. (2020). Selecting appropriate machine learning methods for digital soil mapping. *Applied Mathematical Modelling*, 81, 401–418. <https://doi.org/10.1016/j.apm.2019.12.016>
- Kuhn, M., 2020. caret: Classification and Regression Training.
- Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling, Applied Predictive Modeling. Springer New York, New York, NY. <https://doi.org/10.1007/978-1-4614-6849-3>.
- Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. R News.
- McBratney, A. B., Mendonça-Santos, M. L., & Minasny, B. (2003). On digital soil mapping. *Geoderma*, 117(1–2), 3–52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)
- Mello, D. C., Demattê, J. A. M., Mello, F. A. O., Poppiel, R. R., Silvero, N. E. Q., Safanelli, J. L., Souza, A. B., Di Raimo, L. A. D. L., Rizzo, R., Resende, M. E. B., & Schaefer, C. E. G. R. (2021). Applied gamma-ray spectrometry for evaluating tropical soil processes and attributes. *Geoderma*, 381. <https://doi.org/10.1016/j.geoderma.2020.114736>
- Mello, D. C., Osório, T. F., Veloso, G. V., Lana, M. G., Mello, F. A. O., Di Raimo, L. A. D. L., Schaefer, C. E. G. R., Francelino, M. R., Fernandes-Filho, E. I., & Demattê, J. A. M. (2022). Pedogenetic processes operating at different intensities inferred by geophysical sensors and machine learning algorithms. *Catena*, 216. <https://doi.org/10.1016/j.catena.2022.106370>
- Mello, D. C., Souza, A. B., Mello, F. A. O., Marques, K. P. P., Poppiel, R. R., Belinasso, H., Di Raimo, L. A. D. L., Francelino, M. R., Fernandes-Filho, E. I., Veloso, G. V., Schaefer, C. E. G. R., & Demattê, J. A. M. (2023). Sensor-based field methods for pedology and soil surveys: Protocol suggestions for Brazilian tropical soils. *Geoderma Regional*, 33. <https://doi.org/10.1016/j.geodrs.2023.e00651>
- Mello, D. C., Veloso, G. V., Lana, M. G., Mello, F. A. O., Poppiel, R. R., Cabrero, D. R. O., Di Raimo, L. A. D. L., Schaefer, C. E. G. R., Fernandes-Filho, E. I., Leite, E. P., & Demattê, J. A. M. (2022). A new methodological framework for geophysical sensor combinations associated with machine learning algorithms to understand soil attributes. *Geoscientific Model Development*, 15(3), 1219–1246. <https://doi.org/10.5194/gmd-15-1219-2022>
- Mello, D. C., Veloso, G. V., Mello, M. F., Lana, M. G., de Angeli, I. O., Mello, F. A. O., Siqueira, R. G., Gomes, L. C., Fernandes-Filho, E. I., Schaefer, C. E. G. R., Francelino, M. R., Leite, E. P., Ferreira, T. O., & Demattê, J. A. M. (2025). Integrating proximal geophysical sensing and machine learning for digital soil mapping: Spatial prediction and model evaluation using a small dataset. *Soil Advances*, 3, 100024. <https://doi.org/10.1016/j.soilad.2024.100024>
- Mello, D. C., Veloso, G. V., Moquedace, C. M., de Angeli, I. O., Francelino, M. R., Oliveira, F. S., Souza, J. J. L. L., Gomes, L. C., Schaefer, C. E. G. R., Medeiros-Júnior, E., Fernandes-Filho, E. I., & Demattê, J. A. M. (2023). Chemical weathering detection in the periglacial landscapes of Maritime Antarctica: New approach using geophysical sensors, topographic variables and machine learning algorithms. *Geoderma*, 438. <https://doi.org/10.1016/j.geoderma.2023.116615>
- Mello, D. C., Veloso, G. V., Moquedace, C. M., de Angeli, I. O., Oliveira, F. S., Gomes, L. C., Souza, J. J. L. L., Francelino, M. R., Fernandes-Filho, E. I., Schaefer, C. E. G. R., & Demattê, J. A. M. (2023). Radiometric and magnetic susceptibility characterization of

- soil profiles: Geophysical data and their relationship with Antarctic periglacial processes, pedogenesis, and lithology. *Catena*, 232. <https://doi.org/10.1016/j.catena.2023.107427>
- Mendes, W. de S., Demattê, J. A. M., Minasny, B., Silvero, N. E. Q., Bonfatti, B. R., Safanelli, J. L., Rizzo, R., & Costa, A. C. S. da. (2022). Free iron oxide content in tropical soils predicted by integrative digital mapping. *Soil and Tillage Research*, 219. <https://doi.org/10.1016/j.still.2022.105346>
- Mendes, W. S., Demattê, J. A. M., Bonfatti, B. R., Resende, M. E. B., Campos, L. R., & Costa, A. C. S. (2021). A novel framework to estimate soil mineralogy using soil spectroscopy. *Applied Geochemistry*, 127. <https://doi.org/10.1016/j.apgeochem.2021.104909>
- Mendonça-Santos, M. L., & dos Santos, H. G. (2006). The State of the Art of Brazilian Soil Mapping and Prospects for Digital Soil Mapping. In *Developments in Soil Science* (Vol. 31, Issue C). Elsevier Ltd. [https://doi.org/10.1016/S0166-2481\(06\)31003-3](https://doi.org/10.1016/S0166-2481(06)31003-3)
- 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. (2024). High-resolution mapping of soil carbon stocks in the western Amazon. *Geoderma Regional*, 36. <https://doi.org/10.1016/j.geodrs.2024.e00773>
- Mulder, V. L., Lacoste, M., Richer-de-Forges, A. C., Martin, M. P., & Arrouays, D. (2016). National versus global modelling the 3D distribution of soil organic carbon in mainland France. *Geoderma*, 263, 16–34. <https://doi.org/10.1016/j.geoderma.2015.08.035>
- Pereira, L. F., Moquedace, C. M., Rosolem, G. P. N., de Sousa, M. da C., Francelino, M. R., & Fernandes-Filho, E. I. (2024). Applying Machine Learning Techniques to Model and Map Soil Surface Texture Using Limited Legacy Data. In *Pedometrics in Brazil* (pp. 31–45). https://doi.org/10.1007/978-3-031-64579-2_3
- Polidoro, J. C., Mendonça-Santos, M. D. L., Lumbrreras, J. F., Coelho, M. R., Carvalho-Filho, A., Da Motta, P. E. F., Carvalho-Júnior, W., Araújo-Filho, J. C., Cúrcio, G. R., Correia, J. R., Martins, E. D. S., Spera, S. T., Oliveira, S. R. D. M., Bolfe, E. L., Manzatto, C. V., Tosto, S. G., Venturieri, A., Sa, I. B., Oliveira, V. A., Shinzato, E., Anjos, L. H. C., Valladares, G. S., Ribeiro, J. L., Medeiros, P. S. C., Moreira, F. M. D. S., Silva, L. S. L., Sequinato, L., Aoglio, M. L. D., Dart, R. D. O., 2016. Programa Nacional de Solos do Brasil (PronaSolos), 1st ed. Embrapa Solos, Rio de Janeiro, RJ.
- Rahman, M.M.; Zhang, X.; Ahmed, I.; Iqbal, Z.; Zeraatpisheh, M.; Kanzaki, M.; Xu, M. Remote Sensing-Based Mapping of Senescent Leaf C:N Ratio in the Sundarbans Reserved Forest Using Machine Learning Techniques. *Remote Sens.* 2020, 12, 1375. <https://doi.org/10.3390/rs12091375>
- R Core Team. (2024). *A language and environment for statistical computing*. <https://www.R-project.org/>
- Resende, M., Ker, J. C., de Rezende, S. B., Poggere, G., Zimmer, J. B., Mancini, M., & Curi, N. (2023). *Roteiro pedológico de Minas Gerais: Conhecendo e interpretando os principais solos de diferentes ambientes*. https://www.researchgate.net/publication/370124938_ROTUIRO_PEDOLOGICO_DE_MINAS_GERAIS_Conhecendo_e_interpretando_os_principais_solos_de_diferentes_a_mmbientes
- Rosin, N. A., Demattê, J. A. M., Poppiel, R. R., Silvero, N. E. Q., Rodriguez-Albarracin, H. S., Rosas, J. T. ., Greschuk, L. T., Bellinaso, H., Minasny, B., Gomez, C., Marques-Júnior, J., & Fernandes, K. (2023). Mapping Brazilian soil mineralogy using proximal and remote sensing data. *Geoderma*, 432. <https://doi.org/10.1016/j.geoderma.2023.116413>
- Rosin, N. A., Mello, D. C., Bonfatti, B. R., Hartemink, A. E., Ferreira, T. O., Silvero, N. E. Q., Poppiel, R. R., Mendes, W. S., Veloso, G. V., Francelino, M. R., Alves, M. R., Falcioni, R., & Demattê, J. A. M. (2025). Mapping soil thickness using a mechanistic model and

- machine learning approaches. *CATENA*, 249, 108621. <https://doi.org/10.1016/j.catena.2024.108621>
- Santos, M. do C., Varajão, A. F. D. C., & Yvon, J. (2004). Genesis of clayey bodies in Quadrilátero Ferrífero, Minas Gerais, Brazil. *Catena*, 55(3), 277–291. [https://doi.org/10.1016/S0341-8162\(03\)00106-1](https://doi.org/10.1016/S0341-8162(03)00106-1)
- Shinzato, E. & Carvalho-Filho, A. Projeto APA Sul RMBH: Estudos do meio físico, Pedologia, Belo Horizonte: CPRM/EMBRAPA/SEMAD, 2005.
- Schaefer, C. E. G. R., Cândido, H. C., Corrêa, G. R., Pereira, A., Nunes, J. A., Souza, O. F., Marins, A., Fernandes-Filho, E. I., & Ker, J. C. (2015). Solos desenvolvidos sobre canga ferruginosa no Brasil: Uma revisão crítica e papel ecológico de termiteiros. In F. F. Carmo & L. H. Y. Kamino (Eds.), *Geossistemas ferruginosos no Brasil* (pp. 77–102).
- Schaefer, C. E. G. R., Candido, H. G., Corrêa, G. R., Campos, P. V., Senra, E. O., Gjorup, D. F., & Fernandes-Filho, E. I. (2023). Soils of Campos Rupestres (Rupestrian Grasslands) of the Old Brazilian Mountain Ranges. In *Ecology and Conservation of Mountaintop grasslands in Brazil* (pp. 343–362). https://doi.org/10.1007/978-3-031-19949-3_12
- Schaefer, C. E. G. R., Corrêa, G. R., Candido, H. G., Arruda, D. M., Nunes, J. A., Araujo, R. W., Rodrigues, P. M. S., Fernandes-Filho, E. I., Pereira, A. F. S., Brandão, P. C., & Neri, A. V. (2016). The Physical Environment of Rupestrian Grasslands (Campos Rupestres) in Brazil: Geological, Geomorphological and Pedological Characteristics, and Interplays. In *Ecology and Conservation of Mountaintop grasslands in Brazil* (pp. 15–53). Springer International Publishing. https://doi.org/10.1007/978-3-319-29808-5_2
- Schaefer, C. E. G. R., Fabris, J. D., & Ker, J. C. (2008). Minerals in the clay fraction of Brazilian Latosols (Oxisols): a review. *Clay Minerals*, 43(1), 137–154. <https://doi.org/10.1180/claymin.2008.043.1.11>
- Schliep, K., Hechenbichler, K., 2016. kkn: Weighted k-Nearest Neighbors.
- Schwertmann, U. & Taylor, R.M. (1989) Iron Oxides. In: Dixon, J.B. and Weed, S.B., Eds., Minerals in Soil Environments, Second Edition, Soil Science Society America, Madison.
- Siqueira, R. G., Moquedace, C. M., Fernandes-Filho, E. I., Schaefer, C. E. G. R., Francelino, M. R., Sacramento, I. F., & Michel, R. F. M. (2024). Modelling and prediction of major soil chemical properties with Random Forest: Machine learning as tool to understand soil-environment relationships in Antarctica. *Catena*, 235. <https://doi.org/10.1016/j.catena.2023.107677>
- Siqueira, R. G., Moquedace, C. M., Francelino, M. R., Schaefer, C. E. G. R., & Fernandes-Filho, E. I. (2023a). Machine learning applied for Antarctic soil mapping: Spatial prediction of soil texture for Maritime Antarctica and Northern Antarctic Peninsula. *Geoderma*, 432. <https://doi.org/10.1016/j.geoderma.2023.116405>
- Siqueira, R. G., Moquedace, C. M., Francelino, M. R., Schaefer, C. E. G. R., & Fernandes-Filho, E. I. (2023b). Machine learning applied for Antarctic soil mapping: Spatial prediction of soil texture for Maritime Antarctica and Northern Antarctic Peninsula. *Geoderma*, 432. <https://doi.org/10.1016/j.geoderma.2023.116405>
- Souza, J. J. L. L., Abrahão, W. A. P., de Mello, J. W. V., da Silva, J., da Costa, L. M., & de Oliveira, T. S. (2015). Geochemistry and spatial variability of metal(loid) concentrations in soils of the state of Minas Gerais, Brazil. *Science of the Total Environment*, 505, 338–349. <https://doi.org/10.1016/j.scitotenv.2014.09.098>
- Taghizadeh-Mehrjardi, R., Nabiollahi, K., & Kerry, R. (2016). Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran. *Geoderma*, 266, 98–110. <https://doi.org/10.1016/j.geoderma.2015.12.003>
- Teixeira, P. C., Donagemma, G. K., Fontana, A., & Teixeira, W. G. (2017). *Manual de métodos de análise de solo - Embrapa*.

- ten Caten, A., Dalmolin, R. S. D., Mendonça-Santos, M. L., & Giasson, E. (2012). *Mapeamento digital de classes de solos: características da abordagem brasileira*. <https://doi.org/10.1590/S0103-84782012001100013>
- Viscarra Rossel, R. A., Bui, E. N., De Caritat, P., & McKenzie, N. J. (2010). Mapping iron oxides and the color of Australian soil using visible-near-infrared reflectance spectra. *Journal of Geophysical Research: Earth Surface*, 115(4). <https://doi.org/10.1029/2009JF001645>
- Wadoux, A. M. J. C. (2019). Using deep learning for multivariate mapping of soil with quantified uncertainty. *Geoderma*, 351, 59–70. <https://doi.org/10.1016/j.geoderma.2019.05.012>
- Wadoux, A. M. J. C., Minasny, B., & McBratney, A. B. (2020). Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Science Reviews*, 210, 103359. <https://doi.org/10.1016/j.earscirev.2020.103359>
- Yang, R. M., Zhang, G. L., Liu, F., Lu, Y. Y., Yang, F., Yang, F., Yang, M., Zhao, Y. G., & Li, D. C. (2016). Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem. *Ecological Indicators*, 60, 870–878. <https://doi.org/10.1016/j.ecolind.2015.08.036>
- Zeraatpisheh, M., Ayoubi, S., Jafari, A., Tajik, S., & Finke, P. (2019). Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. *Geoderma*, 338, 445–452. <https://doi.org/10.1016/j.geoderma.2018.09.006>

CONSIDERAÇÕES FINAIS

As técnicas de MDS permitiram construir modelos preditivos, destacados como FPT e mapear os teores de ferro em solos do QF. A partir das avaliações das incertezas, pode-se dizer que o RF foi o melhor modelo por apresentar resultados mais consistentes para as FPT, tanto com e sem a SM. Além disso, ao mapear os teores de ferro, conseguiu captar muito bem as relações entre as covariáveis e gerou os mapas mais condizentes.

Para ambos os capítulos, observou-se a predominância das covariáveis bioclimáticas, como a precipitação e a temperatura, bem como covariáveis morfométricas, com destaque para a altitude. Além disso, as covariáveis da gamaespectrometria se mostraram importantes e recorrentes na seleção dos subconjuntos pelo RFE. Em relação às FPT, a presença de covariáveis pedológicas teve influência elevada, com destaque para a SM, como a variável mais importante. Os modelos construídos se mostraram capazes de determinar os teores de ferro, com a vantagem da não geração de resíduos, como em análises convencionais.

A presença da SM e sua relação direta com os teores de ferro indicam que a inclusão de outras covariáveis ajudou a reduzir os ruídos nos modelos. Em relação ao mapeamento, observou-se que a precipitação foi a variável mais importante e frequentemente selecionada. A presença de água no solo é fundamental para que todas as reações físico-químicas e biológicas ocorram e a dinâmica dos teores de ferro é altamente dependente dessas reações.

Por fim, observou-se que os modelos raramente selecionaram as informações das classificações litoestratigráficas e das pedorochas como variáveis importantes. No entanto, essas covariáveis são essenciais por representarem as fontes naturais de ferro no solo. A análise dos mapas revelou que as maiores concentrações de ferro ocorrem em áreas ferríferas do QF, regiões que abrigam alta biodiversidade e que estão sob constante ameaça devido à atividade minerária. Nesse contexto, todos os mapas gerados constituem dados e informações valiosas para subsidiar decisões voltadas à conservação e à proteção da biodiversidade nas áreas ferríferas do QF.