



UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
INSTITUTO DE FLORESTAS
CURSO DE GRADUAÇÃO EM ENGENHARIA FLORESTAL

JOYCE FERREIRA ALVES DOS SANTOS

**PREDIÇÃO ESPACIAL DAS FRAÇÕES GRANULOMÉTRICAS DO SOLO COM
BASE VARIÁVEIS TOPOGRÁFICAS, ESPECTRAIS E APRENDIZADO DE
MÁQUINA – ESTUDO DE CASO: PROJETO SALITRE - BA**

Prof. Dr. HELENA SARAIVA KOENOW PINHEIRO
Orientador

SEROPÉDICA, RJ
DEZEMBRO – 2025



UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
INSTITUTO DE FLORESTAS
CURSO DE GRADUAÇÃO EM ENGENHARIA FLORESTAL

JOYCE FERREIRA ALVES DOS SANTOS

**PREDIÇÃO ESPACIAL DAS FRAÇÕES GRANULOMÉTRICAS DO SOLO COM
BASE VARIÁVEIS TOPOGRÁFICAS, ESPECTRAIS E APRENDIZADO DE
MÁQUINA – ESTUDO DE CASO: PROJETO SALITRE - BA**

Monografia apresentada ao Curso de Engenharia Florestal, como requisito parcial para a obtenção do Título de Engenheiro Florestal, Instituto de Florestas da Universidade Federal Rural do Rio de Janeiro.

Prof. Dr. HELENA SARAIVA KOENOW PINHEIRO
Orientador

SEROPÉDICA, RJ
DEZEMBRO – 2025



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
DEPARTAMENTO DE SOLOS



HOMOLOGAÇÃO Nº 29 / 2025 - DeptS (12.28.01.00.00.00.33)

Nº do Protocolo: 23083.069932/2025-66

Seropédica-RJ, 03 de dezembro de 2025.

**PREDIÇÃO ESPACIAL DAS FRAÇÕES GRANULOMÉTRICAS DO SOLO COM
BASE EM VARIÁVEIS TOPOGRÁFICAS, ESPECTRAIS E APRENDIZADO DE
MÁQUINA – ESTUDO DE CASO: PROJETO SALITRE- BA**

JOYCE FERREIRA ALVES DOS SANTOS

APROVADA EM: 02/12/2025

BANCA EXAMINADORA:

(Assinado digitalmente em 03/12/2025 11:22)

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

Prof. Dra. HELENA SARAIVA KOENOW PINHEIRO – UFRRJ

Orientador

(Assinado digitalmente em 03/12/2025 12:01)

LEONARDO DURVAL DUARTE GUIMARAES
PROFESSOR ENS BASICO TECN TECNOLOGICO
DeptS (12.28.01.00.00.00.33)
Matrícula: 3720511

Prof. Dr. LEONARDO DURVAL DUARTE GUIMARÃES – UFRRJ

Membro

(Assinado digitalmente em 03/12/2025 15:50)

WALDIR DE CARVALHO JUNIOR
ASSINANTE EXTERNO
CPF: 583.287.696-20

Dr. WALDIR DE CARVALHO JÚNIOR – EMBRAPA-SOLOS

Membro

Visualize o documento original em <https://sipac.ufrj.br/public/documentos/index.jsp>
informando seu número: **29**, ano: **2025**, tipo: **HOMOLOGAÇÃO**, data de emissão: **03/12/2025** e
o código de verificação: **41c3671c40**

RESUMO

O mapeamento digital de solos (MDS) tem se consolidado como uma ferramenta fundamental para estimar atributos do solo a partir de dados ambientais e espectrais, especialmente em regiões com falta de levantamentos detalhados, como o semiárido brasileiro. Diante desse cenário, este estudo avaliou a predição das frações granulométricas do solo (areia, silte e argila) a partir da integração de variáveis topográficas e dados espectrais dos sensores OLI/OLI-2 (Landsat 8 e 9), na área do Projeto Salitre, Bahia. Foram processadas covariáveis topográficas (declividade, curvaturas, índices hidrológicos) e espectrais (NDVI, NDMI, SAVI, OSAVI, BSI e bandas do espectro VIS–NIR–SWIR), associadas a 398 pontos amostrais da camada superficial do solo. Três modelos preditivos foram testados: Random Forest (RF), Regressão Linear Múltipla (RLM) e Árvore de Regressão (AR), sendo que o RF apresentou o melhor desempenho para areia e argila (R^2 de validação entre 0,52 e 0,58), refletindo a forte relação entre textura superficial e resposta espectral, enquanto a RLM apresentou desempenho intermediário e a AR contribuiu principalmente para a interpretação dos limiares ambientais. Os resultados indicam que variáveis espectrais, sobretudo índices de umidade e bandas SWIR, são os melhores preditores das frações granulométricas, ao passo que o silte apresentou baixa previsibilidade em todos os modelos. Assim, o estudo demonstrou o potencial prático do uso integrado de sensoriamento remoto e técnicas de *machine learning* para apoiar o mapeamento das frações granulométricas do solo em ambientes semiáridos, contribuindo para o planejamento agrícola e a gestão sustentável do uso da terra.

Palavras-chave:

Modelo Digital de Solos; sensoriamento remoto; *machine learning*.

ABSTRACT

Digital soil mapping (DSM) has become a key tool for estimating soil properties based on environmental and spectral data, particularly in regions with a lack of detailed soil surveys, such as the Brazilian semi-arid. In this context, this study evaluated the prediction of soil textural fractions (sand, silt, and clay) through the integration of topographic variables and spectral data from the OLI/OLI-2 sensors (Landsat 8 and 9) in the Salitre Project area, Bahia, Brazil. Topographic covariates (slope, curvatures, and hydrological indices) and spectral covariates (NDVI, NDMI, SAVI, OSAVI, BSI, and VIS–NIR–SWIR bands) were processed and associated with 398 soil surface samples. Three predictive models were tested: Random Forest (RF), Multiple Linear Regression (MLR), and Regression Tree (RT). RF showed the best performance for sand and clay (validation R^2 ranging from 0.52 to 0.58), reflecting the strong relationship between surface soil texture and spectral response, while MLR showed intermediate performance and RT mainly contributed to the interpretation of environmental thresholds. The results indicate that spectral variables, particularly moisture-related indices and SWIR bands, are the most effective predictors of soil textural fractions, whereas silt showed low predictability across all models. Overall, this study demonstrates the practical potential of integrating remote sensing data and machine learning techniques to support mapping soil particle size fractions in semi-arid environments, contributing to agricultural planning and sustainable land-use management.

Keywords: Digital Soil Mapping; remote sensing ; *machine learning*.

SUMÁRIO

1. INTRODUÇÃO	1
2. REVISÃO BIBLIOGRÁFICA	2
2.1 Avanços recentes em pedometria e Mapeamento Digital de Solos no Brasil	2
2.2 Importância da granulometria e variabilidade dos solos	3
2.3 Métodos estatísticos e de <i>machine learning</i> aplicados ao MDS	4
3. MATERIAL E MÉTODOS	6
3.1 Caracterização geral da área do estudo de caso	6
3.2 Covariáveis ambientais	15
3.3 Modelagem preditiva da composição granulométrica	19
3.3.1 Seleção e diagnóstico de colinearidade entre covariáveis	19
3.3.2 Modelagem por Random Forest (RF)	20
3.3.3 Modelagem por Árvore de Regressão (AR)	23
3.3.4 Modelagem Regressão Linear Multipla (RLM)	24
4. RESULTADOS E DISCUSSÃO	24
4.1 Distribuição granulométrica e relação com o relevo	24
4.2 Distribuição granulométrica e relação com covariáveis espectrais e topográficas	28
4.3 <i>Random Forest</i>	30
4.4 Regressão Linear Multipla (RLM)	34
4.5 Arvore De Regressão	36
4.6 Comparação entre modelos	39
5. CONCLUSÃO	42
6. REFERÊNCIAS BIBLIOGRÁFICAS	43
APENDICES	50
APENDICE A – Árvores de Regressão para Areia	50
APENDICE B – Árvores de Regressão para Silte	51
APENDICE C – Árvores de Regressão para Argila	52

LISTA DE TABELAS

Tabela 1 - Descrição das Covariáveis Ambientais _____	17
Tabela 2 - Resumo das principais etapas, softwares, ferramentas, plataformas, resoluções e algoritmos utilizados no processamento dos dados _____	18
Tabela 3 - Pacotes utilizados no Random Forest e suas funções _____	21
Tabela 4 - Variáveis significativamente correlacionadas com cada fração granulométrica__	30
Tabela 5 - Resultados numéricos dos modelos RF, RLM e AR_____	41

LISTA DE FIGURAS

Figura 1 - Localização da área de estudo e pontos de amostragem no município de Juazeiro, Estado da Bahia.	8
Figura 2 -Precipitação média anual (1970–2000) na área do Projeto Salitre, Juazeiro, Bahia	9
Figura 3 - Cobertura vegetal da área de estudo no município de Juazeiro, Bahia	10
Figura 4 -Mapa de solos da área de estudo no município de Juazeiro, Bahia	11
Figura 5 - Mapa de declividade da área de estudo no município de Juazeiro, Bahia	12
Figura 6 - Mapa hipsométrico da área de estudo no município de Juazeiro, Bahia.	13
Figura 7 - Mapa litológico da área de estudo no município de Juazeiro, Bahia.	14
Figura 8 - Boxplot dos teores de areia, silte e argila por classes altimétricas e médias.	26
Figura 9 - Matriz de Correlação Spearman para Granulometria x Relevo	27
Figura 10 - Matriz de Correlação Spearman para Covariáveis Ambientais	28
Figura 11 - Gráficos de Resultados do RF para Areia, Silte e Argila	32
Figura 12 - Importância das covariáveis para Areia, Argila e Silte em Random Forest	33
Figura 13 - Gráficos de Resultados do RLM para Areia, Silte e Argila	35
Figura 14 - Gráficos de Resultados do AR para Areia, Silte e Argila	38
Figura 15 - Mapas Preditos dos modelos de RF, RLM e AR para Areia, Silte e Argila	40

1. INTRODUÇÃO

O avanço das geotecnologias tem impulsionado o mapeamento digital de solos (MDS), se consolidando como uma abordagem prática e eficiente para a coleta e interpretação de dados ambientais. Essas tecnologias integram inteligência artificial, algoritmos preditivos e informações espectrais e topográficas que, associadas a modelos matemáticos e estatísticos, permitem a predição de atributos físicos e químicos do solo, ampliando a compreensão da relação solo-paisagem e contribuindo para uma gestão ambiental mais precisa (Pinheiro, 2015; Bhering et al., 2016). Apesar desses avanços, o Brasil ainda carece de mapeamentos detalhados em grande parte do seu território, o que impõe limitações ao planejamento ambiental e ao desenvolvimento sustentável, especialmente em regiões sob intensa pressão de uso do solo (Crespolini e Nascimento, 2021).

Nesse contexto, as regiões semiáridas brasileiras assumem elevada importância ambiental e socioeconômica, uma vez que concentram extensas áreas de uso agrícola e pecuário. Caracterizadas por baixos índices pluviométricos, elevada variabilidade temporal das chuvas e intensa radiação solar, essas regiões estão sujeitas a condições climáticas restritivas que condicionam o uso e a conservação dos recursos naturais. Nesses ambientes, o solo desempenha papel central no suporte aos sistemas produtivos e na manutenção dos ecossistemas, tornando o conhecimento e o mapeamento de seus atributos fundamentais para o planejamento territorial (Lima et al., 2024).

Entretanto, em ambientes semiáridos de relevo predominantemente plano, a predição da granulometria do solo torna-se particularmente desafiadora, em função da complexidade da resposta espectral das frações granulométricas e do baixo contraste topográfico característico dessas paisagens.

No cenário brasileiro, o estudo desenvolvido por Chagas et al. (2016), no Projeto Salitre, evidenciou a aplicabilidade do MDS em regiões semiáridas por meio do uso de algoritmos de *machine learning* e métodos estatísticos. Utilizando dados do sensor TM/Landsat 5 e modelos como *Random Forest* e Regressão Linear Múltipla, os autores obtiveram resultados promissores para a predição da granulometria superficial do solo em áreas planas, sobretudo para as frações de areia e argila. Todavia, apesar de sua relevância metodológica, o estudo se baseou em um conjunto restrito de variáveis espectrais, abrindo possibilidade para expansão metodológica.

Tomando como referência o trabalho de Chagas et al. (2016), o presente estudo propõe a ampliação dessa abordagem. Assim, os objetivos são: (i) adotar dados dos sensores OLI e OLI-2, a bordo dos satélites Landsat 8 e 9, na predição da composição das frações granulométricas do solo; (ii) investigar diferentes combinações de bandas e índices espectrais associados à vegetação, à umidade e ao solo exposto (NDVI, NDMI, BSI, SAVI e OSAVI), bem como razões entre bandas (B4/B7) e bandas isoladas (B3, B4, B5, B6 e B7); (iii) analisar a contribuição de variáveis derivadas de modelos digitais de elevação; e (iv) comparar a robustez e a precisão dos métodos de modelagem Random Forest, Regressão Linear Múltipla e incluir Árvores de Regressão.

2. REVISÃO BIBLIOGRÁFICA

2.1 Avanços recentes em pedometria e Mapeamento Digital de Solos no Brasil

A pedometria tem se consolidado como um campo central da ciência do solo, oferecendo fundamentos quantitativos, qualitativos e computacionais para a compreensão e predição da variabilidade espacial dos atributos do solo. Estudos clássicos estabeleceram suas bases conceituais (McBratney et al., 2003; Lagacherie e McBratney, 2007), as quais têm sido complementadas por abordagens que integram geostatística, *machine learning* e dados provenientes de múltiplas fontes.

Avanços recentes incluem a utilização de sensores de proximidade, espectroscopia vis-NIR, fluorescência portátil de raios-X, drones equipados com LiDAR e sensores orbitais ópticos, permitindo análises mais rápidas, detalhadas dos atributos do solo (Dijair et al., 2020; Carvalho Junior et al., 2020; Marchioro, 2023). Paralelamente, plataformas em nuvem, como o Google Earth Engine, ampliaram a capacidade de processamento de grandes volumes de dados e análises multitemporais, favorecendo a aplicação do Mapeamento Digital de Solos em diferentes escalas (Yang et al., 2022).

Esses avanços têm impulsionado estudos voltados à avaliação da eficácia do MDS na predição remota de atributos do solo. Mgohele et al. (2024) destacam tendências metodológicas recentes. Zhou, Wu e Liu (2022) compararam diferentes cenas orbitais para a estimativa da textura do solo, discutindo a contribuição de bandas espectrais específicas e as limitações associadas à cobertura vegetal. De forma complementar, Mtshawu (2021) analisou a viabilidade e os desafios de dados de sensores orbitais na predição da granulometria da camada superficial do solo.

No contexto brasileiro, como citado anteriormente, o estudo de Chagas et al. (2016) se destaca, desenvolvido no âmbito do Projeto Salitre, demonstrou a aplicabilidade do MDS em regiões semiáridas por meio do uso de algoritmos de *machine learning* e métodos estatísticos. Utilizando dados do sensor Landsat 5 TM e modelos como *Random Forest* e Regressão Linear Múltipla, os autores obtiveram resultados significativos na predição da granulometria superficial do solo em áreas planas, especialmente para as frações de areia e argila.

Mais recentemente, a consolidação da pedometria no Brasil foi reforçada pela publicação da obra *Pedometrics in Brazil*, que reúne aplicações nacionais, discute metodologias adaptadas às condições edafoclimáticas do país e evidencia o protagonismo da comunidade científica brasileira no avanço do MDS (Carvalho Junior et al., 2024).

Diante disso, o mapeamento digital de solos surge como uma solução promissora, integrando diversas tecnologias para otimizar a obtenção e processamento de dados. Esses avanços permitem não apenas a melhoria na gestão territorial, mas também oferecem ferramentas robustas para a predição de características essenciais dos solos. No próximo tópico, exploraremos como o mapeamento digital de solos desempenha um papel crucial na predição de atributos do solo, contribuindo para análises mais eficientes e precisas.

2.2 Importância da granulometria e variabilidade dos solos

Os atributos físicos e químicos do solo são fundamentais para determinar sua qualidade e capacidade de sustentar atividades como agricultura, construção civil e conservação ambiental. Segundo Alves et al. (2021), o mapeamento desses atributos tem se tornado cada vez mais relevante para quantificar e monitorar recursos naturais, permitindo identificar áreas suscetíveis à erosão e orientar práticas conservacionistas.

Entre esses atributos, a granulometria se destaca como elemento-chave para o mapeamento da capacidade de uso e o manejo adequado do solo. Definida pela porcentagem de partículas primárias (areia, silte e argila) presentes na fração mineral (Centeno et al., 2017), ela constitui insumo crucial para modelos hidrológicos, climáticos e ambientais, além de funções de pedotransferência que estimam propriedades como densidade, condutividade hidráulica e capacidade de retenção de água (Ferreira et al., 2022).

A distribuição dessas partículas influencia diretamente a retenção de água e nutrientes, a infiltração e drenagem, a aeração, o teor de carbono orgânico, a capacidade de troca de cátions e a porosidade, além de controlar funções mecânicas essenciais do solo (Bhering et al., 2016).

A variabilidade granulométrica pode ser explicada tanto por fatores intrínsecos, relacionados a processos pedogenéticos (material de origem, intemperismo de minerais, acúmulo de matéria orgânica), quanto por fatores extrínsecos, ligados ao uso e manejo do solo, como práticas agrícolas, espécies vegetais, que influenciam processos de erosão, remoção seletiva de partículas e compactação (Alves et al., 2021).

Essa composição também varia com a profundidade: em geral, a matéria orgânica diminui nas camadas mais profundas, enquanto o teor de argila tende a aumentar. Além disso, os volumes de ar e água oscilam de acordo com a precipitação, irrigação e manejo hídrico (Stamford et al., 2005).

Em escala de paisagem, a distribuição dos solos reflete a interação entre microclima, pedogênese, relevo e processos geológicos superficiais. O relevo exerce papel central por influenciar o fluxo de água e a redistribuição de materiais (Sirtoli et al., 2008) e que pode ser representado por modelos numéricos de superfície, como o Modelo Digital de Elevação (MDE).

Atributos do relevo podem ser derivados de um MDE, sendo classificados em primários (declividade, orientação de vertentes, elevação, perfil de curvatura, comprimento de fluxo) e secundários (índice de umidade, índice de transporte de sedimentos, entre outros), fornecendo parâmetros fundamentais para compreender a dinâmica solo-paisagem.

No contexto semiárido, a dinâmica hídrica limitada acentua o papel das frações granulométricas na organização dos processos superficiais, influenciando diretamente a infiltração, o armazenamento de água e a estabilidade do solo. Araújo et al. (2010) destacam que variações no teor de areia e argila condicionam não apenas a disponibilidade hídrica, mas também o comportamento espectral das superfícies expostas, aspecto fundamental para a modelagem remota de atributos em ambientes com forte variação climática.

Logo, essa complexidade sistêmica reforça a necessidade de modelos capazes de prever essas propriedades no espaço, fundamentando o uso dos métodos estatísticos e de *machine learning* apresentados a seguir.

2.3 Métodos estatísticos e de *machine learning* aplicados ao MDS

O Mapeamento Digital de Solos (MDS) utiliza métodos matemáticos e estatísticos para construir funções de predição espacial de solos com erros autocorrelacionados, que representam as relações quantitativas entre propriedades do solo e características ambientais (Rosa et al., 2012). Sua eficácia, entretanto, depende da qualidade dos dados de entrada, como resolução espacial, covariáveis ambientais explicativas, densidade amostral e capacidade dos modelos em

capturar a variação espacial dos atributos (Dalmolin et al., 2015). Um dos principais desafios é selecionar as variáveis mais relevantes, pois essa escolha impacta diretamente a robustez e a precisão das predições (Alves et al., 2021).

Nesse contexto, os modelos estatísticos e de aprendizado de máquina têm se destacado por sua capacidade de capturar tanto relações lineares quanto padrões complexos e não lineares entre atributos do solo e variáveis ambientais, preenchendo lacunas que métodos tradicionais não conseguem resolver (Escovedo; Koshiyama, 2020). Enquanto abordagens lineares, como a Regressão Linear Múltipla (RLM), oferecem maior interpretabilidade, métodos não paramétricos, como Árvores de Regressão (AR) e Random Forest (RF), proporcionam maior acurácia frente à heterogeneidade dos dados, sendo amplamente empregados em estudos de MDS.

Diversas pesquisas têm demonstrado a eficácia desses métodos: Pereira et al. (2018) aplicaram diferentes algoritmos de *machine learning*, incluindo Árvores de Regressão, para mapear textura e atributos do solo em ambiente tropical, evidenciando a vantagem de modelos não lineares frente à variabilidade espacial; Sayão et al. (2018) utilizaram Regressão Linear Múltipla na predição de atributos físicos do solo a partir de covariáveis ambientais; e Hengl et al. (2017) mostraram que abordagens como o Random Forest ampliam significativamente a precisão e a cobertura espacial de propriedades do solo em mapeamentos digitais.

A Regressão Linear Múltipla (RLM) é um método tradicional que estima uma variável dependente (y) como combinação linear de variáveis independentes presentes na matriz X . É adequada para situações em que as relações são lineares, sem forte ruído ou multicolinearidade. Pode perder eficiência quando há muitas variáveis pouco relevantes, incorporando variância desnecessária ao modelo (Sena et al., 2000).

A Árvore de Regressão (AR) organiza os dados em divisões hierárquicas, partindo de um nó inicial (raiz) até nós terminais (folhas). Cada divisão é definida por regras de decisão baseadas em valores das variáveis preditoras, permitindo classificar padrões complexos de forma simples e interpretável. No entanto, árvores individuais podem apresentar baixa estabilidade e sensibilidade a pequenas variações nos dados (De Melo Figueiredo; De Carvalho, 2006).

O Random Forest (RF), por sua vez, busca superar essas limitações ao combinar muitas árvores construídas de forma independente, a partir de subconjuntos aleatórios de variáveis e amostras. O resultado é obtido pela média dos resultados das árvores, reduzindo o risco de sobreajuste (*overfitting*) e aumentando a precisão do modelo. Trata-se de um método não

paramétrico, robusto e amplamente utilizado em MDS pela sua capacidade de lidar com variáveis de diferentes naturezas e capturar relações não lineares (Carvalho Junior et al., 2016).

Como citado anteriormente, neste estudo serão abordados três métodos: a Regressão Linear Múltipla (RLM), considerada um modelo paramétrico, que assume uma relação linear entre a variável dependente e as variáveis independentes; a Árvore de Regressão (AR) e o Random Forest (RF), ambos modelos não paramétricos, que não pressupõem distribuição específica das variáveis alvo. Esses últimos apresentam vantagens adicionais, como a seleção automática de subconjuntos de variáveis e a capacidade de lidar simultaneamente com dados quantitativos e categóricos, possibilitando integrar, por exemplo, atributos do relevo, índices de sensoriamento remoto, informações geológicas ou classes de uso e cobertura do solo (Ferreira, 2022).

3. MATERIAL E MÉTODOS

3.1 Caracterização geral da área do estudo de caso

O estudo foi realizado na área do Projeto de Irrigação Salitre localizado em Juazeiro, Estado da Bahia. A área selecionada abrange aproximadamente 35.000 hectares e foram coletados 398 perfis de solo, amostrados na camada superficial de 0-20 cm, concedidos ao projeto pela Companhia de Desenvolvimento dos Vales do São Francisco e do Parnaíba (Codevasf) (Fig. 1). Os pontos foram distribuídos espacialmente por toda a área de estudo e posteriormente divididos aleatoriamente em conjuntos de calibração (80%, 318 perfis) e validação (20%, 80 perfis).

Segundo Chagas et al. (2016), a determinação da granulometria das amostras foi realizada em laboratório por meio do método do densímetro (hidrómetro), seguindo as orientações do Manual de Métodos de Análise de Solo da Embrapa (1979). As frações foram separadas conforme os limites granulométricos: areia (2–0,05 mm), silte (0,05–0,002 mm) e argila (< 0,002 mm). O processo utilizou hexametáfosfato de sódio ou hidróxido como agente dispersante, conforme descrito no protocolo analítico.

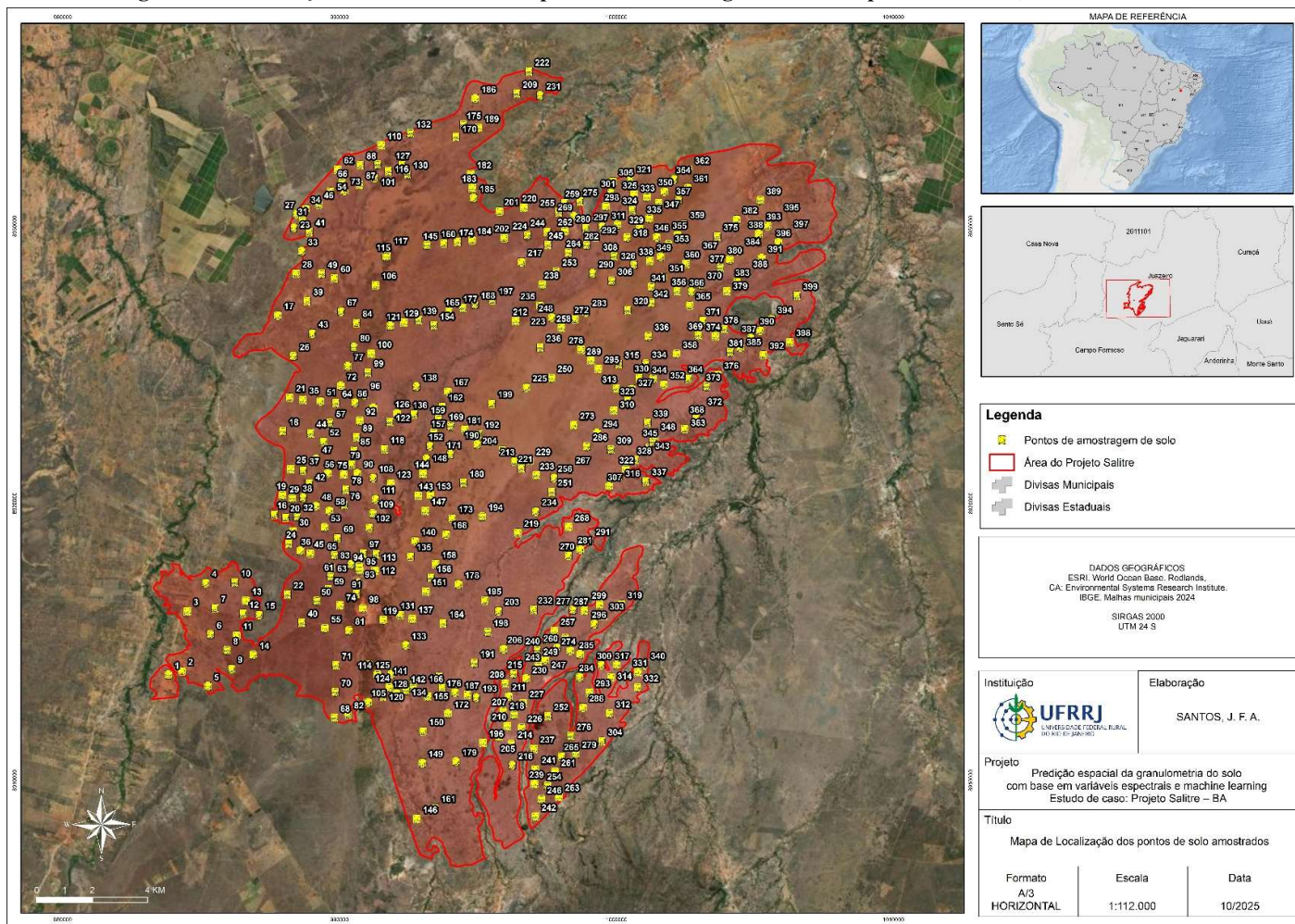
De acordo com a classificação climática de Köppen, o clima da região é BSwb' (semiárido, com invernos secos e verões chuvosos; temperatura média do mês mais frio ≥ 18 °C). A precipitação anual média varia entre aproximadamente 450 e 470 mm (Fig. 2), e a estação chuvosa ocorre entre novembro e abril, sendo março o mês mais úmido, conforme o WorldClim v2.1 (Fick & Hijmans, 2017), que representa médias climatológicas do período 1970–2000.

Segundo o Banco de Dados de Informações Ambientais (BDIA) do Instituto Brasileiro de Geografia e Estatística – IBGE (2023), a área de estudo apresenta predominância das fitofisionomias Savana Estépica Arborizada e Savana Estépica Parque, além de áreas de uso agropecuário (Fig. 3). Do ponto de vista pedológico, os principais solos identificados são Vertissolos e Cambissolos, conforme o Mapa de Solos do Brasil em escala 1:250.000 (IBGE, 2023) (Fig. 4). Essa distribuição reflete a distribuição das frações granulométricas no local: os Vertissolos, geralmente associados às posições mais rebaixadas da paisagem, apresentam maiores teores de argila e forte expansão e contração, influenciando a dinâmica hídrica e a estabilidade superficial; já os Cambissolos, mais comuns em porções elevadas, tendem a apresentar maior conteúdo arenoso, condicionando maior susceptibilidade à erosão e maior variabilidade textural.

A análise da declividade, obtida a partir do Modelo Digital de Elevação (SRTM, resolução espacial de 30 m), permitiu caracterizar o relevo da área (Fig. 5). As declividades variam entre 3% e 14%, com predominância de valores inferiores a 5%, evidenciando relevo majoritariamente plano. Além disso, o mapa hipsométrico da área de estudo (Fig. 6) apresenta altitudes variando entre 410 e 520 m, o que permitiu distinguir três classes altimétricas principais: áreas baixas (410–440 m), intermediária (440–470 m) e áreas elevadas (470–520 m).

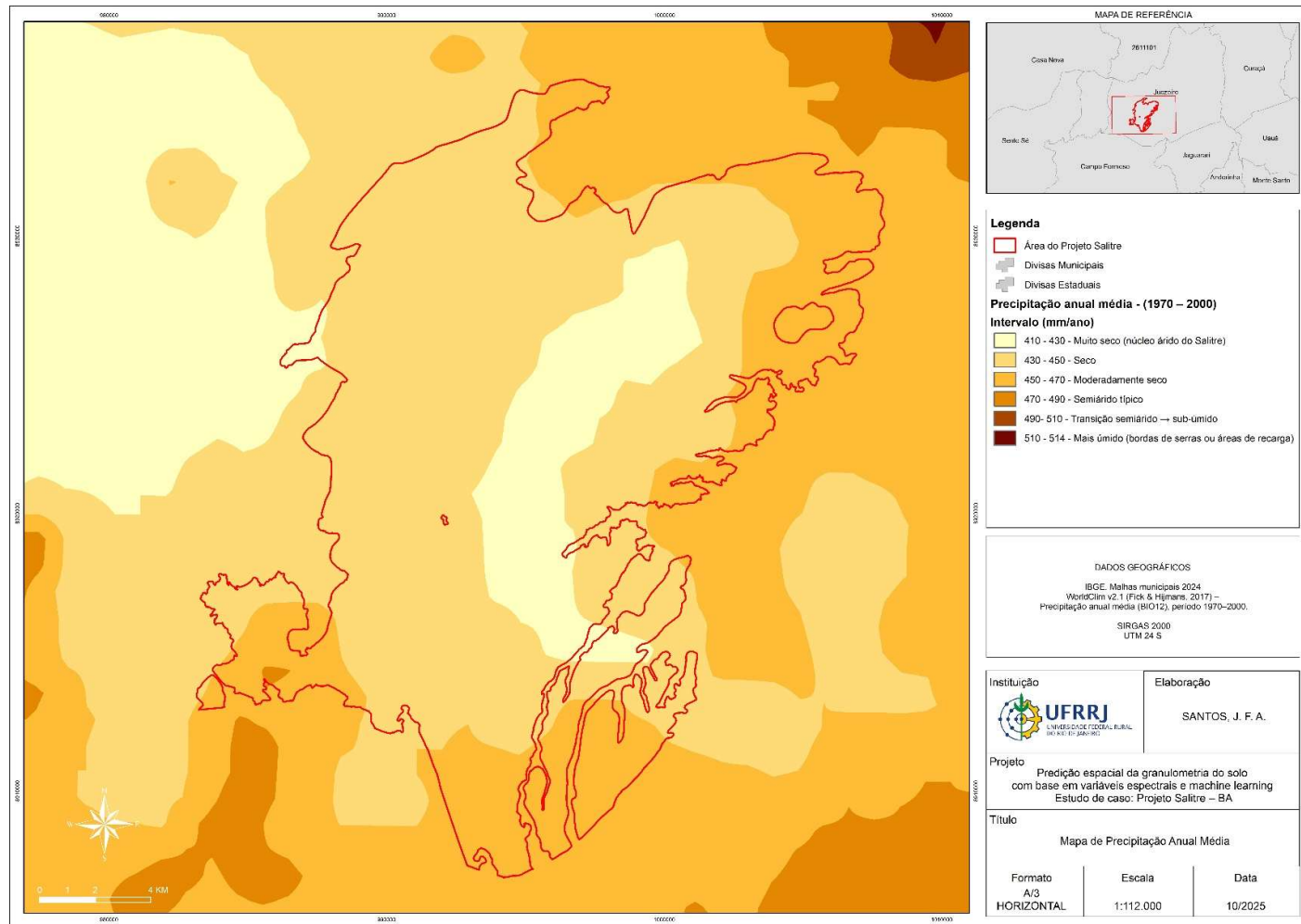
As informações litológicas foram obtidas a partir do Mapa Geológico do Estado da Bahia (CPRM/SGB, 2003), na escala 1:1.000.000 e as unidades identificadas na área incluem majoritariamente crosta calcária (calcrete) e rochas do Complexo metamórfico e metassedimentar, depósitos arenosos e sílticos recentes, e rochas metaultramáficas pontuais (Fig. 7).

Figura 1 - Localização da área de estudo e pontos de amostragem no município de Juazeiro, Estado da Bahia.



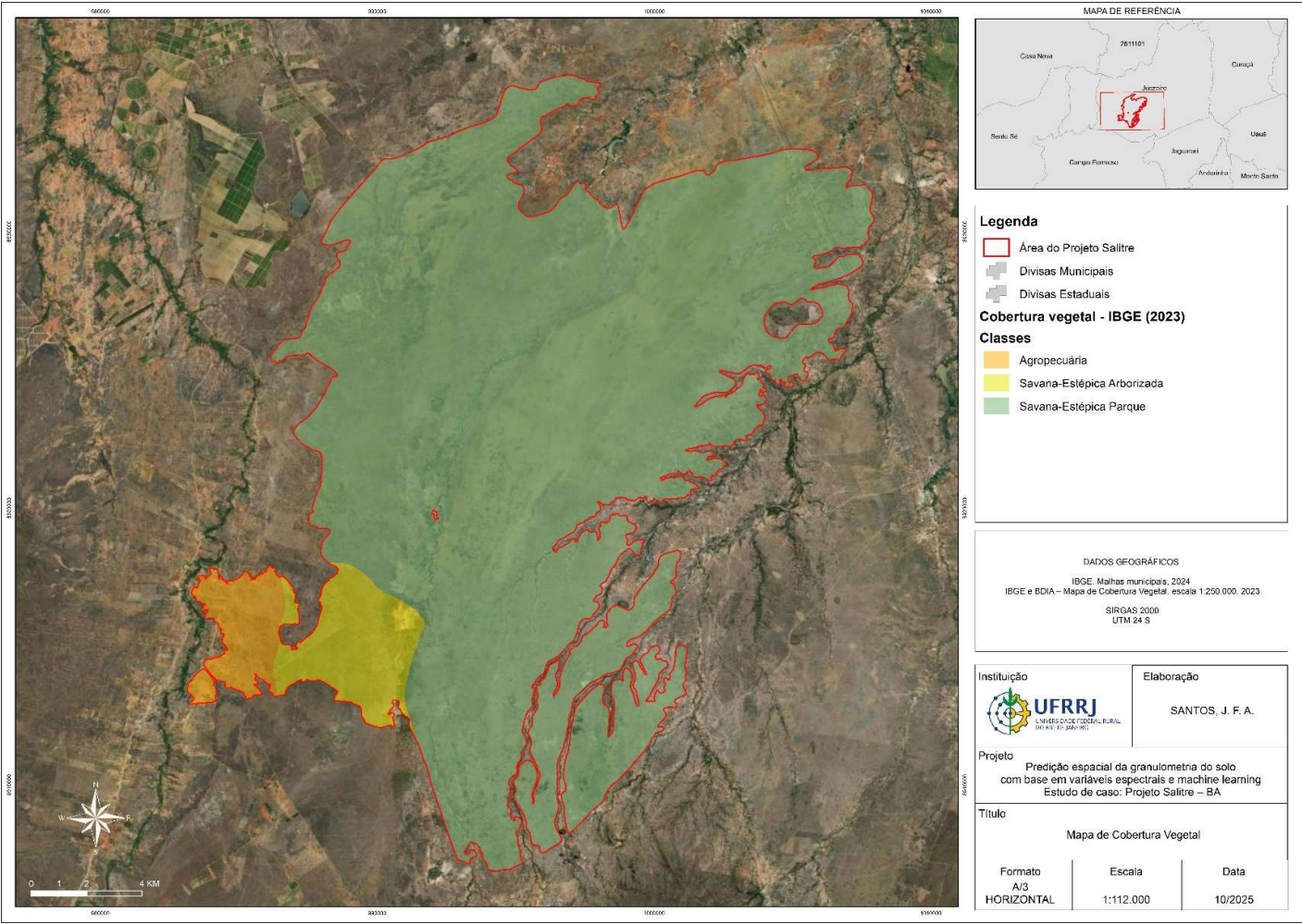
Fonte: IBGE (2023), ESRI World Imagery. **Elaboração:** Santos, J. F. A. (2025)

Figura 2 -Precipitação média anual (1970–2000) na área do Projeto Salitre, Juazeiro, Bahia



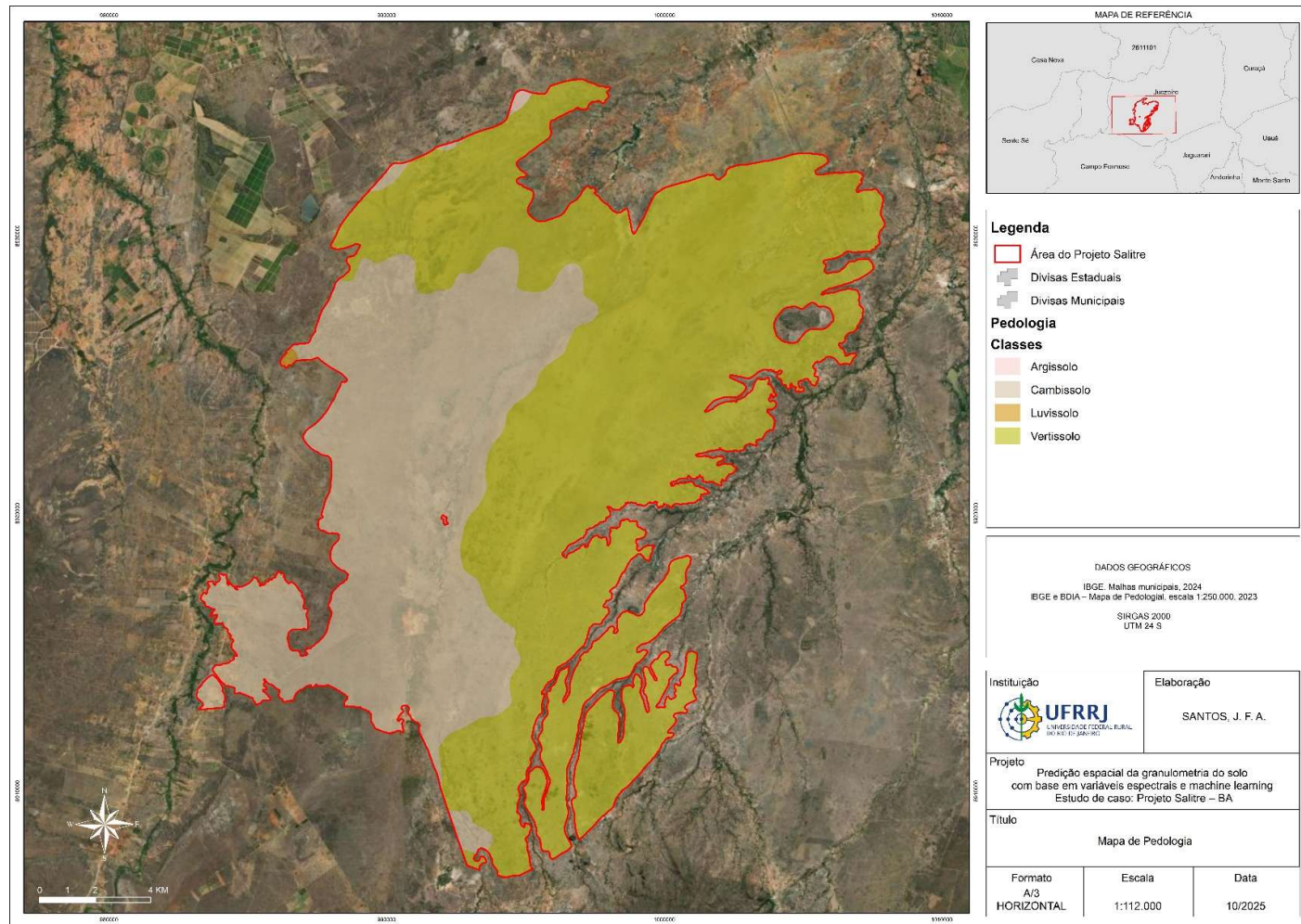
Fonte: Fick, S. E.; Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology, 37(12), 4302–4315. Disponível em: <https://www.worldclim.org/>.

Figura 3 - Cobertura vegetal da área de estudo no município de Juazeiro, Bahia



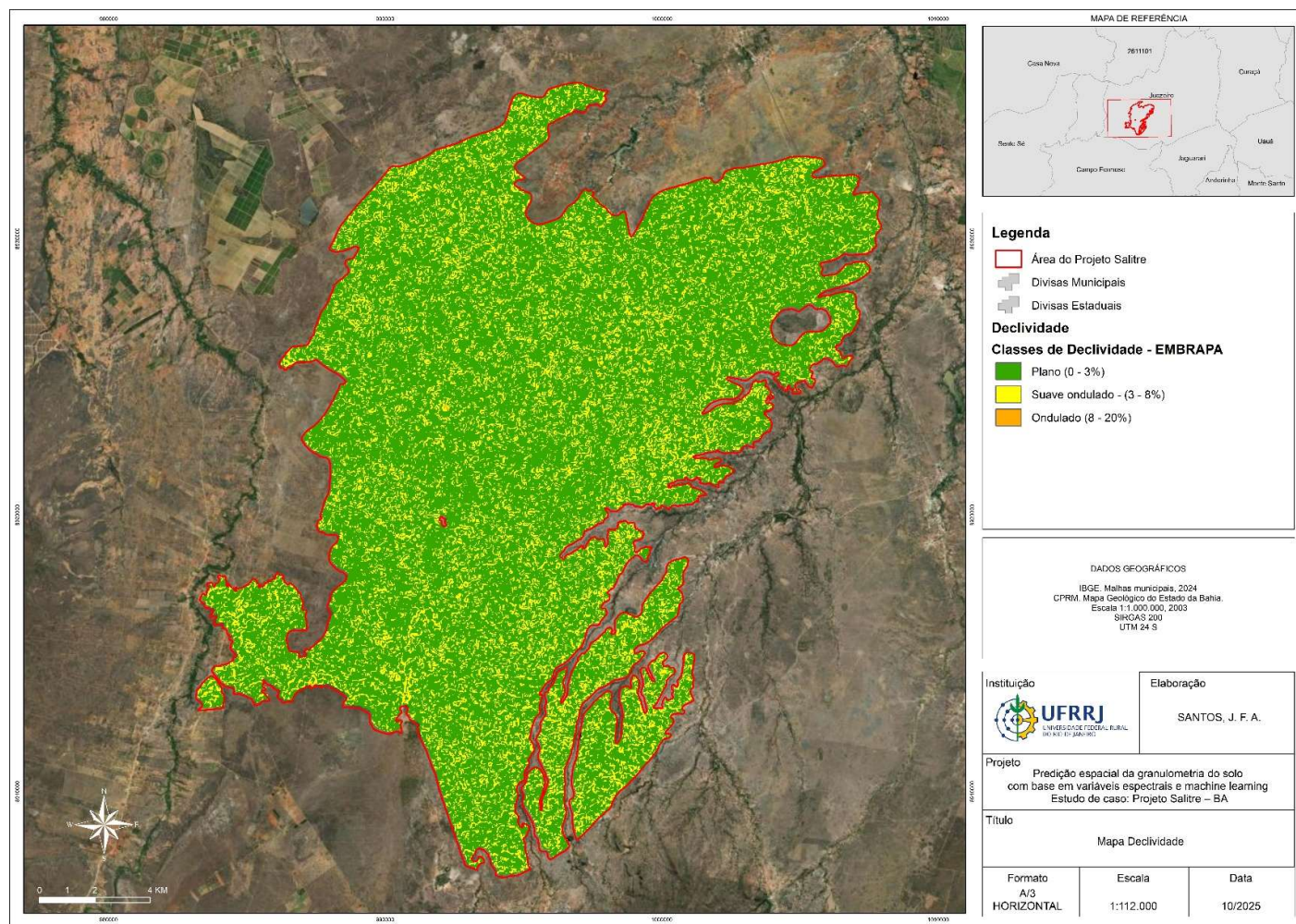
Fonte: Instituto Brasileiro de Geografia e Estatística – IBGE. Banco de Dados de Informações Ambientais (BDIA): Cobertura Vegetal. Rio de Janeiro, 2023. Disponível em: <https://bdiaweb.ibge.gov.br/#/consulta/vegetacao>. Acesso em: 9 out. 2025.

Figura 4 -Mapa de solos da área de estudo no município de Juazeiro, Bahia



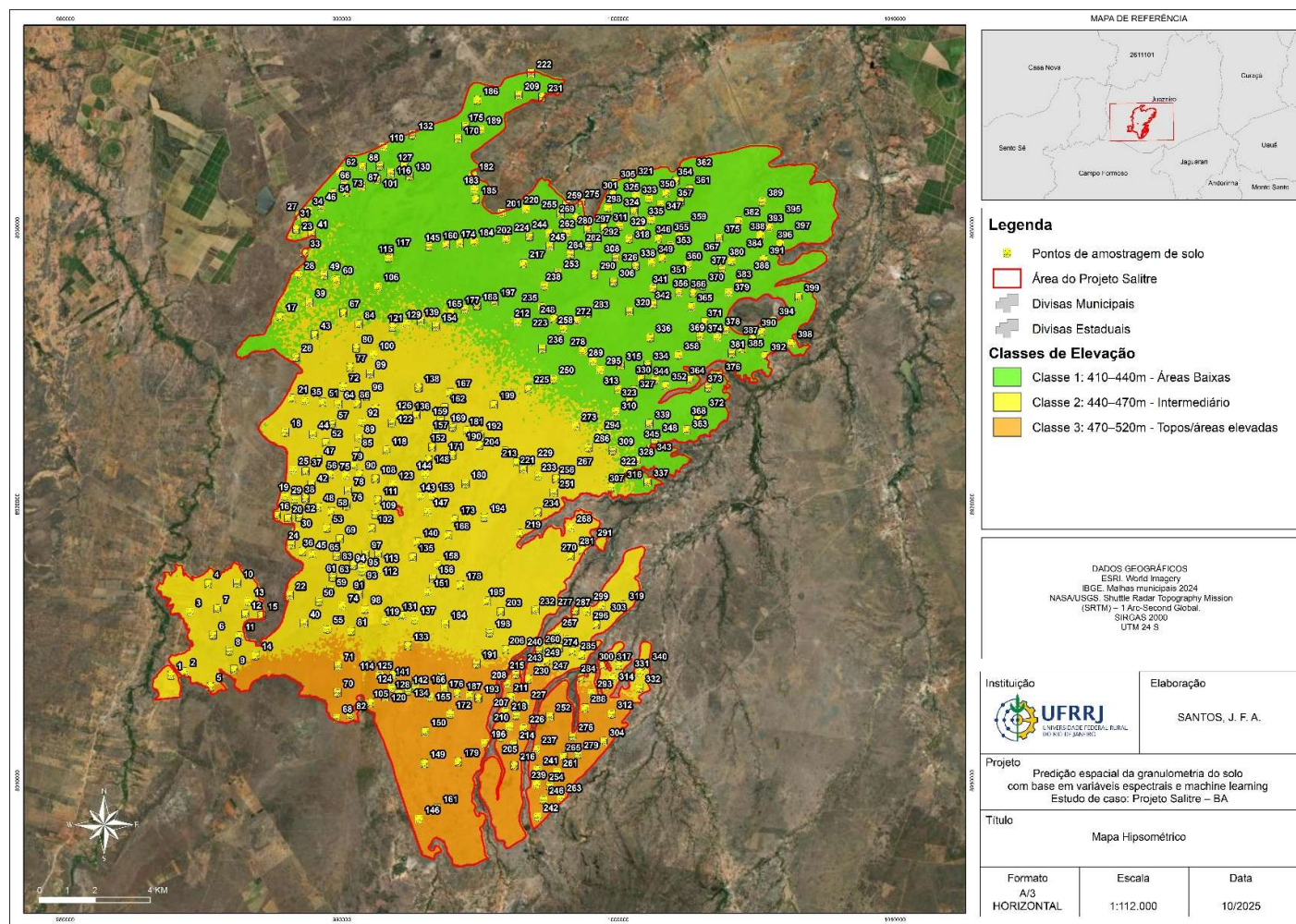
Fonte: Instituto Brasileiro de Geografia e Estatística – IBGE. *Banco de Dados de Informações Ambientais (BDIA): Mapa de Solos do Brasil*. Rio de Janeiro, 2023. Disponível em: <https://bdiaweb.ibge.gov.br/#/consulta/pedologia>. Acesso em: 9 out. 2025.

Figura 5 - Mapa de declividade da área de estudo no município de Juazeiro, Bahia



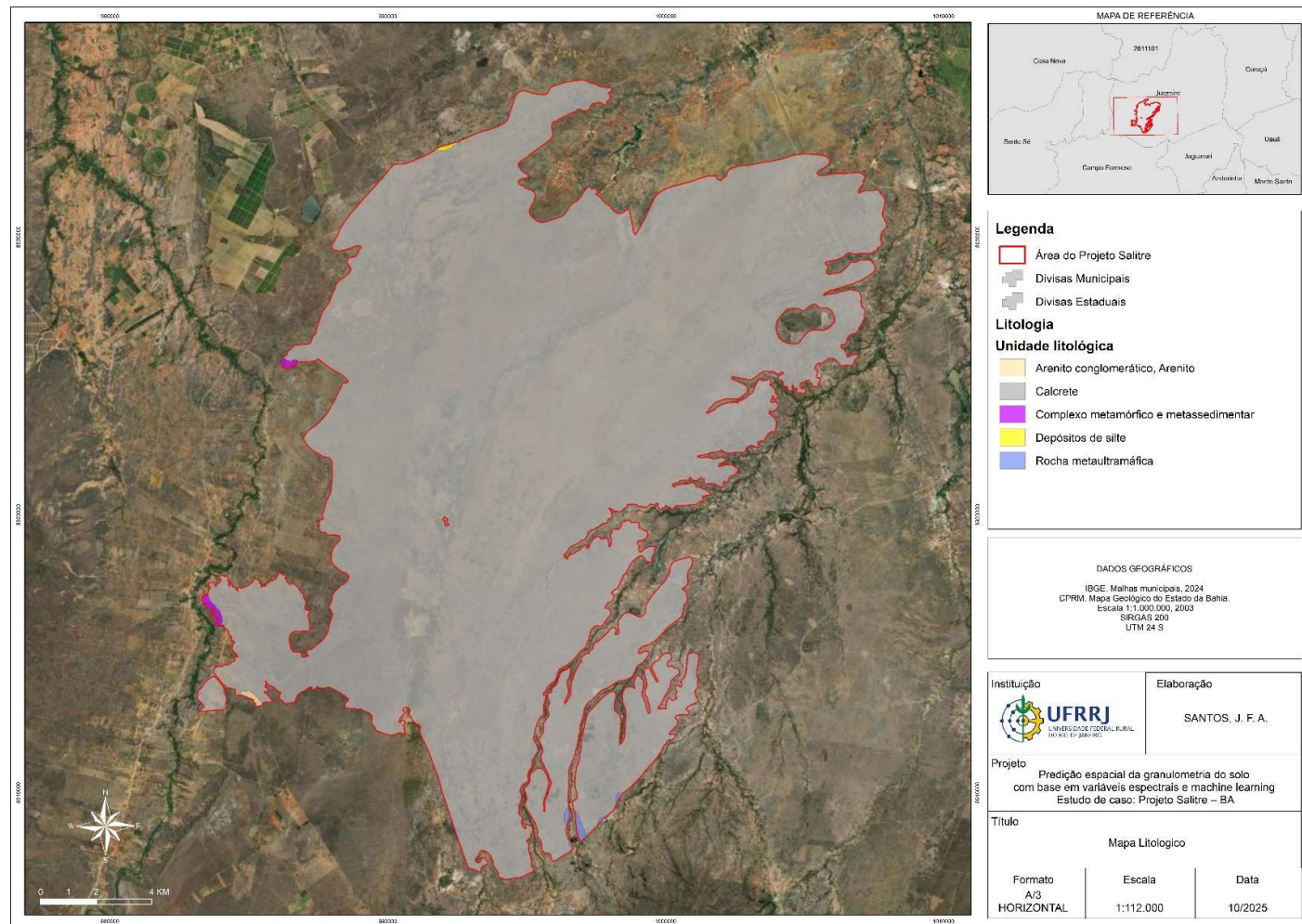
Fonte: NASA/USGS. *Shuttle Radar Topography Mission (SRTM) – 1 Arc-Second Global*. Publicação: 2014. Disponível em: <https://earthexplorer.usgs.gov/>. Acesso em: 9 out. 2025.

Figura 6 - Mapa hipsométrico da área de estudo no município de Juazeiro, Bahia.



Fonte: Dados do Shuttle Radar Topography Mission – SRTM (USGS, 2014). Disponível em: <https://earthexplorer.usgs.gov/>. Acesso em: 9 out. 2025.

Figura 7 - Mapa litológico da área de estudo no município de Juazeiro, Bahia.



Fonte: SILVA, A. M. da; CUNHA, B. C. C. da; LOPES, R. da C.; LIMA, M. I. C. de. *Mapa Geológico do Estado da Bahia*. Escala 1:1.000.000. Brasília: Serviço Geológico do Brasil – CPRM, 2003. Disponível em: <https://rigeo.sgb.gov.br/handle/doc/8665>. Acesso em: 9 out. 2025.

3.2 Covariáveis ambientais

O Modelo Digital de Elevação (MDE) foi obtido a partir da plataforma USGS Earth Explorer com resolução de 30 metros, sendo posteriormente reprojetoado no software ArcGIS 10.8 com a ferramenta *Project Raster* para o sistema de coordenadas SIRGAS 2000 / UTM Zona 24S, recortado para área de estudo com a ferramenta *Extract by mask*. No software SAGA GIS foi corrigido as depressões espúrias (sinks) por meio do algoritmo *Fill Sinks* (Wang & Liu), garantindo a coerência hidrológica do terreno.

A partir do MDE corrigido, ainda no SAGA GIS no módulo *Terrain Analysis – Morphometry / Hydrology* foram derivadas variáveis topográficas de interesse como declividade, curvatura do perfil, curvatura do plano, fator topográfico de comprimento e declive (LS Factor), índice de posição topográfica, acumulação de fluxo e índice de convergência.

Essas variáveis são fundamentais para descrever a dinâmica de relevo, drenagem e potencial de deposição na área. A declividade foi calculada em radianos por ser a unidade mais adequada para uso em equações geomorfológicas e índices compostos (como o LS Factor e o índice de umidade topográfica).

Em seguida, foi utilizada uma cena orbital multiespectral do Landsat 8 OLI/TIRS (Collection 2 Level-2), identificada pelo código “LC08_L2SP_217067_20250328_20250401_02_T1”, obtida em 28 de março de 2025 por meio da plataforma USGS Earth Explorer. A cena apresentava ausência de cobertura de nuvens, favorecendo a extração das informações espectrais. Destaca-se que o mês de março corresponde ao período mais chuvoso na região, condição que influencia a umidade superficial do solo e, portanto, sua assinatura espectral. A cena foi processada no software ArcGIS 10.8.

As bandas espectrais (B2, B4, B5, B6, B7) com resolução de 30 metros foram reprojetoadas para o sistema SIRGAS 2000 / UTM Zona 24S com a ferramenta *Project Raster*, convertidas para o tipo float com a ferramenta *Raster Calculator* e recortadas à área de estudo com *Extract by mask*.

O sensor OLI dos satélites Landsat 8–9 opera em bandas espectrais que vão do visível ao infravermelho de ondas curtas. A banda azul (B2; 0,45–0,51 μm) e a banda verde (B3; 0,53–0,59 μm) registram informações relacionadas à cor, água e solos expostos, enquanto a banda vermelha (B4; 0,64–0,67 μm) destaca a resposta espectral da vegetação e é essencial para análises de uso da terra. A banda do infravermelho próximo (B5; 0,85–0,88 μm) é sensível à estrutura foliar e amplamente utilizada em índices de vegetação. Já as bandas SWIR — B6 (1,57–1,65 μm) e B7 (2,11–2,29 μm) — respondem à umidade do solo e da vegetação, além de variações mineralógicas da superfície. Esse conjunto espectral (B2–B7) fornece ampla capacidade de discriminação ambiental e detalhamento das condições biofísicas da área estudada.

Com a ferramenta *Composite Bands* ainda no ArcGIS 10.8 foram calculados os principais índices espectrais utilizados na literatura pedológica e ambiental, segundo as equações:

- NDVI (Normalized Difference Vegetation Index) = $(B5 - B4) / (B5 + B4)$
- NDMI (Normalized Difference Moisture Index) = $(B5 - B6) / (B5 + B6)$
- SAVI (Soil Adjusted Vegetation Index) = $((B5 - B4) / (B5 + B4 + 0.5)) \times (1 + 0.5)$
- OSAVI (Optimized Soil Adjusted Vegetation Index) = $(B5 - B4) / (B5 + B4 + 0.16)$
- BSI (Bare Soil Index) = $((B6 + B4) - (B5 + B2)) / ((B6 + B4) + (B5 + B2))$

Esses índices foram selecionados por sua sensibilidade à cobertura vegetal, teor de umidade e exposição do solo, variáveis diretamente associadas à dinâmica erosiva e aos padrões de textura superficial.

Posteriormente, aplicou-se a ferramenta “*Extract Multi Values to Points*” no ArcGIS 10.8, permitindo extrair os valores de todas as variáveis (topográficas e espectrais) para pontos amostrais de solo.

Com os dados organizados em planilha, foi construída uma matriz de correlação de Spearman (ρ) (Fig. 10, item 4.2), considerando nível de significância de 0,05, para investigar as relações monotônicas entre as frações granulométricas e as variáveis ambientais e espectrais derivadas.

As tabelas a seguir foram elaboradas para melhor interpretação e justificativa da escolha das covariáveis: A Tabela 1 apresenta a descrição das covariáveis ambientais e a Tabela 2 um resumo dos processos desenvolvidos nessa etapa. A interpretação detalhada das correlações apresentadas será discutida posteriormente no tópico Resultados e Discussão, onde serão analisadas as relações entre as frações granulométricas e as variáveis ambientais e espectrais no contexto da área de estudo. *Tabela*

Tabela 1 - Descrição das Covariáveis Ambientais

Variável	Nome	Descrição
MDE	Modelo Digital de Elevação	Altitude do terreno em cada pixel; base das variáveis topográficas.
SLOPE	Declividade	Inclinação do terreno, influenciando erosão, drenagem e formação do solo.
LS – Factor	Fator LS - comprimento + declividade. (Moore et al. 1991)	Estima potencial erosivo combinando declividade e comprimento de rampa.
SPI	Stream Power Index (Moore et al. 1991)	Mede a energia do fluxo superficial; maior em áreas com maior escoamento.
TPI	Topographic Position Index Guisan et al. (1999), Wilson & Gallant (2000)	Classifica posição relativa: topo, encosta, vale, comparando com vizinhos.
TWI	Topographic Wetness Index - método Standard (Moore et al. 1991)	Indica tendência de acúmulo de água no terreno (umidade).
flow_acc	Flow Accumulation (Top-Down) – Método Multiple Flow Direction. Wilson & Gallant (2000)	Quantidade de pixels que drenam para cada célula; representa convergência hídrica.
conv_ind	Convergence Index Método Aspect Koethe, R. & Lehmeier, F. (1996):	Identifica áreas de convergência/divergência de fluxo superficial.
plan_curv	Plan Curvature Método 9 parameter 2nd order polynom Zevenbergen, L.W., Thorne, C.R. (1987)	Curvatura horizontal que indica convergência ou divergência lateral do terreno.
prof_curv	Profile Curvature Método 9 parameter 2nd order polynom Zevenbergen, L.W., Thorne, C.R. (1987)	Curvatura no sentido do declive; influencia aceleração do fluxo.
BSI	Bare Soil Index Rikimaru, A. ; Miyatake, S. (2002)	Destaca áreas com solo exposto; sensível a textura e matéria orgânica.
OSAVI	Optimized Soil-Adjusted Vegetation Index Rondeaux; Steven; Baret (1996)	Índice de vegetação otimizado para reduzir influência do solo.
SAVI	Soil-Adjusted Vegetation Index Huete (1988)	Índice de vegetação que corrige interferência do solo, útil em vegetação rala.
NDMI	Normalized Difference Moisture Index Adaptado por Jin & Sader (2005)	Estima umidade da vegetação usando NIR e SWIR.
NDVI	Normalized Difference Vegetation Index Rouse Júnior et al. (1974)	Índice clássico de vigor vegetal baseado na reflectância do NIR e vermelho.
b7r	Banda 7 (SWIR-2)	Sensível à mineralogia e umidade do solo; útil em mapeamento geológico.
b6r	Banda 6 (SWIR-1)	Relacionada ao conteúdo de água e minerais secos.
b5r	Banda 5 (NIR)	Essencial para estimar biomassa e vigor da vegetação.
b4r	Banda 4 (Red)	Relacionada à clorofila; componente fundamental do NDVI.
b3r	Banda 3 (Green)	Captura reflectância da vegetação saudável e água rasa.
b2r	Banda 2 (Blue)	Útil para análises de água, aerossóis e correções atmosféricas.

Tabela 2 - Resumo das principais etapas, softwares, ferramentas, plataformas, resoluções e algoritmos utilizados no processamento dos dados

Etapas	Plataforma/ Software	Ferramentas/ Algoritmos Utilizados	Resolução/ Sistema	Descrição/ Observações
Download do MDE e cena Landsat (Sensores OLI/OLI-2) (30 m)	USGS Earth Explorer	–	30 m	Obtenção do MDE bruto (SRTM) e cena Landsat com bandas espectrais.
Reprojeção do MDE e cena Landsat (Sensores OLI/OLI-2)	ArcGIS 10.8	Project Raster	SIRGAS 2000 / UTM 24S	Padronização do sistema de coordenadas.
Recorte da área de estudo (MDE e Cena Orbital)	ArcGIS 10.8	Extract by Mask	–	Ajuste espacial aos limites definidos da área.
Correção de depressões do MDE (sinks)	SAGA GIS	Fill Sinks (Wang & Liu)	–	Correção hidrológica do MDE.
Derivação de variáveis topográficas	SAGA GIS (Terrain Analysis – Morphometry / Hydrology)	Slope, Profile Curvature, Plan Curvature, LS Factor, TWI, TPI, Flow Accumulation, Convergence Index	–	Variáveis relacionadas ao relevo, drenagem, energia e dinâmica erosiva.
Conversão das bandas da cena Landsat para tipo float	ArcGIS 10.8	Raster Calculator	–	Conversão obrigatória para operações matemáticas.
Cálculo de índices espectrais das bandas da cena do Landsat	ArcGIS 10.8	Composite Bands	–	NDVI, NDMI, SAVI, OSAVI, BSI.
Extração de valores para pontos amostrais	ArcGIS 10.8	Extract Multi Values to Points	–	Atribuição dos valores das variáveis a cada ponto de solo.
Análise de correlação e Modelagem Preditiva	R (R Core Team, 2024)	Spearman (ρ), $\alpha = 0,05$ e RF, RLM e AR	–	Avaliação das relações monotônicas entre variáveis e modelagem.

3.3 Modelagem preditiva da composição granulométrica

3.3.1 Seleção e diagnóstico de colinearidade entre covariáveis

Como citado anteriormente, a consolidação dos dados em planilha permitiu a elaboração de uma matriz de correlação de Spearman (ρ) (Fig. 10, item 4.2), considerando nível de significância de 0,05, a fim de identificar padrões monotônicos de associação entre as frações granulométricas e as variáveis ambientais e espectrais derivadas. Essa etapa prévia auxiliou na compreensão das relações estruturais entre os preditores utilizados, fornecendo subsídios para avaliar o comportamento dos modelos diante de possíveis redundâncias informacionais.

A colinearidade é a representação da existência de forte correlação entre duas ou mais variáveis independentes, o que pode gerar redundância de informação e dificultar a estimativa precisa dos coeficientes em modelos estatísticos. Em análises preditivas, a presença de covariáveis altamente correlacionadas tendenciam a variância dos parâmetros e reduzem a interpretação dos modelos lineares, podendo comprometer a estabilidade das previsões (Marcoulides & Raykov, 2018; Chan et al., 2022).

Os modelos preditivos empregados neste estudo apresentam diferentes comportamentos frente à colinearidade entre variáveis independentes. Os algoritmos RF e AR, por serem baseados em árvores de decisão, são menos sensíveis à presença de covariáveis correlacionadas, uma vez que selecionam, a cada divisão, apenas os preditores que mais reduzem a variabilidade interna dos nós, atribuindo maior importância às variáveis com maior poder preditivo. Já a RLM utiliza o método *stepwise backward*, que remove progressivamente as variáveis sem significância estatística, restando apenas aquelas que contribuem de forma mais consistente para a explicação dos teores de areia, silte e argila.

3.3.2 Modelagem por Random Forest (RF)

O conjunto de dados empregado correspondeu à planilha resultante da etapa de extração das covariáveis ambientais e espectrais (Extract Multi Values to Points), contendo as frações granulométricas de areia, silte e argila, além das variáveis derivadas do Modelo Digital de Elevação (MDE) e das bandas espectrais Landsat 8–9 OLI/TIRS (Collection 2 Level-2).

Para cada fração textural, foi ajustado um modelo independente. Os dados foram divididos aleatoriamente em dois subconjuntos: 80% (320 observações) para treinamento e 20% (78 observações) para validação, garantindo independência entre as amostras e representatividade dos valores observados.

O algoritmo foi configurado com 350 árvores de decisão ($n_{tree} = 350$) e 10 variáveis testadas a cada divisão ($m_{try} = 10$), parâmetros definidos com base em testes preliminares e em estudos prévios sobre predição de atributos do solo em ambientes semiáridos (Chagas et al., 2016).

Para melhor compreensão da infraestrutura computacional utilizada, a Tabela 3 a seguir apresenta os principais pacotes empregados na rotina do Random Forest.

Tabela 3 - Pacotes utilizados no *Random Forest* e suas funções

Pacote	Função no Script e no Fluxo de Trabalho	Descrição resumida / Por que é usado
<i>readxl</i>	Leitura da planilha de entrada	Importa arquivos Excel para o R mantendo formatos e tipos.
<i>dplyr</i>	Manipulação e limpeza dos dados.	Seleciona colunas, converte vírgula → ponto, prepara dataset para modelagem.
<i>randomForest</i>	Treinamento do modelo <i>Random Forest</i> e cálculo de importância.	Núcleo da modelagem preditiva; gera o modelo RF, previsões, métricas e importância das covariáveis.
<i>caret</i>	Particionamento da base em treino (80%) e validação (20%).	Permite divisão estratificada garantindo representatividade dos valores do alvo.
<i>openxlsx</i>	Exportação estruturada dos resultados.	Cria planilhas .xlsx com métricas, importância, previsões e tabelas organizadas.
<i>ggplot2</i>	Geração dos gráficos.	Cria visualizações em alta qualidade.
<i>terra</i>	Processamento espacial / geração de mapas raster	Manipula rasters (.tif), calcula previsões espaciais e salva grids preditos. Mais moderno e rápido que raster.
<i>raster</i>	Suporte adicional para rasters	Usado quando alguma função ainda não migrada para terra é necessária. Mantém compatibilidade com fluxos clássicos de GIS.

A avaliação do desempenho do modelo foi realizada por meio de três métricas amplamente utilizadas em modelagem preditiva:

- I. Coeficiente de determinação (R^2) – expressa o grau de ajuste entre valores observados e estimados:

$$R^2 = \left(\frac{\sum (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum (y_i - \bar{y})^2 \sum (\hat{y}_i - \bar{\hat{y}})^2}} \right)^2$$

Onde:

- y_i = valor observado
- $\hat{y}_i$ = valor estimado
- $\bar{y}$ = média dos valores observados
- $\bar{\hat{y}}$ = média dos valores estimados

- II. Raiz do Erro Quadrático Médio (RMSE) – quantifica a magnitude média dos erros de predição:

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

Onde:

- n = número de amostras
- $y_i - \hat{y}_i$ = erro de predição para cada observação

- III. Erro Absoluto Médio (MAE) – representa o desvio médio absoluto entre observações e estimativas:

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

A importância relativa das covariáveis foi avaliada por meio dos indicadores *%IncMSE* (Percentual de Aumento do Erro Quadrático Médio) e *IncNodePurity*, que refletem, respectivamente, a sensibilidade do modelo à exclusão de cada variável e sua contribuição para a pureza dos nós de decisão. Todos os cálculos, gráficos e exportações de resultados foram automatizados no script desenvolvido, garantindo a reprodutibilidade da análise e a organização dos arquivos de saída.

3.3.3 Modelagem por Árvore de Regressão (AR)

A modelagem por Árvore de Regressão (AR) utilizou o mesmo conjunto de pacotes no R, covariáveis ambientais e espectrais descritas anteriormente. Porém, para o ajuste da AR foi adicionado o pacote *rpart*, algoritmo baseado em ANOVA e apropriado para variáveis contínuas.

A ANOVA utiliza o critério dos mínimos quadrados. Ou seja, no *rpart*, o algoritmo avalia todos os possíveis pontos de corte das covariáveis e seleciona aquele que mais reduz a soma dos quadrados dos resíduos (SSE) dentro dos nós. Dessa forma, a árvore realiza divisões apenas quando a partição gera nós mais homogêneos, ou seja, quando os valores observados ficam mais próximos de suas médias internas.

Os dados foram divididos nas mesmas proporções anteriores (treino 80% e validação 20%) e o modelo foi ajustado com parâmetros de complexidade (*cp*) de 0.001 e 0.01, de modo a avaliar diferentes níveis de profundidade e generalização da árvore. A importância relativa das variáveis foi extraída a partir do parâmetro *variable.importance*, permitindo identificar os principais preditores associados à variação dos teores de areia, silte e argila. A avaliação do desempenho seguiu os mesmos critérios adotados para o RF e RLM (R^2 , RMSE e MAE) a fim de possibilitar comparação direta entre os métodos.

3.3.4 Modelagem Regressão Linear Múltipla (RLM)

A modelagem por Regressão Linear Múltipla (RLM) foi realizada no ambiente R, utilizando os mesmos conjuntos de dados e divisão amostral empregados nos modelos RF e AR. Esse método assume relações lineares entre a variável dependente (areia, silte ou argila) e as covariáveis ambientais e espectrais, permitindo estimar a contribuição individual de cada preditor.

O ajuste do modelo foi conduzido por meio do procedimento *stepwise backward*, que remove progressivamente as variáveis sem significância estatística, mantendo apenas aquelas com maior poder explicativo.

O procedimento *stepwise backward* se inicia com o modelo completo, contendo todas as covariáveis e então remove aquelas que apresentam menor contribuição estatística. A cada remoção o modelo é reavaliado, e o processo continua até que permaneçam apenas as variáveis que melhoram o ajuste segundo o critério de seleção adotado. Dessa forma, o método produz um modelo mais consistente, eliminando preditores redundantes ou pouco informativos.

O desempenho da RLM foi avaliado pelas mesmas métricas aplicadas aos demais modelos (R^2 , RMSE com exceção da métrica MAE), possibilitando a comparação direta entre abordagens lineares e não lineares.

4. RESULTADOS E DISCUSSÃO

4.1 Distribuição granulométrica e relação com o relevo

O mapa hipsométrico da área de estudo (Fig. 6, item 3.1) evidencia um relevo predominantemente plano a suavemente ondulado, com altitudes entre 410 e 520 m. Apesar da baixa amplitude altimétrica, foi possível distinguir três classes principais (410–440 m; 440–470 m; 470–520 m), cuja organização espacial contribuiu para derivar variáveis topográficas relevantes aos modelos preditivos.

Mesmo em áreas de relevo pouco expressivo, atributos derivados do MDE (como declividade, curvaturas e índices hidrológicos) desempenharam papel importante por sintetizarem condições de drenagem, redistribuição de materiais e energia do terreno, complementando a interpretação das variáveis espectrais. A análise dos teores de areia, silte e argila por classes altimétricas revelou um gradiente textural sutil, porém consistente ao longo da elevação, mesmo em um relevo predominantemente plano.

Os boxplots com indicação das médias (Fig. 8) mostram que a Classe 1 (410–440 m) apresenta maior variabilidade interna e os maiores teores de argila, acompanhados dos menores valores médios de areia ($263,1 \text{ g}\cdot\text{kg}^{-1}$). Esse padrão sugere condições mais propícias ao acúmulo de partículas finas nas porções mais baixas da paisagem.

Na Classe 2 (440–470 m), observa-se uma redução da argila (média de $442,9 \text{ g}\cdot\text{kg}^{-1}$) e aumento da areia ($346,7 \text{ g}\cdot\text{kg}^{-1}$), enquanto a Classe 3 (470–520 m) apresenta o teor médio de areia mais elevado ($401,7 \text{ g}\cdot\text{kg}^{-1}$) e a menor média de argila ($403,3 \text{ g}\cdot\text{kg}^{-1}$), com amplitude de variação relativamente reduzida. O silte mantém comportamento mais estável entre as classes, com leve declínio à medida que a altitude aumenta.

A matriz de correlação de Spearman ($p < 0,05$) para granulometria versus relevo (Fig. 9) reforça esses padrões, evidenciando correlações negativas entre altitude e argila nas classes superiores, além de correlação positiva entre altitude e areia, especialmente na faixa de 470–520 m.

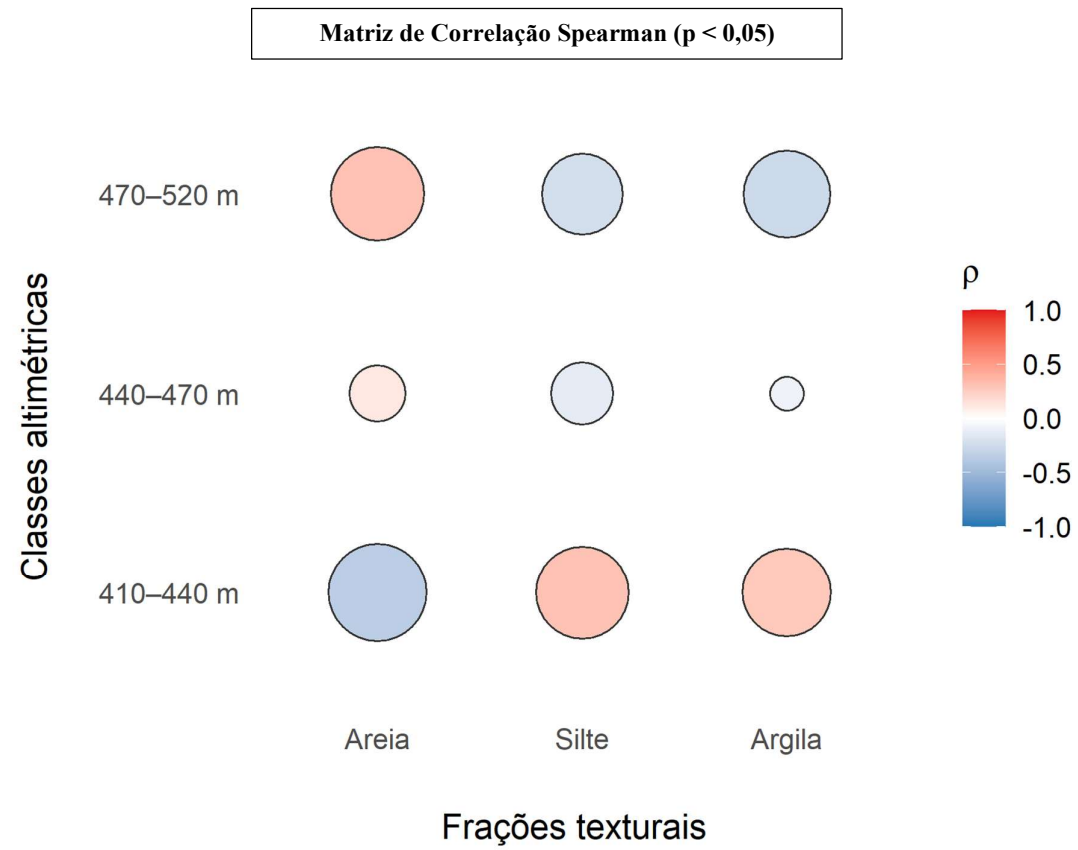
Esses resultados reforçam que, mesmo com diferença altimétrica modesta, a altitude atua como fator organizador da distribuição textural, refletindo gradientes locais de drenagem, energia de relevo e dinâmica de erosão e deposição. Nas porções mais baixas, a maior retenção de água e menor energia do terreno favorecem o acúmulo de partículas finas; enquanto nas regiões mais altas, a drenagem mais eficiente tende a favorecer frações mais arenosas e perfis mais homogêneos.

Essas tendências são coerentes com a distribuição das unidades pedológicas predominantes, onde Vertissolos concentram-se nas áreas mais rebaixadas, associados a ambientes mais úmidos e com maior presença de finos, enquanto Cambissolos tornam-se mais frequentes nas porções elevadas, expressando maior variabilidade litológica e melhor drenagem.

Figura 8 - Boxplot dos teores de areia, silte e argila por classes altimétricas e médias.



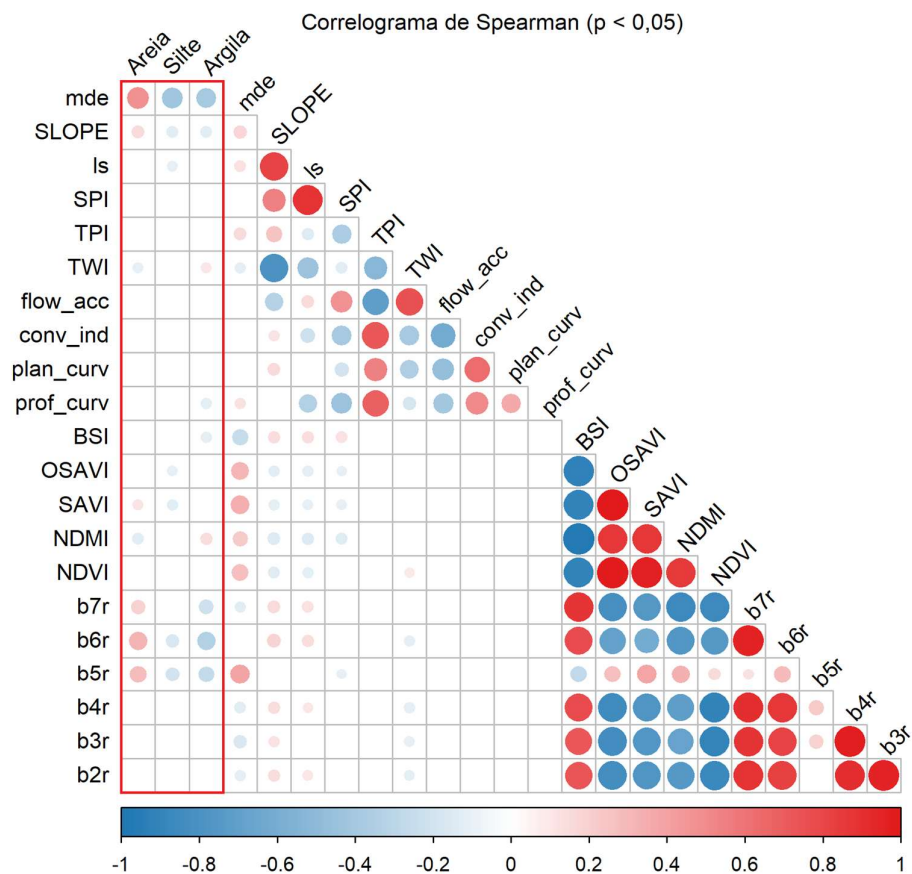
Figura 9 - Matriz de Correlação Spearman para Granulometria x Relevo



4.2 Distribuição granulométrica e relação com covariáveis espectrais e topográficas

A matriz de correlação de Spearman para as covariáveis ambientais ($p < 0,05$) (Fig. 10) demonstrou que as frações granulométricas apresentam relações distintas com as variáveis espectrais e topográficas.

Figura 10 - Matriz de Correlação Spearman para Covariáveis Ambientais



Para a areia, os maiores coeficientes de correlação positiva ocorreram com os índices espectrais associados ao brilho e à resposta no infravermelho de ondas curtas, especialmente BSI, NDMI, SAVI, OSAVI, além das bandas B4, B5, B6 e B7. Esse comportamento é coerente com a literatura, que descreve que solos mais arenosos tendem a apresentar maior reflectância nas faixas VIS–NIR e SWIR devido ao predomínio de minerais silicáticos e à menor capacidade de retenção de água. Esses fatores explicam as correlações moderadas a fortes observadas para a areia em diversas bandas espectrais.

O padrão inverso é observado para a argila: as correlações mais expressivas foram negativas, principalmente com os mesmos índices que favoreceram a predição da areia. Esse resultado também encontra suporte em estudos de espectroscopia de solos, que apontam que materiais argilosos apresentam menor reflectância devido ao aumento da absorção de água, ao maior conteúdo de ferro e às feições de absorção mais pronunciadas no VIS–NIR e SWIR (Demattê & Terra, 2014). Assim, enquanto a areia se associa a maior reflectância, a argila tende a reduzir a intensidade do sinal espectral, o que se reflete em correlações negativas ($\rho < 0$; $p < 0,05$).

O silte, por sua vez, apresentou correlações fracas e pouco consistentes em toda a matriz, com valores discretos de ρ e poucas associações significativas ($p < 0,05$). Esse comportamento é esperado, já que a literatura indica que o silte possui comportamento óptico intermediário e menos distintivo que areia e argila, dificultando sua detecção direta por sensores orbitais (Stenberg et al., 2010). Assim, sua baixa sensibilidade espectral e distribuição menos estruturada na paisagem explicam a ausência de padrões robustos na matriz.

As variáveis topográficas derivadas do MDE (declividade (SLOPE), LS-factor, SPI, TPI, TWI, fluxo acumulado e curvaturas) apresentaram correlações significativas, porém em menor magnitude, tanto positivas quanto negativas, para areia e argila. Esses resultados indicam que o relevo exerce influência textural sutil, mas estatisticamente significativa ($p < 0,05$), coerente com a baixa amplitude altimétrica da área. Ainda assim, as variáveis espectrais demonstraram maior capacidade de discriminar as diferenças texturais superficiais, reforçando seu papel central na modelagem preditiva.

De forma integrada, os resultados confirmam que, no contexto do Projeto Salitre marcado por relevo pouco variável, as covariáveis espectrais foram as mais fortemente associadas às frações granulométricas, enquanto as variáveis de relevo apresentaram contribuição complementar, porém limitada (Tabela 4). Essa estrutura de correlações orienta a expectativa de desempenho dos modelos preditivos discutidos a seguir.

Tabela 4 - Variáveis significativamente correlacionadas com cada fração granulométrica

Fração textural	Variáveis espectrais significativas ($p < 0,05$)	Direção / intensidade	Variáveis topográficas significativas ($p < 0,05$)	Direção/ intensidade
Areia	BSI, OSAVI, SAVI, NDMI, NDVI, Bandas B4, B5, B6, B7	$\rho+$ moderada a forte solos arenosos = maior reflectância	MDE, SLOPE, LS, SPI, TWI, curvaturas	$\rho+$ fraca a moderada
Argila	BSI, OSAVI, SAVI, NDMI, NDVI, Bandas B4, B5, B6, B7	$\rho-$ moderada a forte solos argilosos = menor reflectância	MDE, SLOPE, LS, TWI, plan_curv, prof_curv	$\rho-$ fraca a moderada
Silte	NDVI, SAVI, NDMI (algumas correlações pontuais)	$\rho+$ ou $\rho-$ fraca com padrão pouco definido	SLOPE, SPI (associações isoladas)	ρ fraca / pouco consistente

4.3 Random Forest

Os modelos *Random Forest* (RF) apresentaram bom desempenho geral para estimar os teores de areia, silte e argila a partir das covariáveis espectrais e ambientais, com altos coeficientes de determinação ($R^2 \approx 0,93$) nas etapas de treino e desempenho moderado na validação. Isso demonstra a elevada capacidade do método em capturar relações não lineares entre os atributos do solo e os preditores.

Entre as frações texturais, o melhor desempenho na validação foi obtido para a areia ($R^2 = 0,58$; RMSE = $104,3 \text{ g} \cdot \text{kg}^{-1}$) e para a argila ($R^2 = 0,52$; RMSE = $78,8 \text{ g} \cdot \text{kg}^{-1}$), enquanto o silte apresentou ajuste mais limitado ($R^2 = 0,19$; RMSE = $49,5 \text{ g} \cdot \text{kg}^{-1}$). Essa diferença é coerente com o observado por Chagas et al. (2016), que também reportaram maior acurácia para areia (R^2 de 0,63 e 0,57; RMSE entre 90,8 e $98,0 \text{ g} \cdot \text{kg}^{-1}$) e argila (R^2 de 0,57; RMSE de $73,9 \text{ g} \cdot \text{kg}^{-1}$), e desempenho inferior para silte (R^2 de 0,25; RMSE de $50,2 \text{ g} \cdot \text{kg}^{-1}$).

Os resultados do modelo de validação indicam que, embora o modelo apresente redução de desempenho em relação ao treino, a capacidade de generalização permaneceu boa para areia e argila. Em ambos os casos, os gráficos de observados versus estimados para treino e validação de areia e argila mostram tendência e alinhamento adequado, sugerindo que o RF conseguiu capturar bem os padrões estruturais do terreno e da resposta espectral associada a essas frações.

A melhor performance para areia e argila está relacionada à assinatura espectral mais definida dessas classes e a forte contribuição de variáveis como MDE, bandas SWIR (B6, B7) e índices NDMI/SAVI/OSAVI, que se mostraram informativas no modelo.

Por outro lado, o silte apresentou desempenho mais limitado, padrão também observado na literatura. Essa fração possui maior variabilidade espectral e maior sensibilidade à umidade, matéria orgânica e mistura com areia e argila, fatores que dificultam sua distinção e resultam em menor acurácia nos modelos (Chagas et al., 2016;). Ainda assim, o desempenho encontrado foi coerente com estudos prévios, reforçando a tendência já documentada para ambientes semiáridos.

Esses padrões ficam ainda mais evidentes na Figura 11, que apresenta os gráficos de observados versus estimados para areia, silte e argila. A imagem resume de forma clara o desempenho do *Random Forest*, destacando o melhor ajuste para areia e argila e a maior dispersão observada no silte.

Figura 11 - Gráficos de Resultados do RF para Areia, Silte e Argila

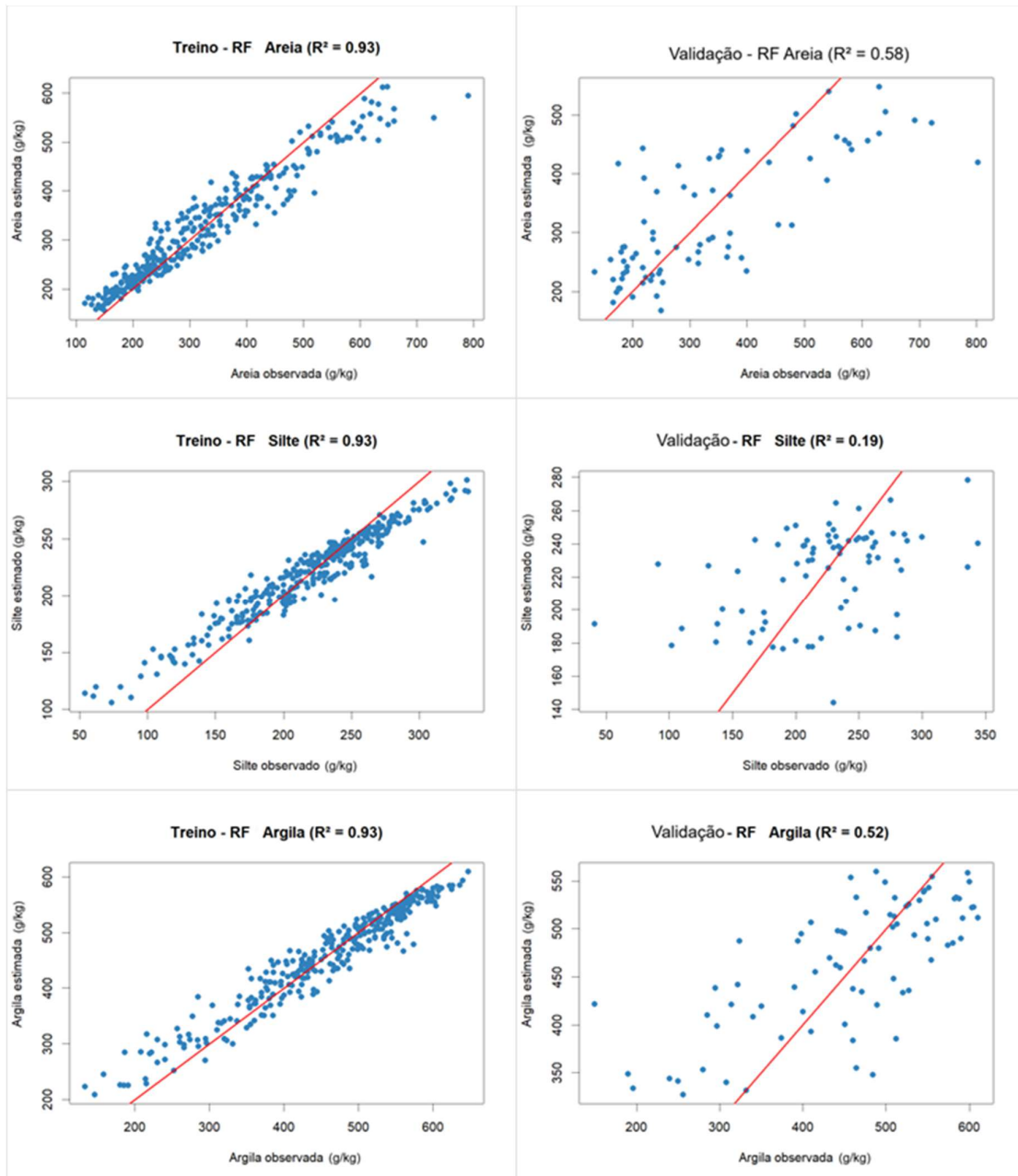
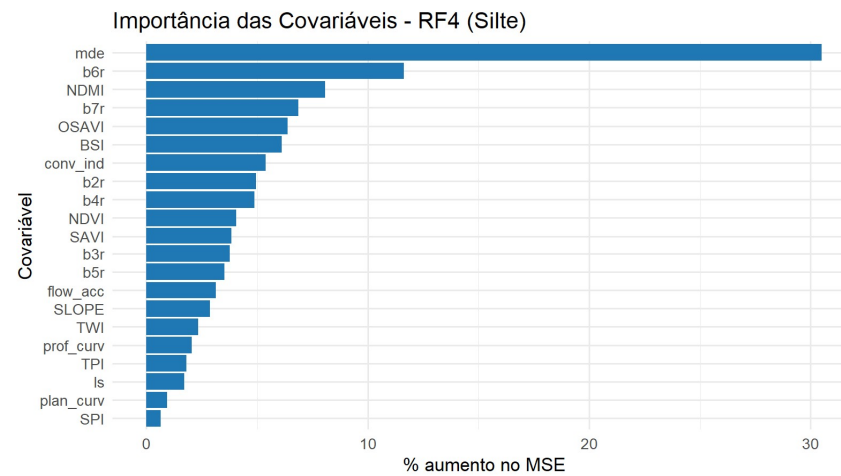
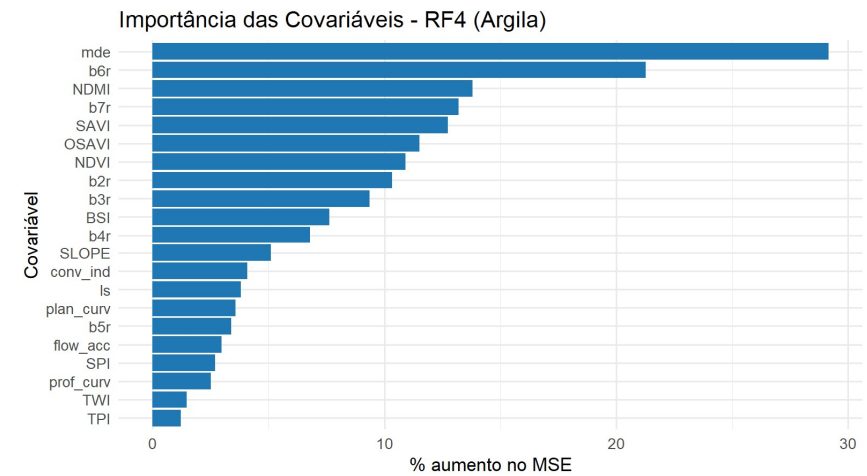
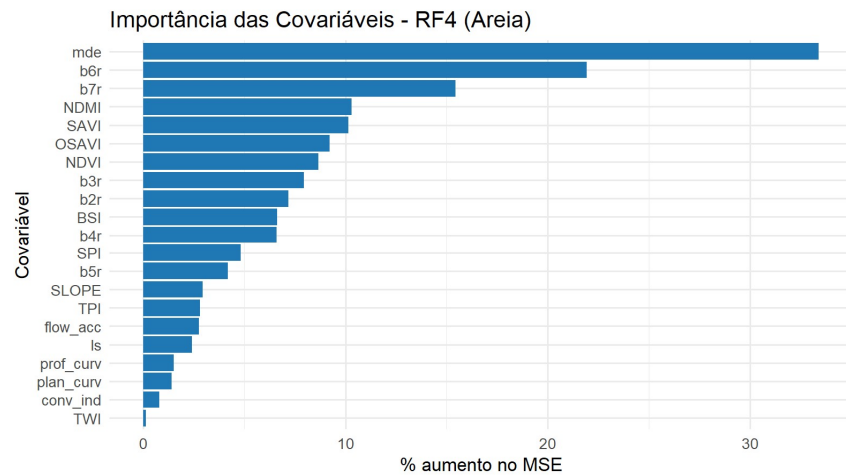


Figura 12 - Importância das covariáveis para Areia, Argila e Silte em Random Forest



4.4 Regressão Linear Múltipla (RLM)

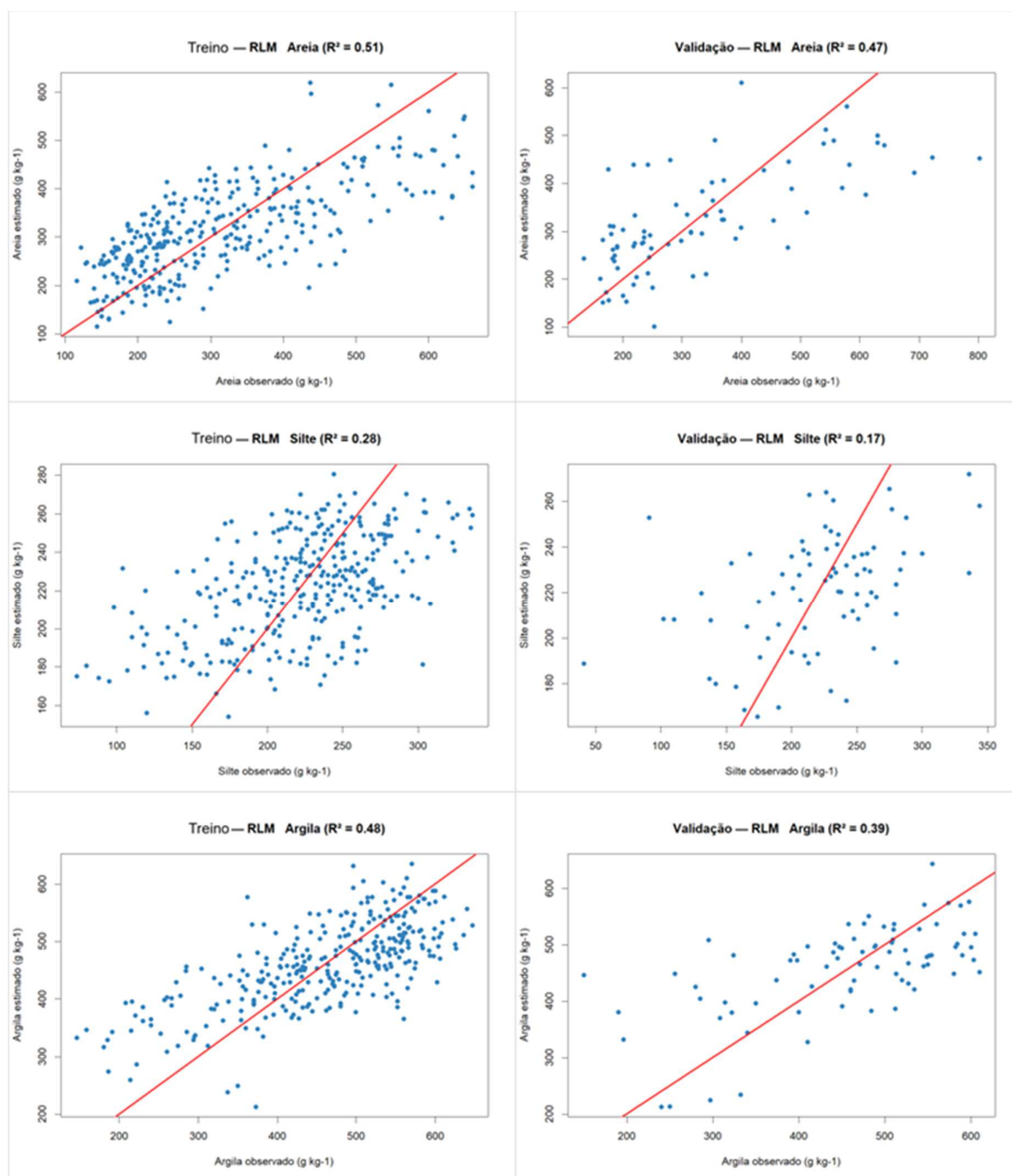
O modelo de Regressão Linear Múltipla (RLM) apresentou desempenho inferior ao *Random Forest*, como esperado para um método linear aplicado a dados ambientais complexos e não lineares. Os valores de R^2 variaram de 0,17 a 0,51 na fase de validação, indicando uma capacidade moderada de explicação da variabilidade dos teores de areia, silte e argila a partir das covariáveis topográficas e espectrais.

Entre as frações, a areia apresentou o melhor desempenho ($R^2 = 0,47$; $RMSE = 115,4 \text{ g}\cdot\text{kg}^{-1}$), seguida pela argila ($R^2 = 0,39$; $RMSE = 87,8 \text{ g}\cdot\text{kg}^{-1}$) e pelo silte, que obteve o menor ajuste ($R^2 = 0,17$; $RMSE = 49,8 \text{ g}\cdot\text{kg}^{-1}$), os resultados são apresentados nos gráficos da Figura 13. Essa hierarquia de acurácia é semelhante a observada nos resultados de Chagas et al. (2016), em que a RLM apresentou R^2 de 0,57 para areia, 0,52 para argila e 0,25 para silte, confirmando a maior previsibilidade dos extremos texturais e reforçando a dificuldade de modelar o silte em ambientes semiáridos.

A pior performance geral da RLM em relação ao RF reflete a natureza predominantemente não linear das relações entre as variáveis espectrais e os atributos do solo, especialmente quando há interação entre fatores topográficos, mineralógicos e de umidade superficial.

Mesmo assim, os resultados obtidos mantêm boa coerência com a literatura, demonstrando que a RLM, embora com suas limitações, continua sendo uma ferramenta útil para identificar tendências gerais e validar a consistência dos preditores selecionados pelos modelos de aprendizado de máquina.

Figura 13 - Gráficos de Resultados do RLM para Areia, Silte e Argila



4.5 Árvore De Regressão

A aplicação do modelo de Árvore de Regressão (AR) representa uma abordagem inédita para os dados do presente estudo, uma vez que Chagas et al. (2016) utilizaram apenas os algoritmos RF e RLM na predição de atributos texturais do solo. A inclusão deste método teve como objetivo avaliar um modelo de estrutura mais simples e interpretável, que permite identificar os limites de decisão mais relevantes na separação dos teores de areia, silte e argila.

Os resultados indicaram desempenho intermediário entre a RLM e o RF, com coeficientes de determinação (R^2) variando de 0,06 a 0,72 e tendência à superestimação no conjunto de treino, refletindo a sensibilidade do método à divisão hierárquica dos dados e à presença de variabilidade espectral dentro de classes texturais. O melhor ajuste foi observado para a argila ($R^2 = 0,30$; RMSE = 95,6 g·kg⁻¹), seguida da areia ($R^2 = 0,26$; RMSE = 142,3 g·kg⁻¹) e do silte ($R^2 = 0,06$; RMSE = 57,8 g·kg⁻¹), confirmando o padrão já observado nos outros modelos, com menor previsibilidade para o silte, os resultados são representados na Figura 14.

Apesar de apresentar menor acurácia em validação quando comparada ao RF, a Árvore de Regressão destacou-se por sua facilidade de interpretação e identificação das variáveis-chave responsáveis pela repartição dos dados, contribuindo para a compreensão dos limiares ambientais que controlam a variação textural no semiárido. Essa característica reforça o potencial do método como ferramenta complementar aos modelos mais complexos, fornecendo suporte à análise física e espectral dos preditores.

Em uma análise mais aprofundada, a AR para o teor de argila mostrou que a altitude, mesmo com pouca amplitude, é a variável de maior importância, separando o conjunto de dados entre áreas mais baixas e mais elevadas. Em ambos os ramos, os índices espectrais de vegetação (SAVI, OSAVI e NDMI) e as bandas do SWIR (b6 e b7) foram os principais responsáveis por refinar as subdivisões, destacando a relação entre maior vigor vegetativo, maior umidade superficial e teores mais elevados de argila. Variáveis topográficas secundárias, como curvatura do terreno e fator LS, também contribuíram para a diferenciação de grupos específicos. De forma geral, o modelo indica que regiões mais úmidas e vegetadas tendem a concentrar maiores teores de argila.

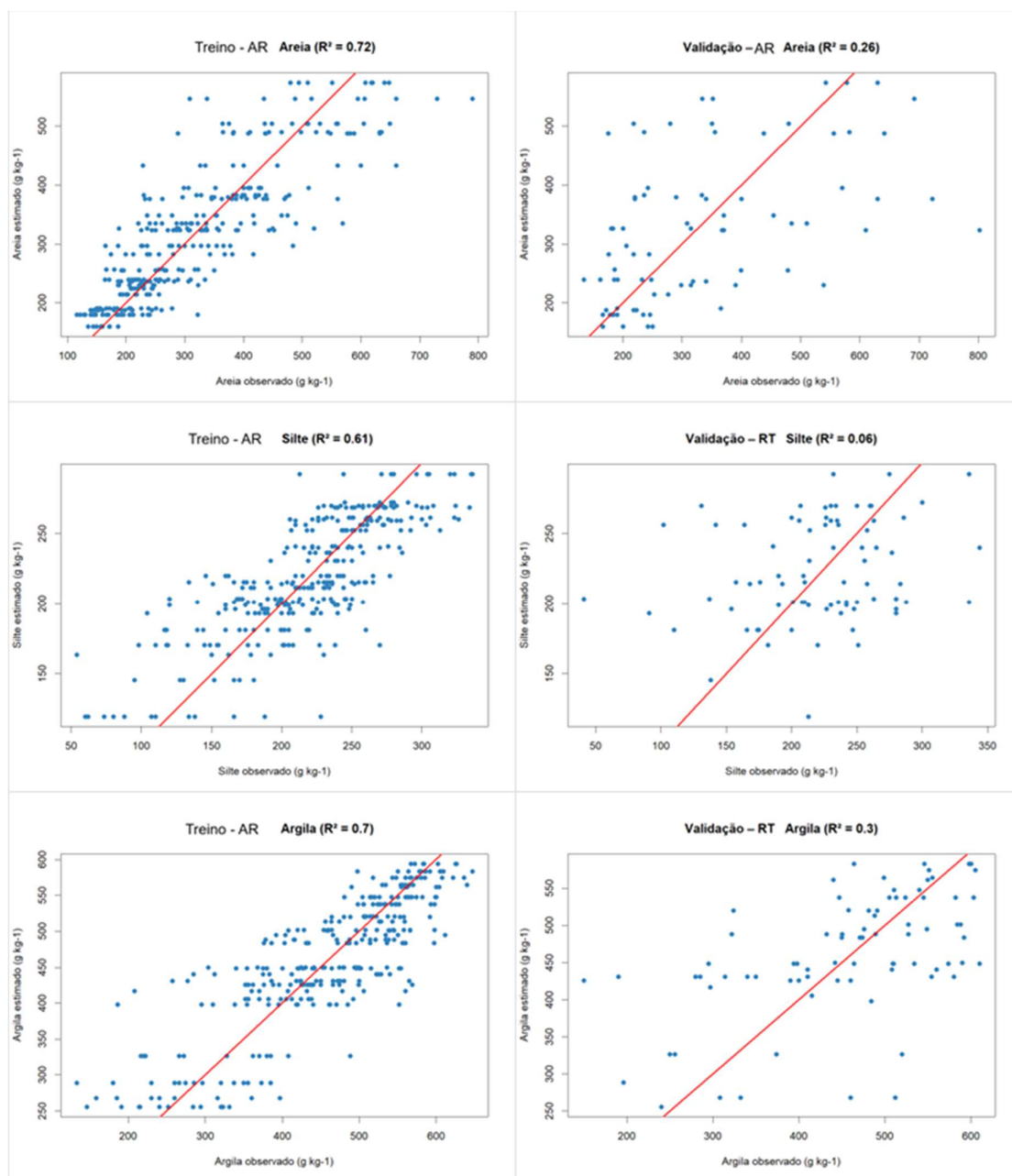
Na Árvore de Regressão referente ao teor de areia, a variável mais influente também foi a altitude (MDE), indicando que a distribuição espacial da areia é fortemente controlada pelo relevo. Em áreas mais baixas ($MDE < 457,5$ m), as divisões subsequentes ocorreram principalmente pelas bandas espectrais SWIR (b4r, b6r) e pelos índices de vegetação, especialmente o b3r, refletindo a sensibilidade desses comprimentos de onda à textura superficial e ao conteúdo hídrico do solo. Já no ramo correspondente às maiores altitudes, o modelo destacou a importância combinada do SAVI, do índice de acúmulo de fluxo (SPI) e do fator LS, sugerindo que processos de redistribuição de materiais pelo relevo e diferenças no vigor da vegetação influenciam a concentração de areia. De modo geral, a árvore indica que solos mais arenosos tendem a ocorrer em locais com menor vigor vegetativo, maior declividade ou menor umidade, enquanto ambientes mais úmidos e vegetados apresentam teores reduzidos de areia.

Na Árvore de Regressão referente ao teor de silte, o modelo destacou novamente a altitude como principal variável discriminante, seguida por indicadores hidrológicos e espectrais. As primeiras divisões ocorreram principalmente pelo TWI (Topographic Wetness Index) e pelo OSAVI, mostrando que a disponibilidade de umidade superficial e o vigor da vegetação exercem forte influência na distribuição do silte. Nos ramos associados a menores valores de TWI, o modelo incorporou variáveis topográficas como declividade (slope) e índices espectrais (b2r), indicando que a redistribuição física de partículas finas está relacionada ao relevo local. Já nos ramos controlados por maiores valores de OSAVI, as subdivisões envolveram bandas SWIR (b6r, b7r), NDVI, BSI e o índice de fluxo acumulado (flow_acc), reforçando a sensibilidade do silte às condições de umidade, erosão e heterogeneidade espectral.

De modo geral, o modelo sugere que solos com maior teor de silte tendem a ocorrer em áreas mais úmidas, vegetadas e com dinâmica hidrológica ativa, enquanto ambientes secos ou mais erodidos apresentam valores mais baixos.

A inclusão da AR amplia o escopo analítico deste estudo, ao introduzir um modelo determinístico e explicativo, capaz de representar relações ambientais relevantes mesmo em cenários de alta heterogeneidade espacial e de limitada disponibilidade amostral. As figuras com as árvores de regressão podem ser encontradas no item Apêndice deste documento.

Figura 14 - Gráficos de Resultados do AR para Areia, Silte e Argila



4.6 Comparação entre modelos

A Tabela 5 resume o desempenho dos três modelos testados (RF, RLM e AR) e evidencia as diferenças na capacidade de estimar as frações granulométricas. O Random Forest apresentou os melhores indicadores gerais, com maior R^2 e menores valores de RMSE na validação para areia ($R^2 = 0,58$; $RMSE = 104,31 \text{ g}\cdot\text{kg}^{-1}$) e argila ($R^2 = 0,52$; $RMSE = 78,81 \text{ g}\cdot\text{kg}^{-1}$). Esse padrão destaca sua maior eficiência em capturar a variabilidade das frações texturais.

A Regressão Linear Múltipla mostrou desempenho intermediário, com redução dos valores de R^2 (entre 0,17 e 0,47) e aumento do erro em comparação ao RF. Apesar disso, manteve relação coerente entre as frações, com melhor ajuste para areia e resultados mais limitados para o silte.

A Árvore de Regressão apresentou os menores valores de R^2 em validação e maior variabilidade dos erros, indicando baixa estabilidade preditiva. Ainda assim, manteve a hierarquia observada nos outros modelos: melhores resultados para argila e areia, e menor ajuste para o silte.

Além dos indicadores numéricos, a comparação espacial entre os mapas preditos pelos três modelos (Figura 15) reforça as diferenças observadas na Tabela 5. O *Random Forest* gerou superfícies mais contínuas com transições suaves, refletindo sua maior capacidade de capturar relações entre covariáveis. Os mapas da RLM, embora apresentem tendência compatível com os gradientes texturais, exibiram regiões de superestimação e subestimação, deixando a representação visual mais homogênea. Já as Árvores de Regressão produziram mapas com mais ruídos/variação abrupta entre pixels, indicando menor estabilidade espacial e maior sensibilidade às covariáveis.

Em conjunto, os produtos espaciais evidenciam visualmente a superioridade do RF, que além de apresentar melhores métricas numéricas, também gerou predições espacialmente mais consistentes e realistas para areia, silte e argila.

Figura 15 - Mapas Preditos dos modelos de RF, RLM e AR para Areia, Silte e Argila

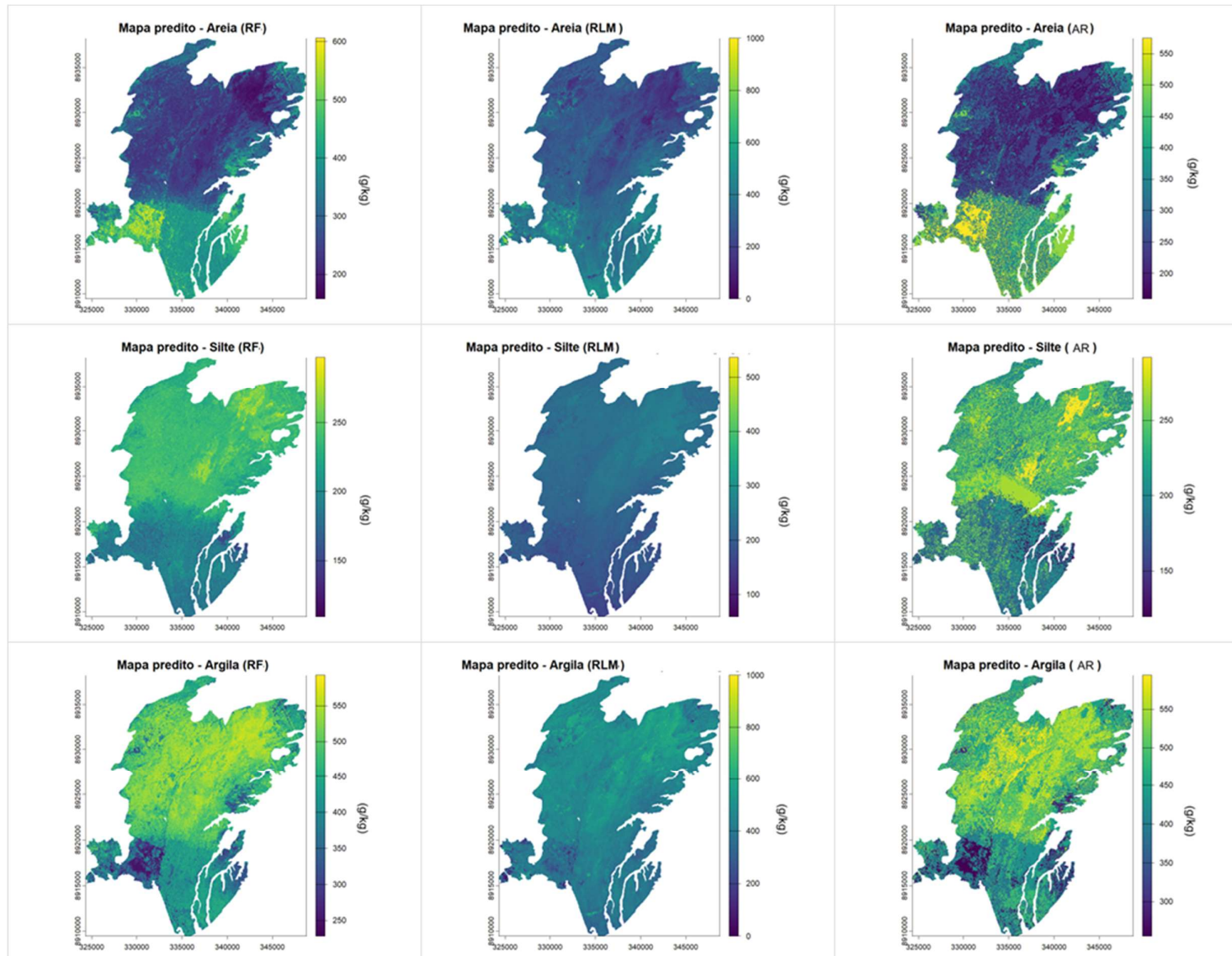


Tabela 5 - Resultados numéricos dos modelos RF, RLM e AR

RF				RLM				AR			
Fração	Conjunto	RMSE (g·kg ⁻¹)	R ²	Fração	Conjunto	RMSE (g·kg ⁻¹)	R ²	Fração	Conjunto	RMSE (g·kg ⁻¹)	R ²
Areia	Treino	40,74	0,93	Areia	Treino	91,14	0,51	Areia	Treino	71,7	0,72
	Validação	104,31	0,58		Validação	115,38	0,47		Validação	142,35	0,26
Silte	Treino	18,67	0,93	Silte	Treino	42,42	0,28	Silte	Treino	32,64	0,61
	Validação	49,47	0,19		Validação	49,8	0,17		Validação	57,84	0,06
Argila	Treino	31,49	0,93	Argila	Treino	76,34	0,48	Argila	Treino	58,81	0,7
	Validação	78,81	0,52		Validação	87,84	0,39		Validação	95,61	0,3

A distribuição textural apresenta um padrão na área de estudo: as porções mais elevadas ao sul concentram maiores teores de areia, enquanto as altitudes mais baixas ao norte acumulam argila. Essa separação é controlada principalmente por três variáveis que aparecem consistentemente entre os modelos: altitude (MDE), índices de umidade (NDMI/TWI) e bandas SWIR (b6r/b7r). Nas áreas mais altas, menor umidade e maior drenagem favorecem a remoção de partículas finas (argila e silte), resultando em solos mais arenosos. Já nas porções baixas, maior fluxo hídrico (TWI alto) e maior retenção de umidade reforçam a deposição de partículas finas, elevando os teores de argila.

5. CONCLUSÃO

A predição espacial da granulometria do solo da área do Projeto Salitre destacou a forte influência das variáveis espectrais na estimativa das frações texturais, sobretudo areia e argila, reforçando a utilidade das bandas SWIR e dos índices NDMI, SAVI e OSAVI na diferenciação de solos em regiões semiáridas. Apesar da baixa amplitude altimétrica da área de estudo, atributos topográficos mostraram efeito complementar, indicando que mesmo pequenos gradientes de relevo podem atuar como organizadores da variação textural.

Entre os métodos avaliados, o *Random Forest* apresentou o melhor desempenho e maior estabilidade, confirmando sua robustez frente à complexidade das relações solo–paisagem. A Regressão Linear Múltipla, embora menos precisa, contribuiu para evidenciar tendências gerais, enquanto a Árvore de Regressão forneceu interpretação clara dos limites ambientais envolvidos nos processos de formação e redistribuição de partículas.

O estudo confirma a viabilidade do uso integrado de dados de sensoriamento remoto, dados topográficos e algoritmos de *machine learning* na predição de atributos do solo, oferecendo um caminho eficiente e com relação custo-benefício favorável para ampliar o detalhamento de informações pedológicas no Brasil. Os resultados obtidos ampliam o entendimento da dinâmica textural dos solos do semiárido e reforçam a importância de investir em abordagens pedométricas para subsidiar o planejamento territorial, manejo sustentável do solo e políticas públicas de desenvolvimento regional.

6. REFERÊNCIAS BIBLIOGRÁFICAS

ALVES, P. F. S.; KONDO, M. K.; FERREIRA, V. G.; SILVA, R. F.; GUERRA, J. V. S. Variabilidade espacial e análise multivariada dos atributos físicos do solo em uma microbacia do rio Pandeiros. *Geo UERJ*, Rio de Janeiro, n. 39, p. e40190, 2021.

ARAÚJO, J. A.; ALBUQUERQUE, A. W.; LIMA, V. L. A.; SOUZA, W. J. Soil water retention and availability in the semiarid region of Brazil. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v. 14, n. 8, p. 826–834, 2010.

BHERING, S. B.; CHAGAS, C. da S.; CARVALHO JUNIOR, W. de; PEREIRA, N. R.; CALDERANO FILHO, B.; PINHEIRO, H. S. K. Mapeamento digital de areia, argila e carbono orgânico por modelos Random Forest sob diferentes resoluções espaciais. *Pesquisa Agropecuária Brasileira*, Brasília, v. 51, n. 9, p. 1359–1370, 2016.

CARVALHO JUNIOR, WALDIR DE; CALDERANO FILHO, BENEDITO; CHAGAS, CLEBER DA SILVA; BHERING, SÉRGIO BORGES; PEREIRA, NILTON RODRIGUES; PINHEIRO, HELENA SARAIVA KOENOW. Regressão linear múltipla e modelo Random Forest para estimar a densidade do solo em áreas montanhosas. *Pesquisa Agropecuária Brasileira*, Brasília, v. 51, n. 9, p. 1428–1437, 2016.

CARVALHO JUNIOR, WALDIR DE; PINHEIRO, HELENA SARAIVA KOENOW; BARBOSA, THERESA ROCCO PEREIRA. Pedometria e mapeamento digital: contribuições na classificação e mapeamento de solos. In: *Resultados Econômicos e de Sustentabilidade nos Sistemas nas Ciências Agrárias*. [S. l.]: [s. n.], p. 43–60, 2020.

CARVALHO JUNIOR, WALDIR DE; PINHEIRO, HELENA SARAIVA KOENOW; CEDDIA, MARCOS BACIS; VALLADARES, GUSTAVO SOUZA. *Pedometrics in Brazil*. Cham: Springer, 2024.

CENTENO, LAINARA NOGUEIRA; SOARES, MARCELO RIBEIRO; SILVA, IVO RIBEIRO. Textura do solo: conceitos e aplicações em solos arenosos. *Revista Brasileira de Engenharia e Sustentabilidade*, Curitiba, v. 4, n. 1, p. 31–37, 2017.

CHAGAS, CÉSAR DA SILVA; CARVALHO JUNIOR, WALDIR DE; BHERING, SÉRGIO BORGES; CALDERANO FILHO, BENEDITO. Spatial prediction of soil surface texture in a semiarid region using Random Forest and multiple linear regressions. *Catena*, Amsterdam, v. 139, p. 232–240, 2016.

CHAN, JIREH YI-LE; LEOW, STEVEN MUN HONG; BEA, KHEAN THYE; CHENG, WAI KHUEN; PHOONG, SEUK WAI; HONG, ZENG-WEI; CHEN, YEN-LIN. Mitigating the multicollinearity problem and its *machine learning* approach: a review. *Mathematics*, Basel, v. 10, n. 1283, p. 1-17, 2022

CRESPOLINI, MATHEUS; NASCIMENTO, PAULO PIRES DO. *PronaSolos: gestão do solo e da água para o desenvolvimento sustentável*. [S. l.]: [s. n.], 2021.

DALMOLIN, RICARDO SIMÃO DINIZ; RUBIN, RONI LUÍS; KAMINSKI, JOÃO; SCHNEIDER, PAULO. Mapeamento digital: nova abordagem em levantamento de solos. *Investigación Agraria: Sistemas y Recursos Forestales*, Madrid, v. 24, n. 2, p. 79–86, 2015.

DE MELO FIGUEIREDO, SILVANA MARIA; DE CARVALHO, LUIZ MARCELO TAVARES. Avaliação da exatidão do mapeamento da cobertura da terra em Capixaba, Acre, utilizando classificação por árvore de decisão. *Cerne*, Lavras, v. 12, n. 1, p. 38–47, 2006.

DEMATTE, J. A. M.; TERRA, F. S. Spectral behavior of soils with different mineralogy and organic matter content. *Geoderma*, v. 217–218, p. 173–183, 2014

DIJAI, T. S. B.; SILVA, F. M.; TEIXEIRA, A. F. D. S.; SILVA, S. H. G.; GUILHERME, L. R. G.; CURI, N. Correção da determinação em campo dos teores de elementos em solos via espectrometria de fluorescência de raios-X portátil. *Ciência e Agrotecnologia*, Lavras, v. 44, p. e002420, 2020.

Embrapa - Serviço Nacional de Levantamento e Conservação do Solo. Manual e métodos de análise do solo. 1979.

ESCOVEDO, TÉRCIO; KOSHIYAMA, ALEXANDRE. *Introdução a Data Science: algoritmos de machine learning e métodos de análise*. São Paulo: Casa do Código, 2020.

FERREIRA, ANA CAROLINA DE SOUZA. *Mapeamento digital de atributos do solo em áreas remotas sob Floresta Amazônica: um estudo de caso na Formação Solimões*. 2022. 99 f. Tese (Doutorado em Agronomia – Ciência do Solo) — Universidade Federal Rural do Rio de Janeiro, Instituto de Agronomia, Seropédica, RJ, 2022.

FICK, S. E.; HIJMANS, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, v. 37, n. 12, p. 4302–4315, 2017. Disponível em: <https://www.worldclim.org/data/worldclim21.html>. Acesso em: 09 out. 2025.

GUISAN, A., WEISS, S.B., WEISS, A.D. GLM versus CCA spatial modeling of plant species distribution. *Plant Ecology* 143: 107-122. 1999.

HENGL, T.; de JESUS, J. M.; MACMILLAN, R. A.; BATJES, N. H.; HEUVELINK, G. B. M.; RIBEIRO, E.; SAMUEL-ROSADO, J.; GUEVARA, M. A.; BROWNE, P.; RANGEL, T.; CHOUDHURY, B. I.; SATOH, M.; SILVA, M.; SIQUEIRA, A.; MENDES, G.; SCHMIDT, K.; MARTIN, M.; AFTERS, R.; RICHARDS, J.; RIVERA, J.; PLEBANI, M.; STRAUSS, J.; WALTER, C.; CROCKER, R.; MUTUKIRA, S.; TAGZOUT, M.; SANZ, G.; MARCHANT, B.; SILVA, C. H. L.; THOMPSON, P.; SHI, X.; MACKAY, A. SoilGrids250m: Global Gridded Soil Information Based on Machine Learning. *PLoS ONE*, v. 12, n. 2, e0169748, 2017.

HUETE, Alfredo R. A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, Amsterdam, v. 25, n. 3, p. 295–309, 1988

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). *Banco de Dados de Informações Ambientais – Geologia, Pedologia e Vegetação: Mapa de Solos e Vegetação do Brasil*, escala 1:250.000. Rio de Janeiro: IBGE, 2023. Disponível em: <https://bdiaweb.ibge.gov.br/#/consulta/geologia>. Acesso em: 08 out. 2025.

JIN, S.; SADER, S. A. Comparison of time series tasseled cap wetness and the normalized difference moisture index in detecting forest disturbances. *Remote Sensing of Environment*, Amsterdam, v. 94, n. 3, p. 364–372, 2005.

LAGACHERIE, PHILIPPE; MCBRATNEY, ALEX BURRELL. Spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping. *Geoderma*, Amsterdam, v. 138, p. 3–10, 2007.

LIMA, ANTONIO Y. V.; CHERUBIN, M. R.; GRESCHUK, L. T.; MUNIZ, G. DE A. M.; CAVALCANTE, D. M.; PEREIRA, A. P. DE A. “Mapping soil health research in the Brazilian Semiarid region: a bibliometric approach.” *Experimental Agriculture*, v. 60, 2024.

MARCHIORO, VINÍCIUS. *Aprendizagem de máquina na definição de zonas homogêneas de manejo para soja*. 2023. 91 f. Tese (Doutorado em Agronomia – Produção Vegetal) — Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal, 2023.

MARCOULIDES, K. M.; RAYKOV, T. Evaluation of variance inflation factors in regression models using latent variable modeling methods. *Educational and Psychological Measurement*, v. 79, n. 5, p. 874–882, 2018.

MCBRATNEY, ALEX B.; MENDONÇA-SANTOS, MARIA DE LOURDES; MINASNY, BUDIMAN. On digital soil mapping. *Geoderma*, Amsterdam, v. 117, p. 3–52, 2003.

MGOHELE, R. N.; TZIOLAS, A. M.; METAXA, G. N.; PAVLATOU-VE, A. K.; TSAKIRIDIS, A. C.; GITAS, I. Z. Prediction of soil texture using remote sensing data: a systematic review. *Frontiers in Remote Sensing*, Lausanne, v. 5, p. 1461537, 2024.

MOORE, I.D., GRAYSON, R.B., LADSON, A.R. Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. *Hydrological Processes*, Vol.5, No.1.1991.

MTSHAWU, BULELANI. *Spatial estimation of surface soil texture using Landsat 8 data*. 2021. Tese (Doutorado em Ciências Ambientais) — University of the Free State, Bloemfontein, 2021.

PEREIRA, R. S.; MENDONÇA-SANTOS, M. L.; TEIXEIRA, W. G.; COELHO, R. M. Digital Soil Mapping Using Machine Learning Algorithms in a Tropical Mountainous Area in Brazil. *Revista Brasileira de Ciência do Solo*, v. 42, e0170260, 2018.

PINHEIRO, HELENA SARAIVA KOENOW. *Métodos de mapeamento digital aplicados na predição de classes e atributos dos solos da bacia hidrográfica do rio Guapi-Macacu, RJ*. 2015. Tese (Doutorado em Agronomia – Ciência do Solo) — Universidade Federal Rural do Rio de Janeiro, Seropédica, 2015.

RONDEAUX, G.; STEVEN, M.; BARET, F. Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, Amsterdam, v. 55, n. 2, p. 95–107, 1996.

ROSA, ADILSON DA SILVA; CARVALHO JUNIOR, WALDIR DE; BHERING, SÉRGIO BORGES; VALLADARES, GUSTAVO SOUZA. *Funções de predição espacial de propriedades do solo*. 2012.

ROUSE JÚNIOR, John W.; HAAS, Robert H.; SCHELL, John A.; DEERING, Donald W. Monitoring vegetation systems in the Great Plains with ERTS. In: *EARTH RESOURCES TECHNOLOGY SATELLITE-1 SYMPOSIUM*, 3., 1973.

SAYÃO, V. M.; LEITE, H. G.; MOREIRA, A. M. Predicting Soil Physical Attributes Using Multiple Linear Regression and Geographic Information Systems. *Revista Brasileira de Ciência do Solo*, v. 42, e0170193, 2018.

SENA, MARCELO M. DE; POPPI, RONEI J.; FRIGHETTO, ROSA T. S.; VALARINI, PEDRO. Avaliação do uso de métodos quimiométricos em análise de solos. *Química Nova*, São Paulo, v. 23, n. 4, p. 547–556, 2000.

SILVA, A. M. DA; CUNHA, B. C. C. DA; LOPES, R. DA C.; LIMA, M. I. C. DE. *Mapa Geológico do Estado da Bahia*, escala 1:1.000.000. Brasília: CPRM, 2003. Disponível em: <https://rigeo.sgb.gov.br/handle/doc/8665>. Acesso em: 09 out. 2025.

SIRTOLI, ANDRÉIA ESTELA; GIRARDI, ROSELI; SIRTOLI, LUIZ FERNANDO; ALVES, MARILICE OENNING CARDOSO; DEDECEK, RENATO ALBERTO. Atributos do relevo derivados de modelo digital de elevação e suas relações com solos. *Scientia Agraria*, Curitiba, v. 9, n. 3, p. 317–329, 2008.

STAMFORD, NEWTON P.; RODRIGUES, JOSÉ J. V.; HECK, RICHARD J.; ANDRADE, DOMINGOS E. G. T. Propriedades físicas e químicas dos solos. In: MICHEREFF, Sami J.; ANDRADE, Domingos E. G. T.; MENEZES, Maria (Orgs.). *Ecologia e Manejo de Patógenos Radiculares em Solos Tropicais*. Recife: Imprensa Universitária da Universidade Federal Rural de Pernambuco, 2005.

STENBERG, B.; VISCARRA ROSSEL, R. A.; MOUAZEN, A. M.; WETTERLIND, J. Visible and near infrared spectroscopy in soil science. *Advances in Agronomy*, v. 107, p. 163–215, 2010.

WILSON, J.P. & GALLANT, J.C. Primary Topographic Attributes. In: Wilson, J.P. & Gallant, J.C. [Eds.]: *Terrain Analysis: Principles and Applications*, John Wiley & Sons, p.51-85.2000

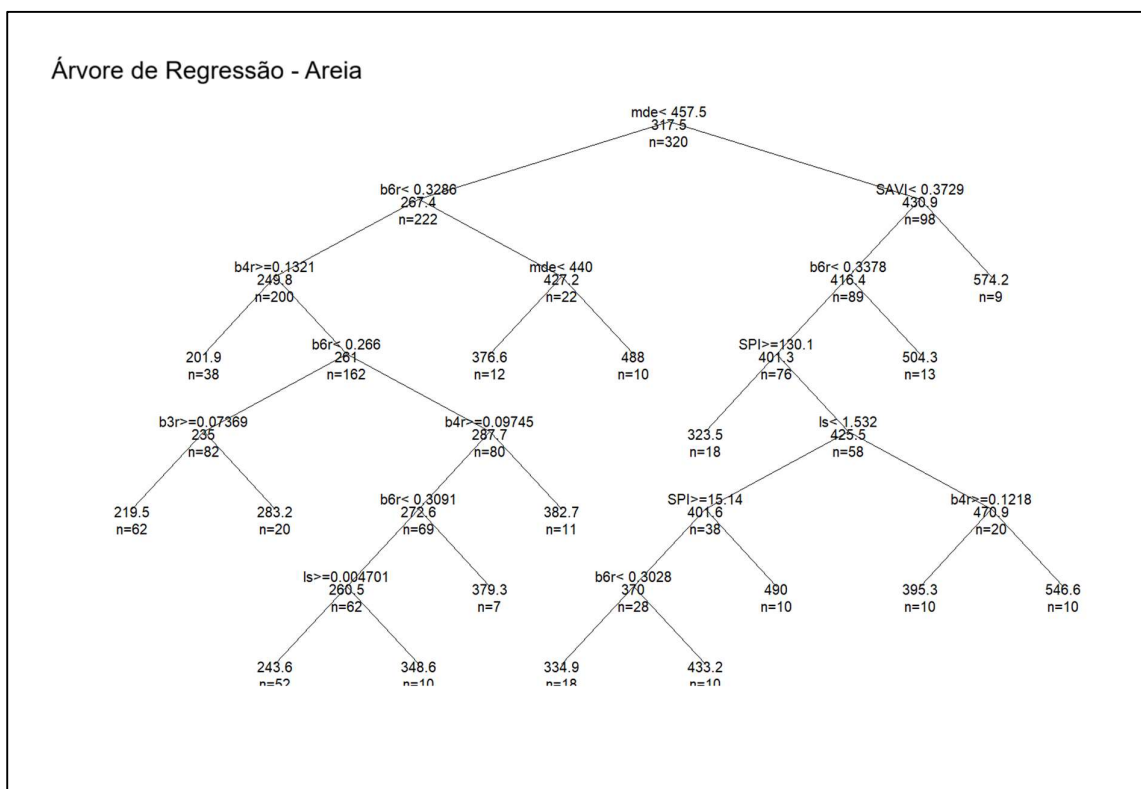
YANG, LIPING; DRISCOL, JOSHUA; SARIGAI, SARIGAI; WU, QIUSHENG; CHEN, HAIFEI; LIPPITT, CHRISTOPHER D. Google Earth Engine and Artificial Intelligence (AI): a comprehensive review. *Remote Sensing*, Basel, v. 14, n. 14, p. 3253, 2022.

ZEVENBERGEN, L.W., THORNE, C.R. Quantitative analysis of land surface topography. *Earth Surface Processes and Landforms*, 12: 47-56.1987.

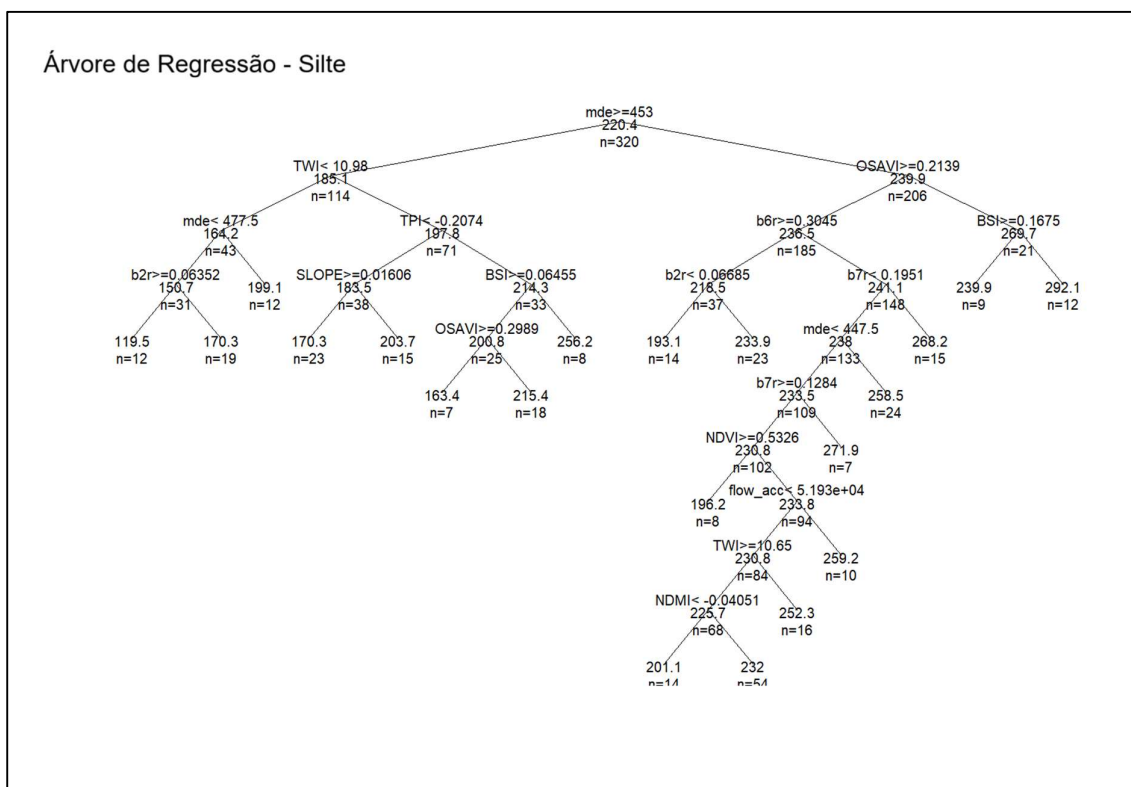
ZHOU, YAN; WU, WEI; LIU, HAN. Exploring the influencing factors in identifying soil texture classes using multitemporal Landsat-8 and Sentinel-2 data. *Remote Sensing*, Basel, v. 14, n. 21, p. 5571, 2022.

APENDICES

APENDICE A – Árvores de Regressão para Areia



APENDICE B – Árvores de Regressão para Silte



APENDICE C – Árvores de Regressão para Argila

