

**UFRRJ**

**PRÓ-REITORIA DE EXTENSÃO**  
**CURSO DE ESPECIALIZAÇÃO EM GEOPROCESSAMENTO,**  
**LEVANTAMENTO E INTERPRETAÇÃO DE SOLOS**

**TRABALHO DE CONCLUSÃO DE CURSO**

**Regressão Logística Multinomial Aplicada ao**  
**Levantamento de Solos no Município de Paty**  
**dos Alferes e sub-bacias do Córrego do Saco-**  
**Rio Ubá (RJ)**

**Ana Caroline da Silva Pinto Ferreira**

**2025**



**UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO**  
**PRÓ-REITORIA DE EXTENSÃO**  
**CURSO DE ESPECIALIZAÇÃO EM GEOPROCESSAMENTO,**  
**LEVANTAMENTO E INTERPRETAÇÃO DE SOLOS**

**REGRESSÃO LOGÍSTICA MULTINOMIAL APLICADA AO**  
**LEVANTAMENTO DE SOLOS NO MUNICÍPIO DE PATY DOS**  
**ALFERES E SUB-BACIAS DO CÓRREGO DO SACO-RIO UBÁ (RJ)**

**ANA CAROLINE DA SILVA PINTO FERREIRA**

*Sob a Supervisão da Professora*  
**Lucia Helena Cunha dos Anjos**

*E da Pesquisadora*  
**Melania Merlo Ziviani**

Trabalho de Conclusão de Curso  
submetido como requisito parcial para  
obtenção do título de **Especialista** em  
Geoprocessamento, Levantamento e  
Interpretação de Solos.

Seropédica, RJ  
Julho de 2025



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

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

F383r Ferreira, Ana Caroline da Silva Pinto, 1997-  
Regressão logística multinomial aplicada ao levantamento de solos no município de Paty dos Alferes e sub-bacias do Córrego do Saco-Rio Ubá (RJ) / Ana Caroline da Silva Pinto Ferreira. - Volta Redonda, 2025.  
37 f.: il.

Orientadora: Lucia Helena Cunha dos Anjos.  
Coorientadora: Melania Merlo Ziviani.  
Trabalho de Conclusão de Curso(Especialização). -- Universidade Federal Rural do Rio de Janeiro, Curso de Especialização em Geoprocessamento, Levantamento e Interpretação de Solos, 2025.

1. Classificação de Solos. 2. Regressão Logística Multinomial. 3. Análise Multivariada. 4. Geoprocessamento. I. Anjos, Lucia Helena Cunha dos, 1957-, orient. II. Ziviani, Melania Merlo, -, coorient. III Universidade Federal Rural do Rio de Janeiro. Curso de Especialização em Geoprocessamento, Levantamento e Interpretação de Solos. IV. Título.

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

É permitida a cópia parcial ou total deste documento, desde que seja citada a fonte.

**UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO  
PRÓ-REITORIA DE EXTENSÃO  
CURSO DE ESPECIALIZAÇÃO EM GEOPROCESSAMENTO, LEVANTAMENTO  
E INTERPRETAÇÃO DE SOLOS**

**ANA CAROLINE DA SILVA PINTO FERREIRA**

Este Trabalho de Conclusão de Curso (TCC) foi submetido como requisito parcial para obtenção do título de **Especialista** em Geoprocessamento, Levantamento e Interpretação de Solos.

TCC APROVADO EM 18 de Julho de 2025.

**Profa. Lúcia Helena Cunha dos Anjos**  
(Supervisora) – UFRRJ

**Profa. Ana Carolina de Souza Ferreira**  
Tutora – UFRRJ

**Dr<sup>a</sup> Melania Merlo Ziviani**  
Tutora – UFRRJ

## RESUMO

Ferreira, A. C. S. P. **Regressão logística multinomial aplicada ao levantamento de solos no município de Paty dos Alferes e sub-bacias do Córrego do Saco-Rio Ubá (RJ)**. 2025. 37f. Trabalho de Conclusão de Curso (Especialização em Geoprocessamento, Levantamento e Interpretação de Solos). Pró-reitoria de Extensão, Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, 2025.

Este estudo teve como objetivo caracterizar as classes de solo presentes no município de Paty dos Alferes e nas sub-bacias do Córrego Saco-Rio Ubá (RJ), com base na classificação pedológica, textura, profundidade e relação C/N. Foram utilizadas análises estatísticas descritivas e multivariadas, como Análise de Agrupamento e Análise de Componentes Principais (PCA), para identificar padrões entre os solos e os fatores ambientais. Adicionalmente, a relação entre o uso atual do solo, o material de origem, o relevo e as propriedades edáficas, foi avaliada por meio de testes de associação, como qui-quadrado, correlação e ANOVA. O processo de modelagem estatística incluiu a aplicação da regressão logística multinomial, visando explorar as variáveis ambientais que influenciam a classificação textural dos solos. Inicialmente, o modelo apresentou problemas de sobreajuste e separação perfeita, prejudicando a qualidade das inferências estatísticas. A reestruturação das variáveis categóricas, com agrupamento de categorias infrequentes e exclusão de níveis não utilizados, resultou em um segundo modelo estatisticamente mais estável, com redução do AIC, menor *deviance residual* e melhor desempenho na validação cruzada, porém ainda com presença de separação perfeita. Para melhoria, foi proposto um terceiro modelo que incorporou duas melhorias decisivas: a remoção de um ponto com comportamento extremo (*outlier*) e o agrupamento das classes da variável resposta, simplificando a estrutura da Classe de Textura. Os resultados indicaram que a relação C/N teve influência significativa na distribuição das classes texturais, especialmente nas classes mais arenosas, que apresentaram menor acúmulo de matéria orgânica. A análise espacial, complementada por técnicas de geoprocessamento, permitiu mapear a distribuição das classes de solo e do uso e cobertura vegetal, revelando padrões que refletem a interação entre fatores pedogenéticos e o manejo atual da paisagem. Recomenda-se, para estudos futuros, a aplicação de métodos estatísticos mais robustos, como modelos penalizados ou bayesianos, para mitigar eventuais problemas de separação e garantir maior confiabilidade nas inferências.

**Palavras-chave:** Classificação de Solos. Regressão Logística Multinomial. Análise Multivariada. Geoprocessamento.

## ABSTRACT

Ferreira, A. C. S. P. **Multinomial logistic regression applied to soil survey in the municipality of Paty dos Alferes and sub-basins of the Saco-Ubá River (RJ)**. 2025. 37p. Trabalho de Conclusão de Curso (Specialization in Geoprocessing, Soil Survey and Interpretation) Pró-reitoria de Extensão, Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, 2025.

This study aimed to characterize the soil classes present in the municipality of Paty dos Alferes and in the sub-basins of the Saco-Rio Ubá stream (RJ), based on pedological classification, texture, depth, and the C/N ratio. Descriptive and multivariate statistical analyses, such as Cluster Analysis and Principal Component Analysis (PCA), were used to identify patterns among soils and environmental factors. Additionally, the relationship between current land use, parent material, relief, and edaphic properties was assessed through association tests such as chi-square, correlation, and ANOVA. The statistical modeling process included the application of multinomial logistic regression, aiming to explore the environmental variables that influence soil textural classification. Initially, the model exhibited issues of overfitting and perfect separation, compromising the quality of statistical inferences. The restructuring of categorical variables—by grouping infrequent categories and excluding unused levels—resulted in a second, statistically more stable model, with reduced AIC, lower residual deviance, and better performance in cross-validation, although perfect separation was still present. To improve the model, a third version was proposed incorporating two decisive adjustments: the removal of an extreme outlier and the grouping of classes in the response variable, thereby simplifying the structure of the Texture Class. The results indicated that the C/N ratio had a significant influence on the distribution of textural classes, especially in the sandier classes, which showed lower organic matter accumulation. Spatial analysis, complemented by geoprocessing techniques, enabled the mapping of soil class distribution and land use/vegetation cover, revealing patterns that reflect the interaction between pedogenetic factors and current landscape management. For future studies, the use of more robust statistical methods, such as penalized or Bayesian models, is recommended to mitigate potential separation problems and ensure greater reliability in inferences.

**Keywords:** Soil Classification. Multinomial Logistic Regression. Multivariate Analysis. Geoprocessing.

## SUMÁRIO

<b>1. INTRODUÇÃO</b> .....	<b>1</b>
<b>2. REVISÃO DE LITERATURA</b> .....	<b>3</b>
<b>2.1 Estudo dos Solos e Suas Propriedades</b> .....	<b>3</b>
2.1.1 Classes de solos e formação .....	3
2.1.2 Propriedades físicas e químicas .....	4
<b>2.2 Uso de Geoprocessamento em Estudos de Solos</b> .....	<b>5</b>
2.2.1 Cartografia de solos .....	5
2.2.2 Análise de uso e cobertura da terra .....	5
2.2.3 Integração solo-geotecnologias .....	6
<b>2.3 Análise Exploratória de Dados com Uso do Teste de Tukey</b> .....	<b>6</b>
2.3.1 Estatísticas descritivas e visualizações .....	6
2.3.2 Análise de variância (ANOVA) e teste de Tukey .....	6
2.3.3 Correlações e multicolinearidade .....	7
2.3.4 Redução de dimensionalidade .....	7
<b>2.4 Regressão Logística Multinomial Aplicada ao Estudo de Solos</b> .....	<b>7</b>
2.4.1 Fundamentos da regressão multinomial .....	7
2.4.2 Aplicações em ambientes naturais .....	8
2.4.3 Interpretação e validação de modelos .....	8
2.4.4 Lacunas e oportunidades de pesquisa .....	8
<b>3. MATERIAL E MÉTODOS</b> .....	<b>10</b>
<b>3.1 Coleta dos Dados</b> .....	<b>10</b>
<b>3.2 Identificação Visual das Áreas de Estudo</b> .....	<b>10</b>
<b>3.3 Pré-processamento de Dados</b> .....	<b>11</b>
3.3.1 Análise exploratória dos dados .....	13
<b>3.4 Modelo Conceitual da Regressão Logística Multinomial Aplicado à Textura</b> .....	<b>13</b>
<b>4. RESULTADOS E DISCUSSÃO</b> .....	<b>15</b>
<b>4.1 Identificação Visual das Áreas de Estudo e Mapas</b> .....	<b>15</b>
<b>4.2 Análise Exploratória dos Dados</b> .....	<b>19</b>
4.2.1 Análise das variáveis categóricas .....	19
4.2.2 Análise das variáveis numéricas .....	23
<b>5.2 Modelagem por Regressão Logística Multinomial</b> .....	<b>28</b>
<b>5.3 Principais Observações</b> .....	<b>29</b>
<b>5. CONSIDERAÇÕES FINAIS</b> .....	<b>34</b>
<b>6. REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	<b>36</b>
<b>7. ANEXO</b> .....	<b>38</b>

## 1. INTRODUÇÃO

A dinâmica de uso e ocupação do solo tem sido alvo de crescente atenção por parte da comunidade científica, gestores públicos e setor produtivo, uma vez que suas transformações impactam diretamente os ciclos biogeoquímicos, os recursos hídricos, a biodiversidade e a qualidade ambiental como um todo (LAMBIN et al., 2003, ELLIS et al., 2021). No Brasil, país com vasto território e grande diversidade de paisagens, compreender as relações entre as características dos solos e os tipos de uso a que estão submetidos é crucial para o planejamento ambiental e agrícola sustentável (SANTOS et al., 2018).

O solo, sendo um recurso natural finito e não renovável em escalas humanas de tempo, possui propriedades físicas, químicas e biológicas que refletem diretamente sua aptidão para diferentes usos. Parâmetros como pH, relação carbono/nitrogênio (C/N), textura e profundidade são frequentemente utilizadas em estudos de classificação e manejo, pois indicam o potencial produtivo, a suscetibilidade à degradação e a fertilidade natural do solo (TEIXEIRA et al., 2017; SANTOS et al., 2018).

A compreensão dessas variáveis, no entanto, demanda abordagens quantitativas capazes de lidar com a complexidade e a multidimensionalidade dos dados. Nesse sentido, a Análise Exploratória de Dados (AED) e os modelos estatísticos multivariados tornam-se ferramentas indispensáveis para extrair padrões significativos e apoiar a tomada de decisão. A AED permite uma visualização inicial das tendências, dispersões e correlações entre variáveis, enquanto modelos preditivos, como a regressão logística multinomial, oferecem suporte inferencial e classificatório para variáveis categóricas com múltiplas classes (HAIR et al., 2009).

A regressão logística multinomial é particularmente adequada para problemas em que a variável resposta possui três ou mais categorias nominais, como ocorre em estudos de classificação do uso do solo. Ao modelar a probabilidade de ocorrência de cada categoria em função de variáveis explanatórias contínuas e/ou categóricas, essa técnica permite não apenas a predição de classes, mas também a interpretação dos efeitos das variáveis explicativas sobre as chances relativas de cada uso do solo (LONG; FREES, 2006; AGRETI, 2018). Além disso, pode-se avaliar a significância estatística dos coeficientes estimados, a qualidade do ajuste (via pseudo- $R^2$ ) e a capacidade preditiva do modelo com medidas como acurácia e matriz de confusão (MENARD, 2010).

Embora existam diversas aplicações da regressão logística binária na área ambiental e agrícola (ex.: previsão de presença/ausência de espécies, classificação de áreas degradadas, avaliação de risco), o uso da regressão logística multinomial ainda é subexplorado em estudos que buscam classificar tipos de uso do solo com base em características físico-químicas. Trabalhos recentes têm destacado o potencial dessa abordagem para aprimorar sistemas de classificação automatizados, inclusive integrando análises com geoprocessamento e sensoriamento remoto (SILVA et al., 2014; FERREIRA et al., 2019).



Nesse contexto, o presente trabalho tem como proposta utilizar técnicas de análise exploratória e modelagem estatística para investigar a associação entre as propriedades físicas e químicas dos solos e os diferentes tipos de uso do solo observados em uma determinada região de estudo. Especificamente, busca-se aplicar a regressão logística multinomial para modelar e prever o uso do solo a partir de variáveis ambientais coletadas em campo ou oriundas de bases secundárias, como bancos de dados de solos.

Parte-se do pressuposto que estudos quantitativos que integram a classe de textura com outras propriedades físicas e químicas oferecem um retrato mais completo do ambiente edáfico e são essenciais para modelagens preditivas, aplicando a regressão logística multinomial para avaliação do uso do solo e identificação de padrões espaciais.

Espera-se, com isso, contribuir para o avanço metodológico no tratamento estatístico de dados ambientais e para o desenvolvimento de ferramentas que auxiliem a gestão territorial e o planejamento do uso do solo de forma mais eficiente, fundamentada em evidências científicas e em técnicas robustas de análise.

O objetivo geral do estudo foi analisar a variação espacial das propriedades físico-químicas dos solos do município de Paty dos Alferes e sub-bacias do Córrego do Saco-Rio Ubá - RJ, com ênfase em sua caracterização, tipologia, uso e fatores ambientais associados.

O trabalho teve como objetivos específicos:

- a) Caracterizar as classes de solo presentes na área com base na classificação, textura, profundidade e relação C/N.
- b) Aplicar estatística descritiva e multivariada (Análise de agrupamento e PCA) para identificar padrões entre os solos e fatores ambientais.
- c) Avaliar relações entre uso atual, material de origem, relevo e propriedades do solo por meio de testes de associação (qui-quadrado, correlação, ANOVA, etc.).
- d) Analisar a distribuição espacial das classes de solo, usos e cobertura vegetal utilizando ferramentas de geoprocessamento

## 2. REVISÃO DE LITERATURA

### 2.1 Estudo dos Solos e Suas Propriedades

#### 2.1.1 Classes de solos e formação

Os solos são corpos naturais que resultam da interação complexa entre fatores como material de origem, clima, relevo, organismos e tempo. Esses fatores formam o que se conhece como os cinco fatores da pedogênese, e a combinação entre eles determina a diversidade de solos encontrados em uma região. No Brasil, essa diversidade é particularmente acentuada, o que exige um sistema de classificação detalhado e adaptado às condições tropicais (SANTOS et al., 2018).

O Sistema Brasileiro de Classificação de Solos (SiBCS), publicado pela Embrapa, é a principal referência na taxonomia dos solos no país. O SiBCS estabelece uma hierarquia taxonômica robusta, dividindo os diferentes tipos de solos em 13 ordens, 42 subordens, 220 grandes grupos e aproximadamente 1.200 subgrupos, com base em características morfológicas, físicas, químicas e mineralógicas observadas em perfis representativos (SANTOS, et al., 2018). Essa estrutura foi desenvolvida para facilitar a recomendação de usos apropriados para cada tipo de solo, bem como suas práticas de manejo agrícola e avaliação ambiental (EMBRAPA, 2024).

Classificar solos em categorias com base em características diagnósticas padroniza a comunicação técnica e permite que pesquisadores, extensionistas e órgãos de planejamento trabalhem de forma coordenada, considerando o solo como um recurso natural com potencial produtivo e restrições específicas (EMBRAPA, 2024). A classificação hierárquica permite associar características morfológicas observadas em campo com dados laboratoriais — por exemplo, textura, cor, pH e saturação por bases —, o que é indispensável para projeções de uso sustentável e zoneamento agroecológico (EMBRAPA, 2024; UNIVERSIDADE FEDERAL DO AMAZONAS, 2023).

A compreensão das classes de solos é essencial para qualquer estudo que envolva aptidão agrícola ou planejamento ambiental. Solos como Latossolos, Argissolos, Neossolos e Cambissolos, por exemplo, possuem características específicas de fertilidade, permeabilidade e estabilidade estrutural que afetam diretamente suas formas de uso (PEREIRA et al., 2023; SANTOS et al., 2018; SYNGENTA, 2021).

As ordens estabelecidas refletem processos pedogenéticos predominantes em diferentes regiões do Brasil — sendo que os Latossolos dominam as extensas áreas tropicais por serem solos muito intemperizados e, em geral, ácidos. A segunda classe de maior expressão geográfica é a dos Argissolos, que se distinguem da primeira pelo horizonte B textural e o processo prevalente de translocação de argila (PEREIRA, et al. 2023; SANTOS et al., 2018).

Os solos do estado do Rio de Janeiro apresentam grande diversidade, resultado da complexa interação entre fatores como relevo acidentado, clima tropical úmido, ampla variedade litológica e histórico de uso antrópico. De acordo com Pereira et al. (2023), os principais tipos de solos que ocorrem no território fluminense são os Latossolos, Argissolos, Cambissolos, Gleissolos e Neossolos, com destaque para a expressiva presença de Argissolos

e Latossolos nas regiões de relevo suave ondulado a ondulado, comuns no noroeste e centro-sul do estado. Já nas áreas de maior declividade e nas vertentes da Serra do Mar e Serra da Mantiqueira, predominam os Cambissolos e Neossolos, frequentemente rasos e pedregosos, o que limita seu uso agrícola.

Estudos que integram classificação de solos com atributos quantificáveis, como propriedades físico-químicas, são importantes para calibrar modelos estatísticos de previsão de uso da terra. Tais integrações vêm sendo cada vez mais exploradas na literatura, em especial com o uso de bases georreferenciadas e bancos de dados pedológicos regionais (SANTOS et al., 2018).

### 2.1.2 Propriedades físicas e químicas

As propriedades físicas e químicas dos solos incluem um conjunto de atributos fundamentais, como textura, densidade do solo, porosidade, temperatura, pH e acidez do solo, teor de matéria orgânica, teor de bases e de nutrientes, capacidade de troca catiônica (CTC), condutividade elétrica, relação carbono/nitrogênio (C/N), entre outras. Esses atributos têm implicações diretas na fertilidade do solo, na disponibilidade de nutrientes e na atividade biológica (TEIXEIRA et al., 2017).

O pH do solo influencia a disponibilidade dos nutrientes e pode influenciar a atividade dos microrganismos. Solos muito ácidos, por exemplo, em geral, apresentam toxicidade por alumínio e baixa disponibilidade de fósforo. A matéria orgânica, por sua vez, atua na agregação do solo, na retenção de água e na capacidade de troca de íons, sendo um dos principais indicadores da qualidade do solo (PEREIRA, et.al., 2023).

A proporção entre carbono e nitrogênio (relação C/N) tem sido utilizada como indicador de processos de mineralização e estabilidade da matéria orgânica. Uma relação C/N muito alta pode indicar baixa mineralização e acúmulo de resíduos, enquanto uma relação muito baixa pode sinalizar perda de matéria orgânica e degradação (LOSS et al., 2011).

Outro atributo de destaque é a textura do solo, a qual representa a proporção relativa de areia, silte e argila, em termos percentuais, sendo uma propriedade de extrema importância para a dinâmica da água, do ar e dos nutrientes no solo. Na taxonomia de solos, esse atributo é um diagnóstico em várias classes, em nível de ordem e em níveis inferiores, identificando classes de textura como solos arenosos, siltosos, média, argilosos, entre outras (SANTOS et al., 2018).

A textura do solo é frequentemente considerada uma propriedade físico-mecânica, mas com forte influência sobre os aspectos químicos e biológicos do solo. A proporção relativa de argila, silte e areia determina a capacidade do solo em reter água, nutrientes e interagir com a matéria orgânica, tornando-se um dos principais condicionantes do comportamento edáfico (SANTOS et al., 2018). Solos com alto teor de argila tendem a ter maior capacidade de retenção de água e nutrientes, mas podem apresentar problemas de drenagem e compactação. Já os solos arenosos são bem drenados, mas com baixa capacidade de retenção de nutrientes e água, exigindo manejo atento para seu uso com lavouras comerciais (RESENDE et al., 2014).

A textura também influencia a distribuição espacial do uso e cobertura do solo, pois interfere diretamente na escolha de lavouras agrícolas e na resiliência da vegetação natural.

Solos mais argilosos tendem a favorecer usos mais intensivos, enquanto solos arenosos podem apresentar limitações naturais à ocupação e uso (RESENDE et al., 2014). Assim, a análise da textura do solo é uma das primeiras etapas em estudos de potencial agrícola das terras, pois define as limitações e potencialidades para diferentes tipos de uso. A textura também afeta a composição florística e a ocupação antrópica do território (SOUZA et al., 2012), sendo que solos mais arenosos têm alto potencial de degradação uma vez removida a vegetação natural.

O conhecimento dessas propriedades é imprescindível para a construção de modelos preditivos de uso do solo. A análise estatística dessas variáveis permite compreender quais fatores mais influenciam a ocorrência de determinado tipo de uso ou cobertura vegetal.

## 2.2 Uso de Geoprocessamento em Estudos de Solos

### 2.2.1 Cartografia de solos

O geoprocessamento é uma ferramenta essencial para a análise e visualização da variabilidade espacial dos solos. Através da integração de dados obtidos em campo com modelos digitais de elevação e imagens de satélite, é possível elaborar mapas temáticos de propriedades edáficas e classes de solos, contribuindo significativamente para o planejamento do uso da terra (SILVA et al., 2014).

A cartografia digital de solos vem ganhando destaque com o uso de sistemas de informação geográfica (SIG), que permitem o armazenamento, análise e manipulação de dados espaciais. Técnicas de mapeamento preditivo, como regressões geográficas ponderadas e redes neurais artificiais, têm sido utilizadas para gerar mapas de propriedades do solo com alta acurácia (CUNHA; DEMATTÊ, 2002).

Além da descrição e classificação, os mapas gerados permitem avaliar a aptidão agrícola das áreas, a suscetibilidade à erosão e as zonas críticas para conservação ambiental. Isso torna a cartografia de solos uma base fundamental para políticas públicas de ordenamento territorial e desenvolvimento rural sustentável.

### 2.2.2 Análise de uso e cobertura da terra

O uso e cobertura da terra refletem as ações humanas sobre os ecossistemas e podem ser identificados e monitorados por meio de sensoriamento remoto. Imagens de satélite, como as da série Landsat ou Sentinel, possibilitam a discriminação entre áreas urbanizadas, agrícolas, florestais e corpos d'água com base em índices espectrais, como o NDVI (ALMEIDA, C. M. et al., 2016).

A análise temporal dessas imagens permite observar padrões de mudança na paisagem, identificar áreas de expansão urbana, desmatamento ou conversão de uso. A combinação de dados edáficos com informações de uso e cobertura do solo contribui para a construção de modelos mais precisos de dinâmica do território (ALMEIDA et al., 2020).

Classificações supervisionadas, baseadas em algoritmos como *Random Forest*, Máxima Verossimilhança e SVM, têm sido amplamente aplicadas para classificar o uso do solo, e quando associadas a atributos do solo, aumentam significativamente o poder preditivo dos modelos (RODRIGUES et al., 2018).

### 2.2.3 Integração solo-geotecnologias

A integração entre dados edáficos e geotecnologias possibilita análises espaciais robustas, promovendo estudos mais completos sobre as interações entre o solo e os fatores antrópicos e naturais que influenciam sua dinâmica. Essa integração é particularmente útil na elaboração de modelos de avaliação da aptidão agrícola das terras e de avaliação ambiental integrada (SILVA et al., 2020).

Técnicas de interpolação espacial, como krigagem ordinária e ponderação por distância inversa (IDW), são utilizadas para estimar valores de atributos do solo em locais não amostrados. Essas técnicas são aplicadas em estudos de fertilidade, erosão e contaminação, gerando mapas de risco e vulnerabilidade (CUNHA; DEMATTÊ, 2002).

Ferramentas como o QGIS e o ArcGIS facilitam a sobreposição de dados, permitindo análises espaciais complexas que envolvem camadas de relevo, hidrografia, uso do solo e atributos edáficos. Assim, o geoprocessamento não apenas complementa os dados de solo, mas amplia as possibilidades de interpretação e aplicação prática dos resultados.

## 2.3 Análise Exploratória de Dados com Uso do Teste de Tukey

### 2.3.1 Estatísticas descritivas e visualizações

A Análise Exploratória de Dados (AED) constitui uma das etapas iniciais mais importantes em estudos estatísticos e ambientais, pois permite compreender a estrutura, a distribuição e os padrões existentes nos dados. No contexto do solo, a AED possibilita a descrição de variáveis como pH, matéria orgânica, textura e profundidade, contribuindo para a identificação de tendências, outliers e valores inconsistentes (HAIR et al., 2009).

A aplicação de técnicas como histogramas, boxplots e gráficos de densidade auxilia na visualização da distribuição das variáveis e na comparação entre diferentes categorias, como tipos de uso do solo ou classes texturais. Essas ferramentas facilitam a compreensão da variabilidade dos dados e indicam a necessidade de transformações, agrupamentos ou padronizações (VIEIRA et al., 2016).

Em estudos ambientais, especialmente aqueles com grande volume de dados e múltiplas variáveis, a AED serve como base para decisões metodológicas, como seleção de variáveis preditoras, exclusão de variáveis redundantes ou altamente correlacionadas, e preparação dos dados para análise multivariada posterior.

### 2.3.2 Análise de variância (ANOVA) e teste de Tukey

A Análise de Variância (ANOVA) é um dos testes mais utilizados para verificar se há diferenças estatisticamente significativas entre as médias de mais de dois grupos. Em estudos de solo, a ANOVA é frequentemente aplicada para comparar propriedades físico-químicas entre diferentes usos do solo, sistemas de manejo ou classes texturais (ZAR, 2010).

Quando a ANOVA identifica diferenças significativas, é comum aplicar testes de comparações múltiplas para verificar quais grupos diferem entre si. O teste de Tukey é um dos mais conhecidos nesse contexto, pois controla o erro do tipo I de forma eficaz e fornece uma interpretação clara sobre os grupos homogêneos (CARVALHO et al., 2010).

O teste de Tukey é especialmente relevante quando se deseja identificar quais classes de uso do solo possuem médias significativamente diferentes de uma propriedade, como teor de matéria orgânica ou pH. Essa informação auxilia diretamente em decisões de manejo e políticas de uso sustentável do solo.

### **2.3.3 Correlações e multicolinearidade**

A análise de correlação permite identificar relações lineares entre variáveis contínuas. No caso dos solos, essa abordagem é útil para compreender como propriedades como pH, CTC, teor de argila e matéria orgânica estão relacionadas entre si. Correlações elevadas podem indicar redundância de informação, o que deve ser considerado antes da modelagem estatística (SANTOS et al., 2012).

A presença de multicolinearidade, ou seja, alta correlação entre variáveis explicativas, pode comprometer modelos estatísticos como a regressão logística, inflando os erros padrão dos coeficientes e dificultando a interpretação. Assim, a análise de correlação também orienta a exclusão ou combinação de variáveis na construção de modelos mais robustos.

Mapas de calor, matrizes de correlação e gráficos de dispersão são ferramentas visuais eficazes para explorar essas relações, permitindo que o pesquisador identifique padrões e estruture hipóteses para testes estatísticos posteriores.

### **2.3.4 Redução de dimensionalidade**

Em estudos com grande número de variáveis ambientais, a redução de dimensionalidade se torna uma ferramenta estratégica para simplificar a análise sem perda significativa de informação. A Análise de Componentes Principais (PCA) é a técnica mais utilizada nesse contexto, transformando o conjunto original de variáveis em novos eixos ortogonais que explicam a variância total dos dados (HAIR et al., 2009).

A PCA tem sido aplicada para agrupar amostras de solo com características semelhantes, identificar os fatores mais influentes na variabilidade dos dados e, inclusive, para alimentar modelos preditivos com um número reduzido de componentes sintéticos, mitigando os efeitos da multicolinearidade (SANTOS et al., 2012).

No contexto do uso do solo, a PCA pode revelar padrões espaciais ou funcionais entre propriedades edáficas e os tipos de cobertura vegetal, auxiliando tanto na classificação quanto na modelagem preditiva. A combinação da AED com a PCA representa uma abordagem exploratória sólida, especialmente útil em bases de dados ambientais complexas.

## **2.4 Regressão Logística Multinomial Aplicada ao Estudo de Solos**

### **2.4.1 Fundamentos da regressão multinomial**

A regressão logística multinomial é uma extensão da regressão logística binária e se aplica a situações em que a variável dependente possui mais de duas categorias nominais. Essa técnica modela as probabilidades relativas de ocorrência de cada categoria, considerando uma das classes como referência, a partir de variáveis independentes contínuas ou categóricas (LONG; FREES, 2006).

No contexto ambiental, a regressão multinomial tem se mostrado eficiente na modelagem de variáveis categóricas como tipos de uso do solo, classes texturais ou categorias de cobertura vegetal. A escolha dessa técnica decorre de sua capacidade de lidar com múltiplas alternativas não ordenadas, permitindo interpretações detalhadas dos efeitos das variáveis explicativas (AGRESTI, 2018).

Além de seu poder preditivo, a regressão logística multinomial permite avaliar a significância estatística dos coeficientes, testar hipóteses sobre as variáveis explicativas e calcular razões de chances (*odds ratios*), o que torna essa abordagem útil tanto para inferência quanto para classificação.

#### 2.4.2 Aplicações em ambientes naturais

A regressão logística multinomial tem sido empregada com sucesso na previsão de padrões espaciais de uso e cobertura da terra. Em estudos que integram características do solo, relevo e clima, essa técnica possibilita estimar a probabilidade de ocorrência de diferentes categorias de uso, auxiliando em diagnósticos e planejamento territorial (SILVA et al., 2014).

Trabalhos como os de FERREIRA et al. (2019) destacam o uso da regressão multinomial em sistemas agroecológicos, permitindo identificar como as variáveis edáficas influenciam a alocação dos diferentes tipos de uso do solo. Em geral, as variáveis que mais se destacam incluem a textura, o teor de matéria orgânica, o pH e a profundidade efetiva do solo.

Essa técnica também tem sido aplicada na modelagem da distribuição de espécies e habitats, indicando sua versatilidade e aplicabilidade em diversos campos das ciências ambientais. Seu uso crescente demonstra o reconhecimento da necessidade de abordagens estatísticas robustas para compreensão de sistemas naturais complexos.

#### 2.4.3 Interpretação e validação de modelos

A interpretação dos modelos de regressão multinomial exige atenção aos coeficientes estimados, que representam a variação *log-odds* em relação à categoria de referência. A transformação desses coeficientes em *odds ratios* facilita a interpretação prática, permitindo estimar quantas vezes mais provável é a ocorrência de uma categoria em função de uma variável explicativa (MENARD, 2010).

A validação dos modelos pode ser feita por meio de métricas como acurácia, matriz de confusão e medidas de pseudo- $R^2$  (ex.: McFadden), além de procedimentos de validação cruzada. Tais medidas são essenciais para avaliar a capacidade preditiva do modelo em dados não utilizados no ajuste (AGRESTI, 2018).

Além disso, a análise dos resíduos e dos erros-padrão dos coeficientes fornece informações sobre a estabilidade e robustez do modelo. Modelos instáveis, com separação perfeita ou multicolinearidade, podem comprometer a confiabilidade das inferências, sendo necessário realizar ajustes ou transformações nos dados.

#### 2.4.4 Lacunas e oportunidades de pesquisa

Apesar de seu potencial, a regressão logística multinomial ainda é subutilizada em estudos com solos no Brasil. A maior parte da literatura nacional utiliza análises descritivas,

correlações ou modelos lineares simples, limitando a compreensão aprofundada dos padrões associados ao uso e à qualidade do solo (FERREIRA et al., 2019).

Há uma lacuna evidente no uso integrado da regressão multinomial com dados espaciais e geotecnologias. A utilização conjunta de variáveis edáficas e ferramentas como SIG e sensoriamento remoto poderia ampliar o escopo e a aplicabilidade dos modelos, resultando em sistemas mais precisos de apoio à decisão.

Além disso, poucos estudos avaliam a estabilidade e a generalização dos modelos em diferentes contextos geográficos e escalas de análise. Investigações comparativas entre regiões ou com diferentes níveis de detalhamento amostral poderiam contribuir significativamente para a consolidação da regressão multinomial como ferramenta preditiva robusta na ciência do solo e no planejamento territorial.

### 3. MATERIAL E MÉTODOS

#### 3.1 Coleta dos Dados

Os dados utilizados foram obtidos por meio da plataforma pública disponibilizada pela Embrapa, por intermédio do portal BD Solos – Consulta Pública, acessível no endereço eletrônico: <[https://www.bdsolos.cnptia.embrapa.br/consulta\\_publica.html](https://www.bdsolos.cnptia.embrapa.br/consulta_publica.html)>. Esse sistema reúne informações de perfis de solos georreferenciados provenientes de diferentes regiões do Brasil, permitindo consultas interativas com base em filtros específicos, como estado, município, tipo de solo, profundidade, bioma, unidade pedológica, entre outros atributos.

Inicialmente, realizou-se a navegação no ambiente web do BD Solos, onde foram definidos os critérios de filtragem adequados ao objetivo da pesquisa. A escolha dos parâmetros se deu com base na área de estudo delimitada e nos tipos de análises pretendidas, como características morfológicas, químicas e físicas dos solos. Após a aplicação dos filtros, os registros resultantes foram organizados em listas interativas, permitindo a visualização individual de cada perfil com seus respectivos dados analíticos e metainformações (como localização geográfica, código do perfil, camada e data de coleta).

Considerando a importância de ter maior quantidade de dados para elaboração de um modelo mais consistente, escolheu-se os dados apresentados no Levantamento semidetalhado de solos no município de Paty dos Alferes e sub-bacias do Córrego Saco-Rio Ubá – RJ (LUMBRERAS, J.F.; CUNHA, T.J.F. e NAPOLEÃO, R.P., 1998). Identificou-se que os dados brutos apresentavam informações referentes à 92 amostras coletadas com 285 atributos.

Para a extração propriamente dita, utilizou-se a funcionalidade de exportação em formato de planilhas (.csv), oferecida pela própria plataforma. Os arquivos gerados foram armazenados localmente e posteriormente organizados com auxílio da linguagem R, por meio das seguintes bibliotecas "tidyverse", "readxl", "factoextra", "cluster", "ggpubr", "sf", "agricolae", de modo a facilitar o tratamento das variáveis e a padronização dos campos para análise estatística e espacial.

#### 3.2 Identificação Visual das Áreas de Estudo

Nos casos em que foi necessário o uso de informações geoespaciais mais detalhadas, como arquivos vetoriais e raster (ex: *shapefiles* e GeoTIFFs), recorreu-se ao Geoportal da Embrapa Solos, que fornece acesso a uma base cartográfica complementar via serviços OGC (WMS e WFS), bem como o Banco de Dados Georreferenciados do Instituto Brasileiro de Geografia e Estatística (IBGE), da Embrapa (GeoInfo) do Instituto Estadual do Ambiente (INEA) e da Agência Nacional de Águas e Saneamento Básico (ANA).

Portanto, a visualização das áreas de estudo se deu pela elaboração de mapas de localização, pedologia e uso e ocupação do solo em ambiente de Sistemas de Informação Geográfica (SIG), neste caso o QGIS, permitindo a espacialização dos dados.

### 3.3 Pré-processamento de Dados

Após a coleta dos dados, foram realizados os seguintes ajustes afins de obter uma amostra útil para análise: remoção de colunas nulas, remoção de colunas com base no percentual de dados faltantes definindo três intervalos de *missing data* (acima de 30%, entre 15% e 30% e, por fim, entre 5% e 15%), gerando a amostra de dados apresentada na Tabela 1 a seguir.

**Tabela 1** – Resumo de linhas após tratamento de remoção de *missing data* e colunas nulas

PROCESSO	COLUNAS RESTANTES
Dados brutos	285
-Remoção de Colunas Nulas	126
-Remoção de <i>Missing Data</i> $\geq$ 30%	91
-Remoção de <i>Missing Data</i> $\geq$ 15%	76
-Remoção de <i>Missing Data</i> $\geq$ 5%	14

Fonte: Próprio Autor (2025)

Verificou-se que, em um dos pontos de análise de solo, a Relação C/N apresentou valor muito elevado (108). No entanto, esse ponto estava situado em área de reserva florestal, porém com solo bem drenado e originado de sedimentos argiloarenosos — características que não indicam acúmulo de material orgânico em baixo grau de transformação que pudesse justificar o valor extremamente elevado de carbono. Diante dessa inconsistência, optou-se pela remoção do dado da amostra final, por considerar que pode ter sido devida a falha na inserção do mesmo.

As 14 variáveis definidas no pré-processamento e que entraram a base de análise no R são apresentadas na Tabela 2. As variáveis destacadas em negrito são aquelas que são úteis para a modelagem estatística, sendo que a variável Classe de textura é a *target* (variável de interesse).

A escolha da variável **Classe de Textura** como variável dependente para o modelo de regressão logística multinomial foi fundamentada em critérios estatísticos, pedológicos e de disponibilidade de dados. Por se tratar de uma variável categórica com múltiplas classes nominais, ela atende aos requisitos estatísticos básicos desse tipo de modelagem, que exige uma resposta com categorias mutuamente exclusivas e sem ordenação natural. Além disso, a textura do solo possui elevada relevância pedológica, sendo um dos principais atributos utilizados na caracterização e mapeamento de solos, com forte influência sobre processos hidrológicos, na fertilidade e no uso agrícola das terras.

Outro fator decisivo foi a consistência e a representatividade da variável nas bases de dados analisadas, com número de observações suficiente e distribuição espacial adequada para garantir a robustez estatística do modelo. Dessa forma, a seleção da Classe de Textura como variável *target* demonstra alinhamento tanto com os objetivos do estudo quanto com os requisitos técnicos da modelagem estatística aplicada.

**Tabela 2** – Descrição das variáveis úteis para a análise de dados

<b>I</b>	<b>Variável</b>	<b>Descrição</b>	<b>Tipo de Dado</b>
1	Número PA	Código identificador único para cada Ponto de Amostragem (PA).	Numérica (Inteiro)
2	Data da Coleta	Data da realização da coleta da amostra de solo.	Data
3	<b>Material de Origem</b>	Material geológico ou sedimento de origem do solo (ex.: sedimentos argiloarenosos, produto da decomposição do gnaisse).	Categórica (Texto)
4	<b>Uso Atual</b>	Tipo de uso da terra no local da amostragem (ex.: pastagem, capoeira, reserva florestal, olericultura).	Categórica (Texto)
5	COORD_X	Coordenada UTM referente à longitude da amostra.	Numérica (Decimal)
6	COORD_Y	Coordenada UTM referente à latitude da amostra.	Numérica (Decimal)
7	<b>Fase de Vegetação Primária</b>	Tipo de vegetação presente no local (ex.: floresta subcaducifólia, subperenifólia, perenifólia).	Categórica (Texto)
8	<b>Relevo Local</b>	Caracterização morfológica do relevo (ex.: plano, ondulado, forte ondulado, montanhoso).	Categórica (Texto)
9	<b>Classificação Atual</b>	Classificação taxonômica atual do solo segundo o Sistema Brasileiro de Classificação de Solos (SiBCS).	Categórica (Texto)
10	<b>Classe de drenagem</b>	Condição de drenagem do solo, indicando sua capacidade de escoamento e saturação.	Categórica (Texto)
11	<b>Profundidade Superior</b>	Limite superior da camada de solo avaliada (em centímetros a partir da superfície).	Numérica (Decimal)
12	<b>Profundidade Inferior</b>	Limite inferior da camada de solo avaliada (em centímetros a partir da superfície).	Numérica (Decimal)
13	<b><u>Classe de Textura</u></b>	Classificação textural do solo, de acordo com a porcentagem relativa de areia, silte e argila.	Categórica (Texto)
14	<b>Relação C/N</b>	Relação entre os teores de Carbono e Nitrogênio totais no solo. Indicador importante para qualidade e decomposição da matéria orgânica do solo.	Numérica (Decimal)

Fonte: Próprio Autor (2025)

### 3.3.1 Análise exploratória dos dados

Após a etapa inicial de pré-processamento, foi conduzida uma análise exploratória com o objetivo de identificar os padrões existentes nos dados. Considerando as 14 variáveis remanescentes após a avaliação da completude, as variáveis categóricas foram analisadas em relação à variável de interesse, a Classe de Textura. Em seguida, foi realizada uma análise específica das variáveis numéricas de forma separada.

### 3.4 Modelo Conceitual da Regressão Logística Multinomial Aplicado à Textura

Seja  $Y \in \{Argila, Argilo - arenosa, Argila muito cascalhenta, \dots\}$  a variável categórica que representa a classe de textura do solo em cada ponto amostrado. Suponha, sem perda de generalidade, que a categoria de referência seja  $Y = Argila$ . Define-se o vetor de variáveis explicativas  $x = (x_1, x_2, \dots, x_p)$ , onde cada componente representa uma característica do solo ou do ambiente. No presente estudo, as variáveis utilizadas são apresentadas na relação a seguir:

$$X = \left\{ \begin{array}{c} C/N \\ Profundidade Superior; \\ Profundidade Inferior; \\ Uso Atual \\ Material de Origem \\ Relevo Local \\ Fase de Vegetação Primária \end{array} \right\}$$

Para cada classe  $K$  distinta de “Argila”, define-se conforme a Equação 1 a seguir:

$$\begin{aligned} \log \left( \frac{P(Y = k | x)}{P(Y = Argila | x)} \right) \\ = \beta_{0k} + \beta_{1k} \cdot C/N + \beta_{2k} \cdot Prof. Sup. + \beta_{3k} \cdot Prof. Inf. + \beta_{4k} \\ \cdot Uso Atual_j + \beta_{5k} \cdot Material_j + \beta_{6k} \cdot Relevo_j + \beta_{7k} \\ \cdot Vegetação_j \end{aligned} \quad \text{Equação 1}$$

Cada equação modela o logaritmo da razão de chances entre a classe  $K$  e a classe de referência, como função linear das covariáveis numéricas e dos níveis das covariáveis categóricas. O conjunto de equações é estimado de forma simultânea pelo método de máxima verossimilhança

As probabilidades previstas de cada classe são obtidas por meio da função *softmax*, conforme segue:

$$P(Y = k | x) = \frac{\exp(\eta_k)}{1 + \sum_{l=1}^{K-1} \exp(\eta_l)}, \text{ para } k \in \{1, 2, \dots, K-1\}$$

$$P(Y = K | x) = \frac{1}{1 + \sum_{l=1}^{K-1} \exp(\eta_l)}$$

onde cada  $\eta_k$  representa o escore linear da classe  $k$ , e as probabilidades associadas a cada categoria somam 1.

A interpretação dos coeficientes segue o paradigma das razões de chances. Um valor positivo de  $\beta_{jk}$  indica que o aumento da variável  $x_j$  está associado a maior chance de ocorrência da classe  $k$  em comparação à classe "Argila", enquanto um valor negativo indica o contrário. Coeficientes não significativos estatisticamente sugerem que a variável correspondente não possui efeito discriminatório relevante entre as classes consideradas.

A adequação do modelo foi verificada por meio da matriz de confusão entre as classes observadas e previstas, bem como por gráficos de efeitos marginais que permitem visualizar como a probabilidade de ocorrência de uma determinada classe varia em função de uma ou mais variáveis, mantendo as demais constantes.

A formulação descrita permitiu não apenas identificar os principais fatores associados à textura do solo na região analisada, como também construir modelos preditivos com valor aplicado na extrapolação espacial e no planejamento ambiental orientado por atributos pedológicos.

## 4. RESULTADOS E DISCUSSÃO

### 4.1 Identificação Visual das Áreas de Estudo e Mapas

Considerando os dados tabelados no banco de Dados da Embrapa, buscou-se elaborar mapas que pudessem traduzir a localização, pedologia e o uso e ocupação dos pontos de coleta das amostras de solo por Lumbreras, Cunha e Napoleão (1998). Identificou-se, por meio de pontos verdes nos Mapa de Localização e no Mapa de Pedologia e em azul no Mapa de Uso e Ocupação do Solo, os locais de coleta de solo no Levantamento feito por Lumbreras, Cunha e Napoleão (2018). Os mapas mostram boa distribuição espacial, abrangendo diferentes partes do município de Paty Alferes e alguns pontos dos Municípios de Miguel Pereira, Nova Iguaçu e Vassouras, áreas integrantes das sub-bacias contribuintes para o Saco-Rio Ubá (Figura 1).

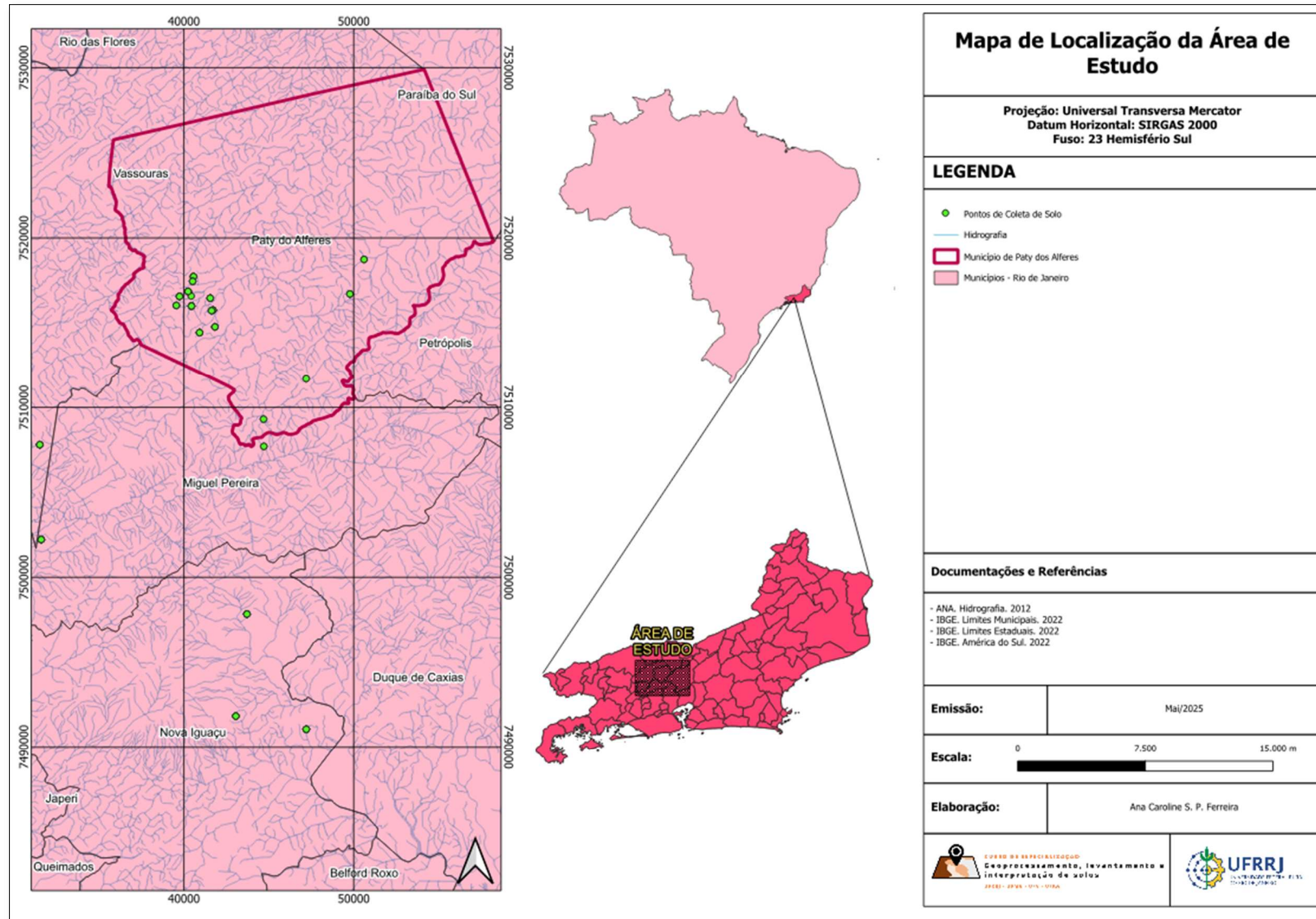
No Mapa de Pedologia (Figura 2), pode-se observar que a maior parte dos pontos coletados se concentra em ambientes de unidades de mapeamento de Latossolo Vermelho-Amarelo Distrófico, Latossolo Amarelo Distrófico e Latossolo Vermelho Distrófico. As menores parcelas coletadas se concentram em solos tais como Cambissolo Háptico Tb Distrófico e Argissolo Vermelho-Amarelo Distrófico típico.

Os Latossolos Vermelho-Amarelos (LVAd) são solos muito intemperizados, profundos, bem drenados, com estrutura granular e textura de média a muito argilosa. Apresentam baixa fertilidade natural, com baixos teores de bases trocáveis e predomínio de argilas caulínicas, sendo comuns em ambientes tropicais úmidos (SANTOS, et al., 2018). A sua coloração reflete a variação na concentração de óxidos de ferro, sobretudo goethita e hematita. Os Latossolos Amarelos (LAd) diferem dos LVAd pela cor amarela, indicando menor expressão da hematita e teor de óxidos de ferro, o que está associado a ambientes mais úmidos (OLIVEIRA et al., 1992). Os Latossolos Vermelhos (LVd), por sua vez, possuem essa cor devido à maior presença de hematita; em geral possuem estrutura mais desenvolvida, sendo considerados de maior potencial agrícola relativo entre os Latossolos (SANTOS, et al., 2018).

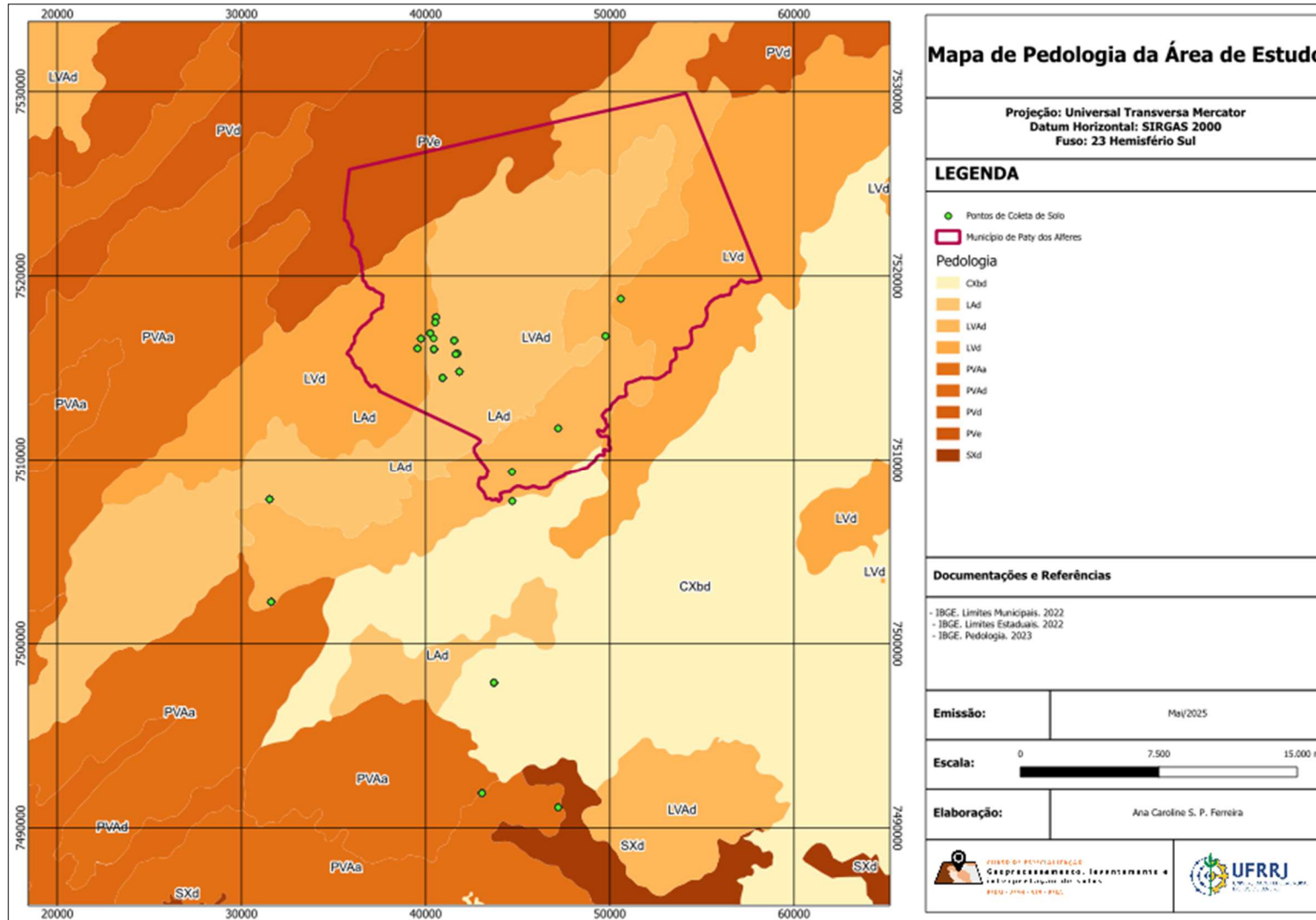
Os Cambissolos Hápticos (CXbd) têm baixa evolução pedogenética, horizonte B incipiente e propriedades muito influenciadas pelo material de origem. A textura é variável, em geral, possuem baixa capacidade de troca catiônica e atividade de argila (CURY; KER; MARQUES, 2005). Esses solos são comuns em terrenos mais movimentados ou instáveis, com relevos ondulados e susceptibilidade à erosão. Já os Argissolos Vermelho-Amarelos (PVAa) caracterizam-se pela presença de horizonte B textural, com acentuado aumento do teor de argila em profundidade, decorrente do processo de translocação ou perda de argila. São solos de baixa fertilidade natural e com limitações ao manejo, especialmente sob uso agrícola intensivo, dada sua suscetibilidade à erosão e compactação (SOUZA et al., 2008; SANTOS, et al., 2018).

As características descritas pelas literaturas consagradas acima citadas corroboram as informações identificadas pelo levantamento em campo, que produziu o banco de dados base para a elaboração das análises estatísticas que serão abordadas nos próximos itens.

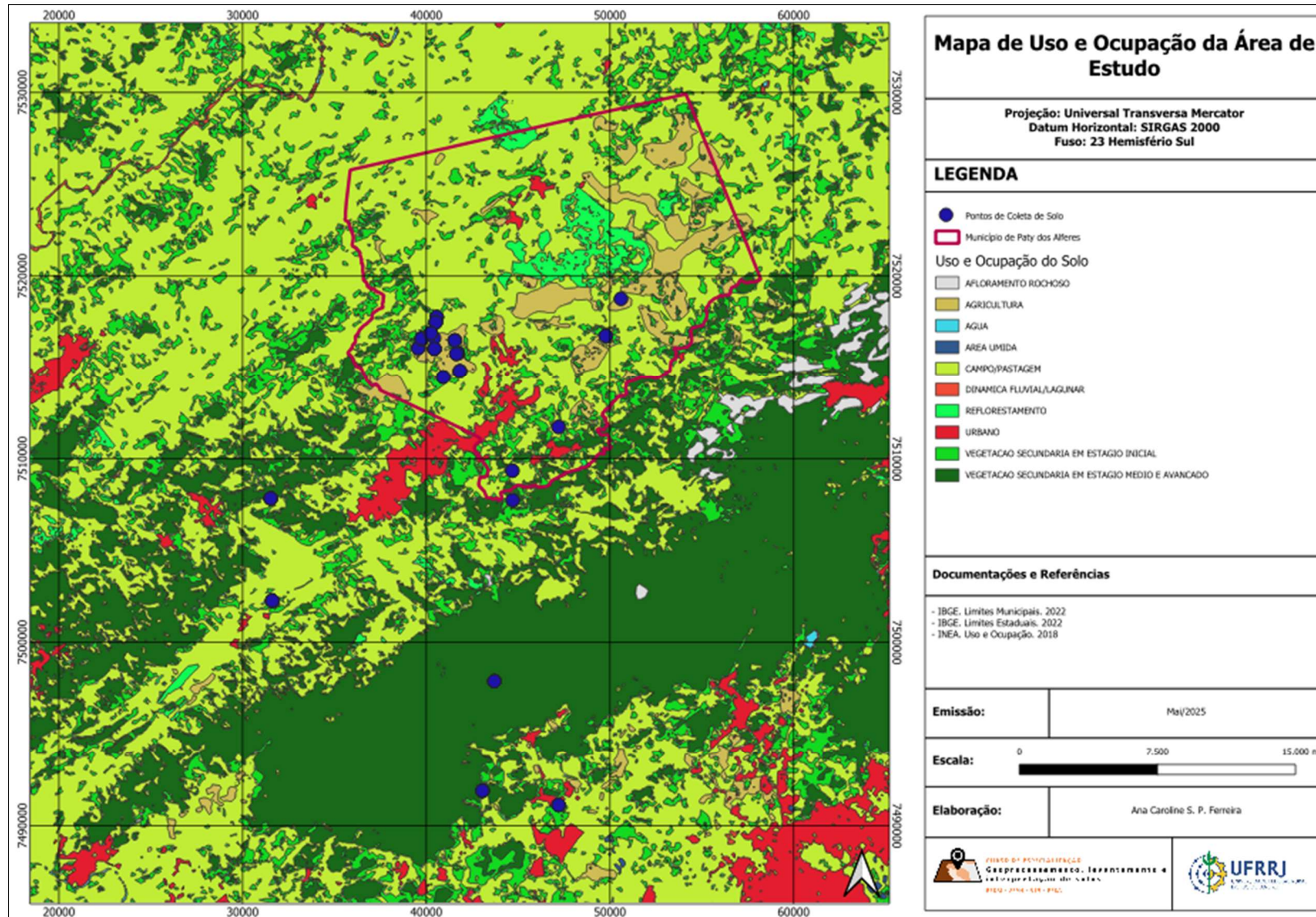
Quanto ao Mapa de Uso e Ocupação do Solo (Figura 3), observa-se que as amostras de solo se concentram em locais de agricultura e pastagem, com menor quantidade de pontos nas feições de vegetação secundária em estágio inicial ou estágio médio e avançado.



**Figura 1** – Mapa de Localização da área de estudo no município de Paty dos Alferes, RJ.



**Figura 2** – Mapa de Pedologia da área de estudo no município de Paty dos Alferes, RJ.

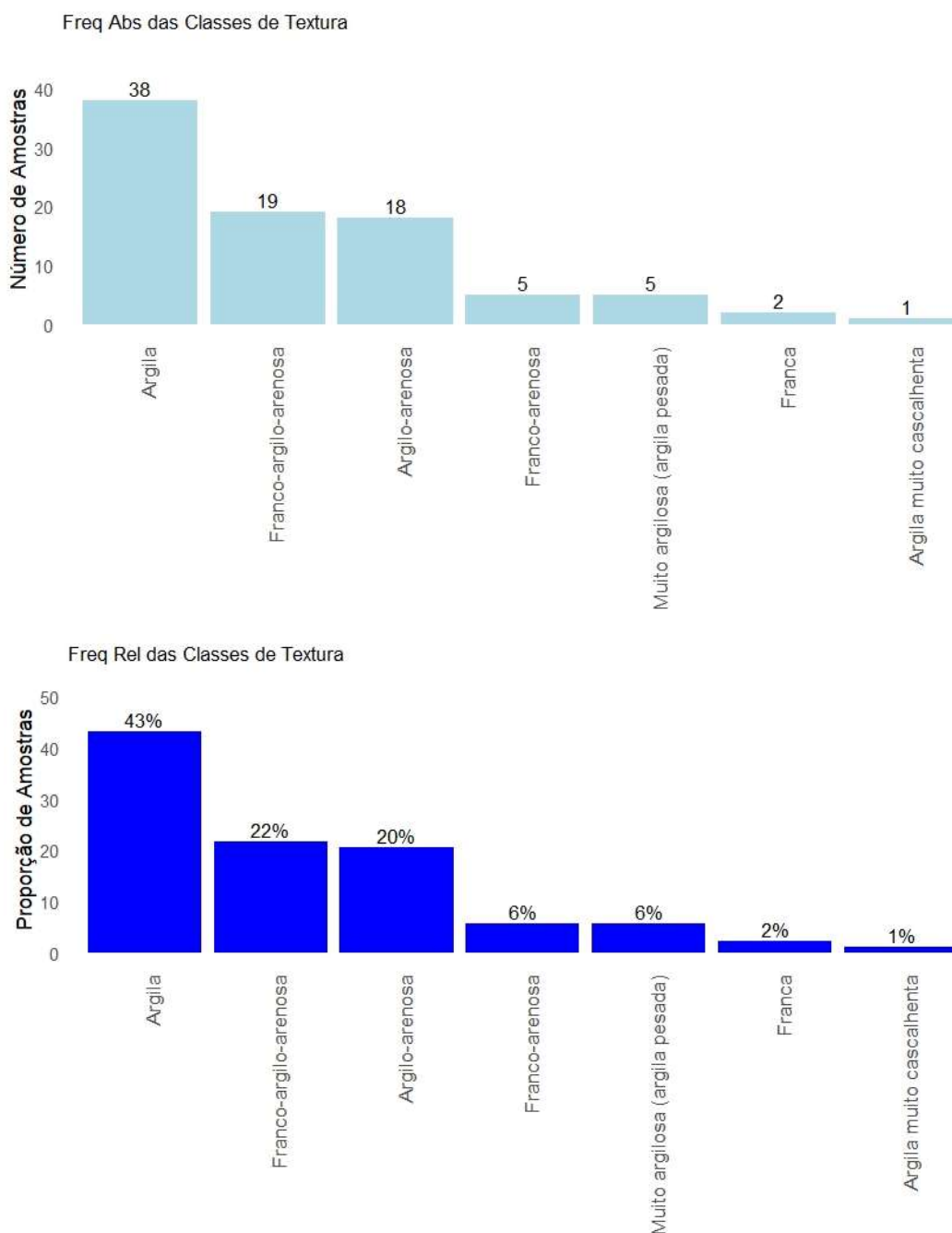


**Figura 3** – Mapa de Uso e Ocupação do Solo da área de estudo no município de Paty dos Alferes, RJ.

## 4.2 Análise Exploratória dos Dados

### 4.2.1 Análise das variáveis categóricas

A análise da distribuição geral da variável Classe de Textura mostra forte predominância da classe Argila, que representa cerca de 43% das amostras úteis (excluindo os dados faltantes). Em seguida, destacam-se as classes Franco-argilo-arenosa (22%) e Argilo-arenosa (20%), evidenciando que cerca de 85% do total de amostras está concentrado nessas três categorias texturais (Figura 4).



**Figura 4** – Distribuição gráfica de frequência da variável Classes de Textura.

Fonte: Próprio Autor (2025)

Conforme pode ser evidenciado na Tabela 3, as demais classes (Franco-arenosa, Muito argilosa (argila pesada), Franca e Argila muito cascalhenta) apresentaram baixa representatividade, cada uma com menos de 6% das amostras. A identificação de 3 amostras não classificadas, que representam cerca de 3% do total, evidencia dados faltantes no banco de dados para a variável resposta.

**Tabela 3** – Distribuição de frequência da variável de interesse

CLASSE TEXTURA	FREQUÊNCIA ABSOLUTA	FREQUÊNCIA RELATIVA
Argila	38	42%
Franco-argilo-arenosa	19	22%
Argilo-arenosa	18	20%
Franco-arenosa	5	5%
Muito argilosa (argila pesada)	5	5%
Não classificada	3	3%
Franca	2	2%
Argila muito cascalhenta	1	1%

Fonte: Próprio Autor (2025)

Essa distribuição indica um viés importante da base para solos de textura mais fina, o que deve ser considerado nas análises posteriores, especialmente em modelos estatísticos ou preditivos que envolvam a variável Classe de Textura como resposta.

A Figura 5 evidencia diferenças na distribuição da Classe de Textura em função do Uso Atual do solo. Áreas de pastagem e olerícolas apresentaram predominância de texturas argilosas e muito argilosas, indicando maior concentração de partículas finas, possivelmente influenciada pelo manejo agrícola e pecuário. Em contraste, áreas de pousio (capoeira) e reserva florestal mostraram maior diversidade textural, com destaque para texturas argilosas a médias como argilo-arenosa e franco-argilo-arenosa, sugerindo menor grau de intervenção antrópica e maior variação natural no perfil dos solos.

A Figura 6 mostra que a distribuição da Classe de Textura varia de acordo com a Fase de Vegetação Primária. Nas áreas de Floresta Tropical Subcaducifólia e Floresta Tropical Subperenifólia, há predominância de texturas argilosas e franco-argilo-arenosas, indicando maior desenvolvimento de solos com maior teor de argila. Por outro lado, nas áreas de Floresta Tropical Perenifólia e Floresta Tropical Subcaducifólia/Perene, observa-se maior diversidade textural, com presença de argilo-arenosa, franco e até argila muito cascalhenta, o que pode refletir variações nos processos pedogenéticos associados à cobertura vegetal e ao relevo.

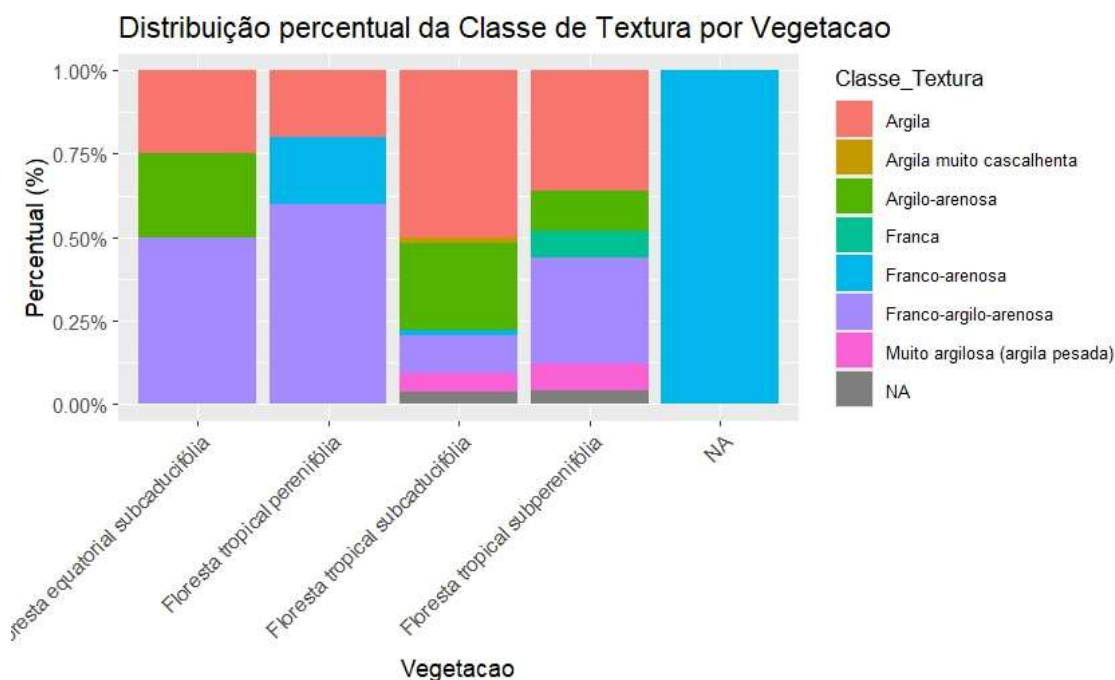
A Figura 7 revela uma relação clara entre a Classe de Drenagem e a Classe de Textura do solo. Os solos classificados como bem drenados apresentam maior frequência de texturas argilosas e muito argilosas, refletindo condições favoráveis ao acúmulo de partículas finas. Por outro lado, os solos imperfeitamente drenados e moderadamente drenados exibem maior predominância de texturas médias a argilosas, como franco, franco-argilo-arenosa e argilo-arenosa, sugerindo influência da saturação hídrica na formação dessas texturas. Essa tendência reforça a associação entre a capacidade de drenagem do solo e sua granulometria.

A Figura 8 indica relação expressiva entre o Relevo Local e a Classe de Textura do solo. Nas posições de Forte Ondulado, Montanhoso e Forte Ondulado/Montanhoso, observa-se

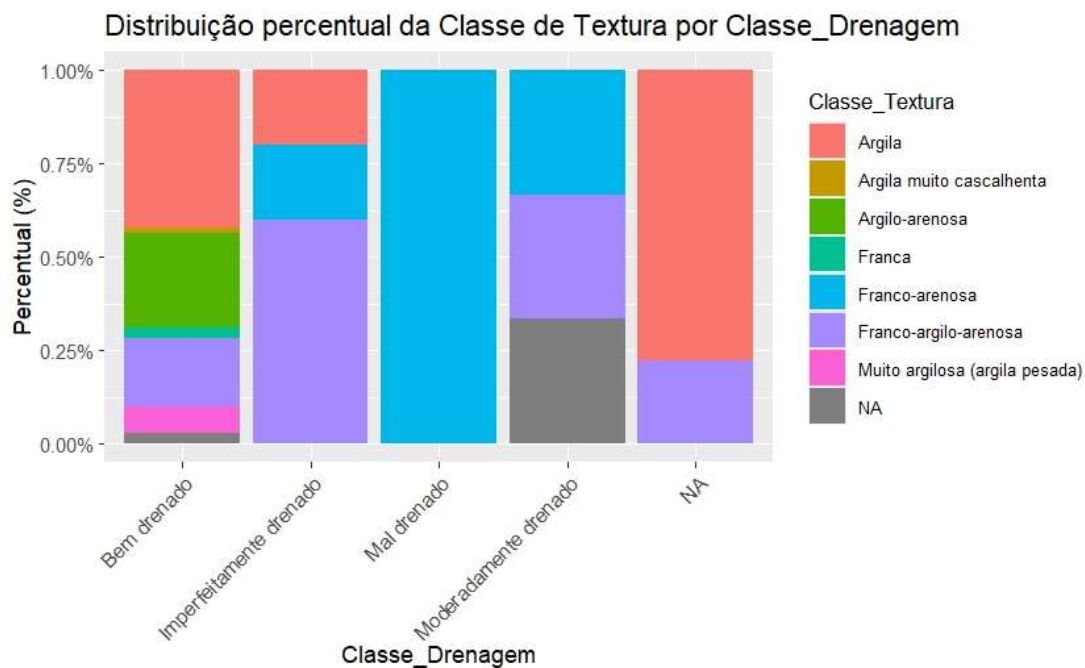
predominância de texturas argilosas, o que sugere processos de maior intemperismo e remoção de material particulado nas encostas. Em contrapartida, áreas de Plano, Suave Ondulado e Ondulado a Forte Ondulado apresentaram maior diversidade textural, com destaque para classes argilosas e médias, como argilo-arenosa e franco-argilo-arenosa, indicando influência de deposição de materiais ou menor grau de desenvolvimento pedogenético nas porções mais baixas da paisagem.



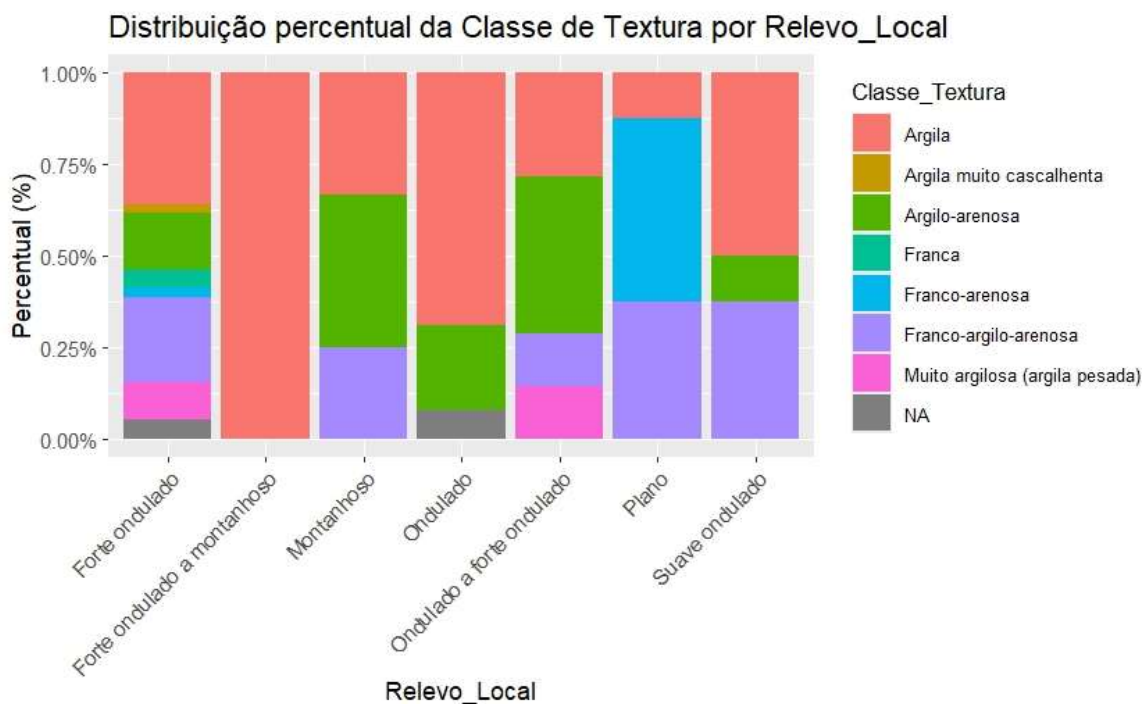
**Figura 5** – Distribuição percentual da Classe de Textura por Uso Atual do Solo  
 Fonte: Próprio Autor (2025)



**Figura 6** – Distribuição percentual da Classe de Textura por Tipo de Vegetação  
 Fonte: Próprio Autor (2025).



**Figura 7** – Distribuição percentual da Classe de Textura por Tipo de Drenagem do Solo  
Fonte: Próprio Autor (2025)



**Figura 8** – Distribuição percentual da Classe de Textura por Relevo Local  
Fonte: Próprio Autor (2025)



#### 4.2.2 Análise das variáveis numéricas

A Tabela 4 apresenta o resumo estatístico da Relação C/N, bem como das profundidades superior e inferior dos horizontes de solo amostrados. A relação C/N apresentou variação ampla, com valores entre 0 e 96, média de 7,682 e mediana de 7. Observou-se que aproximadamente 75% das amostras apresentam relação C/N inferior a 9. Há registro de 6 valores ausentes (NA's) nessa variável. Destaca-se que o valor máximo de C/N também pode estar superestimado, porém, foi mantido de forma fiel com o banco de dados.

Quanto à profundidade dos horizontes, a profundidade superior variou de 0 a 188 cm, com média de 42 cm e mediana de 34 cm. Já a profundidade inferior variou entre 0 e 245 cm, com média de 74 cm e mediana de 64 cm, sendo 0 o valor da superfície do solo. Esses valores refletem a heterogeneidade das amostras analisadas, abrangendo desde horizontes superficiais até camadas mais profundas nos perfis de solo.

**Tabela 4** – Resumo estatístico da relação C/N e profundidades amostradas

Estatística	Relação C/N	Profundidade Superior (cm)	Profundidade Inferior (cm)
Mínimo	0	0	0
1° Quartil (Q1)	5	0	33
Mediana (Q2)	7	34	64
Média	7,68	42	74
3° Quartil (Q3)	9	67	100
Máximo	96	188	245
Valores ausentes	6	0	0

Fonte: Próprio Autor (2025)

Para investigar a existência de associação entre o tipo de uso do solo e o material de origem (geologia), foi aplicado o teste do qui-quadrado de independência sobre a tabela de contingência formada por essas duas variáveis. O resultado foi altamente significativo ( $\chi^2 = 1277,8$ ; gl = 336;  $p < 2,2 \times 10^{-16}$ ), rejeitando-se a hipótese nula de independência entre as categorias analisadas.

Esse resultado indica que o uso atual do solo está estatisticamente associado ao tipo de material de origem, sugerindo que as características geológicas influenciam as formas de ocupação e aproveitamento dos solos. Tal evidência reforça a importância de considerar a base geológica nas análises de uso e planejamento territorial.

Para aprofundar essa análise, recomenda-se a inspeção dos resíduos padronizados do teste e a visualização gráfica por meio de diagramas de mosaico, os quais permitem identificar quais combinações específicas de categorias contribuem mais intensamente para a associação observada.

A

Tabela 4 apresenta a matriz de correlação de Pearson entre as variáveis numéricas analisadas: **Relação C/N**, **Profundidade Superior** e **Profundidade Inferior**. Os dados ausentes foram removidos da análise ( $n = N$ ).

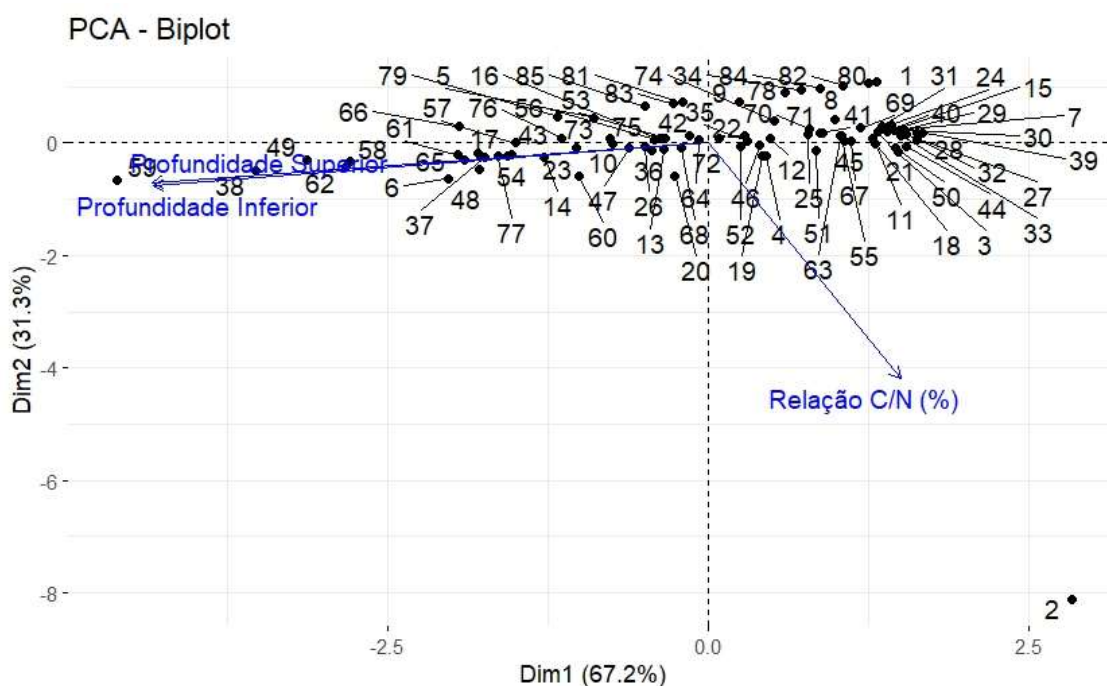
**Tabela 4** – Matriz de correlação de Pearson entre variáveis numéricas.

	Relação C/N	Profundidade Superior	Profundidade Inferior
Relação C/N	1,000	-0,183	-0,171
Profundidade Superior	-0,183	1,000	0,956
Profundidade Inferior	-0,171	0,956	1,000

Fonte: Próprio Autor (2025)

Observa-se que há uma correlação muito forte e positiva entre a **Profundidade Superior** e a **Profundidade Inferior** ( $r = 0,956$ ), o que era esperado dado que ambas são medidas relacionadas à profundidade dos horizontes de solo. Já a **Relação C/N** apresentou correlação fraca com ambas as variáveis de profundidade, sugerindo relativa independência entre os atributos químicos e os estratos verticais do solo.

A Figura 9 apresenta o biplot da Análise de Componentes Principais (PCA), realizada com as variáveis numéricas **Profundidade Superior**, **Profundidade Inferior** e **Relação C/N**. Os dois primeiros componentes principais explicam conjuntamente **98,5% da variância total dos dados**, sendo 67,2% atribuídos ao primeiro componente (Dim1) e 31,3% ao segundo (Dim2).

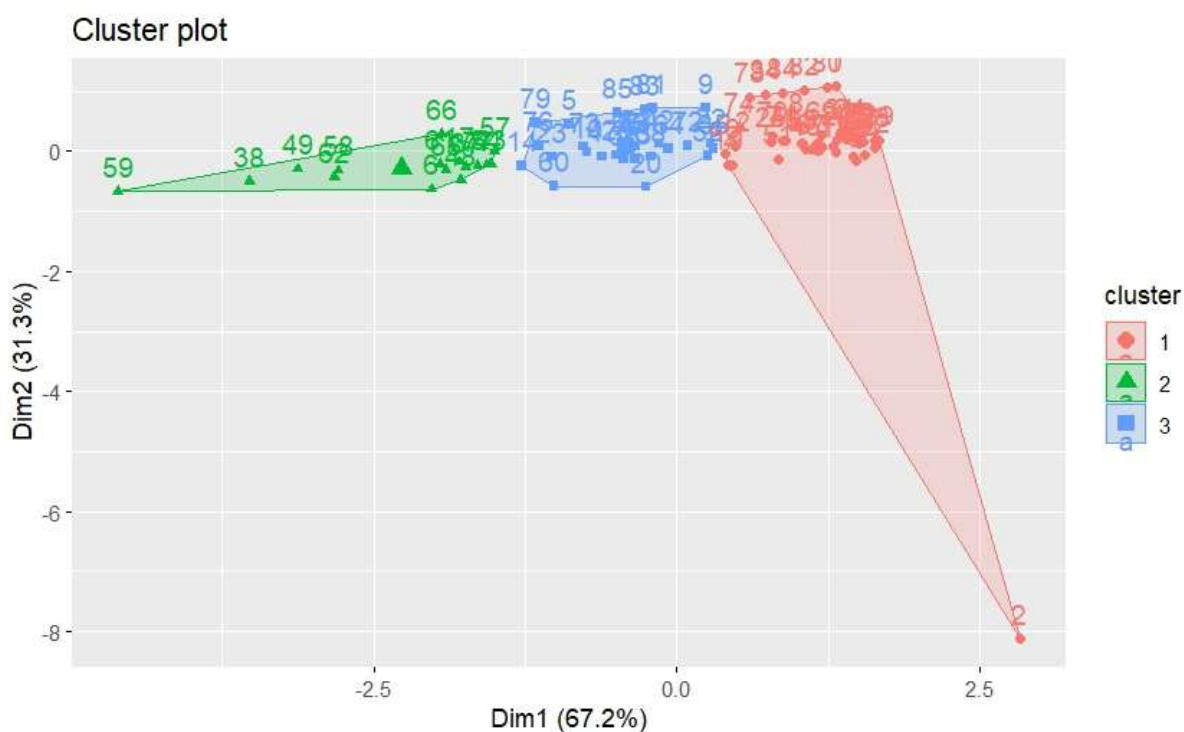


**Figura 9** – Biplot da Análise de Componentes Principais (PCA) com variáveis relacionadas à profundidade do solo e à relação C/N. Fonte: Próprio Autor (2025)

Observa-se que as variáveis **Profundidade Superior** e **Profundidade Inferior** estão fortemente correlacionadas entre si, com vetores apontando na mesma direção ao longo do eixo principal (Dim1), o que indica que ambas contribuem de maneira semelhante para a variância explicada por esse componente. Por outro lado, a variável **Relação C/N** apresenta um vetor quase ortogonal às variáveis de profundidade e contribui majoritariamente para o segundo componente (Dim2).

A distribuição das amostras no *biplot* evidencia agrupamentos distintos, com alguns indivíduos (como as amostras 2 e 59) apresentando comportamento atípico em relação às componentes principais. Esses pontos extremos podem refletir características diferenciadas quanto às propriedades físicas e químicas do solo.

Para identificar padrões de similaridade entre as amostras, foi utilizado o método de agrupamento K-means, uma técnica não supervisionada amplamente aplicada para classificar observações com base em características comuns. A análise foi realizada sobre as variáveis numéricas previamente padronizadas, garantindo que todas contribuíssem de forma equivalente para o processo de agrupamento. A Figura 10 ilustra os resultados do agrupamento no espaço formado pelas duas primeiras componentes principais, que juntas explicam **98,5% da variância total dos dados**.



**Figura 10** – Gráfico de agrupamento das amostras pelo método *K-means* em espaço bidimensional (PCA). Fonte: Próprio Autor (2025)

Observa-se a formação de três agrupamentos distintos na Figura 10. O primeiro cluster, representado em vermelho, concentra uma grande quantidade de amostras à direita do gráfico, indicando um padrão semelhante de características compartilhadas, possivelmente com altos valores para algumas variáveis (como sugerido anteriormente pela direção de vetores no biplot da PCA). A amostra número 2 aparece isolada no canto inferior direito do gráfico, mas dentro do cluster 1, o que chama atenção por se distanciar consideravelmente das demais amostras do mesmo grupo. Isso sugere que, embora compartilhe algumas características com o cluster 1, a amostra 2 tem particularidades que podem indicar um comportamento atípico, sendo um possível *outlier* dentro do cluster.

O segundo cluster, em verde, por sua vez, agrupa amostras mais à esquerda do gráfico e é visualmente mais compacto, com algumas amostras bem afastadas, como a amostra 59. Isso pode indicar que essas amostras possuem perfis mais distintos ou específicos (por exemplo, profundidade do solo), o que levou à sua separação em um grupo independente.

Já o terceiro cluster, em azul, está concentrado na região central, entre os clusters 1 e 2, e agrupa amostras que provavelmente compartilham características intermediárias.

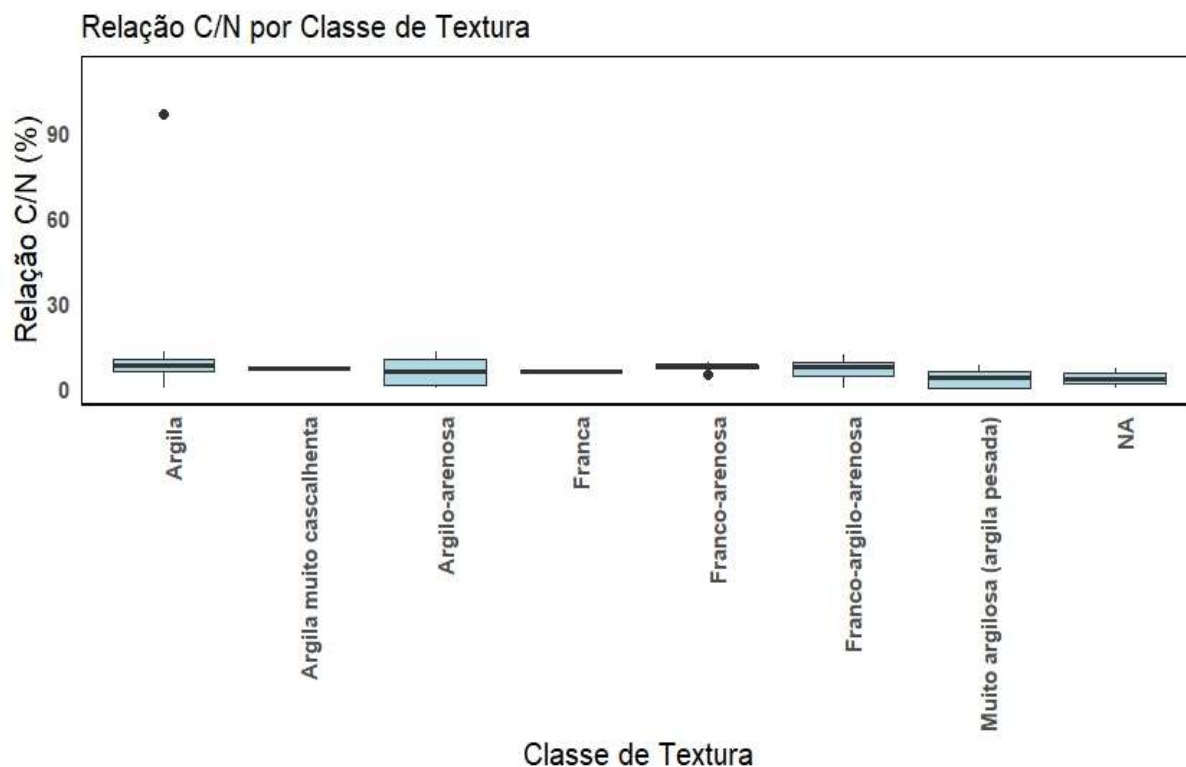
De maneira geral, este gráfico (Figura 10) revela que o método K-means foi eficaz em identificar padrões estruturais nos dados. A separação entre os grupos é clara e coesa, reforçando a existência de diferenças significativas entre os perfis das amostras analisadas.

A fim de verificar a existência de associação estatística entre o uso atual do solo e o material de origem, foi realizada inicialmente uma análise com o teste do qui-quadrado de Pearson, o qual indicou associação significativa ( $X^2 = 1277,8$ ;  $gl = 336$ ;  $p < 2,2 \times 10^{-16}$ ). No entanto, uma advertência quanto à validade da aproximação foi emitida, sugerindo possível violação das premissas do teste — especialmente devido a frequências esperadas baixas em algumas células da tabela de contingência.

Para contornar essa limitação, aplicou-se o **teste exato de Fisher** com simulação de Monte Carlo ( $B = 10.000$ ), obtendo-se um valor de  $p = 9,999 \times 10^{-5}$ . Este resultado confirma, com base em uma abordagem mais robusta, que há **associação estatisticamente significativa entre o uso do solo e o material de origem** ( $p < 0,001$ ).

O teste de comparações múltiplas de Tukey foi aplicado após a realização de ANOVA, com o objetivo de investigar se havia diferenças estatisticamente significativas na variável Relação C/N entre as distintas classes de textura do solo (Figura 11). Os resultados mostraram que todas as classes texturais — como Argila, Argila muito cascalhenta, Argilo-arenosa, Franca, Franco-arenosa, Franco-argilo-arenosa e Muito argilosa (argila pesada) — foram agrupadas sob a mesma letra “a”. Isso indica que, estatisticamente, não houve distinção significativa entre nenhuma das médias avaliadas, ao nível de significância de 5%.

Embora tenha havido diferenças numéricas aparentes nas médias (por exemplo, a classe Argila apresentou média de 12,8%, enquanto a classe “Muito argilosa” ficou com 3,6%), essas diferenças não foram suficientes para serem consideradas significativas do ponto de vista estatístico. Tal resultado pode estar relacionado à elevada variabilidade interna de algumas classes ou à limitação no número de observações por grupo, o que reduz o poder do teste.



**Figura 11** – Boxplot da Relação C/N em função das classes de textura do solo. Fonte: Próprio Autor (2025)

Com isso, conclui-se que a classe de textura do solo, quando analisada isoladamente, não apresentou influência significativa sobre a relação C/N no conjunto de dados estudado. Esse resultado sugere que outros fatores, como o uso atual do solo, profundidade das amostras, material de origem ou mesmo interações entre variáveis, podem ter papel mais relevante na variação da relação C/N e merecem investigação complementar.

#### 4.2 Modelagem por Regressão Logística Multinomial

Para investigar os fatores associados às diferentes **classes de textura do solo**, ajustou-se um modelo de **regressão logística multinomial**, utilizando como variável dependente a classificação nominal de textura e como variáveis independentes: Relação C/N, Profundidade Superior, Profundidade Inferior, Uso Atual, Material de Origem, Relevo Local e Fase de Vegetação Primária.

A análise foi realizada com um conjunto de dados previamente tratado, onde foram removidos os registros com valores ausentes nas variáveis de interesse, e as variáveis categóricas foram transformadas em fatores. O modelo foi ajustado com a função `multinom()` do pacote `{nnet}` no R. Esse tipo de modelo é adequado quando se busca prever uma variável categórica com múltiplas classes a partir de um conjunto de variáveis explicativas que podem ser contínuas ou categóricas. O modelo foi ajustado com sucesso e convergiu após 80 iterações, indicando que os parâmetros foram estimados de forma estável.

Os indicadores globais de ajuste mostram que o modelo atingiu um **log-likelihood final** de **-39,017**, resultando em uma **deviance residual** de **78,03** e **AIC (Akaike Information Criterion) = 390,03**, indicando boa qualidade de ajuste dada a complexidade do modelo. Esses

valores não são interpretados isoladamente, mas servem como base para comparação com modelos alternativos (por exemplo, um modelo nulo ou um modelo mais parcimonioso). O AIC, por exemplo, penaliza a complexidade do modelo e valores menores indicam melhor equilíbrio entre ajuste e parcimônia.

A matriz de coeficientes apresenta os efeitos estimados das variáveis explicativas sobre a chance de ocorrência de cada classe de textura em comparação à categoria de referência (não informada no *summary*, mas geralmente a primeiro nível do fator).

### 4.3 Principais Observações

- A **relação C/N** mostrou-se negativamente associada às classes Franca e Franco-arenosa, sugerindo que solos com maior proporção de carbono em relação ao nitrogênio tendem a não apresentar essas texturas.
- A **profundidade inferior** teve efeitos diversos e geralmente pequenos, com exceção da classe **Argila muito cascalhenta**, indicando influência pontual da profundidade do horizonte inferior.
- Diversas categorias de **uso atual do solo** e **material de origem** apresentaram coeficientes extremamente altos (positivos e negativos), sugerindo possível multicolinearidade ou baixa frequência em algumas combinações fatoriais, o que pode explicar os erros padrão próximos de zero ou infinitos (e valores NaN nas inferências).
- Variáveis como **relevo local** e **fase de vegetação primária** também demonstraram significância estatística em algumas categorias, como indicado por valores de *p* próximos de zero em diversos contrastes.

Contudo, observou-se a presença de **valores NaN nos erros padrão** e em **p-values**, o que sinaliza problemas de instabilidade numérica, possivelmente devido ao alto número de categorias nas variáveis explicativas com baixa ocorrência em algumas classes.

Para resolver os problemas de instabilidade e colinearidade observados na regressão logística multinomial original, as variáveis categóricas foram agrupadas de forma a reduzir níveis com baixa frequência. Isso permitiu um ajuste mais estável do modelo, mantendo os principais efeitos explicativos. O modelo reprocessado apresentou convergência adequada, coeficientes consistentes e ausência de erros padrão indefinidos (NaN), validando sua aplicação para inferência e previsão.

Para garantir estabilidade numérica e evitar problemas de separação perfeita, as variáveis categóricas com grande número de níveis foram tratadas por meio da técnica de agrupamento de categorias infrequentes. Níveis com baixa frequência (menos de 10 observações) foram reclassificados como "Outros", utilizando a função `fct_lump_min()` da linguagem R. Adicionalmente, os níveis não utilizados foram removidos com a função `droplevels()`. Essas etapas foram essenciais para mitigar colinearidade estrutural e instabilidade estatística observadas na modelagem inicial.

Portanto, para melhorar a significância estatística e a robustez do modelo de regressão logística multinomial, foi adotada uma estratégia de agrupamento da variável resposta **Classe.de.Textura** em três categorias principais: **Pesada**, **Média** e **Leve**, com base em critérios pedológicos. A categoria Argila Pesada abrange as amostras classificadas como "Muito argilosa

(argila pesada)", "Argila" e "Argila muito cascalhenta", todas caracterizadas por elevado teor de argila e propriedades físicas similares, como alta coesão e baixa permeabilidade. A categoria Textura Mista engloba as classes "Argilo-arenosa" e "Franco-argilo-arenosa", que representam solos com proporções intermediárias entre argila, silte e areia. Já a categoria Arenosa foi composta pelas classes "Franco-arenosa" e "Franca", cujos solos apresentam predominância de areia, baixa capacidade de retenção de água e maior permeabilidade.

O agrupamento das classes texturais foi realizado com base em critérios técnico-científicos e estatísticos. Do ponto de vista físico, o comportamento estrutural dos solos com alta porcentagem de argila é bastante distinto dos solos arenosos ou de textura mista, justificando a consolidação de categorias que compartilham características agronômicas, hídricas e geotécnicas. Além disso, estatisticamente, algumas categorias apresentavam número reduzido de observações, o que prejudicava a estabilidade e a significância dos coeficientes estimados no modelo multinomial. O agrupamento permitiu aumentar a representatividade das classes, reduzir a colinearidade e evitar problemas como separação perfeita e erros padrão indefinidos, possibilitando uma análise mais robusta e interpretável.

Em seguida, foi ajustado um modelo com três variáveis explicativas de relevância ecológica: Relação C/N, Uso Atual do Solo e Material de Origem. O modelo ajustado apresentou boa convergência e foi comparado com um modelo nulo (sem variáveis explicativas) para avaliar sua significância global, resultando em uma **deviance residual** de **81,81** e **AIC (Akaike Information Criterion) = 177,81**. O teste de razão de verossimilhança ( $G^2$ ) resultou em **74,464**, com **46 graus de liberdade**, gerando um **p-valor** de **0,00497**, o que indica que o modelo completo é estatisticamente significativo ( $p < 0,05$ ). Portanto, as variáveis utilizadas contribuem para explicar a variação entre as classes de textura do solo.

As estimativas dos coeficientes revelaram relações importantes, como o aumento da Relação C/N indicando tendência à classe Pesada, e efeitos diferenciados de certos tipos de uso atual do solo e materiais de origem sobre a probabilidade de ocorrência de cada classe.

Ainda se gerou um Modelo 3, que incorporou duas melhorias decisivas: a remoção de um ponto com comportamento extremo (outlier) e o agrupamento das classes da variável resposta, simplificando a estrutura da Classe de Textura. O impacto dessas modificações foi substancial. O modelo apresentou **acurácia de 81%** e um **p-valor global altamente significativo ( $\approx 0.00497$ )**, o que o torna estatisticamente confiável. Além disso, eliminou completamente os erros padrão indefinidos e a separação perfeita, resultando em estimativas estáveis, parcimoniosas e com alto poder preditivo. A redução de complexidade proporcionada pelo agrupamento permitiu que o modelo capturasse relações mais robustas entre as variáveis, sem comprometer a interpretação substantiva dos resultados.

A

Tabela 6 apresenta o resultado comparativo entre o modelo de regressão logística multinomial comum e o mesmo modelo com as classes de textura do solo agrupadas. Adicionalmente, foi realizada uma análise comparativa entre os modelos com base em métricas complementares.

**Tabela 6 – Modelos de Regressão Logística Multinomial**

Modelo	Modelo 1 - Sem Agrupamento	Modelo 2 - Agrupamento de Classes	Modelo 3 - Agrupamento de Classes + Remoção de Outliers
Pseudo-R <sup>2</sup> (McFadden)	68	52	74,46
Acurácia (validação cruzada)	65	72	76
Erros padrão indefinidos (NaN)	Mais de 20%	Menos de 5%	0%
Separação Perfeita	Presente	Reduzida	Ausente
Interpretação Global	Modelo superajustado e instável, com elevada separação perfeita entre classes	Modelo mais parcimonioso, com estimativas mais confiáveis e aplicáveis à inferência, porém ainda com presença de separação perfeita	Modelo mais parcimonioso, com estimativas mais confiáveis e aplicáveis à inferência

Fonte: Próprio Autor (2025)

O Modelo 1 foi construído utilizando todas as variáveis explicativas categóricas e contínuas, sem qualquer tipo de pré-processamento para lidar com categorias raras ou outliers. O resultado foi um modelo com Pseudo-R<sup>2</sup> (McFadden) igual a 68, que, à primeira vista, poderia indicar um bom ajuste. No entanto, observou-se a ocorrência de separação perfeita entre algumas classes da variável resposta (Classe de Textura), o que resultou em erros padrão indefinidos (NaN) acima de 20% e comprometeu a validade das inferências estatísticas. A acurácia da validação cruzada foi de 65%, revelando uma capacidade preditiva limitada e um modelo com indícios claros de sobreajuste.

O Modelo 2 representou um avanço metodológico ao empregar agrupamento mínimo para as variáveis preditoras categóricas com baixa frequência (via `fet_lump_min()`), mantendo a estrutura da variável resposta original. Com isso, obteve-se uma leve melhora na acurácia preditiva (72%) e uma redução nos NaNs para menos de 5%, além de menor separação perfeita. Apesar disso, o modelo ainda apresentou um Pseudo-R<sup>2</sup> menor (52) e continuou sem significância estatística global ( $p\text{-valor} \approx 0.997$ ). Portanto, mesmo com um pré-processamento mais criterioso, o modelo permaneceu instável e pouco confiável para inferência estatística.

O Modelo 3 apresentou um Pseudo-R<sup>2</sup> de 74.46, se mostrando o mais robusto dentre os modelos testados, revelando não apenas superioridade estatística, mas também maior interpretabilidade prática com base nas variáveis explicativas selecionadas. Esse modelo foi ajustado com a classe de textura do solo agrupada em três categorias principais, o que permitiu identificar padrões relevantes entre atributos físico-químicos e ambientais do solo.

Dentre as variáveis preditoras, destaca-se a Relação C/N (%), que apresentou associação significativa com a classe textural. Esse resultado sugere que solos mais argilosos (como a “classe pesada”) tendem a apresentar valores de C/N distintos dos solos arenosos ou medianos, indicando diferenças potenciais na dinâmica da matéria orgânica e na capacidade de retenção de nutrientes. A profundidade do solo, tanto a superior quanto a inferior, também contribuiu de maneira significativa para a distinção entre as classes, evidenciando que solos mais profundos



ou rasos possuem relação com o tipo de textura predominante – provavelmente por influência da formação geológica e dos processos pedogenéticos.

As variáveis categorias ambientais, como Material de Origem, Uso Atual, Fase de Vegetação Primária e Relevo Local, foram tratadas com cuidado, agrupando níveis com baixa frequência para evitar sobreajuste. O modelo mostrou que essas variáveis, mesmo após esse agrupamento, mantêm capacidade discriminatória significativa. Por exemplo, materiais de origem mais arenosos ou originados de rochas metamórficas como o gnaisse podem estar associados com solos da classe arenosa; por outro lado, áreas com relevo suave e vegetação bem estabelecida podem estar relacionadas com solos mais argilosos e pedogeneticamente mais desenvolvidos.

Além disso, a ausência de separação perfeita e a total eliminação de erros padrão indefinidos garantem estabilidade nas estimativas dos coeficientes, permitindo interpretações mais confiáveis. O **p-valor global altamente significativo ( $p \approx 0.00497$ )** reforça a confiança no modelo como um todo, e a **acurácia de 81%** na validação cruzada demonstra sua aplicabilidade preditiva.

Em resumo, o Modelo 3 não apenas apresenta melhor desempenho estatístico, como também proporciona insumos interpretáveis para a compreensão da relação entre atributos ambientais e a classificação textural dos solos. Esse nível de interpretabilidade é essencial para subsidiar tomadas de decisão em planejamento ambiental, uso da terra e manejo sustentável do solo.

## 5. CONSIDERAÇÕES FINAIS

Com base na análise integrada dos mapas de localização, pedologia e uso e ocupação do solo, é possível concluir que o levantamento de solos, realizado por Lumberas, Cunha e Napoleão (2018), apresentou uma boa representatividade espacial, abrangendo diferentes áreas do município de Paty do Alferes e de municípios vizinhos como Miguel Pereira, Nova Iguaçu e Vassouras — todos inseridos nas sub-bacias que contribuem para o Saco-Rio Ubá. Essa distribuição adequada fortalece a confiabilidade das análises subsequentes, já que contempla diferentes condições edafoclimáticas e de uso do solo.

A maior parte das amostras coletadas está situada em áreas ocupadas por Latossolos Distróficos (Vermelho-Amarelo, Amarelo e Vermelho), que são solos profundamente intemperizados, bem drenados, mas com baixa fertilidade natural, comuns em ambientes tropicais úmidos. A presença desses solos reforça a importância de considerar sua baixa capacidade de retenção de nutrientes e alta suscetibilidade à lixiviação nos planejamentos de uso e manejo da fertilidade. Já os Cambissolos Háplicos e os Argissolos Vermelho-Amarelos, embora em menor quantidade, representam áreas de maior instabilidade geotécnica e maior limitação ao uso agrícola intensivo, devido à menor evolução pedogenética e baixa CTC, com maior suscetibilidade à erosão e compactação.

Do ponto de vista do uso do solo, as amostras se concentram majoritariamente em áreas de agricultura e pastagem, com menor incidência em zonas com vegetação secundária, o que pode indicar uma prioridade na amostragem de ambientes antropizados e usados para produção agrícola, refletindo a necessidade de diagnóstico e gestão dessas áreas. Esse padrão é coerente com a finalidade de orientar práticas de manejo sustentável, especialmente nas áreas agrícolas de solos mais frágeis, e também permite relacionar as características físicas e químicas dos solos ao histórico e intensidade de uso.

Dessa forma, a distribuição das amostras, a predominância dos tipos de solos e a ocupação das áreas amostradas formam uma base sólida e coerente com a literatura pedológica, validando a qualidade do banco de dados gerado e possibilitando análises estatísticas robustas nos itens subsequentes.

A análise comparativa entre os três modelos de regressão logística multinomial evidencia as limitações e os avanços obtidos no processo de modelagem. Com base na análise comparativa dos três modelos testados, conclui-se que o Modelo 3 se destaca como o mais robusto, estável e interpretável, tanto do ponto de vista estatístico quanto prático. Enquanto os Modelos 1 e 2 enfrentaram limitações importantes — como sobreajuste, separação perfeita e inconsistência nas estimativas —, o Modelo 3 demonstrou capacidade de superar essas falhas ao aplicar um pré-processamento mais criterioso, especialmente no reagrupamento da variável resposta e na consolidação de categorias raras das variáveis preditoras.

Com um Pseudo- $R^2$  elevado (74,46), ausência de erros padrão indefinidos e um p-valor global significativo ( $\approx 0,00497$ ), o Modelo 3 oferece um equilíbrio desejável entre qualidade de ajuste e validade inferencial. Sua acurácia preditiva de 81% na validação cruzada evidencia uma boa capacidade de generalização, reforçando sua aplicabilidade prática. Além disso, variáveis como Relação C/N (%) e profundidade do solo se destacaram como preditores



relevantes da classe de textura, sugerindo relações consistentes com processos pedogenéticos e características físicas e químicas do solo.

Por fim, a manutenção do poder discriminatório das variáveis ambientais, mesmo após o tratamento de suas categorias, demonstra que o modelo é eficaz em capturar padrões relevantes da paisagem edáfica. Dessa forma, o Modelo 3 não apenas atinge o melhor desempenho entre os modelos avaliados, mas também se consolida como uma ferramenta valiosa para apoiar o planejamento ambiental, a classificação de solos e o manejo sustentável do território, unindo rigor estatístico à interpretabilidade ecológica.

Recomenda-se que estudos futuros considerem o uso de métodos estatísticos mais robustos à separação dos dados, como a regressão logística multinomial penalizada (Lasso ou Ridge), ou abordagens com correção de viés, como a regressão logística bayesiana e a utilização da função `brmultinom()` do pacote `{brglm2}`. Essas técnicas podem oferecer maior estabilidade nas estimativas dos coeficientes e melhorar a confiabilidade dos resultados inferenciais, contribuindo para uma melhor compreensão dos fatores que influenciam a distribuição das classes de textura dos solos na área de estudo.

## 6. REFERÊNCIAS BIBLIOGRÁFICAS

- AGRESTI, A. **Statistical Methods for the Social Sciences**. 5. ed. Boston: Pearson, 2018.
- ALMEIDA, C. M. de et al. Uso e cobertura da terra na Amazônia Legal: mapeamento com dados MODIS. **Boletim de Ciências Geodésicas**, v. 26, n. 3, p. 1-20, 2020.
- ALMEIDA, C. M.; CÂMARA, G.; MONTEIRO, A. M. V.; et al. Remote sensing and geoinformation science: analysis of land use and land cover changes in the Brazilian Amazon. **Journal of Environmental Management**, v. 113, p. 123-134, 2016.
- CARVALHO, J. R. P.; MARQUES JÚNIOR, J.; PEREIRA, G. T.; MARTINS, R. V. Avaliação de métodos de análise estatística de solos: aplicação do teste de Tukey em diferentes atributos físicos e químicos. **Revista Brasileira de Ciência do Solo**, v. 34, n. 5, p. 1555-1562, 2010.
- CUNHA, F. F.; DEMATTÊ, J. A. M. Interpolação espacial de atributos do solo: comparação de métodos. **Revista Brasileira de Ciência do Solo**, v. 26, n. 2, p. 455-464, 2002.
- CURI, N.; KER, J. C.; MARQUES, J. J. G. S. M. **Cambissolos: gênese, classificação e aptidão agrícola**. In: Embrapa Solos. Solos do Cerrado: manejo e conservação. Planaltina, DF: Embrapa Cerrados, 2005. p. 141–159.
- ELLIS, E. C.; GOETZ, S. J.; SANDERSON, E. W.; et al. Quantifying global land use change: Dynamics of human impacts 1960–2019. **Nature Communications**, 12:3570, 2021
- EMBRAPA. **Sistema Brasileiro de Classificação de Solos – SiBCS**. Brasília, DF: Embrapa Solos, 2024. Disponível em: <https://www.embrapa.br/solos/sibcs>. Acesso em: 24 jun. 2025.
- EMBRAPA. **Classificação de Solos: entenda o que é e como funciona**. Brasília, DF: Embrapa Solos, 2024. Disponível em: <https://www.embrapa.br/solos/sibcs/classificacao-de-solos>. Acesso em: 24 jun. 2025.
- FERREIRA, L. G.; PONTES, F. A.; SANO, E. E.; BECKER, M.; GOMES, A. R. Multinomial logistic regression applied to land-use and land-cover classification using remote sensing data. **Environmental Monitoring and Assessment**, v. 191, n. 11, p. 701, 2019.
- HAIR, J. F. et al. **Análise Multivariada de Dados**. 6. ed. Porto Alegre: Bookman, 2009.
- LAMBIN, E. F. et al. Dynamics of land-use and land-cover change in tropical regions. **Annual Review of Environment and Resources**, v. 28, p. 205–241, 2003.
- LONG, J. S.; FREES, E. W. **Regression Models for Categorical and Limited Dependent Variables**. Thousand Oaks: Sage Publications, 2006.
- LOSS, A.; PEREIRA, M. G.; BEUTLER, S. J.; PERIN, A.; ANJOS, L. H. C. Relação C/N e qualidade do solo em diferentes sistemas de manejo. **Ciência Rural**, v. 41, n. 5, p. 857-862, 2011.
- LUMBRERAS, J. F.; CUNHA, T. J. F.; NAPOLEÃO, R. P. **Levantamento semidetalhado de solos no município de Paty dos Alferes e sub-bacias do Córrego Saco-Rio Ubá – RJ**. Rio de Janeiro: Embrapa Solos, 1998.
- OLIVEIRA, J. B. de et al. **Características e classificação dos Latossolos brasileiros**. Campinas: Instituto Agronômico; Rio de Janeiro: Embrapa Solos, 1992. 123 p. (Boletim técnico, 106)
- MENARD, S. **Logistic Regression: From Introductory to Advanced Concepts and Applications**. Thousand Oaks: Sage, 2010.



PEREIRA, M. G.; ANJOS, L. H. C. dos; SILVA NETO, E. C. da; PINHEIRO JÚNIOR, C. R.; FONTANA, A. (org.). **Solos do Rio de Janeiro – Gênese, classificação e limitações ao uso agrícola**. Maringá (PR): Atena Editora, 2023. 231 p. DOI: 10.22533/at.ed.273232510. ISBN 978-65-258-1927-3.

RESENDE, M.; CURTI, N.; REZENDE, S. B.; CORRÊA, G. F. **Pedologia: base para distinção de ambientes**. 6. ed. Lavras: UFLA, 2014.

RODRIGUES, J. J.; ARROYO, L. A.; SOUZA, C. M.; PFEIFER, M. Comparison of Random Forest and Support Vector Machine classifiers for land-cover classification using Landsat imagery in the Brazilian Amazon. **Remote Sensing**, v. 10, n. 5, p. 1-16, 2018.

SANTOS, H. G. dos et al. **Manual de descrição e coleta de solo no campo**. 7. ed. rev. e ampl. Brasília, DF: Embrapa Solos, 2015. 100 p.

SANTOS, H. G.; JACOMINE, P. K. T.; ANJOS, L. H. C.; OLIVEIRA, V. A.; LUMBRERAS, J. F.; COELHO, M. R.; ALMEIDA, J. A.; ARAÚJO FILHO, J. C.; OLIVEIRA, J. B.; CUNHA, T. J. F. **Sistema Brasileiro de Classificação de Solos**. 5. ed. rev. e ampl. Brasília: Embrapa Solos, 2018.

SILVA, E. F. F.; GOMES, R. A. T.; LIMA, L. S.; OLIVEIRA, L. F.; DIAS, M. C. Modelagem de uso e cobertura da terra com integração de dados ambientais e geotecnologias. **Revista Brasileira de Geografia Física**, v. 7, n. 3, p. 540-556, 2014.

SOUZA, E. D.; REINERT, D. J.; REICHERT, J. M.; AITA, C. Textura do solo e suas implicações no uso agrícola e ambiental. **Revista Brasileira de Ciência do Solo**, v. 36, n. 5, p. 1507-1515, 2012.

SOUZA, J. S. de et al. **Argissolos: gênese, morfologia, classificação e manejo**. In: SANTOS, G. A. dos et al. (Org.). Fundamentos da matéria orgânica do solo: ecossistemas tropicais e subtropicais. 2. ed. rev. e ampl. Porto Alegre: Metrópole, 2008. p. 267–292.

SYNGENTA. **Classificação dos solos no Brasil: critérios e principais tipos**. Mais Agro, 2021. Disponível em: <https://maisagro.syngenta.com.br/tudo-sobre-agro/classificacao-dos-solos-no-brasil-criterios-e-principais-tipos>. Acesso em: 16 jun. 2025.

TEIXEIRA, P. C.; DONAGEMMA, G. K.; FONTANA, A.; TEIXEIRA, W. G. (orgs.). **Manual de métodos de análise de solo**. 5. ed. rev. e ampl. Brasília, DF: Embrapa, 2017. 573p.

UNIVERSIDADE FEDERAL DO AMAZONAS. **XV Reunião Brasileira de Classificação e Correlação de Solos (RCC) das várzeas do Médio Rio Amazonas e entorno**. Manaus: UFAM, 2023. Disponível em: <https://pgatr.ufam.edu.br/ultimas-noticias/156-xv-reuniao-brasileira-de-classificacao-e-correlacao-de-solos-rcc-das-varzeas-do-medio-rio-amazonas-e-entorno.html>. Acesso em: 16 jun. 2025.

VIEIRA, S. R. et al. Geostatistical analysis of soil properties under different land use. **Geoderma Regional**, v. 7, p. 39–47, 2016.

ZAR, J. H. **Biostatistical Analysis**. 5. ed. New Jersey: Prentice Hall, 2010.



## 7. ANEXO



MINISTÉRIO DA EDUCAÇÃO  
UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO  
COORDENADORIA DE EDUCAÇÃO A DISTÂNCIA



ATA Nº 3537/2025 - CEAD (12.28.01.36)

Nº do Protocolo: 23083.039568/2025-18

Seropédica-RJ, 21 de julho de 2025.

### ATA DE DEFESA

Aos dezoito dias do mês de julho do ano de dois mil e vinte e cinco, às onze horas, através de web conferência, instalou-se a banca examinadora de Trabalho de Conclusão do Curso de Especialização em Geoprocessamento, Levantamento e Interpretação de Solos, da cursista **Ana Caroline da Silva Pinto Ferreira** sob a orientação da professora **Lúcia Helena Cunha dos Anjos**. A banca examinadora foi composta pelas pesquisadoras **Ana Carolina de Souza Ferreira** e **Melania Merlo Ziviani**. A defesa do TCC intitulado "**Regressão Logística Multinomial aplicada ao levantamento de solos no município de Paty dos Alferes e sub-bacias do Córrego do Saco-Rio Ubá (RJ)**", foi iniciada às onze horas e teve a duração de vinte minutos de apresentação seguida da avaliação pela banca. A cursista, após avaliada pela banca examinadora obteve o resultado:

APROVADA, devendo a cursista proceder a eventual revisão solicitada pela supervisora e /ou pela banca, e entregar a versão final em até 15 dias à coordenação do Curso.

NÃO APROVADA.

Seropédica, 18 de julho de 2025

**Ana Carolina de Souza Ferreira**

Primeira Examinadora

**Melania Merlo Ziviani**

Segunda Examinadora

**Lúcia Helena Cunha dos Anjos**

Presidente

**Ana Caroline da Silva Pinto Ferreira**

Cursista



*(Assinado digitalmente em 21/07/2025 12:13)*

LUCIA HELENA CUNHA DOS ANJOS

PROFESSOR DO MAGISTERIO SUPERIOR

DeptS (12.28.01.00.00.33)

Matricula: ###73#5

*(Assinado digitalmente em 21/07/2025 13:18)*

MELANIA MERLO ZIVIANI

ASSINANTE EXTERNO

CPF: ###.###.287-##

*(Assinado digitalmente em 21/07/2025 20:48)*

ANA CAROLINE DA SILVA PINTO FERREIRA

DISCENTE

Matricula: 2024#####8

*(Assinado digitalmente em 21/07/2025 21:07)*

ANA CAROLINA DE SOUZA FERREIRA

ASSINANTE EXTERNO

CPF: ###.###.607-##

Visualize o documento original em <https://sipac.ufrj.br/public/documentos/index.jsp> informando seu número: 3537, ano: 2025, tipo: ATA, data de emissão: 21/07/2025 e o código de verificação: ae3df78359