

**UNIVERSIDADE DE SÃO PAULO**

Escola de Engenharia de São Carlos

**Desenvolvimento de funções de pedotransferência para estimativa de teor de carbono em solos brasileiros**

**Ana Carolina Martins Cidin**

Tese (Doutorado) – Programa de Pós-Graduação em Ciências da Engenharia Ambiental – Área de concentração: Ciências da Engenharia Ambiental

UNIVERSIDADE DE SÃO PAULO  
ESCOLA DE ENGENHARIA DE SÃO CARLOS

ANA CAROLINA MARTINS CIDIN

Desenvolvimento de funções de pedotransferência para estimativa de teor de  
carbono em solos brasileiros

São Carlos

2023

ANA CAROLINA MARTINS CIDIN

Desenvolvimento de funções de pedotransferência para estimativa de teor de carbono em solos brasileiros

Tese apresentada à Escola de Engenharia de São Carlos da Universidade de São Paulo, como requisito para a obtenção do Título de Doutora em Ciências.

Orientador: Prof. Dr. Silvio Crestana

VERSÃO CORRIGIDA

São Carlos

2023

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE TRABALHO,  
POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS  
DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pela Biblioteca Prof. Dr. Sérgio Rodrigues Fontes da  
EESC/USP com os dados inseridos pelo(a) autor(a).

C568d Cidin, Ana Carolina Martins  
Desenvolvimento de funções de pedotransferência  
para estimativa de teor de carbono em solos brasileiros  
/ Ana Carolina Martins Cidin; orientador Silvio  
Crestana. São Carlos, 2023.

Tese (Doutorado) - Programa de Pós-Graduação e  
Área de Concentração em Ciências da Engenharia  
Ambiental -- Escola de Engenharia de São Carlos da  
Universidade de São Paulo, 2023.

1. Solos tropicais. 2. Carbono orgânico. 3. Modelo  
linear. 4. Geoestatística. I. Título.

## FOLHA DE JULGAMENTO

Candidata: Engenheira **ANA CAROLINA MARTINS CIDIN**.

Título da tese: "Desenvolvimento de funções de pedotransferência para estimativa de teor de carbono em solos brasileiros".

Data da defesa: 06/12/2023.

### Comissão Julgadora

### Resultado

**Dr. Silvio Crestana**

(Orientador)

(Empresa Brasileira de Pesquisa Agropecuária/EMBRAPA)

APROVADA

**Prof. Associado Frederico Fabio Mauad**

(Escola de Engenharia de São Carlos/EESC-USP)

APROVADA

**Dr. Ednaldo José Ferreira**

(Empresa Brasileira de Pesquisa Agropecuária/EMBRAPA)

APROVADA

**Dr. José Reinaldo da Silva Cabral de Moraes**

(Sombbrero Seguros)

APROVADA

**Dr. Paulino Ribeiro Villas Boas**

(Empresa Brasileira de Pesquisa Agropecuária/EMBRAPA)

APROVADA

Coordenador do Programa de Pós-Graduação em Ciências da Engenharia Ambiental:  
Prof. Tit. **Marcelo Zalot**

Presidente da Comissão de Pós-Graduação:  
Prof. Titular **Carlos De Marqui Junior**

*À minha filha Maria Julia, minha luz e força,  
aos meus pais, Nelson e Mirian,  
pelo incondicional apoio e incentivo...  
...Aos meus irmãos Frederico, Nelsinho e Pedro Paulo,  
DEDICO.*

## AGRADECIMENTOS

Primeiramente a Deus, pela minha vida, pela minha saúde e por sempre atender aos meus anseios.

Ao professor Dr. Silvio Crestana pela oportunidade, orientação e a satisfação de conviver e trabalhar com um exemplo de pessoa e profissional.

À professora Dra. Maria Leonor Assad, que despertou o meu interesse pela pesquisa e foi a maior incentivadora para iniciar este estudo.

À Milene Corso, estatística da Embrapa Instrumentação, pela paciência e incansável determinação em auxiliar nas metodologias e desenvolvimento dos modelos.

Aos colegas da Pós-Graduação Aline Michele Barbosa, José Reinaldo Morais, Nádia Pereira, Alex Watanabe, Kátia e José Otávio Almeida pelo apoio e momentos de descontração.

À minha família, que sempre enviou alegrias e orações. Obrigada Solange, Ricardo, Giovanna, Jeane, Murilo e Heloísa.

Às amigas de longa data, Ariadne Fernandes e Anahi Andreoli, que se mantiveram ao meu lado, mesmo com as mudanças e os percalços da vida.

Ao meu professor da graduação, Dr. Silvano Abreu (*in memoriam*) que me incentivou a buscar o melhor por mais de 20 anos e sempre reconheceu meu esforço e potencial. Estará sempre em meu coração e pensamentos.

Ao Programa de Pós-Graduação em Ciências da Engenharia Ambiental, seus docentes e demais funcionários, pelos conhecimentos adquiridos, compreensão e auxílios concedidos.

À CAPES pela concessão da bolsa de estudos.

## RESUMO

CIDIN, A. C. M. **Desenvolvimento de funções de pedotransferência para estimativa de teor de carbono em solos brasileiros.** 2023. Tese (Doutorado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2023.

O setor agrícola é a maior fonte de emissão de dióxido de carbono ( $\text{CO}_2$ ) no país e a estratégia global é ampliar a adoção de tecnologias agropecuárias sustentáveis, como a redução das emissões de carbono (C) pelas atividades de uso e manejo do solo, promovendo o aumento dos estoques de C no solo. Em razão disso, a necessidade de medir a quantidade de C nos solos tem sido discutida, com o intuito de se obter estimativas mais precisas de valores de C armazenados para compor inventários e subsidiar pesquisas e ampliar a agricultura de baixo C. O objetivo deste trabalho foi desenvolver funções de pedotransferências (PTFs) para estimar os teores de C no solo, a partir de um banco de dados de solos brasileiros. Foram utilizados teores de areia, silte e pH ( $\text{H}_2\text{O}$ ) como variáveis preditoras e as PTFs foram desenvolvidas a partir de modelo linear (LM), modelo linear generalizado (GLM) e modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados (GLMMs), considerando distribuição gama (GAM) e gaussiana (GAU) e função de ligação log e de identidade, respectivamente. Foram desenvolvidas PTFs, considerando a base de dados completa, os dados agrupados por cada bioma brasileiro, os dados agrupados por bioma e classe de solo e os dados agrupados por bioma, classe de solo e uso do solo. As metodologias GLMMs apresentaram melhores resultados, com  $R^2$  ajustado que variou de 0,40 a 0,96, comparados com LM e GLM. As PTFs para Amazônia, Cerrado, Mata Atlântica subgrupos MA2 e MA3 apresentaram melhores resultados desenvolvidas a partir do GLMMs GAM, refletindo em menores valores de RMSE de calibração e validação. As PTFs para Caatinga, Mata Atlântica subgrupo MA1, Pampa e Pantanal apresentaram melhores resultados quando desenvolvidas por GLMMs GAU. Nas amostras agrupadas por biomas e classes de solo, os resultados das PTFs nas metodologias GLMMs foram superiores aos observados nas PTFs desenvolvidas somente com amostras agrupadas por biomas, com exceção da PTF para Latossolo na Amazônia e Luvissolo na Caatinga. As PTFs apresentaram resultados com tendência a superestimar os teores de C no solo.

Palavras-chave: Solos tropicais. Carbono orgânico. Modelo Linear. Geoestatística.



## ABSTRACT

CIDIN, A. C. M. **Development of pedotransfer functions for estimating carbon content in Brazilian soils.** 2023. Tese (Doutorado) – São Carlos School of Engineering, University of São Paulo, 2023.

The agricultural sector is the largest source of carbon dioxide (CO<sub>2</sub>) emissions in the country, and the global strategy is to expand the adoption of sustainable agricultural technologies, such as reducing carbon (C) emissions from land use and management activities, promoting increased soil C stocks. As a result, the need to measure the amount of C in soils has been discussed to obtain more accurate estimates of stored C values for inventory purposes and to support research and expand low-carbon agriculture. The aim of this study was to develop pedotransfer functions (PTFs) for estimate soil C content based on a database of Brazilian soils. Sand, silt, and pH (H<sub>2</sub>O) contents were used as predictor variables, and the PTFs were developed using linear models (LM), generalized linear models (GLM), and generalized linear mixed models with spatially correlated random effects (GLMMs), considering gamma (GAM) and Gaussian (GAU) distributions, and using log and identity link functions, respectively. PTFs were developed for the complete database, grouped by each Brazilian biome, data grouped by biome and soil class, and data grouped by biome, soil class, and land use. The GLMMs methodologies showed better fit, with adjusted R<sup>2</sup> ranging from 0.40 to 0.96, compared to LM and GLM. The PTFs for the Amazon, Cerrado, and Mata Atlântica subgroups MA2 and MA3 performed better when developed using GLMMs GAM, resulting in lower calibration and validation root mean square errors (RMSE). The PTFs for Caatinga, Mata Atlântica subgroup MA1, Pampa, and Pantanal performed better when developed using GLMMs GAU. In the samples grouped by biomes and soil classes, the results of the PTFs using GLMMs methodologies were better than those observed in the PTFs developed only with samples grouped by biomes, except for the PTF for Ferralsols in the Amazon and Luvisols in the Caatinga. The PTFs tended to overestimate soil C content.

Keywords: Tropical soils. Organic carbon. Linear model. Geostatistics

## LISTA DE FIGURAS

|   |    |
|---|----|
| Figura 1- Etapas da metodologia de desenvolvimento de PTFs para estimar teores de C no solo .....   | 28 |
| Figura 2 - Divisão do bioma Mata Atlântica em três partes, de acordo com as dimensões dos paralelos que indicam as latitudes .....                                  | 30 |
| Figura 3 - Distribuição das amostras por biomas brasileiros .....   | 36 |
| Figura 4 - Distribuição das amostras por subgrupos do bioma Mata Atlântica (MA1, MA2, MA3).....   | 37 |
| Figura 5 - Distribuição das amostras por classes de solos brasileiros .....   | 38 |
| Figura 6 - Distribuição das amostras por classes de usos do solo .....  | 38 |
| Figura 7 - Histograma de frequência de dados de C.....  | 39 |
| Figura 8 - Distribuição gráfica dos resíduos dos modelos desenvolvidos para as PTF geral ..   | 43 |
| Figura 9 - Análise gráfica dos resíduos das PTFs desenvolvidas para o bioma Amazônia .....  | 46 |
| Figura 10 - Análise gráfica dos resíduos das PTFs desenvolvidas para o bioma Caatinga .....   | 49 |
| Figura 11 - Análise gráfica dos resíduos das PTFs desenvolvidas para o bioma Cerrado.....   | 51 |
| Figura 12 - Análise gráfica dos resíduos das PTFs desenvolvidas para o bioma Mata Atlântica subgrupo MA1, MA2 e MA3 .....   | 56 |
| Figura 13 - Análise gráfica dos resíduos das PTFs desenvolvidas para o bioma Pampa .....  | 59 |
| Figura 14 - Análise gráfica dos resíduos das PTFs desenvolvidos para o bioma Pantanal .....   | 61 |
| Figura 15 - Esquema de agrupamentos de amostras de biomas por classes de solo para o desenvolvimento das PTFs.....  | 62 |
| Figura 16 - Análise gráfica dos resíduos das PTFs desenvolvidas para o bioma Amazônia agrupados com Argissolo, CN, GEEP e Latossolo.....                            | 66 |
| Figura 17 - Análise gráfica dos resíduos das PTFs desenvolvidas para o bioma Caatinga agrupados com Argissolo, CN, GEEP, Latossolo e Luvissole.....                 | 71 |
| Figura 18 - Análise gráfica dos resíduos das PTFs desenvolvidas para o bioma Cerrado agrupados com Argissolo, CN, GEEP e Latossolo.....                             | 76 |
| Figura 19 - Análise gráfica dos resíduos das PTFs desenvolvidas para o bioma Mata Atlântica, subgrupo MA1, agrupados com Argissolo, CN, Latossolo e Nitossolo ..... | 80 |
| Figura 20 - Análise gráfica dos resíduos das PTFs desenvolvidos para o bioma Mata Atlântica, subgrupo MA2, agrupados com Argissolo e Latossolo .....                | 83 |

|  |    |
|--|----|
| Figura 21 - Análise gráfica dos resíduos das PTFs desenvolvidas para Pastagem em Argissolo, CN, GEEPP e Latossolo no bioma Cerrado ..... | 87 |
|--|----|

## LISTA DE TABELAS

|  |    |
|--|----|
| Tabela 1 - Atributos levantados na literatura utilizados para o desenvolvimento das PTF para estimar teores de C no solo.....  | 27 |
| Tabela 2 - Agrupamento de classes de solo, segundo o guia simplificado do WRB.....   | 29 |
| Tabela 3 - Intervalo de dados de C distribuídos em faixas .....  | 31 |
| Tabela 4 - Estatística descritiva da base de dados (n = 3675 amostras) .....   | 39 |
| Tabela 5 - Estatística descritiva base de dados para o desenvolvimento de PTFs para estimar teores de C no solo.....   | 41 |
| Tabela 6 - Funções de pedotransferência para estimativa de teor de C no solo desenvolvida a partir de metodologias de regressões lineares e desempenhos avaliados pelos critérios estatísticos (PTF geral).....  | 42 |
| Tabela 7 - Estatística descritiva do subconjunto de dados agrupados do bioma Amazônia (n = 675 amostras).....  | 44 |
| Tabela 8 - Funções de pedotransferência para estimativa de teor de C no solo na Amazônia desenvolvida a partir de metodologias de regressões lineares e respectivos desempenhos avaliados pelos critérios estatísticos .....                                       | 45 |
| Tabela 9 - Estatística descritiva da base de dados do bioma Caatinga (n = 717 amostras) .....  | 46 |
| Tabela 10 - Funções de pedotransferência para estimativa de teor de C no solo no bioma Caatinga desenvolvida a partir de metodologias de regressões lineares e respectivos desempenhos avaliados pelos critérios estatísticos.....                                 | 47 |
| Tabela 11 - Estatística descritiva da base de dados do bioma Cerrado (n = 855 amostras).....   | 49 |
| Tabela 12 - Funções de pedotransferência para estimativa de teor de C no solo no bioma Cerrado desenvolvidas a partir de metodologias de regressões lineares e respectivos desempenhos avaliados pelos critérios estatísticos .....                                | 50 |
| Tabela 13 - Estatística descritiva da base de dados do bioma Mata Atlântica subgrupo MA1, MA2 e MA3 .....  | 52 |
| Tabela 14 - Funções de pedotransferência para estimativa de teor de C no solo no bioma Mata Atlântica, subgrupos MA1, MA2, MA3 desenvolvidas a partir de metodologias de regressões lineares e respectivos desempenhos avaliados pelos critérios estatísticos..... | 54 |
| Tabela 15 - Estatística descritiva da base de dados do bioma Pampa (n = 66 amostras) .....   | 57 |

|  |    |
|--|----|
| Tabela 16 - Funções de pedotransferência para estimativa de teor de C no solo no bioma Pampa desenvolvida a partir de metodologias de regressões lineares e respectivos desempenhos avaliados pelos critérios estatísticos .....   | 58 |
| Tabela 17 - Estatística descritiva da base de dados do bioma Pantanal (n = 49 amostras).....   | 59 |
| Tabela 18 - Função de pedotransferência para estimativa de teor de C no solo no bioma Pantanal desenvolvida a partir de metodologias de regressões lineares e respectivos desempenhos avaliados pelos critérios estatísticos .....   | 60 |
| Tabela 19 - Funções de pedotransferência para estimativa de teor de C em diferentes classes de solo do bioma Amazônia desenvolvidas a partir de metodologias de regressões lineares e respectivos desempenhos avaliados pelos critérios estatísticos .....                   | 63 |
| Tabela 20 - Funções de pedotransferência para estimativa de teor de C em diferentes classes de solo do bioma Caatinga desenvolvidas a partir de metodologias de regressões lineares e respectivos desempenhos avaliados pelos critérios estatísticos .....                   | 67 |
| Tabela 21 - Funções de pedotransferência para estimativa de teor de C em diferentes classes de solo do bioma Cerrado desenvolvidas a partir de metodologias de regressões lineares e respectivos desempenhos avaliados pelos critérios estatísticos .....                    | 73 |
| Tabela 22 - Funções de pedotransferência para estimativa de teor de C em diferentes classes de solo do bioma Mata Atlântica subgrupo MA1 desenvolvidas a partir de metodologias de regressões lineares e respectivos desempenhos avaliados pelos critérios estatísticos..... | 78 |
| Tabela 23 - Funções de pedotransferência para estimativa de teor de C em diferentes classes de solo do bioma Mata Atlântica subgrupo MA2 desenvolvidas a partir de metodologias de regressões lineares e respectivos desempenhos avaliados pelos critérios estatísticos..... | 82 |
| Tabela 24 - Funções de pedotransferência para estimativa de teor de C em Pastagem sob diferentes classes de solo do bioma Cerrado desenvolvidas a partir de metodologias de regressões lineares e respectivos desempenhos avaliados pelos critérios estatísticos.....        | 85 |

## SUMÁRIO

|         |   |    |
|---------|---|----|
| 1       | INTRODUÇÃO .....  | 14 |
| 2       | REVISÃO BIBLIOGRÁFICA .....   | 16 |
| 2.1     | Impactos ambientais e o carbono no solo.....  | 16 |
| 2.2     | Metodologias de determinação de c no solo.....  | 17 |
| 2.2.1   | Combustão úmida .....   | 17 |
| 2.2.2   | Combustão seca .....  | 18 |
| 2.2.3   | Métodos espectroscópicos.....   | 19 |
| 2.2.4   | Modelos de simulação.....   | 20 |
| 2.2.5   | Sensoriamento remoto.....   | 21 |
| 2.2.6   | Função de Pedotransferência.....  | 22 |
| 2.2.6.1 | Metodologias de desenvolvimento de Funções de Pedotransferência .....                           | 23 |
| 2.2.6.2 | Função de Pedotransferência para estimativa de teores de C no solo .....                        | 24 |
| 3       | METODOLOGIA .....   | 27 |
| 3.1     | Atualização e organização da base de dados de solos.....  | 27 |
| 3.2     | Padronização e análise crítica dos dados.....   | 29 |
| 3.3     | Estabelecimento de faixas de teores de carbono.....   | 30 |
| 3.4     | Desenvolvimento das funções de pedotransferências .....   | 31 |
| 4       | RESULTADOS.....   | 36 |
| 4.1     | Organização da base de dados .....  | 36 |
| 4.1.1   | Análise Estatística da Base de Dados .....  | 36 |
| 4.2     | Desenvolvimento de Função de Pedotransferência .....  | 40 |
| 4.2.1   | Função de Pedotransferência geral (PTF geral) .....   | 40 |
| 4.2.2   | Função de Pedotransfência para cada bioma brasileiro .....                                      | 43 |
| 4.2.2.1 | Amazônia .....  | 44 |
| 4.2.2.2 | Caatinga .....  | 46 |
| 4.2.2.3 | Cerrado.....  | 49 |
| 4.2.2.4 | Mata Atlântica.....   | 52 |
| 4.2.2.5 | Pampa .....   | 57 |
| 4.2.2.6 | Pantanal.....   | 59 |
| 4.2.3   | Função de Pedotransferência para classes de solo nos biomas brasileiros (Bioma x Solo)<br>..... | 62 |
| 4.2.3.1 | Amazônia agrupado com Argissolo, CN, GEPP e Latossolo .....                                     | 62 |

|   |     |
|---|-----|
| 4.2.3.2 Caatinga agrupado com Argissolo, CN, GEPP, Latossolo e Luvissole .....                                | 67  |
| 4.2.3.3 Cerrado agrupado com Argissolo, CN, GEPP e Latossolo.....   | 72  |
| 4.2.3.4 Mata Atlântica.....   | 77  |
| 4.2.3.5 Mata Atlântica 1 agrupado com Argissolo, CN, Latossolo e Nitossolo.....                               | 77  |
| 4.2.3.6 Mata Atlântica 2 agrupado com Argissolo e Latossolo.....  | 81  |
| 4.2.4 Função de pedotransferência para classe e uso do solo nos biomas brasileiros (Bioma x Solo x Uso).....  | 83  |
| 4.2.4.1 Pastagem no Cerrado agrupado com Argissolo, CN, GEPP e Latossolo .....                                | 84  |
| 5 CONCLUSÃO .....   | 90  |
| SUGESTÕES.....  | 91  |
| REFERÊNCIAS .....   | 92  |
| ANEXO A – Recomendações de uso das PTFs de acordo com a disponibilidade de dados .                            | 102 |
| ANEXO B – Ilustrações da sequência do programa para estimar teores de C a partir das metodologias GLMMs ..... | 103 |

## 1 INTRODUÇÃO

A agricultura brasileira vem se destacando mundialmente nos últimos anos, principalmente em razão da contribuição do país no mercado mundial de alimentos que, em 2020, permitiu alimentar a população brasileira, além de mais de 400 milhões de pessoas no mundo (CONTINI; ARAGÃO, 2021).

Recentemente, em paralelo ao aumento da produtividade das principais *commodities* agrícolas (soja, carnes, milho, algodão e produtos florestais) está o aumento das emissões de gases de efeito estufa (GEE) como metano (CH<sub>4</sub>), óxido nitroso (N<sub>2</sub>O) e, principalmente, o dióxido de carbono (CO<sub>2</sub>), decorrentes da mudança de uso do solo que, em 2020, correspondeu a 46% das emissões brutas de GEE do país (SEEG, 2021).

O setor agrícola é a maior fonte de emissão de CO<sub>2</sub> no país (MCTI, 2021), que é um dos principais GEE causadores do aumento do aquecimento global (IPCC, 2006). O CO<sub>2</sub> tem gerado preocupações em relação ao aumento dos percentuais de concentração na atmosfera desde a década de 1990, em consequência do incremento das emissões oriundas das atividades antropológicas (KUMAR et al., 2022).

A estratégia global é ampliar a adoção de tecnologias agropecuárias sustentáveis, como a redução das emissões de carbono (C) pelas atividades de uso e manejo do solo, promovendo o aumento dos estoques de C no solo, haja vista que o solo tem o potencial de armazenar mais C que a vegetação (COTRUFO; LAVALLEE, 2022).

Em razão disso, a necessidade de medir a quantidade de C nos solos tem sido discutida, com o intuito de se obter estimativas mais precisas de valores de C armazenados, sob as diferentes classes e usos do solo. Essas estimativas podem compor um inventário de C e subsidiar pesquisas e planejamentos de ações, visando o manejo e proteção de áreas com elevado potencial de estocar C no solo e ampliar a agricultura de baixo C (BRASIL, 2012), além de ser pré-requisito para iniciativas privadas de crédito de C.

Para o cálculo do estoque de C no solo, é considerado o teor de C no solo, a densidade do solo (DS) e a profundidade da camada do solo. A determinação dos teores de C no solo pode ocorrer pelos procedimentos via úmida e via seca. A metodologia via úmida, apesar de ser um procedimento acessível, não é preciso e requer fatores de correção para compensar a oxidação parcial da matéria orgânica (MO), além de utilizar reagentes tóxicos que produzem rejeitos de difícil descarte (SHAMRIKOVA et al., 2023). Porém, ainda é um dos procedimentos de rotina em laboratórios de análise de solo mais utilizados no país. Recentemente, analisadores



automatizados têm sido utilizados nas determinações de C via seca, com análises rápidas, confiáveis e sem adição de reagentes tóxicos. Porém, o alto custo e a manutenção desses equipamentos limitam a sua aquisição (NAYAK et al., 2019).

Novas abordagens de simulação de estoques de C têm sido propostas, como os modelos Century (PARTON et al., 1987) e Roth C (COLEMAN; JENKINSON, 1996), utilizados para prever os fluxos de C pelos sistemas agrícolas. Porém, os modelos apresentaram discrepâncias e precisaram de ajustes em ambientes tropicais. As técnicas de mapeamento digital do solo têm sido aplicadas para estimativas de reservas de estoque de C em grandes áreas, porém, é um desafio devido à elevada variabilidade e disponibilidade de dados.

Cada método de determinação de C tem vantagens e limitações. Entretanto, idealizam-se estimativas mais precisas, numerosas, economicamente viáveis e que se ajustem aos princípios de agricultura sustentável. Assim, o desenvolvimento de procedimento alternativo de baixo custo econômico e favorável ambientalmente, combinado com adequada precisão e acurácia, pode ser uma opção para a estimativa de C em solos cultivados com diferentes sistemas de produção. Nesse sentido, as funções de pedotransferência (PTFs) podem ser desenvolvidas para estimar a quantidade de C no solo, a partir de dados de atributos facilmente obtidos em análises de solo (BOUMA, 1989).

Em razão da forte demanda por dados de C, visando estimativas de variações de estoque de C, um modelo de estimativa pode ser de grande utilidade. Diversos estudos foram desenvolvidos com o objetivo de quantificar os estoques de C do solo nos biomas brasileiros, buscando conhecer a variação espaço-temporal (MEDEIROS; SANTOS; MAIA, 2022).

Dessa forma, avalia-se a hipótese de que teores de C do solo podem ser estimados por meio de PTFs desenvolvidas a partir de técnicas de regressão linear, utilizando uma base de dados de solos brasileiros.

O objetivo deste trabalho foi desenvolver PTFs para estimar os teores de C no solo, a partir de uma base de dados de solos brasileiros. Os objetivos específicos foram:

- Atualizar a base de dados organizada por Cidin (2016) para o desenvolvimento de PTFs visando estimar C em solos brasileiros;
- Testar e comparar diferentes metodologias de regressão linear, como modelo linear, modelo linear generalizado e modelo misto linear com efeitos aleatórios espacialmente correlacionados para o desenvolvimento das PTFs;
- Desenvolver PTFs específicas para cada bioma brasileiro;
- Desenvolver PTFs para os biomas combinados com as principais classes e uso dos solos.

## 2 REVISÃO BIBLIOGRÁFICA

### 2.1 Impactos ambientais e o carbono no solo

No auge do processo de industrialização e urbanização do Brasil, concentrado nas décadas de 1920 e 1930, o forte crescimento econômico não incluía o setor agrícola, que não dispunha de muitas tecnologias e apresentava baixas produtividades (CHADDAD, 2015). Com o avanço de políticas públicas de incentivo à exploração das terras e das pesquisas científicas, a agricultura foi sendo modernizada e com a utilização de máquinas, insumos agrícolas e mão de obra, ocorreu a expansão de importantes culturas e, conseqüentemente, da atividade pecuária, consolidando o país como grande produtor e exportador de commodities (MAPA, 2022).

Neste contexto, faz-se presente o uso da terminologia “Agricultura 4.0” para definir o atual status da atividade agropecuária brasileira, que utiliza de tecnologias de precisão, com automação e robótica agrícola, *big data*, modelagem e inteligência artificial para elevar os índices de produtividade, contribuir para a eficiência dos insumos e reduzir os impactos ambientais (CUNHA et al., 2022; ZHAI et al., 2020).

O maior impacto ambiental gerado pela agricultura é o aumento das emissões de GEE na atmosfera, principalmente o dióxido de carbono (CO<sub>2</sub>), provocando grandes mudanças nas paisagens, nos regimes hídricos e na temperatura global, influenciando na produtividade das culturas (SINGH et al., 2023). A 4ª Comunicação Nacional do Brasil à Convenção-Quadro das Nações Unidas sobre Mudança do Clima (MCTI, 2021) apresentou os dados de emissões e remoções de CO<sub>2</sub> do Brasil no ano de 2016. O setor agropecuário contribuiu com 33,2% e o setor de mudança de uso da terra e florestas contribuíram com 27,1% do total das emissões, que foi de 1,476 Tg de CO<sub>2</sub>. No mesmo ano, a remoção de CO<sub>2</sub> foi de quase 613 milhões de toneladas (T), contabilizadas somente do setor de mudança de uso da terra.

A atividade agropecuária, considerada a grande “vilã” do atual cenário de emissões de GEE pode, por outro lado, desempenhar um papel estratégico da mitigação dos impactos gerados pelas mudanças climáticas (FRANCAVIGLIA et al., 2023; MAIA et al., 2022).

A perda de C ocorre quando a vegetação nativa é substituída por áreas cultivadas, de acordo com o sistema de preparo do solo, o sistema de cultivo e condições climáticas (COTRUFO; LAVALLEE, 2022). A maior parte do C existente no sistema solo-planta-atmosfera está concentrado no solo (ADIAHA, 2017). Estudos mostraram que até 1m de profundidades, o solo tem potencial de armazenar três vezes mais C que a vegetação e até o

dobro da quantidade contida na atmosfera terrestre (MENEZES et al., 2021; SMITH et al., 2020). Na camada superficial do solo, a quantidade de C armazenada é próxima à da atmosfera, que é de 760 Pg (LAL et al., 2021).

Além de suas propriedades relacionadas com a qualidade do solo, o C no solo tem potencial para mitigar as mudanças climáticas (CHATTERJEE et al., 2022; MINASNY et al., 2010). Logo, C no solo tem sido o foco de pesquisa nas últimas décadas, visando o conhecimento da variabilidade e espacialização dos estoques de C no solo (CONANT et al., 2011; McNUNN et al., 2020). O C armazenado no solo é um dos indicadores-chave na prestação de serviços ambientais promovidos por boas práticas agrícolas (LEDO et al., 2020; OGLE et al., 2019).

Segundo recomendações do IPCC (2003), no protocolo para amostragem de solo para fins de inventário de C, devem ser testados nove pontos de solo para cada parcela, cada um contendo três profundidades amostradas (0-10 cm, 10-20 cm e 20-30 cm). Estas amostras são necessárias para estudar a variabilidade espacial dos parâmetros do solo para a amostragem inicial. No Brasil, a Embrapa introduziu um protocolo para a quantificação dos estoques de C no solo gerados pela rede de pesquisa PECUS, que é voltada para sistemas agropecuários, mas pode ser aplicado em áreas de produção de grãos e florestas (OLIVEIRA, 2014). Sendo assim, há uma urgente demanda por dados de quantificação de C sob diferentes profundidades do solo, condições de cobertura e uso do solo, localizações de interesse, entre outros.

## 2.2 Metodologias de determinação de c no solo

Para analisar o teor de carbono (C), é necessário converter o CO<sub>2</sub> proveniente de todas as substâncias orgânicas presentes no resíduo. Isso pode ser feito por meio de processos como combustão seca ou úmida, digestão, oxidação e dosagem do agente oxidante remanescente. Essa conversão é realizada utilizando técnicas analíticas como titulometria, volumetria e outros procedimentos similares. (SATO et al., 2014).

### 2.2.1 Combustão úmida

A metodologia precursora de determinação via úmida de teores de C foi desenvolvida por Walkley & Black (WB) em 1934, baseada na combustão úmida que utiliza o dicromato de potássio como agente oxidante. O método Yeomans & Bremner (YB) (1988) é uma variação

da análise de WB. Ainda utilizado nos laboratórios do Brasil, é um método de baixo custo, executado com instrumentos simples e que fornece boa exatidão (TEDESCO et al., 1995); porém, como a oxidação do C não é completa, é necessário aplicar fatores de correção (GATTO et al., 2009; HARDY; DUFFEY, 2017). Outra desvantagem desse método é a presença de resíduos do cromo que podem causar poluição no ambiente e por isso, alguns países deixaram de adotar esta metodologia (SEGNINI et al. 2008; SHAMRIKOVA et al., 2023). No procedimento Mebius (NELSON; SOMMERS, 1996, SATO et al., 2014), semelhante ao método proposto por WB, a metodologia foi modificada adicionando uma fonte externa de energia, descartando a necessidade de aplicação de fatores de correção. Outra variação do método WB é o método colorimétrico, onde a determinação final é feita por espectrofotometria e a medição é feita pela intensidade da cor esverdeada dos íons cromo reduzidos pelo C (QUAGGIO; RAIJ, 1979).

Em estudo de comparação de métodos WB, YB e combustão seca para a determinação do C em solos cultivados com eucalipto, com diferentes profundidades e em diferentes regiões, GATTO et al., 2009 concluíram que os métodos de combustão úmida subestimaram os teores de C em relação ao método de combustão seca em todas as camadas analisadas.

### 2.2.2 Combustão seca

A determinação do C pelo método da combustão seca reúne vantagens em relação aos demais métodos, sendo um procedimento rápido, confiável, não poluente e facilmente automatizado. Porém, os custos de análise e manutenção são elevados, limitando a sua utilização em poucos laboratórios do país (CHATTERJE et al., 2009). Outro método utilizado é conhecido como perda de massa por ignição (PMI), que consiste na determinação gravimétrica do C convertido em CO<sub>2</sub>. Que é estimado com base na massa volatilizada de uma amostra de solo seco submetida a altas temperaturas por meio da ignição (SCHULTE; HOPKINS, 1996). Na análise elementar, equipamentos determinam, simultaneamente, o C, o hidrogênio (H), o nitrogênio (N), o enxofre (S) e o oxigênio (O) presentes nas amostras orgânicas e inorgânicas. Esse método consiste na oxidação das amostras à temperatura de quase 1000°C. Outra opção é o analisador TOC (*total organic carbon*) que mede a quantidade de C total. Sleutel et al. (2007), comparando a determinação de C utilizando o TOC e o analisador elementar, afirmaram que o TOC subestimou em 6% a quantidade de C nas amostras.

Kalembasa e Jenkinson (1973) realizaram o primeiro estudo de comparação entre os métodos a partir de combustão seca e os métodos via úmida. Esses autores concluíram que o método via combustão seca é mais eficiente na recuperação do C, assim como Nelson e Sommers (1996), que consideraram os resultados precisos e exatos. Alguns estudos obtiveram resultados que mostraram que valores de C foram baixos quando analisados pelo método WB, comparados com análise elementar (CONCEIÇÃO et al., 1999; COSER et al., 2012; GATTO et al., 2009; RHEINHEIMER et al., 2008; SILVA et al., 1999) e TOC (SEGNINI et al., 2008).

### 2.2.3 Métodos espectroscópicos

Os métodos espectroscópicos têm sido utilizados para substituir métodos laboratoriais tradicionais, buscando a obtenção de dados para o monitoramento dos sistemas de forma rápida e economicamente viável (FERNANDES et al., 2010). Metodologias como o *Laser Induced Breakdown Spectroscopy* (LIBS) (PASQUINI et al., 2007; VISCARRA ROSSEL et al., 2006); espectroscopia no visível (VIS) (RODRIGUES et al., 2016), infravermelho próximo (NIRS) (MARCHÃO; BECQUER; BRUNET, 2011), já foram aplicadas na determinação dos teores de C em amostras de solo.

Senesi e Senesi (2016), em uma revisão crítica sobre as vantagens e desvantagens de quantificar o carbono orgânico do solo (COS) utilizando o LIBS, afirmaram que a vantagem em relação aos métodos tradicionais, como combustão seca e análise elementar, é a facilidade de coleta e preparação das amostras, curto tempo para análise e a possibilidade de analisar muitas amostras diariamente. Os autores ainda citaram a facilidade de instrumento portátil, que pode ser utilizado *in situ*. Porém, as desvantagens estão relacionadas com as dificuldades instrumentais referentes ao efeito de matriz e demanda por volume de dados para calibração, que podem resultar na redução da intensidade do sinal para elementos em concentrações elevadas e com o fato de que algumas propriedades físicas e químicas dos solos, como estrutura, mineralogia, teor de C inorgânico, umidade, entre outros, podem induzir incertezas nas medições.

Do mesmo modo, Costa et al., 2019 revisaram as particularidades do LIBS em diversas áreas da Ciência e discutiram a utilização do procedimento na agricultura, como a determinação e nutrientes na planta (JULL et al., 2018), a determinação de macro e micronutrientes e contaminantes nos fertilizantes e a determinação do grau de humificação do solo (Ferreira et al., 2014).

Stenio et al. (2022) propuseram um modelo matemático para quantificar o C total em amostras de solos brasileiros com diferentes texturas, contornando as interferências espectrais, a partir do desenvolvimento de um método específico de LIBS para remoção de *outliers* com 6% de remoção de espectro, que refletiu em boa precisão do modelo e baixos valores dos erros.

#### 2.2.4 Modelos de simulação

Novas abordagens para simulação de C no solo têm sido adotadas no Brasil, como os modelos Century (PARTON et al., 1987) e Roth C (COLEMAN; JENKINSON, 1996). O Century funciona em escala de metro quadrado e simula a dinâmica de C, N, S e fósforo (P) para camada de 0-20 cm, usando dados mensais de climatologia. Foi desenvolvido, inicialmente, para ecossistemas de pastagens nos Estados Unidos, mas foi ajustado para incluir florestas e sistemas agrícolas em ambientes de clima temperado (PARTON et al., 1988).

Para aplicação em regiões tropicais, os modelos necessitam de ajustes, de acordo com Leite et al. (2003, 2004a, 2004b) que, simulando o efeito de diferentes formas de adubação do preparo do solo e a distribuição do C no solo em Minas Gerais, observaram discrepâncias nas simulações em decorrência da não adição de dados de mineralogia. Do mesmo modo, Bortolon et al. (2011), utilizando o Century para avaliar a dinâmica do C e N em diferentes manejos no Rio Grande do Sul, obtiveram resultados divergentes e consideraram necessários ajustes na variável mineralogia do modelo, bem como a revisão de alguns parâmetros. Os autores ainda afirmaram que o modelo Roth C não tem a mesma flexibilidade que o Century. Cerri et al. (2007) utilizaram os modelos Roth C e Century para simular as mudanças do C no solo em cronosequência com pastagem em 11 localidades na Região Amazônica e concluíram que os modelos apresentaram divergências atribuídas à diferença de textura do solo, da produtividade das gramíneas e o manejo entre as áreas.

De acordo com Mendonça e Leite (2006), esses modelos têm potencial para simular a dinâmica do C em diversos ecossistemas, mas apresentam limitações para regiões tropicais, como velocidade de ciclagem, mineralogia, profundidades maiores de 20 cm, o teor de lignina permanece constante ao longo do processo de decomposição, a impossibilidade de programar culturas consorciadas e ausência do efeito do pH, fundamental para solos tropicais.

Com o objetivo de calibrar e validar o modelo Century para simular a dinâmica dos estoques e fluxos de C e N no sistema solo-planta em áreas de vegetação nativa de Caatinga

sob diferentes práticas de corte e queima, Althoff et al. (2018) afirmaram que o modelo é sensível ao desmatamento e que foi necessária à sua adaptação para representar esse sistema.

Lee e Viscarra-Rossel (2020) utilizaram as estimativas de frações do C, a partir de PTF, mapeamento digital e sensoriamento proximal, para inicializar o Roth C em amostras na Austrália e avaliaram seu efeito no cálculo do estoque de C (0-30 cm) numa simulação de 30 anos. Segundo os autores, resultados mostram que a inicialização do modelo com sensoriamento proximal e mapas digitais do solo representaram bem a variação espacial do C e dos reservatórios de C, logo, a simulação de alterações nas reservas de C foi mais confiável. Porém, utilização das estimativas PTF resultou em previsões enviesadas que subestimaram as reservas de C durante o período de simulação.

#### 2.2.5 Sensoriamento remoto

A agricultura brasileira vem se destacando na implantação de tecnologias visando obter melhores resultados no que se refere à produtividade animal e agrícola, no manejo do solo e água etc., e para atingir os níveis de desempenho, tem apostado no contínuo avanço científico e tecnológico. De acordo com Formaggio e Sanches (2017), o sensoriamento remoto (SR), definido empiricamente como a aquisição de dados sobre um objeto sem tocá-lo (JENSEN, 2009), tem sido utilizado como meio de informação para monitorar dados de áreas com culturas agrícolas plantadas, estimativas de produtividade, avaliação de eventos climáticos sobre as produções agrícolas e a estimativa de biomassas, produtividades entre outros. Entre os satélites e sensores disponíveis, cita-se o Landsat-8, com os sensores OLI (*Operational Land Imager*), o MSS (*Multispectral Scanner System*), o TM (*Thematic Mapper*), o TIRS (*Thermal Infrared Sensor*), Sentinel-2 e o PRISMA.

Wang et al. (2018) avaliaram os modelos de árvores de regressão impulsionada (BRT) e de *random forest* (RF) na previsão de estoques de C em pastagens na Austrália, a partir de variáveis contínuas disponíveis oriundas de imagens do Landsat-8, sensor OLI. Os autores empregaram a regressão *stepwise* para selecionar as melhores variáveis preditoras, entre dados de clima, topografia, radiometria, vegetação e cobertura fracionada da terra e concluíram que a metodologia RF apresentou o melhor desempenho na estimativa.

Angelopoulou et al., 2019 reuniram estudos que estimaram a quantidade de COS com o uso de técnicas de SR e observaram que a precisão da previsão diminuiu dos Sistemas Aéreos Não Tripulados (UASs) para as plataformas de satélite, embora os avanços nas técnicas de

*machine learning* possam ajudar na geração de melhores modelos de calibração. Os autores listaram que alguns desafios relacionados a correções atmosféricas, radiométricas e geométricas, cobertura vegetal, umidade e rugosidade do solo ainda precisam ser abordados e concluíram que a inclusão de dados de SR por satélite em mapeamentos dos solos tem atendido, de forma satisfatória, às demandas por informações atualizadas sobre propriedades ambientais e atributos do solo de grande interesse.

#### 2.2.6 Função de Pedotransferência

A função de pedotransferência (PTF, da sigla em inglês *Pedotransfer Function*), proposta por Bouma (1989), consiste em um modelo matemático de estimativa de propriedades do solo a partir de outras propriedades mais facilmente mensuráveis.

A proposta foi iniciada numa época em que os levantamentos do solo tinham sido concluídos em muitos países e, dessa forma, foram levantadas questões sobre qual seria a próxima atividade da pedologia. Nesse período, era evidente que esses conhecimentos qualitativos e empíricos podiam ser relevantes na produção de avaliações iniciais em grandes áreas e que procedimentos mais quantitativos seriam necessários para enfrentar as questões modernas de utilização das terras (VAN LOOY et al, 2017).

Mais especificamente, as PTFs têm por objetivo transferir dados estruturais e composicionais dos solos em informação que caracterize o seu funcionamento (permitindo a sua parametrização). Esta informação auxilia a quantificação dos serviços ecossistêmicos do solo, tais como o fornecimento de água e nutrientes às plantas, a regulação do clima e a redução da poluição (VERECKEN et al., 2016).

As PTFs são desenvolvidas para estimar propriedades do solo que não possuem dados de fácil acesso e são aplicadas em uma gama de estudos no mundo, como estimativa de condutividade hidráulica em solos saturados (ZHANG; SCHAAP, 2019), da densidade do solo (PALLADINO et al., 2022), da retenção de água (XU et al., 2021) e de propriedades químicas do solo (ENANG et al., 2022; SONG et al., 2020) A quantidade de C ou MO presente no solo é uma das variáveis mais utilizadas como atributos preditores nos modelos desenvolvidos e são raras as PTFs que não utilizam deste dado (VASCONCELOS et al., 2020; ZACHARIAS; WESSOLEK, 2007)



### 2.2.6.1 Metodologias de desenvolvimento de Funções de Pedotransferência

Pachepsky, Rajkai e Tóth (2015) enumeraram os métodos que podem ser empregados para desenvolver uma PTF. Entre esses métodos estão a regressão linear, modelos lineares generalizados, modelos aditivos generalizados, árvores de regressão e redes neurais. Os autores também destacam que a elaboração de uma PTF requer a disponibilidade de dados básicos obtidos por levantamentos de solos, dados de laboratório e bases de dados organizadas contendo informações das observações dos preditores e dos estimados desejados.

Em uma robusta revisão sobre PTFs existentes e as perspectivas e desafios para o desenvolvimento de novos modelos em todas as áreas de Ciência da Terra, Van Looy et al (2017) listaram as variações nas metodologias, citando pontos fortes e fracos de categorias de qualidades de modelos que podem ser úteis para um utilizador consultar antes de escolher a metodologia de modelação. Os autores reforçam que, como os *softwares* estão em constante evolução, deve-se considerar a seleção da metodologia de acordo com os objetivos do estudo e optar sempre pela transparência, interpretabilidade e facilidade de elaboração e aplicação do modelo.

Os modelos lineares (LM) são os mais utilizados para a geração de PTFs, em razão da sua simplicidade de elaborar e aplicar as funções (LAMICHHANE; KUMAR; WILSON, 2019; QUIN et al., 2022).

Os modelos lineares generalizados (GLM) (NELDER; WEDDERBUR, 1972) foram propostos como alternativa para a variável que não pode ser analisada pelo modelo linear e necessita de maior flexibilidade em relação a distribuição da variável resposta, que deve ser exponencial. A ideia básica é estimar os parâmetros de um modelo linear usando-se o método da máxima verossimilhança baseado na distribuição dos dados. Dessa maneira, pode-se selecionar um modelo que representa melhor os dados e com maior precisão. Os GLM permitem explicar a resposta Y em função de variáveis explicativas e são muito flexíveis devido à possibilidade da utilização de diferentes distribuições para a variável resposta e diferentes formas para a função de ligação (GOTWAY; STROUP 1997).

Os GLM têm somente um componente aleatório, mas podem apresentar, para ter efeitos aleatórios no preditor linear, uma extensão que é conhecida como modelo misto linear generalizado (GLMMs) (GBUR et al., 2012), cuja característica é que as variáveis respostas

são consideradas independentes e estão condicionadas aos valores de um conjunto de variáveis espaciais (DIGGLE et al., 1998; LEDO et al., 2020).

É viável empregar a regressão espacial múltipla utilizando modelos de regressão espacial linear. Essa abordagem permite analisar a relação entre uma variável dependente e várias variáveis independentes, levando em consideração a localização dos dados coletados. Caso essa relação seja confirmada, busca-se ajustar um modelo estatístico capaz de descrever a variável em questão em relação às demais, considerando a localização dos dados. (LAMICHHANE: KUMAR: WILSON, 2019).

O conceito mais importante em análise espacial é a dependência espacial, que é medida através da autocorrelação espacial e expressa numericamente o conceito de dependência espacial. A autocorrelação espacial surge sempre que o valor que uma variável resposta assume num ponto do espaço se correlaciona com os seus valores em localidades próximas, mostrando dependência espacial. A dependência espacial permite estimação mais eficiente e é uma vantagem que permite informar sobre locais não amostrados, a partir de dados em localização ao redor do ponto desejado. (BIAN et al., 2020; ROSSET; FERDY, 2014; WADOUX; MINASNY; MCBRATNEY, 2020). Vários estudos estimaram C no solo utilizando técnicas de GLMMs, como Gomes et al. (2019), que estimaram estoques de C nos solos dos biomas brasileiros; Benke et al. (2020), que desenvolveram PTF para estimar C nos solos na Austrália e Zhang et al. (2023) que compararam metodologias para a estimativa de teores de MO nos solos da China.

#### 2.2.6.2 Função de Pedotransferência para estimativa de teores de C no solo

Inicialmente, as PTFs tinham sido desenvolvidas para estimar propriedades hidráulicas no solo (ARRUDA et al., 1987; MACBRATNEY et al., 2002; MINASNY et al., 1999; TOMASELLA; HOODNETT, 1998; TOMASELLA et al., 2003). Mas, modelos já têm sido aplicados para prever outras propriedades físicas e químicas do solo (BENITES et al., 2007; BERNOUX et al., 1998; BOSCHI et al., 2015; KHALEDIAN et al., 2018; XU et al., 2017; ZINN et al., 2005). Para alguns autores (VAN LOOY et al., 2017), o fator de conversão 1,724 (VAN BEMMELEN, 1890) para estimar C a partir da MO, pode ser considerado como uma das primeiras PTFs elaboradas.

Pesquisas recentes relacionadas com a estimativa de teores de C no solo utilizam diversas metodologias para o desenvolvimento de PTFs. Zin, Lal e Resck (2005)

desenvolveram PTF, a partir de regressão linear, para estimar C em solos do Cerrado altamente intemperizados, com profundidade até 1m, utilizando argila+silte como variável preditora e observaram estimativas precisas e relação linear e correlacionada com C.

Com base no teor de argila e de silte, cor do solo e profundidade, Minasny et al. (2006) desenvolveram PTFs para o predizer o conteúdo de C e DS e aplicaram funções exponenciais para estimar o estoque de C no solo numa extensa região da Austrália.

Em estudo sobre modelagem para estimativa de teores de C para posterior cálculo de estoque de C em Gleissolos e Plintossolos, em regiões com Campo Limpo Úmido no Distrito Federal, França et al. (2014), utilizando as variáveis DS, areia, argila e silte, mostraram que os modelos, desenvolvidos através de regressão não-linear múltipla, apresentaram resultados muito satisfatórios para as classes de solo avaliadas.

Com o objetivo de estimar teores de C com boa exatidão e acurácia em Latossolos no Cerrado, Parizzi (2017) desenvolveu PTF com um conjunto de dados com quase 300 amostras e afirmou que a capacidade de troca de cátions (CTC) e a granulometria são os atributos mais correlacionados com o C, além do pH, soma de bases (SB), saturação por base (V%) e teor de alumínio (Al). Apesar dos bons resultados, a autora reconheceu que mais amostras inseridas no conjunto de dados garantiria modelos mais precisos. Para Somarathna, Minasny e Malone (2017), a precisão da estimativa aumenta quando há um maior número de amostras no banco de dados.

Fernández-Ugalde e Tóth (2017) desenvolveram PTFs para predizer o conteúdo de C em horizontes subsuperficiais da Europa. Segundo os autores, que desenvolveram os modelos por regressões lineares a partir de atributos agrupados em categorias (tipos de solo e uso do solo) e dados de climatologia, as PTFs apresentaram boa capacidade de predição e podem ser aplicadas em larga escala.

Em estudo desenvolvido na Tanzânia, utilizando 100 amostras de solos, Mwangi et al (2019), desenvolveram PTF para estimar C em Cambissolos e Neossolos quartzarênico e flúvico em áreas cultivadas com milho localizadas na planície costeira do país, a partir de modelos de regressão e textura do solo como variáveis preditoras. Os autores concluíram que o uso de uma PTF para estimar C com base na textura do solo é fácil e econômica, apresentou resultados promissores e recomendam mais estudos sobre estimar outras propriedades solo utilizando uma PTF.

Com o propósito de examinar como a relação de metodologias espaciais para estimar C com o número de amostras para calibração do modelo refletem na exatidão das previsões de C

no solo, Somarathna, Minasny e Malone (2017) selecionaram modelos espaciais, entre Regressão Linear Múltipla (RLM), Regressão Geograficamente Ponderada (GWR), Modelo Cúbico, Modelos Mistos Lineares (LMM), Regressão Quantílica de Florestas (QRF) e Regressão de *Machine Learning* Extremo (ELMR) para modelar a distribuição do C no solo. Os autores concluíram que a exatidão da previsão espacial do C do solo depende do número de amostras, independente da metodologia utilizada para desenvolver os modelos, porém, as metodologias MLR, o LMM e o GWR forneceram previsões mais precisas quando foram utilizadas 15 amostras por quilômetros quadrado. Os autores ainda afirmaram que a autocorrelação espacial foi um componente essencial na previsão do teor de C do solo para o referido estudo na Austrália.

Ledo et al. (2020) desenvolveram modelos para prever as mudanças nos estoques de C da transição de cultura anual para cultura perene, através de GLMMs com distribuição tanto gama quanto gaussiana, utilizando uma base de dados global sobre a dinâmica do C em culturas perenes. Os autores consideraram variáveis climáticas, dados de areia, silte, argila, pH e informações relacionadas às culturas como as variáveis preditoras e a taxa de mudança do estoque C como variável resposta, cujos modelos estimaram o aumento de 20% do estoque de C em áreas de transição de cultura anual para cultura perene.

Em artigo de revisão sobre as técnicas de mapeamento digital do solo usadas para mapear a concentração e os estoques de C no solo em pesquisas publicadas entre 2013 e 2019, Lamichhane, Kumar e Wilson, (2019) afirmaram que houve uma progressão dos modelos lineares para as técnicas de *machine learning* (ML) e que, apesar da MLR ter sido o método usado com mais frequência para prever a quantidade de C, a metodologia foi superada por outras técnicas de ML na maioria dos estudos. O *Random Forest* (RF) apresentou um desempenho melhor do que o MLR e outras técnicas de ML na maioria dos estudos comparativos. Outras técnicas comuns e competitivas foram *Cubist*, *Neural Network* (NN), *Boosted Regression Tree* (BRT), *Support Vector Machine* (SVM) e *Geographically Weighted Regression* (GWR). Os autores sugeriram calibrar os algoritmos competitivos usando conjuntos de dados experimentais específicos e validar os modelos em dados externos.

### 3 METODOLOGIA

Para alcançar os objetivos, o trabalho foi desenvolvido em cinco etapas: I) atualização e organização da base de dados, II) padronização e análise crítica dos dados, III) estabelecimento de faixas de C, IV) desenvolvimento das PTFs e V) validação das PTFs (Figura 1).

#### 3.1 Atualização e organização da base de dados de solos

Anteriormente, um banco de dados foi organizado em planilha, com informações de solos brasileiros, extraídas de relatórios de levantamentos de solos, artigos científicos, dissertações e teses publicados até 2015 (CIDIN, 2016). Fez-se necessário adicionar dados mais recentes e, como critério para a atualização da base de dados, foram selecionados trabalhos nos principais periódicos nacionais e internacionais, que apresentassem, primordialmente, dados de atributos físicos e químicos, dados de localização detalhados, de classe e uso do solo e metodologia de determinação de C (Tabela1), compondo a base de dados com 5.207 amostras de solos brasileiros.

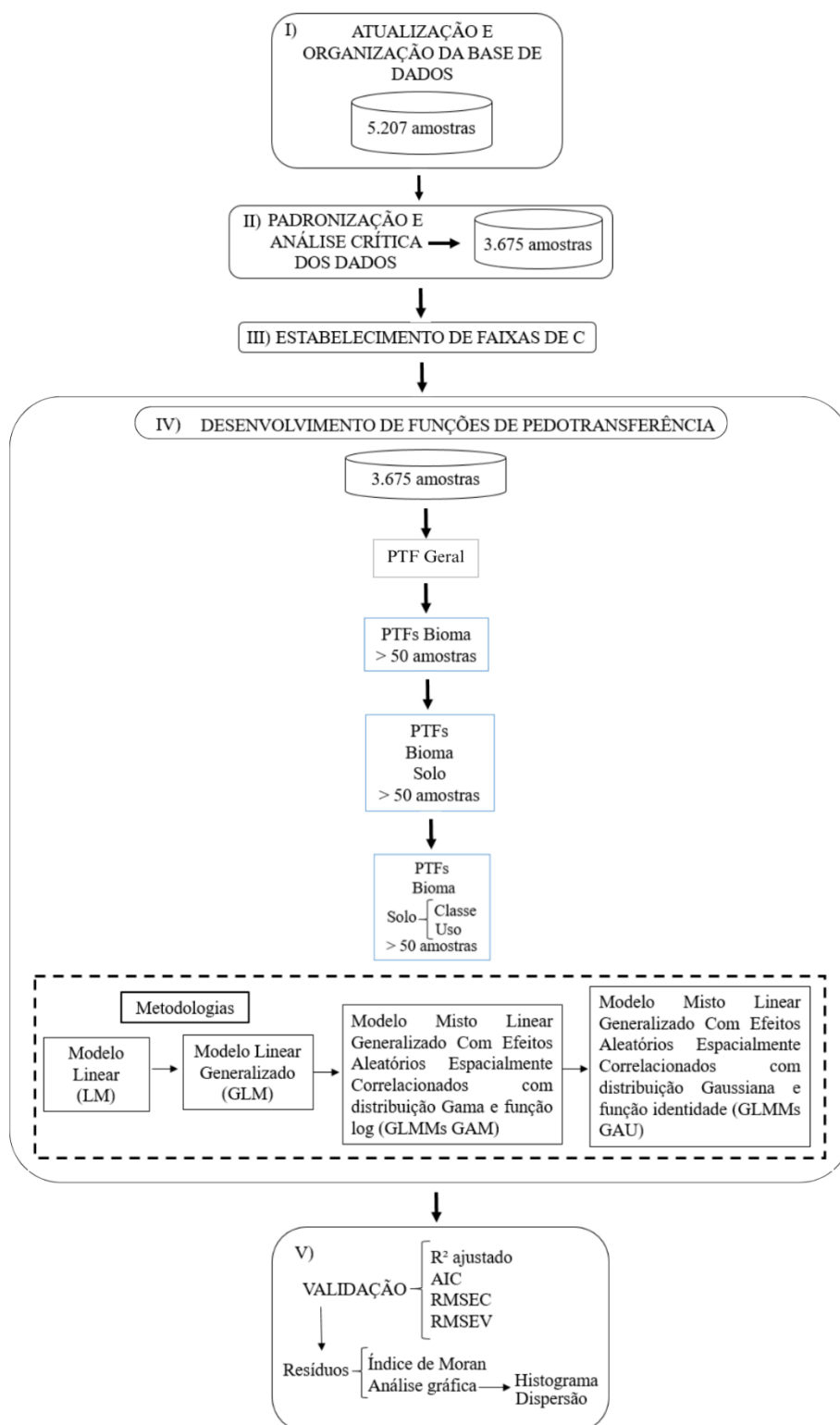
Tabela 1- Atributos levantados na literatura utilizados para o desenvolvimento das PTF para estimar teores de C no solo

| Localização             | Perfil do solo       | Atributos físicos | Atributos químicos  |
|-------------------------|----------------------|-------------------|---------------------|
| Cidade                  | Classificação        | Areia             | Carbono             |
| Estado                  | Uso do solo          | Silte             | Matéria orgânica    |
| Região                  | Profundidade inicial | Argila            | pH H <sub>2</sub> O |
| Coordenadas geográficas | Profundidade final   |                   |                     |
| Bioma                   |                      |                   |                     |

Fonte: Autor (2023).

Em seguida, foi feita uma análise mais detalhada para a verificação da consistência dos valores das variáveis e foram selecionadas somente amostras de interesse para a execução deste trabalho e excluídas amostras com dados não definidos ou inconsistentes.

Figura 1- Etapas da metodologia de desenvolvimento de PTFs para estimar teores de C no solo



AIC = Critério de informação de Akaike; RMSEC = raiz quadrada do erro quadrático médio da calibração; RMSEV = raiz quadrada do erro quadrático médio da validação.

Fonte: Autor (2023).

### 3.2 Padronização e análise crítica dos dados

Foi estabelecido um conjunto de procedimentos para padronização dos dados, tais como:

- Padronização de unidades, expressas de acordo com o Sistema Internacional (SI);
- Conversão de MO em C, utilizando o fator 1,724 de Van Bemmelen;
- Seleção de métodos de determinação de C, que para este estudo, foram utilizados somente métodos de determinação via úmida.
- Classificação do solo: as classes de solos foram definidas conforme SANTOS et al. (2013) e mantido somente o 1º nível de classificação. Em razão do baixo número de amostras de algumas classes de solo presente na base de dados, essas classes de solo foram agrupadas, a partir de características similares, de acordo com o *World Reference Base for Soil Resources* (WRB) (FAO, 2015) (Tabela 2).

Tabela 2 - Agrupamento de classes de solo, segundo o guia simplificado do WRB

| Grupo de solos de referência                          | Principais características   | Sigla |
|---|--|-------|
| Cambissolo<br>Neossolos                               | Solos moderadamente desenvolvidos  | CN    |
| Gleissolo<br>Espodossolo<br>Planossolo<br>Plintossolo | Solos com intensa redistribuição de argila e/ou húmus com ferro e alumínio | GEPP  |

Fonte: Autor (2023).

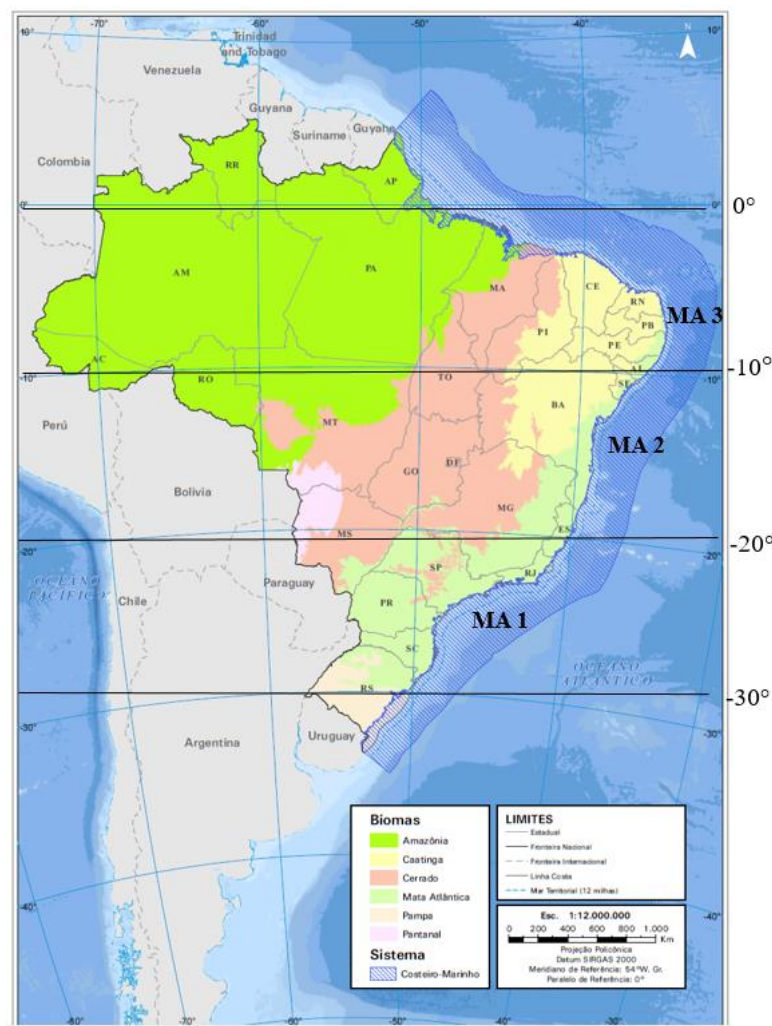
- Uso do solo: foram considerados as principais classes de uso do solo: cultura anual, cultura perene, pastagem, vegetação nativa, sistemas de integrados de produção, floresta plantada e solo sem uso agrícola.
- Coordenadas geográficas: foram utilizadas as coordenadas geográficas de cada município obtidos na base de dados do Instituto Brasileiro de Geografia e Estatística (IBGE) ([https://geoftp.ibge.gov.br/organizacao\\_do\\_territorio/estrutura\\_territorial/localidades](https://geoftp.ibge.gov.br/organizacao_do_territorio/estrutura_territorial/localidades)) em virtude da ausência de localização real do ponto de coleta da maioria das amostras que compõem a base de dados.

Em razão da extensão do bioma Mata Atlântica (MA), as amostras foram agrupadas em três subgrupos (MA1, MA2, MA3) de acordo com as dimensões dos paralelos que indicam as latitudes (Figura 2). Este procedimento se inseriu no esforço de tornar as amostras mais

homogêneas, haja vista não foram utilizadas variáveis agroclimáticas na organização da base de dados e, conseqüentemente, no desenvolvimento das PTFs.

Após a conclusão desta etapa, fez-se uma análise descritiva dos dados para obtenção de dados médios, mínimos, máximos, coeficiente de variação (CV%) e desvio-padrão (DP), verificando a presença de *outliers* e, principalmente, se haviam dados para todas as variáveis necessários para o desenvolvimento das PTFs.

Figura 2 - Divisão do bioma Mata Atlântica em três partes, de acordo com as dimensões dos paralelos que indicam as latitudes



Fonte: Mapa de biomas do IBGE editado pelo próprio autor (2023).

### 3.3 Estabelecimento de faixas de teores de carbono

Esta etapa constituiu em estabelecer faixas de dados de C presentes na base de dados, visando garantir que no momento do sorteio aleatórios de amostras para calibração e validação



das PTFs, cada modelo cobriria a variação dos dados de C. Foram criadas 11 faixas, a partir da distribuição dos dados, conforme descrição na Tabela 3.

Tabela 3 - Intervalo de dados de C distribuídos em faixas

| Faixa | Intervalo de dados de C (g kg <sup>-1</sup> ) | % Frequência |
|-------|---|--------------|
| 1     | ≤ 5   | 12           |
| 2     | >5 e ≤10                                      | 18           |
| 3     | >10 e ≤15                                     | 16           |
| 4     | >15 e ≤20                                     | 15           |
| 5     | > 20 e ≤ 25                                   | 11           |
| 6     | >25 e ≤30                                     | 6            |
| 7     | >30 e ≤35                                     | 6            |
| 8     | >35 e ≤40                                     | 5            |
| 9     | >40 e ≤45                                     | 3            |
| 10    | >45 e ≤ 50                                    | 2            |
| 11    | >50   | 6            |

Fonte: Autor (2023)

### 3.4 Desenvolvimento das funções de pedotransferências

Para o desenvolvimento das PTFs, as variáveis preditoras selecionadas foram areia, silte, argila e pH H<sub>2</sub>O, pois são atributos de baixo custo, fácil determinação e apresentam mais estabilidade ao longo do tempo. Não foram utilizadas variáveis de fertilidade do solo, pois esses atributos, assim como MO ou C, são determinados em análises de rotina de laboratórios de solos.

Para o desenvolvimento das PTFs, a base de dados foi utilizada em 4 etapas:

- Base de dados completa;
- Base com agrupamento dos dados por cada bioma brasileiro;
- Base com agrupamento dos dados por biomas e classes de solo;
- Base com agrupamento dos dados por bioma, classe de solo e uso do solo.

Em todas as etapas, os dados foram, aleatoriamente, divididos para calibração (75%) e validação (25%), com mínimo de 50 amostras em cada agrupamento.

Os modelos foram desenvolvidos através de quatro metodologias: a) modelo linear (*Linear Model* - LM), b) modelo linear generalizado (*Generalized Linear Model* - GLM), c)

modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados (*Generalized Linear Mixed Model With Spatially Correlated Random Effects – GLMMs*), considerando a c1) distribuição Gama com a função de ligação log (GLMMs GAM) e a c2) distribuição Gaussiana e função de ligação identidade (GLMMs GAU).

Os LM são utilizados quando a variável resposta é procedente de uma distribuição oriunda da família das distribuições exponenciais e os GLM são aplicados quando a variável resposta apresenta distribuição exponencial. Os GLMMs são extensões dos GLM que utilizam as coordenadas geográfica no desenvolvimento dos modelos.

a) Modelo Linear

O Modelo Linear (LM) é definido como:

$$Y = Z\beta + \varepsilon \quad (\text{eq. 1})$$

assume-se que  $\varepsilon \sim N_n(0, \sigma^2 I)$

Onde Y corresponde ao vetor de observações  $\{y_1, y_2, \dots, y_n\}^T$ ; Z é a matriz de covariáveis com o primeiro vetor unitário;  $\beta$  é o vetor de parâmetros de efeitos associados às variáveis predictoras; 0 é o vetor nulo de dimensão n; I é a matriz de identidade  $n \times n$ ;  $N_n$  denota a distribuição normal multivariada de dimensão n. Para o vetor de parâmetros  $\beta$ , o estimador de quadrados mínimos é definido como:

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (\text{eq. 2})$$

b) Modelo Linear Generalizado

Os modelos lineares generalizados (GLM), propostos por Nelder e Wedderburn (1972) são uma extensão dos LM que é feita em duas direções: a) a distribuição da variável resposta não precisa ser normal, podendo ser de qualquer distribuição da família exponencial e b) a estrutura de linearidade, pois a função que relaciona o valor esperado e o vetor de covariáveis pode ser de qualquer função diferenciável.

Desse modo, os GLM apresentam a seguinte estrutura:

1) Componente aleatória:  $Y_i$  independentes com distribuição pertencente à família exponencial;

2) Componente estrutural ou sistemática: o valor de  $\mu_i$  se relaciona com o preditor linear na seguinte forma:

$$\mu_i = h(\eta_i) = (z_i^T \beta), \quad \eta_i = g(\mu_i) \quad (\text{eq. 3})$$

Onde  $h$  é uma função monótona e diferenciável;  $g = h^{-1}$  é a função de ligação;  $\beta$  é um vetor de parâmetro de dimensão  $p$  e  $Z_i$  é o vetor de especificação de dimensão  $p$ , função do vetor de covariáveis  $x_i$ .

3) Funções de ligação: são definidas de acordo com a distribuição selecionada para a variável resposta, que podem ser de identidade, log, expoente, logarítmica, entre outras. Os estimadores de máxima verossimilhança são obtidos a partir de equações de máxima verossimilhança, que em geral não têm solução analítica e sua resolução se dá por meio de métodos numéricos.

c) Modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados

O modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados (GLMMs) é uma extensão do GLM e apresenta, além dos efeitos fixos como no LM e GLM, os efeitos aleatórios, que neste estudo, são considerados como os dados de latitude e longitude.

Considerando que observações próximas espacialmente possuem valores parecidos, a autocorrelação espacial rompe com os pressupostos dos modelos clássicos e isso implica na estimação ineficiente dos parâmetros do modelo. Uma maneira de modelar a autocorrelação espacial na variável resposta é considerar que ela resulta de efeitos aleatórios que estão espacialmente correlacionados. Pode ser definida como:

$$Y = X\beta + b + \varepsilon \quad (\text{eq. 4})$$

Com a função de ligação log:

$$\log(Y) = X\beta + b + \varepsilon \quad (\text{eq. 5})$$

Onde  $X$  são as variáveis preditoras em diferentes posições espaciais;  $\beta$  é o vetor de parâmetros de efeitos fixos associados às variáveis preditoras;  $b$  são os parâmetros de efeitos aleatórios em diferentes posições espaciais.

O vetor de valores de  $b$  é, usualmente, representado por  $b = Zv$ , onde  $v$  é o vetor de desvios independentes e  $Z$  é a matriz dos parâmetros estimados.

Para o ajuste deste modelo, conforme Rousset e Ferdy (2014), foi utilizado o pacote SpaMM do R, que usa aproximações clássicas de Laplace para a verossimilhança e o modelo

Matérn para a correlação espacial, cuja vantagem é a flexibilidade dada pela função para modelar o decaimento da correlação, além da suavidade do processo através da modificação de seus parâmetros, dada pela função:

$$M_\nu(x) = x^\nu K_\nu(x) \quad (\text{eq. 6})$$

Onde  $K_\nu$  é a função de Bessel modificada de ordem  $\nu$ , que é o parâmetro de suavidade.

Os parâmetros e seus intervalos de confiança foram estimados por equações de verossimilhança e o procedimento de *bootstrap* para a correção de pequeno viés amostral dos testes. Para a estimação do viés, para cada amostra analisada, 100 novas amostras são simuladas sob a hipótese nula, com parâmetros estimados sob o modelo nulo.

Para a validação dos modelos, foram considerados os critérios de informação Akaike (Akaike, 1974) (AIC) (eq. 7), os indicadores estatísticos que medem a precisão, como coeficiente de determinação ( $R^2$ ) e a acurácia da predição foi dada pela comparação dos valores estimados com os observados a partir do cálculo da raiz quadrada do erro quadrático médio (RMSE, do inglês *Root Mean Square Error*) nos dados de calibração (RMSEC) e validação (RMSEV) (eq. 8).

$$AIC = -2 \ln L(x|\hat{\theta}) + 2(p) \quad (\text{eq. 7})$$

Onde  $L(x|\hat{\theta})$  é a função de máxima verossimilhança e  $p$  é a quantidade de parâmetros do modelo.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (e_i - o_i)^2} \quad (\text{eq. 8})$$

O AIC mede a qualidade e a simplicidade de um modelo estatístico, além de ser uma métrica para comparação e seleção de modelos, em que menores valores de AIC representam uma maior qualidade e simplicidade do modelo. O  $R^2$  indica quanto da variação total do C poderá ser explicada pelas variáveis preditoras presentes em um modelo e o RMSE quantifica a dispersão dos dados medidos em relação aos dados estimados.

A análise dos resíduos é necessária para avaliar o ajuste e qualidade dos modelos. O cálculo do índice de Moran verificou se os resíduos apresentaram autocorrelação espacial, que pode ser positiva ou negativa. A autocorrelação positiva ocorre quando valores semelhantes ocorrem próximos uns dos outros e a negativa ocorre quando valores diferentes ocorrem

próximos uns dos outros. Em seguida, foram elaborados histogramas dos resíduos, que mostram a distribuição dos resíduos para todas as observações e gráficos de dispersão dos resíduos, que mostram a dispersão dos pontos ao longo do eixo da variável independente, indicando se o ajuste subestima ou superestima a variável dependente, pois quanto maior a distância dos pontos observados até a reta, maior o resíduo do modelo.

As análises estatísticas e construção de gráficos foram feitas usando o software estatístico R versão 4.1.2 (R *Core Team*, 2014) e a biblioteca spaMM para ajuste dos modelos mistos lineares generalizados com efeitos aleatórios espacialmente correlacionados (GLMMs) (ROSSET; FERDY, 2014).

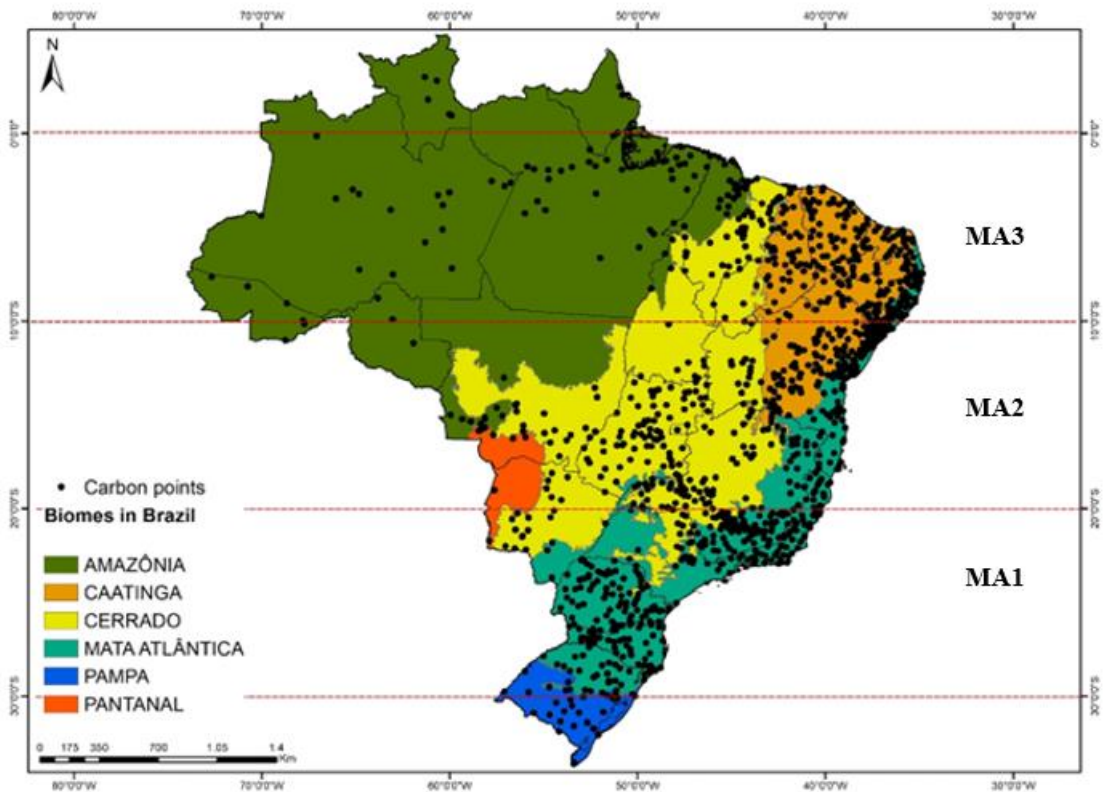
## 4 RESULTADOS

### 4.1 Organização da base de dados

#### 4.1.1 Análise Estatística da Base de Dados

A base de dados, composta por 3.675 amostras de solos brasileiros, com profundidade 0-30 cm contemplou todo o território brasileiro (Figura 3). Na distribuição das amostras por biomas, a Mata Atlântica e Cerrado contribuíram com 36 e 23% das amostras, respectivamente seguidos pelos biomas Amazônia e Caatinga, que contribuíram com 20 e 18% das amostras, respectivamente. Os biomas Pampa e Pantanal contribuíram, juntos, com 3% das amostras presentes na base de dados.

Figura 3 - Distribuição das amostras por biomas brasileiros

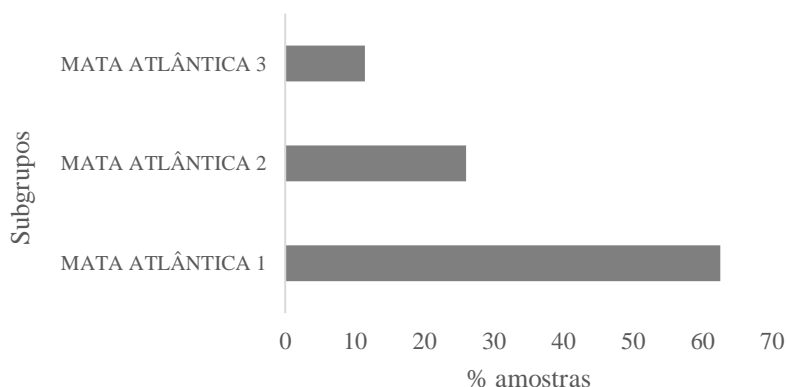


Fonte: Autor (2023).

Em razão da grande extensão do bioma Mata Atlântica, ele foi subdividido de acordo com o intervalo de localização geográfica (Figura 3 e 4), com o intuito de agrupar amostras de

solos com características climáticas similares, haja vista que nenhuma variável climática foi utilizada no desenvolvimento dos modelos. O maior número de amostras corresponde às áreas que compõem o subgrupo MA1, com 826 amostras, seguido de MA2 e MA3 com 341 e 150 amostras, respectivamente.

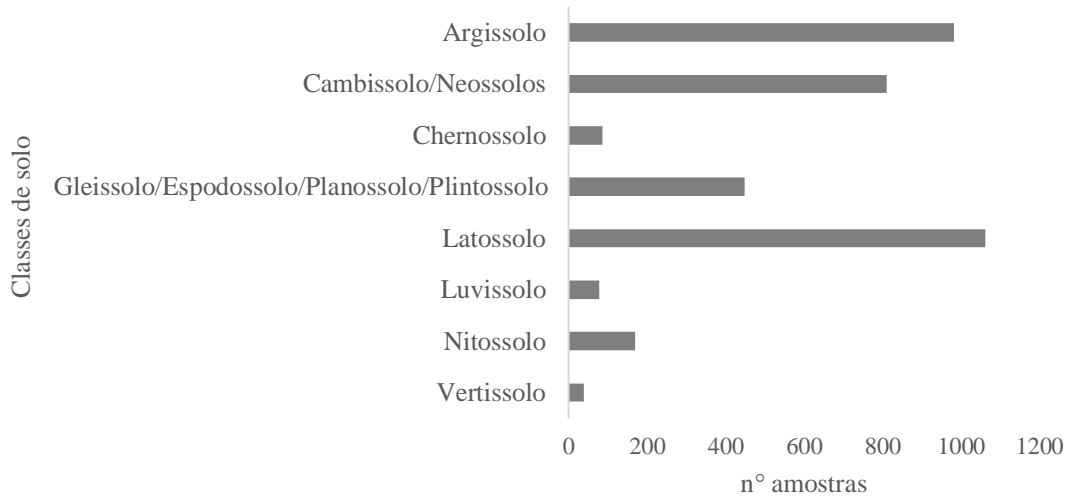
Figura 4 - Distribuição das amostras por subgrupos do bioma Mata Atlântica (MA1, MA2, MA3)



Fonte: Autor (2023)

As principais classes de solos definidas no Sistema Brasileiro de Classificação de Solos (SANTOS et al., 2013) foram representadas na base de dados (Figura 5). Os Latossolos contribuíram com 30% das amostras na base de dados, seguido pelos Argissolos, com 27% das amostras. Conforme Santos et al. (2011) esses dados refletem a distribuição de solos no Brasil, em que Latossolos e Argissolos estão presentes em 58% do território. Vertissolos são as classes de menor expressão, assim como as observações deste estudo, que representa 1% das amostras. O agrupamento das amostras de Cambissolos e Neossolos (CN) e Gleissolos, Espodossolos, Planossolos e Plintossolos (GEPP) contribuíram com 22% e 12% das amostras, respectivamente.

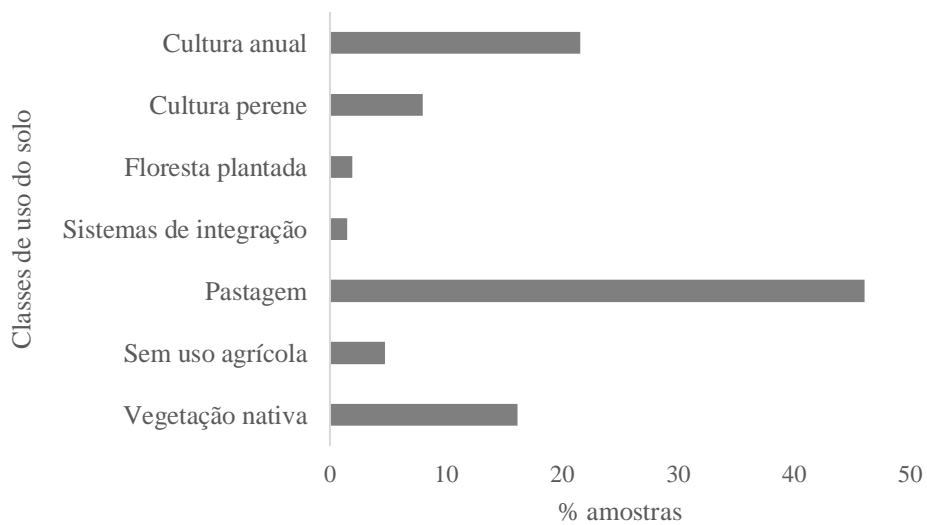
Figura 5 - Distribuição das amostras por classes de solos brasileiros



Fonte: Autor (2023).

As principais classes de uso do solo presentes neste estudo se destacaram com as observações de Pastagem e Cultura anual, com 46% e 22%, respectivamente. Amostras de Vegetação nativa representaram 16% das amostras. Classes de uso do solo com menor representatividade foram Cultura perene (8%), Sem uso agrícola (5%), Floresta plantada (2%) e Sistemas de Integração (1%) (Figura 6).

Figura 6 - Distribuição das amostras por classes de usos do solo



Fonte: Autor (2023).



A análise estatística dos atributos do solo presentes na base de dados estão descritos na Tabela 4. Os dados de granulometria apresentaram elevado DP em razão da presença das diferentes classes de solos, com textura dos solos que variou de arenosa (com teores de areia de até 985 g kg<sup>-1</sup>) a muito siltosa (com teores de silte de até 830 g kg<sup>-1</sup>). Os dados de pH variaram de 2,7 a 8,6, com valor médio de 5,3. Os dados de C, que apresentaram distribuição assimétrica (Figura 7), variaram de 0,4 a 412,9 g kg<sup>-1</sup>, refletindo em elevado CV. Esses dados corroboram com Gomes et al., 2019, que desenvolveram PTFs para estimar estoque de C nos biomas brasileiros. Uma grande base de dados é composta por alta variabilidade dos atributos do solo e nem sempre é possível mensurar a sua confiabilidade (COOPER et al., 2015).

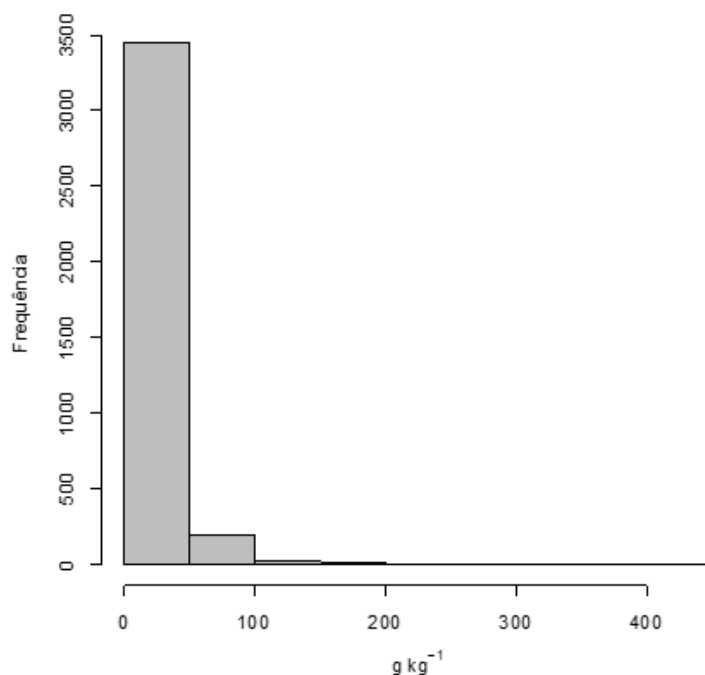
Tabela 4 - Estatística descritiva da base de dados (n = 3675 amostras)

| Variável                     | Média | DP    | CV (%) | Mínimo | Máximo |
|------------------------------|-------|-------|--------|--------|--------|
| C (g kg <sup>-1</sup> )      | 21,5  | 22,1  | 102,4  | 0,4    | 412,9  |
| Areia (g kg <sup>-1</sup> )  | 480,9 | 263   | 54,7   | 2      | 985    |
| Silte (g kg <sup>-1</sup> )  | 200,7 | 136,3 | 67,9   | 0      | 830    |
| Argila (g kg <sup>-1</sup> ) | 318,5 | 206,1 | 64,7   | 10     | 935    |
| pH                           | 5,3   | 0,9   | 16,3   | 2,7    | 8,6    |

n= número de amostras; C = carbono; DP = desvio-padrão; CV = coeficiente de variação.

Fonte: Autor (2023).

Figura 7 - Histograma de frequência de dados de C



## 4.2 Desenvolvimento de Função de Pedotransferência

As PTFs foram geradas a partir de modelo linear (LM), modelo linear generalizado (GLM), modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados, com família Gama e função de ligação log (GLMMS GAM) e com família Gaussiana e função de ligação de identidade (GLMMs GAU). Inicialmente, foi utilizando a base de dados completa para desenvolvimento da PTF geral e, posteriormente, as PTFs foram desenvolvidas a partir de três subconjuntos da base de dados completa: 1) agrupamentos de dados por cada bioma brasileiro; 2) agrupamento de dados por biomas e classes de solo (Bioma x Solo) e 3) agrupamentos de dados por biomas, classes de solo e uso do solo (Bioma x Solo x Uso).

As variáveis preditoras selecionadas foram areia, silte e pH para todas as PTFs. A soma das medidas de areia, argila e silte é igual a 100%, assim, qualquer uma das três variáveis será a combinação linear das outras duas, portanto, não se pode fazer uso das três ao mesmo tempo no modelo, uma vez que causaria multicolinearidade e isso compromete o processo de estimação dos parâmetros, pois geraria múltiplas soluções para o sistema (SOMARATHNA; MINASNY; MALONE, 2017).

As metodologias LM e GLM têm efeitos fixos (areia, silte, pH, classe de solo e uso do solo) e GLMMs têm, além dos efeitos fixos, os efeitos aleatórios (latitude e longitude) que exigem que o cálculo seja feito por um programa que foi desenvolvido para este propósito e está ilustrado no Anexo e será, posteriormente, distribuído aos usuários interessados. Desse modo, as PTFs desenvolvidas a partir de GLMMs apresentadas no texto são apenas ilustrativas para conhecimento geral.

### 4.2.1 Função de Pedotransferência geral (PTF geral)

No desenvolvimento das PTFs para estimar teores de C no solo, a base de dados foi, aleatoriamente, agrupada em conjuntos de dados com 75% das amostras para calibração e 25 % das amostras para validação. As variáveis dos conjuntos de dados utilizados para calibração e validação dos modelos apresentaram valores semelhantes (Tabela 5). O elevado DP para as variáveis areia, silte e argila está relacionado com a alta a variabilidade de classes de solos presentes na base na base de dados. O alto CV do C representa a grande variação de classes e uso do solo.

Tabela 5 - Estatística descritiva base de dados para o desenvolvimento de PTFs para estimar teores de C no solo

| Variável                                     | Média | DP    | CV (%) | Mínimo | Máximo |
|--|-------|-------|--------|--------|--------|
| Conjunto de dados para calibração (n = 2756) |       |       |        |        |        |
| C (g kg <sup>-1</sup> )                      | 21,4  | 21,9  | 102,4  | 0,4    | 412,9  |
| Areia (g kg <sup>-1</sup> )                  | 486   | 259,2 | 53,3   | 5      | 985    |
| Silte (g kg <sup>-1</sup> )                  | 200,3 | 135,9 | 67,8   | 0      | 830    |
| Argila (g kg <sup>-1</sup> )                 | 313,8 | 202,5 | 64,6   | 10     | 905    |
| pH   | 5,3   | 0,9   | 16,3   | 3,3    | 8,6    |
| Conjunto de dados para validação (n = 919)   |       |       |        |        |        |
| C (g kg <sup>-1</sup> )                      | 22,1  | 22,6  | 102,4  | 0,4    | 330,6  |
| Areia (g kg <sup>-1</sup> )                  | 465,7 | 273,8 | 58,8   | 2      | 980    |
| Silte (g kg <sup>-1</sup> )                  | 201,9 | 137,7 | 68,2   | 1      | 790    |
| Argila (g kg <sup>-1</sup> )                 | 332,7 | 215,8 | 64,8   | 10     | 935    |
| pH   | 5,3   | 0,9   | 16,3   | 2,7    | 8,3    |

n = número de amostras; C = carbono; DP = desvio padrão; CV = coeficiente de variação. Fonte: Autor (2023).

No desenvolvimento da PTF com a base de dados composta por 3.675 amostras (Tabela 5), em que 2.756 amostras foram utilizadas para calibrações dos modelos e 919 amostras foram utilizadas para validações das PTFs. Nas metodologias LM e GLM (Tabela 6), os valores dos critérios estatísticos de validação das PTFs foram semelhantes, onde os valores de R<sup>2</sup> ajustado foram baixos, ou seja, as variáveis preditoras explicaram apenas 14% da variância do teor de C, representando baixa correlação entre as variáveis. Os valores de AIC elevados representaram menor qualidade da PTF e o modelo não se ajusta aos dados, refletindo em altos erros de calibração e validação, de 20,2 e 20,9 g kg<sup>-1</sup> no LM e 20,2 e 20,8 g kg<sup>-1</sup> no GLM, respectivamente. As PTFs desenvolvidas a partir das metodologias GLMMs apresentaram maiores valores de R<sup>2</sup> ajustado, de 0,49 e 0,57 no GLMMs GAM e GLMMs GAU, respectivamente, refletindo em menores erros na validação e calibração, com valores de 16 e 18,8 g kg<sup>-1</sup> no GLMMs GAM respectivamente, e 14,7 e 19,2 g kg<sup>-1</sup> no GLMMs GAU para validação e calibração, respectivamente. Esses resultados foram superiores aos observados por Gomes et al. (2019) que estimaram teores de C no solo, para posterior cálculo de estoque de C, utilizando uma base de dados similar a este estudo, contendo 8.227 amostras e classe de solo, índices de vegetação, temperatura e precipitação e dados morfométricos como variáveis preditoras e obtiveram R<sup>2</sup>=0,32 como melhor modelo desenvolvido através de *random forest*.

Para a dependência espacial, calculado a partir de índice de Moran, com valor  $p < 0,001$ , os resíduos das PTFs desenvolvidas a partir de LM e GLM apresentaram dependência espacial, porém, nas metodologias de GLMMs, os resíduos não apresentaram dependência espacial, com maior valor para GLMMs GAU. Os modelos que incorporaram as estruturas espaciais apresentaram melhores resultados pois as coordenadas geográficas foram equacionadas pelos efeitos aleatórios. (Tabela 6).

Tabela 6 - Funções de pedotransferência para estimativa de teor de C no solo desenvolvida a partir de metodologias de regressões lineares e desempenhos avaliados pelos critérios estatísticos (PTF geral)

| Metodologia   | PTFs   |        |           |           |
|---|--|--------|-----------|-----------|
| LM  | $C = 13,04 + 0,043 * \text{areia} - 0,00004 * \text{areia}^2 + 0,128 * \text{silte} - 0,00002 * \text{silte}^2 + 2,60 * \text{pH} + 0,032 * \text{pH}^2 - 0,00023 * (\text{areia} * \text{silte}) - 0,006 * (\text{areia} * \text{pH}) - 0,019 * (\text{silte} * \text{pH}) - 0,00003 * (\text{areia} * \text{silte} * \text{pH})$                 |        |           |           |
| GLM   | $C = 2,60 + 0,0035 * \text{areia} - 0,000003 * \text{areia}^2 + 0,005 * \text{Silte} - 0,000002 * \text{silte}^2 + 0,084 * \text{pH} + 0,002 * \text{pH}^2 - 0,000012 * (\text{areia} * \text{silte}) - 0,00033 * (\text{areia} * \text{pH}) - 0,00065 * (\text{silte} * \text{pH}) + 0,000002 * (\text{areia} * \text{silte} * \text{pH})$        |        |           |           |
| GLMMs GAM   | $C = 2,58 + 0,0028 * \text{areia} - 0,000003 * \text{areia}^2 + 0,004923 * \text{silte} - 0,000004 * \text{silte}^2 + 0,0505 * \text{pH} + 0,0012 * \text{pH}^2 - 0,000007 * (\text{areia} * \text{silte}) - 0,000168 * (\text{areia} * \text{pH}) - 0,00032 * (\text{silte} * \text{pH}) + 0,0000006 * (\text{areia} * \text{silte} * \text{pH})$ |        |           |           |
| GLMMs GAU   | $C = 9,71 + 0,037 * \text{areia} - 0,00004 * \text{areia}^2 + 0,115 * \text{silte} - 0,00004 * \text{silte}^2 + 2,957 * \text{pH} + 0,012 * \text{pH}^2 - 0,0001 * (\text{areia} * \text{silte}) - 0,004 * (\text{areia} * \text{pH}) - 0,014 * (\text{silte} * \text{pH}) + 0,00001 * (\text{areia} * \text{silte} * \text{pH})$                  |        |           |           |
| Critérios estatísticos  | LM   | GLM    | GLMMs GAM | GLMMs GAU |
| R <sup>2</sup> ajustado   | 0,14   | 0,14   | 0,49      | 0,57      |
| AIC   | 24410  | 21274  | 20472     | 24010     |
| RMSEC (n=2756)  | 20,2   | 20,2   | 16        | 14,7      |
| RMSEV (n=919)   | 20,9   | 20,8   | 18,8      | 19,2      |
| Teste de Moran para os resíduos dependência espacial: $p < 0,001$ | <0,001   | <0,001 | 0,028     | 0,544     |

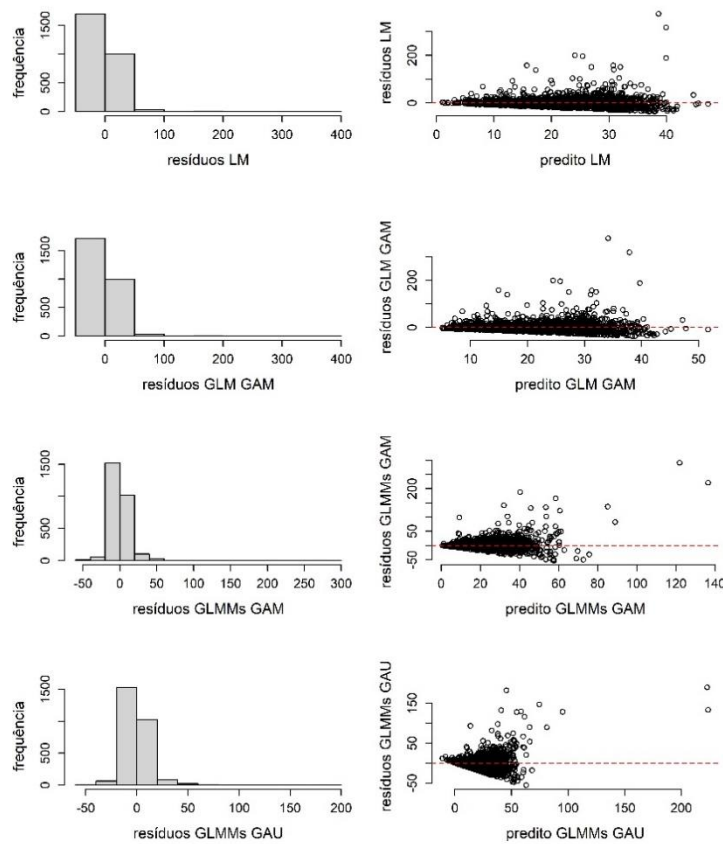
n= número de amostras; C = carbono; AIC = critério de informação Arkaike; RMSEC = raiz quadrada do erro quadrático médio para calibração; RMSEV = raiz quadrada do erro quadrático médio para validação; LM = modelo linear; GLM = modelo linear generalizado; GLMMs GAM = modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados distribuição gama e função de ligação log; GLMMs GAU: modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados distribuição gaussiana e função de ligação identidade.

Fonte: Autor (2023).

Os histogramas dos resíduos e seus respectivos gráficos de dispersão para as quatro metodologias de desenvolvimento das PTFs estão representados na Figura 7. Verificou-se que nas metodologias LM e GLM, a distribuição dos resíduos apresentou comportamentos similares, com dados assimétricos e mesma tendência de dispersão residual. Nas técnicas GLMMs, os resíduos apresentaram similaridades à distribuição normal, com tendência a superestimar os teores de C, em razão da maior frequência dos resíduos próximos de valores negativos.

A metodologia GLMMs GAU gerou PTF com melhor resultado de  $R^2$  ajustado e menor valor de RMSEC, sendo a melhor metodologia para PTFs visando a estimativa de teor de C em solos brasileiros. Porém, o elevado valor de AIC e RMSE para calibração e validação em todas as metodologias refletem na complexidade dos dados se ajustarem aos modelos, sugerindo uma melhor estratificação das amostras.

Figura 8 - Distribuição gráfica dos resíduos dos modelos desenvolvidos para as PTF geral



Fonte: Autor (2023).

Segundo Wang et al. (2021), a avaliação dos resíduos em forma de histogramas é mais adequada, pois quando há muitas observações, somente os gráficos de dispersão podem levar ao risco de má interpretação, em razão da ocorrência de diversos pontos sobrepostos no gráfico.

#### 4.2.2 Função de Pedotransfênciã para cada bioma brasileiro

Foram desenvolvidas PTFs a partir de dados agrupados por biomas brasileiros e apresentados as PTFs para estimativa de teor de C no solo. Para cada agrupamento, o

subconjunto de dados foi, aleatoriamente, dividido em 75% das amostras para calibração e 25% das amostras para validação. Foram considerados agrupamentos com número mínimo de 50 amostras de solo.

#### 4.2.2.1 Amazônia

O subconjunto de dados agrupados do bioma Amazônia, composto por 675 amostras, apresentaram dados de C que variaram de 0,4 a 183 g kg<sup>-1</sup>, com valor médio de 22,4 g kg<sup>-1</sup> (Tabela 7). Dados de granulometria de apresentaram amostras com valores máximos de 980, 790 e 935 g kg<sup>-1</sup>, que correspondem a amostras de areia, silte e argila, respectivamente. Elevados índices de DP e CV configuram a variabilidade de classes de solo presentes no conjunto de dados. O atributo pH apresentou valores que variaram de 3,3 a 7,6 e valor médio de 4,8.

Tabela 7 - Estatística descritiva do subconjunto de dados agrupados do bioma Amazônia (n = 675 amostras)

| Variável                     | Média | DP    | CV%  | Mínimo | Máximo |
|------------------------------|-------|-------|------|--------|--------|
| C (g kg <sup>-1</sup> )      | 22,4  | 21,6  | 96,6 | 0,4    | 183    |
| Areia (g kg <sup>-1</sup> )  | 472,4 | 263,3 | 55,7 | 2      | 980    |
| Silte (g kg <sup>-1</sup> )  | 211,2 | 165,2 | 78,2 | 10     | 790    |
| Argila (g kg <sup>-1</sup> ) | 316,5 | 207,2 | 65,5 | 10     | 935    |
| pH                           | 4,8   | 0,8   | 17,4 | 3,3    | 7,6    |

n = número de amostras; C = carbono; DP = desvio padrão; CV = coeficiente de variação.

Fonte: Autor (2023).

As PTFs desenvolvidas para estimar teores de C nos solos da Amazônia estão descritos na Tabela 8. Nas metodologias LM e GLM, os valores dos critérios estatísticos de validação das PTFs foram semelhantes, onde os valores de R<sup>2</sup> ajustado foram baixos, ou seja, as variáveis preditoras explicaram apenas 16% e 15% da variância do teor de C, respectivamente. Os valores de AIC elevados representaram menor qualidade das PTFs, refletindo em altos erros de calibração e validação, que variaram de 19,5 e 21,4 g kg<sup>-1</sup> no LM e 19,5 e 21,5 g kg<sup>-1</sup> no GLM, respectivamente. As PTFs desenvolvidas a partir das metodologias GLMMs apresentaram maiores valores de R<sup>2</sup> ajustado, de 0,46 e 0,43 no GLMMs GAM e GLMMs GAU, respectivamente, refletindo em menores erros na validação e calibração, com valores de 15,7 e 19,2 g kg<sup>-1</sup> no GLMMs GAM respectivamente, e 16,2 e 19,2 g kg<sup>-1</sup> no GLMMs GAU para validação e calibração, respectivamente.

Para a dependência espacial, calculado a partir de índice de Moran, com valor  $p < 0,001$ , os resíduos dos modelos desenvolvidos a partir GLM apresentaram dependência espacial, porém, os resíduos não apresentaram dependência espacial, com valores de  $p = 0,006$  no LM,  $p = 0,474$  e  $p = 0,845$  nas técnicas GLMMs GAM e GLMMs GAU, respectivamente (Tabela 8).

Tabela 8 - Funções de pedotransferência para estimativa de teor de C no solo na Amazônia desenvolvida a partir de metodologias de regressões lineares e respectivos desempenhos avaliados pelos critérios estatísticos

| Metodologia   | PTFs  |        |           |           |
|---|---|--------|-----------|-----------|
| LM  | $C = 81,51 - 0,036 * \text{areia} - 0,00002 * \text{areia}^2 - 0,032 * \text{silte} - 0,0001 * \text{silte}^2 - 13,94 * \text{pH} + 0,74 * \text{pH}^2 - 0,00011 * (\text{areia} * \text{silte}) + 0,006 * (\text{areia} * \text{pH}) + 0,016 * (\text{silte} * \text{pH}) + 0,00001 * (\text{areia} * \text{silte} * \text{pH})$             |        |           |           |
| GLM   | $C = 5,38 + 0,00067 * \text{areia} - 0,000002 * \text{areia}^2 + 0,00128 * \text{silte} - 0,000003 * \text{silte}^2 - 0,687 * \text{pH} + 0,052 * \text{pH}^2 - 0,00001 * (\text{areia} * \text{silte}) + 0,00003 * (\text{areia} * \text{pH}) + 0,00016 * (\text{silte} * \text{pH}) + 0,000002 * (\text{areia} * \text{silte} * \text{pH})$ |        |           |           |
| GLMMs GAM   | $C = 4,34 + 0,0021 * \text{areia} - 0,000002 * \text{areia}^2 + 0,00347 * \text{silte} - 0,000004 * \text{silte}^2 + 0,524 * \text{pH} + 0,054 * \text{pH}^2 - 0,00001 * (\text{areia} * \text{silte}) - 0,00024 * (\text{areia} * \text{pH}) - 0,00019 * (\text{silte} * \text{pH}) + 0,000002 * (\text{areia} * \text{silte} * \text{pH})$  |        |           |           |
| GLMMs GAU   | $C = 73,46 + 0,0002 * \text{areia} - 0,000042 * \text{areia}^2 - 0,0186 * \text{silte} - 0,000067 * \text{silte}^2 - 14,8 * \text{pH} + 1,04 * \text{pH}^2 - 0,00010 * (\text{areia} * \text{silte}) + 0,0038 * (\text{areia} * \text{pH}) + 0,0142 * (\text{silte} * \text{pH}) + 0,000002 * (\text{areia} * \text{silte} * \text{pH})$      |        |           |           |
| Critérios estatístico   | LM  | GLM    | GLMMs GAM | GLMMs GAU |
| R <sup>2</sup> ajustado   | 0,16  | 0,15   | 0,46      | 0,43      |
| AIC   | 4564  | 4144   | 3850      | 4474      |
| RMSEC (n=517)   | 19,5  | 19,5   | 15,7      | 16,2      |
| RMSEV (n=158)   | 21,4  | 21,5   | 19,4      | 19,2      |
| Teste de Moran para os resíduos dependência espacial: $p < 0,001$ | 0,006   | <0,001 | 0,474     | 0,845     |

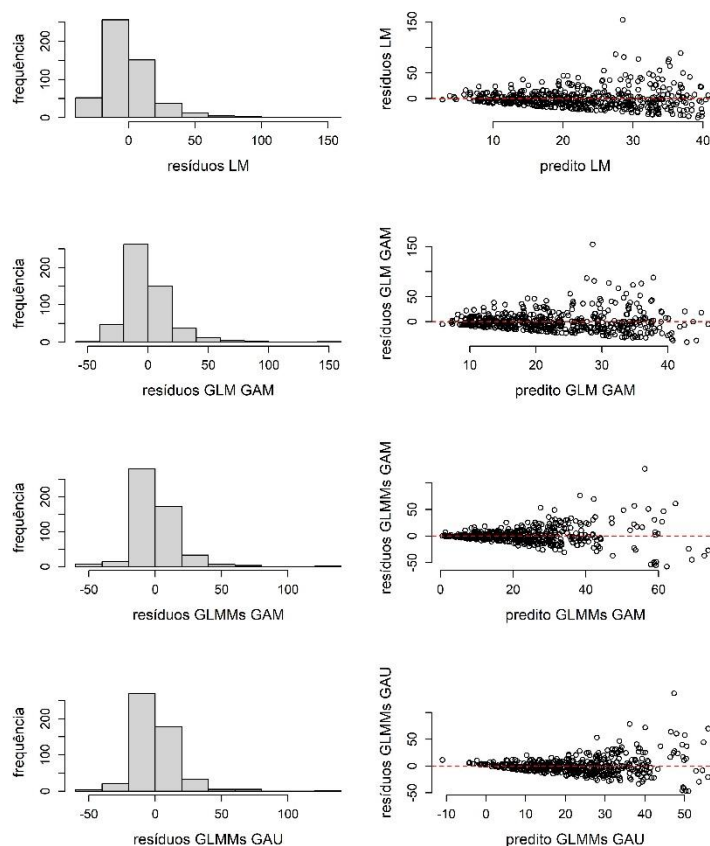
n = número de amostras; C = carbono; AIC = critério de informação Akaike; RMSEC = raiz quadrada do erro quadrático médio para calibração; RMSEV = raiz quadrada do erro quadrático médio para validação; LM = modelo linear; GLM = modelo linear generalizado; GLMMs GAM = modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados distribuição gama e função de ligação log; GLMMs GAU: modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados distribuição gaussiana e função de ligação identidade.

Fonte: Autor (2023).

Os histogramas dos resíduos e seus respectivos gráficos de dispersão para as quatro metodologias de desenvolvimento das PTFs para a Amazônia estão representados na Figura 8. Os resíduos tiveram tendência à distribuição normal em todas as metodologias utilizadas e com a maior frequência dos resíduos negativos, o que representa uma tendência em superestimar os teores de C. Através da dispersão dos resíduos, a metodologia GLMMs GAM foi que apresentou melhor dispersão residual.

A metodologia GLMMs GAM gerou PTF com bons resultados de R<sup>2</sup> ajustado, RMSE e AIC, tornando-se a melhor metodologia para PTFs visando estimar teores de C em solos da Amazônia.

Figura 9 - Análise gráfica dos resíduos das PTFs desenvolvidas para o bioma Amazônia



Fonte: Autor (2023).

#### 4.2.2.2 Caatinga

O subconjunto de dados agrupados do bioma Caatinga, composto por 717 amostras, apresentaram dados de C que variaram de 0,5 a 156,6 g kg<sup>-1</sup>, com valor médio de 14,2 g kg<sup>-1</sup> (Tabela 9). Dados de granulometria de apresentaram amostras com valores máximos de 960, 830 e 800 g kg<sup>-1</sup>, que correspondem a amostras de areia, silte e argila, respectivamente. Elevados índices de DP para granulometria e CV para C configuram a variabilidade de classes de solo presentes no conjunto de dados. O atributo pH apresentou valores que variaram de 3,1 a 8,5 e valor médio de 5,9.

Tabela 9 - Estatística descritiva da base de dados do bioma Caatinga (n = 717 amostras)

| Variável                    | Média | DP    | CV%  | Mínimo | Máximo |
|-----------------------------|-------|-------|------|--------|--------|
| C (g kg <sup>-1</sup> )     | 14,2  | 12,0  | 84,1 | 0,5    | 156,6  |
| Areia (g kg <sup>-1</sup> ) | 608,9 | 207,1 | 34,0 | 20     | 960    |



|                              |       |       |      |     |     |
|------------------------------|-------|-------|------|-----|-----|
| Silte (g kg <sup>-1</sup> )  | 186,9 | 117,0 | 62,6 | 5   | 830 |
| Argila (g kg <sup>-1</sup> ) | 204,1 | 138,9 | 68,0 | 10  | 800 |
| pH                           | 5,9   | 1,0   | 16,8 | 3,1 | 8,5 |

n = número de amostras; C = carbono; DP = desvio padrão; CV = coeficiente de variação.

Fonte: Autor (2023).

As PTFs desenvolvidas para estimar teores de C no solo da Caatinga estão descritos na Tabela 10. Nas metodologias LM e GLM, os valores dos critérios estatísticos de validação do modelo foram semelhantes, onde os valores de R<sup>2</sup> ajustado foram baixos, ou seja, as variáveis preditoras explicaram apenas 27% e 26% da variância do teor de C, respectivamente. Os valores de AIC elevados representaram baixa qualidade das PTF. Os erros de calibração e validação para as técnicas LM e GLM foram iguais, de 8,80 e 14,1 g kg<sup>-1</sup> respectivamente. As PTFs desenvolvidas a partir das metodologias GLMMs apresentaram maiores valores de R<sup>2</sup> ajustado, de 0,42 e 0,61 no GLMMs GAM e GLMMs GAU, respectivamente, refletindo em menores erros na validação e calibração, com valores de 7,9 e 14,1 g kg<sup>-1</sup> no GLMMs GAM e 6,7 e 14,1 g kg<sup>-1</sup> no GLMMs GAU para validação e calibração.

Para a dependência espacial, calculado a partir de índice de Moran, com valor p<0,001, os resíduos dos modelos desenvolvidos a partir GLMMs GAM apresentaram dependência espacial, porém, os resíduos não apresentaram dependência espacial, com valores de p=0,963 no LM, p=0,384 no GLM e p=0,132 no GLMMs GAU (Tabela 10).

Tabela 10 - Funções de pedotransferência para estimativa de teor de C no solo no bioma Caatinga desenvolvida a partir de metodologias de regressões lineares e respectivos desempenhos avaliados pelos critérios estatísticos

| Metodologia             | PTFs   | LM   | GLM  | GLMMs GAM | GLMMs GAU |
|-------------------------|--|------|------|-----------|-----------|
| LM                      | $C=34,93+0,0059*areia-0,000031*areia^2-0,0367*silte-0,00004*silte^2+1,74*pH+0,026*pH^2+0,000001*(areia*silte)+0,0002*(areia*pH)+0,009(silte*pH)-0,000008*(areia*silte*pH)$         |      |      |           |           |
| GLM                     | $C=3,27+0,0028*areia-0,000004*areia^2+0,000004*silte-0,000003*silte^2-0,0915*pH+0,0017*pH^2-0,000004*(areia*silte)-0,00005*(areia*pH)+0,0004(silte*pH)-0,0000001*(areia*silte*pH)$ |      |      |           |           |
| GLMMs GAM               | $C=2,55+0,0035*areia-0,000004*areia^2+0,0024*silte-0,000005*silte^2-0,0396*pH+0,0009*pH^2-0,00001*(areia*silte)-0,00004*(areia*pH)+0,0002(silte*pH)-0,0000001*(areia*silte*pH)$    |      |      |           |           |
| GLMMs GAU               | $C=19,8+0,0227*areia-0,00004*areia^2+0,0066*silte-0,0001*silte^2-0,09*pH+0,0137*pH^2-0,00001*(areia*silte)-0,0005*(areia*pH)+0,0047*(silte*pH)-0,000008*(areia*silte*pH)$          |      |      |           |           |
| Critérios estatísticos  |  |      |      |           |           |
| R <sup>2</sup> ajustado |  | 0,27 | 0,26 | 0,42      | 0,61      |
| AIC                     |  | 3995 | 3652 | 3635      | 3977      |
| RMSEC (n=552)           |  | 8,80 | 8,80 | 7,90      | 6,70      |
| RMSEV (n=165)           |  | 14,1 | 14,1 | 14,1      | 14,1      |

|  |       |       |           |       |
|--|-------|-------|-----------|-------|
| Teste de Moran para os resíduos<br>dependência espacial: $p < 0,001$ | 0,963 | 0,384 | $< 0,001$ | 0,132 |
|--|-------|-------|-----------|-------|

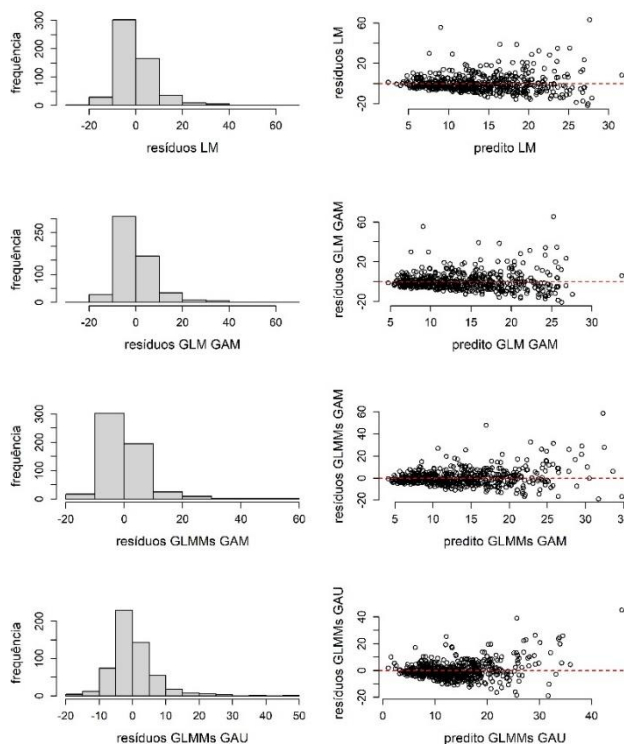
n= número de amostras; C = carbono; AIC = critério de informação Akaike; RMSEC = raiz quadrada do erro quadrático médio para calibração; RMSEV = raiz quadrada do erro quadrático médio para validação; LM = modelo linear; GLM = modelo linear generalizado; GLMMs GAM = modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados distribuição gama e função de ligação log; GLMMs GAU: modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados distribuição gaussiana e função de ligação identidade.

Fonte: Autor (2023).

Os histogramas dos resíduos e seus respectivos gráficos de dispersão para o desenvolvimento das PTFs para a Caatinga estão representados na Figura 9. Em relação aos histogramas, os resíduos tiveram tendência à distribuição normal e com a maior frequência dos resíduos negativos, o que representa uma tendência em superestimar os teores de C, porém, na metodologia GLMMs GAU, a maior frequência ocorreu em classes de valores menores dos resíduos, configurando como a melhor metodologia para a estimativa de C na Caatinga. Através da dispersão dos resíduos, todas as metodologias apresentaram comportamento semelhante, com exceção da metodologia GLMMs GAU que apresentou melhor dispersão residual.

A metodologia GLMMs GAU gerou PTF com os melhores resultados, elevado  $R^2$  ajustado e menores valores de RMSE, configurando como a melhor metodologia para PTFs visando estimar teores de C em solos da Caatinga.

Figura 10 - Análise gráfica dos resíduos das PTFs desenvolvidas para o bioma Caatinga



Fonte: Autor (2023).

#### 4.2.2.3 Cerrado

O subconjunto de dados agrupados do bioma Cerrado, composto por 855 amostras, apresentaram dados de C que variaram de 0,8 a 195,9 g kg<sup>-1</sup>, com valor médio de 20,6 g kg<sup>-1</sup> (Tabela 11). Dados de granulometria de apresentaram amostras com valores máximos de 980, 745 e 860 g kg<sup>-1</sup>, que correspondem a amostras de areia, silte e argila, respectivamente. Elevados índices de DP para a granulometria e CV para o C configuram a variabilidade de classes de solo presentes no conjunto de dados. O atributo pH apresentou valores que variaram de 3,8 a 8,6 e valor médio de 5,4.

Tabela 11 - Estatística descritiva da base de dados do bioma Cerrado (n = 855 amostras)

| Variável                     | Média | DP    | CV%  | Mínimo | Máximo |
|------------------------------|-------|-------|------|--------|--------|
| C (g kg <sup>-1</sup> )      | 20,6  | 18,8  | 91,0 | 0,8    | 195,9  |
| Areia (g kg <sup>-1</sup> )  | 464,0 | 276,5 | 59,6 | 20     | 980    |
| Silte (g kg <sup>-1</sup> )  | 195,8 | 138,1 | 70,5 | 1      | 745    |
| Argila (g kg <sup>-1</sup> ) | 340,5 | 220,3 | 64,7 | 10     | 860    |
| pH                           | 5,4   | 0,7   | 12,9 | 3,8    | 8,6    |

n = número de amostras; C = carbono; DP = desvio padrão; CV = coeficiente de variação. Fonte: Autor (2023).

As PTFs desenvolvidas para estimar teores de C no bioma Cerrado estão descritos na Tabela 12. Nas metodologias LM e GLM, os valores dos critérios estatísticos de validação do modelo foram semelhantes, onde os valores de  $R^2$  ajustado foram baixos, ou seja, as variáveis preditoras explicaram apenas 17% e 19% da variância do teor de C, respectivamente. Os valores de AIC elevados representaram baixa qualidade das PTF. Os erros de calibração e validação para as técnicas LM foram de 16,3 e 18,9 g kg<sup>-1</sup> respectivamente. A metodologia GLM apresentou os maiores valores de RMSE, que foi de 25,1 g kg<sup>-1</sup> de calibração e 27,1 g kg<sup>-1</sup> de validação. As PTFs desenvolvidas a partir das metodologias GLMMs apresentaram valores iguais de  $R^2$  ajustado, que foi de 0,47, e de erros na validação e calibração, com valores de 13,2 e 17,2 g kg<sup>-1</sup>, respectivamente.

Para a dependência espacial, calculado a partir de índice de Moran, com valor  $p=0,255$ , os resíduos dos modelos desenvolvidos a partir GLM apresentaram dependência espacial, porém, os resíduos não apresentaram dependência espacial, com valores de  $p=0,302$  no LM e  $p=0,012$  no GLMMs GAM e  $p=0,663$  no GLMMs GAU, respectivamente (Tabela 12).

Tabela 12 - Funções de pedotransferência para estimativa de teor de C no solo no bioma Cerrado desenvolvidas a partir de metodologias de regressões lineares e respectivos desempenhos avaliados pelos critérios estatísticos

| Metodologia   | PTFs   |        |           |           |
|---|--|--------|-----------|-----------|
| LM  | $C = -2,9 + 0,0482 * \text{areia} - 0,00006 * \text{areia}^2 + 0,1074 * \text{silte} - 0,0001 * \text{silte}^2 + 2,17 * \text{pH} + 0,0862 * \text{pH}^2 - 0,00004 * (\text{areia} * \text{silte}) - 0,0008 * (\text{areia} * \text{pH}) - 0,0030 * (\text{silte} * \text{pH}) - 0,000007 * (\text{areia} * \text{silte} * \text{pH})$       |        |           |           |
| GLM   | $C = 2,9 + 0,00109 * \text{areia} - 0,000004 * \text{areia}^2 + 0,0034 * \text{silte} - 0,00001 * \text{silte}^2 - 0,1924 * \text{pH} + 0,0192 * \text{pH}^2 + 0,00001 * (\text{areia} * \text{silte}) + 0,00034 * (\text{areia} * \text{pH}) + 0,00013 * (\text{silte} * \text{pH}) - 0,000002 * (\text{areia} * \text{silte} * \text{pH})$ |        |           |           |
| GLMMs<br>GAM  | $C = 3,1 + 0,00023 * \text{areia} - 0,000003 * \text{areia}^2 + 0,0039 * \text{silte} - 0,00001 * \text{silte}^2 - 0,2052 * \text{pH} + 0,0193 * \text{pH}^2 + 0,00001 * (\text{areia} * \text{silte}) + 0,0004 * (\text{areia} * \text{pH}) + 0,0002 * (\text{silte} * \text{pH}) - 0,000003 * (\text{areia} * \text{silte} * \text{pH})$   |        |           |           |
| GLMMs<br>GAU  | $C = 3,90 + 0,0255 * \text{areia} - 0,00004 * \text{areia}^2 + 0,1192 * \text{silte} - 0,0001 * \text{silte}^2 + 2,35 * \text{pH} + 0,0528 * \text{pH}^2 - 0,00003 * (\text{areia} * \text{silte}) - 0,0008 * (\text{areia} * \text{pH}) - 0,0055 * (\text{silte} * \text{pH}) - 0,00001 * (\text{areia} * \text{silte} * \text{pH})$        |        |           |           |
| Critérios estatísticos  | LM   | GLM    | GLMMs GAM | GLMMs GAU |
| $R^2$ ajustado  | 0,17   | 0,19   | 0,47      | 0,47      |
| AIC   | 5245   | 4726   | 4578      | 5161      |
| RMSEC (n=620)   | 16,3   | 25,1   | 13,2      | 13,2      |
| RMSEV (n=235)   | 18,9   | 27,1   | 17,2      | 17,4      |
| Teste de Moran para os resíduos dependência espacial: $p=0,255$ | 0,302  | <0,001 | 0,012     | 0,663     |

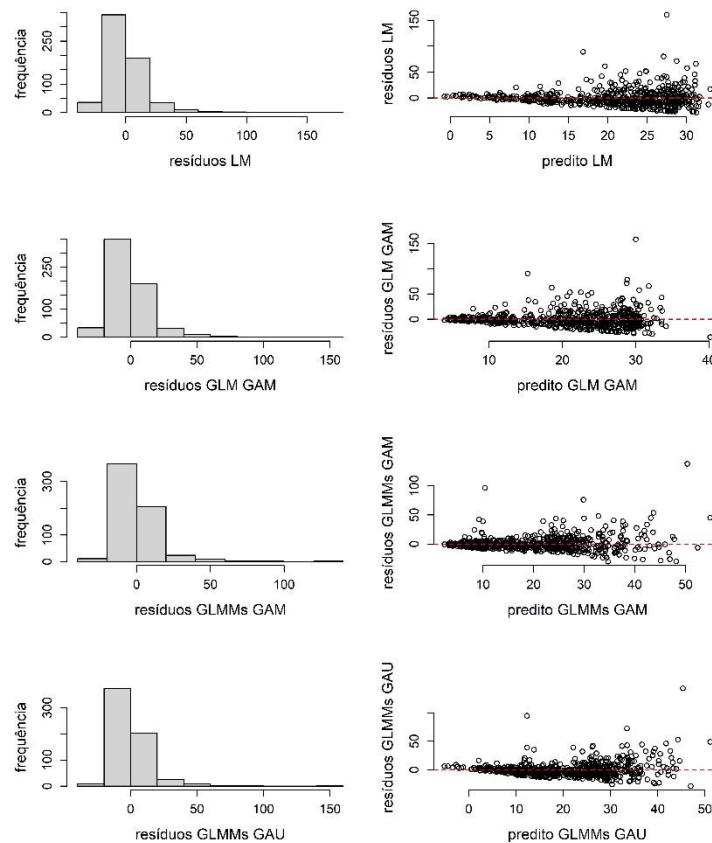
n = número de amostras; C = carbono; AIC = critério de informação Akaike; RMSEC = raiz quadrada do erro quadrático médio para calibração; RMSEV = raiz quadrada do erro quadrático médio para validação; LM = modelo linear; GLM = modelo linear generalizado; GLMMs GAM = modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados distribuição gama e função de ligação log; GLMMs GAU: modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados distribuição gaussiana e função de ligação identidade.

Fonte: Autor (2023).

Os histogramas dos resíduos e seus respectivos gráficos de dispersão para as metodologias de desenvolvimento das PTFs para o Cerrado estão representados na Figura 10. Em relação aos histogramas, nas metodologias LM e GLM, os resíduos apresentaram similaridade à distribuição normal e com a maior frequência dos resíduos negativos, o que representa uma tendência em superestimar os teores de C. Através da dispersão dos resíduos, as metodologias GLMMs apresentaram melhor dispersão residual.

A metodologia GLMMs GAM gerou PTFs com melhores resultados, com elevado  $R^2$  ajustado e menores valores de AIC e RMSE, configurando a melhor metodologia para desenvolver PTF visando estimar teores de C no Cerrado.

Figura 11 - Análise gráfica dos resíduos das PTFs desenvolvidas para o bioma Cerrado



Fonte: Autor (2023).

#### 4.2.2.4 Mata Atlântica

Em razão da extensão do bioma Mata Atlântica, três subconjuntos de dados (MA1, MA2, MA3) foram agrupados considerando as dimensões dos paralelos que indicam as latitudes.

A base de dados do subgrupo da MA1, composta por 821 amostras de solo, apresentou valores de areia, silte e argila que variaram de 5 a 980 g kg<sup>-1</sup>, 0 a 680 g kg<sup>-1</sup> e 10 a 850 g kg<sup>-1</sup> respectivamente (Tabela 13). Teores de C variaram de 0,5 a 412,9 g kg<sup>-1</sup>, com valor médio de 26,1 g kg<sup>-1</sup>. O elevado DP dos dados de granulometria e elevado CV do C correspondem a alta variabilidade de classes e uso do solo. A base de dados do subgrupo da MA2, composta por 341 amostras de solo, apresentou valores de areia, silte e argila que variaram de 15 a 970 g kg<sup>-1</sup>, 10 a 600 g kg<sup>-1</sup> e 10 a 811 g kg<sup>-1</sup> respectivamente (Tabela 13). Teores de C variaram de 1,1 a 330,6 g kg<sup>-1</sup>, com valor médio de 26,6 g kg<sup>-1</sup>. O elevado DP dos dados de granulometria e elevado CV do C correspondem a alta variabilidade de classes e uso do solo. A base de dados do subgrupo da MA3, composta por 150 amostras de solo, apresentou valores de areia, silte e argila que variaram de 20 a 985 g kg<sup>-1</sup>, 0 a 550 g kg<sup>-1</sup> e 10 a 810 g kg<sup>-1</sup> respectivamente (Tabela 13). Teores de C variaram de 0,4 a 81,4 g kg<sup>-1</sup>, com valor médio de 18,4 g kg<sup>-1</sup>. O elevado DP dos dados de granulometria e elevado CV do C correspondem a alta variabilidade de classes e uso do solo.

Tabela 13 - Estatística descritiva da base de dados do bioma Mata Atlântica subgrupo MA1, MA2 e MA3

| MA 1 (n=821 amostras)        |       |       |       |        |        |
|------------------------------|-------|-------|-------|--------|--------|
| Variável                     | Média | DP    | CV%   | Mínimo | Máximo |
| C (g kg <sup>-1</sup> )      | 26,1  | 26,0  | 99,9  | 0,5    | 412,9  |
| Areia (g kg <sup>-1</sup> )  | 352,4 | 253,6 | 72,0  | 5      | 980    |
| Silte (g kg <sup>-1</sup> )  | 224,0 | 131,3 | 58,6  | 0      | 680    |
| Argila (g kg <sup>-1</sup> ) | 423,8 | 203,8 | 48,1  | 10     | 850    |
| pH                           | 5,1   | 0,7   | 13,5  | 3,7    | 8,0    |
| MA 2 (n=341 amostras)        |       |       |       |        |        |
| C (g kg <sup>-1</sup> )      | 26,6  | 29,6  | 111,0 | 1,1    | 330,6  |
| Areia (g kg <sup>-1</sup> )  | 517,1 | 211,8 | 41,0  | 15     | 970    |
| Silte (g kg <sup>-1</sup> )  | 159,0 | 100,1 | 62,9  | 10     | 600    |
| Argila (g kg <sup>-1</sup> ) | 323,9 | 175,4 | 54,2  | 10     | 811    |
| pH                           | 5,2   | 0,8   | 14,7  | 2,7    | 7,8    |
| MA 3 (n=150 amostras)        |       |       |       |        |        |
| C (g kg <sup>-1</sup> )      | 18,4  | 13,0  | 70,5  | 0,4    | 81,4   |

|                              |       |       |      |      |       |
|------------------------------|-------|-------|------|------|-------|
| Areia (g kg <sup>-1</sup> )  | 592,2 | 227,8 | 38,5 | 20,0 | 985,0 |
| Silte (g kg <sup>-1</sup> )  | 171,0 | 119,0 | 69,6 | 0,0  | 550,0 |
| Argila (g kg <sup>-1</sup> ) | 236,9 | 159,8 | 67,5 | 10,0 | 810,0 |
| pH                           | 5,2   | 0,6   | 12,3 | 4,0  | 7,7   |

n = número de amostras; C = carbono; DP = desvio padrão; CV = coeficiente de variação.

Fonte: Autor (2023)

As PTFs desenvolvidas para estimar teores de C no subgrupo MA1, MA 2 e MA3 estão descritas na Tabela 14. Para o subgrupo MA1, os valores dos critérios estatísticos de validação das PTFs nas metodologias LM e GLM foram semelhantes, onde os valores de R<sup>2</sup> ajustado foram baixos, ou seja, as variáveis preditoras explicaram apenas 20% e 14% da variância do teor de C, respectivamente. Os valores de AIC elevados representaram baixa qualidade das PTFs. Os erros de calibração e validação para as técnicas LM foram de 16,3 e 18,9 g kg<sup>-1</sup> respectivamente. A metodologia GLM apresentou os maiores valores de RMSE, que foi de 25,1 g kg<sup>-1</sup> de calibração e 27,1 g kg<sup>-1</sup> de validação. As PTFs desenvolvidas a partir das metodologias GLMMs GAM e GLMMs GAU apresentaram elevados valores de R<sup>2</sup> ajustado, que foi de 0,65 e 0,74 respectivamente, e erros na validação e calibração, com valores de 13,2 e 17,2 g kg<sup>-1</sup>, respectivamente. Para a dependência espacial, calculado a partir do índice de Moran, com valor p=0,255, os resíduos dos modelos desenvolvidos para o subgrupo MA1 a partir de LM apresentou correlação espacial, porém, os resíduos não apresentaram dependência espacial, com valores de p=0,003 no GLM, p=0,488 no GLMMs GAM e p=0,779 no GLMMs GAU (Tabela 14).

As PTFs desenvolvidas para estimar teores de C no subgrupo MA2 estão descritas na Tabela 14. Nas metodologias LM e GLM, os valores dos critérios estatísticos de validação do modelo foram iguais, onde os valores de R<sup>2</sup> ajustado foram baixos e as variáveis preditoras explicaram apenas 0,6% da variância do teor de C, respectivamente. Os valores de AIC elevados representaram baixa qualidade das PTF. Os erros de calibração e validação para as técnicas LM e GLM foram de 23,3 e 35 g kg<sup>-1</sup> e 23,4 e 31,6 g kg<sup>-1</sup> respectivamente. As PTFs desenvolvidas a partir das metodologias GLMMs GAM e GLMMs GAU apresentaram valores de R<sup>2</sup> ajustado de 0,28 e 0,26 respectivamente e valores de RMSE para validação e calibração de 20,7 e 24,4 g kg<sup>-1</sup> e 20,9 e 31,8 g kg<sup>-1</sup> respectivamente. Para a dependência espacial, calculado a partir do índice de Moran, com valor p<0,001, os resíduos dos modelos desenvolvidos para o subgrupo MA2 a partir LM e GLM apresentaram dependência espacial,

porém, os resíduos não apresentaram dependência espacial, com valores de  $p=0,135$  no GLMMs GAM e  $p=0,057$  no GLMMs GAU, respectivamente (Tabela 14).

Tabela 14 - Funções de pedotransferência para estimativa de teor de C no solo no bioma Mata Atlântica, subgrupos MA1, MA2, MA3 desenvolvidas a partir de metodologias de regressões lineares e respectivos desempenhos avaliados pelos critérios estatísticos

| MA 1  |  |        |           |           |
|---|--|--------|-----------|-----------|
| Metodologia   | PTFs   |        |           |           |
| LM  | $C=14,61+0,1140*areia+0,000001*areia^2+0,336*silte+0,0004*silte^2-6,4*pH+3,16*pH^2-0,0002*(areia*silte)-0,0337*(areia*pH)-0,1090*(silte*pH)+0,0001*(areia*silte*pH)$         |        |           |           |
| GLM   | $C=5,19+0,0037*areia-0,000002*areia^2+0,0065*silte+0,000003*silte^2-0,88*pH+0,1090*pH^2-0,00001*(areia*silte)-0,0006*(areia*pH)-0,0015*(silte*pH)+0,000002*(areia*silte*pH)$ |        |           |           |
| GLMMs GAM   | $C=4,83+0,0027*areia-0,000003*areia^2+0,0046*silte+0,0000002*silte^2-0,72*pH+0,08*pH^2-0,00001*(areia*silte)-0,0003*(areia*pH)-0,0007*(silte*pH)+0,0000004*(areia*silte*pH)$ |        |           |           |
| GLMMs GAU   | $C=31,73+0,098*areia-0,00003*areia^2+0,3055*silte+0,0001*silte^2-14,84*pH+2,93*pH^2-0,00004*(areia*silte)-0,0189*(areia*pH)-0,0667*(silte*pH)+0,00001*(areia*silte*pH)$      |        |           |           |
| Critérios estatísticos  | LM   | GLM    | GLMMs GAM | GLMMs GAU |
| R <sup>2</sup> ajustado   | 0,20   | 0,14   | 0,65      | 0,74      |
| AIC   | 5716   | 4915   | 4826      | 5559      |
| RMSEC (n=609)   | 22,6   | 25,7   | 19,7      | 15,2      |
| RMSEV (n=212)   | 25,9   | 37,6   | 49,6      | 20,6      |
| Teste de Moran para os resíduos dependência espacial: $p<0,001$ | <0,001   | 0,003  | 0,488     | 0,779     |
| MA 2  |  |        |           |           |
| Metodologia   | PTFs   |        |           |           |
| LM  | $C=-36,01+0,0753*areia+0,00001*areia^2+0,1646*silte-0,0001*silte^2+11,67*pH+0,47*pH^2+0,0004*(areia*silte)-0,0227*(areia*pH)-0,0243*(silte*pH)-0,0001*(areia*silte*pH)$      |        |           |           |
| GLM   | $C=0,27+0,005*areia-0,000001*areia^2+0,008*silte+0,00001*silte^2+0,5319*pH+0,008*pH^2+0,00001*(areia*silte)-0,001*(areia*pH)-0,001*(silte*pH)-0,000002*(areia*silte*pH)$     |        |           |           |
| GLMMs GAM   | $C=0,75+0,003*areia-0,0000009*areia^2+0,008*silte-0,00001*silte^2+0,45*pH+0,005*pH^2+0,00001*(areia*silte)-0,001*(areia*pH)-0,001*(silte*pH)-0,000003*(areia*silte*pH)$      |        |           |           |
| GLMMs GAU   | $C=-8,45-0,02*areia+0,00001*areia^2+0,081*silte-0,0001*silte^2+7,86*pH+0,24*pH^2+0,001*(areia*silte)-0,006*(areia*pH)-0,007*(silte*pH)-0,0001*(areia*silte*pH)$              |        |           |           |
| Critérios estatísticos  | LM   | GLM    | GLMMs GAM | GLMMs GAU |
| R <sup>2</sup> ajustado   | 0,06   | 0,06   | 0,28      | 0,26      |
| AIC   | 2309   | 2050   | 2025      | 2296      |
| RMSEC (n=250)   | 23,3   | 23,4   | 20,7      | 20,9      |
| RMSEV (n=91)  | 35   | 31,6   | 24,4      | 31,8      |
| Teste de Moran para os resíduos dependência espacial: $p<0,001$ | <0,001   | <0,001 | 0,135     | 0,057     |
| MA 3  |  |        |           |           |
| Metodologia   | PTFs   |        |           |           |
| LM  | $C=125,3-0,08*areia-0,00001*areia^2-0,51*silte+0,0002*silte^2-17,93*pH+0,59*pH^2+0,0005*(areia*silte)+0,01*(areia*pH)+0,08*(silte*pH)-0,0001*(areia*silte*pH)$               |        |           |           |



|   |  |       |           |           |
|---|--|-------|-----------|-----------|
| GLM   | $C = -0,24 + 0,01 * \text{areia} - 0,000004 * \text{areia}^2 + 0,003 * \text{silte} + 0,00001 * \text{silte}^2 + 0,77 * \text{pH} - 0,02 * \text{pH}^2 - 0,00003 * (\text{areia} * \text{silte}) - 0,0013 * (\text{areia} * \text{pH}) - 0,0008 * (\text{silte} * \text{pH}) + 0,000003 * (\text{areia} * \text{silte} * \text{pH})$ |       |           |           |
| GLMMs<br>GAM  | $C = 3,01 + 0,008 * \text{areia} - 0,00001 * \text{areia}^2 - 0,002 * \text{silte} + 0,00001 * \text{silte}^2 - 0,08 * \text{pH} + 0,03 * \text{pH}^2 - 0,00002 * (\text{areia} * \text{silte}) - 0,0008 * (\text{areia} * \text{pH}) - 0,0001 * (\text{silte} * \text{pH}) + 0,000002 * (\text{areia} * \text{silte} * \text{pH})$  |       |           |           |
| GLMMs<br>GAU  | $C = 142,5 - 0,09 * \text{areia} - 0,00002 * \text{areia}^2 - 0,51 * \text{silte} + 0,0001 * \text{silte}^2 - 23,44 * \text{pH} + 0,97 * \text{pH}^2 + 0,0005 * (\text{areia} * \text{silte}) + 0,013 * (\text{areia} * \text{pH}) + 0,08 * (\text{silte} * \text{pH}) - 0,0001 * (\text{areia} * \text{silte} * \text{pH})$         |       |           |           |
| Critérios estatísticos  | LM   | GLM   | GLMMs GAM | GLMMs GAU |
| R <sup>2</sup> ajustado                                       | 0,38   | 0,31  | 0,6       | 0,54      |
| AIC   | 895  | 852   | 839       | 896       |
| RMSEC (n=117)   | 10   | 18,4  | 8,1       | 8,6       |
| RMSEV (n=33)  | 12,1   | 20,5  | 10,7      | 11        |
| Teste de Moran para os resíduos dependência espacial: p=0,091 | 0,933  | 0,575 | 0,966     | 0,524     |

n = número de amostras; C = carbono; AIC = critério de informação Akaike; RMSEC = raiz quadrada do erro quadrático médio para calibração; RMSEV = raiz quadrada do erro quadrático médio para validação; LM = modelo linear; GLM = modelo linear generalizado; GLMMs GAM = modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados distribuição gama e função de ligação log; GLMMs GAU: modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados distribuição gaussiana e função de ligação identidade.

Fonte: Autor (2023)

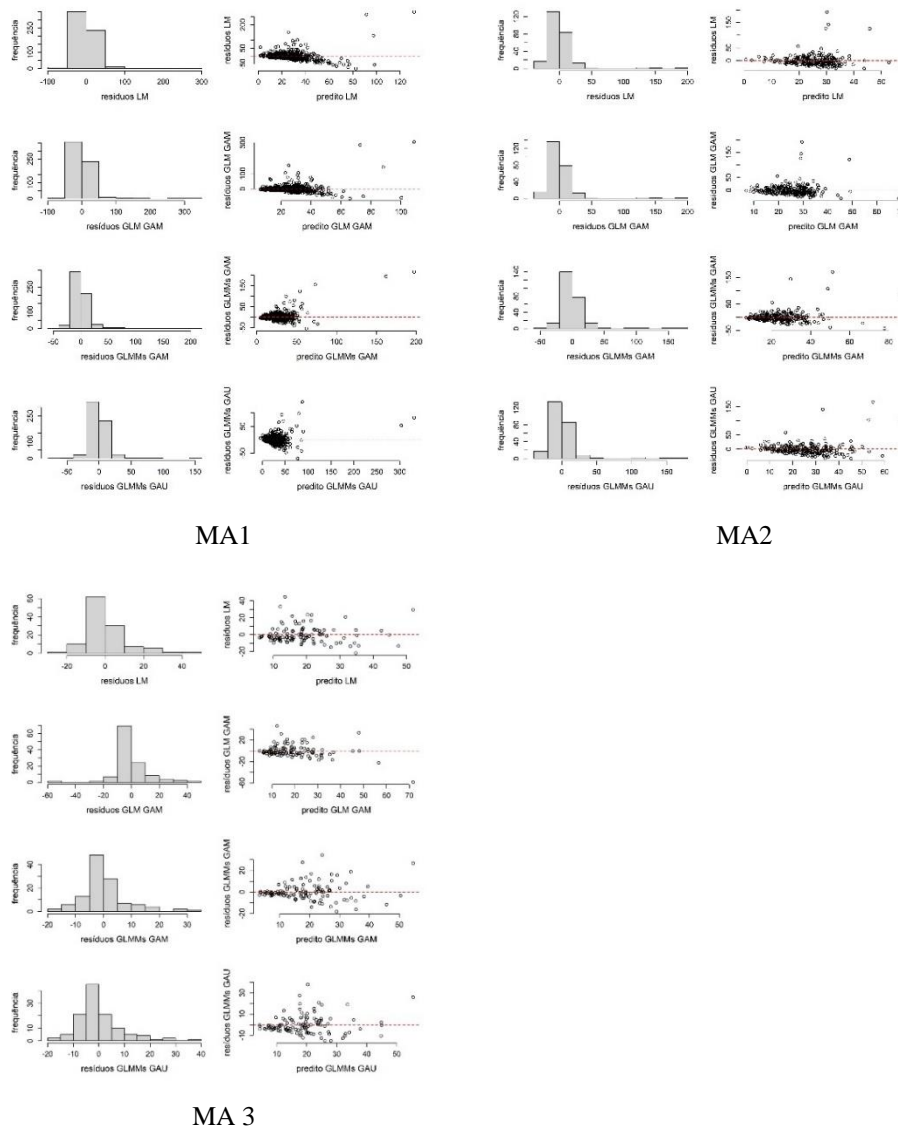
As PTFs desenvolvidas para estimar teores de C no subgrupo MA3 estão descritos na Tabela 14. Nas metodologias LM e GLM, valores de R<sup>2</sup> ajustado foram baixos e as variáveis preditoras explicaram apenas 38 e 31% da variância do teor de C, respectivamente. O menor valor de AIC corresponde a metodologia GLMMs GAM. Os erros de calibração e validação para as técnicas LM e GLM foram de 10 e 12,1 g kg<sup>-1</sup> e 18,4 e 20,5 g kg<sup>-1</sup> respectivamente. As PTFs desenvolvidas a partir das metodologias GLMMs GAM e GLMMs GAU apresentaram valores de R<sup>2</sup> ajustado de 0,6 e 0,54 respectivamente e menores valores de RMSE para validação e calibração de 8,1 e 10,7 g kg<sup>-1</sup> e 8,6 e 11 g kg<sup>-1</sup> respectivamente. Para a dependência espacial, calculado a partir do índice de Moran, com valor p=0,091, os resíduos dos modelos desenvolvidos para o subgrupo MA3 não apresentaram dependência espacial em todas as metodologias, com menor valor de p=0,575 para a metodologia GLM.

Os histogramas dos resíduos e seus respectivos gráficos de dispersão para as quatro metodologias de desenvolvimento das PTFs para o subgrupo MA1 estão representados na Figura 11. Em relação aos histogramas, nas metodologias LM e GLM, os resíduos tiveram distribuição normal e com a maior frequência dos resíduos negativos, o que representa uma tendência em superestimar os teores de C. Através da dispersão dos resíduos, as metodologias GLMMs apresentaram melhor dispersão residual. A análise gráfica dos resíduos do subgrupo MA2 está representada na Figura 11. Nas metodologias LM e GLM os histogramas tiveram distribuição assimétrica e com a maior parte dos resíduos negativos, o que representa uma

tendência em superestimar os teores de C. Nas metodologias GLMMs, a distribuição dos resíduos foi considerada normal, Através da dispersão dos resíduos, as metodologias LM e GLM apresentaram comportamento semelhante, com exceção da metodologia GLMMs, que apresentaram melhor dispersão residual.

A análise gráfica dos resíduos do subgrupo MA3 está representada na Figura 11. Em relação aos histogramas, todas as metodologias apresentaram distribuição normal dos resíduos, com a maior frequência dos resíduos nas classes negativas, o que representa uma tendência em superestimar os teores de C. Mas nas metodologias GLMMs, os resíduos apresentaram maior frequência em valores negativos menores. Através da dispersão dos resíduos, as metodologias GLMMs apresentaram melhor dispersão residual.

Figura 12 - Análise gráfica dos resíduos das PTFs desenvolvidas para o bioma Mata Atlântica subgrupo MA1, MA2 e MA3



Fonte: Autor (2023).

A metodologia GLMMs GAU gerou PTF com os melhores resultados para MA1, configurando a melhor opção para estimar teores de C nos solos das áreas que compõem este subgrupo. Para MA2 e MA3, a metodologia GLMMs GAM gerou a PTF com os melhores resultados e pode ser utilizada para estimar teores de C no solo nas áreas que compõem os respectivos subgrupos.

#### 4.2.2.5 Pampa

O subconjunto de dados agrupados do bioma Pampa, composto por 66 amostras, apresentaram dados de C que variaram de 1,6 a 224 g kg<sup>-1</sup>, com valor médio de 22,3 g kg<sup>-1</sup> (Tabela 15). Dados de granulometria de apresentaram amostras com valores máximos de 900, 580 e 750 g kg<sup>-1</sup>, que correspondem a amostras de areia, silte e argila, respectivamente. Elevados índices de DP para a granulometria e CV para o C configuram a variabilidade de classes de solo presentes no conjunto de dados. O atributo pH apresentou valores que variaram de 3,5 a 6,8 e valor médio de 5,3.

Tabela 15 - Estatística descritiva da base de dados do bioma Pampa (n = 66 amostras)

| Variável                     | Média | DP    | CV%   | Mínimo | Máximo |
|------------------------------|-------|-------|-------|--------|--------|
| C (g kg <sup>-1</sup> )      | 22,3  | 29,5  | 132,6 | 1,6    | 224    |
| Areia (g kg <sup>-1</sup> )  | 472,8 | 237,7 | 50,3  | 15     | 900    |
| Silte (g kg <sup>-1</sup> )  | 278,8 | 134,9 | 48,4  | 20     | 580    |
| Argila (g kg <sup>-1</sup> ) | 248,6 | 148,6 | 59,8  | 20     | 750    |
| pH                           | 5,3   | 0,6   | 11,8  | 3,5    | 6,8    |

n = número de amostras; C = carbono; DP = desvio padrão; CV = coeficiente de variação  
Fonte: Autor (2023)

As PTFs desenvolvidas para estimar teores de C no bioma Pampa estão descritos na Tabela 16. Nas metodologias LM e GLM, os valores de R<sup>2</sup> ajustado foram baixos e as variáveis preditoras explicaram apenas 0,7% e 0,4% da variância do teor de C, respectivamente. Por outro lado, as metodologias GLMMs apresentaram elevados valores de R<sup>2</sup> ajustado, que foi de 89% para o GLMMs GAM e 96% para GLMMs GAU, que também apresentou o menor RMSEC das PTFs. O menor valor de AIC refere-se à metodologia GLMMs GAM.

Para a dependência espacial, calculado a partir de índice de Moran, com valor p=0,440, os resíduos dos modelos desenvolvidos a partir LM, GLM e GLMMs apresentaram

dependência espacial, porém, os resíduos não apresentaram dependência espacial no GLMMs GAM, com valores de  $p=0,759$  (Tabela 16).

Tabela 16 - Funções de pedotransferência para estimativa de teor de C no solo no bioma Pampa desenvolvida a partir de metodologias de regressões lineares e respectivos desempenhos avaliados pelos critérios estatísticos

| Metodologia   | PTFs  |      |           |           |
|---|---|------|-----------|-----------|
| LM  | $C = -745,6 + 1,08 * \text{areia} - 0,0004 * \text{areia}^2 + 1,67 * \text{silte} - 0,001 * \text{silte}^2 + 108,4 * \text{pH} + 0,90 * \text{pH}^2 - 0,0007 * (\text{areia} * \text{silte}) - 0,11 * (\text{areia} * \text{pH}) - 0,18 * (\text{silte} * \text{pH}) - 0,00002 * (\text{areia} * \text{silte} * \text{pH})$       |      |           |           |
| GLM   | $C = -21,7 + 0,04 * \text{areia} - 0,00001 * \text{areia}^2 + 0,06 * \text{silte} - 0,00002 * \text{silte}^2 + 2,94 * \text{pH} + 0,11 * \text{pH}^2 - 0,00002 * (\text{areia} * \text{silte}) - 0,004 * (\text{areia} * \text{pH}) - 0,01 * (\text{silte} * \text{pH}) - 0,000002 * (\text{areia} * \text{silte} * \text{pH})$   |      |           |           |
| GLMMs   | $C = -15,8 + 0,02 * \text{areia} - 0,00001 * \text{areia}^2 + 0,04 * \text{silte} - 0,00002 * \text{silte}^2 + 2,68 * \text{pH} - 0,01 * \text{pH}^2 - 0,000004 * (\text{areia} * \text{silte}) - 0,002 * (\text{areia} * \text{pH}) - 0,003 * (\text{silte} * \text{pH}) - 0,000004 * (\text{areia} * \text{silte} * \text{pH})$ |      |           |           |
| GAM   | $C = -257,7 + 0,21 * \text{areia} - 0,0001 * \text{areia}^2 + 0,37 * \text{silte} - 0,001 * \text{silte}^2 + 50,7 * \text{pH} - 1,81 * \text{pH}^2 + 0,0007 * (\text{areia} * \text{silte}) - 0,008 * (\text{areia} * \text{pH}) + 0,02 * (\text{silte} * \text{pH}) - 0,0002 * (\text{areia} * \text{silte} * \text{pH})$        |      |           |           |
| GLMMs   |   |      |           |           |
| GAU   |   |      |           |           |
| Critérios estatísticos  | LM  | GLM  | GLMMs GAM | GLMMs GAU |
| R <sup>2</sup> ajustado   | 0,07  | 0,04 | 0,89      | 0,96      |
| AIC   | 509,  | 385  | 373       | 485       |
| RMSEC (n=51)  | 12,6  | 9,9  | 9,1       | 6,1       |
| RMSEV (n=15)  | 28,0  | 28,8 | 13,8      | 14,7      |
| Teste de Moran para os resíduos dependência espacial: $p=0,440$ | 0,12  | 0,43 | 0,759     | 0,412     |

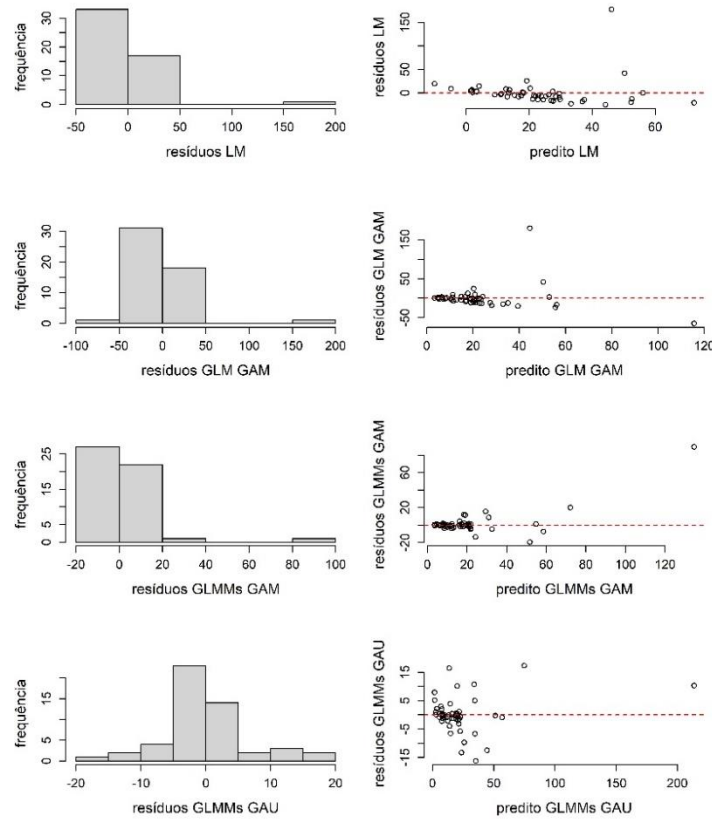
C = carbono; AIC = critério de informação Akaike; RMSEC = raiz quadrada do erro quadrático médio para calibração; RMSEV = raiz quadrada do erro quadrático médio para validação; LM = modelo linear; GLM = modelo linear generalizado; GLMMs GAM = modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados distribuição gama e função de ligação log; GLMMs GAU: modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados distribuição gaussiana e função de ligação identidade.

Fonte: Autor (2023).

Os histogramas dos resíduos e seus respectivos gráficos de dispersão para as quatro metodologias de desenvolvimento das PTFs para o Pampa estão representados na Figura 12. Nas metodologias LM, GLM e GLMMs GAM os histogramas tiveram distribuição assimétrica e com a maior parte dos resíduos negativos, o que representa uma tendência em superestimar os teores de C. Nas metodologias GLMMs, a distribuição dos resíduos foi considerada normal e com maior frequência em menores valores negativos. Através da dispersão dos resíduos, as metodologias LM e GLM apresentaram comportamento semelhante, com exceção da metodologia GLMMs GAU, que apresentou melhor dispersão residual.

A metodologia GLMMs GAU gerou PTF com melhores resultados, com elevado R<sup>2</sup> ajustado e menores valores de RMSE, configurando a melhor metodologia para estimar teores de C em solos do bioma Pampa.

Figura 13 - Análise gráfica dos resíduos das PTFs desenvolvidas para o bioma Pampa



Fonte: Autor (2023).

#### 4.2.2.6 Pantanal

O subconjunto de dados agrupados do bioma Pantanal, composto por 49 amostras, apresentaram dados de C que variaram de 1,7 a 101,8 g kg<sup>-1</sup>, com valor médio de 18,2 g kg<sup>-1</sup> (Tabela 17). Dados de granulometria de apresentaram amostras com valores máximos de 960, 540 e 735 g kg<sup>-1</sup>, que correspondem a amostras de areia, silte e argila, respectivamente. Elevados índices de DP para a granulometria e CV para o C configuram a variabilidade de classes de solo presentes no conjunto de dados. O atributo pH apresentou valores que variaram de 4,2 a 7,9 e valor médio de 5,8.

Tabela 17 - Estatística descritiva da base de dados do bioma Pantanal (n = 49 amostras)

| Variável                    | Média | DP    | CV%  | Mínimo | Máximo |
|-----------------------------|-------|-------|------|--------|--------|
| C (g kg <sup>-1</sup> )     | 18,2  | 18,0  | 98,8 | 1,7    | 101,8  |
| Areia (g kg <sup>-1</sup> ) | 613,9 | 250,5 | 40,8 | 30     | 960    |
| Silte (g kg <sup>-1</sup> ) | 205,3 | 143,2 | 69,7 | 10     | 540    |

|                              |       |       |      |     |     |
|------------------------------|-------|-------|------|-----|-----|
| Argila (g kg <sup>-1</sup> ) | 180,9 | 143,0 | 79,0 | 20  | 735 |
| pH                           | 5,8   | 0,8   | 14,3 | 4,2 | 7,9 |

n = número de amostras; C = carbono; DP = desvio padrão; CV = coeficiente de variação.

Fonte: Autor (2023).

As PTFs desenvolvidas para estimar teores de C no bioma Pantanal estão descritos na Tabela 18. Na metodologia GLM, o valor de R<sup>2</sup> ajustado foi baixo, de 0,44. Por outro lado, as metodologias LM e GLMMs apresentaram elevado R<sup>2</sup> ajustado, que foi de 0,74 para LM e GLMMs GAU, 0,73 para GLMMs GAM. O menor valor de AIC corresponde à metodologia GLM. As metodologias LM e GLMMs GAM apresentaram os menores valores iguais de RMSEC e RMSEV, que foi de 7,9 e 8,5 g kg<sup>-1</sup>, respectivamente.

Para a dependência espacial, calculado a partir de índice de Moran, com valor p=0,640, os resíduos dos modelos desenvolvidos não apresentaram correlação espacial em todas as metodologias analisadas e, em razão disso, as metodologias com os mesmos princípios apresentaram coeficientes iguais nas PTFs.

Tabela 18 - Função de pedotransferência para estimativa de teor de C no solo no bioma Pantanal desenvolvida a partir de metodologias de regressões lineares e respectivos desempenhos avaliados pelos critérios estatísticos

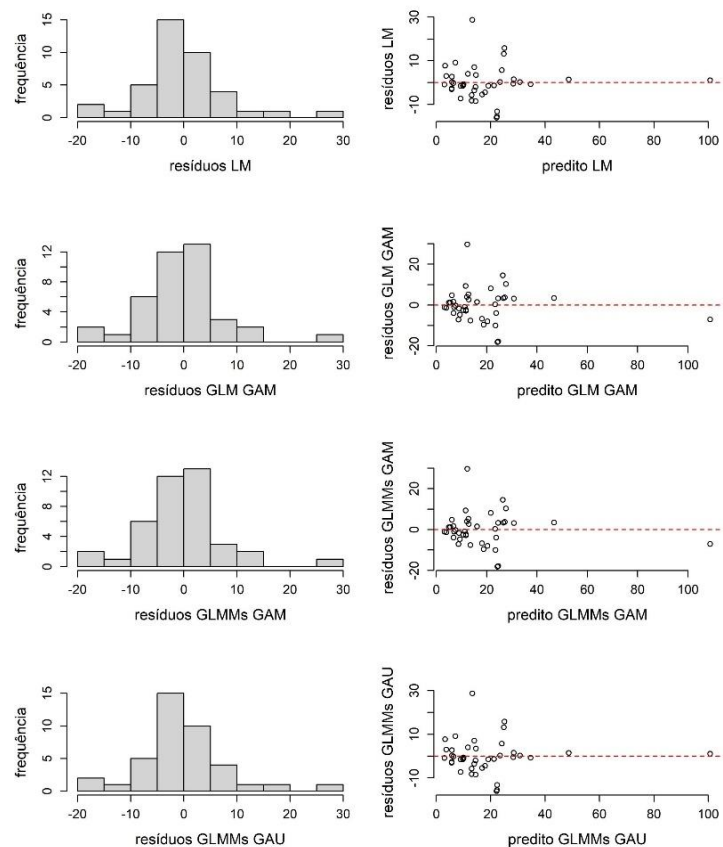
| Metodologia   | PTFs  |       |           |           |
|---|---|-------|-----------|-----------|
| LM  | $C=195,2-0,28*areia-0,00004*areia^2-1,45*silte+0,0006*silte^2+6,09*pH^2-3,33*pH^2+0,0014*(areia*silte)+0,035*(areia*pH)+0,17*(silte*pH)-0,0002*(areia*silte*pH)$    |       |           |           |
| GLM   | $C=8,69-0,004*areia-0,00001*areia^2-0,04*silte+0,00002*silte^2-0,02*pH-0,11*pH^2+0,00003*(areia*silte)+0,001*(areia*pH)+0,005*(silte*pH)-0,00001*(areia*silte*pH)$  |       |           |           |
| GLMMs GAM   | $C=8,69-0,004*areia-0,00001*areia^2-0,04*silte+0,00002*silte^2-0,02*pH-0,11*pH^2+0,00003*(areia*silte)+0,0014*(areia*pH)+0,005*(silte*pH)-0,00001*(areia*silte*pH)$ |       |           |           |
| GLMMs GAU   | $C=195,2-0,28*areia-0,00004*areia^2-1,45*silte+0,001*silte^2+6,09*pH-3,33*pH^2+0,001*(areia*silte)+0,03*(areia*pH)+0,17*(silte*pH)-0,0002*(areia*silte*pH)$         |       |           |           |
| Critérios estatísticos  | LM  | GLM   | GLMMs GAM | GLMMs GAU |
| R <sup>2</sup> ajustado                                       | 0,74  | 0,44  | 0,73      | 0,74      |
| AIC   | 303   | 289   | 295       | 309       |
| RMSEC (n=40)  | 7,9   | 8,2   | 8,2       | 7,9       |
| RMSEV (n=9)   | 8,5   | 9,7   | 9,7       | 8,5       |
| Teste de Moran para os resíduos dependência espacial: p=0,640 | 0,634   | 0,549 | 0,549     | 0,634     |

C = carbono; AIC = critério de informação Akaike; RMSEC = raiz quadrada do erro quadrático médio para calibração; RMSEV = raiz quadrada do erro quadrático médio para validação; LM = modelo linear; GLM = modelo linear generalizado; GLMMs GAM = modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados distribuição gama e função de ligação log; GLMMs GAU: modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados distribuição gaussiana e função de ligação identidade.

Fonte: Autor (2023).

A análise gráfica dos resíduos está representada na Figura 13. Em relação aos histogramas, os resíduos apresentaram distribuição normal em todas as metodologias. Com a maior frequência da distribuição dos resíduos em menores valores negativos, as metodologias LM e GLMMs tendem a superestimar os teores de C preditos e as metodologias GLM e GLMMs GAM tendem a subestimar os teores de C em razão da maior frequência dos resíduos em valores positivos. Através da dispersão dos resíduos, as metodologias LM e GLM apresentaram comportamento semelhante, com exceção da metodologia GLMMs GAU, que apresentou melhor dispersão residual.

Figura 14 - Análise gráfica dos resíduos das PTFs desenvolvidos para o bioma Pantanal



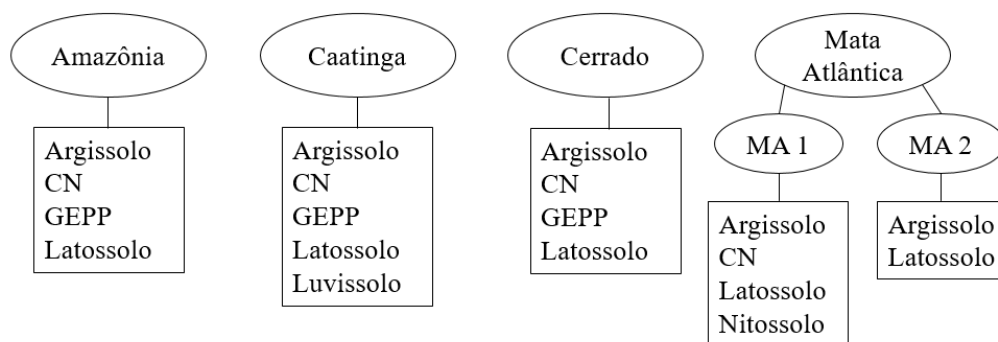
Fonte: Autor (2023).

A metodologia GLMMs GAU gerou a PTF com melhores resultados, com elevado  $R^2$  ajustado e menores valores de RMSE, configurando como a melhor metodologia para estimar teores de C nos solos do bioma Pantanal.

#### 4.2.3 Função de Pedotransferência para classes de solo nos biomas brasileiros (Bioma x Solo)

Nesta etapa, foram agrupados dados de biomas em diferentes classes de solo, com número mínimo de 50 amostras. Dentre as possibilidades, foram desenvolvidas PTFs com a combinação de Bioma x Solo, conforme Figura 14.

Figura 15 - Esquema de agrupamentos de amostras de biomas por classes de solo para o desenvolvimento das PTFs



CN= Cambissolo e Neossolos; GEPP=Gleissolo, Espodossolo, Planossolo e Plintossolo.  
Fonte: Autor (2023)

##### 4.2.3.1 Amazônia agrupado com Argissolo, CN, GEPP e Latossolo

Os modelos desenvolvidos para estimar teores de C em Argissolo, em CN, em GEPP e em Latossolo no bioma Amazônia estão descritos na Tabela 19. Para os Argissolos, nas metodologias LM e GLM, os valores de  $R^2$  ajustado foram baixos e as variáveis preditoras explicaram apenas 26 % e 22% da variância do teor de C, respectivamente. Por outro lado, as metodologias GLMMs apresentaram elevados valores de  $R^2$  ajustado, que foi de 70% para o GLMMs GAM e 67% para GLMMs GAU. O menor valor de AIC refere-se à metodologia GLMMs GAM. Os valores de RMSEC e RMSEV foram similares nas metodologias LM e GLM, que foi de 20,7 e 20,8 g kg<sup>-1</sup> para calibração e 27,9 e 26,7 g kg<sup>-1</sup> para validação, respectivamente. Para a dependência espacial, calculado a partir de índice de Moran, com valor  $p < 0,001$ , os resíduos dos modelos desenvolvidos a partir LM e GLMMs apresentaram dependência espacial, com valores de  $p = 0,602$  para LM,  $p = 0,653$  para GLMMs GAM e  $p = 0,628$  para GLMMs GAU.



Para os CN, nas metodologias LM e GLM, os valores de  $R^2$  ajustado foram baixos e as variáveis preditoras explicaram apenas 23 % e 19% da variância do teor de C, respectivamente. As metodologias GLMMs apresentaram valores melhores de  $R^2$  ajustado, que foi de 0,57 para GLMMs GAM e 0,47 para GLMMs GAU. O menor valor de AIC corresponde à metodologia GLMMs GAM, que apresentou os menores valores RMSEC e RMSEV, que foi de 12,9 e 21,1 g kg<sup>-1</sup>, respectivamente. Para a dependência espacial, calculado a partir de índice de Moran, com valor  $p < 0,029$ , os resíduos dos modelos desenvolvidos apresentaram dependência espacial em todas as metodologias analisadas (Tabela 19).

Tabela 19 - Funções de pedotransferência para estimativa de teor de C em diferentes classes de solo do bioma Amazônia desenvolvidas a partir de metodologias de regressões lineares e respectivos desempenhos avaliados pelos critérios estatísticos

| ARGISSOLO   |   |           |           |           |
|---|---|-----------|-----------|-----------|
| Metodologia   | PTFs  |           |           |           |
| LM  | $C = 203,10 - 0,42 * \text{areia} + 0,0002 * \text{areia}^2 - 0,74 * \text{silte} + 0,0003 * \text{silte}^2 - 5,09 * \text{pH} - 0,41 * \text{pH}^2 + 0,001 * (\text{areia} * \text{silte}) + 0,015 * (\text{areia} * \text{pH}) + 0,07 * (\text{silte} * \text{pH}) - 0,0001 * (\text{areia} * \text{silte} * \text{pH})$              |           |           |           |
| GLM   | $C = 5,69 - 0,01 * \text{areia} - 0,000001 * \text{areia}^2 - 0,02 * \text{silte} + 0,00001 * \text{silte}^2 + 0,32 * \text{pH} - 0,06 * \text{pH}^2 + 0,00002 * (\text{areia} * \text{silte}) + 0,0005 * (\text{areia} * \text{pH}) + 0,002 * (\text{silte} * \text{pH}) - 0,000002 * (\text{areia} * \text{silte} * \text{pH})$       |           |           |           |
| GLMMs<br>GAM  | $C = 1,27 + 0,0003 * \text{areia} - 0,0000002 * \text{areia}^2 - 0,004 * \text{silte} + 0,000002 * \text{silte}^2 + 0,88 * \text{pH} - 0,03 * \text{pH}^2 + 0,00001 * (\text{areia} * \text{silte}) - 0,001 * (\text{areia} * \text{pH}) + 0,00001 * (\text{silte} * \text{pH}) - 0,000002 * (\text{areia} * \text{silte} * \text{pH})$ |           |           |           |
| GLMMs<br>GAU  | $C = 121,8 - 0,25 * \text{areia} + 0,00011 * \text{areia}^2 - 0,39 * \text{silte} + 0,00015 * \text{silte}^2 + 6,08 * \text{pH} - 0,85 * \text{pH}^2 + 0,0005 * (\text{areia} * \text{silte}) + 0,004 * (\text{areia} * \text{pH}) + 0,0323 * (\text{silte} * \text{pH}) - 0,00004 * (\text{areia} * \text{silte} * \text{pH})$         |           |           |           |
| Critérios estatísticos  | LM  | GLM       | GLMMs GAM | GLMMs GAU |
| $R^2$ ajustado  | 0,26  | 0,22      | 0,70      | 0,67      |
| AIC   | 1680  | 1490      | 1364      | 1629      |
| RMSEC (n=186)   | 20,7  | 20,8      | 13,4      | 14        |
| RMSEV (n=47)  | 27,9  | 26,7      | 20,2      | 20,4      |
| Teste de Moran para os resíduos dependência espacial: $p < 0,001$ | 0,602   | $< 0,001$ | 0,653     | 0,628     |
| CN  |   |           |           |           |
| Metodologia   | PTFs  |           |           |           |
| LM  | $C = 409,50 - 0,28 * \text{areia} + 0,000004 * \text{areia}^2 - 0,34 * \text{silte} - 0,0001 * \text{silte}^2 - 109,50 * \text{pH} + 6,79 * \text{pH}^2 + 0,0005 * (\text{areia} * \text{silte}) + 0,05 * (\text{areia} * \text{pH}) + 0,09 * (\text{silte} * \text{pH}) - 0,0001 * (\text{areia} * \text{silte} * \text{pH})$          |           |           |           |
| GLM   | $C = 13,0 - 0,002 * \text{areia} - 0,000003 * \text{areia}^2 - 0,005 * \text{silte} - 0,000004 * \text{silte}^2 - 3,02 * \text{pH} + 0,22 * \text{pH}^2 - 0,000004 * (\text{areia} * \text{silte}) + 0,0005 * (\text{areia} * \text{pH}) + 0,001 * (\text{silte} * \text{pH}) + 0,00000001 * (\text{areia} * \text{silte} * \text{pH})$ |           |           |           |
| GLMMs<br>GAM  | $C = 13,2 - 0,004 * \text{areia} + 0,000001 * \text{areia}^2 - 0,005 * \text{silte} - 0,000002 * \text{silte}^2 - 3,16 * \text{pH} + 0,22 * \text{pH}^2 - 0,000001 * (\text{areia} * \text{silte}) + 0,0004 * (\text{areia} * \text{pH}) + 0,001 * (\text{silte} * \text{pH}) + 0,0000006 * (\text{areia} * \text{silte} * \text{pH})$  |           |           |           |
| GLMMs<br>GAU  | $C = 413,5 - 0,30 * \text{areia} + 0,00003 * \text{areia}^2 - 0,34 * \text{silte} - 0,0001 * \text{silte}^2 - 117,1 * \text{pH} + 7,84 * \text{pH}^2 + 0,0006 * (\text{areia} * \text{silte}) + 0,05 * (\text{areia} * \text{pH}) + 0,08 * (\text{silte} * \text{pH}) - 0,0001 * (\text{areia} * \text{silte} * \text{pH})$             |           |           |           |
| Critérios estatísticos  | LM  | GLM       | GLMMs GAM | GLMMs GAU |

|   |      |       |       |       |
|---|------|-------|-------|-------|
| R <sup>2</sup> ajustado                                       | 0,23 | 0,19  | 0,57  | 0,47  |
| AIC   | 638  | 586   | 560   | 641   |
| RMSEC (n=72)  | 17,2 | 17,6  | 12,9  | 14,4  |
| RMSEV (n=18)  | 23,9 | 23,5  | 21,1  | 21,2  |
| Teste de Moran para os resíduos dependência espacial: p=0,029 | 0,6  | 0,066 | 0,596 | 0,879 |

GEPP

| Metodologia   | PTFs   | LM    | GLM   | GLMMs GAM | GLMMs GAU |
|---|--|-------|-------|-----------|-----------|
| LM  | $C = -39,37 + 0,03 \cdot \text{areia} - 0,00005 \cdot \text{areia}^2 + 0,06 \cdot \text{silte} + 0,00005 \cdot \text{silte}^2 + 19,88 \cdot \text{pH} - 0,83 \cdot \text{pH}^2 + 0,00003 \cdot (\text{areia} \cdot \text{silte}) - 0,01 \cdot (\text{areia} \cdot \text{pH}) - 0,03 \cdot (\text{silte} \cdot \text{pH}) - 0,00001 \cdot (\text{areia} \cdot \text{silte} \cdot \text{pH})$        |       |       |           |           |
| GLM   | $C = -7,19 + 0,001 \cdot \text{areia} - 0,00001 \cdot \text{areia}^2 - 0,0003 \cdot \text{silte} - 0,00001 \cdot \text{silte}^2 + 4,5 \cdot \text{pH} - 0,58 \cdot \text{pH}^2 - 0,00002 \cdot (\text{areia} \cdot \text{silte}) + 0,001 \cdot (\text{areia} \cdot \text{pH}) + 0,002 \cdot (\text{silte} \cdot \text{pH}) + 0,000002 \cdot (\text{areia} \cdot \text{silte} \cdot \text{pH})$     |       |       |           |           |
| GLMMs GAM   | $C = -3,25 + 0,005 \cdot \text{areia} - 0,000004 \cdot \text{areia}^2 + 0,0038 \cdot \text{silte} - 0,000001 \cdot \text{silte}^2 + 2,39 \cdot \text{pH} - 0,24 \cdot \text{pH}^2 - 0,00002 \cdot (\text{areia} \cdot \text{silte}) - 0,0004 \cdot (\text{areia} \cdot \text{pH}) - 0,001 \cdot (\text{silte} \cdot \text{pH}) + 0,000003 \cdot (\text{areia} \cdot \text{silte} \cdot \text{pH})$ |       |       |           |           |
| GLMMs GAU   | $C = 9,35 + 0,03 \cdot \text{areia} - 0,00003 \cdot \text{areia}^2 + 0,05 \cdot \text{silte} + 0,0001 \cdot \text{silte}^2 + 1,95 \cdot \text{pH} + 1,18 \cdot \text{pH}^2 + 0,00008 \cdot (\text{areia} \cdot \text{silte}) - 0,01 \cdot (\text{areia} \cdot \text{pH}) - 0,03 \cdot (\text{silte} \cdot \text{pH}) - 0,00001 \cdot (\text{areia} \cdot \text{silte} \cdot \text{pH})$            |       |       |           |           |
| Critérios estatísticos  |  |       |       |           |           |
| R <sup>2</sup> ajustado                                       |  | 0,09  | 0,04  | 0,58      | 0,38      |
| AIC   |  | 465   | 421   | 401       | 469       |
| RMSEC (n=58)  |  | 10,9  | 11,3  | 7,7       | 9,2       |
| RMSEV (n=23)  |  | 10    | 9,4   | 9,7       | 9,3       |
| Teste de Moran para os resíduos dependência espacial: p=0,846 |  | 0,737 | 0,032 | 0,948     | 0,82      |

LATOSSOLO

| Metodologia  | PTFs  | LM    | GLM   | GLMMs GAM | GLMMs GAU |
|--|---|-------|-------|-----------|-----------|
| LM   | $C = 156,1 - 0,17 \cdot \text{areia} + 0,00003 \cdot \text{areia}^2 - 0,33 \cdot \text{silte} - 0,0003 \cdot \text{silte}^2 - 41,1 \cdot \text{pH} + 2,36 \cdot \text{pH}^2 + 0,0008 \cdot (\text{areia} \cdot \text{silte}) + 0,032 \cdot (\text{areia} \cdot \text{pH}) + 0,12 \cdot (\text{silte} \cdot \text{pH}) - 0,0002 \cdot (\text{areia} \cdot \text{silte} \cdot \text{pH})$           |       |       |           |           |
| GLM  | $C = 11,65 - 0,01 \cdot \text{areia} + 0,000001 \cdot \text{areia}^2 - 0,02 \cdot \text{silte} - 0,00002 \cdot \text{silte}^2 - 2,81 \cdot \text{pH} + 0,17 \cdot \text{pH}^2 + 0,00005 \cdot (\text{areia} \cdot \text{silte}) + 0,002 \cdot (\text{areia} \cdot \text{pH}) + 0,01 \cdot (\text{silte} \cdot \text{pH}) - 0,00002 \cdot (\text{areia} \cdot \text{silte} \cdot \text{pH})$       |       |       |           |           |
| GLMMs GAM  | $C = 5,10 - 0,003 \cdot \text{areia} - 0,0000002 \cdot \text{areia}^2 + 0,003 \cdot \text{silte} - 0,00002 \cdot \text{silte}^2 - 0,71 \cdot \text{pH} + 0,04 \cdot \text{pH}^2 - 0,00001 \cdot (\text{areia} \cdot \text{silte}) + 0,001 \cdot (\text{areia} \cdot \text{pH}) + 0,0012 \cdot (\text{silte} \cdot \text{pH}) + 0,0000004 \cdot (\text{areia} \cdot \text{silte} \cdot \text{pH})$ |       |       |           |           |
| GLMMs GAU  | $C = 91,3 - 0,091 \cdot \text{areia} - 0,000001 \cdot \text{areia}^2 - 0,105 \cdot \text{silte} - 0,0003 \cdot \text{silte}^2 - 20,6 \cdot \text{pH} + 0,92 \cdot \text{pH}^2 + 0,0002 \cdot (\text{areia} \cdot \text{silte}) + 0,0192 \cdot (\text{areia} \cdot \text{pH}) + 0,0652 \cdot (\text{silte} \cdot \text{pH}) - 0,00009 \cdot (\text{areia} \cdot \text{silte} \cdot \text{pH})$     |       |       |           |           |
| Critérios estatísticos                                       |   |       |       |           |           |
| R <sup>2</sup> ajustado                                      |   | 0,14  | 0,13  | 0,4       | 0,32      |
| AIC  |   | 1473  | 1355  | 1327      | 1473      |
| RMSEC (n=168)  |   | 18,1  | 18,9  | 15,2      | 16,2      |
| RMSEV (n=58)   |   | 17,4  | 24,5  | 17,1      | 16,7      |
| Teste de Moran para os resíduos dependência espacial: p=0,02 |   | 0,262 | 0,072 | 0,791     | 0,924     |

C = carbono; AIC = critério de informação Akaike; RMSEC = raiz quadrada do erro quadrático médio para calibração; RMSEV = raiz quadrada do erro quadrático médio para validação; LM = modelo linear; GLM = modelo linear generalizado; GLMMs GAM = modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados distribuição gama e função de ligação log; GLMMs GAU: modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados distribuição gaussiana e função de ligação identidade.

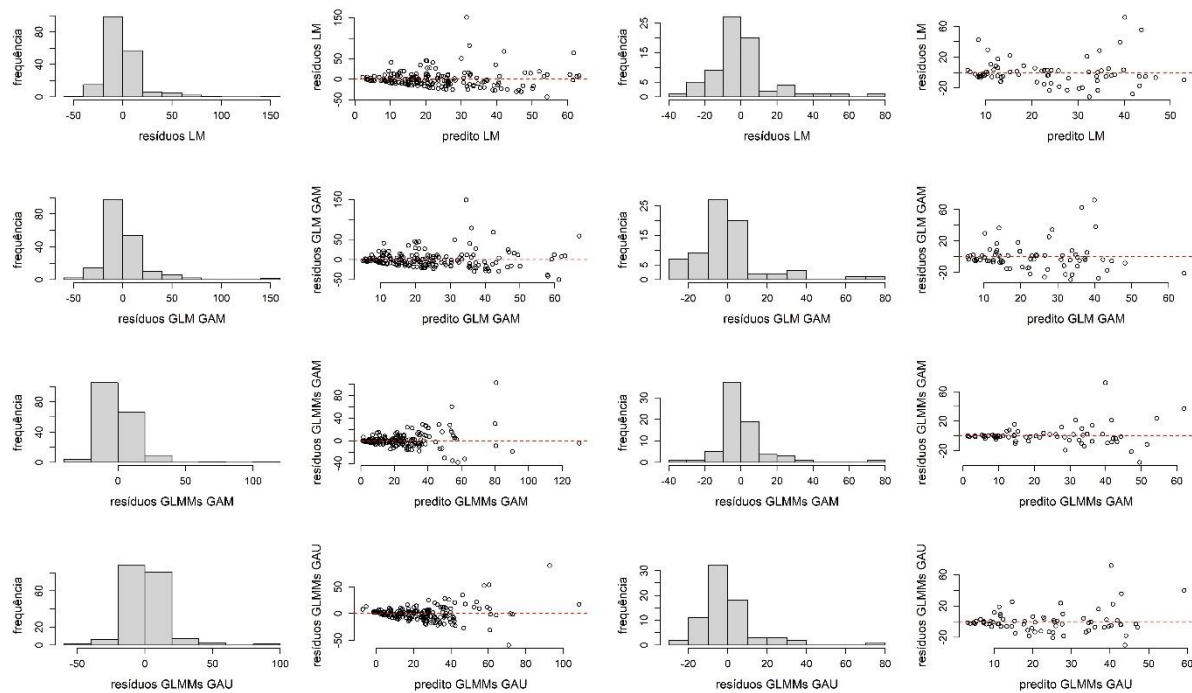
Fonte: Autor

Para os GEPP (Tabela 19), nas metodologias LM e GLM, os valores de  $R^2$  ajustado foram baixos e as variáveis preditoras explicaram apenas 0,9 % e 0,4% da variância do teor de C, respectivamente. As metodologias GLMMs apresentaram valores melhores de  $R^2$  ajustado, que foi de 0,58 para GLMMs GAM e 0,38 para GLMMs GAU. O menor valor de AIC corresponde à metodologia GLMMs GAM, que apresentou os menores valores RMSEC e RMSEV, que foi de 7,7 e 9,7 g kg<sup>-1</sup>, respectivamente. Para a dependência espacial, calculado a partir de índice de Moran, com valor  $p < 0,846$ , os resíduos dos modelos desenvolvidos a partir de GLM não apresentaram dependência espacial, porém, o maior valor de  $p$  refere-se a metodologia GLMMs GAM, com  $p = 0,948$ .

Para os Latossolos (Tabela 19), nas metodologias LM e GLM, os valores de  $R^2$  ajustado foram baixos e as variáveis preditoras explicaram apenas 14 % e 13% da variância do teor de C, respectivamente. As metodologias GLMMs apresentaram valores de  $R^2$  ajustado de 0,4 para GLMMs GAM e 0,32 para GLMMs GAU. O menor valor de AIC corresponde à metodologia GLMMs GAM, que apresentou os menores valores RMSEC e RMSEV, que foi de 15,2 e 17,1 g kg<sup>-1</sup>, respectivamente. Para a dependência espacial, calculado a partir de índice de Moran, com valor  $p = 0,02$ , os resíduos dos modelos desenvolvidos apresentaram dependência espacial em todas as metodologias analisadas, com maior valor para GLMMs GAU, com  $p = 0,924$  (Tabela 19).

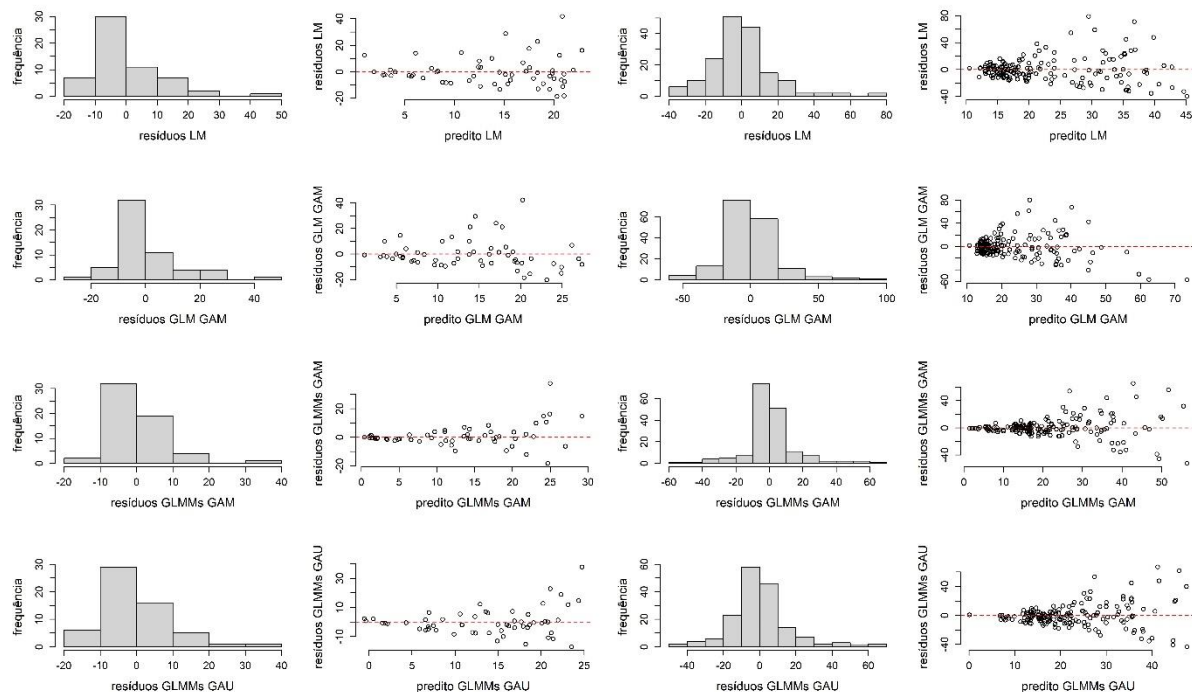
A análise gráfica dos resíduos está representada na Figura 15. Em todos os agrupamentos das amostras de Argissolo, CN, GEPP e Latossolos na Amazônia, os resíduos apresentaram tendência a superestimar os valores de C, em razão da maior frequência dos dados nas classes de valores negativos e a metodologia GLMMs GAM apresentou melhor dispersão residual, configurando como a melhor sugestão para desenvolver PTFs para essas classes de solo no bioma Amazônia, com elevados valores de  $R^2$  ajustado e menores valores de RMSE.

Figura 16 - Análise gráfica dos resíduos das PTFs desenvolvidas para o bioma Amazônia agrupados com Argissolo, CN, GEEP e Latossolo



Amazônia x Argissolo

Amazônia x CN



Amazônia x GEEP

Amazônia x Latossolo

Fonte: Autor (2023).

#### 4.2.3.2 Caatinga agrupado com Argissolo, CN, GEPP, Latossolo e Luvisolo

As PTFs desenvolvidas para estimar teores de C em diferentes classes de solo do bioma Caatinga estão descritos na Tabela 20. Para os Argissolos, nas metodologias LM e GLM, os valores de R<sup>2</sup> ajustado foram baixos e as variáveis preditoras explicaram apenas 34 % e 31% da variância do teor de C, respectivamente. Por outro lado, as metodologias GLMMs apresentaram elevados valores de R<sup>2</sup> ajustado, que foi de 48% para o GLMMs GAM e 88% para GLMMs GAU. O menor valor de AIC refere-se à metodologia GLM. Os valores de RMSEC e RMSEV foram similares nas metodologias LM e GLM, que foi de 6,3 g kg<sup>-1</sup> para calibração e 22,6 e 22,8 g kg<sup>-1</sup> para validação, respectivamente. Os menores valores de RMSE foram da metodologia GLMMs GAU, com 3 g kg<sup>-1</sup> para calibração e 22,3 g kg<sup>-1</sup> para validação. Para a dependência espacial, calculado a partir de índice de Moran, com valor p<0,001, os resíduos dos modelos desenvolvidos a partir LM, GLM e GLMMs GAU apresentaram dependência espacial, com valores de p=0,6, p=0,247 e p=0,872, respectivamente.

Para os CN (Tabela 20), nas metodologias LM e GLM, os valores de R<sup>2</sup> ajustado foram baixos e as variáveis preditoras explicaram apenas 23 % e 21% da variância do teor de C, respectivamente. As metodologias GLMMs apresentaram melhores valores de R<sup>2</sup> ajustado, que foi de 0,46 para GLMMs GAM e 0,61 para GLMMs GAU. As metodologias GLM e GLMMs GAM apresentaram o menor valor de AIC. A metodologia GLMMs GAU apresentou os menores valores RMSEC e RMSEV, que foi de 6,3 e 7 g kg<sup>-1</sup>, respectivamente. Para a dependência espacial, calculado a partir de índice de Moran, com valor p<0,001, apenas a metodologia GLMMs GAM não apresentou resíduos com dependência espacial e o maior valor refere-se a GLMMs GAU, com p=0,872.

Tabela 20 - Funções de pedotransferência para estimativa de teor de C em diferentes classes de solo do bioma Caatinga desenvolvidas a partir de metodologias de regressões lineares e respectivos desempenhos avaliados pelos critérios estatísticos

| ARGISSOLO    |   |
|--------------|---|
| Metodologia  | PTFs  |
| LM           | $C = -97,6 + 0,22 * \text{areia} - 0,0001 * \text{areia}^2 + 0,93 * \text{silte} - 0,001 * \text{silte}^2 + 8,11 * \text{pH} - 0,80 * \text{pH}^2 - 0,001 * (\text{areia} * \text{silte}) - 0,003 * (\text{areia} * \text{pH}) - 0,07 * (\text{silte} * \text{pH}) + 0,0002 * (\text{areia} * \text{silte} * \text{pH})$              |
| GLM          | $C = -4,51 + 0,016 * \text{areia} - 0,00001 * \text{areia}^2 + 0,0447 * \text{silte} - 0,00005 * \text{silte}^2 + 0,37 * \text{pH} - 0,0812 * \text{pH}^2 - 0,0001 * (\text{areia} * \text{silte}) + 0,0004 * (\text{areia} * \text{pH}) - 0,0008 * (\text{silte} * \text{pH}) + 0,00001 * (\text{areia} * \text{silte} * \text{pH})$ |
| GLMMs<br>GAM | $C = -2,49 + 0,013 * \text{areia} - 0,00001 * \text{areia}^2 + 0,034 * \text{silte} - 0,00004 * \text{silte}^2 + 0,11 * \text{pH} - 0,085 * \text{pH}^2 - 0,0001 * (\text{areia} * \text{silte}) + 0,0008 * (\text{areia} * \text{pH}) + 0,0007 * (\text{silte} * \text{pH}) + 0,000003 * (\text{areia} * \text{silte} * \text{pH})$  |
| GLMMs<br>GAU | $C = -92,42 + 0,24 * \text{areia} - 0,0001 * \text{areia}^2 + 0,92 * \text{silte} - 0,001 * \text{silte}^2 + 3,85 * \text{pH} - 0,41 * \text{pH}^2 - 0,001 * (\text{areia} * \text{silte}) - 0,004 * (\text{areia} * \text{pH}) - 0,07 * (\text{silte} * \text{pH}) + 0,0001 * (\text{areia} * \text{silte} * \text{pH})$             |

| Critérios estatísticos  | LM   | GLM   | GLMMs GAM | GLMMs GAU |
|---|------|-------|-----------|-----------|
| R <sup>2</sup> ajustado                                       | 0,34 | 0,31  | 0,42      | 0,88      |
| AIC   | 955  | 911   | 913       | 951       |
| RMSEC (n=143)   | 6,3  | 6,3   | 5,9       | 3,0       |
| RMSEV (n=50)  | 22,6 | 22,8  | 22,4      | 22,3      |
| Teste de Moran para os resíduos dependência espacial: p<0,001 | 0,6  | 0,247 | <0,001    | 0,872     |

CN

| Metodologia | PTFs  |
|-------------|---|
| LM          | $C = -17,09 + 0,0639 * \text{areia} - 0,00001 * \text{areia}^2 + 0,115 * \text{silte} - 0,0001 * \text{silte}^2 + 0,08 * \text{pH} + 0,71 * \text{pH}^2 + 0,0001 * \text{areia} * \text{silte} - 0,01 * (\text{areia} * \text{pH}) - 0,0082 * (\text{silte} * \text{pH}) + 0,00002 * (\text{areia} * \text{silte} * \text{pH})$             |
| GLM         | $C = -1,23 + 0,011 * \text{areia} - 0,000005 * \text{areia}^2 + 0,014 * \text{silte} - 0,00001 * \text{silte}^2 + 0,009 * \text{pH} + 0,048 * \text{pH}^2 - 0,00001 * (\text{areia} * \text{silte}) - 0,001 * (\text{areia} * \text{pH}) - 0,001 * (\text{silte} * \text{pH}) - 0,0000001 * (\text{areia} * \text{silte} * \text{pH})$      |
| GLMMs GAM   | $C = -2,06 + 0,01 * \text{areia} - 0,00001 * \text{areia}^2 + 0,017340 * \text{silte} - 0,00001 * \text{silte}^2 + 0,04826 * \text{pH} + 0,044 * \text{pH}^2 - 0,00001 * (\text{areia} * \text{silte}) - 0,00087 * (\text{areia} * \text{pH}) - 0,001 * (\text{silte} * \text{pH}) + 0,0000001 * (\text{areia} * \text{silte} * \text{pH})$ |
| GLMMs GAU   | $C = -29,3 + 0,10 * \text{areia} - 0,00004 * \text{areia}^2 + 0,171 * \text{silte} - 0,00011 * \text{silte}^2 - 0,67370 * \text{pH} + 0,789 * \text{pH}^2 + 0,00001 * (\text{areia} * \text{silte}) - 0,0095 * (\text{areia} * \text{pH}) - 0,011 * (\text{silte} * \text{pH}) - 0,000018 * (\text{areia} * \text{silte} * \text{pH})$      |

| Critérios estatísticos  | LM    | GLM   | GLMMs GAM | GLMMs GAU |
|---|-------|-------|-----------|-----------|
| R <sup>2</sup> ajustado                                       | 0,23  | 0,21  | 0,46      | 0,61      |
| AIC   | 1231  | 1113  | 1113      | 1232      |
| RMSEC (n=72)  | 8,0   | 7,7   | 7,1       | 6,3       |
| RMSEV (n=18)  | 8,4   | 8,5   | 7,2       | 7,0       |
| Teste de Moran para os resíduos dependência espacial: p=0,267 | 0,058 | 0,003 | 0,223     | 0,618     |

GEPP

| Metodologia | PTFs   |
|-------------|--|
| LM          | $C = 266,4 - 0,09 * \text{areia} - 0,0001 * \text{areia}^2 - 0,19 * \text{silte} - 0,0001 * \text{silte}^2 - 64,9 * \text{pH} + 3,5 * \text{pH}^2 + 0,00003 * (\text{areia} * \text{silte}) + 0,02 * (\text{areia} * \text{pH}) + 0,042 * (\text{silte} * \text{pH}) - 0,00002 * (\text{areia} * \text{silte} * \text{pH})$        |
| GLM         | $C = 14,3 + 0,01 * \text{areia} - 0,00001 * \text{areia}^2 + 0,001 * \text{silte} - 0,00001 * \text{silte}^2 - 4,2 * \text{pH} + 0,27 * \text{pH}^2 - 0,00001 * (\text{areia} * \text{silte}) + 0,001 * (\text{areia} * \text{pH}) + 0,002 * (\text{silte} * \text{pH}) - 0,000001 * (\text{areia} * \text{silte} * \text{pH})$    |
| GLMMs GAM   | $C = 9,3 + 0,009 * \text{areia} - 0,00001 * \text{areia}^2 + 0,010 * \text{silte} - 0,00001 * \text{silte}^2 - 3,1 * \text{pH} + 0,18 * \text{pH}^2 - 0,00001 * (\text{areia} * \text{silte}) + 0,0010 * (\text{areia} * \text{pH}) + 0,0013 * (\text{silte} * \text{pH}) - 0,000002 * (\text{areia} * \text{silte} * \text{pH})$  |
| GLMMs GAU   | $C = 204,5 + 0,007 * \text{areia} - 0,00012 * \text{areia}^2 + 0,004 * \text{silte} - 0,00019 * \text{silte}^2 - 61,2 * \text{pH} + 3,32 * \text{pH}^2 - 0,00004 * (\text{areia} * \text{silte}) + 0,025 * (\text{areia} * \text{pH}) + 0,036 * (\text{silte} * \text{pH}) - 0,000037 * (\text{areia} * \text{silte} * \text{pH})$ |

| Critérios estatísticos  | LM    | GLM   | GLMMs GAM | GLMMs GAU |
|---|-------|-------|-----------|-----------|
| R <sup>2</sup> ajustado                                       | 0,43  | 0,36  | 0,72      | 0,92      |
| AIC   | 579   | 537   | 542       | 576       |
| RMSEC (n=84)  | 6,6   | 7,3   | 4,7       | 2,6       |
| RMSEV (n=27)  | 6,5   | 6,1   | 6         | 7,9       |
| Teste de Moran para os resíduos dependência espacial: p=0,547 | 0,446 | 0,403 | 0,771     | 0,153     |

LATOSSOLO

| Metodologia | PTFs  |
|-------------|---|
| LM          | $C = -27,7 - 0,04 * \text{areia} - 0,00001 * \text{areia}^2 - 0,43440 * \text{silte} + 0,001 * \text{silte}^2 + 35,0 * \text{pH} - 1,53 * \text{pH}^2 + 0,001 * (\text{areia} * \text{silte}) - 0,0114 * (\text{areia} * \text{pH}) - 0,04 * (\text{silte} * \text{pH}) - 0,0001 * (\text{areia} * \text{silte} * \text{pH})$ |

| GLM   | $C=2,3-0,005*areia-0,000002*areia^2-0,027*silte+0,00003*silte^2+1,3*pH-0,11*pH^2+0,00005*(areia*silte)+0,0004*(areia*pH)+0,0014*(silte*pH)-0,000007*(areia*silte*pH)$ |       |           |           |
|---|---|-------|-----------|-----------|
| GLMMs<br>GAM  | $C=2,8-0,003*areia-0,00002*areia^2-0,02*silte+0,00002*silte^2+0,88*pH-0,06*pH^2+0,00004*(areia*silte)+0,0002*(areia*pH)+0,0006*(silte*pH)-0,000006*(areia*silte*pH)$  |       |           |           |
| GLMMs<br>GAU  | $C=-6,81+0,04*areia-0,00003*areia^2+0,10*silte+0,001*silte^2+9,11*pH+0,65*pH^2+0,0001*(areia*silte)-0,01*(areia*pH)-0,07*(silte*pH)+0,00001*(areia*silte*pH)$         |       |           |           |
| Critérios estatísticos  | LM  | GLM   | GLMMs GAM | GLMMs GAU |
| R <sup>2</sup> ajustado                                       | 0,37  | 0,36  | 0,68      | 0,95      |
| AIC   | 561   | 518   | 523       | 552       |
| RMSEC (n=76)  | 8,3   | 8,3   | 6,4       | 2,4       |
| RMSEV (n=26)  | 8,2   | 8,0   | 8,1       | 10,9      |
| Teste de Moran para os resíduos dependência espacial: p<0,001 | 0,002   | 0,128 | 0,66      | 0,388     |

#### LUVISSOLO

| Metodologia   | PTFs   |       |           |           |
|---|--|-------|-----------|-----------|
| LM  | $C=171,9-0,05*areia-0,0001*areia^2-0,32*silte+0,00008*silte^2-26,4*pH+0,007*pH^2-0,0002*(areia*silte)+0,0249*(areia*pH)+0,05*(silte*pH)+0,00002*(areia*silte)$       |       |           |           |
| GLM   | $C=9,7+0,003*areia-0,00001*areia^2-0,02*silte+0,00001*silte^2-1,27*pH+0,003*pH^2-0,00002*(areia*silte)+0,0008*(areia*pH)+0,002*(silte*pH)+0,000004*(areia*silte*pH)$ |       |           |           |
| GLMMs<br>GAM  | $C=9,7+0,003*areia-0,00001*areia^2-0,02*silte+0,00001*silte^2-1,27*pH+0,003*pH^2-0,00002*(areia*silte)+0,001*(areia*pH)+0,002*(silte*pH)+0,000004*(areia*silte*pH)$  |       |           |           |
| GLMMs<br>GAU  | $C=162,9-0,09*areia-0,0001*areia^2-0,38*silte+0,0002*silte^2-20,8*pH-0,06*pH^2+0,00002*(areia*silte)+0,025*(areia*pH)+0,042*(silte*pH)+0,000002*(areia*silte*pH)$    |       |           |           |
| Critérios estatísticos  | LM   | GLM   | GLMMs GAM | GLMMs GAU |
| R <sup>2</sup> ajustado                                       | 0,22   | 0,19  | 0,23      | 0,59      |
| AIC   | 345  | 334   | 340       | 350       |
| RMSEC (n=46)  | 7,9  | 7,9   | 7,9       | 5,9       |
| RMSEV (n=9)   | 11,9   | 13,2  | 13,2      | 12,3      |
| Teste de Moran para os resíduos dependência espacial: p=0,762 | 0,183  | 0,256 | 0,256     | 0,942     |

n= número de amostras; C = carbono; AIC = critério de informação Arkaike; RMSEC = raiz quadrada do erro quadrático médio para calibração; RMSEV = raiz quadrada do erro quadrático médio para validação; LM = modelo linear; GLM = modelo linear generalizado; GLMMs GAM = modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados distribuição gama e função de ligação log; GLMMs GAU: modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados distribuição gaussiana e função de ligação identidade.

Fonte: Autor

Para os GEPP (Tabela 20), nas metodologias LM e GLM, os valores de R<sup>2</sup> ajustado foram 0,43 e 0,36 respectivamente. As metodologias GLMMs apresentaram valores elevados de R<sup>2</sup> ajustado, que foi de 0,72 para GLMMs GAM e 0,92 para GLMMs GAU, que apresentou menor RMSEC, que foi de 2,6 g kg<sup>-1</sup>. O menor valor de AIC corresponde à metodologia GLM. Para a dependência espacial, calculado a partir de índice de Moran, com valor p=0,547, os resíduos dos modelos desenvolvidos apresentaram dependência espacial apenas na metodologia GLMMs GAM, com p=0,771.

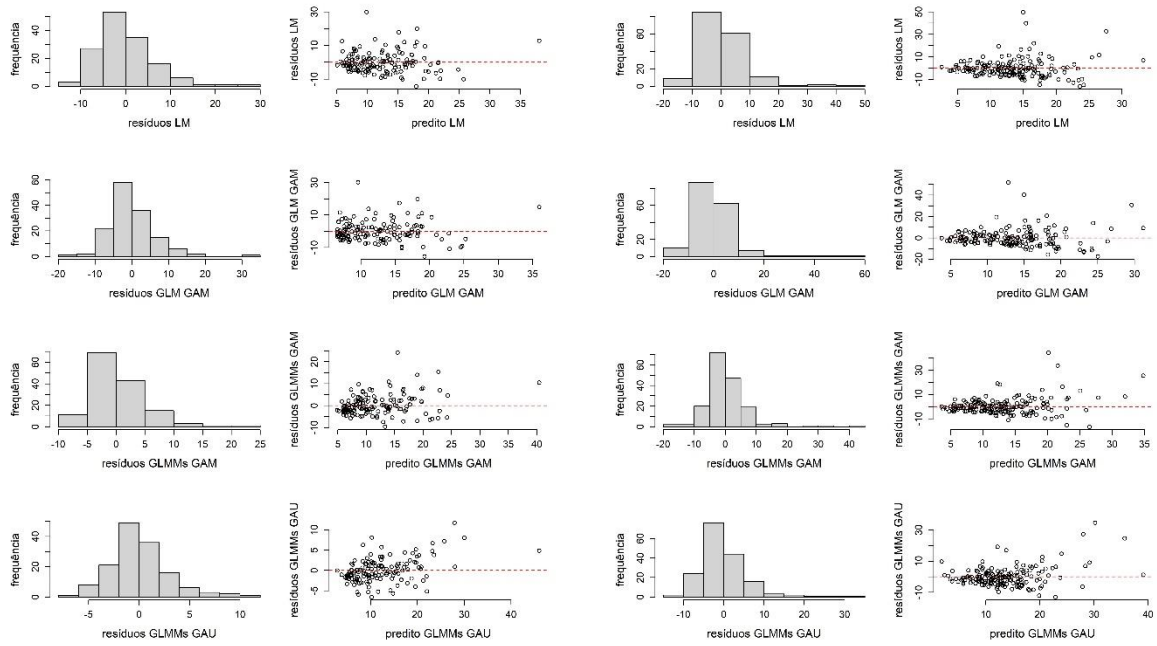
Para os Latossolos (Tabela 20) nas metodologias LM e GLM, os valores de  $R^2$  ajustado foram 0,37 e 0,36 respectivamente. As metodologias GLMMs apresentaram valores elevados de  $R^2$  ajustado, que foi de 0,68 para GLMMs GAM e 0,95 para GLMMs GAU, que apresentou menor RMSEC, que foi de 2,4 g kg<sup>-1</sup>. O menor valor de AIC corresponde à metodologia GLM. Para a dependência espacial, calculado a partir de índice de Moran, com valor  $p < 0,001$ , os resíduos dos modelos desenvolvidos apresentaram dependência espacial em todas as metodologias analisadas, com maior valor para GLLMs GAU, com  $p = 0,66$ .

Para os Luvisolos (Tabela 20) nas metodologias LM e GLM, os valores de  $R^2$  ajustado foram 0,22 e 0,19 respectivamente. As metodologias GLMMs apresentaram valores de  $R^2$  ajustado de 0,23 para GLMMs GAM e 0,59 para GLMMs GAU, que apresentou menor RMSEC, que foi de 5,9 g kg<sup>-1</sup>. O menor valor de AIC corresponde à metodologia GLM. Para a dependência espacial, calculado a partir de índice de Moran, com valor  $p = 0,762$ , os resíduos dos modelos desenvolvidos não apresentaram dependência espacial, com exceção do GLMMs GAU que apresentou dependência espacial, com  $p = 0,942$ .

A análise gráfica dos resíduos está representada na Figura 16. Em todos os agrupamentos de Argissolo, CN, GEEP, Latossolos e Luvisolo na Caatinga, os resíduos apresentaram tendência a superestimar os valores de C, em razão da maior frequência dos dados nas classes de valores negativos e a metodologia GLMMs GAU apresentou melhor dispersão residual, configurando como a melhor sugestão para desenvolver PTFs nessas classes de solo no bioma Caatinga, com elevados valores de  $R^2$  ajustado e menores valores de RMSE para calibração e validação.

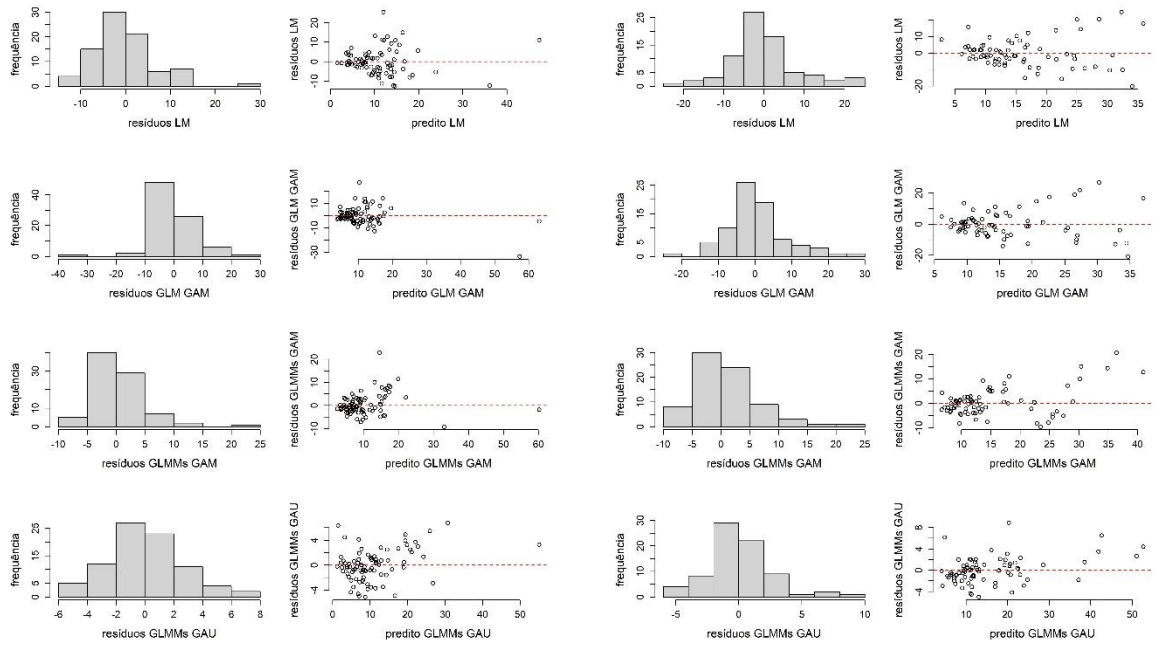


Figura 17 - Análise gráfica dos resíduos das PTFs desenvolvidas para o bioma Caatinga agrupados com Argissolo, CN, GEEP, Latossolo e Luvissole



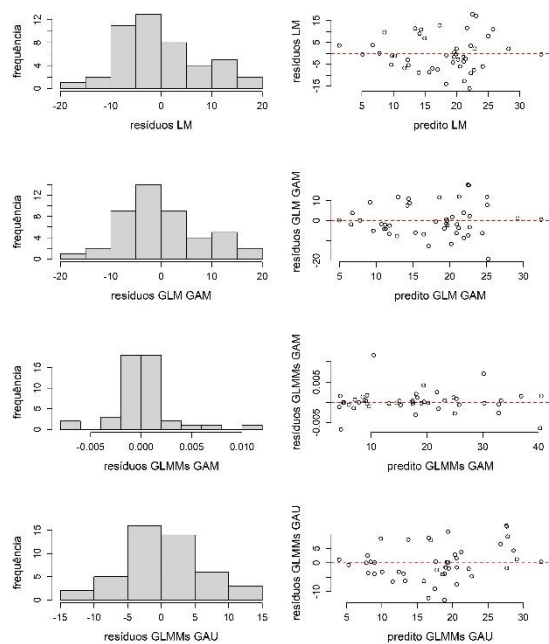
Caatinga x Argissolo

Caatinga x CN



Caatinga x GEEP

Caatinga x Latossolo



Caatinga x Luvissolo

Fonte: Autor (2023).

#### 4.2.3.3 Cerrado agrupado com Argissolo, CN, GEPP e Latossolo

As PTFs desenvolvidos para estimar teores de C em diferentes classes de solo do bioma Cerrado estão descritos na Tabela 21. Para os Argissolos nas metodologias LM e GLM, os valores de  $R^2$  ajustado foram baixos e variáveis predictoras explicaram apenas 20 e 19% da variância do teor de C, respectivamente. As metodologias GLMMs apresentaram valores de  $R^2$  ajustado de 0,51 em GLMMs GAM e 0,72 em GLMMs GAU, que apresentou os menores valores de RMSE, de 8,7 g kg<sup>-1</sup> de calibração e 8,9 8,7 g kg<sup>-1</sup> de validação. O menor valor de AIC corresponde à metodologia GLM. Para a dependência espacial, calculado a partir de índice de Moran, com valor p=0,004, os resíduos apresentaram dependência espacial em todas as metodologias.

Para os CN (Tabela 21), nas metodologias LM e GLM, os valores de  $R^2$  ajustado foram baixos e as variáveis predictoras explicaram apenas 34 % e 30% da variância do teor de C, respectivamente. As metodologias GLMMs apresentaram valores iguais de  $R^2$  ajustado, que foi de 0,85. AS metodologias GLM e GLMMs GAM apresentaram o menor valor de AIC. A metodologia GLMMs GAU apresentou o menor valore RMSEC, que foi de 6,5g kg<sup>-1</sup>. Para a dependência espacial, calculado a partir de índice de Moran, com valor p=0,919, os resíduos

dos modelos desenvolvidos não apresentaram dependência espacial em todas as metodologias analisadas e o maior valor refere-se a metodologia GAM com valor de  $p=0,875$ .

Tabela 21 - Funções de pedotransferência para estimativa de teor de C em diferentes classes de solo do bioma Cerrado desenvolvidas a partir de metodologias de regressões lineares e respectivos desempenhos avaliados pelos critérios estatísticos

| ARGISSOLO   |  |        |           |           |
|---|--|--------|-----------|-----------|
| Metodologia   | PTFs   |        |           |           |
| LM  | $C = 203,4 - 0,10 * \text{areia} - 0,0001 * \text{areia}^2 - 0,26 * \text{silte} - 0,0003 * \text{silte}^2 - 46,6 * \text{pH} + 1,13 * \text{pH}^2 + 0,0002 * (\text{areia} * \text{silte}) + 0,04 * (\text{areia} * \text{pH}) + 0,09 * (\text{silte} * \text{pH}) - 0,0001 * (\text{areia} * \text{silte} * \text{pH})$              |        |           |           |
| GLM   | $C = 13,4 - 0,003 * \text{areia} - 0,00001 * \text{areia}^2 - 0,01 * \text{silte} - 0,00002 * \text{silte}^2 - 3,0 * \text{pH} + 0,11 * \text{pH}^2 + 0,000003 * (\text{areia} * \text{silte}) + 0,002 * (\text{areia} * \text{pH}) + 0,005 * (\text{silte} * \text{pH}) - 0,000003 * (\text{areia} * \text{silte} * \text{pH})$       |        |           |           |
| GLMMs   | $C = 9,99 - 0,002 * \text{areia} - 0,00001 * \text{areia}^2 - 0,01006 * \text{silte} - 0,00001 * \text{silte}^2 - 1,97 * \text{pH} + 0,07 * \text{pH}^2 + 0,00001 * (\text{areia} * \text{silte}) + 0,0017 * (\text{areia} * \text{pH}) + 0,004 * (\text{silte} * \text{pH}) - 0,000004 * (\text{areia} * \text{silte} * \text{pH})$   |        |           |           |
| GAM   |  |        |           |           |
| GLMMs   | $C = 118,0 - 0,095 * \text{areia} - 0,00009 * \text{areia}^2 - 0,18 * \text{silte} - 0,0003 * \text{silte}^2 - 23,4 * \text{pH} + 0,19 * \text{pH}^2 + 0,0002 * (\text{areia} * \text{silte}) + 0,033 * (\text{areia} * \text{pH}) + 0,068 * (\text{silte} * \text{pH}) - 0,00008 * (\text{areia} * \text{silte} * \text{pH})$         |        |           |           |
| GAU   |  |        |           |           |
| Critérios estatísticos  | LM   | GLM    | GLMMs GAM | GLMMs GAU |
| R <sup>2</sup> ajustado   | 0,2  | 0,19   | 0,51      | 0,72      |
| AIC   | 889  | 806    | 807       | 889       |
| RMSEC (n=107)   | 9,5  | 9,8    | 9,1       | 8,7       |
| RMSEV (n= 44)   | 13,8   | 14,1   | 11,2      | 8,9       |
| Teste de Moran para os resíduos dependência espacial: $p=0,004$ | 0,456  | 0,717  | 0,257     | 0,484     |
| CN  |  |        |           |           |
| Metodologia   | PTFs   |        |           |           |
| LM  | $C = -207,7 + 0,213 * \text{areia} - 0,00001 * \text{areia}^2 + 0,36 * \text{silte} + 0,00001 * \text{silte}^2 + 49,2 * \text{pH} - 0,66 * \text{pH}^2 - 0,0002 * (\text{areia} * \text{silte}) - 0,045 * (\text{areia} * \text{pH}) - 0,077 * (\text{silte} * \text{pH}) + 0,00006 * (\text{areia} * \text{silte} * \text{pH})$       |        |           |           |
| GLM   | $C = -1,8 + 0,004 * \text{areia} - 0,000002 * \text{areia}^2 + 0,008 * \text{silte} + 0,0000005 * \text{silte}^2 + 0,81 * \text{pH} + 0,03 * \text{pH}^2 + 0,00001 * (\text{areia} * \text{silte}) - 0,0008 * (\text{areia} * \text{pH}) - 0,002 * (\text{silte} * \text{pH}) - 0,0000004 * (\text{areia} * \text{silte} * \text{pH})$ |        |           |           |
| GLMMs   | $C = -0,7 + 0,003 * \text{areia} - 0,000003 * \text{areia}^2 + 0,01 * \text{silte} - 0,000001 * \text{silte}^2 + 0,7 * \text{pH} + 0,01 * \text{pH}^2 + 0,000003 * (\text{areia} * \text{silte}) - 0,0005 * (\text{areia} * \text{pH}) - 0,001 * (\text{silte} * \text{pH}) - 0,0000002 * (\text{areia} * \text{silte} * \text{pH})$   |        |           |           |
| GAM   |  |        |           |           |
| GLMMs   | $C = -208,2 + 0,205 * \text{areia} - 0,00002 * \text{areia}^2 + 0,40 * \text{silte} + 0,000003 * \text{silte}^2 + 52,8 * \text{pH} - 0,92 * \text{pH}^2 - 0,0002 * (\text{areia} * \text{silte}) - 0,0445 * (\text{areia} * \text{pH}) - 0,085 * (\text{silte} * \text{pH}) + 0,000061 * (\text{areia} * \text{silte} * \text{pH})$    |        |           |           |
| GAU   |  |        |           |           |
| Critérios estatísticos  | LM   | GLM    | GLMMs GAM | GLMMs GAU |
| R <sup>2</sup> ajustado   | 0,34   | 0,3    | 0,85      | 0,85      |
| AIC   | 1186   | 1047   | 999       | 1166      |
| RMSEC (n=146)   | 12,9   | 13     | 6,8       | 6,5       |
| RMSEV (n= 54)   | 11,8   | 11,6   | 9,9       | 10,7      |
| Teste de Moran para os resíduos dependência espacial: $p=0,919$ | 0,167  | <0,001 | 0,875     | 0,838     |
| GEPP  |  |        |           |           |
| Metodologia   | PTFs   |        |           |           |
| LM  | $C = 112,3 + 0,06 * \text{areia} - 0,0001 * \text{areia}^2 + 0,36 * \text{silte} - 0,0004 * \text{silte}^2 - 37,84 * \text{pH} + 1,91 * \text{pH}^2 - 0,0002 * (\text{areia} * \text{silte}) + 0,02 * (\text{areia} * \text{pH}) + 0,004 * (\text{silte} * \text{pH}) - 0,00004 * (\text{areia} * \text{silte} * \text{pH})$           |        |           |           |

|   |   |        |           |           |
|---|---|--------|-----------|-----------|
| GLM   | $C=2,3+0,005*areia-0,00001*areia^2+0,01*silte-0,00001*silte^2-0,08*pH-0,02*pH^2+0,00001*(areia*silte)+0,0002*(areia*pH)+0,001*(silte*pH)-0,000004*(areia*silte*pH)$   |        |           |           |
| GLMMs<br>GAM  | $C=2,31+0,004*areia-0,00001*areia^2+0,01*silte-0,00002*silte^2-0,07*pH-0,04*pH^2-0,000001*(areia*silte)+0,001*(areia*pH)+0,0008*(silte*pH)-0,000003*(areia*silte*pH)$ |        |           |           |
| GLMMs<br>GAU  | $C=91,7+0,064*areia-0,00010*areia^2+0,52*silte-0,0005*silte^2-39,2*pH+1,91*pH^2-0,00045*(areia*silte)+0,018*(areia*pH)-0,002*(silte*pH)-0,000002*(areia*silte*pH)$    |        |           |           |
| Critérios estatísticos  | LM  | GLM    | GLMMs GAM | GLMMs GAU |
| R <sup>2</sup> ajustado                                       | 0,2   | 0,16   | 0,52      | 0,51      |
| AIC   | 718   | 590    | 585       | 721       |
| RMSEC (n= 146)  | 24,7  | 23,8   | 19,8      | 19,8      |
| RMSEV (n=54)  | 42,5  | 43,2   | 39,9      | 34,6      |
| Teste de Moran para os resíduos dependência espacial: p=0,147 | 0,024   | 0,063  | 0,563     | 0,796     |
| LATOSSOLO   |   |        |           |           |
| Metodologia   | PTFs  |        |           |           |
| LM  | $C= -39,7+0,18*areia-0,00004*areia^2+0,65*silte+0,0003*silte^2+7,8*pH+1,5*pH^2-0,002*(areia*silte)-0,04*(areia*pH)-0,16*(silte*pH)+0,0004*(areia*silte*pH)$           |        |           |           |
| GLM   | $C=1,6+0,01*areia-0,000003*areia^2+0,05*silte+0,00001*silte^2-0,60*pH+0,20*pH^2-0,0001*(areia*silte)-0,002*(areia*pH)-0,01*(silte*pH)+ 0,00002*(areia*silte*pH)$      |        |           |           |
| GLMMs<br>GAM  | $C=-0,3+0,002*areia-0,000003*areia^2+0,02*silte-0,00001*silte^2+0,92*pH-0,06*pH^2-0,00002*(areia*silte)-0,0002*(areia*pH)-0,002*(silte*pH)+0,000004*(areia*silte*pH)$ |        |           |           |
| GLMMs<br>GAU  | $C=-79,7+0,045*areia-0,00002*areia^2+0,16*silte+0,00005*silte^2+37,6*pH-2,90*pH^2-0,0005*(areia*silte)-0,012*(areia*pH)-0,041*(silte*pH)+ 0,00012*(areia*silte*pH)$   |        |           |           |
| Critérios estatísticos  | LM  | GLM    | GLMMs GAM | GLMMs GAU |
| R <sup>2</sup> ajustado                                       | 0,17  | 0,16   | 0,62      | 0,66      |
| AIC   | 2015  | 1916   | 1809      | 1933      |
| RMSEC (n= 250)  | 13,0  | 13,1   | 8,8       | 8,5       |
| RMSEV (n=101)   | 14,8  | 21,7   | 12,6      | 12,5      |
| Teste de Moran para os resíduos dependência espacial: p<0,001 | <0,001  | <0,001 | 0,001     | 0,227     |

C = carbono; AIC = critério de informação Akaike; RMSEC = raiz quadrada do erro quadrático médio para calibração; RMSEV = raiz quadrada do erro quadrático médio para validação; LM = modelo linear; GLM = modelo linear generalizado; GLMMs GAM = modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados distribuição gama e função de ligação log; GLMMs GAU: modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados distribuição gaussiana e função de ligação identidade.

Fonte: Autor (2023).

Para os GEPP (Tabela 21), nas metodologias LM e GLM, os valores de R<sup>2</sup> ajustado foram 0,2 e 0,16 respectivamente. As metodologias GLMMs apresentaram valores elevados de R<sup>2</sup> ajustado, que foi de 0,52 para GLMMs GAM e 0,51 para GLMMs GAU. As metodologias GLMMs apresentaram os menores valores de RMSE, com 19,8 g kg<sup>-1</sup> para calibração e 39,9 e 34,6 g kg<sup>-1</sup> para validação nos GLMMs GAM e GLMMs GAU, respectivamente. O menor valor de AIC corresponde à metodologia GLM. Para a dependência espacial, calculado a partir de índice de Moran, com valor p=0,147, os resíduos dos modelos desenvolvidos a partir de

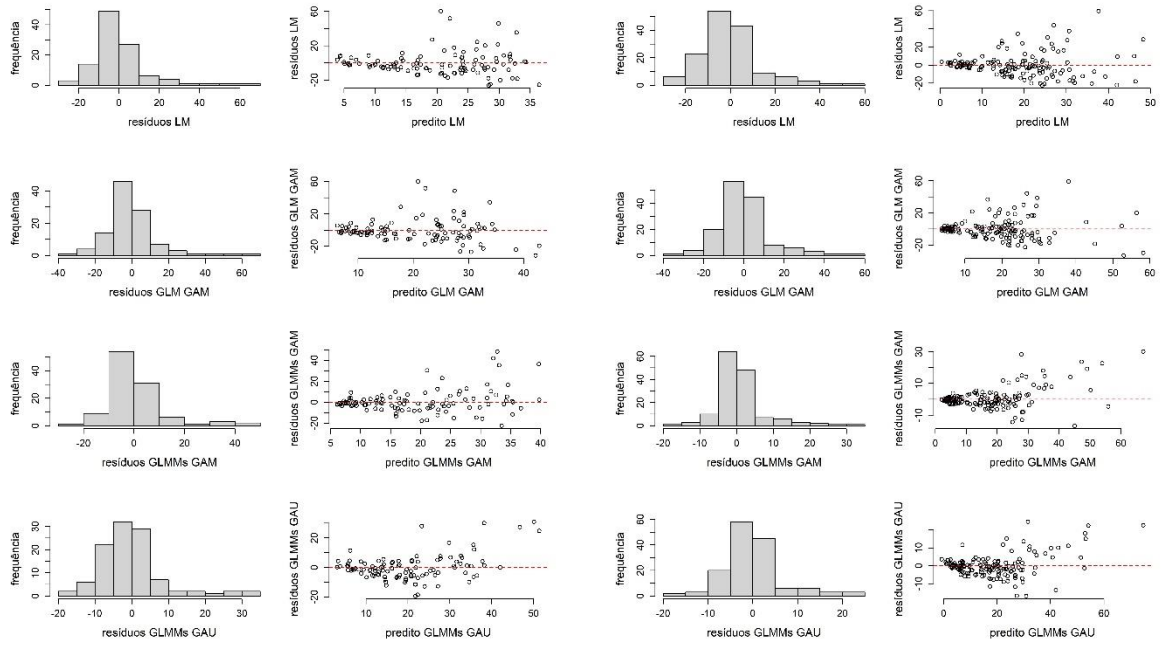
GLMMs apresentaram autocorrelação espacial, com valor de  $p=0,563$  para GLMMs GAM e  $p=0,796$  para GLMMs GAU.

França et al. (2014) desenvolveram PTFs, para estimar teor de C em Gleissolos e Plintossolos em áreas de Campo Limpo Úmido, localizados na região central do Cerrado, a partir de regressão linear e DS e granulometria como variáveis preditoras, e, com 56 amostras para cada tipo de solo, observaram elevados valores de  $R^2$  ajustado de 0,88 e 0,77 e baixos valores de desvio quadrático médio para os solos analisados, respectivamente.

Para os Latossolos (Tabela 21) nas metodologias LM e GLM, os valores de  $R^2$  ajustado foram 0,17 e 0,16 respectivamente. As metodologias GLMMs apresentaram valores elevados de  $R^2$  ajustado, que foi de 0,62 para GLMMs GAM e 0,66 para GLMMs GAU, que apresentou menor RMSEC e RMSEV, que foi de 8,5 e 12,5 g kg<sup>-1</sup>. O menor valor de AIC corresponde à metodologia GLMMs GAM. Para a dependência espacial, calculado a partir de índice de Moran, com valor  $p<0,001$ , os resíduos dos modelos desenvolvidos não apresentaram autocorrelação espacial em todas as metodologias analisadas, com exceção do GLMMs GAM, com valor de  $p=0,001$  e GLMMs GAU com valor de  $p=0,227$ . Os resultados obtidos através da metodologia LM diferem dos valores de  $R^2$ , que variou de 0,40 a 0,44, e de RMSE, que foi de 4,5 g kg<sup>-1</sup> observados por Parizzi (2017), que estimou teor C em Latossolos no Cerrado, profundidades 0-30cm e com dados de granulometria como variáveis de entrada. Na profundidade 30-100 cm, a PTF desenvolvida com variáveis da granulometria teve elevados valores de predição. A autora ainda afirmou que o manejo e uso do solo podem interferir na modelagem superficial de C no solo.

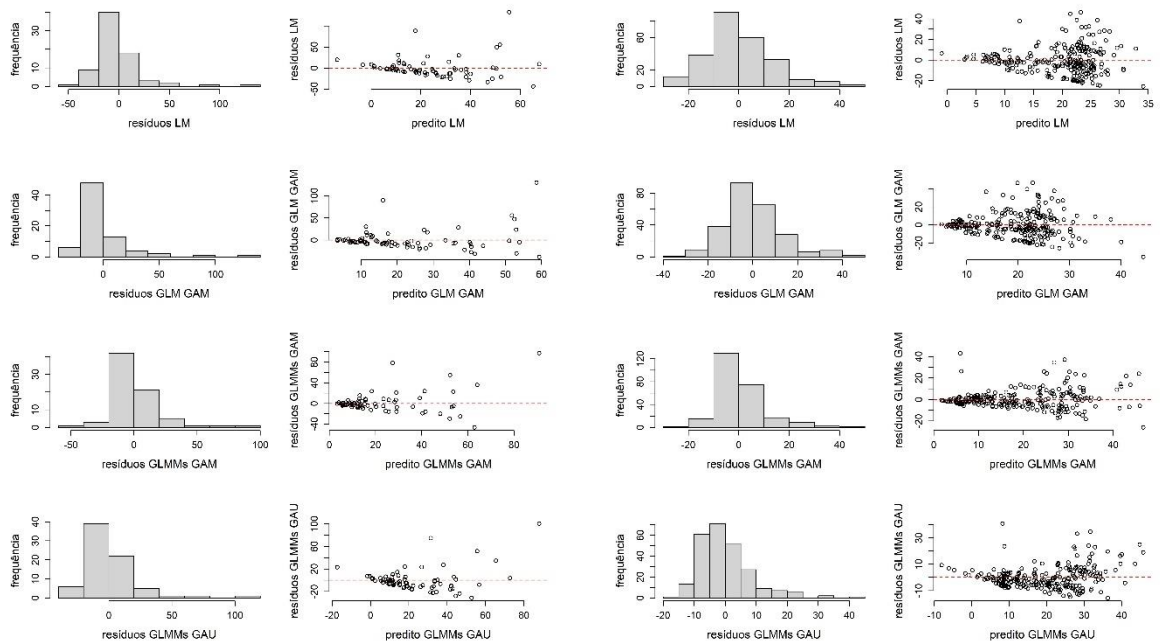
A análise gráfica dos resíduos está representada na Figura 17. Em todos os agrupamentos das amostras de Argissolo, CN, GEEP e Latossolos no Cerrado, os resíduos apresentaram tendência a superestimar os valores de C, em razão da maior frequência dos dados nas classes de valores negativos e a metodologia GLMMs GAM apresentou melhor dispersão residual, configurando como a melhor sugestão para desenvolver PTFs para estimar C em Argissolo, CN e GEEP, com elevados valores de  $R^2$  ajustado e menores valores de RMSE. GLMMs GAU foi a melhor metodologia para estimar teores de C em Latossolos no Cerrado, com elevado valor de  $R^2$  ajustado e menores valores de RMSE para calibração e validação.

Figura 18 - Análise gráfica dos resíduos das PTFs desenvolvidas para o bioma Cerrado agrupados com Argissolo, CN, GEEP e Latossolo



Cerrado x Argissolo

Cerrado x CN



Cerrado x GEEP

Cerrado x Latossolo

Fonte: Autor (2023).

#### 4.2.3.4 Mata Atlântica

As PTFs desenvolvidas para estimar teores de C em diferentes classes de solo do bioma Mata Atlântica foram gerados a partir do agrupamento de diferentes classes de solo nos subgrupos MA1 e MA2. O subgrupo MA 3 não apresentou observações suficiente para agrupar amostras por diferentes classes de solo.

#### 4.2.3.5 Mata Atlântica 1 agrupado com Argissolo, CN, Latossolo e Nitossolo

As desenvolvidas para estimar teores de C em diferentes classes de solo do bioma Mata Atlântica subgrupo MA1 estão descritos na Tabela 22. Para os Argissolos nas metodologias LM e GLM, os valores de  $R^2$  ajustado foram baixos e as variáveis preditoras explicaram apenas 40% e 37% da variância do teor de C, respectivamente. As metodologias GLMMs apresentaram valores elevados de  $R^2$  ajustado, que foi de 0,63 para GLMMs GAM e 0,81 GLMMs GAU. O menor valor de AIC corresponde à metodologia GLM e GLMMs GAM. O menor RMSEC corresponde a metodologia GLMMs GAU, que foi de 5,7 g kg<sup>-1</sup>. Para a dependência espacial, calculado a partir de índice de Moran, com valor  $p=0,914$ , os resíduos dos modelos desenvolvidos não apresentaram dependência espacial em todas as metodologias analisadas, com maior valor de  $p=0,694$  para GLMMs GAU.

Para os CN (Tabela 22), nas metodologias LM e GLM, os valores de  $R^2$  ajustado foram baixos e as variáveis preditoras explicaram apenas 23 % e 12% da variância do teor de C, respectivamente. As metodologias GLMMs apresentaram valores elevados de  $R^2$  ajustado, que foi de 0,85 para GLMMs GAM e 0,89 para GLMMs GAU. A metodologia GLMMs GAM apresentou o menor valor de AIC. A metodologia GLMMs GAU apresentou o menor valor RMSE, que foi de 15,1 g kg<sup>-1</sup> na calibração e 23,7 g kg<sup>-1</sup> na validação. Para a dependência espacial, calculado a partir de índice de Moran, com valor  $p=0,602$ , apenas na metodologia GLMMs GAM, os resíduos não apresentaram dependência espacial, com valor de  $p=0,098$  e o maior valor refere-se a metodologia GLM, com  $p=0,97$ .

Tabela 22 - Funções de pedotransferência para estimativa de teor de C em diferentes classes de solo do bioma Mata Atlântica subgrupo MA1 desenvolvidas a partir de metodologias de regressões lineares e respectivos desempenhos avaliados pelos critérios estatísticos

| ARGISSOLO   |  |       |           |           |
|---|--|-------|-----------|-----------|
| Metodologia   | PTFs   |       |           |           |
| LM  | $C=141,1-0,10*areia-0,00001*areia^2-0,24*silte+0,0001*silte^2-24,8*pH+1,42*pH^2+0,0003*(areia*silte)+0,01*(areia*pH)+0,03*(silte*pH)-0,00005*(areia*silte*pH)$             |       |           |           |
| GLM   | $C=7,6-0,0006*areia-0,000003*areia^2-0,004*silte+0,000004*silte^2-1,07*pH+0,08*pH^2-0,00001*(areia*silte)+0,00003*(areia*pH)-0,00001*(silte*pH)+0,000002*(areia*silte*pH)$ |       |           |           |
| GLMMs   | $C=4,8+0,002*areia-0,000002*areia^2+0,002*silte+0,00001*silte^2-0,4*pH+0,07*pH^2-0,00002*(areia*silte)-0,001*(areia*pH)-0,001*(silte*pH)+0,000004*(areia*silte*pH)$        |       |           |           |
| GAM   |  |       |           |           |
| GLMMs   | $C=9,8+0,03*areia-0,000012*areia^2+0,06*silte+0,00009*silte^2+3,6*pH+0,8*pH^2+0,00002*(areia*silte)-0,014*(areia*pH)-0,03*(silte*pH)+0,00001*(areia*silte*pH)$             |       |           |           |
| GAU   |  |       |           |           |
| Critérios estatísticos  | LM   | GLM   | GLMMs GAM | GLMMs GAU |
| R <sup>2</sup> ajustado                                       | 0,4  | 0,37  | 0,63      | 0,81      |
| AIC   | 986  | 932   | 932       | 979       |
| RMSEC (n=130)   | 9,8  | 9,9   | 7,8       | 5,7       |
| RMSEV (n=59)  | 9,4  | 10,3  | 10,6      | 10,4      |
| Teste de Moran para os resíduos dependência espacial: p=0,914 | 0,003  | 0,085 | 0,315     | 0,694     |
| CN  |  |       |           |           |
| Metodologia   | PTFs   |       |           |           |
| LM  | $C=169,4+0,39*areia+0,00002*areia^2+0,72*silte+0,0007*silte^2+28,7*pH+5,1*pH^2-0,0005*(areia*silte)-0,099*(areia*pH)-0,24*(silte*pH)+0,0002*(areia*silte*pH)$              |       |           |           |
| GLM   | $C=-1,0+0,01*areia-0,000002*areia^2+0,02*silte+0,000002*silte^2+0,6*pH+0,05*pH^2-0,00002*(areia*silte)-0,002*(areia*pH)-0,003*(silte*pH)+0,000003*(areia*silte*pH)$        |       |           |           |
| GLMMs   | $C=0,7+0,01*areia-0,000003*areia^2+0,01*silte-0,000004*silte^2+0,08*pH+0,04*pH^2-0,00001*(areia*silte)-0,001*(areia*pH)-0,001*(silte*pH)+0,000001*(areia*silte*pH)$        |       |           |           |
| GAM   |  |       |           |           |
| GLMMs   | $C=-165,8+0,34*areia-0,00007*areia^2+0,7*silte-0,00008*silte^2+16,3*pH+2,19*pH^2-0,0001*(areia*silte)-0,046*(areia*pH)-0,10*(silte*pH)-0,000001*(areia*silte*pH)$          |       |           |           |
| GAU   |  |       |           |           |
| Critérios estatísticos  | LM   | GLM   | GLMMs GAM | GLMMs GAU |
| R <sup>2</sup> ajustado                                       | 0,23   | 0,12  | 0,85      | 0,89      |
| AIC   | 1805   | 1511  | 1497      | 1746      |
| RMSEC (n=176)   | 25   | 16,8  | 16,3      | 15,1      |
| RMSEV (n=48)  | 38,1   | 37,7  | 24,9      | 23,7      |
| Teste de Moran para os resíduos dependência espacial: p=0,602 | 0,144  | 0,97  | 0,098     | 0,744     |
| LATOSSOLO   |  |       |           |           |
| Metodologia   | PTFs   |       |           |           |
| LM  | $C=71,5+0,001*areia+0,19*silte-0,0001*silte^2-17,6*pH+1,8*pH^2-0,00003*(areia*silte)-0,0037*(areia*pH)-0,03*(silte*pH)+0,000002*(areia*silte*pH)$                          |       |           |           |
| GLM   | $C=6,5-0,001*areia+0,004*silte-0,00001*silte^2-1,1*pH+0,1*pH^2-0,000001*(areia*silte)-0,0001*(areia*pH)-0,0003*(silte*pH)+0,000001*(areia*silte*pH)$                       |       |           |           |
| GLMMs   | $C=5,0+0,00001*areia+0,005*silte-0,000003*silte^2-0,7*pH+0,08*pH^2-0,000006*(areia*silte)-0,0002*(areia*pH)-0,0009*(silte*pH)+0,000001*(areia*silte*pH)$                   |       |           |           |
| GAM   |  |       |           |           |
| GLMMs   | $C=42,8+0,0025*areia+0,17*silte-0,00004*silte^2-5,6*pH+0,7*pH^2-0,0004*(areia*silte)-0,0045*(areia*pH)-0,027*(silte*pH)+0,00008*(areia*silte*pH)$                          |       |           |           |
| GAU   |  |       |           |           |



| Critérios estatísticos  | LM    | GLM         | GLMMs GAM | GLMMs GAU |
|---|-------|-------------|-----------|-----------|
| R <sup>2</sup> ajustado                                       | 0,13  | 0,11        | 0,53      | 0,57      |
| AIC   | 1614  | 1554        | 1518      | 1595      |
| RMSEC (n=193)   | 15    | 15,1        | 11,2      | 10,9      |
| RMSEV (n= 63)   | 16,7  | <u>50,5</u> | 21,1      | 14,2      |
| Teste de Moran para os resíduos dependência espacial: p<0,001 | 0,035 | 0,038       | 0,232     | 0,361     |

NITOSSOLO

| Metodologia | PTFs   |
|-------------|--|
| LM          | $C=282,4-0,5*areia-0,47*silte+0,00004*silte^2-59,3*pH+3,3*pH^2+0,002*(areia*silte)+0,083*(areia*pH)+0,07*(silte*pH)-0,0004*(areia*silte*pH)$         |
| GLM         | $C= 11,5-0,02*areia-0,019*silte+0,000003*silte^2-1,8*pH+0,1*pH^2+0,0001*(areia*silte)+0,004*(areia*pH)+0,003*(silte*pH)-0,00002*(areia*silte*pH)$    |
| GLMMs GAM   | $C= 6,6-0,01*areia-0,009*silte-0,000001*silte^2-0,7*pH+0,04*pH^2+0,00004*(areia*silte)+0,0017*(areia*pH)+0,0015*(silte*pH)-0,00001*(areia*silte*pH)$ |
| GLMMs GAU   | $C=183,1-0,35*areia-0,30*silte-0,000003*silte^2-32,4*pH+1,63*pH^2+0,001*(areia*silte)+0,06*(areia*pH)+0,04*(silte*pH)-0,0002*(areia*silte*pH)$       |

| Critérios estatísticos  | LM    | GLM   | GLMMs GAM | GLMMs GAU |
|---|-------|-------|-----------|-----------|
| R <sup>2</sup> ajustado                                       | 0,29  | 0,28  | 0,77      | 0,7       |
| AIC   | 458   | 439   | 436       | 461       |
| RMSEC (n=55)  | 12,7  | 13,3  | 7,6       | 8,6       |
| RMSEV (n=19)  | 11,4  | 12,3  | 11        | 10,5      |
| Teste de Moran para os resíduos dependência espacial: p=0,025 | 0,913 | 0,804 | 0,881     | 0,958     |

n = número de amostras; C = carbono; AIC = critério de informação Akaike; RMSEC = raiz quadrada do erro quadrático médio para calibração; RMSEV = raiz quadrada do erro quadrático médio para validação; LM = modelo linear; GLM = modelo linear generalizado; GLMMs GAM = modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados distribuição gama e função de ligação log; GLMMs GAU: modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados distribuição gaussiana e função de ligação identidade.

Fonte: Autor (2023).

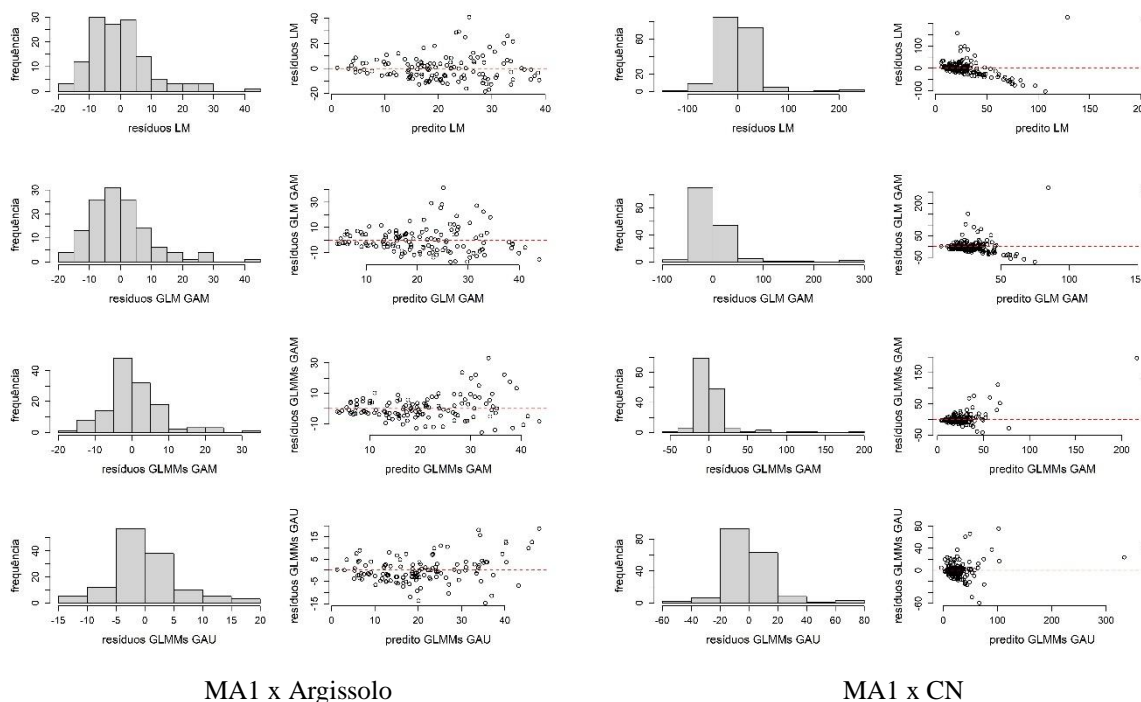
Para os Latossolos (Tabela 22) nas metodologias LM e GLM, os valores de R<sup>2</sup> ajustado foram 0,13 e 0,11 respectivamente. As metodologias GLMMs apresentaram valores elevados de R<sup>2</sup> ajustado, que foi de 0,53 para GLMMs GAM e 0,57 para GLMMs GAU, que apresentou menor RMSEC e RMSEV, que foi de 10,9 e 14,2 g kg<sup>-1</sup>, respectivamente. O menor valor de AIC corresponde à metodologia GLMMs GAM. Para a dependência espacial, calculado a partir de índice de Moran, com valor p<0,001, os resíduos dos modelos desenvolvidos apresentaram autocorrelação espacial em todas as metodologias analisadas, com maior valor para GLMMs GAU, com p=0,361.

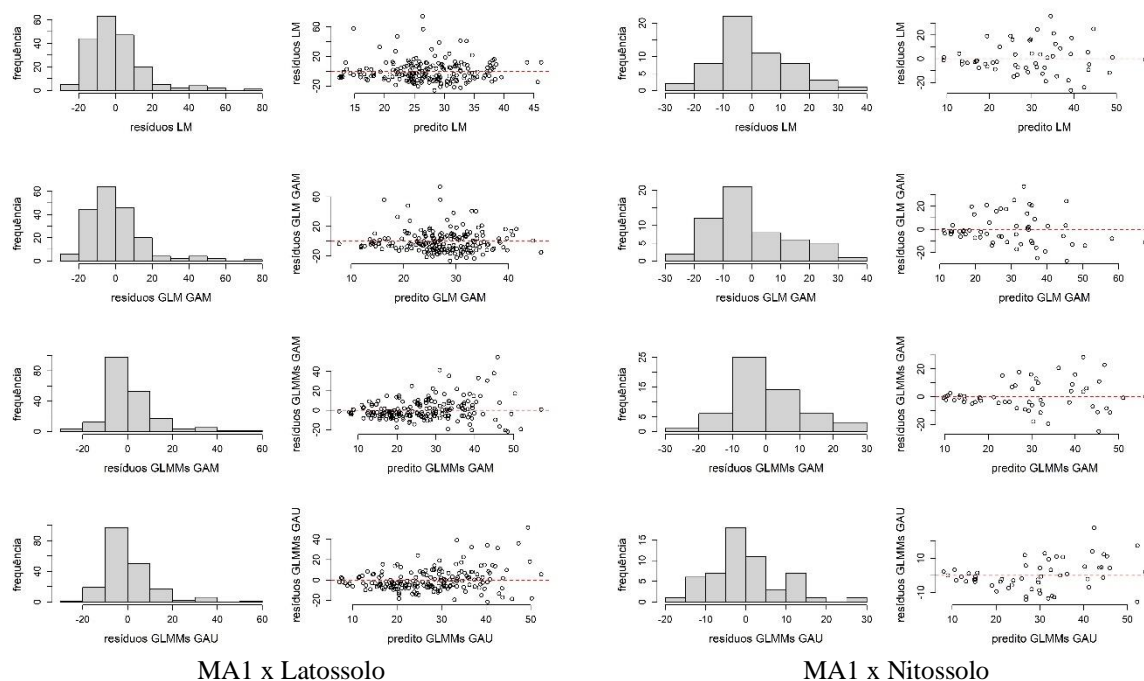
Para os Nitossolos (Tabela 22) nas metodologias LM e GLM, os valores de R<sup>2</sup> ajustado foram 0,29 e 0,28 respectivamente. As metodologias GLMMs apresentaram valores elevados

de  $R^2$  ajustado, que foi de 0,77 para GLMMs GAM e 0,7 para GLMMs GAU. A metodologia GLMMs GAM apresentou o menor valor de RMSEC, que foi de 7,6 g kg<sup>-1</sup> e o menor valor de AIC. Para a dependência espacial, calculado a partir de índice de Moran, com valor  $p=0,025$ , os resíduos dos modelos desenvolvidos apresentaram autocorrelação espacial em todas as metodologias analisadas, com maior valor para GLMMs GAU, com  $p=0,958$ .

A análise gráfica dos resíduos está representada na Figura 18. No agrupamentos das amostras de Argissolo, CN, Latossolo e Nitossolo no subgrupo MA1, os resíduos apresentaram distribuição normal, com exceção do agrupamento MA1 x CN, e tendência a superestimar os valores de C em todas as metodologias, em razão da maior frequência dos dados nas classes de valores negativos e a metodologia GLMMs GAU apresentou melhor dispersão residual, configurando como a melhor sugestão para desenvolver PTFs para estimar C em Argissolo, CN e Latossolo, com elevados valores de  $R^2$  ajustado e menores valores de RMSE. GLMMs GAU foi a melhor metodologia para estimar teores de C em Nitossolos em áreas que compõem o subgrupo MA1, com elevado valor de  $R^2$  ajustado e menores valores de RMSE para calibração e validação.

Figura 19 - Análise gráfica dos resíduos das PTFs desenvolvidas para o bioma Mata Atlântica, subgrupo MA1, agrupados com Argissolo, CN, Latossolo e Nitossolo





Fonte: Autor (2023)

#### 4.2.3.6 Mata Atlântica 2 agrupado com Argissolo e Latossolo

As PTFs desenvolvidos para estimar teores de C em Argissolos e Latossolos do bioma Mata Atlântica subgrupo MA2 estão descritos na Tabela 23. Para os Argissolos nas metodologias LM e GLM, os valores de  $R^2$  ajustado foram baixos e as variáveis preditoras explicaram apenas 11% e 0,9% da variância do teor de C, respectivamente. As metodologias GLMMs apresentaram valores de  $R^2$  ajustado de 0,28 para GLMMs GAM e 0,77 GLMMs GAU. O menor valor de AIC corresponde a metodologia GLM. O menor RMSEC corresponde a metodologia GLMMs GAU, que foi de 6,3 g kg<sup>-1</sup>. Para a dependência espacial, calculado a partir de índice de Moran, com valor  $p=0,113$ , os resíduos dos modelos desenvolvidos apresentaram autocorrelação espacial em todas as metodologias analisadas, com exceção do LM, com valor de  $p=0,03$ .

Para os Latossolos (Tabela 23) nas metodologias LM e GLM, os valores de  $R^2$  ajustado foram 0,08 e 0,07 respectivamente. As metodologias GLMMs apresentaram melhores valores de  $R^2$  ajustado, que foi de 0,47 para GLMMs GAM e 0,53 para GLMMs GAU, que apresentou menor RMSEC e RMSEV, que foi de 8,9 e 13,5 g kg<sup>-1</sup>, respectivamente. O menor valor de AIC corresponde à metodologia GLMMs GAM. Para a dependência espacial, calculado a partir de índice de Moran, com valor  $p<0,001$ , os resíduos dos modelos desenvolvidos pelas

metodologias GLMMs apresentaram autocorrelação espacial, com valor de  $p=0,412$  para GLMMs GAM e  $p=0,568$  para GLMMs GAU.

Tabela 23 - Funções de pedotransferência para estimativa de teor de C em diferentes classes de solo do bioma Mata Atlântica subgrupo MA2 desenvolvidas a partir de metodologias de regressões lineares e respectivos desempenhos avaliados pelos critérios estatísticos

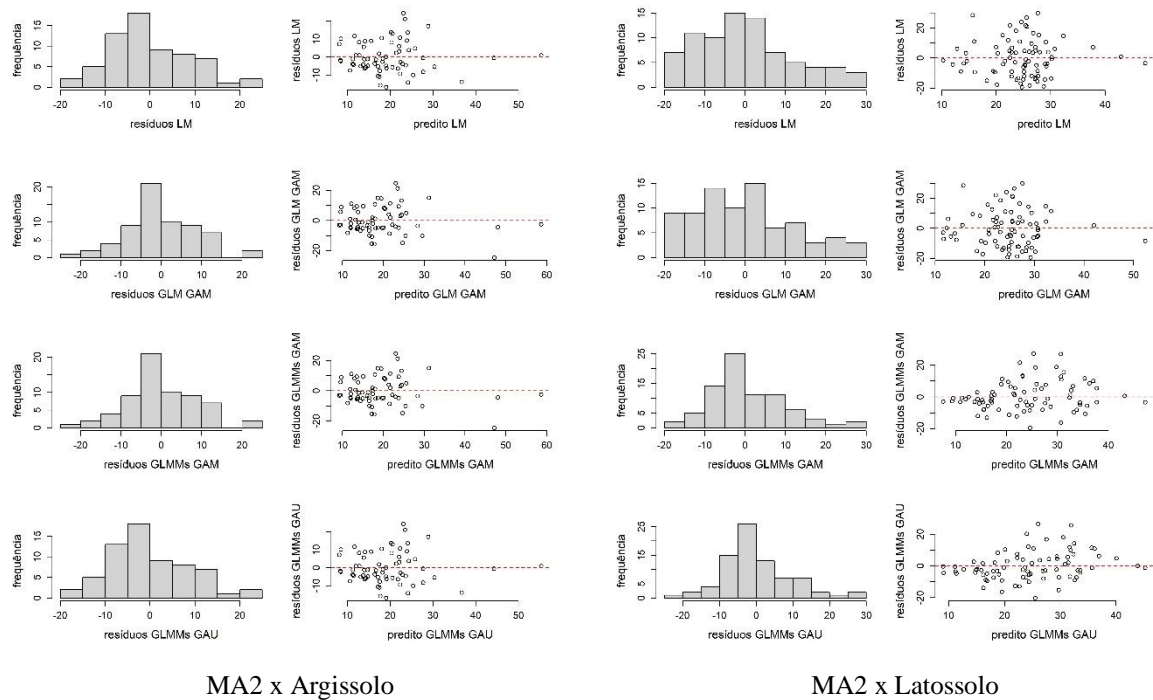
| ARGISSOLO   |  |        |           |           |
|---|--|--------|-----------|-----------|
| Metodologia   | PTFs   |        |           |           |
| LM  | $C = 5,5 + 0,18 * \text{areia} - 0,00003 * \text{areia}^2 + 0,41 * \text{silte} - 0,0001 * \text{silte}^2 - 17,8 * \text{pH} + 3,3 * \text{pH}^2 - 0,0003 * (\text{areia} * \text{silte}) - 0,03 * (\text{areia} * \text{pH}) - 0,05 * (\text{silte} * \text{pH}) + 0,00004 * (\text{areia} * \text{silte} * \text{pH})$         |        |           |           |
| GLM   | $C = 2,7 + 0,01 * \text{areia} - 0,000002 * \text{areia}^2 + 0,02 * \text{silte} - 0,00001 * \text{silte}^2 - 1,1 * \text{pH} + 0,2 * \text{pH}^2 - 0,00001 * (\text{areia} * \text{silte}) - 0,001 * (\text{areia} * \text{pH}) - 0,002 * (\text{silte} * \text{pH}) + 0,000001 * (\text{areia} * \text{silte} * \text{pH})$    |        |           |           |
| GLMMs<br>GAM  | $C = -1,5 + 0,01 * \text{areia} - 0,000002 * \text{areia}^2 + 0,03 * \text{silte} - 0,00001 * \text{silte}^2 - 0,2 * \text{pH} + 0,1 * \text{pH}^2 - 0,00001 * (\text{areia} * \text{silte}) - 0,002 * (\text{areia} * \text{pH}) - 0,003 * (\text{silte} * \text{pH}) + 0,0000004 * (\text{areia} * \text{silte} * \text{pH})$  |        |           |           |
| GLMMs<br>GAU  | $C = -27,8 + 0,18 * \text{areia} - 0,000005 * \text{areia}^2 + 0,49 * \text{silte} - 0,0002 * \text{silte}^2 - 10,3 * \text{pH} + 3,3 * \text{pH}^2 - 0,0002 * (\text{areia} * \text{silte}) - 0,03 * (\text{areia} * \text{pH}) - 0,06 * (\text{silte} * \text{pH}) + 0,000008 * (\text{areia} * \text{silte} * \text{pH})$     |        |           |           |
| Critérios estatísticos  | LM   | GLM    | GLMMs GAM | GLMMs GAU |
| R <sup>2</sup> ajustado   | 0,11   | 0,09   | 0,28      | 0,77      |
| AIC   | 734  | 715    | 716       | 734       |
| RMSEC (n=93)  | 11   | 11,3   | 10        | 6,3       |
| RMSEV (n=28)  | 11,2   | 10,8   | 10,2      | 10,9      |
| Teste de Moran para os resíduos dependência espacial: $p=0,113$   | 0,03   | 0,17   | 0,242     | 0,568     |
| LATOSSOLO   |  |        |           |           |
| Metodologia   | PTFs   |        |           |           |
| LM  | $C = 29,4 + 0,09 * \text{areia} - 0,00007 * \text{areia}^2 + 0,29 * \text{silte} - 0,00016 * \text{silte}^2 - 3,2 * \text{pH} + 0,23 * \text{pH}^2 - 0,0009 * (\text{areia} * \text{silte}) - 0,01 * (\text{areia} * \text{pH}) - 0,05 * (\text{silte} * \text{pH}) + 0,0001 * (\text{areia} * \text{silte} * \text{pH})$        |        |           |           |
| GLM   | $C = 2,1 + 0,01 * \text{areia} - 0,000004 * \text{areia}^2 + 0,02 * \text{silte} - 0,000004 * \text{silte}^2 + 0,2 * \text{pH} + 0,002 * \text{pH}^2 - 0,0001 * (\text{areia} * \text{silte}) - 0,001 * (\text{areia} * \text{pH}) - 0,003 * (\text{silte} * \text{pH}) + 0,00001 * (\text{areia} * \text{silte} * \text{pH})$   |        |           |           |
| GLMMs<br>GAM  | $C = 1,10 + 0,01 * \text{areia} - 0,000004 * \text{areia}^2 + 0,02 * \text{silte} - 0,00001 * \text{silte}^2 + 0,31 * \text{pH} + 0,008 * \text{pH}^2 - 0,00005 * (\text{areia} * \text{silte}) - 0,001 * (\text{areia} * \text{pH}) - 0,003 * (\text{silte} * \text{pH}) + 0,00001 * (\text{areia} * \text{silte} * \text{pH})$ |        |           |           |
| GLMMs<br>GAU  | $C = 11,5 + 0,09 * \text{areia} - 0,0000 * \text{areia}^2 + 0,37 * \text{silte} - 0,00025 * \text{silte}^2 + 1,29 * \text{pH} + 0,19 * \text{pH}^2 - 0,0008 * (\text{areia} * \text{silte}) - 0,011 * (\text{areia} * \text{pH}) - 0,06 * (\text{silte} * \text{pH}) + 0,00014 * (\text{areia} * \text{silte} * \text{pH})$      |        |           |           |
| Critérios estatísticos  | LM   | GLM    | GLMMs GAM | GLMMs GAU |
| R <sup>2</sup> ajustado   | 0,08   | 0,07   | 0,47      | 0,53      |
| AIC   | 649  | 645    | 643       | 647       |
| RMSEC (n=80)  | 12   | 12,1   | 9,3       | 8,9       |
| RMSEV (n=26)  | 15,2   | 15,4   | 13,7      | 13,5      |
| Teste de Moran para os resíduos dependência espacial: $p < 0,001$ | <0,001   | <0,001 | 0,412     | 0,284     |

n = número de amostras; C = carbono; AIC = critério de informação Arkaike; RMSEC = raiz quadrada do erro quadrático médio para calibração; RMSEV = raiz quadrada do erro quadrático médio para validação; LM = modelo linear; GLM = modelo linear generalizado; GLMMs GAM = modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados distribuição gama e função de ligação log; GLMMs GAU: modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados distribuição gaussiana e função de ligação identidade.

Fonte: Autor (2023).

A análise gráfica dos resíduos está representada na Figura 19. Nos agrupamentos das amostras de Argissolos e Latossolos no subgrupo MA2, os resíduos apresentaram distribuição normal e tendência a superestimar os valores de C em todas as metodologias, em razão da maior frequência dos dados nas classes de valores negativos e a metodologia GLMMs GAU apresentou melhor dispersão residual, configurando como a melhor sugestão para desenvolver PTFs para estimar C em Argissolos e Latossolos em áreas que compõem o subgrupo MA2, com elevados valores de  $R^2$  ajustado e menores valores de RMSE para validação e calibração.

Figura 20 - Análise gráfica dos resíduos das PTFs desenvolvidos para o bioma Mata Atlântica, subgrupo MA2, agrupados com Argissolo e Latossolo



MA2 x Argissolo

MA2 x Latossolo

Fonte: Autor (2023).

#### 4.2.4 Função de pedotransferência para classe e uso do solo nos biomas brasileiros (Bioma x Solo x Uso)

Dentre todas as possibilidades de agrupamentos de amostras de biomas, classe de solo e uso do solo, os agrupamentos Cerrado e Pastagem, combinados com Argissolo, CN, GEPP e Latossolo foram os que apresentaram número mínimo de amostras para o desenvolvimento das PTFs.

#### 4.2.4.1 Pastagem no Cerrado agrupado com Argissolo, CN, GEPP e Latossolo

As PTFs desenvolvidas para estimar teores de C em Argissolo, CN, GEPP e Latossolo sob pastagem no bioma Cerrado estão descritos na Tabela 24. Para os Argissolos nas metodologias LM e GLM, os valores de  $R^2$  ajustado foram baixos e as variáveis preditoras explicaram apenas 28% e 24% da variância do teor de C, respectivamente. As metodologias GLMMs apresentaram valores de  $R^2$  ajustado de 0,51 para GLMMs GAM e 0,76 para GLMMs GAU. O menor valor de AIC corresponde a metodologia GLM. O menor RMSEC corresponde a metodologia GLMMs GAU, que foi de 8,2 g kg<sup>-1</sup>. Para a dependência espacial, calculado a partir de índice de Moran, com valor  $p < 0,001$ , os resíduos dos modelos desenvolvidos apresentaram autocorrelação espacial em todas as metodologias analisadas, com maior valor para GLMMs, com valor de  $p = 0,255$ .

Para os CN (Tabela 24) nas metodologias LM e GLM, os valores de  $R^2$  ajustado foram baixos e as variáveis preditoras explicaram apenas 35% e 31% da variância do teor de C, respectivamente. As metodologias GLMMs apresentaram valores de  $R^2$  ajustado de 0,87 para GLMMs GAM e 0,89 para GLMMs GAU. O menor valor de AIC corresponde a metodologia GLMMs GAM. O menor RMSEC corresponde a metodologia GLMMs GAU, que foi de 5,6 g kg<sup>-1</sup>. Para a dependência espacial, calculado a partir de índice de Moran, com valor  $p = 0,743$ , os resíduos dos modelos desenvolvidos apresentaram autocorrelação espacial nas metodologias GLMMs GAM e GLMMs GAU, com maior valor de  $p = 0,96$  para GLMMs GAU.

Para os GEPP (tabela 24) nas metodologias LM e GLM, os valores de  $R^2$  ajustado foram baixos e as variáveis preditoras explicaram apenas 25% e 20% da variância do teor de C, respectivamente. As metodologias GLMMs apresentaram valores de  $R^2$  ajustado de 0,43 para GLMMs GAM e 0,27 para GLMMs GAU. O menor valor de AIC corresponde a metodologia GLM. O menor RMSEC corresponde a metodologia GLMMs GAM, que foi de 17,5 g kg<sup>-1</sup>. Para a dependência espacial, calculado a partir de índice de Moran, com valor  $p = 0,510$ , os resíduos dos modelos desenvolvidos apresentaram autocorrelação espacial nas metodologias GLMMs GAM e GLMMs GAU, com maior valor de  $p = 0,799$  para GLMMs GAM.

Para os Latossolos (Tabela 24), nas metodologias LM e GLM, os valores de  $R^2$  ajustado foram baixos e as variáveis preditoras explicaram apenas 20% e 19% da variância do teor de C, respectivamente. As metodologias GLMMs apresentaram valores de  $R^2$  ajustado de 0,69 para GLMMs GAM e 0,70 para GLMMs GAU. O menor valor de AIC corresponde a metodologia GLMMs GAM. Os menores RMSEC e RMSEV corresponde a metodologia GLMMs GAU,

que foi de 7,3 e 12,8 g kg<sup>-1</sup> respectivamente. Para a dependência espacial, calculado a partir de índice de Moran, com valor p<0,001, os resíduos dos modelos desenvolvidos apresentaram autocorrelação espacial nas metodologias GLMMs GAM e GLMMs GAU, com maior valor de p=0,943 para GLMMs GAU. Estes resultados diferem de Parizzi (2017), que observou R<sup>2</sup> ajustado entre 0,65 e 0,70 para LM e baixos valores de RMSE na estimativa de teor de C em Latossolo sob pastagem plantada no Cerrado, utilizando dados de granulometria e CTC como variáveis preditoras dos modelos.

Tabela 24 - Funções de pedotransferência para estimativa de teor de C em Pastagem sob diferentes classes de solo do bioma Cerrado desenvolvidas a partir de metodologias de regressões lineares e respectivos desempenhos avaliados pelos critérios estatísticos

| ARGISSOLO   |  |       |           |           |
|---|--|-------|-----------|-----------|
| Metodologia   | PTFs   |       |           |           |
| LM  | $C=569,9-0,6*areia+0,000004*areia^2-1,55*silte+0,0004*silte^2-89,3*pH+1,01*pH^2+0,0009*(areia*silte)+0,09*(areia*pH)+0,21*(silte*pH)-0,0001*(areia*silte*pH)$              |       |           |           |
| GLM   | $C=24,1-0,02*areia+-0,000003*areia^2-0,05*silte+0,00001*silte^2-4,0*pH+0,10*pH^2+0,00002*(areia*silte)+0,003*(areia*pH)+0,008*(silte*pH)-0,000003*(areia*silte*pH)$        |       |           |           |
| GLMMs<br>GAM  | $C=22,5-0,02*areia-0,000002*areia^2-0,05*silte+0,00001*silte^2-3,9*pH+0,13*pH^2+0,00003*(areia*silte)+0,003*(areia*pH)+0,007*(silte*pH)-0,000005*(areia*silte*pH)$         |       |           |           |
| GLMMs<br>GAU  | $C=686,9-0,81*areia+0,00007*areia^2-1,94*silte+0,00063*silte^2-104,2*pH+1,46*pH^2+0,001*areia*silte+0,105*(areia*pH)+0,24*(silte*pH)-0,0002*(areia*silte*pH)$              |       |           |           |
| Critérios estatísticos  | LM   | GLM   | GLMMs GAM | GLMMs GAU |
| R <sup>2</sup> ajustado                                       | 0,28   | 0,24  | 0,51      | 0,76      |
| AIC   | 628  | 561   | 562       | 626       |
| RMSEC (n=75)  | 13,6   | 13,8  | 11,4      | 8,2       |
| RMSEV (n=28)  | 15,5   | 21,7  | 14,9      | 18,5      |
| Teste de Moran para os resíduos dependência espacial: p<0,001 | 0,026  | 0,188 | 0,169     | 0,255     |
| CN  |  |       |           |           |
| Metodologia   | PTFs   |       |           |           |
| LM  | $C=199,3+0,21*areia-0,00002*areia^2+0,31*silte-0,00003*silte^2+51,3*pH-1,36*pH^2-0,0004*(areia*silte)-0,043*(areia*pH)-0,06*(silte*pH)+0,0001*(areia*silte*pH)$            |       |           |           |
| GLM   | $C=1,5+0,002*areia-0,000002*areia^2+0,0003*silte-0,000003*silte^2+0,2*pH+0,02*pH^2+0,000004*(areia*silte)-0,0003*(areia*pH)-0,0001*(silte*pH)-0,00000003*(areia*silte*pH)$ |       |           |           |
| GLMMs<br>GAM  | $C=0,1+0,003*areia-0,000003*areia^2+0,006*silte-0,000001*silte^2+0,6*pH+0,02*pH^2+0,000003*(areia*silte)-0,0003*(areia*pH)-0,001*(silte*pH)-0,0000004*(areia*silte*pH)$    |       |           |           |
| GLMMs<br>GAU  | $C=-236,3+0,22*areia-0,000009*areia^2+0,46*silte+0,00002*silte^2+61,1*pH-1,06*pH^2-0,0003*(areia*silte)-0,05*(areia*pH)-0,102*(silte*pH)+0,00008*(areia*silte*pH)$         |       |           |           |
| Critérios estatísticos  | LM   | GLM   | GLMMs GAM | GLMMs GAU |
| R <sup>2</sup> ajustado                                       | 0,35   | 0,31  | 0,87      | 0,89      |
| AIC   | 912  | 788   | 758       | 898       |
| RMSEC (n=112)   | 12,7   | 13,0  | 6,6       | 5,6       |
| RMSEV (n=39)  | 12,2   | 11,9  | 10,6      | 11,6      |

|   |  |        |           |           |
|---|--|--------|-----------|-----------|
| Teste de Moran para os resíduos dependência espacial: $p=0,743$ | 0,1  | <0,001 | 0,838     | 0,96      |
| GEPP  |  |        |           |           |
| Metodologia   | PTFs   |        |           |           |
| LM  | $C=10,9-0,06*areia+0,00005*areia^2+0,41*silte-0,00013*silte^2+18,6*pH-1,36*pH^2-0,0006*(areia*silte)-0,01*(areia*pH)-0,07*(silte*pH)+0,00013*(areia*silte*pH)$           |        |           |           |
| GLM   | $C=-0,008+0,0015*areia-0,000001*areia^2+0,016*silte-0,00001*silte^2+1,2*pH-0,09*pH^2-0,00001*(areia*silte)-0,0004*(areia*pH)-0,002*(silte*pH)+0,000002*(areia*silte*pH)$ |        |           |           |
| GLMMs GAM   | $C=-2,47+0,005*areia-0,000006*areia^2+0,02*silte-0,00001*silte^2+1,7*pH-0,15*pH^2-0,00002*(areia*silte)-0,0001*(areia*pH)-0,002*(silte*pH)+0,000002*(areia*silte*pH)$    |        |           |           |
| GLMMs GAU   | $C=10,2-0,05*areia+0,00005*areia^2+0,42*silte-0,00014*silte^2+17,3*pH-1,23*pH^2-0,0006*(areia*silte)-0,0113*(areia*pH)-0,071*(silte*pH)+0,0001*(areia*silte*pH)$         |        |           |           |
| Critérios estatísticos  | LM   | GLM    | GLMMs GAM | GLMMs GAU |
| R <sup>2</sup> ajustado   | 0,25   | 0,20   | 0,43      | 0,27      |
| AIC   | 491  | 418    | 419       | 497       |
| RMSEC (n=53)  | 19,9   | 20,1   | 17,5      | 19,7      |
| RMSEV (n=21)  | 51,2   | 51,1   | 50,8      | 51        |
| Teste de Moran para os resíduos dependência espacial: $p=0,510$ | 0,453  | 0,425  | 0,799     | 0,592     |
| LATOSSOLO   |  |        |           |           |
| Metodologia   | PTFs   |        |           |           |
| LM  | $C=32,8+0,15*areia-0,00002*areia^2+0,65*silte+0,0004*silte^2-19,7*pH+4,61*pH^2-0,0014*(areia*silte)-0,04*(areia*pH)-0,18*(silte*pH)+0,0003*(areia*silte*pH)$             |        |           |           |
| GLM   | $C=3,6+0,01*areia-0,000002*areia^2+0,06*silte+0,00002*silte^2-1,6*pH+0,34*pH^2-0,0001*(areia*silte)-0,002*(areia*pH)-0,01*(silte*pH)+0,00002*(areia*silte*pH)$           |        |           |           |
| GLMMs GAM   | $C=1,3+0,001*areia-0,000002*areia^2+0,01*silte-0,000001*silte^2+0,3*pH+0,01*pH^2-0,00002*(areia*silte)-0,0001*(areia*pH)-0,002*(silte*pH)+0,000003*(areia*silte*pH)$     |        |           |           |
| GLMMs GAU   | $C=-14,7+0,02*areia-0,000004*areia^2+0,15*silte+0,0002*silte^2+13,1*pH-0,31*pH^2-0,0006*(areia*silte)-0,011*(areia*pH)-0,057*(silte*pH)+0,00014*(areia*silte*pH)$        |        |           |           |
| Critérios estatísticos  | LM   | GLM    | GLMMs GAM | GLMMs GAU |
| R <sup>2</sup> ajustado   | 0,20   | 0,19   | 0,69      | 0,70      |
| AIC   | 1385   | 1320   | 1246      | 1324      |
| RMSEC (n=250)   | 11,6   | 11,7   | 7,4       | 7,3       |
| RMSEV (n=101)   | 15,1   | 15,3   | 13,0      | 12,8      |
| Teste de Moran para os resíduos dependência espacial: $p<0,001$ | <0,001   | <0,001 | 0,303     | 0,943     |

N= número de amostras; C = carbono; AIC = critério de informação Akaike; RMSEC = raiz quadrada do erro quadrático médio para calibração; RMSEV = raiz quadrada do erro quadrático médio para validação; LM = modelo linear; GLM = modelo linear generalizado; GLMMs GAM = modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados distribuição gama e função de ligação log; GLMMs GAU: modelo misto linear generalizado com efeitos aleatórios espacialmente correlacionados distribuição gaussiana e função de ligação identidade.

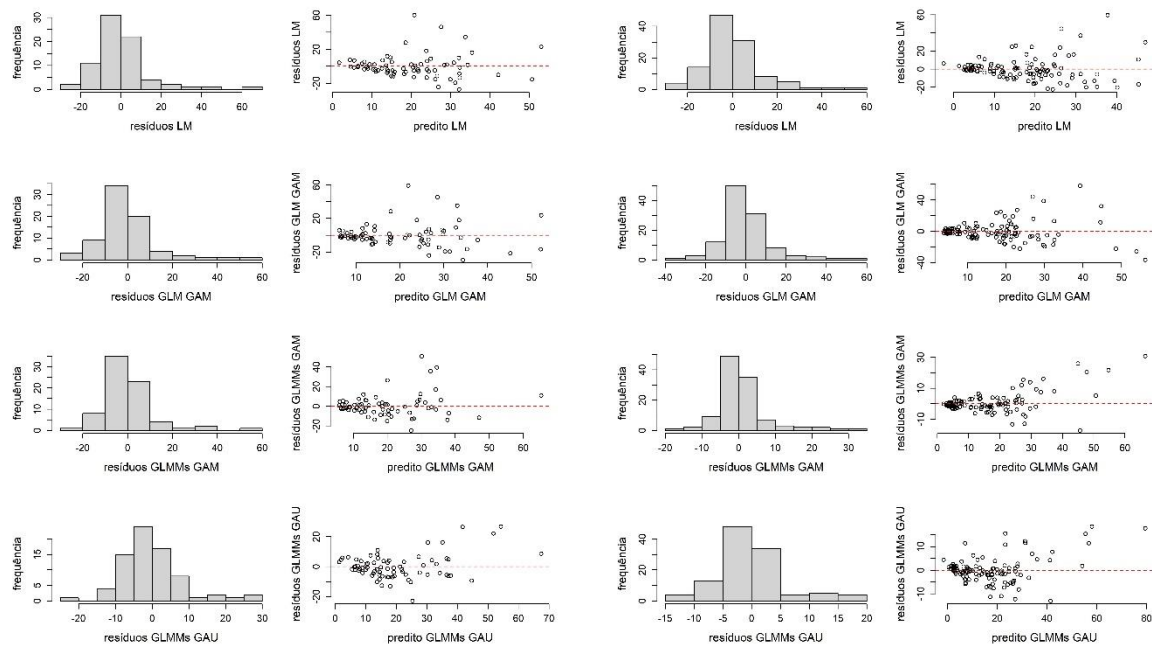
Fonte: Autor (2023).

A análise gráfica dos resíduos está representada na Figura 20. Nos agrupamentos das amostras de Pastagem em Argissolos, CN e Latossolos no Cerrado, os resíduos apresentaram distribuição normal e tendência a superestimar os valores de C em todas as metodologias, com



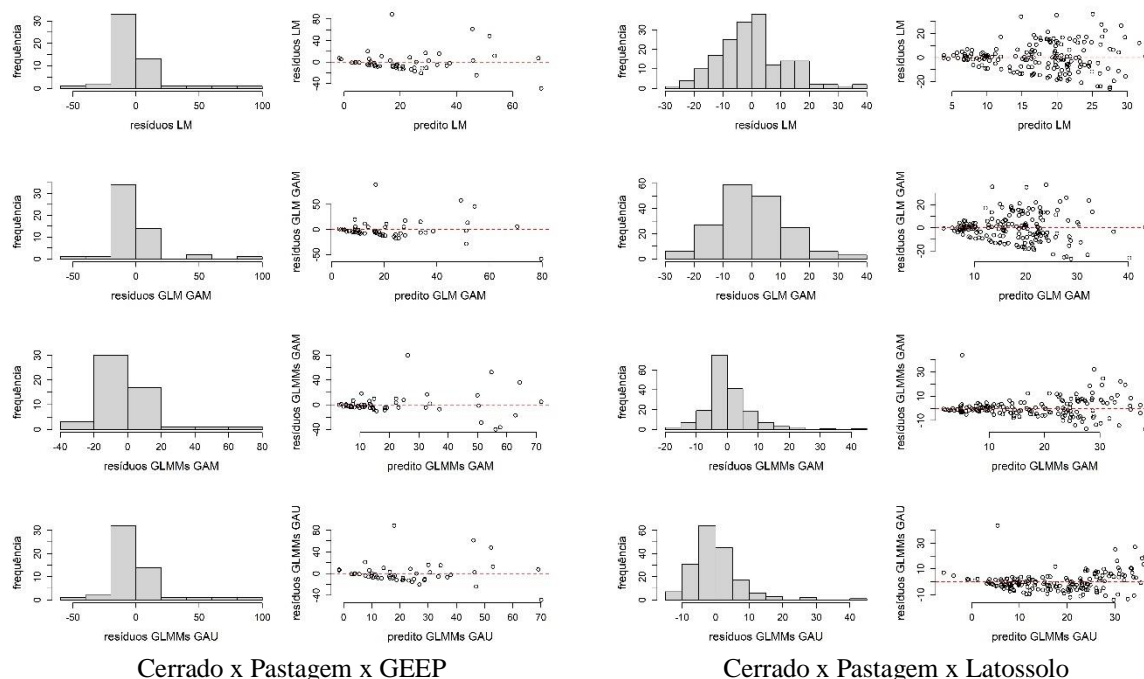
exceção de LM em CN, em razão da maior frequência dos dados nas classes de valores negativos e a metodologia GLMMs GAU apresentou melhor dispersão residual, configurando como a melhor sugestão para desenvolver PTFs para estimar C em Pastagens nos Argissolos, CN e Latossolos no Cerrado, com elevados valores de  $R^2$  ajustado e menores valores de RMSE para validação e calibração. Para os GEEP, GLMMs GAM apresentou melhores resultados, com elevados valores de  $R^2$  ajustado, menores RMSE para validação e calibração e melhor distribuição residual.

Figura 21 - Análise gráfica dos resíduos das PTFs desenvolvidas para Pastagem em Argissolo, CN, GEEPP e Latossolo no bioma Cerrado



Cerrado x Pastagem x Argissolo

Cerrado x Pastagem x CN



Fonte: Autor (2023).

É preciso salientar a dificuldade em encontrar trabalhos na mesma linha de estudo desta pesquisa. PTF para estimar teores C em diferentes classes e uso do solo ainda não é uma metodologia consagrada em estudos que se referem à qualidade do solo ou, mais especificamente, como estratégia de mitigação das emissões de CO<sub>2</sub> pela agropecuária brasileira.

Fez-se o esforço de tentar reunir amostras de solo para validação externa das PTFs, mas a maioria dos estudos fez análises em profundidades diferentes da adotada nesta pesquisa e não foi possível obter um número significativo de amostras para todos os biomas. A validação externa das melhores PTFs confirmaria o bom desempenho dos modelos e sua aplicabilidade em outros estudos.

Observou-se que a estratégia em agrupar as amostras por biomas e, especificamente no bioma Mata Atlântica, onde foram organizados subgrupos com base na latitude, buscando uma similaridade na relação do clima com o teor de C no solo se mostrou eficiente, refletindo em melhores resultados das PTFs de cada bioma, comparado com a PTF geral. O solo é um reservatório provisório do C, resultado do equilíbrio entre adição e redução da MO, influenciado, entre outros fatores, pelo manejo e uso do solo, temperatura e precipitação, que estão diretamente relacionados com a velocidade de decomposição da MO (KUMAR et al., 2022; MURPHY, 2015). Alguns autores afirmaram que dados de climatologia devem ser

inseridos como atributos preditores em PTFs que buscam estimar teores de C no solo (ARRUDA et al., 2017; FALLON; SMITH, 2010; FERNADEZ-UGALD; TOTH, 2017; LEDO et al., 2020; VEREECKEN et al., 2016).

As PTFs que apresentaram resíduos com dependência espacial, em razão da autocorrelação espacial, geraram estimativas mais eficientes. Em situação dessa natureza, pode-se optar por excluir a amostra que anteriormente mostrou dependência ou considerar um método que analise a autocorrelação espacial (LAMICHHANE; KUMAR; WILSON, 2019; LEDO et al., 2020).

## 5 CONCLUSÃO

As PTFs desenvolvidas a partir de metodologias de regressão espacial linear, os GLMMs, apresentaram melhores resultados em todos os biomas brasileiros, quando comparados com LM e GLM. As PTFs para Amazônia, Cerrado, Mata Atlântica subgrupos MA2 e MA3 apresentaram melhores resultados desenvolvidas a partir do GLMMs GAM, resultando em menores valores de RMSE de calibração e validação. As PTFs para Caatinga, Mata Atlântica subgrupo MA1, Pampa e Pantanal apresentaram melhores resultados quando desenvolvidas por GLMMs GAU.

Quando as amostras foram agrupadas por biomas e classes de solo, os resultados das PTFs nas metodologias GLMMs foram superiores aos observados nas PTFs desenvolvidas somente com amostras agrupadas por biomas, com exceção da PTF para Latossolo na Amazônia e Luvissole na Caatinga. Para o Cerrado, dados de Pastagem agrupados com classes de solo apresentaram resultados superiores aos agrupamentos anteriores, com exceção dos GEEP.

Diante do que foi proposto, desenvolver PTFs utilizando dados de textura e pH do solo rendeu, de modo geral, resultados com elevados valores de RMSE e tendências a superestimas as estimativas de C. As dificuldades encontradas referem-se à ausência de estudos mais detalhados para os biomas brasileiros.

## SUGESTÕES

A partir da presente pesquisa, sugerem-se estudos visando o aperfeiçoamento do desenvolvimento de funções de pedotransferência para estimativa de teores de C no solo:

- Organizar uma nova base de dados de solos, com amostras mais detalhadas e recentes, e inserir dados de temperatura e precipitação, em razão da estreita relação dessas variáveis com a quantidade de C disponível no solo, além de serem utilizadas nos principais estudos consultados e, posteriormente, desenvolver PTFs por classes de solo e uso do solo.
- Fazer um estudo comparativo de determinação dos teores de C no solo com dados observados e dados estimados pelos melhores modelos utilizando dados externos, visando aperfeiçoar e validar as PTFs e incentivar o uso dos modelos por outros pesquisadores.
- Testar outras metodologias para desenvolver PTFs, como redes neurais e *random forest*, que apresentaram bons resultados em estudos de estimativas de atributos do solo.
- Desenvolver um aplicativo utilizando as variáveis preditoras, obtidas em análises laboratoriais de baixo custo e as coordenadas geográficas da área de interesse, facilmente disponibilizadas nos dispositivos com acesso à internet e necessárias para os GLMMs.
- Fazer uma análise de custo e viabilidade de determinação de teores de C nas diferentes metodologias citadas nesta pesquisa, visando reforçar a afirmação de que PTF têm o melhor custo-benefício entre todas as técnicas de quantificação de C ou MO.

## REFERÊNCIAS

- ADIAHA, M. S. The role of organic matter in tropical soil productivity. **World Scientific News**, v. 86, p. 1-66, 2017.
- ALTHOFF, T. D. et al. Adaptation of the century model to simulate C and N dynamics of Caatinga dry forest before and after deforestation. **Agriculture, Ecosystems & Environmental**, v. 254, p. 26–34, 2018. Doi: 10.1016/j.agee.2017.11.016.
- AMUNDSON, R. et al. Soil science in the time of climate mitigation. **Biogeochemistry**, v. 161, p. 47 – 58, 2022. Doi:10.1007/s10533-022-00952-6.
- ANGELOPOULOU, T. et al. Remote sensing techniques for soil organic carbon estimation: a review. **Remote Sensing**, v. 676, 2019.
- AKAIKE, H. A new look at the statistical model identification, in IEEE Transactions on Automatic Control, v. 19, p. 716-723, 1974.
- ARRUDA, D. M. et al. Combining climatic and soil properties better predicts covers of Brazilian biomes. **Science Natural**, 2017. Doi: 10.1007/s00114-017-1456-6
- ARRUDA, F.B.; ZULLO JR, J.; OLIVEIRA, J.B. Parâmetros de solo para o cálculo da água disponível com base na textura do solo. **Revista Brasileira de Ciência do Solo**, v.11, p.11-15, 1987.
- BARROS, H. S.; FEARNSTIDE, P. M. Soil carbon is decreasing under “undisturbed” Amazonian forest. **Soil Science Society of America Journal**, v. 83, 1779–1785, 2019. Doi.10.2136/sssaj2018.12.0489.
- BERNOUX, M et al. Bulk densities of Brazilian Amazon soils related to other soil properties. **Soil Science Society of America Journal**, v.62, n.3, p.743-749, 1998.
- BIAN, Z. et al. Applying statistical methods to map soil organic carbon of agricultural lands in northeastern coastal areas of China. **Archives of Agronomy and Soil Science**, v. 66, p. 532-544, 2020. Doi: 10.1080/03650340.2019.1626983.
- BRASIL. Ministério Da Agricultura, Pecuária e Abastecimento (MAPA). Plano setorial de mitigação e de adaptação às mudanças climáticas para a consolidação de uma economia de baixa emissão de carbono na agricultura: plano ABC (Agricultura de Baixa Emissão de Carbono). Brasília: MAPA/ACS, 2012.
- BENITES, V. et al. Pedotransfer functions for estimating soil bulk density from existing soil survey report in Brazil. **Geoderma**, v. 139, p. 90-97, 2007.
- BENKE, K. K. et al. Development of pedotransfer functions by machine learning for prediction of soil electrical conductivity and organic carbon content. **Geoderma**, v. 366, p. 114210, 2020.
- BORTOLON, E.S.O. et al. Validation of the Century model to estimate the impact of agriculture on soil organic carbon in Southern Brazil. **Geoderma**, v.167-168, p. 156-166, 2011.

BOSCHI R. et al. (2015). Efeito de estimativa de densidade do solo no cálculo de estoque de carbono. **Revista Ciência, Tecnologia & Ambiente**, v. 2, p. 8-16, 2015.

BOUMA, J. Using soil survey data for quantitative land evaluation. **Advances in Soil Science**, v.9, p.177-213, 1989.

CERRI, C. C.; CERRI, C. E. P. Agricultura e aquecimento global. Boletim Informativo. **Sociedade Brasileira de Ciência do Solo**, v. 23, p. 40-44, 2007.

CERRI, C.E.P. et al. Simulating SOC changes in 11 land use change chronosequences from the Brazilian Amazon with RothC and Century models. **Agriculture, Ecosystems & Environment**, v.122, p. 46-57, 2011.

CHADDAD, F. **The economics and organization of Brazilian agriculture: Recent evolution and productivity gains**. Cambridge: Academic Press, 2015.

CHATTERJEE, A. et al. Evaluation of different soil carbon determination methods. **Critical Reviews in Plant Science**, v.28, p.164-178, 2009.

CIDIN, A.C.M. **Estoque de carbono em solos brasileiros e potencial de contribuição de mitigação de emissões de gases de efeito estufa**. Dissertação (Mestrado em Agricultura e Ambiente) - Universidade Federal de São Carlos, Araras, São Paulo, 2016.

COLLEMAN, K.; JENKINSON, D.S. RothC 26.3- a model for the turnover of carbon in soil. In: POLWSON, T.S.; SMITH, P.; SMITH, J.U. (Eds.). **Evaluation of soil organic matter models using existing longterm datasets**. Heidelberg: Springer-Verlag, v.38, p.237-246, 1996.

CONANT, R. T. et al. Measuring and monitoring soil organic carbon stocks in agricultural lands for climate mitigation. **Front Ecol Environ**, v .9, p.169–173, 2011. Doi:10.1890/090153

CONCEIÇÃO, M. et al. **Estudo comparativo de métodos de determinação do teor de matéria orgânica em organossolos**. Rio de Janeiro, Embrapa, 1999. 4p. (Comunicado Técnico, 2).

CONTINI, E.; ARAGÃO, A. O agro brasileiro alimenta 800 milhões de pessoas. Brasília: Embrapa, 2021. Disponível em: <<https://www.gov.br/pt-br/noticias/agricultura-e-pecuaria/2021/03/participacao-brasileira-saltou-de-us-20-6-bilhoes-para-us-100-bilhoes/populacao-alimentada-pelo-brasil.pdf>>. Visualizado em outubro de 2022.

COOPER, M. et al. A national soil profile database for Brazil available to international scientists. **Soil Science Society of America Journal**, v. 69, p. 649-652, 2005.

COSER, T.R. et al. Recuperação de carbono obtida por três métodos em frações da matéria orgânica de latossolo, sob consórcio milho-forrageiras, no Cerrado. **Bioscience Journal**, v.28, n.1, p.91-97, 2012.

COSTA, V. C. et al. Laser induced-breakdown spectroscopy (LIBS): histórico, fundamentos, aplicações e potencialidades. **Química Nova**, v. 42, p. 527-545, 2019.

COTRUFO, M. F.; LAVALLEE, J. M. Soil organic matter formation, persistence, and functioning: A synthesis of current understanding to inform its conservation and regeneration. **Adv. Agron**, v.172, p. 1–66, 2022.

DIGGLE, P. et al. Model-based geostatistics. **Applied Statistics**, v. 47, p. 299-350, 1998.

ENANG, R. K. et al. Pedotransfer function for cation exchange capacity estimation in highly weathered soil of the tropical highlands of NW Cameroon. **Geoderma Regional**, v. 29, e00514, 2022.

ESWARAN, H. et al. Organic carbon in soils of the world. **Soil Science of America Journal**, v. 57, p. 192–194, 1993.

FALLOON, P.; SMITH, P. Modelling soil carbon dynamics. In: L. Werner, M.B. Kutsch, and A. Heinemeyer, editors, **Soil carbon dynamics: An integrated methodology**. Cambridge Univ. Press, Cambridge, UK. p. 221–244, 2010.

FAO. Food and Agriculture Organization of the United Nations. Status of the World's Soil Resources. Technical Summary, 2015. Disponível em: <http://www.fao.org/3/a-i5126e.pdf>

FERNANDES, F. et al. **Uso de espectrometria de refletância no infravermelho próximo (NIRS) na análise de carbono de Neossolos do Pantanal**. Comunicado Técnico, ISSN 1981-7231, 2010.

FERNANDEZ-UGALDE, O.; TÓTH, G. Pedotransfer function for predicting organic carbon in subsurface horizons of European soils. **European Journal of Soil Science**, 17. 2017. Doi: 10.1111/ejss.12464.

FERREIRA, E. C, et al. Novel estimation of the humification degree of soil organic matter by laser-induced breakdown spectroscopy. **Spectrochimica Acta Part B: Atomic Spectroscopy**, v. 99, p. 76-81, 2014.

FORMAGGIO, A. R.; SANCHES, I. D. **Sensoriamento remoto em agricultura**. Oficina de Textos, São Paulo, SP. 2017.

FRANÇA, A. M. S. et al. Estimates for carbon stocks in soil under humid grassland areas in the Federal District of Brazil. **Open Journal of Ecology**, v. 4, p. 777-787, 2014. Doi:10.4236/oje.2014.41306.

FRANCAVIGLIA, R. et al. Conservation agriculture and soil organic carbon: Principles, processes, practices and policy options. **Soil Systems**, 2023. Doi.org/10.3390/soilsystems7010017.

GATTO, A. et al. Comparação de métodos de determinação do carbono orgânico em solos cultivados com eucalipto. **Revista Brasileira da Ciência do Solo**, v.33, p.735-740, 2009.

GBUR, E. E. et al. **Analysis of generalized linear mixed models in the agricultural and natural resources sciences**. John Wiley & Sons, 2020.



GOMES, L. V. et al. Modelling and mapping soil organic carbon stocks in Brazil. **Geoderma**, v. 340, p. 337–350, 2019. Doi: 10.1016/j.geoderma.2019.01.007.

GOTWAY, C. A.; STROUP, W. W. A generalized linear model approach to spatial data analysis and prediction. **Journal of Agricultural, Biological, and Environmental Statistics**, v. 2, p. 157-178, 1997.

HARDY, B.; DUFREY, J. E. The resistance of centennial soil charcoal to the “Walkley-Black” oxidation. **Geoderma**, v. 303, p. 37-43, 2017.

HODNETT, M. G.; TOMASELLA, J. Marked differences between van Genuchten soil waterretention parameters for temperate and tropical soils: A new waterretention pedotransfer functions developed for tropical soils. **Geoderma**, v. 108, p. 155-180, 2007.

IPCC. Intergovernmental Panel on Climate Change. IPCC guidelines for national greenhouse gas inventories. Prepared by the National Greenhouse Gas Inventories Programme. Institute for Global Environmental Strategies. Geneva, Switzerland, 2006.

IPCC. Intergovernmental Panel on Climate Change. Penman J., M. Gytarsky, T. Hiraiishi, T. Krug, D. Kruger, R. Pipatti, L. Buendia, K. Miwa, T. Ngara, K. Tanabe and F. Wagner (Eds). Good Practice Guidance for Land Use, Land Use Change and Forestry. IPCC/OECD/IEA/IGES, Hayama, Japan, 2003.

JENSEN, J. R. Sensoriamento Remoto do Ambiente: uma perspectiva em ambientes terrestres. Ed. Parênteses, São José dos Campos, SP. 2009.

JIMENEZ R. R.; LADHA, J. K. Automated elemental analysis: A rapid and reliable but expensive measurement of total carbon and nitrogen in plant and soil samples. **Communications in Soil Science and Plant Analysis**, v. 24, p. 1897-1924, 1993.

JULL, H. et al. Nutrient quantification in fresh and dried mixtures of ryegrass and clover leaves using laser-induced breakdown spectroscopy. **Precision Agriculture**, v. 19, p. 823-839, 2018.

KALEMBASA, S.J.; JENKINSON, D.S. A comparative study of titrimetric and gravimetric methods for the determination of organic carbon in soil. **Journal of the Science of Food and Agriculture**, v.24, n. 9, p.1085–1090, 1973.

KHALEDIAN, Y. et al. Developing global pedotransfer functions to estimate available soil phosphorus. **Science of Total Environmental**, v. 644, p. 1110-1116, 2018.

LAL, R. et al. The role of soil in regulation of climate. **Philosophical Transaction of Royal Society B**, 376: 20210084, 2021. Doi: 10.1098/rstb.2021.0084.

LAMICHHANE S. et al. Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. **Geoderma**, v. 352, p. 395-413, 2019.

LEDO, A. et al. Changes in soil organic carbon under perennial crops. **Global Change Biology**, v. 26, p. 4158–4168, 2020.

LEE, J.; VISCARA-ROSSEL, R. Soil carbon simulation cofounded by different pool initialization. **Nutrient Cycling in Agroecosyst**, v. 116, p. 245–255, 2020.

LEITE, F. F. G. D. et al. Greenhouse gas emissions and carbon sequestration associated with Integrated Crop-Livestock-Forestry (ICLF) systems. **Environmental Reviews**, 2023. Doi:10.1139/er-2022-0095.

LEITE, L.F.C; MENDONÇA, E.S. Modelo Century de dinâmica da matéria orgânica do solo: equações e pressupostos. **Ciência Rural**, v. 33, n. 4, p. 679-686, 2003.

LEITE, L.F.C. et al. Simulating trends in soil organic carbon of an Acrisol under no-tillage and disc-plow systems using the Century model. **Geoderma**, v. 120, n. 3-4, p.283-295, 2004a.

LEITE, L.F.C.; MENDONÇA, E.S.; MACHADO, P.L.O.A. Simulation of organic matter dynamics in na argisol under mineral and organic fertilization with the Century model. **Revista Brasileira de Ciência do Solo**, v.28, n.2, p.347-358, 2004b.

MALLIK, S. et al. Mapping and prediction of soil organic carbon by an advanced geostatistical technique using remote sensing and terrain data. **Geocarto International**, 37, 2198-2214, 2020.

MAIA, S. M. F. et al. Potential of no-till agriculture as a nature-based solution for climate-change mitigation in Brazil. **Soil & Tillage Research**, v. 220, 105368, 2022.

MAPA. Ministério da Agricultura pecuária e Abastecimento. **Projeções do Agronegócio. Brasil 2021/22 a 2031/32**. Brasília, DF. 2022. Disponível em: <https://www.gov.br/agricultura/pt-br/assuntos/noticias/producao-de-graos-deve-crescer-36-8-nos-proximos-dez-anos/PROJEESDOAGRONEGCIO20212022a203132.pdf>. Visualizado em dezembro de 2022.

MARCHÃO, R.L.; BECQUER, T.; BRUNET, D. **Predição dos teores de carbono e nitrogênio do solo utilizando espectroscopia de infravermelho próximo**. Planaltina: Embrapa Cerrados, 2011, 21 p. (Boletim de Pesquisa e Desenvolvimento).

McBRATNEY, A.B. et al. From pedotransfer functions to soil inference systems. **Geoderma**, v.109, p.41-73, 2002.

McCULLAGH, P. **Introduction to Generalized Linear Models**. In: Kotz, S., Johnson, N.L. (eds) *Breakthroughs in Statistics*. Springer Series in Statistics. Springer, New York, NY, 1992.

McNUNN, G. et al. Climate smart agriculture opportunities for mitigation soil greenhouse gas emissions across U. S. Corn-Belt. **Journal of Cleaner Production**, v. 168, 122240, 2020.

MCTI. MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES. Quarta Comunicação Nacional do Brasil à Convenção-Quadro das Nações Unidas sobre Mudança do Clima - Volume IV. Relatório de Referência: Setor Agropecuário, subsetores Solos Manejados, Calagem e Aplicação de Uréia. Brasília: Ministério da Ciência, Tecnologia e Inovação, 148 p, 2020.

MEDEIROS, A. S.; SANTOS, T. C.; MAIA, S. M. F. Effect of long-term and soil depth on soil carbon stocks after conversion from native vegetation to conventional tillage system in Brazil. **Soil & Tillage Research**, 219, 105336, 2022.

MENDONÇA, E.S.; LEITE, L.F.C. Modelagem matemática e simulação da dinâmica da matéria orgânica do solo. In: ROSCOE, R.; MERCANTE, F.M.; SALTON, J.C. (Ed.). **Dinâmica da matéria orgânica do solo em sistemas conservacionistas: modelagem matemática e métodos auxiliares**. Dourados, MS, 2006. Cap.4. p.75-106.

MENEZES, R. S. C. et al. Soil vegetation carbon stocks after land-use changes in a seasonally dry tropical forest. **Geoderma**, v. 390, 114943, 2020.

MINASNY, B.; HARTEMINK, A.E. Predicting soil properties in the tropics. **Earth Science Reviews**, v. 106, p. 52-62, 2011.

MINASNY, B., SULARMAN, Y., MCBRATNEY, A. B. Is Soil carbon disappearing? The dynamic of soil carbon in Java. **Global Change Biology**, 2010.

MINASNY, B. et al. Prediction and digital mapping of soil carbon storage in the Lower Namoi Valley. **Soil Research**, v. 44, n. 3, p. 233-244, 2006.

MURPHY, B. Key soil function properties affected by soil organic matter – evidence from published literature. **Soil Changes Matters**, 2014. Doi:10.1088/1755-1315/25/1/012008.

MWANGO, S. B. et al. The use of pedo-transfer functions for estimating soil organic carbon contents in maize cropland ecosystem in the Coastal Plains of Tanzania. **Catena**, 172, 163-169, 2019.

NAYAK, A. K. et al. Current and emerging methodologies for estimating carbon sequestration in agricultural soils: A review. **Science of the Total Environment**, v. 665, p. 890-912, 2019.

NELDER, J. A.; WEDDERBURN, R. W. Generalized linear models. **Journal of the Royal Statistical Society: Series A (General)**, v. 135, p. 370–384, 1972.

NELSON, D.W.; SOMMERS, L.E. Total carbon, organic carbon, and organic matter. In: BLACK, C.A., ed. Methods of soil analysis. Chemical methods. Madison, **Soil Science of America/American Society of Agronomy**. Part 3, p.961-1010, 1996.

OGLE, S. M. et al. Climate and soil characteristics determine where no-till management can store carbon in soils and mitigate greenhouse gas emissions. **Scientific Reports**, v. 9, 11665, 2019. Doi: 10.1038/s41598-019-47861-7.

OLIVEIRA, P. P. A. **Protocolo para quantificação dos estoques de carbono do solo da rede de pesquisa Pecus** São Carlos: Embrapa Pecuária-Sudeste, 2014. 20p. (Embrapa Pecuária-Sudeste. Documentos, 116).

PACHEPSKY, A. et al. Pedotransfer in soil physics: trends and outlook. A review. **Agrokem Talajtan**, v. 64, p. 339-360, 2015. Doi:10.1556/0088.2015.64.2.3.

PALLADINO, M. et al. Developing pedotransfer functions for predicting soil bulk density in Campania. **Geoderma**, v. 412, p. 115726, 2022.

PARIZZI, T. N. T. **Modelagem para predição de carbono orgânico em áreas naturais e antropizadas do cerrado brasileiro**. Tese (Doutorado em Engenharia Agrícola) – Universidade Estadual de Campinas, Campinas, 2017.

PARTON, W.J. et al. Analysis of factors controlling soil organic matter levels in great plains grasslands. **Soil Science Society America Journal**, v.51, p.1173-1179, 1987.

PARTON, W.J.; STEWART, J.W. B.; COLE, C.V. Dynamics of C, N, P and S in grassland soils: a model. *Biogeochemistry*, Springer, v.5, n.1, p.109–131, 1988.

PASQUINI, C. et al. Laser Induced Breakdown Spectroscopy. **Journal of the Brazilian Chemical Society**, v.18, n.3, p.463-512, 2007.

PAUSTIAN, K. et al. Quantifying carbon for agricultural soil management: from the current status toward a global soil information system. **Carbon Management**, v. 10, pp. 567-587, 2019.

PELLISSARI, R. et al **Possibilities for carbon capture, utilization, and storage in brazil**. in: Peyerl, D., Relva, S., Da Silva, V. (eds) *Energy Transition in Brazil. The Latin American Studies Book Series*. Springer, Cham, 2023. Doi:10.1007/978-3-031-21033-4\_9.

QUAGGIO, J.A.; RAIJ, B.V. Comparação de métodos rápidos para a determinação da matéria orgânica em solos. **Revista Brasileira da Ciência do Solo**, v. 3, p.184-187, 1979.

QUIN, L. et al. Evaluation of pedotransfer functions for predicting particle density of soils with low organic matter contents. **Geoderma**, 416, 2022. Doi: 10.1016/j.geoderma.2022.115812.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: The R Foundation for Statistical Computing, 2014.

RHEINHEIMER, D.S. et al. Comparação de métodos de determinação de carbono orgânico total no solo. **Revista Brasileira da Ciência do Solo**, v.32, p.435-440, 2008.

RODRIGUES, M.J.M. et al. Espectropia no infravermelho próximo para a quantificação de carbono em solos da bacia do Acre. **Biota Amazônia**, v.6, n.1, p.119-124, 2016.

ROUSSET, F.; FERDY, J. B. Testing environmental and genetic effects in the presence of spatial autocorrelation. **Ecography**, v. 37, p. 781-790, 2014. Doi:10.1111/ecog.00566.

RUDIYANTO et al Pedotransfer functions for estimating soil hydraulic properties from saturation to drynes. **Geoderma**, v. 403, p. 115-194, 2021.

SHAMRIKOVA, E. V. et al. Which method to choose for measurement of oranic and inorganic carbon content in carbonate-rich soils? Advantages and disadvantages of dry and wet chemistry. **Catena**, 228, 107151, 2023. Doi:10.1016/j.catena.2023.107151.

SATO, J. H. et al. Methods of soil carbon determination in Brazilian savannah soils. **Scientia Agricola**, v. 71, p. 302-308, 2014.

SANTOS, H. G. dos et al. *Sistemas Brasileiro de Classificação de Solos*. 3 ed. Ver. Ampl. Brasília, DF: Embrapa, 353p. 2013.

SANTOS, H. G. dos et al. *O novo Mapa de Solos do Brasil: Legenda atualizada*. Embrapa Solos, 2011. (Documento n. 130).

SCHULTE, E.E.; HOPKINS, B.G. Estimation of soil organic matter by weight loss-on-ignition. In: MAGDOFF, F.R.; TABATABAI, M.A. & HANLON Jr., E.A., eds. *Soil organic matter: Analysis and interpretation*. Madison, **American Society of Agronomy**, p.21-31, 1996.

SENESI, G. S.; SENESI, N. Laser-induced breakdown spectroscopy (LIBS) to measure quantitatively soil carbon with emphasis on soil organic carbon. A review. **Analytica Chimica Acta**, v. 938, p. 7-17, 2016.

SEEG. Sistema de Estimativas de Emissões e Remoções de Gases de Efeito Estufa. Análise das emissões brasileiras de gases de efeito estufa e suas implicações para as metas climáticas do Brasil: 1970-2020. 2021. Disponível em: [https://seeg-br.s3.amazonaws.com/Documentos%20Analiticos/SEEG\\_9/OC\\_03\\_relatorio\\_2021\\_FINAL.pdf](https://seeg-br.s3.amazonaws.com/Documentos%20Analiticos/SEEG_9/OC_03_relatorio_2021_FINAL.pdf). Acesso em novembro de 2022.

SEGNINI, A. et al. Estudo comparativo de métodos para a determinação da concentração de carbono em solos com altos teores de Fe (Latosolos). **Química Nova**, v.31, p.94-97, 2008.

SHAMRIKOVA, E. V. et al. Which method to choose for measurement of organic and inorganic carbon content in carbonate-rich soils? Advantages and disadvantages of dry and wet chemistry. **Catena**, v. 228, 107151, 2023.

SINGH, S. et al. Carbon offset potential of biochar based straw management under rice- wheat system along Indo-Gangetic Plains of India. **Science of Total Environment**, 2023.

SLEUTEL, S. et al. Quantification of organic carbon in soils: a comparison of methodologies and assessment of the carbon content of organic matter. **Communications in Soil Science and Plant Analysis**, v.38, p.2647-2657, 2007.

SMITH, P. et al. (2020). How to measure, report and verify soil carbon change to realize the potential of soil carbon sequestration for atmospheric greenhouse gas removal. **Global Change Biology**, v. 26, p. 219-241, 2020. Doi: 10.1111/gcb.14815.

SOMARATHNA, P. D. S. N. et al. (2017). More data or a better model? Figuring out what matters most for the spatial prediction of soil carbon. **Soil Science Society of America Journal**, 81, p. 1413-1426, 2017. Doi:10.2136/sssaj2016.11.0376.

SONG, X. D. et al. Can pedotransfer function based on environmental variables improve soil total nutrient mapping at a regional scale? **Soil & Tillage Research**, v. 202, 104672, 2020.

STENIO, K. et al. Carbon quantification in soils with different textures using laser-induced breakdown spectroscopy: spectral interference correction and use of a 3D plane model. **Analytical Methods**, v. 14, n. 42, p. 4219-4229, 2022.

SUN, W. et al. Climate drives global soil carbon sequestration and crop yield changes under conservation agriculture. **Global Change Biology**, v. 26, p.3325–3335, 2020.

TEDESCO, M.J. et al. **Análises de solo, plantas e outros materiais**. 2.ed. Porto Alegre: UFRGS, 1995. 174p.

TOMASELLA, J. et al. Comparison of two approximation techniques to develop pedotransfer functions for Brazilian soil. **Soil Science Society of America Journal**, v.67, p.1085-1092, 2003.

TOMASELLA, J.; HODNETT, M.G. Estimating soil water retention characteristics from limited data in Brazilian Amazonia. **Soil Science**, v.163, n.3, p.190-202, 1998.

TÓTH, B. et al. (2015). New generation of hydraulic pedotransfer functions for Europe. **Europe Journal of Soil Science**, v. 66, p. 226–238, 2015. Doi:10.1111/ejss.12192

VALANI, G.P. et al. Soil quality assessments in integrated crop–livestock–forest systems: A review. **Soil Use and Management**, v. 37, p. 22– 36, 2021. Doi: 10.1111/sum.12667.

VAN LOOY, K. et al. Pedotransfer functions in Earth system science: challenges and perspectives. **Reviews of Geophysics**, v. 55, p.1199-1256, 2017. Doi:10.1002/2017RG000581.

VASCONCELOS, J. C. S. et al. A new regression model for bimodal data and applications in agriculture. **Journal of Applied Statistics**, v. 48, p. 349-372, 2020. doi:10.1080/02664763.2020.1723503.

VERECKEN, H. et al. Modeling soil processes: review, key challenges and new perspectives. **Vadose Zone Journal | Advancing Critical Zone Science**, v. 15, 2016.

VISCARRA ROSSEL, R.A. et al. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. **Geoderma**, v.131, p.59-75, 2006.

VITHARANA, U. W. A.; MISHRA, U., MAPA, R. B. National soil organic carbon estimates can improve global estimates. **Geoderma**, v. 337, p. 55-64, 2019.

XU, C. et al., Developing pedotransfer functions to estimate the S-index for indicating soil quality. **Ecological Indicators**, v. 83, p. 338-345, 2014.

XU, X. et al. Pedotransfer functions for estimating soil water retention properties of northern China agricultural soils: Development and needs. **Irrigation and Drainage**, v. 70, n. 4, p. 593-608, 2021.

WADOUX, A. M.J-C.; MINASNY, B.; MCBRATNEY, A. B. Machine learning for digital soil mapping: Applications, challenges and suggested solutions. **Earth-Science Reviews**, v. 210, p. 103359, 2020.

WALKLEY, A.; BLACK, I.A. An examination of the Degtjareff method for determining soil organic matter, and proposed modification of the chromic acid titration method. **Soil Science**, v. 37 p. 29-38, 1934.

WANG, B. et al. Estimating soil organic carbon stocks using different modelling techniques in the semi-arid rangelands of eastern Australia. **Ecological Indicators**, v.88, p. 425-438, 2018.

WANG, X. et al. High precision error prediction algorithm based on ridge regression predictor for reversible data hiding. **IEEE Signal Processing Letters**, v. 28, p. 1125-1129, 2021.

YEOMANS, J.C.; BREMNER, J.M. A rapid and precise method for routine determination of organic carbon in soil. **Communications in Soil Science and Plant Analysis**, v.19, p.1467-1476, 1988.

ZACHARIAS, S.; WESSOLEK, G. Excluding organic matter content from pedotransfer predictors of soil water retention. **Soil Science Society of America Journal**, v. 71, p. 43-50, 2007.

ZHAI, Z. et al. Decision supports systems for agriculture 4.0: surveys and challenges. **Computers and Electronic in Agriculture**, v. 170, 105256, 2020.

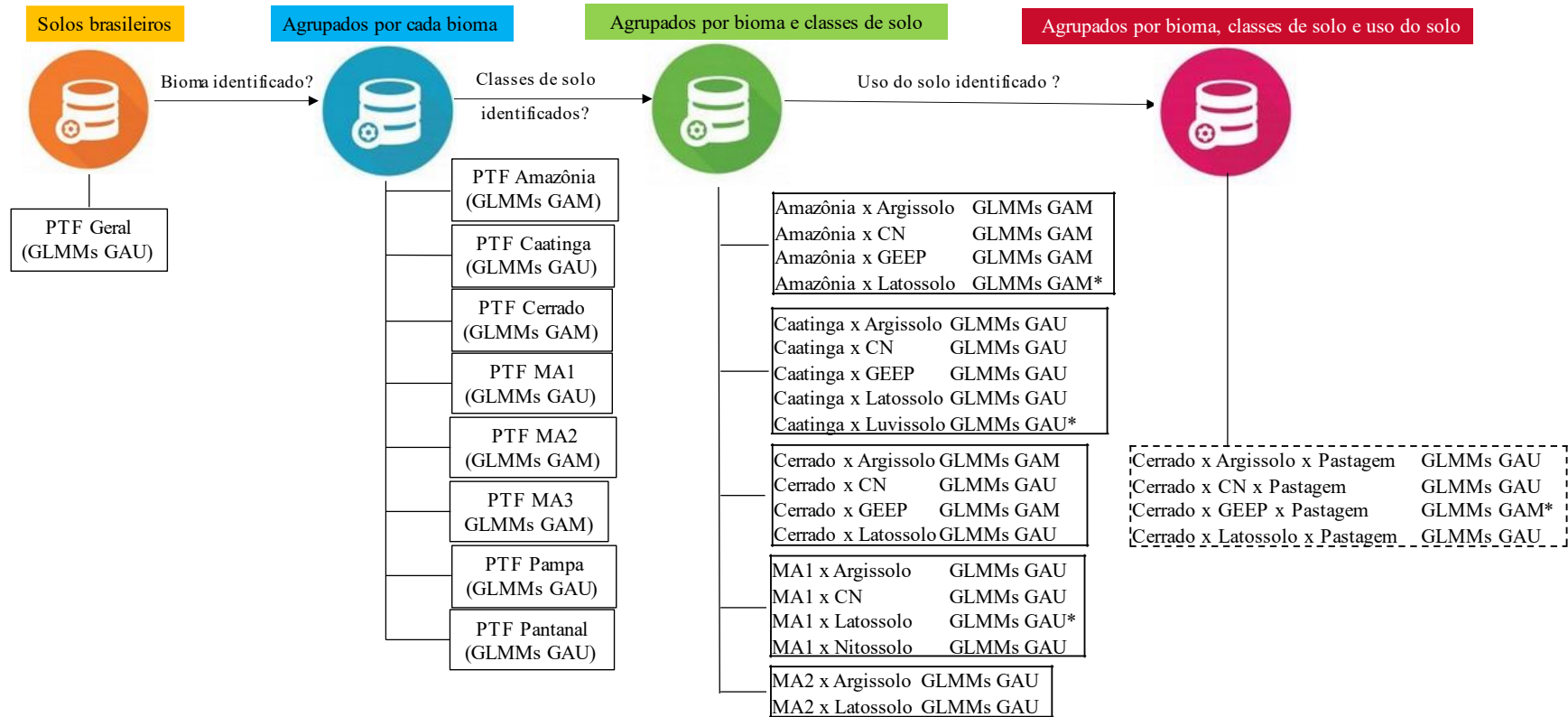
ZHANG, M. W. et al. Application of generalized linear geostatistical model for regional soil organic mapping: the effect of sampling density. **Geoderma**, v. 433, 116446, 2023.

ZHANG, Y.; SCHAAP, M. G. Estimation of saturated hydraulic conductivity with pedotransfer function. **Journal of Hydrology**, v. 575, p. 1011-1030, 2019.

ZHANG, Z. et al. Exploring the inter-decadal variability of soil organic carbon in China. **Catena**, v. 230, 107242, 2023.

ZINN, Y. L. et al. Texture and organic carbon relations described by a profile pedotransfer function for Brazilian Cerrado soils. **Geoderma**, v.127, p. 168-173, 2005.

## ANEXO A – RECOMENDAÇÕES DE USO DAS PTFs DE ACORDO COM A DISPONIBILIDADE DE DADOS

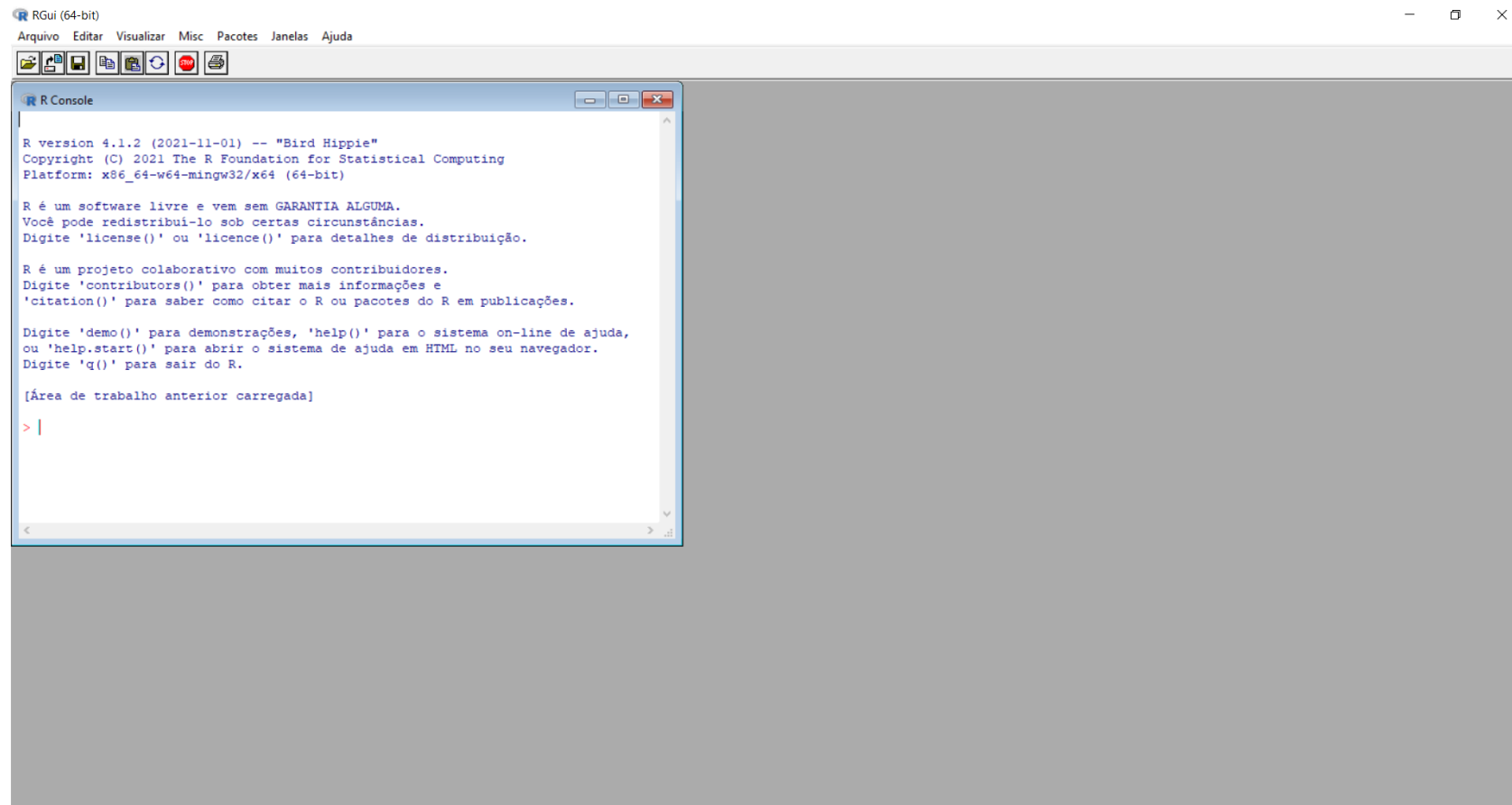


\* Melhor opção refere -se à PTF do bioma ou agrupamento anterior



## ANEXO B – ILUSTRAÇÕES DA SEQUÊNCIA DO PROGRAMA PARA ESTIMAR TEORES DE C A PARTIR DAS METODOLOGIAS GLMMS

1 – É necessária a instalação do *software* R



Fonte: <https://www.r-project.org/> (2023)

2 – O usuário receberá um arquivo compactado com todas as planilhas necessárias para a estimativa de C através da metodologia GLMMs.

| Nome                | Tamanho    | Comprimido | Tipo                  | Modificado        | CRC32     |
|---------------------|------------|------------|-----------------------|-------------------|-----------|
| Pasta de arquivos   |            |            |                       |                   |           |
| ..                  |            |            |                       |                   |           |
| ~\$dados.xlsx       | 165        | 66         | Planilha do Micros... | 11/07/2023 10:... | CCFA787A  |
| ~\$predicao.xlsx    | 165        | 66         | Planilha do Micros... | 11/07/2023 14:... | CCFA787A  |
| dados.xlsx          | 9.650      | 7.090      | Planilha do Micros... | 13/07/2023 17:... | A5AA1254  |
| m4.01-geral.rda     | 45.372.746 | 45.092.102 | Arquivo RDA           | 13/07/2023 10:... | AF5DBD6D  |
| m4.03-A-ARGI-U.rda  | 285.965    | 282.445    | Arquivo RDA           | 13/07/2023 10:... | 3F97E6A7  |
| m4.07-A-GEPP.rda    | 199.748    | 197.762    | Arquivo RDA           | 13/07/2023 10:... | 79D777A1  |
| m4.09-A-LATO.rda    | 272.882    | 260.033    | Arquivo RDA           | 13/07/2023 10:... | BC1B81A8  |
| m4.11-A.rda         | 823.365    | 805.655    | Arquivo RDA           | 13/07/2023 10:... | 527334FC  |
| m4.12-CA-ARGI-U.rda | 267.498    | 257.717    | Arquivo RDA           | 13/07/2023 10:... | CCDE4CE8  |
| m4.15-CA-CN.rda     | 370.244    | 363.408    | Arquivo RDA           | 13/07/2023 10:... | 0C2ED4A1  |
| m4.17-CA-GEPP.rda   | 409.281    | 405.548    | Arquivo RDA           | 13/07/2023 10:... | 5E20140D  |
| m4.18-CA-LATO.rda   | 355.808    | 347.972    | Arquivo RDA           | 13/07/2023 10:... | 16216EE7  |
| m4.19-CA-LUVI.rda   | 230.250    | 228.851    | Arquivo RDA           | 13/07/2023 10:... | 2B90BA7D  |
| m4.21-CA.rda        | 3.876.853  | 3.856.327  | Arquivo RDA           | 13/07/2023 10:... | 8D73612A  |
| m4.23-CE-ARGI.rda   | 437.064    | 432.682    | Arquivo RDA           | 13/07/2023 10:... | AA322FE3  |
| m4.25-CE-CN.rda     | 557.399    | 548.980    | Arquivo RDA           | 13/07/2023 10:... | 070A1946  |
| m4.30-CE-LATO.rda   | 970.393    | 951.015    | Arquivo RDA           | 13/07/2023 10:... | 46BE8107  |
| m4.33-CE.rda        | 3.097.581  | 3.074.678  | Arquivo RDA           | 13/07/2023 10:... | 89014E24  |
| m4.35-MA1-ARGI.rda  | 284.069    | 280.099    | Arquivo RDA           | 13/07/2023 10:... | DBF967C5  |
| m4.41-MA1-LATO.rda  | 619.117    | 613.567    | Arquivo RDA           | 13/07/2023 10:... | 7506643F  |
| m4.44-MA1-NITO.rda  | 215.734    | 214.187    | Arquivo RDA           | 13/07/2023 10:... | 6871DA39  |
| m4.45-MA1-SU.rda    | 915.460    | 892.486    | Arquivo RDA           | 13/07/2023 10:... | 678640AA  |
| m4.48-MA2-ARGI.rda  | 384.733    | 381.088    | Arquivo RDA           | 13/07/2023 10:... | 1CE04075  |
| m4.50-MA2-LATO.rda  | 314.121    | 307.159    | Arquivo RDA           | 13/07/2023 10:... | EC1DCD... |
| m4.51-MA2-SU.rda    | 744.408    | 738.191    | Arquivo RDA           | 13/07/2023 11:... | B0E4A1F8  |
| m4.54-MA3.rda       | 388.515    | 377.329    | Arquivo RDA           | 13/07/2023 10:... | F879EBFC  |
| m4.55-PAMPA.rda     | 195.260    | 193.070    | Arquivo RDA           | 13/07/2023 10:... | F36F6867  |
| m4.56-PANTANAL.rda  | 148.631    | 148.176    | Arquivo RDA           | 13/07/2023 10:... | C02C668E  |
| programa_predicao.R | 4.106      | 833        | Arquivo R             | 13/07/2023 16:... | 15030911  |

Fonte: Autor (2023).

### 3 – Preencher a planilha “dados” com as informações dos atributos preditores disponíveis.

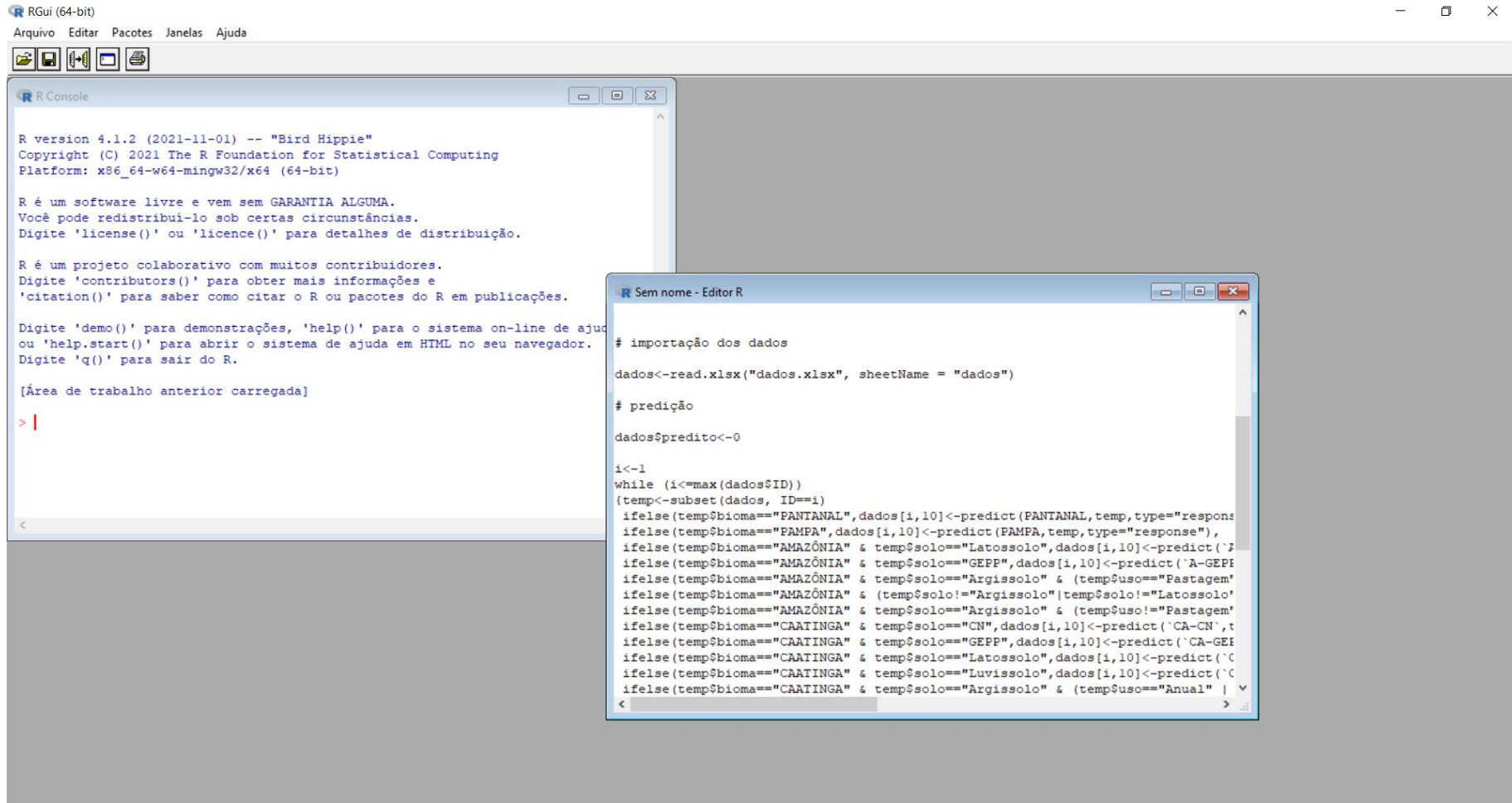
The screenshot shows an Excel spreadsheet with the following data:

| ID | solo      | uso         | bioma          | areia | silte | ph_h2o | longitude | latitude |
|----|-----------|-------------|----------------|-------|-------|--------|-----------|----------|
| 1  | Latossolo | S/ uso agr. | SEM INFORMAÇÃO | 715   | 164   | 4,6    | -63,0267  | -7,51222 |
| 2  | Argissolo | Veg. nativa | AMAZÔNIA       | 180   | 542   | 4,4    | -64,7948  | -7,26195 |
| 3  | GEPP      | Pastagem    | CAATINGA       | 23    | 492   | 4,1    | -60,6206  | -3,29828 |

The dropdown menu for the 'bioma' column includes the following options: MATA ATLÂNTICA 3, MATA ATLÂNTICA 2, CERRADO, MATA ATLÂNTICA 1, PANTANAL, PAMPA, and SEM INFORMAÇÃO.

Fonte: Planilha do excel editada pelo autor (2023)

#### 4 – Rodar o script de interesse.



#### 5 – O dado estimado de C estará na planilha “predição\_C”



***EESC • USP***