Faculdade de Agronomia e Engenharia Florestal
Mestrado em Maneio e Conservação da Biodiversidade
Dissertação de Mestrado

Modelagem do Carbono Orgânico do Solo nas Formações Florestais em Moçambique

Autor: Engº. Sérgio Simão João
Supervisor: Professor Doutor Almeida Sitoe

Maputo, Julho de 2026

SÉRGIO SIMÃO JOÃO

Modelagem do Carbono Orgânico do Solo das Formações Florestais em Moçambique

Dissertação apresentada à Universidade Eduardo Mondlane - Faculdade de Agronomia e Engenharia Florestal, como parte das exigências do Programa de Pós-Graduação em Maneio e Conservação da Biodiversidade, para a obtenção do título de Mestre em Maneio e Conservação da Biodiversidade.

Supervisor: Professor Doutor Almeida Sitoe

Maputo, Julho de 2026

DECLARAÇÃO DE HONRA

Eu, **Sérgio Simão João**, declaro por minha honra que o presente trabalho para obtenção do grau de Mestre em Maneio e Conservação da Biodiversidade, com o tema: **Modelagem do Carbono Orgânico do Solo das Formações Florestais em Moçambique**, foi por mim elaborado, onde os resultados conferem a realidade do campo, com a devida instrução do supervisor, tendo sido feitas as devidas análises dos dados e consultas bibliográficas que se encontram devidamente citadas nas referências bibliográficas.

Declaro ainda que, o presente trabalho nunca foi apresentado por nenhum outro estudante em condições similares para obtenção de qualquer outro grau académico.

O Autor:

_____

(Sérgio Simão João)

Dedicatória

A todos que acreditam que a escola é a chave primária para vencer as barreiras (sociais e económicas) impostas pela conjuntura local ou global. Que este trabalho sirva de fonte de renovação das vossas forças para vencerem no mundo, considerando as tribulações as quais somos sujeitos mesmo antes de nascermos.

AGRADECIMENTOS

À Deus pelo dom da vida, saúde e proteção;

Ao financiador nomeadamente, a Agência Sueca de Cooperação Internacional para o Desenvolvimento (ASDI), através da Universidade Eduardo Mondlane- Faculdade de Agronomia e Engenharia Florestal.

Ao meu Supervisor, na pessoa do Professor Almeida Sitoe, pela tutoria na elaboração do presente trabalho, bem como pela motivação para a continuação dos estudos.

A Unidade de Monitoria, Relatório e Verificação (MRV) do Fundo Nacional de Desenvolvimento Sustentável (FNDS), na pessoa do Engº. Aristides Muhate (MSc.) por ter me colocado o desafio de produzir o mapa de carbono orgânico do solo em ecossistemas florestais, visto que, constituía uma das lacunas que o país dispunha para os projectos REDD+, bem como por ter disponibilizado os dados para o estudo. E

A todos que de forma directa ou indirecta contribuíram quer na componente curricular, quer na componente de dissertação.

ÍNDICE

DECLARAÇÃO DE HONRA ii

Dedicatória iii

AGRADECIMENTOS iv

LISTA DE ACRÓNIMOS E ABREVIATURAS vii

LISTA DE TABELAS viii

LISTA DE FIGURAS viii

I. INTRODUÇÃO 1

1.1. Contextualização 1

1.2. Problema e justificativa do estudo 4

1.3. Objectivos 7

1.3.1. Geral 7

1.3.2. Específicos 7

1.4. Questões de estudo 7

II. REVISÃO BIBLIOGRÁFICA 8

2.1. Características e distribuição dos solos em Moçambique 8

Fonte: Adaptado pelo autor. 14

2.2. Formações Florestais e Stock de Carbono em Moçambique 14

2.3. Carbono orgânico do solo (COS) 16

2.3.1. Conceito e importância 16

2.3.2. Variáveis que influenciam o carbono orgânico do solo 18

2.3.3. Estimativa de carbono orgânico do solo 21

2.4. Modelagem de parâmetros do solo através de técnicas de Machine Learning (ML) 24

III. MATERIAL E MÉTODO 28

3.1. Descrição da área de estudo 28

3.1.1. Localização da área de estudo 28
3.1.2. Clima e Hidrologia 29
3.1.3. Topografia e solos 29
3.1.4. Vegetação 30
3.2. Método 31
3.2.1. Desenho Amostral 31
3.2.2. Levantamento de dados (carbono orgânico do solo) 34
3.2.3. Processamento e Análise de dados 35
IV. RESULTADOS E DISCUSSÃO 49
4.1. Stocks de Carbono Orgânico do Solo 49
4.1.1. Estatística descritiva dos dados 49
4.1.2. Stocks de Carbono Orgânico do Solo Por Tipo Florestal 50
4.2. Modelo de Carbono Orgânico do Solo 55
4.3. Distribuição espacial de carbono orgânico do solo 59
4.3.1. Mapa de distribuição espacial do COS nos diferentes tipos florestais 59
4.3.2. Acurácia do mapa de carbono orgânico do solo 65
V. CONCLUSÕES E RECOMENDAÇÕES 69
5.1. Conclusões 69
5.2. Recomendações 70
VI. REFERÊNCIAS BIBLIOGRÁFICAS 71

LISTA DE ACRÓNIMOS E ABREVIATURAS

% – Percentagem

AGB – Biomassa Acima do Solo

BGB – Biomassa Abaixo do Solo

CO_2eq. - Dióxido de Carbono Equivalente

COS - Carbono Orgânico do Solo

GEE - Gases de Efeito Estufa

ha - Hectares

IV - Quarto

km - Quilómetros

m - Metros

mm - Milímetros

NDC - Nationally Determined Contribuition (Contribuições Nacionalmente Determinadas)

^{o}C - Graus Celsius

PMA - Precipitação Média Anual

ppm - Parte por milhões

Pt - Pentagrama (mil milhões de toneladas de carbono)

tonC - toneladas de carbono

LISTA DE TABELAS

Tabela 1: Relação entre o tipo de solo e o potencial de carbono orgânico do solo 13
Tabela 2: Número de amostras viáveis por cada tipo florestal 33
Tabela 3: Correlações entre as variáveis preditoras quantitativas e a variável dependente..... 43
Tabela 4: Estatística descritiva do COS e suas covariantes da amostra de treino 45
Tabela 5: Estatística descritiva do COS e suas covariantes da amostra de validação 45
Tabela 6: Estatística descritiva dos dados colectados..... 49
Tabela 7: Teores médios de carbono orgânico do solo (COS) por tipo de floresta 50
Tabela 8: Resumo dos parâmetros do modelo de "*random forest*" estimado..... 55
Tabela 9: Resumo dos resultados do ajuste do modelo de "*random forest*" 55
Tabela 10: Estatística descritiva do erro médio de predição de carbono orgânico do solo 66

LISTA DE FIGURAS

Figura 1: Limites geográficos da República de Moçambique 28
Figura 2: Distribuição das unidades amostrais nos tipos florestais 32
Figura 3: Ilustração da unidade amostral (cluster)..... 34
Figura 4: Unidade amostral para colecta de amostra de carbono orgânico do solo..... 35
Figura 5: Separatrizes e boxplot de COS do grupo (esquerda) e subgrupos (direita) dos dados originais..... 36
Figura 6: Histograma de COS do grupo (esquerda) e subgrupos (direita) dos dados originais.... 37
Figura 7: Histograma e testes de verificação dos pressupostos de ANOVA dos resíduos dos dados originais 38
Figura 8: Separatrizes e boxplot de COS do grupo (esquerda) e subgrupos (direita) dos dados transformados (Box-Cox) 38
Figura 9: Histograma de COS do grupo (esquerda) e subgrupos (direita) dos dados transformados (Box-Cox) 39
Figura 10: Histograma e testes de verificação dos pressupostos de ANOVA dos resíduos dos dados transformados (Box-Cox)..... 39

Figura 11: Gráfico de densidade das amostras de treinamento e validação do COS 44
Figura 12: Comparações estatísticas dos teores médios de carbono orgânico do solo entre os estratos florestais através do teste de *TukeyHSD* com ajuste de BH 51
Figura 13: Comparação de *stock* de carbono entre tipos de reservatório por tipo de floresta 54
Figura 14: Relação entre o erro médio quadrático (MSE) e o número de árvores no modelo de *random forest* (verificação da estabilização do erro)..... 56
Figura 15: Importância relativa das variáveis preditoras no modelo de *Random Forest* (%IncMSE) 58
Figura 16: Importância relativa das variáveis das preditoras do modelo de RF (IncNodePurity) 59
Figura 17: Padrão de distribuição do carbono orgânico do solo nos diferentes tipos florestais de Moçambique 60
Figura 18: Padrão de distribuição do carbono orgânico do solo na floresta Decídua incluindo Miombo..... 61
Figura 19: Padrão de distribuição do carbono orgânico do solo na floresta Sempre Verde 62
Figura 20: Padrão de distribuição do carbono orgânico do solo na floresta de Mecrusse..... 63
Figura 21: Padrão de distribuição do carbono orgânico do solo na floresta de Mopane 64
Figura 22: Espacialização do erro de predição (*spatial bubble*)..... 67

RESUMO

Com o crescente interesse dos países signatários do Acordo de Paris em contribuir para o alcance e sustento da meta de **"**valor líquido zero de emissões antrópicas de CO_2eq**."**, aumenta também a relevância de incorporar o carbono orgânico do solo (COS) nas Contribuições Nacionalmente Determinadas (NDCs**)**. Contudo, a escassez de parâmetros nacionais tem sido a grande barreira para a incorporação desse reservatório. Para ajudar na disponibilidade de parâmetros nacionais de COS, foi realizado o presente estudo, cujo objetivo visa modelar a distribuição espacial dos stocks de carbono orgânico do solo nas diferentes formações florestais de Moçambique utilizando técnicas de aprendizagem de automática, visando apoiar o seu mapeamento e gesto sustentável dos recursos naturais frente as mudanças climáticas. Provenientes do IV Inventário Florestal Nacional de Moçambique, os dados usados foram colectados usando a técnica de amostragem aleatória estratificada, a uma profundidade de 0-30 cm e o teor de carbono ao nível do laboratório foi determinado usando o método de *Walkley-Black*, a posterior, convertidos em COS/ha; Estatística F (anova) e teste de *TukeyHSD* à 5% foram usados para avaliar as diferenças de COS entre os estratos. Para a modelagem de COS foi usado o algoritmo *Random Forest* e foram usadas as variáveis biomassa, área basal, precipitação, volume de madeira, radiação solar, evapotranspiração, humidade do solo, vegetação, radiação solar, aspecto, declive, altitude, % de argila, areia e limo e tipo de solo; e métricas estatísticas (ME, RMSE, MAE e R^2) foram usadas para a validação do mapa de COS produzido. A abordagem convencional (dados observados), estimou COS máximo na floresta Sempre Verde (49,63 tonC/ha) e mínimo na floresta de Mopane (35,08 tonC/ha) onde a diferença pode estar associada a diferença na taxa de incorporação da matéria orgânica no solo. A abordagem de *Machine learning* através do algoritmo de *Random Forest* estimou um modelo composto por 500 árvores, 17 variáveis ambientais consideradas para divisão dos nós em cada etapa de treinamento e explicando cerca de 80% da variação do COS. O mapa produzido através da modelagem estimou COS máximo na Floresta Sempre Verde (45,42 tonC/ha) e mínimo no Mopane (34,80 tonC/ha) e precisão aceitável (ME=1,8%, MAE=15,3%, RMSE=23,4% e R=88%). A variação das médias entre as abordagens (r=0,95) pode estar relacionada aos erros de estimativas, heterogeneidade e variação espacial à curta escala das covariantes. Embora o avanço significativo na melhoria das estimativas, há necessidade de se aperfeiçoar as estimativas através de incorporação de mais amostras e covariantes.

Palavras-chave: *Estratos, Random Forest, Factores de emissão, Machine learning.*

ABSTRACT

With the growing interest of countries that are signatories to the Paris Agreement in contributing to the achievement and maintenance of the goal of **"**net zero anthropogenic CO_2eq emissions**"**, the relevance of incorporating soil organic carbon (SOC) into their Nationally Determined Contributions (NDCs) has also increased. However, the scarcity of national parameters has been a major barrier to the inclusion of this reservoir. To support the availability of national SOC parameters, the present study was conducted with the objective of modeling the spatial distribution of soil organic carbon stocks across different forest formations in Mozambique using machine learning techniques, aiming to assist in mapping and sustainable management of natural resources in the face of climate change. The data, derived from the IV National Forest Inventory of Mozambique, were collected using stratified random sampling at a depth of 0–30 cm. Carbon content was determined in the laboratory using the Walkley-Black method and subsequently converted into SOC/ha. F-statistics (ANOVA) and the TukeyHSD test at 5% were applied to assess differences in SOC among strata. For SOC modeling, the Random Forest algorithm was employed, with predictor variables including biomass, basal area, precipitation, wood volume, solar radiation, evapotranspiration, soil moisture, vegetation, aspect, slope, altitude, soil texture (% clay, sand, silt), and soil type. Statistical metrics (ME, RMSE, MAE, and R^2) were used to validate the SOC map produced. The conventional approach (observed data) estimated maximum SOC in Evergreen Forest (49,63 tonC/ha) and minimum in Mopane Forest (35,08 tonC/ha), where differences may be associated with varying rates of organic matter incorporation into the soil. The machine learning approach (Random Forest) produced a model with 500 trees and 17 environmental variables considered at each split, explaining about 80% of SOC variation. The resulting map estimated maximum SOC in Evergreen Forest (45,42 tonC/ha) and minimum in Mopane (34,80 tonC/ha), with acceptable accuracy (ME = 1,8%; MAE = 15,3%; RMSE = 23,4%; R = 88%). The variation in means between approaches (r = 0,95) may be related to estimation errors, heterogeneity, and short-scale spatial variation of covariates. Despite significant advances in improving estimates, there remains a need to refine them through the incorporation of additional samples and covariates.

Keywords*: Strata, Random Forest, Emission factors, Machine learning.*

I. INTRODUÇÃO

1.1.Contextualização

O carbono orgânico do solo (COS) constitui um dos principais reservatórios de carbono dos ecossistemas terrestres e desempenha um papel fundamental no funcionamento dos ecossistemas florestais. A sua dinâmica esta directamente relacionada com a entrada e decomposição da matéria orgânica, com os processos de ciclagem de nutrientes e com as trocas de carbono entre o solo, a vegetação e a atmosfera. Por esta razão a quantificação do COS é relevante não apenas para a compreensão do funcionamento dos ecossistemas, mas também para a avaliação da capacidade dos solos em armazenar carbono e contribuir para a mitigação das mudanças climáticas (FAO, 2017; Trivedi et al., 2018).

Nos ecossistemas florestais, o carbono encontra-se distribuído entre diferentes reservatórios, incluindo a biomassa acima e abaixo do solo e o próprio solo. Embora a biomassa arbórea seja frequentemente considerada nos inventários de carbono, o COS constituí um reservatório de elevada importância, podendo representar uma parcela substancial do carbono total armazenado nos ecossistemas florestais. A sua conservação e aumento podem, portanto, contribuir para a manutenção dos serviços ecossistémicos, associados aos solos e para os esforços de adaptação e mitigação às mudanças climáticas (Burton & Pregitzer, 2008; FAO, 2017, 2018).

A importância do COS torna-se particularmente relevante no contexto das mudanças climáticas e dos compromissos internacionais de redução das emissões de gases de efeito estufa (GEE). No âmbito do acordo de Paris, os países signatários assumiram compromissos nacionais de redução das emissões, materializados nas Contribuições Nacionalmente Determinadas (NDC's). Neste contexto, o sector florestal desempenha um papel estratégico por meio da redução das emissões provenientes do desmatamento e da degradação florestal e da conservação e aumento dos stocks de carbono, incluindo através das iniciativas de Redução de Emissões por Desmatamento e Degradação Florestal e conservação, maneio sustentável das florestas e aumento do stock de carbono (REDD+) (Berry et al., 2013; United Nations, 2015).

Moçambique como parte signatária do Acordo de Paris, submeteu a sua primeira NDC para o período de 2020-2025. Entretanto, no processo de preparação da NDC para o sector florestal, foi

identificada uma importante lacuna relacionada com a disponibilidade de parâmetros nacionais representativos do COS. Esta limitação resulta, em particular, da inexistência de factores de emissão específicos para este reservatório, uma vez que o IV Inventário Florestal Nacional de Moçambique, concentrou-se nos factores de emissão associados ao reservatório de carbono arbóreo (MITADER, 2018a; Mozambique, 2021).

A lacuna de informação sobre o COS é particularmente relevante em um país como Moçambique, onde ocorrem diferentes formações florestais com características ambientais distinta. O país apresenta formações como floresta sempre-verde, floresta de Mecrusse, decídua, incluindo Miombo e Mopane, que se distribuem ao longo do gradiente de precipitação, temperatura, altitude, solos e outras condições ambientais. Essas diferenças condicionam a produção e incorporação de matéria orgânica do solo e consequentemente, podem determinar diferenças nos stocks de COS entre formações florestais (FNDS, 2019; Timberlake & Gumbo, 2011).

A variabilidade do COS não ocorre apenas entre formações florestais, podendo também ocorrer dentro de uma mesma formação em função da variação espacial das condições ambientais. Factores como a precipitação, temperatura, humidade do solo, textura, tipo de solo, topografia, altitude, declive e características da vegetação influenciam a produção, decomposição, estabilização e armazenamento da matéria orgânica do solo. A interacção desses factores, produz uma elevada heterogeneidade espacial das propriedades do solo, podendo resultar em diferenças nos stocks de COS mesmo em áreas pertencentes a mesma formação florestal (Bhunia et al., 2019; Ramesh et al., 2019; Yao et al., 2023).

Em Moçambique, esta variabilidade é evidenciada pelos diferentes valores de COS já reportados para as formações florestais. Estudos realizados em áreas de Miombo, Mopane e floresta sempre-verde demonstram diferenças consideráveis nos stocks de COS, mesmo dentro de uma mesma formação florestal. Essa variação sugere que a utilização de valores médios generalizados pode não representar adequadamente a distribuição espacial do COS no território nacional, reforçando a necessidade de métodos capazes de incorporar a heterogeneidade ambiental na estimativa deste reservatório (Dimande, 2018; Ribeiro et al., 2013; Ryan et al., 2011; Sitoe et al., 2009).

Para reduzir a lacuna de informação existente, o IV Inventário Florestal Nacional incluiu a recolha de amostras de solo destinadas à determinação de propriedades relacionadas com o carbono orgânico do solo. Contudo, devido à morosidade dos procedimentos laboratoriais, os resultados dessas análises não foram incorporados no relatório do inventário. Os dados posteriormente disponibilizados constituem, assim, uma oportunidade para estimar os stocks de COS das diferentes formações florestais de Moçambique e produzir informação espacialmente explícita sobre este reservatório (MITADER, 2018a, 2018b).

Embora os dados observados permitam estimar os stocks médios de COS por formação florestal através de métodos convencionais, esta abordagem apresenta limitações quando se pretende representar a distribuição espacial de uma propriedade que apresenta elevada heterogeneidade. A utilização directa dos valores observados pode não captar adequadamente a variação espacial das condições ambientais entre os pontos amostrados, podendo resultar em estimativas menos precisas em áreas não amostradas. Estas limitações são particularmente importantes quando o objectivo é produzir uma representação espacial do COS para todo o território ocupado pelas formações florestais (Bhunia et al., 2019; MITADER, 2018a).

Neste contexto, técnicas de Machine Learning apresentam-se como uma alternativa para melhorar a predição do COS em locais onde não existem observações directas. Estas técnicas permitem combinar os valores de COS obtidos no campo com variáveis ambientais potencialmente explicativas, explorando relações complexas entre a variável resposta e os seus preditores. A capacidade de modelar relações não lineares e integrar simultaneamente diferentes variáveis ambientais torna o Machine Learning particularmente adequado para estudos de predição espacial das propriedades do solo (Blesslin Sheeba et al., 2022; Mahmoudzadeh et al., 2020; Yang et al., 2021).

Assim, a utilização de um modelo de Machine Learning para o COS permite passar de uma estimativa baseada exclusivamente nos pontos amostrados para uma abordagem capaz de predizer os valores de carbono em locais não amostrados, considerando a influência conjunta de variáveis bióticas e abióticas. No presente estudo, essa abordagem é particularmente pertinente porque permite investigar quais factores ambientais contribuem para explicar a variação espacial do COS

e, posteriormente, utilizar o modelo ajustado para representar espacialmente a distribuição deste reservatório nas formações florestais de Moçambique.

Deste modo, o presente estudo concentra-se na avaliação da variabilidade espacial do carbono orgânico do solo nas formações florestais de Moçambique, partindo de três necessidades complementares: quantificar os stocks de COS nas diferentes formações florestais, identificar e utilizar factores ambientais capazes de explicar a sua variação espacial e produzir uma representação espacial do COS através de modelagem. A resposta a estas necessidades poderá contribuir para reduzir a lacuna de factores de emissão de COS no país e gerar informação relevante para a contabilização do carbono florestal, para as NDCs de Moçambique e para iniciativas de redução de emissões por desmatamento e degradação florestal, conservação, maneio florestal sustentável e sequestro de carbono (REDD+) relacionadas com a adaptação e mitigação às mudanças climáticas.

1.2.Problema e justificativa do estudo

Apesar da importância do carbono orgânico do solo (COS) para o armazenamento de carbono nos ecossistemas florestais e para a mitigação e adaptação às mudanças climáticas, Moçambique ainda apresenta uma importante lacuna de informação sobre a quantidade e a distribuição espacial deste reservatório de carbono. A principal limitação está relacionada à escassez de factores de emissão de COS representativos das diferentes formações florestais existentes no país, o que dificulta a sua incorporação adequada nos processos nacionais de contabilização de carbono e nas Contribuições Nacionalmente Determinadas (NDCs) (MITADER, 2018a; Mozambique, 2021).

A existência desta lacuna é particularmente relevante porque as formações florestais de Moçambique apresentam condições ambientais distintas, capazes de influenciar a acumulação e a distribuição do carbono no solo. A variação de factores como precipitação, altitude, características do solo, topografia, humidade, vegetação e produtividade pode resultar em diferenças significativas nos stocks de COS entre formações florestais e entre diferentes locais dentro da mesma formação. Consequentemente, a utilização de valores médios para representar extensas áreas pode não reflectir adequadamente a variabilidade espacial do COS (Bhunia et al., 2019).

O IV Inventário Florestal Nacional de Moçambique constituiu uma importante fonte de dados para reduzir esta lacuna, uma vez que incluiu a recolha de amostras de solo para determinar propriedades relacionadas com o carbono orgânico do solo. Contudo, devido à morosidade das análises laboratoriais, os resultados dessas amostras não foram incorporados no relatório final do inventário. Assim, embora actualmente existam dados que podem ser utilizados para estimar os stocks de COS, a informação disponível ainda não se encontra plenamente traduzida em factores de emissão e informação espacialmente explícita para as diferentes formações florestais do país (MITADER, 2018a, 2018b).

Adicionalmente, a estimativa convencional baseada exclusivamente nos valores observados nos pontos de amostragem apresenta limitações para representar a distribuição espacial do COS. A heterogeneidade das condições ambientais e a variação das propriedades do solo em pequenas escalas podem fazer com que os valores observados num determinado ponto não representem adequadamente áreas não amostradas. Consequentemente, embora esta abordagem seja útil para determinar stocks médios por formação florestal, ela apresenta limitações quando se pretende produzir uma estimativa espacial contínua do COS em áreas extensas (Bhunia et al., 2019).

Diante desta limitação, surge a necessidade de utilizar métodos que permitam relacionar os valores de COS observados no campo com as variáveis ambientais responsáveis pela sua distribuição. Os algoritmos de Machine Learning constituem uma alternativa metodológica para este efeito, uma vez que permitem integrar simultaneamente diferentes variáveis ambientais e explorar relações complexas, incluindo relações não lineares, entre essas variáveis e o COS. Desta forma, os valores observados nos locais amostrados podem ser utilizados para ajustar modelos capazes de predizer o COS em locais onde não existem observações directas (Blesslin Sheeba et al., 2022; Mahmoudzadeh et al., 2020; Yang et al., 2021).

No caso específico de Moçambique, a aplicação desta abordagem permite ultrapassar a limitação de uma representação baseada apenas em médias por formação florestal, incorporando a variabilidade espacial dos factores ambientais associados ao COS. A utilização de variáveis como biomassa, área basal, volume de madeira, precipitação, radiação solar, evapotranspiração, humidade do solo, altitude, declive, textura e tipo de solo permite investigar quais factores

contribuem para explicar a distribuição do COS e melhorar a sua predição em áreas não amostradas (Austin, 2007; Bird et al., 2001; Lal, 2021).

Neste sentido, a lacuna não se limita à ausência de valores de carbono orgânico do solo, mas envolve também a insuficiência de informação sobre como esse carbono se distribui espacialmente e quais factores ambientais explicam essa distribuição. É precisamente nesta lacuna que se enquadra o presente estudo, ao procurar estimar os stocks de COS nas diferentes formações florestais, ajustar um modelo de Machine Learning para a sua predição e, posteriormente, modelar a distribuição espacial do COS em Moçambique.

A realização deste estudo justifica-se, portanto, pela necessidade de produzir informação quantitativa e espacialmente explícita sobre um reservatório de carbono ainda insuficientemente representado nos mecanismos nacionais de contabilização. A estimativa dos stocks por formação florestal permitirá caracterizar a magnitude do COS, enquanto o modelo de Machine Learning permitirá explorar os factores associados à sua variabilidade e melhorar a predição para áreas não amostradas. O mapa resultante poderá, por sua vez, constituir uma ferramenta de apoio à avaliação e monitoria dos stocks de carbono e ao desenvolvimento de iniciativas relacionadas com REDD+ e com as estratégias nacionais de adaptação e mitigação às mudanças climáticas.

Assim, o estudo procura responder a uma lacuna científica e técnica específica: a insuficiência de conhecimento sobre a magnitude, os factores associados e a distribuição espacial do carbono orgânico do solo nas diferentes formações florestais de Moçambique. Ao integrar dados de campo provenientes do IV Inventário Florestal Nacional com variáveis ambientais e técnicas de Machine Learning, pretende-se gerar uma base de informação mais robusta para a quantificação e representação espacial deste reservatório de carbono, contribuindo para a melhoria dos factores de emissão nacionais e para uma maior consideração do COS nas políticas e mecanismos de mitigação e adaptação às mudanças climáticas.

1.3. Objectivos

1.3.1. Geral

- Modelar a distribuição espacial dos stocks de carbono orgânico do solo nas diferentes formações florestais de Moçambique utilizando técnicas de aprendizagem de automática, visando apoiar o seu mapeamento e gesto sustentável dos recursos naturais frente as mudanças climáticas.

1.3.2. Específicos

- Estimar e comparar os *stocks* de carbono orgânico do solo em diferentes estratos (formações florestais) de Moçambique;
- Ajustar modelo de *Machine Learning* (*random forest*) para a predição do carbono orgânico do solo nas diferentes formações florestais de Moçambique; e
- Produzir mapas de distribuição espacial dos stocks de carbono orgânico do solo utilizando preditivo ajustado.

1.4. Questões de estudo

- Quais são os *stocks* médios de carbono orgânico do solo nas diferentes formações florestais de Moçambique?
- Como se encontra distribuído espacialmente o carbono orgânico do solo nas diferentes formações florestais de Moçambique?

II. REVISÃO BIBLIOGRÁFICA

2.1.Características e distribuição dos solos em Moçambique

O solo é uma camada superior da crosta terrestre, formada por partículas minerais, matéria orgânica, água e microrganismos. É a fonte primária para o suporte e crescimento das plantas, visto que, é do solo que as plantas obtêm nutrientes essenciais para o seu crescimento (FAO, 2008). O solo, deriva da combinação das condições geológicas, hídricas, morfológicas, climáticas e idade da rocha, matriz que varia ao longo do território nacional (CNDS, 2002), e isso ocasiona uma variação espacial à curta escala das propriedades do solo (Bhunia *et al*., 2019), fazendo com que, existam em Moçambique a semelhança do resto do mundo, uma grande variedade de solos (CNDS, 2002).

Gouveia & Azevedo (1949), agruparam provisoriamente a variedade de solos de Moçambique em seis grupos característicos, sendo: solos pedalféricos, solos pedocálinos, solos calomórficos, solos halomórficos, solos hidromórficos e solos aluvionares.

i. Solos pedalféricos

Os solos pedalféricos subdividem-se em solos vermelhos, solos cor da laranja, alaranjados e amarelos, solos cinzentos, solos do planalto dos Macondes, solos vermelhos arenosos dos Urrongas, e solos da faixa arenosa costeira.

a) **Solos pedalféricos vermelhos**- solos de cor vermelha mais ou mesmo intensa, sujeitos ao processo geológico de laterização (desgaste das rochas, lixiviação, acumulo de óxidos de ferro e alumínio, cimentação e formação da crosta) existindo quatro variantes deste tipo de solo, nomeadamente: solos vermelhos lateríticos (solos profundos, vermelho intenso em todo perfil, textura argilosa, friáveis, pH baixo, raras concreções ferruginosas, predominantes em regiões altas e chuvosas “Metonia, Ilé, Alto Molocué, Gurué, Milange, Tacuane, Angónia, Barué, Manica, Chimoio, Gorongosa e Namaacha”); solos cinzentos lateríticos ferruginosos (similares aos solos vermelhos lateríticos, textura variável, abundância de concreções ferruginosas, frequentes nas baixas e planaltos médios de Niassa e Zambézia, podendo também ser encontrados nas zonas altas de Manica e Sofala); solos cinzentos lateríticos com bancadas (apresentam a profundidade variável uma camada de

concreções e cimentadas formando bancadas, que podem ser tipo pisolítico ou celular, as camadas superficiais podem ser completamente expostas); e outros solos vermelhos (solos vermelhos sujeitos ao processo de laterização mas formados a partir de materiais originários provenientes de outras rochas que não as ígneas e metamórficas ácidas, perfil similar aos solos vermelhos lateríticos sendo esses mais compactos, encontrados em pequenas manchas isoladas da província de Niassa) (Gouveia & Azevedo, 1949).

b) **Solos pedalféricos cor de laranja, alaranjados e amarelos**- solos relacionados com solos vermelhos, ocorrendo em sucessão movida provocada pelas condições de drenagem (solos vermelhos em pontos mais elevados da cota, seguidos pelos solos laranja, depois alaranjados, e por fim os solos amarelos em cotas mais baixas), formando catenas, e ocorrendo em Niassa e Zambézia (Gouveia & Azevedo, 1949).

c) **Solos pedalféricos cinzentos**- solos sujeitos ao processo de laterização, encontrados em zonas baixas, de difícil drenagem, encontrando-se associados aos solos vermelhos, cor de laranja, alaranjados e amarelos, sendo os últimos da sucessão catenária. Estes solos dividem-se em: solos cinzentos com horizonte gley (ocorrem em áreas planas, de curta extensão na transição entre encostas e as baixas, caraterizada pela primeira camada cinzenta mais ou menos escura, assentando-se sobre uma rocha parda acinzentada ou amarelada, compacta, com concreções ferruginosas, encontrados em Niassa); solos cinzentos lateríticos ferruginosas (primeira canada de cor cinzenta escura, de espessura variável, delgada, assente a outra rocha espessa de cor cinzenta clara ou esbranquiçada, com concreções ferruginosas, textura variável, encontrando-se em áreas de drenagem da província do Niassa e Zambézia); e solos cinzentos lateríticos com bancadas (solos parecidos com os solos cinzentos lateríticos ferruginosos, mas com concreções soldadas formando bancadas, de difícil drenagem em relação aos solos cinzentos lateríticos ferruginosos) (Gouveia & Azevedo, 1949).

d) **Solos pedalféricos do planalto dos Macondes**- solos formados a partir do material originário proveniente de grés do Cretácico, sujeitos a uma intensa lixiviação, de cores claras, textura ligeira, friáveis a firmes, muito permeáveis, ocorrendo no planalto de Mueda e Macomia (Gouveia & Azevedo, 1949).

e) **Solos pedalféricos vermelhos arenosos Urrongas-** solos vermelhos, de textura arenosa fina a franco-arenosa, tonalidade acastanhada e escurecida pela presença de matéria orgânica, encontrados no planalto de Urrongas (Gouveia & Azevedo, 1949).

f) **Solos pedalféricos de faixa arenosa costeira-** ocorrem na faixa quase ininterrupta de solos arenosos que vai de Ponta de Ouro à foz do Rovuma, de natureza arenosa a franco-areosa com tendência ligeira a uma textura fina e tonalidades mais escuras a medida que se desloca do interior para a costa, encontram-se nessa transição solos arenosos claros (cinzentos e amarelos), solos arenosos avermelhados, solos arenosos acastanhados, e dunas litorais (Gouveia & Azevedo, 1949).

ii. Solos pedocálicos

Os solos pedocálicos subdividem-se em dois grupos de solos, nomeadamente: os solos de terra negras e cinzentas pedocálicas (solos de barro negro da Moamba e solos cinzentos do Guijá), e os solos pardos (pardo-avermelhado e pardo acinzentados) das regiões áridas e semiáridas.

a) **Solos negros da Moamba** (variantes dos solos de terras negras e cinzentas pedocálicas) - são solos associados aos solos vermelhos sobre basalto da Moamba. Em alguns casos, esses solos têm nas primeiras camadas, camadas espessas (30-60 cm), nitidamente negras, fortemente argilosas, granulosas, muito fendilhadas e presença de nódulos de calcários, e apresentam-se de coloração castanho-avermelhada. As camadas menos espessa (0,70-100 cm), são fortemente argilosas, estrutura prismática, compactas, fendilhadas e com presença de nódulos de calcários. Em outros casos, os solos têm uma coloração escura em todas as camadas e espessura que varia entre 0,5 a 100 cm. Esses solos criam condições para o desenvolvimento de vegetação espinhosa (*Acacia* spp. e *Dicrostachys* sp.), dominada por gramíneas no seu sub-bosque (Gouveia & Azevedo, 1949).

b) **Solos cinzentos do Guijá** (variantes de terras negras e cinzentas pedocálicas) – são solos encontrados em áreas não aluvionares (managa), e tem como caracteristica, a coloração cinzenta, presença de concreções calcareas. A camada superfial é menos espessa (15-30 cm), arenosa e solta, assenta sobre terra cinzenta levemente mais clara, argilo-arenosa, compacta, cimentada nalguns perfis; a segunda camada é mais espessa (podendo chegar a

1,15 m), caraterizada pela presença de coloração cinzenta, argilo-arenosa a argilosa, rica em concreções calcárea, e de coloração cinzenta. Esses solos são encontrados em Magude, Guijá, Chibuto, Panda, Inharrime e Manhiça, predominando nestes solos, formacões florestais com predominância de *Screrocarya caffra*, *Lonchocarpus* sp., *Acacia litakunensis*, *Acacia senegal* e *Acacia arabica* (Gouveia & Azevedo, 1949).

c) **Os solos pardos (pardo-avermelhado e pardo acinzentados) das regiões áridas e semiáridas**- apresentam um horizonte superficial de 30 cm de espessura, pardo-acinzentado, fraco-arenoso, friável, de tonalidade levemente avermelhada, argiloso, compacto, com concreções calcáreas, seguido por uma camada parda, levemente mais clara, argilosa e rica em concreções calcáreas. São solos encontrados na margem do Limpopo, Massangena, Funhalouro e Mabote, ocorrendo nestes solos, formações gregarias de *Androstachys* sp (Gouveia & Azevedo, 1949).

iii. Solos calomórficos

São solos vermelhos, argilosos e compactos, com a camada superficial variando de 15 a 30 cm de espessura, cinzento-avermelhada a castanho-avermelhada, geralmente argilo-arenosa. A camada que segue a superficial é geralmente espessa, aparecendo elementos grosseiros e cascalho calcário a 60 cm de profundidade, são encontrados apenas no distrito de Guijá (Gouveia & Azevedo, 1949).

iv. Solos halomórficos

São solos de fendilhamento superficial, com presença de manchas salinas esbranquiçadas, ocupando áreas de drenagem deficiente e alagados, encontrando-se com frequência em pequenas manchas dissecadas na área ocupada pelos solos cinzentos do Guijá. Os solos salgados da orla marítima que se encontram nos estuários do rio Incomáti e Limpopo e algumas lagoas salgadas são algumas variantes dos solos holomórficos. Ocorrem neste tipo de solos, uma vegetação herbácea caraterizada pela presença de plantas halóficas, com ocorrência ocasional de árvores (Gouveia & Azevedo, 1949).

v. Solos hidromórficos

Os solos hidromórficos subdividem-se em 3 grupos de solos, os solos de machongos, solos de vlei, e solos argilosos das baixas.

a) **Solos de machongos** (variantes dos solos hidromorficos) - são solos cinzentos muito escuros a negros, ricos em matéria orgânica, fofos, de textura variando entre arenosa a argilosa e com abundância de água, que impede a rápida decomposição da matéria orgânica, características essas que as agrupas em solos de machongos puros, arenosos e argilosos. Esses solos ocorrem ao longo das faixas dos cursos de água e podem armazenar até cerca de 42% de matéria orgânica, sendo maioritariamente encontrados na província de Maputo e Inhambane (Gouveia & Azevedo, 1949).
b) **Os solos vlei** (variantes dos solos hidromorficos) - são solos negros argilosos, compactos, fendilhados e com nódulos calcáreos e concreções ferruginosas, que correm em baixas alagadas durante a época chuvosas e que servem de vias de drenagem de cursos de água permanentes (Gouveia & Azevedo, 1949).
c) **Solos argilosos das baixas** (variantes dos solos hidromorficos) - são solos cinzentos, quase negros, extraordinariamente argilosos e compactos, muito fendilhados, que ocupam cotas inferiores (zona de transição dos solos cinzentos do Guijá para as áreas aluvionares dos vales do Limpopo e do Elefante). Ocorrem neste tipo de solos, povoamentos puros ou quase puros de *Acacia xanthophloea*, e vegetação arbórea e arbustiva de *Tamarindus sp., Phoenix sp.*, e *Hyphaene sp* (Gouveia & Azevedo, 1949).

vi. Solos aluvionares

São solos formados a partir de material carregado pelos cursos de água e depositados nas suas margens após as cheias, sendo as maiores aluviões encontrados nos vales do Limpopo, Incomáti, Umbelúzi, do Rio dos Elefantes, Maputo, Save e Mazinchopes. Esses solos apresentam uma grande variabilidade de um local para o outro, significando variação na disposição, espessura e natureza das camadas dos perfis. A título de exemplo, a medida que se afasta dos cursos de água, as camadas de areia (franco-arenosas e arenosas-limosas) vão cedendo gradualmente o seu lugar a

outras camadas argilo-arenosas e argilosas, o que justifica a presença de solos mais argilosos quando se encontra mais distante dos curos de água (Gouveia & Azevedo, 1949).

Os diferentes tipos de solos que ocorrem em Moçambique influenciam directamente no stock de carbono orgânico do solo nos ecossistemas florestais, facto sustentado por Buczko *et al*. (2023) que afirmam que, a variabilidade dos stocks de carbono orgânico são directamente relacionados com os factores do solo (tipo de solo, textura do solo e pH do solo), factores esses que, influenciam no tipo de vegetação que se pode estabelecer e desenvolver em uma determinada região e, segundo (Kuznetsova, 2022), a vegetação é a principal fonte de matéria orgânica no solo.

Com base na combinação das características descritas acima para cada tipo de solo, o potencial de acúmulo de carbono orgânico do solo em cada um dos tipos de solos que ocorre em Moçambique, pode ser resumido em três classes: baixa, média e alta (Gouveia & Azevedo, 1949), conforme apresentado na tabela abaixo (Tabela 1).

Tabela 1: Relação entre o tipo de solo e o potencial de carbono orgânico do solo

Tipo de Solo	**Subtipos Principais**	**Potencial de Acúmulo de COS**
Pedalféricos	Solos vermelhos lateríticos	Alto
	Solos cinzentos lateríticos ferruginosos	Médio
	Solos cinzentos lateríticos com bancadas	Médio
	Outros solos vermelhos não ácidos	Médio
	Solos alaranjados e amarelos em catenas	Médio
	Solos cinzentos com horizonte gley	Médio à Alto
	Solos do planalto dos Macondes	Baixo à Médio
	Solos vermelhos arenosos dos Urrongas	Baixo à Médio
	Solos arenosos da faixa costeira	Baixo
Pedocálicos	Solos negros da Moamba	Alto
	Solos cinzentos do Guijá	Médio
	Solos pardos das regiões áridas e semiáridas	Baixo à Médio
Calomórficos	-	Médio
Halomórficos	-	Baixo

Hidromórficos	Solos de machongos	Muito Alto
	Solos vlei	Alto
	Solos argilosos das baixas	Alto
Aluvionares	-	Médio à Alto

Fonte: Adaptado pelo autor.

2.2.Formações Florestais e Stock de Carbono em Moçambique

Segundo a FAO, a floresta é definida como sendo uma área de pelo menos 0,5 hectares, com árvores maiores que 5 m de altura e cobertura de copa superior a 10%. No mundo, as florestas cobrem cerce de 1/3 da superfície da terra e ocorrem em quatro principais domínios climáticos, sendo o boreal, o temperado, subtropical e tropical (FAO, 2022).

Embora Moçambique tenha adoptado uma definição de floresta proposta por Noa & Falcão, 2016, na qual consideram floresta à terras geográficas com uma área mínima de 1ha, composta por árvores com potencial de atingir 3 m de altura na maturidade e com cobertura de copa maior que 30%, o mesmo não se pode dizer quanto a classificação das formações florestais, o que pode ser mostrado pela inconstância ao tempo da classificação das formações vegetais, que vão desde a classificação de Gomes e Sousa (1957), passando Wild Fernandes e Barbosa (1983), Malleux (1980), Saquete (1994), Marzolli (2007), JICA (2013) e por fim a classificação do FNDS (2019).

O Fundo Nacional de Desenvolvimento Sustentável, na sua última publicação nacional com a classificação das florestas, classifica as formações florestais de Moçambique em floresta de mangal, floresta de Mecrusse, floresta sempre-verde, floresta de montanha, floresta decídua (incluindo o Miombo), e floresta de Mopane (FNDS, 2019).

A floresta de mangal é uma formação que cresce na área entre-marés nas zonas tropical e sub-tropical, ocupando costas abrigadas e lugares com pouca influência da energia das marés e em Moçambique é composta essencialmente pelas espécies *Avicennia marina*, *Bruguiera gymnhoriza*, *Ceriops tagal*, *Heritiera litoralis*, *Lumnitzera racemosa*, *Rhizophora mucronata*, *Sonneratia alba* e *Xylocarpus granatum* (Macamo & Sitoe, 2016).

A floresta de Mecrusse é um tipo de vegetação que ocorre de forma gregária e tem a *Androstachys johnsonii* como a principal espécie, sendo encontrada nas províncias de Gaza e Inhambane. É uma vegetação que se caracteriza pela baixa diversidade de espécies arbóreas e pela uniformidade em termos de indivíduos em diâmetro e altura (Bila *et al*., 2018).

A floresta sempre-verde é uma formação florestal composta pela floresta de galeria, floresta de montanha e floresta sempre verde. A floresta de galeria é uma formação florestal que ocorre nas faixas estreitas dos cursos de água, podendo sofrer secas na estação seca e inundações na estação chuvosa, desde que esteja disponível um lençol freático permanente, sendo compostas pelas espécies *Spirostachys africana, Schinziophyton rautaneniie, Milletia stuhlmanni, Friesodielsia obovatae* e *Strychnus sp* (Marzoli, 2007).

A floresta de montanha é uma formação que ocorre em altitudes entre 1200 à 1600 m com PMA entre 1700 à 2000 mm ou em locais onde há mudança brusca de altitude (300 a 2500 m), sendo encontradas nas regiões de Chimanimani, Gurué, Milange, Vumba, Gorongosa, Namuli, Tamassa, planalto de Mueda e entre outras zonas altas (Marzoli, 2007; Sitoe, 2001).

A floresta sempre verde, é uma formação florestal intermediária entre a floresta tropical perene e a floresta decídua, caracterizada por uma camada superior das árvores desfolhadas durante parte do ano e camada inferior das árvores com folhagem perene (Dittus, 1977).

A floresta decídua é uma formação florestal que apresenta espécies caducifólias ou decídua (Lamprecht, 1990), composta pela floresta de Miombo e floresta indiferenciada. A floresta de Miombo é uma formação caracterizada pela presença das espécies do género *Brachystegia, Julbernardia* e *Isoberlina*, estando essa formação localizada em zonas com PMA acima dos 700 mm e em solos pobres, sendo a mesma dividida em Miombo húmido e seco (Timberlake & Gumbo, 2011). É a principal formação florestal de Moçambique, estendendo-se desde o extremo norte do País no rio Rovuma ao rio Limpopo, sendo mais predominante no norte do País cobrindo extensas áreas da província de Niassa, Nampula e Cabo Delgado (Ribeiro *et al*., 2002).

A floresta de Mopane faz parte do grupo das decíduas e é uma formação dominada pela espécie *Colophospermum mopane* (Timberlake & Gumbo, 2011). Para além da espécie dominante, a *Colophospermum mopane*, existem algumas espécies associadas a este ecossistema, com destaque

para *Acacia exuvialis*, *Combretum apiculatum*, *Combretum imberbe* e *Commiphora sp.* (Sitoe, 2001) e ocupa predominantemente a zona climática árida e semi-árida do Sul do Rio Save (Timberlake & Gumbo, 2011).

A floresta de Mopane é o segundo tipo de vegetação mais extenso depois do Miombo, que se desenvolve em regiões quentes e secas, de baixa altitude (200 a 600 m) e ocorre em solos argilosos, mas pode também ocorrer em solos aluviais entre outro (Bila & Mabjaia, 2012).

Essas florestas providenciam uma variedade de benefícios ambientais, económicos, sociais e culturais. Os benefícios incluem o fornecimento de produtos florestais madeireiros (madeira, lenha) e não madeireiros (frutas, peixes, carne de caça, insectos comestíveis, cera de abelha, remédio e mel), fornecimento de serviços ecossistémicos (regulação dos fluxos e qualidade da água, absorção máxima da chuva, preservação das bacias hidrográficas, clima e protecção do solo contra a erosão, oxigénio, sequestro do dióxido de carbono, mitigação do aquecimento global), ecoturismo e recreação, entre outros (Aju *et al*., 2015; Timberlake & Gumbo, 2011).

No actual contexto das alterações climáticas, essas florestas são tidas como principais reservatórios de carbono, pois, desempenham um papel importante no ciclo global do carbono, funcionando como fonte de emissão através da degradação e desmatamento florestal, assim como um sumidouro através do processo de captura e armazenamento de biomassa pelo processo de fotossíntese (FAO, 2022). Em Moçambique, todas as formações florestais, com excepção da floresta de mangal, tem um *stock* de carbono estimado em 1419 milhões de toneladas de carbono arbóreo acima e abaixo do solo, dos quais 60% se encontram na floresta decídua e o resto divide-se entre as outras formações florestais exceptuando a floresta de mangal (MITADER, 2018a).

2.3.Carbono orgânico do solo (COS)

2.3.1. Conceito e importância

O carbono é um elemento químico do grupo 14 da tabela periódica (Oshiro, 2010), que é a principal base da vida terreste (British Society of Soil Science, 2021), isso por que, participa na composição química de todos compostos orgânicos (Rosa *et al.*, 2003), que são constituintes essenciais de toda

a matéria viva e são fundamentais na respiração, fotossíntese e regulação do clima (Martins *et al*., 2003).

O carbono por ser um dos elementos mais comuns do universo, ele é encontrado praticamente em todos reservatórios do planeta terra, neste caso, nos oceanos, atmosfera, formações geológicas (carbono fóssil e mineral) e nos ecossistemas terrestres (biota e solos) (British Society of Soil Science, 2021; Corning *et al*., 2016).

No ecossistema terrestre, o carbono está presente em maiores quantidades nos compostos inorgânicos e menores quantidades nos compostos orgânicos (biomassa marinha e terrestre), representando cerca de 0,05% do total de carbono nos ecossistemas terrestres (British Society of Soil Science, 2021). Os oceanos constituem os maiores reservatórios de carbono (38500 Gt), seguido pelos solos (1720 Gt), atmosfera (755 Gt) e por fim pela biomassa (florestas e outras formações vegetais) com cerca de 650 Gt.

Do total do carbono disponível na biomassa, cerca de 90% é encontrado nos ecossistemas florestais (Barbosa *et al*., 2013; British Society of Soil Science, 2021; Corning *et al*., 2016), daí a razão das florestas serem consideradas como um compartimento vital (sumidouro de carbono) (Hammarstrand & Särnberger, 2013). Dentro dos ecossistemas florestais, o carbono encontra-se distribuído de forma irregular, podendo a componente arbórea armazenar até 60% do total de carbono das florestas (Curtis, 2008).

O stock de carbono no solo é cerca de três vezes maior do que na atmosfera, em torno de 2300 Gt (2,3 trilhões de toneladas) até uma profundidade de três metros e 1500 Gt no primeiro metro (British Society of Soil Science, 2021). O stock de carbono no solo encontra-se na forma orgânica e inorgânica, ou seja, carbono inorgânico e orgânico do solo (Xiao, 2015). O carbono orgânico do solo, é parte do ciclo global de carbono que envolve o movimento do carbono através do solo, vegetação, oceano e atmosfera, o que o torna um reservatório variável de carbono (FAO, 2017). O carbono orgânico do solo está sujeito a perdas devido à erosão do solo. Estudos estimam que a erosão em regiões áridas provoca emissões de 0,21 a 0,26 Pg de carbono por ano (Timberlake & Gumbo, 2011).

O ciclo global de carbono é um geossistema que consiste na transferência de carbono em diferentes formas (trocas gasosas provenientes da respiração, fotossíntese, queima de biomassa, intemperismo e precipitação de carbonato de cálcio) e o solo (pedosfera), sendo esse classificado em geológico e biológico (Barbosa *et al*., 2013; FAO, 2017; Trivedi *et al*., 2018). Nos ecossistemas florestais, as trocas de carbono entre a atmosfera e a biosfera ocorrem através da fotossíntese e respiração. Através da fotossíntese, a energia solar e o dióxido de carbono (CO_2) atmosférico são convertidos e armazenados na forma de biomassa (energia química vegetal) e pela respiração, parte do carbono contido na biomassa é reemitido para a atmosfera (FAO, 2017; NCASI, 2013).

A queda das folhas, galhos, flores, frutos e cascas resultantes da renovação dos constituintes das árvores, que caracterizam o crescimento das árvores, proporciona a dinâmica de carbono, na qual, parte do carbono é incorporado na atmosfera em forma de CO_2 e o resto é incorporado no solo na forma de húmus estável (FAO, 2017; Trivedi *et al*., 2018).

Dependendo da taxa de emissão e remoção do carbono, o fluxo de carbono entre a atmosfera e a biosfera pode estar em equilíbrio ou em desequilíbrio (Brown *et al*., 1989; Trivedi *et al*., 2018). O ciclo de carbono encontra-se em equilíbrio, quando as quantidades de carbono emitidas para a atmosfera são iguais às quantidades removidas da atmosfera. O ciclo de carbono entra em desequilíbrio quando o carbono presente nos combustíveis fósseis, na biomassa e no solo é liberto e emitido através da actividade humana para a atmosfera, ou seja, "acréscimo na atmosfera de uma quantidade de carbono que não fazia parte por milhões de anos" (NCASI, 2013; Trivedi *et al*., 2018).

2.3.2. Variáveis que influenciam o carbono orgânico do solo

Embora o carbono ocorra em todos reservatórios do planeta terra (oceanos, atmosfera, formações geológicas, e ecossistemas terrestres (British Society of Soil Science, 2021; Corning *et al*., 2016), o teor de carbono orgânico do solo disponível no solo é influenciado por factores climáticos e pelos factores relacionados com o solo (Ramesh *et al*., 2019; Yao *et al*., 2023), e estes factores influenciam a variação espacial do carbono.

Os factores climáticos incluem, a temperatura, precipitação, humidade, evapotranspiração, enquanto os factores do solo incluem, a textura do solo (limo, argila e areia), densidade do solo,

mineralogia, porosidade, tipo de solo, pH do solo, humidade do solo, estrutura do solo, comunidade microbiana do solo, topografia, e altitude (Ramesh *et al*., 2019). Esses factores de forma conjunta ou isolada tem potencial de influenciar a ciclagem de nutrientes e matéria orgânica no solo, pois controlam o potencial de sequestro de carbono dos ecossistemas.

Ao nível do solo, a temperatura tem papel de estabelecer o equilíbrio entre as adições e exclusões do carbono orgânico do solo. As adições se referem a incorporação da matéria orgânica no solo através do processo de decomposição da matéria orgânica, enquanto as exclusões se referem a emissão do dióxido de carbono, metano e lixiviação de compostos de carbono. A temperatura ajuda a acelerar o crescimento vegetal (disponibilidade de matéria orgânica), bem como intensificar a decomposição da matéria orgânica através do aumento da actividade microbiana (aumento exponencial da taxa de decomposição com o aumento da temperatura) (Ramesh *et al*., 2019).

A precipitação tem influência directa e indirecta na dinâmica do carbono orgânico do solo nos ecossistemas florestais. O aumento da precipitação proporciona condições para o crescimento da vegetação que ocasiona o aumento na produção da biomassa abaixo e acima do solo, e dependendo da taxa de decomposição (actividade microbiana) grande parte desta biomassa é incorporada no solo em forma de húmus. Outro impacto da precipitação é a alteração dos processos biológicos que levam a mudança no pH do solo, disponibilidade de nutrientes, intemperismo do solo e mineralogia do solo (Ramesh et *al*., 2019).

A humidade do solo que é directamente proporcional ao aumento da precipitação, tem potencial de influenciar os níveis de carbono orgânico do solo, visto que, influencia os ciclos biogeoquímicos. Solos com menos humidade (água) proporcionam aumento da disponibilidade dos microrganismos, ocasionando a libertação de dióxido de carbono por meio da decomposição microbiana do carbono orgânico do solo, enquanto solos com mais humidade (água) proporcionam a redução da taxa de difusão de oxigénio para os locais de reação e promove a decomposição anaeróbica e reduz a taxa de mineralização (Ramesh *et al*., 2019).

Os tipos de solo regulam as variações dos teores de carbono orgânico do solo, visto que, a produção da biomassa é determinada pelos níveis de nutrientes, textura e outras propriedades de mineralogia contidos em cada tipo de solo. Solos com altos níveis de nutrientes produzem mais biomassa, ocasionando maior armazenamento do carbono orgânico do solo. Os histossolos têm maior

concentração do carbono orgânico do solo, seguidos de forma decrescente pelos Inceptissolos, Latossolos, Alfissolos, Aridissolos, Entissolos e Vertissolos. Ao nível dos ecossistemas florestais, altos teores de carbono orgânico do solo são encontrados em solos pretos que são ricos em argila (Ramesh *et al*., 2019).

A textura do solo definida como distribuição percentual das partículas de areia, limo e argila (fração mineral), tem importante papel na estabilização (retenção) e sequestro do carbono orgânico do solo. Os solos arenosos a franco-arenosos contêm baixos níveis de carbono, que se deve a baixa capacidade de protecção da matéria orgânica que resulta do défice da argila e limo. Solos com alto teor de argila e limo contêm altos níveis de carbono orgânico do solo que se deve a alta proteção que a matéria orgânica está sujeita, o que lhes permite maior resistência aos ataques microbiano do carbono e consequente maior armazenamento de carbono. O aumento dos teores de argila proporciona aumento proporcional da matéria orgânica no solo, primeiro porque cria dificuldade de decomposição da matéria orgânica, segundo, porque aumenta a probabilidade de formação de agregados, cujo papel é de proteger a matéria orgânica da mineralização. O papel da argila na protecção do carbono orgânico do solo deve-se a presença de fracções de minerais primários (quartzo, feldspatos), filossilicatos (caulinita, esmectitas), óxidos e hidróxidos metálicos (goethita , ferrihidrita), minerais de aluminossilicato de ordem de curto alcance (alofano, imogolita), e carbonatos (Ramesh *et al*., 2019).

O pH do solo afecta a dinâmica do carbono orgânico do solo através da alteração na decomposição do carbono orgânico do solo pelos processos físicos, químicos e biológicos. Nos processos físico-químicos, o pH tem potencial de afectar a solubilidade de substâncias húmidas do carbono orgânico do solo por meio de reações de protonação/ desprotanação, onde a quantidade de carbono orgânico do solo dissolvido, aumenta em função do aumento do pH do solo induzido por cal, e diminui com a alta deposição ácida. Nos processos biológicos, o pH tem potencial de alterar a decomposição do carbono orgânico, isto porque afecta o crescimento dos microrganismos responsáveis pelas actividades microbianas, sendo pH óptimo para estabelecimento e desenvolvimento de microrganismo o que se encontram da faixa de 6,5 a 7,5 (Ramesh *et al*., 2019).

A topografia através da altitude, declive e aspecto, influência a dinâmica do carbono orgânico do solo, isso porque tem potencial de afectar as variáveis temperatura e precipitação, que como

descritas acima, tem influência directa no carbono orgânico do solo. A topografia através do declive contribui para a redistribuição da matéria orgânica ao longo da paisagem (erosão), fazendo com que regiões baixas tenham maior concentração da matéria orgânica em relação ao topo de um planalto ou montanha. O mesmo acontece com o aspecto (topografia), na qual maior concentração de matéria orgânica tende a ser encontra em encostas voltadas para o Norte (hemisfério Norte) em relação as encostas voltas para o Sul, devendo-se pela diferença na temperatura que afecta de forma directa a ação microbiana. A semelhança do declive e aspecto, a altitude promove alteração das variáveis climáticas (precipitação e temperatura), fazendo com que os teores de carbono orgânico sofram alteração também (aumento do carbono orgânico do solo com o aumento da altitude) (Ramesh *et al*., 2019).

As diversas formas de interação dos factores acima descritos tornam o solo um ambiente muito variável em termos de estrutura e composição (heterogéneos), fazendo com que os teores de carbono tenham uma variação a curta escala (Bhunia *et al*., 2019), o que explica a variação dos teores de carbono orgânico encontrados no mesmo ecossistema, conforme descrito na secção 2.3.3.

2.3.3. Estimativa de carbono orgânico do solo

A quantificação de carbono nos ecossistemas florestais exige a estimativa da matéria orgânica contida no solo (Sitoe *et al*., 2009). Devido a presença de quantidades substanciais de carbono armazenado no solo, a estimativa de carbono nestas componentes do ecossistema florestal (solo) sempre se justifica (Pearson *et al*., 2005).

Para a quantificação do carbono orgânico do solo, três variáveis devem ser medidas: a profundidade, a densidade do solo (obtida da relação entre o peso seco e volume do cilindro conhecido) e as concentrações de carbono orgânico ao nível do laboratório (Pearson *et al*., 2005).

O levantamento de dados para a determinação de *stocks* de carbono orgânico do solo é feito através de técnicas de amostragem. A amostragem é um processo pelo qual, parte representativa (amostra) da população é estudada a fim de permitir a generalização ou extrapolação para a população ou área de interesse. Na amostragem, bases estatísticas são usadas para estimar as incertezas (erros),

garantindo assim a verificação da precisão das estimativas realizadas (Pearson *et al*., 2005; Schlegel *et al*., 2001).

A amostragem pode ser realizada através dos métodos probabilísticos ou através de métodos não probabilísticos. Nos métodos probabilísticos, as unidades amostrais têm a mesma probabilidade de fazerem parte da amostra, sendo inverso para o método não probabilístico (Tito *et al*., 2008). Em ambos métodos podem ser usadas ou unidades amostrais de área variável, ou unidades de área fixa (permanentes ou temporárias), ambas seleccionadas de forma aleatória ou sistemática (Pearson *et al*., 2005).

Nos métodos de amostragem, a representatividade das unidades amostrais e a redução do erro de estimativas, consegue-se através da estratificação da área de estudo, podendo a estratificação ser segundo o tipo de vegetação, solo ou topografia. Mas em estudos de carbono, apenas a biomassa aérea (vegetação) é usada como critério de estratificação, isto deve-se a forte correlação entre a biomassa aérea e o stock de carbono (Schlegel *et al*., 2001).

Para além da estratificação, o tamanho da amostra (intensidade amostral), o tipo, número e a dimensão das parcelas, também exercem influência sobre a representatividade e precisão do estudo (Schlegel *et al*., 2001). O tamanho da amostra define-se em função da precisão requerida e normalmente varia entre 80 a 95% (Schlegel *et al*., 2001). O tipo, o número e a dimensão das unidades amostrais são definidas em função do uso de terra, custos envolvidos, precisão e informação requerida (Tito *et al*., 2008).

Em estudos de carbono orgânico do solo associado a estudos de biomassa arbórea, recomenda-se que as parcelas para levantamento de dados estejam localizadas no centro das parcelas rectangulares destinadas para o levantamento de dados de vegetação arbórea (Tito *et al*., 2008). Recomenda-se que sejam obtidas quatro subamostras nos quatro quadrantes da parcela rectangular que deveram depois ser harmonizadas para constituírem uma única unidade amostral e a mesma será usada para determinar os teores de carbono ao nível do laboratório; para além dessas amostras, uma outra amostra não perturbada deverá ser retirada no centro da parcela e a mesma será usada para determinar a densidade do solo (Pearson *et al*., 2005). As amostras devem ser retiradas a uma profundidade de pelo menos 30 cm, isto porque é considerada a profundidade com maior

concentração do carbono, que deve-se a maior actividade biológica na superfície do solo (Ravindranath & Ostwald, 2007).

Embora haja padronização dos métodos de estimativa de carbono orgânico do solo, existem variações dos teores de carbono orgânico do solo, facto que se deve a diferença das características das formações, conforme descrito na secção 2.2.

Para a floresta de Miombo foram encontrados os seguintes teores de carbono orgânico do solo, 34,72 tonC/ha em Niassa (Ribeiro *et al*., 2013); 57,9 tonC/ha e 44,9 tonC/ha na floresta intacta e pousio, respectivamente na província de Sofala (Ryan *et al*., 2011); 73,04 tonC/ha na província de Sofala (João, 2016); 39,30 tonC/ha, 62,62 tonC/ha e 48,93 tonC/ha na coutada 9/13, floresta comunitária e concessão florestal, respectivamente na província de Manica (Dimande, 2018); 31,4 tonC/ha na província de Manica (Sitoe *et al*., 2009). Os teores de carbono orgânico do solo encontrados na floresta de Miombo apresentam uma variação de 41,64 tonC/ha, com mínimo de 31,4 tonC/ha e máximo de 73,04 tonC/ha.

Na floresta de Mopane foram encontrados os seguintes teores médios de carbono orgânico do solo, 47,34 tonC/ha na província de Manica (Dimande, 2018); 60,7 tonC/ha na província de Manica (Boma, 2017); 18,97 tonC/ha na província de Manica (Lisboa, 2017); e 24,6 tonC/ha e 16,2 tonC/ha na província de Gaza (Cambule *et al*., 2014). Os teores de carbono orgânico do solo encontrados na floresta de Mopane apresentam uma variação de 44,5 tonC/ha, com mínimo de 16,2 tonC/ha e máximo de 60,7 tonC/ha.

Na floresta sempre-verde foram encontrados os seguintes teores de carbono orgânico do solo, 91,94 tonC/ha na província de Manica (Dimande, 2018); 94,6 tonC/ha na província de Manica (Dimande, 2018). Os teores de carbono orgânico do solo reportados na floresta sempre verde apresentam uma variação de 2,66 tonC/ha.

As variações de teores de carbono orgânico do solo descritos acima dentro dos tipos florestais sustentam a ideia da variação espacial à curta escala dos teores de carbono orgânico do solo, que se deve a variação das componentes ambientais (Bhunia *et al*., 2019), isto porque os valores reportados variaram consoante o tipo de regime de uso (coutada, concessão florestal, floresta

comunitária), estado de conservação (intacto, perturbado) das formações florestais (Dimande, 2018; João, 2016).

2.4.Modelagem de parâmetros do solo através de técnicas de Machine Learning (ML)

Nos dias actuais cresce a preocupação no uso da abordagem convencional para a estimativa do carbono orgânico do solo, porque essa abordagem pode levar a produção de estimativas menos precisas devido a factores como a falta de homogeneidade espacial das unidades amostrais, variação espacial à curta escala das variáveis ambientais dentro da unidade de observação e erros de estimativas (Bhunia *et al*., 2019).

A modelagem através das técnicas de *machine learning* tem sido apontada como uma das soluções para fazer face às limitações da abordagem convencional, visto que, é uma abordagem que combina variáveis ambientais e observações de campo para a predição da variável de interesse em locais não amostrados (Blesslin Sheeba *et al*., 2022).

O *machine learning* é consequência natural da interseção da ciência da computação e a estatística (Mohammed *et al*., 2016). O *machine learning* é o processo de entrada de dados nos sistemas de computador, tornando este capaz de processar e executar a actividade no futuro sem que seja programado ou alimentado com dados semelhantes ou extras (Sullivan, 2017).

O *machine learning* ajuda a entender como os animais e humanos aprendem, ajuda na extração de relações e correlações escondidas entre um elevado número de dados, assim como melhora o trabalho de projecto de máquinas já existentes (Mohammed *et al*., 2016), visto que a quantidade de conhecimento disponível sobre determinadas tarefas pode ser muito grande para a codificação explícita por humanos, assim, as máquinas que aprendem esse conhecimento gradualmente, podem ser capazes de capturar mais do que os humanos desejam (Nilsson, 1998).

Essencialmente o *machine learning* é dividido em *machine learning* supervisionado, *machine learning* não supervisionado, *machine learning* semi-supervisionado, e *machine learning* por reforço. O *machine learning* supervisionado, também conhecido como aprendizado indutivo, é considerado o mais avançado e maduro de todos os tipos de *machine learning* (Sullivan, 2017). O *machine learning* supervisionado tem como objectivo inferir uma função ou mapeamento, a partir

dos dados de treinamento rotulados (Alafandy *et al*., 2022; Mohammed *et al*., 2016). As máquinas podem ser usadas também para a rotulagem, pois, o julgamento humano é mais caro que o das máquinas, embora se saiba que os dados rotulados manualmente são um recurso precioso e confiável para a aprendizagem supervisionada.

a) ***Machine learning*** **supervisionado -** contempla dois grandes grupos de técnicas, a regressão e a classificação (Mohammed *et al*., 2016). As duas técnicas contemplam um leque de algoritmos, dos quais podemos destacar, o *regression model, decission tree, support vector machine, artificial neural network, Ensemble Learning e random forest*, que são algoritmos amplamente usados na predição das propriedades do solo e têm a vantagem de poderem modelar correlações não lineares complexas entre variáveis ambientais e as propriedades do solo *(*Mahmoudzadeh *et al*., 2020), embora cada um produza níveis diferentes de incertezas associadas as estimativas (Nabiollahi *et al*., 2019). O *random forest* é um algoritmo de ML usado para resolver problemas relacionados com a classificação e regressão (Mahmoudzadeh *et al*., 2020). O uso do *random forest* tem ganhado cada vez mais espaço, e isso pode estar associado a facilidade que o algoritmo tem em lidar com dados de alta dimensão (dados categóricos, binários, contínuos, dados em falta), mas também a eficiência do algoritmo na descriminação da classificação (Ali *et al*., 2012).

A regressão é um algoritmo supervisionado de *machine learning,* normalmente usada em modelagem de sereis temporais, precisão e definição de relação entre variáveis (Benos *et al*., 2021), para a obtenção de uma função de previsão de uma variável dependente com base em variáveis independentes (Benos *et al*., 2021; Graybill & Iyer, 1994; Sullivan, 2017).

Dos diversos tipos de regressão, os mais comuns usados no *machine learning* são: a regressão linear, regressão polinomial, regressão de lasso, regressão logística, regressão bayesiana linear, regressão de mínimos quadrados ordinários (OLSR), regressão gradual, splines de regressão adaptativa multivariada (MARS) e a suavização de gráfico de dispersão estimada localmente (LOESS) (Sullivan, 2017).

A classificação é um algoritmo supervisionado de *machine learning* que se difere da regressão através da função hipótese, visto que, a função custo e os parâmetros do modelo actualizado são

quase os mesmos na classificação e diferentes na regressão. Nos modelos de aprendizado de classificação, as entradas e saídas do conjunto de dados devem possuir uma relação matemática, porém as saídas são previstas dentro de uma função discreta (Alafandy *et al*., 2022; Deuschle, 2019). Na classificação, a função de hipótese pode ser calculada para cada entrada no conjunto de dados, usando os parâmetros do modelo e os recursos de entrada, sendo que, a função mais utilizada é a função de sigmoide (Deuschle, 2019).

b) ***Machine learning*** **não supervisionado -** é uma técnica que tomam em conta dados de treinamentos, tendo como foco encontrar uma estrutura oculta nos dados (Mohammed *et al*., 2016). O *machine learning* não supervisionado é prático em situações de indisponibilidade de fundos para a rotulagem manual ou quando os dados assim o exijam, ou seja, natureza inerente dos dados (Benos *et al*., 2021), dado que existem numerosos dispositivos de recolha de dados, colocando a recolha em um ritmo sem precedentes, possibilitando uma variedade de dados a uma velocidade e volume numeroso (Alafandy *et al*., 2022).

c) ***Machine learning*** **semi-supervionado**- é uma técnica que mistura dados classificados e não classificados, sendo apropriado em situação em que a maioria dos dados rotulados são escassos e os não rotulados são abundantes (Mohammed *et al*., 2016). O *machine learning* semi-supervionado objectiva aprender um modelo que irá prever classes de dados de teste futuros, melhor do que o modelo gerado (Sullivan, 2017).

d) ***Machine learning*** **por reforço**- é uma técnica que visa utilizar observações colhidas na interação com o ambiente, buscando realizar acções que maximizem a recompensa ou minimizem o risco, onde, o estado de entrada é observado pelo agente, a função de tomada de decisão é usada para fazer com que o agente execute uma acção e após a execução da acção, o agente recebe uma recompensa ou reforço do ambiente e as informações do estado e da acção sobre a recompensa são armazenadas (Mohammed *et al*., 2016). As informações armazenadas, a política para um estado específico em termos de acção podem ser ajustadas, o que ajuda na tomada de decisão ideal para o agente (Nilsson, 1998).

Apesar das algumas lacunas relacionadas com a seleção de algoritmos de *machine learning* em estudos das propriedades de solos, há um registo significativo do uso das técnicas de *machine learning* em estudos relacionados ás propriedades do solo (Benos *et al*., 2021; Liu *et al*., 2018; Mahmoudzadeh *et al*., 2020; Nabiollahi *et al*., 2019) e são exemplos desses estudos os seguintes:

- Estudo conduzido por Yang *et al*. (2021), cujo objectivo era a previsão do teor de carbono orgânico do solo em escala regional usando variáveis fenológicas baseadas em satélites. Os autores exploraram a eficácia das variáveis fenológicas MODIS MCD12Q2 de dez anos para a previsão de SOC com um modelo de rede neural convolucional (CNN), concluíram que, a adição de variáveis fenológicas possibilita a obtenção de maior precisão, provando assim que, as métricas fenológicas da superfície eram preditores eficazes e a CNN é um método promissor para o mapeamento do solo em escala regional.
- Estudo conduzido por Gruszczyński (2019), cujo objectivo era buscar predizer as propriedades do solo através dos modelos de aprendizado de máquina, reduzindo o tempo e custo de análise do solo. O autor baseou-se na resposta espectral de amostras de solo na faixa do infravermelho, usando os modelos de regressão *stepwise* tendo como propriedades, as propriedades ao quadrado e produtos de propriedades variáveis explicativas.
- Estudo conduzido por Bhunia *et al*. (2019), cujo objectivo foi fazer predição do carbono orgânico do solo e seu mapeamento usando análise de regressão e dados do sensoriamento remoto no *R* e *GIS*.
- Estudo conduzido por Nabiollahi *et al*. (2019), cujo objectivo era avaliar os stocks de carbono orgânico do solo sob diversos cenários de mudança de uso e cobertura de terra usando modelos de *Random Forest* no Irão.

III. MATERIAL E MÉTODO

3.1. Descrição da área de estudo

3.1.1. Localização da área de estudo

O estudo foi levado a cabo em todas formações florestais (excluindo a floresta de Mangal) das 10 províncias de Moçambique, nomeadamente, Niassa, Cabo Delgado, Nampula, Zambézia, Tete, Manica, Sofala, Inhambane, Gaza e Maputo, o que corresponde a uma extensão de cerca de 32 milhões de hectares de florestas (FNDS, 2019). A não inclusão da floresta de Mangal deveu-se a falta de amostra de solo, isso porque o IV Inventário florestal Nacional de Moçambique não incluiu a floresta de Mangal no seu desenho amostral (população alvo.

Moçambique tem uma superfície de 799380 km^2 e localiza-se na faixa sul-oriental de África entre os paralelos 10º 27’ e 26º 52’ de latitude Sul e entre os meridianos 30º 12’ e 40º 51’ de longitude Este, limitado ao Norte pela Tanzânia, ao Sul pela África do Sul, a Oeste pelo Malawi, Zâmbia, Zimbabwe e Swazilândia e a Oeste pelo Oceano Índico (INE, 1997). A figura 1 ilustra a localização e delimitação das províncias tomadas como áreas de estudo, como acima mencionadas e as duas províncias não inclusas no estudo.

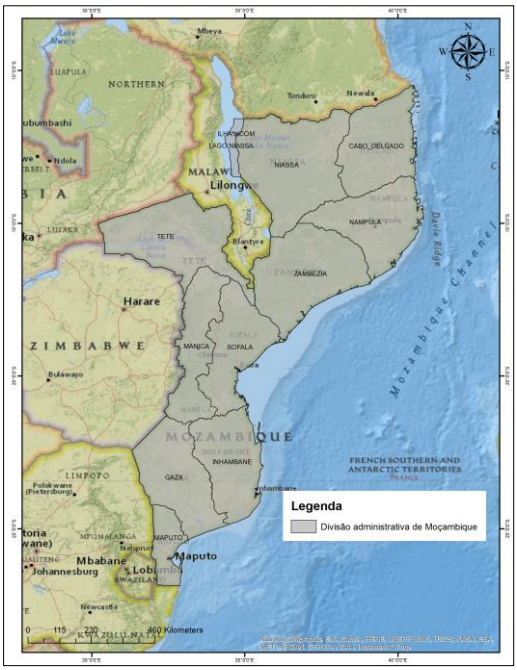


Figura 1: Limites geográficos da República de Moçambique

Fonte: CENACARTA

3.1.2. Clima e Hidrologia

Por ter maior extensão do território localizado na zona intertropical, existem em Moçambique quatro tipos climáticos, que são o clima tropical húmido, tropical seco, clima semiárido e clima modificado pela altitude. O clima tropical húmido é o clima predominante, que se caracteriza por duas estações, a estação quente e húmida (novembro-março), e a estação fresca e seca (abril-outubro) (Milhano, 2008). As temperaturas médias anuais variam entre 24 a 34ºC no verão (estação húmida) e 20 a 22 ºC no inverso (estação seca), podendo chegar a valores extremos de 25 a 27 ºC na zona do vale do Zambeze no verão (CVM, 2022).

A precipitação segue diversos padrões, ao longo da costa, a mesma varia de 800 a 900 mm/ano, com alguns focos húmidos adicionais nas zonas do litoral Sudoeste (SO)e Nordeste (NE). De maneira geral, para a região Sul do país, a precipitação diminui à medida que se desloca do litoral para o interior chegando a valores como 300 mm/ano, havendo alguns focos de aumento da mesma a medida que se aproxima de cadeias montanhosas. Diferente da região Sul, as regiões Centro e Norte apresentam valores mais altos de precipitação, havendo regiões que chegam a registar valores de precipitação de 1800 mm/ano (regiões do delta do Zambeze) (CNDS, 2002; CVM, 2022).

Moçambique apresenta uma larga rede hidrográfica constituída por 104 bacias hidrográficas, sendo o rio Zambeze o maior de todos. Muitas das bacias hidrográficas têm suas nascentes nos países vizinhos, o que lhes torna cursos de água internacionais e são exemplos desses cursos, o rio Zambeze, Limpopo, Rovuma, Save, Incomáti, Maputo, Púnguè, Búzi, Umbelúzi, Tembe e Fúti. Apesar da extensa rede hidrográfica, o país enfrenta problemas de escassez de água, o que se deve a má distribuição do caudal dos rios ao longo do ano, pois a maioria das bacias hidrográficas são de regime irregular e altamente sazonais (apenas três terços de caudal elevado por ano) (Milhano, 2008).

3.1.3. Topografia e solos

Existem em Moçambique três tipos de relevo, o de planície, planalto e de montanha, cujo gradiente de relevo evolui da costa (planície) para o interior (montanha), passando pelo planalto. Embora o padrão da evolução do gradiente, existem algumas excepções em que em algumas planícies

ocorrem montanhas. Em Moçambique, a planície (altitude máxima de 200 m) é homogénea e se estende ao longo de todo litoral (Rovuma-Ponta do Ouro), numa extensão de cerca de 40% do território, sendo a maior extensão encontrada nas províncias de Sofala, Inhambane e Gaza (CNDS, 2002).

De acordo com dados provenientes do modelo digital de elevação, o planalto (altitude mínima de 201 m e máxima de 1000 m) ocupa cerca de 56% do território, sendo de vez em quando interrompida por montanhas e a sua maior extensão encontra-se na região norte e centro, existindo pequenas faixas nas províncias de Maputo e Gaza; já as montanhas (altitude acima de 1000 m) ocorrem nas zonas Centro e Norte e encontram-se agrupadas em cadeias montanhosas (cadeia montanhosa de Maniamba-Amaramba "Niassa", formações de Chire-Namuli "Zambézia", Maciço de Chimanimani "Manica" e cadeia dos montes Libombos "Maputo e Gaza", ocupando cerca de 4% do território nacional (CNDS, 2002).

Devido as condições geológicas e climáticas, existe em Moçambique uma grande variedade de solos, cujo detalhe encontra-se descrito no ponto 2.1 deste documento. Na região Norte (rochas do Pré-câmbrico e altas precipitações) predominam solos argilosos que variam de franco argilosos avermelhados (susceptíveis a erosão) à argilosos vermelho acastanhado (menos susceptíveis a erosão). No litoral da região Norte e Centro (rochas do Fanerozoico) predomina solos arenosos de dunas costeiras e fluviais, havendo extensão de solos franco argilosos, arenosos acastanhados ao sul de Tete e ao longo da bacia do Zambeze; na região Sul, predominam solos arenosos de baixa fertilidade e baixo poder de retenção de água, existem também solos delgados e pouco profundos (pouco apto para agricultura) ao longo da cadeia dos montes Libombos. Em todas regiões, ao longo dos cursos de água, ocorrem solos fluviais de alta fertilidade (CNDS, 2002).

3.1.4. Vegetação

Segundo FNDS (2019), a vegetação de Moçambique é composta por florestas, pradarias e vegetação herbácea inundada. A área de floresta está estimada em cerca de 32 milhões de hectares, a de pradaria em cerca de 21 milhões de hectares, e a de vegetação herbácea inundada está estimada em cerca de 3 milhões de hectares. As florestais subdividem-se em: floresta de mangal, floresta de

Mecrusse, floresta de montanha, floresta sempre-verde, floresta de Mopane, floresta decídua, e florestas plantadas de espécies exóticas de rápido crescimento.

A floresta de Mangal descrita no ponto 2.2 deste documento, é uma formação que se encontra distribuída ao longo da costa de Moçambique, concretamente nas províncias de Cabo delgado, Nampula, Zambézia, Sofala, Inhambane, Gaza e Maputo, e tem uma extensão estimada de cerca de 270 mil hectares (FNDS, 2019).

A floresta de Mecrusse descrita no ponto 2.2 deste documento, é uma formação com cerca de 411 mil hectares de extensão, encontrada nas províncias de Gaza e Inhambane, podendo ocorrer algumas manchas no sul da província de Manica (Machaze), logo no limite com a província de Gaza e Inhambane (FNDS, 2019).

A floresta de montanha, é uma variante da floresta sempre-verde, encontrada nas formações montanhosas das províncias a norte do rio Save, concretamente nas províncias de Sofala, Manica, Tete, Zambézia, Nampula e Niassa. Essa formação cobre uma extensão de cerca de 275 mil hectares. A floresta sempre-verde ocupa uma extensão de cerca de 2.4 milhões de hectares, e encontram-se distribuída em todas as províncias do país (FNDS, 2019).

A floresta de Mopane ocupa uma extensão de cerca de 16.8 milhões de hectares, e ocorre maioritariamente na província de Gaza e Tete, e em pequenas manchas nas províncias de Maputo, Inhambane, Manica e Sofala. A floresta semi-decídua ocupa uma extensão de cerca de 26.5 milhões de hectares, e encontram-se distribuída em todas as províncias do país (FNDS, 2019).

3.2.Método

3.2.1. Desenho Amostral

Os dados foram levantados considerando a floresta (Noa & Falcão, 2016) como população alvo, e para fazer face a integração do inventário florestal com o sensoriamento remoto (análise de mudança de uso e cobertura de terra: desmatamento), os diferentes tipos florestais foram agregados em quatro classes de florestas com base nas características similares, assim quatro estratos foram estabelecidos a saber, a floresta decídua, floresta sempre verde (inclui a floresta de montanha e galeria), floresta de Mecrusse e floresta de Mopane. Foi considerado como desenho amostral, a

amostragem aleatória estratificada restrita (grelha sistemática como critério de restrição), considerando a floresta semi-decídua, floresta sempre verde, Mecrusse e Mopane como estrato, conforme ilustrado na figura um abaixo (Figura 2).

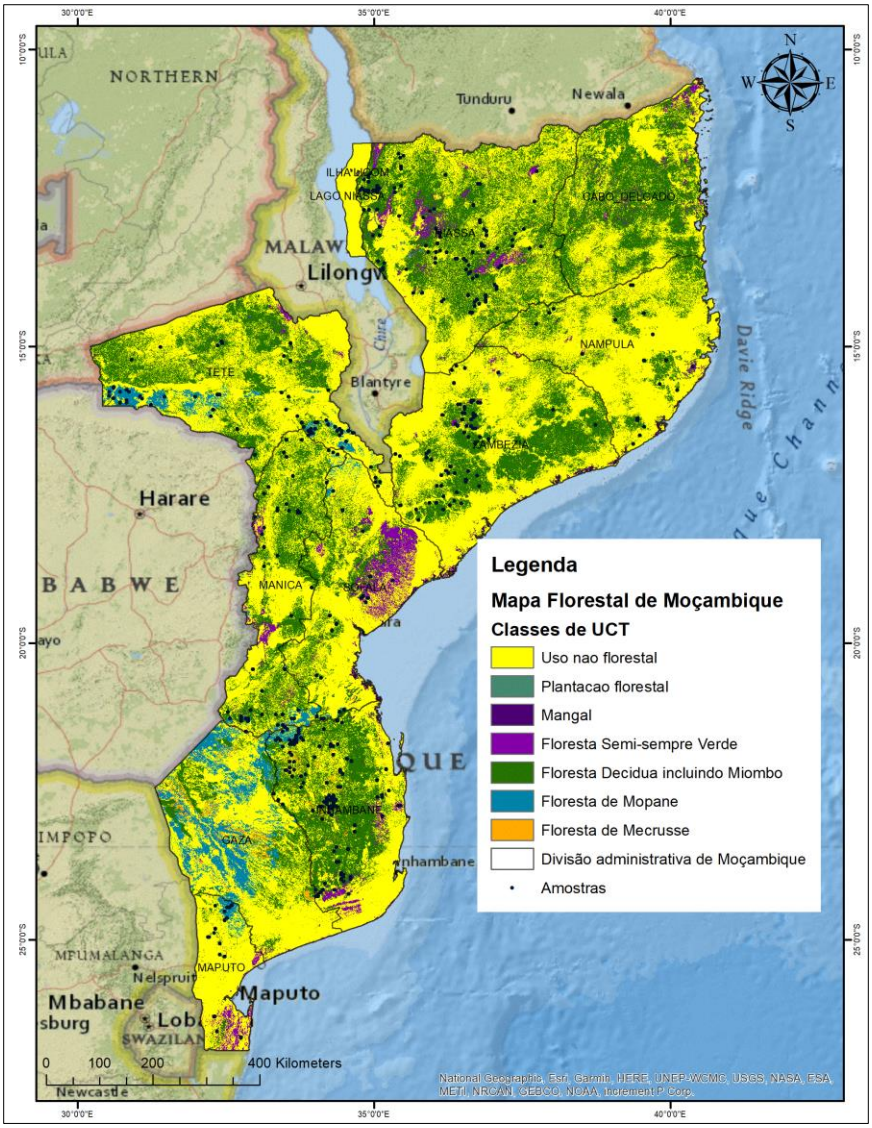


Figura 2: Distribuição das unidades amostrais nos tipos florestais

Fonte: IV Inventário Florestal Nacional de Moçambique.

O tamanho da amostra foi estimado usando a alocação óptima considerando um erro relativo de amostragem de 10% para o volume total e um nível de significância de 5% e dentro dos estratos o tamanho das unidades amostrais foi alocado usando a alocação proporcional (MITADER, 2018a). Do total das unidades amostrais do inventário nas oito províncias (2260), apenas 1720 amostras

(tabela 2) mostraram-se viáveis para fazer parte deste estudo, o resto foram descartadas por apresentarem problemas relacionadas com o rastreio das mesmas.

Tabela 2: Número de amostras viáveis por cada tipo florestal

Tipos florestais	**Área (ha)**	**Parcelas viáveis (n_h)**
Floresta decídua	20 682 838,00	845
Floresta sempre verde	6 989 275,00	420
Floresta de Mecrusse	843 213,00	190
Floresta de Mopane	3 178 546,00	264
Total	31 693 872,00	1 720

Fonte: IV Inventário Florestal Nacional de Moçambique.

Para o levantamento de dados de campo foi usado o cluster como unidade amostral, que é um aglomerado de 4 parcelas. O cluster usado foi de formato quadrado (100 m x 100 m) composto por parcelas rectangulares (20 m x 50 m) localizadas nos quatro quadrantes do cluster e separadas uma da outra por uma distância de 40 m. Por sua vez, cada parcela foi composta por 4 sub-parcelas de 25 m x 10 m de dimensões (A, B, C e D), também denominados de blocos (Figura 3) (MITADER, 2018b).

Dados de campo mostraram que o estrato transcendia a unidade amostral, ou seja, foi encontrado mais de um estrato (tipo florestal) em cada cluster, facto que fez com que a parcela fosse considerada como unidade amostral para efeitos de analises.

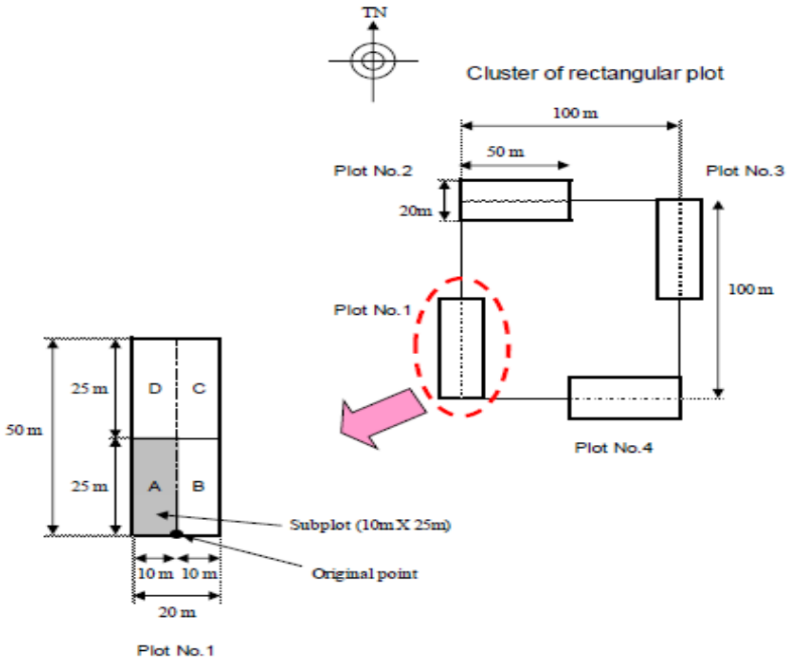


Figura 3: Ilustração da unidade amostral (cluster)

Fonte: IV Inventário Florestal Nacional de Moçambique.

3.2.2. Levantamento de dados (carbono orgânico do solo)

Para a estimativa do carbono orgânico do solo foram colectados dois tipos de amostras de solo, amostras perturbadas e não perturbadas, ambas obtidas a uma profundidade máxima de 30 cm. A escolha dessa profundidade se deve ao facto de ser onde existe maior concentração de carbono orgânico, que resulta da maior actividade biológica nesse intervalo de profundidade (Ravindranath & Ostwald, 2007).

As amostras perturbadas e não perturbadas foram obtidas no centro de cada parcela em número de uma para cada uma parcela (Figura 4) (MITADER, 2018b).

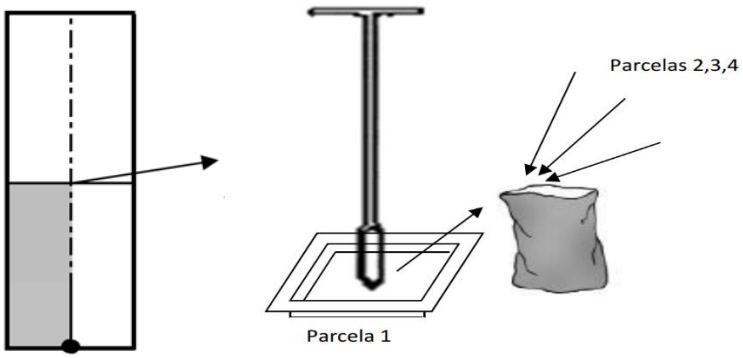


Figura 4: Unidade amostral para colecta de amostra de carbono orgânico do solo

Fonte: IV Inventário Florestal Nacional de Moçambique.

As amostras perturbadas foram usadas para determinar os teores de carbono orgânico do solo e as amostras não perturbadas foram usadas para determinar a densidade aparente do solo e a fracção de massa do fragmento mineral grosseiro, tudo isso ao nível do laboratório (Pearson *et al*., 2005). Como forma de facilitar o rastreio durante e após as análises laboratoriais, todas as amostras foram identificadas por um código específico.

3.2.3. Processamento e Análise de dados

3.2.3.1.Determinação dos Teores do Carbono Orgânico do Solo

Os teores de carbono orgânico do solo foram determinados ao nível do laboratório de solo da Faculdade e Agronomia e Engenharia Florestal da Universidade Eduardo Mondlane. Para a determinação do teor de carbono orgânico do solo recorreu-se ao método de combustão húmida (*Walkley-Black*), que consiste na reacção de oxirredução do dicromato ($Cr_2O_7^{2-}$) com a matéria orgânica do solo em meio do ácido sulfúrico, seguido da titulação do dicromato remanescente com uma solução de Fe (II) (Fontana & de Campos, 2017). Antes da aplicação do método de combustão húmida, as amostras foram submetidas ao pré-tratamento, e este pré-tratamento consistiu na secagem das amostras, destorroamento das amostras e retirada de partículas grosseiras com diâmetro maior que 2mm, e pesagem das amostras. O pré-tratamento foi efectuado para permitir que apenas o carbono medido seja apenas da fração orgânica do solo. Para a estimativa do carbono orgânico do solo por unidade de área recorreu-se a expressão abaixo (FAO, 2018).

$$COS\ [ton.C/ha) = \mathrm{OC}i \times \mathrm{BD}i \times (1 - \mathrm{gG}i) \times ti \times 0.1 \qquad (1)$$

Onde: *COSi [tonC/ha]* é o stock de carbono orgânico do solo expresso em toneladas por unidade de área (hectare) observados na unidade amostral i; OC_i é o teor de carbono orgânico do solo determinado ao nível do laboratório observado na unidade amostral i expresso em percentagem; BD_i é a massa do solo por volume total da amostra do solo (densidade do solo) observado na unidade amostral i, expresso em g/cm³; gG_i é a fracção de massa do fragmento mineral grosseiro observado na unidade amostral i; t_i é a profundidade da obtenção das amostras de solo na unidade

amostral i, e para este este estudo foi usado uma profundidade de 30 cm; e 0,1 é o factor de conversão de mg C/cm para tonC/ha.

3.2.3.2.Análise exploratória dos dados

Para a presente pesquisa, os dados foram primeiramente submetidos a uma análise exploratória, objectivando detectar possíveis anomalias (valores extremos, inconsistências ou erro de digitação) e assim proceder-se com possíveis correções caso se aplicasse, o que permite uma melhor compressão do comportamento dos mesmos (Favero & Belfiore, 2017; Frost, 2019). A análise exploratória compreendeu a realização das estatísticas univariada e bivariada, essa última usada na fase do ajuste do modelo de predição de carbono orgânico do solo.

A condução da estatística univariada visou representar o comportamento da variável carbono orgânico do solo por meio de seus valores centrais e não centrais, suas dispersões ou formas de distribuição dos seus valores em torno da média. A estatística univariada foi avaliada através da estatística descritiva, ou seja, através da média, amplitudes, separatrizes, boxplot e histograma (Favero & Belfiore, 2017; Frost, 2019).

As figuras 5 e 6 ilustram a distribuição (box-plot, histograma) dos dados originais do estudo (carbono orgânico do solo).

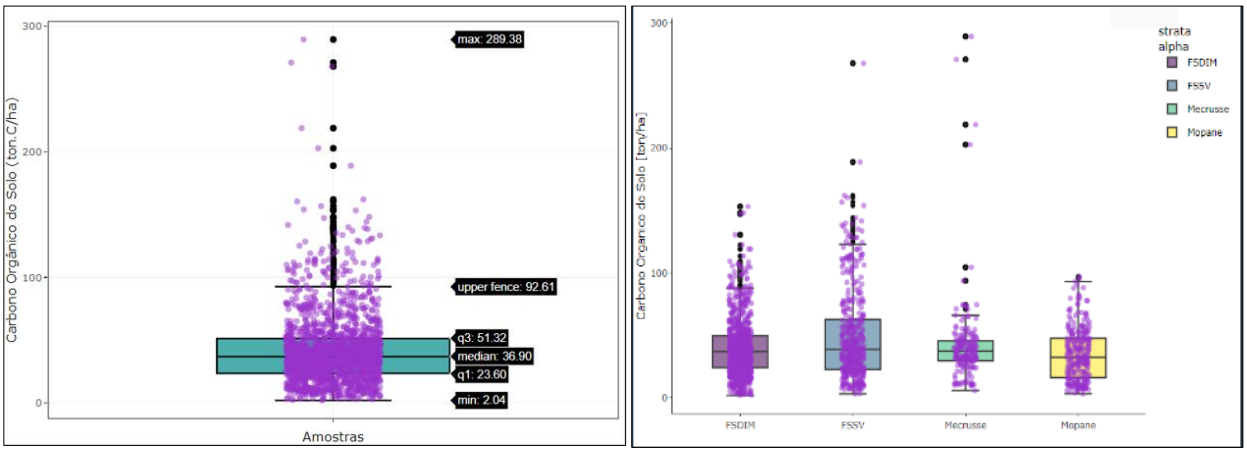


Figura 5: Separatrizes e boxplot de COS do grupo (esquerda) e subgrupos (direita) dos dados originais

Fonte: Presente estudo.

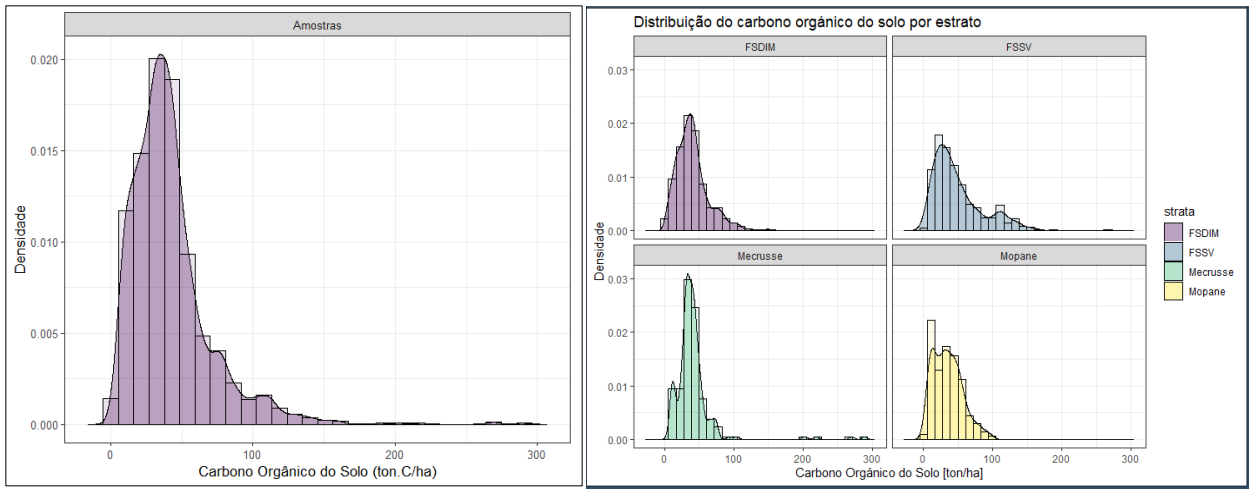


Figura 6: Histograma de COS do grupo (esquerda) e subgrupos (direita) dos dados originais

Fonte: Presente estudo.

A estatística possui dois métodos de análises de dados que são os testes paramétricos e os testes não paramétricos (livres de distribuição ou pressupostos). A escolha de um método em detrimento do outro passa pela satisfação de certos pressupostos e para o caso específico dos testes paramétricos a sua aplicação implica a satisfação dos pressupostos de independência dos dados, aleatoriedade, amostra retirada de populações com determinada distribuição (normalidade e homocedasticidade dos dados) e variáveis medidas em escalas que permitam operações aritméticas (Favero & Belfiore, 2017; Sheskin, 2003).

Tendo em conta que os dados deste estudo são independentes e aleatórios (desenho amostral), os resíduos dos dados originais foram submetidos apenas a verificação dos pressupostos de normalidade (*Shapiro-Francia: n>30 samples*) e homocedasticidade (*Brensch-Padan*). Resultados do teste de *Shapiro-Francia* e *Brensch-Padan* (Figura 8) mostraram uma evidente violação dos pressupostos de normalidade e heterocedasticidade para ANOVA entre tipos florestais.

A figura 7 ilustra os resultados da avaliação da satisfação dos pressupostos de ANOVA da variável carbono orgânico do solo, considerando o estrato florestal, dos resíduos dos dados originais provenientes do campo. Os resultados mostrados são o histograma dos resíduos do modelo de ANOVA, e testes de aderência de normalidade de *Shapiro-Francia* e heterocedasticidade de *Breusch-Pagan*.

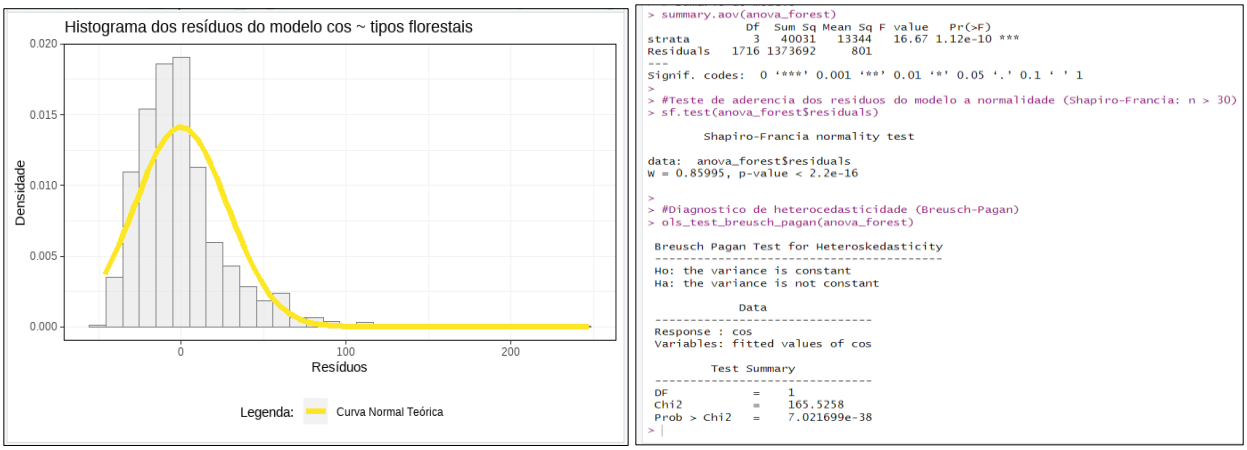


Figura 7: Histograma e testes de verificação dos pressupostos de ANOVA dos resíduos dos dados originais

Fonte: Presente estudo.

Para se ultrapassar o problema da, não normalidade e heterocedasticidade e assim seguir-se com os testes paramétricos (anova e teste post hoc), os dados foram submetidos a transformação Box-Cox, objectivando induzir a normalidade e homocedasticidade dos mesmos (Box & Cox, 1982; Mangiafico, 2015). Os dados transformados foram novamente submetidos a análise exploratória.

As figuras 8 e 9 ilustram a distribuição (box-plot, histograma) dos dados transformados (carbono orgânico do solo). Com base nos gráficos observa-se uma ligeira melhoria da distribuição dos dados, ou seja, há redução dos *outliers* e os gráficos visualmente aproximam-se mais da distribuição normal teórica.

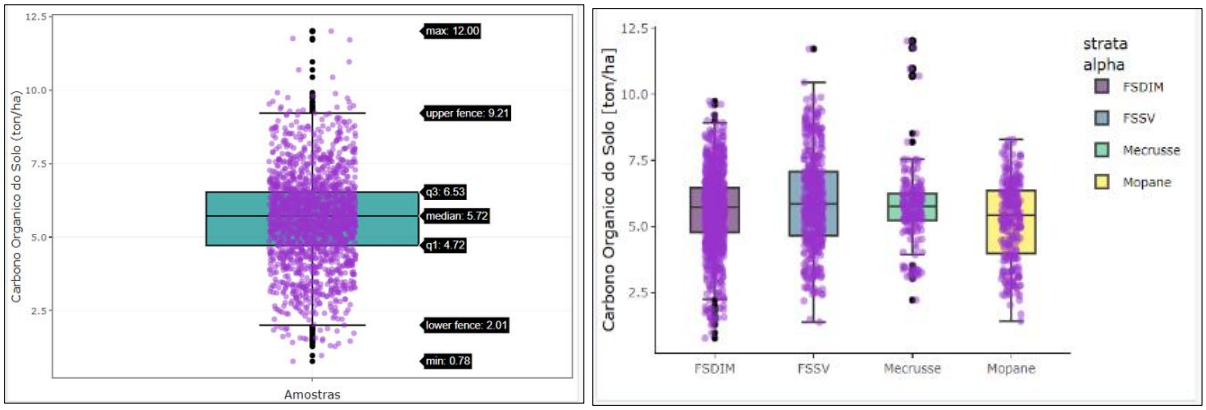


Figura 8: Separatrizes e boxplot de COS do grupo (esquerda) e subgrupos (direita) dos dados transformados (Box-Cox)

Fonte: Presente estudo.

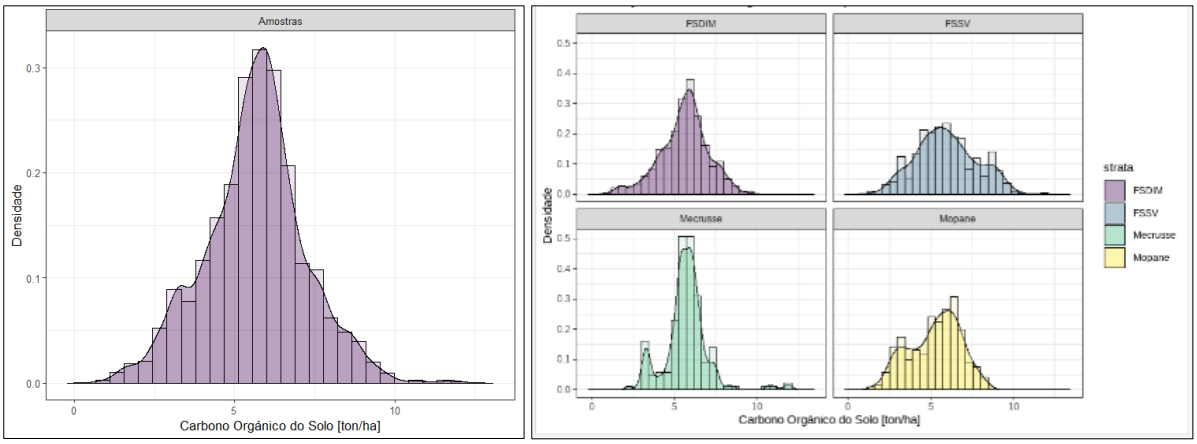


Figura 9: Histograma de COS do grupo (esquerda) e subgrupos (direita) dos dados transformados (Box-Cox)

Fonte: Presente estudo.

A semelhança dos dados originais, os resíduos dos dados transformados foram também submetidos a verificação dos pressupostos de normalidade (*Shapiro-Francia*) e homocedasticidade (*Brensch-Padan*). Resultados dos testes mostraram uma violação dos dois pressupostos para o modelo de ANOVA (Figura 10).

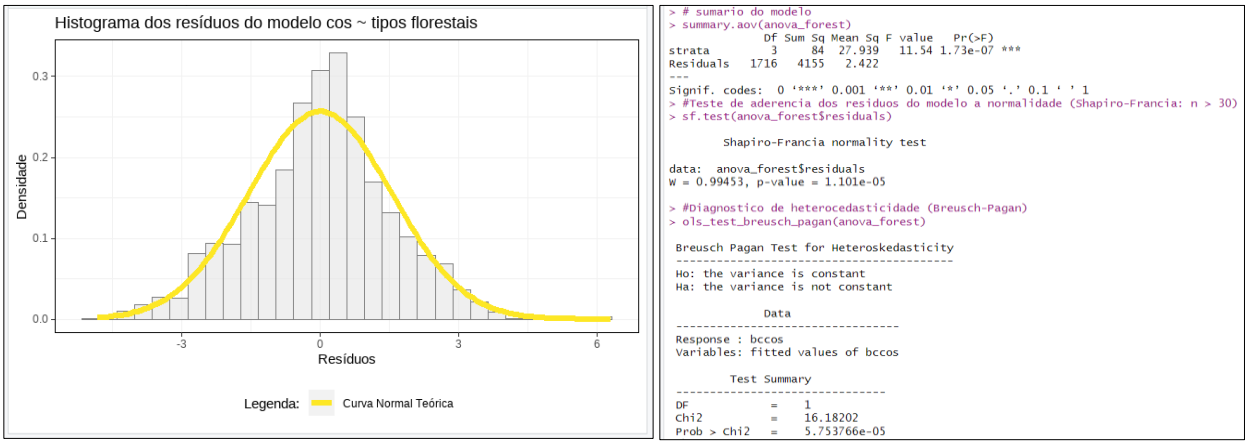


Figura 10: Histograma e testes de verificação dos pressupostos de ANOVA dos resíduos dos dados transformados (Box-Cox)

Fonte: Presente estudo.

À primeira vista, a violação dos pressupostos sugere a utilização de testes não-paramétricos por se tratarem de testes que não dependem do pressuposto de normalidade dos dados (livre de pressuposto de distribuição), equivalentes aos testes paramétricos e neste caso tratando-se de um

estudo com mais de duas variáveis, a anova de um factor deve ser substituído pelo teste de *Kruskal-Wallis* que é o seu equivalente (Field, 2018; Mangiafico, 2015; Trawinski *et al*., 2012).

Embora seja uma alternativa para o problema de não normalidade e heterocedasticidade, os testes não-paramétricos são tidos por alguns como sendo menos poderosos que os testes paramétricos e isso deve-se ao facto de muitos serem testes que se baseiam na mediana (*Kruskal-Wallis*) e cujos dados originais são transformados em dados ordenados (*rank*) condicionando a perda de informação original (Field, 2018; Mangiafico, 2015; Mcdonald, 2014; Sheskin, 2003; Zwick & Marascuilo, 1984).

O facto de os testes não-paramétricos serem menos robustos (nunca virão a ser melhor que os paramétricos), condicionou em parte a escolha dos testes paramétricos em detrimento dos não paramétrico para análise estatística dos dados deste estudo (Field, 2018). Este facto foi mostrado em diversas simulações executadas por Mcdonald (2014), em uma variedade de dados de distribuição anormal na qual encontrou que a proporção de falsos positivos é sempre em torno de 0,05 ou um pouco menor, exactamente como deveria ser, tendo o autor recomendado a utilização da ANOVA de um factor em detrimento do teste de *Kruskal-Wallis*, que por sinal, pode vir a ser pior que ANOVA quando os dados também são heterocedásticos.

Outra razão da escolha dos testes paramétricos (ANOVA, post hoc) deve-se ao facto destes não serem extremamente sensíveis aos desvios dos pressupostos (os dados não precisam ser perfeitamente normais e homocedásticos), pese embora sempre haja necessidade de cautela na rejeição da hipótese nula se o valor de *P* não estiver muito abaixo de 0,05 (Mcdonald, 2014).

Para além das razões acima citadas, simulações feitas pela equipa da MINITAB mostraram que, os testes paramétricos podem apresentar um bom desempenho com distribuições assimétricas e não normais quando se trabalha com dados contínuos que não sejam normais, desde que o tamanho da amostra seja suficientemente grande para cada grupo (Minitab, 2023).

O desempenho é explicado pelo teorema do limite central (Sheskin, 2003; Trawinski *et al*., 2012) que afirma que, a medida que o tamanho amostral aumenta, a distribuição da média segue uma distribuição normal, mesmo que a distribuição subjacente da variável original não seja normal (Frost, 2019). Assim, optou-se em usar os testes paramétricos (mais apropriados) devido ao seu

poder estatístico, fora o facto de serem mais sensíveis, principalmente se tratando de um estudo em que a amostra em cada grupo é suficientemente grande (n > 100) (Sheskin, 2003; Trawinski *et al*., 2012).

3.2.3.3.Comparações estatísticas

Para efeitos da análise das diferenças estatísticas entre os grupos (estratos florestais), recorreu-se ao teste de análise de variância de um factor (*anova one-way*) à 5% de significância, que é indicado para dados observacionais (amostras não repetidas) e de um único factor (Mangiafico, 2015; Mcdonald, 2014; Sheskin, 2003).

Embora a violação dos pressupostos (Figura 8 e 11), a anova foi utilizada porque ela admite uso de dados que não sejam perfeitamente normais e homocedásticos (Mcdonald, 2014) e para procurar aproximar os dados a uma distribuição normal, os mesmos foram submetidos a transformação Box-Cox e visivelmente mostraram apresentar melhor aproximação a distribuição normal que os dados originais (Figura 11).

Da análise de variância obteve-se um valor de probabilidade muito abaixo de 0,05 (diferenças significativas entre os tratamentos), o que em termos estatísticos mostra menos preocupação com relação a probabilidade de se cometer o erro do tipo I (rejeitar a hipótese nula quando ela é verdadeira) que poderia surgir caso o valor de *P* não estivesse muito abaixo de 0,05. Dado que, a anova mostrou diferenças estatísticas significativas entre os grupos (estratos), o teste de post hoc de comparação múltipla de *TukeyHSD* (diferenças significativas honestas) à 5% de significância foi utilizado para identificar quais tratamentos diferiam entre si.

A escolha do teste de *TukeyHSD* deve-se ao facto de ser o teste de post hoc recomendado quando o teste de anova evidencia a existência de diferenças estatísticas significativas entre os tratamentos, pois, o mesmo mostrou melhor desempenho no controlo do erro (Zwick & Marascuilo, 1984).

Quando se procede com os testes de comparações múltiplas, normalmente corre-se o risco de rejeitar incorrectamente pelo menos uma hipótese nula (erro do tipo I), risco esse que aumenta com o aumento de número de comparações (Mcdonald, 2014) e para controlar esse problema

existem na literatura diversos métodos de ajuste do valor de *p* que ajudam a reduzir a propagação do erro do tipo I.

Para corrigir os valores de *p* produzidos pelo teste de comparação múltipla de *TukeyHSD* (evitar a inflação do erro tipo I), foi aplicado o método de correcção de Benjamim & Hochberg "*BH* ou *fdr*" e a escolha deste método em detrimentos dos demais deve-se ao facto de ser o método mais poderoso quando comparados aos demais (Trawinski et al., 2012).

3.2.3.4.Estimativa de modelo para predição de carbono orgânico do solo

Para o estudo, considerou-se variáveis preditoras para o ajuste do modelo, todas as variáveis ambientais com potencial de influenciar ou explicar o carbono orgânico do solo (variável dependente), que é a variável resposta (Lal, 2021). As variáveis ambientais foram selecionadas recorrendo o conhecimento dos processos biofísicos (permitem a consideração e seleção de preditores ambientais), conhecimento dos factores eco-fisiológicos e a disponibilidade e experiência de que as variáveis independentes se correlacionam com as variáveis dependentes, neste caso o carbono orgânicos do solo (Austin, 2007; Bird *et al*., 2001). Foi também usada a análise de correlação (coeficiente de correlação de Pearson), cujo objectivo foi de avaliar o grau de associação entre as variáveis (Favero & Belfiore, 2017).

Foram eleitas como variáveis ambientais (variável dependente) candidatas a fazerem parte dos modelos, 17 variáveis, que são: biomassa abaixo do solo (BGB), biomassa acima do solo (AGB), volume total de madeira (VT), densidade árborea (Nha), área basal (Gha), precipitação média anual (PMA), radiação solar (RS), aspecto, altitude, declive, evapotranspiração, humidade do solo (HS), areia, argila, limo, tipo de solo, e tipos florestais. As variáveis biomassa, área basal e volume foram obtidos do IV Inventário Florestal Nacional de Moçambique (MITADER, 2018a); os teores de argila, limo e areia são originários do presente estudo; as variáveis altitude, aspecto e declive são produtos do modelo digital de elevação (ASTER DEM), obtido no site da NASA; os dados de PMA, radiação solar e humidade do solo foram também obtidos no site da NASA; os dados de solos foram obtidos no site da FAO; e os dados dos tipos florestais foram obtidos no site da unidade de MRV do FNDS. Os graus de associação com o carbono orgânico do solo são apresentados na tabela 3 abaixo, que ilustra o grau de associação entre carbono orgânico do solo e as variáveis

explicativas quantitativas. Os valores entre parenteses são negativos, o que significa que a correlação entre as duas variáveis é negativa; por sua vez, o * representa a significância estatística da correlação.

Tabela 3: Correlações entre as variáveis preditoras quantitativas e a variável dependente

Variável dependente	**Variável explicativa (covariantes)**	**Unidade**	**r (coeficiente de correlação)**	**Valor do P (*p-value*)**
Carbono orgânico do solo	Teor de areia	%	(0.21)	0.00*
	Teor de argila	%	0.12	0.00*
	Teor de limo	%	0.38	0.00*
	Densidade arbórea	Arv/ha	(0.16)	0.00*
	Área basal	m^2/ha	0.10	0.00*
	Volume total	m^3/ha	0.16	0.00*
	Biomassa abaixo do solo	ton./ha	0.19	0.00*
	Biomassa acima do solo	ton./ha	0.21	0.00*
	Aspecto	º	0.08	0.05
	Altitude	m	0.03	1.00
	Declive	%	0.41	0.00*
	Evapotranspiração	mm/ano	(0.20)	0.00*
	Humidade do solo	%	(0.06)	0.29
	Precipitação média anual	mm	0.25	0.00*
	Radiação solar	W/m²	(0.19)	0.00*

*Nota: *p<0,05*

Fonte: Presente estudo.

As variáveis tipo do solo e tipos florestais por se tratarem de variáveis qualitativas não foi possível a análise da correlação com o carbono orgânico do solo, assim, a sua inclusão como variáveis candidatas do modelo foi baseando-se no resultado da análise de variância (*anova one-way*), cujos valores de *p-value* foram todos inferiores a ao nível de significância de 5%, mostrando que pelo menos uma das categorias inclusas nas variáveis tem influencia no carbono orgânico do solo.

Antes do processo de estimativa do modelo de predição do carbono orgânico do solo, as amostras foram divididas de forma aleatória em amostras de treino e amostras de validação (Figura 11). As amostras de treino correspondem a 75% total dos dados (n=1290) e as mesmas foram usadas para a estimativa do modelo de predição de COS, enquanto as amostras de validação que correspondem

os remanescentes 25% (n=430)foram usadas para avaliar a performance do modelo estimado (FAO, 2018; Zhang *et al*., 2022). A divisão da amostra em amostra de treino e amostra de treino foi feita de forma aleatória através do software estatísticos *R-statistic*.

A figura 11 abaixo ilustra o gráfico de densidade das amostras de treino e amostras de validação da variável carbono orgânico do solo. As estatísticas descritivas das amostras de treino e amostras de validação para as variáveis preditoras e dependentes são apresentadas nas tabelas 4 e 5 abaixo.

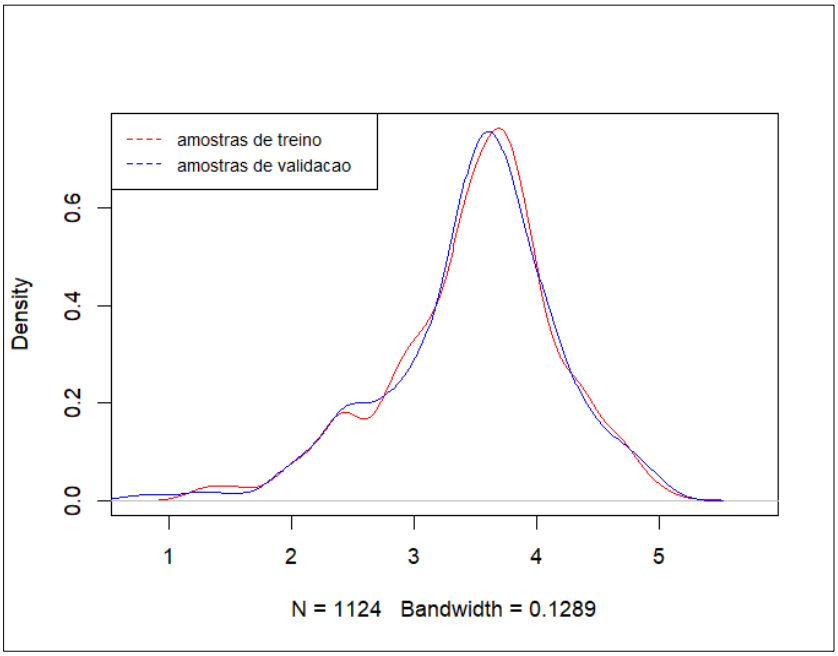


Figura 11: Gráfico de densidade das amostras de treinamento e validação do COS

Fonte: Presente estudo.

As tabelas 4 e 5 abaixo ilustram o resumo da estatística descritiva das variáveis preditoras e dependentes (COS).

Tabela 4: Estatística descritiva do COS e suas covariantes da amostra de treino

Parâmetro	**COS**	**Areia**	**Argila**	**Limo**	**Densidade**	**Área basal**	**Vol. total**	**AGB**	**BGB**	**Aspecto**	**Altitude**	**Declive**	**Evapotranspiração**	**Humidade do solo**	**Radiação solar**	**PMA**
	ton/ha	%	%	%	N/ha	m_2/ha	m^3/ha	ton/ha	ton/ha	º	m	%	mm/ano	%	W/m^2	mm/ano
Mínimo	2,4	0	0	0	134,7	1,1	4,4	4,1	2,7	(1,0)	16,0	0	1422,0	13,0	16365,0	536,0
1st quartil	23,6	68,8	9,8	3,4	425,3	8,4	47,0	49,4	17,8	66,04	133,0	0,6	1558,0	30,0	17147,0	691,0
Mediana	37,0	78,1	11,7	8,2	551,2	10,8	69,6	67,9	24,0	160,5	246,0	1,4	1593,0	34,0	17575,0	895,0
Média	41,4	73,6	13,0	10,1	607,8	11,4	74,8	74,0	25,1	167,9	376,7	2,7	1597,0	38,6	17598,0	963,2
3st quartil	51,1	84,9	15,5	14,1	744,3	13,3	88,8	92,6	30,1	258,7	630,2	2,9	1630,0	43,0	17877,0	1159,2
Máximo	218,8	91,5	38,8	54,6	1792,5	25,9	246,9	211,2	68,7	355,8	1457,0	28,8	1784,0	97,0	19322,0	1974,0

Fonte: Presente estudo.

Tabela 5: Estatística descritiva do COS e suas covariantes da amostra de validação

Parâmetro	**COS**	**Areia**	**Argila**	**Limo**	**Densidade**	**Área basal**	**Vol. total**	**AGB**	**BGB**	**Aspecto**	**Altitude**	**Declive**	**Evapotranspiração**	**Humidade do solo**	**Radiação solar**	**PMA**
	ton/ha	%	%	%	N/ha	m2/ha	m3/ha	ton/ha	ton/ha	º	m	%	mm/ano	%	W/m^2	mm/ano
Mínimo	3,3	0	0	0	134,9	1,1	4,4	4,1	2,7	(1,0)	31,0	0	1422,0	13,0	16545,0	536,0
1st quartil	23,6	69,4	9,6	3,2	430,5	8,2	47,7	46,6	18,0	81,7	129,0	0,6	1558,0	30,0	17138,0	695,0
Mediana	37,0	78,0	11,8	8,2	557,3	11,0	69,5	67,6	24,0	159,4	221,0	1,2	1593,0	37,0	17591,0	862,0
Média	41,9	73,5	12,7	9,9	625,6	11,2	73,3	73,4	25,2	168,5	358,9	2,4	1596,0	39,2	17606,0	946,3
3st quartil	51,3	84,5	15,0	14,0	770,0	13,4	87,1	91,1	30,4	258,7	608,0	2,5	1630,0	43,0	17891,0	1122,0
Máximo	202,8	91,5	38,8	54,8	1760,3	27,5	251,0	211,2	70,7	359,3	1223,0	23,8	1784,0	73,0	19322,0	1968,0

Fonte: Presente estudo.

Para a estimativa do modelo de "*random forest*" para a predição do carbono orgânico do solo partiu-se do modelo teórico de *random forest* (FAO, 2018; Zhao & CenYonghua, 2014) representado pela equação 2. A escolha do "*random forest*" deve-se ao facto de ter-se mostrado mais eficiente na predição de carbono orgânico do solo nas formações florestais de Moçambique (João, 2024).

$$f(x) = \frac{1}{\mathrm{B}}\sum_{b=1}^{B} T_b(x) \qquad (2)$$

Onde: x é o vector de variáveis explicativas (altitude, vegetação, etc.); $T_b(x)$ é a previsão da árvore b; e $f(x)$ é a previsão final do Random Forest (valor continuo).

Uma vez definido o modelo teórico, fez-se o ajuste dos parâmetros do modelo, cujo objetivo foi encontrar a combinação de hiperparâmetros que pudessem produzir melhor desempenho do modelo (evitar "*overfitting*" e sob ajuste), ou seja, encontrar o melhor parâmetro de "*mtry*" e decidir o número apropriado de árvores para o ajuste do modelo, e isso fez-se recorrendo a estratégia de validação cruzada com um padrão de 10 divisões ("*10-fold cross-validation*") da amostra de treino (FAO, 2018), e isso foi feito recorrendo a biblioteca "*caret*" do "*R-statistic*" (FAO, 2018; Piri Sahragard & Pahlavan-Rad, 2020; Zhao & CenYonghua, 2014).

3.2.3.5.Modelagem do carbono orgânico do solo

O modelo de RF estimado foi a posterior usado em conjunto com as variáveis preditoras para fazer a predição do COS em locais não amostrados a quando da atividade de campo, e deste processo resultou um mapa com a distribuição espacial do COS nos ecossistemas florestais em toda área de estudo (FAO, 2018). Embora o modelo tenha sido estimado considerando a validação do mesmo através da estratégia da validação-cruzada, houve necessidade de uma validação independente do mapa produzido (FAO, 2018), o que foi feito na fase de comparação da performance dos algoritmos na estimativa do COS.

3.2.3.6.Validação das predições de COS

Dado que todo o processo de estimativas está associado a certo nível de incertezas estatísticas, houve necessidade de se validar as estimativas de COS no mapa produzido e isso foi realizado através da análise da performance da modelagem de COS.

A validação da modelagem é um processo em que valores estimados do mapa são comparados com os obtidos no campo e para tal é usado um conjunto de dados independentes (não usados na estimativa do modelo) (Bangroo *et al*., 2020; Bhunia *et al*., 2019; FAO, 2018; Hengl *et al*., 2004; Masson-Delmotte *et al*., 2019; Mondal *et al*., 2016; Yang *et al*., 2021) e para este estudo essa amostra independente foi de 25%, conforme proposto por FAO (2018).

Para a avaliação da performance da modelagem foram usados os seguintes parâmetros estatísticos: o erro médio (ME), raiz quadrada do erro médio (RMSE), erro médio absoluto (MAE), coeficiente de determinação ajustado (Bangroo *et al*., 2020; Bhunia *et al*., 2019; FAO, 2018; Hengl *et al*., 2004; Masson-Delmotte *et al*., 2019; Mondal *et al*., 2016; Yang *et al*., 2021).

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |X_i - Y_i| \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n}} \sum_{i=1}^{n} (X_i - Y_i)^2 \quad (4)$$

$$R^2 = \frac{\sum_{i=1}^{n} (X_i - \hat{Y}_i)^2}{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2} \quad (5)$$

Onde: *Xi* e *Yi* são valores de carbono orgânico de solo estimados e observados, respectivamente; MAE é erro médio absoluto; RMSE é a raiz do erro médio quadrático; R^2 é o coeficiente de determinação e *n* é o número de amostras.

IV. RESULTADOS E DISCUSSÃO

4.1. Stocks de Carbono Orgânico do Solo

4.1.1. Estatística descritiva dos dados

A Tabela 6 abaixo ilustra o resumo da estatística univariada da variável carbono orgânico do solo (COS), densidade de solo (ρb) e teor de carbono (C) determinado ao nível do laboratório. Os dados são apresentados por tipo florestal e população dos dados colectados ao nível de estudo. Com base na tabela, observou-se uma variação significativa (amplitude de 287,34 tonC/ha, e desvio padrão de 28,68 tonC/ha) dos valores observados ao nível do campo.

Tabela 6: Estatística descritiva dos dados colectados

Parâmetro estatístico	População			Floresta decídua			Floresta sempre verde			Floresta de Mecrusse			Floresta de Mopane		
	ρb (Kg/m3)	*C (%)*	*COS (ton/ha)*	*ρb (Kg/m3)*	*C (%)*	*COS (ton/ha)*	*ρb (Kg/m3)*	*C (%)*	*COS (ton/ha)*	*ρb (Kg/m3)*	*C (%)*	*COS (ton/ha)*	*ρb (Kg/m3)*	*C (%)*	*COS (ton/ha)*
Média	1379,06	1,20	41,48	1,351.77	1.16	40.15	1,296.42	1.56	49.11	1,532.80	1.01	39.47	1,408.85	0.91	35.08
Erro padrão	5,85	0,06	0,65	10.80	0.06	0.80	13.92	0.12	1.74	11.39	0.13	1.76	15.41	0.07	1.29
Mediana	1392,00	0,87	36,83	1,376.71	1.04	37.04	1,298.50	1.17	38.82	1,551.71	0.88	37.48	1,435.45	0.82	32.49
Moda	1543,00	0,91	47,34	-	-	-	-	-	-	-	-	-	-	-	-
Desvio padrão	243,49	1,29	26,97	271.84	0.71	23.30	255.22	1.25	35.61	150.71	0.92	23.96	243.11	0.57	20.95
Variância	59285,78	1,66	727,61	73,898.36	0.50	543.10	65,139.22	1.57	1,267.84	22,714.86	0.85	574.30	59,102.76	0.32	438.81
Curtose	0,87	13,89	4,20	0.84	0.44	2.19	0.55	6.54	0.85	-0.44	30.66	27.42	0.03	0.08	-0.07
Assimetria	-0,57	3,30	1,63	-0.58	0.88	1.19	-0.04	1.98	1.20	-0.27	5.06	4.14	-0.41	0.75	0.64
Amplitude	1814,00	9,81	216,75	1,814.00	3.25	151.22	1,653.00	8.00	185.60	691.33	6.48	212.85	1,364.69	2.32	93.82
Mínimo	362,00	0,09	2,04	362.00	0.09	2.04	529.00	0.14	3.30	1,200.67	0.12	5.94	631.31	0.10	3.40
Máximo	2176,00	9,90	218,79	2,176.00	3.34	153.26	2,182.00	8.14	188.90	1,892.00	6.60	218.79	1,996.00	2.42	97.21

Fonte: Presente estudo.

Embora a média em casos de amplitudes muito altas não represente melhor a população, a mediana encontrada se aproxima da média da população dos dados levantados, isto porque ela se aproxima do valor médio encontrado, o mesmo em relação a moda.

4.1.2. Stocks de Carbono Orgânico do Solo Por Tipo Florestal

A análise dos teores de carbono orgânico do solo considerando os tipos florestais usados no IV Inventário Florestal Nacional de Moçambique (Miombo, floresta sempre verde, Mopane e Mecrusse), mostrou haver variação dos teores médios de COS entre as formações florestais. A floresta Sempre Verde apresentou o maior teor médio de carbono orgânico do solo (49,11 tonC/ha), seguindo pela floresta decídua (40,15 tonC/ha), floresta de Mecrusse (39,36 tonC/ha) e por último na floresta de Mopane (35,08 tonC/ha), conforme ilustrado na tabela 7 abaixo.

A tabela ilustra os teores médios de carbono orgânico do solo (COS) e suas respectivas incertezas (erro padrão da média e intervalo de confiança) por tipo florestal. O COS está expresso em toneladas por unidade de área (tonC/ha). O intervalo de confiança foi determinado a um nível de significância de 5%, o equivalente a 95% de confiança.

Tabela 7: Teores médios de carbono orgânico do solo (COS) por tipo de floresta

Tipo Florestal	**Média [tonC/ha]**	**Erro padrão [tonC/ha]**	**Intervalo de confiança [+/- tonC/ha]**
Floresta decídua	40,15	0,80	38,58 – 41,72
Floresta sempre verde	49,11	1,74	45,7 – 52,52
Mecrusse	39,36	1,74	35,95 – 42,77
Mopane	35,08	1,29	32,55 – 37,61
Média Nacional	**41,78**	**0,69**	**40,43 – 43,13**

Fonte: Presente estudo.

O teor de carbono orgânico do solo encontrado na floresta Sempre Verde mostrou ser superior em 18%, 20% e 29% em relação ao encontrado na floresta decídua (incluindo Miombo), Mecrusse e Mopane, respectivamente, facto validado pela análise de variância (ANOVA) que evidenciou a existência de pelo menos uma diferença estatística entre as médias do carbono orgânico do solo entre os tipos florestais (*p-value* < 0,05).

Comparações estatísticas através do teste de comparação de médias de *TukeyHSD* (*Honest Significant Differences*) à 5% de significância, mostraram não haver diferenças estatísticas significativas da média de carbono orgânico do solo entre a floresta Sempre Verde e a floresta de

mecrusse, e entre a floresta decídua incluindo o Miombo e a floresta de Mecrusse, pese embora que a floresta Sempre Verde se tenha mostrado superior em 20% em relação a floresta de mecrusse.

Diferenças estatísticas significativas da média de carbono orgânico do solo foram encontradas entre a floresta Sempre Verde e as florestas de decíduas incluindo o Miombo e Mopane; diferenças estatísticas significativas da média de carbono orgânico do solo entre a floresta decídua e a floresta de Mopane, e diferenças estatísticas significativas da média de carbono orgânico do solo entre a floresta de Mecrusse e a floresta de Mopane, conforme ilustrado na figura 12 abaixo.

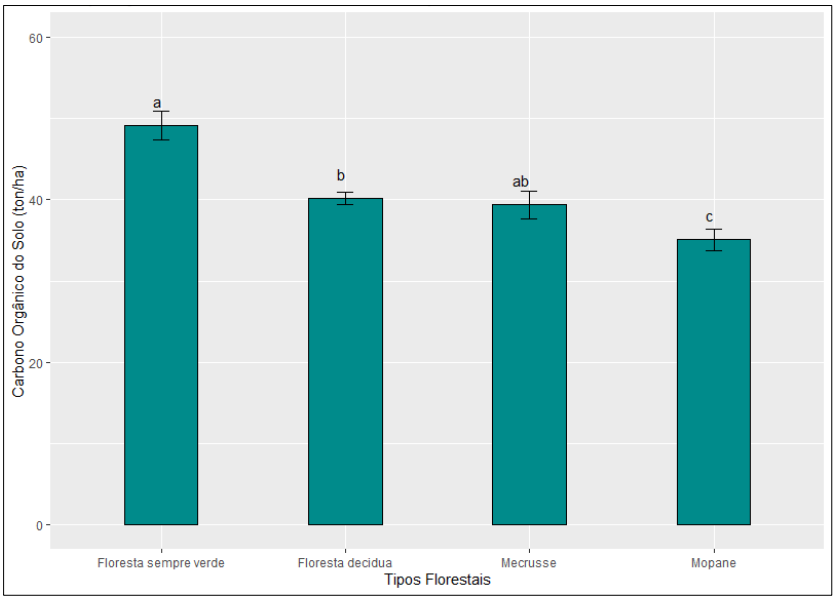


Figura 12: Comparações estatísticas dos teores médios de carbono orgânico do solo entre os estratos florestais através do teste de *TukeyHSD* com ajuste de BH

Fonte: Presente estudo.

As diferenças dos teores médios entre os teores médios de carbono orgânico do solo entre as formações florestais podem estar associadas a diferença na taxa de incorporação da matéria orgânica morta (folhas, galhos, flores, frutos, cascas) e entre as formações florestais (João, 2016) que são directamente proporcionais à intensidade da actividade biológica dos microrganismos de solo, visto que segundo Ravindranath & Ostwald (2007) a maior concentração do carbono está em função da concentração da maior actividade biológica no solo.

Factores como temperatura, precipitação, textura do solo, geomorfologia e distúrbios naturais e antropogénicos, podem também contribuir para a variação dos teores de carbono orgânico do solo, sendo que segundo Bird *et al*. (2001) e Mayer *et al*. (2020) esses factores afectam os reservatórios de carbono orgânico do solo.

Apesar da floresta de Mopane ocorrer de forma geral em solos ricos em argila (Timberlake & Gumbo, 2011) proporcionando proteção física da matéria orgânica (Silva *et al*., 2014), o menor teor de carbono orgânico do solo nessa formação deve-se em parte a menor disponibilidade de matéria morta na superfície do solo, visto que a floresta de Mopane por ocorrer em regiões quentes e secas (Bila & Mabjaia, 2012), acabam sofrendo queimadas descontroladas que destroem grande parte da matéria orgânica morta depositada acima do solo, para além do facto de ser uma formação florestal que é caraterizada pela escassez do estrato herbáceo.

Pode também estar por detrás da diferença dos teores médios de COS, a diferença da intensidade de perturbação que cada formação está sujeita em relação à outra, visto que isso dita a velocidade de mineralização da matéria orgânica do solo e consequente emissão de carbono do solo (Arevalo *et al*., 2002).

Estudos de estimativa de carbono orgânico do solo na floresta de Miombo mostraram diferenças dos teores médios de carbono orgânico do solo entre eles. Williams *et al*. (2008) encontrou uma média de 44,9 tonC/ha e 57,9 tonC/ha para Miombo perturbado e intacto, respectivamente, em Nhambita; João (2016) no Miombo da província de Sofala encontrou uma média de 73,04 tonC/ha

Ryan *et al*. (2011) no Miombo de Nhambita encontrou uma média de 76,2 tonC/ha; Ribeiro *et al*. (2013) encontram uma média de 37,72 tonC/ha no Miombo da Reserva Nacional do Niassa; Pienimäki (2014) encontrou uma média de 31,49 tonC/ha e 37,03 tonC/ha no Miombo aberto e fechado, respectivamente, isso na província do Niassa; Dimande (2018) encontrou uma média de 48,93 tonC/ha, 39,30 tonC/ha e 61,62 tonC/ha na concessão florestal, área de uso múltiplo e floresta comunitária respectivamente, todas pertencentes a floresta de Miombo.

As estimativas de carbono orgânico do solo deste estudo na floresta de Miombo (40,15 +/- 1,57) quando comparadas com os resultados encontrados por diversos actores (acima citados) no mesmo ecossistema florestal do país observam-se diferenças que vão desde pequenas a grandes nos valores

encontrados. Estas diferenças podem ser explicadas pela vasta extensão desse ecossistema que ocasiona ligeira variação do mesmo (Timberlake & Gumbo, 2011) proporcionando a ocorrência de cerca de 7 variantes de Miombo no país (Guedes, 2017), isso aliado ao facto de quase todos os outros estudos que antecedem ao presente estudo foram circunscritos a regiões geográficas específicas, explicando em grande parte as diferenças dos teores de carbono, visto que, não houve possibilidade nestes estudos em captarem todas as variantes do Miombo de Moçambique, diferente do presente estudo que com excepção do Miombo de Cabo Delgado, todo outro Miombo foi captado.

O valor de carbono orgânico do solo encontrado neste estudo na floresta de Mopane (35,08 +/- 2,53) difere do valor encontrado por Dimande (2018) na floresta de Mopane (47,34 tonC/ha) da província de Manica, (Boma, 2017) na floresta de Mopane da província de Manica (60,7 tonC/ha), e do encontrado por Cambule *et al*. (2014) na floresta de Mopane da província de Gaza (24,6 tonC/ha e 16,2 tonC/ha).

A semelhança do Miombo, a diferença dos teores de carbono orgânico do solo encontrado neste estudo com os demais estudos pode ser atribuída ao facto do presente estudo ter coberto mais variantes de Mopane em relação aos demais, considerando que, segundo FNDS (2019) o Mopane é a segunda formação mais extensa de Moçambique, depois do Miombo.

O valor de carbono orgânico do solo encontrado neste estudo na floresta de Mecrusse (41,90 +/- 4,87 tonC/ha) não difere em termos estatísticos (intervalo de confiança) do valor encontrado por Magalhães (2017) na floresta de Mecrusse das província de Gaza e Inhambane (39,66 +/- 2,1 tonC/ha). Embora o presente estudo não tenha abrangido a província de Gaza devido a escassez de amostras de solo, a não diferença dos teores médios de carbono entre os dois estudos pode ser explicada pela baixa diversidade biológica e uniformidade dos indivíduos (FNDS, 2019), resultando em cerca de 74% da floresta de Mecrusse ocorrendo em um único tipo de solo (arenosolos e ferráliticos) e 20% ocorrendo em solos estagnados (Magalhães, 2017).

Dimande (2018) encontrou para a floresta Sempre Verde uma média de 71,10 tonC/ha, valor esse que diferente em grande magnitude com o encontrado no presente estudo (49,63 +/- 3,55 tonC/ha). Essa diferença pode também ser atribuída a mais cobertura das variantes da floresta sempre verde neste estudo em relação ao estudo de Dimande (2018) que apenas foi circunscrita a província de

Manica, visto que a floresta sempre verde varia em função da altitude (floresta de montanha), precipitação (floresta de galeria, Miombo húmido), entre outras formações (FNDS, 2019).

Comparação de reservatórios de carbono do presente estudo (carbono orgânico do solo) com o relatório do IV Inventário Florestal Nacional (carbono arbóreo acima e abaixo do solo), mostrou que, o reservatório arbóreo de carbono é que mais contribui para o *stock* de carbono (tonC/ha) nas formações florestais em Moçambique, isto porque, o carbono arbóreo (acima e abaixo do solo) expresso em tonC/ha, representa cerca de 50,5, 55,3 e 54,2% do total de carbono dos dois reservatórios (solo e árvores) na floresta decídua incluindo Miombo, floresta sempre verde e floresta de Mecrusse, respectivamente; enquanto que na floresta de Mopane o carbono arbóreo (acima e abaixo do solo) expresso em tonC/ha, representa cerca de 43,9% do total de carbono dos dois reservatórios (solo e árvores), conforme ilustrado na figura 13 abaixo.

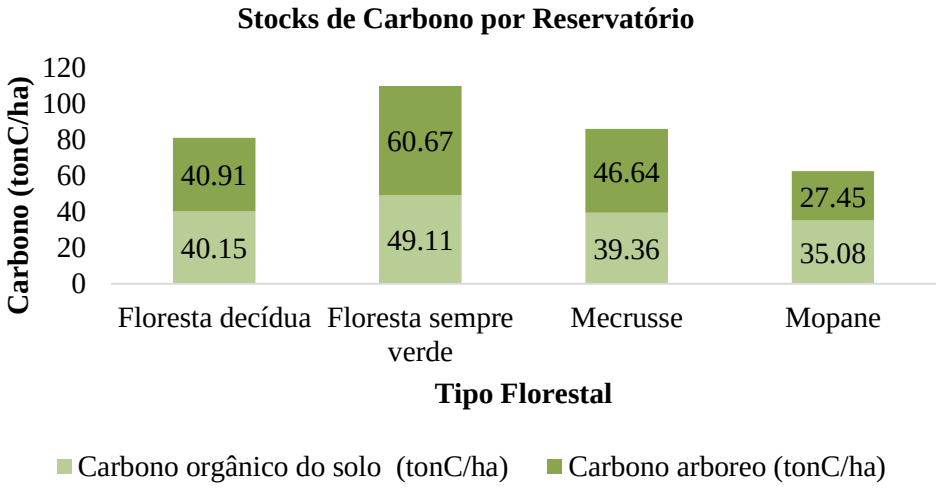


Figura 13: Comparação de *stock* de carbono entre tipos de reservatório por tipo de floresta

Fonte (COS): Presente estudo.
Fonte (Carbono arbóreo): (MITADER, 2018a)

A diferença do teor médio de carbono entre os reservatórios nos ecossistemas florestais era de se prever, dado que, o carbono encontra-se distribuído de forma irregular nos ecossistemas florestais (Burton & Pregitzer, 2008), sendo que, o maior *stock* é encontrado no reservatório do solo das florestas tropicais secas, e maior *stock* na componente arbórea na floresta tropical húmida (Mayer *et al*., 2020).

Apesar da baixa densidade arbórea e consequente baixo teor de carbono arbóreo na floresta de Mopane (MITADER, 2018a), a maior expressão do carbono orgânico do solo em relação ao carbono arbóreo pode ser explicada pela maior concentração de argila na textura dos solos dessas formações que, contribuem bastante para a maior proteção física da matéria orgânica (Silva *et al*., 2014), contra o rompimento estável do carbono do solo. Resultados de maior *stock* de carbono no reservatório solo na floresta de Mopane também foi encontrado por Dimande (2018) na província de Manica na qual o solo representou 74% do total de carbono no Mopane na área de estudo.

Apesar de se esperar que floresta de Mopane tenha maior *stock* de carbono no reservatório solo que se deve ao facto de ser uma tropical seca (Mayer *et al*., 2020), estudo feito por Boma (2017) no mesmo tipo florestal (Mopane) encontrou maior reservatório de carbono no reservatório arbóreo (52% em relação ao total).

4.2.Modelo de Carbono Orgânico do Solo

O modelo de *random forest* estimado, cujos parâmetros são apresentados nas Tabelas 8 e 9, é um modelo de *random forest* do tipo regressão, com um total de 500 árvores (ver Figura 5 que mostra a tendência de estabilização do modelo com o aumento das árvores), e 17 variáveis consideradas (selecionadas aleatoriamente) para dividir o nó em cada etapa de treinamento do modelo. O modelo estimado explica 80.36% da variância, indicando um bom ajuste do modelo (coeficiente de determinação próximo de 1), e tem quadrado médio dos resíduos quadráticos ("RMSE") de 134.21.

Tabela 8: Resumo dos parâmetros do modelo de "*random forest*" estimado

Parâmetros	**Valores**
Tipo de *random forest*	Regressão
Número de árvores	500
No. de variáveis testadas em cada nó	17
Quadrado medio de resíduos	134,21
% Variância explicada	80,36

Fonte: Presente estudo.

Tabela 9: Resumo dos resultados do ajuste do modelo de "r*andom forest*"

Mtry	RMSE	Rsquared	MAE	RMSESD	RsquaredSD	MAESD
2	15,141	0,735	10,901	3,102	0,092	0,923
*17	11,645	0,807	7,246	3,466	0,095	0,769
33	11,789	0,800	7,314	3.470	0,096	0,803

** melhor modelo ajustado*

Fonte: Presente estudo.

O modelo de *random forest* estimado (*mtry = 17*) indica uma acurácia considerada aceitável para a modelagem de COS, isso porque, o modelo apresentou um coeficiente de determinação (R^2) próximo de 1 que é bom (uma forte correlação entre as variáveis preditoras e as propriedades do solo 'COS'), uma raiz do erro quadrático médio normalizado (0,0537) menor que 0,4 que é o máximo aceitável em estudos de modelagem de propriedades do solo (Hengl *et al.,* 2004), e um erro relativo médio de 18%.

Para além dos aspectos acima descritos, com base na figura 14 o modelo ajustado mostrou tendência de estabilização do erro médio quadrático (MSE) à medida que se aumentava o número de árvores, tendo chegado a sua estabilização quando o modelo atingiu 500 árvores, ponto a partir do qual qualquer aumento no número de árvores não significou nenhuma melhoria significativa no desempenho do modelo de *random forest* estimado.

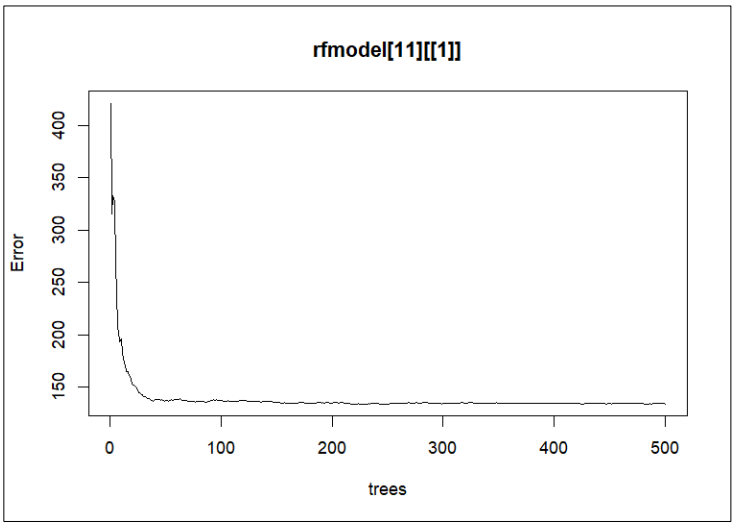


Figura 14: Relação entre o erro médio quadrático (MSE) e o número de árvores no modelo de *random forest* (verificação da estabilização do erro)

Fonte: Presente estudo.

A acurácia encontrada no modelo de *random forest* pode ter sido influenciada pelas amostras de treino e pelas variáveis consideradas na estimativa do modelo, isso porque, segundo Hengl *et al*. (2004), a acurácia de modelos está diretamente relacionada à variabilidade da variável resposta (COS), do tamanho da amostra (amostras menores provocam overfitting do modelo) e das variáveis preditoras incluídas no modelo (variáveis irrelevantes ou altamente correlacionadas podem levar a acurácia do modelo).

Embora todas as variáveis incluídas no modelo tenham mostrado uma influência significativa na acurácia do modelo, nem todas elas tiveram a mesma importância no modelo de *random forest* estimado. Isso foi demostrado através da análise dos parâmetros, Percentual de Aumento no Erro Quadrático Médio (%IncMSE: *Percentage Increase in Mean Squared Error*), e Percentual de Aumento na Pureza dos Nós (%IncNodePurity: *Percentage Increase in Node Purity*), ilustrados nas figuras 15 e 16 abaixo, cujo objetivo de estimativa é fornecer uma indicação de quão significativo é o impacto de cada variável na capacidade do modelo de fazer previsões precisas.

O Percentual de Aumento no Erro Quadrático Médio (%IncMSE) é uma métrica de importância de variáveis usada em modelos de *Random Forest* para regressão, e indica o quanto o desempenho do modelo piora quando os valores de uma variável são destruídos. Valores altos de %IncMSE significam que a variável é muito importante e que sua remoção implica o aumento significativo do erro de predição; enquanto valores baixos ou negativos de %IncMSE significa variável menos importante e que a remoção da mesma pode significar aumento no desempenho do modelo ou nenhuma melhoria.

A métrica Percentual de Aumento na Pureza dos Nós (%IncNodePurity) quantifica o impacto que cada variável tem na melhoria da homogeneidade dos grupos formados durante as divisões dos nós no modelo de random forest. Valores elevados de %IncNodePurity indicam que a variável contribui significativamente para a redução da impureza, enquanto valores baixos (próximos de zero) significam limitada influência no modelo. Comparado ao Percentual de Aumento na Pureza dos Nós (%IncNodePurity), o Percentual de Aumento no Erro Quadrático Médio (%IncMSE) é a métrica mais robusta, pois, esta avalia directamente o impacto da variável na capacidade preditiva do modelo.

Com base na Figura 15, observou-se que a medida que se sai da variável limo para a variável Solos de dunas costeiras, a importância relativa das variáveis para o parâmetro %IncMSE vai decrescendo, isso significa que, a variável limo é a mais importante no modelo e sua possível exclusão causaria um aumento significativo do erro médio quadrado (MSE), o mesmo não se pode esperar com a remoção da variável Solos de dunas costeiras que é a menos importante no modelo, e sua possível remoção não traria um impacto significativo no desempenho do modelo estimado. A baixa importância da variável solo no geral pode ser explicado pelo efeito da esparsidade dos dados (poucas amostras das subcategorias) (Friedman *et al*., 2001).

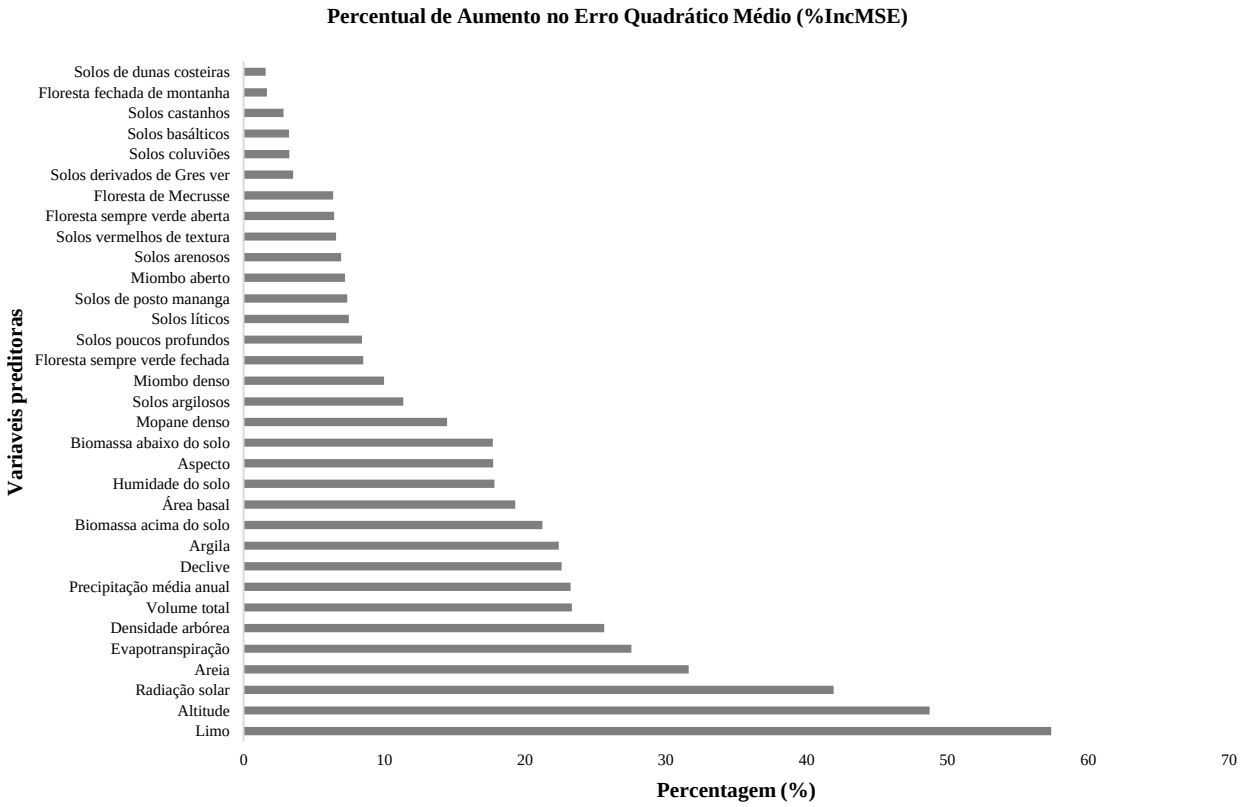


Figura 15: Importância relativa das variáveis preditoras no modelo de *Random Forest* (%IncMSE)

Fonte: Presente estudo.

Para a métrica %IncNodePurity (Figura 16) que é usada para perceber como cada variável contribui para a estrutura de decisão do modelo e na pureza dos nós da árvore, não se verificou um padrão linear de decréscimo da importância relativa das variáveis. Com base na tabela observou-se que a variável limo continua sendo a mais importante para esta métrica, o que sugere que a

mesma seja muito importante na separação das amostras em grupos homogéneos, o que ajuda bastante no desempenho do modelo na predição de COS.

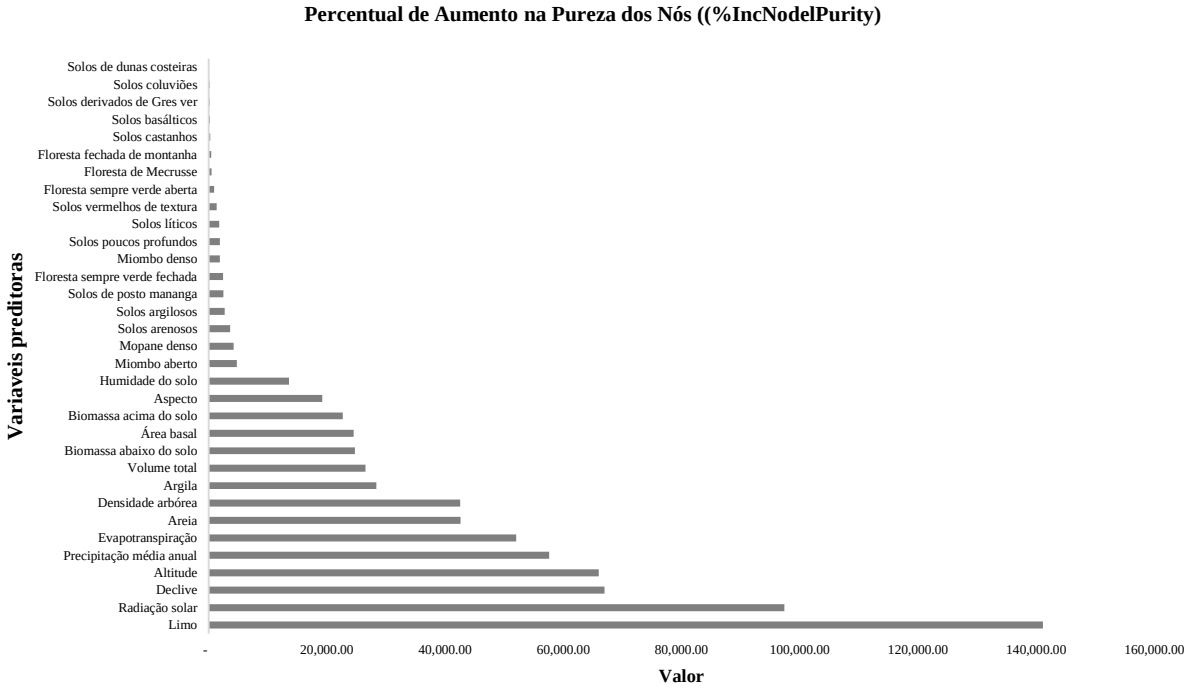


Figura 16: Importância relativa das variáveis das preditoras do modelo de RF (IncNodePurity)

Fonte: Presente estudo.

4.3.Distribuição espacial de carbono orgânico do solo

4.3.1. Mapa de distribuição espacial do COS nos diferentes tipos florestais

O mapa abaixo ilustra a distribuição espacial do carbono orgânico do solo produzido através da técnica *Random Forest*, cujos parâmetros da equação foram apresentados na secção 4.2 deste estudo. Os teores de carbono orgânico do solo estão expressos em toneladas de carbono por hectare (tonC/ha).

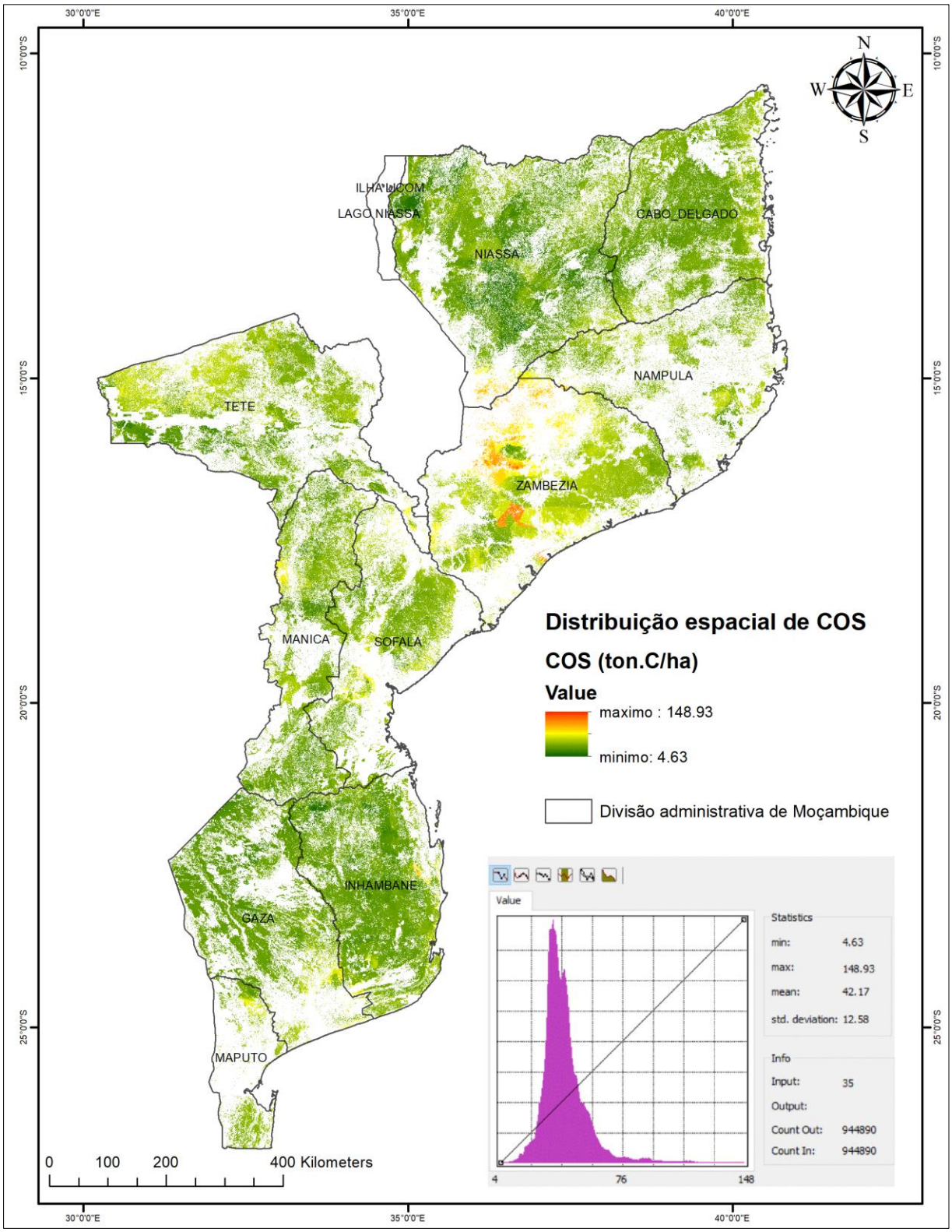


Figura 17: Padrão de distribuição do carbono orgânico do solo nos diferentes tipos florestais de Moçambique

Fonte: Presente estudo.

Com base no mapa de COS produzido através da modelagem e a área florestal por estrato, a média ponderada em função do tamanho (área) dos estratos de carbono orgânico do solo nos ecossistemas florestais encontrada foi de 42,22 tonC/ha, variando de 4,63 tonC/ha (mínimo) à 148,93 tonC/ha (máximo). A média encontrada através da modelagem (42,22 tonC/ha) é superior em cerca de 1,06% da encontrada através do método convencional (41,78 tonC/ha).

Os mapas abaixo ilustram a distribuição espacial do carbono orgânico do solo produzido através da técnica *Random Forest* nos diferentes tipos florestais (floresta decídua incluindo Miombo, floresta Sempre Verde, floresta de Mopane, e floresta de Mecrusse). Os teores de carbono orgânico do solo estão expressos em toneladas de carbono por hectare (tonC/ha).

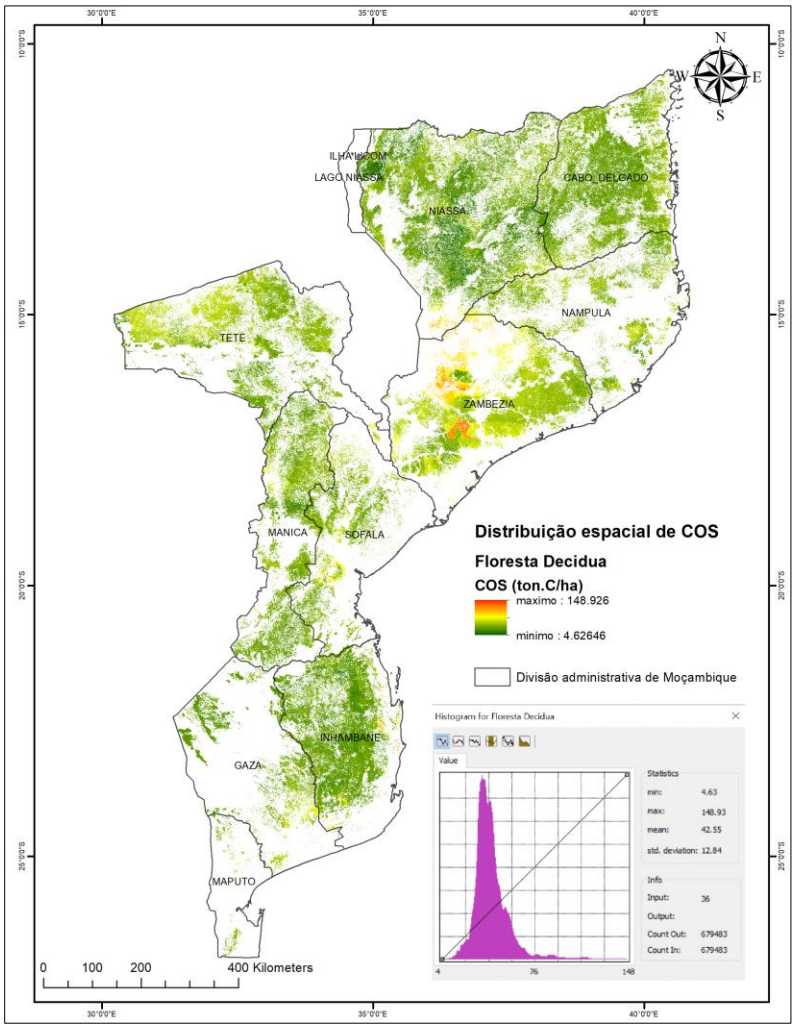


Figura 18: Padrão de distribuição do carbono orgânico do solo na floresta Decídua incluindo Miombo

Fonte: Presente estudo.

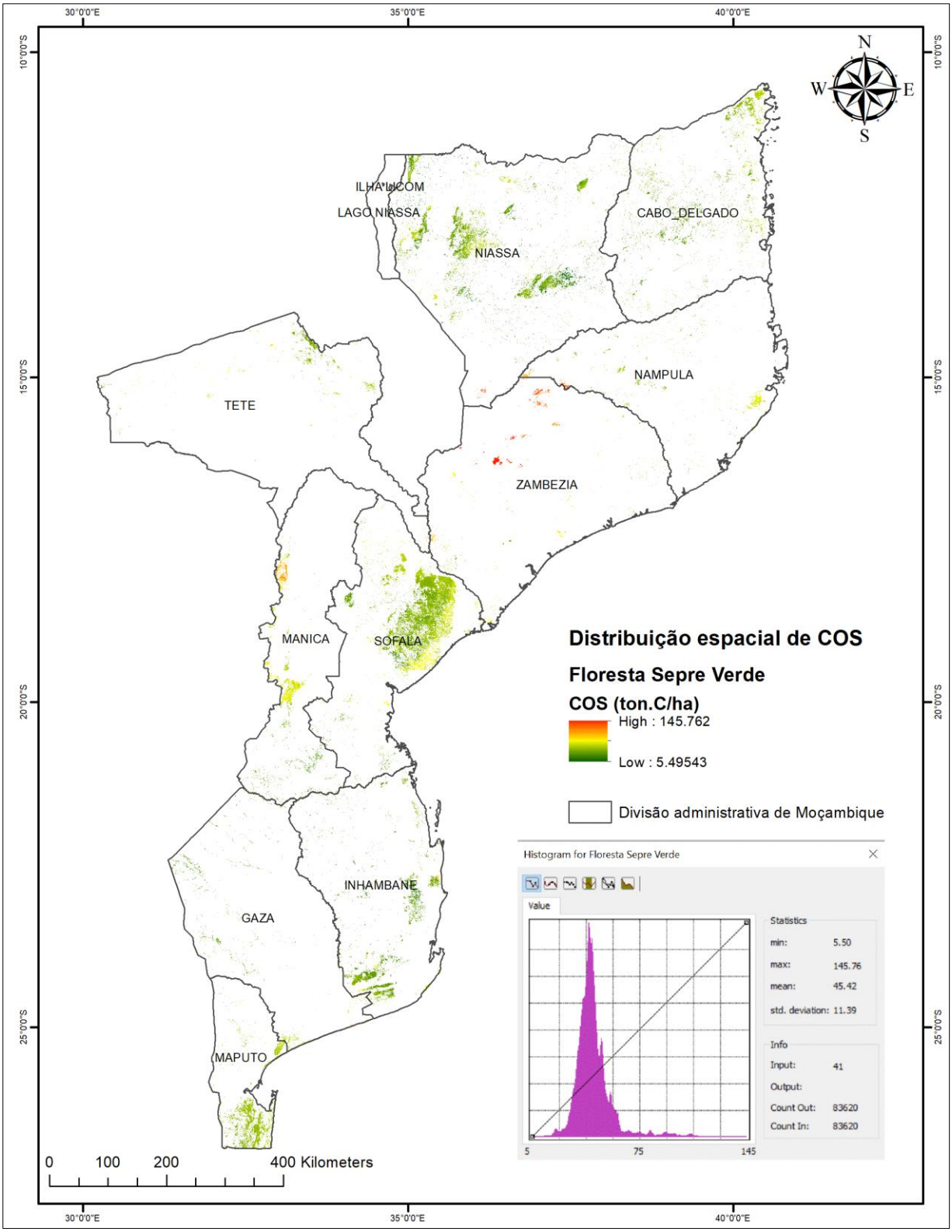


Figura 19: Padrão de distribuição do carbono orgânico do solo na floresta Sempre Verde

Fonte: Presente estudo.

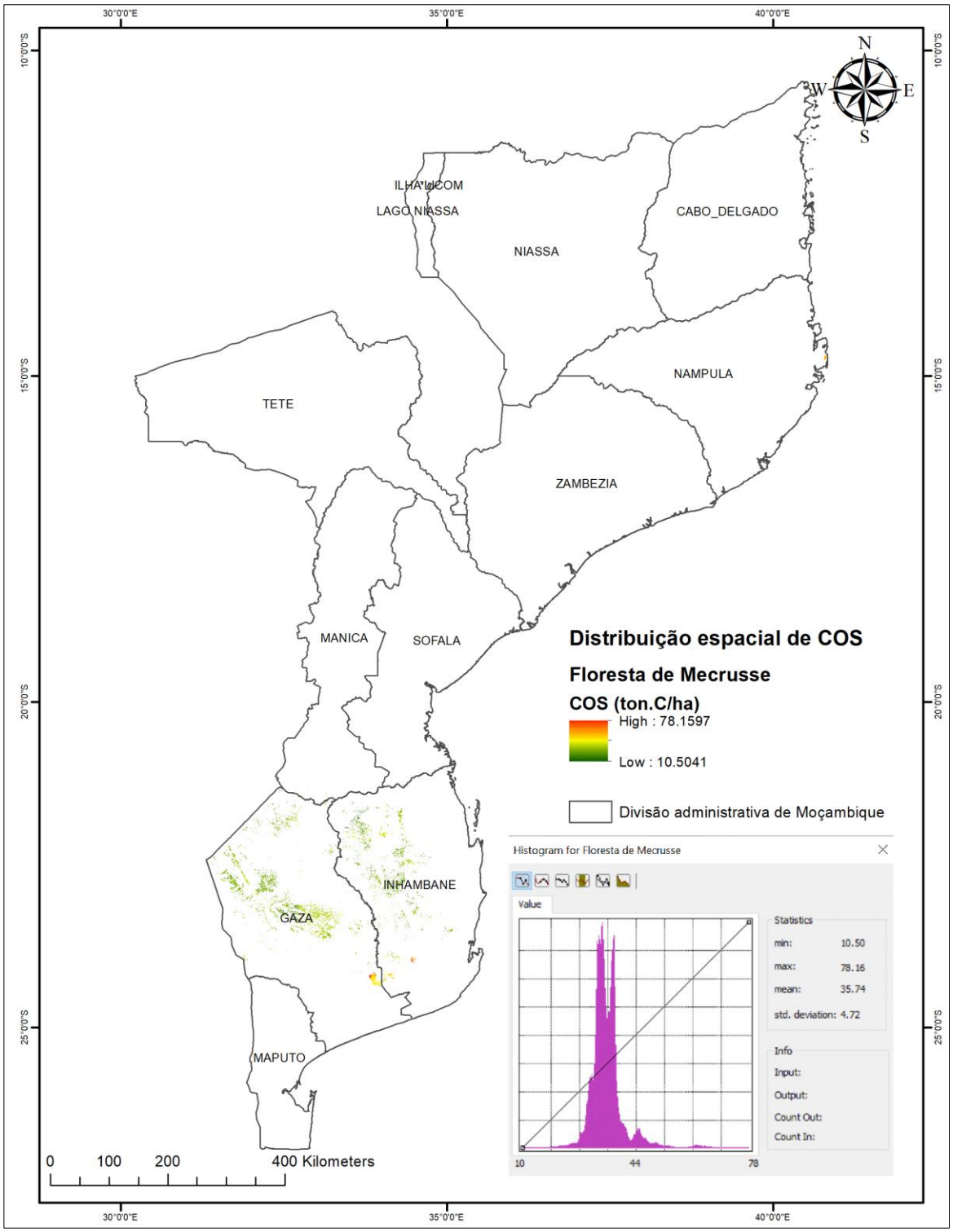


Figura 20: Padrão de distribuição do carbono orgânico do solo na floresta de Mecrusse

Fonte: Presente estudo.

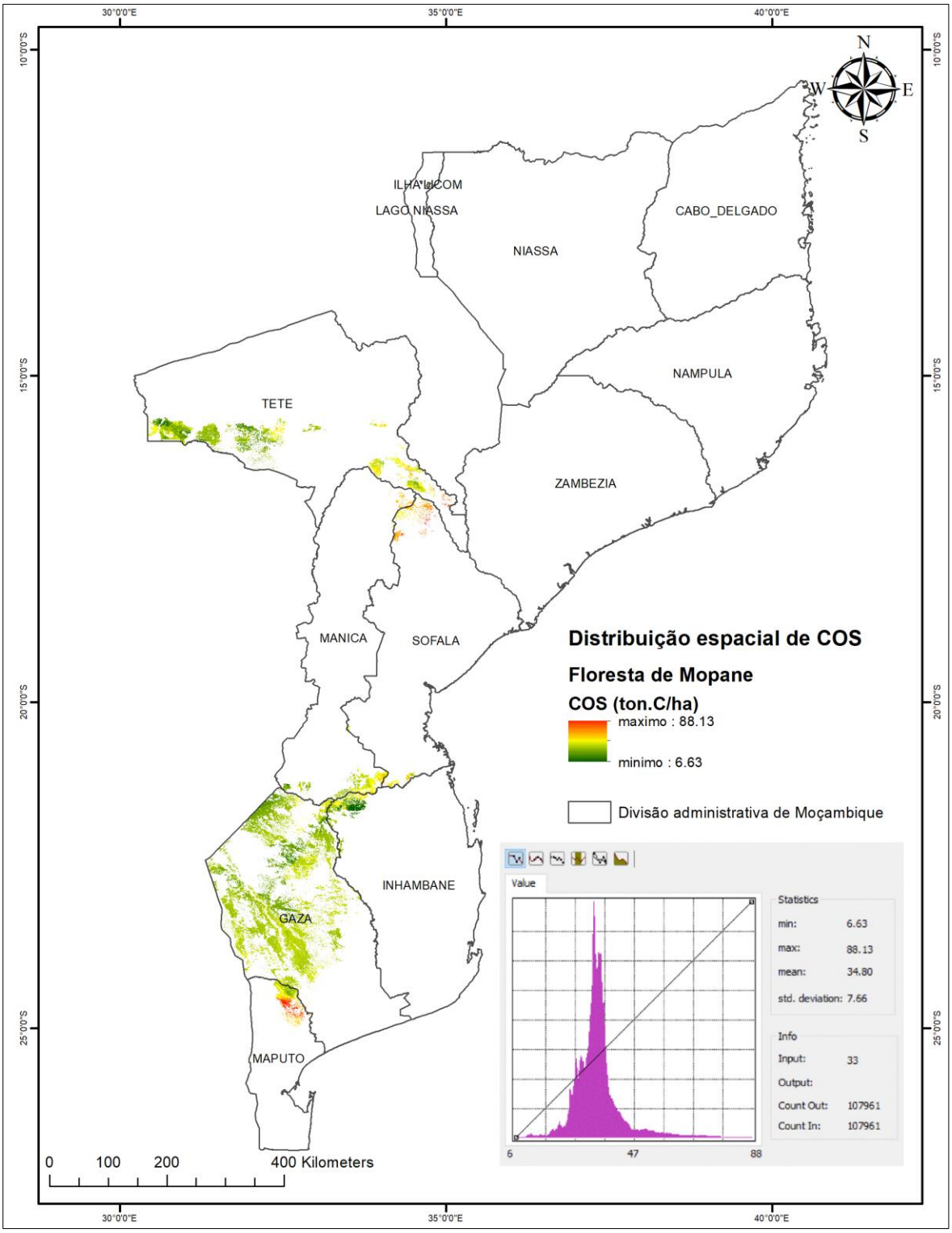


Figura 21: Padrão de distribuição do carbono orgânico do solo na floresta de Mopane

Fonte: Presente estudo.

A nível dos estratos (Figura 18 e 21), a modelagem produziu para a floresta Sempre Verde uma média de 45,42 tonC/ha contra a média de 49,63 tonC/ha obtido do método convencional de carbono orgânico do solo. Na floresta decídua incluindo o Miombo a modelagem produziu uma média de 42,55 tonC/ha contra a média de 40,15 tonC/ha obtido do método convencional de carbono orgânico do solo. Na floresta de Mecrusse a modelagem produziu uma média de 35,74 tonC/ha contra a média de 41,90 tonC/ha obtido do método convencional de carbono orgânico do solo. E na floresta de Mopane a modelagem produziu uma média de 34,80 tonC/ha contra a média de 35,08 tonC/ha obtido do método convencional de carbono orgânico do solo.

Apesar de existirem reservas no uso do método convencional na estimativa de carbono orgânico do solo devido as incertezas que podem advir do uso deste método (falta de homogeneidade espacial na unidade de observação, variação espacial à curta escala das variáveis ambientais dentro da unidade de observação e erros de estimativas) (Bhunia *et al*., 2019), a comparação das médias estimadas através dos dois métodos mostraram não serem muito diferentes para a população (floresta), visto que, a diferença entre elas foi apenas de 1,06%, sendo maior para a média proveniente do método de estimativa através da modelagem. A não diferença pode significar que, as amostras de campo foram representativas aos parâmetros do solo que influenciam no carbono orgânico do solo.

Ao nível dos estratos, registou-se uma diferença de 0,8%, 6,0%, 7,5% e 9,2% na floresta de Mopane, floresta decídua, floresta sempre verde e floresta de Mecrusse, respectivamente, entre os teores de carbono orgânico do solo obtido do método convencional dos obtidos da modelagem. A semelhança da população, todos os tipos florestais com excepção da floresta de Mecrusse produziram diferenças não expressivas de carbono orgânico do solo entre o método convencional e a modelagem através do *Random Forest*, isto porque as diferenças foram todas abaixo de 10%.

4.3.2. Acurácia do mapa de carbono orgânico do solo

A validação cruzada usada para avaliar a exactidão global da predição produziu um erro médio de predição (ME) de -0,79 tonC/ha que, olhando para o sinal negativo indica que o mesmo tende a superestimar em média cerca de 0,79 tonC/ha para cada local não amostrado (Altoé, 2017). Embora a tendência do modelo em superestimar o carbono orgânico do solo, a mesma

superestimativa tende a ser insignificante, isto porque, a superestimativa representa menos de 2% em relação ao valor de carbono orgânico do solo estimado através da modelagem.

Apesar do erro médio de predição ter-se mostrado insignificante, a estatística descritiva dos erros de predição (Tabela 10) mostraram que, o modelo tem alguns problemas de subestimar (sinal positivo) e superestimar (sinal negativo) o carbono orgânico do solo em locais específicos do país, caso gritante foram encontrados nas províncias da Zambézia e Tete (ver a figura 20), isso pode significar que as variáveis que mostraram ser significativas no modelo ajustado, não são eficientes na predição do carbono orgânico do solo nestes locais, ou deve ter havido problema com as amostras nestes locais.

Através da estatística descritiva mostrada na tabela 10 e figura 20, observou-se que, metade dos dados produziram um erro máximo de 0,1% (mediana), e 75% (terceiro quartil) com um erro máximo de 11,0%, o que significa que, as estimativas produzidas são boas, principalmente para o contexto em que se trabalha com dados de solos que tem variação enorme a curta escala espacial.

Tabela 10: Estatística descritiva do erro médio de predição de carbono orgânico do solo

Parâmetro	**COS [tonC/ha]**
Mínimo	-65.35
Primeiro quartil	-4,12
Mediana	-0.06
Média (ME)	-0,79
Terceiro quartil	4,72
Máximo	25,62

Fonte: Presente estudo.

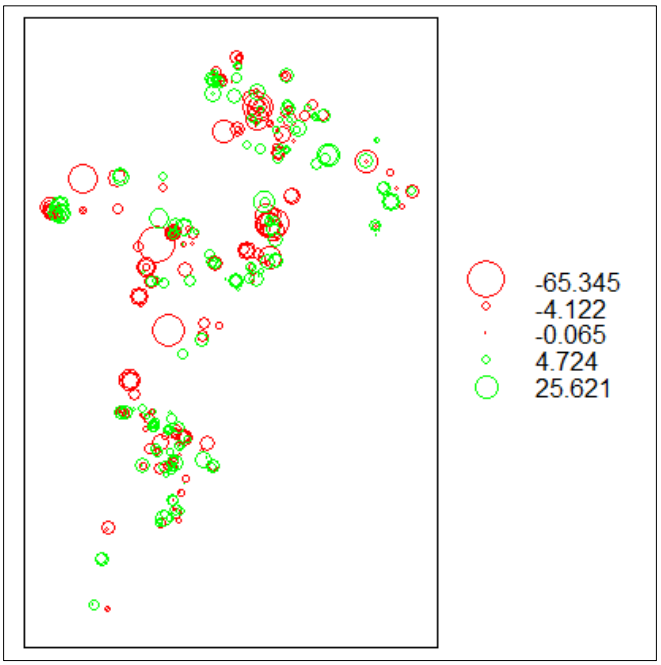


Figura 22: Espacialização do erro de predição (*spatial bubble*)

Fonte: Presente estudo.

Para além do erro de predição, a validação cruzada produziu um erro médio absoluto (MAE) de 6,59 tonC/ha, que corresponde a uma magnitude de erro (desvio em relação ao real valor) de 15,3% do real valor. A magnitude de erro de 15,3% encontrado neste estudo é igual a magnitude de 15,3% encontrado por Wang *et al*. (2018) e inferior a magnitude de 19,8% e 20,76% encontrado por Altoé (2017) e Zhang *et al*. (2022), respectivamente, em estudos de predição de carbono orgânico de solo usando algoritmos de regressão.

A raiz do erro médio quadrático (RMSE) de 10,06 tonC/ha produzido pela validação cruzada, corresponde a um desvio percentual de 23,4%, o que significa que, o modelo produziu uma variância de 76,6%, ou seja, o modelo capta 76,6% da variação dos dados. O desvio percentual encontrado de 23,4% é superior ao encontrado por Altoé (2017) e Wang *et al*. (2018), 13,9% e 19,4%, respectivamente, e inferior (27.61%) ao encontrado por *Zhang et al*. (2022) em estudos similares de predição de carbono orgânico de solo.

Outro parâmetro resultante da validação cruzada foi o coeficiente de determinação ajustado (R^2), que foi de 0,88 que significa que, o modelo é capaz de explicar até 88% dos dados observados (Figura 21). O coeficiente de determinação encontrado neste estudo é superior ao encontrado por

Altoé (2017), Wang *et al*. (2018) e Zhang *et al*. (2022), 0,7, 0,62, e 0,56, respectivamente, em estudos similares de predição de carbono orgânico de solo.

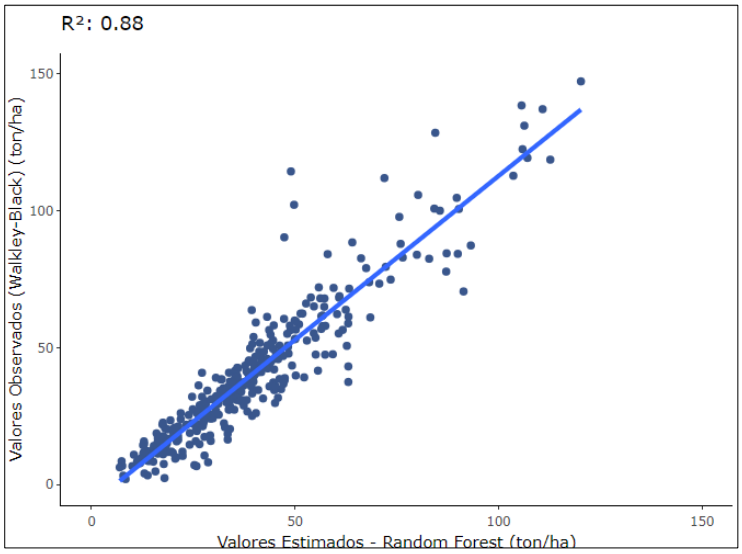


Figura 21: Coeficiente de determinação entre os valores observados e estimados

Fonte: Presente estudo.

Resultados encontrados neste estudo quando comparados aos encontrados por FAO (2020) em Moçambique, mostraram ser muito superiores, o que reforça a acurácia aceitável para o contexto de estudo das propriedades do solo, visto que, usando o algoritmo *Random Forest* de *Machine Learning*, para a mesma área de estudo, encontraram uma acurácia de 4,4 para o erro médio (ME) e 0,31 para o coeficiente de determinação ajustado.

V. CONCLUSÕES E RECOMENDAÇÕES

5.1.Conclusões

Os objectivos específicos do presente estudo conduziram às seguintes conclusões:

Ao nível dos tipos florestais, o carbono orgânico do solo encontra-se distribuído de forma irregular, estando o maior *stock* por unidade de área na floresta Sempre Verde (49,11 tonC/ha), floresta decídua (40,15 tonC/ha) e floresta de Mecrusse (39,36 tonC/ha), e o menor stock por unidade de área encontrado na floresta de Mopane, esta última que representa cerca de 2/3 do carbono orgânico encontrado na floresta sempre verde.

Do total do carbono por unidade de área encontrado nos diferentes tipos florestais (componente arbórea e solo), o carbono orgânico do solo é o maior contribuinte apenas na floresta de Mopane, enquanto o carbono arbóreo (tronco e raízes) é o maior contribuinte apenas na floresta de Mecrusse e floresta Sempre Verde. Na floresta decídua incluindo o Miombo, ambos reservatórios (carbono arbóreo e orgânico do solo) tem a mesma comparticipação no carbono total.

Dentre vários modelos de random forest possíveis, o modelo de random forest do tipo regressão que considera 17 variáveis na divisão das etapas de treinamento foi que mostrou ser mais eficiente na predição do carbono orgânico do solo, isto por considerar uma diversidade suficiente de variáveis (*mtry = 17*) sem causar problemas de overfitting, o que faz com que o mesmo consiga explicar cerca de 80% da variação de carbono orgânico do solo na área de estudo. As variáveis limo, altitude, radiação solar, areia e evapotranspiração são as cinco variáveis mais importantes, cuja remoção poderá causar perda da performance do modelo na predição do carbono orgânico do solo.

Espacialmente, os teores de carbono orgânico do solo variam em larga escala (4,63 a 148,93 tonC/ha) ao longo das formações florestais de Moçambique e isso se deve a variação abrupta dos factores que afectam o teor de carbono orgânico do solo. Para o país como todo, há uma ligeira diferença da média de carbono orgânico do solo proveniente da modelagem (3%) em relação à média proveniente do método convencional (pontos observados), mas essa diferença já não é ligeira quando se analisa ao nível dos tipos florestais, pois, pode chegar aos 10,4% (Mecrusse).

O mapa de carbono orgânico do solo produzido através da modelagem tem validade estatística, o que significa que pode ser usado para fins relacionados com projectos de REDD+, isto porque a validação cruzada através do erro médio (ME=1,8%), erro médio absoluto (MAE = 15,3%), raiz do erro médio quadrático (RMSE = 23,4%) e coeficiente de determinação (R=0,88) mostrou que os valores encontrados estão dentro dos intervalos aceitáveis quando se trata de estudos relacionados com as propriedades de solos.

5.2.Recomendações

Os resultados conjugados com a revisão bibliográfica conduziram as seguintes recomendações:

- ✓ Procurar condições para que amostras de solo sejam obtidas nas províncias de Gaza e Cabo Delgado como forma a aumentar a precisão das estimativas nestas províncias. A alternativa menos onerosa seria encontrar amostras de solo de outros estudos feitos nestas duas províncias para servirem ao menos como amostras de validação.
- ✓ Identificar mais variáveis ambientais que possam ser incluídas como forma de melhorar a precisão das estimativas, principalmente para regiões em que os erros de estimativas foram grosseiros.
- ✓ Modelar outros parâmetros relacionados com as propriedades do solo, como é o caso da fertilidade, nitrogénio e fosforo nos ecossistemas florestais de Moçambique.

VI. REFERÊNCIAS BIBLIOGRÁFICAS

Aju, P. C., Iwuchukwu, J. J. I., & Ibe, C. C. (2015). Our Forests, Our Environment, Our Sustainable Livelihoods. *European Journal of Academic Essays*, *2*(4), 6–19. www.euroessays.org

Alafandy, K. A., Omara, H., Lazaar, M., & Al Achhab, M. (2022). *Machine Learning* (pp. 83–113). https://doi.org/10.4018/978-1-7998-9831-3.ch005

Ali, J., Khan, R., Ahmad, N., & Maqsood, I. (2012). Random Forests and Decision Trees. *International Journal of Computer Science Issues*, *9*(5), 272–278. www.IJCSI.org

Altoé, T. F. (2017). *Modelagem E Espacialização Do Estoque De Carbono Em Fragmentos NATIVOS DE MINAS GERAIS EM FUNÇÃO DE FATORES EDAFOCLIMÁTICOS*. Universidade Federal de Lavra.

Arevalo, L. A., Alegre, J. C., & Vilcahuaman, L. J. M. (2002). Metodologia para Estimar o Estoque de Carbono em Diferentes Sistemas de Uso da Terra. *Embrapa*, 1–38.

Austin, M. (2007). Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling*, *200*(1–2), 1–19. https://doi.org/10.1016/j.ecolmodel.2006.07.005

Bangroo, S. A., Najar, G. R., Achin, E., & Truong, P. N. (2020). Application of predictor variables in spatial quanti fi cation of soil organic carbon and total nitrogen using regression kriging in the North Kashmir forest Himalayas. *Catena*, *193*(September 2019), 104632. https://doi.org/10.1016/j.catena.2020.104632

Barbosa, R. R. N., Silva, A. A., Galvão, M., Galvão, A. R. A., & Neto, C. F. O. (2013). Produção e sequestro de carbono na atmosfer. *Enciclopédia Biosfera*, *9*(16), 1783–1798. http://www.conhecer.org.br/enciclop/2013a/agrarias/producao e sequestro.pdf

Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine Learning in Agriculture: A Comprehensive Updated Review. *Sensors*, *21*(11), 3758. https://doi.org/10.3390/s21113758

Berry, N. J., Harley, R., & Ryan, C. M. (2013). Enabling communities to benefit from REDD+: Pragmatic assessment of carbon benefits. *Carbon Management*, *4*(6), 571–573. https://doi.org/10.4155/cmt.13.62

Bhunia, G. S., Shit, P. K., Pourghasemi, H. R., & Edalat, M. (2019). Prediction of Soil Organic Carbon and its Mapping Using Regression Analyses and Remote Sensing Data in GIS and R. In *Spatial Modeling in GIS and R for Earth and Environmental Sciences* (pp. 429–450). Elsevier. https://doi.org/10.1016/b978-0-12-815226-3.00019-3

Bila, J. M., & Mabjaia, N. (2012). Crescimento e fitossociologia de uma floresta com Colophospermum mopane, em Mabalane, Província de Gaza, Moçambique. *Pesquisa Florestal Brasileira*, *32*(72), 421–427. https://doi.org/10.4336/2012.pfb.32.72.421

Bila, J. M., Sanquetta, C. R., Paula, A., Corte, D., José, L., & Freitas, M. De. (2018). *Distribuição diamétrica e principais espécies arbóreas presentes nos ecossistemas de Miombo , Mopane e Mecrusse em Moçambique*. https://doi.org/10.4336/2018.pfb.38e201701523

Bird, M., Santrùcková, H., Lloyd, J., & Veenendaal, E. (2001). Global Soil Organic Carbon Pool. *Global Biogeochemical Cycles in the Climate System*, *December*, 185–199. https://doi.org/10.1016/b978-012631260-7/50016-9

Blesslin Sheeba, T., Anand, L. D. V., Manohar, G., Selvan, S., Wilfred, C. B., Muthukumar, K., Padmavathy, S., Ramesh Kumar, P., & Asfaw, B. T. (2022). Machine Learning Algorithm for Soil Analysis and Classification of Micronutrients in IoT-Enabled Automated Farms. *Journal of Nanomaterials*, *2022*, 1–7. https://doi.org/10.1155/2022/5343965

Boma, A. (2017). *Estoques de carbono de florestas nativas e de plantações florestais de Eucalyptus Pinus na Província de Manica, Centro de Moçambique*. Universidade Eduardo Mondlane.

Box, G. E., & Cox, D. R. (1982). An analysis of transformations revisited, rebutted. *Journal of the American Statistical Association*, *77*(377), 209–210. https://doi.org/10.1080/01621459.1982.10477788

British Society of Soil Science. (2021). *SCIENCE NOTE : SOIL CARBON*. 1–8.

Brown, S., Gillespie, A. J. R., & Lugo, A. E. (1989). Biomass estimation methods for tropical forests with applications to forest inventory data. *Forest Science*, *35*(4), 881–902.

Buczko, U., García, R., Harmuth, J., Kalbe, J., Scharnweber, T., Stoll, A., Wilmking, M., & Jurasinski, G. (2023). Soil and vegetation factors affecting carbon storage in a coastal forest in NE Germany. *Geoderma Regional*, *33*, e00629. https://doi.org/10.1016/j.geodrs.2023.e00629

Burton, A. J., & Pregitzer, K. S. (2008). Measuring Forest Floor, Mineral Soil, and Root Carbon Stocks. In *Field Measurements for Forest Carbon Monitoring* (pp. 129–142). Springer Netherlands. https://doi.org/10.1007/978-1-4020-8506-2_10

Cambule, A. H., Rossiter, D. G., Stoorvogel, J. J., & Smaling, E. M. A. (2014). Soil organic carbon stocks in the Limpopo National Park, Mozambique: Amount, spatial distribution and uncertainty. *Geoderma*, *213*, 46–56. https://doi.org/10.1016/j.geoderma.2013.07.015

CNDS. (2002). *Relatório Sobre a Avaliação do Grau de Implementação da Agenda 21 em Moçambique*. http://sistemas.mda.gov.br/condraf/arquivos/1372414655.pdf

Corning, E., Sadeghpour, A., Ketterings, Q., & Czymmek, K. (2016). The Carbon Cycle and Soil Organic Carbon. *Agronomy Fact Sheet Series*, *C*, 2.

Curtis, P. S. (2008). Estimating Aboveground Carbon in Live and Standing Dead Trees. In *Field Measurements for Forest Carbon Monitoring* (pp. 39–44). Springer Netherlands. https://doi.org/10.1007/978-1-4020-8506-2_4

CVM. (2022). *Mozambique* (Vol. 2020).

Deuschle, W. J. (2019). *Undergraduate Fundamentals of Machine Learning*. Harvard College.

Dimande, F. (2018). *Estimativa de Estoque de Carbono nos Ecossistemas Florestais da Província de Manica Estimativa de Estoque de Carbono nos Ecossistemas Florestais da Província de Manica*.

Dittus, W. P. J. (1977). The Ecology of a Semi-Evergreen Forest Community in Sri Lanka. *Biotropica*, *9*(4), 268. https://doi.org/10.2307/2388145

FAO. (2008). *Guide to laboratory establishment for plant nutrient analysi*. FAO (FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS).

FAO. (2017). Soil Organic Carbon: the hidden potential. In *Nature Reviews Genetics* (Vol. 9, Issue 4). Food and Agriculture Organization of the United Nations. https://doi.org/10.1038/nrg2350

FAO. (2018). *Soil Organic Carbon Mapping Cookbook* (2nd ed., Vol. 1).

FAO. (2020). Global Soil Organic Carbon Map Version 1.5: Technical Report. In *Global Soil Organic Carbon Map (GSOCmap) Version 1.5*. https://doi.org/10.4060/ca7597en

FAO. (2022). *The State of the World's Forests 2022*. FAO. https://doi.org/10.4060/cb9360en

Favero, L. P., & Belfiore, P. (2017). *MANUAL DE ANALISE DE DADOS: Estatística e Modelagem Multivariada com Excel, SPSS e Stata*.

Field, A. (2018). *Discovering statistics using IBN SPSS statistics* (J. Seaman (ed.); 5th ed.). SAGE. https://doi.org/10.1234/12345678

FNDS. (2019). *Mapa de Cobertura Florestal de Moçambique 2016*.

Fontana, A., & de Campos, D. V. B. (2017). Carbono Organico. In P. C. Teixeira, G. K. Donagemma, A. Fontana, & W. G. Teixeira (Eds.), *MANUAL DE METODOS DE ANALISE DE SOLO* (3rd ed., pp. 360–367). Embrapa. http://www.cse.edu.uy/sites/www.cse.edu.uy/files/documentos/Liccom_Camejo_2011-07-28.pdf

Friedman, J., Trevor, H., & Tibshinani, R. (2001). *The elements os statistical learning*.

Frost, J. (2019). Introduction to Statistics: an intuitive guide for analyzing data and unlocking discoveries. In *Technometrics* (1st ed.). Statistics By Jim Publishing. https://doi.org/10.1080/00401706.1968.10490641

GOUVEIA, D. H. G., & AZEVEDO, A. L. (1949). *Caraterísticas e distribuição dos solos de Moçambique*. Centro de Investigacão Cientifica Algodoeira.

Graybill, F. A., & Iyer, H. K. (1994). Regression Analysis: Concepts and Applications. *Duxbury Pr*, 1–699.

Gruszczyński, S. (2019). Prediction of soil properties with machine learning models based on the spectral response of soil samples in the near infrared range. *Soil Science Annual*, *70*(4), 298–313. https://doi.org/10.2478/ssa-2019-0027

Guedes, B. (2017). *Carbon Stocks and Dynamics in Miombo, Pinus and Eucalyptus Forests in Mozambique*. Swedish University of Agricultural Sciences.

Hammarstrand, L., & Särnberger, A. (2013). *Comparative evaluation of two forest systems under different management regimes in Miombo woodlands - A case study in Kitulangalo area, Tanzania*. *2013*, 74.

Hengl, T., Heuvelink, G. B. M., & Stein, A. (2004). A generic framework for spatial prediction of soil variablesm based on regression-kriging. *Geoderma*, *120*, 75–93. https://doi.org/10.1016/j.geoderma.2003.08.018

INE. (1997). *Inquerito demografico e de saúde*.

João, S. S. (2016). *Avaliação da composição florística, estrutura horizontal e estoques de carbono em duas formações florestais na província de Sofala*. Eduardo Mondlenae University.

João, S. S. (2024). *Performance dos Algoritmos Supervisionados de Machine Learning na Predição do Carbono Orgânico do Solo nas Florestas Tropicais Secas: Estudo de Caso de Moçambique*. Universidade de Sao Paulo.

Kuznetsova, A. I. (2022). *INFLUENCE OF VEGETATION ON SOIL CARBON STOCKS IN FORESTS (REVIEW)*. *5*(4), 1–35. https://doi.org/10.31509/2658-607x-202251-100

Lal, R. (2021). *Soil Organic Carbon and Feeding the Future*. CRC Press.

https://doi.org/10.1201/9781003243090

Lamprecht, H. (1990). *Silvicultura nos Trópicos*.

Lisboa, S. N. (2017). *Estimativa de Biomassa e Carbono na Floresta Sempre Verde de Montanha da Reserva Florestal de Moribane*. Universidade Eduardo Mondlane.

Liu, Z., Peng, C., Work, T., Candau, J.-N., DesRochers, A., & Kneeshaw, D. (2018). Application of machine-learning methods in forest ecology: recent progress and future challenges. *Environmental Reviews*, *26*(4), 339–350. https://doi.org/10.1139/er-2018-0034

Macamo, C., & Sitoe, A. (2016). *Relatório de Governação Ambiental 2016 - Governação e gestão de mangais em Moçambique.* 63.

Magalhães, T. M. (2017). Carbon stocks in necromass and soil pools of a Mozambican tropical dry forest under different disturbance regimes. *Biomass and Bioenergy*, *105*, 373–380. https://doi.org/10.1016/j.biombioe.2017.07.023

Mahmoudzadeh, H., Matinfar, H. R., Taghizadeh-Mehrjardi, R., & Kerry, R. (2020). Spatial prediction of soil organic carbon using machine learning techniques in western Iran. *Geoderma Regional*, *21*, e00260. https://doi.org/10.1016/j.geodrs.2020.e00260

Mangiafico, S. S. (2015). *An R Companion for the Handbook of Biological Sta s cs Analysis of Covariance*. *August 2016*, 1–8.

Marzoli, A. (2007). *Inventário Florestal Nacional -Avaliação Integrada Das Florestas de Moçambique*.

Masson-Delmotte, V., Zhai, P., Pörtner, H.-O., Roberts, D., Skea, J., Shukla, P. R., Pirani, A., Moufouma-Okia, W., Péan, C., Pidcock, R., Connors, S., Matthews, J. B. R., Chen, Y., Zhou, X., Gomis, M. I., Lonnoy, E., Maycock, T., Tignor, M., & Waterfield, T. (2019). *Global warming of 1.5°C: An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*. www.environmentalgraphiti.org

Mayer, M., Prescott, C. E., Abaker, W. E. A., Augusto, L., Cécillon, L., Ferreira, G. W. D., James, J., Jandl, R., Katzensteiner, K., Laclau, J. P., Laganière, J., Nouvellon, Y., Paré, D., Stanturf, J. A., Vanguelova, E. I., & Vesterdal, L. (2020). Influence of forest management activities on soil organic carbon stocks: A knowledge synthesis. *Forest Ecology and Management*, *466*(March), 118127. https://doi.org/10.1016/j.foreco.2020.118127

Mcdonald, J. H. (2014). *H a n d b o o k o f bi o l o g i c a l s t a t i s t i c s* (3rd ed.). SPARKY HOUSE PUBLISHING.

Milhano, A. P. (2008). *GESTÃO DOS RECURSOS HÍDRICOS EM MOÇAMBIQUE: Gaza-Limpopo*.

Minitab. (2023). *What to do with nonnormal data*. https://support.minitab.com/en-us/minitab/21/help-and-how-to/statistics/basic-statistics/supporting-topics/normality/what-to-do-with-nonnormal-data/

MITADER. (2018a). *Inventario Florestal Nacional*.

MITADER. (2018b). *MANUAL DE PROCEDIMENTOS DO IV INVENTARIO FLORESTAL NACIONAL 2016. v 0.0*. 92.

Mohammed, M., Khan, M. B., & Bashier, E. B. M. (2016). *Machine Learning*. CRC Press. https://doi.org/10.1201/9781315371658

Mondal, A., Khare, D., Kundu, S., & Mondal, S. (2016). Spatial soil organic carbon (SOC) prediction by regression kriging using remote sensing data. *The Egyptian Journal of Remote Sensing and Space Sciences*. https://doi.org/10.1016/j.ejrs.2016.06.004

Mozambique, G. of. (2021). *Updated First National Determined Contribution of Mozambique*.

Nabiollahi, K., Eskandari, S., Taghizadeh-Mehrjardi, R., Kerry, R., & Triantafilis, J. (2019). Assessing soil organic carbon stocks under land-use change scenarios using random forest models. *Carbon Management*, *10*(1), 63–77. https://doi.org/10.1080/17583004.2018.1553434

NCASI. (2013). A review of biomass carbon accounting methods and implications. *NCASI Technical Bulletin*, *1015*, 1–52.

Nilsson, N. J. (1998). *INTRODUCTION TO MACHINE LEARNING AN EARLY DRAFT OF A PROPOSED TEXTBOOK Department of Computer Science*. Stanford University.

Noa, M., & Falcão, P. M. (2016). *Definição de Florestas no âmbito do REDD+*. https://www.biofund.org.mz/wp-content/uploads/2018/06/1528267239-Relatorio definicao de floresta V5_19.10.2016.pdf

Pearson, T., Walker, S., & Brown, S. (2005). Sourcebook for Land use, Land-use change and forestry projects. *Winrock International and the BioCarbon Fund of the World Bank 57 (2005)*, *21*(3), 64. http://wbcarbonfinance.org/docs/Background_LULUCF_Sourcebook_compressed.pdf

Pienimäki, A. (2014). *Carbon stocks for six different land use classes in Niassa province , Mozambique*. https://core.ac.uk/download/pdf/33725366.pdf

Piri Sahragard, H., & Pahlavan-Rad, M. R. (2020). Prediction of Soil Properties Using Random Forest with Sparse Data in a Semi-Active Volcanic Mountain. *Eurasian Soil Science*, *53*(9), 1222–1233. https://doi.org/10.1134/S1064229320090136

Ramesh, T., Bolan, N. S., Kirkham, M. B., Wijesekara, H., Kanchikerimath, M., Srinivasa Rao, C., Sandeep, S., Rinklebe, J., Ok, Y. S., Choudhury, B. U., Wang, H., Tang, C., Wang, X., Song, Z., & Freeman II, O. W. (2019). *Soil organic carbon dynamics: Impact of land use changes and management practices: A review* (pp. 1–107). https://doi.org/10.1016/bs.agron.2019.02.001

Ravindranath, N. H., & Ostwald, M. (2007). Carbon inventory methods: Handbook for greenhouse gas inventory, carbon mitigation and roundwood production projects. In *Advances in Global Change Research* (Vol. 29). Springer US. https://doi.org/10.1007/978-1-4020-6547-7

Ribeiro, N., Matos, C., Moura, I., Washington-allen, R., & Ribeiro, A. (2013). *Monitoring vegetation dynamics and carbon stock density in miombo woodlands*. 1–9.

Ribeiro, N., Sitoe, A. A., Guedes, B. S., & Staiss, C. (2002). Manual de silvicultura tropical. *Revista Do Instituto de Medicina Tropical de S*, *35*(4), 130.

Ryan, C. M., Williams, M., & Grace, J. (2011). Above- and belowground carbon stocks in a miombo woodland landscape of mozambique. *Biotropica*, *43*(4), 423–432. https://doi.org/10.1111/j.1744-7429.2010.00713.x

Schlegel, Gayoso, & Guerra. (2001). Manual de procedimientos para inventarios de carbono en ecosistemas forestales. In *Proyecto FONDEF D98I1076*. Universidad Austral De Chile.

Sheskin, D. J. (2003). Handbook of Parametric and Nonparametric Statistical Procedures. In *Handbook of Parametric and Nonparametric Statistical Procedures*. https://doi.org/10.4324/9780203489536

Silva, A. S., Silva, I. de F. da, Bandeira, L. B., Dias, B. de O., & Silva Neto, L. de F. da. (2014). Argila e matéria orgânica e seus efeitos na agregação em diferentes usos do solo. *Ciência Rural*, *44*(10), 1783–1789. https://doi.org/10.1590/0103-8478cr20130789

Sitoe, A. (2001). *Bases ecológicas para agronomia e silvicultura*.

Sitoe, A., Guedes, B., Machoco, R., Cambule, A., Cunhete, D., Nacamo, E., & Guiamba, R. (2009). *Baseline carbon estimation in Dombe , Manica Biofuel production area Mozambique Principle Energy*.

Sullivan, W. (2017). *Machine Learning For BeginnersGuide Algorithms:Supervised & Unsupervised LearningDecision Tree & Random ForestIntroduction*. Healthy Pragmatic Solutions Inc.

Timberlake, J., & Gumbo, D. (2011). The Dry forests and woodlands of Africa: managing for products and services. In *Choice Reviews Online* (Vol. 48, Issue 06). https://doi.org/10.5860/choice.48-3271

Tito, M. R., León, M. C., & Porro, R. (2008). Guía para la Determinación de Carbono en Pequeñas Propiedades Rurales. In *Consorcio Iniciativa Amazónica (IA) y Centro Mundial Agroforestal (ICRAF)*. ICRAF)/Consórcio Iniciativa Amazônica (IA).

Trawinski, B., Smetek, M., Telec, Z., & Lasota, T. (2012). Nonparametric statistical analysis for multiple comparison of machine learning regression algorithms. *International Journal of Applied Mathematics and Computer Science*, *22*(4), 867–881. https://doi.org/10.2478/v10006-012-0064-z

Trivedi, P., Singh, B. P., & Singh, B. K. (2018). Soil carbon: Introduction, importance, status, threat, and mitigation. In *Soil Carbon Storage: Modulators, Mechanisms and Modeling*. Elsevier Inc. https://doi.org/10.1016/B978-0-12-812766-7.00001-9

United Nations. (2015). *ADOPTION OF THE PARIS AGREEMENT - Paris Agreement*.

Wang, S., Adhikari, K., Wang, Q., Jin, X., & Li, H. (2018). Role of environmental variables in the spatial distribution of soil carbon (C), nitrogen (N), and C:N ratio from the northeastern coastal agroecosystems in China. *Ecological Indicators*, *84*(August 2017), 263–272. https://doi.org/10.1016/j.ecolind.2017.08.046

Williams, M., Ryan, C. M., Rees, R. M., Sambane, E., Fernando, J., & Grace, J. (2008). Carbon sequestration and biodiversity of re-growing miombo woodlands in Mozambique. *Forest Ecology and Management*, *254*(2), 145–155. https://doi.org/10.1016/j.foreco.2007.07.033

Xiao, C. (2015). *Soil Organic Carbon Storage (Sequestration) Principles and Management. Potential Role for Recycled Organic Materials in Agricultural Soils of Washington State* (Issue 15).

Yang, L., Cai, Y., Zhang, L., Guo, M., Li, A., & Zhou, C. (2021). A deep learning method to predict soil organic carbon content at a regional scale using satellite-based phenology variables. *International Journal of Applied Earth Observation and Geoinformation*, *102*, 102428. https://doi.org/10.1016/j.jag.2021.102428

Yao, Y., Dai, Q., Gao, R., Yi, X., Wang, Y., & Hu, Z. (2023). Characteristics and factors influencing soil organic carbon composition by vegetation type in spoil heaps. *Frontiers in Plant Science*, *14*. https://doi.org/10.3389/fpls.2023.1240217

Zhang, W. chun, Wan, H. shuang, Zhou, M. hou, Wu, W., & Liu, H. bin. (2022). Soil total and organic carbon mapping and uncertainty analysis using machine learning techniques.

Ecological Indicators, *143*(August), 109420. https://doi.org/10.1016/j.ecolind.2022.109420

Zhao, Y., & CenYonghua. (2014). *Data Mining Applications with R*.

Zwick, R., & Marascuilo, L. A. (1984). Selection of pairwise multiple comparison procedures for parametric and nonparametric analysis of variance models. *Psychological Bulletin*, *95*(1), 148–155. https://doi.org/10.1037/0033-2909.95.1.148