UNIVERSIDADE ESTADUAL PAULISTA (UNESP)
FACULDADE DE ENGENHARIA
CÂMPUS DE ILHA SOLTEIRA

MELISSA ALEXANDRE SANTOS

MODELAGEM MULTIVARIADA E INTELIGÊNCIA ARTIFICIAL PARA A PREDIÇÃO DE INDICADORES DE QUALIDADE DO SOLO NA RECUPERAÇÃO DE LATOSSOLOS DEGRADADOS SOB CULTIVO DE *Mabea fistulifera* E *Eucalyptus urograndis*

Ilha Solteira
2026

UNIVERSIDADE ESTADUAL PAULISTA (UNESP)
FACULDADE DE ENGENHARIA
CÂMPUS DE ILHA SOLTEIRA

MELISSA ALEXANDRE SANTOS

MODELAGEM MULTIVARIADA E INTELIGÊNCIA ARTIFICIAL PARA A PREDIÇÃO DE INDICADORES DE QUALIDADE DO SOLO NA RECUPERAÇÃO DE LATOSSOLOS DEGRADADOS SOB CULTIVO DE *Mabea fistulifera* E *Eucalyptus urograndis*

Tese apresentada à Faculdade de Engenharia de Ilha Solteira, Universidade Estadual Paulista (UNESP), como parte dos requisitos para obtenção do título de Doutora em Agronomia.
Especialidade: Sistemas de Produção.

Orientadora: Prof.ª Drª. Carolina dos Santos Batista Bonini

Coorientadora: Prof.ª Drª. Marlene Cristina Alves

Ilha Solteira
2026

FICHA CATALOGRÁFICA
Desenvolvida pela Diretoria Técnica de Biblioteca e Documentação

S237m

Santos, Melissa Alexandre.
Modelagem multivariada e inteligência artificial para a predição de indicadores de qualidade do solo na recuperação de latossolos degradados sob cultivo de *Mabea fistulifera* e *Eucalyptus urograndis* / Melissa Alexandre Santos. -- Ilha Solteira: [s.n.], 2026
113 f.

Tese (doutorado) - Universidade Estadual Paulista. Faculdade de Engenharia de Ilha Solteira. Área de conhecimento: Sistemas de Produção, 2026

Orientadora: Carolina dos Santos Batista Bonini
Coorientadora: Marlene Cristina Alves
Inclui bibliografia

1. Análise multivariada. 2. Aprendizagem de máquina. 3. Funcionalidade edáfica. 4. Infiltração de água. 5. Resiliência ambiental.

Elaborada por Raiane da Silva Santos - CRB-8/9999

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Ilha Solteira

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: Modelagem Multivariada e Inteligência Artificial para a Predição de Indicadores de Qualidade do Solo na Recuperação de Latossolos Degradados sob Cultivo de Mabea Fistulifera e Eucaliptus urograndis

AUTORA: MELISSA ALEXANDRE SANTOS
ORIENTADORA: CAROLINA DOS SANTOS BATISTA BONINI
ORIENTADOR DE MOBILIDADE: OSVALDO SALAZAR GUERRERO (Internacional)

Aprovada como parte das exigências para obtenção do Título de Doutora em Agronomia, especialidade: Sistemas de Produção pela Comissão Examinadora:

gov.br Documento assinado digitalmente
CAROLINA DOS SANTOS BATISTA BONINI
Data: 04/02/2026 09:46:50-0300
Verifique em https://validar.iti.gov.br

Profa. Dra. CAROLINA DOS SANTOS BATISTA BONINI (Participaçao Virtual)
Departamento de Produção Vegetal / UNESP / Câmpus de Dracena - FCAT

gov.br Documento assinado digitalmente
PAULO RENATO MATOS LOPES
Data: 04/02/2026 09:41:02-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. PAULO RENATO MATOS LOPES (Participaçao Virtual)
Departamento de Produção Vegetal / UNESP / Câmpus de Dracena - FCAT

gov.br Documento assinado digitalmente
THAIS SOTO BONI
Data: 04/02/2026 09:26:02-0300
Verifique em https://validar.iti.gov.br

Profa. Dra. THAÍS SOTO BONI (Participaçao Virtual)
Departamento de Fitossanidade, Engenharia Rural e Solos / UNESP / Câmpus de Ilha Solteira - FEIS

Firmado digitalmente por
Dr. Borja Velázquez Martí

Prof. Dr. BORJA VELAZQUEZ MARTI (Participaçao Virtual)
Departamento de Ingeniería Rural y Agroalimentaria / Universidad Politècnica de València (Espanha)

gov.br Documento assinado digitalmente
DIEGO HENRIQUE SILVA DE SOUZA
Data: 03/02/2026 17:31:40-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. DIEGO HENRIQUE SILVA DE SOUZA (Participaçao Virtual)
Departamento de Engenharia Agrícola / Universidade Federal Rural do Pernambuco - UFRPE

Ilha Solteira, 30 de janeiro de 2026.

IMPACTO POTENCIAL DESTA PESQUISA

A pesquisa apresenta contribuições científicas e práticas de grande relevância para a restauração de solos tropicais, especialmente Latossolos severamente degradados. Ao integrar técnicas avançadas de modelagem multivariada e inteligência artificial (ACP + RNP), o estudo oferece um método robusto para interpretar a evolução de atributos edáficos de forma integrada, superando limitações de abordagens unidimensionais tradicionalmente utilizadas em avaliações de recuperação.

Do ponto de vista ambiental, os resultados demonstram que o uso de espécies nativas, como *Mabea fistulifera*, aliado a doses adequadas de composto orgânico, favorece a reconstrução da estrutura do solo, aumentando a macroporosidade, a infiltração de água, a capacidade de troca catiônica e o conteúdo de matéria orgânica. Esses processos contribuem significativamente para o restabelecimento da funcionalidade ecológica, influenciando positivamente o balanço hídrico, a ciclagem de nutrientes e o sequestro de carbono.

O impacto social e econômico também é significativo, pois a metodologia proposta pode subsidiar políticas públicas voltadas à recuperação de áreas degradadas por hidrelétricas, mineração, expansão agrícola e outros tipos de degradação. A aplicação prática deste modelo analítico em propriedades rurais, assentamentos, áreas de preservação ou terras coletivas pode reduzir custos ao indicar combinações de espécies e insumos com maior eficiência comprovada.

Além disso, o modelo ACP+RNP pode ser replicado em outros biomas e tipos de solo, ampliando seu potencial internacional. A pesquisa contribui para os Objetivos do Desenvolvimento Sustentável (ODS), especialmente os ODS 13 (Ação Climática) e 15 (Vida Terrestre), fortalecendo práticas agroecológicas e restauradoras. Assim, o trabalho oferece uma contribuição estratégica para a ciência do solo, para a gestão ambiental e para comunidades que dependem da recuperação produtiva e ecológica de seus territórios.

POTENTIAL IMPACT OF THIS RESEARCH

This research provides significant scientific and practical contributions to the restoration of tropical soils, particularly severely degraded Ferralsols. By integrating advanced multivariate modeling and artificial intelligence techniques (PCA + PNN), the study offers a robust method for interpreting soil recovery dynamics in an integrated manner, overcoming limitations of traditional univariate approaches.

Environmentally, the findings demonstrate that the use of native species such as *Mabea fistulifera*, combined with adequate amounts of organic compost, helps restore soil structure, increasing macroporosity, water infiltration, cation exchange capacity, and organic matter content. These improvements significantly contribute to reestablishing ecological functionality, enhancing hydrological regulation, nutrient cycling, and carbon sequestration.

The study also has important social and economic implications. The proposed analytical model can support public policies for recovering areas degraded by hydroelectric construction, mining, agricultural expansion, and other disturbances. Its application in rural properties, conservation areas, settlements, and community territories can help reduce restoration costs by identifying the most efficient combinations of species and soil amendments.

Moreover, the PCA+PNN model can be replicated across different biomes and soil types, increasing its international relevance. The research contributes to Sustainable Development Goals, particularly SDG 13 (Climate Action) and SDG 15 (Life on Land), strengthening ecological and agroecological restoration practices. Overall, this work provides a strategic contribution to soil science, environmental management, and communities seeking to restore ecological and productive functions.

IMPACTO POTENCIAL DE ESTA INVESTIGACIÓN

Esta investigación aporta contribuciones científicas y prácticas de gran relevancia para la restauración de suelos tropicales, especialmente Ferralsoles severamente degradados. La integración de técnicas avanzadas de modelado multivariado e inteligencia artificial (ACP + RNP) permite una interpretación robusta e integrada de la evolución del suelo, superando las limitaciones de enfoques univariados tradicionales.

Desde una perspectiva ambiental, los resultados muestran que el uso de especies nativas como Mabea fistulifera, junto con dosis adecuadas de compost orgánico, favorece la reconstrucción de la estructura del suelo, aumentando la macroporosidad, la infiltración de agua, la capacidad de intercambio catiónico y la materia orgánica. Estos procesos contribuyen al restablecimiento de la funcionalidad ecológica, mejorando la regulación hídrica, la ciclicación de nutrientes y el secuestro de carbono.

A nivel social y económico, el modelo propuesto puede apoyar políticas públicas de recuperación de áreas degradadas por represas hidroeléctricas, minería, expansión agrícola y otros procesos de deterioro ambiental. Su aplicación práctica en predios rurales, áreas protegidas, asentamientos o territorios colectivos puede reducir costos al señalar las combinaciones más eficientes de especies y enmiendas.

Asimismo, el modelo ACP+RNP puede replicarse en diversos biomas y tipos de suelo, ampliando su alcance internacional. La investigación contribuye a los Objetivos de Desarrollo Sostenible, especialmente los ODS 13 (Acción Climática) y 15 (Vida de Ecosistemas Terrestres), fortaleciendo prácticas de restauración ecológica y agroecológica. En conjunto, este trabajo constituye una contribución estratégica para la ciencia del suelo, la gestión ambiental y las comunidades que buscan recuperar la funcionalidad ecológica y productiva de sus territorios.

À minha avó Gegé (*in memoriam*),

Por cada gargalhada e pelo cuidado com seu jardim (seus netos e flores).

Às mulheres do campo e da ciência,

que lutaram e lutam para que hoje este e outros trabalhos possam ser lidos e feitos por outras mulheres.

Dedico.

AGRADECIMENTOS

Antes de tudo, agradeço às forças que têm me regido nesse caminho; independentemente de seu nome, minha gratidão.

Aos meus pais, Jô e Severo, por apostarem cada ficha em mim e por me amarem incondicionalmente.

À minha família, que sempre me apoiou e deu suporte a todas as escolhas que fiz para seguir meu caminho. Em especial, às minhas irmãs Milene e Regiane, minhas paixões eternas, que aquecem meu coração e me fortalecem, e às minhas irmãs Sabrina e Valentina.

Aos meus sobrinhos Samuel e Davi pelo dom da vida e do amor.

Às minhas primas maravilhosas, Nete, Kell, Lene e Ruthe pelo carinho e confiança no meu potencial.

Aos meus tios Gina e Cleo pelo amor e zelo.

À minha tia Glorinha, que sempre me apoiou e ajudou no que era necessário, meu muito obrigada tia.

Aos meus animais Tião, Didi e Fumaça (*in memoriam*), Joca, Neguinha, Zeus, aos bibis Didica, Mask e Pepe, às gatitas Tiê, Cora e Marrie, e ao meu novo laranja maluco, Lóki, que foram suporte emocional fundamental para que eu pudesse vivenciar essa experiência da pós-graduação com tranquilidade e me proporcionaram a melhor família que alguém poderia ter.

À minha orientadora, Carol Bonini, que, com amor e paciência, me ensinou muito mais do que Ciência do Solo, mas também a delicadeza do cuidado, do amor e da gentileza — além de ter suportado esta maluquinha que vos escreve.

Às minhas mães do coração, Lucia Helena Romitelli e Maria Silvia Romitelli, que foram mais do que anjos em minha vida: foram o estímulo na busca pelo conhecimento que me trouxe até aqui.

Ao Sato, que foi e continua sendo um pai com sua presença e amor, às vezes a gente só precisa entender que nem todas as pessoas vão nos machucar, obrigada papa por esses 21 anos de contato e cuidado.

Aos professores Prof. Dr. Paulo Lopes e Prof. Dr. Vitor Corrêa, pelo suporte prestado à pesquisa.

Aos meus amigos Laura Nantes, Aline Yukari, Lucas Alves, José Augusto, Gabi Freitas e Asser, pelos dias de alegria e loucura e por todo o companheirismo ao longo desses dez anos de UNESP.

Aos amigos Vitor Cruz, Aline Marchetti, Gabriel Lunardelli e Jorge Delfim, bem como aos colegas do GENAP, que muito me apoiaram nas coletas, no laboratório e também nos momentos de celebração.

Aos assistentes administrativos, aos técnicos de campo e laboratório, especialmente à Andréia da Silva Pepeliskof, que me auxiliou durante as análises químicas, ao Alan, Gustavo F., Edna (raiiiiinhaaaaaaaaaa <3), Renan B. e ao Carlos G., que sempre foram amáveis e solícitos no que fosse necessário, vocês são massa demais.

Ao bonde das princesas (trabalhadoras do setor terceirizado) Diana, Marcilene, Simone, Mari, Regina e Cris por todos os beijos e abraços de bom dia durante esses 10 anos, vocês tornaram meus dias mais leves e com certeza levarei no coração cada palavra bonita e de incentivo que vocês me presentearam (nossa, me debulho em lágrimas nesse momento, porque vocês são realmente especiais e importantes para mim).

Às amigas que fiz ao longo da vida e que revelam a beleza de viver: Dani Carla (dona da melhor gargalhada do mundo), Mel Isabel (que compartilha comigo do niilismo de forma cômica) e Fran Ribas (a maluquete das tecnologias).

Às Betes, ao Anão e ao Murilo, que me acolheram e me deram um lar quando precisei; à Xuca maluca e ao Colgate, pelas risadas compartilhadas.

Às minhas amigas Iza Okuma e Nayr Pereira, pelo amor e por todas as vezes em que precisei de suas palavras para lembrar que o amor está em tudo, inclusive dentro de mim. Eu não tenho palavras para agradecer todas as vezes que me escutaram e tiraram a dor do meu peito quando o mundo desabava sobre minha cabeça.

A mis amigos chilenos, Luis Rojo y su mamá Victoria, que fueron verdaderos ángeles cuando yo necesitaba un techo; a Basti y Felipe, por soportar mis millones de memes en IG y juntos ilusionarnos con un viaje a Lençóis Maranhenses; a Caro y Coni, por compartir el mejor departamento y las noches más memorables de piscola y fiestas; a Benko, Cano, Vicho, Imara, Javi, Yasna y Paul, por los mejores terrenos y aventuras en el Sur.

A los amigos en España, especialmente Aya, Andrés, Gabriel, Davi, Max, Alan, Luis, Aldo, Salah y Mohamed.

A la familia más tierna y amable de toda España, Mavi, Diego y mi sobrino precioso Jackie. ¡Qué regalo tan grande me ha dado Tuti! ¡Los amo muchísimo, amigos!

أشكر أيضاً عزيز على قوله لي هذه الكلمات الثمينة وعلى قراءة هالة طاقتي بهذه الطريقة الكريمة.

A mis profesores Borja Velázquez y Osvaldo Salazar, por todo el conocimiento que me han transmitido y por su enorme paciencia. También a Don Hugo y al profesor Óscar Seguel, quienes me enseñaron que el conocimiento no debe separarnos, sino unirnos, y que la humildad es un verdadero regalo ("El desorden es el progreso").

A todos os que contribuíram, direta ou indiretamente, para que meus sonhos se concretizassem e que continuam vibrando positivamente por mim.

À UNESP – Faculdade de Engenharia de Ilha Solteira (FEIS) e à Faculdade de Ciências Agrárias e Tecnológicas (FCAT), por meio de seus professores, técnicos de campo, laboratório e administrativos, que contribuíram para a realização deste trabalho.

Por fim, agradeço à minha avó Gegé, que, rindo e chorando ao mesmo tempo, me fez compreender a efemeridade da vida e a importância de não deixar nada para depois, pois o momento presente é o único que verdadeiramente existe e me ensinou que minha intensidade, nossa intensidade, no cuidado e amor, nunca serão um castigo, mas uma dádiva.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES – Código de Financiamento 001, por meio da concessão de bolsa de doutorado DS.

“Terra não é fonte de lucro,

nem dádiva social,

nem objeto de especulação

ou modalidade de poupança.

***Terra é a base vital da humanidade**.”*

Ana Primavesi

APRESENTAÇÃO

O presente trabalho teve como objetivo avaliar, por meio de ferramentas multivariadas e de modelos de inteligência artificial, os processos de recuperação de Latossolos degradados submetidos a diferentes estratégias de revegetação e adição de insumos orgânicos. A pesquisa integra abordagens analíticas avançadas, com destaque para a Análise de Componentes Principais (ACP) e Redes Neurais Artificiais, especialmente as Redes Neurais Probabilísticas (PNN), visando a predição e a interpretação de indicadores físicos, químicos e biológicos associados à qualidade do solo em áreas submetidas à restauração ecológica.

A tese está estruturada em dois estudos complementares. O primeiro "**APPLICATION OF PRINCIPAL COMPONENT ANALYSIS AND PROBABILISTIC NEURAL NETWORKS IN FERRALSOLS RECOVERY EVALUATION THROUGH PLANTING OF *Mabea fistulifera* AND *Eucalyptus urograndis***" aborda a aplicação de modelos multivariados na avaliação da recuperação do solo a partir do plantio de *Mabea fistulifera* (espécie nativa) e *Eucalyptus urograndis* (espécie exótica), sob diferentes doses de composto orgânico, em Latossolos com histórico de degradação. Esse estudo busca identificar variáveis-chave e padrões de resposta dos atributos do solo decorrentes das distintas combinações de espécie e manejo.

O segundo estudo "**ARTIFICIAL NEURAL NETWORKS FOR PREDICTING SOIL QUALITY INDICATORS BASED ON AGGREGATE STRUCTURE**" concentra-se na predição de indicadores de qualidade do solo a partir da estrutura de agregados, utilizando redes neurais artificiais como ferramenta de modelagem não linear. A análise visa compreender a contribuição da estabilidade e organização dos agregados do solo na determinação de sua funcionalidade ecológica, explorando a capacidade das redes neurais de capturar relações complexas entre variáveis estruturais e parâmetros de qualidade.

De forma integrada, o projeto busca avançar na compreensão dos mecanismos de recuperação de Latossolos degradados, oferecendo subsídios científicos para o planejamento de práticas de restauração ecológica e manejo sustentável. A combinação de análise multivariada e modelos de inteligência artificial representa uma abordagem inovadora, com elevado potencial para aprimorar a interpretação de dados complexos e fortalecer a utilização de indicadores robustos para a avaliação da qualidade do solo.

PRESENTATION

The present reserach aimed to evaluate, through multivariate tools and artificial intelligence models, the recovery processes of degraded Ferralsols subjected to different revegetation strategies and the addition of organic inputs. The research integrates advanced analytical approaches, with emphasis on Principal Component Analysis (PCA) and Artificial Neural Networks, particularly Probabilistic Neural Networks (PNN), to support the prediction and interpretation of physical, chemical, and biological indicators associated with soil quality in areas under ecological restoration.

The thesis is structured into two complementary studies. The first, **"APPLICATION OF PRINCIPAL COMPONENT ANALYSIS AND PROBABILISTIC NEURAL NETWORKS IN FERRALSOLS RECOVERY EVALUATION THROUGH PLANTING OF *Mabea fistulifera* AND *Eucalyptus urograndis***," addresses the application of multivariate models to assess soil recovery based on the planting of *Mabea fistulifera* (native species) and *Eucalyptus urograndis* (exotic species), under different doses of organic compost in Ferralsols with a history of degradation. This study aims to identify key variables and response patterns of soil attributes resulting from distinct species and management combinations.

The second study, **"ARTIFICIAL NEURAL NETWORKS FOR PREDICTING SOIL QUALITY INDICATORS BASED ON AGGREGATE STRUCTURE,"** focuses on predicting soil quality indicators from aggregate structure using artificial neural networks as a nonlinear modeling tool. The analysis seeks to understand the contribution of soil aggregate stability and organization to ecological functionality, exploring the networks' capacity to capture complex relationships between structural variables and quality parameters.

Taken together, the project advances the understanding of recovery mechanisms in degraded Ferralsols, offering scientific support for the planning of ecological restoration practices and sustainable management. The combination of multivariate analysis and artificial intelligence models represents an innovative approach with high potential to enhance the interpretation of complex datasets and strengthen the use of robust indicators for assessing soil quality.

PRESENTACIÓN

El presente trabajo tuvo como objetivo evaluar, mediante herramientas multivariadas y modelos de inteligencia artificial, los procesos de recuperación de Ferralsoles degradados sometidos a diferentes estrategias de revegetación y adición de insumos orgánicos. La investigación integra enfoques analíticos avanzados, con énfasis en el Análisis de Componentes Principales (ACP) y Redes Neuronales Artificiales, particularmente las Redes Neuronales Probabilísticas (PNN), con el fin de predecir e interpretar indicadores físicos, químicos y biológicos asociados a la calidad del suelo en áreas sometidas a restauración ecológica.

La tesis está estructurada en dos estudios complementarios. El primero, **"APPLICATION OF PRINCIPAL COMPONENT ANALYSIS AND PROBABILISTIC NEURAL NETWORKS IN FERRALSOLS RECOVERY EVALUATION THROUGH PLANTING OF *Mabea fistulifera* AND *Eucalyptus urograndis***," aborda la aplicación de modelos multivariados para evaluar la recuperación del suelo a partir del cultivo de *Mabea fistulifera* (especie nativa) y *Eucalyptus urograndis* (especie exótica), bajo diferentes dosis de compost orgánico en Ferralsoles con historial de degradación. Este estudio busca identificar variables clave y patrones de respuesta de los atributos del suelo derivados de las distintas combinaciones de especie y manejo.

El segundo estudio, **"ARTIFICIAL NEURAL NETWORKS FOR PREDICTING SOIL QUALITY INDICATORS BASED ON AGGREGATE STRUCTURE,"** se centra en la predicción de indicadores de calidad del suelo a partir de la estructura de agregados, utilizando redes neuronales artificiales como herramienta de modelación no lineal. El análisis pretende comprender la contribución de la estabilidad y organización de los agregados del suelo en la determinación de su funcionalidad ecológica, explorando la capacidad de las redes para capturar relaciones complejas entre variables estructurales y parámetros de calidad.

De manera integrada, el proyecto busca avanzar en la comprensión de los mecanismos de recuperación de Ferralsoles degradados, ofreciendo aportes científicos para la planificación de prácticas de restauración ecológica y manejo sostenible. La combinación de análisis multivariado y modelos de inteligencia artificial constituye un enfoque innovador, con alto potencial para mejorar la interpretación de datos complejos y fortalecer el uso de indicadores robustos en la evaluación de la calidad del suelo.

RESUMO

A degradação de Latossolos tropicais representa um desafio crítico para programas de recuperação ambiental, especialmente em áreas submetidas a perturbações severas, como a construção de hidrelétricas. Neste contexto, este trabalho avaliou a recuperação física e química de um Latossolo Vermelho distrófico degradado, por meio do plantio de *Mabea fistulifera* (espécie nativa) e *Eucalyptus urograndis* (espécie exótica), associados a doses crescentes de composto orgânico e adubação mineral. O experimento foi conduzido em área degradada, considerada como “área de empréstimo” fornecido para Usina Hidrelétrica de Ilha Solteira, representando um cenário real de perda estrutural e funcional do solo. Foram monitorados atributos físicos (densidade, porosidade total, macro e microporosidade, distribuição de poros, infiltração de água) e químicos (matéria orgânica, pH, acidez potencial, CTC, Ca, Mg, K, P, saturação por bases). Para interpretar a complexidade dos dados, foi empregada uma abordagem integrada de Análise de Componentes Principais (ACP) e Redes Neurais Probabilísticas (RNP), permitindo sintetizar variáveis e identificar probabilisticamente os tratamentos mais eficientes na recuperação do solo. Os resultados indicaram que *Mabea fistulifera*, especialmente quando combinada com 20 Mg ha^{-1} de composto orgânico, promoveu melhorias substanciais na estrutura e fertilidade do solo, com aumento de matéria orgânica, macroporosidade, capacidade de troca catiônica e infiltração, e redução da densidade do solo, aproximando o sistema de condições edáficas consideradas ótimas pelo modelo ACP+RNP. *Eucalyptus urograndis* apresentou desempenho inferior, sobretudo nos atributos físicos. O estudo evidencia que a integração entre espécies nativas, insumos orgânicos e modelagem multivariada é eficaz para restaurar a funcionalidade de Latossolos degradados, oferecendo subsídios técnicos robustos para programas de recuperação em ambientes tropicais.

Palavras-chave: análise multivariada; aprendizagem de máquina; funcionalidade edáfica; infiltração de água; resiliência ambiental.

ABSTRACT

The degradation of tropical Oxisol represents a critical challenge for environmental restoration programs, especially in areas subjected to severe disturbances such as hydroelectric construction. In this context, the present study evaluated the physical and chemical recovery of a degraded dystrophic Red Oxisol through the planting of *Mabea fistulifera* (native species) and *Eucalyptus urograndis* (exotic species), combined with increasing doses of organic compost and mineral fertilization. The experiment was conducted in a degraded area classified as a “borrow area” used for the Ilha Solteira Hydroelectric Plant, representing a real scenario of structural and functional soil loss. Physical (bulk density, total porosity, macro- and microporosity, pore-size distribution, water infiltration) and chemical attributes (organic matter, pH, potential acidity, CEC, Ca, Mg, K, P, base saturation) were monitored. To interpret the complexity of the dataset, an integrated approach combining Principal Component Analysis (PCA) and Probabilistic Neural Networks (PNN) was applied, allowing variable synthesis and probabilistic identification of the most efficient treatments for soil recovery. The results showed that *Mabea fistulife*ra, especially when combined with 20 Mg ha^{-1} of organic compost, promoted substantial improvements in soil structure and fertility, with increases in organic matter, macroporosity, cation exchange capacity, and water infiltration, and reductions in bulk density, bringing the system closer to the optimal edaphic conditions defined by the PCA+PNN model. *Eucalyptus urograndis* exhibited lower performance, particularly in physical attributes. The study demonstrates that the integration of native species, organic inputs, and multivariate modeling is an effective strategy for restoring the functionality of degraded Oxisol, providing robust technical support for restoration programs in tropical environments.

Keywords: environmental resilience; machine learning; multivariate analysis; soil functionality; water infiltration.

RESUMEN

La degradación de Ferralsoles tropicales representa un desafío crítico para los programas de recuperación ambiental, especialmente en áreas sometidas a perturbaciones severas, como la construcción de represas hidroeléctricas. En este contexto, este estudio evaluó la recuperación física y química de un Ferralsol Rojo distrófico degradado mediante el cultivo de *Mabea fistulifera* (especie nativa) y *Eucalyptus urograndis* (especie exótica), combinados con dosis crecientes de compost orgánico y fertilización mineral. El experimento se llevó a cabo en un área degradada clasificada como “área de préstamo” utilizada para la Central Hidroeléctrica de Ilha Solteira, representando un escenario real de pérdida estructural y funcional del suelo. Se monitorearon atributos físicos (densidad aparente, porosidad total, macro y microporosidad, distribución de poros, infiltración de agua) y químicos (materia orgánica, pH, acidez potencial, CIC, Ca, Mg, K, P, saturación de bases). Para interpretar la complejidad del conjunto de datos, se aplicó un enfoque integrado basado en el Análisis de Componentes Principales (ACP) y Redes Neuronales Probabilísticas (RNP), lo que permitió sintetizar variables e identificar probabilísticamente los tratamientos más eficientes para la recuperación del suelo. Los resultados indicaron que *Mabea fistulifera*, especialmente cuando se combinó con 20 Mg ha^{-1} de compost orgánico, promovió mejoras sustanciales en la estructura y fertilidad del suelo, con aumentos en la materia orgánica, macroporosidad, capacidad de intercambio catiónico e infiltración de agua, y reducción de la densidad aparente, aproximando el sistema a condiciones edáficas óptimas según el modelo ACP+RNP. *Eucalyptus urograndis* presentó un desempeño inferior, especialmente en los atributos físicos. El estudio demuestra que la integración de especies nativas, enmiendas orgánicas y modelado multivariado constituye una estrategia eficaz para restaurar la funcionalidad de Ferralsoles degradados, proporcionando un soporte técnico sólido para programas de recuperación en ambientes tropicales.

Palabras clave: aprendizaje automático; análisis multivariante; funcionalidad edáfica; infiltración de agua; resiliencia ambiental.

LISTA DE FIGURAS

CAPÍTULO II

LISTA DE TABELAS

CAPÍTULO I

CAPÍTULO II

SUMÁRIO

1 CAPÍTULO INICIAL

1.1 INTRODUÇÃO GERAL

A degradação dos solos em decorrência de ações antrópicas intensivas tornou-se um dos maiores desafios ambientais nas últimas décadas. Esse cenário é particularmente crítico em regiões tropicais, onde as condições climáticas e a fragilidade estrutural dos solos favorecem perdas aceleradas de matéria orgânica, erosão e redução da fertilidade (Lal, 2004). Entre os casos mais emblemáticos no Brasil destacam-se as áreas degradadas pela instalação de grandes empreendimentos hidrelétricos, onde a retirada da camada superficial para movimentação de terras expõe o subsolo e resulta em ambientes com baixa resiliência ecológica, elevada acidez, compactação e severa limitação ao desenvolvimento vegetal (Giacomo, 2013; Souto Filho *et al.*, 2017).

A recuperação desses ambientes exige estratégias de manejo capazes de restaurar simultaneamente os atributos físicos, químicos e biológicos do solo, visto que a funcionalidade edáfica depende da interação desses componentes (Anderson, 1988; Brady; Weil, 2008). Nesse contexto, a revegetação tem se consolidado como alternativa eficaz para promover a reabilitação do solo, sobretudo quando associada à aplicação de corretivos e de fontes orgânicas capazes de favorecer a ciclagem de nutrientes e o incremento da matéria orgânica (Wendling *et al.*, 2019).

Estudos de longo prazo demonstram que sistemas florestais restauradores podem aumentar a estabilidade de agregados, melhorar a porosidade, intensificar a atividade microbiana e recuperar gradualmente a fertilidade (Lourencetti, 2023). O uso de compostos orgânicos provenientes de resíduos industriais ou agroflorestais tem ganhado destaque nesse processo por acelerar a recuperação química e biológica do solo, ampliando os teores de carbono orgânico, a capacidade de troca catiônica e a disponibilidade de nutrientes às plantas (Parrotta, 1992; Bonini *et al.*, 2025).

A seleção de espécies florestais adaptadas ao ambiente degradado desempenha papel decisivo para o sucesso da restauração ecológica. Nesse sentido, a combinação entre *Eucalyptus urograndis*, espécie exótica de rápido crescimento que promove sombreamento, deposição de serrapilheira e proteção física do solo, e *Mabea fistulifera*, espécie nativa do Cerrado com elevada tolerância a ambientes pobres e capacidade de acúmulo de biomassa e nutrientes na superfície, tem se destacado como estratégia eficiente para recompor solos tropicalizados severamente degradados.

A interação entre essas espécies, associada à adubação mineral e à aplicação de diferentes doses de composto orgânico, tem demonstrado potencial para acelerar a recuperação ambiental ao melhorar substancialmente os atributos químicos e físicos do solo, incluindo incrementos nos teores de matéria orgânica, porosidade, capacidade de troca catiônica e retenção de água (Matos, 2022; Lourencetti, 2023).

O monitoramento da recuperação de áreas degradadas tem evoluído significativamente com o uso de técnicas estatísticas multivariadas, como Análise de Componentes Principais (ACP) e Redes Neurais Probabilísticas (RNP), que permitem interpretar a contribuição conjunta de múltiplos atributos do solo e identificar quais variáveis são mais sensíveis às mudanças promovidas pelo manejo (Jolliffe, 2002). Esse tipo de abordagem fornece um panorama integrado da qualidade do solo ao longo do processo de restauração, facilitando a comparação entre tratamentos e subsidiando decisões técnicas.

Diante desse panorama, a combinação de espécies florestais com adubação orgânica e mineral, analisada sob a perspectiva de modelos multivariados, representa uma estratégia promissora para reconstruir a funcionalidade ecológica de solos degradados. Para além da reabilitação produtiva, essa abordagem contribui para a restauração ambiental, o aumento do estoque de carbono, a ciclagem de nutrientes e a construção de ecossistemas mais resilientes, oferecendo bases científicas para a replicação de projetos de recuperação em larga escala.

1.2 OBJETIVOS

1.2.1 Objetivo geral

Avaliar a recuperação física e química de um Latossolo degradado sob diferentes doses de composto orgânico e espécies arbóreas, utilizando modelagem multivariada e inteligência artificial (ACP + RNP) para identificar os tratamentos mais eficazes na reabilitação da qualidade do solo.

1.2.2 Objetivos específicos

- Caracterizar os atributos físicos e químicos do solo após 14 anos de restauração sob *Mabea fistulifera* e *Eucalyptus urograndis*.
- Aplicar a Análise de Componentes Principais (ACP) para identificar variáveis determinantes na evolução da qualidade do solo.
- Utilizar Redes Neurais Probabilísticas (RNP) para classificar tratamentos e definir zonas ótimas de recuperação.
- Comparar a eficiência de doses crescentes de composto orgânico e adubação mineral sobre a funcionalidade do solo.
- Determinar o papel relativo de espécies nativas e exóticas na melhoria dos atributos edáficos.
- Integrar resultados físico-químicos em um modelo preditivo de qualidade do solo para apoiar práticas de restauração.
- Gerar recomendações técnicas aplicáveis a programas de recuperação de áreas degradadas em Latossolos tropicais.

1.3 HIPÓTESE

Supõe-se que a combinação entre o plantio de *Mabea fistulifera* e a aplicação de doses progressivas de composto orgânico, com ênfase em 20 Mg ha^{-1}, desencadeia processos sinérgicos de reconstrução da estrutura e da fertilidade de Latossolos degradados. Essa interação promoveria aumentos significativos em matéria orgânica, macroporosidade e capacidade de troca catiônica, além da redução da densidade do solo, conduzindo o ecossistema edáfico a um estado funcional mais próximo das condições ótimas de recuperação estimadas pelo modelo multivariado ACP+RNP.

1.4 MATERIAL E MÉTODOS

O projeto de pesquisa ocorreu no período entre 2023 e 2024. O experimento foi conduzido na Fazenda de Ensino, Pesquisa e Extensão da Unesp, Faculdade de

Engenharia, Campus de Ilha Solteira/SP, no município de Selvíria/MS (Figura 4). A fazenda está situada na região noroeste do Estado de São Paulo, entre as coordenadas 20°22' de latitude Sul e 51°22' de longitude Oeste, e altitude em torno de 327 metros acima do nível do mar, onde a temperatura média anual é de 23,5 °C e a precipitação média anual de 1370 mm. O solo característico da região de estudo é um textura franco argilo-arenosa e rico em sesquióxidos, e classificado como Latossolo Vermelho distrófico (Santos *et al.*, 2018).

1.4.1 Caracterização inicial da área experimental

A área estudada é uma área degradada, onde foi removida uma camada de 8,6 metros do perfil do solo original, sendo que o subsolo da área em estudo estava exposto desde 1969. A área de empréstimo surgiu da retirada de solo para a terraplanagem e fundação da barragem da usina Hidrelétrica de Ilha Solteira – SP, que teve o início de sua construção na década de 60 (Alves; Nascimento; Souza, 2012).

1.4.2 Delineamento experimental e tratamentos

O experimento consistiu em 12 tratamentos (Tabela 2), com 4 repetições por tratamento. No total, foram estabelecidas 48 parcelas de 15 m × 12 m, distribuídas em 4 blocos (um bloco por repetição), cada um com área de 2.160 m². Em cada parcela experimental, foram plantadas 40 árvores da mesma espécie, com espaçamento de 1,5 m entre plantas e 3 m entre linhas. Assim, cada bloco foi constituído por 240 árvores de *Eucalyptus urograndis* (clone "h-17") e 240 árvores de *Mabea fistulifera* Mart. (produzidas a partir de sementes). A distância entre os blocos foi de 3 m.

1.4.3 Avaliações

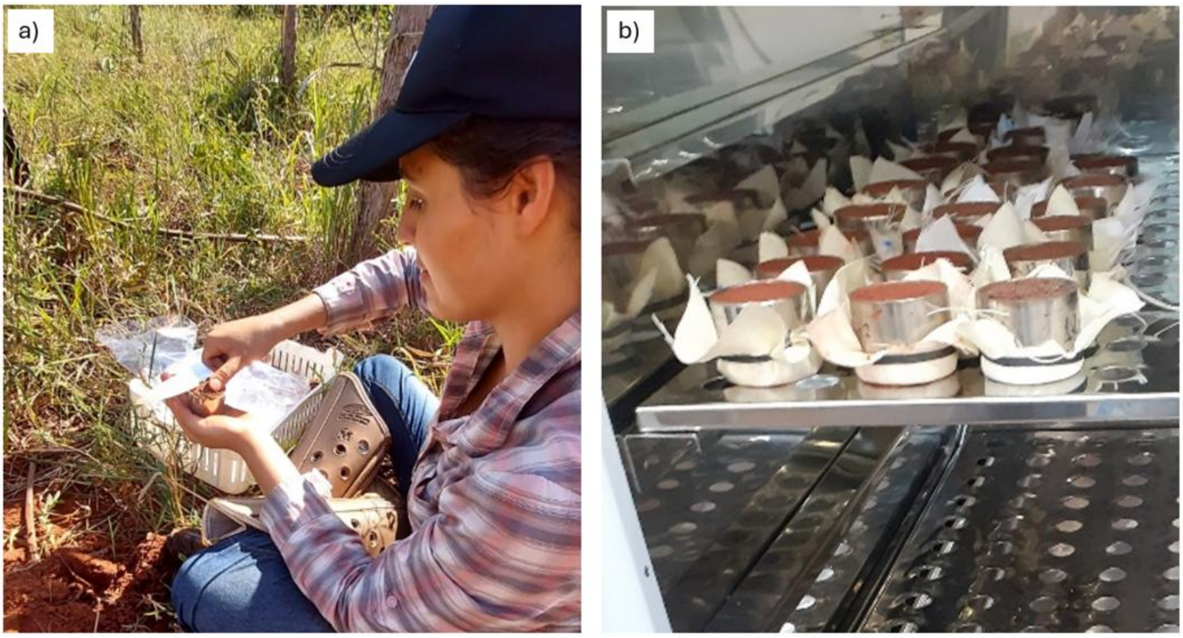
Foram avaliados atributos físicos e químicos do solo. Para as análises de solo foram coletadas amostras indeformadas e deformadas de solo, em três camadas de solo: 0,00–0,05; 0,05–0,10 e de 0,10–0,20 m.

As avaliações do solo foram: densidade do solo, porosidade total, macro e microporosidade do solo, taxa de infiltração, estabilidade de agregados do solo, diâmetro médio ponderado, matéria orgânica, pH e CTC.

1.4.4 Densidade do solo

A densidade do solo é um parâmetro fundamental para avaliar a qualidade física do solo, sendo um indicador direto de sua compactação. Ela é influenciada por processos naturais e pedogenéticos, mas a intervenção humana, com gestão da terra e das culturas, desempenha um papel crucial na alteração desse parâmetro (Haruna *et al.*, 2020). Para um solo de textura franco-argilosa, a densidade de solo se encontra entre 1,4 - 1,5g/cm³, acima disso é caracterizado como uma densidade crítica, a qual dificulta ou até impede o crescimento e desenvolvimento de raízes. Cada solo possui uma densidade crítica, e o principal fator determinante é sua classe (Cintra; Mielniczuk, 1983; Reichert; Reinert; Braida, 2023; Rosenberg, 1964). A obtenção da densidade do solo envolve duas etapas: massa e volume da amostra, e foi determinada pelo método do anel volumétrico (Teixeira *et al.*, 2017) (Fig. 1).

Figura 1 – Coleta e preparo de amostras para avaliação da densidade do solo



Fonte: Autoria própria.
Legenda: a) coleta de amostra indeformável em anel volumétrico, e b) secagem de anéis volumétricos para medição de densidade do solo.

1.4.5 Porosidade total, Microporosidade e Macroporosidade do solo

A porosidade total (Pt) foi determinada pelo método de saturação das amostras com água, considerando-se o volume total de poros ocupado após saturação completa. As amostras indeformadas, acondicionadas em cilindros volumétricos, tiveram suas superfícies superior e inferior niveladas. Na base do cilindro foi fixado um tecido permeável, previamente pesado, preso com liga elástica, constituindo o conjunto amostra–cilindro–tecido–liga.

O conjunto foi vedado com tampa de alumínio e colocado em bandeja plástica, à qual se adicionou água até aproximadamente 1 cm de altura da base do anel metálico. A saturação ocorreu por ascensão capilar até que a água atingisse o topo da amostra. Posteriormente, elevou-se o nível da água até próximo à borda do cilindro, mantendo-se o sistema em repouso por 12 horas.

Após esse período, o conjunto saturado foi pesado e, em seguida, levado à estufa a 105 °C até massa constante. Depois da secagem, as amostras foram resfriadas em dessecador e novamente pesadas. A porosidade total foi calculada conforme:

$$Pt = \frac{[(a-b)-(c-d)]}{e}$$

Onde:

Pt – Porosidade total, em $m^3.m^{-3}$

a – massa do conjunto amostra-cilindro-tecido-liga saturado, em kg.

b – massa do conjunto amostra-cilindro-tecido-liga seco a 105 °C, em kg.

c – massa do conjunto cilindro-tecido-liga saturado, em kg.

d – massa do conjunto cilindro-tecido-liga seco a 105 ºC, em kg.

e – volume total da amostra, em m^3.

O volume total da amostra corresponde ao volume interno do cilindro (Vc), estimado por:

$$V_c = \pi . r^2 . h$$

Onde:

V_c – volume do cilindro, em m3. **r** – raio do cilindro, em m. **h** – altura do cilindro, em m.

A microporosidade (Mi) foi obtida pelo método da mesa de tensão, utilizando-se uma coluna de água correspondente a 6 kPa (0,60 m). As amostras indeformadas foram previamente saturadas e submetidas à tensão até atingirem equilíbrio hídrico. Após estabilização, determinou-se por pesagem a massa de água retida. A microporosidade foi calculada por:

$$M_i = \frac{(a-b)}{c}$$

Onde:

Mi – microporosidade, em m3 m-3. **a** – massa do solo seco + água retida, após equilíbrio com um potencial de 6 kPa (60 cm de coluna de água), em g.

b – massa do solo seco a 105 ºC, em g.

c – volume total da amostra, em cm3 (nesse caso, assume-se que o volume total da amostra é igual ao volume do cilindro).

E a macroporosidade foi calculada por diferença entre a porosidade total e a microporosidade, a seguir:

$$M_a = \left(Pt - M_i\right)$$

Onde:

Ma – macroporosidade, em m3 m-3.

Pt – porosidade total, em m3 m-3.

Mi –microporosidade, em m^3 m^{-3}

Todos esses atributos seguiram a metodologia de TEIXEIRA *et al*. (2017).

1.4.6 Estabilidade de agregados (via úmida)

As partículas minerais e orgânicas do solo se organizam em arranjos estruturais que constituem sua estrutura. A associação dessas partículas em unidades maiores dá origem aos agregados, formados pelo processo de agregação, cuja dimensão, formato e grau de estabilidade variam conforme as forças de coesão e adesão atuantes nos pontos de contato entre as partículas (Lepsch, 2021).

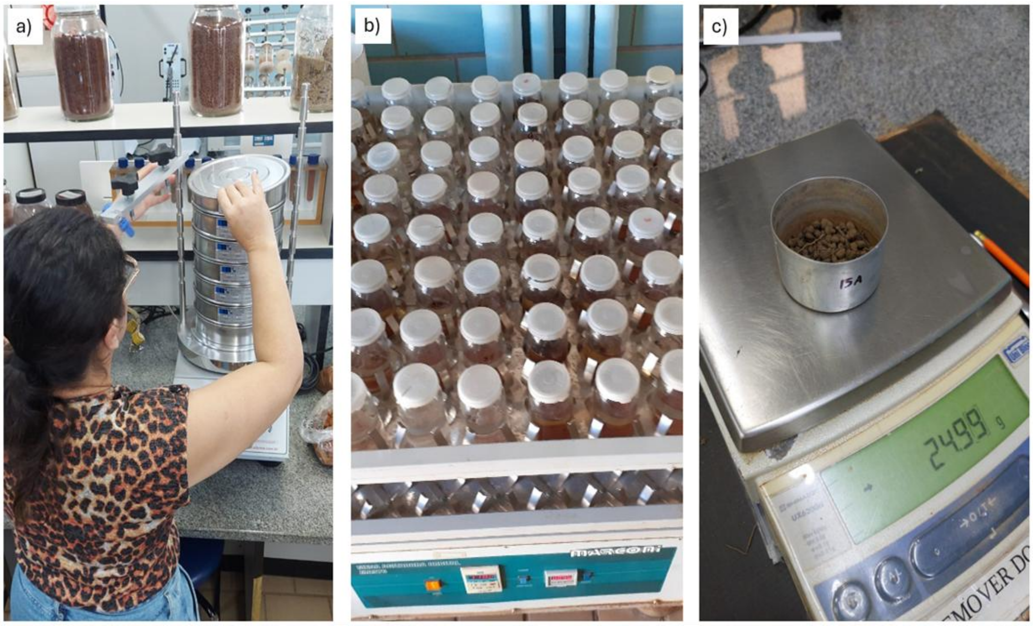
A qualidade estrutural do solo está intimamente relacionada às características de seus agregados, influenciando processos como infiltração e retenção de água, resistência mecânica ao crescimento radicular e suscetibilidade à erosão (Souza *et al.*, 2024). A estabilidade dos agregados expressa a capacidade dessas unidades estruturais de resistirem à desagregação, sendo fundamental para a manutenção da produtividade agrícola e para a mitigação de processos erosivos e de degradação ambiental (Amézketa, 1999).

Entre os principais indicadores utilizados para avaliar a qualidade da agregação destacam-se o diâmetro médio ponderado (DMP) e o índice de estabilidade de agregados (IEA). O DMP representa a proporção relativa de agregados distribuídos em diferentes classes de tamanho, sendo expresso em milímetros. Valores mais elevados indicam maior participação de agregados de maior diâmetro, condição associada a maior estabilidade estrutural, melhor infiltração e drenagem de água e maior resistência à erosão hídrica. O IEA, por sua vez, quantifica o grau de estabilidade da agregação do solo, variando de 0 a 100%. Valores iguais ou superiores a 70% são indicativos de elevada estabilidade estrutural (Castro Filho *et al.*, 2002; Castro Filho; Logan, 1991; Castro Filho; Muzilli; Podanoschi, 1998).

A estabilidade dos agregados exerce papel determinante no controle da erosão hídrica, especialmente em solos tropicais ácidos, pois a erodibilidade está diretamente relacionada à capacidade do solo de manter seus agregados íntegros sob ação da água. A manutenção de macroporos responsáveis por altas taxas de infiltração e adequada aeração depende dessa estabilidade estrutural. Adicionalmente, a agregação favorece o sequestro e a proteção do carbono no solo, enquanto a degradação estrutural contribui para a redução da matéria orgânica e das reservas de carbono (Castro Filho; Logan, 1991; Castro Filho; Muzilli; Podanoschi, 1998; Kemper; Rosenau, 1986; Lal; Kimble; Follett, 1997; Roth *et al.*, 1986).

A distribuição do tamanho dos agregados e sua estabilidade em água foram determinadas conforme o procedimento descrito por Angers, Bullock e Mehuys (2008).

Figura 2 – Preparo de amostras para avaliação de estabilidade de agregados



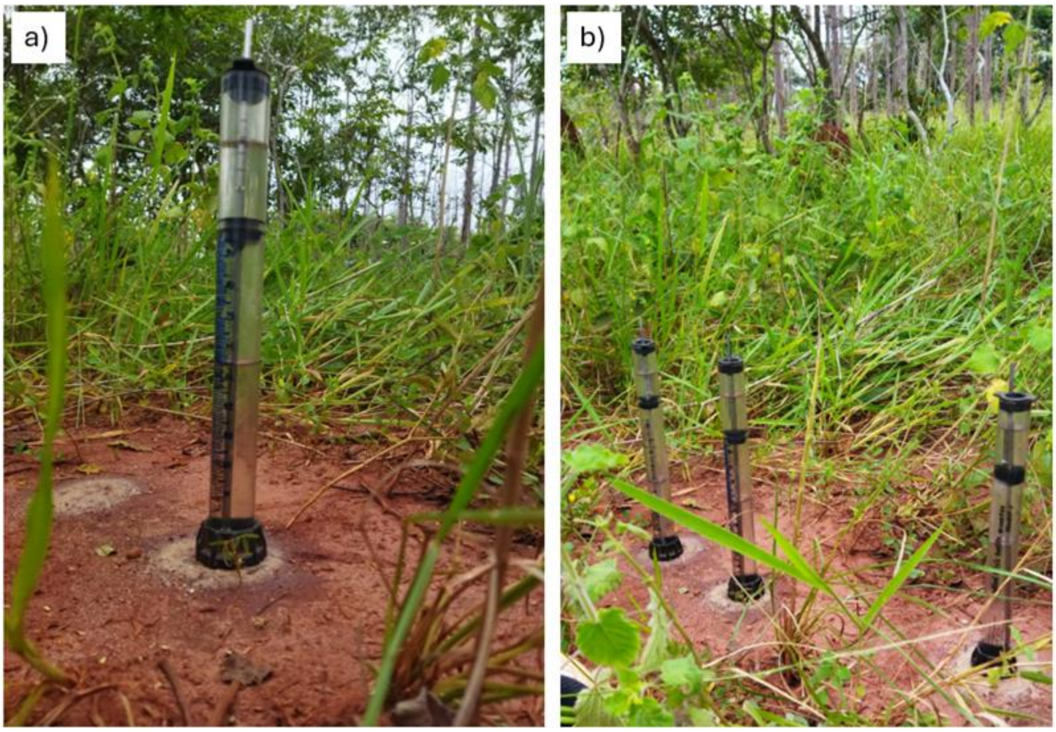
Fonte: Autoria própria.

1.4.7 Infiltração de água no solo

A infiltração de água representa um dos principais atributos hidrológicos do solo, influenciando diretamente o escoamento superficial, a lixiviação de nutrientes, os processos erosivos e a disponibilidade de água para plantas e reservatórios subterrâneos. Essa dinâmica é controlada por fatores em escala regional, como a topografia, e por condições locais, como umidade, teor de matéria orgânica, mineralogia, textura, estrutura e nível de compactação do solo (Francos *et al*., 2023).

Foram realizadas três medições da taxa de infiltração em cada parcela experimental, empregando-se o mini infiltrômetro de disco conforme o procedimento descrito por Zhang (1997) (Figura 3).

Figura 3 – Análise de infiltração de água no solo à campo



Fonte: Autoria própria.
Legenda: a) e b) análise de infiltração de água no solo utilizando mini infiltrômetro de disco.

1.4.8 Fertilidade do solo

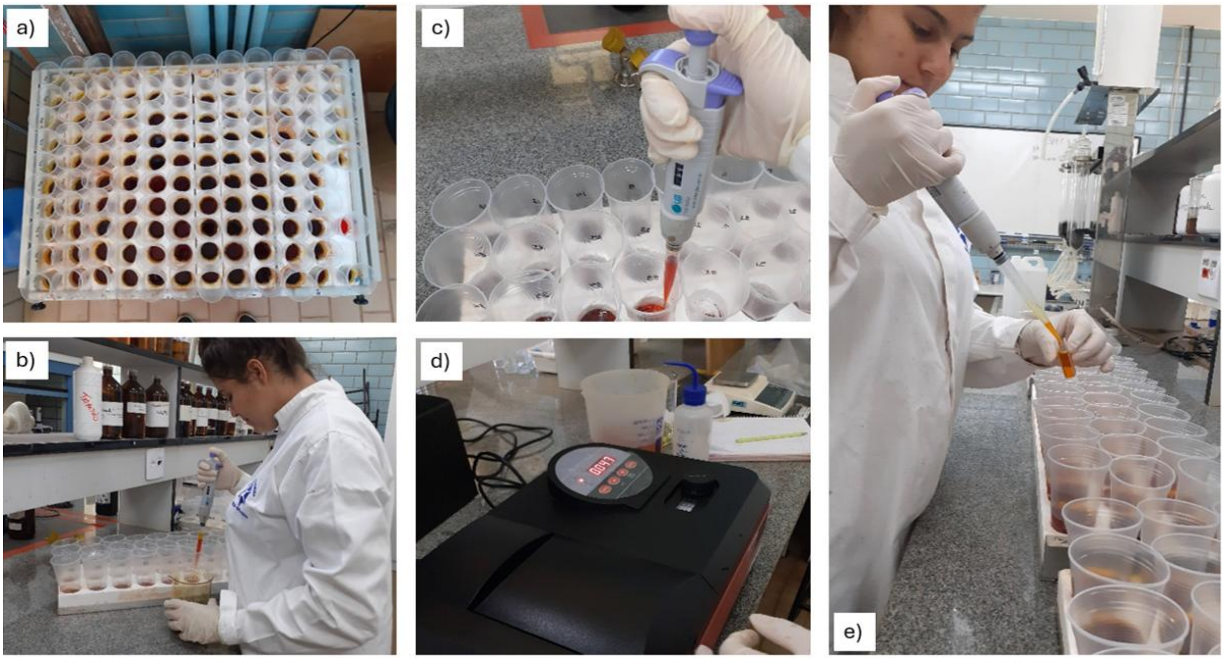
A avaliação da fertilidade do solo seguiu os procedimentos metodológicos estabelecidos por Raij *et al*. (2001). Foram determinados os valores de pH em solução de cloreto de cálcio, a acidez potencial (H + Al) a pH 7,0 e a capacidade de troca catiônica, calculada pela soma de bases (SB) acrescida de (H + Al), conforme a expressão CTC = SB + (H + Al).

1.4.9 Teor de matéria orgânica

A matéria orgânica do solo (MOS) é considerada um dos principais indicadores de qualidade e funcionalidade edáfica, pois atua como agente cimentante na formação de agregados, contribui para a redução da densidade e aumenta a elasticidade e a resiliência estrutural, promovendo melhorias nas propriedades físicas do solo. Dessa forma, práticas de manejo devem priorizar sua manutenção ou incremento (Blanco-Canqui; Benjamin, 2015; Liu *et al*., 2006).

Para solos de textura média, os teores adequados de MOS situam-se entre 16 e 30 g dm^{-3} (Cantarella *et al*., 2022). A determinação do teor de matéria orgânica foi realizada conforme a metodologia descrita por Raij *et al*. (2001) (Figura 4).

Figura 4 – Análise de teor de matéria orgânica do solo



Fonte: Autoria própria.
Legenda: a) agitação das amostras; b) e c) pipetagem da solução (dicromato de sódio + ácido sulfúrico) nas amostras; d) leitura por espectofotometria visível; e) pipetagem para leitura no espectofotômetro visível

1.4.10 Análise de dados

Os dados provenientes dos dois artigos que compõem esta tese foram submetidos a procedimentos estatísticos adequados às respectivas abordagens experimentais e analíticas. Inicialmente, todos os conjuntos de dados foram avaliados quanto aos pressupostos de normalidade dos resíduos e homogeneidade das variâncias.

No primeiro artigo, as análises estatísticas foram realizadas utilizando o software Statgraphics. Os dados foram submetidos à análise de variância (ANOVA) e, quando constatado efeito significativo dos tratamentos, procedeu-se à comparação de médias por meio de testes estatísticos apropriados, adotando-se o nível de significância de 5%.

No segundo artigo, as análises foram conduzidas no ambiente computacional MATLAB, no qual foram aplicados métodos estatísticos e matemáticos compatíveis com a estrutura e complexidade dos dados. Foram empregadas análises multivariadas e rotinas computacionais específicas para a avaliação das relações entre as variáveis, identificação de padrões e interpretação integrada dos resultados.

Em ambos os estudos, as análises foram realizadas de forma a assegurar consistência estatística, reprodutibilidade dos resultados e suporte quantitativo às interpretações discutidas ao longo da tese.

REFERÊNCIAS

ALVES, M. C.; NASCIMENTO, V. DO; SOUZA, Z. M. DE. Recuperação em área de empréstimo usada para construção de usina hidrelétrica. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 16, n. 8, p. 887–893, ago. 2012.

AMÉZKETA, E. Soil Aggregate Stability: A Review. **Journal of Sustainable Agriculture**, v. 14, n. 2–3, p. 83–151, 16 jul. 1999.

ANDERSON, J. M. Spatiotemporal effects of invertebrates on soil processes. **Biology and Fertility of Soils**, v. 6, p. 216–227, 1988. DOI: 10.1007/BF00260818.

ANGERS, D. A.; BULLOCK, M. S.; MEHUYS, G. R. Aggregate stability to water. *In*: CARTER, M. R.; GREGORICH, E. G. (ed.). **Soil sampling and methods of analysis**. 2. ed. Boca Raton: CRC Press; Canadian Society of Soil Science, 2008. p. 811–819.

BLANCO-CANQUI, H.; BENJAMIN, J. G. Impacts of soil organic carbon on soil physical behavior. *In:* LOGSDON, S.; BERLI, M.; HORN, R. (ed.). **Advances in Agricultural Systems Modeling**. Madison: American Society of Agronomy; Soil Science Society of America, 2013. p. 11–40.

BONINI, C. S. B.; DELFIM, J. J.; LOURENCETTI, J.; NANTES, L. S.; SANTOS, M. A.; SOUZA, J. A. L. Residual effects of cover crops and soil amendments on carbon content and stock in a degraded soil. **Engenharia Agrícola**, v. 45, n. spe 1, e20240188, 2025. DOI: 10.1590/1809-4430-Eng.Agric.v45nespe120240188/2025.

BRADY, N. C.; WEIL, R. R. **The nature and properties of soils**. 14. ed. Upper Saddle River: Pearson Prentice Hall, 2008.

CANTARELLA, H.; QUAGGIO, J. A.; MATTOS JR, D.; BOARETTO, R. M.; RAIJ, B. V. **Boletim 100:** recomendações de adubação e calagem para o estado de São Paulo. Campinas: Instituto Agronômico (IAC), 2022.

CASTRO FILHO, C.; LOGAN, T. J. Liming effects on the stability and erodibility of some brazilian oxisols. **Soil Science Society of America Journal**, v. 55, n. 5, p. 1407–1413, set. 1991.

CASTRO FILHO, C.; LOURENÇO, A.; DE F. GUIMARÃES, M.; FONSECA, I. C. B. Aggregate stability under different soil management systems in a red latosol in the state of Parana, Brazil. **Soil and Tillage Research**, v. 65, n. 1, p. 45–51, abr. 2002.

CASTRO FILHO, C.; MUZILLI, O.; PODANOSCHI, A. L. Estabilidade dos agregados e sua relação com o teor de carbono orgânico num latossolo roxo distrófico, em função de sistemas de plantio, rotações de culturas e métodos de preparo das amostras. **Revista Brasileira de Ciência do Solo**, v. 22, n. 3, p. 527–538, set. 1998.

CINTRA, F. L. D.; MIELNICZUK, J. Potencial de algumas espécies para a recuperação de solos com propriedades físicas degradadas. **Revista Brasileira de Ciência do Solo**. v. 7, p. 197–201, 1983.

GIACOMO, R. G. **Recuperação do solo usando composto produzido com resíduo da extração de celulose.** 2013. Tese (Doutorado) – Universidade Estadual Paulista, Ilha Solteira, 2013.

HARUNA, S. I.; ANDERSON, S. H.; UDAWATTA, R. P.; GANTZER, C. J.; PHILLIPS, N. C.; CUI, S.; GAO, Y. Improving soil physical properties through the use of cover crops: A review. **Agrosystems, Geosciences & Environment**, v. 3, n. 1, p. e20105, jan. 2020.

JOLLIFFE, I. T.; CADIMA, J. Principal component analysis: a review and recent developments. **Philosophical Transactions of the Royal Society A**, v. 374, art. 20150202, 2016. DOI: 10.1098/rsta.2015.0202.

KEMPER, W. D.; ROSENAU, R. C. Aggregate stability and size distribution. *In:* **Methods of soil analysis: physical and mineralogical methods**. Madison: Soil Science Society of America, 1986. p. 425–442.

LAL, R. Soil carbon sequestration to mitigate climate change. **Geoderma**, v. 123, n. 1–2, p. 1–22, nov. 2004.

LAL, R.; KIMBLE, J.; FOLLETT, R. F. Pedospheric processes and the carbon cycle. *In:* **Methods for assessment of soil degradation**. Boca Raton: CRC Press, 1997. p. 1–8.

LEPSCH, I. F. **19 lições de pedologia**. 2. ed. São Paulo, SP: Oficina de Textos, 2021.

LOURENCETTI, J.; BONINI, C. dos S. B.; ANDREOTTI, M.; ALVES, M. C.; BONINI NETO, A.; SANTOS, M. A.; BARRETTO, V. C. de M.; DE FIGUEREDO, R. W. R. Evolution of soil chemical fertility in an area under recovery for 30 years with anthropic intervention. **Sustainability**, v. 15, art. 10344, 2023. DOI: 10.3390/su151310344.

MATOS, A. M. S. **Qualidade de um Argissolo Vermelho-Amarelo distrófico em sistema ILPF.** 2022. 125 f. Dissertação (Mestrado em Agronomia) – Universidade Estadual Paulista, Ilha Solteira, 2022.

PARROTTA, J. A. The role of plantation forests in rehabilitating degraded tropical ecosystems. **Agriculture, Ecosystems and Environment**, v. 41, p. 115–133, 1992.

RAIJ, B. VAN; ANDRADE, J. C. DE; CANTARELLA, H.; QUAGGIO, J. A. **Análise química para avaliação da fertilidade de solos tropicais**. Campinas: Instituto Agronômico (IAC), 2001.

REICHERT, J. M.; REINERT, D. J.; BRAIDA, J. A. Qualidade dos solos e sustentabilidade de sistemas agrícolas. **Ciência & Ambiente**. v. 27, p. 29–48, 2023.

ROSENBERG, N. J. Response of plants to the physical effects of soil compaction. *In:* **Advances in Agronomy**. v. 16, p. 181–196, 1964.

ROTH, C. H.; PAVAN, M. A.; CHAVES, J. C. D.; MEYER, B.; FREDE, H. G. Efeito das aplicações de calcário e gesso sobre a estabilidade de agregados e infiltrabilidade de água em um Latossolo Roxo cultivado com cafeeiros. **Revista Brasileira de Ciência do Solo**. v. 10, p. 163–166, 1986.

SANTOS, H. G.; JACOMINE, P. K. T.; ANJOS, L. H. C.; OLIVEIRA, V. A.; LUMBRERAS, J. F.; COELHO, M. R.; ALMEIDA, J. A.; ARAÚJO FILHO, J. C.; OLIVEIRA, J. B.; CUNHA, T. J. F. **Sistema brasileiro de classificação de solos**. Brasília: Embrapa, 2018.

SOUTO FILHO, S. N.; ALVES, M. C.; MONREAL, C. M.; BONINI, C. dos S. B. Nanoparticles and nanostructure morphology of a Red Latosol in rehabilitation. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 21, p. 530–536, 2017. DOI: 10.1590/1807-1929/agriambi.v21n8p530-536.

TEIXEIRA, P. C.; DONAGEMMA, G. K.; FONTANA, A.; TEIXEIRA, W. G. **Manual de métodos de análise de solo**. 3. ed. Brasília: Embrapa, 2017.

WENDLING, M.; *et al*. Effect of species identity and diversity on biomass production and its stability in cover crop mixtures. **Agriculture, Ecosystems & Environment,** v. 281, p. 81–91, 2019. DOI: 10.1016/j.agee.2019.04.032.

2 CAPÍTULO I - Application of Principal Component Analysis and Probabilistic Neural Networks in Ferralsols Recovery Evaluation through planting of Mabea fistulifera and Eucalyptus urograndis

Melissa Alexandre Santos[1], Jorge João Delfim[2], Laura Silva Nantes[2], José Augusto Liberato de Souza[1], Gabriel Augusto da Silva Lunardelli[2], Dayara Vivian Alvares[2], Carolina dos Santos Batista Bonini[2], Borja Velázquez-Martí[3]

Abstract

This study presents an innovative assessment model for analyzing the evolution of degraded soils subjected to different reclamation strategies. The proposal combines statistical and artificial intelligence tools to jointly integrate multiple physical and chemical soil properties, allowing for a more synthetic view of the processes. The model uses principal component analysis to synthesize information on the most relevant variables and subsequently applies probabilistic neural networks to identify the most likely values of the principal components when a given soil reclamation treatment is applied. Once the optimal ranges for a successfully reclaimed soil have been defined, the developed numerical methods are applied, defining an area considered optimal in the principal component diagram. The most appropriate treatment is considered to be the one most likely to occupy the optimal area. The methodology was applied to a dystrophic Red Oxisol degraded by the construction of the Ilha Solteira Hydroelectric Plant in Brazil, where a long-term experiment with two tree species (*Eucalyptus urograndis* and *Mabea fistulifera*) and different doses of organic and mineral fertilization was established in 2010. The results show that the combination of *Mabea fistulifera* with the application of 20 Mg ha^{-1} of compost significantly improves organic matter, porosity, and cation exchange capacity in the surface soil horizons, generating more favorable conditions for plant growth in the long term. Beyond the specific results, this multivariate model represents a useful tool for evaluating the effectiveness of long-term soil restoration programs, providing objective criteria that can guide decision-making in projects for ecological recovery and sustainable management of degraded lands.

Keywords: Soil reclamation; Oxisols; Principal Component Analysis (PCA); Probabilistic Neural Networks (PNN).

[1] Graduate Program in Agronomy, Faculty of Engineering, São Paulo State University (Unesp), Ilha Solteira 15385-000, SP, Brazil.
[2] Department of Plant Production, Faculty of Agrarian and Technological Sciences, São Paulo State University (Unesp), Dracena 17915-899, SP, Brazil.
[3] Departamento de Ingeniería Rural y Agroalimentaria, Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain

2.1 INTRODUCTION

Since the 1960s, the construction of large hydroelectric plants in Brazil has driven energy development, but has also generated large areas degraded by the extraction of materials for the projects. These soils, devoid of vegetation and with a low capacity for natural regeneration, face serious challenges in recovering their structure, fertility, and ecological functionality (Bonini *et al.*, 2023; Giacomo, 2013; Souto Filho *et al.*, 2017).

Degradation causes profound local impacts, such as erosion, biodiversity loss, and the alteration of hydrological cycles. Restoring these areas requires strategies that combine physical, chemical, and biological amendments, as well as assessment tools capable of integrating multiple soil indicators (Soto Boni *et al.*, 2022). Particularly in tropical regions, the usefulness of physical and chemical indicators, along with practices such as the use of compost or adapted tree species, has been highlighted for monitoring and guiding recovery (Lourencetti *et al.*, 2023a; Suzuki *et al.*, 2022).

Soil is a complex system in which properties such as organic matter, water retention, structure, and nutrient availability interact interdependently. Therefore, monitoring its recovery requires multivariate approaches that capture these interactions. Statistical and artificial intelligence methods, such as Principal Component Analysis (PCA) and Artificial Neural Networks (ANN), offer advantages in this regard: the former allows reducing data complexity and highlighting relevant patterns, while ANNs capture nonlinear relationships and improve predictive capacity (Bonini Neto *et al.*, 2019; Dolácio *et al.*, 2020; Greenacre *et al.*, 2022; Sabattini *et al.*, 2019). For example, improvements in organic matter can affect water retention capacity and soil structure, while pH or the concentration of certain minerals can regulate nutrient availability for plants and microorganisms (McCauley *et al.*, 2009; Shoumik *et al.*, 2025). This interdependence means that soil behavior in response to a change is the result of a network of complex relationships, where no variable acts independently or can be considered the sole explanation for the observed phenomenon.

Therefore, evaluating the evolution of soil in these contexts requires integrated analysis tools, such as multivariable modeling, the use of soil quality indices, or analysis with combinations of mathematical models. These methods allow for capturing interactions between the different dimensions of the system, providing a more complete and realistic understanding of recovery processes.

In the context of degraded soil restoration, technological tools such as Principal Component Analysis (PCA) and Artificial Neural Networks (ANN) have shown great potential. PCA allows for the dimensionality reduction of large data sets, identifying critical patterns without significant information loss (Jolliffe; Cadima, 2016; Palazzolo *et al.*, 2023). ANN offers the ability to model non-linear relationships, providing greater precision in predicting soil attributes and optimizing the management of natural resources (Bonini Neto *et al.*, 2019; Sun *et al.*, 2022). The combination of these approaches allows for solid analyses, integrating artificial intelligence and multivariate statistics, representing a significant advance in soil science.

Methods like PCA are often used in specific studies, while the practical application of artificial intelligence in soil management and conservation remains in its infancy (Lu *et al.*, 2022; Palazzolo *et al.*, 2023). This gap reinforces the need to develop interdisciplinary approaches that combine technological innovation with traditional knowledge, aiming to promote effective and lasting recovery of degraded areas.

Within this context, *Mabea fistulifera* Mart. (Euphorbiaceae) emerges as a promising species for ecological rehabilitation. It is a pioneer, deciduous, and fast-growing tree widely distributed across Brazilian biomes—including the Atlantic Forest, Amazon, Cerrado, and Caatinga—and is particularly well adapted to sandy soils and disturbed environments in the North, Northeast, Central-West, and Southeast regions. The species reaches 4–12 meters in height, develops a dense crown, and has leaves of approximately 12 cm. Its fruits open explosively, promoting zoochoric dispersal. Flowering occurs mainly between February and June, offering abundant nectar and pollen that attract diverse fauna, including primates, birds, bats, and marsupials (Assumpção, 1981; Ferrari; Strier, 1992; Vieira *et al.*, 1991; Olmos; Boulhosa, 2000).

M. fistulifera has multiple uses, including light timber production, beekeeping potential, extractive applications for biodiesel, medicinal uses, and seed commercialization. Owing to its hardiness and capacity to colonize open and degraded environments, the species is recommended for ecological restoration programs. It contributes to soil protection, organic matter accumulation, nutrient cycling, and biological activity—processes essential for environmental recovery (Perin, 2001; Duda *et al.*, 2003). Its high floral resource availability and the resulting attraction of pollinators and seed dispersers promote recolonization and plant succession, making *M. fistulifera* a strategic species for the rehabilitation of degraded areas by improving soil conditions and facilitating key ecological processes (Leal Filho; Borges, 1992).

Objective of the study: To propose an innovative methodology based on multivariate analysis using Principal Component Analysis (PCA), complemented by Probabilistic Neural Networks (ANN), as a system for evaluating the evolution of the physical and chemical properties of degraded soils. This methodology was applied to assess the recovery of a degraded Oxisol subjected to different doses of mineral fertilization and compost, combined with the planting of native and exotic tree species (Mabea fistulifera and Eucalyptus urograndis), providing solid criteria to guide ecological restoration programs and the sustainable management of degraded areas.

2.2 MATERIALS AND METHODS

2.2.1 Experimental design

The experiment was carried out at the Teaching, Research and Extension Farm of UNESP, in Selvíria, MS, Brazil, located at an altitude of 327 m, between coordinates 51°22' W and 20°22' S, on the bank of the Paraná River (Figure 1). The region has a humid tropical climate (Aw), with rainy season in summer and dry season in winter, average annual precipitation of 1370 mm, average temperature of 23.5°C and relative air humidity between 70% and 80% (Koeppen, 1948).

Figure 1. Location of the degraded area (A) and the average layer of decapitated soil (B) at the Teaching, Research, and Extension Farm of the Faculty of Engineering, Ilha Solteira Campus.

Source: Giacomo, 2013.

The experimental area, characterized as Oxisol (Typic Hapludox) according to USDA Soil Taxonomy (Soil Survey Staff, 2022), and Ferralsol following the WRB classification (IUSS Working Group WRB, 2022) that was decapitated for the construction of the Ilha Solteira Hydroelectric Plant (1969), was subjected to an experiment starting in 2010, using a randomized block design with subdivided plots.

To mechanically decompact the soil, in December 2009, a subsoiling was done transversally to a depth of 0.40 m, followed by light harrowing. In February 2010, nutritional amendments were applied to the experimental plots, either compost or mineral fertilization, according to the chosen treatment. The amendments were manually distributed in each plot, ensuring uniform application. The materials were then incorporated into the soil using deep plowing, covering the entire treated area, aiming to favor its integration and availability for crops.

For planting, another subsoiling was done with a single rod to 0.50 m, where the plants were introduced. The introduction of two tree species in the plots was evaluated: one exotic (*Eucalyptus urograndis*) and one native to the Cerrado (*Mabea fistulifera* Mart.) combined with mineral fertilizer and five different doses of organic compost resulting from the composting of cellulosic waste (Garcia; Corrêa, 2024).

For mineral fertilization, the recommendations for *Mabea fistulifera* from the native seedling nursery of Companhia Energética de São Paulo (CESP), Jupiá unit, were followed, 100 g of 8-28-16 formula per plant was applied, corresponding to 166.70 kg ha^{-1}, followed by fertilization at 60 days with 48.8 g of urea per plant (81.45 kg ha^{-1}) and 16.70 g of potassium chloride (KCl) per plant (27.80 kg ha^{-1}).

For organic fertilization, a compost composed of a mixture of sludge, sand, limestone, ash, and other residues from the industrial cellulose extraction process was used (Table 1). It underwent 30 days of composting in outdoor rows, with periodic mechanical turning. The organic compost used was supplied by the Composting Center of the Ambitec Group (International Paper Unit, Mogi Guaçu, SP).

Table 1. Compost characterization from cellulose production waste, used as a nutrient source in the experiment conducted at the Teaching, Research, and Extension Farm of the Faculty of Engineering, located in the municipality of Selvíria, MS.

Chemical attribute	Unit	Value
pH	---	9.5
C/N		29.7
C organic	$g\ kg^{-1}$	186
N		6.3
P		2.4
K		5.95
Ca		86.9
Mg		3.8
S		1.8
B	$mg\ kg^{-1}$	30.3
Cu		14.3
Fe		5458
Mn		845
Zn		27.9
Na		1348

Source: Giacomo, 2013

The experiment consisted of 12 treatments (Table 2) with 4 repetitions per treatment. In total, there were 48 plots of 15 m x 12 m, divided into 4 blocks (one block per repetition) of 2160 m² each. In each experimental plot, 40 trees of the same species were planted with a planting spacing of 1.5 m between plants and 3 m between rows.

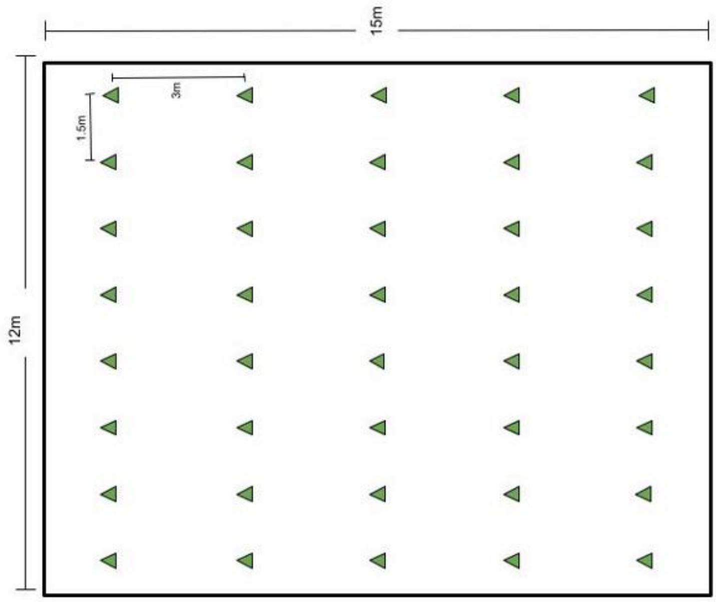
Therefore, each block consisted of 240 *Eucalyptus urograndis* trees (clone "h-17") and 240 *Mabea fistulifera* Mart. trees (produced from seeds). The distance between each block was 3 m.

Figure 2. Placement of trees on the plot

X	X	X	X	X
X	1	2	3	X
X	4	5	6	X
X	7	8	9	X
X	10	11	12	X
X	13	14	15	X
X	16	17	18	X
X	X	X	X	X

1-18 usable area

X – borders



Source: Elaborated by the author

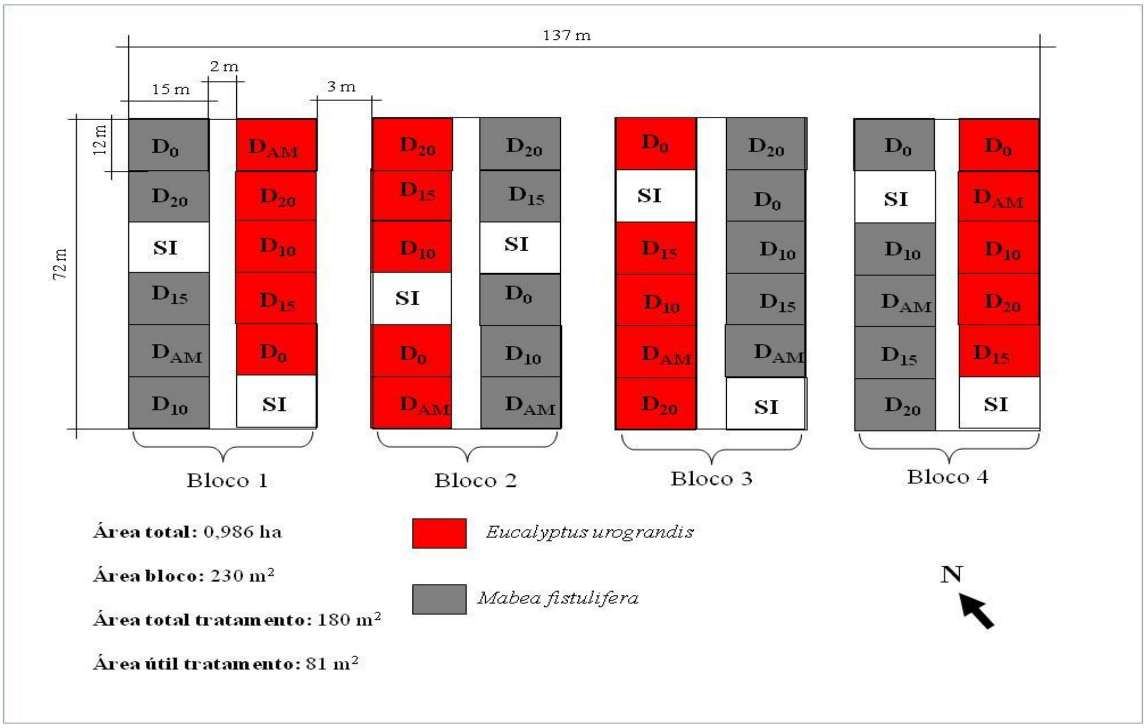
Table 2. Definition of treatments

Treatment		Plant species	Fertilizer
T1	D0 EU	*Eucalyptus urograndis*	D0 - Areas with initial tillage but no fertilization
T2	D10 EU	*Eucalyptus urograndis*	D10 - Areas with initial tillage and 10 Mg ha-1 of compost
T3	D15 EU	*Eucalyptus urograndis*	D15 - Areas with initial tillage and 15 Mg ha-1 of compost
T4	D20 EU	*Eucalyptus urograndis*	D20 - Areas with initial tillage and 20 Mg ha-1 of compost
T5	DAM EU	*Eucalyptus urograndis*	DAM - Areas with initial tillage and mineral fertilization
T6	SI EU	area without intervention *Eucalyptus urograndis*	SI - Without any physical or chemical management
T7	D0 CP	*Mabea fistulifera* Mart.	D0 - Areas with initial tillage but no fertilization

T8	D10 CP	*Mabea fistulifera* Mart.	D10 - Areas with initial tillage and 10 Mg ha-1 of compost
T9	D15 CP	*Mabea fistulifera* Mart.	D15 - Areas with initial tillage and 15 Mg ha-1 of compost
T10	D20 CP	*Mabea fistulifera* Mart.	D20 - Areas with initial tillage and 20 Mg ha-1 of compost
T11	DAM CP	*Mabea fistulifera* Mart.	DAM - Areas with initial tillage and mineral fertilization
T12	SI CP	area without intervention *Mabea fistulifera*	SI - Without any physical or chemical management

Source: Elaborated by the author

Figure 3. Sketch of the experiment implemented at the Fazenda de Ensino e Pesquisa da Faculdade de Engenharia, Selvíria.



Source: Giacomo, 2013

2.2.2 Soil Analysis

A characterization of the area was conducted in 2009 before the start of the experiment. The bulk density, silt content, sand and clay content/texture, and chemical analysis of the soil down to 0.40 m were analyzed.

Table 3. Theoretical values of phosphorus (P), potassium (K), calcium (Ca), magnesium (Mg), hydrogen + aluminum (H+Al), aluminum (Al), organic matter; cation exchange capacity (CEC), V% value (V), pH of the soil before the implementation of the experiment, Selvíria, MS, 2009.

Depth	pH	Ca	Mg	K	Al	H + Al	CTC	P - Resina	V	MO
m				mmol$_c$.dm^{-3}				mg.dm^{-3}	%	g.dm^{-3}
0.0 – 0.05	5.4	10	10	2.2	0	16	38.2	3	58	14
0.05 – 0.10	5.4	9	8	1.3	0	15	33.3	3	55	12
0.10 – 0.20	5.6	7	5	0.7	0	13	25.7	3	49	8
0.20 – 0.40	6.1	6	5	0.7	0	12	23.7	3	49	5

Source: Giacomo, 2013

Table 4. Values of bulk density (BD), sand, silt, clay and soil textural class before the implementation of the experiment, Selvíria, MS, 2009.

Depth	BD	Sand	Silt	Clay	Textural Class
m	kg.dm^{-3}	--------------g.kg^{-1}-------------			
0.00 - 0.05	1.67	645	84	271	*Sandy clay loam*
0.05 - 0.10	1.58	635	76	289	*Sandy clay loam*
0.10 - 0.20	1.65	633	89	278	*Sandy clay loam*
0.20 - 0.40	1.70	616	90	294	*Sandy clay loam*

Source: Giacomo, 2013

In February 2023, soil samples were collected at three depths (0.00–0.05; 0.05–0.10; 0.10–0.20 m). The samples were taken from the central area, shaded in Figure 2. For soil analyses in each plot, three samples were taken at each of the three depths. This means that each plot has nine soil analyses. This means that each treatment has 432 analyses.

12 treatments x 4 plots/treatment x 3 depths/plot x 3 replicates = 432 analyses.

The following physical properties were measured:

— *Bulk Density* (kg/dm³): The determination of bulk density involves two steps: mass and volume of the sample, and it was determined using the volumetric ring method (Teixeira *et al.*, 2017).
— *Total Porosity, Macro and Microporosity* ($m^3 \cdot m^{-3}$): Total porosity was determined by soil saturation (volume of total pores occupied by water); microporosity was measured using the tension table method with a water column of 0.060 kPa; and macroporosity was calculated as the difference between total porosity and microporosity, according to (Teixeira *et al.*, 2017).
— *Water Infiltration* (mm/h): Three evaluations were conducted in each experimental plot using the minidisk infiltrometer (Zhang, 1997). The device was read every 30 seconds, adjusting to a water suction of 0.05 m until a constant infiltration rate was reached, and the value was represented by the arithmetic mean of these measurements.
— The chemical properties measured in each soil layer were as follows:
— Organic Matter: The analysis of organic matter content, CEC (Cation Exchange Capacity), and CECEffective was carried out according to the methodology described by (Raij *et al.,* 2001).
— CEC and CECEffective: The CEC ($mmolc \cdot dm^{-3}$) refers to the total capacity of the soil to retain cations, influenced by texture, organic matter, and mineral composition. The Effective CEC ($mmolc \cdot dm^{-3}$) reflects the cation exchange capacity under specific field conditions, considering the natural pH ($mmolc \cdot dm^{-3}$) and ionic strength. It was determined through extraction solutions that efficiently displace exchangeable cations (Teixeira et al., 2017). The Effective CEC presents a practical evaluation of nutrient availability under real environmental conditions, which is crucial for soil management and sustainable agricultural practices. On the other hand, CEC provides a theoretical measure of the cation retention potential, considering pH 7.0 in the laboratory (Demattê; Demattê, 2024)

Figure 4 presents the timeline used to monitor soil evolution in the experiment. It is important to note that there is a considerable gap of 13 years between the establishment of the experiment in 2010 and the data analysis in 2023. This long period of time has been carefully considered, as it offers a unique and valuable perspective on the long-term effects of the treatments applied for soil restoration.

This temporal aspect, while a strength in terms of evaluating the sustainability and effectiveness of restoration practices over time, is also recognized as a potential limitation. Over the 13 years, factors such as annual climatic variations and changes in soil management practices could have influenced the observed results, affecting the consistency or magnitude of the treatment effects.

However, rather than perceiving this time lag as a drawback, we believe that the long follow-up time adds significant value to our analysis. Observing the effects over more than a decade provides a more robust and realistic view of how restoration strategies perform over the long term, rather than relying solely on short-term results that may not reflect the full dynamics of soil restoration processes.

Therefore, while we recognize that this long interval may have introduced some external variations that affected the results, we interpret this as reinforcing the applicability and relevance of our findings for long-term restoration programs, providing a solid basis for assessing the viability of soil management practices in the future.

Figure 4. Timeline used to monitor soil evolution in the experiment.

Source: Elaborated by the author

2.2.3 Mathematical Modeling

The data collected in February 2023 were analyzed using Principal Component Analysis (PCA), combined with Probabilistic Neural Networks (PNN).

Initially, each soil variable x_i (bulk density, total porosity, macro and microporosity, etc.) is standardized using equation (1). That is, by subtracting the mean of the variable ($\bar{x}$), and dividing by the standard deviation of the variable (σ_x).

$$Z_{xi} = \frac{x_i - \bar{x}}{\sigma_x} \tag{1}$$

The principal components are obtained as a linear combination of the standardized physical and chemical variables of the soils, in such a way that they explain the maximum variability of the population, equations (2) and (3) where the β_{1i} are the weights of the principal component 1 ($PC1$) for each variable x_i, and the β_{2i} are the weights of the principal component 2 ($PC2$) for each variable x_i.

$$PC1 = \beta_{11} \cdot Z_{x1} + \beta_{12} \cdot Z_{x2} + \beta_{13} \cdot Z_{x3} + \cdots + \beta_{1n} \cdot Z_{xn} \quad (2)$$

$$PC2 = \beta_{21} \cdot Z_{x1} + \beta_{22} \cdot Z_{x2} + \beta_{23} \cdot Z_{x3} + \cdots + \beta_{2n} \cdot Z_{xn} \quad (3)$$

Each soil sample I analyzed from a given treatment (T_k) is represented by a point on the Cartesian diagram formed by $\overrightarrow{PC_i}$ $(PC1_i, PC2_i)$. The values of the principal components are entered into a probabilistic neural network that provides the areas where the values of $\overrightarrow{PC_i}$ of a given treatment are most likely. This probability is calculated with equation (4).

$$P\left(\overrightarrow{PC}_i \mid T_k\right) = \frac{1}{(2\pi)^{d/2}\sigma^d} \sum_{j=1}^{r} exp\left(- \frac{\left\| \overrightarrow{PC_i} - \overrightarrow{PC_{jk}} \right\|^2}{2\sigma^2} \right) \quad (4)$$

T_k represents each treatment, $T_k \in \{T_1, T_2, T_3, \dots, T_{12}\}$. On the other hand, $\overrightarrow{CP_i}$ is each point in the diagram $(PC1_i, PC2_i)$. $\overrightarrow{CP_{jk}}$ are the principal components obtained from the jth of the samples analyzed for treatment T_k; that is, $\left(PC1_{jk}, PC2_{jk}\right)$ are used as comparison standards. σ is the standard deviation of the samples corresponding to treatment T_k, which is considered 1 because the variables are normalized. d is the dimension of the data, 2. r is the number of samples for each treatment T_k.

Each point in the two-phase diagram $(PC1_i, PC2_i)$ is classified based on maximum likelihood: A new observation $\vec{x}_i$ is classified into the class with the highest probability, such that each treatment is defined by an area of highest probability:

$$Treatment\ Area\ \ T_k = \arg max\left[P(T_k)P(\overrightarrow{PC_i}|T_k)\right]$$

Subsequently, the optimal ranges expected for a successful treatment are defined. Each range is determined by limits corresponding to the minimum and maximum values that would define a successful recovery according to a variable. The limit values of each range considered optimal corresponding to a variable are standardized using equation (5) where Z_{x-opt} is the value of the standardized variable, x_{opt} is the value considered optimal in variable x, $\bar{x}$ is the mean of the values obtained from variable x, σ_x is the standard deviation of the values obtained from variable x.

$$Z_{x-opt} = \frac{x_{opt} - \bar{x}}{\sigma_x} \quad (5)$$

The ranges between the minimum and maximum successful optimum define an area of the two-phase principal components diagram.

$$PC1_{opt} = \beta_{11} \cdot Z_{x1_{opt}} + \beta_{12} \cdot Z_{x2_{opt}} + \beta_{13} \cdot Z_{x3_{opt}} + \cdots + \beta_{1n} \cdot Z_{xn_{opt}} \quad (2)$$

$$PC2_{opt} = \beta_{21} \cdot Z_{x1_{opt}} + \beta_{22} \cdot Z_{x2_{opt}} + \beta_{23} \cdot Z_{x3_{opt}} + \cdots + \beta_{2n} \cdot Z_{xn_{opt}} \quad (3)$$

Each area of greatest probability T_k is characterized by the coordinates $(PC1_k, PC2_k)$, which will be located in the two-phase diagram at a certain distance from the centroid of the area considered optimal. If the centroid of the area considered optimal is defined as $(PC1_{opt}, PC2_{opt})$, the distance of a plot to the optimum is given by equation (6).

$$d_j = \sqrt{\left(PC1_{opt} - PC1_k\right)^2 + \left(PC2_{opt} - PC2_k\right)^2} \quad (6)$$

Determining the distance of each plot from the optimum allows for several evaluation possibilities. For example, positively evaluating treatments whose sum of distances to the optimum is lower. It also allows for analysis of variance, determining whether there are significant differences in distances between treatments with a 95% confidence level.

The results obtained for each treatment are evaluated according to their proximity to the optimal area.

2.2.4 Integration of PCA and PNN

PCA is used to reduce the dimensionality of the data before inputting it into the PNN network. This improves computational efficiency and reduces noise in the data, retaining only the most relevant features.

Data reduction with PCA: The original data x_i is transformed into Z_i, where $\overrightarrow{PC_j}$ represents the selected principal components with lower dimensionality.
Training and prediction with PNN: The PNN is trained with the transformed data Z_i to generate a predictive model.

The flow is as follows:

$$\overrightarrow{x_i} \rightarrow Z_i \xrightarrow{PCA} CPC_i \xrightarrow{PNN} Classification/Prediction$$

2.3 RESULTS AND DISCUSSION

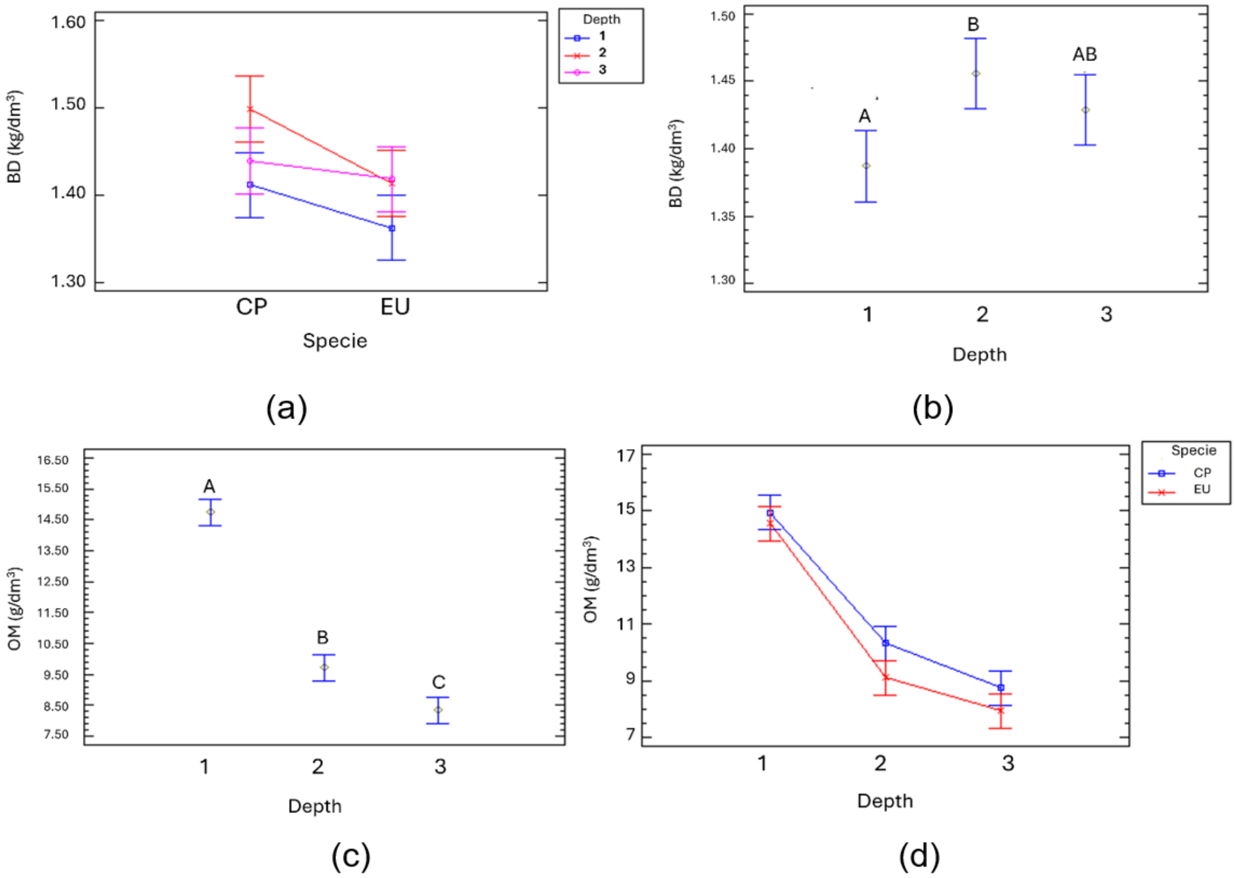
2.3.1 Statistical description of the studied variables

The supplementary material for this article provides a statistical description of the variables measured in each treatment. Based on these results, an analysis of the factors studied is presented: soil horizon, planted species, fertilizer rate applied, and initial soil tillage.

Table 2 of complementary material shows the results of the analysis of variance to detect whether each factor or their interactions have influenced the soil variables studied. It can be observed that at the different depths where samples were taken, there are differences in the variables organic matter (OM), cation exchange capacity (CEC), macroporosity (MA), and bulk density (BD). It is noteworthy that the behavior has been similar for the two tree species planted (Figure 4a). That is, no interaction between species and depth is detected (Figure 4a shows the overlap of the Least Significant Difference (LSD) intervals at a 95% confidence level). In Figure 4b, it can be observed that the organic matter content decreases as the depth increases. On the other hand, the bulk density is lower at the soil surface, being higher at depths 2 and 3, which do not show significant differences between them (Figure 4c).

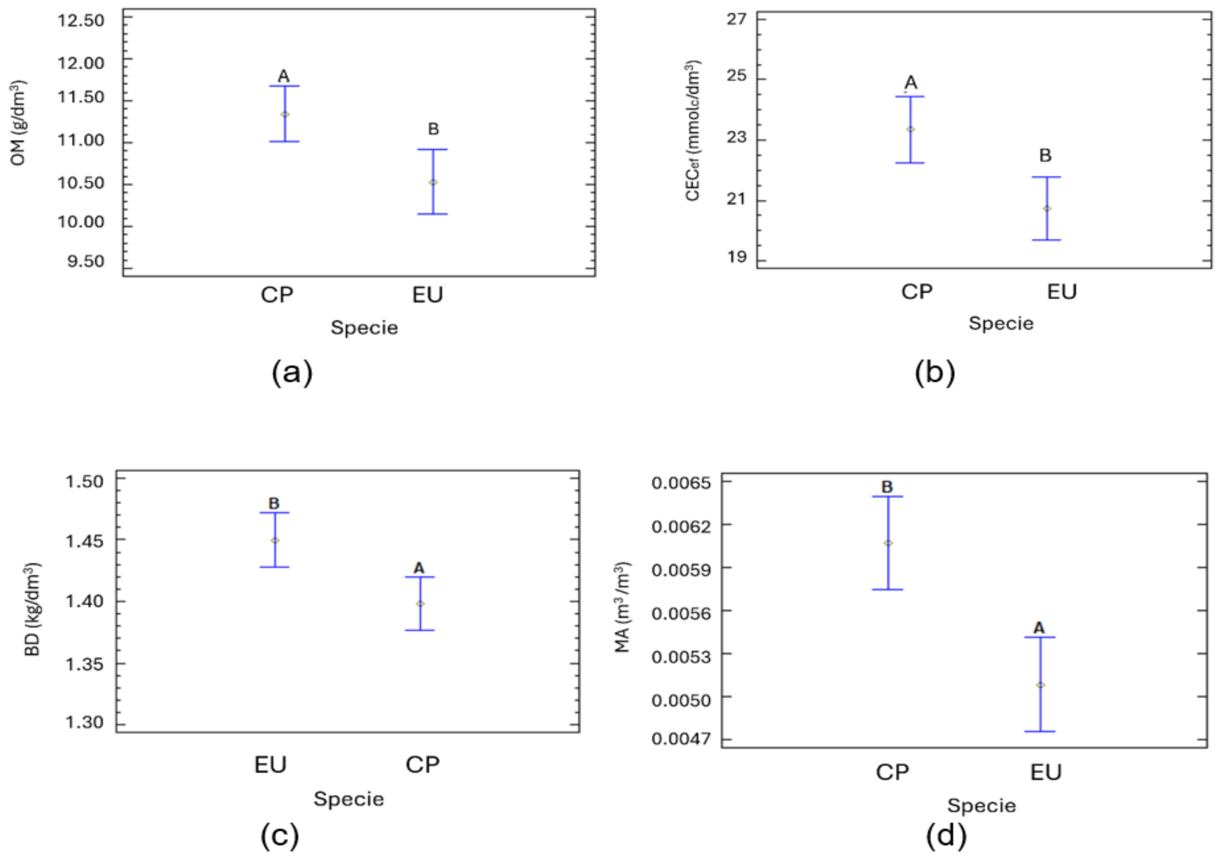
It is also noteworthy in Table 2 of complementary that the species used for the recovery of the system has globally influenced the physical and chemical variables of the soil. In Figure 5a, it can be seen that *Mabea fistulifera* plantations maintain a higher cation exchange capacity in the soil, as well as organic matter (Figure 5b). Additionally, this species provides greater macroporosity, which directly influences bulk density and infiltration rate (Figures 5c and 5d).

Figure 5. LSD intervals (95% confidence level) to analyze the influence of sampling depth (N=432): (a) depth and species interaction within the content of organic matter OM; (b) in bulk density BD; (c) in organic matter OM; (d) depth and species interaction within the content of organic matter OM.



Source: Elaborated by the author

Figure 6. LSD intervals (95% confidence level) to analyze the influence of the species (N=432): (a) in the content of organic matter; (b) in the cation exchange capacity; (c) in the bulk density; (d) in macroporosity.



Source: Elaborated by the author

Regarding the analysis of the influence of the soil amendment treatments on the measured variables, Table 5 shows the analysis of homogeneous groups. It can be observed that the treatments D0, areas with initial soil tillage but no fertilization, developed with lower organic matter. On the contrary, the treatments D20, areas with initial soil tillage and 20 Mg·ha^{-1} of compost, developed with higher organic matter content. Therefore, it can be concluded that initial fertilization with compost influences the evolution of organic matter content over time, as this influence is still detectable after 15 years. It is noteworthy that the *Mabea fistulifera* treatments generally provided better organic matter values than those performed with eucalyptus.

Regarding bulk density, the treatments do not provide any specific trend, as D0, areas with initial soil tillage but no fertilization, show similar values to the D20 treatments for both *Mabea fistulifera* and eucalyptus. In Table 5, the treatments with the lowest values of bulk density and macroporosity have been highlighted, and these are evenly distributed in the table. This does not allow drawing conclusions about the influence of the treatments on bulk density. It is noteworthy that there is direct correlation between organic matter content and CEC, as well as inverse correlation between bulk density and macroporosity.

Table 5. Homogeneous Groups

Treatment	OM	CEC	BD	MA
	$g \cdot dm^{-3}$	$mmol_c \cdot dm^{-3}$	$kg \cdot dm^{-3}$	$m^3.m^{-3}$
D0 EU	7.20 a	33.20 ab	1.37 a	0.065 cde
D0 CP	7.47 a	31.40 a	1.47 abc	0.038 a
SI EU	8.47 b	35.03 bcd	1.54 abc	0.055 bcd
D15 EU	8.74 b	34.43 abc	1.58 bc	0.043 ab
DAM EU	8.81 b	54.43 g	1.35 a	0.080 ef
D10 EU	8.90 b	37.53 cde	1.54 abc	0.035 a
DAM CP	8.90 b	33.16 ab	1.36 a	0.080 ef
D10 CP	9.83 c	38.43 de	1.67 c	0.060 cd
SI CP	10.26 c	37.07 cde	1.54 abc	0.035 a
D20 CP	12.51 d	39.68 e	1.35 a	0.070 de
D20 EU	12.53 d	54.43 g	1.35 a	0.050 abc
D15 CP	13.03 d	43.20 f	1.58 bc	0.090 f

Source: Elaborated by the author

2.3.2 Definition of the Optimal Values Sought

Table 6 shows the values considered optimal in a soil recovery process.

Bulk density is a fundamental parameter in evaluating physical quality, serving as a direct indicator of its compaction. High bulk density values can reduce total porosity, negatively affecting water infiltration and aeration, both essential for root development. In silty-clay soils, densities above 1.5 $g \cdot cm^{-3}$ are considered critical and may restrict root growth (Reichert *et al.*, 2023; Demattê; Demattê, 2024; Nantes *et al.*, 2025).

Macroporosity, referring to pores larger than 0.05 mm in diameter, is crucial for air and water movement in the soil. Macroporosity values lower than 0.10 $m^3 \cdot m^{-3}$ are considered limiting for root development (Demattê; Demattê, 2024). Works such as Rosa *et al*., (2012); Severiano *et al*., (2011); Suzuki *et al*., (2022) suggest that an ideal soil should have approximately 17% macroporosity and 33% microporosity, totaling 50% pore space. The reduction of macroporosity, often resulting from compaction, compromises oxygen diffusion, which is essential for root respiration (Demattê; Demattê, 2024; Nantes *et al.*, 2025) (Demattê; Demattê, 2024; Nantes *et al*., 2025).

Water infiltration in the soil is directly influenced by porosity and soil structure. Infiltration rates higher than 25 $mm.h^{-1}$ are considered adequate for silty-clay soils, promoting aquifer recharge and water availability for plants (Pinheiro *et al.*, 2009). Soil compaction can significantly reduce this rate, increasing surface runoff and susceptibility to erosion.

Soil organic matter (OM) plays a vital role in improving soil physical and chemical properties. It contributes to the formation of stable aggregates, increasing porosity and water retention capacity, in addition to providing cation exchange sites, increasing the soil's cation exchange capacity (CEC) (Blanco-Canqui; Benjamin, 2015; Liu *et al.*, 2006). In medium-textured soils, OM levels between 16 and 30 g/dm^3 are considered adequate (Cantarella *et al.*, 2022; Lourencetti *et al.*, 2023b).

Cation exchange capacity (CEC) is an indicator of soil fertility, reflecting its ability to retain and make essential cations available to plants. In Red Oxisols, CEC is generally low due to the prevalence of low-activity clay minerals and low organic matter levels. CEC values lower than 6.0 $cmolc \cdot dm^{-3}$ are considered critical and may limit nutrient availability for plants. Studies indicate that adding organic matter and liming can increase CEC in Red Oxisols, improving fertility (Chaves *et al.*, 2012).

The interrelationship between these attributes is evident, as soil compaction increases its bulk density, reducing macroporosity, and consequently compromising water infiltration and aeration. A decrease in macroporosity harms oxygen diffusion and water movement in the soil, both essential for root development. Reduced infiltration rates worsen surface runoff and erosion, in addition to limiting water availability for plants. Organic matter, in turn, plays a central role in improving soil structure, increasing porosity, water infiltration, and cation exchange capacity (CEC), while reducing bulk density. A higher CEC, benefiting from increased organic matter, strengthens the retention and availability of nutrients, essential for healthy and sustainable plant growth.

Table 6. Values considered to be optimal in a soil recovery process.

		Minimum	Maximum	Population average	Population desviation	Minimum Z_{x-min}	Maximum Z_{x-max}
Physical properties	Soil bulk density ($kg \cdot dm^{-3}$) (BD)	1.00	1.65	1.42	0.14	-3.14	1.67
	Total porosity ($m^{-3}.m^{-3}$) (PT)	0.30	0.55	0.43	0.04	-0.59	5.39
	Macroporosity ($m^{-3}.m^{-3}$) (MA)	0.17	0.50	0.06	0.02	5.60	0.44
	Microporosity ($m^{-3}.m^{-3}$) (MI)	0.25	0.33	0.27	0.04	-0.52	1.70
	Water infiltration ($mm \cdot h^{-1}$) (TI)	40.00	100.00	27.44	7.17	1.75	10.12
Chemical properties	Organic matter (%) ($g \cdot dm^{-3}$) (OM)	20.00	30.00	10.94	3.49	2.59	5.45
	CEC ($cmol_c.dm^{-3}$)	5.00	50.00	27.50	9.27	9.27	1.18
	$CEC_{efective}$ ($cmol_c.dm^{-3}$)	5.00	15.00	22.27	9.43	-1.83	-0.77

Source: Elaborated by the author

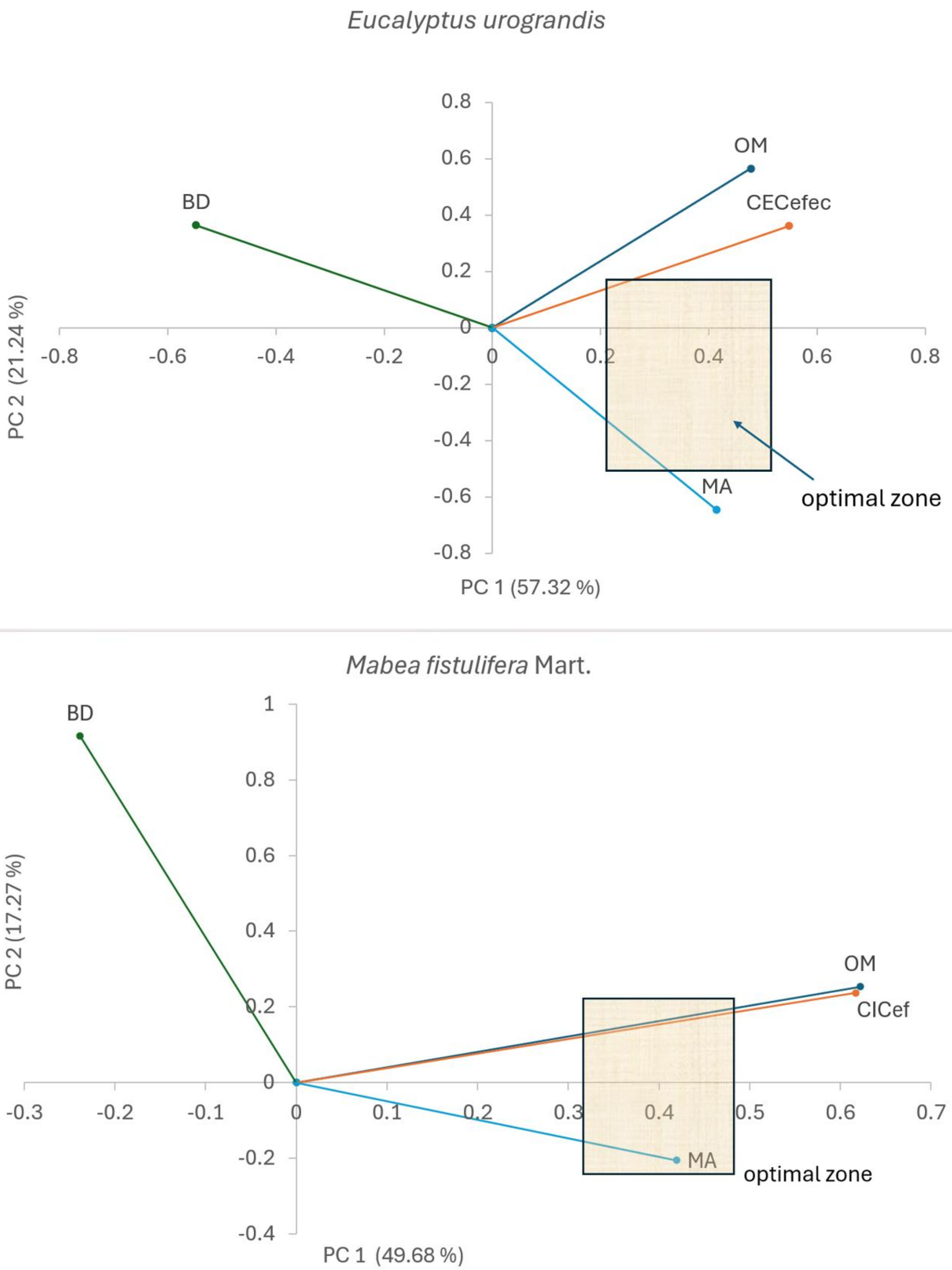
$$Z_x = \frac{x_{opt} - \bar{x}}{\sigma_x}$$

The proposed optimal ranges for bulk density, total porosity, macroporosity, microporosity, organic matter, $CEC_{effective}$ CEC, and infiltration rate were chosen based on critical limits and functional responses reported in the literature for Oxisols (Ferralsols). In Oxisols — soils with low mineral CEC, high susceptibility to acidity, and a strong dependence on organic matter for charge retention and aggregate formation—small improvements in OM and physical structure trigger proportional gains in infiltration, aeration, and nutrient availability (Demattê; Demattê, 2024). Therefore, we define ranges that reflect practical recovery conditions (e.g., OM = 20–30 $g \cdot dm^{-3}$; MA = 0.17–0.50 $m^3 \cdot m^{-3}$; PT = 0.30–0.55 $m^3 \cdot m^{-3}$; BD = 1.00–1.65 $kg \cdot dm^{-3}$; functional CTCef = 5–15 $cmolc \cdot dm^{-3}$; TI >25 $mm \cdot h^{-1}$, with 40–100 $mm \cdot h^{-1}$ as a reference for well-structured profiles). These ranges are operationally useful because they represent integrated conditions under which the plant can efficiently exploit the profile, nutrient cycling is activated, and recovery mechanisms are self-reinforcing. The practical validity of these targets is corroborated by the results of the experiment, in which subsoiling combined with 20 $Mg \cdot ha^{-1}$ of compost and planting of *Mabea fistulifera* shifted the plots to the optimal area defined by the ACP + PNN model.

2.3.3 Principal Component Analysis

For the most appropriate interpretation of the principal components, this statistical technique was applied to the variables that were influenced by the factors studied according to Table 6. In Figure 6, bifactorial plots of the variables BD, MA, effective CEC, and OM are shown. These plots represent the weight of each variable in the first two principal components, which together explain most of the data variability. There is a concordance between organic matter content and cation exchange capacity, as well as an antagonism between bulk density and macroporosity. According to the diagrams in Figure 6, high values of organic matter and cation exchange capacity are associated with high values of principal component 1 (*PC1*) and a moderate increase in principal component 2 (*PC2*). High bulk density is associated with very low values of component 1 and high values of principal component 2. On the other hand, high macroporosity is associated with high values of component 1 and low values of component 2. This means that the optimal areas would be located in the represented zones.

Figure 7. Weight of the variables in the principal components *PC1* and *PC2*



Source: Elaborated by the author

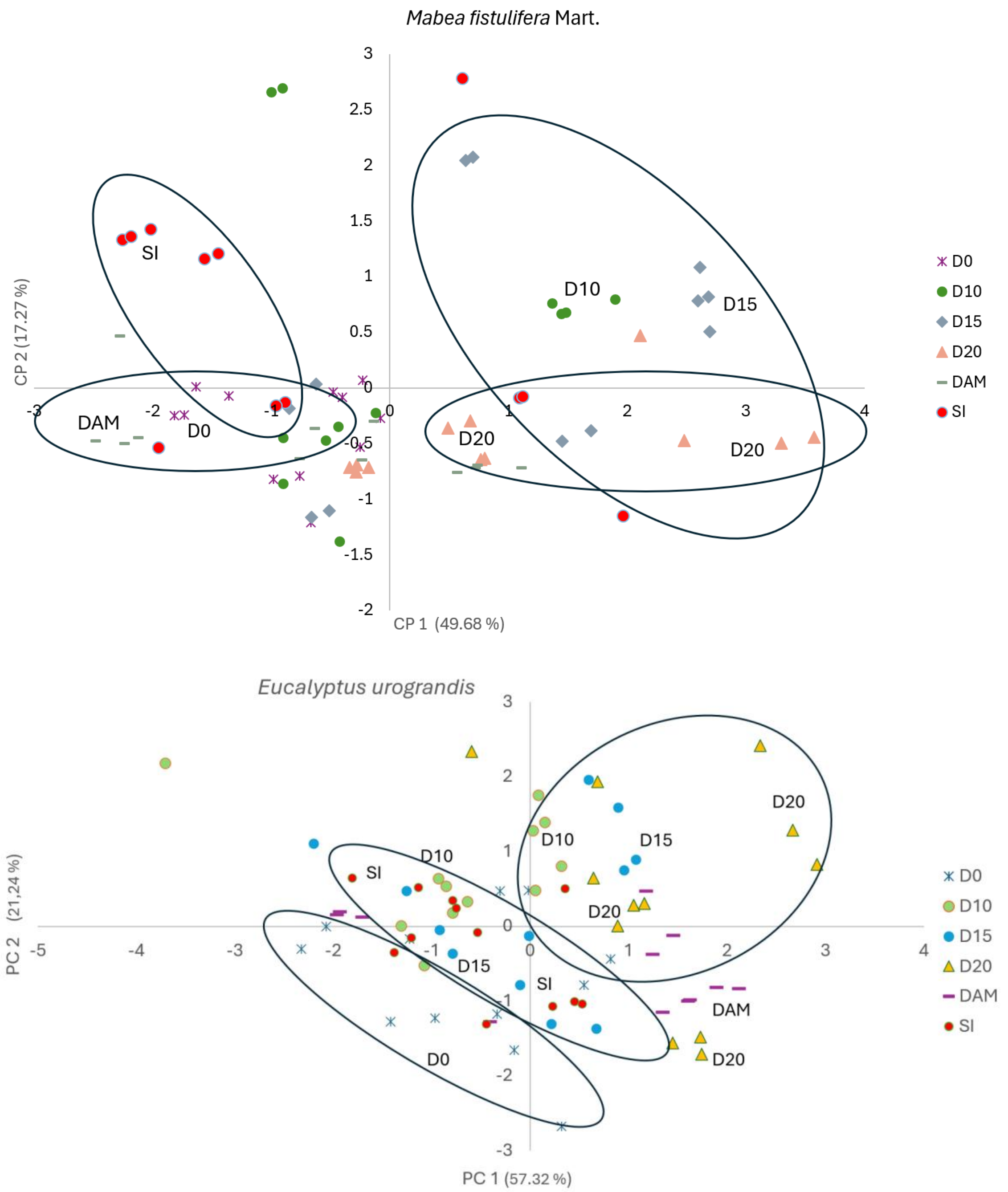
In Figure 7, the representation of each treatment according to the principal components obtained by applying the weights to the standardized soil characterizing variables can be observed. It can be seen that in the plots where *Mabea* was cultivated (Figure 8a), several zones can be discriminated according to the applied treatment. The plots with treatment D20 predominantly have positive *PC1* coordinates, which means better organic matter content, and negative *PC2* coordinates, representing lower bulk density and higher macroporosity. This means better nutrient retention capacity, increased aeration, and better conditions for root development.

In contrast, the plots with SI treatments are located in the second quadrant. This means they had negative *PC1* and positive *PC2* values. This indicates that these plots, having received no tillage or any type of organic or mineral amendment, evolved with the lowest organic matter content and the highest bulk density, which provides the worst conditions for plant development.

The treatments D10 and D15 in the *Mabea* plots are located in the first quadrant, meaning they have positive *PC1* and *PC2* values. This indicates that the organic matter content can be considered high, but the bulk density is also greater than that obtained in the D20 treatments.

The treatments D0 and DAM are located in the third quadrant, meaning they have negative *PC1* and *PC2* values. This indicates that the organic matter content is the lowest, although the bulk density may be moderate, with some macroporosity.

Figure 8. Analysis of treatments in the principal components *PC1* and *PC2*



Source: Elaborated by the author

When analyzing the scatter diagram of the principal components in the set of plots cultivated with eucalyptus, the discrimination according to the treatments becomes much more difficult. In general, three more or less homogeneous areas can be differentiated, grouping various treatments. First, treatments D20, D15, and D10 are mostly located in the first quadrant, i.e., with positive *PC1* and *PC2* values. This means higher organic matter content and cation exchange capacity, but with relatively higher bulk density compared to the other treatments. Conversely, treatments D0 are located in the third quadrant, i.e., with negative *PC1* and *PC2* values. This means lower organic matter content and cation exchange capacity, but with relatively more moderate bulk density. Lastly, treatments SI, DAM, along with some plots of D10 and D15, are located along a diagonal in the second and fourth quadrants, but are quite centered on the coordinate axes. This indicates that the effect of eucalyptus cultivation moderates the impact of the treatments, making significant differences between them less apparent.

The conclusions drawn from the analysis are as follows: In the plots with *Mabea*, treatment D20 improves organic matter and macroporosity, favoring plant development. The SI treatment shows poorer conditions due to the lack of tillage and amendments. Treatments D10 and D15 have high organic matter content but with higher bulk density compared to D20. Treatments D0 and DAM have low organic matter content and moderate bulk density. In the plots with eucalyptus the differences between treatments are less noticeable. D20, D15, and D10 have more organic matter and higher cation exchange capacity, but also higher bulk density. Treatment D0 has less organic matter but with moderate bulk density. Treatments SI and DAM show more homogeneous effects due to the eucalyptus cultivation.

In treatments with *Eucalyptus*, the smallest difference observed between the conditions evaluated could be attributed to the functional and ecophysiological characteristics of the species. *Eucalyptus* develops a deep and highly exploitative root system, capable of accessing subsurface horizons of the soil, which reduces the influence of physical-chemical modifications generated by treatments on the surface layer. As a result, its high capacity for absorbing water and nutrients, together with the production of slowly decomposing hobs, limits the return of organic matter and the renewal of the surface structure of the soil. This behavior favors homogenization of building properties between treatments, especially in the short term. Consequently, even though the highest dosages (D20, D15 and D10) are associated with greater organic matter content and cationic exchange capacity, the influence of *Eucalyptus* attenuates the differences induced by management. These results reflect the exotic character of the species, aimed at rapid recovery of vegetation cover and erosion control, but with a less differentiating effect on soil dynamics in comparison with *Mabea fistulifera*.

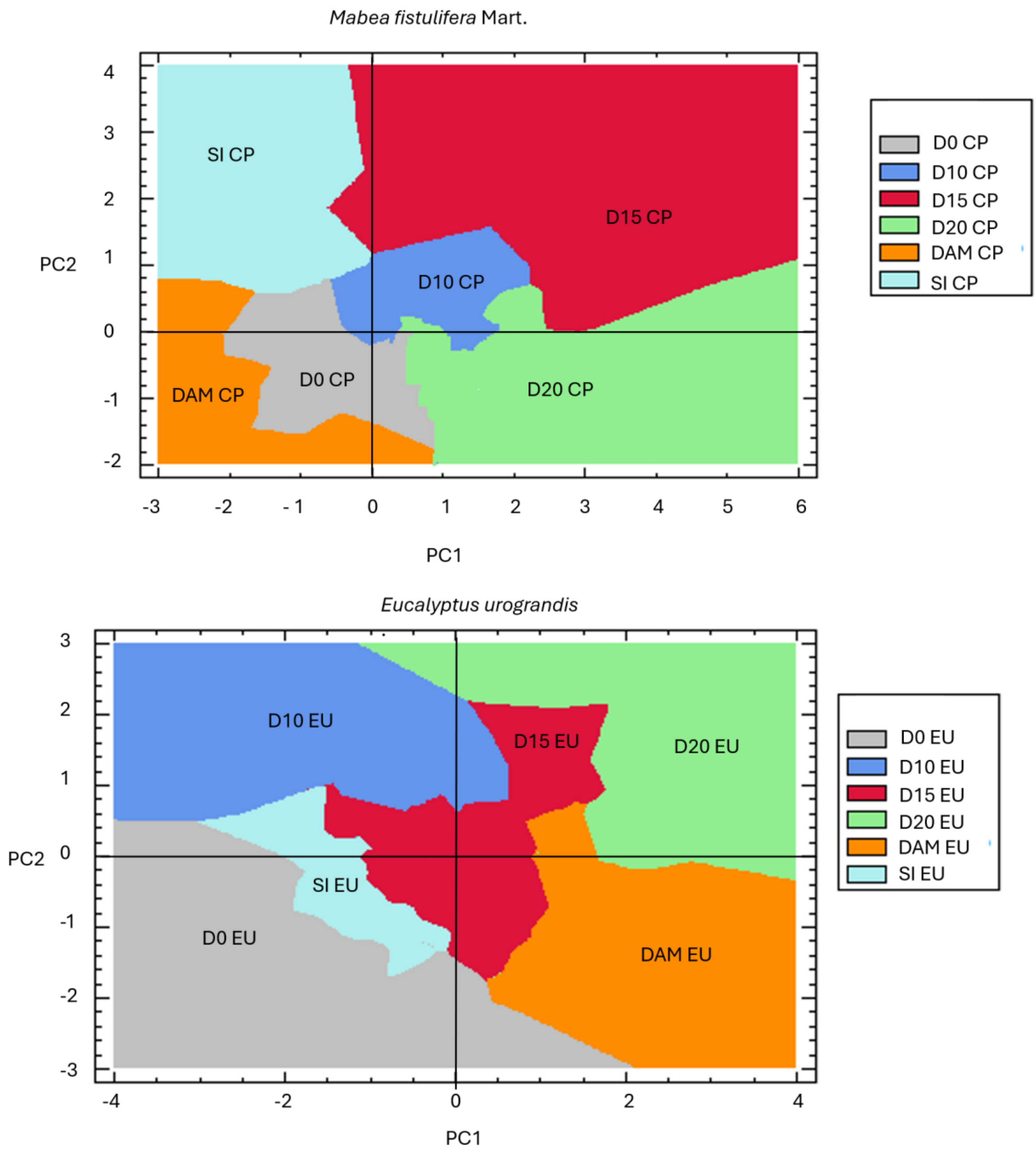
In the case of *Mabea fistulifera*, its superiority over Eucalyptus is explained mechanically by its greater contribution and recycling of organic matter through the leaf litter and fine roots, which increases the capacity for cation exchange and improves the structure of the soil. This is reflected in positive values of *PC1*, associated with greater organic matter and macroporosity, and lower apparent bulk density. However, Eucalyptus, despite being an exotic species with a deeper root system and lower surface input of residues, tends to generate denser and less porous soil over time, which reduces the differentiation between treatments in the analysis of main components. Therefore, the PCA shows clearer and more favorable discrimination in plots with *Mabea*, where management and plant-soil interactions promote more balanced and favorable physical and chemical conditions for ecological restoration.

The large dispersion obtained in the treatments characterized by the principal components *PC1* and *PC2* means that improvement and refinement methods should be applied. This is why, as presented below, a refinement based on probabilistic neural networks is discussed.

2.3.4 Refinement of the principal components with PNN

In Figure 8, it is observed that the discrimination between the different treatments is significantly more precise when probabilistic neural networks are employed compared to the exclusive use of the principal components. To improve the system's accuracy, in an initial stage, outlier points from each treatment were eliminated. This elimination was based on the distance of the data from the average principal component of each group, which helped reduce noise in the classification. Values with a distance greater than the average but 2 times the typical deviation of the distance, eliminating 5% of the data most crippled by the average. Subsequently, each remaining point was assigned a treatment classification based on the highest probability estimated by the probabilistic neural network model. This approach not only optimizes the separation between treatments but also demonstrates the effectiveness of probabilistic neural networks in improving data discrimination, in contrast to the simple dimensionality reduction provided by the principal components.

Figure 9. Refinement through the application of probabilistic neural networks to the principal components *PC1* and *PC2*



Source: Elaborated by the author

In Figure 9a, corresponding to the cultivation of *Mabea*, a clear ordering of the treatments based on the dose of organic compost applied is observed. As the applied dose increases, the treatments tend to progressively move further to the right in the diagram. This shift is associated with an increase in the first principal component, which reflects a higher concentration of organic matter in the soil as well as a greater cation exchange capacity.

It is important to note that the treatments analyzed were applied 14 years before the evaluation. Therefore, the analysis conducted highlights the influence of the treatments on the evolution of the soils intended for recovery. The increase in the dose of organic compost seems to improve soil quality, as the first principal component effectively captures these variations in its composition.

In Figure 8b, corresponding to the eucalyptus cultivation, the classification of the treatments shows a less organized distribution compared to *Mabea*. Treatment D20 is mainly concentrated in the first quadrant, characterized by high organic matter content and greater soil compaction. In contrast, treatment D0 is located in the third quadrant, associated with lower organic matter content and less compaction. Meanwhile, treatment DAM stands out for its high organic matter content and reduced compaction. This distribution suggests that the evolution of treatments in eucalyptus follows a more dispersed pattern compared to *Mabea*. A determining factor could be the initial development of the eucalyptus root system, stimulated by the application of higher doses of compost and mineral fertilization. This effect might have contributed to reducing bulk density, even in situations of lower organic matter content, which would explain the lack of a clear correlation between these parameters.

2.3.5 Obtaining the optimal area

In Table 7, the weights obtained for the calculation of each component in each of the evaluated species are presented. The weights applied to each of the standardized variables allow us to determine the position of a sample in the scatter diagram. For example, in the cultivation of *Mabea*, the first and second principal components would be calculated with equations (6) and (7).

Table 7. Weights of the main components (β_i)

	Mabea		*Eucalyptus*	
	PC 1	PC 2	PC 1	PC 2
OM ($g.dm^{-3}$)	0.622	0.254	0.478	0.564
CECef ($mmolc.dm^{-3}$)	0.617	0.237	0.548	0.361
BD ($kg.dm^{-3}$)	-0.238	0.915	-0.547	0.364
MA ($m^3.m^{-3}$)	0.419	-0.205	0.415	-0.647

Source: Elaborated by the author

Equation 2

$$PC1 = 0.622 \cdot Z_{OM} - 0.617 \cdot Z_{CICef} - 0.238 \cdot Z_{BD} + 0.419 \cdot Z_{MA}$$

Equation 3

$$PC2 = 0.254 \cdot Z_{OM} + 0.237 \cdot Z_{CICef} + 0.915 \cdot Z_{BD} - 0.205 \cdot Z_{MA}$$

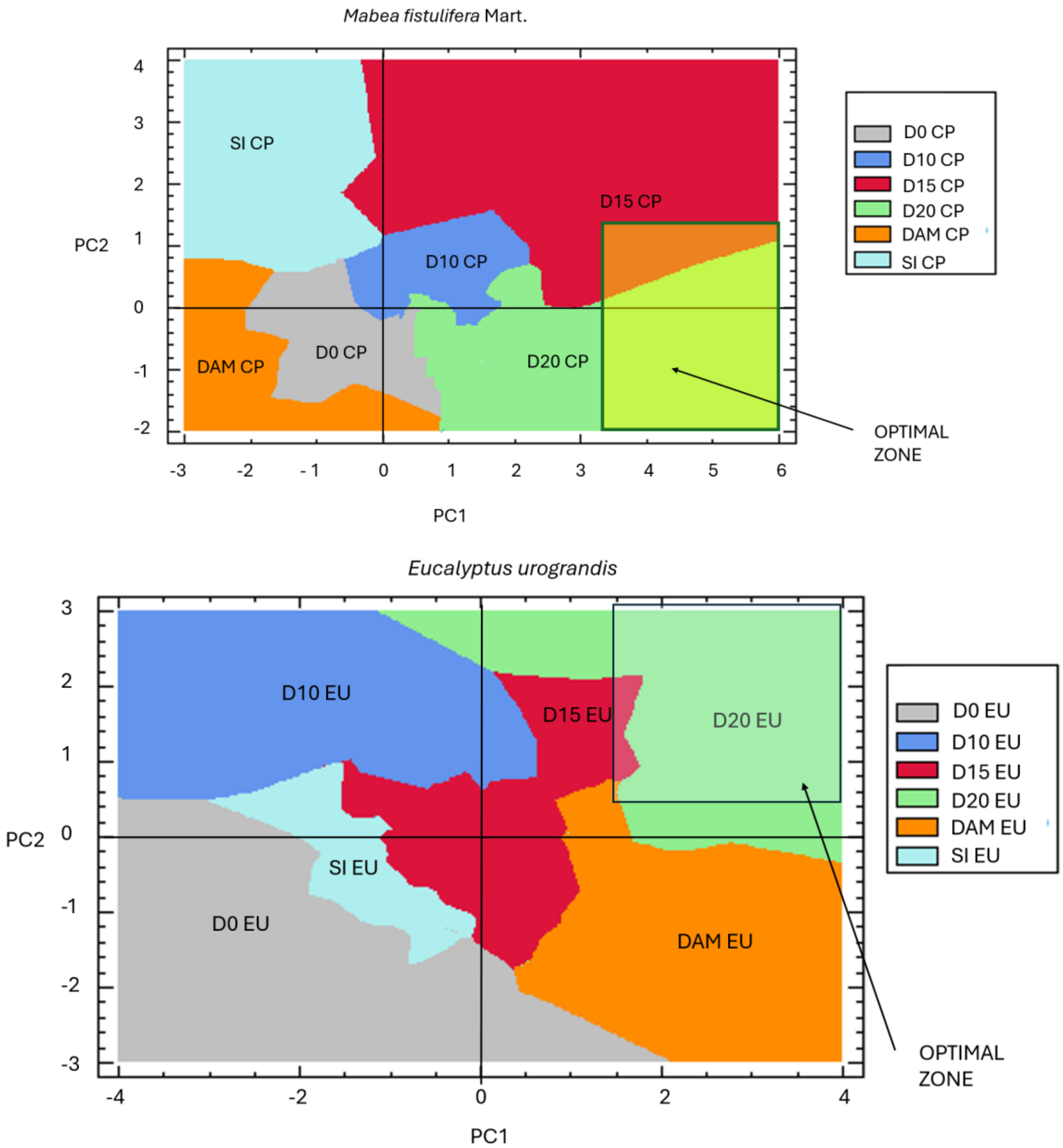
Applying the equations (2) and (3) to the standardized values of the situation considered optimal (Table 7). Let us obtain in the dispersion diagram the area considered objective. characterized by the values from Table 7. The average value of *PC1* and *PC2* of the optimal values obtained by providing the centroids of the optimal area in each species

Table 8. Areas considered optimal in the main components dispersion diagram

	Mabea fistulifera		*Eucalyptus*	
	PC1	*PC2*	*PC1*	*PC2*
Minimum Optimum	0.701	-2.389	1.430	0.477
Maximum Optimum	5.716	3.166	4.200	3.346
Average	3.208	0.389	2.815	1.911

Source: Elaborated by the author

Figure 10. Optimal zone in PNN classification



Source: Elaborated by the author

2.3.6 Calculation of distances

Table 9 shows the sums of the distances of the repetitions of each treatment to the optimal centroid. It can be seen that the minimum distance to the center of the area considered optimal corresponds to the D20 treatments, and subsequently to the D15 treatment. When analyzing variance over distances (Table 10) it is observed that there is no significant difference between D20 and D15 treatments, and these are significantly different from the rest of the treatments. However, in *Eucalyptus*, D20 is significantly closer to the optimum than the rest as there are no differences between them.

Table 9. Sume of the distances of the repetitions of each treatment to the optimal centroid.

	Sum of distances to the optimal centroid *Mabea*	Sum of distances to the optimal centroid *Eucalyptus*
D0	16.63 b	17.70 b
D10	13.24 b	15.16 b
D15	9.56 a	13.92 b
D20	9.78 a	9.34 a
DAM	16.08 b	14.11 b
SI	16.42 b	16.06 b

Source: Elaborated by the author

In Table 10, it is shown the variation analysis comparing all treatments according to their distance from the optimal centroid.

Table 10. Variance analysis of the distance to the optimal centroid in each horizon for each treatment.

Treatment	*Mabea* Medium distance	*Eucalyptus* Medium distance
D0	4.15951	4.42436
D10	3.31029	3.78943
D15	2.39236	3.47975
D20	2.44487	2.33503
DAM	4.02008	3.52697
SI	4.10709	4.01582

Source: Elaborated by the author

Even though the highest dosages (D20, D15 and D10) are associated with greater organic matter content and cation exchange capacity, the influence of Eucalyptus attenuates the differences induced by management. These results reflect the exotic character of the species, aimed at rapid recovery of vegetation cover and erosion control, but with a less differentiating effect on soil dynamics in comparison with *Mabea fistulifera.*

2.4 CONCLUSIONS

In this work, the application of a new methodology for evaluating soil evolution in recovery processes based on principal component analysis, complemented with probabilistic neural networks, has been demonstrated. This methodology allows for the global evaluation of all physical and chemical variables of the soils.

According to our results, the effects of applying tillage to promote plant growth combined with the application of a nutritional amendment, introducing eucalyptus or *Mabea* trees, have different effects depending on the soil depth.

The benefits to soil properties in the surface horizons are greater when compost is applied, and proportional to the applied dose.

The improvement of soil properties in the surface horizons is significantly greater when a compost dose of 20 Mg per hectare is used in combination with the planting of *Mabea fistulifera* trees.

In the *Mabea* plots, treatment D20 improves organic matter and macroporosity, favoring plant development. The SI treatment shows worse conditions due to lack of tillage and amendments. Treatments D10 and D15 have high organic matter content, but with higher bulk density than D20. Treatments D0 and DAM have low organic matter content and moderate bulk density.

In the *Eucalyptus* plots, the differences between treatments are less noticeable. D20, D15, and D10 have more organic matter and greater cation exchange capacity but also higher bulk density. Treatment D0 has less organic matter but with moderate bulk density. The SI and DAM treatments show more homogeneous effects due to eucalyptus cultivation.

The joint application of Principal Component Analysis (PCA) and Probabilistic Neuronal Networks (RNN) represents a significant advance in the interpretation of the evolution of degraded soils subjected to recovery treatments. This multivariable approach allows you to simultaneously integrate and analyze different physical and chemical properties of the soil, providing a more complete and detailed view of the restoration processes.

The use of PCA allows to reduce the complexity of data, synthesizing the information of multiple variables into main components that capture the underlying relationships between the properties of the data. By combining this technique with probabilistic neural networks, it is possible to discriminate in a more precise way the optimal values of each variable based on the different treatments applied, which facilitates a more robust evaluation of its impact on soil recovery.

This on-the-ground methodology improves precision in the interpretation of results, but also provides a powerful tool for making informed decisions in ecological restoration programs. The ability to jointly evaluate the interactions between the different variables of the soil allows us to identify complex patterns and relationships that, otherwise, could go unnoticed in univariant approaches. In short, the integration of PCA and RNN optimizes the understanding of soil evolution and improves the capacity to guide more effective and sustainable long-term management strategies.

Funding Statement

This work has been carried out within the framework of the IBEROMASA Network (719RT0586) of the Ibero-American Program of Science and Technology for Development (CYTED). Funding for open access charge: CRUE-Universitat Politècnica de València.

This work was supported by the Coordination for Improvement of Higher Education Personnel (CAPES-Brazil, Financing Code: 001); and the Universitat Politècnica de València, Spain.

CRediT authorship contribution statement

Conceptualization: BVM, CBB. Methodology: MAS, BVM, CBB. Investigation: MAS, JJD, LSN, DVA, JALS, GASL. Formal analysis: BVM, MAS Data curation: BVM, CBB. Writing—original draft preparation: MAS, BVM Writing—review and editing: BVM, CBB. Supervision, validation, funding acquisition, project administration, and resources: CBB, BVM.

Declaration of Competing Interest

The authors declare no conflicts of interest. The sponsors had no role in the design, execution, interpretation, or writing of the study.

Data availability

The datasets generated and analyzed during this study are available in the UNESP Institutional Repository, in the UPV Institutional Repository, or from the corresponding author upon reasonable request.

Acknowledgments

We also thank the Universitat Politècnica de València, Spain, Departamento de Ingeniería Rural y Agroalimentaria, for providing research infrastructure and technical support during the academic exchange developed under the Institutional Sandwich Doctorate Program, supervised by Prof. Dr. Borja Velázquez-Martí, specialist in Agricultural Engineering and Agro-Environmental Systems.

Informed Consent Statement

This study did not involve human or animal subjects.

Institutional Review Board Statement

Authors confirm that the manuscript has not been submitted to journal for simultaneous consideration and has not been previously published. Results collection, selection, and processing were performed personally.

REFERENCES

ASSUMPÇÃO, C. T. de. *Cebus apella and Brachyteles arachnoides (Cebidae) as potential pollinators of Mabea fistulifera (Euphorbiaceae).* **Journal of Mammalogy**, v. 62, n. 2, p. 386–388, 1981.

BLANCO-CANQUI, H.; BENJAMIN, J. G. Impacts of soil organic carbon on soil physical behavior. *In:* LOGSDON, S.; BERLI, M.; HORN, R. (ed.). **Advances in Agricultural Systems Modeling**. Madison: American Society of Agronomy; Soil Science Society of America, 2013. p. 11–40. DOI: 10.2134/advagricsystmodel3.c2.

BONINI NETO, A.; BONINI, C. S. B.; REIS, A. R.; PIAZENTIN, J. C.; COLETTA, L. F. S.; PUTTI, F. F.; HEINRICHS, R.; MOREIRA, A. Automatic recovery estimation of degraded soils by artificial neural networks in function of chemical and physical attributes in Brazilian Savannah soil. **Communications in Soil Science and Plant Analysis**, v. 50, p. 1785–1798, 2019. DOI: 10.1080/00103624.2019.1635144.

BONINI, C. dos S. B.; MACIEL, T. M. de S.; MOREIRA, B. R. de A.; CHITERO, J. G. M.; HENRIQUE, R. L. P.; ALVES, M. C. Long-term integrated systems of green manure and pasture significantly recover the macrofauna of degraded soil in the Brazilian Savannah. **Soil Systems**, v. 7, art. 56, 2023. DOI: 10.3390/soilsystems7020056.

CANTARELLA, H.; QUAGGIO, J. A.; MATTOS JUNIOR, D.; BOARETTO, R. M.; VAN RAIJ, B. **Boletim 100:** Recomendações de adubação e calagem para o estado de São Paulo. Campinas: Instituto Agronômico, 2022.

CHAVES, A. A. A.; LACERDA, M. P. C.; GOEDERT, W. J.; RAMOS, M. L. G.; KATO, E. Indicadores de qualidade de Latossolo Vermelho sob diferentes usos. **Pesquisa Agropecuária Tropical**, v. 42, p. 446–454, 2012. DOI: 10.1590/S1983-40632012000400002.

CINTRA, F. L. D.; MIELNICZUK, J. Potencial de algumas espécies vegetais para a recuperação de solos com propriedades físicas degradadas. **Revista Brasileira de Ciência do Solo**, v. 7, p. 197–201, 1983.

DEMATTÊ, J. L. I.; DEMATTÊ, J. A. M. **Manejo de solos arenosos:** fundamentos e aplicações. Piracicaba: LSO-ESALQ, 2024.

DOLÁCIO, C. J. F.; OLIVEIRA, R. S.; NAKAJIMA, N. Y.; TAVARES JÚNIOR, I. da S.; ROCHA, J. E. C. da; EBLING, Â. A.; GAMA, M. A. P. Integration of principal component analysis and artificial neural network to modeling productive capacity of eucalypt stands from biophysical attributes. **Forest Ecology and Management**, v. 460, art. 117862, 2020. DOI: 10.1016/j.foreco.2019.117862.

DUDA, G. P.; GUERRA, J. G. M.; MONTEIRO, M. T.; DE-POLLI, H.; TEIXEIRA, M. G. Perennial herbaceous legumes as live soil mulches and their effects on C, N and P of the microbial biomass. **Scientia Agricola**, v. 60, n. 1, p. 139–147, 2003. DOI: 10.1590/S0103-90162003000100021.

EDMEADES, D. C.; CLINTON, O. E. A simple rapid method for the measurement of exchangeable cations and effective cation exchange capacity. **Communications in Soil Science and Plant Analysis**, v. 12, p. 683–695, 1981.

ERICKSON, A. E. Tillage effects on soil aeration. *In:* VAN DOREN, D. M.; ALLMARAS, R. R.; LINDEN, D. R.; WHISLER, F. D. (ed.). **Predicting tillage effects on soil physical properties and processes**. Madison: American Society of Agronomy, 1985. p. 91–104.

FERRARI, S. F.; STRIER, K. B. Exploitation of Mabea fistulifera nectar by marmosets and muriquis in south-east Brazil. **Journal of Tropical Ecology**, v. 8, n. 3, p. 225–239, 1992. DOI: 10.1017/S0266467400006428.

GARCIA, A. R. de M.; CORRÊA, R. S. Revegetation of areas degraded by mining in the cerrado: the use of sewage sludge in soil rehabilitation. **Disciplinarum Scientia: Ciências Naturais e Tecnológicas**, v. 25, p. 123–144, 2024. DOI: 10.37779/nt.v25i2.4907.

GIACOMO, R. G. Recuperação do solo usando composto produzido com resíduo da extração de celulose. 2013. Tese (Doutorado em Ciência e Tecnologia de Materiais) – Universidade Estadual Paulista, Ilha Solteira, 2013.

GRABLE, A. R.; SIEMER, E. G. Effects of bulk density, aggregate size, and soil water suction on oxygen diffusion, redox potentials, and elongation of corn roots. **Soil Science Society of America Journal**, v. 32, p. 180–186, 1968. DOI: 10.2136/sssaj1968.03615995003200020011x.

GREENACRE, M.; GROENEN, P. J. F.; HASTIE, T.; D'ENZA, A. I.; MARKOS, A.; TUZHILINA, E. Principal component analysis. **Nature Reviews Methods Primers**, v. 2, art. 100, 2022. DOI: 10.1038/s43586-022-00184-w.

IUSS WORKING GROUP WRB. **World Reference Base for Soil Resources: international soil classification system for naming soils and creating legends for soil maps**. 4. ed. Vienna: International Union of Soil Sciences, 2022.

JOLLIFFE, I. T.; CADIMA, J. Principal component analysis: a review and recent developments. **Philosophical Transactions of the Royal Society A**, v. 374, art. 20150202, 2016. DOI: 10.1098/rsta.2015.0202.

KOEPPEN, W. **Climatología:** con un estudio de los climas de la tierra. México: Fondo de Cultura Económica, 1948.

LEAL FILHO, N.; BORGES, E. E. L. Influência da temperatura e da luz na germinação de sementes de canudo-de-pito (*Mabea fistulifera* Mart.). **Revista Brasileira de Sementes**, v. 14, n. 1, p. 57–60, 1992.

LIU, X.; HERBERT, S. J.; HASHEMI, A. M.; ZHANG, X.; DING, G. Effects of agricultural management on soil organic matter and carbon transformation: a review. **Plant, Soil and Environment**, v. 52, p. 531–543, 2006. DOI: 10.17221/3544-PSE.

LORENZI, H. **Árvores brasileiras:** manual de identificação e cultivo de plantas arbóreas nativas do Brasil. 4. ed. Nova Odessa: Instituto Plantarum, 2002.

LOURENCETTI, J.; BONINI, C. dos S. B.; ANDREOTTI, M.; ALVES, M. C.; BONINI NETO, A.; SANTOS, M. A.; BARRETTO, V. C. de M.; DE FIGUEREDO, R. W. R. Evolution of soil chemical fertility in an area under recovery for 30 years with anthropic intervention. **Sustainability**, v. 15, art. 10344, 2023. DOI: 10.3390/su151310344.

LU, M.; LIU, Y.; LIU, G. Precise prediction of soil organic matter in soils planted with a variety of crops through hybrid methods. **Computers and Electronics in Agriculture**, v. 200, art. 107246, 2022. DOI: 10.1016/j.compag.2022.107246.
MCCAULEY, A.; JONES, C.; JACOBSEN, J. Soil pH and organic matter. **Nutrient Management Module**, n. 8, p. 1–12, 2009.

NANTES, L. S.; BONINI, C. dos S. B.; SANTOS, M. A.; DA SILVA ALVES, L.; LOURENCETTI, J.; BONINI NETO, A.; DE SOUZA, J. A. L.; DA SILVA LUNARDELLI, G. A.; SILVA, M. B.; ALVARES, D. V.; PAZ-GONZALEZ, A. Exposed subsoil under recovery for 30 years: influence of management strategies on soil physical attributes and organic matter. **Soil Systems**, v. 9, art. 17, 2025. DOI: 10.3390/soilsystems9010017.

OLMOS, F.; BOULHOSA, R. L. P. A meeting of opportunists: birds and other visitors to *Mabea fistulifera* inflorescences. **Ararajuba**, v. 8, n. 2, p. 93–98, 2000.

PALAZZOLO, N.; PERES, D. J.; CREACO, E.; CANCELLIERE, A. Using principal component analysis to incorporate multi-layer soil moisture information in hydrometeorological thresholds for landslide prediction. **Natural Hazards and Earth System Sciences**, v. 23, p. 279–291, 2023. DOI: 10.5194/nhess-23-279-2023.

PERIN, A. **Desempenho de leguminosas herbáceas perenes com potencial de utilização para cobertura viva e seus efeitos sobre atributos físicos do solo**. 2001. Dissertação (Mestrado) – Universidade Federal Rural do Rio de Janeiro, Seropédica, 2001.

PINHEIRO, A.; TEIXEIRA, L.; KAUFMANN, V. Capacidade de infiltração de água em solos sob diferentes usos e práticas de manejo agrícola. **Ambiente & Água**, v. 4, p. 188–199, 2009. DOI: 10.4136/ambi-agua.97.

RAIJ, B. VAN; ANDRADE, J. C. DE; CANTARELLA, H.; QUAGGIO, J. A. **Análise química para avaliação da fertilidade de solos tropicais**. Campinas: Instituto Agronômico, 2001.

REICHERT, J. M.; REINERT, D. J.; BRAIDA, J. A. Qualidade dos solos e sustentabilidade de sistemas agrícolas. **Ciência & Ambiente**, v. 27, p. 29–48, 2023.

ROSA, H. A.; VELOSO, G.; SECCO, D.; RUFFATO, G. G.; MARCONDES, L. C.; SANTOS, R. F. Efeitos do tempo de adoção do sistema plantio direto em atributos físicos de um Latossolo argiloso. **Revista Cultivando o Saber**, v. 5, n. 2, p. 165–174, 2012.

ROSENBERG, N. J. Response of plants to the physical effects of soil compaction. **Advances in Agronomy**, v. 16, p. 181–196, 1964. DOI: 10.1016/S0065-2113(08)60024-3.

SABATTINI, J. A.; SABATTINI, R. A.; URTEAGA-OMAR, A. F.; VELÁZQUEZ-MARTÍ, B. Classification of successional stages in native forests of the Argentine Spinal through neural networks. **Land Degradation & Development**, v. 30, p. 2064–2072, 2019. DOI: 10.1002/ldr.3409.

SEVERIANO, E. da C.; OLIVEIRA, G. C. de; DIAS JÚNIOR, M. de S.; COSTA, K. A. de P.; SILVA, F. G.; FERREIRA FILHO, S. M. Structural changes in latosols of the cerrado region: relationships between soil physical properties and least limiting water range. **Revista Brasileira de Ciência do Solo**, v. 35, p. 773–782, 2011. DOI: 10.1590/S0100-06832011000300013.

SHOUMIK, B. A.; KHAN, M. Z.; BŁOŃSKA, E.; LASOTA, J. Dynamics of soil organic carbon and total nitrogen in particulate and mineral-associated organic matter fractions under different land use patterns across Europe. **Agriculture, Ecosystems & Environment**, v. 381, art. 109411, 2025. DOI: 10.1016/j.agee.2024.109411.

SOIL SURVEY STAFF. **Keys to Soil Taxonomy**. 13. ed. Washington, DC: USDA-Natural Resources Conservation Service, 2022.

SOTO BONI, T.; PUJOL PEREIRA, E. I.; AVELINO SANTOS, A.; RODRIGUES CASSIOLATO, A. M.; MALTONI, K. L. Biomass residues improve soil chemical and biological properties reestablishing native species in an exposed subsoil in Brazilian Cerrado. **PLoS One**, v. 17, art. e0270215, 2022. DOI: 10.1371/journal.pone.0270215.

SOUTO FILHO, S. N.; ALVES, M. C.; MONREAL, C. M.; BONINI, C. dos S. B. Nanoparticles and nanostructure morphology of a Red Latosol in rehabilitation. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 21, p. 530–536, 2017. DOI: 10.1590/1807-1929/agriambi.v21n8p530-536.

SUN, M.; LI, Q.; JIANG, X.; YE, T.; LI, X.; NIU, B. Estimation of soil salt content and organic matter on arable land in the Yellow River Delta by combining UAV hyperspectral and Landsat-8 multispectral imagery. **Sensors**, v. 22, art. 3990, 2022. DOI: 10.3390/s22113990.

SUZUKI, L. E. A. S.; REINERT, D. J.; ALVES, M. C.; REICHERT, J. M. Critical limits for soybean and black bean root growth based on macroporosity and penetrability for soils with distinct texture and management systems. **Sustainability**, v. 14, art. 2958, 2022. DOI: 10.3390/su14052958.

TEIXEIRA, P. C.; DONAGEMMA, G. K.; FONTANA, A.; TEIXEIRA, W. G. **Manual de métodos de análise de solo**. Brasília: Embrapa, 2017.

THOMASSON, A. J. Towards an objective classification of soil structure. **Journal of Soil Science**, v. 29, p. 38–46, 1978. DOI: 10.1111/j.1365-2389.1978.tb02029.x.

VIEIRA, M. F.; CARVALHO-OKANO, R. M. Pollination biology of *Mabea fistulifera* in southeastern Brazil. **Biotropica**, v. 28, n. 1, p. 61–68, 1996. DOI: 10.2307/2388771.

VIEIRA, M. F.; CARVALHO-OKANO, R. M.; SAZIMA, M. The common opossum (*Didelphis marsupialis*) as a pollinator of *Mabea fistulifera*. **Ciência e Cultura**, v. 43, n. 5, p. 390–393, 1991.

VIEIRA, M. F.; MATTOS, G. T.; CARVALHO-OKANO, R. M. *Mabea fistulifera* na alimentação de aves na região de Viçosa, Minas Gerais, Brasil. **Iheringia, Série Zoologia**, v. 73, p. 65–68, 1992.

ZHANG, R. Determination of soil sorptivity and hydraulic conductivity from the disk infiltrometer. **Soil Science Society of America Journal**, v. 61, p. 1024–1030, 1997. DOI: 10.2136/sssaj1997.03615995006100040005x.

3 CAPÍTULO II - ARTIFICIAL NEURAL NETWORKS FOR PREDICTING SOIL QUALITY INDICATORS BASED ON AGGREGATE STRUCTURE

Melissa Alexandre Santos[a], Borja Velázquez-Martí[b], Alfredo Bonini Neto[c], Laura Silva Nantes[a], Carolina dos Santos Batista Bonini[d]

Highlights:

1. **Development of predictive models based on artificial neural networks (ANN) to assess critical soil properties** such as infiltration rate, total cation exchange capacity (CEC), effective CEC, and soil organic matter (SOM) content.
2. **Innovative integration of artificial intelligence and soil science** to predict key soil quality indicators using easily measurable structural variables, such as aggregate size and mean weight diameter (MWD).
3. **Significant improvement in the fit of predictive models**, with determination coefficients (R^2) reaching values of up to 0.90 by transforming output variables and reducing the number of predictive variables.
4. **Potential application in sustainable agriculture**, offering a quick and costeffective alternative to evaluate soil quality without relying on expensive laboratory analyses.
5. **Demonstration of ANN's ability to generalize across soils with different characteristics**, opening new possibilities for efficient and precise soil management and data-driven agronomic decision-making.

[a] Graduate Program in Agronomy, Faculty of Engineering, São Paulo State University (Unesp), Ilha Solteira 15385-000, SP, Brazil.
[b] Departamento de Ingeniería Rural y Agroalimentaria. Universitat Politècnica de València. Camino de Vera s/n. 46022 Valencia (España). borvemar@dmta.upv.es
[c] Department of Biosystems Engineering, São Paulo State University (Unesp), School of Sciences and Engineering, Tupa, Brazil.
[d] Department of Plant Production, São Paulo State University (Unesp), College of Agricultural and Technological Sciences, Dracena, Brazil.

ABSTRACT

This work describes the development and validation of Artificial Neural Network (ANN) models capable of predicting key soil properties —infiltration rate, total Cation Exchange Capacity (CEC), effective CEC, and Soil Organic Matter (SOM) content — based on physical variables such as aggregate size and mean weight diameter (MWD). Various ANN architectures were trained. The training was carried out using the Levenberg-Marquardt algorithm, employing cross-validation and an early stopping strategy to prevent overfitting. The modeling of soil (edaphic) parameters using Artificial Neural Networks (ANNs) based on soil structural variables showed a clear gradient of improvement in this study as data preprocessing and output variable selection were adjusted. Initially, the architectures with two hidden layers and between 2 and 10 neurons per layer achieved relatively modest coefficients of determination ($R^2 \approx$ 0,63–0,66) when simultaneously predicting infiltration, total CEC, effective CEC, and organic matter. Subsequently, the transformation of the output variables using the square root raised the R^2 values to R^2 a $\approx$ 0,70–0,72, and restricting the problem to the prediction of the square root of CEC and organic matter allowed for substantially superior fits ($R^2 \approx$ 0.87–0.90) with no evidence of overfitting. This approach demonstrates the possibility of accurately estimating relevant soil (edaphic) parameters using easily measurable structural variables, representing a notable advance for soil monitoring. Furthermore, it opens up new possibilities for more efficient and sustainable agriculture by integrating artificial intelligence with pedological knowledge, facilitating better-founded agronomic decisions and reducing the dependence on costly and prolonged laboratory analyses.

Key words: soil aggregates, artificial intelligence, soil structure

3.1 INTRODUCTION

The understanding of the relationships between soil structure and its functional properties is essential for agronomy, soil science, and environmental management. In this context, predicting the content of organic matter, infiltration, and cation exchange capacity (CEC) from aggregate size is of great importance, as these parameters are determinant for fertility, soil health, and the sustainability of agricultural ecosystems (Mehta *et al*., 2024; Amini *et al*., 2005; Tang *et al*., 2009).

Soil structure is formed by the organization of mineral particles and organic matter into aggregates, generated through physical, chemical, and biological processes. The size and stability of the aggregates directly condition the edaphic structure, influencing porosity, infiltration, aeration, and nutrient retention, while reducing compaction and erosion (Abu-Hamdeh *et al*., 2006; Lipiec *et al*., 2007; Oliverio *et al*., 2021; Nantes *et al*., 2025). The organization of these aggregates determines the distribution and mobility of water and nutrients in the soil profile, affecting both the dynamics of organic matter and the CEC (Manrique *et al*., 1991; Fooladmand, 2008; Bonini *et al*., 2025).

Organic matter plays a central role, as it not only acts as a binding agent in the formation and stabilization of aggregates (Okolo *et al*., 2020), but also contributes to nutrient retention and the balance of the soil microbial community (Ding *et al*., 2013; Mhete *et al*., 2020). Soils with a higher proportion of large and stable aggregates tend to accumulate more organic carbon, which is protected from microbial decomposition by micropores (Álvaro-Fuentes *et al*., 2009). This opens up the possibility of predicting carbon content and fertility using simple physical variables, reducing the need for laboratory analyses (Souza *et al*., 2024).

The size and stability of the aggregates also decisively influence infiltration, as larger structures promote macroporosity, favoring the downward flow of water and reducing runoff and erosion (Amézketa, 1999; Dorji *et al*., 2020). However, phenomena such as air slaking generate variability even under similar management conditions (Korenkova; Urik, 2018). Hence the importance of characterizing the soil's physical attributes for efficient water management (Bronick; Lal, 2005).

The CEC, in turn, expresses the soil's capacity to retain and supply cations such as calcium, magnesium, potassium, and sodium. This property is strongly influenced by organic matter and the formation of stable aggregates, which increase the active surface area through negatively charged functional groups. Multivariate models and neural networks have proven effective in predicting CEC based on variables such as texture and carbon, thereby improving the assessment of agricultural suitability (Kashi *et al*., 2014).

Among the parameters used to evaluate structural stability, the mean weight diameter (MWD) and the aggregate stability index (ASI) stand out. The MWD reflects the relative proportion of large aggregates, with high values being indicative of greater resistance to disaggregation (Castro Filho *et al*., 2002). The ASI, on the other hand, measures the percentage of stable aggregates remaining after agitation in water, with values above 70% being desirable (Castro Filho; Logan, 1991; Castro Filho *et al*., 1998).

In highly weathered tropical soils, these parameters acquire particular relevance, since the preservation of macroporosity and preferential water flow paths is fundamental for controlling erodibility. Structural stability also favors carbon sequestration, while its degradation accelerates the mineralization of organic matter and the emission of CO_2 (Kemper; Rosenau, 1986; Lal *et al*., 1997; Castro Filho; Logan, 1991; Castro Filho *et al*., 1998).

Comparisons between managed areas and native vegetation areas show that the former usually present a higher concentration of microaggregates. Nevertheless, soils under native vegetation, due to the greater input of roots and organic matter, promote the formation of stable aggregates and the accumulation of carbon (Silva, 2019; Kochiieru *et al*., 2020; Haydu-Houdeshell *et al*., 2018). Thus, aggregation is consolidated as a key indicator of soil quality, integrating physical, chemical, and biological processes that determine its conservation.

3.1.1 Predictive Models and Management Efficiency

The use of predictive models based on aggregate size allows agronomists and natural resource managers to perform faster and more economical estimations of soil quality. Instead of conducting multiple costly laboratory analyses, valuable information can be obtained for decision-making related to improving soil quality, fertilizer application, the selection of suitable crops, or the implementation of sustainable management practices.

Estimating complete soil characteristics from soil aggregate size can be a key tool in the sustainable management of agricultural soils, especially in practices such as regenerative agriculture or agroecology, where the long-term increase in soil stability is sought. Knowing that aggregate size is related to characteristics such as infiltration, CEC, and organic matter, it is possible to optimize tillage practices, the use of compost, or the implementation of cover crops to improve these factors without relying excessively on external inputs.

Applying neural networks to predict organic matter content, infiltration, and cation exchange capacity (CEC) from soil aggregate size is an innovative approach that combines soil science with advanced machine learning techniques. Neural networks are particularly suitable for this type of problem because they can model complex and non-linear relationships between input variables (such as aggregate size) and desired outputs (such as organic matter content, infiltration, and CEC) without requiring an explicit specification of the mathematical function relating them.

The objective is to develop a predictive model that allows the estimation of critical soil characteristics such as organic matter, infiltration, and CEC based on soil aggregate size. These parameters are fundamental for the sustainable management of agricultural soils and their long-term health. The idea is to train a neural network with data on different aggregate sizes and their corresponding values for organic matter, CEC, and infiltration, so that the network learns the relationships between these variables and can make predictions for new soils.

3.2 MATERIALS AND METHODS

3.2.1 Experiment Delimitation

The study was conducted at the Teaching and Research Farm of the Faculty of Engineering, in Selvíria, Mato Grosso do Sul (MS), Brazil, located at an altitude of 327 m, between the coordinates 20°22' S and 51°22' W, on the banks of the Paraná River (Figure 1). The region's climate is humid tropical (Aw), with a rainy summer and dry winter season, an average annual precipitation of 1370 mm, an average temperature of 23.5 °C, and relative air humidity ranging between 70% and 80% (Köppen, 1948). The soil was classified as Oxisol (Typic Hapludox) according to USDA Soil Taxonomy (Soil Survey Staff, 2022), and Ferralsol following the WRB classification (IUSS Working

Group WRB, 2022), which was decapitated during the construction of the Ilha Solteira Hydroelectric Power Plant in 1969.

Figure 1. Location of the degraded area (A) and average decapitated soil layer (B), at the Faculty of Engineering Teaching, Research, and Extension Farm, Ilha Solteira Campus.

Source: Giacomo, 2013.

The preparation of the experimental area began in December 2009, with transversal subsoiling at a depth of 0.40 m, followed by light harrowing. In February 2010, nutritional corrections were carried out by applying organic compost or mineral fertilization, according to the treatments. The inputs were distributed manually and incorporated by deep plowing, in order to ensure greater homogeneity and nutrient availability. For planting, a new subsoiling was performed with a single shank at a depth of 0.50 m, where the seedlings were introduced.

Two tree species were evaluated: *Eucalyptus urograndis* (clone h-17, exotic species) and *Mabea fistulifera* Mart. (native to the Cerrado, derived from seeds). Both were combined with mineral fertilization and five different doses of organic compost produced from cellulosic waste (Garcia; Corrêa, 2024).

For mineral fertilization, the recommendations for *Mabea fistulifera* from the native seedling nursery of Companhia Energética de São Paulo (CESP), Jupiá unit, were followed, 100 g of 8-28-16 formula per plant was applied, corresponding to 166.70 kg Page 8 of 81 ha^{-1}, followed by fertilization at 60 days with 48.8 g of urea per plant (81.45 kg ha^{-1}) and 16.70 g of potassium chloride (KCl) per plant (27.80 kg ha^{-1}). For organic fertilization, a compost composed of a mixture of sludge, sand, limestone, ash, and other residues from the industrial cellulose extraction process was used. It underwent 30 days of composting in outdoor rows, with periodic mechanical turning.

The organic compost used was supplied by the Composting Center of the Ambitec Group (International Paper Unit, Mogi Guaçu, SP).

The experiment consisted of 12 treatments (Table 2) with 4 repetitions per treatment. In total, there were 48 plots of 15 m x 12 m, divided into 4 blocks (one block per repetition) of 2160 m² each. In each experimental plot, 40 trees of the same species were planted with a planting spacing of 1.5 m between plants and 3 m between rows.

3.2.2 Measured Soil Variables

In each soil sample, the percentages of aggregates corresponding to the following particle size fractions were measured: less than 0.25 mm, between 0.25 and 0.5 mm, between 0.5 and 1 mm, between 1 and 2 mm, between 2 and 4 mm, and greater than 4 mm.

The measurements were performed using a wet sieving process, following standard procedures for aggregate stability analysis. For this, each air-dried soil sample was placed in a column system of sieves, ordered from largest to smallest mesh aperture (4 mm, 2 mm, 1 mm, 0.5 mm, and 0.25 mm). The set of sieves was subjected to controlled mechanical shaking for a fixed time, which allowed the aggregates to be separated according to their size. Afterwards, the samples were dispersed in a sodium hydroxide solution to quantify only the soil aggregate portion, separating the mechanically trapped sand.

After sieving, the material retained in each sieve—as well as the fine fraction smaller than 0.25 mm that passed through the smallest mesh—was collected and weighed individually using a precision balance. Finally, the percentages of each fraction were calculated relative to the total sample weight, thereby obtaining the aggregate size distribution of the soil.

The percentages obtained were used as inputs in the neural network analysis.

Soil quality assessments were performed based on the following indicators:

- Water Infiltration (mm/h): Three evaluations were conducted in each experimental plot using the minidisk infiltrometer (Zhang, 1997).
- Soil Organic Matter (SOM): analyzed according to Raij *et al*. (2001). It is considered the central indicator of soil quality and health, as it provides organic binding agents, reduces bulk density, and improves structural resilience. For medium-textured soils,

the SOM content ranges between 16 and 30 g/dm³ (Blanco-Canqui; Benjamin, 2015; Liu *et al*., 2006; Cantarella *et al*., 2022).

- Cation Exchange Capacity (CEC): determined according to Raij *et al*. (2001), encompassing both the total CEC, measured at pH 7.0 and considered as the theoretical potential for cation retention (Hoddinott, 1983), and the effective CEC, which reflects the actual field conditions, being determined from extracting solutions at natural pH (Edmeades; Clinton, 1981).

The data used for the analyses were obtained directly from field measurements and laboratory determinations, supplemented by historical records of land use and the environmental conditions of the region.

The values of these variables were used as outputs of the neural network.

3.2.3 Neural Network Design

To predict the soil parameters based on aggregate size, an Artificial Neural Network (ANN) was used, which consists of several layers of interconnected nodes:

a. Input: The first layer of the network has a number of nodes equal to the number of features to be used as inputs. In this case, the main inputs will be the fractions of the soil aggregate sizes.
b. Hidden Layers: In neural networks, the hidden layers are where the data representations are learned. In this work, the goal was to select an appropriate number of neurons in the hidden layers to prevent overfitting or underfitting. The activation function used was Pureline to ensure the models can learn non-linear relationships between the inputs and the outputs.

The **Purelin** function performs a linear transformation on a neuron's input; that is, the output is directly proportional to the input without any non-linearity. Mathematically, it is defined as:

$$f(x) = x$$

That is, the neuron's output will be equal to the input multiplied by a determined weight, without any additional modification.

c. Output: The output layer is the set of neurons that represents the predicted values for organic matter content, infiltration, and CEC (multitask learning).

3.2.4 Neural Network Training

To train the network, a labeled dataset was used, that is, a dataset in which the aggregate size and the corresponding values for organic matter, infiltration, and CEC are already known. The training process consisted of:

- Forward Propagation: The input values (aggregate size) are passed through the network, and predictions are obtained.
- Error Calculation: The error is calculated by comparing the network's predictions with the actual output values measured in the laboratory.
- Backpropagation: The network's weights are adjusted to minimize the error using the backpropagation and optimization algorithm, such as the gradient descent algorithm.

This process is repeated over multiple iterations (epochs) until the model converges and the prediction error is minimized.

3.2.5 Model Evaluation

Once the network was trained, the performance of each analyzed network architecture was evaluated using a test dataset that had not been used in training. The metrics used to evaluate the network performance were:

- Mean Squared Error (MSE): It measures the average difference between the predictions and the actual values.

$$MSE = \frac{1}{N}\sum_{i=1}^{N} (t_i - y_i)^2$$

- Coefficient of Determination (R^2): It measures the proportion of the variance in the data that is explained by the model.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

Where:

- $SS_{res} = \sum(t - y)^2$→sum of squared errors

- $SS_{tot} = \sum(t - t^{-})^2 \rightarrow$ total variance of the actual data
- t are the measured values
- t^{-} is the average of the measured values
- y are the predicted values

3.3 RESULTS

In Table 1, the statistical description of all variables used as inputs in the neural network models is presented: soil mass fraction less than 0,25 mm ($f_{<0,25}$), soil mass fraction between 0,25 and 0,5 mm ($f_{0,25-0,5}$), soil mass fraction between 0,5 and 1 mm ($f_{0,5-1}$),soil mass fraction between 1 and 2 mm (f_{1-2}), soil mass fraction between 2 and 4 mm (f_{2-4}), and soil mass fraction greater than 4 mm ($f_{>4}$). These variables, corresponding to the measurements performed on soil aggregates and the mean weight diameter (MWD), were analyzed in each of the treatments defined in the experiment. The table includes indicators such as the minimum value, maximum value, mean, standard deviation, and coefficient of variation, which allows for the identification of the amplitude of the existing variability between treatments and the evaluation of the consistency of the measurements. This statistical characterization is fundamental, as it provides a clear view of the initial edaphic structure and how the different treatments can influence the distribution of soil aggregates—essential inputs for training the neural networks.

In particular, Table 2 offers the statistical description of the variables used as outputs in the neural networks, including the infiltration rate (T_{inf}), the total cation exchange capacity (CEC), the effective cation exchange capacity (CEC_{ef}) and the soil organic matter content (SOM). As in the input variables, the table details parameters such as extreme values, the mean, and the dispersion, allowing for the identification of relevant differences between treatments. This information is essential for evaluating the complexity of the system, since the chemical and functional properties of the soil respond differently according to the experimental conditions. Furthermore, these values help interpret the degree of heterogeneity and the magnitude of the changes that the neural networks must model.

The standardized skewness and kurtosis are particularly relevant here, as they serve to determine if the sample follows a normal distribution. Values of these statistics outside the range of -2 to +2 indicate significant deviations from normality, which would tend to invalidate any statistical test with reference to the standard deviation. It can be observed that the standardized skewness and standardized kurtosis values are within the expected range for data originating from a normal distribution.

Table 1. Statistical characterization of aggregate distribution in the soil (input variables).

	MWD (mm)	$f_{>4}$	f_{2-4}	f_{1-2}	$f_{0.5-1}$	$f_{0.25-0.5}$	$f_{<0.25}$
N	144	144	144	144	144	144	144
Average	3.27	54.08	13.89	3.22	4.49	5.46	18.86
Standard Deviation (SD)	1.07	23.63	6.37	2.09	4.14	5.35	16.09
Coefficient of Variation (CV)	0.33	0.44	0.46	0.65	0.92	0.98	0.85
Minimum	1.01	3.70	2.06	0.73	0.26	0.36	0.48
Maximum	4.72	87.38	35.45	10.34	17.97	23.91	64.22
Range	3.71	83.68	33.39	9.62	17.71	23.55	63.74
Standardized Skewness	-1.65	-1.79	1.89	1.61	1.64	1.42	1.75
Standardized Kurtosis	-1.91	-1.36	1.16	1.69	1.26	2.41	0.16

Source: Elaborated by the author

Table 2. Statistical characterization of soil fertility (output variables).

	T_{inf} (cm/h)	SOM (g.dm-3)	CEC (mmolc dm-3)	CEC_{ef} (mmolc dm-3)
N	144	144	144	144
Average	30.83	10.94	39.03	22.27
Standard Deviation (SD)	10.72	3.49	9.27	9.43
Coefficient of Variation (CV)	0.35	0.32	0.24	0.42
Minimum	14.15	4.36	25.20	10.20
Maximum	66.67	18.40	69.10	54.10
Range	52.52	14.04	43.90	43.90
Standardized Skewness	1.57	1.54	1.63	1.18
Standardized Kurtosis	1.89	-1.90	1.16	1.14

Source: Elaborated by the author

3.3.1 Neural Network Analysis

In this study, the artificial neural network (ANN) models were built using seven variables associated with the distribution and stability of soil aggregates as input data, including the mean weight diameter (MWD), a key indicator of edaphic structure. As output variables, the system sought to predict four fundamental properties for evaluating soil quality: the infiltration rate, the total cation exchange capacity (CEC), the effective CEC (ECEC), and the organic matter content (OM).

With the aim of improving the precision of the predictions, different modeling strategies were evaluated, experimenting with both different network architectures and the transformation of input variables. In a first approach, an architecture with two hidden layers was tested, varying the number of neurons per layer between 2 and 10 to identify configurations that would optimize the model's performance.

To evaluate the quality of the results, two central metrics were used: the coefficient of determination (R^2), which expresses the extent to which the predicted values explain the variability of the observed values, and the mean squared error (MSE), which quantifies the average magnitude of the error in the predictions during the validation phase.

The results, summarized in Table 3, show that the R^2 values obtained with this first strategy were relatively modest, ranging between 0.63 and 0.66, which indicates a limited model fit. Furthermore, it was confirmed that increasing the number of neurons in the hidden layers did not produce significant improvements in any of the evaluated indicators. This suggests that, under this initial configuration, the network did not adequately manage to capture the non-linear relationships between the soil structural variables and the functional properties that were sought to be predicted.

These observations justified the need to explore alternative configurations and more advanced data preprocessing strategies, aimed at improving the model's capacity to describe the complex physico-chemical dynamics of the soil.

Table 3. Statistical indicators of the neural network evaluation for predicting T_{inf}, CEC, CEC_{ef} y SOM.

Number of neurons in layer 1	Number of neurons in layer 2	MSE (Mean Squared Error)	R^2 Training	R^2 Validation	R^2 test	R^2 Complete
2	2	63.56	0.652	0.736	0.596	0.654
2	4	66.54	0.696	0.535	0.516	0.637
2	6	64.32	0.652	0.748	0.560	0.649
2	8	65.35	0.665	0.587	0.598	0.644
2	10	63.74	0.680	0.583	0.602	0.653
4	2	63.58	0.660	0.716	0.575	0.653
4	4	65.54	0.684	0.560	0.568	0.643
4	6	64.04	0.665	0.614	0.623	0.651
4	8	65.21	0.660	0.636	0.586	0.645
4	10	66.78	0.661	0.611	0.550	0.636
6	2	63.95	0.665	0.615	0.622	0.651
6	4	63.69	0.670	0.588	0.642	0.653
6	6	64.80	0.667	0.552	0.664	0.647
6	8	64.63	0.645	0.677	0.628	0.648
6	10	67.04	0.641	0.633	0.597	0.635
8	2	66.50	0.665	0.570	0.574	0.637
8	4	66.07	0.680	0.598	0.519	0.640
8	6	65.17	0.683	0.520	0.630	0.645
8	8	63.56	0.658	0.617	0.666	0.654
8	10	64.52	0.652	0.606	0.681	0.648
10	2	63.36	0.639	0.654	0.746	0.655
10	4	68.87	0.670	0.441	0.556	0.625
10	6	70.66	0.626	0.679	0.530	0.615
10	8	64.97	0.680	0.611	0.524	0.646
10	10	65.64	0.670	0.621	0.548	0.642

Source: Elaborated by the author

Table 4 presents the results corresponding to the evaluation of diverse neural network architectures applied to the dataset in which the output variables were transformed using the square root. This preprocessing strategy was implemented with the objective of reducing dispersion, stabilizing variance, and making it easier for the network to better capture the non-linear relationships between soil aggregates and the functional properties analyzed.

The performance indicators show a notable improvement in the coefficients of determination (R^2) compared to the models developed without transformation. In this phase, the R^2 values were located in a range of 0.70 to 0.72, which evidences a considerably more robust fit between the predicted and observed values.

This increase in model performance suggests that the square root transformation allowed for reducing the asymmetry and heterogeneity present in the original output variables, strengthening the network's ability to distinguish relevant patterns in the data, and decreasing the prediction error, which is reflected in more favorable metrics.

Table 4. Statistical indicators of the neural network evaluation for predicting $\sqrt{T_{inf}}$, $\sqrt{CEC}$, $\sqrt{CEC_{ef}}$ y $\sqrt{SOM}$.

Number of neurons in layer 1	Number of neurons in layer 2	MSE (Mean Squared Error)	R^2 Training	R^2 Validation	R^2 test	R^2 Complete
2	2	0.520	0.730	0.706	0.652	0.714
2	4	0.565	0.715	0.641	0.629	0.689
2	6	0.514	0.743	0.621	0.703	0.717
2	8	0.536	0.714	0.657	0.705	0.705
2	10	0.590	0.719	0.575	0.614	0.675
4	2	0.518	0.712	0.715	0.725	0.715
4	4	0.510	0.734	0.703	0.666	0.719
4	6	0.564	0.684	0.762	0.643	0.690
4	8	0.513	0.731	0.678	0.699	0.718
4	10	0.517	0.721	0.741	0.668	0.715
6	2	0.514	0.721	0.730	0.681	0.717
6	4	0.534	0.741	0.656	0.621	0.706
6	6	0.521	0.729	0.712	0.635	0.713
6	8	0.510	0.739	0.676	0.675	0.719
6	10	0.528	0.742	0.564	0.740	0.710
8	2	0.602	0.667	0.691	0.648	0.669
8	4	0.516	0.765	0.707	0.531	0.716
8	6	0.535	0.726	0.664	0.648	0.705
8	8	0.511	0.738	0.630	0.725	0.719
8	10	0.527	0.735	0.685	0.618	0.710
10	2	0.518	0.718	0.726	0.687	0.715
10	4	0.508	0.735	0.698	0.674	0.721
10	6	0.522	0.721	0.739	0.640	0.713
10	8	0.517	0.727	0.713	0.654	0.716
10	10	0.533	0.722	0.664	0.668	0.706

Source: Elaborated by the author

Upon evidencing that the infiltration rate showed a weaker relationship with the input variables—which could limit the network's capacity to precisely fit the complete model—a new test was chosen, reducing the number of output variables. In this stage, the neural network was trained only to predict the square root of the cation exchange capacity (CEC) and the organic matter content (OM), two properties closely linked to the physical structure and chemical quality of the soil, and which also showed greater coherence with the analyzed aggregate patterns.

The results of this strategy, presented in Table 5, show a significant improvement in model performance: the coefficients of determination (R^2) were situated in a range of 0.87 to 0.90, values superior to those achieved in the previous trials. This improvement suggests that the reduction in the number of output variables allowed the neural network to concentrate its learning capacity on more consistent and less noisy relationships within the dataset.

This increase in fit is explained because:

- CEC and organic matter are more directly related to soil structural stability and aggregate size, which facilitates their modeling from the input variables.
- The exclusion of the infiltration rate, whose relationship with aggregates is more complex and is influenced by multiple hydrodynamic factors not included in the model, reduced the residual variability.
- With fewer simultaneous objectives, the network was better able to optimize the synaptic weights, capturing non-linear patterns with greater precision.

In synthesis, this result confirms that the adequate selection of output variables—aligned with the explanatory capabilities of the dataset—is a key step to improve the predictive efficacy of neural models applied to the characterization of edaphic properties.

Table 5. Statistical indicators of the neural network evaluation for predicting $\sqrt{CEC}$ y $\sqrt{SOM}$.

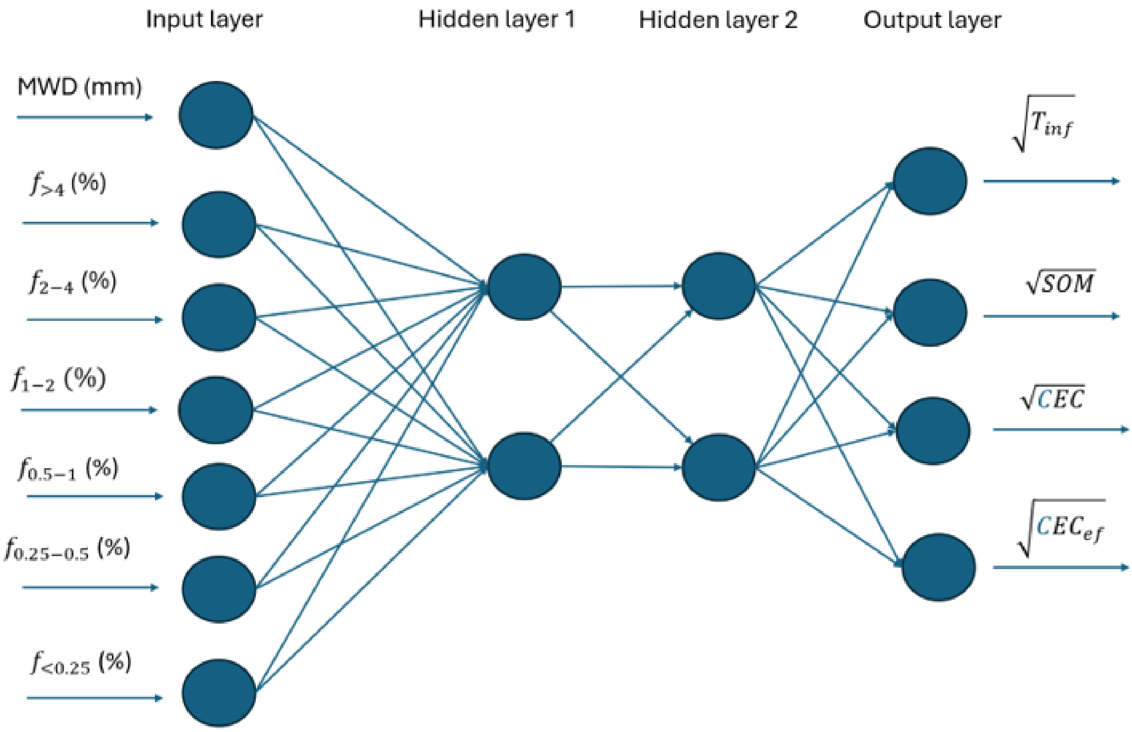
Number of neurons in layer 1	Number of neurons in layer 2	MSE (Mean Squared Error)	R^2 Training	R^2 Validation	R^2 test	R^2 Complete
2	2	0.257	0.897	0.879	0.932	0.899
2	4	0.321	0.872	0.912	0.847	0.874
2	6	0.291	0.898	0.878	0.847	0.886
2	8	0.257	0.902	0.921	0.864	0.899
2	10	0.267	0.903	0.864	0.891	0.895
4	2	0.261	0.902	0.872	0.903	0.898
4	4	0.267	0.905	0.906	0.844	0.895
4	6	0.258	0.902	0.898	0.881	0.899
4	8	0.256	0.907	0.872	0.893	0.899
4	10	0.258	0.910	0.862	0.888	0.899
6	2	0.255	0.897	0.896	0.920	0.900
6	4	0.265	0.900	0.884	0.889	0.896
6	6	0.258	0.913	0.845	0.887	0.899
6	8	0.264	0.913	0.844	0.885	0.896
6	10	0.256	0.901	0.900	0.892	0.899
8	2	0.264	0.891	0.928	0.886	0.896
8	4	0.272	0.900	0.861	0.898	0.893
8	6	0.257	0.902	0.878	0.903	0.899
8	8	0.258	0.900	0.891	0.902	0.899
8	10	0.255	0.898	0.910	0.897	0.900
10	2	0.260	0.903	0.888	0.886	0.898
10	4	0.268	0.913	0.865	0.839	0.895
10	6	0.261	0.907	0.902	0.847	0.898
10	8	0.274	0.905	0.869	0.866	0.893
10	10	0.258	0.898	0.918	0.884	0.899

Source: Elaborated by the author

Networks with a greater number of layers were tested, with the number of neurons ranging between 2 and 10. These networks with more complex architectures did not improve the predictions.

Since it has been demonstrated that increasing the number of neurons in the two hidden layers used does not significantly improve the predictions, the decision was made to evaluate networks with two neurons in each layer (Figure 1), whose results are shown in Table 6.

Figure 2. Structure of the tested neural network with 2 hidden layers with 2 neurons each for predicting $\sqrt{T_{inf}}$, $\sqrt{SOM}, \sqrt{CEC}$ y $\sqrt{CEC_{ef}}$.



Source: Elaborated by the author

Table 6. Weights and bias of the 2-hidden layer network for the outputs $\sqrt{T_{inf}}$, $\sqrt{SOM}, \sqrt{CEC}$ y $\sqrt{CEC_{ef}}$.

	Soil Input Variables							
	MWD (mm)	$f_{>4}$ (%)	f_{2-4} (%)	f_{1-2}(%)	$f_{0,5-1}$ (%)	$f_{0,25-0,5}$ (%)	$f_{<0,25}$(%)	B_{input}
N11	-0,273	1,403	0,336	0,533	0,353	-1,057	1,804	-0,447
N12	-0,330	1,043	1,576	-0,324	1,844	1,725	1,034	-0,299

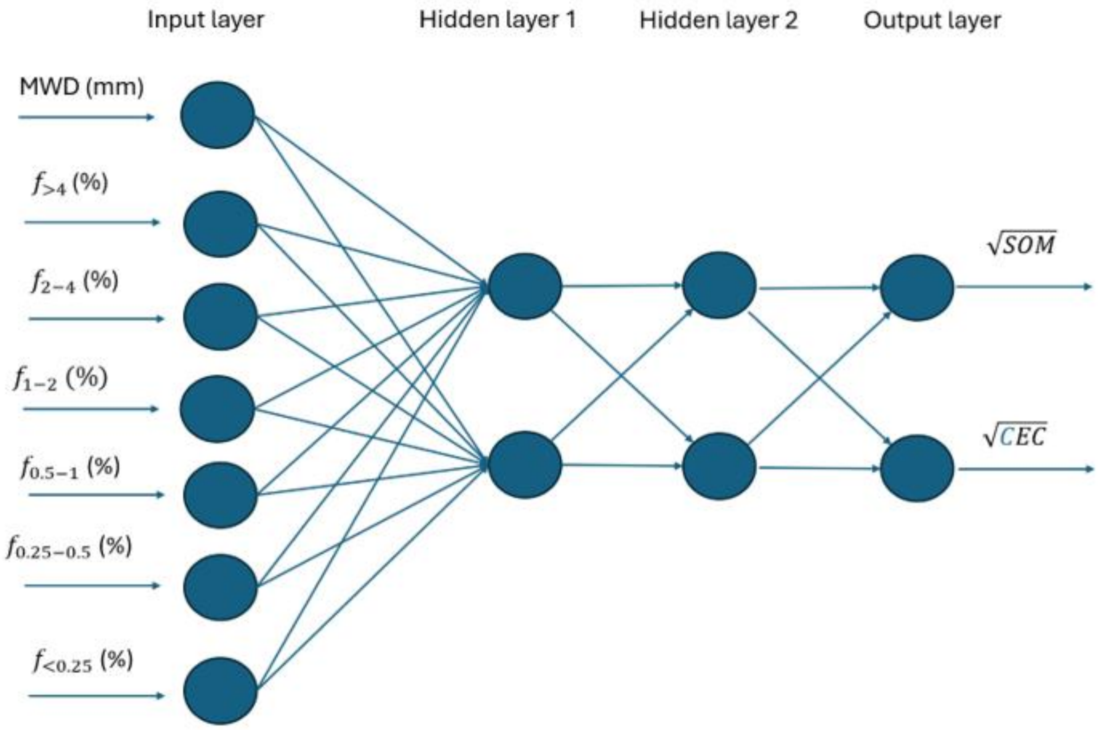
	N11	N12	B_2
N21	-3,308	2,591	2,033
N22	0,793	-3,285	2,958

	N21	N22	B_{output}
NO1	0,302	-0,766	-0,412
NO2	-1,625	1,821	-0,747
NO3	-1,104	0,861	-1,130
NO4	-0,836	0,397	-1,146

Source: Elaborated by the author

Figure 2 shows the neural network used to evaluate the prediction of the output variables $\sqrt{MO}$ and $\sqrt{CEC}$. The results are shown in Table 7.

Figure 3. Structure of the tested neural network with 2 hidden layers with 2 neurons each for predicting $\sqrt{MO}$ y $\sqrt{CEC}$



Source: Elaborated by the author

Table 7. Weights of the 2-hidden layer network for the outputs $\sqrt{MO}$ and $\sqrt{CEC}$.

Soil Input Variables								
	MWD (mm)	$f_{>4}$ (%)	f_{2-4} (%)	f_{1-2}(%)	$f_{0,5-1}$ (%)	$f_{0,25-0,5}$ (%)	$f_{<0,25}$(%)	B_{input}
N11	-0,763	1,404	0,833	0,039	13,169	0,781	-2,649	8,953
N12	-2,839	-1,497	-2,379	-0,085	-2,908	-0,661	-7,035	-2,220

	N11	N12	B_2
N21	-3,308	2,591	2,033
N22	0,793	-3,285	2,958

	N21	N22	B_{output}
NO1	-0,089	-1,317	0,868
NO2	-0,400	-1,460	0,835

Source: Elaborated by the author

These networks stood out for their capacity for generalization and precision in predicting the edaphic variables of interest, which demonstrates the potential of using artificial neural networks in the modeling of processes related to soil structural dynamics.

3.4 DISCUSSION

When developing predictive models using neural networks to estimate soil quality, based on aggregate size, it is essential to understand and address the inherent challenges of this type of approach. While neural networks offer great advantages for modeling complex relationships, they also present certain obstacles that can affect the accuracy and applicability of the results. These challenges focus mainly on data quality, model generalization, and the interaction with other relevant variables that also impact soil properties.

Data quality is one of the most critical factors when training a predictive model. In an environment such as soil analysis, where variables are inherently complex and variable, the precision of the model largely depends on the quality and representativeness of the data used during training. If the input data—for example, aggregate sizes—is contaminated by noise or is insufficient in terms of quantity or diversity, the model's predictions can be highly inaccurate.

Noise in the data can arise from various sources, such as errors in measurement methods, fluctuations in environmental conditions during sampling, or inconsistencies in data collection procedures. Noisy data could distort the pattern that the neural network attempts to learn, leading it to formulate incorrect relationships between the variables. This not only reduces the accuracy of the predictions but can also diminish the model's reliability.

On the other hand, data insufficiency is another significant challenge. A model trained with a limited number of samples or without adequate representation of the variability in soils (in terms of pH, texture, aggregate size, etc.) could overfit to the specific characteristics of the available data, without being able to generalize to other conditions or regions. It is for this reason that having a broad, diverse, and representative dataset is fundamental for training a neural network that is truly useful in practical scenarios. The collection of high-quality data is, therefore, an essential step in creating reliable and robust models.

3.4.1 Generalization

One of the main challenges of any machine learning model, including neural networks, is generalization. Generalization refers to the model's ability to make accurate predictions not only on the data it has been trained on but also on new data it has never seen before. This is a fundamental challenge because, in practice, the model is expected to be applied in diverse situations and with new data that may have different characteristics than the training data.

If the model over-adjusts to the specific training data, it can fall into overfitting, a problem where the neural network excessively memorizes the data, learning patterns that only exist in that particular set. Although this approach can yield good results on the training set, the model loses the ability to generalize to new data, which compromises its practical utility. Overfitting can lead to incorrect predictions when the model is applied to soils or conditions different from those represented in the original dataset.

To avoid this problem, regularization and cross-validation techniques are usually employed during training, which help ensure that the model does not learn irrelevant or trivial patterns. Cross-validation, for example, allows the model's performance to be evaluated on subsets of data that have not been used in its training, which helps verify that the neural network has truly learned general patterns that apply to a wide range of data.

3.4.2 Interaction with Other Variables

Although aggregate size is an important variable when predicting soil characteristics, such as CEC, infiltration, and organic matter, it should not be considered in isolation. Soils are complex multivariate systems, where multiple factors interact with each other, and each influences the physical, chemical, and biological properties of the soil. Therefore, aggregate size cannot be fully explained without considering other variables.

Factors such as pH, soil texture (the proportion of sand, clay, and silt), temperature, or even the local climate have a direct impact on the formation, stability, and functionality of aggregates, as well as on the soil's capacity to retain nutrients and water.

For example, a soil with an acidic pH may have a greater difficulty in forming stable aggregates, which in turn affects the infiltration capacity and nutrient retention. Likewise, soil texture influences the way aggregates group together, as clay soils tend to form more stable aggregates than sandy soils.

The interaction between these variables can be complex, and neural networks must be sophisticated enough to capture these non-linear relationships. However, multicollinearity (when two or more variables are strongly correlated) can complicate this process, as the model may have difficulty discriminating between the individual contributions of each variable.

For example, if aggregate size and pH are correlated in the data, the model may have difficulty determining which of these variables has a more significant influence on CEC or infiltration.

To address this difficulty, it is crucial to include a wide variety of variables in the model, and not just aggregate size. Additionally, it may be necessary to perform a preliminary analysis of the correlations between variables and apply dimensionality reduction techniques, such as Principal Component Analysis (PCA), to simplify the model without losing precision. An approach that considers the interaction between multiple variables can significantly improve predictions, making the model more robust and applicable to a wider range of conditions.

These obtained results are consistent with recent literature that underscores both the potential of ANNs to estimate soil properties and the critical importance of preprocessing and the selection of target variables.

3.4.3 Relationship Between Aggregates, CEC, and Organic Matter

The high R^2 values obtained when modeling CEC and organic matter from the mean weight diameter (MWD) and aggregate size distribution confirm that the soil's physical structure is a good predictor of its chemical functioning.

In agricultural and forestry systems, numerous studies have shown that macroaggregates (> 2 mm) tend to concentrate higher organic carbon and contribute to the stabilization of organic matter and the increase of effective CEC, reinforcing the role of aggregation as a soil quality indicator (Mechri *et al.*, 2023).

In this work, the strong predictive capacity for CEC and organic matter suggests that the information contained in the aggregate distribution and the MWD captures a good part of the processes of physical protection of carbon, organo-mineral complexation, and increase of the active specific surface area.

Studies based on ANNs and other machine learning algorithms have shown similar results when using textural and carbon variables to predict CEC in different pedoclimatic contexts, with high R^2 values and moderate prediction errors, which supports the robustness of the adopted approach (Emamgholizadeh *et al*., 2023).

3.4.4 ANN Performance and Comparison with Other Approaches

The evolution of the model's performance—from $R^2 \approx 0.65$ to values approaching 0.90 through variable transformation and reduction of output dimensions—is consistent with what has been reported in the literature for complex edaphic problems. Studies applying ANNs to predict soil organic matter based on topographic or land-use variables indicate substantial improvements in explanatory capacity when the network architecture is carefully adjusted and transformations are applied to reduce asymmetry and heterogeneity in the data (Guo *et al*., 2013).

Similarly, recent reviews on AI applied to soil analysis highlight that ANN performance depends strongly on: (i) the quality and representativeness of the training set, (ii) the selection of activation functions suitable for nonlinear relationships, and (iii) the use of normalization and regularization techniques to prevent overfitting (Awais *et al*., 2023). In this study, input-variable normalization, cross-validation, and early stopping contributed to achieving stable models with a sound balance between fit and generalization.

The fact that increasing the number of neurons and the architectural complexity did not improve predictions reflects another typical outcome in edaphic regression problems: greater complexity does not necessarily imply better performance when the amount of data is limited or experimental noise is significant. Several studies show that relatively simple models (few layers and neurons) can match or even surpass deeper architectures, provided that input variables are well selected and properly preprocessed (Maino *et al*., 2024).

3.4.5 Difficulties in Predicting Infiltration

In contrast to CEC and organic matter, the infiltration rate proved to be the most difficult variable to model, reducing the overall performance when included as a joint output.

This is consistent with recent studies on infiltration prediction using AI, where even comparing multiple models (ANN, SVM, GMDH, MARS) the authors highlight a strong sensitivity of infiltration to local hydrodynamic factors—pore structure, macropore continuity, surface crusting, initial moisture—that are not always represented in the available input variables (Mehta *et al*., 2024).

Thus, although soil aggregation clearly influences infiltration, the dynamics of flow in the field also depend on antecedent moisture, the presence of biopores, localized compaction, and surface roughness, among other factors. The omission of these variables can explain both the lower correlation observed between aggregates and infiltration and the improvement of the model upon removing this variable from the outputs.

Studies in Mediterranean agroecosystems have shown that hillslope position, cover management, and aggregate size strongly interact to simultaneously control erosion, infiltration, and soil quality, which reinforces the idea that infiltration is an emergent parameter of multiple processes (Brito *et al*., 2025).

3.4.6 Implications for Soil Quality Evaluation and Management

From an applied standpoint, the capacity to estimate CEC and organic matter with high precision from relatively simple and low-cost measurements (dry sieving, MWD calculation) opens the door to rapid soil quality monitoring systems in extensive areas. Literature from MDPI and Springer converges in pointing to aggregation, organic matter, and CEC as core indicators of soil quality, especially in contexts of sustainable agriculture and soil conservation (Mechri *et al*., 2023).

In this sense, the results of this work suggest that:

- It is possible to integrate simple structural measures with AI models to derive soil fertility and health indicators without the need for exhaustive chemical analyses in all plots.

- Spatial maps of aggregation could be used as a proxy for the distribution of CEC and organic matter, facilitating management zoning, the definition of differentiated doses of fertilizers and amendments, and the evaluation of practices such as conservation agriculture or the incorporation of organic residues.
- In degraded soils, such as those studied, where the reconstruction of structure and the increase of organic matter are priority objectives, ANNs can help identify management combinations (organic/mineral fertilization, tree species, planting density) that maximize structural and chemical benefits.

3.4.7 Limitations and Research Perspectives

Nevertheless, the study presents some limitations that must be considered when extrapolating the results. Firstly, the models were trained with data originating from a single soil type (Red Latosol/Decapitated Ferralsol) and under specific climatic and management conditions, which may restrict generalization to other contexts. Multicentric studies that combine soils from different countries have shown that the transfer of models for CEC or other properties between regions requires recalibration or hybrid learning and ensemble strategies to maintain accuracy (Khaledian *et al*., 2017).

Secondly, although progress was made in improving the model through variable transformations and output selection, other machine learning algorithms (e.g., Random Forest, Gradient Boosting, SVM) were not systematically explored. These algorithms have shown, in some cases, performance equal to or superior to that of ANNs for soil hydraulic and chemical properties (Mehta *et al*., 2024). Direct comparisons between different algorithms, based on error metrics and generalization capacity, could help define more clearly in which cases ANNs prove to be the most advantageous option.

Finally, the incorporation of additional variables—such as pH, detailed texture, biological indicators, or topographical and soil use information—could improve the prediction of more complex parameters like infiltration, as well as further refine the estimation of CEC and organic matter. In this sense, recent advances in the integration of remote sensing data, proximal sensors, and AI models offer a promising framework to expand the approach presented here (Maino *et al*., 2024).

3.5 CONCLUSION

In this work, it has been demonstrated that once the artificial neural network (ANN) models are trained and validated, their capacity to predict key edaphic parameters accurately and efficiently, based on soil structural variables, specifically aggregate size and mean weight diameter (MWD), is verified. In this study, seven input variables were used, corresponding to the aggregate size distribution and the mean weight diameter (MWD), and four output variables: infiltration rate, total cation exchange capacity (CEC), effective CEC, and soil organic matter content (SOM).

The training process involved diverse ANN architectures, and those that presented the best performance in terms of the coefficient of determination (R^2) were selected. These networks showed a high fit between the predicted and observed values, with no signs of overfitting, which supports their robustness and reliability.

The possibility of predicting key soil properties solely based on easily observable physical characteristics, such as aggregate distribution, represents a significant advancement for monitoring and managing soil quality. This methodology allows for reducing the need for costly and prolonged laboratory analyses, offering an agile alternative for evaluating soil health. Furthermore, its practical applications are multiple: from the optimization of fertilization and irrigation strategies, adjusted to the infiltration capacity and nutrient availability, to the design of sustainable management plans that prioritize the long-term conservation of organic matter and edaphic functionality.

In this sense, the use of ANNs applied to soil science proves to be a powerful tool for informed agronomic decision-making. By integrating artificial intelligence approaches with pedological knowledge, a promising field is opened for the development of more sustainable, efficient, and resilient agricultural practices in the face of contemporary challenges in agriculture.

REFERENCES

ABU-HAMDEH, N. H.; ABO-QUDAIS, S. A.; OTHMAN, A. M. Effect of soil aggregate size on infiltration and erosion characteristics. **European Journal of Soil Science**, v. 57, n. 5, p. 609–616, 2006. DOI: 10.1111/j.1365-2389.2005.00743.x.

ÁLVARO-FUENTES, J.; CANTERO-MARTÍNEZ, C.; LÓPEZ, M. V.; PAUSTIAN, K.; DENEF, K.; STEWART, C. E.; ARRÚE, J. L. Soil aggregation and soil organic carbon stabilization: effects of management in semiarid Mediterranean agroecosystems. **Soil Science Society of America Journal**, v. 73, n. 5, p. 1519–1529, 2009. DOI: 10.2136/sssaj2008.0333.

AMÉZKETA, E. Soil aggregate stability: a review. **Journal of Sustainable Agriculture**, v. 14, n. 2–3, p. 83–151, 1999. DOI: 10.1300/J064v14n02_08.

AMINI, M.; ABBASPOUR, K. C.; KHADEMI, H.; FATHIANPOUR, N.; AFYUNI, M.; SCHULIN, R. Neural network models to predict cation exchange capacity in arid regions of Iran. **European Journal of Soil Science**, v. 56, n. 4, p. 551–559, 2005. DOI: 10.1111/j.1365-2389.2005.00698.x.

AWAIS, M.; NAQVI, S. M. Z. A.; ZHANG, H.; LI, L.; ZHANG, W.; AWWAD, F. A.; ISMAIL, E. A. A.; KHAN, M. I.; RAGHAVAN, V.; HU, J. AI and machine learning for soil analysis: an assessment of sustainable agricultural practices. **Bioresources and Bioprocessing**, v. 10, n. 1, art. 90, 2023. DOI: 10.1186/s40643-023-00710-y.

BLANCO-CANQUI, H.; BENJAMIN, J. G. Impacts of soil organic carbon on soil physical behavior. *In:* LOGSDON, S.; BERLI, M.; HORN, R. (ed.). **Advances in Agricultural Systems Modeling**. Madison: American Society of Agronomy; Soil Science Society of America, 2013. p. 11–40. DOI: 10.2134/advagricsystmodel3.c2.

BONINI, C. S. B.; DELFIM, J. J.; LOURENCETTI, J.; NANTES, L. S.; SANTOS, M. A.; SOUZA, J. A. L. Residual effects of cover crops and soil amendments on carbon content and stock in a degraded soil. **Engenharia Agrícola**, v. 45, n. spe 1, e20240188, 2025. DOI: 10.1590/1809-4430-Eng.Agric.v45nespe120240188/2025.

BRITO, L. da C.; SEVERO, E. M.; JIMENEZ, P. A. J.; SILVA, A. O.; AVANZI, J. C.; SANTOS, D.; CARNEIRO, M. A. C.; SILVA, M. L. N. Soil quality indicators and water erosion in olive groves (*Olea europaea* L.) under different vegetation cover management. **Soil Systems**, v. 9, n. 3, art. 96, 2025. DOI: 10.3390/soilsystems9030096.

BRONICK, C. J.; LAL, R. Soil structure and management: a review. **Geoderma**, v. 124, n. 1–2, p. 3–22, 2005. DOI: 10.1016/j.geoderma.2004.03.005.

CANTARELLA, H.; QUAGGIO, J. A.; MATTOS JUNIOR, D.; BOARETTO, R. M.; RAIJ, B. VAN. **Boletim 100:** recomendações de adubação e calagem para o estado de São Paulo. Campinas: Instituto Agronômico, 2022.

CASTRO FILHO, C.; LOURENÇO, A.; GUIMARÃES, M. DE F.; FONSECA, I. C. B. Aggregate stability under different soil management systems in a red latosol in the state of Paraná, Brazil. **Soil and Tillage Research**, v. 65, n. 1, p. 45–51, 2002. DOI: 10.1016/S0167-1987(01)00275-6.

CASTRO FILHO, C.; MUZILLI, O.; PODANOSCHI, A. L. Estabilidade dos agregados e sua relação com o teor de carbono orgânico num Latossolo Roxo distrófico, em função de sistemas de plantio, rotações de culturas e métodos de preparo das amostras. **Revista Brasileira de Ciência do Solo**, v. 22, n. 3, p. 527–538, 1998. DOI: 10.1590/S0100-06831998000300019.

CASTRO, C.; LOGAN, T. J. Liming effects on the stability and erodibility of some Brazilian Oxisols. **Soil Science Society of America Journal**, v. 55, n. 5, p. 1407–1413, 1991. DOI: 10.2136/sssaj1991.03615995005500050034x.

DING, G. C.; PRONK, G. J.; BABIN, D.; HEUER, H.; HEISTER, K.; KÖGEL-KNABNER, I.; SMALLA, K. Mineral composition and charcoal determine the bacterial community structure in artificial soil. **FEMS Microbiology Ecology**, v. 86, n. 1, p. 15–25, 2013.

DORJI, T.; FIELD, D. J.; ODEH, I. O. A. Soil aggregate stability and aggregate-associated organic carbon under different land use or land cover types. **Soil Use and Management**, v. 35, n. 4, p. 670–681, 2019. DOI: 10.1111/sum.12549.

EDMEADES, D. C.; CLINTON, O. E. A simple rapid method for the measurement of exchangeable cations and effective cation exchange capacity. **Communications in Soil Science and Plant Analysis**, v. 12, n. 7, p. 683–695, 1981. DOI: 10.1080/00103628109367184.

EMAMGHOLIZADEH, S.; BAZOOBANDI, A.; MOHAMMADI, B.; GHORBANI, H.; AMEL SADEGHI, M. Prediction of soil cation exchange capacity using enhanced machine learning approaches in the southern region of the Caspian Sea. **Ain Shams Engineering Journal**, v. 14, n. 2, art. 101876, 2023. DOI: 10.1016/j.asej.2022.101876.

FOOLADMAND, H. R. Estimating cation exchange capacity using soil textural data and soil organic matter content: a case study for the south of Iran. **Archives of Agronomy and Soil Science**, v. 54, n. 4, p. 381–386, 2008. DOI: 10.1080/03650340802158193.

GARCIA, A. R. DE M.; CORRÊA, R. S. Revegetation of areas degraded by mining in the cerrado: the use of sewage sludge in soil rehabilitation. **Disciplinarum Scientia: Ciências Naturais e Tecnológicas**, v. 25, p. 123–144, 2024. DOI: 10.37779/nt.v25i2.4907.

GIÁCOMO, R. G.; ALVES, M. C.; CAMARA, R.; PEREIRA, M. G.; FILHO, S. N. S.; DE MORAES, M. L. T. Performance of species *Mabea fistulifera* and *Eucalyptus urograndis* with use of cellulose residue in degraded areas. **Floresta**, v. 49, n. 2, p. 363–372, 2019. DOI: 10.5380/rf.v49i2.58256.

GIACOMO, R. G. Recuperação do solo usando composto produzido com resíduo da extração de celulose. 2013. Tese (Doutorado em Ciência e Tecnologia de Materiais) – Universidade Estadual Paulista, Ilha Solteira, 2013.

GUO, P.-T.; WU, W.; SHENG, Q.-K.; LI, M.-F.; LIU, H.-B.; WANG, Z.-Y. Prediction of soil organic matter using artificial neural network and topographic indicators in hilly areas. **Nutrient Cycling in Agroecosystems**, v. 95, n. 3, p. 333–344, 2013. DOI: 10.1007/s10705-013-9566-9.

HAYDU-HOUDESHELL, C.-A.; et al. Soil aggregate stability under chaparral species in southern California. **Geoderma**, v. 310, p. 201–208, 2018.

HODDINOTT, K. B. A method for evaluating the soil's effective cation exchange capacity and exchangeable base content using silver-thiourea. *In:* **Hazardous and Industrial Solid Waste Testing**. ASTM, 1983. DOI: 10.1520/STP33543S.

IUSS WORKING GROUP WRB. **World Reference Base for Soil Resources: international soil classification system for naming soils and creating legends for soil maps**. 4. ed. Vienna: IUSS, 2022.

KASHI, H.; EMAMGHOLIZADEH, S.; GHORBANI, H. Estimation of soil infiltration and cation exchange capacity based on multiple regression, ANN (RBF, MLP), and ANFIS models. **Communications in Soil Science and Plant Analysis**, v. 45, n. 9, p. 1195–1213, 2014. DOI: 10.1080/00103624.2013.874029.

KEMPER, W. D.; ROSENAU, R. C. Aggregate stability and size distribution. *In:* **Methods of soil analysis: physical and mineralogical methods**. Madison: Soil Science Society of America, 1986. p. 425–442. DOI: 10.2136/sssabookser5.1.2ed.c17.

KHALEDIAN, Y.; BREVIK, E. C.; PEREIRA, P.; CERDÀ, A.; FATTAH, M. A.; TAZIKEH, H. Modeling soil cation exchange capacity in multiple countries. **Catena**, v. 158, p. 194–200, 2017. DOI: 10.1016/j.catena.2017.07.002.

KOCHIIERU, M.; FEIZIENE, D.; FEIZA, V.; VOLUNGEVICIUS, J.; VELYKIS, A.; SLEPETIENE, A.; DEVEIKYTE, I.; SEIBUTIS, V. Freezing-thawing impact on aggregate stability as affected by land management, soil genesis and soil chemical and physical quality. **Soil and Tillage Research**, 2020.

KORENKOVA, L.; URIK, M. Infiltration variability in agricultural soil aggregates caused by air slaking. **Eurasian Soil Science**, v. 51, n. 4, p. 428–433, 2018. DOI: 10.1134/S1064229318040087.

LAL, R.; KIMBLE, J.; FOLLETT, R. F. Pedospheric processes and the carbon cycle. *In:* **Methods for assessment of soil degradation**. Boca Raton: CRC Press, 1997. p. 1–8.

LIPIEC, J.; WALCZAK, R.; WITKOWSKA-WALCZAK, B.; NOSALEWICZ, A.; SŁOWIŃSKA-JURKIEWICZ, A.; SŁAWIŃSKI, C. The effect of aggregate size on water retention and pore structure of two silt loam soils of different genesis. **Soil and Tillage Research**, v. 97, n. 2, p. 239–249, 2007. DOI: 10.1016/j.still.2007.10.001.

LIU, X.; HERBERT, S. J.; HASHEMI, A. M.; ZHANG, X.; DING, G. Effects of agricultural management on soil organic matter and carbon transformation: a review. **Plant, Soil and Environment**, v. 52, n. 12, p. 531–543, 2006. DOI: 10.17221/3544-PSE.

LOZANO OLIVÉRIO, G.; DOS SANTOS BATISTA BONINI, C.; FERNANDA DIAS SOUZA, J.; LUÍS SANCHEZ PERUSSO, R.; MEIRELLES, G. C.; ANDRIGHETTO, C.; LUPATINI, G. C.; MATEUS, G. P.; LAPAZ, A. M.; HEINRICHS, R.; MOREIRA, A. Water infiltration, resistance to penetration and soil moisture in integrated agricultural yield systems over time. **Communications in Soil Science and Plant Analysis**, v. 53, n. 3, p. 327–336, 2022. DOI: 10.1080/00103624.2021.1993888.

MAINO, A.; ALBERI, M.; BARBAGLI, A.; CHIARELLI, E.; COLONNA, T.; FRANCESCHI, M.; GALLORINI, F.; GUASTALDI, E.; LOPANE, N.; MANTOVANI, F.; PETRONE, D.; PIERINI, S.; RAPTIS, K. G. C.; STRATI, V.; XHIXHA, G. A deep neural network for predicting soil texture using airborne radiometric data. **Radiation Physics and Chemistry**, v. 221, art. 111767, 2024. DOI: 10.1016/j.radphyschem.2024.111767.

MANRIQUE, L. A.; JONES, C. A.; DYKE, P. T. Predicting cation-exchange capacity from soil physical and chemical properties. **Soil Science Society of America Journal**, v. 55, n. 3, p. 787–794, 1991. DOI: 10.2136/sssaj1991.03615995005500030026x.

MECHRI, M.; BOUAJILA, K.; ZIADI, N.; RAZA, T.; BELTAYEF, H.; EASH, N. S.; JEDIDI, N.; GHARBI, A.; HASHEM, A.; ABD_ALLAH, E. F. Soil aggregation and associated organic carbon and total nitrogen in a sandy loam soil under long-term tillage effects. **Agronomy**, v. 13, n. 10, art. 2520, 2023. DOI: 10.3390/agronomy13102520.

MEHTA, V.; HASANVAND, S.; SEPAHVAND, A.; SIHAG, P.; BEIRANVAND, N.; SINGH, B. A benchmark comparison of AI-based modeling of soil infiltration rates. **Journal of Hydroinformatics**, v. 26, n. 12, p. 3060–3079, 2024. DOI: 10.2166/hydro.2024.086.

NANTES, L. S.; BONINI, C. D. S. B.; SANTOS, M. A.; DA SILVA ALVES, L.; LOURENCETTI, J.; BONINI NETO, A.; DE SOUZA, J. A. L.; DA SILVA LUNARDELLI, G. A.; SILVA, M. B.; ALVARES, D. V.; PAZ-GONZALEZ, A. Exposed subsoil under recovery for 30 years: influence of management strategies on soil physical attributes and organic matter. **Soil Systems**, v. 9, n. 1, art. 17, 2025. DOI: 10.3390/soilsystems9010017.

OKOLO, C. C.; GEBRESAMUEL, G.; ZENEBE, A.; HAILE, M.; EZE, P. N. Accumulation of organic carbon in various soil aggregate sizes under different land-use systems in a semi-arid environment. **Agriculture, Ecosystems & Environment**, v. 297, art. 106924, 2020. DOI: 10.1016/j.agee.2020.106924.

OLIVERIO, A. M.; GEISEN, S.; DELGADO-BAQUERIZO, M.; MAESTRE, F. T.; TURNER, B. L.; FIERER, N. The global-scale distributions of soil protists and their contributions to belowground systems. **Science Advances**, v. 6, n. 4, eaax8787, 2020. DOI: 10.1126/sciadv.aax8787.

RAIJ, B. VAN; ANDRADE, J. C. DE; CANTARELLA, H.; QUAGGIO, J. A. **Análise química para avaliação da fertilidade de solos tropicais**. Campinas: Instituto Agronômico, 2001.

SILVA, F. R. F.; SCHMITT, E. Thermal comfort and grazing behavior of Girolando heifers in integrated crop-livestock (ICL) and crop-livestock-forest (ICLF) systems. **Acta Scientiarum. Animal Sciences**, v. 41, p. 1–10, 2019.

SINQUINI DE SOUZA, R.; SOUSA DE MORAIS, I.; ROSSET, J. S.; DE MELO RODRIGUES, T.; LOSS, A.; PEREIRA, M. G. Aggregation as soil quality indicator in areas under different uses and managements. **Farming System(s)**, v. 2, n. 2, art. 100082, 2024. DOI: 10.1016/j.farsys.2024.100082.

SOIL SURVEY STAFF. **Keys to Soil Taxonomy**. 13. ed. Washington, DC: USDA-Natural Resources Conservation Service, 2022.

TANG, L.; ZENG, G.; NOURBAKHSH, F.; SHEN, G. L. Artificial neural network approach for predicting cation exchange capacity in soil based on physico-chemical properties. **Environmental Engineering Science**, v. 26, n. 1, p. 139–146, 2009. DOI: 10.1089/ees.2007.0238.

ZHANG, R. Determination of soil sorptivity and hydraulic conductivity from the disk infiltrometer. **Soil Science Society of America Journal**, v. 61, n. 4, p. 1024–1030, 1997. DOI: 10.2136/sssaj1997.03615995006100040005x.

4 CONCLUSÃO GERAL

O presente trabalho atingiu plenamente seu objetivo geral de avaliar a recuperação física e química de um Latossolo degradado utilizando modelagem multivariada e inteligência artificial (ACP + RNP/ANN) para identificar os tratamentos mais eficazes. A integração dos dois estudos complementares demonstrou, de forma robusta e inovadora, o potencial sinérgico entre práticas de manejo baseadas em espécies nativas e a aplicação de ferramentas analíticas avançadas para a predição e interpretação da qualidade do solo.

O Capítulo I evidenciou que o plantio da espécie nativa *Mabea fistulifera*, combinado com a dose de 20 Mg ha^{-1} de composto orgânico (D20), foi o tratamento mais eficiente na recuperação do solo degradado. Este manejo promoveu melhorias substanciais nos atributos edáficos, elevando os teores de matéria orgânica, a macroporosidade, a capacidade de troca catiônica e a taxa de infiltração de água, enquanto reduziu a densidade do solo. A aplicação da Análise de Componentes Principais (ACP) integrada às Redes Neurais Probabilísticas (RNP) permitiu uma avaliação multidimensional e probabilística, identificando de forma objetiva a "zona ótima" de recuperação e validando a superioridade do tratamento D20 com *Mabea fistulifera* sobre o *Eucalyptus urograndis*.

Complementarmente, o Capítulo II confirmou o alto potencial das Redes Neurais Artificiais (ANN) como ferramentas preditivas na ciência do solo. Os modelos desenvolvidos demonstraram capacidade de generalização para estimar indicadores-chave de qualidade, como a capacidade de troca catiônica (CTC) e o teor de matéria orgânica, a partir de variáveis estruturais facilmente mensuráveis (distribuição e diâmetro médio ponderado dos agregados). A otimização do modelo, com transformação das variáveis de saída e foco em predições específicas, resultou em coeficientes de determinação (R^2) superiores a 0,90, oferecendo uma alternativa precisa e de menor custo para o monitoramento da qualidade do solo.

Em síntese, esta tese fornece um aporte científico e metodológico estratégico para a restauração de Latossolos tropicais degradados. Conclui-se que: (i) a combinação de *Mabea fistulifera* com adubação orgânica (20 Mg ha^{-1}) é uma prática de manejo altamente recomendável; e (ii) a integração da ACP com redes neurais (RNP e ANN) constitui uma abordagem poderosa para sintetizar dados complexos e gerar modelos preditivos robustos. Estas ferramentas permitem uma avaliação integrada da recuperação e oferecem suporte decisório para políticas públicas e programas de restauração em larga escala, contribuindo diretamente para os Objetivos do Desenvolvimento Sustentável, em particular a Ação Climática (ODS 13) e a Vida Terrestre (ODS 15). Para futuros avanços, recomenda-se a validação dos modelos em diferentes condições pedoclimáticas e a inclusão de variáveis biológicas e topográficas para aprimorar a predição de parâmetros hidrodinâmicos, como a infiltração.

APÊNDICE A - PRODUÇÃO ACADÊMICA DESENVOLVIDA DURANTE O DOUTORADO

Período do doutorado: agosto de 2022 a janeiro de 2026

ARTIGOS CIENTÍFICOS PUBLICADOS EM PERIÓDICOS 2025

Publicados:

BONINI, C. S. B.; DELFIM, J. J.; LOURENCETTI, J.; NANTES, L. S.; SANTOS, M. A.; SOUZA, J. A. L. Residual effects of cover crops and soil amendments on carbon content and stock in a degraded soil. **Engenharia Agrícola (Jaboticabal)**, v. 45, p. xx, 2025.

NANTES, L. S.; BONINI, C. S. B.; SANTOS, M. A.; DA SILVA ALVES, L.; LOURENCETTI, J.; BONINI NETO, A.; SOUZA, J. A. L.; LUNARDELLI, G. A. S.; SILVA, M. B.; ALVARES, D. V.; PAZ-GONZÁLEZ, A. Exposed subsoil under recovery for 30 years: influence of management strategies on soil physical attributes and organic matter. **Soil Systems**, v. 9, p. 17, 2025.

Submetidos:

— APPLICATION OF PRINCIPAL COMPONENT ANALYSIS AND PROBABILISTIC NEURAL NETWORKS IN FERRALSOLS RECOVERY EVALUATION THROUGH PLANTING OF *Mabea fistulifera* AND *Eucalyptus urograndis*

— ARTIFICIAL NEURAL NETWORKS FOR PREDICTING SOIL QUALITY INDICATORS BASED ON AGGREGATE STRUCTURE.

CAPÍTULOS DE LIVROS PUBLICADOS

SANTOS, M. A.; BONINI, C. S. B.; SOUZA, J. A. L.; BARRETTO, V. C. M.; HEINRICHS, R. Restauração da qualidade do solo utilizando sistema agroflorestal. *In:* MEMBRIVE, C. M. B.; BERNARDES, E. M.; ROSAS, F. S.; FONSECA, R. (Org.).

IMAST 2023: IV International Meeting of Agrarian Science and Technology. p. 106–123, 2023.

BOGAS, C. L. P.; BARRETTO, V. C. M.; CRUZ, V. H.; BONINI, C. S. B.; ARAUJO, M. S.; FRIAS, Y. A.; SANTOS, M. A.; NANTES, L. S.; ANJOS, L. V. S.; SOUZA, J. A. L. Eficiência de uso do boro no desenvolvimento de mogno-africano em solo arenoso. *In:* **Ciência do Solo: o avanço da ciência no Brasil.** p. 23–34, 2023.

BONINI, C. S. B.; SANTOS, M. A.; MATOS, A. M. S.; HEINRICHS, R.; BONINI NETO, A.; COSTA, N. R. Qualidade física do solo em sistemas conservacionistas. *In*: MEMBRIVE, C. M. B. *et al*. (Org.). **IMAST 2021 – International Meeting of Agrarian Science and Technology.** p. 122–137, 2021.

BONINI, C. S. B.; BONINI NETO, A.; COSTA, N. R.; OLIVERIO, G. L.; SANTOS, M. A.; SOUZA, C. T. Qualidade física de solos tropicais. *In*: POLYCARPO, G. V. V. *et al*. (Org.). **IMAST 2018 – Atualidades nas Ciências Agrárias: mudanças, tendências e impactos**. p. 102–124, 2018.

PARTICIPAÇÃO EM CONGRESSOS

Congresso Latino-Americano de Ciência do Solo (CLACS 2025)
Santa Cruz de la Sierra, Bolívia – 24 a 27 de junho de 2025.
Participação com **apresentação de trabalho na modalidade pôster**.

Congresso Brasileiro de Agroecologia (CBA 2025)
Juazeiro, Bahia, Brasil – 15 a 18 de outubro de 2025.
Participação com **apresentação oral de trabalho científico**.

COORIENTAÇÕES EM PROJETOS DE INICIAÇÃO CIENTÍFICA

2022

- **Miriã Ferreira da Silva.** *Infiltração de água no solo em sistema agroflorestal com base agroecológica*. Iniciação Científica (Ensino Médio) – Orlando Guirado Braga, Pró-Reitoria de Pesquisa. **Coorientadora**.

- **Dayara Vivian Alvares.** *Compactação de um Argissolo Vermelho Amarelo cultivado com hortaliças em sistema de plantio direto*. Iniciação Científica (Ensino Médio) – Escola Estadual 9 de Julho, CNPq. **Coorientadora**.

- **Daiane Aparecida Costa Silva.** *Porosidade do solo em sistema agroflorestal com base agroecológica*. Iniciação Científica (Ensino Médio) – E. E. Dom Lúcio Antunes, CNPq. **Coorientadora**.

2023

- **Gabriel Carvalho Santos.** *Agregados via seca em milho cultivado em sistemas de consorciação, doses de nitrogênio e inoculação com Azospirillum brasilense*. Iniciação Científica – Etec Profª Carmelina Barbosa, CNPq. **Coorientadora**.

- **Dayara Vivian Alvares.** *Estabilidade da estrutura de um Latossolo Vermelho decapitado sob intervenção antrópica*. Iniciação Científica – Universidade Estadual Paulista (UNESP). **Coorientadora**.

- **Dienifer Valeska Alvares.** *Grau de floculação de um Latossolo Vermelho degradado em recuperação*. Iniciação Científica (Ensino Médio) – Escola Estadual 9 de Julho, CNPq. **Coorientadora**.

- **Danielly Assiria Soares Lopes.** *Porosidade de um Argissolo Vermelho distrófico após efeito residual do consórcio e inoculação de feijoeiro*. Iniciação Científica – Etec Profª Carmelina Barbosa, CNPq. **Coorientadora**.

- **Yasmin Alves da Cruz.** *Estabilidade de agregados (via seca e úmida) de um Latossolo Vermelho degradado em recuperação*. Iniciação Científica – Etec Profª Carmelina Barbosa, CNPq. **Coorientadora**.

PROJETOS EM ANDAMENTO

2024

- **Yasmin Alves da Cruz.** *Granulometria de um Latossolo Vermelho degradado em recuperação*. Iniciação Científica – Etec Profª Carmelina Barbosa, CNPq. **Coorientadora**.

- **Maria Clara Aquino Silva.** *Agregação de um solo arenoso cultivado com feijão em sucessão de blends de plantas de cobertura*. Iniciação Científica (Ensino Médio) – Escola Estadual 9 de Julho, Pró-Reitoria de Pesquisa. **Coorientadora**.

- **Heloísa Silva Vilar.** *Distribuição do tamanho de poros de um solo arenoso cultivado com feijão em sucessão de blends de plantas de cobertura*. Iniciação Científica (Ensino Médio) – Escola Estadual 9 de Julho, Pró-Reitoria de Pesquisa. **Coorientadora**.

INTERCÂMBIOS ACADÊMICOS INTERNACIONAIS

Universidad de Chile – Campus Antumapu
Período: 19 de abril a 05 de junho de 2024.
Modalidade: mobilidade acadêmica internacional no âmbito da **Asociación de Universidades Grupo Montevideo (AUGM)**.
Orientador no exterior: Prof. Dr. Osvaldo Salazar.

Universitat Politècnica de València (UPV)
Período: novembro de 2024 a junho de 2025.
Modalidade: estágio doutoral internacional, com fomento **CAPES – Programa Institucional de Internacionalização (PrInt)**.
Orientador no exterior: Prof. Dr. Borja Velázquez Martí.

ATIVIDADE DOCENTE MINISTRADA NO EXTERIOR

Aula ministrada sobre projeto de pesquisa desenvolvido pela doutoranda intitulado *Qualidade dos solos sob manejo agroecológico em sistemas agroflorestais*, com base em resultados do mestrado.

Instituição: Universidad de Chile – Programa de Pós-Graduação em Agronomia.
Carga horária: 2 horas.
Ano: 2024.

Este anexo reúne a produção científica, acadêmica e formativa desenvolvida ao longo do doutorado, evidenciando a integração entre pesquisa, extensão, formação de recursos humanos e internacionalização.