



**PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
DOUTORADO EM AGRONOMIA**

RONALDO TOSHIKI OIKAWA

**APRENDIZAGEM DE MÁQUINA E ANÁLISE DE COMPONENTES PRINCIPAIS
NA AVALIAÇÃO DE QUALIDADE DO SOLO COM PASTAGENS**

Presidente Prudente - SP
2023



**PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
DOUTORADO EM AGRONOMIA**

RONALDO TOSHIKI OIKAWA

**APRENDIZAGEM DE MÁQUINA E ANÁLISE DE COMPONENTES PRINCIPAIS
NA AVALIAÇÃO DE QUALIDADE DO SOLO COM PASTAGENS**

Tese de Doutorado apresentada à Pró-Reitoria de Pesquisa e Pós-Graduação, Universidade do Oeste Paulista, como parte dos requisitos para obtenção do título de Doutor em Agronomia – Área de concentração: Produção vegetal.

Orientador: Dr. Fábio Fernando de Araújo

Coorientador: Dr. Lucas Prado Osco

Coorientadora: Dra. Ana Paula Marques Ramos

630
O39a

Oikawa, Ronaldo Toshiaki.

Aprendizagem de máquina e análise de componentes principais na avaliação de qualidade do solo com pastagens / Ronaldo Toshiaki Oikawa. – Presidente Prudente, 2023.

102 f.: il.

Tese (Doutorado em Agronomia) - Universidade do Oeste Paulista – Unoeste, Presidente Prudente, SP, 2023.

Bibliografia.

Orientador: Fábio Fernando de Araújo

1. Aprendizado de máquina. 2. Enzimas microbiológicas. 3. Modelo estatístico. 4. Física do solo. 5. Química do solo. I. Título.

ATESTADO

ATESTAMOS, para os devidos fins, que o(a) Sr(a) **RONALDO TOSHIKI OIKAWA**, Aluno(a) Regular do Programa de Pós-Graduação em "AGRONOMIA", nível: **DOUTORADO**, da Universidade do Oeste Paulista – UNOESTE, em Presidente Prudente/SP, submeteu-se ao Exame da Tese "APRENDIZAGEM DE MÁQUINA E ANÁLISE DE COMPONENTES PRINCIPAIS NA AVALIAÇÃO DE QUALIDADE DO SOLO COM PASTAGENS" e foi considerado(a) **Aprovado(a)** para a obtenção do título de "Doutor(a) em Agronomia", Área de Concentração: "Produção Vegetal", no dia 26 de Outubro de 2023. Fizeram parte da banca examinadora:

FABIO FERNANDO DE ARAÚJO (Orientador)

UNOESTE – Universidade do Oeste Paulista / Presidente Prudente (SP)

RITA DE CÁSSIA LIMA MAZZUCHELLI

UNOESTE – Universidade do Oeste Paulista / Presidente Prudente (SP)

CARLOS SÉRGIO TIRITAN

UNOESTE – Universidade do Oeste Paulista / Presidente Prudente (SP)

GELCI CARLOS LUPATINI

UNESP – Universidade Estadual Paulista Júlio de Mesquita Filho – Dracena (SP)

ALMIR OLIVETTE ARTERO

UNESP – Universidade Estadual Paulista Júlio de Mesquita Filho – Presidente Prudente (SP)

ATESTAMOS, ainda, que a expedição do diploma está condicionada ao cumprimento das exigências do Regulamento Interno do Programa.

Presidente Prudente, 26 de outubro de 2023.

Prof. Dr. Fábio Rafael Echer
Coordenador do Programa

DEDICATÓRIA

Dedico esta tese à presença divina que, como a luz que permeia todas as coisas, iluminou meu caminho ao longo desta jornada de descobrimento e aprendizado. Agradeço a Deus por ser a fonte inesgotável de inspiração, fortaleza e sabedoria que me guiou nos momentos mais desafiadores e nas alegrias mais singelas.

À minha amada família, cujo amor incondicional é o alicerce que sustenta meus sonhos, dedico este trabalho. Cada conquista é um reflexo da força que encontro em seus abraços e no apoio constante que me impulsiona a ir além. Agradeço por serem minha âncora nos mares tempestuosos da vida acadêmica.

Aos amigos, que ao longo dessa jornada se tornaram minha segunda família, agradeço pela camaradagem, pelo estímulo mútuo e pela partilha de risos e desafios. Cada passo dado foi mais leve por causa da amizade que nos une.

À ciência da aprendizagem de máquina, que se revelou como uma ferramenta poderosa na compreensão de padrões complexos, expresso minha admiração. Este trabalho é uma homenagem à fusão entre a tecnologia e a mente humana, uma busca incessante pelo conhecimento que transcende os limites convencionais.

À Terra, aos solos que são a base da vida, as análises físicas, químicas e biológicas empreendidas nesta pesquisa, que contribuiu para uma compreensão mais profunda e holística da interação entre o solo e seus habitantes. O Índice de Qualidade do Solo, que sugere a qualidade do solo. E que as descobertas aqui apresentadas possam influenciar positivamente a maneira como percebemos, conservamos e valorizamos esse recurso precioso que sustenta toda forma de vida.

Que esta tese seja mais uma contribuição ao conhecimento científico, mas um testemunho de gratidão e reverência àquilo que é divino, familiar, amigável, tecnológico e, acima de tudo, essencial para a existência humana no planeta.

AGRADECIMENTOS

A conclusão desta jornada de doutorado é marcada por uma profunda gratidão que transcende as páginas desta tese. Expresso meus agradecimentos a todos aqueles cujo apoio e orientação foram fundamentais para a realização deste trabalho.

Primeiramente, agradeço a Deus por ser a fonte de sabedoria que iluminou meu caminho ao longo desta jornada acadêmica. Sua graça e orientação foram meu alicerce, dando-me força nos desafios e celebrando comigo as singelas conquistas.

À minha família, expresso meu amor e gratidão. Vocês são minha âncora, inspiração e fonte constante de apoio. Cada conquista alcançada é compartilhada com vocês, e este título é tão de vocês quanto meu.

Aos meus amigos e colegas de jornada acadêmica, agradeço pela camaradagem, pelas trocas intelectuais e pelo apoio mútuo. As discussões e desafios partilhados tornaram esta jornada mais rica e significativa.

Ao meu orientador Prof. Dr. Fábio Fernando Araújo, expresso minha profunda gratidão. Sua orientação sábia, paciência e incentivo foram cruciais para o desenvolvimento deste trabalho. Suas contribuições foram além da pesquisa, pois desde o início acreditou e me proporcionou a chance de participar do programa de pós-graduação em agronomia, pois não é minha área de formação.

Aos coorientadores Prof. Dr. Lucas Prado Osco e Prof. Dr^a. Ana Paula Marques Ramos, por ajudar muito com todo o conhecimento e paciência.

A parceria do Prof. Dr. Juliano Carlos Calonego e da Universidade Estadual de São Paulo – Unesp de Botucatu, que me recebeu de braços abertos e auxiliou em muitas análises físicas do solo.

As (os) companheiras (os) de laboratório, Dr^a. Lorryne guimarães bavaresco, Dr^a. Caroline Honorato Rocha, Me. Vilma Polidorio, Me. Ivan Komuro (doutorando), Me. Renan Barros.

Ao Diego, auxiliar de laboratório que nos dias de hoje está fora do Brasil com sua família, por acreditar no meu potencial e ter a paciência de ensinar química para um aluno de ciências da computação.

Aos professores e colegas de departamento, agradeço pela rica experiência acadêmica, aulas, debates e contribuições fundamentais para a minha formação.

Por fim, dedico este trabalho à busca constante por conhecimento e ao desejo de contribuir para o avanço da ciência e da sociedade. Que este estudo seja uma pequena luz em meio à vastidão do conhecimento, inspirando outros a continuarem explorando e questionando.

Obrigado a todos que, de alguma forma, contribuíram para este trabalho da minha jornada acadêmica. Este é um marco que celebramos juntos.

"Quando pedi dinheiro, Deus me presenteou com trabalho, forjando em mim a habilidade de conquistar cada centavo. Solicitei força, e Ele, com sabedoria divina, me agraciou com o dom do esforço, revelando que a verdadeira fortaleza reside na determinação. Ao clamar por paciência, Deus, em sua sabedoria suprema, me confiou desafios, compreendendo que é na superação dos problemas que cultivamos a verdadeira paciência. Roguei inteligência à Deus, e Ele me presenteou com a oportunidade de aprender, de adquirir conhecimento e sabedoria. Ele guia por caminhos de descoberta, proporcionando a clareza de pensamento para compreender o mundo ao redor. A inteligência que Ele concede não é apenas um dom, mas sim a capacidade de explorar, questionar e crescer intelectualmente, enriquecendo não apenas a mente, mas também a alma."

- Autor desconhecido

"A vida é feita de escolhas. Quando você dá um passo à frente, inevitavelmente alguma coisa fica para trás. Escolher é renunciar. Viver é, a cada momento, abrir mão de uma opção em favor de outra." - Augusto Cury

RESUMO

O solo pode ser analisado sob a ótica de várias perspectivas, sendo elas enzimas do solo, β -glucosidase, celulase, xilanase e amilase são importantes para a decomposição de matéria orgânica, respiração do solo, que envolvem os microrganismos que liberam CO₂ durante a decomposição, sendo processo crucial para o ciclo do carbono. Para abordar essa área em um modelo matemático, pode-se considerar medir as taxas de atividade enzimática e respiração do solo e correlacioná-las com variáveis ambientais e de manejo, outro elemento é o quociente microbiano (qMic) e quociente metabólico (qCO₂). Que avalia a eficiência dos microrganismos em utilizar carbono orgânico disponível (qMic). Já o qCO₂, relaciona a respiração do solo com a biomassa microbiana, sendo um indicador de eficiência na utilização de carbono orgânico. Este projeto tem como objetivo criar um modelo matemático pode envolver a correlação destes quocientes com variáveis ambientais para entender os fatores que influenciam a atividade microbiana, atributos químicos e físicos do solo. No capítulo 2, é trabalhado o modelo principal componente análise (PCA) que auxilia na redução e complexidade do modelo e gerando métricas apropriadas como F1-score, facilitando a interpretação e o processamento. Já no capítulo 1, para alcançar a máxima generalização e acurácia dos resultados são trabalhados o aprendizado de máquina não supervisionado, Mapas Auto-organizáveis (SOM), com a estratégia de reduzir a dimensionalidade da matriz de valores, a Floresta Aleatória (RF), para minimizar os atributos, e na última cadeia de avaliação o ranqueamento desses atributos, para melhor disponibilizar e formar novas visões para a interpretação. Todos os dados, passam por um pré-processamento submetidos em cada modelo buscando a máxima generalização, para que consiga atender o maior número de tipos de solo, e a máxima acurácia. Em todos os modelos utilizados há importância de cada característica em cada modelo para identificar quais atributos devem ser mantidos ou descartados. Ao combinar essas abordagens, isso ajuda a otimizar os recursos disponíveis e a simplificar o processo de análise e tomada de decisão.

Palavras-chave: enzimas microbiológicas, aprendizado de máquina, modelo estatístico, física do solo, química do solo.

ABSTRACT

Soil can be analyzed from several perspectives, such as soil enzymes, β -glucosidase, cellulase, xylanase and amylase, which are important for the decomposition of organic matter, soil respiration, which involve microorganisms that release CO₂ during decomposition, being a crucial process for the carbon cycle. To address this area in a mathematical model, one can consider measuring the rates of enzyme activity and soil respiration and correlating them with environmental and management variables, another element is the microbial quotient (qMic) and metabolic quotient (qCO₂). Which evaluates the efficiency of microorganisms in using available organic carbon (qMic). qCO₂, on the other hand, relates soil respiration to microbial biomass, being an indicator of efficiency in the use of organic carbon. This project aims to create a mathematical model that can involve the correlation of these quotients with environmental variables to understand the factors that influence microbial activity, chemical and physical attributes of the soil. In chapter 1, the main component analysis (PCA) model is worked on, which helps to reduce the complexity of the model and generate appropriate metrics such as F1-score, facilitating interpretation and processing. In chapter 2, to achieve maximum generalization and accuracy of results, unsupervised machine learning, Self-Organizing Maps (SOM), is used with the strategy of reducing the dimensionality of the matrix of values, the Random Forest (RF), to minimize the attributes, and in the last evaluation chain the ranking of these attributes, to better make available and form new views for interpretation. All data undergoes pre-processing submitted in each model seeking maximum generalization, so that it can cover the largest number of soil types, and maximum accuracy. In all models used, each characteristic is important in each model to identify which attributes should be maintained or discarded. By combining these approaches, it helps to optimize available resources and simplify the analysis and decision-making process.

Keywords: microbiological enzymes, machine learning, statistical model, soil physics, soil chemistry.

LISTA DE TABELAS

Tabela 1 - Resultados da análise de componentes principais (PCA) em 2019	91
Tabela 2 - Resultados da análise de componentes principais (PCA) em 2021	92
Tabela 3 - Coeficientes de correlação de Spearman (r^2) dos diferentes índices de qualidade do solo (IQS) versus desenvolvimento da massa seca da raiz(MSR) e parte aérea (MSPA) da alface.	95

LISTA DE FIGURAS

Figura 1 - Localização das áreas selecionadas no município de Caiuá, SP, pontos amarelo pastagem sem manejo, pontos verdes mata de preservação permanente, e em vermelho área de ILP.....	49
Figura 2 - O fluxograma demonstra os métodos implementados nesta pesquisa.	53
Figura 3 - Figura do box plot de todos os modelos de aprendizagem de máquina utilizados no projeto, demonstrando os resultados dos algoritmos com validação cruzada e elencando seus desempenhos.	57
Figura 4 - Matriz de confusão do algoritmo Random Forest (RF) na etapa de treinamento com os dados de 2019.	58
Figura 5 - Importância dos 26 atributos de solo avaliados pelo Random Forest.	59
Figura 6 - Distribuição por box plots das classes estabelecidas pelo algoritmo RF com base no desenvolvimento da planta indicadora. Em todos os atributos da alfaca Emergência, MFPA, MFR obtiveram cluster excelente, muito bom e bom.....	60
Figura 7 - Correlação das variáveis de crescimento das plantas utilizadas no modelo RF.....	61
Figura 8 - Matriz de confusão para o algoritmo RF na etapa de validação com os dados de 2021.....	62
Figura 9 - Localização das áreas selecionadas no município de Caiuá, SP, sendo em amarelo pastagem com manejo e sem manejo, em verde área de APP e vermelho ILP.	77

LISTA DE SIGLAS

APP	- Área de Preservação Permanente
Ca	- Cálcio
CaCl ²	- Cloreto de Cálcio
CBM	- Carbono da Biomassa Microbiana
CTC	- Capacidade de Troca Catiônica
CO ²	- Dióxido de carbono
COM	- Carbono de Partículas Minerais
COP	- Carbono Particulado Orgânico
COT	- Carbono Total Orgânico
FDA	- Hidrólise de Diacetato de Fluoresceína
GB	- Modelo de Aprendizado de Máquina Gradient Boosting
GERM	- Germinação
H ⁺ AI	- Acidez Potencial
ILP	- Integração Lavoura Pecuária
K	- Potássio
Knn	- K-vizinho mais próximo
LDA	- Análise Discriminante Linear
LR	- Regressão Logística
MFPA	- Matéria Fresca da Parte Aérea
MFR	- Matéria Fresca da Raiz
MLP	- Rede Neural Multi Layer Perceptron
Mg	- Magnésio
MO	- Matéria Orgânica
MSPA	- Matéria Seca da Parte Aérea
MSR	- Matéria Seca da Raiz
NBM	- Nitrogênio da Biomassa Microbiana
NDVI	- Índice de Vegetação com Diferença Normalizada
Na	- Sódio
NB	- Algoritmo de Classificação Navie Bayes
P	- Fósforo

PAST CM	- Pastagem Contínua Com Manejo
PAST SM	- Pastagem Contínua Sem Manejo
pH	- Acidez do Solo
qCO^2	- Quociente metabólico do solo
$qMic$	- Quociente microbiano
RF	- Floresta Randômica
RT	- Árvore de Regressão
SB	- Soma de Bases
SOM	- Mapa auto organizável
SVM	- Máquina de Suporte de Vetor
V%	- Porcentagem de Saturação por Bases

SUMÁRIO

1 INTRODUÇÃO GERAL	16
Importância das pastagens e pecuária no Brasil.....	18
Atributos Químicos do Solo.....	19
Atributos Físicos do Solo	22
Atributos Microbiológicos do Solo	24
Análise de Aprendizado de Máquina.....	28
Mapas Auto-organizáveis, Aprendizado Não Supervisionado	29
Análise Discriminante Linear.....	30
Árvore de Decisão.....	31
Floresta Aleatória (<i>Random Forest</i>).....	32
Máquina de Vetores de Suporte.....	32
Avaliação de Atributos, Classificação e Ranqueamento	33
REFERÊNCIAS	35
2 CAPÍTULO 1	45
INTRODUÇÃO	47
Material e Método.....	48
Área de Estudo.....	48
Avaliação da Planta Indicadora	51
Processamento dos Dados	52
Resultados	55
Discussão.....	62
Conclusões.....	65
REFERÊNCIAS	66
3 CAPÍTULO 2	71
INTRODUÇÃO	73
Material e Método.....	75
Área de Estudo.....	75
Análises de atributos químicos, físicos e biológicos do solo	77
Atributos químicos.....	78
Atributos microbiológicos.....	78
Atributos físicos	78
Análise com planta indicadora.....	79
Determinação do índice de qualidade do solo (IQS).....	80
Seleção de indicadores: Conjunto mínimo de dados (CMD).....	80
Transformação dos indicadores do CMD	81
Integração dos indicadores do CMD no SQI.	82
Validação do IQS	83

Análise estatística	83
Resultados	83
Discussão.....	95
Conclusões.....	97
REFERÊNCIAS	98

1 INTRODUÇÃO GERAL

Para a sustentabilidade agrícola e segurança alimentar o solo é uma ferramenta importante para garantir esses requisitos segundo a FAO (2011), pois em seu relatório sobre o estado da fertilidade do solo no mundo, a organização destaca a importância da análise do solo para a tomada de decisões relacionadas à gestão do solo, água, à seleção de culturas e variedades, e o desenvolvimento de estratégias de manejo integrado para mitigar pragas e doenças (FAO 2015a, b, c; Santos; Ferreira, 2022). Essas e outras práticas de análise do solo são fundamentais para a preservação do meio ambiente, já que permite avaliar o impacto das atividades agrícolas na qualidade do solo e na biodiversidade. De acordo com o estudo publicado por Ibrahim *et al.* (2022), a análise do solo pode ser utilizada para avaliar a eficácia de práticas de conservação, como a adição de matéria orgânica e a rotação de culturas, na redução da erosão e na melhoria da qualidade do solo.

As análises químicas, físicas e biológicas do solo são indispensáveis para o sucesso da agricultura e para a preservação do balanceamento ambiental, por esse motivo é importante que os agricultores, técnicos e pesquisadores tenham acesso a informações atualizadas sobre a qualidade do solo, para garantir a sustentabilidade da produção e a conservação dos recursos do solo.

A análise química do solo indica diversos atributos importantes para a avaliação da fertilidade do solo e para a recomendação de práticas de manejo adequadas. Dentre os principais atributos, é possível citar a porcentagem de saturação por bases (V%), matéria orgânica do solo (MOS), os teores de alumínio trocável (Al), cálcio (Ca), magnésio (Mg), potássio (K) e fósforo (P), o pH e a capacidade de troca catiônica (CTC) (Singh *et al.*, 2022).

A análise física do solo, aborda características como textura, porosidade, densidade e permeabilidade, que influenciam diretamente na capacidade do solo de reter água e nutrientes, além de afetar a disponibilidade de oxigênio para as raízes das plantas.

De acordo com o estudo dos autores Meena, Lal e Yadav (2020), a análise física do solo é importante para a avaliação da qualidade do solo em áreas degradadas e para a tomada de decisões em relação ao manejo do solo. Os

diferentes compartimentos de carbono no solo também podem ser avaliados através da análise física do solo, fornecendo informações importantes sobre a qualidade do solo e sua capacidade de armazenamento de carbono (Cao *et al.*, 2021). O carbono pode estar presente na fração mineral do solo, como também na fração orgânica do solo, desta forma o carbono total é a soma desses dois componentes (Janzen *et al.*, 1992; Baldock; Skjemstad, 2000; Schöning; Kögel-Knabner, 2006).

Microrganismos incluem todos os organismos unicelulares e portanto, são extremamente diversos, dessa forma a análise biológica do solo indica diversos atributos importantes para a avaliação da qualidade do solo e da atividade microbiana presente nele. Dentre os principais parâmetros, pode-se citar a respiração do solo, o teor de nitrogênio da biomassa microbiana, atividade de enzimas como a arilsulfatase, β -glucosidase fosfatase, quociente microbiano (q_{Mic}) e o quociente metabólico (q_{CO_2}) (Ashraf; Waqas; Rahman, 2022).

Os microrganismos do solo desempenham papel fundamental no ciclo biogeoquímico do fósforo (P), nitrogênio (N), carbono (C), enxofre (S), potássio (K), entre outros, que são indispensáveis para as plantas. Mediante o fluxo de P pela biomassa microbiana, a solubilização do P inorgânico, a mineralização do P orgânico, há também o carbono decomposto da matéria orgânica no solo como contribuição (Moreira; Siqueira, 2002).

O enxofre envolve a transformação de compostos de enxofre orgânicos e inorgânicos, incluindo a redução de sulfatos e sulfetos, que por processos complexos converte para outras formas de enxofre, já o nitrogênio é absorvido e fixado da atmosfera, nitrificação, desnitrificação e a amonificação às plantas (Moreira; Siqueira, 2002).

Os constituintes da fração orgânica do solo, como a biomassa microbiana, embora quantitativamente pouco representada, é de grande significância, visto que os produtos do seu metabolismo constituem, uma das principais fontes do nitrogênio mineral e fósforo para as plantas (Colodro *et al.*, 2007). Neste sentido a hipótese desse estudo, busca empregar análise de componentes principais e aprendizagem de máquina utilizando banco de dados compostos por análises de atributos químicos, físicos, biológicos do solo e avaliação da planta bioindicadora, para estimar a redução das análises laboratoriais do solo. As áreas de pastagens com diferentes manejos possibilitam à pesquisa avaliar resultados que auxiliem em indicadores de qualidade do solo. O mesmo se justifica pela necessidade de novas

tecnologias e modelos de aprendizagem de máquina para áreas de pastagens em solos arenosos com limitações edafoclimáticas.

A pesquisa tem cunho de oferecer aos especialistas agrônomos e produtores, visões já consagradas e novas perspectivas e possibilidades nas tomadas de decisão em relação ao solo, com o objetivo de avaliar, comparar e sugerir novas formas de classificar os indicadores de qualidade do solo, demonstrando o que quais componentes devem ser levados em consideração no momento.

Importância das pastagens e pecuária no Brasil

A pecuária é uma das atividades econômicas mais importantes do Brasil, sendo responsável por cerca de 8% do PIB agropecuário do país. Dentre as atividades relacionadas à pecuária, a produção de forragem é fundamental para garantir a alimentação animal e, conseqüentemente, a produtividade do setor (Pereira-Flores *et al.*, 2023; Brasil, 2021; Projeto Mapbiomas, 2021).

Existem diversas espécies forrageiras utilizadas no Brasil, cada uma com suas particularidades em relação à adaptação das condições climáticas, tipos de solo, produção de massa, valor nutritivo e palatabilidade do gado (Pereira-Flores *et al.*, 2023). Entre as espécies mais utilizadas no Brasil estão o capim-braquiária (*Brachiaria brizantha*), o capim-tanzânia (*Panicum maximum*), o capim-mombaça (*Panicum maximum*) e a aveia preta (*Avena strigosa*). Cada uma das espécies citadas, possui características específicas que as tornam adequadas para determinadas regiões e condições de produção (Bourscheidt, 2021; Pereira-Flores *et al.*, 2023). O capim-braquiária é uma espécie forrageira muito utilizada na pecuária brasileira e destaca-se pela alta adaptabilidade e produtividade, além de ser resistente à seca e a doenças. Outra característica é a boa relação entre folhas e colmos, o que favorece sua digestibilidade pelo gado, além de ser rica em proteína bruta, energia e minerais, tais características favorece a rentabilidade da produção animal em sistemas de pastoreio extensivo (Bourscheidt, 2021; Pereira-Flores *et al.*, 2023).

Com essa espécie adequada ao clima do Brasil, as análises laboratoriais são indispensáveis para analisar o estado atual do solo, dessa forma, há um grande

esforço na coleta de dados e gastos financeiros para realizar muitas análises que podem não ser relevantes se analisados separadamente, ou seja, há uma gama de análises laboratoriais físicas, químicas e microbiológicas que em conjunto, podem demonstrar novas perspectivas dos indicadores do solo. Assim, muitos autores como Pereira-flores *et al.*, (2023), Santos e Ferreira, (2022) e Ashraf, Waqas e Rahman, (2022) retratam a importância e contribuição para os atributos indicadores de qualidade do solo.

Atributos Químicos do Solo

A análise química do solo utilizada na agricultura para avaliar a fertilidade e a capacidade produtiva do solo, realizada em laboratório, permite determinar as concentrações de nutrientes essenciais às plantas, como fósforo, potássio, cálcio, magnésio, além de outros elementos importantes como alumínio, ferro, manganês, pH do solo, capacidade de troca catiônica (CTC), e o teor de matéria orgânica (Brasil *et al.*, 2020; Ghosh *et al.*, 2020). Os resultados da análise química do solo são utilizados para determinar as doses de nutrientes que devem ser aplicadas no solo através da correção e adubação, visando suprir possíveis deficiências ou excessos. Essa prática auxilia a maximizar a absorção de nutrientes pelas plantas e, conseqüentemente, aumentar a produtividade agrícola (Brasil *et al.*, 2020; Meena, *et al.*, 2020).

O atributo pH influencia a disponibilidade de nutrientes às plantas e a atividade dos micro-organismos presentes no solo, sendo o atributo que mensura a acidez ou alcalinidade do solo, definido como o logaritmo negativo da concentração de íons hidrogênio (H^+) presentes no solo. A escala de pH varia de 0 a 14, sendo que o pH de 7 é considerado neutro, valores abaixo de 7 indicam acidez e valores acima de 7 indicam alcalinidade, porém de acordo com a literatura, valores entre 4 e 5, são ideais para o cultivo (Teixeira *et al.*, 2017; Angeletti, *et al.*, 2021; Kim *et al.*, 2021). A acidez do solo pode ser causada por diferentes fatores, incluindo a presença de íons de hidrogênio (H^+) e alumínio (Al^{3+}) na solução do solo, bem como pela presença de materiais orgânicos ácidos em decomposição. Solos ácidos geralmente apresentam baixa disponibilidade de nutrientes essenciais às plantas, como cálcio (Ca), magnésio (Mg) e fósforo (P), além de terem maior solubilidade de

metais tóxicos, como alumínio (Al) e manganês (Mn) em excesso, que podem afetar negativamente o desenvolvimento das plantas (Teixeira *et al.*, 2017).

A análise do pH do solo é fundamental para determinar a correção da acidez do solo, visando aumentar a disponibilidade de nutrientes às plantas e melhorar a produtividade agrícola, utilizando a calagem, que consiste na aplicação de calcário para neutralizar a acidez do solo, esta quantidade de calcário a ser aplicada depende do teor de acidez do solo e da cultura a ser cultivada (Teixeira *et al.*, 2017; Liu *et al.*, 2020).

O teor de fósforo solúvel no solo tem grande importância pois contribui no crescimento e desenvolvimento das plantas, uma vez que desempenha várias funções importantes em processos bioquímicos, como a transferência de energia celular, a síntese de ácidos nucleicos e a regulação do metabolismo vegetal (Cakmak; Mclaughlin; White, 2017; Calderano *et al.*, 2011). A disponibilidade do fósforo no solo é influenciada por fatores como o pH, a textura, a mineralogia e a quantidade de matéria orgânica, que afetam a capacidade de retenção do fósforo no solo. Dessa forma, é interessante considerar esses fatores ao interpretar os resultados da análise química do fósforo no solo (Calderano *et al.*, 2011; Liu *et al.*, 2020; Samson *et al.*, 2020; Bindraban; Dimkpa; Pandey, 2020).

O H^+ e Al^{3+} são cátions que podem ser trocados por outros cátions na superfície das partículas de solo, como cálcio (Ca^{2+}) e magnésio (Mg^{2+}). Quando o solo está muito ácido, o H^+ e o Al^{3+} tendem a predominar, o que pode afetar negativamente o crescimento das plantas. A saturação por alumínio, que é a porcentagem de carga da capacidade de troca catiônica (CTC) ocupada por Al^{3+} , indica a toxicidade do alumínio às plantas (Sparks, 2003; Brady; Weil, 2008; Kölln; Silva; Paludetto, 2023).

O potássio (K) é essencial para o desenvolvimento das plantas, sendo responsável por diversos processos fisiológicos, tais como a regulação do movimento de água e nutrientes nas células vegetais, o aumento da resistência a estresses ambientais e o favorecimento da fotossíntese (Mpelasoka *et al.*, 2003; Franco; Delgado, 2022). A deficiência de K no solo pode afetar negativamente a produtividade e qualidade das culturas. A análise química do K no solo é importante para avaliar a disponibilidade desse nutriente, para orientar a adubação adequada, de forma a maximizar a produtividade das culturas (Huang *et al.*, 2022; Kölln; Silva;

Paludetto, 2023; Biliás *et al.*, 2023). A elevada concentração de K no solo pode levar à toxicidade e reduzir a disponibilidade de outros nutrientes, como o cálcio (Ca) e o magnésio (Mg), além de favorecer a absorção excessiva de sódio (Na) pelas plantas (Franco; Delgado, 2022; Biliás *et al.*, 2023).

O cálcio (Ca) é um dos nutrientes que está presente em diversos processos fisiológicos, como a formação de paredes celulares e a regulação do crescimento e desenvolvimento vegetal, juntamente com o Magnésio (Mg) e K. O ciclo do cálcio no solo é complexo e envolve diversas formas de interação com outros elementos químicos e organismos presentes no ambiente edáfico (White; Broadley, 2003; Mühlbachová *et al.*, 2020; Babasaheb *et al.*, 2022). O Ca pode estar presente em diferentes formas químicas no solo, como cátions solúveis, compostos orgânicos, minerais primários e secundários, entre outros, a disponibilidade de cálcio às plantas depende de diversos fatores, como pH do solo, teor de matéria orgânica, atividade microbiana e textura do solo (White; Broadley, 2003; Thor, 2019; Rayne; Aula, 2020).

A matéria orgânica (MO) é um componente importante no solo pois controla a dinâmica no solo, como a decomposição da matéria, a adsorção de minerais e a estabilização química, sendo fundamental na dinâmica da matéria orgânica no solo, que decompõe a matéria orgânica em compostos mais simples para serem utilizados pelas plantas (Samson *et al.*, 2020; Angeletti, 2021). Esse processo é importante para a ciclagem de nutrientes no solo e também influencia a quantidade de carbono orgânico no solo. Já a absorção de minerais do solo ocorre quando a matéria orgânica se liga quimicamente a superfícies de minerais do solo, como argilas e óxidos de ferro e alumínio (Chen *et al.*, 2003; Angeletti 2021; Kim *et al.*, 2021). Quando a matéria orgânica se liga quimicamente a compostos orgânicos e inorgânicos no solo, ocorre a estabilização do solo, formando estruturas moleculares mais complexas e resistentes à decomposição microbiana. Esses compostos incluem polímeros orgânicos, complexos metálicos e hidróxidos de ferro e alumínio.

O efeito resultante dessa adsorção e absorção, é a estabilização química, sendo fundamental na formação de estoques de carbono orgânico no solo a longo prazo, já que essas estruturas químicas são menos suscetíveis à decomposição microbiana (Chen *et al.*, 2003; Audette *et al.*, 2021; Kim *et al.*, 2021).

A soma de bases trocáveis (SB) de um solo, argila ou húmus representa a soma dos teores de cátions permutáveis, exceto H^+ e Al^3 + ($SB = Ca^{2+} + Mg^{2+} + K^+$), sendo um excelente indicativo das condições gerais de fertilidade do solo, assim a soma de bases (SB) indica a capacidade de troca catiônica (CTC) do solo, ou seja, a capacidade do solo em reter e fornecer nutrientes importantes para as plantas, como cálcio (Ca), magnésio (Mg), potássio (K) e sódio (Na) (Ronquim, 2010). Com esse fornecimento de nutrientes a SB é a soma das concentrações de cátions básicos trocáveis na fase sólida do solo. Esses cátions são importantes para o crescimento e desenvolvimento das plantas, e sua disponibilidade no solo pode afetar diretamente a produtividade das culturas (Ronquim, 2010).

A capacidade de troca catiônica (CTC) e a saturação de base (V%) são duas medidas da análise química do solo, que estão relacionadas à fertilidade e à qualidade do solo em reter cátions, e é determinada pela soma das concentrações de cátions trocáveis na fase sólida do solo, como cálcio (Ca), magnésio (Mg), potássio (K) e sódio (Na). Já a V% é a proporção, em porcentagem da CTC ocupada pelos cátions básicos (Ca, Mg, K e Na) (Franzluebbers, 2010). De acordo com Lindsay e Norvell (2020), nos solos ácidos, a CTC é geralmente baixa, o que indica uma menor capacidade do solo em reter nutrientes e uma maior necessidade de correção do pH. Nesse caso, a V% também pode ser baixa, indicando uma menor disponibilidade de nutrientes básicos. A correção do pH pode aumentar a CTC e a V%, melhorando a disponibilidade de nutrientes do solo (Franzluebbers, 2010).

Atributos Físicos do Solo

A análise física do solo é utilizada na caracterização e classificação dos solos, compreende a determinação de propriedades físicas, como textura, estrutura, densidade, porosidade, permeabilidade, carbono, entre outras. A partir dessas informações, é possível compreender melhor as condições físicas do solo e sua capacidade para o crescimento de plantas, além de auxiliar na escolha das práticas de manejo adequadas para cada tipo de solo (Rego *et al.*, 2023).

Diversos estudos têm sido realizados para aprimorar as técnicas de análise física do solo, utilizando métodos cada vez mais precisos e eficientes, dessa forma o estudo realizado por Rego *et al.* (2023), investigou a influência da compactação do

solo na qualidade física de um latossolo vermelho, que indicou que a compactação afetou negativamente a porosidade do solo, aumentando sua densidade e reduzindo sua permeabilidade. Isso ressalta a importância de práticas de manejo que evitem a compactação do solo, como o uso de técnicas de plantio direto, plantas com raízes vigorosas e a rotação de culturas (Barros *et al.*, 2020; He *et al.*, 2021).

A análise física do solo auxilia na escolha das práticas de manejo adequadas, pois o solo pode estar com menor porosidade e maior compactação, ou baixa/alta retenção de água (Silva *et al.*, 2019; He *et al.*, 2021; Rego *et al.*, 2023). A textura do solo, que é a proporção relativa de partículas de areia, silte e argila presentes no solo, define as propriedades físicas do solo, como a capacidade de retenção de água e nutrientes, a permeabilidade e a estrutura do solo (Silva *et al.*, 2019; Kipper *et al.*, 2020).

A avaliação de silte, areia total e argila é utilizada para investigar a porosidade e textura do solo. Para isso, é realizada a separação das partículas presentes no solo de acordo com o tamanho e a massa das partículas granuladas, utilizando técnicas como a sedimentação e a centrifugação. O resultado dessa técnica são proporções relativas de cada tipo de partícula que são determinadas e utilizadas para classificar a textura e granulação do solo (Leal *et al.*, 2020).

O estudo de Kipper *et al.* (2020), avaliou a textura do solo em áreas de cultivo de soja no Brasil, que indicaram que as áreas com maior proporção de argila no solo, apresentam maior produtividade de soja, sugerindo práticas de manejo adequado nessas áreas. Já o trabalho de Leal *et al.* (2020), combinou a análise de textura do solo com a avaliação da porosidade e da permeabilidade do solo em áreas de produção de milho, sugerindo que a maior porosidade e permeabilidade apresentaram maior produtividade de milho (Rego *et al.*, 2023).

O carbono orgânico associado às partículas minerais (COM) representa um reservatório mais estável de C no solo enquanto que o carbono associado as partículas vegetais (COP) representa um reservatório de entrada de C no solo. A soma dos dois representa o carbono orgânico total (COT), refere-se à quantidade total de carbono presente na matéria orgânica do solo, incluindo restos de plantas, matéria orgânica em decomposição e outros componentes orgânicos (Stevens *et al.*, 2014; Silva *et al.*, 2019).

A análise física do solo pode indicar a quantidade e a distribuição do carbono em diferentes camadas do solo, incluindo a camada superficial (0-10 cm), a compactação do solo, que pode limitar a penetração das raízes das plantas e a aeração do solo, afetar a atividade microbiana e a decomposição da matéria orgânica. Já a erosão do solo pode remover a camada superficial do solo rica em carbono, reduzindo a fertilidade.

As práticas de manejo do solo como a adição de material orgânico ou o uso de técnicas de conservação, podem aumentar a quantidade de carbono na camada superficial do solo. Uma possível sugestão é a incorporação de resíduos vegetais que pode aumentar a biomassa microbiana e decomposição da matéria orgânica, aumentando o carbono total e médio do solo (Silva *et al.*, 2019).

Atributos Microbiológicos do Solo

A microbiologia do solo é uma área de estudos que investiga os microrganismos e suas interações com o solo e meio ambiente. Esses microrganismos são importantes para as funções do solo e para a produção agrícola, pois desempenham papéis fundamentais na ciclagem de nutrientes, na decomposição da matéria orgânica, na fixação de nitrogênio e na promoção do crescimento das plantas (Gururani *et al.*, 2013; Li *et al.*, 2023).

A análise física e química do solo são métodos trivialmente utilizado para avaliar os atributos de qualidade do solo e determinar suas características, como a textura, a capacidade de retenção de água, pH e a presença de nutrientes essenciais. No entanto, a análise microbiológica do solo é igualmente importante, pois os microrganismos presentes no solo afetam diretamente a decomposição da matéria, qualidade e a fertilidade do solo (Yadav *et al.*, 2023).

Os microrganismos do solo podem ser afetados por diversas características físicas e químicas do solo, como o teor de umidade, a disponibilidade de nutrientes, o pH e a presença de metais pesados. Por sua vez, a atividade microbiana no solo pode afetar a disponibilidade de nutrientes às plantas, como também, a formação e estabilidade dos agregados do solo e a sua capacidade de armazenar carbono (Li *et al.*, 2023; Yadav *et al.*, 2023).

O Carbono da Biomassa Microbiana (CBM) é um indicador da quantidade de carbono microbiano presente no solo, responsável pela decomposição da matéria orgânica e pela ciclagem de nutrientes, usado como uma medida da atividade biológica do solo e da disponibilidade de nutrientes para as plantas (Collins *et al.*, 2020). O CBM é influenciado por fatores como o pH do solo, umidade, temperatura e a disponibilidade de nutrientes, com isso é avaliada a qualidade do solo, bem como para monitorar a eficácia de práticas de manejo, que aumentam a adição de matéria orgânica. Da mesma forma, o uso excessivo de fertilizantes minerais ou a compactação do solo podem reduzir a biomassa microbiana e o CBM (Joergensen, 1996; Collins *et al.*, 2020; Li *et al.*, 2021).

A determinação do nitrogênio da biomassa microbiana (NBM) avalia a capacidade do solo em metabolizar e mineralizar compostos orgânicos para estimar a quantidade de N disponível às plantas. Além disso, o NBM pode ser usado como um indicador da qualidade do solo e a eficácia do manejo adotado. Outra forma de avaliação é a verificação dos resultados da adição de fertilizantes nitrogenados (Yang *et al.*, 2022; Yang *et al.*, 2023; Lupwayi *et al.*, 2023). Estudos têm demonstrado a importância do NBM em diversas situações, como na recuperação de solos degradados, na avaliação de práticas agrícolas sustentáveis e na compreensão dos efeitos de mudanças climáticas na atividade microbiana (Silva *et al.*, 2019; Liu *et al.*, 2020).

Dentro das análises enzimáticas do solo pode ser destacada a desidrogenase, que é uma enzima presente no solo que está relacionada com a atividade microbiana e pode ser usada como um indicador da atividade biológica. A atividade da desidrogenase é afetada pela quantidade de nitrogênio disponível no solo, uma vez que o nitrogênio é um nutriente essencial à atividade microbiana. Quanto maior a disponibilidade de nitrogênio, maior tende a ser a atividade da desidrogenase, que é frequentemente usada como um indicador da biomassa microbiana, uma vez que ambas estão relacionadas à atividade microbiana (George *et al.*, 2021).

Outra análise é a arilsulfatase é uma enzima que catalisa a hidrólise de ésteres de sulfato de arilas em presença de água, liberando um grupo aril e um ânion sulfato, sendo produzida por microrganismos e tem sido associada à fertilidade do solo e à ciclagem de nutrientes (Ferreira *et al.*, 2022). O estudo

publicado por Chaer *et al.* (2023), relatou que a atividade da arilsulfatase e β -glucosidase em solos agrícolas brasileiros vêm sendo maior em solos bem manejados. Outra relação da arilsulfatase está relacionada com a mineralização do enxofre orgânico no solo, que é um processo que consome carbono e gera CO₂ por meio da respiração microbiana. Além disso, a arilsulfatase também está relacionada com a disponibilidade de nitrogênio no solo, uma vez que o enxofre é um nutriente essencial aos microrganismos fixadores de nitrogênio. Muitas vezes solos degradados ou com baixa atividade biológica tendem a apresentar uma menor atividade de arilsulfatase (Chaer *et al.*, 2023).

Ainda dentro das análises do solo o diacetato de fluoresceína (FDA) é uma substância corante fluorescente que é frequentemente utilizado como um indicador da atividade biológica do solo, quando aplicado ao solo em análise, o FDA é convertido em fluoresceína por meio da ação de enzimas produzidas por microrganismos (Santos *et al.*, 2022). A fluoresceína é uma molécula que emite fluorescência, a intensidade da fluorescência é proporcional à quantidade de microrganismos ativos no solo, uma vez que esses microrganismos são os responsáveis pela conversão do FDA em fluoresceína. Assim, a técnica do FDA é utilizada como um indicador rápido (Ferreira *et al.*, 2022).

Outra análise é a fosfatase ácida, que é uma enzima que catalisa a hidrólise do grupo fosfato em compostos orgânicos e inorgânicos, liberando fosfato inorgânico (Pi) no solo, tal enzima é produzida por microrganismos do solo e por raízes de plantas, sendo importante para a ciclagem de fósforo, nutriente indispensável para o desenvolvimento e crescimento das plantas (Ashraf; Waqas; Rahman, 2022). A atividade da fosfatase no solo tem sido amplamente utilizada como um indicador da qualidade do solo, pois reflete a capacidade dos microrganismos do solo de decompor e mineralizar compostos orgânicos. Estudos têm mostrado que a atividade da fosfatase pode ser afetada por diversos fatores, como a textura do solo, o pH, o teor de matéria orgânica, a presença de metais pesados e a adição de fertilizantes em excesso (Ghosh *et al.*, 2020).

A β -glucosidase é uma enzima presente no solo e produzida por microrganismos, como bactérias, fungos e actinomicetos. Essa enzima é responsável pela hidrólise de ligações β -glicosídicas em compostos como celobiose, lactose e outras moléculas de celulose e hemicelulose, liberando glicose e outros

açúcares (Chen *et al.*, 2019; Ashraf; Waqas; Rahman, 2022; Chaer *et al.*, 2023). A atividade da β -glucosidase reflete a capacidade dos microrganismos do solo de decompor a matéria orgânica e de transformar nutrientes em formas disponíveis às plantas, além de estar relacionada com outras enzimas do solo, como a celulase, a xilanase e a amilase, que também estão envolvidas na decomposição da matéria orgânica e na ciclagem de nutrientes (Zhang *et al.*, 2021; Chaer *et al.*, 2023). Um estudo realizado por Chen *et al.* (2019) avaliou a influência da β -glucosidase na degradação da palha do arroz, que mostraram o aumento significativo da atividade da β -glucosidase e interação com a celulase, xilanase e de outras enzimas do solo.

A respiração do solo é um processo pelo qual os microrganismos do solo, como bactérias e fungos, liberam dióxido de carbono (CO_2), como produto residual durante o processo de decomposição da matéria orgânica presente no solo, esse processo é importante para a manutenção do ciclo do carbono no solo e na atmosfera (Co; Hug, 2021). A respiração do solo ocorre em duas etapas: a respiração aeróbia e a respiração anaeróbia. Na respiração aeróbia, que ocorre em condições de presença de oxigênio no solo, os microrganismos utilizam o oxigênio para oxidar a matéria orgânica e produzem dióxido de carbono como subproduto. Já a respiração anaeróbia ocorre em condições de ausência de oxigênio, ou seja, utiliza outros compostos como nitrato e sulfato, para oxidar a matéria orgânica e produzir dióxido de carbono (George *et al.*, 2021). Essa taxa de respiração do solo pode variar de acordo com fatores ambientais, como a umidade do solo, a temperatura, a disponibilidade de nutrientes e a presença de substâncias tóxicas, sendo que a taxa de respiração é maior em solos úmidos e quentes, onde há maior atividade microbiana, porém, é influenciada pela atividade da desidrogenase, da arilsulfatase da hidrólise do FDA, CBM e NBM, uma vez que todas estão relacionadas à atividade microbiana do solo (Cairo-Cairo *et al.*, 2023).

O quociente microbiano (q_{Mic}) é uma medida da eficiência dos microrganismos do solo em utilizar o carbono orgânico disponível para o crescimento e manutenção da biomassa microbiana, sendo definido como a relação entre o carbono da biomassa microbiana e o carbono orgânico do solo e ciclagem de nutrientes, podendo avaliar a qualidade do solo (Ashraf; Waqas; Rahman, 2022).

O quociente metabólico do solo (q_{CO_2}) é uma medida que relaciona a respiração do solo com a biomassa microbiana do solo, sendo a quantidade de

dióxido de carbono (CO₂) emitida pelo solo por unidade de carbono da biomassa microbiana. O $q\text{CO}_2$ é um indicador da atividade microbiana no solo, que pode ser utilizado para avaliar a eficiência dos microrganismos em utilizar o carbono orgânico disponível no solo para a respiração e produção de energia (Dynarski; Bossio; Scow, 2020). A relação do $q\text{CO}_2$ com o solo é um processo chave no ciclo global do carbono e no sequestro de carbono atmosférico pelos solos, que liberam CO₂ na atmosfera, como resultado da sua atividade metabólica do solo, que é influenciada por vários fatores, incluindo temperatura, umidade, disponibilidade de nutrientes e qualidade do substrato (Sharma, 2022).

Análise de Aprendizado de Máquina

O surgimento do aprendizado de máquina está diretamente associado ao desenvolvimento de big data e tecnologias de computação de alto desempenho, que possibilitam novas oportunidades para a ciência de dados em diversos domínios multidisciplinares, incluindo a agro tecnologia (Mehra, *et al.*, 2016). Nessa área da inteligência artificial, que se concentra em criar algoritmos capazes de aprender a partir de dados, são tratados dois principais tipos de aprendizado de máquina: supervisionado e não supervisionado. Nesse texto, são utilizados esses dois tipos de aprendizado e suas aplicações (Alpaydin, 2010). O aprendizado de máquina supervisionado é aquele em que o algoritmo é alimentado com um conjunto de dados rotulados. Esses rótulos indicam o que o algoritmo deve aprender a prever ou classificar, um algoritmo de classificação de atributos supervisionado pode ser treinado para distinguir objetos alvos como resultados, com cada atributo sendo rotulado como resultado alvo. À medida que o algoritmo é treinado com mais dados, ele aprende a identificar padrões que permitem que ele faça previsões precisas sobre novos dados (Jeyaraj; Nadar, 2019). Já o aprendizado de máquina não supervisionado não utiliza dados rotulados, em vez disso, o algoritmo tenta encontrar padrões e estruturas nos dados por conta própria, um algoritmo de *clustering* não supervisionado, agrupa-se os dados em grupos com base em suas características, sem ter conhecimento prévio de quais grupos existem (Kumar *et al.*, 2019).

Ambos os tipos de aprendizado de máquina têm aplicações em diversas áreas, desde reconhecimento de fala até previsão de chuva e vendas. O aprendizado supervisionado é frequentemente usado em tarefas de classificação, enquanto o aprendizado não supervisionado é comumente usado em análise exploratória de dados e identificação de anomalias (He; Garcia, 2009; Behmann *et al.*, 2015; Zair; Chemseddine; Benazzouz, 2018).

Mapas Auto-organizáveis, Aprendizado Não Supervisionado

Self-Organizing Maps (SOM) ou Mapas Auto organizáveis, é uma técnica de aprendizado não supervisionado que tem como objetivo descobrir padrões em dados complexos, baseados em uma rede neural artificial que é organizada em forma de grade (Begum *et al.*, 2021).

Cada neurônio na rede representa um ponto na grade e é conectado a todos os neurônios adjacentes, que possuem pesos que são inicializados aleatoriamente e são ajustados durante o treinamento para representar as características dos dados de entrada (Chlingaryan; Sukkarieh; Whelan, 2018). Durante o treinamento, os dados de entrada são apresentados à rede, e cada neurônio na rede é ativado com base na sua distância em relação aos dados de entrada. Os neurônios que estão mais próximos dos dados de entrada são ativados com mais intensidade do que os neurônios que estão mais distantes. A medida que a rede é treinada, os neurônios que são ativados com mais frequência se tornam especializados em representar certas características dos dados de entrada (Haykin, 2010).

O processo de treinamento em mapas auto organizáveis envolve ajustar os pesos dos neurônios na rede para que os neurônios ativados com mais frequência sejam aqueles que estão mais próximos dos dados de entrada. Isso é feito através de um processo iterativo que envolve a atualização dos pesos dos neurônios com base na distância entre os dados de entrada e os pesos dos neurônios na rede. Esse processo de treinamento permite que a rede aprenda uma representação comprimida dos dados de entrada, ou seja, redução de dimensionalidade, clustering, detecção de outliers, reconhecimento de padrões e análise de imagens e processamento de sinais (Begum *et al.*, 2021).

Chlingaryan, Sukkariéh e Whelan (2018) mostraram que a combinação de modelagem de dados e métodos de agrupamento pode melhorar significativamente a qualidade do design da zona de gerenciamento, permitindo a separação dos *clusters* e a formação de zonas de manejo com base na fusão de dados NDVI e parâmetros do solo. Especificamente, eles compararam o algoritmo de *clustering k-means* com o mapa de auto-organização (SOM) para delinear os mapas da zona de gerenciamento para a aplicação de taxa variável N (Zair; Chemseddine; Benazzouz, 2018).

Análise Discriminante Linear

A Análise Discriminante Linear (LDA) é uma técnica de redução de dimensionalidade, também utilizada como etapa de pré-processamento em aplicações de *Machine Learning* e classificação de padrões. O objetivo do LDA é projetar os recursos em um espaço dimensional superior em um espaço dimensional inferior, a fim de evitar a maldição da dimensionalidade e também reduzir recursos e custos dimensionais (Adenaíke *et al.*, 2022). Para trabalhar com o LDA é necessário seguir três passos:

Primeiramente, é necessário calcular a separabilidade entre classes, que é a distância entre as médias de diferentes classes, utilizando a seguinte equação, chamada variância entre classes.

$$S_b = \sum_{i=1}^g N_i (\bar{X}_i - \bar{X}) (\bar{X}_i - \bar{X})^T$$

Equação I

O segundo passo é calcular a distância entre a média e a amostra de cada classe, chamado também de variação dentro da classe.

$$S_b = \sum_{i=1}^g (N_i - 1) S_i = \sum_{i=1}^g \sum_{j=1}^{X_i} (X_{ij} - \bar{X}_i) (X_{ij} - \bar{X})^T$$

Equação II

Finalmente, é construído o espaço de dimensão inferior que maximiza a variação entre as classes e minimiza a variação dentro da classe. P é considerado como a projeção espacial de menor dimensão, também chamada de critério de Fisher (Janabi; Kadhim, 2018).

$$P_{lda} = arg \max_P \left| \frac{P^T S_b P}{P^T S_w P} \right| \quad \text{Equação III}$$

Árvore de Decisão

O algoritmo de árvore de decisão C4.5 é um dos algoritmos utilizados, que constrói árvores de decisão a partir de um conjunto de dados de treinamento usando o conceito de entropia. O *dataset* de treinamento é um conjunto de amostras já classificadas que consiste em um vetor p -dimensional, onde representa valores de atributos ou características da amostra, bem como a categoria ou classe a que pertence (Sousa Mendes, 2022).

Em cada nó da árvore, o algoritmo C4.5 escolhe o atributo de dados que particiona com mais eficiência seu conjunto de amostras em subconjuntos que tendem a uma categoria ou outra. O critério de particionamento é o ganho de informação normalizado (diferença de entropia). O atributo com o maior ganho de informação normalizada é escolhido para tomar a decisão, então o algoritmo C4.5 repete o passo anterior nas partições menores (Ahmed; Razak, 2022).

Este algoritmo possui alguns itens importantes como, todas as amostras do conjunto pertencem a uma mesma categoria, quando este caso ocorre, o algoritmo simplesmente cria um nó folha para a árvore de decisão e escolhe a categoria em questão. Nenhum dos recursos fornece ganho de informação. Nesse caso, o algoritmo C4.5 cria um nó de decisão na árvore usando o valor esperado; Instâncias inéditas. Novamente, o algoritmo C4.5 cria um nó de decisão na árvore usando o valor esperado (Demisse; Tadesse; Bayissa, 2017; Ahmed; Razak, 2022).

Floresta Aleatória (*Random Forest*)

O algoritmo random forest (RF) é um algoritmo de aprendizado supervisionado, que cria uma floresta de forma aleatória, combinando-se um conjunto de árvores de decisão, na maioria dos casos treinados com o método de *bagging* (Li *et al.*, 2020). A ideia principal do método *bagging* é que a combinação de modelos de aprendizado melhora o resultado geral, ou seja, o algoritmo *random forest* cria várias árvores de decisão e as combina para obter uma previsão mais precisa e estável.

A RF funciona da seguinte forma (Liu; Motoda, 2007; Stańczyk; Zielosko; Jain, 2018). Para cada árvore a ser construída, uma amostra diferente com reposição é retirada dos dados de treinamento (amostra de *bootstrap*). Aproximadamente 1/3 das instâncias não são usadas para construir uma árvore. Essas instâncias são chamadas *out-of-bag* (OOB). O tamanho da amostra é o mesmo do banco de dados original. A cada passo da construção da árvore, um subconjunto diferente de n variáveis é selecionado aleatoriamente. Um número N menor que o total de N variáveis é definido. O valor padrão para tarefas de classificação é tipicamente $n \sim \sqrt{N}$ (Demisse; Tadesse; Bayissa, 2017; Stańczyk; Zielosko; Jain, 2018). A melhor divisão entre essas n variáveis é escolhida para o nó atual, em contraste com a construção típica da árvore de decisão que seleciona a melhor divisão entre todas as variáveis.

Máquina de Vetores de Suporte

O modelo de máquina de vetores de suporte (SVM) constrói um classificador de acordo com um conjunto de padrões identificados por ele nos exemplos de treinamento, sendo a classificação conhecida. Assim, existe um conjunto de classificadores lineares que separam duas ou mais classes, mas maximizam a margem de separação (distância da instância mais próxima ao hiperplano de separação das duas classes em questão). O hiperplano com margem máxima é chamado de hiperplano ótimo, será objeto da busca do treinamento do classificador (Al-Mejibli; Alwan; Abd Dhafar, 2020).

O objetivo é produzir um classificador que funcione adequadamente com exemplos desconhecidos, ou seja, exemplos que não foram aplicados durante o treinamento, adquirindo assim a capacidade de prever as saídas de novas entradas futuras. Neste modelo SVM, também foi utilizado o *kernel* núcleo gaussiano (RBF), que, ao contrário do *kernel* linear, permite resolver problemas originalmente não linearmente separáveis, através do mapeamento para um espaço maior. Existem dois parâmetros que podem ser variados em busca de um melhor resultado para o aprendizado do classificador, são eles: γ (gama) e C (custo) (Ghojogh; Karray; Crowley, 2019; Perisetty *et al.*, 2020).

Avaliação de Atributos, Classificação e Ranqueamento

A pesquisa trabalhou uma sequência como demonstrada na Figura 2 do capítulo 1, sendo que a avaliação individual dos atributos é uma etapa importante no processo de modelagem de *Machine Learning*, pois ajuda a identificar quais características têm maior influência na previsão do modelo e a eliminar atributos irrelevantes ou redundantes, melhorando a precisão e eficiência do modelo (Song *et al.*, 2022; Sugianela; Ahmad, 2020).

O algoritmo de Random Forest é um modelo de aprendizado de máquina que usa árvores de decisão aleatórias para construir sua estrutura, sua avaliação de atributos com Random Forest envolve a importância de cada atributo no modelo final. Isto é realizado calculando o ganho de informação médio de cada atributo em todas as árvores de decisão da floresta, assim a importância de um atributo é então classificada em ordem decrescente (Maleki *et al.*, 2022).

Já o *Ranker* é uma abordagem que envolve a criação de um modelo de classificação separado para cada atributo, sendo que, cada modelo é treinado para prever se um exemplo pertence ou não a uma determinada classe, com base apenas em um único atributo. A avaliação de atributos com *Ranker* envolve classificar os atributos com base no desempenho de seus respectivos modelos de classificação (Kumar, 2018; Abdullah; Abdulazeez; Sallow, 2021; Daly *et al.*, 2023).

Ambas as abordagens foram utilizadas para avaliar a importância dos atributos em um modelo de classificação. O algoritmo Random Forest pode lidar com atributos correlacionados e ser flexível ao *overfitting*, enquanto o Ranker pode

identificar interações complexas entre muitos atributos, auxiliando na interpretação e pode fornecer informações mais detalhadas sobre a importância dos atributos (Maleki *et al.*, 2022; Daly, *et al.*, 2023).

Independentemente da abordagem escolhida, a avaliação de atributos é uma etapa crítica na construção de modelos de *machine learning*, que auxilia a garantir que o modelo seja preciso e eficiente.

REFERÊNCIAS

- ANGELETTI, C.; MONACI, E.; GIANNETTA, B.; POLVERIGIANI, S.; VISCHETTI, C. Soil organic matter content and chemical composition under two rotation management systems in a Mediterranean climate. **Pedosphere**, v. 31, n. 6, p. 903-911, 2021. [https://doi.org/10.1016/S1002-0160\(21\)60032-2](https://doi.org/10.1016/S1002-0160(21)60032-2)
- ABDULLAH, D. M.; ABDULAZEEZ, A. M.; SALLOW, A. B. Lung cancer prediction and classification based on correlation selection method using machine learning techniques. **Qubahan Academic Journal**, v. 1, n. 2, p. 141-149, 2021. DOI: 10.48161/qaj.v1n2a58
- ADENAIKE, A. S.; OLOYE, O. S.; EMMANUEL, H. O.; BELLO, K. O.; IKEOBI, C. N. O. Comparison of linear discriminant analysis, support vector machine and artificial neural network in classifying Nigerian local turkeys based on plumage colours using biometric traits. **Journal of Agriculture and Rural Development in the Tropics and Subtropics (JARTS)**, v. 123, n. 2, p. 197-204, 2022.
- ALPAYDIN, E. **Introduction to machine learning**. 2. ed. MIT press, 2010. Disponível em: [http://nuir.nkumbauniversity.ac.ug/xmlui/bitstream/handle/20.500.12383/1421/Introduction%20to%20Machine%20Learning,%20Second%20Edition%20\(Adaptive%20Computation%20and%20Machine%20Learning\)%20\(%20PDFDrive%20\).pdf?sequence=1](http://nuir.nkumbauniversity.ac.ug/xmlui/bitstream/handle/20.500.12383/1421/Introduction%20to%20Machine%20Learning,%20Second%20Edition%20(Adaptive%20Computation%20and%20Machine%20Learning)%20(%20PDFDrive%20).pdf?sequence=1) . Acessado em: 10 abr. 2023.
- AHMED, A.; ZAKIUDDIN; RAZAK, T. A. Implementation of multilayer neural network with decision tree model for classification of soil type and suggesting suitable crop cultivation using machine learning technique. **Journal of Algebraic Statistics**, v. 13, n. 2, p. 2872-2878, 2022.
- AL-MEJIBLI, I. S.; ALWAN, J. K.; ABD DHAFAR, H. The effect of gamma value on support vector machine performance with different kernels. **International Journal of Electrical and Computer Engineering**, v. 10, n. 5, p. 5497, 2020.
- ASHRAF, M. N.; WAQAS, M. A.; RAHMAN, S. Microbial metabolic quotient is a dynamic indicator of soil health: trends, implications and perspectives. **Eurasian Soil Science**, v. 55, n. 12, p. 1794-1803, 2022. <https://doi.org/10.1134/S1064229322700119>
- AUDETTE, Y.; CONGREVES, K. A.; SCHNEIDER, K.; ZARO, G. C.; NUNES, A. L.; ZHANG, H.; VORONEY, R. P. The effect of agroecosystem management on the distribution of C functional groups in soil organic matter: A review. **Biology and Fertility of Soils**, v. 57, p. 881-894, 2021. <https://doi.org/10.1007/s00374-021-01591-z>
- BABASAHEB, M.; MATSAGAR, K. C.-W. Agricultural waste-derived biochar for environmental management. In: Tsang, D.C.W.; Ok, Y.S. (eds). **Biochar in agriculture for achieving sustainable development goals**. Academic Press, 2022. Cap.1, p.3-13. <https://doi.org/10.1016/B978-0-323-85343-9.00026-4>.

- BALDOCK, J. A.; SKJEMSTAD, J. O. Role of the soil matrix and minerals in protecting natural organic materials against biological attack. **Organic Geochemistry**, v. 31, n. 7-8, p.697-710, 2000. [https://doi.org/10.1016/S0146-6380\(00\)00049-8](https://doi.org/10.1016/S0146-6380(00)00049-8)
- BARROS, J. A.; DE MEDEIROS, E. V.; DA COSTA, D. P.; DUDA, G. P.; DE SOUSA LIMA, J. R.; DOS SANTOS, U. J.; HAMMECKER, C. Human disturbance affects enzyme activity, microbial biomass and organic carbon in tropical dry sub-humid pasture and forest soils. **Archives of Agronomy and Soil Science**, v. 66, n. 4, p. 458-472, 2020. <https://doi.org/10.1080/03650340.2019.1622095>
- BEHMANN, J., MAHLEIN, AK., RUMPF, T. A review of advanced machine learning methods for the detection of biotic stress in precision crop protection. **Precision Agriculture**, v. 16, p. 239–260, 2015. <https://doi.org/10.1007/s11119-014-9372-7>
- BILIAS, F.; KALDERIS, D.; RICHARDSON, C.; BARBAYIANNIS, N.; GASPARATOS, D. Biochar application as a soil potassium management strategy: A review. **Science of The Total Environment**, v. 858, p. 159782, 2023. <https://doi.org/10.1016/j.scitotenv.2022.159782>
- BINDRABAN, P.S.; DIMKPA, C.O.; PANDEY, R. Exploring phosphorus fertilizers and fertilization strategies for improved human and environmental health. **Biology and Fertility of Soils**, v. 56, p. 299–317, 2020. <https://doi.org/10.1007/s00374-019-01430-2>
- BEGUM, M.; DAS, B. C.; HOSSAIN, M. Z.; SAHA, A.; PAPRY, K. A. An improved Kohonen self-organizing map clustering algorithm for high-dimensional data sets. **Indonesian Journal of Electrical Engineering and Computer Science**, v. 24, n. 1, p. 600-610, 2021. DOI: <http://doi.org/10.11591/ijeecs.v24.i1.pp600-610>
- BOURSCHEIDT, M. L. B.; DO CARMO, K. B.; PEDREIRA, B. C.; BERBER, G. D.; FERREIRA, A. Biodiversidade de bactérias antagonicas a fungos associados à síndrome da morte do capim-marandu (*Brachiaria brizantha*). 2021. Disponível em: <http://www.alice.cnptia.embrapa.br/alice/handle/doc/1139575>. Acesso em: 15 abr. 2023.
- BRASIL, E. C.; ARAUJO, S. M. B.; VIEGAS, I. DE J. M.; DANTAS, R. C. R. **Métodos de análise do solo e representação dos resultados**. Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/218393/1/LV-RecomendacaoSolo-2020-57-61.pdf>. Acesso em 15 abr. 2023.
- BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. **Agricultura brasileira em números 2021**. Disponível em: https://www.gov.br/agricultura/pt-br/assuntos/publicacoes-e-relatorios/boletim-anual-agricultura-em-numeros/BAN_2021_web.pdf. Acesso em: 15 de abril de 2023.
- BRADY, N. C.; WEIL, R. R.; WEIL, R. R. **The nature and properties of soils**. Upper Saddle River, NJ: Prentice Hall, 2008. v. 13, p. 662-710.

CAIRO-CAIRO, P.; DIAZ-MARTIN, B.; MACHADO-DE-ARMAS, J.; RODRIGUEZ-LOPEZ, O. Effects of poultry manure on structure and some indicators of fertility in tropical soils. **Archives of Agronomy and Soil Science**, p. 1-11, 2023. <https://doi.org/10.1080/03650340.2023.2171020>

CALDERANO, S. B.; DONAGEMMA, G.; DE CAMPOS, D. V. B.; TEIXEIRA, W.; VIANA, J.; DONAGEMMA K. G.; VIANA J. H. M. **Manual de métodos de análise de solo**. 2011. Disponível em: <https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/990374/1/ManualdeMtodosdeAnilisedeSolo.pdf>. Acesso em: 15 abr. 2023.

CAKMAK, I.; MCLAUGHLIN, M.; WHITE, P. Zinc for better crop production and human health. **Plant Soil**, v. 411, p. 1–4, 2017. <https://doi.org/10.1007/s11104-016-3166-9>

CAO, S.; ZHOU, Y.; ZHOU, Y.; ZHOU, X.; ZHOU, W. Soil organic carbon and soil aggregate stability associated with aggregate fractions in a chronosequence of citrus orchards plantations. **Journal of Environmental Management**, v. 293, p. 112847, 2021. <https://doi.org/10.1016/j.jenvman.2021.112847>

CHAER, G. M.; MENDES, I. C.; DANTAS, O. D.; MALAQUIAS, J. V.; DOS REIS JUNIOR, F. B.; OLIVEIRA, M. I. L. Evaluating C trends in clayey Cerrado Oxisols using a four-quadrant model based on specific arylsulfatase and β -glucosidase activities. **Applied Soil Ecology**, v.183, p. 104742, 2023. <https://doi.org/10.1016/j.apsoil.2022.104742>

CHEN, C. R.; XU Z, H.; BLUMFIELD, T. J.; HUGHES, J. M. Soil microbial biomass during the early establishment of hoop pine plantation: seasonal variation and impacts of site preparation. **Forest Ecology and Management**, v. 186, p. 213–225, 2003. [https://doi.org/10.1016/S0378-1127\(03\)00275-5](https://doi.org/10.1016/S0378-1127(03)00275-5)

CHEN, C. H.; YAO, J. Y.; YANG, B.; LEE, H. L.; YUAN, S. F.; HSIEH, H. Y.; LIANG, P. H. Engineer multi-functional cellulase/xylanase/ β -glucosidase with improved efficacy to degrade rice straw. **Bioresource Technology Reports**, v. 5, p. 170-177, 2019. <https://doi.org/10.1016/j.biteb.2019.01.008>

CHLINGARYAN, A.; SUKKARIEH, S.; WHELAN, B. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review. **Computers and Electronics in Agriculture**, v. 151, p. 61-69. <https://doi.org/10.1016/j.compag.2018.05.012>, 2018.

COLLINS, A.; JOLANTA, K.; STEVEN, F. T.; FENTON, O.; GRZEGORZ, M.; EWA, S. Restoration of soil quality using biochar and brown coal waste. **A Review, Science of The Total Environment**, v. 722, p. 137852, 2020. <https://doi.org/10.1016/j.scitotenv.2020.137852>.

CO, R.; HUG, L. A. Prediction, enrichment and isolation identify a responsive, competitive community of cellulolytic microorganisms from a municipal landfill. **FEMS Microbiology Ecology**, v. 97, n. 5, fiab065, 2021. <https://doi.org/10.1093/femsec/fiab065>

DALY, E. J.; Kim, K.; HERNANDEZ-RAMIREZ, G.; KLIMCHUK, K. The response of soil physical quality parameters to a perennial grain crop. **Agriculture, Ecosystems & Environment**, v. 343, p. 108265, 2023.
<https://doi.org/10.1016/j.agee.2022.108265>

DEMISSE, G. B.; TADESSE, T.; BAYISSA, Y. Data mining attribute selection approach for drought modelling: A case study for greater horn of Africa. **arXiv**, v. 7, n. 4, 2017. DOI: 10.5121/ijdkp.2017.7401

DYNARSKI, K. A.; BOSSIO, D. A.; SCOW, K. M. Dynamic stability of soil carbon: Reassessing the “permanence” of soil carbon sequestration. **Frontiers in Environmental Science**, v. 8, p. 514701, 2020.
<https://doi.org/10.3389/fenvs.2020.514701>

FAO. **The state of the world's land and water resources for food and agriculture**. 2011. Disponível em: <https://www.fao.org/3/i1688e/i1688e.pdf>

FAO. Latin America. **World's Soil Resources. Guidelines for Soil Description**. Rome: Food and Agriculture Organization of the United Nations, 2015a. Disponível em: <http://www.fao.org/3/i3794en/l3794EN.pdf>

FAO. AMERICA, Latin. **World's Soil Resources**. Global Soil Partnership. 2015b. Disponível em: <https://www.fao.org/3/bc601e/bc601e.pdf>

FAO. Latin America. **World's Soil Resources**. Global Soil Partnership - Action Programme. 2015c Disponível em:
https://www.fao.org/fileadmin/user_upload/GSP/tenth_PA/GSP_Action_Framework_FINAL.pdf

FRANCO, M. F. S.; DELGADO, E. U. A. Relação entre adubação e qualidade dos produtos agrícolas. **Research, Society and Development**, v. 11, n. 4, p. e36311427562-e36311427562, 2022. <https://doi.org/10.33448/rsd-v11i4.27562>

FRANZLUEBBERS, A. J. Achieving soil organic carbon sequestration with conservation agricultural systems in the southeastern United States. **Soil Science Society of America Journal**, v. 74, n. 2, p. 347-357, 2010.
<https://doi.org/10.2136/sssaj2009.0079>

FERREIRA DE LIMA, F.; DUDA, G. P.; MEDEIROS, E.; MARQUES, M. C.; COSTA, D. P.; HAMMECKER, C.; SANTOS, U. J. Land use regulates microbial biomass and activity in highly degraded soil from Brazilian dry tropical forest. **Archives of Agronomy and Soil Science**, p. 1-16, 2022.
<https://doi.org/10.1080/03650340.2022.2099845>

GEORGE, P. B.; COELHO, K. P.; CREER, S.; LEBRON, I.; ROBINSON, D. A.; JONES, D. L. Decoupled richness of generalist anaerobes and sulphate-reducing bacteria is driven by pH across land uses in temperate soils. **European Journal of Soil Science**, v. 72, n. 6, p. 2445-2456, 2021. <https://doi.org/10.1111/ejss.13040>

GHOSH, A.; SINGH, A. B.; KUMAR, R. V.; MANNA, M. C.; BHATTACHARYYA, R.; RAHMAN, M. M.; MISRA, S. Soil enzymes and microbial elemental stoichiometry as bio-indicators of soil quality in diverse cropping systems and nutrient management practices of Indian Vertisols. **Applied Soil Ecology**, v. 145, p. 103304, 2020. <https://doi.org/10.1016/j.apsoil.2019.06.007>

GURURANI, M. A.; UPADHYAYA, C. P.; BASKAR, V.; VENKATESH, J.; NOOKARAJU, A.; PARK, S. W. Plant growth-promoting rhizobacteria enhance abiotic stress tolerance in *Solanum tuberosum* through inducing changes in the expression of ROS-scavenging enzymes and improved photosynthetic performance. **Journal of Plant Growth Regulation**, v. 32, n. 2, p. 245-258, 2013. <https://doi.org/10.1007/s00344-012-9292-6>

HAYKIN, S. **Redes neurais**: princípios e práticas. Bookman, 2010.

HE, Y.; WU, Z.; ZHAO, T.; YANG, H.; ALI, W.; CHEN, J. Different plant species exhibit contrasting root traits and penetration to variation in soil bulk density of clayey red soil. **Agronomy Journal**, v. 114, n. 1, p. 867-877, 2021. <https://doi.org/10.1002/agj2.20972>

HUANG, W.; LIN, M.; LIAO, J.; LI, A.; TSEWANG, W.; CHEN, X.; ZHENG, P. Effects of potassium deficiency on the growth of tea (*Camelia sinensis*) and strategies for optimizing potassium levels in soil: a critical review. **Horticulturae**, v. 8, n. 7, p. 660, 2022. <https://doi.org/10.3390/horticulturae8070660>

IBRAHIMI, K.; ATTIA, K. B.; AMAMI, R.; AMÉRICO-PINHEIRO, J. H. P.; SHER, F. Assessment of three decades treated wastewater impact on soil quality in semi-arid agroecosystem. **Journal of the Saudi Society of Agricultural Sciences**, v. 21, n. 8, p. 525-535, 2022. <https://doi.org/10.1016/j.jssas.2022.03.002>

JANZEN, H. H.; CAMPBELL, C. A.; BRANDT, S. A.; LAFOND, G. P.; TOWNLEY-SMITH, L. Light fraction organic matter in soils from longterm crop rotations. **Soil Science Society of America Journal**, v. 56, n. 6, p. 1799-1806, 1992. <https://doi.org/10.2136/sssaj1992.03615995005600060025x>

JEYARAJ, P.R.; NADAR, S. E. R. Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm. **Journal of Cancer Research and Clinical Oncology**, v. 145, p. 829–837, 2019. <https://doi.org/10.1007/s00432-018-02834-7>

JOERGENSEN, R. G. The fumigation-extraction method to estimate soil microbial biomass: calibration of the kEC value. **Soil Biology and Biochemistry**, v. 28, n. 1, p. 25-31, 1996. DOI: [https://doi.org/10.1016/0038-0717\(95\)00102-6](https://doi.org/10.1016/0038-0717(95)00102-6)

KIM, J.; LEE, J.; YANG, Y.; YUN, J.; DING, W.; YUAN, J.; KANG, H. Microbial decomposition of soil organic matter determined by edaphic characteristics of mangrove forests in East Asia. **Science of the Total Environment**, v. 763, p. 142972, 2021. <https://doi.org/10.1016/j.scitotenv.2020.142972>

KIPPER, L. M.; ZUFFO, A. M.; STEFANELLO, L. S.; SANTI, A. L.; REINERT, D. J. Soil texture and soybean yield in southern Brazil. **Ciência Rural**, v. 50, n. 5, p. e20190533 (2020). <https://doi.org/10.1590/0103-8478cr20190533>

KÖLLN, O. T.; SILVA, S. R.; PALUDETTO, A. **O solo: a base para a produção vegetal**. 2023. Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/doc/1150999/1/EBOOK-Sistemas-para-producao-agropecuaria-19122022-15-60.pdf>. Acesso em: 14 abr. 2023

KUMAR, N.; RUSTUM, R.; SHANKAR, V.; ADELOYE, A. J. Self-organizing map estimator for the crop water stress index. **Computers and Electronics in Agriculture**, v. 187, p. 106232, 2021. <https://doi.org/10.1016/j.compag.2021.106232>

LEAL, D. A. G.; OLIVEIRA, F. A.; SILVA, A. S.; PEREIRA, M. G.; REIS, E. F. Soil physical quality indicators under different crop management systems. **Engenharia Agrícola**, v. 40, n. 4, p. 473-480, 2020. <https://doi.org/10.1590/1809-4430-Eng.Agric.v40n4p473-480/2020>

LI, M.; WEI, Y.; YIN, Y.; ZHU, W.; BAI, X.; ZHOU, Y. Characteristics of soil physicochemical properties and microbial community of mulberry (*Morus alba* L.) and Alfalfa (*Medicago sativa* L.) intercropping system in northwest liaoning. **Microorganisms**, v. 11, n. 1, p. 114, 2023. <https://doi.org/10.3390/microorganisms11010114>

LI, Y.; WANG, C.; GAO, S.; WANG, P.; QIU, J.; SHANG, S. Impacts of simulated nitrogen deposition on soil enzyme activity in a northern temperate forest ecosystem depend on the form and level of added nitrogen. **European Journal of Soil Biology**, v. 103, p. 103287, 2021. <https://doi.org/10.1016/j.ejsobi.2021.103287>

LINDSAY, W. L.; NORVELL, W. A. Development of a DTPA soil test for zinc, iron, manganese, and copper. **Soil Science Society of America Journal**, v. 84, n. 6, p. 835-854, 2020. <https://doi.org/10.2136/sssaj1978.03615995004200030009x>

LIU, T.; WU, X.; LI, H.; ALHARBI, H.; WANG, J.; DANG, P.; YAN, W. Soil organic matter, nitrogen and pH driven change in bacterial community following forest conversion. **Forest Ecology and Management**, v. 477, p. 118473, 2020. <https://doi.org/10.1016/j.foreco.2020.118473>

LIU, H.; MOTODA, H. **Computational methods of feature selection**. Boca Raton: Chapman & Hall, 2007.

LUPWAYI, N. Z.; ELLERT, B. H.; BREMER, E.; SMITH, E. G.; PETRI, R. M.; NEILSON, J. A.; JANZEN, H. H. Ramifications of crop residue loading for soil microbial community composition, activity and nutrient supply. **Soil Use and Management**, v. 39, n. 1, p. 402-414, 2023. <https://doi.org/10.1111/sum.12834>

MALEKI, S.; ZERAATPISHEH, M.; KARIMI, A.; SAREBAN, G.; WANG, L. Assessing Variation of Soil Quality in Agroecosystem in an Arid Environment Using Digital Soil Mapping. **Agronomy**, v. 12, n. 3, p. 578, 2022. <https://doi.org/10.3390/agronomy12030578>

MEENA, R. S.; LAL, R.; YADAV, G. S. Long-term impacts of topsoil depth and amendments on soil physical and hydrological properties of an Alfisol in central Ohio, USA. **Geoderma**, v. 363, p. 114164, 2020. <https://doi.org/10.1016/j.geoderma.2019.114164>

MENDES, I. D. C.; CHAER, G. M.; REIS JUNIOR, F. B.; SOUSA, D. M. G.; SILVA, O. D. D.; OLIVEIRA, M.; MALAQUIAS, J. V. **Tecnologia BioAS**: uma maneira simples e eficiente de avaliar a saúde do solo. 2021. <http://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1133109>

MEHRA, L. K. *et al.* Predicting pre-planting risk of *Stagonospora nodorum* blotch in winter wheat using machine learning models. **Frontiers in Plant Science**, v. 7, p. 390, 2016. <https://doi.org/10.3389/fpls.2016.00390>

MPELASOKA, B. S.; SCHACHTMAN, D. P.; TREEBY, M. T.; THOMAS, M. R. A review of potassium nutrition in grapevines with special emphasis on berry accumulation. **Australian Journal of Grape and Wine Research**, v. 9, n. 3, p. 154-168, 2003. <https://doi.org/10.1111/j.1755-0238.2003.tb00265.x>

MÜHLBACHOVÁ, G.; ČERMÁK, P.; KÁŠ, M.; VAVERA, R.; PECHOVÁ, M.; MARKOVÁ, K. Boron content in soils under increasing magnesium and sulphur doses in a field experiment. **Plant, Soil and Environment**, v. 66, n. 7, p. 366-373, 2020. DOI: 10.17221/221/2020-PSE

PEREIRA-FLORES, M. E.; JUSTINO, F.; RODRIGUES, J. M.; BOEHRINGER, D.; MELO, A. A. M.; CURSI, A. G.; RUIZ-VERA, U. M. Seasonal climate impact on Brazilian pasture (*Brachiaria brizantha* cv Marandu): growth rate, CO₂ efflux, and irrigation strategies. **Theoretical and Applied Climatology**, v. 151, n. 1, p. 651-666, 2023. <https://doi.org/10.1007/s00704-022-04295-y>

PERISETTY, A.; BODEMPUDI, S. T.; SHAIK, P. R.; KUMAR, B. P. Classification of hyperspectral images using edge preserving filter and nonlinear support vector machine (SVM). *In*: INTERNATIONAL CONFERENCE ON INTELLIGENT COMPUTING AND CONTROL SYSTEMS (ICICCS), 4., 2020. p. 1050-1054. **Anais [...]**, IEEE. doi:10.1109/ICICCS48265.2020.9121035.

PROJETO MAPBIOMAS. **Mapeamento anual de cobertura e uso da terra do Brasil**. 2021. Col. 6. Disponível em: https://mapbiomas-br-site.s3.amazonaws.com/Fact_Sheet_PASTAGEM_13.10.2021_ok_ALTA.pdf. Acesso em: 10 abr. 2023.

RAYNE, N.; AULA, L. Livestock manure and the impacts on soil health: a review. **Soil Systems**, v. 4, n. 4, p. 64, 2020. <https://doi.org/10.3390/soilsystems4040064>

REGO, C. A. R. D. M.; OLIVEIRA, P. S. R. D.; MUNIZ, L. C.; ROSSET, J. S.; MATTEI, E.; COSTA, B. P.; PEREIRA, M. G. Chemical, physical, and biological properties of soil with pastures recovered by integration crop-livestock system in Eastern Amazon. **Revista Brasileira de Ciência do Solo**, v. 47, p. e0220094, 2023. <https://doi.org/10.36783/18069657rbcS20220094>

RONQUIM, C. C. **Conceitos de fertilidade do solo e manejo adequado para as regiões tropicais**. 2010. Disponível em: <https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/882598/1/BPD8.pdf>. Acesso em: 14 abr. 2023.

SANTOS, G. O.; FERREIRA, M. A. A. Growth of urochloa irrigated in clayey and sandy soil submitted to nutritional management in the cerrado. **Revista Engenharia na Agricultura-REVENG**, v. 30, p. 111-126, 2022. DOI: <https://doi.org/10.13083/reveng.v30i1.13733>

SANTOS, F. D.; OLIVEIRA, M. W.; OLIVEIRA, T. B. A.; SOARES, E. C.; ASSIS, W. O.; SILVA, R. N. Dry matter allocation and chemical composition of brachiaria brizantha and decumbens 45 days after emergence Alocação da matéria seca e composição química das braquiárias brizantha e decumbens, aos 45 dias após emergência. **Brazilian Journal of Development**, v. 8, n. 5, p. 37050-37061, 2022. DOI:10.34117/bjdv8n5-286

SAMSON, M. É.; CHANTIGNY, M. H.; VANASSE, A.; MENASSERI-AUBRY, S.; ANGERS, D. A. Coarse mineral-associated organic matter is a pivotal fraction for SOM formation and is sensitive to the quality of organic inputs. **Soil Biology and Biochemistry**, v. 149, p. 107935, 2020. <https://doi.org/10.1016/j.soilbio.2020.107935>

SILVA, A. M. M.; RAMOS, M. L. G.; NASCIMENTO, R. S. D. M. P. D.; SILVA, A. D. N.; SILVA, S. B.; CARDOSO, E. J. B. N.; PAULA, A. M. D. Soil quality indicators under management systems in a Quilombola community in the Brazilian Cerrado. **Scientia Agricola**, v. 76, p. 518-526, 2019. <https://doi.org/10.1590/1678-992X-2018-0008>

SILVA, M. **Capim-piatã em sistema de integração pecuária-floresta: revisão de literatura**. 2022. Disponível em: <https://repositorio.ifgoiano.edu.br/handle/prefix/2721>. Acesso em: 15 abr. 2023.

SINGH, H.; NORTHUP, B. K.; RICE, C. W.; PRASAD, P. V. Biochar applications influence soil physical and chemical properties, microbial diversity, and crop productivity: a meta-analysis. **Biochar**, v. 4, n. 1, p.8, 2022. <https://doi.org/10.1007/s42773-022-00138-1>

STEVENS, F.; BOGAERT, P.; VAN OOST, K.; DOETTERL, S.; VAN WESEMAEL, B. Regional-scale characterization of the geomorphic control of the spatial distribution of soil organic carbon in cropland. **European Journal of Soil Science**, v. 65, n. 4, p. 539-552, 2014. <https://doi.org/10.1111/ejss.12153>

STAŃCZYK, U.; ZIELOSKO, B.; JAIN, L. C. (eds.). Advances in feature selection for data and pattern recognition. **Springer International Publishing**, 2018.

SCHÖNING, I.; KÖGEL-KNABNER, I. Chemical composition of young and old carbon pools throughout Cambisol and Luvisol profiles under forests. **Soil Biology and Biochemistry**, v. 38, n. 8, p. 2411-2424, 2006. <https://doi.org/10.1016/j.soilbio.2006.03.005>

SMITH, S. E.; JAKOBSEN, I.; GRONLUND, M.; SMITH, F. A. Roles of arbuscular mycorrhizas in plant phosphorus nutrition: interactions between pathways of phosphorus uptake in arbuscular mycorrhizal roots have important implications for understanding and manipulating plant phosphorus acquisition. (Update on Arbuscular Mycorrhizas and Phosphorus Nutrition)(Report). **Plant Physiology**, v. 156, p. 1050, 2011. <https://doi.org/10.1104/pp.111.174581>.

SPARKS, D. L. The chemistry of soil acidity. *In: Environmental soil chemistry*. 2. ed. Academic Press, 2003. Cap. 9, p. 267-283. <https://doi.org/10.1016/B978-012656446-4/50009-8>.

SHARMA, A. R. (Ed.). **Conservation agriculture in india: a paradigm shift for sustainable production**. 1. ed. Routledge, 2022. <https://doi.org/10.4324/9781003292487>

SONG W.; ZHANG,H.; ZHAO R.; WU, K., Li, X., NIU, B.; LI, J. Study on cultivated land quality evaluation from the perspective of farmland ecosystems. **Ecological Indicators**, v. 139, p. 108959, 2022. <https://doi.org/10.1016/j.ecolind.2022.108959>

SUGIANELA, Y.; AHMAD, T. Pearson correlation attribute evaluation-based feature selection for intrusion detection system. *In: CONFERENCE: 2020 INTERNATIONAL CONFERENCE ON SMART TECHNOLOGY AND APPLICATIONS (ICOSTA). Proceedings [...]*. 2020. p. 1–5. DOI: 10.1109/ICoSTA48221.2020.1570613717

TEIXEIRA, P. C.; DONAGEMMA, G. K.; FONTANA, A.; TEIXEIRA, W. G. (eds.). **Manual de métodos de análise de solo**. 3. ed. rev. ampl. Brasília: Embrapa, 2017. pt. 2, cap. 1, p. 199-202. Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/172308/1/Pt-2-Cap-1-pH-do-solo.pdf>. Acesso em: 15 abr. 2023.

THOR, K.. Calcium—nutrient and messenger. **Frontiers in plant science**, v. 10, n. 440, 2019. <https://doi.org/10.3389/fpls.2019.00440>

WHITE, P. J.; BROADLEY, M. R. Calcium in plants. **Annals of botany**, v. 92, n. 4, p.487-511, 2003. <https://doi.org/10.1093/aob/mcg164>

YADAV, G. K.; DADHICH, S. K.; YADAV, R. K.; KUMAR, R.; DOBARIA, J.; PARAY, B. A.; RAVINDRAN, B. Impact of biomass recycling and fertilization on soil microbiological characteristics and wheat productivity in semi-arid environment. **Agronomy**, v. 13, n. 4, p. 1054, 2023. <https://doi.org/10.3390/agronomy13041054>

YANG, R.; YANG, Z.; YANG, S.; CHEN, L. L.; XIN, J.; XU, L.; LI, Z. Nitrogen inhibitors improve soil ecosystem multifunctionality by enhancing soil quality and alleviating microbial nitrogen limitation. **Science of The Total Environment**, p. 163238, 2023. <https://doi.org/10.1016/j.scitotenv.2023.163238>

YANG, Y.; CHEN, X.; LIU, L.; LI, T.; DOU, Y.; QIAO, J.; CHANG, S. X. Nitrogen fertilization weakens the linkage between soil carbon and microbial diversity: a global meta-analysis. **Global Change Biology**, v. 28, n. 21, p. 6446-6461, 2022. <https://doi.org/10.1111/gcb.16361>

ZAIR, M.; RAHMOUNE, C.; BENAZZOUZ, D. Multi-fault diagnosis of rolling bearing using fuzzy entropy of empirical mode decomposition, principal component analysis, and SOM neural network. **Journal of Mechanical Engineering Science**, v. 233, n. 9, p. 3317-3328, 2019. <https://doi.org/10.1177/0954406218805510>

ZHANG, C.; WANG, X.; JIANG, S.; ZHOU, M.; LI, F.; BI, X.; LIU, J. Heavy metal pollution caused by cyanide gold leaching: a case study of gold tailings in central China. **Environmental Science and Pollution Research**, v. 28, p. 29231-29240, 2021. <https://doi.org/10.1007/s11356-021-12728-w>

2 CAPÍTULO 1

Aprendizagem de máquina para avaliação de qualidade de áreas com pastagens usando atributos de solo

RESUMO

A análise de atributos do solo é realizada para avaliar sua aptidão produtiva e amplamente praticada de longa data. Assim, as análises químicas, físicas e biológica do solo são considerados relevantes, e a interpretação das qualidades necessita de vários especialistas, ou seja, devido à complexidade e variabilidade dos elementos dos solos. O objetivo do estudo é apresentar uma modelagem para qualidade do solo de áreas de pastagem usando abordagens de aprendizado de máquina combinadas com mapas auto-organizáveis e técnicas de classificação e ranqueamento. O estudo foi desenvolvido usando uma abordagem multi temporal com coletas do solo, em pontos georeferenciados, em três áreas com pastagens (perene com adubação e calagem, perene sem manejo, integrada com soja e uma área de preservação permanente). Amostras de solo foram coletadas nas áreas em 2019 e 2021, posteriormente foram realizadas, em laboratório, análises de atributos químicos, físicos e biológicos. No total foram avaliados 26 atributos de solo. O conjunto de dados obtidos em 2019 serviram de treinamento e o conjunto de 2021 serviu como validação para os algoritmos de aprendizagem de máquina. Foram testados vários algoritmos, entre eles: Random forest, Decision Tree, Support Vector Machine, Logistic Regression e Extra-Tree. O algoritmo Random Forest apresentou 97% de acurácia destacando-se entre eles. Com base nos resultados obtidos verificou-se que o teor de matéria orgânica do solo foi o atributo de maior importância na modelagem empregada e conclui-se que é possível utilizar os atributos do solo e a alface como planta indicadora (*Lactuca sativa*) para classificar e monitorar a qualidade de solo em áreas de pastagens.

Palavras-chave: qualidade do solo, modelos não-paramétricos, microbiologia do solo, fertilidade do solo.

Machine learning for quality evaluation of pasture areas using soil attributes

ABSTRACT

Soil attribute analysis is carried out to assess its long-standing and widely practiced productive suitability. Thus, the chemical, physical and biological analyzes of the soil are considered relevant, and the interpretation of the qualities requires several specialists, that is, due to the complexity and variability of the soil elements. The goal of the study is to present a model for soil quality in pasture areas using machine learning approaches combined with self-organizing maps and classification and ranking techniques. The study was developed using a multitemporal approach with soil collections, at georeferenced points, in three areas with pastures (perennial with fertilization and liming, perennial without management, integrated with soybeans and a permanent preservation area). Soil samples were collected in the areas in 2019 and 2021, analyzes of chemical, physical and biological attributes were subsequently carried out in the laboratory. In total, 26 soil attributes were evaluated. The data set obtained in 2019 served as training and the 2021 set served as validation for the machine learning algorithms. Several algorithms were tested, including: Random forest, Decision Tree, Support Vector Machine, Logistic Regression and Extra-Tree. The Random Forest algorithm presented 97% accuracy, standing out among them. Based on the From the results obtained, it was found that the soil organic matter content was the most important attribute in the modeling employed and it was concluded that it is possible to use soil attributes and lettuce as an indicator plant (*Lactuca sativa*) to classify and monitor the quality soil in pasture areas.

Keywords: soil quality, non-parametric models, soil microbiology, soil fertility.

INTRODUÇÃO

A necessidade de indicadores de qualidade do solo provocadas pelo manejo do solo, se faz necessário pois fundamenta a análise dos fatores que comprometem o uso sustentável dos recursos naturais, permitindo a obtenção de índices de qualidade que, por sua vez, favorecem uma avaliação integrada dos atributos físicos, químicos e biológicos (Silva; Azevedo). Dessa forma a qualidade do solo pode ser medida diretamente, mas é inferida pela avaliação de atributos dentro das propriedades físicas, químicas e biológicas.

A qualidade do solo já foi prevista por vários tipos de abordagens estatísticas, porém, existe uma lacuna nos estudos que abordam a mineração de dados por meio de mapeamento digital do solo. Isso inclui Random Forest (RF). Redes Neurais Artificiais (ANN). Árvores de Regressão (RT), Regressão Logística (LR) e suporte de máquinas de vetores (SVM) (Kovačević; Bajat; Gajić, 2010). A abordagem usual da pesquisa é avaliar vários modelos e escolher o de melhor desempenho (Brungard *et al.*, 2021). Mas, cada modelo tem suas próprias vantagens e desvantagens para uma determinada região (Górecki; Krzyśko, 2015).

O Método baseado em mapas auto-organizáveis procuram manter os vizinhos, ou seja, utilizam uma grade regular de “unidades” na qual os objetos são mapeados, criando o formato de *clusters*. Já abordagens baseadas em aprendizado de máquina estão ganhando popularidade na análise de dados científicos e são usadas em muitos métodos que envolvem predição, ajudando pequenos e grandes produtores agrícolas. Desta forma, a combinação de diferentes conjuntos de técnicas baseadas em modelagem e análise de dados pode auxiliar a resolver problemas relacionados à determinação da qualidade do solo.

O Random Forest (RF) foi utilizado para trabalhar com aprendizado de máquina supervisionado e classificação regressiva (Sheng *et al.*, 2015; Tohry *et al.*, 2022). O RF é um método multivariado não paramétrico, ou seja, os dados da amostra não precisam seguir uma distribuição normal e múltiplas variáveis podem ser trabalhadas ao mesmo tempo, de forma que seja possível melhorar a precisão da classificação e determinar a importância das variáveis existentes em uma classificação (Dong *et al.*, 2020). Este algoritmo é baseado no método estatístico de árvore de decisão, onde a classificação ou o valor de uma variável é previsto com

base em múltiplas variáveis de entrada, o poder de análise é amplificado pela combinação de camadas mais profundas de nós, o RF emprega árvores de decisão mais profundas durante o processo de treinamento, como o próprio o nome indica, é uma “floresta” de árvores (Dong, 2020). Cada nó na árvore de decisão trabalha em um subconjunto aleatório de dados do conjunto para calcular a saída.

Cada amostra é utilizada para construir uma árvore de decisão individual através do método estatístico de *bootstrap* com reposição (Salomón *et al.*, 2017; Lee; Ullah; Wang, 2020). O resultado é obtido a partir da maioria dos votos para classificação ou média para regressão (Dong, 2020; Abedini *et al.*, 2019).

A revisão da literatura indica que nem todos os fatores considerados condicionantes têm um efeito positivo ou o mesmo efeito na ocorrência do desenvolvimento do solo (Lisboa *et al.*, 2019). Portanto, alguns atributos foram removidos da modelagem.

Até a presente data, não foram encontrados estudos que combinem resultados de análises química, física, biológica e desenvolvimento de planta indicadora com intuito de modelar a qualidade do solo no ambiente de aprendizado de máquina. Nesse contexto, o objetivo do estudo é apresentar uma modelagem para qualidade do solo de áreas de pastagem usando abordagens de aprendizado de máquina combinadas com mapas auto-organizáveis e técnicas de classificação e ranqueamento.

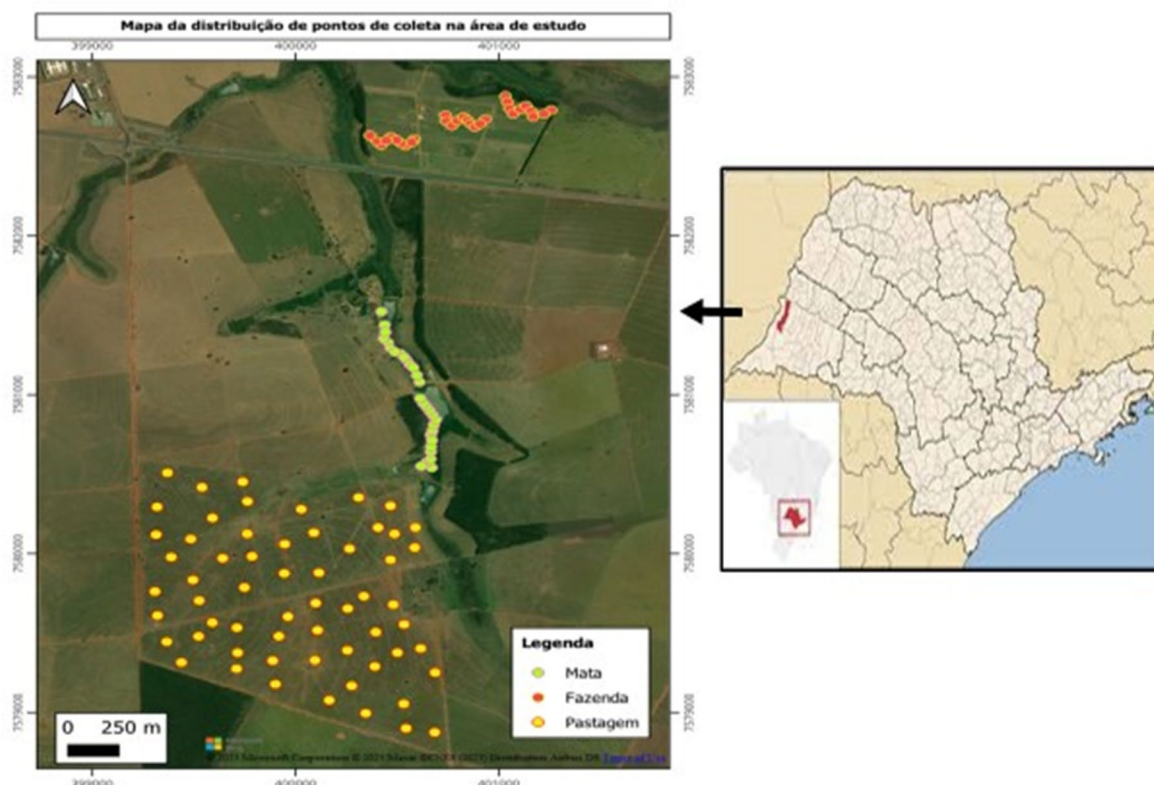
Material e Método

Área de Estudo

As áreas de estudo selecionadas para coleta de amostras de solo estão localizadas no município de Caiuá, no Estado de São Paulo (21°53'06.2"S 51°57'56.4"W e 21°51'23,3"S 51°57'42,5"W) foram realizadas duas coletas efetivadas nos meses Outubro de 2019, Setembro de 2021. Foram estabelecidas quatro áreas com manejo distinto para coleta de solo. As áreas foram denominadas como de pastagem contínua com manejo de adubação e calagem (PAST CM), pastagem contínua sem manejo (PAST SM), pastagem em sistema de integração

lavouira pecuária (ILP) e área de preservação permanente (APP) (Figura 1). Em cada área de manejo foram estabelecidos 30 pontos de coleta georreferenciados, tendo no total 120 pontos de coleta de amostra de solo no total.

Figura 1 - Localização das áreas selecionadas no município de Caiuá, SP, pontos amarelo pastagem sem manejo, pontos verdes mata de preservação permanente, e em vermelho área de ILP.



O solo das áreas é classificado como argissolo e nas áreas de pastagem o cultivo de *Urochloa brizantha*. As amostras de solo foram coletadas com trado manual na profundidade de 0 a 10 cm. Em cada área foram georreferenciados 30 pontos para coleta de amostras, sendo que cada amostra continha cerca de 2 kg de solo, resultado da coleta de cinco subamostras para melhor homogeneização e representação do ponto.

Análises de atributos químicos, físicos e biológicos do solo

Neste estudo foram estabelecidos a utilização de dez atributos de fertilidade de solo (químicos); seis atributos físicos do solo e dez atributos microbiológicos do solo, totalizando no final 26 atributos de solo analisados.

Análises químicas

As amostras de solo coletadas foram fracionadas e parte delas foram encaminhadas ao laboratório para análises de fertilidade do solo (Malavolta *et al.* 2002), sendo determinados: pH em Cloreto de Cálcio (CaCl_2), Matéria Orgânica, Fósforo (P), Acidez Potencial (H+Al), Potássio (K), Cálcio (Ca), Magnésio (Mg), Soma de Bases (SB), Capacidade de Troca Catiônica (CTC) e Porcentagem de Saturação por Bases (V%).

Análises Microbiológicas

Para as análises microbiológicas, outra parte das amostras foram encaminhadas para o laboratório microbiológico, e passaram por peneiramento e adequação de umidade para condição de capacidade de campo, em seguida foram encaminhadas para análises biológicas de: carbono e nitrogênio da biomassa microbiana (Ferreira *et al.*, 1999; Caldeira *et al.*, 2000); atividade enzimática da desidrogenase (Van Os; Ginkel, 2001) fosfatase ácida, glicosidase, arilsulfatase: (Deng, Tabatabai, 1994); respiração basal do solo (Jenkinson; Powlson, 1976); atividade da hidrólise de diacetato de fluoresceína (Chen *et al.*, 1988), quociente metabólico, ($q\text{CO}_2$) calculado a partir da razão entre o C-CO₂ liberado pela respiração do solo pelo carbono da biomassa microbiana (Anderson; Domsch, 1993) e microbiano, calculado a partir da relação entre o carbono da biomassa microbiana com carbono orgânico total do solo (Alves *et al.*, 2011; Sparling, 1992).

Análises Físicas

As análises físicas do solo foram processadas pelo fracionamento granulométrico realizado por meio de peneiramento e sedimentação. A granulometria do solo foi determinada no Laboratório utilizando-se o método da pipeta, conforme descrito em Embrapa (1997). Foram determinados o teor de argila,

partículas com diâmetro $< 0,002\text{mm}$, silte partículas com diâmetro entre $0,002$ e $0,05\text{mm}$ e areia partículas com diâmetro entre $0,05$ e 2mm .

Parte das amostras foi encaminhada para determinação da estabilidade dos agregados, sendo utilizado o método descrito por Kemper (1986) e Chepil (1965), com modificações propostas por Silva e Mielniczuk (1997) (Silva *et al.*, 2021). Esse método consiste no fracionamento do solo para excluir fragmentos vegetais, outros resíduos, pedras e brita retida na peneira, de modo que o volume total da amostra ultrapasse a malha de $9,51\text{ mm}$. Para obtenção dos agregados e determinação da estabilidade, utilizou-se o método descrito por Kemper e Chepil (1965).

Com alterações propostas por Silva e Mielniczuk (1997), a separação dos agregados em classes de tamanho foi feita pela desagregação e peneiramento em meio úmido. Foram obtidas três classes de agregados: macroagregados diâmetro $> 2,00\text{mm}$, mesoagregados diâmetro entre $2,00$ e $0,25\text{mm}$ e microagregados diâmetro entre $0,25$ e $0,05\text{mm}$. Para as determinações dos teores de C total, uma alíquota de cada classe de agregados do solo obtida foi totalmente moída e passada em peneiras de malha de $0,150\text{mm}$, com posterior pesagem de $0,150\text{g}$ dessa alíquota para análise em analisador elementar automático

A determinação dos teores de carbono particulado (COP), associado a mineral (COM) e carbono total (COT) em cada classe de agregados é importante para avaliar o sequestro de carbono no solo e entender como a matéria orgânica é distribuída em diferentes frações do solo. A análise foi realizada no analisador automático elementar, que é um equipamento que utiliza a técnica de combustão para determinar a concentração de elementos químicos presentes em uma amostra (Eivazi; Mullings; Banks, 2018).

Avaliação da Planta Indicadora

A alface Crespa *Lucy Brown* (*Lactuca sativa*) foi utilizada como planta indicadora de qualidade do solo, as amostras de solo (300 g) foram acondicionadas em vasos com a finalidade de avaliar a emergência e desenvolvimento da alface. Os vasos foram alocados dentro de câmara de crescimento de plantas (Fitotron® SGC 120; Weiss Technik UK Ltd., Loughborough, UK) por 20 dias a $29/23\text{ }^{\circ}\text{C}$ sob um

ciclo de 14 h de luz / 10 h de escuridão e com manutenção de 60% de umidade relativa no seu interior.

Foram semeadas dez sementes por vaso em solo úmido, a avaliação de emergência da planta foi realizada após sete dias, e deixou-se uma planta por vaso que foram conduzidas na câmara durante 20 dias. Após esse período avaliou-se a produção de massa fresca e massa seca da raiz, parte aérea.

Os dados foram tabulados na sequência dos resultados das análises químicas, físicas, biológicas e a planta, na sequencia agrupa-se todos os dados para formarem variáveis diretas e indiretas, o modelo de agrupamento de mapa de auto-organização, balanceado e atribuído a sua área PAST CM, PAST SM, ILP e APP, não perdendo sua referência original.

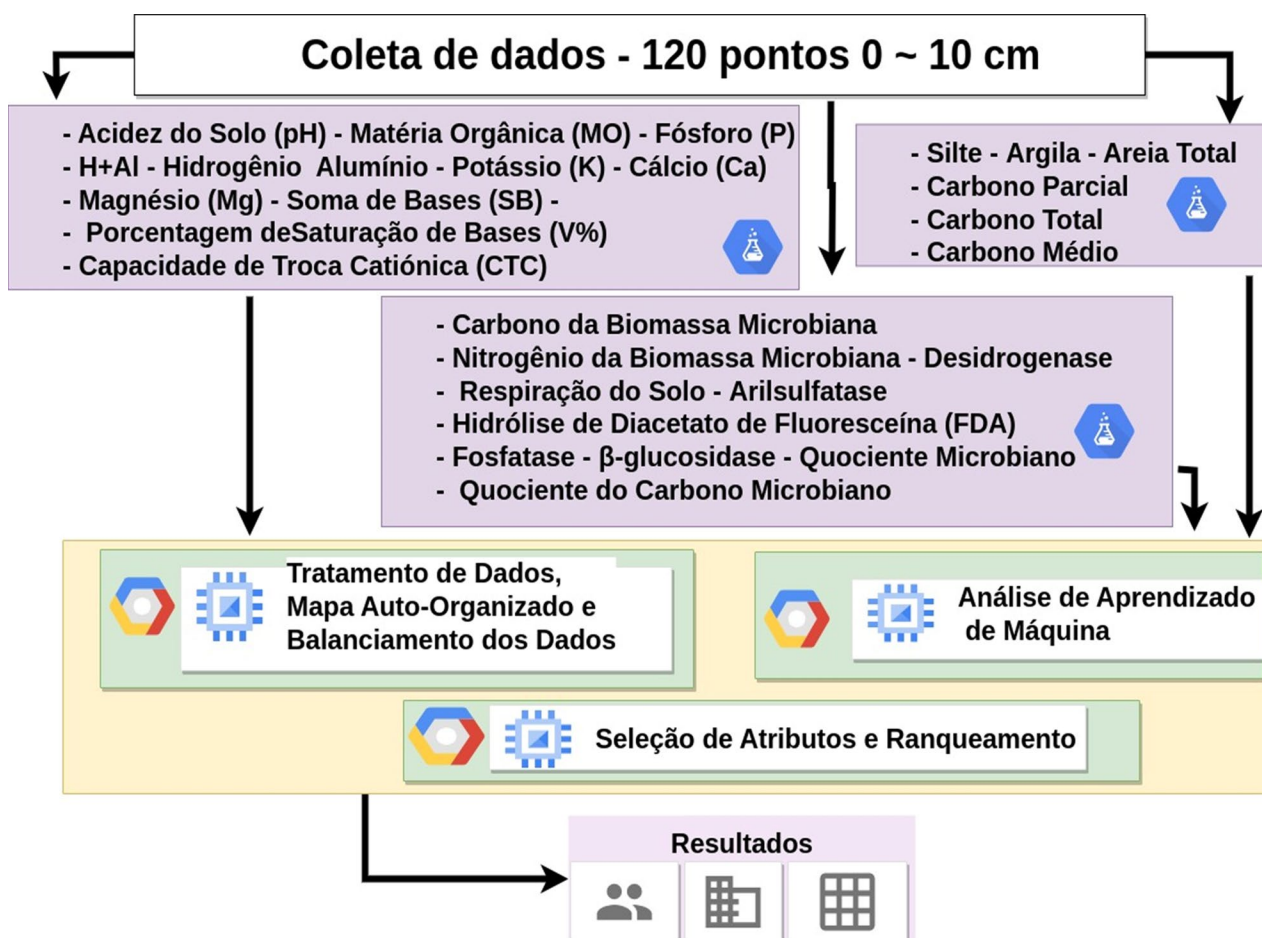
Processamento dos Dados

O presente estudo foi analisado utilizando a linguagem R em conjunto com a interface RStudio 2023.03.0+386 "Cherry Blossom" Release (3c53477afb13ab959aeb5b34df1f10c237b256c3, 2023-03-09) for Ubuntu Bionic Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) rstudio/2023.03.0+386 Chrome/108.0.5359.179 Electron/22.0.3 Safari/537.36, bem como a plataforma Weka 3.8.6., para realização dos cálculos (Janabi; Kadhim, 2018).

O projeto foi estruturado em quatro fases principais, a saber: Fase 1) descrição da área e do método de amostragem e análises laboratoriais realizados; Fase 2) organização e preparação dos dados para análise; Fase 3) implementação de técnicas de aprendizado de máquina para análise dos dados; e Fase 4) avaliação das variáveis de entrada e determinação de sua importância relativa para a obtenção do melhor método geral, com base em Ramos *et al.* (2022) e Osco *et al.* (2022).

As etapas mencionadas, encontram-se resumidas na figura 2 sendo organizada de acordo com outros estudos já realizados (Tellaeché *et al.*, 2007; Polyak *et al.*, 2018; Van *et al.*, 2019).

Figura 2 - O fluxograma demonstra os métodos implementados nesta pesquisa.



O pré-processamento de dados é uma etapa importante no processo de modelagem e análise de dados, especialmente quando se lida com dados desequilibrados, como o caso em questão. O desbalanceamento de classes pode afetar a capacidade do modelo em prever a classe minoritária, uma vez que ele pode ter uma tendência a favorecer a classe majoritária. Corrigir o desbalanceamento de dados tem sido objeto de estudos de vários pesquisadores como Krawczyk (2016), Sun, Wong e Kamel (2009), Zhou *et al.*, (2009) e He e Garcia (2009).

Para balancear a distribuição das classes no *dataset* de treinamento, o *Random Oversampling* e o *Random subsampling*, funcionam fundamentalmente, escolhendo objetos aleatoriamente e removendo-os da classe majoritária do *dataset* de treinamento. As quantidades de seleções são variadas, porém, em geral é capaz

de criar uma relação entre as classes majoritária e minoritária que é de 1:1, representa-se a relação entre a menor e o maior grupo ao final do processo (Abedini *et al.*, 2019).

Os valores encontrados nos diferentes atributos do solo, foram submetidos ao modelo matemático mapa de auto-organização (SOM) e depois o balanço das variáveis alvo da planta indicadora: emergência, matéria seca da parte aérea (MSPA), matéria seca da raiz (MSR), matéria seca matéria fresca da parte aérea (MFPA) e matéria fresca da raiz com o auxílio do software WEKA versão 3.8.6 (MFR) (Yang; Meng; Meng, 2019).

Foram utilizados nove algoritmos para o desenvolvimento desta pesquisa, incluindo Regressão Logística (LR), Análise Discriminante Linear (LDA), K vizinhos mais próximos (KNN), Árvore de Decisão (CART), Random Forest (RF), Navie Bayes NB (NB), Máquina de vetor de suporte (SVM), Gradient Boosting (GB) e Perceptron Multicamada (MLP).

Os estudos recentes de técnicas baseadas em aprendizado de máquina utilizam diferentes proporções para dividir os dados de treinamento e de validação. Nesse estudo estabeleceu-se que o banco de dados de 2019 com amostragem de 120 pontos, foi utilizado como treinamento, e o banco de dados de 2021 com a mesma amostragem foi utilizado como validação.

Foram realizados testes com diferentes divisões de classes para os classificadores utilizados, optando-se pelo uso de cinco classes (ruim, regular, bom, muito bom e excelente). As classes foram divididas em intervalos iguais e foram calculadas com base nos valores obtidos pelo desenvolvimento da planta indicadora.

O objetivo dessa etapa foi verificar o nível de detalhamento do desenvolvimento da semente e crescimento da alface, sem prejudicar o desempenho do algoritmo. A escolha da quantidade de classes é de importância considerável, pois classificações com poucas classes podem não refletir adequadamente a variação dos dados, enquanto classificações com muitas classes podem ser mais complexas e demandar mais trabalho computacional para serem processadas (Madala; Chandrasekaran; Bunk, 2023).

A avaliação do desempenho de cada algoritmo foi realizada através da construção da matriz de confusão. A partir dessa matriz, várias métricas foram

extraídas, como o *F-score*, que é a média harmônica entre os valores de precisão e *recall*. Essa métrica é útil para avaliar o desempenho do modelo de maneira geral, com base nos dados obtidos com as amostras de solo (Zhou *et al.*, 2021).

Na última etapa foi utilizado o modelo de aprendizagem de máquina que obteve melhor performance para avaliar a contribuição individual de cada atributo, identificando o ranqueamento dos atributos avaliados. Esse processo permitiu identificar a importância relativa de cada atributo de entrada no resultado do modelo.

Ao analisar o impacto individual dos atributos, é possível entender como eles influenciam a eficácia do algoritmo e direcionar esforços para refinar e aprimorar a seleção de características, visando melhorar ainda mais o desempenho do modelo.

A matriz de confusão é uma tabela que compara a classificação real com a classificação prevista pelo modelo, permitindo avaliar a acurácia, precisão, recall e F1-score da classificação. Já o R^2 indica a proporção da variação total da variável dependente que é explicada pelas variáveis independentes, sendo uma medida de quão bem o modelo se ajusta aos dados. A validação cruzada é uma técnica que divide o conjunto de dados em subconjuntos de treinamento e validação, permitindo avaliar a capacidade do modelo de generalizar dados não conhecidos previamente (Thiele-Bruhn *et al.*, 2020).

Resultados

A utilização dos valores de resultado da acurácia, foram avaliados utilizando os nove algoritmos de teste para treinamento, utilizando-se os dados obtidos com as análises obtidas em 2019 para comparação e identificação do melhor modelo (Figura 3).

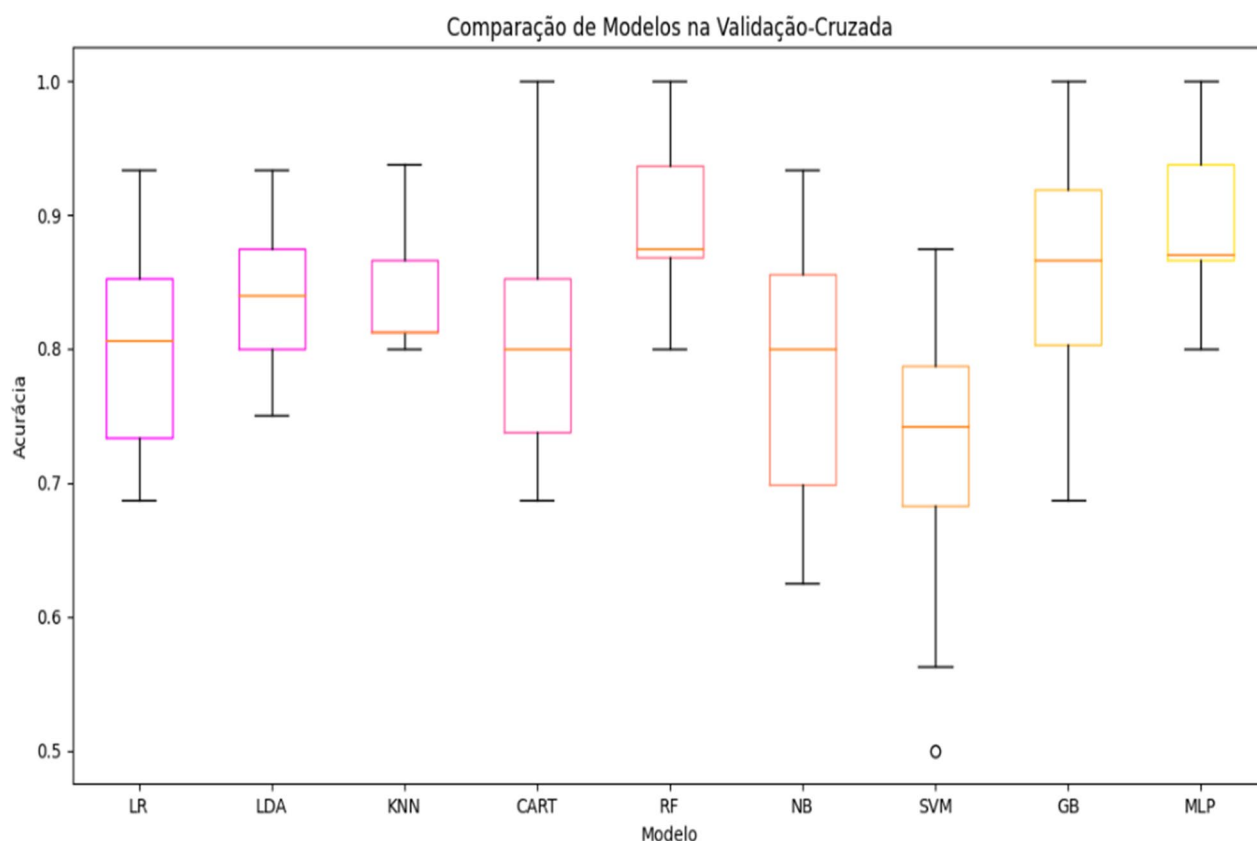
Para cada algoritmo, os valores de acurácia mensurados foram apresentados com o respectivo desvio padrão. Observa-se que o Random forest foi o algoritmo que apresentou o melhor desempenho apresentando uma acurácia de 97%.

Para que os resultados dos modelos matemáticos tenham validade, é realizado o teste de validade e robustez dos modelos desta pesquisa, são eles a

validação cruzada (*cross-validation*) e a matriz de confusão (*confusion matrix*). Esses testes de validação e robustez dos modelos são utilizados para garantir que os subconjuntos nunca foram apresentados para os modelos e verificar se há coerência nos resultados, Figura 3.

A validação cruzada ajuda a evitar problemas de sobre ajuste (*overfitting*) e sub ajuste (*underfitting*) ao ajustar um modelo a um conjunto específico de dados de treinamento. Ao avaliar o modelo em conjuntos de testes diferentes, a validação cruzada ajuda a garantir que o modelo seja capaz de generalizar para novos conjuntos de dados e não esteja super ajustado ou sub ajustado ao conjunto de dados de treinamento específico. A figura a seguir mostra graficamente no eixo X o modelo que melhor conseguiu classificar os atributos da alface em relação aos atributos do solo, o eixo Y demonstra a acurácia dos modelos, sendo que o modelo RF teve o melhor desempenho.

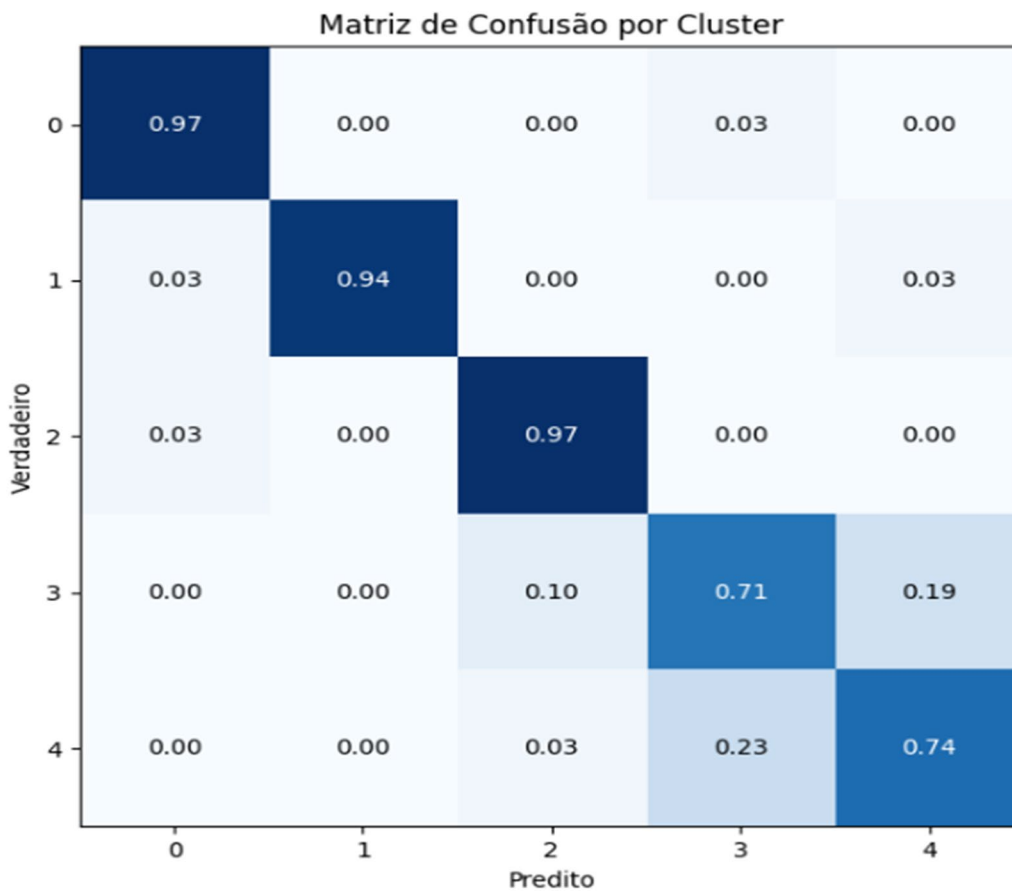
Figura 3 - Figura do box plot de todos os modelos de aprendizagem de máquina utilizados no projeto, demonstrando os resultados dos algoritmos com validação cruzada e elencando seus desempenhos.



A Figura 4 , demonstra o aprendizado do RF, com 70% dos dados e aplicado a validação cruzada, a característica de utilizar 70% para treino e 30% para validação é justamente para não haver o super aprendizado (*overfitting*), ou seja, causar vícios no aprendizado e sim, generalizar o aprendizado suficiente para conseguir entender qualquer tipo de solo que esteja em condições semelhantes aos solos amostrados.

O resultado da matriz de confusão indica boa acurácia em todos os solos classificados, no eixo Y, zero como ruim (97%), 1 como regular (94%), 2 como bom (97%), 2 como muito bom (71%) e quatro como excelente (74%) nos valores preditos no eixo X.

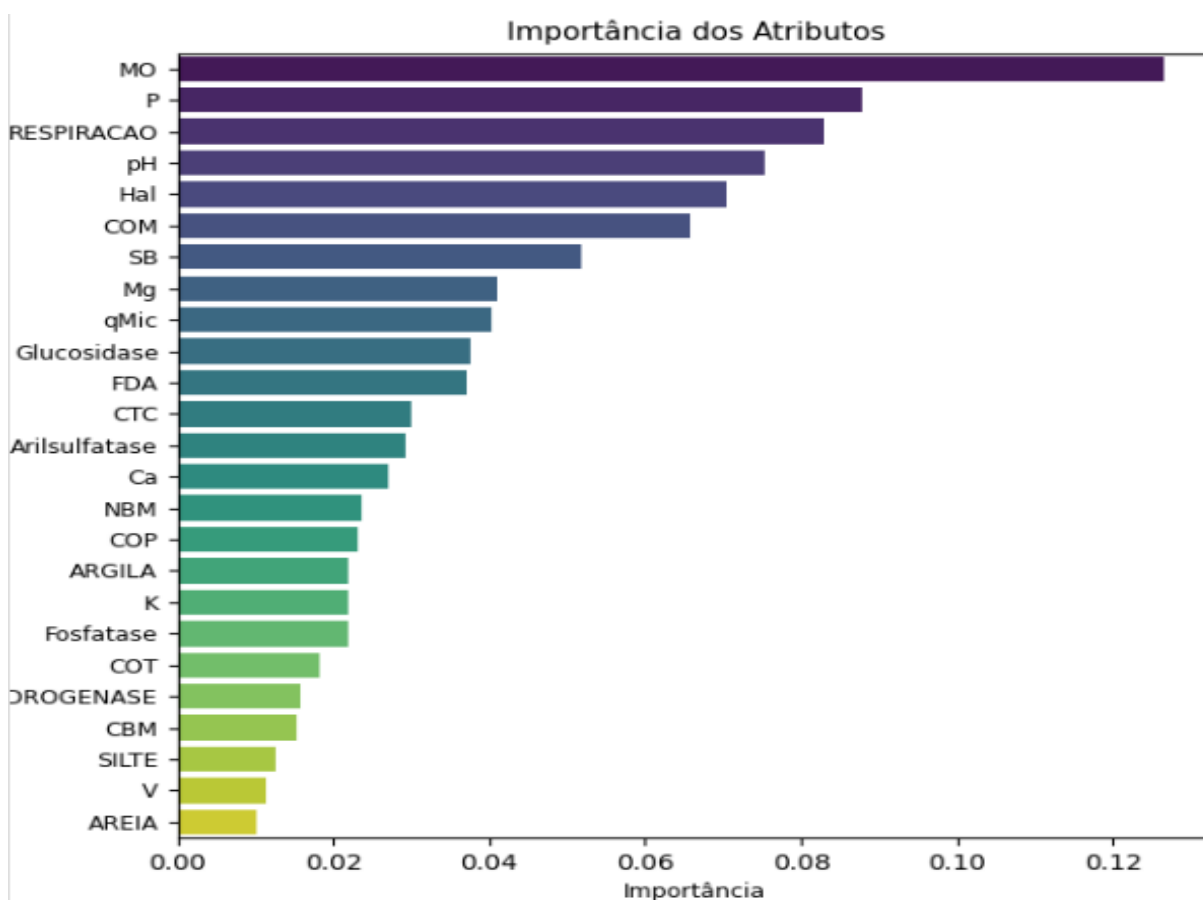
Figura 4 - Matriz de confusão do algoritmo Random Forest (RF) na etapa de treinamento com os dados de 2019.



Fonte: O autor.

A próxima figura demonstra a contribuição dos melhores atributos de solo utilizados no algoritmo Random Forest (RF) balanceado. Neste gráfico, os atributos são classificados em ordem decrescente de importância, com o mais importante no topo. As barras coloridas representam a importância relativa de cada comprimento do eixo X, sendo que cores mais escuras indicam maior importância e cores mais claras indicam menor importância.

Figura 5 - Importância dos 26 atributos de solo avaliados pelo Random Forest.

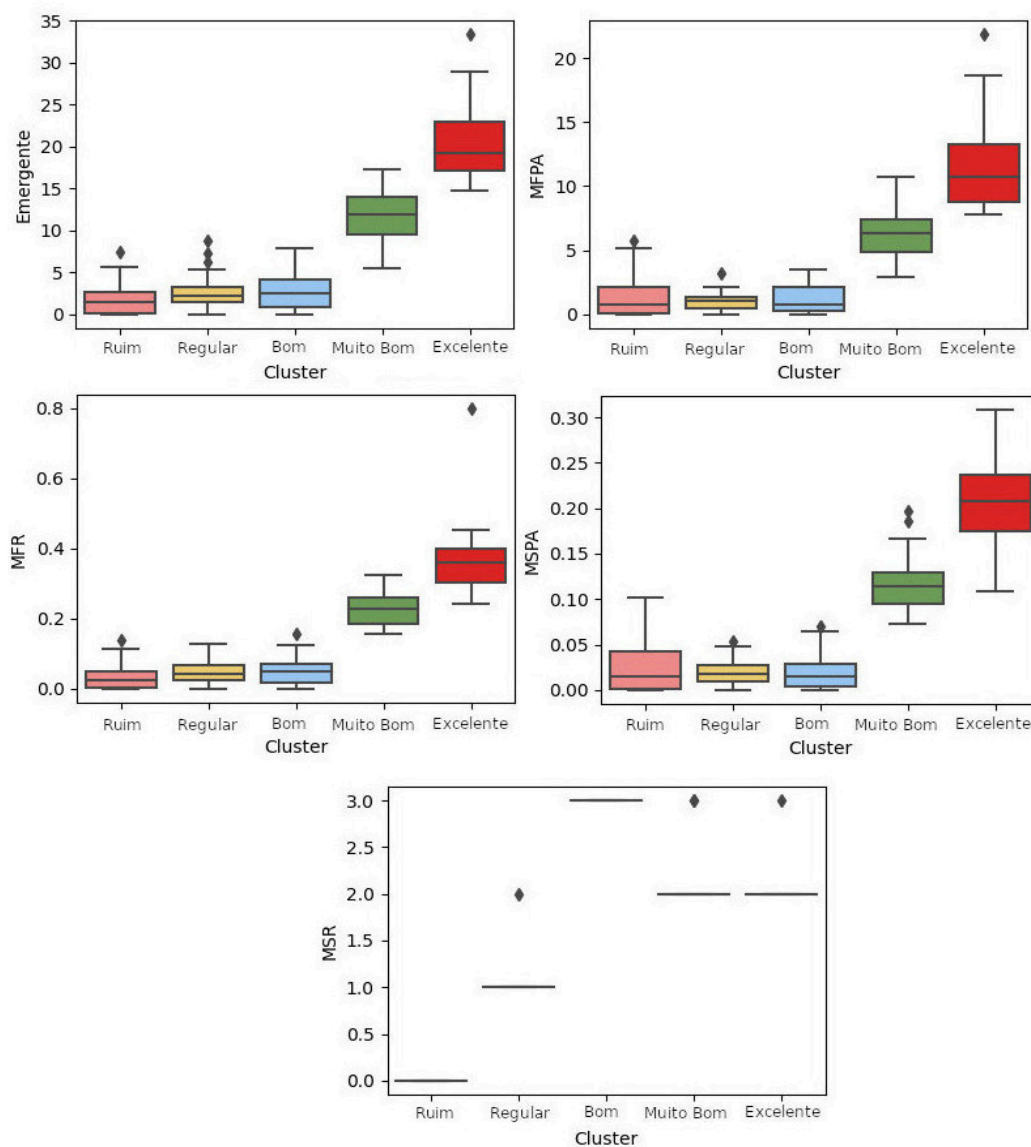


A Figura 6 apresenta *boxplots* para a visualização e análises estatísticas das cinco variáveis de crescimento da planta indicadora utilizada dentro das cinco classes (Ruim, Regular, Bom, Muito Bom e Excelente) estabelecidas, o eixo X representa as classes e o eixo Y representa as variáveis da planta (emergência da planta, matéria fresca da parte aérea, matéria fresca da raiz, matéria seca da parte aérea e matéria seca da raiz).

Os gráficos de *boxplot* mostram que as variáveis da planta revelaram diferentes padrões de resposta aos classificadores, que ajudam a garantir uma boa discriminação entre as classes de desenvolvimento da alface usando a classificação da *Random Forest* (RF). A emergência de sementes, MFPA, MSPA e MSR apresentou menor desvio ou erro em suas médias apresentadas no modelo proposto.

Em todas as variáveis utilizadas foram obtidos desenvolvimentos expressivos da alface exceto no gráfico MSR.

Figura 6 - Distribuição por box plots das classes estabelecidas pelo algoritmo RF com base no desenvolvimento da planta indicadora. Em todos os atributos da alface Emergência, MFPA, MFR obtiveram cluster excelente, muito bom e bom.



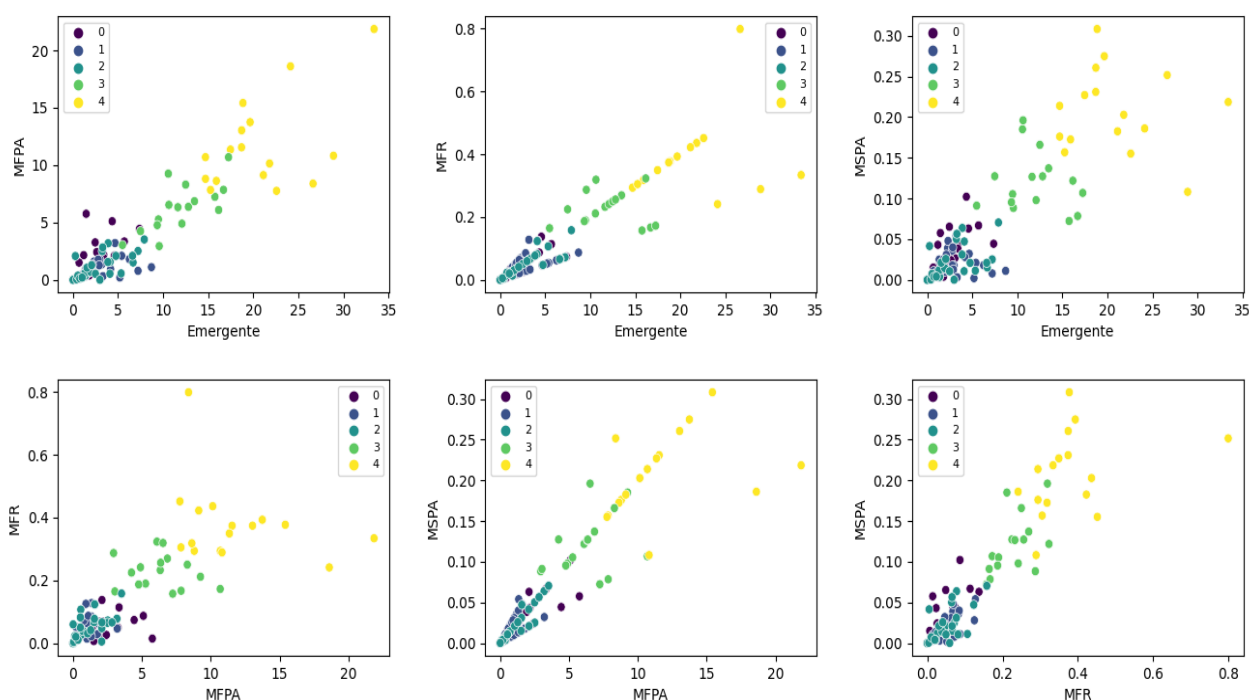
Ao correlacionar as variáveis na Figura 7, e examinar a dispersão dos pontos em cada classe, é possível observar a variabilidade dos dados em relação à correlação das variáveis. No entanto, a distribuição dos pontos pode não estar diretamente relacionada à intensidade da correlação entre as variáveis, e sim com sua classe.

No contexto de aprendizado de máquina, é importante considerar a dispersão dos pontos em cada classe, independentemente da correlação entre as variáveis. Uma dispersão maior pode indicar uma variabilidade maior nos dados

indicando *outliers*. Por outro lado, uma dispersão menor pode indicar maior sobreposição entre as classes e dificuldade na separação das mesmas como segue na figura seguinte.

Outro ponto a ser destacado é a capacidade de separar em cores as classes determinadas no modelo RF, ou seja, o modelo tem a capacidade de separar as classes zero, um, dois, três e quatro em cores e suas correlações, quanto mais alinhadas e próxima em direção ao zero de Z e Y maior a correlação.

Figura 7 - Correlação das variáveis de crescimento das plantas utilizadas no modelo RF.



Uma vez que os modelos de aprendizado de máquina foram treinados com os dados de 2019 e escolhido o modelo que melhor obteve acurácia (RF), os dados de 2021 são utilizados para a avaliação do modelo e sua acurácia, demonstrado na matriz de confusão Figura 8.

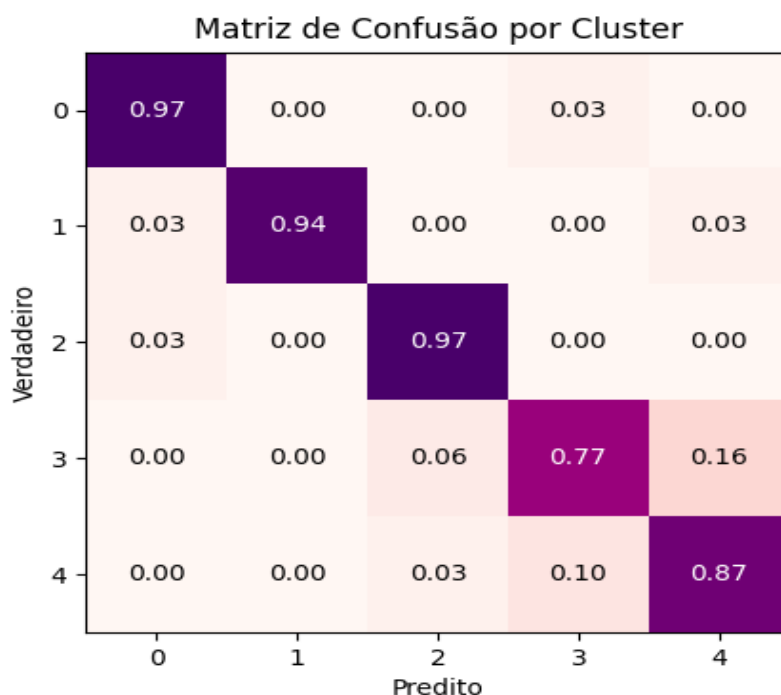
O modelo RF que melhor se ajustou aos dados de 2019 apresentou novamente grande acurácia e sensibilidade de ajuste aos dados de 2021, demonstrando robustez para identificar erros de classificação e determinar o balanceamento entre a taxa de falsos positivos e falsos negativos. Essas métricas

podem auxiliar na tomada de decisões e na escolha de ajustes ou melhorias no modelo para alcançar melhores resultados no manejo do solo.

No eixo Y, observa-se zero como variável de classificação “ruim” (97% de acertos), um como variável de classificação “regular” (94% de acertos), dois como “bom” (97% de acertos), três como “muito bom” (77% de acertos) e quatro como “excelente” (87% de acertos) nos valores preditos eixo X.

Como o modelo teve bom desempenho na predição, robustez e generalização dos dados, sugere-se que em uma próxima análise laboratorial dos atributos químico, físico, biológico do solo, e desenvolvimento da planta, não seja necessário a realização de todas as 26 análises laboratoriais, economizando recurso financeiro e temporal.

Figura 8 - Matriz de confusão para o algoritmo RF na etapa de validação com os dados de 2021.



As variáveis que foram retiradas e chegaram nesses resultados foram, areia, V%, Silte, CBM, Desidrogenase, Carbono Total, sendo que são indicativos de qualidades do solo.

Discussão

De forma geral os resultados encontrados nesse estudo demonstram-se que a classificação de padrões de crescimento da planta indicadora pelo RF é altamente sensível a características dos atributos de solo utilizados como dados de treinamento, incluindo tamanho da amostra e proporções de classe. Usando a estratégia de usar o primeiro ano da coleta como treinamento reduziu o erro de avaliação independente. De acordo com outros classificadores, tamanhos de amostra de treinamento são recomendados para comparar e avaliar a precisão da classificação e estabilidade contra o modelo RF.

Outra forma de verificar a robustez do modelo RF em relação ao conjunto de dados é a avaliação do R^2 , MAE, MSE e RMSE em função dos ajustes de dados de 2019 e 2021, como segue. $R^2 = 0.92$, MAE = 0.14, MSE = 0.18, RMSE = 0.42.

O R^2 é utilizado em modelos de regressões da área de agronomia, sendo o Coeficiente de Determinação, que visa expressar a quantidade da variância dos dados que é explicada pelo modelo construído. O Erro Quadrático Médio (MSE) consiste na diferença entre o valor predito pelo modelo e o valor real, eleva-se o resultado ao quadrado, faz-se a mesma coisa com todos os outros pontos, soma-os, e divide-se pelo número de elementos preditos. Quanto maior esse número, pior o modelo (Marques *et al.*, 2020; Osco *et al.*, 2022).

Já a Raiz do erro quadrático médio (RMSE) entra como uma forma de melhorar a interpretabilidade da métrica, acertando a unidade. Entretanto, essa medida, assim como o MSE, penaliza predições muito distantes da real. O Erro Absoluto Médio (MAE) consiste na média das distâncias entre valores preditos e reais, sendo uma métrica sólida para modelos que devem prever muitos dados ou dados sazonais, nas quais prever a tendência e sazonalidade dos números é mais importante do que os valores absolutos de cada dia do desenvolvimento da planta (Marques *et al.*, 2020).

Este estudo investigou a possibilidade de minimizar a quantidade de coletas de solo a serem realizadas ao longo dos anos, pois auxilia na redução de custos de análises melhorando a assertividade nas recomendações de manejo ou mesmo nas correções do solo. Os resultados mostrados na Figura 7 indicam que o modelo auto-

organizável dos *clusters* selecionados obtiveram resultados significativos em relação a predição do desenvolvimento da planta no ano de 2019 e 2021.

Com os resultados desse estudo verifica-se a necessidade buscar algoritmos baseados em comportamento não linear, como aprendizado de máquina, para prever variáveis biométricas da planta como a emergência, MFPA, MFR, MSPA. Neste estudo, demonstramos a capacidade de diferentes algoritmos de ML para prever variáveis biométricas, com base em coletas e análises físicas, químicas e microbiológicas do solo.

Os modelos de Regressão Logística (LR), tem apresentado bons resultados em vários estudos, porem comparando máquina de suporte de vetor (SVM), análise discriminante Linear (LDA), K-Neighbors (KNN), Perceptron Multicamada (MLP), Naive Bayes (NB), árvore de decisão (CART), Gradient Boosting (GB) e floresta aleatória (RF), o modelo que demonstrou melhor desempenho foi o RF (Ghojogh; Crowley, 2019; Ghojogh; Karray; Crowley, 2019; Santos; Dias, 2022; Seba *et al.*, 2023; Arvind; Abhishek, 2022; Antoniadis *et al.*, 2021).

Durante o processo de avaliação dos modelos de aprendizado de máquina desta pesquisa, observou-se percepções equivalentes ao trabalho de Antoniadis *et al.* (2021), no modelo de Randon Forest (RF), que retrata a precisão, reduzir o *overfitting* e *underfitting*, uma vez que combina várias árvores independentes para fazer previsões, e a estabilidade, sendo menos sensível a pequenas variações nos dados de treinamento, o que a torna mais estável.

De acordo com a revisão da literatura, nota-se que existem poucos estudos relacionados à estimação das variáveis físicas, químicas e microbiológicas do solo em função do desenvolvimento da planta. Aqui, foi aplicado algoritmos de ML em dados do solo em função ao desenvolvimento da planta, no entanto, já foram conduzidos estudos que usam uma abordagem semelhante para prever outras variáveis, como o teor de nutrientes foliares em pomares de frutas cítricas e rendimento em cultivares de milho (Marques *et al.*, 2020; Osco *et al.*, 2022).

A utilização de validação cruzada (cross-validation) e matriz de confusão são importantes, para melhorar a confiança nos resultados dos algoritmos, de acordo com Thiele-Bruhn *et al.*, (2020) e Zhou *et al.*, (2021), a matriz de confusão atinge várias métricas de resultado, como o *F-score*, que é a média harmônica entre os

valores de precisão e *recall*. Na validação cruzada, Thiele-Bruhn *et al.* (2020) relata que esta técnica que divide o conjunto de dados em subconjuntos de treinamento e validação, permitindo avaliar a capacidade do modelo de generalizar para dados não vistos previamente.

Os resultados obtidos deste estudo demonstram bons resultados, robustez, sensibilidade e balanceamento entre as taxas de erro, porém futuras pesquisas são necessárias para validar os resultados deste estudo em diferentes tipos de solos e condições climáticas, a fim de fornecer uma base mais sólida para a utilização desses atributos em diferentes regiões e culturas. Além disso, estudos adicionais podem ser realizados para identificar novos atributos que possam ser relevantes para avaliar a qualidade do solo.

Conclusões

Utilizando-se aprendizagem de máquina com modelo de mapas auto-organizáveis foi possível separar em *clusters* uma classificação para qualidade de solo com pastagens a partir de um conjunto de atributos do solo.

Foi ranqueado os atributos que mais contribuíram para a qualidade do solo chegando-se a matéria orgânica do solo como o atributo de maior importância;

O modelo *Random Forest* foi o mais eficaz e que se ajustou melhor com os dados do solo conseguindo acurácia de até 97%.

REFERÊNCIAS

- ABEDINI, M.; GHASEMIAN, B.; SHIRZADI, A.; SHAHABI, H.; CHAPI, K.; PHAM, B. T.; TIEN BUI, D. A novel hybrid approach of bayesian logistic regression and its ensembles for landslide susceptibility assessment. **Geocarto International**, v. 34, n. 13, p. 1427-1457, 2019.
- ANTONIADIS, A.; LAMBERT-LACROIX, S.; POGGI, J.-M. Random forests for global sensitivity analysis: a selective review. **Reliability Engineering & System Safety**, v. 206, p. 107312, 2021. <https://doi.org/10.1016/j.ress.2020.107312>
- ARVIND, K. G.; ABHISHEK, B. A review on cyberstalking detection using machine learning techniques: current trends and future direction. **International Journal of Engineering Trends and Technology**, v. 70, n. 3, p. 95-107, 2022. <https://doi.org/10.14445/22315381/IJETT-V70I3P211>
- BRUNGARD, C.; NAUMAN, T.; DUNIWAY, M.; VEBLEN, K.; NEHRING, K.; WHITE, D.; ANCHANG, J. Regional ensemble modeling reduces uncertainty for digital soil mapping. **Geoderma**, v. 397, p. 114998, 2021. <https://doi.org/10.1016/j.geoderma.2021.114998>
- CHEPIL, W. S. Transport of soil and snow by wind. *In: Agricultural meteorology*. Boston, MA: American Meteorological Society, 1965. p. 123-132.
- CHEN, H. C.; BROWN, J. H.; MORELL, J. L.; HUANG, C. M. Synthetic magainin analogues with improved antimicrobial activity. **Febs Letters**, v. 236, n. 2, p.462-466, 1988. [https://doi.org/10.1016/0014-5793\(88\)80077-2](https://doi.org/10.1016/0014-5793(88)80077-2)
- DENG, S. P.; TABATABAI, M. A. Cellulase activity of soils. **Soil Biology and Biochemistry**, v. 26, n. 10, p. 1347-1354, 1994. [https://doi.org/10.1016/0038-0717\(94\)90216-X](https://doi.org/10.1016/0038-0717(94)90216-X)
- DONG, X. *et al.* A survey on ensemble learning. **Frontiers of Computer Science**, v. 14, p. 241–258, 2020. <https://doi.org/10.1007/s11704-019-8208-z>
- EMBRAPA. **Manual de métodos de análise do solo**. Rio de Janeiro: Embrapa, 1997. 212p.
- EIVAZI, F.; MULLINGS, N.; BANKS, M. L. Effect of select surfactants on activities of soil enzymes involved in nutrient cycling. **Communications in Soil Science and Plant Analysis**, v. 49, n. 3, p. 371-379, 2018. <https://doi.org/10.1080/00103624.2018.1427263>
- FERREIRA, A. S.; CAMARGO, F. A. O.; VIDOR, C. Use of microwave radiation to evaluate soil microbial biomass. **Revista Brasileira de Ciência do Solo**, v. 23, p. 991-996, 1999.
- GHOJOGH, B.; CROWLEY, M. Linear and quadratic discriminant analysis: tutorial. **arXiv preprint arXiv:1906.02590**, 2019. <https://doi.org/10.48550/arXiv.1906.02590>

GHOJOGH, B.; KARRAY, F.; CROWLEY, M. Fisher and kernel Fisher discriminant analysis: tutorial. **arXiv preprint arXiv:1906.09436**, 2019. <https://doi.org/10.48550/arXiv.1906.09436>

GÓRECKI, T.; KRZYŚKO, M. Regression methods for combining multiple classifiers. **Communications in Statistics-Simulation and Computation**, v. 44, n. 3, p. 739-755, 2015. <https://doi.org/10.1080/03610918.2013.794286>

HAYKIN, S. **Redes neurais**: princípios e práticas. Bookman, 2010.

HE, H.; GARCIA, E. A. Learning from imbalanced data. **IEEE Transactions on Knowledge and Data Engineering**, v. 21, n. 9, p. 1263-1284, 2009.

JANABI K. B. A.; KADHIM R. 'Weka' Feature Selection - bad results - data reduction techniques: a comparative study for attribute selection methods. **International Journal of Advanced Computer Science and Information Technology**, v. 8, n. 1, p. 1–13, 2018.

JENKINSON, D. S.; POWLSON, D. S. The effects of biocidal treatments on metabolism in soil—V: a method for measuring soil biomass. **Soil Biology and Biochemistry**, v. 8, n. 3, p. 209-213, 1976.

KEMPER, W. D.; CHEPIL, W. S. Size distribution of aggregates. *In*: **Methods of soil analysis**: part 1 physical and mineralogical properties, including statistics of measurement and sampling. 1965. Cap. 39, p. 499-510. <https://doi.org/10.2134/agronmonogr9.1.c39>

KEMPER, W. D.; ROSENAU, R. C. Aggregate stability and size distribution. **Methods of soil analysis**: part 17 physical and mineralogical methods. 1986. Cap. 5, p. 425-442. Disponível em: <https://eprints.nwisrl.ars.usda.gov/id/eprint/732/3/585.pdf>

KRAWCZYK, B. Learning from imbalanced data: open challenges and future directions. **Progress in Artificial Intelligence**, v. 5, n. 4, p. 221-232, 2016.

KOVAČEVIĆ, M.; BAJAT, B.; GAJIĆ, B. Soil type classification and estimation of soil properties using support vector machines. **Geoderma**, v. 154, n. 3-4, p. 340-347, 2010. <https://doi.org/10.1016/j.geoderma.2009.11.005>

LEE, T. H.; ULLAH, A.; WANG, R. Bootstrap aggregating and random forest. *In*: **Macroeconomic forecasting in the era of big data**. Cham: Springer, 2020. p. 389-429.

LISBOA, I. P.; CHERUBIN, M. R.; SATIRO, L. S.; SIQUEIRA-NETO, M.; LIMA, R. P.; GMACH, M. R.; CERRI, C. E. Applying Soil Management Assessment Framework (SMAF) on short-term sugarcane straw removal in Brazil. **Industrial Crops and Products**, v. 129, n. 175-184, 2019.

MADALA, V. C.; CHANDRASEKARAN, S.; BUNK, J. CNNs avoid curse of dimensionality by learning on patches. **IEEE Open Journal of Signal Processing**, 2023. DOI: 10.1109/OJSP.2023.3270082

MARQUES RAMOS, A. P.; PRADO OSCO, L.; ELIS GARCIA FURUYA, D.; NUNES GONÇALVES, W.; CORDEIRO SANTANA, D.; PEREIRA RIBEIRO TEODORO, L.; ANTONIO DA SILVA JUNIOR, C.; FERNANDO CAPRISTO-SILVA, G.; LI, J.; HENRIQUE ROJO BAIO, F. A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. **Computers and Electronics in Agriculture**, v. 178, p. 105791, 2020. <https://doi.org/10.1016/j.compag.2020.105791>

LI, J.; BAIO, H. R. F., MARCATO, J.; TEODORO, E. P.; PISTORI, H. A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. **Computers and Electronics in Agriculture**, v. 178, 2020. <https://doi.org/10.1016/j.compag.2020.105791>

OSCO, L. P.; FURUYA, G. D. E.; FURUYA, G., M. T.; CORRÊA, D. V.; GONÇALVES, W. N.; JUNIOR, J. M.; BORGES, M.; BLASSIOLI-MORAES, M. C.; MICHEREFF, F. M. F.; SANTOS, A. M. F.; LAUMANN, R. A.; LISENBERG, V.; RAMOS, M. A. P.; JORGE, L. A. C. An impact analysis of pre-processing techniques in spectroscopy data to classify insect-damaged in soybean plants with machine and deep learning methods. **Infrared Physics and Technology**, v. 123, 2022. <https://doi.org/10.1016/j.infrared.2022.104203>

OSCO, L. P.; RAMOS, A. P. M.; FAITA, P. M. M.; MORIYA, É. A. S.; IMAI, N. N.; ESTRABIS, N.; IANCZYK, F.; ARAÚJO, F. F.; LIESENBERG, V.; CASTRO, J. L.A. A machine learning framework to predict nutrient content in valencia-orange leaf hyperspectral measurements. **Remote Sensing**, v. 12, p. 906, 2020. <https://doi.org/10.3390/rs12060906>

POLYAK, Y. M.; BAKINA, L. G.; CHUGUNOVA, M. V.; MAYACHKINA, N. V.; GERASIMOV, A. O.; BURE, V. M. Effect of remediation strategies on biological activity of oil-contaminated soil-A field study. **International Biodeterioration & Biodegradation**, v. 126, p.57-68, 2018. <https://doi.org/10.1016/j.ibiod.2017.10.004>

RAMOS, A. P. M.; GOMES, F. D. G.; PINHEIRO, M. M. F.; FURUYA, D. E. G.; GONÇALVES, W. N.; JUNIOR, J. M.; MICHEREFF, M. F. F.; BLASSIOLI-MORAES, M. C.; BORGES, M.; ALAUMANN, R. A.; LIESENBERG, V.; JORGE, L. A. C.; OSCO, L. P. Detecting the attack of the fall armyworm (*Spodoptera frugiperda*) in cotton plants with machine learning and spectral measurements. **Precision Agriculture**, v. 23, n. 2, p.470–491,2022. <https://doi.org/10.1007/s11119-021-09845-4>

SALOMÓN, R. L.; RODRÍGUEZ-CALCERRADA, J.; STAUDT, M. Carbon losses from respiration and emission of volatile organic compounds—the overlooked side of tree carbon budgets. *In*: GIL-PELEGRÍN, E.; PEGUERO-PINA, J.; SANCHO-KNAPIK, D. (eds). **Oaks Physiological Ecology exploring the functional diversity of genus Quercus L.** Cham: Springer, 2017. https://doi.org/10.1007/978-3-319-69099-5_10

SILVA, I.; MIELNICZUK, J. Ação do sistema radicular de plantas na formação e estabilização de agregados do solo. **Revista Brasileira de Ciência do Solo**, v. 21, n. 1, p. 113-117, 1997.

SILVA, I. M.; ROMERO, D. J.; GUIMARÃES, C. C. B.; ALVES, M. R.; OSCO, L. P.; SOUZA, A. B.; SILVA, A. P.; DEMATTÊ, J. A. M. Readily dispersible clay in soils from different Brazilian regions by visible, near, and mid-infrared spectral data. **International Journal of Remote Sensing**, v. 42, n. 18, p. 6945–6962, 2021. <https://doi.org/10.1080/01431161.2021.1948625>

SANTOS, K.; DIAS, J. P.; AMADO, C. A literature review of machine learning algorithms for crash injury severity prediction. **Journal of Safety Research**, v. 80, p. 254-269, 2022. <https://doi.org/10.1016/j.jsr.2021.12.007>

SEBA, P. A.; SELVAKUMARAN, R.; RAJ, D. Predictive analytics for fake currency detection. *In: International Conference on Intelligent Vision and Computing (ICIVC 2022)*. **Proceedings [...]**. Cham: Springer Nature Switzerland, 2023. v. 1, p. 128-137. DOI: 10.1007/978-3-031-31164-2_11

SHENG, K.; DONG, W.; KONG, Y.; MEI, X.; Li, J.; WANG, C.; HU, B. G. Evaluating the quality of face alignment without ground truth. **Computer Graphics Forum**, v. 34, n. 7, p. 213-223, oct. 2015.

SUN, Y.; WONG, A. K.; KAMEL, M. S. Classification of imbalanced data: a review. **International Journal of Pattern Recognition and Artificial Intelligence**, v. 23, n. 04, p. 687-719, 2009.

TELLAECHE, A. *et al.* A vision-based classifier in precision agriculture combining bayes and support vector machines. **IEEE International Symposium on Intelligent Signal Processing**, 2007. DOI: 10.1109/WISP.2007.4447561.

THIELE-BRUHN, S.; SCHLOTTER, M.; WILKE, B. M.; BEAUDETTE, L. A.; MARTIN-LAURENT, F.; CHEVIRON, N.; RÖMBKE, J. Identification of new microbial functional standards for soil quality assessment. **Soil**, v. 6, n. 1, p. 17-34, 2020. <https://doi.org/10.5194/soil-6-17-2020>

TOHRY, A.; JAFARI, M.; FARAHANI, M.; MANTHOURI, M.; CHELGANI, S. C. Variable importance assessments of an innovative industrial-scale magnetic separator for processing of iron ore tailings. **Mineral Processing and Extractive Metallurgy**, v. 131, n. 2, p. 122-129, 2022.

VAN, L.; CREAMER, R. E.; CLUZEAU, D.; DEBELJAK, M.; GATTI, F.; HENRIKSEN, C. B.; Rutgers, M. Modeling of soil functions for assessing soil quality: soil biodiversity and habitat provisioning. **Frontiers in Environmental Science**, v. 7, n. 113, 2019. <https://doi.org/10.3389/fenvs.2019.00113>

VAN, O. G. J.; VAN GINKEL, J. H. Suppression of Pythium root rot in bulbous Iris in relation to biomass and activity of the soil microflora. **Soil Biology and Biochemistry**, v. 33, n. 11, p. 1447-1454, 2001. [https://doi.org/10.1016/S0038-0717\(01\)00053-0](https://doi.org/10.1016/S0038-0717(01)00053-0)

ZHOU, J.; HUANG, S.; WANG, M.; QIU, Y. Performance evaluation of hybrid GA–SVM and GWO–SVM models to predict earthquake-induced liquefaction potential of soil: a multi-dataset investigation. **Engineering with Computers**, p. 1-19, 2021. <https://doi.org/10.1007/s00366-021-01418-3>

ZHOU, X.; LIU, H.; ZHU, S.; WANG, X.; WANG, H. Research on matching of the remote sensing image matching based on histogram invariant moments and genetic algorithms. *In: International Conference on Information Engineering and Computer Science, IEEE*, p. 1-4, 2009. DOI: 10.1109/ICIECS.2009.5362849

YANG, X.; MENG, L.; MENG, F. Combination of self-organizing map and parallel factor analysis to characterize the evolution of fluorescent dissolved organic matter in a full-scale landfill leachate treatment plant. **Science of the Total Environment**, v. 654, p. 1187-1195, 2019. <https://doi.org/10.1016/j.scitotenv.2018.11.135>

3 CAPÍTULO 2

Comparação de índices de qualidade do solo de pastagens por análise de componentes principais e validados com planta indicadora

RESUMO

O principal uso dado ao solo brasileiro é a pastagem, que ocupa cerca de 150 milhões de hectares. Existe uma demanda em avaliar a qualidade dos solos agrícolas de forma geral nas próximas décadas. Os desenvolvimentos de métodos confiáveis sobre a qualidade do solo são necessários, e o índice de qualidade do solo (IQS) é um deles, portanto, o objetivo deste estudo foi calcular o IQS do solo com diferentes manejos para examinar sua relação com desenvolvimento de uma planta indicadora como validadora, assumindo que os usos do solo afetam o solo e seu desenvolvimento de forma diferente. A partir de solos com diferentes usos (áreas de preservação e de pastagens perenes e integradas) análise de componentes foi usada para selecionar um conjunto mínimo de dados (CMD) para gerar IQS de 26 atributos do solo coletadas em áreas no município de Caiuá, SP. Foram realizadas duas campanhas para coleta de solo em 2019 e 2021. Utilizando-se o modelo de componentes principais foram selecionados atributos do solo que foram determinados em laboratório. Três IQS foram calculados de cada conjunto mínimo de dados por meio de equações de pontuação lineares, não lineares e por treinamento. Na conclusão os IQS gerados foram validados pelo crescimento da planta indicadora. O IQS foi afetado pelo manejo de solo nas diferentes áreas de pastagens. Indicando o sistema ILP como de maior valor obtido nas duas avaliações. O IQS calculado usando métodos não lineares foram mais significativamente correlacionados com o desenvolvimento da alfaca, indicando que é relativamente fácil de usar. A estratégia de usar o IQS validado pode ser considerado como o melhor e mais fácil modelo devido ser considerado como um procedimento relativamente menos dispendioso de tempo.

Palavras-chave: modelo linear, modelo não linear, fertilidade do solo, microbiologia do solo

ABSTRACT

The main use given to Brazilian soil is pasture: it occupies around 150 million hectares. There is a demand to assess the quality of agricultural soils in general in the coming decades. The development of reliable methods on soil quality is necessary, and the soil quality index (SQI) is one of them. of the most used. Therefore, the objective of this study was to calculate the IQS area of pastures with different managements to examine its relationship with the development of an indicator plant for validation, assuming that land uses affect soil qualities differently. From soils with different uses (preservation areas and perennial and integrated pastures) component analysis was used to select a minimum data set (MDS) to generate IQS of 26 soil attributes collected in areas in the municipality of Caiuá, SP . Two soil collection campaigns were carried out in 2019 and 2021. Using PCA, soil attributes that were determined in the laboratory were selected, which were used in the calculation of the IQS. Three IQS were calculated from each MDS through linear and non-linear scoring equations and through training. In the end, the generated IQS were validated by the growth of the indicator plant. The IQS was affected by soil management in different pasture areas. Indicating the ILP system as having the highest value obtained in both evaluations. The IQS calculated using non-linear methods were more significantly correlated with lettuce development, indicating that it is relatively easy and easy to use. The strategy of using the validated IQS can be considered as the best and easiest model because it is considered as a relatively less time-consuming procedure.

Key words: *Urochloa brizantha*, soil use, soil fertility, soil microbiology

INTRODUÇÃO

A área de pastagem com espécies cultivadas no Brasil ocupa cerca de 154 milhões de hectares (IBGE, 2020). O gênero *Urochloa*, conhecido também como braquiárias, são largamente utilizadas em pastagens na América Tropical onde a *Urochloa brizantha* apresenta boa digestibilidade pelo gado, além de ser rica em proteína bruta, energia e minerais, tais características favorece a rentabilidade da produção animal em sistemas de pastoreio extensivo ou intensivo (Bourscheidt, 2021; Silva, 2022; Pereira-Flores *et al.*, 2023).

Devido ao solo ser um importante componente relacionado à produção vegetal, a conservação ou a melhoria da sua qualidade é importante para a sustentação dessa atividade produtiva (Chaer; Tótola, 2007). A qualidade do solo, de acordo com Doran e Parkin (1994) Nortcliff (2002), Sparling e Schipper (2004) é definida pela sua capacidade de funcionar dentro das fronteiras dos ecossistemas para sustentar a produtividade biológica, manter a qualidade ambiental e promover o crescimento das plantas e animais. Marzaioli *et al.* (2010) relataram que a avaliação das práticas de manejo realizadas no solo, pode definir indicadores de qualidade para comparar as condições do solo ao longo do tempo ou entre diferentes locais. Segundo Sánchez-Navarro (2015) os índices de qualidade do solo são obtidos através da integração de diferentes variáveis que representam o conjunto de propriedades do solo, fornecendo informações sobre o estado do solo.

Estudos sobre qualidade do solo, com frequência, utilizam funções estatísticas para agrupar variáveis, com posterior indicação daquelas que possam melhor representar a qualidade do solo (Brejda *et al.*, 2000; Anderson, 2003; Sparling; Schipper, 2004; Tziolas, 2022). De acordo com Stocking (2003) a qualidade do solo é um conceito funcional complexo, porem pode ser inferida a partir de características do solo, ou seja, uma variedade de parâmetros ou indicadores da terra, que é utilizado para estimar a qualidade do solo. No entanto, a qualidade do solo está frequentemente relacionada ao manejo da terra, bem como às características do solo (Silva *et al.*, 1997; Silva *et al.*, 2003; Dick; Stott, 2001).

Modelos matemático para estimar o índice de qualidade do solo (IQS) a partir da análise integrada de diferentes atributos do solo foram propostos por diferentes autores (Karlen *et al.*, 1997; Andrews; Karlen; Mitchell, 2002). O IQS foi

concebido para que o manejo do solo não seja direcionado apenas para aumento da produtividade, mas também para promover sustentabilidade ambiental. Assim, um IQS apropriado pode ter três objetivos: qualidade ambiental, sustentabilidade agrônômica e viabilidade socioeconômica (Seybold *et al.*, 1997; Andrews; Karlen; Mitchell, 2002). Para maior confiabilidade do IQS como índice de sustentabilidade agrônômica e ambiental é importante que estes se correlacionem com avaliações da produção vegetal, como produção de biomassa vegetal, grãos, madeira, etc (Mukherjee; Lal, 2014).

Os índices de qualidade do solo (IQS) são usados para selecionar e integrar a qualidade do solo e indicadores em um único índice, servir como uma ferramenta de gestão para fornecer aos gestores de terras com todas as informações mais importantes para facilitar tomada de decisão em questões de manejo de agroecossistemas (Andrews; Carrol, 2001). IQS são calculados a partir de um conjunto mínimo de dados (CMD) formado por indicadores selecionados por meio de métodos estatísticos de um conjunto inicial de atributos das propriedades físicas, químicas e biológicas do solo. A análise de componentes principais (PCA) é o método estatístico mais utilizado para selecionar os indicadores em um CMD (Bünemann *et al.*, 2018). No entanto, observou-se em alguns casos que esta técnica é incapaz de selecionar indicadores que reflitam as diferenças na qualidade de solos sob diferentes sistemas de manejo (Askari; Holden, 2014; Hamidi Nehrani *et al.*, 2020).

Os cálculos de IQS são complexos e apresentam algumas limitações, nesse sentido os estudos tem buscado cada vez mais a concepção de IQS confiáveis utilizando-se da comparação de vários métodos disponíveis (Mukherje; Lal, 2014). Assim, o embasamento deste estudo foi comparar os IQS estimados por duas vias diferentes usando duas campanhas de amostragem em dois anos, analisando-se também a possibilidade de treinamento e validação dos dados utilizando anos distintos. Outro caminho de inovação nesse estudo foi avaliar a hipótese de que IQS estimados possam ter relação com o crescimento vegetal, utilizando proposta com uso de plantas bioindicadoras. Para isso foi escolhido a alface, como uma planta indicadora, que apresenta crescimento rápido, e tem sido indicada como espécie vegetal com grande capacidade de monitoramento, devido à sua capacidade de fornecer informações quantitativas sobre a qualidade do meio ambiente (Markert,

2007), sendo por isso que as organizações internacionais a recomendam para testes que avaliam a toxicidade de poluentes ambientais (Ocde, 2006; ISO, 2012)

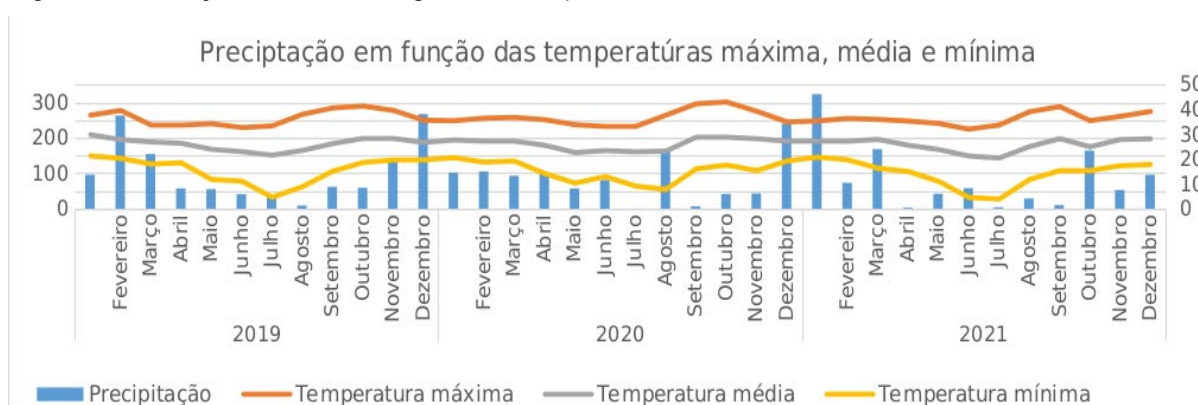
A validação do IQS pelo rendimento de culturas ainda é escassa, já que a maioria dos estudos se concentram em atributos ambientais do solo como variável fim, com avaliação apenas da camada superficial do solo (Fernandes *et al.*, 2011; Navas *et al.*, 2011). No entanto, há outro desafio em curso nesse estudo que é também de validar o IQS a partir de índices calculados apenas com dados de superfície (0-10 cm), isto pode fugir um pouco da realidade entre a qualidade do solo e rendimento da cultura porque o sistema radicular pode se estender para camadas mais profundas (Brady; Weil, 1984). Sendo que a maioria dos estudos já apresentados visam a obtenção de IQS a partir avaliações em camadas mais profundas do solo (Mukherje; Lal, 2014; Martín-Sanz *et al.*, 2022; Andrews; Karlen; Mitchell, 2002). Assim, o objetivo deste estudo foi calcular o IQS de área de pastagens cultivadas com diferentes manejos, usando IQS a partir de análises multivariadas, utilizando-se atributos físicos, químicos e biológicos, com validação pelo desenvolvimento de planta indicadora (Monteiro; Gama-Rodrigues *et al.*, 2004; Iwata *et al.*, 2010).

Material e Método

Área de Estudo

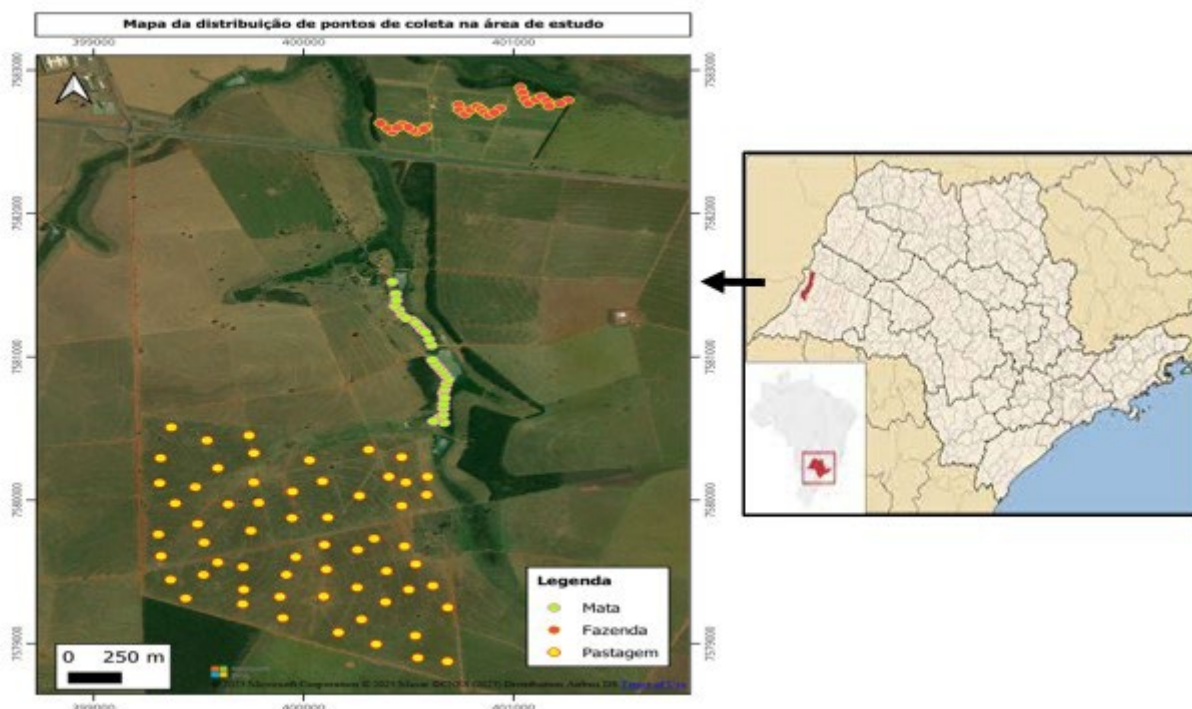
As áreas de estudo selecionadas para coleta de amostras de solo são localizadas no município de Caiuá, no Estado de São Paulo (21°53'06.2"S 51°57'56.4"W e 21°51'23,3"S 51°57'42,5"W). Na condução do estudo foram realizadas duas campanhas de coletas de solo anuais efetivadas nos meses de outubro de 2019 e setembro de 2021. Os dados mensais referentes às temperaturas e precipitação pluvial foram coletados durante a condução do estudo e estão apresentados na figura 1.

Figura 1 - Condições climáticas registradas no período do estudo.



Foram estabelecidas quatro áreas com manejo distinto para avaliação no estudo. Três áreas com pastagens (*Urochloa brizantha*) e uma área de preservação permanente como referência. As áreas foram denominadas como de pastagem contínua com manejo de adubação e calagem (PAST CM), pastagem contínua sem manejo (PAST SM), pastagem em sistema de integração lavoura pecuária (ILP) e área de preservação permanente (APP) (Figura 1). As áreas de pastagens e integração estão estabelecidas nas áreas há mais de seis anos. Todas áreas foram corrigidas e adubadas na implantação, na área de pastagem com adubação e manejo (PAST CM), foi realizada correção do solo e adubação em 2018, de acordo com recomendado por Gosai *et al.* (2021) e Pavanellie Araújo (2010). Em cada área de manejo foram estabelecidos 30 pontos de coleta georeferenciados, tendo no total 120 pontos de coleta de amostra de solo no total que foram avaliados em 2019 e 2021.

Figura 9 - Localização das áreas selecionadas no município de Caiuá, SP, sendo em amarelo pastagem com manejo e sem manejo, em verde área de APP e vermelho ILP.



A área de estudo está localizada a 330 m de altitude e possui clima classificado como clima tropical com estação seca (Aw), de acordo com a classificação de Köppen-Geiger (1936). Média anual temperatura é de 25° C e a precipitação total anual é de 1154 mm. O solo da área experimental foi classificado com Latossolo Vermelho distrófico textura arenosa (Santos *et al.*, 2022). As amostras de solo foram coletadas com trado manual profundidade de 0 a 10 cm. Em cada área foram georeferenciados 30 pontos para coleta de amostras, sendo que cada amostra, cerca de 2 Kg de solo, foi resultante da coleta de cinco subamostras para melhor homogeneização e representação do ponto.

Análises de atributos químicos, físicos e biológicos do solo

Para garantir a construção de índices de qualidade de solo representativos com a maior capacidade possível para discriminar entre usos, os solos foram caracterizados conforme descrito abaixo por meio de 26 atributos, químicos biológicos e físicos relacionados com a funcionalidade do solo. Nesse sentido foram selecionados dez atributos de fertilidade de solo (químicos); dez atributos

microbiológicos do solo e seis atributos físicos do solo. Totalizando no final 26 atributos de solo analisados. Também foi utilizada alface, como indicadora para avaliação da qualidade do solo.

Análises químicas

As amostras de solo coletadas foram fracionadas e parte delas foram encaminhadas ao laboratório para análises de fertilidade do solo (Malavolta *et al.*, 1997), sendo determinados: pH em Cloreto de Cálcio (CaCl₂), Matéria Orgânica, Fósforo (P), Acidez Potencial (H+Al), Potássio (K), Cálcio (Ca), Magnésio (Mg), Soma de Bases (SB), Capacidade de Troca Catiônica (CTC) e Porcentagem de Saturação por Bases (V%).

Análises microbiológicas

Parte das amostras foram encaminhadas para o laboratório, onde passaram por peneiramento e adequação de umidade para condição de capacidade de campo e seguida foram destinadas, para análises biológicas de: carbono e nitrogênio da biomassa microbiana (Kloeppe *et al.*, 1999; Ferreira *et al.*, 1999; Tedesco *et al.*, 1995); atividade enzimática da desidrogenase (Van Os; Ginkel, 2001) fosfatase ácida, glicosidase e arilsulfatase: (Tabatabai, 1994); respiração basal do solo (Jenkinson; Powlson, 1976); atividade da hidrólise de diacetato de fluoresceína (Chen *et al.*, 1988), quociente metabólico (calculado a partir da razão entre o C-CO₂ liberado pela respiração do solo pelo carbono da biomassa microbiana) (Anderson; Domsch, 1993) e microbiano (calculado a partir da relação entre o carbono da biomassa microbiana com carbono orgânico total do solo) (Agertt *et al.*, 2016; Sparling, 1992; Zinn *et al.*, 2002).

Análises físicas

Parte das amostras coletadas foram processadas pelo fracionamento granulométrico realizado por meio de peneiramento e/ou sedimentação. A granulometria do solo foi determinada utilizando-se o método da pipeta, conforme

descrito em Embrapa (1997). Foram determinados o teor de argila (partículas com diâmetro < 0,002mm), silte (partículas com diâmetro entre 0,002 e 0,05mm) e areia (partículas com diâmetro entre 0,05 e 2mm).

Parte das amostras foram encaminhadas para determinação da estabilidade dos agregados, sendo utilizado o método descrito por Kemper (1986) e Chepil (1965), com modificações propostas por Silva e Mielniczuk (1997) (Silva *et al.*, 2021). Esse método consiste inicialmente no peneiramento do solo para excluir fragmentos vegetais, outros resíduos, pedras e brita retida na peneira, de modo que o volume total da amostra ultrapasse a malha de 9,51 mm. Para obtenção dos agregados e determinação da estabilidade, utilizou-se o método descrito por Kemper e Chepil (1965), com alterações propostas por Silva e Mielniczuk (1997). A separação dos agregados em classes de tamanho foi feita pela desagregação e peneiramento em meio úmido. Foram obtidas três classes de agregados: macroagregados (diâmetro > 2,00mm), mesoagregados (diâmetro entre 2,00 e 0,25mm) e microagregados (diâmetro entre 0,25 e 0,05mm). Para as determinações dos teores de C total, uma alíquota de cada classe de agregados do solo obtida foi totalmente moída e passada em peneiras de malha de 0,150mm, com posterior pesagem de 0,150g dessa alíquota para análise em analisador elementar automático (Barros *et al.*, 2010).

A determinação dos teores de carbono particulado (COP), associado a mineral (COM) e carbono total (COT) em cada classe de agregados é importante para avaliar o sequestro de carbono no solo e entender como a matéria orgânica é distribuída em diferentes frações do solo. A análise foi realizada no analisador automático elementar, que é um equipamento que utiliza a técnica de combustão para determinar a concentração de elementos químicos presentes em uma amostra (Eivazi; Mullings; Banks, 2018).

Análise com planta indicadora

Para esta avaliação foi realizado bioensaio utilizando a alface (*Lactuca sativa*) com planta indicadora de qualidade do solo (referencia). Parte das amostras de solo (300 g) foram acondicionadas em vasos com a finalidade de avaliar a emergência e desenvolvimento da alface. Os vasos foram alocados dentro de

câmara de crescimento de plantas (Fitotron® SGC 120; Weiss Technik UK Ltd., Loughborough, UK) por 20 dias a 29/23 ° C sob um ciclo de 14 h de luz / 10 h de escuridão e com manutenção de 60% de umidade relativa no seu interior. O solo foi umedecido previamente e foram semeadas dez sementes por vaso. Após sete dias fez-se a avaliação de emergência de plantas e deixou-se uma planta por vaso que foram conduzidas na câmara durante 20 dias. Após esse período a planta (Alface) foi analisada pelas produções de massa fresca e massa seca da raiz e parte aérea. A variedade escolhida foi a Alface Crespa Lucy Brown.

Os dados foram tabulados na sequência dos resultados das análises químicas, físicas, biológicas atribuído a sua área PAST CM, PAST SM, ILP e APP, nos dois anos avaliados.

Determinação do índice de qualidade do solo (IQS)

O conjunto inicial de 120 pontos de amostragem foi avaliado em 2019 e 2021 de forma independente. O IQS foi determinado seguindo três etapas: 1) os indicadores representativos no conjunto mínimo de dados (CMD) foram selecionados do conjunto total de atributos do solo analisados; 2). Os indicadores selecionados (CMD) foram transformadas em escores utilizando-se transformação linear e não linear (sgmoidal); e 3) os indicadores foram integrados para formam o índice de qualidade do solo (IQS). Também foi feita uma abordagem denominando o conjunto de dados de 2019, com 120 amostras como de calibração definindo-se o CMD escores e IQS, e outro conjunto de 2021, com resultados nas 120 amostras como de verificação utilizando o modelo de 2019 (Masto *et al.*, 2007).

Seleção de indicadores: Conjunto mínimo de dados (CMD)

Duas diferentes abordagens foram usadas em todas as 120 amostras no conjunto de calibração, independentemente do tipo de uso da área, para selecionar as variáveis que compõem o MDS: i) componente principal (PCA), pois esta é a metodologia mais utilizada (Bhunja *et al.*, 2021; Askari e Holden, 2014), e ii) a mais nova análise de rede (NTA). A seleção do CMD foi obtida a partir de análise multivariada com determinação de componentes principais (PCA) de acordo com

Andrews e outros (2002), considerando o componente principal (PC) com autovetor $\lambda > 1$ e que explicam pelo menos 5 % da variância acumulada total. A matriz de componentes foi obtido via rotação Varimax, que minimiza o número de atributos com valores de carregamento mais altos em cada PC facilitando a interpretação (Peris *et al.*, 2008). Os atributos foram selecionados a partir de seu autovalor absoluto em cada PC escolhido, juntamente com outros atributos que diferiam desse valor em 10 % (Lenka *et al.*, 2022; Andrews *et al.*, 2002). Para reduzir a redundância dos atributos no CMD, quando vários atributos em um PC preencheram as condições anteriores foi utilizada análise de correlação de Pearson ($P < 0,05$) para eliminar atributos correlacionados (Chattopadhyay; Hassanzadeh; Pasha, 2020). No caso de correlações, os atributos com os maiores valores de carga foram selecionados como indicadores, e na ausência de correlações entre os atributos no mesmo PC, todas as variáveis foram selecionadas como indicadores (Caldeira *et al.*, 2000).

Transformação dos indicadores do CMD

Foram utilizadas duas transformações diferentes no CMD: lineares (L) e não lineares (NL); empregadas para reduzir o efeito de escala entre os diferentes indicadores do CMD e normalizar todos os seus valores para o intervalo [0,1]. A transformação linear (L) (Andrews *et al.*, 2002) do CMD considerara se o valor mais alto do indicador seja benéfico para a qualidade do solo (“mais é melhor”), caso em que a Equação (1) foi usado; e se um valor menor do indicador fosse benéfico para a qualidade do solo (“menos é melhor”), então a Equação (2) foi usada. Para indicadores com um faixa ótima (“mid-point”), a transformação foi construída usando As Equações (1) e (2) são normalizações e igualando os valores dos indicadores na faixa ideal para um valor de 1.

$$SL = \frac{x - x_{Min}}{x_{Max} - x_{Min}} \quad (1)$$

$$SL = 1 - \frac{(x - x_{Min})}{(x_{Max} - x_{Min})} \quad (2)$$

Em ambas as equações, x é o valor do indicador, x_{Max} é o valor máximo do indicador em todas as amostras analisadas, e x_{Min} é o valor mínimo do indicador em todas as amostras analisadas.

Para transformação não linear (NL) utilizou-se o procedimento de obtenção de Curvas sigmoidais com os resultados dos atributos escolhidos no CMD de acordo com recomendado por Bastida *et al.*, (2006) e Hussain *et al.*, (1999) seguindo a Equação (3):

$$\text{SNL} = a / (1 + (x/x_0)^b) \quad (3)$$

Sendo o valor máximo obtido pela curva sigmoideal (neste estudo $a = 1$), x é o valor do indicador, x_0 é o valor médio deste indicador em todas as amostras analisadas, e b é um coeficiente igual a $-2,5$ para os indicadores “mais é melhor” e $2,5$ para o “menos é melhor”. Os indicadores com valores ótimos (“ponto médio”) foram transformados usando uma função por partes: um valor de pontuação de 1 para todos os valores do indicador (x) dentro da faixa ideal, e as transformações de pontuação “mais é melhor” ou “menos é melhor” foram usados quando os valores do indicador diferiram do ótimo.

Integração dos indicadores do CMD no SQI

Os escores lineares e não lineares dos indicadores foram integradas em um único índice de forma aditiva ponderada (SQI-W), seguindo a Equação (5):

$$\text{SQI-W} = \sum n W_i * S_i \quad (5)$$

Onde S_i é o escore linear (L) ou não linear (NL) transformada para cada indicador, n é o número de indicadores MDS e W_i é o peso de cada indicador. No caso do CMD selecionado pelo PCA, W_i é um valor entre 0 e 1 que corresponde ao peso de cada indicador calculado de acordo com Equação (6).

$$W_i = (\% \text{Var CP}_i) / (\% \text{Var Total}) \quad (6)$$

Onde $\% \text{Var CP}_i$ é a porcentagem de variância explicada pelo CP para indicador i , $\% \text{VarTotal}$ é a porcentagem da variância explicada por todos os CPs no CMD e n é o número máximo de CPs selecionados.

Validação do IQS

Os IQSs estimados a partir de três métodos diferentes foram validados pelo crescimento da alface em cada ano de avaliação calculando-se coeficientes de correlação (Andrews *et al.*, 2002). Para o caso da ausência de normalidade foi utilizado a correlação de Spearman ($p < 0,05$) como método não paramétrico.

Análise estatística

Yaseen (2021) testou todas as variáveis, sendo assim, nesta análise PCA foram testadas todas as variáveis quanto à normalidade e homogeneidade de variância por meio do teste de Kolmogorov-Smirnov, e avaliação visual de histogramas. Análise de variância (ANOVA) de um fator foi usada para estabelecer qualquer diferença entre os atributos do solo devido aos vários usos do solo, e para identificar se algum dos SQI distinguiu entre os usos do solo no estudo. As análises estatísticas foram feitas usando o software SPSS v.23.0.

Resultados

Os atributos do solo das quatro áreas de estudo, avaliadas em 2019 e 2021, foram agrupados pelas propriedades do solo, e apresentadas nas figuras 3, 4 e 5. Verificou-se mudanças consideráveis nos resultados de alguns atributos comparando-se 2019 com 2021.

Foi observado que nos atributos de fertilidade de solo os valores foram maiores na área de integração-Lavoura-Pecuária (ILP), destacando-se valores consideráveis em atributos como fósforo solúvel e soma de bases, comparando-se com os teres nos outros sistemas. A área de preservação permanente (APP) apresentou maior concentração de matéria orgânica e alumínio e menor valor de pH, cálcio e saturação de bases, principalmente, comparando-se com as áreas agrícolas nos dois anos avaliados (Figura 3).

Nos atributos de microbiologia do solo destaca-se os maiores valores de arilsulfatase na APP e de desidrogenase nas pastagens, nos dois anos avaliados. Foi também observado que o coeficiente metabólico aumentou bastante no segundo ano da avaliação realizada na área de preservação permanente. Foi verificado também aumento nas atividades de fosfatase e glicosidase nos tratamentos com pastagens, comparando-se 2019 com 2021 (Figura 4).

Constatou-se nos atributos de física do solo, maiores teores de argila nas áreas de pastagens. No tocante ao acúmulo de carbono orgânico no solo observou-se que nas áreas de APP houve acúmulo de carbono associado a minerais e particulados, destacando-se maior diferença em 2019.

Figura 3 - Atributos de fertilidade do solo em 2019 e 2021 nas áreas de estudo. APP: área de preservação permanente; ILP Integração lavoura pecuária; PAST CM: pastagem com manejo; PAST SM: pastagem sem manejo.

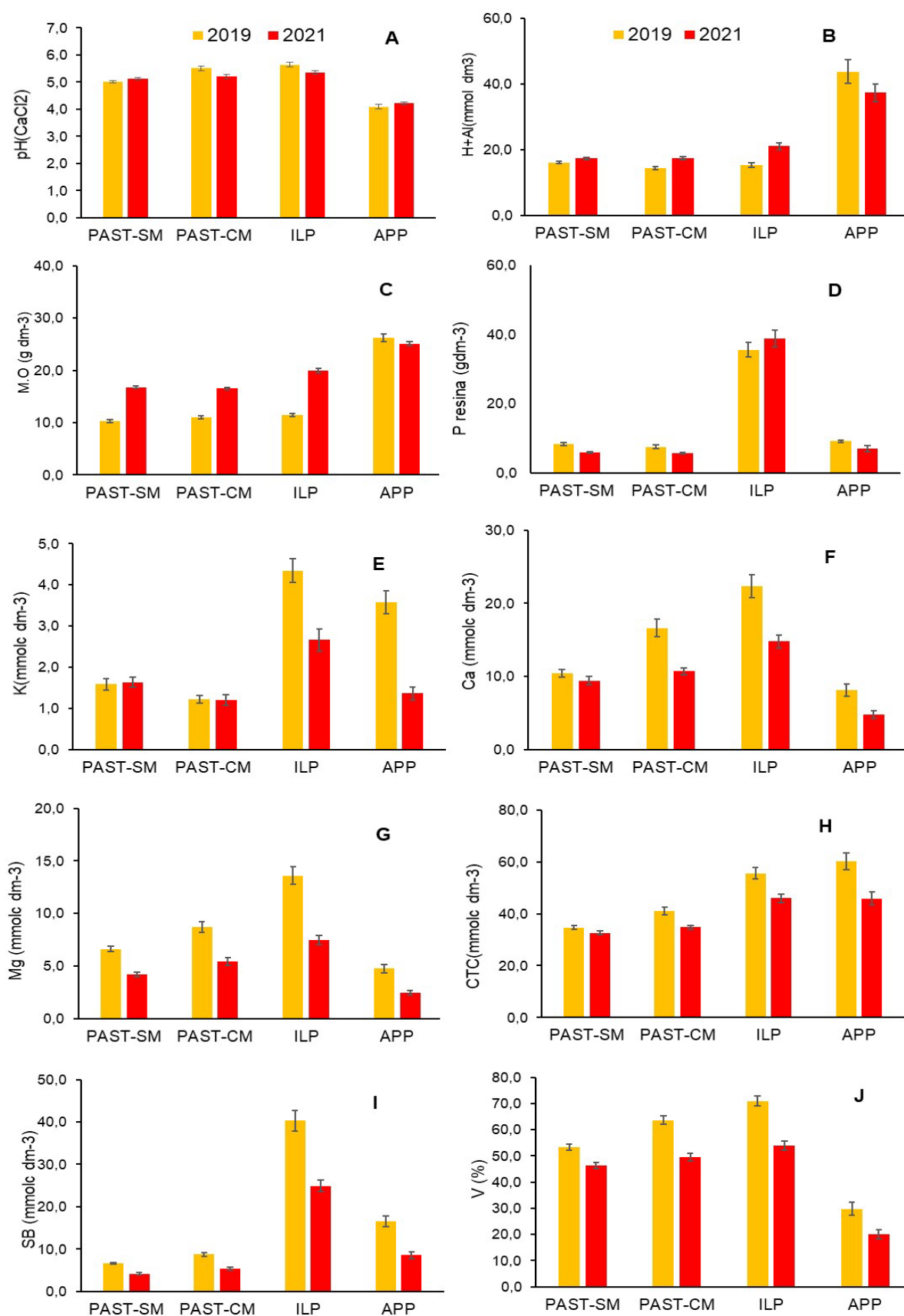


Figura 4 - Atributos de microbiologia do solo em 2019 e 2021 nas diferentes áreas de estudo. APP: área de preservação permanente; ILP Integração lavoura pecuária; PAST CM: pastagem com manejo; PAST SM: pastagem sem manejo.

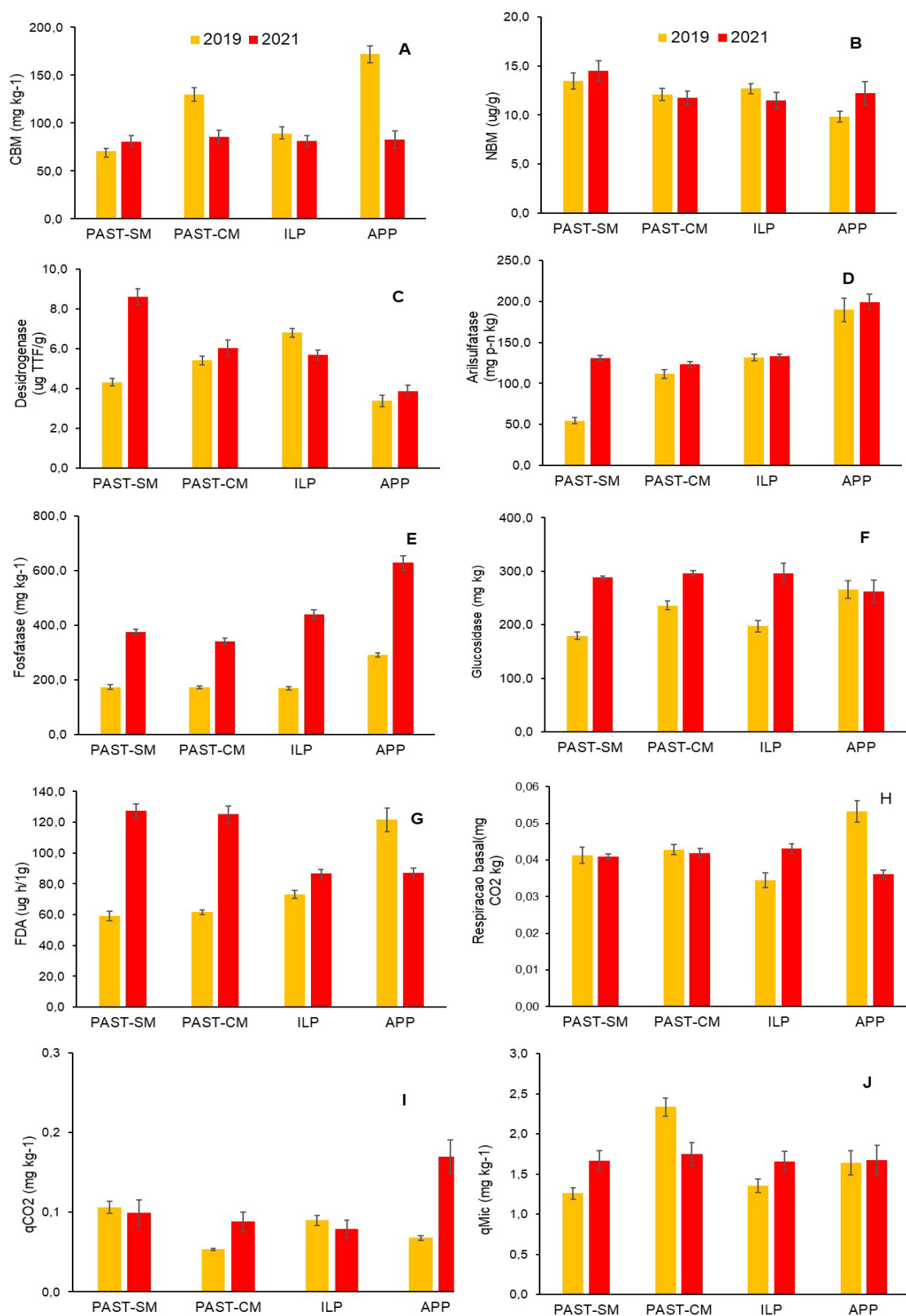
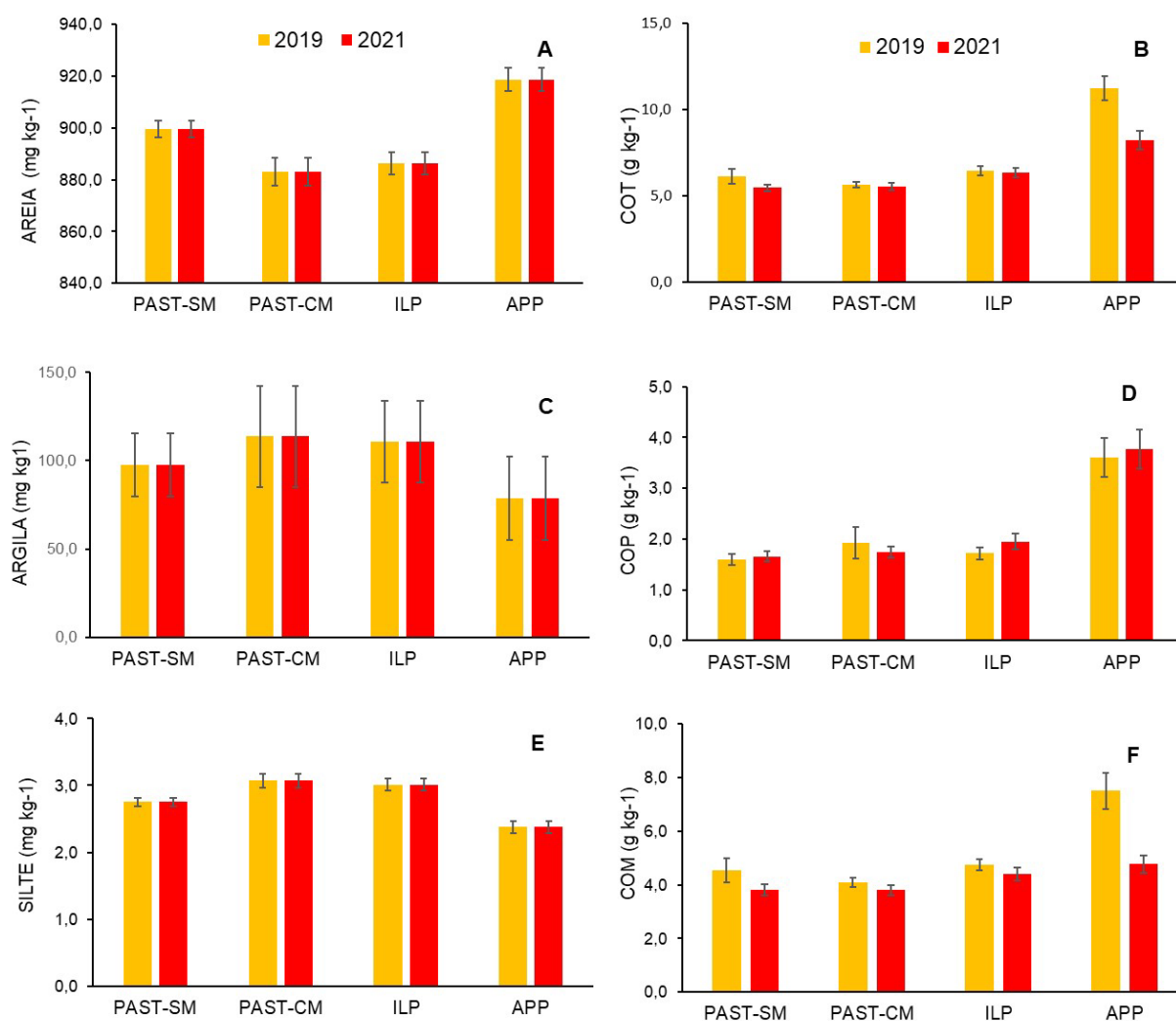
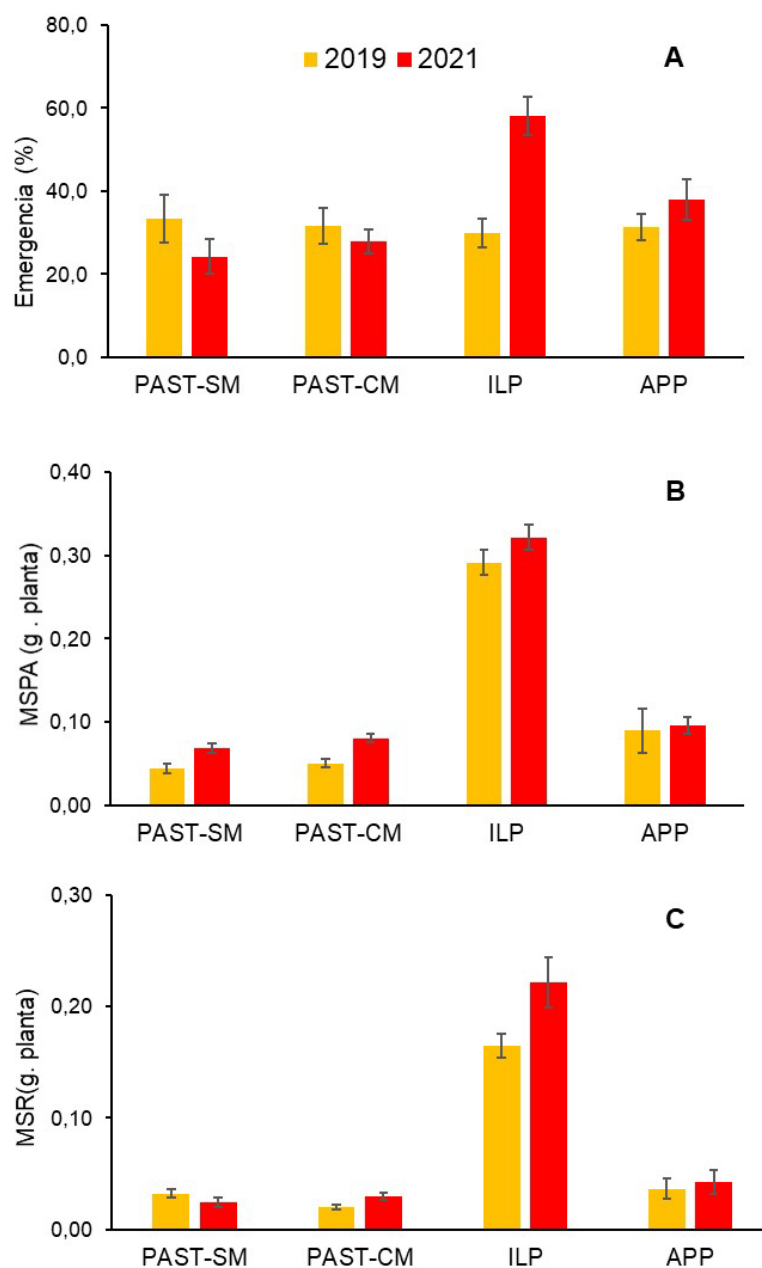


Figura 5 - Atributos físicos do solo em 2019 e 2021 nas diferentes áreas de estudo. APP: área de preservação permanente; ILP Integração lavoura pecuária; PAST CM: pastagem com manejo; PAST SM: pastagem sem manejo.



No tocante a avaliação da alface como planta indicadora, observou-se que todas as variáveis analisadas foram importantes para classificar as áreas, verificando-se maior crescimento da cultura na área de ILP e maior igualdade no desempenho das outras áreas avaliadas (Figura 5). A maioria dos tratamentos apresentou tendência de melhoria do crescimento da planta na segunda avaliação realizada, no ano de 2021.

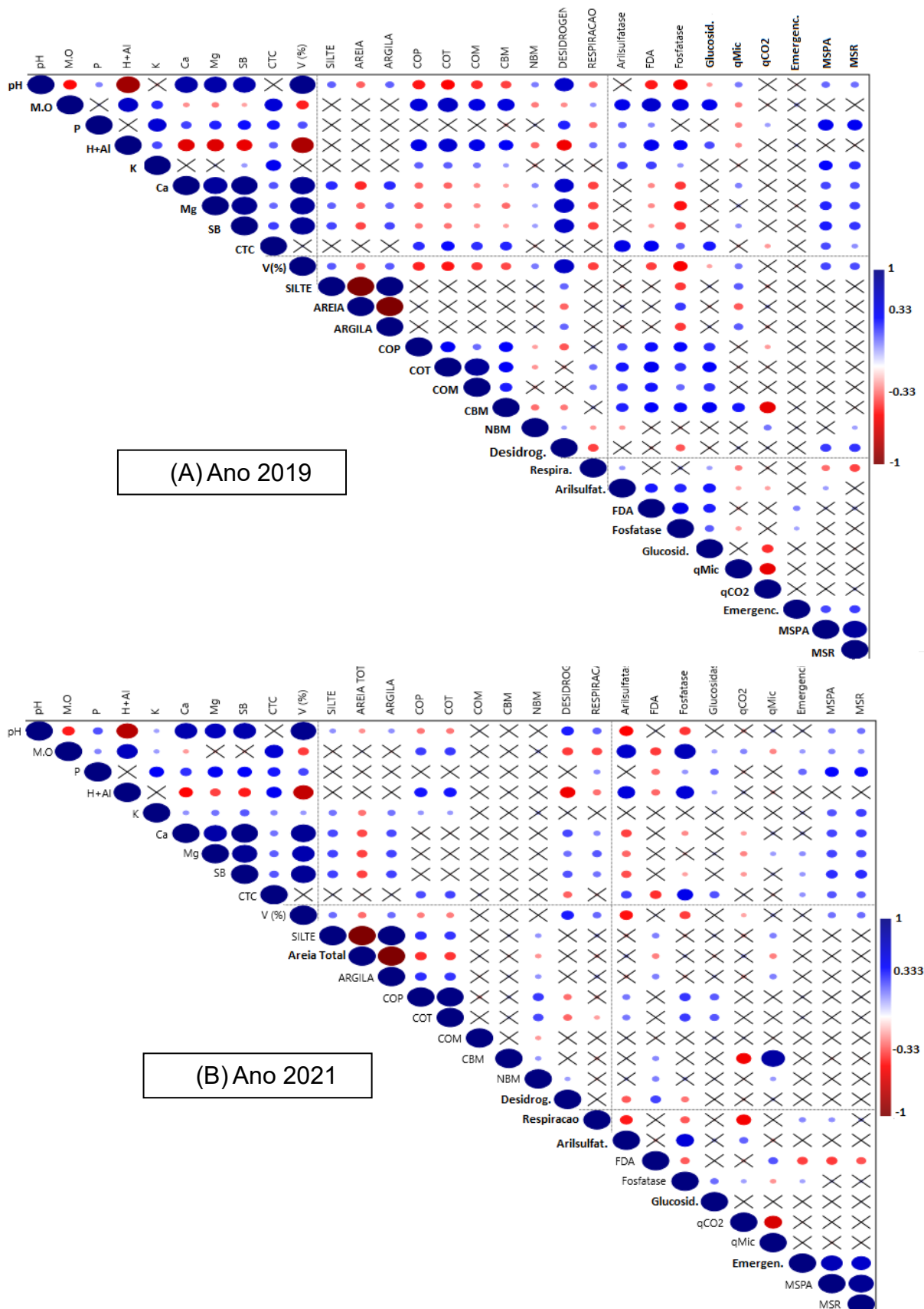
Figura 6 - Emergência de plantas, massa seca da raiz e parte aérea da alfaca em 2019 e 2021 nas diferentes áreas de estudo. APP: área de preservação permanente; ILP Integração lavoura pecuária; PAST CM: pastagem com manejo; PAST SM: pastagem sem manejo.



Na Figura 7 foram apresentadas as matrizes de correlação obtidas nos dois anos avaliados. As mesmas também foram importantes para seleção de atributos que apresentavam correlação significativa entre eles.

Na figura a seguir, apresenta-se as matrizes de correlação obtidas nos

Figura 7 - Matrizes de correlação obtidas nos dois anos avaliados "A" 2019 e "B" 2021, quanto mais azul mais próximo de 1, tem-se maior correlação entre as variáveis.



dois anos avaliados “A” 2019 e “B” 2021. As mesmas também foram importantes para seleção de atributos que apresentavam correlação significativa entre eles.

A partir das análises de componentes principais com os atributos do solo verificou-se que no primeiro ano de avaliação (2019) seis CPs foram obtidos no PCA com autovalores acima de 1 e variância de pelo menos 5%. Os atributos do solo com maior carregamento nos componentes principais, sem nenhuma correlação foram: matéria orgânica (CP 1), soma de bases (CP 2), teor de argila (CP 3). Coeficiente microbiano- q_{Mic} (CP 4), potássio (CP 5) e nitrogênio da biomassa microbiana (CP 6) (Tabela 1).

No segundo ano da avaliação (2021) sete CP foram obtidos com autovalores acima de 1 e variância de pelo menos 5%. Os atributos do solo com maior carregamento nos componentes principais, sem nenhuma correlação foram: matéria orgânica (CP 1), soma de bases (CP 2), teor de argila (CP 3). Carbono da biomassa microbiana (CP 4), diacetato de fluoresceína (CP 5), glucosidase (CP 6), e nitrogênio da biomassa microbiana (CP 7) (Tabela 2).

Em resumo, os pesos dos indicadores selecionados pelo CMD - PCA calculados de acordo a variância total e de cada CP conforme descrito na Equação (6) são mostrados na Tabela 1 e 2 e foram subsequentemente usados para calcular o IQS. O resultado de todo esse processo foi um total de seis SQI, diferenciados com base sobre o tipo de transformação (uma por dados lineares e outra por não lineares) e pela validação de 2019 por 2021 (Figuras 7 a 9).

Tabela 1 - Resultados da análise de componentes principais (PCA) de 2019

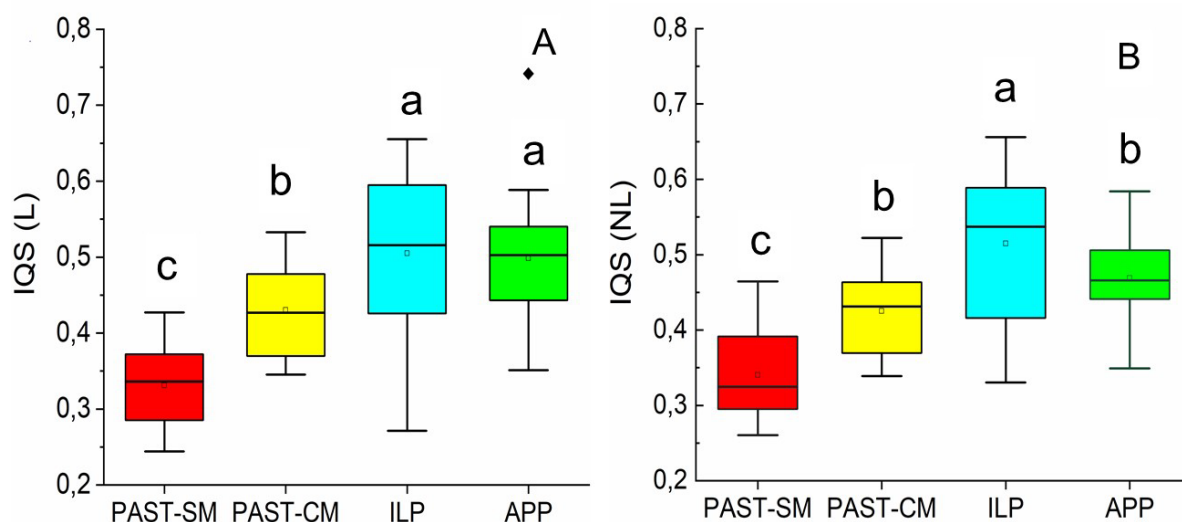
Atributos do solo	Componente principal					
	CP 1	CP 2	CP 3	CP 4	CP 5	CP 6
pH	-0.56	0.75				
M.O	<u>0.89</u>					
P		0.68				
H+Al	0.86					
K					<u>0.82</u>	
Ca		0.93				
Mg		0.95				
SB		<u>0.95</u>				
CTC	0.82					
V (%)	-0.56	0.79				
SILTE			0.97			
AREIA			-0.97			
ARGILA			<u>0.98</u>			
COP	0.57					
COT	0.83					
COM	0.71					
CBM	0.68			0.57		
NBM						<u>0.87</u>
DHA		0.75				
Resp						
ARIL	0.74					
FDA	0.77					
FOSF	0.68					
GLUC	0.82					
qMic				<u>0.94</u>		
qCO2				-0.63		
Valor próprio	7.55	5.88	3.2	2.05	1.44	1.33
Variância (%)	29.0	22.6	12.3	7.89	5.5	5.11
Variância acumulada(%)	29.0	51.7	63.98	71.9	77.4	82.5

Tabela 2 - Resultados da análise de componentes principais (PCA) de 2021

Atributos do solo	Componente principal						
	CP 1	CP 2	CP 3	CP 4	CP 5	CP 6	CP 7
pH	-0.54	0.73					
M.O	<u>0.94</u>						
P		0.70					
H+Al	0.82						
K							
Ca		0.91					
Mg		0.89					
SB		<u>0.96</u>					
CTC	0.81						
V (%)	-0.55	0.79					
SILTE			0.94				
AREIA			-0.94				
ARGILA			<u>0.95</u>				
COP						0.54	0.50
COT						0.55	
COM							<u>-0.74</u>
CBM				<u>0.97</u>			
NBM					0.62		
DHA					0.55		
Resp	-0.54						
ARIL	0.76						
FDA					<u>0.63</u>		
FOSF	0.84						
GLUC						<u>0.83</u>	
qMic				-0.72			
qCO2				0.96			
Valor próprio	5.54	5.07	3.35	2.41	1.77	1.47	1.32
Variância (%)	21.3	19.5	12.9	9.26	6.83	5.65	5.1
Variância acumulada(%)	21.3	40.8	53.7	62.9	69.8	75.4	80.5

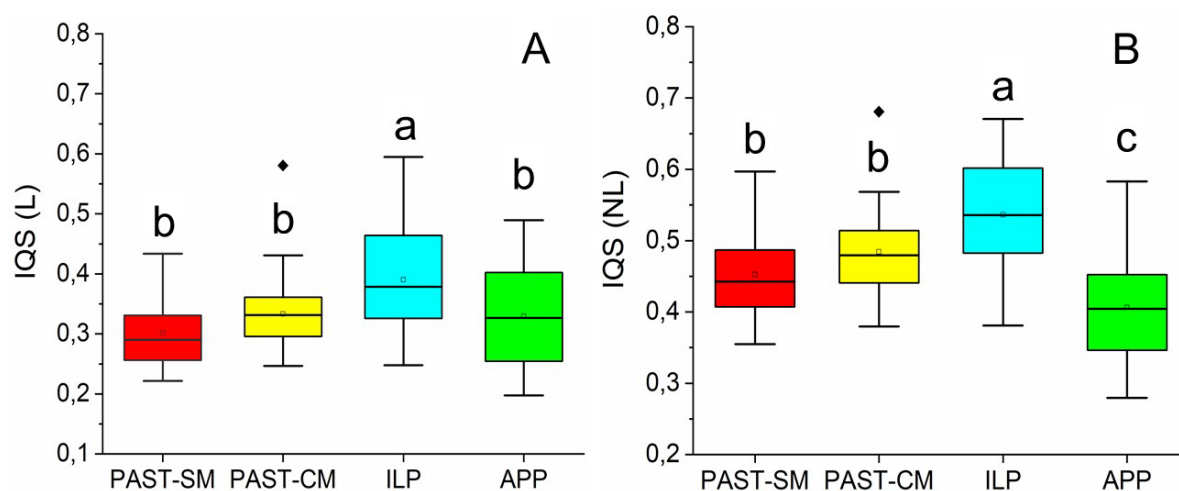
A partir da seleção de atributos pelo PCA e utilizando transformação dos escores no conjunto mínimo de dados estabelecidos verificou-se que na avaliação de 2019, quando se utilizou a transformação não linear dos dados no CMD, o IQS do ILP diferenciou-se do APP o que não ocorreu quando não foi executada a transformação (figura 8). A área de pastagem sem manejo foi classificada com a área de menor qualidade do solo.

Figura 8 - Índice de qualidade do solo (IQS), obtido em 2019, em áreas de pastagem obtido a partir de transformação linear (A) e não linear (B) de atributos do solo selecionados por PCA. APP: área de preservação permanente; ILP Integração lavoura pecuária; PAST



Em 2021 ocorreu a diferenciação do IQS do ILP das outras áreas pelos dois métodos de transformação dos dados, mas as áreas de pastagens perenes tiveram desempenho semelhante, entretanto foram diferenciadas da área de preservação quando se utilizou a transformação não linear (Figura 9).

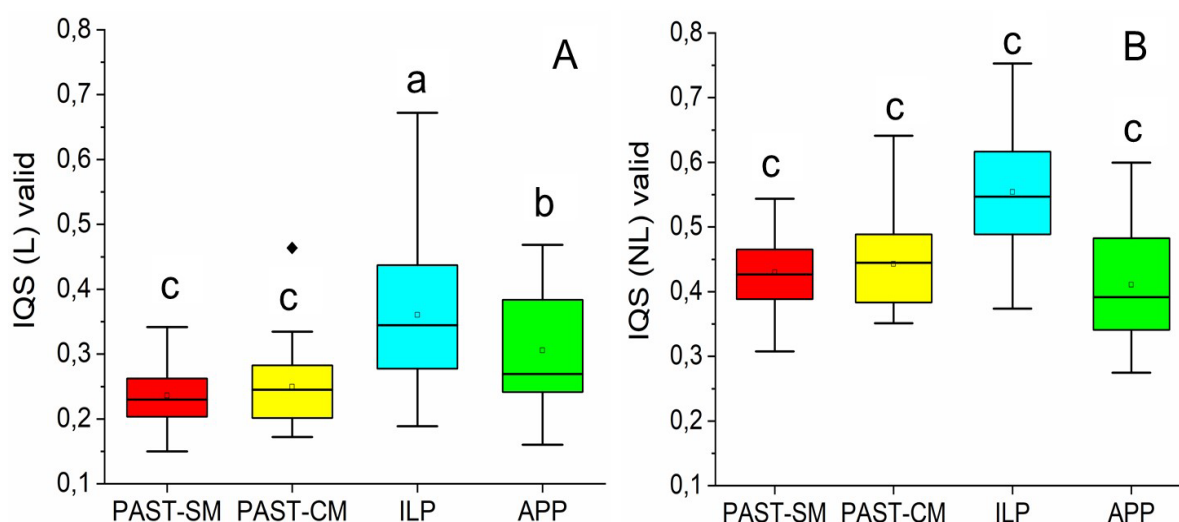
Figura 9 - Índice de qualidade do solo (IQS), obtido em 2021, em áreas de pastagem obtido a partir de transformação linear (A) e não linear (B) de atributos do solo selecionados por PCA. APP: área de preservação permanente; ILP Integração lavoura pecuária; PAST



Utilizando-se a validação de dados de 2021 pelo IQS de 2019 verificou-se que também ocorreu classificação das áreas de forma semelhante, apontado a área de ILP ainda como a de melhor desempenho, contudo verificou-se que a área de preservação se destacou como de melhor desempenho na comparação com áreas de pastagem perene pelo método linear (Figura 10).

A APP: área de preservação permanente; ILP Integração lavoura pecuária; PAST CM: pastagem com manejo; PAST SM: pastagem sem manejo no IQS NL, não houve diferença estatística, porém no IQS L, houve diferença em PAST CM, ILP e APP (Figura 10).

Figura 10 - Índice de qualidade do solo (IQS), obtido em 2019 com validação em 2021 em áreas de pastagem obtido a partir de transformação linear (A) e não linear (B) de atributos do solo selecionados por PCA com validação do IQS de 2019 em 2021.



A validação dos IQS pelo desenvolvimento da alface também mostrou que o IQS calculado pela maioria dos métodos utilizados foi efetivo na correlação dos valores encontrados (Tabela 3). Mas pode ser destacado que o método da transformação dos dados de forma não linear foi mais adequado para todos os modelos empregados com valores de p com alta significância ($p < 0,001$) na validação de 2021.

Tabela 3 - Coeficientes de correlação de Spearman (r_2) dos diferentes índices de qualidade do solo (IQS) versus desenvolvimento da massa seca da raiz (MSR) e parte aérea (MSPA) da alfaca.

	MSPA 2019	MSR 2019	MSPA 2021	MSR 2021
IQS L 2019	0.32*** (1)	0.15 ^{n.s.}	0.37***	0.33***
IQS NL 2019	0.35***	0.19*	0.40***	0.40***
IQS L 2021			0.16 ^{n.s.}	0.21*
IQS NL 2021			0.21*	0.26**
IQS L VALID 2021			0.26**	0.27**
IQS NL VALID 2021			0.30***	0.33***

(1) n.s.- não significativo; * $p < .05$; ** $p < .01$; *** $p < .001$

Discussão

Verificou-se mudanças consideráveis em alguns atributos na comparação dos resultados em cada ano avaliado (Figuras 3 a 5). Isto pode ser explicado em parte as condições climáticas que diferiram bastante em cada período avaliado, destacando-se que no inverno 2021 teve a ocorrência de geadas nas áreas avaliadas, o que não ocorreu em 2019. Verificou-se alterações consideráveis em alguns atributos, principalmente no tocante as perdas verificadas nos teores de carbono orgânico no solo da área de preservação. Isto pode ser em parte confirmado pelo aumento em cerca de 100% no coeficiente metabólico, mostrando que área estava com maior estresse resultando em maior perda de carbono para a atmosfera. O carbono orgânico relacionado aos minerais foi mais afetado com redução em 2021, mostrando que mesmo sendo mais estável que o particulado pode ser utilizado pelos microrganismos do solo em condições de estresses, principalmente na camada de 0-10 cm (Evangelista *et al.*, 2013; Melo *et al.*, 2016).

Os resultados encontrados nesse estudo confirmam a eficácia de se utilizar índices calculados por modelos estatísticos como o PCA para diferenciar áreas agrícolas com diferentes manejos (Murkherje; Lal, 2014; Karlen *et al.*, 1997; Andrews; Karlen; Mitchell, 2002). Salientando-se que um dos objetivos deste estudo foi verificar a possibilidade de usar um IQS calculado anteriormente (2019) como treinamento e validação do mesmo com dados coletados em 2021 nos mesmos

pontos dentro das áreas. Destacando-se que já foi feita essa abordagem de treinamento e validação de dados para o IQS, usando aprendizagem de máquina, mas utilizando o conjunto de dados obtido num único ano de coleta (Martín-Sanz *et al.*, 2022). Verificou-se que o IQS obtido pela validação pode ser utilizado como estratégia para avaliação de solos com pastagens em sistemas, com condições edafoclimáticas semelhantes, entretanto foi observado que fatores climáticos podem causar alterações consideráveis, que precisam também serem objeto de avaliação em novos estudos, pois podem interferir nos resultados dos modelos empregados.

Para determinar a adequação deste tipo de abordagem para esta tarefa, a classificação de atributos, indicadores e SQI obtidos por PCA foram comparados. O PCA em 2019 identificou seis indicadores, em ordem de maior para menor importância considerando a porcentagem de variância explicada: M.O, S.B., ARGILA, q_{Mic} , K, e NBM. Já em 2021 foi selecionado sete indicadores pelo PCA: M.O, S.B., ARGILA, CBM, FDA, GLUC e COM. Verificou-se que matéria orgânica tem um papel fundamental na qualidade do solo devido à sua influência em um amplo conjunto de propriedades biológicas, físicas e químicas do solo (Hamidi Nehrani *et al.*, 2020), também se destaca a soma de bases que representa um grande compartimento na fertilidade do solo, relacionado principalmente com o equilíbrio iônico no solo.

Os modelos empregados nos dois anos de avaliação foram eficazes para classificar e diferenciar estatisticamente as áreas com diferentes manejos chegando-se a IQS com validação pelo uso do crescimento da alfaca como planta indicadora, sabe-se que essa cultura é muito sensível a mudanças ambientais sendo utilizadas também, como referência para estudo ecotoxicológicos (Markert, 2007). Para o emprego da técnica de bioensaio faz-se necessária a seleção de espécies vegetais que exibam ou alteração do acúmulo de matéria seca e do crescimento inicial (Marchesan *et al.*, 2011; Diesel *et al.*, 2012). Diante destes resultados obtidos e com a maior rapidez para realização destes bioensaios, pode então ser sugerido o uso de plantas bioindicadoras, com a alfaca, para validação ambiental e agrônômica de outros índices de qualidade de solo propostos em novos estudos.

Os IQS obtidos conseguiram diferenciar as áreas com índices de qualidade distintos, mas abaixo de 0,6. De acordo com Fathizad *et al.* (2020) áreas classificadas por IQS baseadas em PCA são consideradas boas quando apresentam

valores acima de 0,6 e são consideradas muito fracas quando apresentam valores abaixo de 0,4. Pode ser destacado que área de integração lavoura pecuária obteve índice de qualidade próximo de 0,6. Isto confirma que o sistema de integração pode ser o mais recomendado para o cultivo de pastagens em solos arenosos nas regiões tropicais, pois consegue manter a qualidade do sistema com valores mais elevados quando comparados com o sistema de cultivo perene. Em estudo comparativo com área de pastagem introduzida na região amazônica, verificou-se que o sistema de integração lavoura pecuária floresta (ILPF) também manteve o IQS acima de 6 semelhante a área de mata nativa.

Verifica-se nesse estudo que o IQS calculado foi validado pela planta indicadora, utilizada como parâmetro de cultivo agrícola e ambiental. Deve-se sempre atentar para a variabilidade de dados e influência das condições climáticas nas avaliações de campo. Considerando que este estudo avaliou e validou o IQS em dois anos, com desempenho satisfatório, pode ser indicado que a estratégia do modelo empregado abre caminho para novos estudos na área de qualidade de solo, buscando facilitar a classificação das áreas com métodos menos onerosos para o produtor rural.

Conclusões

O atributos químicos e biológicos do solo apresentam relativa variabilidade na avaliação temporal com influência do manejo aplicado e condições climáticas.

O IQS foi afetado pelo manejo de solo nas diferentes áreas de pastagens. Indicando o sistema ILP como de maior valor obtido nas duas avaliações.

O IQS calculado usando métodos não lineares foram mais significativamente correlacionados com o desenvolvimento da alface, indicando que é relativamente pouco utilizado porém simples utilização.

A estratégia de usar o IQS validado pode ser considerado como o melhor e mais fácil modelo devido ser considerado como um procedimento relativamente menos dispendioso de tempo.

REFERÊNCIAS

- ALVES, T. D. S.; CAMPOS, L. L.; ELIAS NETO, N.; MATSUOKA, M.; LOUREIRO, M. F. Biomassa e atividade microbiana de solo sob vegetação nativa e diferentes sistemas de manejos. **Acta Scientiarum. Agronomy**, v. 33, p. 341-347, 2011. <https://doi.org/10.4025/actasciagron.v33i2.4841>
- ANDREWS, S. S.; FLORA, C. B.; MITCHELL, J. P.; KARLEN, D. L. Growers' perceptions and acceptance of soil quality indices. **Geoderma**, p. 114, n. 3-4, p. 187-213, 2003. [https://doi.org/10.1016/S0016-7061\(03\)00041-7](https://doi.org/10.1016/S0016-7061(03)00041-7)
- ANDREWS, S.; KARLEN, D.; MITCHELL, J. A comparison of soil quality indexing methods for vegetable production systems in Northern California. **Agriculture, Ecosystems & Environment**, v. 90, p. 25–45, 2002.
- AGERTT, V. A.; BONEZ, P. C.; ROSSI, G. G.; FLORES, V. D. C.; SIQUEIRA, F. D. S.; MIZDAL, C. R.; CAMPOS, M. M. A. Identification of antimicrobial activity among new sulfonamide metal complexes for combating rapidly growing mycobacteria. **BioMetals**, v. 29, p. 807-816, 2016. <https://doi.org/10.1007/s10534-016-9951-3>
- ASKARI, M. S.; HOLDEN, N. M. Indices for quantitative evaluation of soil quality under grassland management. **Geoderma**, v. 230, p. 131-142, 2014. <https://doi.org/10.1016/j.geoderma.2014.04.019>
- BRADY, N. C.; WEIL, R. R. **The nature and properties of soils**. New York:Macmillan, 1984.
- BARROS, Y. J.; MELO, V. D. F.; DIONÍSIO, J. A.; OLIVEIRA, E. B. D.; CARON, L.; KUMMER, L.; SOUZA, L. C. D. P. Indicadores de qualidade de solos em área de mineração e metalurgia de chumbo: I-microrganismos. **Revista Brasileira de Ciência do Solo**, v. 34, p. 1397-1411, 2010. <https://doi.org/10.1590/S0100-06832010000100008>
- BHUNIA, S.; BHOWMIK, A.; MALLICK, R.; MUKHERJE, J. Agronomic efficiency of animal-derived organic fertilizers and their effects on biology and fertility of soil: a review. **Agronomy**, v. 11, n. 5, p. 823, 2021. <https://doi.org/10.3390/agronomy11050823>
- BOURSCHEIDT, M. L. B.; CARMO, K. B.; PEDREIRA, B. C.; BERBER, G. D. C. M.; FERREIRA, A. Biodiversidade de bactérias antagônicas a fungos associados à síndrome da morte do capim-marandu (*Brachiaria brizantha*). **Nativa**, v. 9, n. 5, p. 454-459, 2021. DOI: 10.31413/nativa.v9i5.12666
- BÜNEMANN, E. K.; BONGIORNO, G.; BAI, Z.; CREAMER, R. E.; DEYN, G.; GOEDE, R.; BRUSSAARD, L. Soil quality—a critical review. **Soil Biology and Biochemistry**, v. 120, p. 105-125, 2018. <https://doi.org/10.1016/j.soilbio.2018.01.030>
- CALDEIRA, M. V. W.; SCHUMACHER, M. V.; TEDESCO, N.; SANTOS, E. M. D. Ciclagem de nutrientes em *Acacia mearnsii* DE WILD. V. Quantificação do conteúdo de nutrientes na biomassa aérea de *Acacia mearnsii* DE WILD. procedênci

australiana. **Ciência Rural**, v. 30, p. 977-982, 2000. <https://doi.org/10.1590/S0103-84782000000600009>

CHAER, G. M.; MENDES, I. C.; DANTAS, O. D.; MALAQUIAS, J. V.; REIS JUNIOR, F. B.; OLIVEIRA, M. I. L.. Evaluating C trends in clayey Cerrado Oxisols using a four-quadrant model based on specific arylsulfatase and β -glucosidase activities. **Applied Soil Ecology**, v. 183, p. 104742, 2023. <https://doi.org/10.1016/j.apsoil.2022.104742>

CHEPIL, W. S. Transport of soil and snow by wind. *In: Agricultural meteorology* Boston, MA: American Meteorological Society, 1965. p. 123-132.

CHATTOPADHYAY, A.; HASSANZADEH, P.; PASHA, S. Predicting clustered weather patterns: A test case for applications of convolutional neural networks to spatio-temporal climate data. **Scientific Reports**, v. 10, n. 1, p. 1317, 2020. <https://doi.org/10.1038/s41598-020-57897-9>

DORAN, J. W.; PARKIN, T. B. Defining and assessing soil quality. **SSSA Special Publication**, v. 35, p. 3–3, 1994.

DIESEL, F.; TREZZI, M. M.; PAZUCH, D.; XAVIER, E.; ROSIN, D.; PAGNONCELLI, F. Seleção de espécies da família Cucurbitaceae e Chenopodiaceae para indicação da presença de saflufenacil no solo. **Pesticidas: Revista de Ecotoxicologia e Meio Ambiente**, 2012.

EMBRAPA. **Manual de métodos de análise do solo**. Rio de Janeiro: Embrapa, 1997. 212p.

EIVAZI, F.; MULLINGS, N.; BANKS, M. L. Effect of select surfactants on activities of soil enzymes involved in nutrient cycling. **Communications in Soil Science and Plant Analysis**, v. 49, n. 3, p. 371-379, 2018. <https://doi.org/10.1080/00103624.2018.1427263>

EVANGELISTA, C. R.; PARTELLI, F. L.; BRITO FERREIRA, E. P.; PIRES, F. R. Atributos microbiológicos do solo na cultura da cana-de-açúcar sob manejo orgânico e convencional. **Ciências Agrárias**, v. 34, n. 4, p. 1549-1562, 2013. DOI: <https://doi.org/10.5433/1679-0359.2013v34n4p1549>

FATHIZAD, H.; ARDAKANI, M.A.H.; HEUNG, B.; SODAIEZADEH, H.; RAHMANI, A.; FATHABADI, A.; SCHOLTEN, T.; TAGHIZADEH-MEHRJARDI, R. Spatio-temporal dynamic of soil quality in the central Iranian desert modelled with machine learning and digital soil assessment techniques. **Ecological Indicators**, v. 118, p. 106736, 2020.

FERNANDES, J. C.; GAMERO, C.A.; RODRIGUES, J.G.L.; MIRAS-AVALOS, J. M. Determination of the quality index of a Paleudult under sunflower culture and different management systems. **Soil and Tillage Research**, v. 112, p. 167–174, 2011.

GOSAI, D.; NAAYAK, R. JAYSWAL, H.; PATEL, A. Crop recommendation system using machine learning. **International Journal of Scientific Research in Computer Science, Engineering and Information Technology**, v. 7, n. 3, p. 558-569, 2021.

IBGE. **Monitoramento da cobertura de uso da terra do Brasil**. 2020. Disponível em https://www.ibge.gov.br/apps/monitoramento_cobertura_uso_terra/v1/#/home . Acesso em: 10 fev. 2024.

IWATA, B. D. F.; LEITE, L.; CAMPOS, L.; ARAUJO, A.; COSTA, C. D. N. **Dinâmica sazonal do carbono orgânico total e da biomassa microbiana em sistemas agroflorestais e agricultura de corte e queima no cerrado**. 2010. Disponível em: <https://www.embrapa.br/busca-de-publicacoes/-/publicacao/881151/dinamica-sazonal-do-carbono-organico-total-e-da-biomassa-microbiana-em-sistemas-agroflorestais-e-agricultura-de-corte-e-queima-no-cerrado>. Acesso em: 10 fev. 2024.

JENKINSON, D. S.; POWLSON, D. S. The effects of biocidal treatments on metabolism in soil—V: a method for measuring soil biomass. **Soil biology and Biochemistry**, v. 8, n. 3, p. 209-213 (1976). [https://doi.org/10.1016/0038-0717\(76\)90005-5](https://doi.org/10.1016/0038-0717(76)90005-5)

KARLEN, D. L.; DITZLER, C. A.; ANDREWS, S. S. Soil quality: why and how?. **Geoderma**, v. 114, n. 3-4, p. 145-156, 2003. [https://doi.org/10.1016/S0016-7061\(03\)00039-9](https://doi.org/10.1016/S0016-7061(03)00039-9)

KEMPER, W. D.; CHEPIL, W. S. Size distribution of aggregates. *In: Methods of soil analysis: part 1 physical and mineralogical properties, including statistics of measurement and sampling*. 1965. Cap.39, p. 499-510. <https://doi.org/10.2134/agronmonogr9.1.c39>

KLOEPPE, J. W.; RODRIGUEZ-KABANA, R.; ZEHNDER, A. W.; MURPHY, J. F.; SIKORA, E.; FERNANDEZ, C. Plant root-bacterial interactions in biological control of soilborne diseases and potential extension to systemic and foliar diseases. **Australasian Plant Pathology**, v. 28, n.1, p. 21-26, 1999. <https://doi.org/10.1071/AP99003>

KÖPPEN, W. **Classificação climática de Köppen-Geiger**. 1936.

LENKA, N. K.; MEENA, B. P.; LAL, R.; KHANDAGLE, A.; LENKA, S.; SHIRALE, A. O. Comparing four indexing approaches to define soil quality in an intensively cropped region of Northern India. **Frontiers in Environmental Science**, v. 10, p. 865473, 2022. <https://doi.org/10.3389/fenvs.2022.865473>

MALAVOLTA, E.; FAVARIN, J. L.; MALAVOLTA, M.; CABRAL, C. P.; HEINRICHS, R.; SILVEIRA, J. S. M. Repartição de nutrientes nos ramos, folhas e flores do cafeeiro. **Pesquisa agropecuária brasileira**, v. 37, p. 1017-1022, 2002. <https://doi.org/10.1590/S0100-204X2002000700016>

MARTÍN-SANZ, J.; SANTIAGO-MARTÍN, A.; VALVERDE-ASENJO, I.; QUINTANA-NIETO, J. R.; GONZÁLEZ-HUECAS, C.; LÓPEZ-LAFUENTE, A. L. Comparison of soil quality indexes calculated by network and principal component analysis for carbonated soils under different uses. **Ecological Indicators**, v. 143, 2022.

MARZAIOLI, R.; D'ASCOLI, R.; PASCALE, R. A.; RUTIGLIANO, F. A. Soil quality in a Mediterranean area of Southern Italy as related to different land use types. **Applied**

Soil Ecology, v. 44, n. 3, p. 205-212, 2010.
<https://doi.org/10.1016/j.apsoil.2009.12.007>

MARKERT, B. Definitions and principles for bioindication and biomonitoring of trace metals in the environment. **Journal of Trace Elements in Medicine and Biology**, v. 21, n. (suppl. 1), p. 77-82, 2007.

MARCHESAN, E. D.; DEDORDI, G.; REZZI, M. M.; VIDAL, R. A.; DICK, D. P. Seleção de espécies bioindicadoras para uso em bioensaios de lixiviação e persistência de atrazina no solo. **Pesticidas: Revista de Ecotoxicologia e Meio Ambiente**, n. 21, 2011.

MASTO, R. E.; CHHONKAR, P. K.; SINGH, D.; PATRA, A. K. Soil quality response to long-term nutrient and crop management on a semi-arid Inceptisol. **Agriculture, Ecosystems & Environment**, v. 118, n. (1-4), p. 130-142, 2007.
<https://doi.org/10.1016/j.agee.2006.05.008>

MELO, G. B.; PEREIRA, M. G.; PERIN, A.; GUARESCHI, R. F.; SOARES, P. F. C. Estoques e frações da matéria orgânica do solo sob os sistemas plantio direto e convencional de repolho. **Pesquisa Agropecuária Brasileira**, v. 51, p. 1511-1519, 2016. <https://doi.org/10.1590/S0100-204X2016000900050>

MONTEIRO, M. T.; GAMA-RODRIGUES, E. F. Carbono, nitrogênio e atividade da biomassa microbiana em diferentes estruturas de serapilheira de uma floresta natural. **Revista Brasileira de Ciência do Solo**, v. 28, p. 819-826, 2004.
<https://doi.org/10.1590/S0100-06832004000500004>

MUKHERJEE, A.; LAL, R. Comparison of soil quality index using three methods. **PLoS ONE**, v. 9, n. 8, p. e105981, 2014. doi:10.1371/journal.pone.0105981

NAVAS, M.; BENITO, M.; RODRIGUEZ, I.; MASAGUER, A. Effect of five forage legume covers on soil quality at the Eastern plains of Venezuela. **Applied Soil Ecology**, v. 49, p. 242–249, 2011.

NEHRANI, S. H.; ASKARI, M. S.; SAADAT, S.; DELAVAR, M. A.; TAHERI, M.; HOLDEN, N. M. Quantification of soil quality under semi-arid agriculture in the northwest of Iran. **Ecological indicators**, v. 108, p. 105770, 2020.
<https://doi.org/10.1016/j.ecolind.2019.105770>

NORTCLIFF, S. Standardisation of soil quality attributes. **Agriculture, Ecosystems & Environment**, v. 88, p. 161–168, 2002.

PAVANELLI, L. E.; ARAÚJO, F. F. D. Parâmetros químicos e biológicos indicadores de qualidade de solo sob cultivo de braquiárias e soja no oeste paulista. **Revista Ceres**, v. 57, p. 118-124, 2010. <https://doi.org/10.1590/S0034-737X2010000100019>

PEREIRA-FLORES, M. E.; JUSTINO, F.; RODRIGUES, J. M.; BOEHRINGER, D.; MELO, A. A. M.; CURSI, A. G.; RUIZ-VERA, U. M. Seasonal climate impact on Brazilian pasture (*Brachiaria brizantha* cv Marandu): growth rate, CO₂ efflux, and

irrigation strategies. **Theoretical and Applied Climatology**, v. 151, n. (1-2), p. 651-666, 2023. <https://doi.org/10.1007/s00704-022-04295-y>

SANTOS, K.; DIAS, J. P.; AMADO, C. A literature review of machine learning algorithms for crash injury severity prediction. **Journal of Safety Research**, v. 80, p. 254-269, 2022. <https://doi.org/10.1016/j.jsr.2021.12.007>

SÁNCHEZ-NAVARRO, A.; GIL-VÁZQUEZ, J. M.; DELGADO-INIESTA, M. J.; MARÍN-SANLEANDRO, P.; BLANCO-BERNARDEAU, A.; ORTIZ-SILLA, R. Establishing an index and identification of limiting parameters for characterizing soil quality in Mediterranean ecosystems. **Catena**, v. 131, p. 35-45, 2015. <https://doi.org/10.1016/j.catena.2015.02.023>

SILVA, I. M.; ROMERO, D. J.; GUIMARÃES, C. C. B.; ALVES, M. R.; OSCO, L. P.; SOUZA, A. B. E.; SILVA, A. P.; DEMATTÊ, J. A. M. Readily dispersible clay in soils from different Brazilian regions by visible, near, and mid-infrared spectral data. **International Journal of Remote Sensing**, v. 42, n. 18, p. 6945–6962, 2021. <https://doi.org/10.1080/01431161.2021.1948625>

SILVA, J. J. C.; SOUZA, R. M.; FONTANELA, E.; PRATES, E. D.; LIMA, A. C. R. Monitoramento da qualidade de 7 solo hidromórfico através de indicadores biológicos. **Desenvolvimento de protocolo, ii DII ILIDifil III W**, n. 117, 2003. Disponível em: <https://core.ac.uk/download/pdf/45514149.pdf#page=119>. Acesso em: 20 jul 2023.

SILVA, I.; MIELNICZUK, J. Ação do sistema radicular de plantas na formação e estabilização de agregados do solo. **Revista Brasileira de Ciência do Solo**, v. 21, n. 1, p. 113-117, 1997. Disponível em: <http://pascal-francis.inist.fr/vibad/index.php?action=getRecordDetail&idt=2704315>. Acesso em: 10 jan. 2023.

STOCKING, M. A.. Tropical soils and food security: the next 50 years. **Science**, 302(5649), 1356-1359 (2003). DOI: 10.1126/science.1088579

TZIOLAS, N. **Faculty of agriculture and natural environment school of agriculture** . 2022. Thesis (Doctoral) - Aristotle University of Thessaloniki, 2022. Disponível em: https://web.archive.org/web/20220919193741id_/http://ikee.lib.auth.gr/record/341727/files/TZIOLAS_thesis_final.pdf. Acesso em: 10 mar. 2023.

YASEEN, Z. M. An insight into machine learning models era in simulating soil, water bodies and adsorption heavy metals: review, challenges and solutions. **Chemosphere**, v. 277, p. 130126, 2021. <https://doi.org/10.1016/j.chemosphere.2021.130126>

ZINN, Y. L.; RESCK, D. V.; SILVA, J. E. Soil organic carbon as affected by afforestation with Eucalyptus and Pinus in the Cerrado region of Brazil. **Forest Ecology and Management**, v. 166, n. (1-3), p. 285-294, 2002. [https://doi.org/10.1016/S0378-1127\(01\)00682-X](https://doi.org/10.1016/S0378-1127(01)00682-X)