

UNIVERSIDADE FEDERAL DE VIÇOSA

**Desenvolvimento de um sistema de baixo custo para aquisição do espectro de
reflectância do solo**

Thiago Furtado de Oliveira
Doctor Scientiae

**VIÇOSA - MINAS GERAIS
2023**

THIAGO FURTADO DE OLIVEIRA

Desenvolvimento de um sistema de baixo custo para aquisição do espectro de reflectância do solo

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Engenharia Agrícola, para obtenção do título de *Doctor Scientiae*.

Orientador: Domingos S. M. Valente

Coorientadores: Francisco de A. de C. Pinto
Flora M. de Melo Villar
Daniel M. de Queiroz

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

O48d
2023

Oliveira, Thiago Furtado, 1987-

Desenvolvimento de um sistema de baixo custo para
aquisição do espectro de reflectância do solo / Thiago Furtado
Oliveira. – Viçosa, MG, 2023.

1 tese eletrônica (98 f.): il. (algumas color.).

Orientador: Domingos Sárvio Magalhães Valente.

Tese (doutorado) - Universidade Federal de Viçosa,
Departamento de Engenharia Agrícola, 2023.

Referências bibliográficas: f. 46-53.

DOI: <https://doi.org/10.47328/ufvbbt.2025.503>

Modo de acesso: World Wide Web.

1. Agricultura de precisão. 2. Aprendizado do computador.
3. Espectroscopia de reflectância. 4. Solos - Análise. I. Valente,
Domingos Sárvio Magalhães, 1978-. II. Universidade Federal de
Viçosa. Departamento de Engenharia Agrícola. Programa de
Pós-Graduação em Engenharia Agrícola. III. Título.

CDD 22. ed. 631.3

THIAGO FURTADO DE OLIVEIRA

Desenvolvimento de um sistema de baixo custo para aquisição do espectro de reflectância do solo

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Engenharia Agrícola, para obtenção do título de *Doctor Scientiae*.

APROVADA: 16 de junho de 2023.

Assentimento:

Thiago Furtado de Oliveira
Autor

Domingos Sarvio Magalhaes Valente
Orientador

Essa tese foi assinada digitalmente pelo autor em 18/08/2025 às 13:40:34 e pelo orientador em 18/08/2025 às 13:54:31. As assinaturas têm validade legal, conforme o disposto na Medida Provisória 2.200-2/2001 e na Resolução nº 37/2012 do CONARQ. Para conferir a autenticidade, acesse <https://siadoc.ufv.br/validar-documento>. No campo 'Código de registro', informe o código **BBOO.OX65.NJBG** e clique no botão 'Validar documento'.

Dedico esta tese em memória de minha mãe, que sempre me inspirou a seguir em frente e perseguir meus objetivos. Sua memória sempre estará presente em meu coração.

AGRADECIMENTOS

Aos meus pais, Maria Aparecida Furtado e Amaro Bezerra de Oliveira, minha gratidão por sua confiança, apoio, carinho e amor incondicional.

A minha esposa Daniele de Barros Gomes Furtado, agradeço o companheirismo, compreensão, alegria, apoio incondicional e motivação, sem a qual toda a jornada seria muito mais difícil.

A minha filha Alice, por trazer alegria e diversão durante momentos desafiadores e por inspirar-me a tornar o mundo um lugar melhor.

Ao professor Domingos Sarvio Magalhães Valente, por sua orientação, dedicação, ensinamentos, confiança e ajuda.

A Universidade Federal de Viçosa, e ao Programa de Pós-Graduação em Engenharia Agrícola, pela oportunidade e todo aprendizado proporcionado.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa de estudos.

Aos professores Rodrigo Sinaidi Zandonadi (UFMT) e Alexandre da Silva Adão (IFMG) pela conceção das amostras de solo e resultados de análises de laboratório.

Aos amigos do Laboratório de Mecanização Agrícola, pelos bons momentos de alegria e descontração, sempre dispostos a ajudar no que preciso fosse.

Este trabalho foi realizado com o apoio das seguintes agências de pesquisa brasileiras: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

RESUMO

OLIVEIRA, Thiago Furtado de, D.Sc., Universidade Federal de Viçosa, junho de 2023. **Desenvolvimento de um sistema de baixo custo para aquisição do espectro de reflectância do solo.** Orientador: Domingos Sarvio Magalhaes Valente. Coorientadores: Francisco de Assis de Carvalho Pinto, Flora Maria de Melo Villar e Daniel Marcal de Queiroz.

Esta tese teve como objetivo estudar a utilização de um espectrorradiômetro de baixo custo para fornecer suporte às técnicas de agricultura de precisão, visando o manejo da fertilidade do solo. A tese foi dividida em dois capítulos, nos quais foram investigados o uso da reflectância de amostras de solo obtidas com o espectrorradiômetro de baixo custo como entrada em modelos de aprendizado de máquina (ML). O primeiro capítulo teve o objetivo de avaliar o uso das reflectâncias como entrada de modelos supervisionados de ML para predição de atributos do solo. Foram utilizadas 417 amostras de solo provenientes de duas áreas agrícolas, uma localizada em Barbacena/MG (Área 1) e outra em Tabaporã/MT (Área 2). As reflectâncias na faixa de 410 a 1.600 nm foram coletadas utilizando amostras de solo seca e saturada. Após a correção do espalhamento utilizando o método Standard Normal Variate (SNV), foram calculados índices espectrais do tipo índices por diferença (DI), índices normalizados (NI) e índices por diferença normalizada (NDI). Os índices mais significativos para predição de cada atributo foram selecionados utilizando o algoritmo de seleção genética. Os índices selecionados foram utilizados como entrada dos modelos de ML para a predição dos atributos químicos das amostras de solo. Foram avaliados os modelos Random Forest (RF), Support Vector Machine Regression (SVM), Partial Least Squares Regressor (PLSR) e Multi-layer Perceptron (MLP). Os atributos do solo estudados foram o pH em água, fósforo disponível (P), potássio (K), cálcio (Ca^{2+}), magnésio (Mg^{2+}), capacidade de troca de cátions efetiva (t), capacidade de troca de cátions potencial (T), fósforo remanescente (Prem) e matéria orgânica (MO). Durante a validação cruzada dos dados de teste, os modelos apresentaram coeficiente de determinação (R^2) variando de 0,50 a 0,68 e desvio residual da predição (RPD) entre 1,42 e 1,80. Foi possível obter estimativas satisfatórias para Ca^{2+} , t e Prem em ambas as áreas de estudo, bem como para pH, K, T e MO nas amostras da Área 2 e Mg^{2+} nas amostras da Área 1. Assim podemos notar que o tipo de solo tem impacto direto na performance do espectrorradiômetro, evidenciando que os modelos gerados são dependentes desses fatores. No segundo capítulo, o objetivo foi avaliar se as assinaturas espectrais obtidas pelo sensor podem ser utilizadas para definir zonas de manejo (ZM) baseadas na fertilidade do solo. Para isso, foram selecionadas 112

amostras de solo georreferenciadas, presentes em uma área agrícola de 123 ha localizada em Tabaporã/MT. As assinaturas espectrais foram utilizadas para cálculo de índices espectrais. Os índices do tipo NI, DI e NDI foram calculados utilizando 16 comprimentos de onda relacionados com compostos orgânicos e minerais do solo descritos na literatura. A seguir, os índices foram utilizados como entrada na análise de componentes principais. Foram geradas ZM utilizando a primeira componente principal (PC1) e ZM utilizando o conjunto das cinco primeiras componentes (PC1 a 5), por meio do algoritmo Fuzzy K-Means. O erro calculado em relação à média das ZM geradas foi comparado com os erros gerados por diferentes metodologias de amostragem do solo. As metodologias de amostragem utilizadas para comparação foram amostragem em grid com 112 amostras (Grid 1), amostragem em grid com 30 amostras (Grid 2), o método convencional e definição de ZM aleatórias. Os erros foram calculados para as variáveis pH, Ca²⁺, Mg²⁺, K, t, T, P, Prem, MO e Argila. As ZMs geradas com a PC1 e com as PC1 a 5 apresentaram erros menores do que a definição aleatória de ZMs. Os erros obtidos pelas ZM foram menores que os da amostragem convencional e similares aos erros da utilização do Grid 2. As zonas de manejo criadas pela metodologia proposta permitiram a separação de zonas homogêneas dos atributos do solo com erros menores em comparação com a divisão da área de forma aleatória, e próximos ao erro do Grid 2. Como é necessária apenas uma amostra composta para caracterizar cada ZM, o uso do espectrorradiômetro possibilitou reduzir o número de amostras necessárias para delimitar as zonas de manejo, sem perda na qualidade dos resultados quando comparado com um grid menos denso.

Palavras-chave: aprendizagem de máquina; espectrometria do solo; agricultura de precisão.

ABSTRACT

OLIVEIRA, Thiago Furtado de, D.Sc., Universidade Federal de Viçosa, June, 2023.
Development of a low-cost system for the soil reflectance spectrum acquiring.
Adviser: Domingos Sarvio Magalhaes Valente. Co-advisers: Francisco de Assis de Carvalho Pinto, Flora Maria de Melo Villar and Daniel Marcal de Queiroz.

This thesis aims to study the use of a low-cost spectroradiometer to support precision agriculture techniques for soil fertility management. The thesis was divided into two chapters, in which the use of reflectance from soil samples obtained with the low-cost spectroradiometer as input in machine learning (ML) models was investigated. The first chapter aimed to evaluate the use of reflectance as input in supervised ML models for predicting soil attributes. A total of 417 soil samples from two agricultural areas, one located in Barbacena/MG (Area 1) and another in Tabaporã/MT (Area 2), were used. Reflectance in the range of 410 nm to 1600 nm was collected using dried and saturated soil samples. After scatter correction using the Standard Normal Variate (SNV) method, spectral indices of the type of difference indices (DI), normalized indices (NI), and normalized difference indices (NDI) were calculated. The most significant indices for predicting each attribute were selected using the genetic selection algorithm. The selected indices were used as input for the ML models. The Random Forest (RF), Support Vector Machine Regression (SVM), Partial Least Squares Regressor (PLSR), and Multi-layer Perceptron (MLP) models were evaluated. The soil attributes studied were water pH, available phosphorus (P), potassium (K), calcium (Ca), magnesium (Mg), effective cation exchange capacity (CTC), potential cation exchange capacity (CTCP), remaining phosphorus (Prem), and organic matter (OM). During the cross-validation of the test data, the models presented R^2 ranging from 0.50 to 0.68 and RPD between 1.42 and 1.80. Satisfactory predictions were possible for Ca, CTC, and Prem in both study areas, as well as for pH, K, CTCP, and MO in samples from Area 2 and Mg in samples from Area 1. The results indicate that the embedded system was able to make accurate predictions of soil attributes and performs similarly to conventional spectroradiometers. In the second chapter, the objective was to evaluate whether the spectral signatures obtained by the sensor can be used to define soil fertility management zones (MZ). For this purpose, 112 georeferenced soil samples from a 123-ha agricultural area located in Tabaporã/MT were selected. The spectral signatures were used to calculate spectral indices. The indices of the NI, DI, and NDI types were calculated using 16 wavelengths related to organic and mineral compounds of the soil described in the literature. The indices were then

used as input for principal component analysis. SMZs were generated using the first principal component (PC1) and MZs using the set of the first five components (PC1 to 5) using the Fuzzy K-Means algorithm. The resulting error associated with the generated MZs was compared to errors generated by different soil sampling methodologies. The sampling methodologies used for comparison were grid sampling with 112 samples (Grid 1), grid sampling with 30 samples (Grid 2), conventional sampling, and random MZ definition. The errors were calculated for the variables pH, Ca, Mg, K, CTC, CTCP, P, Prem, MO, and Clay. The MZs generated with PC1 and with PC1 to 5 had smaller errors than the random MZ definition. The errors obtained by the MZs were smaller than those of conventional sampling and like the errors of using Grid 2. The soil management zones created by the proposed methodology allowed the separation of homogeneous soil attribute zones with smaller errors compared to random area division and close to the error of Grid 2. The use of the embedded system made it possible to reduce the number of necessary samples to define the SMZs without loss of result quality.

Keywords: machine learning; soil spectrometry; precision agriculture.

LISTA DE FIGURAS

Figura 1 – Sensor <i>spectral modelo AS7265x Triad</i> (a), <i>DLP® NIRScan™ Nano</i> (b), módulo desenvolvido para acoplamento dos sensores, amostra e fontes de iluminação(c) e vista superior do módulo com suas partes constituintes (d).....	29
Figura 2 - Fluxograma mostrando o fluxo de informações (azul), energia (vermelho) e movimento (preto) no sistema embarcado.....	31
Figura 3 - Mapas das áreas amostradas: em Barbacena/MG, com 417 amostras (a); e em Tabaporã/MT, com 151 amostras (b).	32
Figura 4 – Espectrorradiômetro de baixo custo (a); detalhe mostrando o branco padrão, cubeta com amostra de solo e local de inserção das amostras no espectrorradiômetro (b); cubetas com amostras de solo seco, solo saturado e branco padrão (c); e software desenvolvido para coleta e visualização de dados (d).....	33
Figura 5 – Fluxograma mostrando os passos para criação do banco de dados utilizado para desenvolvimento e validação dos modelos de <i>Machine Learning</i>	35
Figura 6 - Fluxograma mostrando o pré-processamento utilizado, a seleção de variáveis, o ajuste dos hiper parâmetros e a avaliação das métricas para seleção do melhor modelo para cada variável predita.	37
Figura 7 - Matriz de correlação dos atributos do solo para os resultados das análises de laboratório.....	39
Figura 8 - Assinaturas espectrais das 417 amostras de solo da Área 1 e das 151 amostras da Área 2 secas e saturadas. (a) valores de reflectância obtidos com o sensor <i>Triad Spectral</i> , e (b) espectros obtidos com o <i>NIRScan Nano</i> após pré-processamento com o SNV. Nas cores claras estão os espectros individuais e em escuro os espectros médios.....	40
Figura 8 – Comprimentos de ondas selecionados para o modelo MLPR para predição do pH nas amostras da Área 2 (a) e importância das variáveis no modelo (b).	44
Figura 9 – Comprimentos de ondas selecionados para o modelo RF para predição do potássio disponível nas amostras da Área 2 (a) e dez variáveis de maior importância no modelo RF(b).	46
Figura 10 - Comprimentos de onda selecionados para predição do cálcio pelo modelo SVR para o MT (a) valores preditos vs. observados para os dados de treino e teste dos modelos gerados (b) e as dez variáveis de maior importância no modelo SVR(c).	48

Figura 11 - Comprimentos de onda selecionados para predição do cálcio pelo modelo RF para Área 1 (a) valores preditos vs. observados para os dados de treino e teste dos modelos gerados (b) e as dez variáveis de maior importância no modelo RF(c).	49
Figura 12 - Comprimentos de onda selecionados para predição do magnésio pelo modelo RF para Área 1 (a) valores preditos vs. observados para os dados de treino e teste dos modelos gerados (b) e as dez variáveis de maior importância no modelo RF (c).	51
Figura 13 - Comprimentos de onda selecionados para predição da capacidade de troca de cátions pelo modelo SVR para Área 2 (a) valores preditos vs. observados para os dados de treino e teste dos modelos gerados (b) e as dez variáveis de maior importância no modelo SVR (c).	53
Figura 14 - Comprimentos de onda selecionados para predição da capacidade de troca de cátions pelo modelo RF para Área 1 (a) valores preditos vs. observados para os dados de treino e teste dos modelos gerados (b) e as dez variáveis de maior importância no modelo RF (c).	55
Figura 15 - Comprimentos de onda selecionados para predição da capacidade de troca de cátions potencial pelo modelo RF para Área 2 (a) valores preditos vs. observados para os dados de treino e teste dos modelos gerados (b) e as oito variáveis de maior importância no modelo RF (c).	57
Figura 16 - Comprimentos de onda selecionados para predição do fósforo remanescente pelo modelo SVR para Área 2 (a) valores preditos vs. observados para os dados de treino e teste dos modelos gerados (b) e as dez variáveis de maior importância no modelo SVR (c).	59
Figura 17 - Comprimentos de onda selecionados para predição do fósforo remanescente pelo modelo RF para Área 1 (a) valores preditos vs. observados para os dados de treino e teste dos modelos gerados (b) e as dez variáveis de maior importância no modelo RF (c).	61
Figura 18 - Comprimentos de onda selecionados para predição da matéria orgânica pelo modelo SVR para Área 2 (a) valores preditos vs. observados para os dados de treino e teste dos modelos gerados (b) e as dez variáveis de maior importância no modelo SVR (c).	62
Figura 19 – Área de estudo localizada em Tabaporã/MT (polígonos em azul) e localização das amostras de solo coletadas (Grid 1).	75
Figura 20 – Espectrorradiômetro de baixo custo AISpec fechado para transporte (a), AISpec aberto e em funcionamento (b), cubetas com amostras de solo e branco padrão (c) e software desenvolvido para coleta e visualização de dados (d).	76
Figura 21 – Fluxograma mostrando desde a leitura das amostras até a criação do banco de dados utilizado nos algoritmos de cluster.	77
Figura 22 - Comprimentos de ondas utilizados para compor os índices espectrais e espectros de reflectância.	78

Figura 23 – Fluxograma para geração das classes.....	79
Figura 24 – Amostras utilizadas para o Grid 2 (30 amostras).	81
Figura 25 – Variância explicada pelas componentes principais obtidas na análise de componentes principais (PCA).....	83
Figura 26 – Mapas interpolados por krigagem ordinária para as componentes principais um a cinco.	84
Figura 27 - Mapas dos atributos físico-químicos interpolados por krigagem ordinária utilizando o Grid 1.....	86
Figura 28 - Mapas dos atributos físico-químicos interpolados por krigagem ordinária utilizando o Grid 2.....	87
Figura 29 – Mapas com as zonas de manejo geradas utilizando a componente principal um como entrada do algoritmo fuzzy K-Means.	88
Figura 30 – Mapas com as zonas de manejo geradas utilizando as componentes principais 1 a 5 como entrada do algoritmo fuzzy K-Means.	89
Figura 31 - Comparação entre os erros calculados para cada método de amostragem utilizado. Comparação entre os métodos Grid 1, Grid 2, Convencional e zonas de manejo geradas com a PC1 (a). Comparação entre os métodos Grid 1, Grid 2, Convencional e zonas de manejo geradas com as PC1, 2, 3, 4 e 5 (b).....	91

LISTA DE TABELAS

CAPÍTULO I

Tabela 1: Estatística descritiva dos resultados das análises laboratoriais para as amostras de Minas Gerais (Área 1) e Mato Grosso (Área 2).	38
Tabela 2 - Resumo das métricas para os modelos gerados para predição dos atributos do solo, em negrito está destacada o melhor modelo para o atributo em questão.	42
Tabela 3: Estatística descritiva dos resultados das análises laboratoriais para as amostras de solo utilizadas.....	83
Tabela 4: Índices de Moran e p-valor para os atributos físico-químicos do solo em relação ao Grid 1 e ao Grid 2.....	85
Tabela 5: Porcentagem das vezes em que a simples divisão aleatória das zonas de manejo repetida 2000 vezes, resultou em erros menores que a utilização da metodologia proposta na divisão das zonas de manejo.....	90

SUMÁRIO

INTRODUÇÃO GERAL	14
REFERÊNCIAS	19
CAPÍTULO I.....	25
RESUMO.....	25
ABSTRACT	26
1 INTRODUÇÃO.....	27
2 MATERIAL E MÉTODOS.....	29
3 RESULTADOS E DISCUSSÃO	37
4 CONCLUSÃO.....	64
REFERÊNCIAS	64
CAPÍTULO II.....	70
RESUMO.....	70
ABSTRACT	72
1 INTRODUÇÃO.....	73
2 MATERIAL E MÉTODOS.....	74
3 RESULTADOS E DISCUSSÃO	81
4 CONCLUSÃO.....	92
REFERÊNCIAS	93

INTRODUÇÃO GERAL

O conhecimento da variabilidade espacial das propriedades físico-químicas do solo é fundamental para otimizar o uso de insumos na agricultura. As fontes desta variabilidade podem ser classificadas em variações sistemáticas e variações aleatórias (Couto; Stein; Klamt, 1997). Variações sistemáticas são aquelas decorrentes da forma do terreno (montes, vales, planícies, planaltos, etc.), dos elementos geomórficos, das condições que deram origem ao solo (material de origem, posição topográfica, fatores biológicos e condições do clima) e a interação entre estas três. Variações aleatórias podem ocorrer devido a ações antropogênicas como forma de preparo do solo, manejo, adubação, tipo de cultivo, utilização de técnicas conservacionistas, etc. podem ocorrer também devido à influência da fauna e condições climáticas (Couto; Stein; Klamt, 1997).

Devido aos fenômenos que originam as variações sistemáticas, estas ocorrem em escala regional. As variações advindas destas fontes por vezes representam a menor parte das variações observadas a nível local ou de campo (Couto; Stein; Klamt, 1997). Portanto, em escala de campo, os fatores que mais determinam a variabilidade são o tipo de solo, topografia, cultura anterior, práticas de manejo do solo, orientação das linhas de cultivo, formas de aplicação dos fertilizantes e compactação (Mallarino; Wittry, 2004). A mobilidade dos nutrientes do solo em conjunto com a técnica de fertilização utilizada pode também influenciar diretamente a variabilidade espacial do nutriente. Em solos oxídicos, a mobilidade limitada do fósforo ocorre principalmente por difusão, resultando na concentração preferencial desse nutriente nas linhas de semeadura quando se utiliza técnicas de fertilização em linha (Han *et al.*, 1994; Novais *et al.*, 2007). Os nutrientes que se movimentam por fluxo de massa, como o nitrogênio, não apresentam esta tendência de concentração. A fim de gerir estas variabilidades, existem diversas metodologias que a agricultura de precisão (AP) utiliza para amostrar e modelar a variação espacial destes atributos.

A geoestatística tem sido amplamente adotada na AP para estudar e representar a variabilidade espacial das características do solo. Por meio de processos de interpolação é possível estimar os valores das variáveis de interesse entre os pontos amostrados (Lauzon *et al.*, 2005). Porém, as propriedades do solo apresentam diferentes distribuições espaciais em diferentes escalas, assim, a confiabilidade da interpolação está intimamente relacionada com a característica estudada e a densidade amostral coletada (Lauzon *et al.*, 2005; Mallarino; Wittry, 2004).

A densidade do grid amostral necessária para representar a variabilidade das características do solo é dependente da característica estudada e da estratégia de amostragem utilizada, e pode ser diferente entre talhões (Han *et al.*, 1994). Esta conclusão é reforçada pelas diferentes densidades amostrais necessárias para representar a variabilidade espacial apresentada por diferentes estudos. Mallarino e Wittry (2004) em estudo entre diferentes métodos de amostragem para determinação do fósforo, potássio, pH e matéria orgânica, concluem que a amostragem em grid com cinco amostras por hectare apresentou melhor resultados. Lauzon *et al.* (2005) estudando a variabilidade do fósforo total, potássio total e pH, em 23 campos diferentes, concluíram ser necessário um grid amostral de 30 m x 30 m ou menor para representar fielmente a variabilidade. Rodrigues *et al.* (2016) utilizaram um grid amostral regular de 30 m x 30 m para caracterizar a variabilidade espacial de diversas propriedades do solo em uma área com cultivo de cana-de-açúcar. Já Bernardi *et al.* (2016) estudando a variabilidade espacial das propriedades do solo em área de pastagem utilizaram um grid irregular com 10 amostras por hectare. Segundo os autores, valores de dependência espacial classificados como baixos indicam que a densidade amostral não foi suficientemente densa para caracterizar a variabilidade espacial da propriedade do solo. Na região sul do país por exemplo, é comum a utilização de grids que variam de 100 x 100m a 173 x 173m. A adoção de baixas densidades amostrais muitas vezes ocorre devido a critérios econômicos ou práticos, o que acaba por ignorar princípios geoestatísticos, impactando negativamente a exatidão dos resultados (Cherubin *et al.*, 2016).

Como mostrado, há diferentes causas para a existência das variações a nível de campo, o que implica nas diferenças no padrão de distribuição da variabilidade de local para local. Assim, não existe um padrão de amostragem capaz de se ajustar de forma ótima para todos os casos. Aumentar a densidade do grid de amostragem georreferenciada de solo pode contribuir para a redução das incertezas nas predições, no entanto, é importante considerar que em regiões com alta uniformidade no campo, essa abordagem pode se tornar menos viável, resultando em excesso de amostras dentro da região. De maneira análoga, em regiões com elevada variabilidade, a subamostragem pode ser um problema.

Uma forma de conhecer previamente a variabilidade presente em um campo, seria utilizar variáveis que se relacionem com atributos de interesse e sejam fáceis e baratas de serem amostradas, como por exemplo características topográficas, mapas de vegetação, condutividade elétrica aparente, reflectância, entre outros. Ao utilizar essas variáveis indiretas, é possível identificar áreas potenciais de maior variabilidade, direcionando a amostragem de forma mais eficiente e garantindo uma representação adequada da variabilidade espacial dos atributos do

solo. Essa abordagem complementar e estratégica na coleta de dados em estudos de agricultura de precisão pode auxiliar na obtenção de informações mais precisas e confiáveis (Fortes *et al.*, 2015; Lawrence *et al.*, 2020).

A espectrometria é uma metodologia que se destaca como ferramenta analítica, pois o espectro VIS-NIR-MIR é sensível a diferentes componentes minerais e orgânicos presentes no solo. Devido a cada componente impactar uma região diferente do espectro eletromagnético, é possível obter, de forma simultânea, informações sobre diferentes compostos do solo. O que também possibilita a utilização de modelos multivariados para predição dos atributos do solo. Aliado a isso, a facilidade de obter os espectros sem a necessidade de utilizar reagentes químicos e o baixo custo de obtenção do espectro de refletância por amostra justificam o interesse no uso desta metodologia (Stenberg *et al.*, 2010a).

A espectrometria do solo quantifica a radiação eletromagnética refletida/absorvida por amostras de solo para comprimentos de ondas compreendidos entre de 400 nm a 50.000 nm (VIS-NIR-MIR) (Graham R. Hunt; John W. Salisbury, 1970; ISO, 2007). O fenômeno de refletância ocorre devido a tipos específicos de vibração e rotação entre átomos de uma mesma molécula excitados por comprimentos de ondas específicos. Tais moléculas podem vibrar de formas diferentes. A vibração pode ocorrer pelo alongamento da ligação entre os átomos ao longo do eixo da ligação, ou pela flexão da ligação atômica em deslocamentos que acontecem em ângulo reto com eixo da ligação. Quando ocorre uma alteração no momento dipolo da molécula decorrente da vibração, há então a absorção da radiação no comprimento de onda que provocou a vibração (Graham R. Hunt; John W. Salisbury, 1970). O tipo e o número de vibrações que podem ocorrer em uma molécula são dependentes da massa atômica dos átomos, da geometria da molécula e das forças intratômicas existentes (Graham R. Hunt; John W. Salisbury, 1970).

Dentro do espectro VIS-NIR-MIR ocorrem a excitação das ligações atômicas em frequências fundamentais e suas harmônicas. As frequências fundamentais de ligações químicas, presentes alguns argilominerais e componentes da matéria orgânica (MO), estão presentes no MIR (3000 a 50.000 nm). Na região do NIR (780 a 3.000 nm) estão presentes a combinação de frequências fundamentais e harmônicas (de frequências fundamentais presentes no MIR), estas frequências apresentam menor intensidade de absorção e maior largura de banda que suas frequências fundamentais. Na região do VIS a elevada energia da radiação faz com que o processo de absorção/reflexão seja dominado pela excitação eletrônica (Stenberg *et al.*, 2010a). Apesar da região do VIS não apresentar harmônicas relacionadas com os constituintes

dos solos, a absorção em alguns comprimentos de ondas está relacionada a minerais contendo Fe como a hematita e a goethita e cromóforos que constituem a MO (Stenberg *et al.*, 2010a).

Diversos estudos realizados na última década constataam a influência dos argilominerais e componentes da matéria orgânica nas assinaturas espectrais de diferentes tipos de solos. A absorção da radiação eletromagnética devido a vibrações das ligações químicas existentes em constituintes da MO ocorre nos comprimentos de onda de 1420 nm, 1930 nm e 2200 nm, devido ligações C-H, C=O e N-H (Pudelko; Chodak, 2020; Rossel; Behrens, 2010). No MIR, os picos de absorção devido aos compostos orgânicos se dão nos comprimentos de onda de 5714 a 7143 nm e 7143 a 10526 nm (Metzger *et al.*, 2020), 3390 a 3484 nm devido a absorção por grupamentos C-H, 2941 nm promovido por ligações N-H e 1600 a 1400 nm em ligações de compostos alifáticos (Hutengs *et al.*, 2019).

Compostos inorgânicos apresentam picos de absorção em 2703 nm, 2740 nm, 4000 nm, 5000 a 6060 nm, 9174 a 10638 nm e 11236 nm como a caulinita, argilas 2:1, carbonatos, caulinita/quartzo, areia/quartzo e carbonatos respectivamente (Metzger *et al.*, 2020). Coutinho *et al.* (2019) atribuíram os picos de absorção nos comprimentos de onda do NIR, entre 1900-2500 nm, às frequências fundamentais dos oxihidróxidos de Fe e Al (8197 a 8873 nm). Picos de absorção no NIR em 1400 e 1900 nm estão também relacionados a presença de água, as ligações metal-OH provocam absorção entre 2200 e 2300 nm (Hutengs *et al.*, 2019). Estes autores também evidenciam as frequências fundamentais dos grupamentos OH presentes na água estão entre 2778 e 2702 nm, e os silicatos apresentam frequências fundamentais em 7812 e 9346 nm.

Portanto, as assinaturas espectrais dos solos são sensíveis aos argilominerais e componentes das frações da MO presentes no solo. Os comprimentos de onda mais sensíveis estão no MIR e correspondem as frequências fundamentais de vibração entre átomos de uma mesma molécula, e os harmônicos destas frequências fundamentais estão presentes no NIR. Sendo assim, a aplicação da espectrometria para quantificar e qualificar atributos do solo está limitada aos atributos, fração argila, MO, e propriedades a estes correlacionadas. O alto custo dos fotodiodos que operam no MIR eleva o custo de tais equipamentos quando comparado aos equipamentos que operam no VIS-NIR (THORLABS INC., 2020a, 2020b), além de serem em sua grande maioria, desenvolvidos para operar em bancadas de laboratório, inviabilizando seu uso em campo.

Devido às técnicas de espectrometria resultarem em um amplo espaço dimensional, com muitas variáveis correlacionadas e por muitas vezes apresentar um relacionamento não linear com a variável predita, o uso de algoritmos de ML vêm sendo cada vez mais aplicados para

treinamento de modelos preditivos aplicando a espectrometria (Padarian; Minasny; McBratney, 2020) e pré-processamento dos dados. Dentre os modelos de ML mais frequentemente utilizados em estudos sobre espectrometria do solo, podemos destacar o *Partial Least Squares* (PLS), *Support Vector Machine* (SVM), *Artificial Neural Network* (ANN), *Random Forest* (RF), *Principal Component Regression* (PCR), e *Multiple Linear Regression* (MLR) (Barra *et al.*, 2021; Padarian; Minasny; McBratney, 2020).

Embora a espectrometria VIS-NIR seja uma técnica que resulta em menor custo com análises, são necessários sensores (espectrorradiômetros) de alto custo para aquisição dos espectros, que posteriormente precisam ser processados utilizando modelos complexos para extração das informações. Atualmente, no entanto, existem sensores de baixo custo para obtenção da reflectância, bem como computadores de placa única pequenos o suficiente para serem embutidos em um sistema portátil e que podem ser utilizados para o processamento dos dados. Este sistema pode desempenhar um papel fundamental na agricultura de precisão. Por exemplo, pode ser empregado para predição dos atributos do solo, identificar áreas com características similares, guiar o processo de coleta de amostras de solo e atuar como covariável na modelagem da variabilidade. Essas funcionalidades permitem uma abordagem mais assertiva na gestão da variabilidade espacial do solo.

Portanto, o objetivo foi o desenvolvimento e avaliação de um sistema portátil de baixo custo para realizar a aquisição e o processamento dos espectros VIS-NIR em amostras de solo utilizando algoritmos de machine learning. Para tanto, esta tese foi dividida em dois capítulos. O CAPÍTULO I descreve o desenvolvimento do sistema sensor e avalia a capacidade preditiva de modelos de machine learning que utilizam como entrada valores de reflectância obtidos com o sistema desenvolvido para a predição de atributos físico-químicos do solo, enquanto o CAPÍTULO II avalia a capacidade de geração de zonas de manejo homogêneas para amostragem dos atributos do solo, utilizando como entrada as informações espectrais obtidas por meio do equipamento desenvolvido.

REFERÊNCIAS

- ABDI, Dalel *et al.* **Nutrient Management & Soil & Plant Analysis Predicting Soil Phosphorus-Related Properties Using Near-Infrared Reflectance Spectroscopy** Léon-Étienne Parent *Soil Sci. Soc. Am. J.* [S. l.: s. n.], 2012. Disponível em: www.soils.org/publications/sssaj. .
- ALTMANN, André *et al.* Permutation importance: A corrected feature importance measure. **Bioinformatics**, [s. l.], v. 26, n. 10, p. 1340–1347, 2010.
- AMS. **AS7265x Datasheet**. [S. l.: s. n.], 2018. Disponível em: asm.com.
- AURENHAMMER, Franz. Voronoi diagrams—a survey of a fundamental geometric data structure. **ACM Computing Surveys (CSUR)**, [s. l.], v. 23, n. 3, p. 345–405, 1991. Disponível em: Acesso em: 20 mar. 2023.
- BARRA, Issam *et al.* **Soil spectroscopy with the use of chemometrics, machine learning and pre-processing techniques in soil diagnosis: Recent advances—A review**. [S. l.]: Elsevier B.V., 2021. Disponível em: Acesso em: 29 jun. 2022.
- BEAGLEBOARD.ORG FOUNDATION. **BeagleBone Black**. [S. l.], 2023. Disponível em: <https://docs.beagleboard.org/latest/boards/beaglebone/black/>. Acesso em: 20 mar. 2023.
- BERNARDI, A. C.C. *et al.* Spatial variability of soil properties and yield of a grazed alfalfa pasture in Brazil. **Precision Agriculture**, [s. l.], v. 17, n. 6, p. 737–752, 2016.
- BZDOK, Danilo; ALTMAN, Naomi; KRZYWINSKI, Martin. THIS MONTH Statistics versus machine learning. **Nature Publishing Group**, [s. l.], v. 15, 2018. Disponível em: Acesso em: 26 jul. 2022.
- CHANG, C. W. *et al.* Near-infrared reflectance spectroscopy - Principal components regression analyses of soil properties. **Soil Science Society of America Journal**, [s. l.], v. 65, n. 2, p. 480–490, 2001.
- CHERUBIN, Maurício Roberto *et al.* Amostragem de solo na agricultura de precisão. In: **AGRICULTURA DE PRECISÃO NO RIO GRANDE DO SUL**. [S. l.: s. n.], 2016. p. 79–98.
- CLINGENSMITH, C. M.; GRUNWALD, S.; WANI, S. P. Evaluation of calibration subsetting and new chemometric methods on the spectral prediction of key soil properties in a data-limited environment. **European Journal of Soil Science**, [s. l.], v. 70, n. 1, p. 107–126, 2019. Disponível em: Acesso em: 4 jul. 2022.
- COUTINHO, Marcos A.N. *et al.* Influence of soil sample preparation on the quantification of NPK content via spectroscopy. **Geoderma**, [s. l.], v. 338, n. July 2018, p. 401–409, 2019. Disponível em: <https://doi.org/10.1016/j.geoderma.2018.12.021>.

COUTO, E. Guimarães; STEIN, A.; KLAMT, E. Large area spatial variability of soil chemical properties in central Brazil. **Agriculture, Ecosystems and Environment**, [s. l.], v. 66, n. 2, p. 139–152, 1997.

DE VARGAS, Jocelina P.R. *et al.* Application forms and types of soil acidity corrective: Changes in depth chemical attributes in long term period experiment. **Soil and Tillage Research**, [s. l.], v. 185, p. 47–60, 2019. Disponível em: <https://doi.org/10.1016/j.still.2018.08.014>. Acesso em: 30 jun. 2022.

DUNN, B. W. *et al.* The potential of near-infrared reflectance spectroscopy for soil analysis - A case study from the Riverine Plain of south-eastern Australia. **Australian Journal of Experimental Agriculture**, [s. l.], v. 42, n. 5, p. 607–614, 2002.

FÁBREGA, José R. *et al.* Modeling short-term soil-water distribution of aromatic amines. **Environmental Science and Technology**, [s. l.], v. 32, n. 18, p. 2788–2794, 1998. Disponível em: doi.org/10.1021/es9802394. Acesso em: 4 jan. 2023.

FAGERIA, N. K.; BALIGAR, V. C. Chapter 7 Ameliorating Soil Acidity of Tropical Oxisols by Liming For Sustainable Crop Production. **Advances in Agronomy**, [s. l.], v. 99, p. 345–399, 2008. Disponível em: Acesso em: 19 jul. 2022.

FORTES, R. *et al.* A methodology based on apparent electrical conductivity and guided soil samples to improve irrigation zoning. **Precision Agriculture**, [s. l.], v. 16, n. 4, p. 441–454, 2015.

GOMEZ, Cécile; LAGACHERIE, Philippe; COULOUMA, Guillaume. Regional predictions of eight common soil properties and their spatial structures from hyperspectral Vis-NIR data. **Geoderma**, [s. l.], v. 189–190, p. 176–185, 2012. Disponível em: www.elsevier.com/locate/geoderma. Acesso em: 3 jul. 2022.

GRAHAM R. HUNT; JOHN W. SALISBURY. **Visible and Near Infrared Spectra of Minerals and Rocks: I Silicate Minerals**. [S. l.: s. n.], 1970.

HAN, S *et al.* Cell size selection for site-specific crop management. **Transactions of the ASAE**, [s. l.], v. 37, n. 1, p. 19–26, 1994.

HENGL, Tomislav; ROSSITER, David G.; STEIN, Alfred. Soil sampling strategies for spatial prediction by correlation with auxiliary maps. **Australian Journal of Soil Research**, [s. l.], v. 41, n. 8, p. 1403–1422, 2003.

HUTENGs, Christopher *et al.* In situ and laboratory soil spectroscopy with portable visible-to-near-infrared and mid-infrared instruments for the assessment of organic carbon in soils. **Geoderma**, [s. l.], v. 355, n. August, p. 113900, 2019. Disponível em: <https://doi.org/10.1016/j.geoderma.2019.113900>.

ISO, B S I. 20473: 2007 Optics and photonics. **Spectral bands**, [s. l.], p. 10, 2007.

KAISER, M.; ELLERBROCK, R. H.; GERKE, H. H. Cation Exchange Capacity and Composition of Soluble Soil Organic Matter Fractions. **Soil Science Society of America Journal**, [s. l.], v. 72, n. 5, p. 1278–1285, 2008.

LAUZON, John D *et al.* Spatial variability of soil test phosphorus, potassium, and pH of Ontario soils. **Agronomy Journal**, [s. l.], v. 97, n. 2, p. 524–532, 2005. Disponível em: <https://www.researchgate.net/publication/237348952>. Acesso em: 3 dez. 2019.

LAWRENCE, Patrick G. *et al.* Guiding soil sampling strategies using classical and spatial statistics: A review. **Agronomy Journal**, [s. l.], v. 112, n. 1, p. 493–510, 2020.

LIU, Fei Tony; TING, Kai Ming; ZHOU, Zhi Hua. Isolation forest. *In:* , 2008. **Proceedings - IEEE International Conference on Data Mining, ICDM**. [S. l.: s. n.], 2008. p. 413–422.

LIU, Fei Tony; TING, Kai Ming; ZHOU, Zhi-Hua. **Isolation-based Anomaly Detection**. [S. l.: s. n.], 2012.

MALLARINO, Antonio P; WITTRY, David J. Efficacy of grid and zone soil sampling approaches for site-specific assessment of phosphorus, potassium, pH, and organic matter. **Precision Agriculture**, [s. l.], v. 5, n. 2, p. 131–144, 2004. Disponível em: Acesso em: 3 dez. 2019.

MAYRINK, G. O. *et al.* Determination of chemical soil properties using diffuse reflectance and ion-exchange resins. **Precision Agriculture**, [s. l.], v. 20, n. 3, p. 541–561, 2018. Disponível em: <https://doi.org/10.1007/s11119-018-9597-y>.

METZGER, Konrad *et al.* Mid-infrared spectroscopy as an alternative to laboratory extraction for the determination of lime requirement in tillage soils. **Geoderma**, [s. l.], v. 364, n. December 2019, p. 114171, 2020. Disponível em: <https://doi.org/10.1016/j.geoderma.2020.114171>.

NAWAR, Said *et al.* Delineation of Soil Management Zones for Variable-Rate Fertilization: A Review. *In: ADVANCES IN AGRONOMY*. [S. l.]: Academic Press Inc., 2017. v. 143, p. 175–245.

NGUYEN, T.T.; JANIK, L.J.; RAUPACH, M. Diffuse Reflectance Infrared Fourier Transform (DRIFT) Spectroscopy in Soil Studies. [s. l.], n. 1982, p. 49–67, 1991.

NOVAIS, Roberto Ferreira *et al.* **Fertilidade do solo**. [S. l.: s. n.], 2007.

PADARIAN, Jose; MINASNY, Budiman; MCBRATNEY, Alex B. Machine learning and soil sciences: A review aided by machine learning tools. **SOIL**, [s. l.], v. 6, n. 1, p. 35–52, 2020. Disponível em: Acesso em: 29 jun. 2022.

PARFITT, R. L.; GILTRAP, D. J.; WHITTON, J. S. Contribution of organic matter and clay minerals to the cation exchange capacity of soils. **Communications in Soil Science and Plant Analysis**, [s. l.], v. 26, n. 9–10, p. 1343–1355, 1995.

PEREIRA, Gustavo Willam *et al.* Smart-Map: An Open-Source QGIS Plugin for Digital Mapping Using Machine Learning Techniques and Ordinary Kriging. **Agronomy**, [s. l.], v. 12, n. 6, p. 1350, 2022. Disponível em: <https://www.mdpi.com/2073-4395/12/6/1350>.

PUDEŁKO, Anna; CHODAK, Marcin. Estimation of total nitrogen and organic carbon contents in mine soils with NIR reflectance spectroscopy and various chemometric methods. **Geoderma**, [s. l.], v. 368, n. February, 2020.

QGIS DEVELOPMENT TEAM. **QGIS Geographic Information System**. [S. l.]: QGIS Association, 2023. Disponível em: <https://www.qgis.org>. Acesso em: 9 maio 2023.

RADOČAJ, Dorijan *et al.* The effect of soil sampling density and spatial autocorrelation on interpolation accuracy of chemical soil properties in arable cropland. **Agronomy**, [s. l.], v. 11, n. 12, 2021.

RODRIGUES, Khalil de Menezes *et al.* Spatial Variability in Soil Fertility and Particle Size and Their Effects on Sugarcane Yield. **Sugar Tech**, [s. l.], v. 18, n. 1, p. 39–48, 2016.

ROSSEL, R. A. Viscarra; BEHRENS, T. Using data mining to model and interpret soil diffuse reflectance spectra. **Geoderma**, [s. l.], v. 158, n. 1–2, p. 46–54, 2010. Disponível em: <http://dx.doi.org/10.1016/j.geoderma.2009.12.025>.

SILVA, Elisângela Benedet *et al.* A Regional Legacy Soil Dataset for Prediction of Sand and Clay Content with Vis-Nir-Swir, in Southern Brazil. **Revista Brasileira de Ciência do Solo**, [s. l.], v. 43, n. August, 2019. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-06832019000100304&tlng=en.

STENBERG, Bo *et al.* Visible and Near Infrared Spectroscopy in Soil Science. **Advances in Agronomy**, [s. l.], v. 107, n. C, p. 163–215, 2010a.

STENBERG, Bo *et al.* Visible and Near Infrared Spectroscopy in Soil Science. [s. l.], v. 107, p. 163–215, 2010b. Disponível em: <http://dx.doi.org/10.1016/S0065-2113>. Acesso em: 26 jun. 2022.

TERRA, Fabrício S. *et al.* Soil fertility assessment by Vis-NIR spectroscopy: Predicting soil functioning rather than availability indices. **Geoderma**, [s. l.], v. 337, n. 3, p. 368–374, 2019. Disponível em: <https://doi.org/10.1007/s11119-018-9597-y>.

TEXAS, Instruments. DLP® NIRscan™ Nano EVM User's Guid. **Texas Instruments Incorporated**, [s. l.], n. June 2015, 2015. Disponível em: <http://www.ti.com/tool/dlpnirnanoevm>.

THORLABS INC. **Detectors Fotocondutores para o Infravermelho**. [S. l.], 2020a. Disponível em: https://www.thorlabs.com/newgrouppage9.cfm?objectgroup_id=6479. Acesso em: 18 abr. 2020.

THORLABS INC. **MIR Photovoltaic Detectors, HgCdTe (MCT)**. [S. l.], 2020b. Disponível em: https://www.thorlabs.com/newgrouppage9.cfm?objectgroup_id=11319. Acesso em: 18 abr. 2020.

VALENTE, Sárvio *et al.* Definition of management zones in coffee production fields based on apparent soil electrical conductivity. **Scientia Agricola**, [s. l.], v. 69, n. 3, p. 173–179, 2012a.

VALENTE, Sárvio *et al.* The relationship between apparent soil electrical conductivity and soil properties. [s. l.], p. 683–690, 2012b. Disponível em: www.ccarevista.ufc.br.

VAN ROSSUM, Guido; DRAKE, Fred L. **Python 3 Reference Manual**. Scotts Valley, CA: CreateSpace, 2009.

VISCARRA ROSSEL, R. A. *et al.* Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. **Geoderma**, [s. l.], v. 131, n. 1–2, p. 59–75, 2006.

WIJewardane, Nuwan K.; GE, Yufeng; MORGAN, Cristine L.S. Moisture insensitive prediction of soil properties from VNIR reflectance spectra based on external parameter orthogonalization. **Geoderma**, [s. l.], v. 267, n. January, p. 92–101, 2016. Disponível em: <http://dx.doi.org/10.1016/j.geoderma.2015.12.014>.

WILDING, L. P. Spatial variability: it's documentation, accommodation and implication to soil surveys. *In:* , 1985, Las Vegas. (J Bouma & D. R Nielsen, Org.) **Soil Spatial Variability**. Pudoc. Las Vegas: [s. n.], 1985. p. 166–194. Disponível em: Acesso em: 30 jun. 2022.

XU, Shengxiang *et al.* Comparison of multivariate methods for estimating selected soil properties from intact soil cores of paddy fields by Vis–NIR spectroscopy. **Geoderma**, [s. l.], v. 310, n. September 2017, p. 29–43, 2018a.

XU, Shengxiang *et al.* Comparison of multivariate methods for estimating selected soil properties from intact soil cores of paddy fields by Vis–NIR spectroscopy. **Geoderma**, [s. l.], v. 310, p. 29–43, 2018b.

ZEAITER, M.; RUTLEDGE, D. Preprocessing Methods. **Comprehensive Chemometrics**, [s. l.], v. 3, p. 121–231, 2009. Disponível em: Acesso em: 28 jun. 2022.

CAPÍTULO I

DESENVOLVIMENTO DE UM SISTEMA PORTÁTIL PARA PREDIÇÃO DE ATRIBUTOS FÍSICO-QUÍMICOS DO SOLO POR MEIO DE ESPECTRORRADIOMETRIA

RESUMO

A análise de solo é uma ferramenta importante para a tomada de decisão em atividades agrícolas. A qualidade dos modelos utilizados para descrever a variabilidade dos atributos do solo está diretamente relacionada à quantidade de amostras coletadas e às técnicas de modelagem utilizadas. A obtenção de um número suficientemente alto de amostras de solo para representar de forma satisfatória essa variabilidade é um desafio enfrentado na Agricultura de Precisão (AP). A espectrometria é um método rápido de análise que permite a obtenção de vários atributos do solo em uma única leitura do espectro. Nesse contexto, a espectrometria pode ser uma ferramenta valiosa na predição de atributos do solo, possibilitando aprimorar os modelos gerados, ao guiar a coleta de amostras, ou sendo utilizada como covariável nos processos de modelagem, o que pode reduzir os custos sem comprometer a qualidade dos modelos. O objetivo foi desenvolver um sistema portátil para aquisição dos espectros de reflectância e utilizá-los para gerar e avaliar modelos de aprendizado de máquina para a predição de atributos do solo, empregando um espectrorradiômetro portátil de baixo custo. Foram utilizadas 417 amostras de solo coletadas em Minas Gerais (MG) e Mato Grosso (MT). Os espectros de reflectância das amostras de solo foram pré-processados e utilizados para o cálculo de índices espectrais. Este banco de dados foi utilizado para treinamento e validação dos modelos *Support Vector Machine* (SVM), *Random Forest* (RF), *Partial Least Square Regressor* (PLSR) e *Multi-Layer Perceptron* (MLP). Os modelos apresentaram R^2 entre 0,50 e 0,68 e RPD entre 1,42 e 1,80 para os dados de teste. Foi possível prever satisfatoriamente o Ca, CTC e o P_{rem} nas duas áreas de estudo, o pH, K, CTC_p e MO para as amostras de MT e o Mg para as amostras de MG, possivelmente devido aos diferentes materiais de origem e processos de formação dos diferentes solos analisados. O equipamento desenvolvido foi capaz de realizar a predição de atributos do solo.

Palavras-chave: Solo, Espectrometria, Machine Learning, Agricultura de Precisão, Atributos do Solo.

ABSTRACT

Soil analysis is a crucial tool for decision-making in agricultural activities. The quality of models used to describe soil attribute variability is directly related to the number of samples collected and the modeling techniques employed. Obtaining a sufficiently high number of soil samples to adequately represent this variability is a challenge faced in Precision Agriculture (PA). Spectrometry, a rapid analysis method, enables the acquisition of multiple soil attributes in a single spectrum reading. In this context, spectrometry can be a valuable tool in predicting soil attributes, allowing for the refinement of generated models by guiding sample collection or serving as a covariate in modeling processes, thereby reducing costs without compromising model quality. The objective was to develop a portable system for acquiring reflectance spectra and to utilize them to generate and evaluate machine learning models for predicting soil attributes, employing a low-cost portable spectroradiometer. A total of 417 soil samples collected in Minas Gerais (MG) and Mato Grosso (MT) were used. The reflectance spectra of the soil samples were pre-processed and used to calculate spectral indices. This database was employed to train and validate Support Vector Machine (SVM), Random Forest (RF), Partial Least Square Regressor (PLSR), and Multi-Layer Perceptron (MLP) models. The models exhibited R^2 values between 0.50 and 0.68 and RPD values between 1.42 and 1.80 for the test data. Satisfactory predictions were achieved for Ca, CTC, and Prem in both study areas, as well as pH, K, CTCP, and MO for MT samples, and Mg for MG samples. This may be attributed to the diverse origin materials and formation processes of the analyzed soils. The developed equipment successfully performed the prediction of soil attributes.

Keywords: Soil, Spectrometry, Machine Learning, Precision Agriculture, Soil Attributes.

1 INTRODUÇÃO

Um dos grandes desafios da agricultura moderna é o aumento da produtividade com eficiência econômica e ambiental. Para que uma cultura possa expressar todo seu potencial produtivo, é necessário realizar a correção da fertilidade e do pH do solo (Fageria; Baligar, 2008). Esta correção é realizada em função dos níveis dos atributos físico-químicos. No entanto, tais características apresentam variabilidade espacial, o que abre oportunidade para utilização de técnicas de Agricultura de Precisão (AP). Porém, a aplicação dessas técnicas está condicionada ao conhecimento da variabilidade espacial desses atributos. Assim, geralmente é necessária uma elevada quantidade de amostras georreferenciadas para a geração dos mapas de variabilidade espacial.

A geração de mapas por meio de interpolação é amplamente utilizada para caracterizar a variabilidade espacial das características do solo. A qualidade destes mapas depende diretamente da densidade amostral. Estudos indicam que para caracterizar de forma fiel a variabilidade espacial dos atributos do solo, muitas vezes são necessários grids amostrais densos. Para o fósforo disponível, potássio, pH e matéria orgânica por exemplo, seria necessário um grid com distâncias mínimas entre amostras de aproximadamente 30 a 45 metros (Lauzon et al., 2005; Mallarino & Wittry, 2004).

A necessidade de usar elevadas densidades amostrais para caracterização da variabilidade espacial dos atributos do solo resultam em elevado custo com análises laboratoriais, o que pode levar os produtores e profissionais a adotarem um número de amostras insuficientes para geração de mapas confiáveis. Técnicas que utilizam variáveis facilmente mensuradas e correlacionadas com atributos do solo, como a condutividade elétrica e a espectrometria, podem ser úteis pois permitem trabalhar com maiores densidades amostrais. Conhecendo a variabilidade espacial desta variável, é possível criar zonas de manejo, ou ainda, guiar a amostragem do solo. Assim, é possível reduzir o número de amostras necessárias para utilização de técnicas de AP (Nawar *et al.*, 2017).

A espectrometria de amostras de solo utilizando os comprimentos de onda do visível e infravermelho próximo tem sido utilizada com sucesso na predição dos atributos físico-químicos do solo. Estudos com a utilização da espectrometria no visível e infravermelho próximo (VIS-NIR), para predição dos atributos do solo, apresentaram R^2 da ordem de 0,65 a 0,88, dependendo do atributo estudado e das metodologias utilizadas (Silva *et al.*, 2019; Terra *et al.*, 2019; Xu *et al.*, 2018b), o que indica a viabilidade da metodologia nesta aplicação. A espectrometria por ser um método de análise rápida, não usar reagentes químicos e possibilitar

a obtenção de múltiplos atributos em uma única leitura do espectro, vem chamando atenção (Mayrink *et al.*, 2018). Essa abordagem aumenta a segurança e reduz custos e tempo de análise. No entanto, é um método que resulta em um grande número de variáveis obtidas. Portanto, há a necessidade da utilização de modelos multivariados, seleção de variáveis mais importantes para predição dos atributos do solo.

Atualmente, a utilização de modelos de *Machine Learning* (ML) aliados a técnicas quimiométricas vêm ganhando destaque na predição das propriedades do solo utilizando espectrometria VIS-NIR por meio de modelos multivariados (Barra *et al.*, 2021). Modelos de ML precisam de menos suposições sobre geração e distribuição dos dados. Assim, os modelos de ML são úteis na espectrometria pois os conjuntos de dados gerados possuem grande número de variáveis de entrada em relação ao número de amostras. Isso resulta em um conjunto de dados grande, com elevada colinearidade entre as variáveis (Bzdok; Altman; Krzywinski, 2018).

Apesar da facilidade advinda do uso da espectrometria VIS-NIR para a predição dos atributos do solo, os sensores utilizados são geralmente caros. Além do alto custo, tais equipamentos geralmente são utilizados em laboratórios, inviabilizando sua utilização em campo. Tais equipamentos não são capazes de gerar ou aplicar os modelos multivariados para predição de atributos do solo. Os recentes avanços tecnológicos permitiram a miniaturização e redução no custo de produção de diversos tipos de sensores e processadores. Existem hoje no mercado diversos espectrômetros e kits de desenvolvimento portáteis com baixo custo, o que torna possível o desenvolvimento de sistemas que utilizam espectrometria e que sejam capazes de aplicar os modelos gerados. Assim, há a necessidade de desenvolvimento de instrumentos portáteis capazes de aplicar os modelos multivariados, especialmente para predições realizadas em campo. Portanto, o objetivo deste trabalho foi o desenvolvimento e avaliação de um sistema portátil para aquisição da reflectância de amostras de solo para geração dos modelos de *Machine Learning* de predição de atributos do solo.

2 MATERIAL E MÉTODOS

O sistema portátil proposto possui dois sensores controlados por um computador de placa única, que também é responsável pelo armazenamento e processamento das informações, exibição da interface gráfica em uma tela touchscreen e acionamento das fontes de luz. A estrutura para acoplamento dos sensores, fontes de luz e posicionamento das cubetas foi impressa em impressora 3D (Figura 1c).

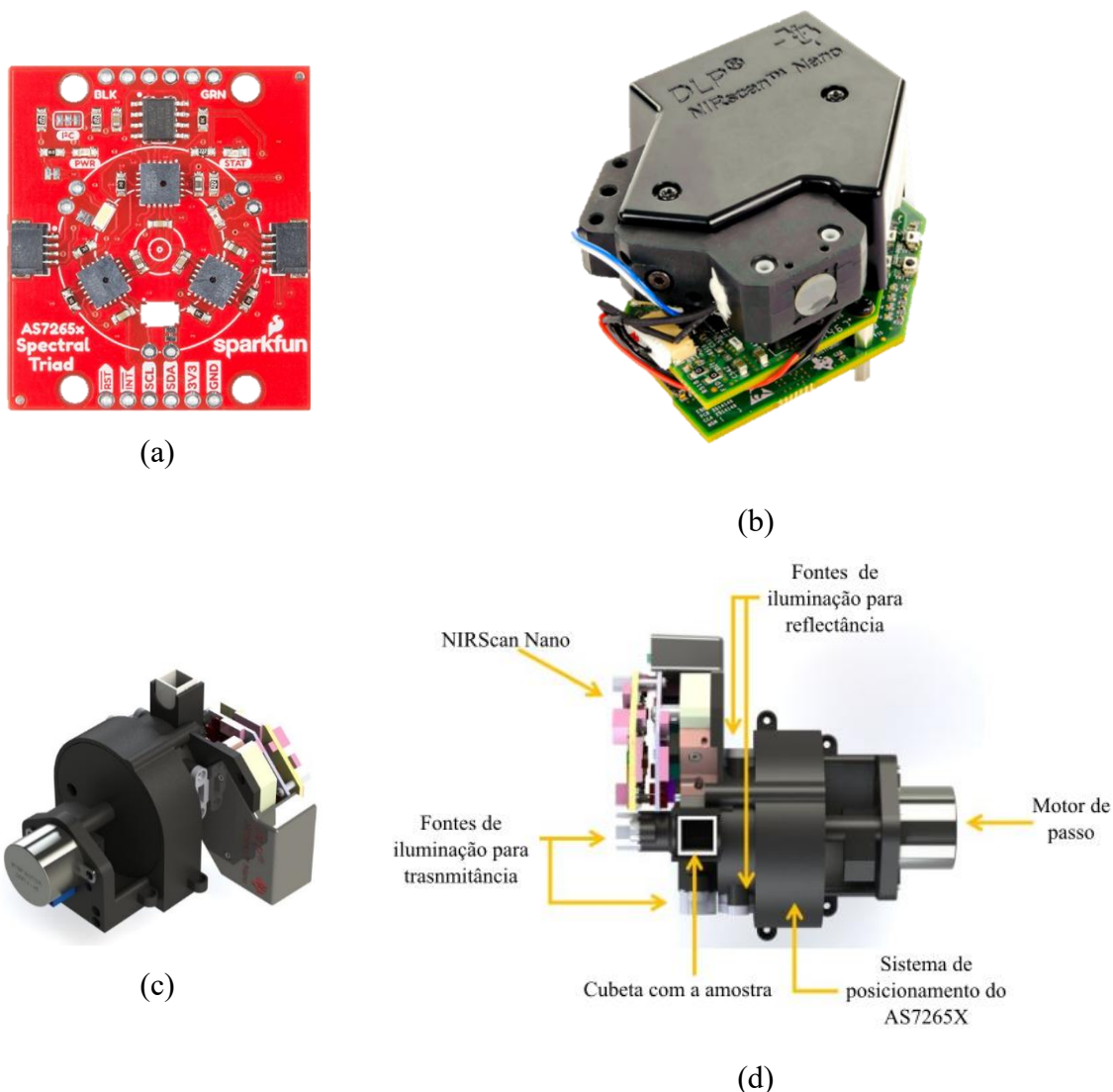


Figura 1 – Sensor *spectral* modelo AS7265x Triad (a), DLP® NIRScan™ Nano (b), módulo desenvolvido para acoplamento dos sensores, amostra e fontes de iluminação(c) e vista superior do módulo com suas partes constituintes (d).

O sistema foi construído com base nos módulos sensores *Spectral Triad* produzido pela *SparkFun* (Figura 1a) e o DLP® NIRScan™ Nano produzido pela *Texas Instruments* (Figura 1b). O sensor produzido pela *SparkFun* é composto por três fotodiodos AS7265x sensíveis a

seis faixas espectrais cada, totalizando 18 canais compreendidos entre 410 e 940 nm. Devido a disposição dos sensores na placa, não seria possível alinhar os três fotodiodos com a amostra de solo de forma simultânea. Assim houve a necessidade de desenvolver um sistema para o alinhamento dos fotodiodos com as amostras de solo. Este sistema de posicionamento foi desenvolvido com a utilização de um motor de passo responsável por rotacionar toda a placa do *Spectral Triad*, alinhando um fotodiodo por vez com a amostra.

O *DLP® NIRScan™ Nano* é um módulo de avaliação que utiliza a tecnologia *DLP®* para redirecionar o espectro de luz recebido para um fotodiodo do tipo InGaAs, sensível aos comprimentos de onda compreendidos entre 900 e 1700 nm. Devido as características construtivas e de funcionamento, este sensor pode ser configurado para apresentar diferentes resoluções, faixa das bandas e intervalo de comprimentos de onda. Para este estudo o sensor foi configurado para resultar em 228 comprimentos de ondas mensurados, compreendidos entre 900 e 1600 nm. Neste caso, cada banda apresentou em média 3 nm de faixa.

Como fonte de radiação eletromagnética foram utilizadas quatro lâmpadas halógenas 12 volts com 5 watts de potência, de tal forma que o sistema de aquisição de transmitância contou com uma lâmpada para cada sensor, posicionada do lado oposto da cubeta (Figura 1d). Devido à disposição dos três leds na placa do *Triad Spectroscopy Sensor* não permitir a iluminação da amostra de forma simultânea, foi desenvolvido um sistema de iluminação com duas lâmpadas halógenas para aquisição da reflectância com este sensor. Para manter a tensão de alimentação das lâmpadas constante, foi utilizado um regulador de tensão *step-up*. O acionamento das lâmpadas foi realizado por um módulo com quatro relés, comandado pelo software. Além das lâmpadas, o sistema conta com uma bateria de lítio de 12 volts e 4000 mAh, suficiente para mantê-lo em funcionamento por aproximadamente oito horas.

O computador de placa única utilizado foi uma *BeagleBone Black*, responsável por hospedar o sistema operacional *Debian* e o software para a configuração dos sensores, a aquisição de dados, o controle dos sistemas de iluminação e a interface gráfica. A interface gráfica foi desenvolvida no *Qt designer* e o código para controle do sistema e processamento das informações foi desenvolvido utilizando a linguagem de programação *Python 3*.

A Figura 2 mostra o fluxo de energia e de informações do sistema portátil desenvolvido.

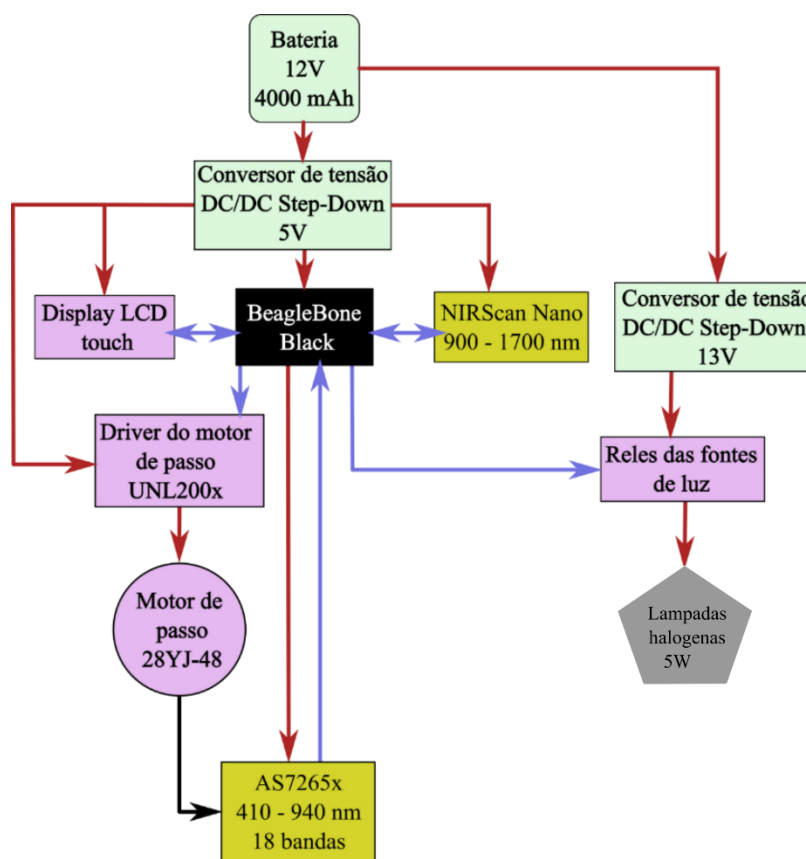


Figura 2 - Fluxograma mostrando o fluxo de informações (azul), energia (vermelho) e movimento (preto) no sistema portátil.

Para o treinamento e validação dos modelos preditivos, foram utilizadas 568 amostras de solo. Dessas, 417 foram coletadas por Adão et al. (2022) no município de Barbacena/MG (Área 1). Essas amostras originaram-se de áreas com diferentes sistemas de cultivo (plantio direto, cultivo mínimo e cultivo convencional), onde foram coletadas a 0-30 cm de profundidade. Cada amostra foi composta por 8 subamostras coletadas num raio de 10m. O solo foi classificado como Cambissolo háplico. A coleta foi realizada utilizando um trado do tipo holandês

Na Área 2 (Tabaporã/MT), foram coletadas 151 amostras de solo em dois talhões, um com cultivo convencional e outro com plantio direto. As amostras foram coletadas a 0-20 cm de profundidade, sendo que cada amostra foi composta por 10 a 15 subamostras coletadas num raio de 10m do ponto amostrado. O solo dos talhões foi classificado como Latossolo Vermelho-amarelo Distrófico (Figura 3). A coleta foi realizada utilizando um trado do tipo calador.

As amostras foram submetidas à secagem ao ar e, posteriormente, peneiradas com uma malha de 2 mm. Uma parte dessas amostras foi destinada à obtenção dos espectros de reflectância, enquanto outra parte foi encaminhada ao laboratório de análises de solo para a realização das análises padrão. As análises incluíram pH em água, fósforo disponível (P), potássio (K), cálcio (Ca^{2+}), magnésio (Mg^{2+}), capacidade de troca de cátions efetiva (t),

capacidade de troca de cátions potencial (T), fósforo remanescente (P-rem) e matéria orgânica (MO).

Os métodos específicos de análise foram os seguintes: para os elementos P e K, utilizou-se o Extrator Mehlich-1; para os íons Ca^{2+} e Mg^{2+} , foi empregado o extrator KCl a 1 mol/L; e a determinação da matéria orgânica (MO) foi realizada multiplicando-se o carbono orgânico por 1,724, utilizando o método Walkley-Black. A análise estatística descritiva dos atributos adquiridos em cada localidade foi conduzida com o propósito de identificar a variabilidade presente em cada área de estudo.

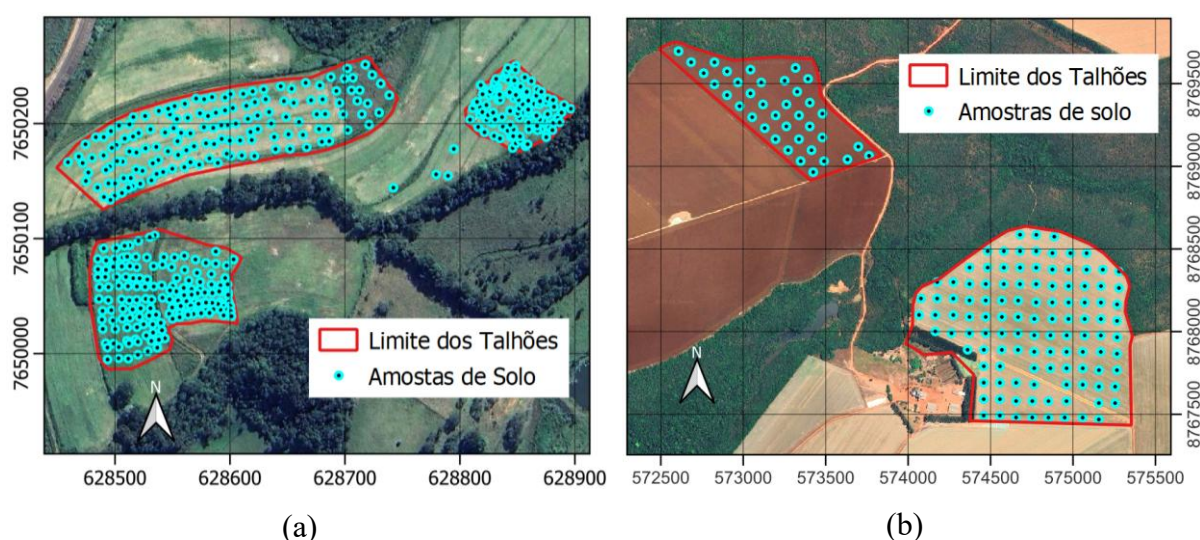


Figura 3 - Mapas das áreas amostradas: em Barbacena/MG, com 417 amostras (a); e em Tabaporã/MT, com 151 amostras (b).

A calibração dos sensores do sistema portátil foi realizada utilizando uma cubeta preenchida com sulfato de bário (BaSO_4) compactado. Esta cubeta foi utilizada como branco padrão. O processo de calibração consistiu em obter os valores de número digital para o branco padrão, considerando valores de refletância iguais a um para a leitura do branco padrão. Assim os valores de refletância das amostras foram obtidos dividindo os valores de número digital das amostras, pelo número digital do branco padrão para cada comprimento de onda de forma individual (Equação 1). Devido a alteração no espectro emitido pelas lâmpadas halógenas variar com a temperatura das mesmas, o processo de calibração foi realizado sempre que o equipamento era ligado e a cada 10 minutos com o equipamento funcionando.

$$R_i^{\text{Amostra}} = \frac{DN_i^{\text{Amostra}}}{DN_i^{\text{Branco}}} \quad \text{Equação 1}$$

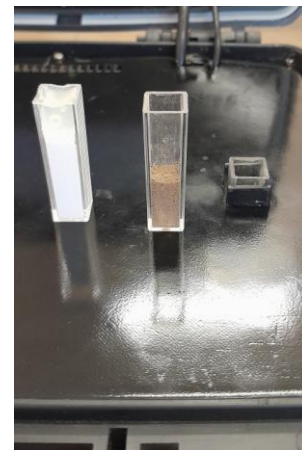
em que:

- $R_i^{Amostra}$ - Reflectância da amostra no comprimento de onda i (adimensional);
- $DN_i^{Amostra}$ - Número digital lido no comprimento de onda i da amostra, (adimensional); e
- DN_i^{Branco} - Número digital lido no comprimento de onda i do branco padrão.

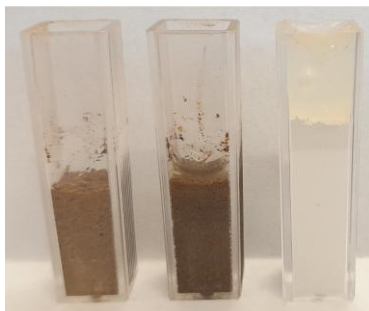
Os espectros de reflectância foram adquiridos utilizando 2 cm³ de solo, contidos em cubetas de poliestireno de 4,5 cm³ (Figura 4c). Para cada amostra de solo, foram obtidos quatro espectros de reflectância com a amostra seca. Posteriormente, as amostras foram saturadas com 1,5 ml de água destilada (Figura 4c), homogeneizadas, e duas novas leituras foram realizadas para cada amostra saturada. Dessa forma, o espectro de cada amostra no estado seco foi composto pela média de quatro leituras. O espectro da amostra saturada foi obtido a partir da média de duas leituras, devido à estabilidade das leituras das amostras saturadas, não justificando a necessidade de aumentar o número de leituras para tais amostras.



(a)



(b)



(c)



(d)

Figura 4 – Espectroradiômetro de baixo custo (a); detalhe mostrando o branco padrão, cubeta com amostra de solo e local de inserção das amostras no espectrorradiômetro (b); cubetas com amostras de solo seco, solo saturado e branco padrão (c); e software desenvolvido para coleta e visualização de dados (d).

Para a correção dos efeitos do espalhamento multiplicativo e do tamanho das partículas nos espectros, que comumente ocorre em amostras sólidas, foi utilizado o método Standard Normal Variate (SNV) (Equação 2) (Zeaiter; Rutledge, 2009). Após a aplicação do SNV, os espectros obtidos das amostras secas e saturadas foram combinados e, em seguida, foram gerados índices espectrais combinando os comprimentos de onda dois a dois. Foram produzidos índices por diferença, índices normalizados e índices por diferença normalizada, conforme apresentado nas Equações 3, 4 e 5 respectivamente.

$$x_{i,j}^{SNV} = \frac{x_{i,j} - \bar{x}_i}{\sqrt{\frac{\sum_{j=1}^p (x_{i,j} - \bar{x}_i)^2}{p - 1}}} \quad \text{Equação 2}$$

em que:

$x_{i,j}^{SNV}$ - elemento do espectro transformado (adimensional);

$x_{i,j}$ - elemento correspondente do espectro original i no comprimento de onda j (adimensional);

$\bar{x}_i$ - média do espectro i (adimensional);

p - número de comprimentos de onda no espectro.

$$ID = B_2 - B_1 \quad \text{Equação 3}$$

$$IN = \frac{B_2}{B_1} \quad \text{Equação 4}$$

$$IDN = \frac{B_2 - B_1}{B_2 + B_1} \quad \text{Equação 5}$$

em que:

ID , IN e IDN - índices por diferença, normalizados e por diferença normalizada (adimensionais);

B_1 e B_2 - bandas utilizadas nos índices (adimensional).

O banco de dados utilizado para o treinamento dos modelos de *Machine Learning* foi composto pelos espectros das amostras secas e saturadas corrigidos pelo SNV e os índices calculados. A saída dos modelos foram os teores de nutrientes disponíveis no solo. A Figura 5 mostra os processos desde a aquisição dos espectros até a criação do banco de dados utilizado para treinamento e validação dos modelos de *Machine Learning*.

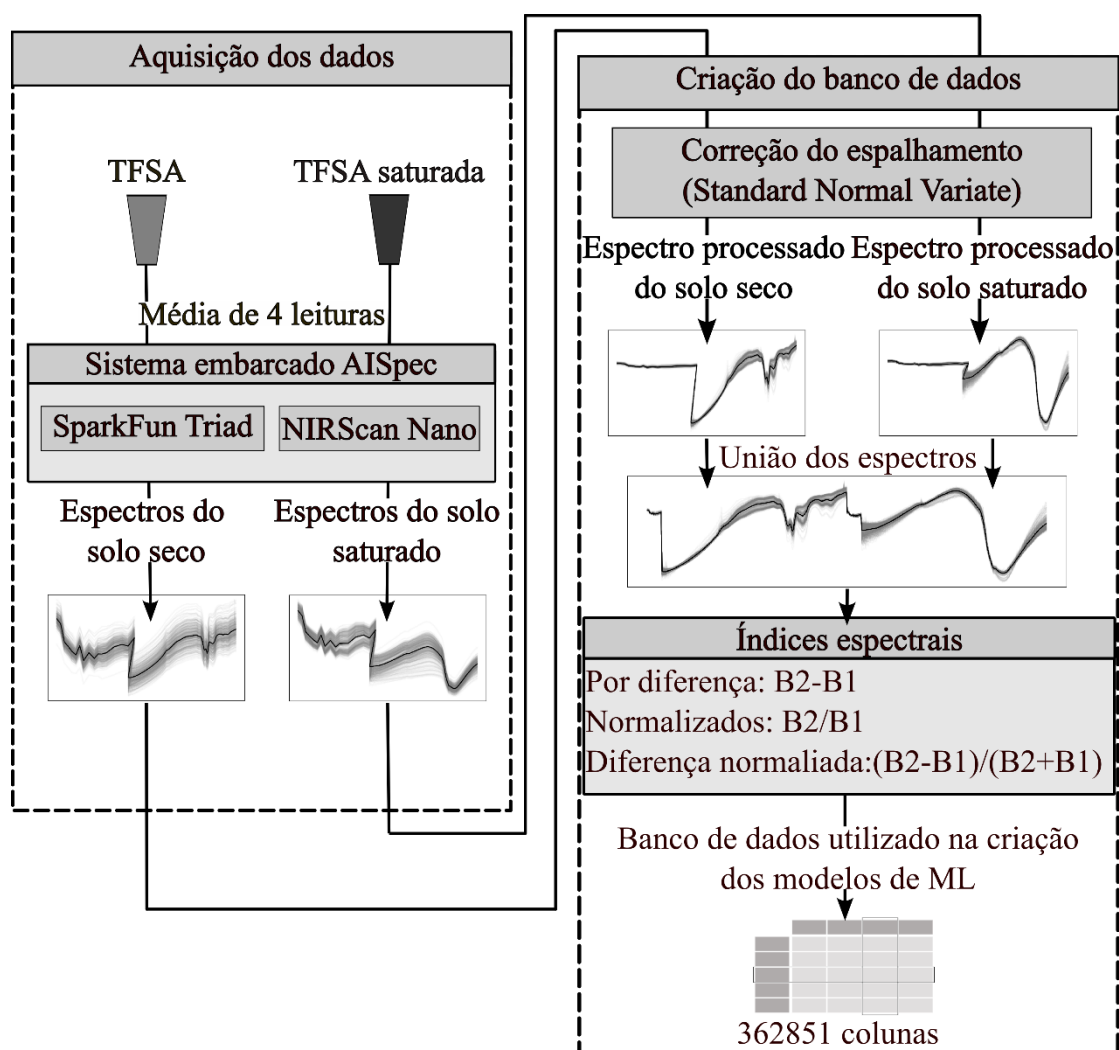


Figura 5 – Fluxograma mostrando os passos para criação do banco de dados utilizado para desenvolvimento e validação dos modelos de *Machine Learning*.

O banco de dados gerado foi dividido de tal forma que 85% dos dados foram utilizados para seleção de variáveis e treinamento dos modelos e 15% dos dados utilizados para validação e avaliação dos modelos. O pré-processamento dos dados incluiu a remoção de dados anômalos por meio do algoritmo *Isolation Forest*, (Liu; Ting; Zhou, 2008, 2012) e em seguida as variáveis foram escalonadas a fim de deslocar a média para zero e apresentar desvio padrão igual a um.

Foram treinados os modelos Random Forest (RF), Support Vector Machine Regression (SVM), Partial Least Squares Regressor (PLSR) e Multi-layer Perceptron (MLP), além de um modelo Stacking que utilizou como entrada as saídas dos modelos citados anteriormente. Em função da grande quantidade de variáveis, a seleção das variáveis de maior importância para cada um dos modelos foi realizada por meio de um algoritmo de seleção genética com validação cruzada utilizando a biblioteca sklearn. Os parâmetros utilizados foram cinco folds, 60 gerações, 100 indivíduos e número máximo de variáveis igual a 300. Após a seleção das variáveis, os hiperparâmetros foram ajustados utilizando otimização Bayesiana usando o

processo Gaussiano. Tanto a seleção das variáveis quanto o ajuste de hiperparâmetros foram realizados utilizando validação cruzada do tipo K-fold com cinco folds. Foi utilizada como métrica de decisão a raiz quadrada do erro médio quadrático (RMSE) e foram realizados separadamente para cada variável predita. As variáveis de maior importância para os modelos foram determinadas utilizando o método *Permutation Feature Importance* (Altmann et al., 2010).

As métricas utilizadas para avaliação dos modelos foram o RMSE, o coeficiente de determinação (R^2) e a razão entre predição e desvio (RPD). As métricas foram avaliadas tanto para os dados de treinamento utilizando validação cruzada com cinco *folds*, quanto para os dados de teste. O melhor modelo gerado para cada variável foi determinado com base nas métricas dos dados de teste.

Os processos, desde a divisão dos dados em dados de treino e teste até a seleção do melhor modelo, estão exemplificados na Figura 6.

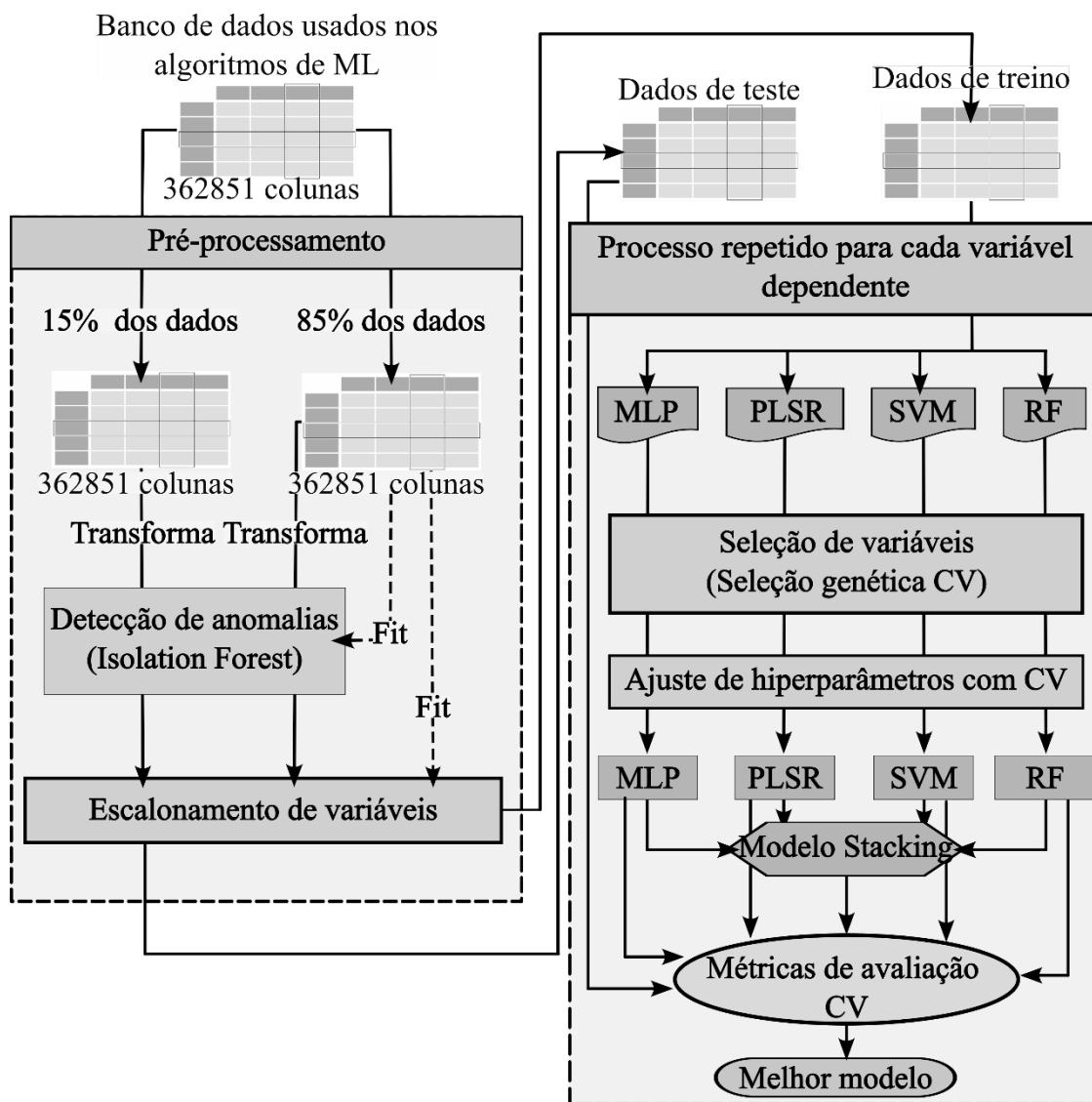


Figura 6 - Fluxograma mostrando o pré-processamento utilizado, a seleção de variáveis, o ajuste dos hiper parâmetros e a avaliação das métricas para seleção do melhor modelo para cada variável predita.

3 RESULTADOS E DISCUSSÃO

A Tabela 1 apresenta a estatística descritiva dos resultados obtidos nas análises de laboratório para as 417 amostras de solo coletadas na Área 1 e para as 151 coletadas na Área 2, utilizadas na calibração e validação dos modelos de ML. Em geral, as amostras coletadas na Área 1 apresentaram maior variabilidade dos nutrientes quando comparada as amostras coletadas na Área 2, possivelmente devido aos diferentes usos do solo e sistemas de cultivos presentes nas áreas de coleta na Área 1, além do maior número de amostras. Devido à

proximidade entre as áreas onde as amostras foram coletadas tanto na Área 2 quanto na Área 1, seria pouco provável tal variabilidade ser proporcionada por diferentes materiais de origem.

Para as amostras da Área 1, os atributos P, K, Ca^{2+} , Mg^{2+} e t, apresentaram elevada variabilidade ($\text{CV} \geq 35\%$), a T apresentou variabilidade moderada ($15\% \leq \text{CV} < 35\%$) e o pH e P_{rem} baixa variabilidade ($\text{CV} < 15\%$) (WILDING, 1985) citado por (Xu *et al.*, 2018a). Para as amostras da Área 2, os atributos com elevada variabilidade foram o P, K e P_{rem} , variabilidade moderada Ca^{2+} , Mg^{2+} , SB, t, T, e MO, e com baixa variabilidade o pH. Estes resultados indicam que há variabilidade nas áreas de estudo, tanto na Área 1 quanto na Área 2, o que aumenta as chances de obter modelos baseados em espectrometria com boa performance (Abdi *et al.*, 2012)

Tabela 1: Estatística descritiva dos resultados das análises laboratoriais para as amostras de Minas Gerais (Área 1) e Mato Grosso (Área 2).

	Média		Mediana		Desvio padrão		CV (%)		Mínimo		Máximo	
	MT	MG	MT	MG	MT	MG	MT	MG	MT	MG	MT	MG
pH (H_2O)	6,18	5,56	6,14	5,55	0,29	0,4	4,72	7,28	5,76	4,37	7,78	6,78
P	15,26	20,97	11,70	12,88	10,16	25,08	66,54	119,6	1,10	2,05	56,60	194,52
K	58,77	55,41	64,00	47,1	41,62	31,18	70,82	56,27	7,00	12,1	199,00	219,4
Ca^{2+}	2,69	1,17	2,57	1,03	0,70	0,66	26,15	56,62	0,99	0,11	5,82	3,81
Mg^{2+}	0,90	0,29	0,88	0,25	0,23	0,19	25,95	63,89	0,43	0,03	1,68	1,04
P_{rem}	24,43	25,29	21,90	25,72	12,26	3,53	50,20	13,94	11,10	16,32	147,40	33,88
t	3,74	1,86	3,53	1,69	0,93	1,01	24,85	54,16	1,45	0,81	6,46	16,66
T	6,88	6,12	6,80	6,07	1,11	0,92	16,16	15,12	3,25	3,39	9,82	12,3
MO	2,65	-	2,66	-	0,79	-	29,83	-	0,80	-	4,79	-

CV – coeficiente de variação (%); P – fósforo pelo método da resina (mg/dm^3); K – potássio (mg/dm^3); Ca^{2+} – cálcio (cmol/dm^3); Mg^{2+} – magnésio (cmol/dm^3); P_{rem} – fósforo remanescente (mg/L); t – capacidade de troca de cátions efetiva (cmol/dm^3); T – capacidade de troca de cátions potencial (cmol/dm^3); MO – matéria orgânica (dag/kg).

Entre as correlações presentes nos dados das análises de laboratório, podemos destacar a correlação positiva entre o Ca^{2+} vs. Mg^{2+} , Ca^{2+} e Mg^{2+} vs. T, Ca^{2+} e Mg^{2+} vs. pH e as correlações negativas entre Ca^{2+} , Mg^{2+} vs. saturação por alumínio (Figura 7). Possivelmente devido ao processo de correção de acidez do solo com calcário dolomítico, resultando em um aumento no pH, Ca, Mg e uma redução do alumínio trocável e saturação por alumínio (de Vargas *et al.*, 2019). A elevada correlação entre variáveis dos dados de treinamento pode fazer com que os modelos gerados apresentem resultados significativos, mesmo que o modelo não tenha a capacidade de prever determinada variável, por estar prevendo alguma outra variável correlacionada. Os atributos K e P apresentaram baixa correlação com as demais variáveis.

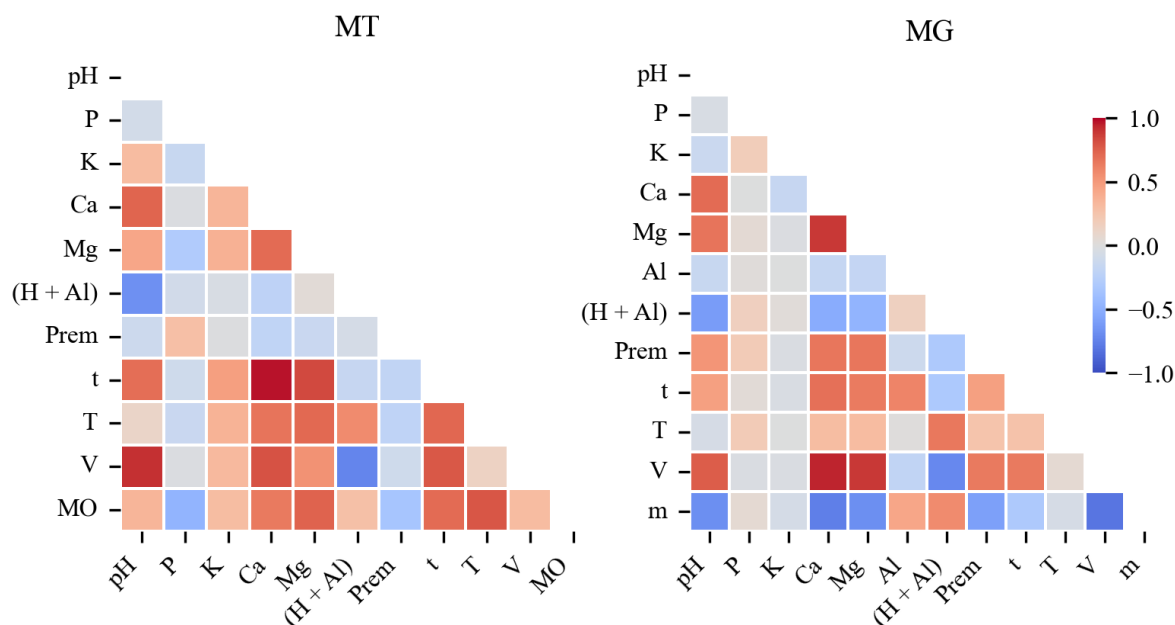


Figura 7 - Matriz de correlação dos atributos do solo para os resultados das análises de laboratório.

Os espectros das amostras saturadas apresentaram menores valores de reflectância no visível quando comparados aos espectros das amostras secas. Esta redução ocorreu de forma uniforme em todo o espectro captado com o sensor *Triad Spectrometry*, sem que a forma do espectro fosse alterada (Figura 8a). Observando a diferença entre os espectros das amostras secas e saturadas na região do NIR, obtido com o *NIRScan Nano*, nota-se maior ruído próximo ao comprimento de 900 nm e um pico de absorção próximo a 1450 nm (Figura 8b) possivelmente relacionados com absorção do grupamento O-H da água (Nguyen; Janik; Raupach, 1991; Stenberg *et al.*, 2010b).

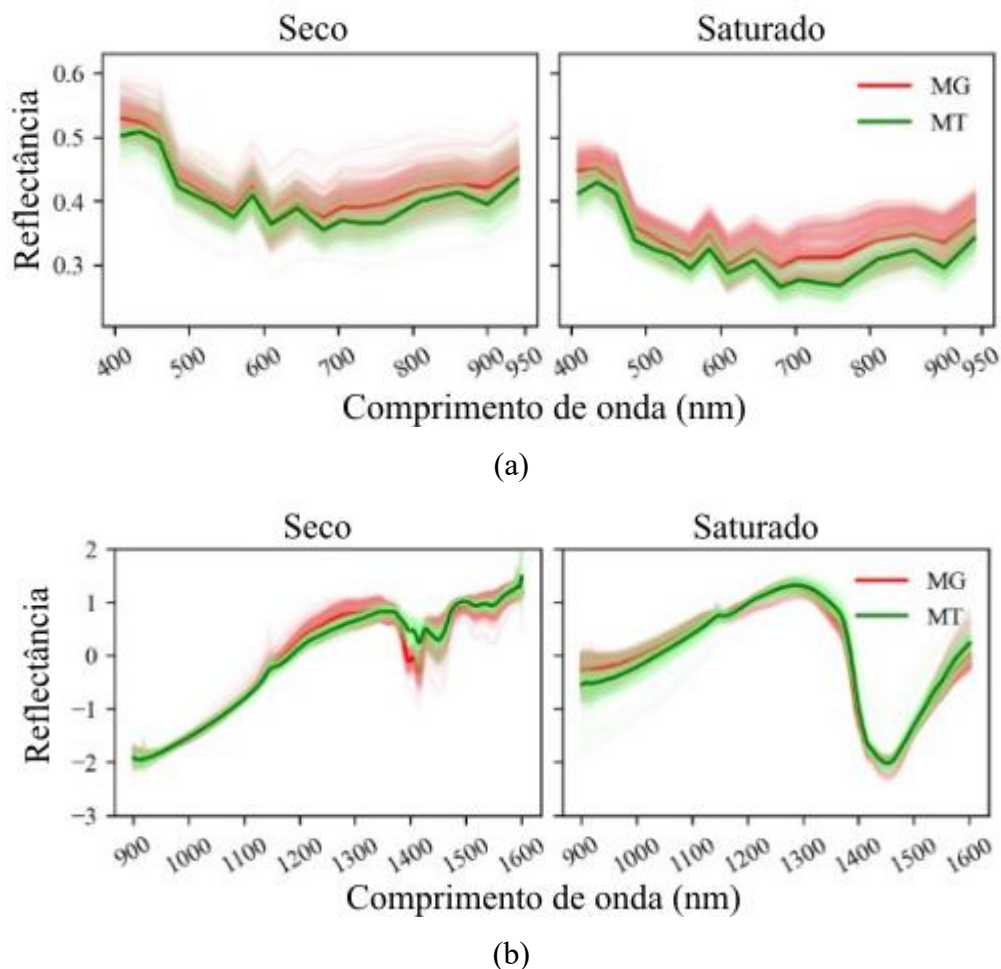


Figura 8 - Assinaturas espectrais das 417 amostras de solo da Área 1 e das 151 amostras da Área 2 secas e saturadas. (a) valores de reflectância obtidos com o sensor *Triad Spectral*, e (b) espectros obtidos com o *NIRScan Nano* após pré-processamento com o SNV. Nas cores claras estão os espectros individuais e em escuro os espectros médios.

Segundo (Chang *et al.*, 2001) os modelos que utilizam NIR para predição de atributos do solo podem ser classificados quanto a sua capacidade preditiva como excelentes com RPD maior que 2, aceitáveis com RPD entre 1,4 e 2 e modelos não confiáveis com RPD menor que 1,4.

Para os atributos P e K, os modelos gerados para Área 1 não foram capazes de estimar o valor dos atributos. Nestes casos os valores de R^2 foram próximos a zero e RPD de aproximadamente um (Tabela 2). Para as amostras da Área 2 o modelo RF apresentou resultados aceitáveis para o K com RPD e R^2 de respectivamente 1,52 e 0,55. Porém não foi capaz de realizar a predição do P disponível.

Na literatura são encontrados diferentes valores para as métricas dos modelos de predição do P disponível. Comparando estudos que utilizaram mesmo tratamento nas amostras de solo, são encontrados modelos com valores de RPD e R^2 variando entre 1,17 até 2,95 e 0,22 até 0,83

respectivamente (Dunn *et al.*, 2002; Wijewardane; Ge; Morgan, 2016; Xu *et al.*, 2018b). O P disponível dos solos se encontra ligados a oxidróxidos de ferro e alumínio (Coutinho *et al.*, 2019). Dessa forma, o P é quantificado de forma indireta pelo espectro NIR, que é sensível a estes oxidróxidos nos comprimentos de onda entre 1900 e 2500 nm e no infravermelho médio (MIR) nos números de onda de 1220 a 1127 cm^{-1} . Portanto, a incapacidade do modelo para predição do P pode estar relacionada ao alcance máximo do sistema portátil ser 1700 nm, não contemplando a região do espectro sensível ao P O que indica que o sensor não é adequado para mensurar este elemento diretamente em amostras de solo.

Tabela 2 - Resumo das métricas para os modelos gerados para predição dos atributos do solo, em negrito está destacada o melhor modelo para o atributo em questão.

Atributo	Métrica	Área 1					Área 2				
		SVR	PLSR	RF	MLPR	Stacking	SVR	PLSR	RF	MLPR	Stacking
pH	RMSE	0,36	0,38	0,34	0,38	0,36	0,18	0,32	0,18	0,16	0,17
	R ²	0,31	0,24	0,40	0,22	0,31	0,37	0,19	0,33	0,51	0,43
	RPD	1,20	1,15	1,29	1,13	1,21	1,28	0,72	1,23	1,44	1,34
P	RMSE	25,27	16,48	25,20	33,89	24,39	6,83	9,86	5,96	20,51	12,18
	R ²	0,00	0,00	0,00	0,00	0,01	0,23	0,12	0,42	0,00	0,11
	RPD	0,95	1,00	0,98	0,72	1,01	1,16	0,80	1,33	0,38	0,65
K	RMSE	35,08	33,63	33,02	33,43	33,37	34,09	45,20	31,26	165,56	73,45
	R ²	0,05	0,03	0,07	0,05	0,05	0,47	0,07	0,55	0,10	0,29
	RPD	0,98	1,02	1,04	1,03	1,03	1,40	1,05	1,52	0,28	0,65
Ca	RMSE	0,48	0,54	0,47	0,80	0,56	0,40	2,87	0,44	0,51	0,40
	R ²	0,61	0,48	0,60	0,26	0,44	0,53	0,14	0,40	0,21	0,52
	RPD	1,60	1,39	1,60	0,94	1,35	1,48	0,20	1,31	1,15	1,47
Mg	RMSE	0,15	0,16	0,15	0,15	0,21	0,17	0,81	0,17	0,97	0,32
	R ²	0,49	0,43	0,50	0,48	0,21	0,41	0,12	0,43	0,00	0,14
	RPD	1,41	1,33	1,42	1,39	0,99	1,32	0,28	1,35	0,24	0,72
t	RMSE	0,53	0,58	0,52	7,99	0,52	0,51	2,14	0,60	7,85	1,06
	R ²	0,55	0,48	0,57	0,00	0,57	0,62	0,20	0,50	0,00	0,05
	RPD	1,49	1,38	1,53	0,09	1,52	1,65	0,40	1,43	0,10	0,80
T	RMSE	0,83	0,79	0,77	0,80	0,85	0,86	0,90	0,57	3,14	0,96
	R ²	0,06	0,16	0,19	0,12	0,01	0,17	0,10	0,64	0,14	0,18
	RPD	1,03	1,09	1,11	1,06	1,00	1,12	1,07	1,69	0,31	1,00
P _{rem}	RMSE	2,12	2,24	1,97	2,22	1,94	3,18	3,36	3,44	7,27	10,22
	R ²	0,60	0,56	0,66	0,57	0,67	0,66	0,62	0,61	0,46	0,22
	RPD	1,59	1,50	1,72	1,52	1,74	1,75	1,66	1,62	0,77	0,55
MO	RMSE	-	-	-	-	-	0,41	1,14	0,48	31,86	1,07
	R ²	-	-	-	-	-	0,68	0,30	0,56	0,00	0,19
	RPD	-	-	-	-	-	1,80	0,65	1,53	0,02	0,69

SVR – support vector regressor; PLSR – partial least squares regressor; RF – random forest; MLPR – multi layer perceptron regressor; RMSE – raiz quadrada do erro médio quadrático; R² – coeficiente de determinação; RPD – razão entre predição e desvio; P – fósforo disponível; K – potássio; Ca – cálcio; Mg – magnésio; P-rem – fósforo remanescente; t – capacidade de troca de cátions efetiva; T – capacidade de troca de cátions potencial; MO – matéria orgânica.

A predição do pH do solo para as amostras da Área 2 apresentaram resultados aceitáveis, com valores de RMSE, R² e RPD de 0,16, 0,51 e 1,44 respectivamente, obtidos pelo modelo MLPR. Para as amostras da Área 1 os modelos para predição não foram capazes de predizer de forma aceitável o pH do solo, sendo RF o modelo que apresentou melhor performance, com valores de RMSE, R² e RPD iguais a 0,34, 0,40 e 1,29 respectivamente (Tabela 2). Na literatura encontramos modelos com métricas variando de 0,37, 0,31 e 1,21 para RMSE, R² e RPD

respectivamente, até valores de R^2 de 0,74 (Clingensmith; Grunwald; Wani, 2019; Gomez; Lagacherie; Coulouma, 2012; Viscarra Rossel *et al.*, 2006). Portanto, os resultados obtidos neste trabalho são condizentes com valores encontrados por outros autores. A ampla diferença entre valores encontrados na literatura pode estar relacionada aos diferentes sensores utilizados, diferentes métodos para aquisição dos espectros, processamento dos dados e características das áreas amostradas.

O algoritmo de seleção genética resultou na seleção de nove índices como entrada do modelo MLPR para pH nas amostras da Área 2. Destes índices um foi do tipo índice por diferença normalizada, três índices normalizados e quatro índices por diferença. Foram utilizadas um total de 18 bandas, sendo 10 provenientes das amostras secas e oito das amostras saturadas para compor os índices selecionados (Figura 9a).

Observando na Figura 9 os comprimentos de ondas que foram selecionados, podemos notar valores próximos a 1500, 1524, 1000 e, 1100 nm, relacionados a compostos orgânicos como a amina, amidos e compostos aromáticos. Também foi selecionado o comprimento de onda de 1418 nm, próximo aos comprimentos de onda sensíveis a caulinita e a hidroxilas presentes nos óxidos de ferro (Nguyen; Janik; Raupach, 1991; Rossel; Behrens, 2010).

O índice de maior impacto no modelo MLPR foi o índice normalizado utilizando os comprimentos de onda de 976 e 1548 nm (Figura 9c), ambos de amostras secas, estes valores estão próximos a valores sensíveis a aminas e amidos, compostos de pH básico e que possuem sua sorção ao solo determinadas pelo pH do meio (FÁBREGA *et al.*, 1998). Os resultados preditos versus observados são mostrados na Figura 9b.

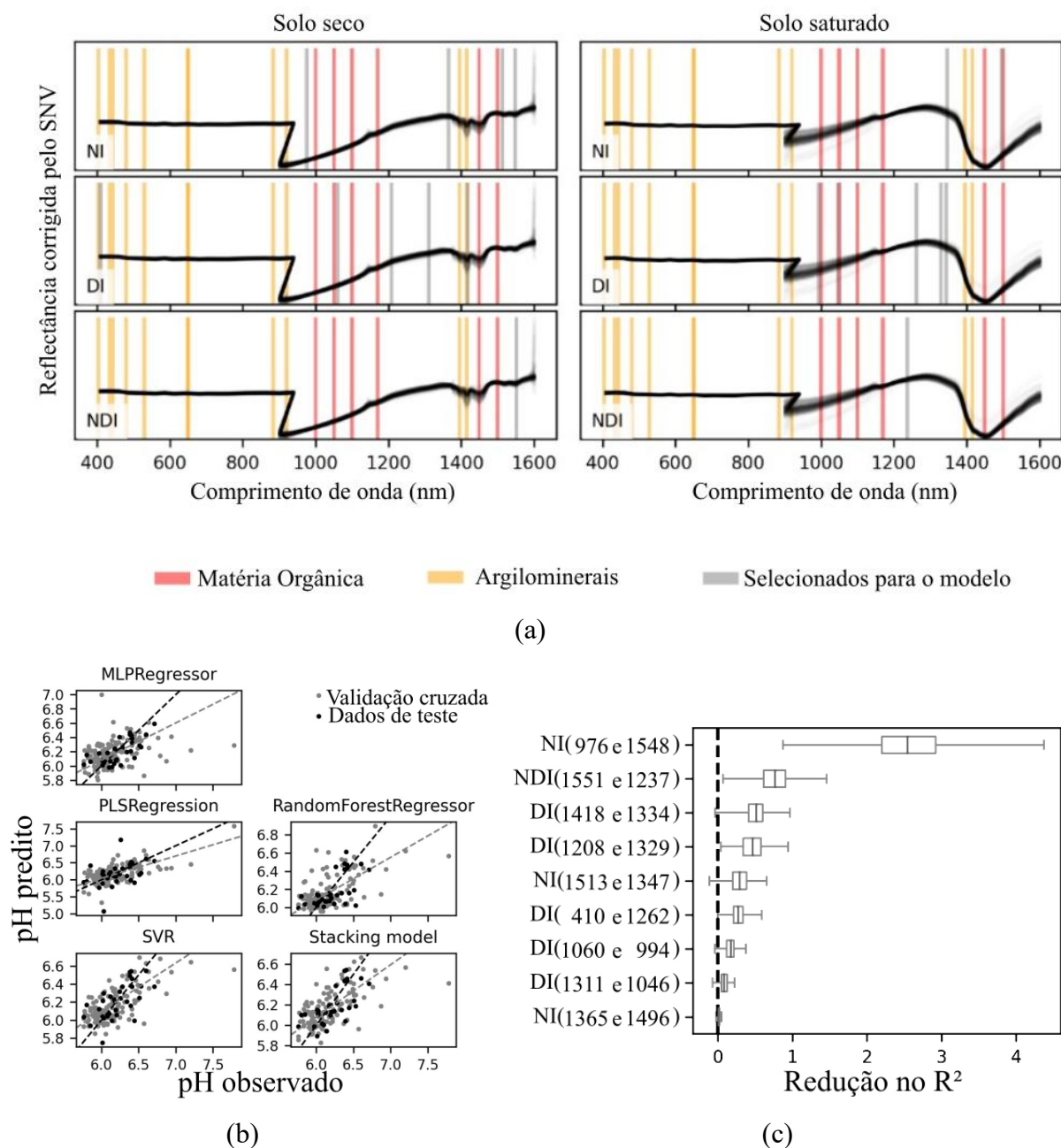


Figura 9 – Comprimentos de ondas seleccionados para o modelo MLPR para predição do pH nas amostras da Área 2 (a), valores preditos vs. observados para os dados de treino e teste dos modelos gerados (b) e importância das variáveis no modelo (c).

Dentre os modelos de predição gerados para o potássio disponível para as amostras da Área 2, o modelo RF apresentou a melhor performance, com valores de RMSE, R^2 e RPD iguais a 31,26, 0,55 e 1,52 respectivamente (Tabela 2), indicando capacidade de predição aceitável. Para este modelo foram seleccionadas 184 variáveis, sendo 61 índices normalizados, 54 índices por diferença, 68 índices por diferença normalizada, além de uma banda. No total foram utilizados 259 comprimentos de onda, sendo 124 comprimentos de onda de amostras secas e

135 de amostras saturadas (Figura 10a). Foram selecionados comprimentos de onda abrangendo todo o espectro, com maior intensidade a partir de 960 nm, indicando que o sensor NIRScan Nano apresenta maior importância para o modelo. No visível foram selecionados comprimentos de ondas próximos aos sensíveis aos óxidos de ferro (Nguyen; Janik; Raupach, 1991; Rossel; Behrens, 2010). As variáveis de maior importância para o modelo (Figura 10c) apresentam comprimentos de onda próximos a 560 e 950 nm (comprimentos de onda relacionados aos óxidos de ferro), e 1450 nm relacionado tanto aos óxidos de ferro quanto a componentes da matéria orgânica. Os valores preditos pelo modelo versus os valores observados são mostrados na Figura 10b. Para a Área 1 os modelos gerados apresentaram resultados insatisfatórios na predição do potássio disponível.

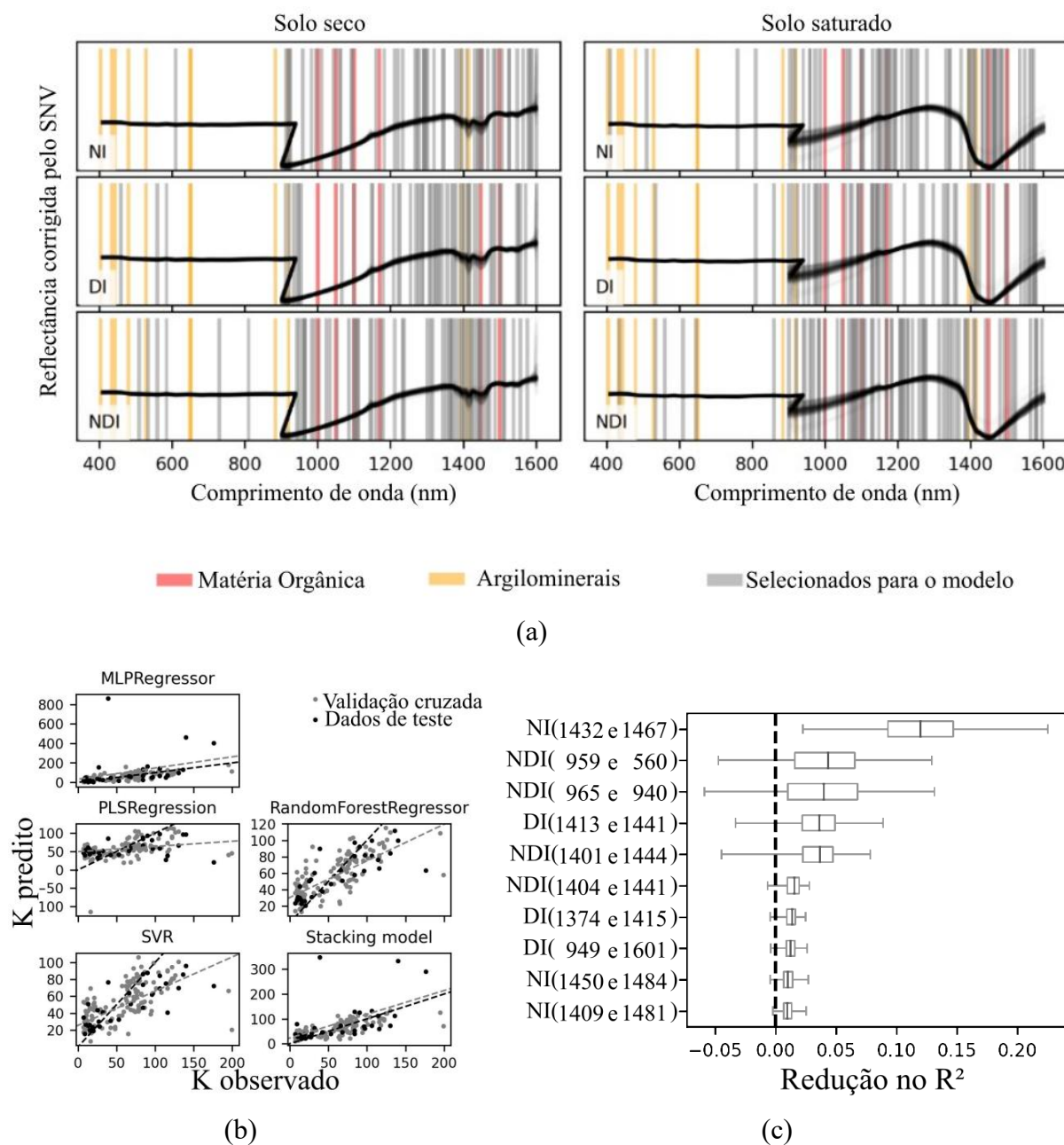
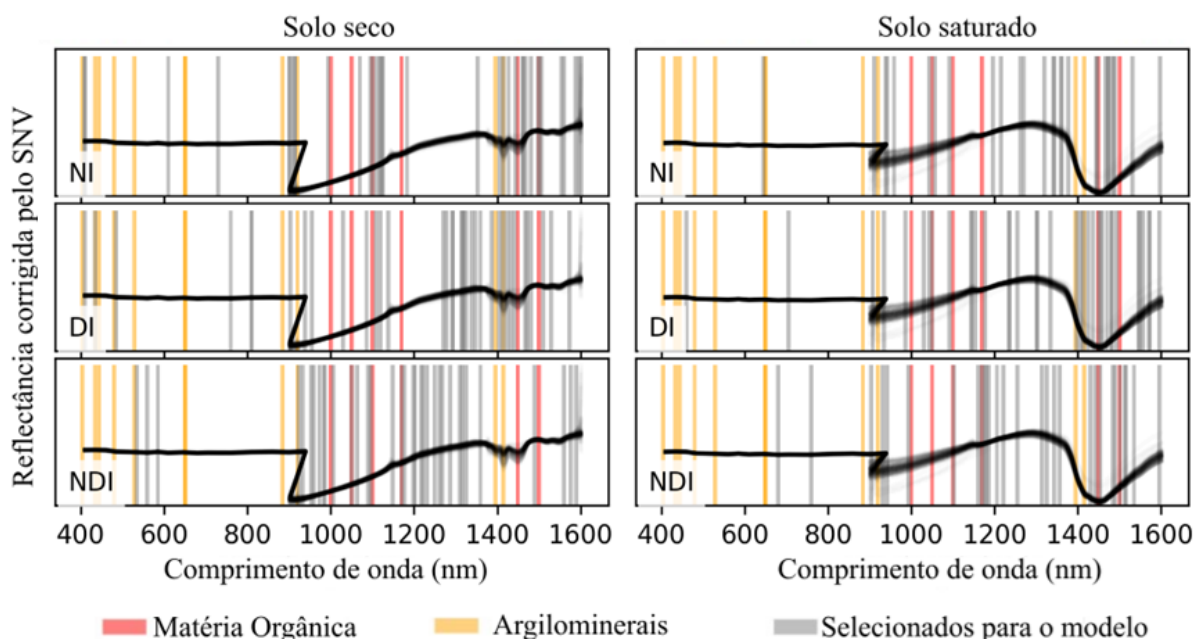


Figura 10 – Comprimentos de ondas seleccionados para o modelo RF para predição do potássio disponível nas amostras da Área 2 (a), valores preditos vs. observados para os dados de treino e teste dos modelos gerados (b) e dez variáveis de maior importância no modelo RF(c).

Os modelos de predição do cálcio apresentaram resultados satisfatórios para as duas áreas de estudo. Para as amostras da Área 2 o modelo SVR foi o que apresentou melhores resultados, com valores de RMSE, R^2 e RPD iguais à 0,40, 0,53, 1,48. O melhor modelo para a predição do cálcio nas amostras da Área 1 foi o RF, com métricas iguais a 0,47, 0,60 e 1,60 para o RMSE, R^2 e RPD respectivamente (Tabela 2).

A seleção genética resultou na utilização de 100 índices espectrais para modelo SVR gerado para as amostras da Área 2. Sendo 31 índices normalizados, 37 índices por diferença e

32 índices por diferença normalizada, utilizando um total de 91 bandas dos espectros das amostras secas e 82 bandas dos espectros das amostras saturadas. A maioria dos comprimentos de ondas selecionados pertencem ao NIR, com alguns comprimentos de onda selecionados no visível onde podemos notar a seleção de comprimentos de onda próximos aos sensíveis aos argilominerais e óxidos de ferro (Figura 11a). No NIR podemos notar que algumas variáveis selecionadas coincidem com comprimentos de onda sensíveis aos componentes da matéria orgânica, e em menor quantidade foram selecionados comprimentos de ondas próximos aos sensíveis aos argilominerais (Figura 11a).



(a)

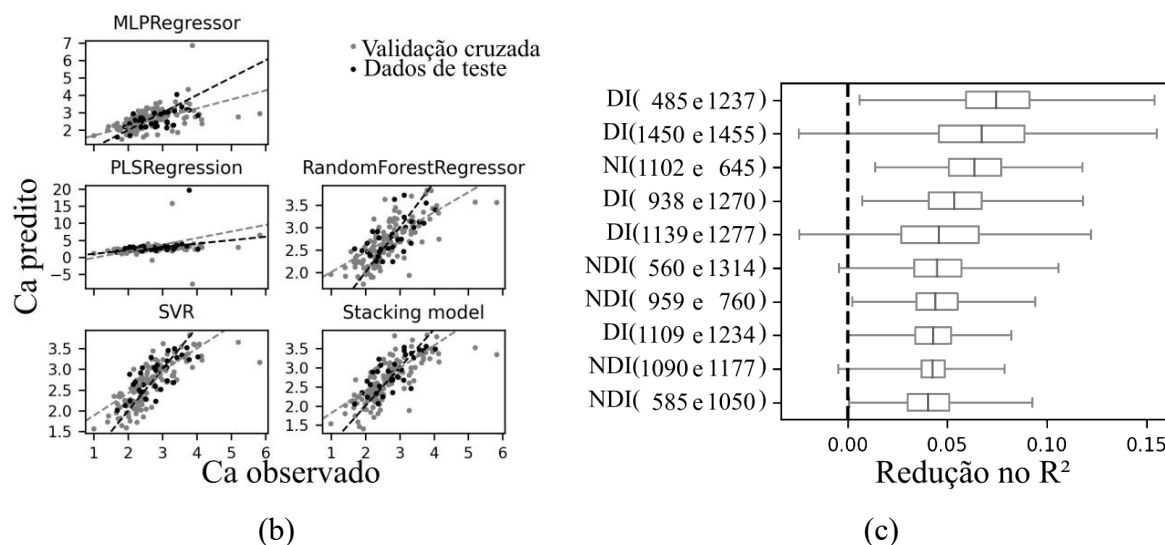


Figura 11 - Comprimentos de onda selecionados para predição do cálcio pelo modelo SVR para o MT (a), valores preditos vs. observados para os dados de treino e teste dos modelos gerados (b) e as dez variáveis de maior importância no modelo SVR(c).

As variáveis de maior importância para o modelo SVR para predição do Ca nas amostras da Área 2 (Figura 11c) são formadas por índices contendo os comprimentos de onda 485nm que é próximo ao comprimento de onda sensível a goethita, 645nm que está próximo a comprimentos de ondas sensível a Hematita, 1455 nm sensível a moléculas de água presentes nos óxidos de ferro e o comprimento de onda de 1102nm que está relacionado com a presença de compostos aromáticos presentes na MO do solo (Nguyen; Janik; Raupach, 1991; Rossel; Behrens, 2010). O comparativo entre valores preditos e observados pelos modelos gerados para o Ca nas amostras de MT são mostrados na Figura 11b.

O modelo RF gerado para a predição do cálcio na Área 1 selecionou um total de 144 índices espectrais, dos quais 56 são índices normalizados, 46 índices por diferença e 42 índices por diferença normalizada, no total foram utilizadas 213 bandas, sendo 105 bandas dos espectros de amostras secas e 108 bandas dos espectros de amostras saturadas (Figura 12a).

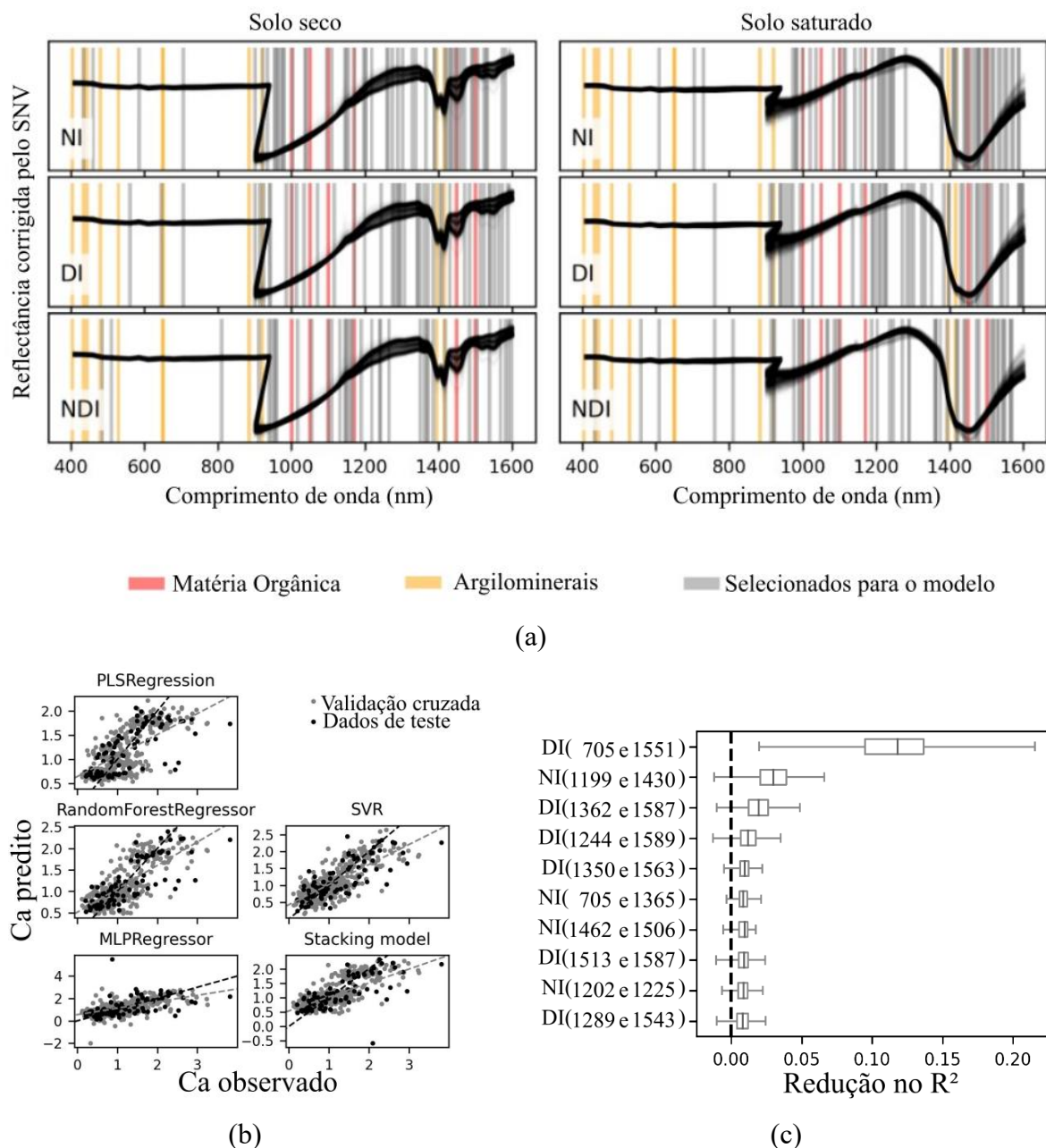


Figura 12 - Comprimentos de onda selecionados para predição do cálcio pelo modelo RF para Área 1 (a) valores preditos vs. observados para os dados de treino e teste dos modelos gerados (b) e as dez variáveis de maior importância no modelo RF(c).

Para este modelo também podemos observar uma maior quantidade de comprimentos de ondas selecionados no NIR, sendo que alguns se sobrepõem a comprimentos de onda relacionados aos compostos orgânicos. No espectro visível foram selecionados os comprimentos de ondas sensíveis a hematita e goethita 440, 650 e 529nm (Nguyen; Janik; Raupach, 1991; Rossel; Behrens, 2010). O índice de maior impacto no modelo foi do tipo índice por diferença, que utilizou os comprimentos de onda 705 e 1551nm (Figura 12c). Os valores preditos e observados para todos os modelos de predição de Ca gerados para MG estão na Figura 12b.

Tanto o modelo para predição do Ca gerado para a Área 1 quanto o gerado para Área 2 apresentaram grande parte dos comprimentos de ondas selecionados no NIR, indicando a maior importância do NIR nos modelos de predição (Figura 11a e Figura 12a). Observamos também que o modelo gerado para Área 2 apresenta uma distribuição da importância das variáveis mais distribuída, enquanto o modelo RF selecionado para amostras da Área 1 o índice por diferença das bandas 750nm, 1151 nm foi o principal preditor do modelo. (Figura 11c e Figura 12c).

Somente para as amostras da Área 1 foi possível gerar modelos de predição do Mg confiáveis, sendo que o modelo RF apresentou melhor performance, com RMSE, R^2 e RPD iguais à 0,15, 0,50 e 1,42 respectivamente. O algoritmo de seleção genética resultou na seleção de 58 índices, sendo, 25 índices normalizados, 18 índices por diferença e 15 índices por diferença normalizada. Estes índices utilizaram um total de 102 comprimentos de onda, dos quais 51 são do espectro de amostras secas, e 51 dos espectros de amostras saturadas (Figura 13a). Foram selecionados comprimentos de onda próximos a 480 e 920nm, que são relacionados a presença de goethita. Comprimentos de onda próximos de 1000, 1050 e 1100nm também foram selecionados, e estão relacionados a presença de aminas, carboidratos e compostos aromáticos respectivamente. Podemos destacar também a seleção de comprimentos de onda próximos a 930, 1400 e 1455nm, correspondentes a absorção por vibração das ligações de moléculas de água e hidroxilas associadas aos óxidos de ferro (Nguyen; Janik; Raupach, 1991; Rossel; Behrens, 2010). A variável de maior importância foi o índice por diferença normalizada entre as bandas 1286nm do espectro das amostras secas e de 1341nm dos espectros das amostras saturadas (Figura 13c). Este índice bem como outros presentes entre as dez variáveis de maior importância é composto por uma combinação de espectro das amostras seca com espectro das amostras úmidas, indicando que a utilização da relação entre os espectros secos e úmidos pode trazer informações relativas à quantidade deste nutriente no solo. Os valores preditos e observados para todos os modelos gerados estão mostrados na Figura 13b.

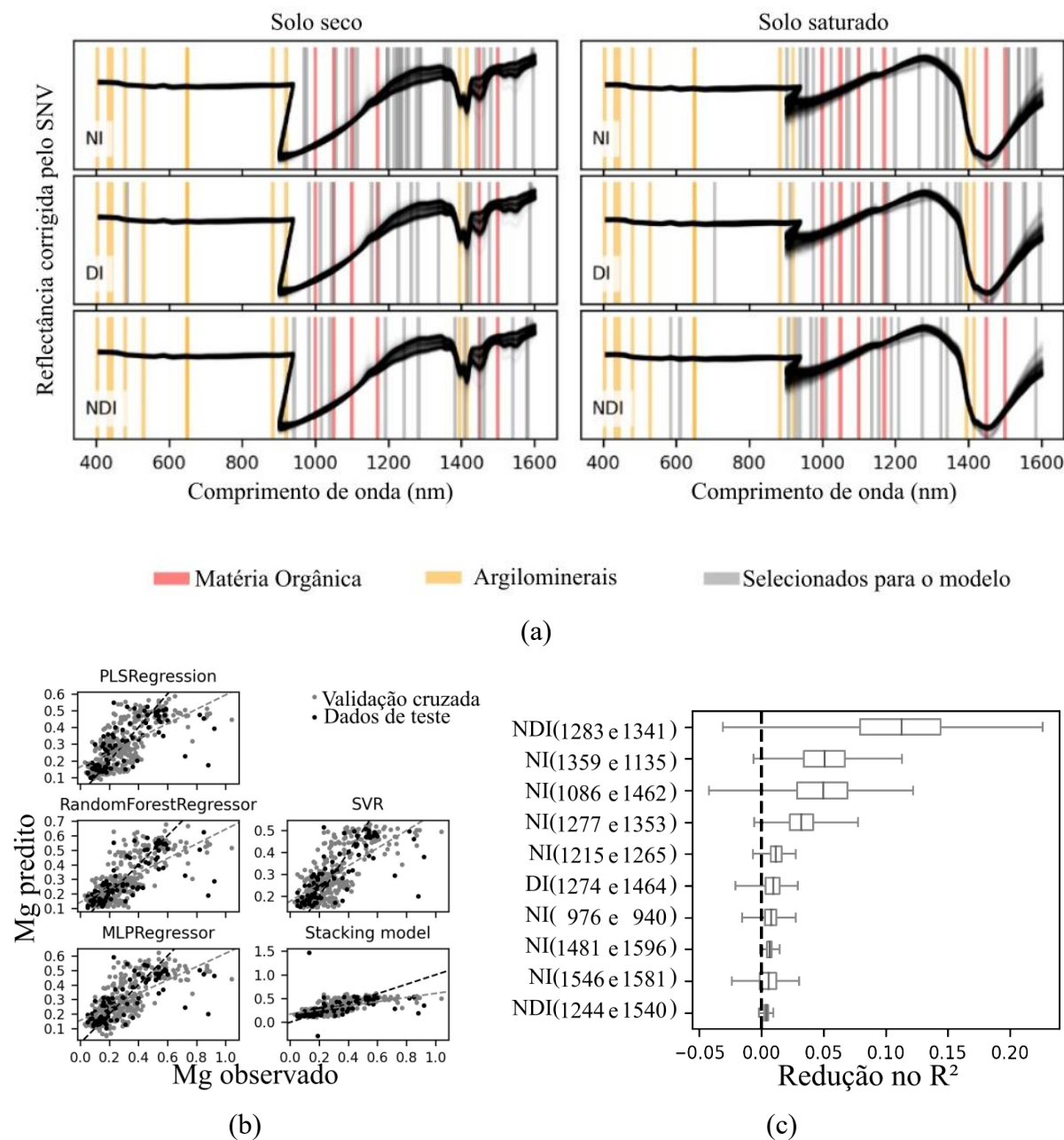


Figura 13 - Comprimentos de onda selecionados para predição do magnésio pelo modelo RF para Área 1 (a) valores preditos vs. observados para os dados de treino e teste dos modelos gerados (b) e as dez variáveis de maior importância no modelo RF (c).

Para a Área 2 a predição da t os modelos os modelos SVR e RF apresentaram valores de RPD acima de 1,40, sendo que o SVR teve melhor performance com RMSE, R^2 e RPD iguais a 0,51, 0,62 e 1,65 respectivamente. Para as amostras da Área 1 os modelos RF e SVR apresentaram resultados satisfatórios, com o RF apresentando melhores métricas, com RMSE, R^2 e RPD iguais a 0,52, 0,57 e 1,53 respectivamente.

O modelo SVR gerado para predição da t nas amostras da Área 2 utilizou 58 índices espectrais, dos quais 18 foram índices normalizados, 14 foram índices por diferença e 26 índices

por diferença normalizada. Estes índices foram compostos por um total de 106 comprimentos de onda, sendo que 46 foram dos espectros de amostras secas e 60 dos espectros de amostras úmidas. Os comprimentos de ondas selecionados que estão próximos aos valores tipicamente associados aos compostos orgânicos foram 1000, 1100, 1170, 1449 e 1500nm (Figura 14a). Comparando estes comprimentos de onda com as variáveis de maior importância (Figura 14c), podemos notar a presença do comprimento de onda de 1498nm no índice de maior importância. Este comprimento de onda está próximo ao pico de absorção associado as aminas presentes na MO do solo (Nguyen; Janik; Raupach, 1991; Rossel; Behrens, 2010). A seleção desta variável pode indicar que parte da CTC do solo é dependente dos componentes da MO presente, uma vez que para valores de pH próximos a CTC é mais influenciada pela fração orgânica no solo (Kaiser; Ellerbrock; Gerke, 2008; Parfitt; Giltrap; Whitton, 1995). Os valores preditos versus observados para todos os modelos gerados para predição da CTC na Área 2 são mostrados na Figura 14b.

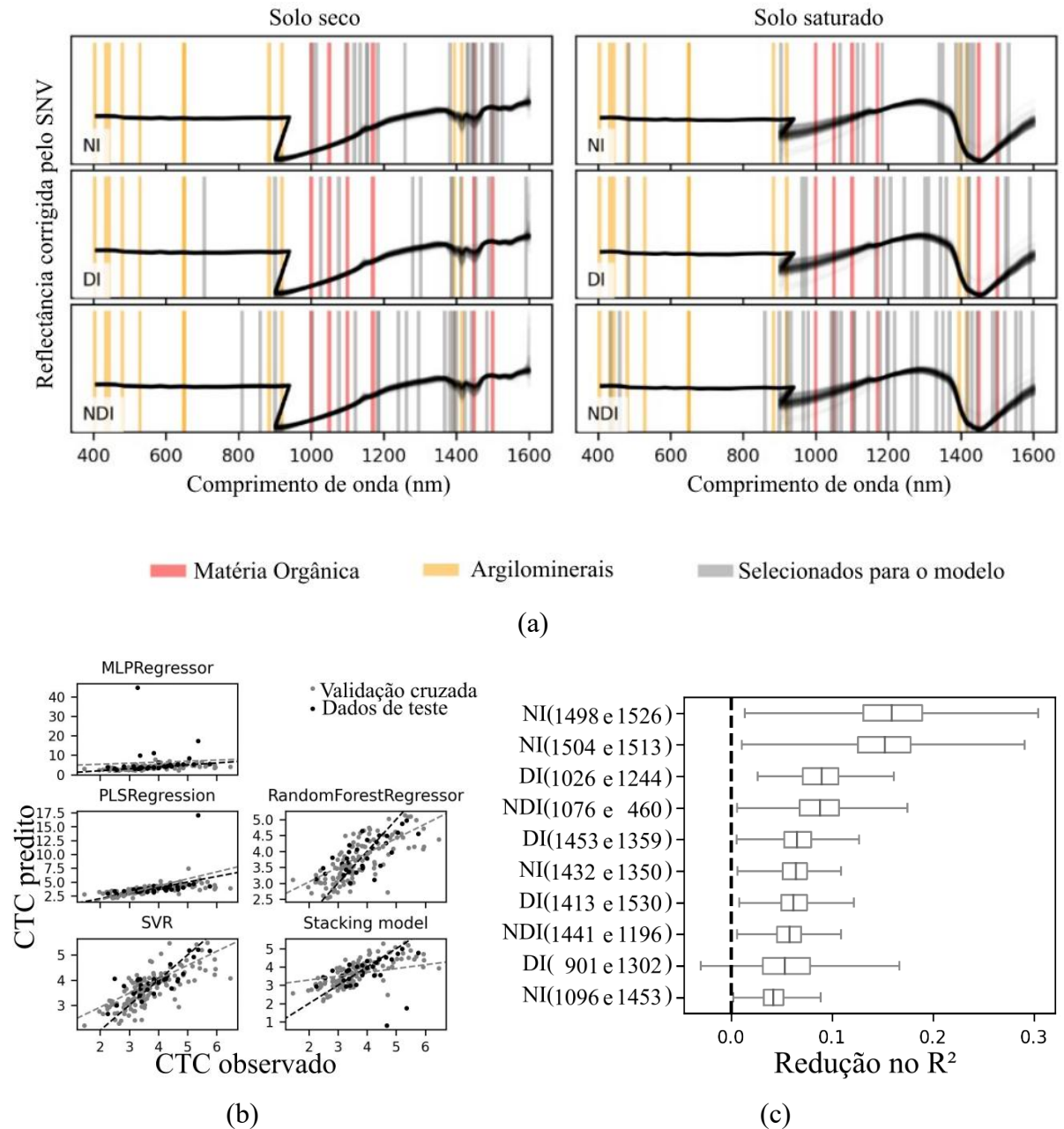


Figura 14 - Comprimentos de onda selecionados para predição da capacidade de troca de cátions pelo modelo SVR para Área 2 (a) valores preditos vs. observados para os dados de treino e teste dos modelos gerados (b) e as dez variáveis de maior importância no modelo SVR (c).

O modelo RF para predição da CTC nas amostras da Área 1 utilizou um total de 113 índices, dos quais 42 foram índices normalizados, 36 índices por diferença e 35 índices por diferença normalizada. No total foram selecionados 170 comprimentos de onda, sendo 83 dos espectros de amostras secas e 87 dos espectros de amostras saturadas (Figura 15a). Podemos notar a seleção de comprimentos de onda no espectro visível que estão relacionados com os argilominerais do solo em 404nm, 434nm, 444nm, e 529 nm que estão relacionados com os óxidos de ferro presentes nos solos.(Nguyen; Janik; Raupach, 1991; Rossel; Behrens, 2010).

No espectro NIR podemos notar a seleção de comprimentos de onda relacionados a argilominerais e componentes da matéria orgânica do solo. As dez variáveis de maior importância para o modelo são mostradas na Figura 15c, Podemos notar a seleção de quatro índices compostos por combinações de comprimentos de onda dos espectros das amostras secas com comprimentos de onda dos espectros das amostras saturadas. A variável de maior importância para o modelo foi o índice por diferença entre as bandas 1283nm do espectro de amostras secas e 1338nm do espectro de amostras úmidas, indicando que a combinação dos espectros pode conter informações importantes para o modelo.

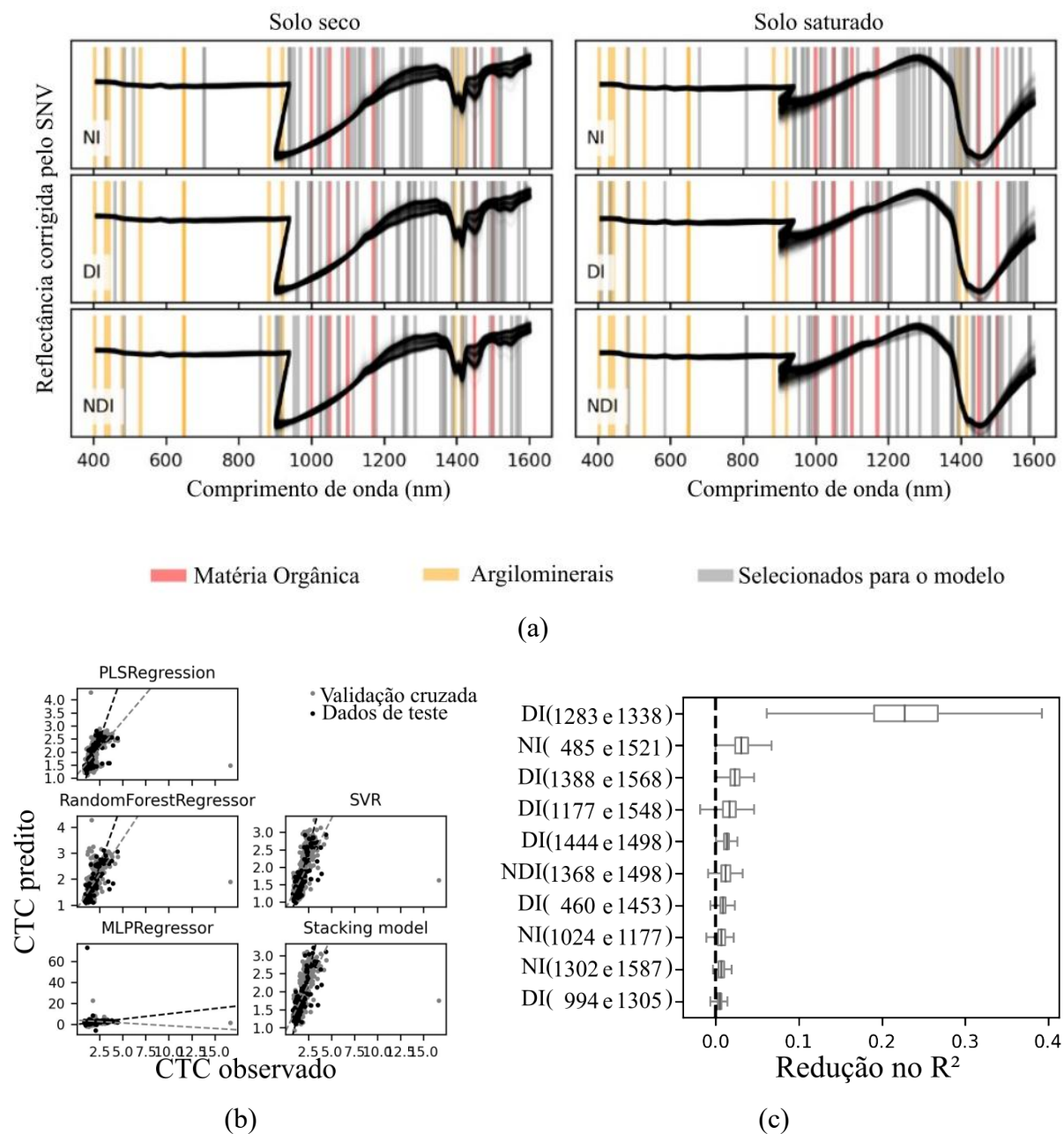


Figura 15 - Comprimentos de onda selecionados para predição da capacidade de troca de cátions pelo modelo RF para Área 1 (a) valores preditos vs. observados para os dados de treino e teste dos modelos gerados (b) e as dez variáveis de maior importância no modelo RF (c).

Somente para as amostras da Área 2 foi possível gerar um modelo com métricas aceitáveis para predição da CTC potencial do solo. O único modelo com valores de RPD maiores que 1,40 foi o RF, com RMSE, R^2 e RPD iguais a 0,57, 0,64 e 1,69 respectivamente (Tabela 2). Este modelo utilizou como entrada 103 índices espectrais, sendo 31 índices normalizados, 35 índices por diferença e 36 índices por diferença normalizada. Para compor os índices foram utilizadas 163 bandas, 81 do espectro de amostras secas e 82 do espectro de amostras saturadas (Figura 16a). Foram selecionados ao todo nove comprimentos de onda no visível, com destaque para

os comprimentos de onda de 529nm e 480nm (Figura 16a), tipicamente relacionados a absorção pela hematita e goethita respectivamente. No NIR podemos notar uma ampla seleção de comprimentos de onda (Figura 16a), com alguns comprimentos de onda sendo tipicamente relacionados aos componentes da MO do solo e a presença de água e hidroxilas associadas a argilominerais (Nguyen; Janik; Raupach, 1991; Rossel; Behrens, 2010). Neste modelo também podemos notar a presença de índices compostos por combinações dos comprimentos de onda dos espectros de amostras secas e do espectro de amostras saturadas entre as variáveis de maior importância para o modelo (Figura 16c). dentre os comprimentos de onda que compõem os índices de maior impacto no modelo, podemos destacar os comprimentos de 1401nm e 1377nm tipicamente associados a absorção pelas hidroxilas presentes nos óxidos de ferro, o comprimento de onda de 1455nm associado a absorção pelas moléculas de água nos óxidos de ferro e o comprimento de onda de 1446nm tipicamente associado a absorção por ácidos carboxílicos presentes na MO (Nguyen; Janik; Raupach, 1991; Rossel; Behrens, 2010).

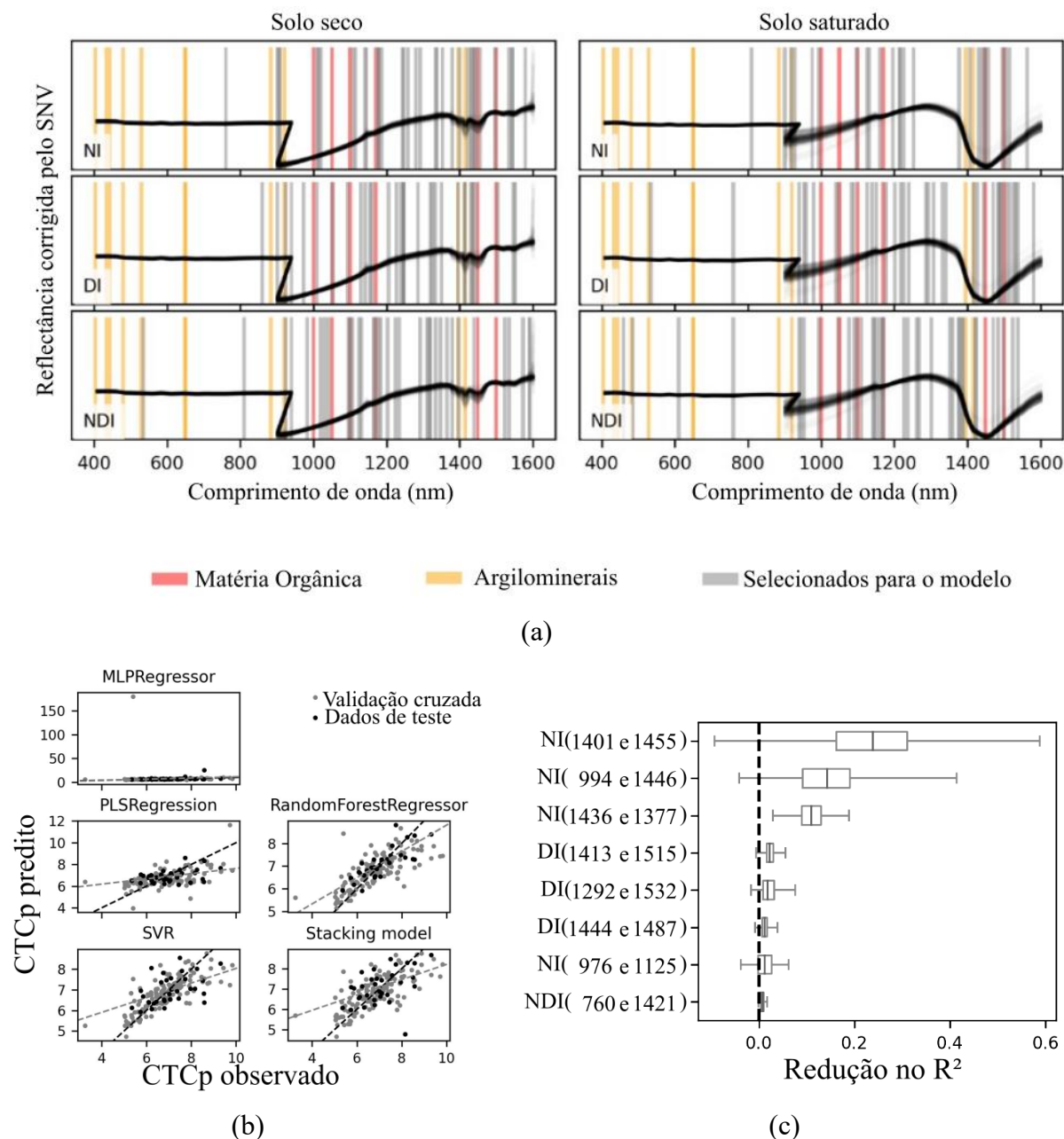


Figura 16 - Comprimentos de onda selecionados para predição da capacidade de troca de cátions potencial pelo modelo RF para Área 2 (a) valores preditos vs. observados para os dados de treino e teste dos modelos gerados (b) e as oito variáveis de maior importância no modelo RF (c).

Para a predição do P-rem todos os modelos gerados para as Área 1 e 2 apresentaram RPD maior que 1,40, com exceção do modelo MLPR para as amostras da Área 2 que apresentou elevado erro. O melhor modelo para predição do P-rem para as amostras da Área 2 foi o SVR, com valores de RMSE, R^2 e RPD iguais a 3,18, 0,66 e 1,75 respectivamente. Para a Área 1 o melhor modelo foi o RF, com métricas iguais a 1,97, 0,66 e 1,72 respectivamente.

O modelo SVR para predição do P-rem para as amostras da Área 2 utilizou um total de 76 índices, sendo 26 índices normalizados, 25 índices por diferença e 24 índices por diferença normalizada. Foram utilizados um total de 126 comprimentos de onda para formar os índices utilizados, dos quais 63 são dos espectros das amostras secas e 63 para os espectros das amostras saturadas (Figura 17a). Este modelo selecionou somente um comprimento de onda no VIS e o restante no NIR. As variáveis com maior importância no modelo são mostradas na Figura 17c, metade das variáveis foram índices compostos por comprimentos de onda dos espectros secos e saturados. A variável de maior importância foi o índice por diferença normalizada entre as bandas 1487nm do espectro das amostras secas e 980nm do espectro das amostras saturadas. Dentre os comprimentos de onda de maior importância, podemos notar a presença de comprimentos de ondas relacionados a presença de componentes da MO e dos argilominerais no solo como os comprimentos de onda de 1170nm, 1401nm e 1504nm (Nguyen; Janik; Raupach, 1991; Rossel; Behrens, 2010).

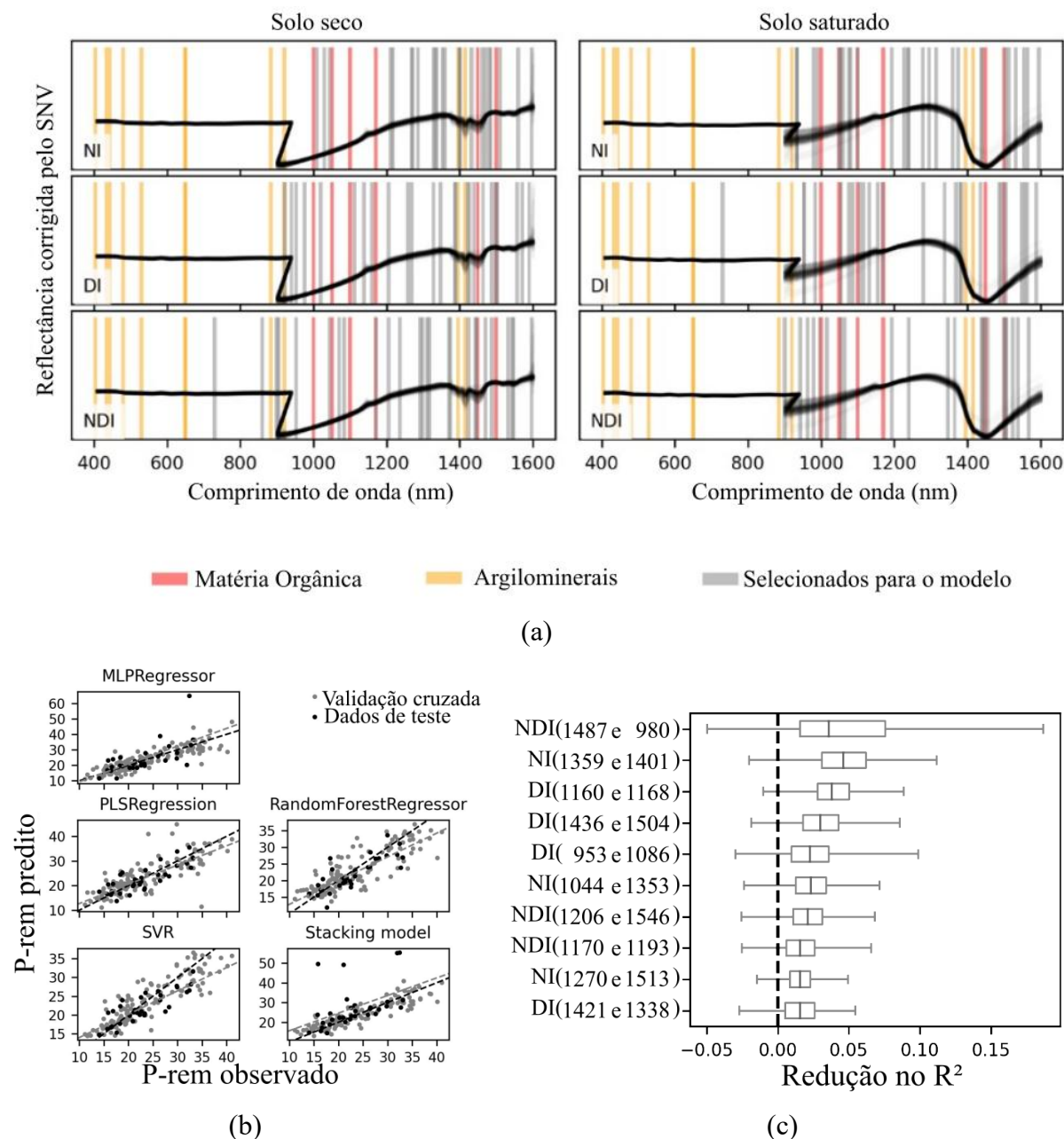


Figura 17 - Comprimentos de onda selecionados para predição do fósforo remanescente pelo modelo SVR para Área 2 (a) valores preditos vs. observados para os dados de treino e teste dos modelos gerados (b) e as dez variáveis de maior importância no modelo SVR (c).

O modelo RF para predição do P-rem na Área 1 selecionou um total de 165 índices, sendo 59 índices normalizados, 56 índices por diferença e 50 índices por diferença normalizada. Para compor estes índices foram utilizados 127 comprimentos de onda dos espectros das amostras secas e 118 comprimentos de ondas das amostras saturadas, resultando em um total de 245 comprimentos de onda utilizados (Figura 18a). Grande parte dos comprimentos de onda selecionados para compor os índices pertencem ao NIR, cobrindo quase todo o espectro, incluindo comprimentos de ondas relacionados a compostos orgânicos e presença de água e

hidroxila em argilominerais. No espectro VIS foram selecionados 12 comprimentos de onda, com destaque para os que sobrepõem valores tipicamente relacionados aos argilominerais do solo (529nm, 404nm e 434nm). Dentre os índices de maior importância para o modelo (Figura 18c) podemos destacar os comprimentos de onda relacionados aos argilominerais e óxidos de ferro 1374nm, 1392nm e 1455nm, bem como o comprimento de onda de 1056nm tipicamente associado a presença de carboidratos na MO (Nguyen; Janik; Raupach, 1991; Rossel; Behrens, 2010).

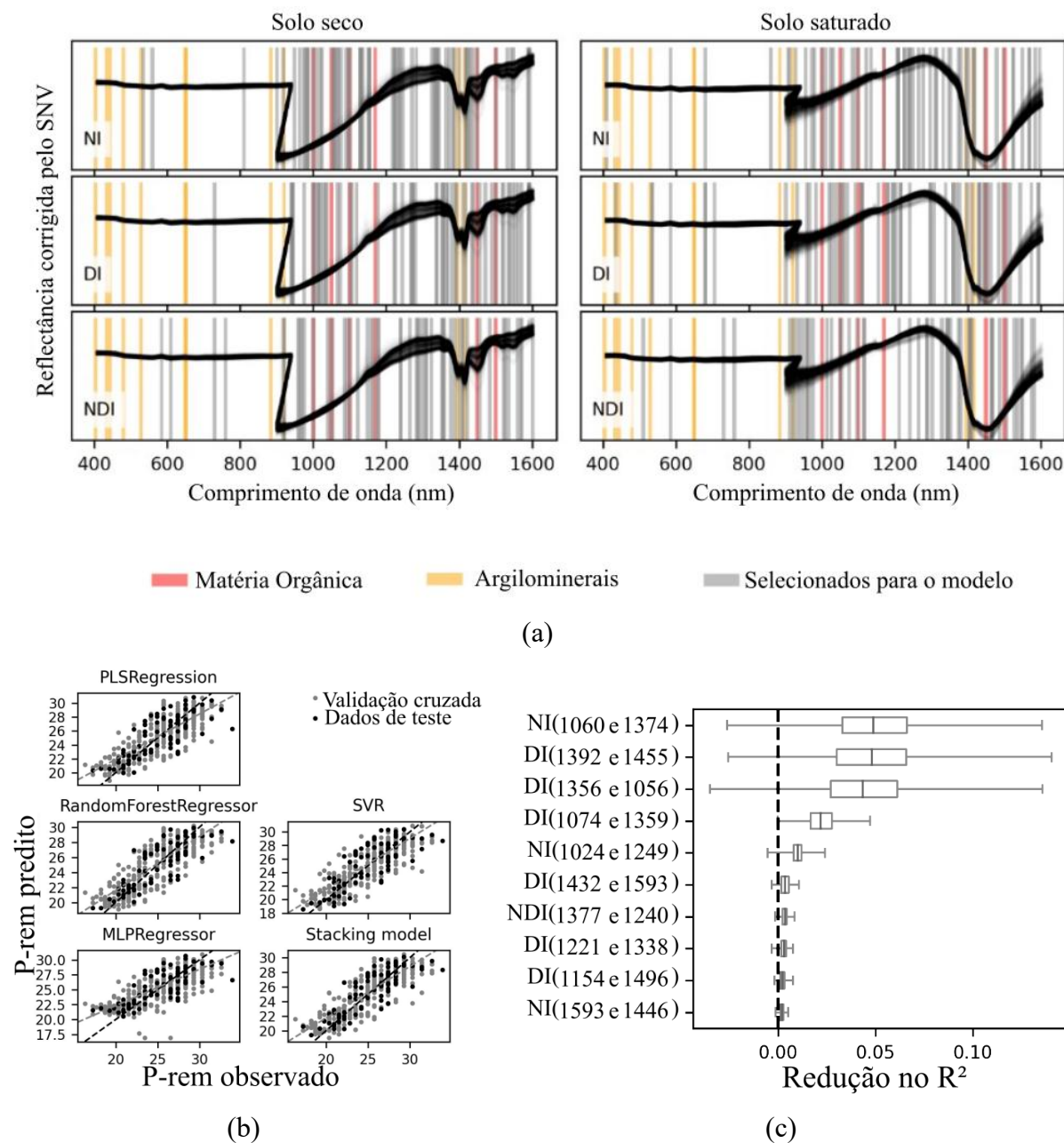


Figura 18 - Comprimentos de onda selecionados para predição do fósforo remanescente pelo modelo RF para Área 1 (a) valores preditos vs. observados para os dados de treino e teste dos modelos gerados (b) e as dez variáveis de maior importância no modelo RF (c).

Como as análises laboratoriais para MO foram realizadas somente para as amostras da Área 2, apenas os modelos para Área 2 foram gerados. O modelo com melhor performance na predição da MO foi o SVR, com valores de RMSE, R^2 e RPD iguais a 0,41, 0,68 e 1,80 respectivamente. Este modelo utilizou um total de 136 índices espectrais, sendo que 48 foram índices normalizados, 45 índices por diferença e 43 índices por diferença normalizada. Para compor estes índices foram utilizados 106 comprimentos de onda dos espectros das amostras secas e 100 comprimentos de onda os espectros de amostras saturadas, resultando em um total

de 206 comprimentos de onda utilizados (Figura 19a). Dentre as variáveis de maior importância para o modelo estão índices compostos por comprimentos de ondas tipicamente relacionados a presença de compostos da matéria orgânica do solo como os comprimentos de onda de 1141nm, 1148nm, 1501nm e 1535nm (Nguyen; Janik; Raupach, 1991; Rossel; Behrens, 2010). No visível estão presentes entre as variáveis de maior importância os comprimentos de onda no verde (560nm) e no laranja (610nm) (Figura 19c).

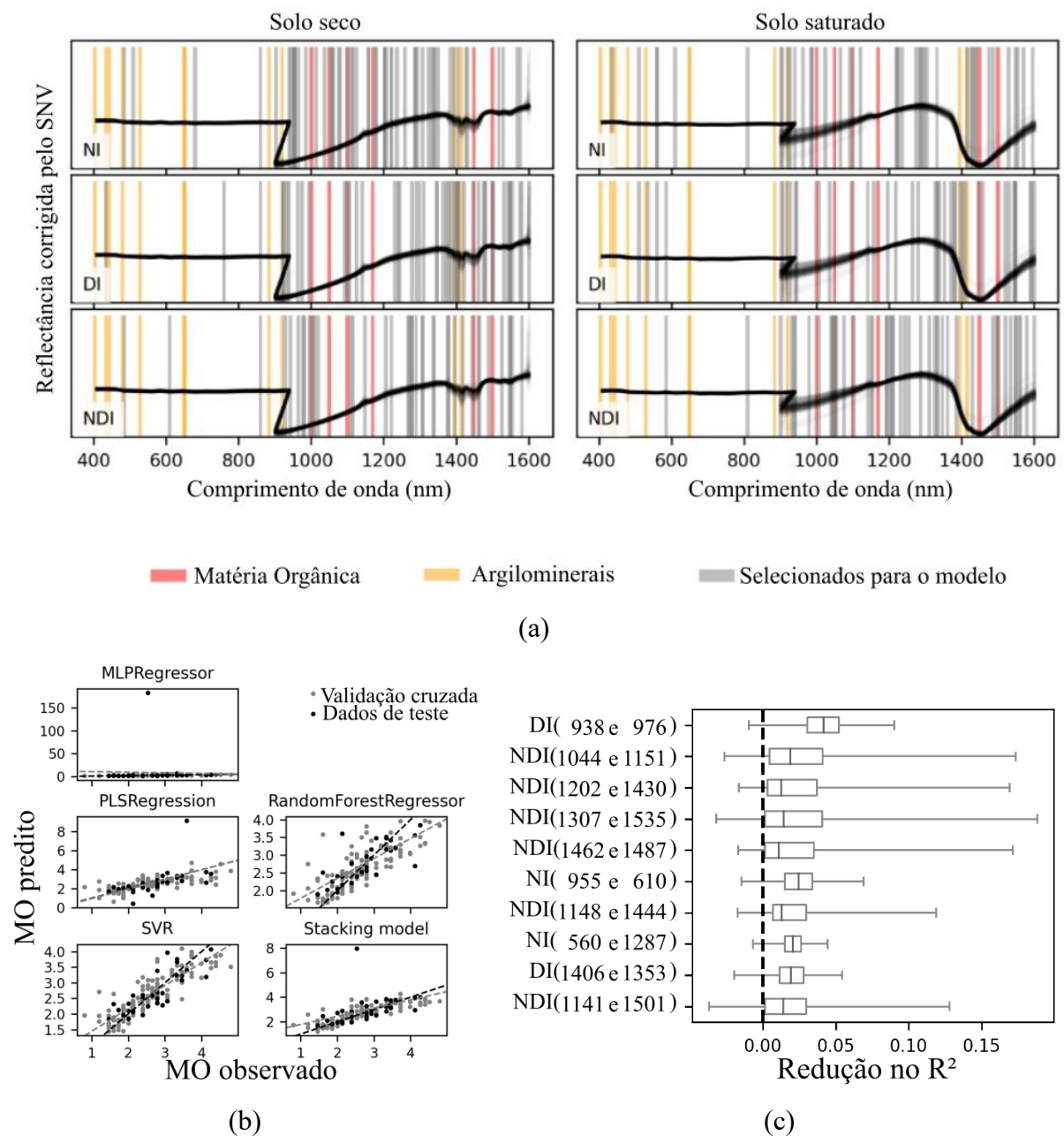


Figura 19 - Comprimentos de onda selecionados para predição da matéria orgânica pelo modelo SVR para Área 2 (a) valores preditos vs. observados para os dados de treino e teste dos modelos gerados (b) e as dez variáveis de maior importância no modelo SVR (c).

Todos os modelos com capacidade preditiva aceitáveis (RPD maior que 1,4) tiveram índices espectrais como variáveis selecionadas pelo algoritmo genético, não sendo selecionado nenhum comprimento de onda puro. Este resultado é possivelmente devido a maior quantidade de informação contida nos índices espectrais quando comparados com os comprimentos de onda separados. Todos os modelos apresentaram entre as variáveis de maiores importâncias comprimentos de onda provenientes dos espectros de amostras secas e amostras saturadas, indicando que a combinação dos espectros resultou em ganho para os modelos. É possível que a menor correlação entre a reflectância dos comprimentos de onda das amostras secas e comprimentos de onda das amostras saturadas contenha menos redundância de informações, o que ajudaria a reduzir problemas de colinearidade entre as variáveis selecionadas. Todos os modelos com capacidade preditiva aceitáveis apresentaram equilíbrio entre o número de índices selecionados de cada tipo (NI, DI e NDI), o que parece indicar que não existe um tipo específico de índice que carregue a maior parte das informações. Foi observado em todos os modelos uma quantidade similar de comprimentos de ondas utilizados provenientes dos espectros de amostras secas e espectro de amostras saturadas, o que pode indicar que a utilização do espectro de amostras saturadas pode trazer ganhos aos modelos.

Os modelos gerados apresentaram entre as variáveis selecionadas índices que contêm como ao menos uma de suas bandas comprimentos de onda relacionados com diferentes argilominerais (oxihidróxidos de ferro e alumínio) e componentes da matéria orgânica do solo. Possivelmente devido a correlação do teor dos atributos do solo com os componentes minerais e orgânicos presentes nas amostras, o que pode indicar que os modelos quantificam de forma indireta determinados atributos do solo. As diferentes variáveis selecionadas, bem como as variáveis de maior importância dos modelos de predição de um mesmo atributo das Áreas 1 e 2 foram diferentes. Isto pode ocorrer possivelmente devido a diferenças nos materiais de origem, processos de formação do solo e manejo utilizado, que impactam a informação contida em cada comprimento de onda, o que torna difícil a geração de um modelo único para as áreas.

Os modelos gerados para Área 2 foram capazes de prever de forma aceitável os atributos pH, Ca, K, CTC, CTC_p, P-rem e MO. Para Área 1 os modelos gerados para os atributos Ca, Mg, CTC e P-rem apresentaram resultados aceitáveis. Apesar da capacidade de predição dos atributos, os modelos gerados podem apresentar capacidade preditiva relacionada às características do local de obtenção das amostras, e.g. tipo de solo, material de origem, clima e relevo. Assim estudos futuros são necessários para inferir sobre a capacidade de generalização dos modelos.

Como as métricas dos modelos gerados para predição dos atributos estão próximas aos valores encontrados na literatura para predição dos atributos (onde foram utilizados equipamentos de laboratório) (Xu *et al.*, 2018b).

4 CONCLUSÃO

O emprego das informações espectrais adquiridas por meio do sistema desenvolvido aplicadas a modelos de Machine Learning possibilitou a predição dos teores de cálcio, magnésio, fósforo remanescente, soma de bases, saturação por base e capacidade de troca de cátions efetiva, bem como a estimativa da matéria orgânica em amostras de solo.

O algoritmo de seleção genética de variáveis selecionou somente índices espectrais, os índices foram formados por combinações de comprimentos de ondas dos espectros secos e saturados, o que indica melhoria na capacidade preditiva dos modelos ao utilizar tal conjunto de dados.

REFERÊNCIAS

ABDI, Dalel *et al.* **Nutrient Management & Soil & Plant Analysis Predicting Soil Phosphorus-Related Properties Using Near-Infrared Reflectance Spectroscopy** Léon-Étienne Parent *Soil Sci. Soc. Am. J.* [S. l.: s. n.], 2012. Disponível em: www.soils.org/publications/sssaj. .

ALTMANN, André *et al.* Permutation importance: A corrected feature importance measure. **Bioinformatics**, [s. l.], v. 26, n. 10, p. 1340–1347, 2010.

AMS. **AS7265x Datasheet**. [S. l.: s. n.], 2018. Disponível em: asm.com.

AURENHAMMER, Franz. Voronoi diagrams—a survey of a fundamental geometric data structure. **ACM Computing Surveys (CSUR)**, [s. l.], v. 23, n. 3, p. 345–405, 1991. Disponível em: Acesso em: 20 mar. 2023.

BARRA, Issam *et al.* **Soil spectroscopy with the use of chemometrics, machine learning and pre-processing techniques in soil diagnosis: Recent advances—A review**. [S. l.]: Elsevier B.V., 2021. Disponível em: Acesso em: 29 jun. 2022.

BEAGLEBOARD.ORG FOUNDATION. **BeagleBone Black**. [S. l.], 2023. Disponível em: <https://docs.beagleboard.org/latest/boards/beaglebone/black/>. Acesso em: 20 mar. 2023.

BERNARDI, A. C.C. *et al.* Spatial variability of soil properties and yield of a grazed alfalfa pasture in Brazil. **Precision Agriculture**, [s. l.], v. 17, n. 6, p. 737–752, 2016.

BZDOK, Danilo; ALTMAN, Naomi; KRZYWINSKI, Martin. THIS MONTH Statistics versus machine learning. **Nature Publishing Group**, [s. l.], v. 15, 2018. Disponível em: Acesso em: 26 jul. 2022.

CHANG, C. W. *et al.* Near-infrared reflectance spectroscopy - Principal components regression analyses of soil properties. **Soil Science Society of America Journal**, [s. l.], v. 65, n. 2, p. 480–490, 2001.

CHERUBIN, Maurício Roberto *et al.* Amostragem de solo na agricultura de precisão. In: AGRICULTURA DE PRECISÃO NO RIO GRANDE DO SUL. [S. l.: s. n.], 2016. p. 79–98.

CLINGENSMITH, C. M.; GRUNWALD, S.; WANI, S. P. Evaluation of calibration subsetting and new chemometric methods on the spectral prediction of key soil properties in a data-limited environment. **European Journal of Soil Science**, [s. l.], v. 70, n. 1, p. 107–126, 2019. Disponível em: Acesso em: 4 jul. 2022.

COUTINHO, Marcos A.N. *et al.* Influence of soil sample preparation on the quantification of NPK content via spectroscopy. **Geoderma**, [s. l.], v. 338, n. July 2018, p. 401–409, 2019. Disponível em: <https://doi.org/10.1016/j.geoderma.2018.12.021>.

COUTO, E. Guimarães; STEIN, A.; KLAMT, E. Large area spatial variability of soil chemical properties in central Brazil. **Agriculture, Ecosystems and Environment**, [s. l.], v. 66, n. 2, p. 139–152, 1997.

DE VARGAS, Jocelina P.R. *et al.* Application forms and types of soil acidity corrective: Changes in depth chemical attributes in long term period experiment. **Soil and Tillage Research**, [s. l.], v. 185, p. 47–60, 2019. Disponível em: <https://doi.org/10.1016/j.still.2018.08.014>. Acesso em: 30 jun. 2022.

DUNN, B. W. *et al.* The potential of near-infrared reflectance spectroscopy for soil analysis - A case study from the Riverine Plain of south-eastern Australia. **Australian Journal of Experimental Agriculture**, [s. l.], v. 42, n. 5, p. 607–614, 2002.

FÁBREGA, José R. *et al.* Modeling short-term soil-water distribution of aromatic amines. **Environmental Science and Technology**, [s. l.], v. 32, n. 18, p. 2788–2794, 1998. Disponível em: doi.org/10.1021/es9802394. Acesso em: 4 jan. 2023.

FAGERIA, N. K.; BALIGAR, V. C. Chapter 7 Ameliorating Soil Acidity of Tropical Oxisols by Liming For Sustainable Crop Production. **Advances in Agronomy**, [s. l.], v. 99, p. 345–399, 2008. Disponível em: Acesso em: 19 jul. 2022.

FORTES, R. *et al.* A methodology based on apparent electrical conductivity and guided soil samples to improve irrigation zoning. **Precision Agriculture**, [s. l.], v. 16, n. 4, p. 441–454, 2015.

GOMEZ, Cécile; LAGACHERIE, Philippe; COULOUMA, Guillaume. Regional predictions of eight common soil properties and their spatial structures from hyperspectral Vis-NIR data. **Geoderma**, [s. l.], v. 189–190, p. 176–185, 2012. Disponível em: www.elsevier.com/locate/geoderma. Acesso em: 3 jul. 2022.

GRAHAM R. HUNT; JOHN W. SALISBURY. **Visible and Near Infrared Spectra of Minerals and Rocks: I Silicate Minerals**. [S. l.: s. n.], 1970.

HAN, S *et al.* Cell size selection for site-specific crop management. **Transactions of the ASAE**, [s. l.], v. 37, n. 1, p. 19–26, 1994.

HENGL, Tomislav; ROSSITER, David G.; STEIN, Alfred. Soil sampling strategies for spatial prediction by correlation with auxiliary maps. **Australian Journal of Soil Research**, [s. l.], v. 41, n. 8, p. 1403–1422, 2003.

HUTENGs, Christopher *et al.* In situ and laboratory soil spectroscopy with portable visible-to-near-infrared and mid-infrared instruments for the assessment of organic carbon in soils. **Geoderma**, [s. l.], v. 355, n. August, p. 113900, 2019. Disponível em: <https://doi.org/10.1016/j.geoderma.2019.113900>.

ISO, B S I. 20473: 2007 Optics and photonics. **Spectral bands**, [s. l.], p. 10, 2007.

KAISER, M.; ELLERBROCK, R. H.; GERKE, H. H. Cation Exchange Capacity and Composition of Soluble Soil Organic Matter Fractions. **Soil Science Society of America Journal**, [s. l.], v. 72, n. 5, p. 1278–1285, 2008.

LAUZON, John D *et al.* Spatial variability of soil test phosphorus, potassium, and pH of Ontario soils. **Agronomy Journal**, [s. l.], v. 97, n. 2, p. 524–532, 2005. Disponível em: <https://www.researchgate.net/publication/237348952>. Acesso em: 3 dez. 2019.

LAWRENCE, Patrick G. *et al.* Guiding soil sampling strategies using classical and spatial statistics: A review. **Agronomy Journal**, [s. l.], v. 112, n. 1, p. 493–510, 2020.

LIU, Fei Tony; TING, Kai Ming; ZHOU, Zhi Hua. Isolation forest. *In:* , 2008. **Proceedings - IEEE International Conference on Data Mining, ICDM**. [S. l.: s. n.], 2008. p. 413–422.

LIU, Fei Tony; TING, Kai Ming; ZHOU, Zhi-Hua. **Isolation-based Anomaly Detection**. [S. l.: s. n.], 2012.

MALLARINO, Antonio P; WITTRY, David J. Efficacy of grid and zone soil sampling approaches for site-specific assessment of phosphorus, potassium, pH, and organic matter.

Precision Agriculture, [s. l.], v. 5, n. 2, p. 131–144, 2004. Disponível em: Acesso em: 3 dez. 2019.

MAYRINK, G. O. *et al.* Determination of chemical soil properties using diffuse reflectance and ion-exchange resins. **Precision Agriculture**, [s. l.], v. 20, n. 3, p. 541–561, 2018. Disponível em: <https://doi.org/10.1007/s11119-018-9597-y>.

METZGER, Konrad *et al.* Mid-infrared spectroscopy as an alternative to laboratory extraction for the determination of lime requirement in tillage soils. **Geoderma**, [s. l.], v. 364, n. December 2019, p. 114171, 2020. Disponível em: <https://doi.org/10.1016/j.geoderma.2020.114171>.

NAWAR, Said *et al.* Delineation of Soil Management Zones for Variable-Rate Fertilization: A Review. *In: ADVANCES IN AGRONOMY*. [S. l.]: Academic Press Inc., 2017. v. 143, p. 175–245.

NGUYEN, T.T.; JANIK, L.J.; RAUPACH, M. Diffuse Reflectance Infrared Fourier Transform (DRIFT) Spectroscopy in Soil Studies. [s. l.], n. 1982, p. 49–67, 1991.

NOVAIS, Roberto Ferreira *et al.* **Fertilidade do solo**. [S. l.: s. n.], 2007.

PADARIAN, Jose; MINASNY, Budiman; MCBRATNEY, Alex B. Machine learning and soil sciences: A review aided by machine learning tools. **SOIL**, [s. l.], v. 6, n. 1, p. 35–52, 2020. Disponível em: Acesso em: 29 jun. 2022.

PARFITT, R. L.; GILTRAP, D. J.; WHITTON, J. S. Contribution of organic matter and clay minerals to the cation exchange capacity of soils. **Communications in Soil Science and Plant Analysis**, [s. l.], v. 26, n. 9–10, p. 1343–1355, 1995.

PEREIRA, Gustavo Willam *et al.* Smart-Map: An Open-Source QGIS Plugin for Digital Mapping Using Machine Learning Techniques and Ordinary Kriging. **Agronomy**, [s. l.], v. 12, n. 6, p. 1350, 2022. Disponível em: <https://www.mdpi.com/2073-4395/12/6/1350>.

PUDEŁKO, Anna; CHODAK, Marcin. Estimation of total nitrogen and organic carbon contents in mine soils with NIR reflectance spectroscopy and various chemometric methods. **Geoderma**, [s. l.], v. 368, n. February, 2020.

QGIS DEVELOPMENT TEAM. **QGIS Geographic Information System**. [S. l.]: QGIS Association, 2023. Disponível em: <https://www.qgis.org>. Acesso em: 9 maio 2023.

RADOČAJ, Dorijan *et al.* The effect of soil sampling density and spatial autocorrelation on interpolation accuracy of chemical soil properties in arable cropland. **Agronomy**, [s. l.], v. 11, n. 12, 2021.

RODRIGUES, Khalil de Menezes *et al.* Spatial Variability in Soil Fertility and Particle Size and Their Effects on Sugarcane Yield. **Sugar Tech**, [s. l.], v. 18, n. 1, p. 39–48, 2016.

ROSSEL, R. A. Viscarra; BEHRENS, T. Using data mining to model and interpret soil diffuse reflectance spectra. **Geoderma**, [s. l.], v. 158, n. 1–2, p. 46–54, 2010. Disponível em: <http://dx.doi.org/10.1016/j.geoderma.2009.12.025>.

SILVA, Elisângela Benedet *et al.* A Regional Legacy Soil Dataset for Prediction of Sand and Clay Content with Vis-Nir-Swir, in Southern Brazil. **Revista Brasileira de Ciência do Solo**, [s. l.], v. 43, n. August, 2019. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-06832019000100304&tlng=en.

STENBERG, Bo *et al.* Visible and Near Infrared Spectroscopy in Soil Science. **Advances in Agronomy**, [s. l.], v. 107, n. C, p. 163–215, 2010a.

STENBERG, Bo *et al.* Visible and Near Infrared Spectroscopy in Soil Science. [s. l.], v. 107, p. 163–215, 2010b. Disponível em: <http://dx.doi.org/10.1016/S0065-2113>. Acesso em: 26 jun. 2022.

TERRA, Fabrício S. *et al.* Soil fertility assessment by Vis-NIR spectroscopy: Predicting soil functioning rather than availability indices. **Geoderma**, [s. l.], v. 337, n. 3, p. 368–374, 2019. Disponível em: <https://doi.org/10.1007/s11119-018-9597-y>.

TEXAS, Instruments. DLP® NIRscan™ Nano EVM User's Guid. **Texas Instruments Incorporated**, [s. l.], n. June 2015, 2015. Disponível em: <http://www.ti.com/tool/dlpnirnanoevm>.

THORLABS INC. **Detetores Fotocondutores para o Infravermelho**. [S. l.], 2020a. Disponível em: https://www.thorlabs.com/newgrouppage9.cfm?objectgroup_id=6479. Acesso em: 18 abr. 2020.

THORLABS INC. **MIR Photovoltaic Detectors, HgCdTe (MCT)**. [S. l.], 2020b. Disponível em: https://www.thorlabs.com/newgrouppage9.cfm?objectgroup_id=11319. Acesso em: 18 abr. 2020.

VALENTE, Sárvio *et al.* Definition of management zones in coffee production fields based on apparent soil electrical conductivity. **Scientia Agricola**, [s. l.], v. 69, n. 3, p. 173–179, 2012a.

VALENTE, Sárvio *et al.* The relationship between apparent soil electrical conductivity and soil properties. [s. l.], p. 683–690, 2012b. Disponível em: www.ccarevista.ufc.br.

VAN ROSSUM, Guido; DRAKE, Fred L. **Python 3 Reference Manual**. Scotts Valley, CA: CreateSpace, 2009.

VISCARRA ROSSEL, R. A. *et al.* Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. **Geoderma**, [s. l.], v. 131, n. 1–2, p. 59–75, 2006.

WIJEWARDANE, Nuwan K.; GE, Yufeng; MORGAN, Cristine L.S. Moisture insensitive prediction of soil properties from VNIR reflectance spectra based on external parameter orthogonalization. **Geoderma**, [s. l.], v. 267, n. January, p. 92–101, 2016. Disponível em: <http://dx.doi.org/10.1016/j.geoderma.2015.12.014>.

WILDING, L. P. Spatial variability: it's documentation, accommodation and implication to soil surveys. *In:* , 1985, Las Vegas. (J Bouma & D. R Nielsen, Org.) **Soil Spatial Variability**. Pudoc. Las Vegas: [s. n.], 1985. p. 166–194. Disponível em: Acesso em: 30 jun. 2022.

XU, Shengxiang *et al.* Comparison of multivariate methods for estimating selected soil properties from intact soil cores of paddy fields by Vis–NIR spectroscopy. **Geoderma**, [s. l.], v. 310, n. September 2017, p. 29–43, 2018a.

XU, Shengxiang *et al.* Comparison of multivariate methods for estimating selected soil properties from intact soil cores of paddy fields by Vis–NIR spectroscopy. **Geoderma**, [s. l.], v. 310, p. 29–43, 2018b.

ZEAITER, M.; RUTLEDGE, D. Preprocessing Methods. **Comprehensive Chemometrics**, [s. l.], v. 3, p. 121–231, 2009. Disponível em: Acesso em: 28 jun. 2022.

CAPÍTULO II

GERAÇÃO DE ZONAS DE MANEJO PARA ATRIBUTOS FÍSICO-QUÍMICOS DO SOLO UTILIZANDO ESPECTROMETRIA

RESUMO

Um dos desafios da agricultura moderna é utilizar de maneira eficiente os recursos de produção, incluindo a gestão da fertilidade do solo, que é influenciada por sua heterogeneidade espacial. Uma das formas de gerir a variabilidade espacial da fertilidade do solo é a definição de zonas de manejo (ZM) que apresentem características de fertilidade semelhantes. As ZMs podem ser definidas pela utilização de algoritmos de classificação não supervisionada, os quais não necessitam de amostras para calibração. Deste modo é possível reduzir o número de amostras de solo necessárias para caracterização da variabilidade espacial, se comparada com amostragem em grid. Os algoritmos de classificação não supervisionada podem utilizar como entrada variáveis de fácil mensuração que sejam relacionadas a disponibilidade de nutrientes nos solos. O que viabiliza utilizar um grid amostral denso com esta variável, e assim gerar ZM mais detalhadas. A espectrometria se destaca por ser um método de análise rápido com baixo custo, e pode ser utilizada para estimar diversas propriedades do solo em uma única leitura. Porém os equipamentos utilizados são caros e muitas vezes restritos a laboratórios, o que dificulta sua aplicação em campo. Este estudo avaliou o uso de um sensor portátil de baixo custo para delimitar ZM para a fertilidade do solo utilizando espectrometria. O estudo foi conduzido em uma área de 123 hectares, onde foram coletadas 112 amostras de solo. As amostras foram utilizadas para aquisição da reflectância utilizando o espectrorradiômetro de baixo custo AISpec e enviadas para laboratório para análise do pH, Ca^{2+} , Mg^{2+} , K, t, T, P, P_{rem} , MO e Argila. Os espectros foram pré-processados e utilizados como entrada na análise de componentes principais. Foram geradas ZM utilizando a componente principal um (PC1) e utilizando as componentes principais de um a cinco (PC1 a 5). Essas ZM foram comparadas com o método convencional de amostragem e com mapas gerados pela amostragem em grid, utilizando um grid com 112 amostras (Grid 1) e um grid com 30 amostras (Grid 2). Tanto as ZM geradas com a PC1 e com as PC1 a 5 apresentaram erros menores que a definição aleatória de ZMs. Os erros obtidos foram próximos aos erros da utilização do Grid 2. As zonas de manejo geradas pela metodologia proposta foram capazes de separar zonas homogêneas dos atributos

do solo com erros menores em comparação com a divisão da área de forma aleatória e próximos ao erro do Grid 2. O uso do sistema portátil pode reduzir o número de amostras necessárias para caracterizar a variabilidade espacial quando comparado com a amostragem em grid, pois permitiu delimitar as zonas de manejo cujo atributos são homogêneos, sem perda na qualidade dos resultados.

Palavras-chave: Zonas de Manejo, Espectrometria, Solo, Aprendizado não-supervisionado, Agricultura de precisão.

ABSTRACT

One of the challenges in modern agriculture is the efficient utilization of production resources, including soil fertility management, which is influenced by spatial heterogeneity. One way to address the spatial variability of soil fertility is by defining management zones (MZ) with similar fertility characteristics. MZs can be established using unsupervised classification algorithms that do not require calibration samples. This approach enables a reduction in the number of soil samples needed for spatial variability characterization compared to grid sampling. Unsupervised classification algorithms can utilize easily measurable variables related to nutrient availability in soils as input. This allows for the use of a dense sample grid with this variable, creating more detailed MZs. Spectrometry stands out as a rapid, low-cost analysis method that can estimate various soil properties in a single reading. However, the equipment used is often expensive and limited to laboratories, posing challenges for field applications. This study assessed the use of a low-cost onboard sensor to delineate MZs for soil fertility using spectrometry. The research was conducted in a 123-hectare area, where 112 soil samples were collected. The samples were used for reflectance acquisition using the low-cost AISpec spectroradiometer and sent to the laboratory for analysis of pH, Ca^{2+} , Mg^{2+} , K, t, T, P, Prem, MO, and Clay. The spectra were pre-processed and used as input for principal component analysis. MZs were generated using the first principal component (PC1) and the first five principal components (PC1 to 5). These MZs were compared with the conventional sampling method and maps generated by grid sampling, using a grid with 112 samples (Grid 1) and a grid with 30 samples (Grid 2). Both MZs generated with PC1 and PC1 to 5 showed smaller errors than randomly defined MZs. The errors obtained were close to the errors of using Grid 2. The management zones generated by the proposed methodology were able to separate homogeneous soil attribute zones with smaller errors compared to randomly dividing the area and close to the error of Grid 2. The use of the onboard system can reduce the number of samples needed to characterize spatial variability compared to grid sampling, allowing for the delineation of management zones with homogeneous attributes without compromising result quality.

Keywords: Management Zones, Spectrometry, Soil, Unsupervised Learning, Precision Agriculture.

1 INTRODUÇÃO

Atender à crescente demanda por alimentos, preservando os recursos naturais, constitui um desafio para a agricultura contemporânea. Nesse contexto, é imperativo aumentar a eficiência da produção sem expandir as áreas agrícolas. Para alcançar tal objetivo, torna-se essencial otimizar o uso de insumos, minimizando os impactos ambientais e econômicos. Essa otimização requer uma compreensão aprofundada da variabilidade espacial no ambiente de produção, a fim de estabelecer recomendações precisas para corretivos e fertilizantes em taxas variadas.

Para conhecer a variabilidade espacial do solo, é preciso coletar amostras georreferenciadas em toda a área de interesse. Assim é possível gerar mapas de fertilidade, representando os nutrientes disponíveis para as culturas. Os mapas de distribuição espacial dos nutrientes são cruciais para aplicação de técnicas de agricultura de precisão. Porém, a geração de mapas confiáveis pode ser onerosa, uma vez que a sua qualidade é dependente da quantidade de amostras coletadas e a posição da geográfica das amostras (Nawar et al., 2017).

Para se reduzir a quantidade de amostras enviadas ao laboratório, pode-se utilizar variáveis de fácil mensuração que sejam correlacionadas com atributos de fertilidade (e.g. condutividade elétrica aparente, espectros de refletância, produtividade, dentre outras) para ter conhecimento prévio sobre variabilidade da fertilidade do solo. A partir destas informações, é possível delimitar zonas de manejo utilizando-se algoritmos de *machine learning* de aprendizado não-supervisionado.

Com esse tipo de algoritmo é possível utilizar camadas de informações de fácil obtenção para identificar regiões similares (clusters). Após definidos os clusters (zonas de manejo) pode-se coletar amostras de solo para cada zona de manejo (ZM). Assim, é possível reduzir o número de amostras necessárias para caracterizar a variabilidade espacial da fertilidade do solo, permitindo realizar a correção da fertilidade para cada ZM de forma específica (Hengl; Rossiter; Stein, 2003; Nawar et al., 2017; Valente et al., 2012a, 2012b).

Diversos estudos apresentam resultados satisfatórios ao utilizar a espectrometria para geração e modelos de predição dos atributos do solo, o que indica a sensibilidade do espectro aos compostos presentes no solo (HUTENGs et al., 2019; METZGER et al., 2020; PUDEŁKO; CHODAK, 2020; TERRA et al., 2019). Por ser uma metodologia rápida, não utilizar reagentes, apresentar baixo custo por amostra e se relacionar com compostos do solo, a espectrometria é uma ferramenta promissora para definição de ZM para a fertilidade do solo (Nawar et al., 2017). Apesar de ser um método de análise rápido, a espectrometria utiliza espectrorradiômetros, que

são equipamentos caros e muitas vezes são usados exclusivamente em bancadas de laboratórios. Além disso, tais equipamentos são incapazes de treinar ou aplicar os modelos de *machine learning*.

Os recentes avanços tecnológicos permitiram a miniaturização e redução no custo de produção de diversos tipos de sensores e processadores. Assim seria possível a utilização de um equipamento de baixo custo para coleta de informações necessárias para definir ZM para a fertilidade do solo por meio de aprendizagem de máquina não-supervisionado.

A utilização de modelos de aprendizado não supervisionado utilizando espectrometria, se destaca por não precisar de treinamento prévio. Portanto, não há a necessidade de coleta de amostras de solo para análises laboratoriais para delimitação das zonas de manejo. Além disso, diversos fatores dificultam a geração de modelos supervisionados generalistas para os atributos do solo usando espectrometria, havendo a necessidade de calibração nos modelos para cada localidade específica

Isso possibilitaria a redução do número de amostras de solo necessárias para caracterização da variabilidade espacial de uma área, uma vez que modelos não-supervisionados não necessitam de amostras para seu treinamento. Tal equipamento poderá contribuir para a utilização de técnicas de AP ajudando a gerar ZM para a aplicação de fertilizantes a taxa variada, reduzindo o custo com amostras laboratoriais. Portanto o objetivo deste trabalho foi avaliar o uso de um sistema sensor de baixo custo, para obter as reflectâncias de amostras de solo e delimitar zonas de manejo para a fertilidade do solo utilizando algoritmos de classificação não-supervisionados.

2 MATERIAL E MÉTODOS

O estudo foi conduzido em uma área de uso agrícola com 123 ha localizada em Tabaporã/MT, onde foram coletadas 112 amostras de solo (Figura 20). As amostras foram coletadas de 0 a 20 cm de profundidade, cada amostra foi composta de 10 a 15 subamostras coletadas em um raio de 10m. As amostras coletadas foram peneiradas em peneira de malha com 2 mm e colocadas para secar ao ar, a fim de eliminar torrões e reduzir a umidade para uso no equipamento. A seguir uma fração de terra fina seca ao ar (TFSA) de cada amostra foi enviada para laboratório para as análises padrões de pH, cálcio (Ca^{2+}), magnésio (Mg^{2+}), capacidade de troca de cátions (t), capacidade de troca de cátions potencial (T) potássio (K), fósforo disponível (P), fósforo remanescente (P_{rem}), matéria orgânica (MO) e argila. Com o

objetivo de avaliar a variabilidade dos atributos nas amostras, foi conduzida uma análise utilizando a estatística descritiva dos atributos amostrados.

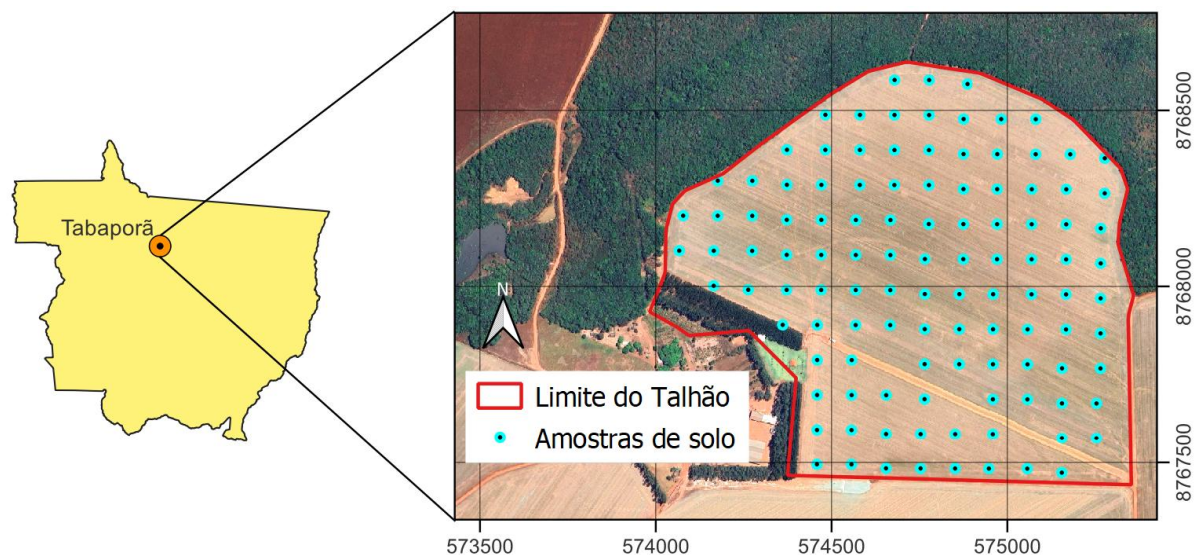


Figura 20 – Área de estudo localizada em Tabaporã/MT e localização das amostras de solo coletadas (Grid 1).

Parte das amostras peneiradas e secas foram utilizadas para obtenção dos espectros de reflectância utilizando, um espectrorradiômetro de baixo custo desenvolvido para este trabalho. O equipamento desenvolvido faz a integração de dois sensores comerciais por meio de um computador de placa única *BeagleBone Black* (BeagleBoard.org Foundation, 2023), os sensores utilizados foram o *Triad Spectroscopy Sensor* (AMS, 2018) produzido pela *SparkFun*, sensível a 18 bandas localizadas entre 410 nm e 940 nm, e o *DLP® NIRScan™ Nano* (Texas, 2015) produzido pela *Texas Instruments* sensível aos comprimentos de onda de 900 nm a 1700 nm.

Para a aquisição dos espectros de reflectância foram utilizadas cubetas de poliestireno de 4,5 cm³, preenchidas com 2 cm³ de TFSA. As cubetas foram inseridas no equipamento desenvolvido, onde foram realizadas quatro repetições agitando as amostras entre cada leitura (Figura 21). O espectro de reflectância de cada amostra foi composto pela média das quatro repetições. Após aquisição dos espectros das amostras secas, as amostras foram saturadas com água destilada, e em seguida foram realizadas duas leituras em cada amostra para compor o espectro da amostra saturada.

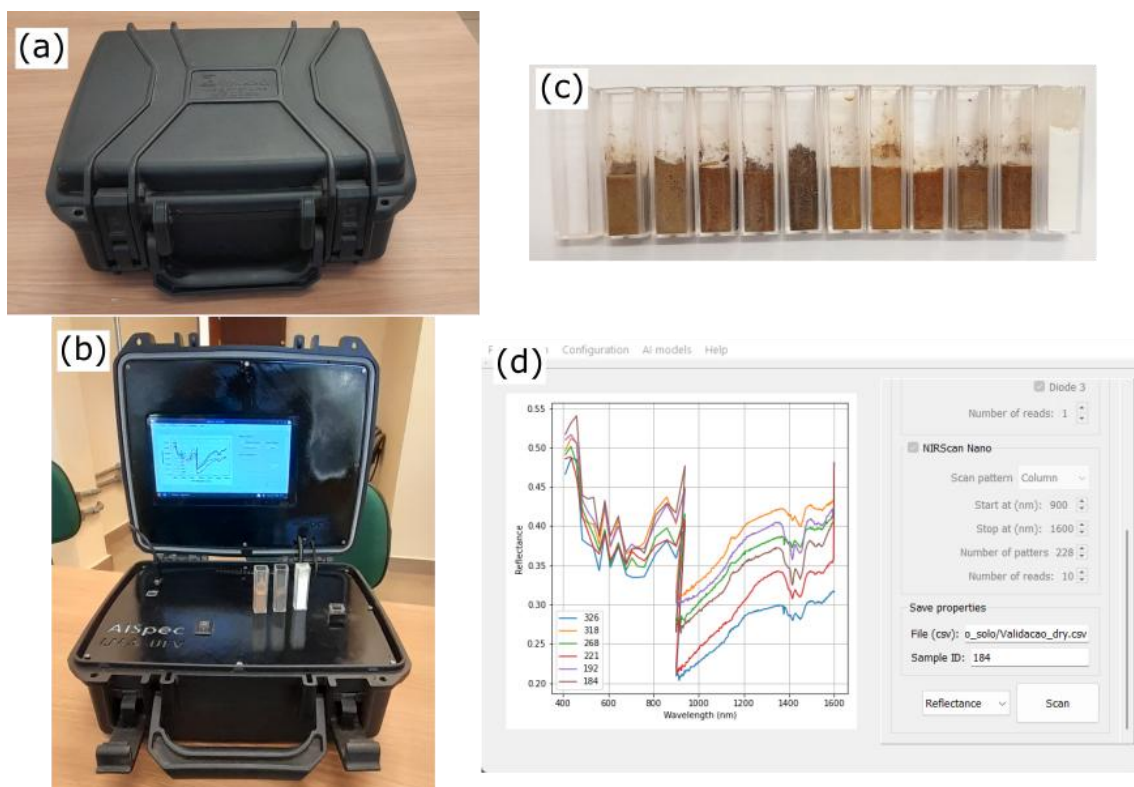


Figura 21 – Espectrorradiômetro de baixo custo AISpec fechado para transporte (a), AISpec aberto e em funcionamento (b), cubetas com amostras de solo e branco padrão (c) e software desenvolvido para coleta e visualização de dados (d).

A fim de corrigir os efeitos de espalhamento da radiação devidos aos diferentes tamanhos de partículas, foi utilizado o método *Standard Normal Variate* (SNV) (Zeaiter; Rutledge, 2009), esta correção foi realizada apenas para os comprimentos de onda obtidos com o sensor *NIRScan Nano*. Após correção pelo SNV os espectros das amostras secas e das amostras saturadas foram unidos em um mesmo banco de dados, resultando em um total de 492 comprimentos de onda. Este banco de dados foi utilizado para os modelos de classificação não-supervisionada.

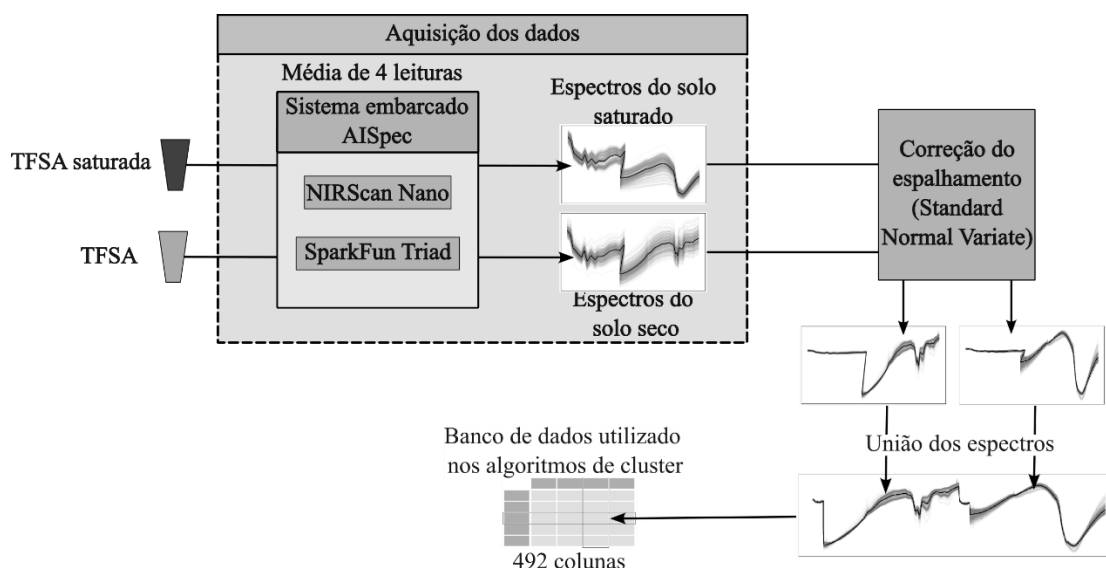


Figura 22 – Fluxograma mostrando desde a leitura das amostras até a criação do banco de dados utilizado nos algoritmos de cluster.

Para geração das zonas de manejo foi utilizado o algoritmo de clusterização *Fuzzy K-Means*, disponível no plug-in *Smart-Map* (Pereira *et al.*, 2022) do software QGIS (QGIS Development Team, 2023). Como visto no Capítulo 1, os índices espectrais foram mais importantes para os modelos. Portanto o algoritmo desenvolvido em *python 3*, utilizou como entrada um banco de dados composto pelas combinações das cinco primeiras componentes principais (PC) obtidas da análise de componentes principais (PCA) dos índices espectrais. Os índices espectrais do tipo índices por diferença, índices normalizados e índices por diferença normalizada basearam-se em comprimentos de onda conhecidos por estarem relacionados aos componentes dos argilominerais e compostos orgânicos presentes no solo (Nguyen; Janik; Raupach, 1991; Rossel; Behrens, 2010) (Figura 23).

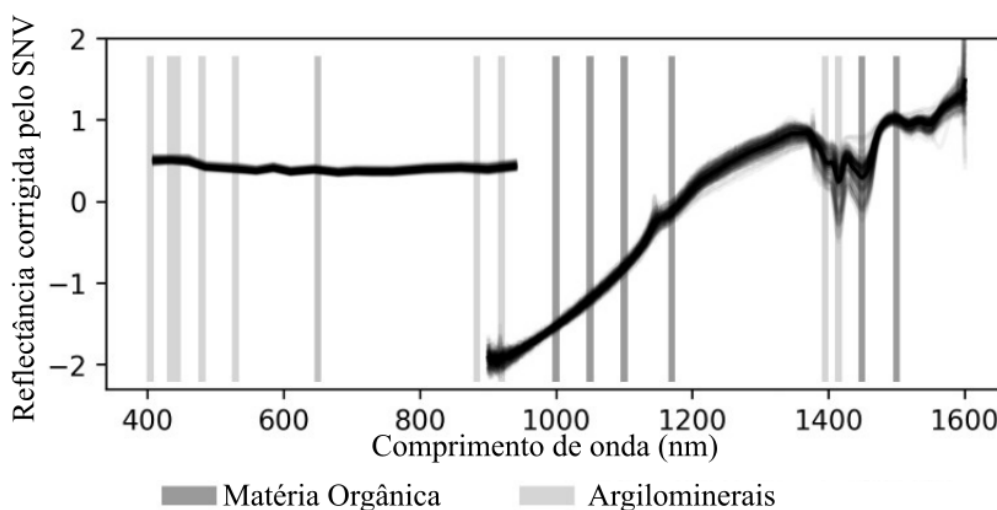


Figura 23 - Comprimentos de ondas utilizados para compor os índices espectrais e espectros de reflectância das 112 amostras.

Os índices utilizados foram do tipo índice por diferença, índices normalizados e índices por diferença normalizada, de acordo com as equações 1, 2 e 3 respectivamente. Foram gerados todos os índices possíveis utilizando a combinação dois a dois dos comprimentos de ondas selecionados. Resultando em um total de 3978 índices.

$$ID = B_2 - B_1 \quad \text{Equação 6}$$

$$IN = \frac{B_2}{B_1} \quad \text{Equação 7}$$

$$IDN = \frac{B_2 - B_1}{B_2 + B_1} \quad \text{Equação 8}$$

em que:

ID , IN e IDN - índices por diferença, normalizados e por diferença normalizada;

B_1 e B_2 - bandas utilizadas nos índices.

O fluxograma da metodologia proposta para gerar as classes pode ser visto na Figura 24. Essa metodologia utiliza o banco de dados de reflectâncias corrigidas pelo SNV e filtradas para incluir apenas os comprimentos de onda selecionados e mostrados na Figura 23.

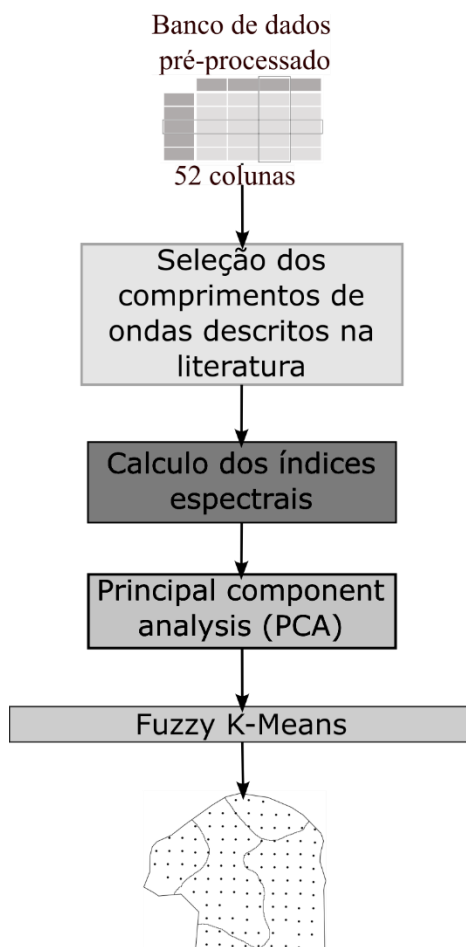


Figura 24 – Fluxograma para geração das classes.

Após a definição das classes foi realizada uma filtragem para excluir classes com áreas demasiadamente pequenas, as quais teriam menos de 5 amostras de solo presentes, para evitar um número reduzidos de amostras para o cálculo do erro. Para isso, as classes com menos de 5 ha foram desfeitas, sendo incorporada automaticamente à classe vizinha que apresentasse maior limite em comum com a área a ser excluída. As zonas de manejo foram estabelecidas considerando cada área individual como uma Zona de Manejo (ZM), independentemente da classe à qual foi atribuída. Dessa forma, as classes compostas por mais de um polígono foram subdivididas, de modo que cada polígono representasse uma ZM distinta. Essa abordagem foi adotada devido à possibilidade de que as semelhanças nas assinaturas espectrais das amostras pudessem ser resultado de fatores diversos em áreas distintas.

Com o objetivo de avaliar se a utilização do sensor em conjunto com a metodologia proposta foi capaz de gerar zonas de manejo (ZM) mais homogêneas do que a divisão da área em ZM aleatórias, o erro estimado das ZM foi comparado com a distribuição dos erros para dois mil cenários com ZM geradas aleatoriamente. Para gerar as ZM de forma aleatórias, foram

alocados pontos com coordenadas aleatórias no interior da área e, em seguida, foram gerados os polígonos por meio do diagrama de Voronoi (Aurenhammer, 1991) utilizando os pontos aleatórios e os limites da área. Dessa forma, foi possível gerar a distribuição do erro para cada variável físico-química do solo em função do número de ZM geradas de forma aleatória. Essa análise permitiu verificar se a metodologia proposta foi eficaz em gerar ZM mais homogêneas em comparação com a abordagem aleatória.

Tanto as ZM geradas com o uso do sensor quanto para os cenários com ZM geradas de forma aleatória, erro médio estimado para cada ZM foi calculado segundo equação Eq. 9. O erro foi calculado para cada variável físico-química do solo. Para cada ZM gerada foi realizada uma validação do tipo *leave-one-out* retirando uma amostra e a comparando com a média das amostras restantes no interior da ZM, utilizando todas as amostras do Grid 1. Este processo foi realizado utilizando a linguagem de programação *python 3*. O processo foi repetido para cada amostra da ZM dentro de cada ZM gerada, sendo considerado como o erro médio a média dos erros calculados para cada ZM.

$$Erro_{ZM} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n} \quad \text{Eq. 9}$$

em que:

$Erro_{ZM}$ – erro estimado da variável do solo na ZM gerada;

n – número de pontos no interior da zona de manejo;

x_i – valor da variável do solo no ponto i ;

$\bar{x}$ – valor médio para a variável do solo considerando todos os pontos com exceção do ponto i .

Com o objetivo de comparar a metodologia proposta com o cenário onde as ZM foram geradas aleatoriamente, foi analisada a porcentagem de erros do cenário aleatório com valores inferiores ao erro gerado pelas ZM utilizando o sensor. Assim, conseguimos calcular a porcentagem de vezes em que a metodologia proposta apresentou resultados superiores à abordagem de definição aleatória das Zonas de Manejo.

A fim de comparar as ZM geradas com as metodologias mais utilizadas na amostragem de solos, foram calculados os erros associados a amostragem em grid com uma amostra por hectare (Grid 1), em grid com uma amostra a cada quatro hectare (Grid2) obtido pela remoção de amostras do Grid 1 e o método convencional (Conv), que foi considerado como a média das amostras do Grid 1.

A determinação do erro decorrente da utilização do Grid 1 na produção de mapas contínuos dos atributos físico-químicos do solo foi realizada por meio da validação cruzada dos mapas gerados pela krigagem ordinária (OK) para cada variável. O procedimento de krigagem ordinária foi aplicado aos atributos que exibiram índice de Moran significativo ao nível de um por cento, indicando a presença de autocorrelação espacial.

Para o Grid 2 o erro foi calculado por meio da validação cruzada para as amostras da Figura 25. As amostras que foram removidas do Grid 1 para formação do Grid 2 também foram utilizadas para o cálculo do erro, por meio do cálculo do erro absoluto dos valores dos atributos físico-químicos das amostras com os valores de pixels gerados na mesma posição geográfica pela OK utilizando o Grid 2.

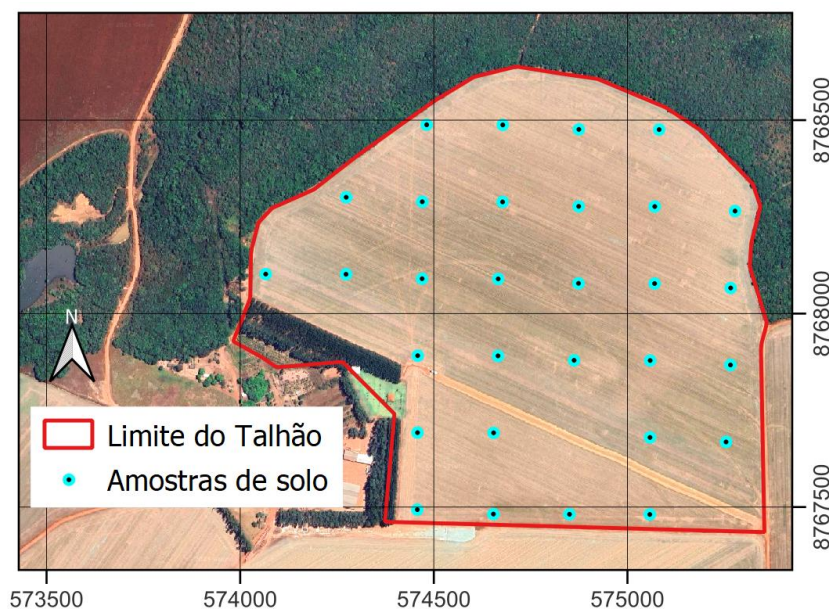


Figura 25 – Amostras utilizadas para o Grid 2 (30 amostras).

O erro da utilização do método Conv foi calculado segundo a Eq. 4 considerando uma única ZM, toda a área de estudo. Para todos os cálculos dos erros dos métodos abordados (ZM, Grid 1, Grid 2 e Conv) foram utilizadas todas as amostras do Grid 1. Todos os cálculos dos erros, bem como a PCA foram realizados utilizando a linguagem de programação Python 3 (VAN ROSSUM; DRAKE, 2009).

3 RESULTADOS E DISCUSSÃO

A estatística descritiva das 112 amostras de solo utilizadas neste estudo está mostrada na Tabela 3. De acordo com (WILDING, 1985) citado por (Xu *et al.*, 2018a), atributos que apresentam coeficiente de variação maior que 35% apresentam elevada variabilidade, atributos

com CV entre 15 e 35% apresentam variabilidade moderada e atributos com CV menores que 15% apresentam baixa variabilidade. Neste trabalho os atributos P, K e P_{rem} apresentaram elevada variabilidade, os atributos Ca^{2+} , Mg^{2+} , t e T, MO e argila apresentaram variabilidade moderada e o pH apresentou baixa variabilidade. Os resultados encontrados neste estudo demonstram que a área avaliada apresenta variabilidade espacial, o que pode aumentar a eficiência de modelos baseados em espectrometria e suas respectivas performances, conforme mencionado por Abdi et al. (2012).

Tabela 3: Estatística descritiva dos resultados das análises laboratoriais para as amostras de solo utilizadas.

	Média	Mediana	Desvio padrão	CV (%)	Mínimo	Máximo
pH (H ₂ O)	6,11	6,07	0,28	4,55	5,76	7,78
P	17,91	15,15	10,40	58,04	3,40	56,60
K	47,85	31,50	40,89	85,46	7,00	199,00
Ca ²⁺	2,55	2,42	0,70	27,57	0,99	5,82
Mg ²⁺	0,84	0,82	0,22	25,96	0,43	1,68
t	3,52	3,39	0,88	24,95	1,45	6,46
T	6,80	6,75	1,18	17,30	3,25	9,82
MO	2,45	2,39	0,77	31,33	0,80	4,79
P _{rem}	26,13	24,65	13,72	52,52	11,10	147,40
Argila	35,65	34,55	11,91	33,42	16,30	62,00

CV – coeficiente de variação (%); P – fósforo pelo método da resina (mg/dm³); K – potássio (mg/dm³); Ca – cálcio (cmol_c/dm³); Mg – magnésio (cmol_c/dm³); P-rem – fósforo remanescente (mg/L); t – CTC efetiva (cmol_c/dm³); T – CTC potencial (cmol_c/dm³); MO – matéria orgânica (dag/kg).

Observando a variância explicada pelas componentes principais (PCs) (Figura 26), podemos perceber que a PC1 explicou aproximadamente 40% da variação total do banco de dados dos índices espectrais. Há uma queda acentuada na porcentagem de variação explicada entre a primeira e a segunda PC, e uma redução menos acentuada na porcentagem de variação explicada pelas demais PCs a partir da PC2. A variância explicada acumulada até a PC5 foi de aproximadamente 70% (Figura 26).

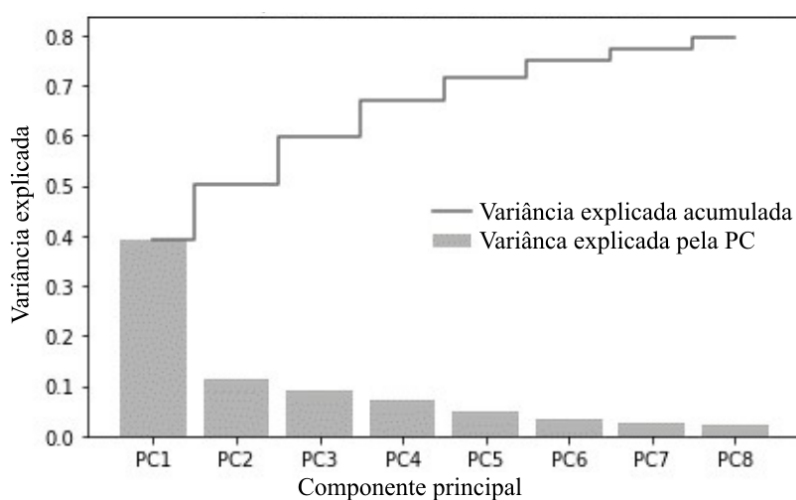


Figura 26 – Variância explicada pelas componentes principais obtidas na análise de componentes principais (PCA).

Todas as PCs estudadas apresentaram p-valor para o índice de Moran global significativo ao nível de 0,01, indicando forte autocorrelação espacial. Os mapas interpolados por OK são apresentados na Figura 27. Podemos notar que tanto a PC1 quanto a PC2 mostram uma área ao norte com valores discrepantes com os do entorno.

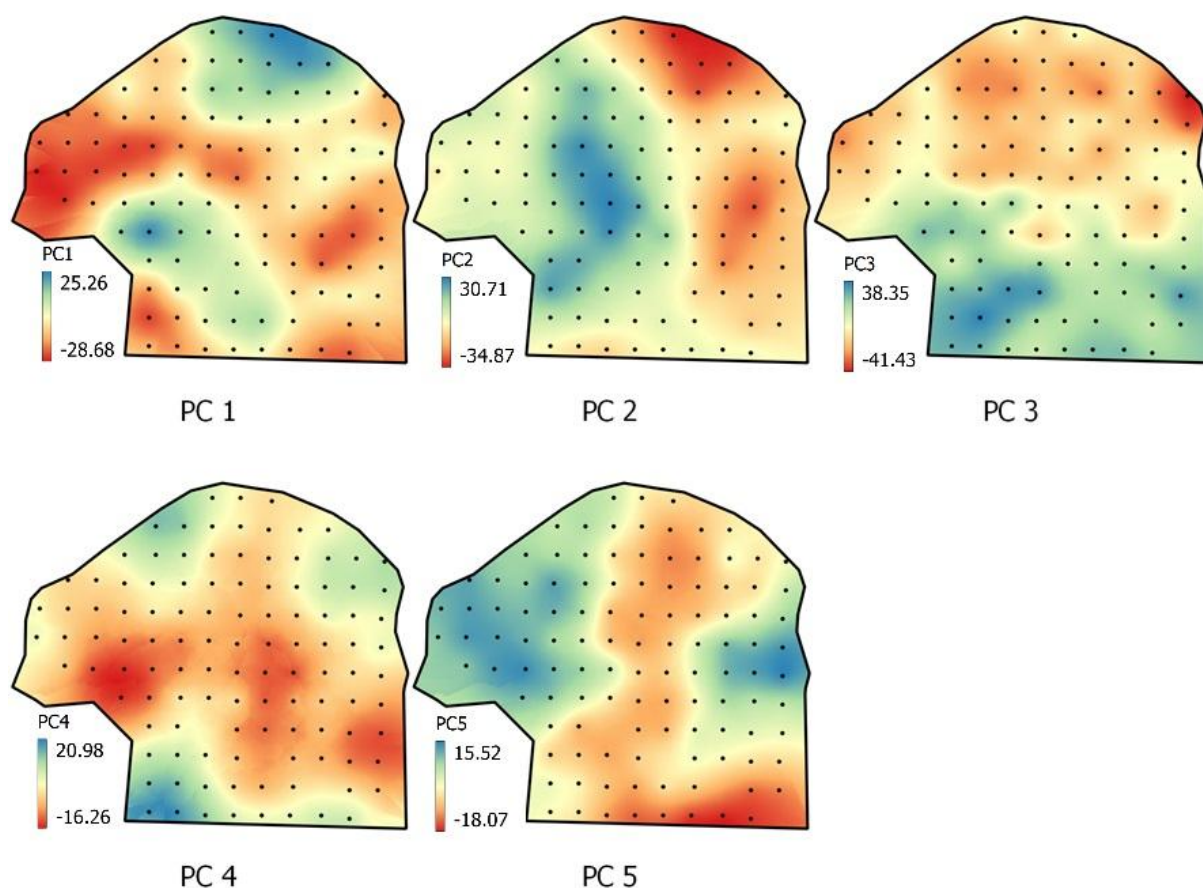


Figura 27 – Mapas interpolados por krigagem ordinária para as componentes principais um a cinco.

Os valores do índice de Moran para cada variável físico-química dos Grid 1 e Grid 2 e seus respectivos p-valores são mostrados na Tabela 4. Podemos observar que o Grid 1 apresentou elevada autocorrelação espacial para todas os atributos estudados, indicados pelos valores do índice de Moran próximos a um e p-valores menores que 1%. Comparado ao Grid 1, o Grid 2 resultou em valores do índice de Moran menores, com os atributos pH, Mg e P com p-valores maiores que 1%. Apesar da significância do índice de Moran apresentar p-valores significantes no Grid 2, ainda assim os modelos podem ser prejudicados, uma vez que a acurácia da interpolação está primariamente relacionada a número de amostras, com a autocorrelação espacial apresentando menor impacto nos resultados (Radočaj et al., 2021).

Tabela 4: Índices de Moran e p-valor para os atributos físico-químicos do solo em relação ao Grid 1 e ao Grid 2

	Grid 1		Grid 2	
	Índice de Moran	p-value	Índice de Moran	p-value
pH	0,848	0,001	0,627	0,073
Ca	0,881	0,001	0,735	0,003
Mg	0,676	0,001	0,643	0,044
K	0,906	0,001	0,815	0,001
P	0,85	0,001	0,642	0,022
P _{rem}	0,951	0,001	0,728	0,002
t	0,735	0,001	0,725	0,005
T	0,888	0,001	0,738	0,001
MO	0,902	0,001	0,739	0,002
Argila	0,958	0,001	0,776	0,001

CV – coeficiente de variação (%); P – fósforo pelo método da resina (mg/dm³); K – potássio (mg/dm³); Ca – cálcio (cmol/dm³); Mg – magnésio (cmol/dm³); P-rem – fósforo remanescente (mg/L); t – CTC efetiva (cmol/dm³); T – CTC potencial (cmol/dm³); MO – matéria orgânica (dag/kg).

Os resultados das interpolações por OK no Grid 1 e Grid 2 são apresentados nas Figura 28 e Figura 29, respectivamente. Observa-se que a OK utilizando o Grid 2 conseguiu identificar as áreas com valores altos e baixos dos atributos de maneira semelhante ao Grid 1, embora com menor resolução.

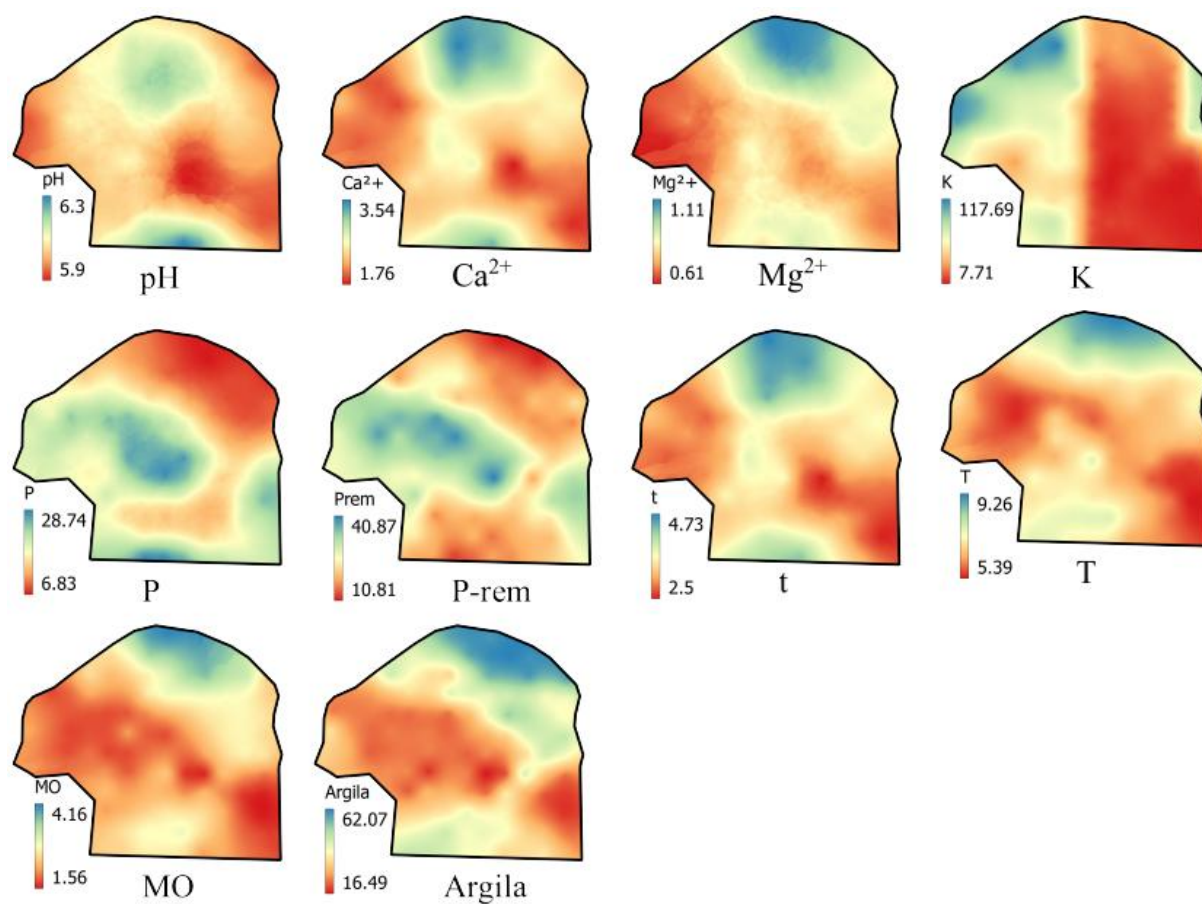


Figura 28 - Mapas dos atributos físico-químicos interpolados por krigagem ordinária utilizando o Grid 1. As unidades utilizadas para cada variável foram: mg/dm³ para o fósforo e potássio, cmol/dm³ para o cálcio, magnésio, CTC efetiva e potencial, mg/L para o fósforo remanescente e dag/kg para a matéria orgânica.

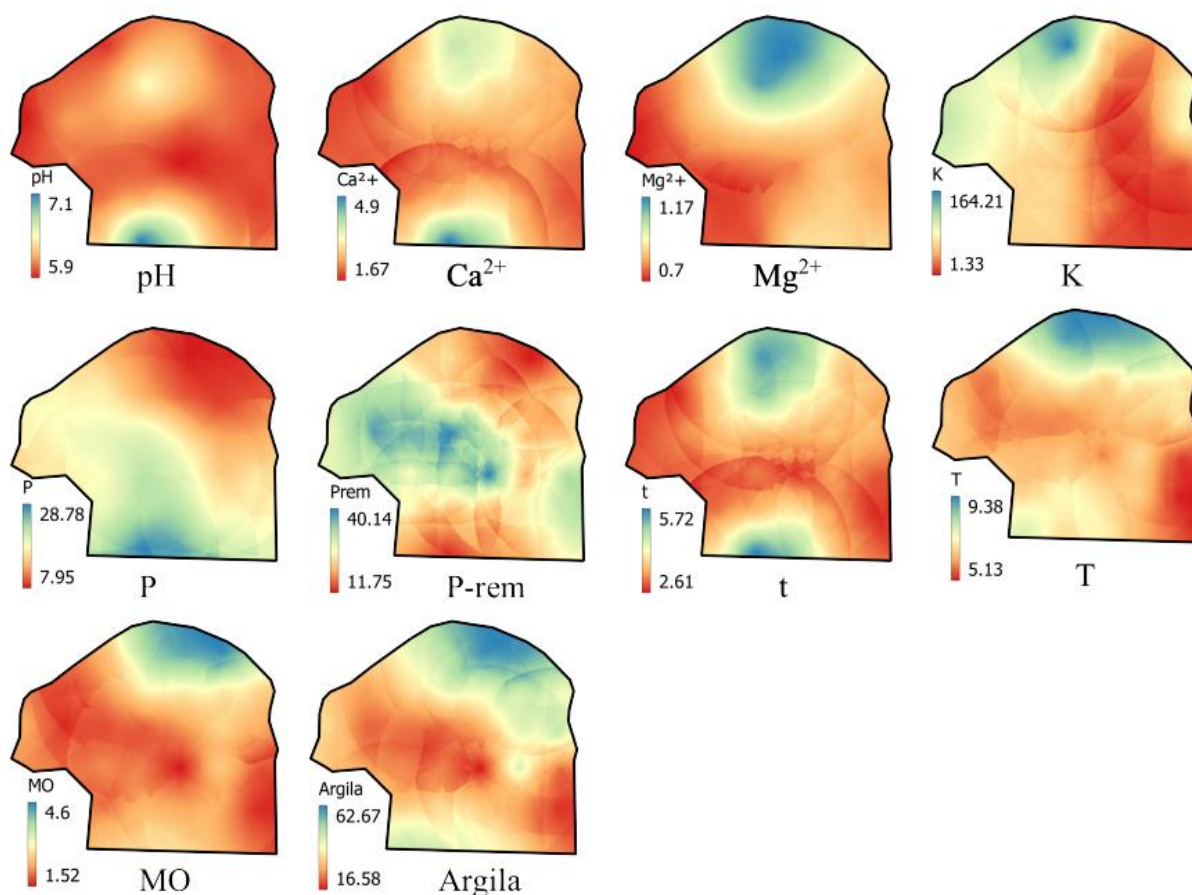


Figura 29 - Mapas dos atributos físico-químicos interpolados por krigagem ordinária utilizando o Grid 2. As unidades utilizadas para cada variável foram: mg/dm³ para o fósforo e potássio, cmol/dm³ para o cálcio, magnésio, CTC efetiva e potencial, mg/L para o fósforo remanescente e dag/kg para a matéria orgânica.

As ZM geradas por meio do algoritmo *Fuzzy K-Means* utilizando a PC1 é mostrada na Figura 30. Os números de classes utilizados como entrada do algoritmo foram três, quatro, cinco, seis e sete classes, o que resultou em mapas com cinco, seis, sete e nove ZM respectivamente. Analisando os valores do *fuzzy performance index* (FPI) e a *normalized classification entropy* (NCE), o número ótimo de classes foi igual a três.

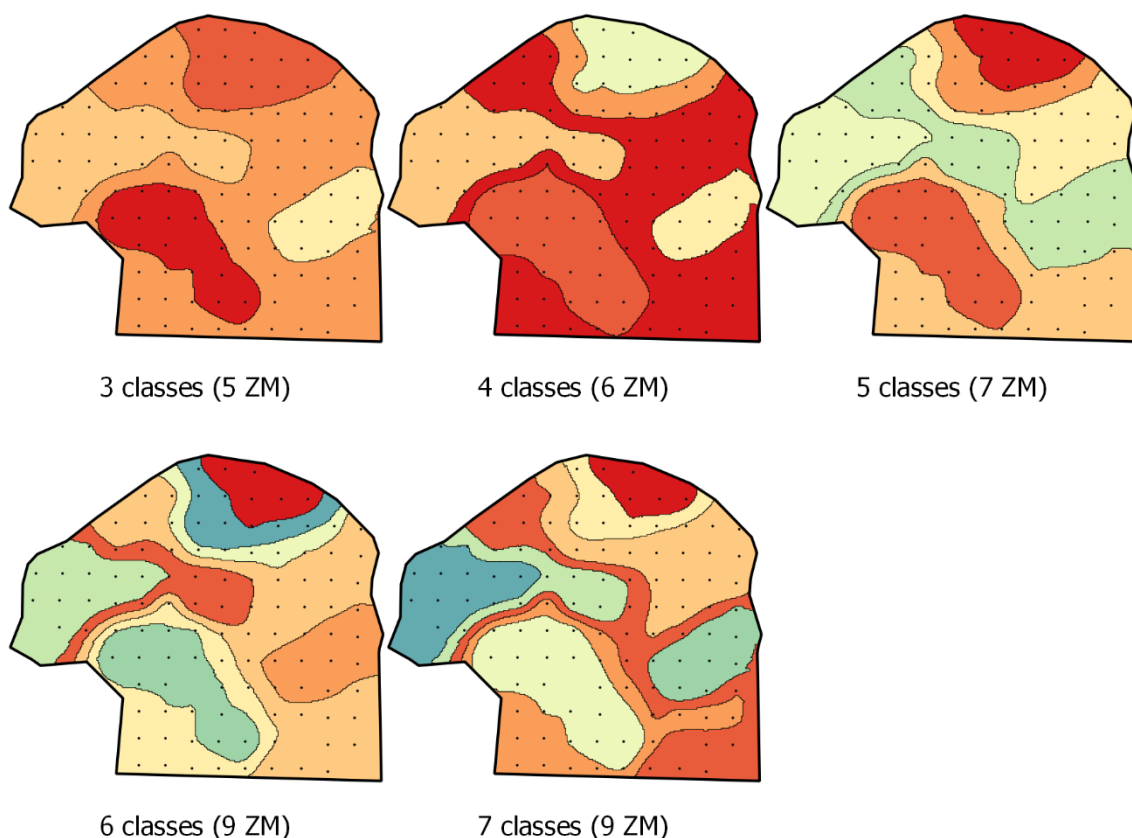


Figura 30 – Mapas com as zonas de manejo geradas utilizando a componente principal um como entrada do algoritmo fuzzy K-Means.

Na Figura 31 são apresentados os mapas com as Zonas de Manejo (ZM) geradas pelo algoritmo Fuzzy K-Means, utilizando as PC1 a 5 como entrada. Foram testados diferentes números de classes, sendo três, quatro, cinco, seis e sete classes, resultando em mapas com cinco, seis, sete, nove e nove ZM, respectivamente. A análise dos valores FPI e da NCE indicou que o número ótimo de classes foi igual a nove.

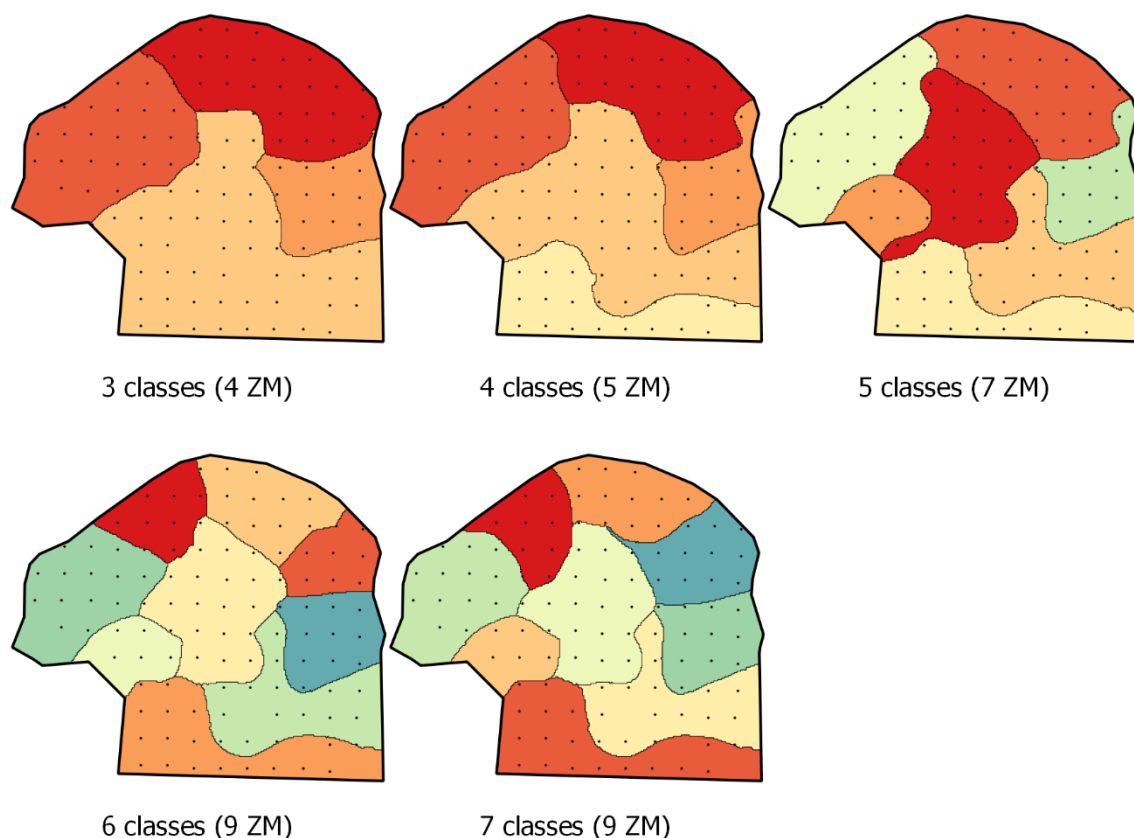


Figura 31 – Mapas com as zonas de manejo geradas utilizando as componentes principais 1 a 5 como entrada do algoritmo fuzzy K-Means.

Comparando os erros médios absolutos das ZMs geradas pela metodologia proposta com o cenário onde as ZM foram geradas aleatoriamente, podemos notar que para a maioria dos atributos físico-químicos os erros gerados pela metodologia proposta foram menores que os gerados pelo cenário com ZMs aleatórias (Tabela 5). Indicando que a definição das ZMs utilizando as PCs dos índices espectrais foi sensível a variação espacial dos atributos.

O erro na estimativa dos níveis de potássio (K) utilizando as Zonas de Manejo (ZMs) geradas pela primeira componente principal (PC1) foi comparado com o erro obtido no cenário onde as ZM foram geradas de forma aleatória. Os resultados mostram que o erro associado às ZMs da PC1 foi maior do que aquele gerado pelo cenário com ZM aleatórias. Aproximadamente 99% das ZMs geradas aleatoriamente apresentaram menor erro na quantificação dos níveis de K em comparação com as ZMs da PC1. Este erro elevado pode estar relacionado à variação da distribuição espacial do K (Figura 28), que ocorre no sentido Leste-Oeste, enquanto a divisão das ZMs pela PC1 favoreceu ZMs em sentido preferencialmente perpendicular à variação do K (Figura 30).

Tabela 5: Porcentagem das vezes em que a simples divisão aleatória das zonas de manejo repetida 2000 vezes, resultou em erros menores que a utilização da metodologia proposta na divisão das zonas de manejo.

	Classes (ZM)	pH	P	K	Ca	Mg	t	T	MO	P _{rem}	Arg.
			----- % -----								
PC 1	3(5)	0,2	76,5	95,4	11,3	29,3	25,1	46,8	41,8	4,3	39,4
	4(6)	4,5	40,7	99,4	17,1	59,2	50,2	10,1	30,4	17,8	49,1
	5(7)	46,7	20,5	99,7	22,4	10,8	32,1	1,4	0,5	25,6	11,7
	7(9)	10,1	49,5	99,4	12,1	25,2	21,7	5,1	15,6	16,3	57,5
	6(9)	13,7	47,6	99,9	18,4	49,7	36,7	5,7	20,1	20,0	44,2
PC 1, 2, 3, 4 e 5	3(4)	2,9	14,1	56,7	5,9	8,6	8,8	15,5	14,8	6,7	27,3
	4(5)	63,7	12,4	81,6	64,6	30,5	65,7	30,3	30,0	56,4	23,1
	5(7)	59,8	3,4	58,2	37,8	52,8	48,2	21,7	27,8	43,2	10,2
	6(9)	41,4	0,2	67,7	39,0	19,9	39,3	14,8	27,5	9,0	16,8
	7(9)	42,3	1,1	78,1	15,8	12,7	17,1	3,3	10,3	19,5	18,2

P – fósforo pelo método da resina; K – potássio; Ca – cálcio; Mg – magnésio; P_{rem} – fosforo remanescente; t – capacidade de troca de cátions efetiva; T – capacidade de troca de cátions potencial; MO – matéria orgânica; Arg – argila.

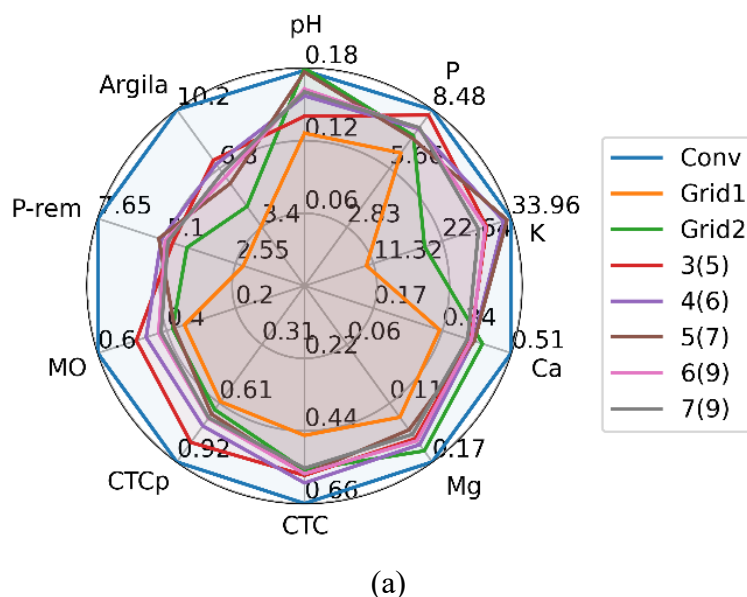
Para todos os atributos com exceção do K, o uso da PC1 para delimitação das ZM apresentou erros absolutos menores que o cenário com ZM aleatórias. O aumento do número de classes/ZM proporcionou melhoras para algumas variáveis e piora em outras, não indicando nenhum padrão para relação entre o aumento do número de classes e melhoria dos resultados (Tabela 5).

As ZM geradas com as PCs de 1 a 5 de forma geral apresentaram melhores resultados que a utilização somente da PC1, com destaque para os atributos P, K e Argila. Para as ZM geradas utilizando as PCs 1 a 5, o aumento do número de classes/ZM a partir de 5 ZM, apresentou uma tendência na redução do erro para as variáveis pH, P, Ca, Mg, t, T, MO, P_{rem} e Argila (Tabela 5).

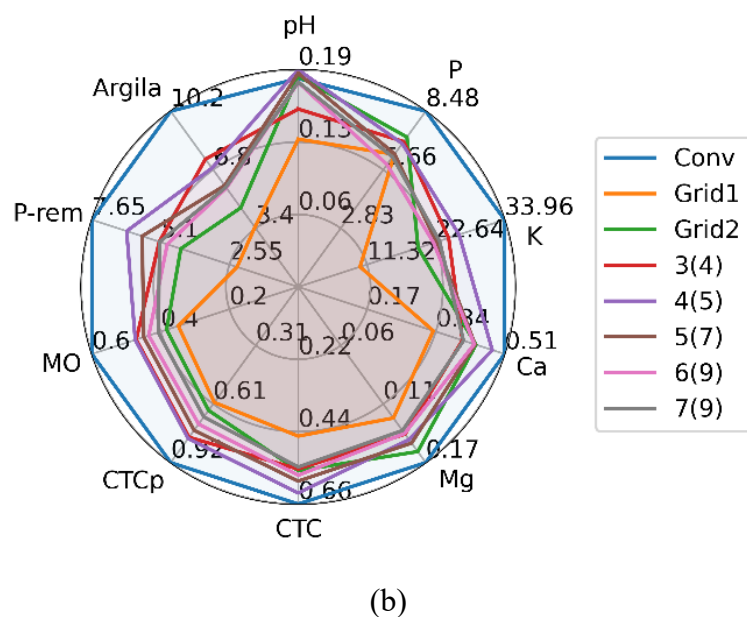
Na Figura 32 é apresentada a comparação dos erros absolutos entre as metodologias de amostragem do solo abordadas. Podemos notar que todas as metodologias apresentaram erros absolutos menores que a amostragem convencional. A utilização do Grid 1 foi a metodologia que apresentou menor erro médio absoluto para todos os atributos físico-químicos do solo comparada com as ZM geradas pela PC1 (Figura 32a).

A utilização das PCs dos índices para delimitação das ZM apresentou erros absolutos médios maiores que o Grid 1. Comparadas com o Grid 2 as ZM geradas utilizando a PC1 apresentaram menores erros absolutos médios para as variáveis pH, t, Mg e Ca. Para os atributos P, MO e T as ZM apresentaram valores de erro absoluto médio próximos ao Grid 2. As maiores

diferenças dos erros entre o Grid 2 e as ZM geradas a partir da PC1 foram para as variáveis K, e de forma menos acentuada para o P_{rem} e Argila (Figura 32a).



(a)



(b)

Figura 32 - Comparação entre os erros calculados para cada método de amostragem utilizado. Comparação entre os métodos Grid 1, Grid 2, Convencional e zonas de manejo geradas com a PC1 (a). Comparação entre os métodos Grid 1, Grid 2, Convencional e zonas de manejo geradas com as PC1, 2, 3, 4 e 5 (b)

As ZM geradas utilizando as PCs de 1 a 5 apresentaram erros absolutos médios próximo aos valores obtidos pelo Grid 1 para os atributos P, MO e T. Comparadas com o Grid 2, as ZM apresentaram menores erros para Mg, Ca, P e Argila. Em comparação as ZM geradas pela

utilização da PC1, as ZM geradas pelas PCs de 1 a 5 apresentaram erros médios absolutos semelhantes (Figura 32a e Figura 32b). Porém podemos observar um menor erro para os atributos P e K, possivelmente devido a utilização das PCs 2, 4 e 5 (Figura 27) que apresentam distribuição espacial mais relacionada com o mapa de distribuição do K e do P (Figura 28).

Observando os erros médio absolutos para todos os atributos, podemos notar que os maiores erros foram do método convencional de amostragem, o que era esperado, uma vez que a área de estudo apresenta elevado coeficiente de variação para as amostras analisadas, além dos atributos apresentarem dependência espacial (Tabela 3 e Tabela 4).

Tanto as ZM geradas utilizando a PC1 quanto as PCs de 1 a 5 mostraram ser capazes de identificar zonas onde os tributos físico-químicos são mais homogêneos, pois apresentaram erros menores que a definição de ZM aleatórias (Tabela 5).

A metodologia testada para a definição das ZM resultou na divisão da área de estudo entre quatro e nove ZM, o que implicaria na necessidade de menos amostras enviadas ao laboratório se comparada com go Grid 1 (com 112 amostras) e o Grid 2 (com 30 amostras), sem que haja um aumento considerável nos erros. De fato, para alguns atributos do solo, a utilização das ZM resultou em menores erros que o Grid 2 (composto por um maior número de amostras).

Assim, a metodologia proposta foi capaz de definir de ZM com finalidade de gerir a fertilidade do solo, utilizando um menor número de amostras. Estudos futuros podem também avaliar o potencial de utilização do sistema portátil para metodologias que visem guiar a amostragem do solo afim de gerar mapas contínuos da distribuição de seus atributos utilizando índices espectrais.

4 CONCLUSÃO

As zonas de manejo geradas com informações espectrais do sensor desenvolvido foram capazes de separar zonas homogêneas dos atributos do solo, uma vez que apresentam erros menores se comparado com divisão da área de forma aleatória. Os erros observados foram semelhantes aos erros da utilização do grid de menor densidade, porém utilizando um menor número de amostras.

O aumento do número de zonas de manejo geradas não levou a ganhos consistentes, uma vez que resultou em aumento dos erros absolutos médios para alguns atributos e redução do erro para outros.

A utilização do sensor desenvolvido apresenta potencial para delimitação de zonas de manejo para gestão da fertilidade do solo, reduzindo o número de amostras necessárias se comparado com grids amostrais de baixa densidade.

REFERÊNCIAS

ABDI, Dalel *et al.* **Nutrient Management & Soil & Plant Analysis Predicting Soil Phosphorus-Related Properties Using Near-Infrared Reflectance Spectroscopy** Léon-Étienne Parent *Soil Sci. Soc. Am. J.* [S. l.: s. n.], 2012. Disponível em: www.soils.org/publications/sssaj. .

ALTMANN, André *et al.* Permutation importance: A corrected feature importance measure. **Bioinformatics**, [s. l.], v. 26, n. 10, p. 1340–1347, 2010.

AMS. **AS7265x Datasheet**. [S. l.: s. n.], 2018. Disponível em: asm.com.

AURENHAMMER, Franz. Voronoi diagrams—a survey of a fundamental geometric data structure. **ACM Computing Surveys (CSUR)**, [s. l.], v. 23, n. 3, p. 345–405, 1991. Disponível em: Acesso em: 20 mar. 2023.

BARRA, Issam *et al.* **Soil spectroscopy with the use of chemometrics, machine learning and pre-processing techniques in soil diagnosis: Recent advances—A review**. [S. l.]: Elsevier B.V., 2021. Disponível em: Acesso em: 29 jun. 2022.

BEAGLEBOARD.ORG FOUNDATION. **BeagleBone Black**. [S. l.], 2023. Disponível em: <https://docs.beagleboard.org/latest/boards/beaglebone/black/>. Acesso em: 20 mar. 2023.

BERNARDI, A. C.C. *et al.* Spatial variability of soil properties and yield of a grazed alfalfa pasture in Brazil. **Precision Agriculture**, [s. l.], v. 17, n. 6, p. 737–752, 2016.

BZDOK, Danilo; ALTMAN, Naomi; KRZYWINSKI, Martin. THIS MONTH Statistics versus machine learning. **Nature Publishing Group**, [s. l.], v. 15, 2018. Disponível em: Acesso em: 26 jul. 2022.

CHANG, C. W. *et al.* Near-infrared reflectance spectroscopy - Principal components regression analyses of soil properties. **Soil Science Society of America Journal**, [s. l.], v. 65, n. 2, p. 480–490, 2001.

CHERUBIN, Maurício Roberto *et al.* Amostragem de solo na agricultura de precisão. In: **AGRICULTURA DE PRECISÃO NO RIO GRANDE DO SUL**. [S. l.: s. n.], 2016. p. 79–98.

CLINGENSMITH, C. M.; GRUNWALD, S.; WANI, S. P. Evaluation of calibration subsetting and new chemometric methods on the spectral prediction of key soil properties in a

data-limited environment. **European Journal of Soil Science**, [s. l.], v. 70, n. 1, p. 107–126, 2019. Disponível em: Acesso em: 4 jul. 2022.

COUTINHO, Marcos A.N. *et al.* Influence of soil sample preparation on the quantification of NPK content via spectroscopy. **Geoderma**, [s. l.], v. 338, n. July 2018, p. 401–409, 2019. Disponível em: <https://doi.org/10.1016/j.geoderma.2018.12.021>.

COUTO, E. Guimarães; STEIN, A.; KLAMT, E. Large area spatial variability of soil chemical properties in central Brazil. **Agriculture, Ecosystems and Environment**, [s. l.], v. 66, n. 2, p. 139–152, 1997.

DE VARGAS, Jocelina P.R. *et al.* Application forms and types of soil acidity corrective: Changes in depth chemical attributes in long term period experiment. **Soil and Tillage Research**, [s. l.], v. 185, p. 47–60, 2019. Disponível em: <https://doi.org/10.1016/j.still.2018.08.014>. Acesso em: 30 jun. 2022.

DUNN, B. W. *et al.* The potential of near-infrared reflectance spectroscopy for soil analysis - A case study from the Riverine Plain of south-eastern Australia. **Australian Journal of Experimental Agriculture**, [s. l.], v. 42, n. 5, p. 607–614, 2002.

FÁBREGA, José R. *et al.* Modeling short-term soil-water distribution of aromatic amines. **Environmental Science and Technology**, [s. l.], v. 32, n. 18, p. 2788–2794, 1998. Disponível em: doi.org/10.1021/es9802394. Acesso em: 4 jan. 2023.

FAGERIA, N. K.; BALIGAR, V. C. Chapter 7 Ameliorating Soil Acidity of Tropical Oxisols by Liming For Sustainable Crop Production. **Advances in Agronomy**, [s. l.], v. 99, p. 345–399, 2008. Disponível em: Acesso em: 19 jul. 2022.

FORTES, R. *et al.* A methodology based on apparent electrical conductivity and guided soil samples to improve irrigation zoning. **Precision Agriculture**, [s. l.], v. 16, n. 4, p. 441–454, 2015.

GOMEZ, Cécile; LAGACHERIE, Philippe; COULOUMA, Guillaume. Regional predictions of eight common soil properties and their spatial structures from hyperspectral Vis-NIR data. **Geoderma**, [s. l.], v. 189–190, p. 176–185, 2012. Disponível em: www.elsevier.com/locate/geoderma. Acesso em: 3 jul. 2022.

GRAHAM R. HUNT; JOHN W. SALISBURY. **Visible and Near Infrared Spectra of Minerals and Rocks: I Silicate Minerals**. [S. l.: s. n.], 1970.

HAN, S *et al.* Cell size selection for site-specific crop management. **Transactions of the ASAE**, [s. l.], v. 37, n. 1, p. 19–26, 1994.

HENGL, Tomislav; ROSSITER, David G.; STEIN, Alfred. Soil sampling strategies for spatial prediction by correlation with auxiliary maps. **Australian Journal of Soil Research**, [s. l.], v. 41, n. 8, p. 1403–1422, 2003.

HUTENGs, Christopher *et al.* In situ and laboratory soil spectroscopy with portable visible-to-near-infrared and mid-infrared instruments for the assessment of organic carbon in soils. **Geoderma**, [s. l.], v. 355, n. August, p. 113900, 2019. Disponível em: <https://doi.org/10.1016/j.geoderma.2019.113900>.

ISO, B S I. 20473: 2007 Optics and photonics. **Spectral bands**, [s. l.], p. 10, 2007.

KAISER, M.; ELLERBROCK, R. H.; GERKE, H. H. Cation Exchange Capacity and Composition of Soluble Soil Organic Matter Fractions. **Soil Science Society of America Journal**, [s. l.], v. 72, n. 5, p. 1278–1285, 2008.

LAUZON, John D *et al.* Spatial variability of soil test phosphorus, potassium, and pH of Ontario soils. **Agronomy Journal**, [s. l.], v. 97, n. 2, p. 524–532, 2005. Disponível em: <https://www.researchgate.net/publication/237348952>. Acesso em: 3 dez. 2019.

LAWRENCE, Patrick G. *et al.* Guiding soil sampling strategies using classical and spatial statistics: A review. **Agronomy Journal**, [s. l.], v. 112, n. 1, p. 493–510, 2020.

LIU, Fei Tony; TING, Kai Ming; ZHOU, Zhi Hua. Isolation forest. *In:* , 2008. **Proceedings - IEEE International Conference on Data Mining, ICDM**. [S. l.: s. n.], 2008. p. 413–422.

LIU, Fei Tony; TING, Kai Ming; ZHOU, Zhi-Hua. **Isolation-based Anomaly Detection**. [S. l.: s. n.], 2012.

MALLARINO, Antonio P; WITTRY, David J. Efficacy of grid and zone soil sampling approaches for site-specific assessment of phosphorus, potassium, pH, and organic matter. **Precision Agriculture**, [s. l.], v. 5, n. 2, p. 131–144, 2004. Disponível em: Acesso em: 3 dez. 2019.

MAYRINK, G. O. *et al.* Determination of chemical soil properties using diffuse reflectance and ion-exchange resins. **Precision Agriculture**, [s. l.], v. 20, n. 3, p. 541–561, 2018. Disponível em: <https://doi.org/10.1007/s11119-018-9597-y>.

METZGER, Konrad *et al.* Mid-infrared spectroscopy as an alternative to laboratory extraction for the determination of lime requirement in tillage soils. **Geoderma**, [s. l.], v. 364, n. December 2019, p. 114171, 2020. Disponível em: <https://doi.org/10.1016/j.geoderma.2020.114171>.

NAWAR, Said *et al.* Delineation of Soil Management Zones for Variable-Rate Fertilization: A Review. *In: ADVANCES IN AGRONOMY*. [S. l.]: Academic Press Inc., 2017. v. 143, p. 175–245.

NGUYEN, T.T.; JANIK, L.J.; RAUPACH, M. Diffuse Reflectance Infrared Fourier Transform (DRIFT) Spectroscopy in Soil Studies. [s. l.], n. 1982, p. 49–67, 1991.

NOVAIS, Roberto Ferreira *et al.* **Fertilidade do solo**. [S. l.: s. n.], 2007.

PADARIAN, Jose; MINASNY, Budiman; MCBRATNEY, Alex B. Machine learning and soil sciences: A review aided by machine learning tools. **SOIL**, [s. l.], v. 6, n. 1, p. 35–52, 2020. Disponível em: Acesso em: 29 jun. 2022.

PARFITT, R. L.; GILTRAP, D. J.; WHITTON, J. S. Contribution of organic matter and clay minerals to the cation exchange capacity of soils. **Communications in Soil Science and Plant Analysis**, [s. l.], v. 26, n. 9–10, p. 1343–1355, 1995.

PEREIRA, Gustavo Willam *et al.* Smart-Map: An Open-Source QGIS Plugin for Digital Mapping Using Machine Learning Techniques and Ordinary Kriging. **Agronomy**, [s. l.], v. 12, n. 6, p. 1350, 2022. Disponível em: <https://www.mdpi.com/2073-4395/12/6/1350>.

PUDEŁKO, Anna; CHODAK, Marcin. Estimation of total nitrogen and organic carbon contents in mine soils with NIR reflectance spectroscopy and various chemometric methods. **Geoderma**, [s. l.], v. 368, n. February, 2020.

QGIS DEVELOPMENT TEAM. **QGIS Geographic Information System**. [S. l.]: QGIS Association, 2023. Disponível em: <https://www.qgis.org>. Acesso em: 9 maio 2023.

RADOČAJ, Dorijan *et al.* The effect of soil sampling density and spatial autocorrelation on interpolation accuracy of chemical soil properties in arable cropland. **Agronomy**, [s. l.], v. 11, n. 12, 2021.

RODRIGUES, Khalil de Menezes *et al.* Spatial Variability in Soil Fertility and Particle Size and Their Effects on Sugarcane Yield. **Sugar Tech**, [s. l.], v. 18, n. 1, p. 39–48, 2016.

ROSSEL, R. A.Viscarra; BEHRENS, T. Using data mining to model and interpret soil diffuse reflectance spectra. **Geoderma**, [s. l.], v. 158, n. 1–2, p. 46–54, 2010. Disponível em: <http://dx.doi.org/10.1016/j.geoderma.2009.12.025>.

SILVA, Elisângela Benedet *et al.* A Regional Legacy Soil Dataset for Prediction of Sand and Clay Content with Vis-Nir-Swir, in Southern Brazil. **Revista Brasileira de Ciência do Solo**, [s. l.], v. 43, n. August, 2019. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-06832019000100304&tlng=en.

STENBERG, Bo *et al.* Visible and Near Infrared Spectroscopy in Soil Science. **Advances in Agronomy**, [s. l.], v. 107, n. C, p. 163–215, 2010a.

STENBERG, Bo *et al.* Visible and Near Infrared Spectroscopy in Soil Science. [s. l.], v. 107, p. 163–215, 2010b. Disponível em: <http://dx.doi.org/10.1016/S0065-2113>. Acesso em: 26 jun. 2022.

TERRA, Fabrício S. *et al.* Soil fertility assessment by Vis-NIR spectroscopy: Predicting soil functioning rather than availability indices. **Geoderma**, [s. l.], v. 337, n. 3, p. 368–374, 2019. Disponível em: <https://doi.org/10.1007/s11119-018-9597-y>.

TEXAS, Instruments. DLP® NIRscan™ Nano EVM User's Guid. **Texas Instruments Incorporated**, [s. l.], n. June 2015, 2015. Disponível em: <http://www.ti.com/tool/dlpnirnanoevm>.

THORLABS INC. **Detectores Fotocondutores para o Infravermelho**. [S. l.], 2020a. Disponível em: https://www.thorlabs.com/newgrouppage9.cfm?objectgroup_id=6479. Acesso em: 18 abr. 2020.

THORLABS INC. **MIR Photovoltaic Detectors, HgCdTe (MCT)**. [S. l.], 2020b. Disponível em: https://www.thorlabs.com/newgrouppage9.cfm?objectgroup_id=11319. Acesso em: 18 abr. 2020.

VALENTE, Sárvio *et al.* Definition of management zones in coffee production fields based on apparent soil electrical conductivity. **Scientia Agricola**, [s. l.], v. 69, n. 3, p. 173–179, 2012a.

VALENTE, Sárvio *et al.* The relationship between apparent soil electrical conductivity and soil properties. [s. l.], p. 683–690, 2012b. Disponível em: www.ccarevista.ufc.br.

VAN ROSSUM, Guido; DRAKE, Fred L. **Python 3 Reference Manual**. Scotts Valley, CA: CreateSpace, 2009.

VISCARRA ROSSEL, R. A. *et al.* Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. **Geoderma**, [s. l.], v. 131, n. 1–2, p. 59–75, 2006.

WIJewardane, Nuwan K.; GE, Yufeng; MORGAN, Cristine L.S. Moisture insensitive prediction of soil properties from VNIR reflectance spectra based on external parameter orthogonalization. **Geoderma**, [s. l.], v. 267, n. January, p. 92–101, 2016. Disponível em: <http://dx.doi.org/10.1016/j.geoderma.2015.12.014>.

WILDING, L. P. Spatial variability: it's documentation, accommodation and implication to soil surveys. *In*: , 1985, Las Vegas. (J Bouma & D. R Nielsen, Org.) **Soil Spatial**

Variability. Pudoc. Las Vegas: [s. n.], 1985. p. 166–194. Disponível em: Acesso em: 30 jun. 2022.

XU, Shengxiang *et al.* Comparison of multivariate methods for estimating selected soil properties from intact soil cores of paddy fields by Vis–NIR spectroscopy. **Geoderma**, [s. l.], v. 310, n. September 2017, p. 29–43, 2018a.

XU, Shengxiang *et al.* Comparison of multivariate methods for estimating selected soil properties from intact soil cores of paddy fields by Vis–NIR spectroscopy. **Geoderma**, [s. l.], v. 310, p. 29–43, 2018b.

ZEAITER, M.; RUTLEDGE, D. Preprocessing Methods. **Comprehensive Chemometrics**, [s. l.], v. 3, p. 121–231, 2009. Disponível em: Acesso em: 28 jun. 2022.