

USO DE APRENDIZADO DE MÁQUINA PARA AVALIAR EFEITOS CAUSADOS POR UMIDADE NA ESPECTROSCOPIA DO SOLO

GABRIEL MOVIO NALIN¹, RICARDO CANAL FILHO², JOSÉ PAULO MOLIN³

¹ Graduando em Engenharia Agrônoma, Depto. De Engenharia de Biosistemas, Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba – SP. Fone: (019)996579066-9066, gabrielnalin@usp.br

² Engenheiro agrônomo, Doutorando, Depto. De Engenharia de Biosistemas, Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba – SP. Fone: (019)3447-8579

³ Engenheiro Agrícola, Professor Titular, Depto. De Engenharia de Biosistemas, Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba – SP. Fone: (019)3447-8579

Apresentado no
Congresso Brasileiro de Agricultura de Precisão e Digital - ConBAP 2024
Porto Alegre, RS, 02 a 04 de julho de 2024

RESUMO: A espectroscopia de refletância no infravermelho próximo e visível (VNIR) é amplamente estudada como método rápido e econômico para inferir quantitativamente as propriedades do solo, visando melhorar a densidade amostral e possibilitar recomendações mais precisas para a produção agrícola. Este estudo avaliou a interferência da umidade na determinação de sete atributos-chave de fertilidade do solo (argila, areia, matéria orgânica, capacidade de troca catiônica, cálcio, magnésio e potássio), utilizando 72 amostras de solo coletadas em campo. Os dados brutos foram empregados para calibrar modelos preditivos via regressão de componentes principais (PCR). Verificou-se que, para parâmetros como matéria orgânica, capacidade de troca catiônica, potássio, cálcio e magnésio, as validações da análise úmida mostraram valores de R^2 superiores em comparação com a análise seca. Os resultados sugerem que a umidade, inicialmente considerada interferente nos espectros NIR, não necessariamente prejudica o desempenho dos modelos de previsão. Além disso, a utilização da regressão de componentes principais emerge como uma alternativa viável para o tratamento prévio das amostras de solo, o que elimina a necessidade de procedimentos como a homogeneização de partículas e superfícies, bem como a moagem e peneiramento.

PALAVRAS-CHAVE: aprendizado de máquina, espectroscopia no infravermelho próximo; quimiometria.

USE OF MACHINE LEARNING TO ASSESS EFFECTS CAUSED BY MOISTURE IN SOIL SPECTROSCOPY

ABSTRACT: Reflectance spectroscopy in the near and visible infrared (VNIR) is widely used as a fast and economical method to quantitatively infer soil properties, aiming to improve sampling density and enable more accurate recommendations for agricultural production. This study evaluated the interference of moisture in the determination of seven key soil fertility attributes (clay, sand, organic matter, cation exchange capacity, calcium, magnesium and potassium.), using 72 soil samples collected in the field. The raw data was used to calibrate predictive models via principal component regression (PCR). It was found that, for parameters such as organic matter, cation exchange capacity, potassium, calcium and magnesium, the wet analysis validations showed higher R^2 values compared to the dry analysis. The results suggest that humidity, initially considered an interfering factor in NIR spectra, does not necessarily impair the performance of prediction models. Furthermore, the utilization of principal component regression emerges as a viable alternative for the pre-treatment of soil samples, which eliminates the need for procedures such as particle and surface homogenization, as well as grinding and sieving.

KEYWORDS: chemometrics, machine learning, near-infrared spectroscopy.

INTRODUÇÃO: A análise de solo desempenha um papel crucial no diagnóstico das condições edáficas, fornecendo uma visão abrangente dos atributos químicos, físicos e biológicos do solo. Essa prática é fundamental para avaliar a capacidade do solo de fornecer nutrientes essenciais às plantas, sua textura, sua capacidade de retenção de água e outros aspectos relevantes para a produção agrícola. Para garantir a precisão e representatividade dos resultados da análise de solo, é essencial adotar procedimentos adequados de coleta

de amostras. A recomendação dominante é a coleta de uma amostra por hectare, uma prática que os pesquisadores identificaram como a quantidade mínima necessária para manter a significância estatística dos dados obtidos. As densidades de amostragem comumente utilizadas na maioria das lavouras demonstram ser ineficazes na captura adequada da variabilidade dos atributos (WOLLENHAUPT et al., 1994; MONTANARI et al., 2012; CHERUBIN et al., 2014; CHERUBIN et al., 2015). Como resultado, a representatividade dos dados é comprometida, afetando a precisão das recomendações de fertilizantes e corretivos a serem aplicados no solo. Uma alternativa para análise de solo convencional é o uso da espectroscopia de reflectância. A espectroscopia de reflectância no visível e no infravermelho próximo (VNIR, 350–2500 nm) é uma técnica rápida, econômica e não destrutiva para determinar atributos do solo. Além disso, exige apenas uma preparação mínima das amostras e não gera resíduos químicos perigosos (STENBERG et al., 2010). Apesar de ser um conceito interessante, o maior obstáculo para a implementação de bibliotecas espectrais de solo em aplicações de campo reside na diversidade de níveis de umidade do solo. Muitos autores relatam que a umidade gera ruído no sinal coletado pelo sensor, reduzindo o desempenho preditivo de modelos de aprendizado de máquina aplicados aos dados. Com isso, um pré-processamento das amostras, que normalmente envolve secagem, moagem e peneiramento é frequentemente aplicado a amostras de solo antes da coleta espectral. Neste sentido, este trabalho se propõe a avaliar a interferência causada pela umidade de amostras, com preparação mínima antes da coleta de espectros NIR, na predição de atributos do solo por modelos de aprendizado de máquina.

MATERIAL E MÉTODOS: A área utilizada para realização do experimento, de aproximadamente 6,0 ha, está localizada no campus da ESALQ/USP, em Piracicaba, SP. A textura do solo da área de estudo é classificada como franco-arenosa, com teor médio de argila variado entre 10 e 13%. Em novembro de 2021, foram coletadas 72 amostras de solo na área, que foram enumeradas de 1 a 72, e subdivididas em dois grupos para serem submetidas para coleta espectral e para análise em laboratório tradicional. Para realizar a coleta espectral em bancada, o espectro do infravermelho próximo (NIR) das amostras de solo foi obtido enquanto elas estavam em condições de umidade de campo. Em seguida, as amostras foram pesadas e colocadas em uma estufa com ventilação forçada a uma temperatura de 105°C por 72 horas para secagem. Após o período de secagem, as amostras foram pesadas novamente para determinar a umidade de campo no momento da primeira coleta espectral, por meio da Equação 1.

$$\theta (\%) = \frac{m_u}{m_s} \times 100 = \frac{m_u - m_s}{m_s} \times 100 \quad (1)$$

em que θ : umidade gravimétrica do solo, em porcentagem; m_u : massa de solo úmido; m_s : massa de solo seco.

Depois, o espectro NIR foi novamente coletado, mantendo a condição em que as amostras saíram do campo (sem moagem e peneiramento), mas com o solo seco. O segundo grupo foi encaminhado para um laboratório responsável por realizar as análises de solo. Os atributos analisados neste estudo foram a matéria orgânica (MO) e a capacidade de troca catiônica (CTC), bem como os teores de potássio (K), cálcio (Ca) e magnésio (Mg) trocáveis no solo. Para a determinação da matéria orgânica, foi empregado o método de oxidação. Já a capacidade de troca catiônica foi calculada considerando as bases mais a acidez total do solo. Os teores de potássio, cálcio e magnésio trocáveis foram analisados utilizando o método de extração com resina.

A aquisição dos dados espectrais foi realizada com o espectrômetro MicroNIR OnSite-W (Viavi Solutions Inc., CA, EUA) com resolução espectral de 908,1-1676,2 nm, com leituras a cada 6,2 nm, resultando em 125 diferentes comprimentos de onda. Os espectros foram coletados na base do compartimento, através de uma janela de safira, transportados por um cabo USB e convertidos para transmissão por meio de um cabo Ethernet, sendo gravados em um computador portátil conectado ao conjunto (Figura 1). Uma placa de 99% reflectância foi usada como referência do branco (máxima reflectância), e o próprio equipamento possui medição interna de referência do preto (mínima reflectância).

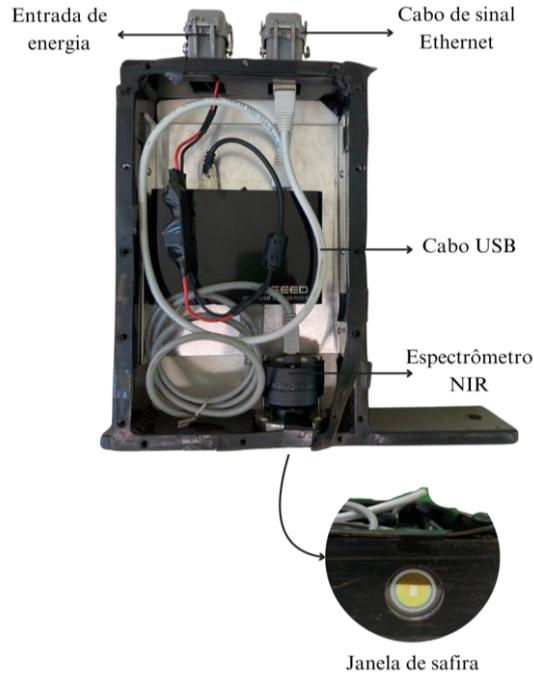


FIGURA 1. Espectrômetro MicroNIR OnSite-W e seus componentes (Viavi Solutions Inc., CA, EUA).

O processamento dos dados espectrais ocorreu no ambiente de programação do software RStudio (RSTUDIO TEAM, 2023). Os dados brutos coletados pelo sensor foram utilizados para calibrar modelos preditivos via regressão de componentes principais (PCR), uma técnica muito aplicada a dados multivariados como espectros NIR do solo, pois consegue reduzir dados ruidosos, redundantes e esparsos (AGARWAL et al., 2021; CHANG et al., 2001). A PCR ocorre em três etapas: (a) uma análise de componentes principais (PCA) é aplicada em toda a matriz de dados; (b) uma regressão linear é usada para obter o vetor de coeficientes de regressão estimados; (c) os autovetores da PCA são usados para obter os estimadores da PCR (Equação 2):

$$\widehat{\beta}_k = v_k \widehat{\delta}_k \quad (2)$$

em que β = estimador da PCR, k pertence de $\{1, \dots, p\}$, p = número de covariáveis, V = conjunto ortonormal de autovetores, δ = vetor de coeficientes dos regressores estimados.

Os modelos preditivos foram calibrados e posteriormente avaliados pelo processo de validação independente, i.e., utilizando amostras que não estiveram presentes na calibração do modelo. Para isso, as 72 amostras de solo foram divididas na proporção de 70% para calibração e 30% para validação, utilizando a técnica de validação cruzada k -fold com $k = 10$. Esta abordagem é recomendada para a avaliação de modelos de aprendizado de máquina, pois tende a fornecer uma estimativa mais imparcial ou conservadora da capacidade do modelo em comparação com outros métodos (JUNG et al., 2018). Basicamente, o processo envolve as seguintes etapas: 1) embaralhamento aleatório dos dados; 2) divisão dos dados em k grupos; 3) para cada grupo, um grupo é mantido como conjunto de validação, enquanto $k-1$ grupos restantes são combinados e utilizados como conjunto de treinamento para ajustar um modelo, que é então avaliado no conjunto de validação e descartado, repetindo-se esse processo para todos os grupos; 4) a habilidade do modelo é resumida pela média das avaliações realizadas. Isso garante que cada amostra seja utilizada uma vez no conjunto de validação e $k-1$ vezes no conjunto de treinamento, proporcionando uma estimativa mais imparcial da capacidade preditiva do modelo. As métricas utilizadas para avaliação dos modelos preditivos foram o coeficiente de determinação (R^2), a raiz do erro quadrático médio (RMSE) e o erro médio absoluto (MAE). Quanto maior o valor do R^2 , e menores o RMSE e MAE, subentende-se que melhor foi o desempenho dos modelos.

Características espectrais dos espectros úmidos e secos, como intensidade, morfologia e feições de absorção, também foram comparadas a fim de avaliar a influência da umidade nos espectros NIR, e avaliar possíveis

interferências notadas pelos modelos preditivos. Para isso, foram utilizadas técnicas de pré-processamento para comparação entre os dois grupos, como a variação normal padrão (SNV), que é um dos métodos de correção de espalhamento mais comuns utilizados em espectros NIR, proposto para remover a interferência multiplicativa do tamanho da partícula (BARNES et al., 1989). A SNV permite centralizar os espectros no valor zero, o que facilita a comparação entre dois grupos coletados em diferentes momentos, pela variação na intensidade e nas feições de absorção a partir desse valor nulo.

RESULTADOS E DISCUSSÃO: Os resultados das validações dos modelos preditivos construídos com base em análises úmidas e secas estão resumidos na Tabela 1. Os parâmetros de MO, CTC, K, Ca e Mg, as validações da análise úmida revelaram valores de R² superiores em comparação com a análise seca. Um R² mais elevado indica um desempenho superior do modelo, uma vez que este parâmetro indica a porcentagem da variância dos dados observados que foi explicada pelo modelo (dados preditos). Adicionalmente, o modelo preditivo relacionado à análise úmida exibe maiores valores de RMSE para CTC e Ca, bem como valores de MAE superiores para MO, CTC e Ca, em comparação com a análise seca. No entanto, a análise seca revelou configurações otimizadas para MO, K e Mg em termos de RMSE, com valores menores, mas registrou valores maiores de MAE para K e Mg.

TABELA 1. Resultados de validação dos modelos de regressão via componentes principais validação utilizando validação cruzada k-fold.

	Análise úmida					Análise seca						
	min	max	a	pc	R ²	RMSE	MAE	NC	R ²	RMSE	MAE	
MO	12,0	- 35,0	23,0	7	0,86	2,29	1,92	5	0,74	2,76	1,85	
CTC	44,5	- 68,4	23,9	5	0,83	4,54	3,36	4	0,64	3,43	2,79	
K	0,4	- 5,0	4,6	5	0,64	0,72	0,61	8	0,46	0,86	0,72	
Ca	10,0	- 34,0	24,0	8	0,50	2,65	2,09	2	0,40	2,10	1,56	
Mg	5,0	- 22,0	17,0	5	0,32	1,53	1,31	2	0,19	1,66	1,43	

min, max, a: mínimo, máximo, e amplitude dos valores observados no banco de dados; pc: número de componentes principais utilizados na regressão; R²: coeficiente de determinação; RMSE: raiz do erro quadrático médio; MAE: erro médio absoluto; MO: matéria orgânica, g dm⁻³; CTC: capacidade de troca catiônica, mmol_c dm⁻³; K: potássio trocável, mmol_c dm⁻³; Ca: cálcio trocável, mmol_c dm⁻³; Mg: magnésio trocável, mmol_c dm⁻³.

A análise do espectro médio de reflectância das 72 amostras de solo secas comparadas ao espectro médio dessas mesmas amostras úmidas comprova que, com a maior presença de água, houve maior absorção da luz NIR (Figura 2). Isso corrobora com o descrito na literatura, observando que a água tende a reduzir a intensidade do espectro (NOCITA et al., 2013; MORELLOS et al., 2016; YI-PING WANG et al., 2020). A análise de morfologia e características espectrais entre os espectros de análise úmida e seca demonstrou diferenças que se repetiram entre as 72 amostras analisadas. Essas diferenças estão ilustradas na Figura 3. A intensidade das feições de absorção aumentou consideravelmente (Figura 3.A). Onde essas feições tiveram intensidade aumentada, houve ainda um prolongamento nos vales formados, principalmente entre 1400 e 1500 nm (Figura 3.B). Entretanto, os espectros conservaram interações solo-espectro entre a análise úmida e seca, com feições, vales e picos de absorção nos mesmos comprimentos de onda. Após 1400 nm, entretanto, é notável que a taxa de reflectância aumentou muito mais no espectro úmido do que no espectro de análise seca (Figuras 3.C e 3.D).

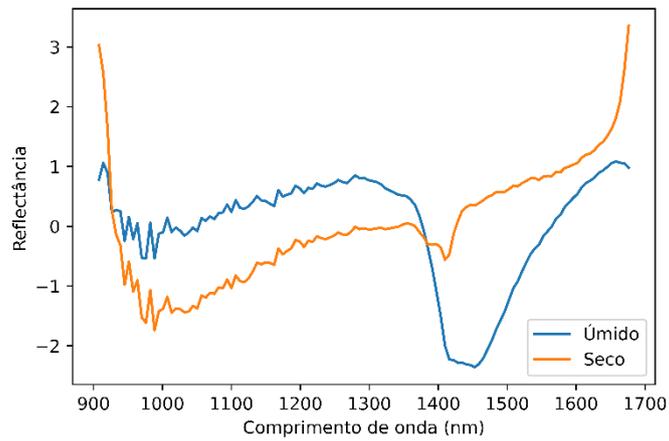


FIGURA 2. Espectro médio das 72 amostras de solo seco e úmido.

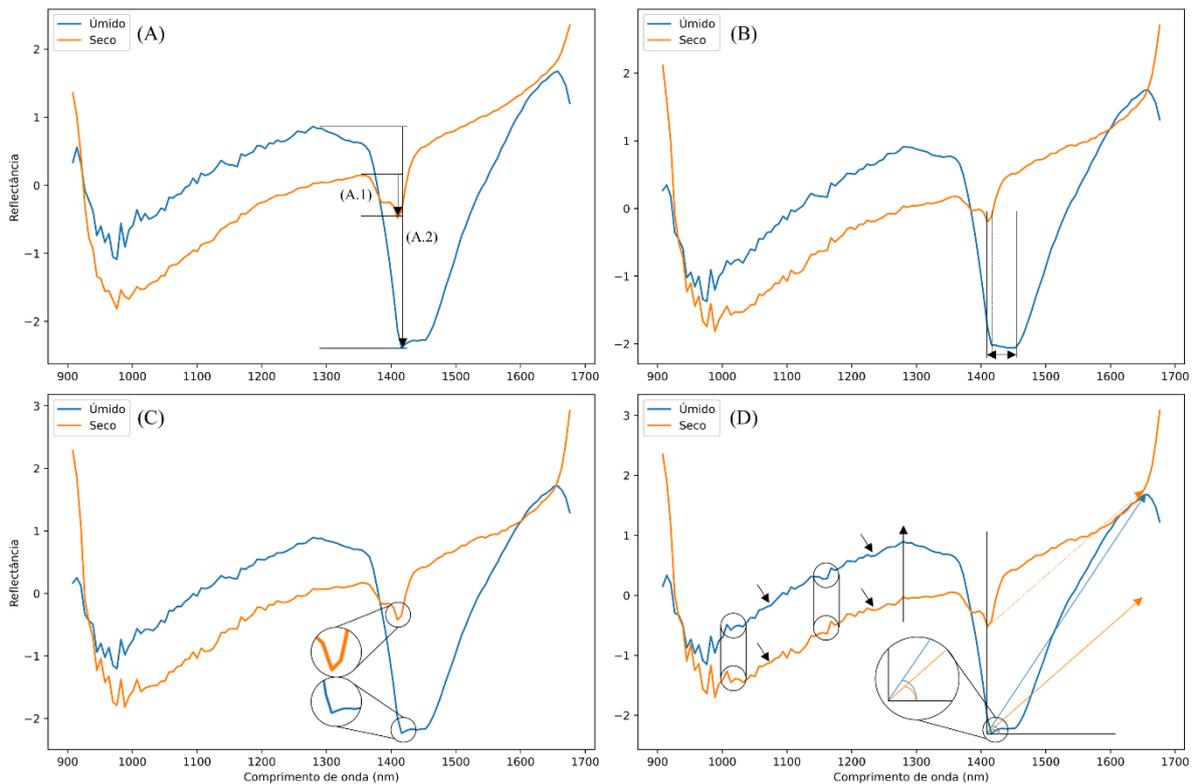


FIGURA 3. Comparação entre quatro amostras de solo e seus respectivos espectros úmidos e secos, destacando feições de absorção, intensidade e morfologia semelhantes e discrepantes entre os dois grupos.

Os valores de R^2 mais próximos de 1 obtidos nas validações dos modelos preditivos elaborados com base em análises úmidas podem ser atribuídos aos efeitos do umedecimento nos espectros de refletância. Quando o solo está molhado, sua capacidade de reflexão da luz diminui em comparação com quando está seco. Isso ocorre porque a presença de água, tanto na superfície quanto nos poros do solo, altera a maneira como a luz é dispersa. A água modifica o índice de refração ao redor das partículas do solo, passando de ar para água, resultando em menos contraste e mais dispersão direta da luz. Isso faz com que a luz percorra um caminho mais longo antes de ser reemitida, aumentando assim a absorção pelo solo (YI-PING WANG et al., 2020).

É importante ressaltar que os resultados obtidos neste estudo diferem do que alguns autores relatam. A umidade do solo é, por vezes, tratada como indesejável para construir modelos preditivos a partir de espectros NIR (STENBERG et al., 2010; NOCITA et al., 2015). Ainda, outro ponto a ser destacado é o de que, mesmo com a coleta espectral sendo realizada sem pré-tratamento das amostras de solo quanto a uniformização de partículas e superfície, pela moagem e peneiramento do solo, os modelos preditivos foram capazes de prever de forma robusta os atributos do solo. A fim de evitar os efeitos da umidade e de uniformizar a superfície da amostra para coleta, autores recorrem a técnicas morosas de secagem, moagem e peneiramento (DEMATTÊ et al., 2017; CEZAR et al., 2019). Além de levar tempo e dispendar mão de obra para tais atividades,

aumentando o tempo entre a amostra chegar ao laboratório e a análise estar disponível ao produtor, os resultados reportados mostram que nem sempre isso pode representar ganhos no desempenho preditivo dos modelos. Isso corrobora com o encontrado por YI-PING WANG et al. (2020), que também reportaram melhores modelos preditivos de MO com um nível adequado de umidade no solo no momento da coleta espectral NIR, em detrimento da coleta com o solo seco.

Os efeitos observados no presente estudo podem ser divididos entre dois fatores: (i) a técnica de modelagem preditiva, utilizando um modelo de redução de dimensionalidade; e (ii) a interação específica da água com o espectro NIR e com feições relacionadas à MO. A modelagem preditiva com a técnica de PCR provê a redução da dimensionalidade do espectro. As 125 variáveis originais, inseridas na calibração do modelo, são transformadas em combinações lineares, conhecidas como componentes principais da variância dos dados. Dessa forma, a PCR utiliza uma extração de variáveis de forma a preservar a estrutura original de variância, o que perde poucas informações do banco de dados original, garantindo ainda a redução de dados ruidosos, esparsos e covariáveis, que reduziriam o desempenho do modelo (JOLLIFFE et al., 2011; AGARWAL et al., 2019; VELLIANGIRI et al., 2019). Entende-se, portanto, que a PCR foi capaz de atuar como um pré-processamento de espectros, sendo capaz de automatizar o processo de limpeza dos dados que normalmente é realizado por técnicas de escolha manual de pré-processamento espectral. Esse efeito benéfico da técnica também é reportado por outros autores para a predição de atributos do solo por espectros NIR (CANAL FILHO, 2023) e de raio-x de fluorescência (WEI et al., 2022).

A interação da água com o espectro NIR e com feições relacionadas à MO também pode explicar os resultados obtidos nesse estudo. Água e argila são conhecidas por ter um efeito de aumento nas feições de absorção do espectro NIR, e redução da intensidade de reflectância do espectro (TERRA et al., 2018; YI-PING WANG et al., 2020). Ainda, as feições de água e argila no NIR podem se sobrepor em sobretons observados entre 1400 e 1500 nm (STENBERG et al., 2010; NOCITA et al., 2015; PASQUINI et al., 2018). Entretanto, em uma área de textura franco-arenosa como a do presente estudo, com teores médios de argila entre 10 e 13%, sabe-se que a MO pode desempenhar papel fundamental na física e fertilidade do solo. A MO também possui, naturalmente, propriedades elevadas de absorção de água (SHEPERD et al., 2002; WEIL & BRADY, 2017). Dessa forma, sugere-se que essas interações da MO com a água, e a resposta espectral semelhante, de redução da intensidade de reflectância e aumento da intensidade das feições de absorção, foi o mecanismo utilizado pelos modelos preditivos que permitiram a melhora do desempenho a partir de espectros úmidos, ao invés de espectros secos.

Para os outros atributos preditos neste trabalho, CTC, K, Ca e Mg, o mecanismo foi, provavelmente, o mesmo da MO, e por isso os resultados reportados demonstram tendência semelhante. A CTC e os nutrientes de plantas não possuem feições de absorção no NIR (STENBERG et al., 2010; NOCITA et al., 2015), portanto, sua predição com essa região espectral depende de calibrações indiretas, em locais onde existe uma covariância entre um atributo de resposta primária (argila, MO, água). Ainda, a MO pode representar até 90% da CTC em solos de textura arenosa (WEIL & BRADY, 2017). A partir daí, uma vez que a CTC é diretamente proporcional à quantidade de cargas negativas do solo, que atraem os cátions de K, Ca e Mg, compreende-se o mecanismo de predição desses atributos no presente estudo, e a resposta ao umedecimento dos espectros semelhante ao observado para os modelos de MO.

Os resultados reportados neste trabalho indicam que nem sempre a umidade, tratada como ruído na espectroscopia NIR de solo, terá efeito negativo no desempenho de modelos preditivos. Muitas vezes, ainda, ela pode ter um efeito sinérgico com a predição de determinados atributos, o que corrobora com outros trabalhos da literatura. Uma vez que a moagem e peneiramento também não foram necessárias para os resultados satisfatórios apresentados neste trabalho, sugere-se que essas etapas de pré-processamento de amostras e espectros podem ser avaliadas e, algumas vezes, substituídas por técnicas que lidem com isso de forma automática (como a PCR). Isso pode ser importante para o crescimento da espectroscopia NIR na análise de solo em laboratório

CONCLUSÃO: Este trabalho avaliou a interferência causada pela umidade em amostras de campo, com preparação mínima antes da coleta de espectros NIR, na predição de atributos do solo por modelos de aprendizado de máquina. Os resultados apresentados sugerem que a presença de umidade, inicialmente considerada como interferência nos espectros NIR do solo, não necessariamente prejudica o desempenho dos modelos de previsão. Em algumas situações, observa-se uma interação positiva entre a umidade e a predição de certos atributos, em consonância com descobertas anteriores documentadas na literatura científica. Adicionalmente, foi constatado que certas técnicas, como a regressão de componentes principais, podem ser empregadas como método de pré-processamento das amostras. Isso pode eliminar a necessidade de tratamentos prévios das amostras de solo, tais como uniformização de partículas e superfície, bem como moagem e

peneiramento do solo. Ainda, pode representar um avanço positivo para o desenvolvimento da técnica em sensoriamento proximal do solo, coletando espectros diretamente no campo, onde a umidade, granulometria e superfície amostral são fatores mais difíceis de controlar.

AGRADECIMENTOS: À Spectral Solutions Comércio e Serviços Ltda. pela parceria que ofereceu o sensor utilizado neste estudo

REFERÊNCIAS:

- AGARWAL, A.; SHAH, D.; SHEN, D.; SONG, D. On Robustness of Principal Component Regression. In: **Advances in Neural Information Processing Systems**. NeurIPS: New Orleans, LA, USA, 2021.
- BARNES, R. J.; DHANOA, M. S.; LISTER, S. J. Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra. **Applied Spectroscopy**, v. 43, n. 5, p. 772-777, 1989.
- CANAL FILHO, R. Online Near-Infrared Spectroscopy for Soil Attributes Prediction. Master's thesis, Universidade de São Paulo, 2023.
- CANAL FILHO, R. et al. Soil Attributes Mapping with Online Near-Infrared Spectroscopy Requires Spatio-Temporal Local Calibrations. **AgriEngineering**, v. 5, n. 3, p. 1163-1177, 1 set. 2023.
- CEZAR, E.; NANNI, M. R.; GUERRERO, C.; da SILVA JUNIOR, C. A.; CRUCIOL, L. G. T.; CHICATI, M. L.; SILVA, G. F. C. Organic Matter and Sand Estimates by Spectoradiometry: Strategies for the Development of Models with Applicability at a Local Scale. **Geoderma**, v. 340, p. 224-233, 2019.
- CHANG, C.W.; LAIRD, D.A.; MAUSBACH, M.J.; HURBURGH, C.R. Near-Infrared Reflectance Spectroscopy-Principal Components Regression Analyses of Soil Properties. **Soil Science Society of America Journal**, v. 65, p. 480-490, 2001.
- DEMATTÊ, J. A. M.; RAMIREZ-LOPEZ, L.; MARQUES, K. P. P.; RODELLA, A. A. Chemometric Soil Analysis on the Determination of Specific Bands for the Detection of Magnesium and Potassium by Spectroscopy. **Geoderma**, v. 288, p. 8-22, 2017.
- JOLLIFFE, I. Principal Component Analysis. **Springer Berlin Heidelberg**, 2011, pp. 1094-1096.
- MARSTON et al. Effects of External Factors on Soil Reflectance Measured On-The-Go and Assessment of Potential Spectral Correction through Orthogonalisation and Standardisation Procedures. **Geoderma**, v. 177, p. 19-36, 1 abr. 2018.
- MORELLOS, A. et al. Machine Learning Based Prediction of Soil Total Nitrogen, Organic Carbon and Moisture Content by Using VIS-NIR Spectroscopy. **Biosystems Engineering**, v. 152, p. 104-116, dez. 2016.
- NOCITA, M. et al. Prediction of Soil Organic Carbon for Different Levels of Soil Moisture Using Vis-NIR Spectroscopy. **Geoderma**, v. 199, p. 37-42, maio 2013.
- NOCITA, M. et al. Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring. **Advances in Agronomy**, v. 132, p. 139-159, 2015.
- PASQUINI, C. Near Infrared Spectroscopy: A Mature Analytical Technique with New Perspectives – A Review. **Analytica Chimica Acta**, v. 1026, p. 8-36, out. 2018.
- RICKLEMYER, R. S.; BROWN, D. J. On-the-Go VisNIR: Potential and Limitations for Mapping Soil Clay and Organic Carbon. **Computers and Electronics in Agriculture**, v. 70, n. 1, p. 209-216, jan. 2010.
- RSTUDIO TEAM. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA. URL <http://www.rstudio.com/>, 2020.
- SHEPHERD, M. A.; HARRISON, R.; WEBB, J. Managing Soil Organic Matter-Implications for Soil Structure on Organic Farms. **Soil Use and Management**, v. 18, p. 284-292, 2002.
- STENBERG, B. Effects of Soil Sample Pretreatments and Standardised Rewetting as Interacted with Sand Classes on Vis-NIR Predictions of Clay and Soil Organic Carbon. **Geoderma**, v. 158, n. 1-2, p. 15-22, ago. 2010.
- STENBERG, B.; VISCARRA ROSSEL, R.A.; MOUAZEN, A.M.; WETTERLIND, J. Visible and Near Infrared Spectroscopy in Soil Science. In: **Advances in Agronomy**. Elsevier: Amsterdam, The Netherlands, 2010; Volume 107.
- TERRA, F. S.; DEMATTÊ, J. A. M.; VISCARRA ROSSEL, R. A. Proximal Spectral Sensing in Pedological Assessments: Vis-NIR Spectra for Soil Classification Based on Weathering and Pedogenesis. **Geoderma**, v. 318, 2018.
- VELLIANGIRI, S.; ALAGUMUTHUKRISHNAN, S.; THANKUMAR J. S. I. A Review of Dimensionality Reduction Techniques for Efficient Computation. **Procedia Computer Science**, v. 165, p. 104-111, 2019.
- VISCARRA ROSSEL, R. A. et al. Visible, Near Infrared, Mid Infrared or Combined Diffuse Reflectance Spectroscopy for Simultaneous Assessment of Various Soil Properties. **Geoderma**, v. 131, n. 1-2, p. 59-75, mar. 2006.

WANG, Y.-P. et al. Effect of Wetting on the Determination of Soil Organic Matter Content Using Visible and Near-Infrared Spectrometer. **Geoderma**, v. 376, p. 114528, out. 2020.

WEI, M. C. F.; CANAL FILHO, R.; TAVARES, T. R.; MOLIN, J. P.; VIEIRA, A. M. C. Dimensionality Reduction Statistical Models for Soil Attribute Prediction Based on Raw Spectral Data. **AI**, v. 3, n. 4, p. 809-819, 2022.

WEIL, R.R.; BRADY, N.C. The Nature and Properties of Soils. 15th Edition, Pearson, New York, 2017.

WIJEWARDANE, N. K.; GE, Y.; MORGAN, C. L. S. Moisture Insensitive Prediction of Soil Properties from VNIR Reflectance Spectra Based on External Parameter Orthogonalization. **Geoderma**, v. 267, p. 92–101, abr. 2016.