

AYMBIRÉ ANGELETTI DA FONSECA

**AVANÇOS NO USO DA ESPECTROSCOPIA NO INFRAVERMELHO-PRÓXIMO
PARA O MONITORAMENTO DA MATÉRIA ORGÂNICA DO SOLO E SUAS
FRAÇÕES**

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

Orientadora: Emanuelle Mercês Barros Soares

Coorientador: Celio Pasquini

**VIÇOSA - MINAS GERAIS
2022**

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

T

F676a
2022

Fonseca, Aymbiré Angeletti da, 1989-
Avanços no uso da espectroscopia no
infravermelho-próximo para o monitoramento da matéria
orgânica do solo e suas frações / Aymbiré Angeletti da Fonseca.
– Viçosa, MG, 2022.

1 tese eletrônica (97 f.): il. (algumas color.).

Texto em português e inglês.

Orientador: Emanuelle Mercês Barros Soares.

Tese (doutorado) - Universidade Federal de Viçosa,
Departamento de Solos, 2022.

Inclui bibliografia.

DOI: <https://doi.org/10.47328/ufvbbt.2022.276>

Modo de acesso: World Wide Web.

1. Solos - Amostragem. 2. Humus. 3. Espectrofotômetro.
4. Análise de regressão. I. Soares, Emanuelle Mercês Barros,
1979-. II. Universidade Federal de Viçosa. Departamento de
Solos. Programa de Pós-Graduação em Solos e Nutrição de
Plantas. III. Título.

CDD 22. ed. 631.417

AYMBIRÉ ANGELETTI DA FONSECA

**AVANÇOS NO USO DA ESPECTROSCOPIA NO INFRAVERMELHO-PRÓXIMO
PARA O MONITORAMENTO DA MATÉRIA ORGÂNICA DO SOLO E SUAS
FRAÇÕES**

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

APROVADA: 22 de fevereiro de 2022

Assentimento:



Aymbiré Angeletti da Fonseca
Autor



Emanuelle Mercês Barros Soares
Orientadora

A Deus

Aos meus pais, Maria da Penha Angeletti e Aymbiré Francisco Almeida da Fonseca

Aos meus avós Mário, Thereza, Francisco e Therezinha

Dedico

AGRADECIMENTOS

Agradeço primeiramente a Deus, por me dar a força necessária para superar os momentos de dificuldade e por conseguir finalizar mais esta etapa da minha vida.

Agradeço aos meus pais por todo o apoio e incentivo dados durante a minha vida e por todos os esforços feitos para promover minha educação. Agradeço aos meus irmãos, Bruno e Gustavo, pelo apoio, amizade e carinho. Ao caçula (Gustavo), agradeço também por contribuir com o seu vasto conhecimento para o desenvolvimento de uma etapa extremamente importante para o desenvolvimento deste trabalho. Agradeço também a Tia Maria, pelos seus conselhos, dicas e apoio sempre que precisei.

A Thais Sarandy, uma mulher sem igual, com quem tenho o prazer de compartilhar os meus dias e as vitórias conquistadas em minha vida, agradeço pelo amor e companheirismo incondicionais, pela dedicação e apoio dado, não apenas no desenvolvimento deste trabalho, mas também em minha vida.

Agradeço a minha orientadora, Professora Dra. Emanuelle Mercês Barros Soares, pela confiança em mim depositada, pela compreensão nos momentos difíceis e por todos os ensinamentos transmitidos durante estes quatro anos.

Agradeço imensamente ao meu Coorientador, Professor Dr. Celio Pasquini, por todos os ensinamentos transmitidos, pela disponibilidade e paciência para ajudar a superar cada dificuldade encontrada no desenvolvimento deste trabalho e por todo o apoio dado nos momentos em que precisei.

Ao Professor Dr. Renato Ribeiro Passos, agradeço por todo o apoio, conselhos, amizade e por me dar a oportunidade de iniciar minha trajetória acadêmica sob sua orientação.

Aos grandes amigos e compadres, Lindomar de Souza Machado e Cynthia Meiry Machado, agradeço pela amizade, conselhos e dicas mesmo com a distância. Aos amigos Klever, Raphael, Ésio, Ricardo, Cristian, Bruno, Hend, Antônio, Ranyer, Bernardo, Daniele, João e Humberto pelo companheirismo e pelos conhecimentos adquiridos.

Agradeço ao professor Dr. Jarbas Jose Rodrigues Rohwedder, e a todos os integrantes do Laboratório de Automação e Instrumentação em Química por todo apoio durante as análises realizadas no Instituto de Química (UNICAMP). Agradeço também ao professor Dr. Luiz Antônio dos Santos Dias e ao Guilherme por todo o apoio durante as análises realizadas no Laboratório de Melhoramento de Oleaginosas (UFV).

Por fim, agradeço ao Departamento de Solos e ao Programa de Pós-Graduação em Solos e Nutrição de Plantas da UFV pela oportunidade de realização do curso; a todos os demais professores, profissionais e amigos que contribuíram para a minha formação e com quem tive a satisfação de conviver durante este período; e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão da bolsa de estudos.

RESUMO

FONSECA, Aymbiré Angeletti da, D.Sc., Universidade Federal de Viçosa, fevereiro de 2022. **Avanços no uso da espectroscopia no infravermelho-próximo para o monitoramento da matéria orgânica do solo e suas frações.** Orientadora: Emanuelle Mercês Barros Soares. Coorientador: Celio Pasquini.

A espectroscopia na região do infravermelho-próximo – NIR (*Near-infrared*) é uma técnica analítica rápida, não destrutiva, de fácil uso, altamente reprodutiva e ambientalmente correta, que pode ser utilizada para o monitoramento da matéria orgânica do solo, um atributo que pode beneficiar diversos serviços ecossistêmicos providos pelo solo. No presente trabalho, aspectos metodológicos que envolvem o uso de espectrofotômetros NIR compactos para determinação de carbono orgânico do solo (COS) foram avaliados, assim como o potencial do uso da espectroscopia NIR (NIRS) para a determinação de C, N, $\delta^{13}\text{C}$ e $\delta^{15}\text{N}$ em frações da matéria orgânica do solo (matéria orgânica particulada e matéria orgânica associada aos minerais). Os resultados indicaram que o aumento da área da amostra sondada pelo espectrofotômetro NIR compacto afeta positivamente a representatividade dos espectros de amostras de solo e proporciona ganhos na acurácia dos modelos de determinação de COS. O uso de um sistema de movimentação automático programado para mover, axial e radialmente, os frascos contendo as amostras de solo sobre o feixe de sondagem ótica de um instrumento compacto aumentou a representatividade dos espectros do solo e resultou em ganhos de até 400 % na acurácia dos modelos. Constatou-se ainda que a performance dos modelos de determinação de COS, desenvolvidos com instrumentos compactos e usando conjuntos de dados amplos e heterogêneos, pode ser beneficiada pela moagem do solo. Contudo, nestas condições, o uso de modelos de regressão local, construídos utilizando apenas as amostras do conjunto de calibração consideradas semelhantes à amostra que se deseja avaliar usando a NIRS, se mostrou uma melhor alternativa. Quando os modelos locais foram construídos utilizando amostras de solo secas em estufa e finamente moídas, foram obtidos os melhores resultados, sendo observada redução no erro de predição de até 15 %. Os resultados demonstraram também que a NIRS pode ser utilizada para caracterização rápida e não-destrutiva da matéria orgânica do solo. Além de fornecer informações sobre os teores totais de C e N, foi possível inferir os teores de C, N e $\delta^{13}\text{C}$ na matéria orgânica particulada e na matéria orgânica associada aos minerais.

Palavras-chave: Espectrofotômetros NIR compactos. Modelos de regressão local. Pré-processamento de amostras de solo. Matéria orgânica particulada. Matéria orgânica associada aos minerais.

ABSTRACT

FONSECA, Aymbiré Angeletti da, D.Sc., Universidade Federal de Viçosa, February, 2022. **Advances in the use of near-infrared spectroscopy for monitoring soil organic matter and its fractions.** Advisor: Emanuelle Mercês Barros Soares. Co-advisor: Celio Pasquini.

Near-infrared spectroscopy (NIR) is a rapid, non-destructive, easy-to-use, highly reproducible and environmentally friendly analytical technique that can be used for monitoring soil organic matter, an attribute that can benefit several ecosystem services provided by soil. In the present work, methodological aspects involving the use of compact NIR spectrophotometers to determine soil organic carbon (SOC) were evaluated, as well as the potential of using NIR spectroscopy (NIRS) to determine C, N, $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ in fractions of soil organic matter (particulate organic matter and mineral-associated organic matter). Through the results, it can be observed that increasing the sample area probed by a compact NIR spectrophotometer positively affects the representativeness of the soil sample spectra, which can increase the accuracy of SOC determination models. Using an automatic system to move, axially and radially, the flasks containing the soil samples over the optical probe beam during spectral data acquisition increase the representativeness of the soil spectra acquired using a compact spectrophotometer and has resulted in gains of up to 400 % in model accuracy. It was also found that the performance of SOC determination models, developed with compact instruments and using large and heterogeneous data sets, can be improved when ground soil samples are used. However, when compact instruments are used under these conditions, the use of local regression models, built using only the samples from the calibration set with similar spectra to the sample being evaluated, has proven to be a better option. The best performance of compact spectrophotometers was obtained using local calibration and finely ground oven-dried soil samples. This protocol provided a reduction of up to 15% in the prediction error. The results also demonstrated that NIRS can be used for rapid and non-destructive characterisation of soil organic matter. In addition to providing information on total C and N contents, it was possible to infer the C, N and $\delta^{13}\text{C}$ contents in particulate organic matter and mineral-associated organic matter.

Keywords: Compact NIR spectrophotometers. Local regression models. Soil sample pre-processing. Particulate organic matter. Mineral-associated organic matter.

SUMÁRIO

Introdução geral	10
Referências	15
Capítulo 1	19
Effect of the sample measurement representativeness on soil carbon determination using near-infrared compact spectrophotometers	19
1. Introduction	20
2. Materials and Methods	22
2.1. Soil samples	22
2.2. Reference values of soil organic carbon	23
2.3. NIR spectrophotometer	23
2.4. Sample Measurement Device	23
2.5. Data acquisition and data processing	23
2.6. Evaluation of reproducibility and representativeness	24
2.7. Soil carbon determination by NIRS	25
3. Results	26
3.1. NIR spectra reproducibility and representativeness	26
3.2. Soil carbon determination by NIRS	34
4. Discussion	39
5. Conclusion	43
References	43
Capítulo 2	48
Performance de espectrofotômetros NIR compactos na determinação de carbono orgânico do solo em escala regional: efeitos do pré-processamento das amostras e uso de modelos locais	48
1. Introdução	49
2. Materiais e métodos	52
2.1. Amostras de solo	52
2.2. Determinação dos teores de carbono	52
2.3. Espectros NIR	53
2.3.1. Preparo das amostras de solo antes da aquisição dos espectros	53
2.3.2. Instrumentação e escaneamento das amostras	54
2.3.3. Pré-processamento dos espectros	54
2.3.4. Remoção de outliers e definição dos conjuntos de calibração e validação	56
2.3.5. Desenvolvimento dos modelos de regressão globais	56
2.3.6. Desenvolvimento dos modelos de regressão locais	57
3. Resultados	58
3.1. Teores de COS	58
3.2. Dados espectrais	58
3.3. Determinação de COS usando modelos globais	60
3.4. Determinação de COS usando modelos locais	62
4. Discussão	64
4.1. Outliers	64
4.2. Dados espectrais	65
4.3. Determinação de COS usando modelos globais	66
4.4. Determinação de COS usando modelos locais	67
5. Conclusão	69
Referências	70

Capítulo 3	75
Spectral information supporting near-infrared spectroscopy to estimate C, N, $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ in soil organic matter fractions	75
1. Introduction	76
2. Material and methods	77
2.1. Soil samples	77
2.2. Reference values for carbon and nitrogen	78
2.3. Near-infrared spectra	79
3. Results	82
3.1. NIR spectra of soil.....	82
3.2. Exploratory analysis of soil spectra.....	83
3.3. Performance of PLSR models.....	83
3.3.1. Carbon.....	83
3.3.2. Nitrogen	86
3.3.3. Ratio between stable isotopes of carbon and nitrogen.....	87
4. Discussion.....	87
4.1. Soil spectra and exploratory analysis	87
4.2. PLSR models.....	89
4.2.1. Carbon.....	89
4.2.2. Nitrogen	90
4.2.3. Ratio between stable isotopes of carbon and nitrogen.....	91
5. Conclusion	92
References	92
Conclusões finais	97

Introdução geral

O solo representa o maior reservatório terrestre de carbono (C), com estoques estimados de aproximadamente 2770 Pg de C nos três (3) metros superficiais (JACKSON et al., 2017). Deste total, 1300-1500 Pg de C estão na camada de 0-1 m de profundidade e, conseqüentemente, estão mais susceptíveis a rápidas alterações em função da atividade antrópica (HARDEN et al., 2018). Aumentar os teores de C orgânico do solo (COS) é um dos maiores desafios do século (MINASNY et al., 2017), dada a importância deste constituinte para diversos serviços ecossistêmicos prestados pelo solo, dentre os quais se destacam o sequestro de C e o fornecimento de alimentos (PROUT et al., 2020). Manejar o solo com o objetivo de aumentar os teores de COS requer o monitoramento frequente deste atributo, o que, por sua vez, requer métodos analíticos capazes de fornecer resultados de forma rápida, fácil e sem gerar resíduos nocivos ao ambiente (DEMATTE et al., 2019b).

O método padrão de determinação de COS é a combustão seca. Porém, a determinação por Walkley-Black é atualmente o método mais utilizado nos laboratórios de análise de solo (VISCARRA ROSSEL et al., 2016), por não demandar instrumentos específicos e de alto valor aquisitivo. A oxidação via úmida proposta por Walkley-Black consiste na oxidação do COS pela adição de uma solução de $K_2Cr_2O_7$ ($0,167 \text{ mol L}^{-1}$) e H_2SO_4 concentrado (WALKLEY; BLACK, 1934). Apenas no Brasil, 4 milhões de determinações de COS são feitas todos os anos (SANTANA; SOUZA; POPPI, 2019). Supondo que 75 % destas determinações são feitas por Walkley-Black, e considerando que a cada amostra de solo analisada por este método são gerados 150 mL de resíduos úmidos contendo cromo (SANTANA; SOUZA; POPPI, 2019), aproximadamente $450.000 \text{ L ano}^{-1}$ de resíduos tóxicos são gerados apenas nesta determinação. Globalmente, este volume pode chegar a $67.500.000 \text{ L ano}^{-1}$, dado que mais de 600 milhões de amostras de solo são analisadas anualmente (DEMATTE et al., 2019b). Além de gerar este grande volume de resíduos nocivos ao ambiente, a oxidação via úmida, assim como outros métodos comumente utilizados nos laboratórios de rotina, possuem limitações que podem prejudicar a repetibilidade e reprodutibilidade dos resultados analíticos, tais como, variações na qualidade dos reagentes utilizados e vulnerabilidade a erros operacionais devido ao elevado número de amostras avaliadas simultaneamente (O'ROURKE; HOLDEN, 2011).

O uso de técnicas espectroscópicas por reflectância difusa na Ciência do Solo tem crescido exponencialmente nas últimas décadas, principalmente por serem métodos analíticos não destrutivos, rápidos, altamente reprodutivos, de fácil uso e com baixo custo operacional

(GUERRERO; ROSSEL; MOUAZEN, 2010; LI; VISCARRA ROSSEL; WEBSTER, 2022). Entre estes métodos, a espectroscopia na região do infravermelho-próximo – NIRS (*near-infrared spectroscopy*), associada ou não à região do visível (Vis-NIRS e NIRS, respectivamente), é a que vem demonstrando maior potencial para ser implementada nos laboratórios de análise de solo como uma alternativa para determinação dos teores de COS (AHMADI et al., 2021; DEMATTÊ et al., 2019a; LI; VISCARRA ROSSEL; WEBSTER, 2022; VISCARRA ROSSEL et al., 2016), principalmente por fornecer bons resultados e por demandar apenas um preparo mínimo das amostras de solo (LI; VISCARRA ROSSEL; WEBSTER, 2022). Contudo, o valor de aquisição destes espectrofotômetros ainda é alto (em geral, maiores que US\$ 40,000.00), o que pode limitar a adoção desta técnica nos laboratórios de rotina e em pequenos grupos de pesquisa.

Os avanços recentes no desenvolvimento de dispositivos eletrônicos e computadores miniaturizados têm possibilitado a produção de diferentes tipos de espectrofotômetros NIR compactos (BEC et al., 2020; PASQUINI, 2018). Embora forneçam informações em uma faixa espectral menor do que a de espectrofotômetros NIR convencionais e Vis-NIR, os instrumentos compactos também podem ser utilizados para quantificar os teores de COS de forma acurada (BARTHÈS et al., 2019; KARYOTIS et al., 2021; MURA et al., 2019; SHARIFIFAR et al., 2019a; SORIANO-DISLA et al., 2017), sendo o baixo custo de aquisição destes espectrofotômetros um importante diferencial para a sua escolha, mesmo quando se deseja trabalhar em condições de laboratório (BARTHÈS et al., 2019). No entanto, os equipamentos compactos apresentam diferenças importantes em relação aos espectrofotômetros convencionais, assim, para implementar o uso destes instrumentos nos laboratórios de análise de solo é necessário definir, primeiramente, os aspectos metodológicos de seu uso.

Uma das maiores dificuldades que podem ser encontradas no uso de espectrofotômetros NIR compactos em matrizes complexas, tal como solo, é a sua reduzida capacidade de sondar a amostra durante a aquisição dos espectros (SANTOS et al., 2013). Esta limitação, intrínseca aos equipamentos compactos, pode prejudicar a representatividade dos dados espectrais adquiridos e, conseqüentemente, a performance dos modelos de predição (PASQUINI; HESPANHOL, 2021). Uma maneira efetiva, de baixo custo e rápida de aumentar a representatividade dos espectros adquiridos com espectrofotômetros NIR compactos é o uso de sistemas automatizados capazes de mover os frascos contendo as amostras sobre o feixe de sondagem ótica durante as medições (PASQUINI; HESPANHOL, 2021). Considerando que este sistema proporciona um aumento na área escaneada pelos equipamentos compactos, o seu

uso pode aumentar a representatividade e a reprodutibilidade dos espectros adquiridos, o que, por sua vez, pode melhorar o desempenho dos espectrofotômetros compactos na determinação de COS.

Embora o uso do sistema de movimentação de amostras possa melhorar a qualidade dos espectros adquiridos, para implementar o uso da NIRS nos laboratórios de rotina é necessário construir modelos regionais, que contemplem diversos tipos de solo. Este processo não é simples, mesmo com o uso de espectrofotômetros convencionais. O solo é uma matriz complexa, gerando espectros que variam amplamente em função da sua constituição mineralógica e de suas características físicas (BARTHÈS et al., 2006; FANG et al., 2018; MILTZ; DON, 2012; VISCARRA ROSSEL; BEHRENS, 2010). Em conjuntos de dados amplos e heterogêneos, estas variações dificultam a modelagem das informações espectrais relacionadas ao COS (HONG et al., 2018), prejudicando a acurácia dos modelos desenvolvidos (LIU et al., 2019). As principais estratégias que podem ser utilizadas para melhorar o desempenho da NIRS nestas condições é o preparo do solo antes da aquisição dos espectros e a estratificação do conjunto de dados para o desenvolvimento de modelos locais.

Idealmente, os espectros NIR são adquiridos usando um processamento mínimo das amostras de solo, utilizando amostras secas ao ar e peneiradas a 2 mm. No entanto, o preparo das amostras antes da aquisição dos espectros, feito através da secagem e moagem do solo, reduz a heterogeneidade física e química do material e minimiza a dominância da água nos espectros NIR (BARTHÈS et al., 2006; MILTZ; DON, 2012; TAMBURINI et al., 2017). Embora os efeitos deste processo ainda sejam controversos, no que diz respeito à performance de modelos desenvolvidos com espectrofotômetros convencionais (BARTHÈS et al., 2006; MILTZ; DON, 2012; TAMBURINI et al., 2017), a capacidade de sondagem destes instrumentos, assim como a sua resolução espectral, é superior à de espectrofotômetros compactos. Considerando as limitações intrínsecas aos equipamentos compactos, é provável que os benefícios do processamento do solo sejam mais evidentes nos modelos desenvolvidos com o uso destes espectrofotômetros.

Outra estratégia que pode ser utilizada para melhorar a performance da NIRS na determinação de COS, em escala regional, é o uso de modelos locais, desenvolvidos a partir da estratificação dos espectros do conjunto de calibração (JACONI; DON; FREIBAUER, 2017; LIU et al., 2019; MOURA-BUENO et al., 2019). A acurácia dos modelos de determinação de COS reduz com o aumento da heterogeneidade do banco de dados utilizado (HONG et al., 2018; VISCARRA ROSSEL et al., 2016), portanto, o uso de modelos locais, desenvolvidos

utilizando apenas o grupo de amostras do conjunto de calibração que apresentam maior similaridade com a amostra que se deseja avaliar, pode proporcionar melhores resultados do que o uso de modelos “globais”, desenvolvidos com o uso de todas as amostras do conjunto de calibração (JACONI; DON; FREIBAUER, 2017; LIU et al., 2019; MOURA-BUENO et al., 2019). Diferentes critérios de similaridade vêm sendo propostos, sempre proporcionando ganhos importantes na acurácia dos modelos de predição de COS desenvolvidos com instrumentos convencionais (JACONI; DON; FREIBAUER, 2017; LIU et al., 2019; MOURA-BUENO et al., 2019), mas, até o presente momento, não existem trabalhos que reportem os efeitos desta prática na performance dos modelos desenvolvidos usando espectrofotômetros NIR compactos.

A definição de todos estes aspectos metodológicos que estão relacionados à otimização do uso de espectrofotômetros NIR compactos, além de melhorar o desempenho destes equipamentos na determinação de COS, podem expandir as aplicações da NIRS em estudos relacionados à matéria orgânica do solo (MOS). A exemplo disto, a quantificação dos teores de C e nitrogênio (N), bem como a composição isotópica destes elementos ($\delta^{13}\text{C}$ and $\delta^{15}\text{N}$, respectivamente), na matéria orgânica particulada – POM (*particulate fraction of MOS*) e na matéria orgânica associada aos minerais – MAOM (*mineral-associated fraction of MOS*), pode contribuir para o melhor entendimento das transformações destes elementos no solo (LAVALLEE; SOONG; COTRUFO, 2019). No entanto, os métodos analíticos atualmente utilizados nestas determinações são laboriosos e demorados, o que, por muitas vezes, limita o seu uso em aplicações de larga escala. A NIRS pode ser utilizada para mensurar diferentes atributos do solo simultaneamente (AHMADI et al., 2021). Portanto, o mesmo espectro utilizado para estimar os teores de COS pode ser utilizado para quantificar o C, N, $\delta^{13}\text{C}$ e $\delta^{15}\text{N}$ nas diferentes frações da MOS, desde que exista uma correlação consistente entre estes atributos e os espectros NIR (MCBRIDE, 2021). Na literatura, existem alguns trabalhos com o uso da NIRS na determinação de C e N na POM e MAOM (BARTHÈS et al., 2008; JACONI et al., 2019; ST. LUCE et al., 2014) e de $\delta^{13}\text{C}$ e $\delta^{15}\text{N}$ na MOS (FUENTES et al., 2012; KANG; GAO; YU, 2017; WINOWIECKI et al., 2017). No entanto, não foram encontrados trabalhos que relatem as informações espectrais que podem respaldar o uso da NIRS nestas determinações, o que é necessário para demonstrar que existe um fundamento químico na aplicabilidade deste método analítico proposto (MCBRIDE, 2021). Assim, para avaliar o potencial dos espectrofotômetros NIR compactos na quantificação de C, N, $\delta^{13}\text{C}$ e $\delta^{15}\text{N}$ em frações da MOS é necessário, primeiramente, encontrar e entender as regiões dos espectros NIR que estão

relacionadas às variações destes atributos no solo e que podem dar suporte e significância aos modelos de regressão.

Considerando todo o contexto apresentado, os objetivos do presente trabalho foram: (i) entender como o aumento da área sondada pelos espectrofotômetros NIR compactos pode influenciar a representatividade dos espectros de amostras de solo e qual é a implicação disto para a performance destes instrumentos na determinação dos teores de COS; (ii) avaliar os efeitos do tipo de preparo da amostra de solo antes da aquisição dos espectros NIR e do uso de modelos de regressão local no desempenho de dois espectrofotômetros NIR compactos, que trabalham em diferentes faixas do espectro NIR, na determinação dos teores de COS; (iii) construir modelos de regressão, baseados nos espectros NIR, para determinar os teores de C, N, $\delta^{13}\text{C}$ e $\delta^{15}\text{N}$ nas frações POM e MAOM da MOS e identificar as informações espectrais que suportam estes modelos. Para atingir estes objetivos, três estudos foram desenvolvidos em paralelo e os seus resultados são apresentados em três capítulos. Capítulo 1: Efeito da representatividade na medição da amostra na determinação de carbono do solo usando espectrofotômetros no infravermelho-próximo compactos (*Effect of the sample measurement representativeness on soil carbon determination using near-infrared compact spectrophotometers*); Capítulo 2: Performance de espectrofotômetros NIR compactos na determinação de COS em escala regional: Efeitos do pré-processamento das amostras e uso de modelos locais; Capítulo 3: Informações espectrais que suportam o uso da espectroscopia no infravermelho-próximo para determinar C, N, $\delta^{13}\text{C}$ e $\delta^{15}\text{N}$ nas frações da matéria orgânica do solo (*Spectral information supporting near-infrared spectroscopy to estimate C, N, $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ in soil organic matter fractions*).

Referências

AHMADI, A. et al. Soil properties prediction for precision agriculture using visible and near-infrared spectroscopy: A systematic review and meta-analysis. **Agronomy**, v. 11, n. 3, 2021.

BARTHÈS, B. G. et al. Determination of total carbon and nitrogen content in a range of tropical soils using near infrared spectroscopy: Influence of replication and sample grinding and drying. **Journal of Near Infrared Spectroscopy**, v. 14, n. 5, p. 341–348, 2006.

BARTHÈS, B. G. et al. Determining the distributions of soil carbon and nitrogen in particle size fractions using near-infrared reflectance spectrum of bulk soil samples. **Soil Biology and Biochemistry**, v. 40, n. 6, p. 1533–1537, 2008.

BARTHÈS, B. G. et al. Performance comparison between a miniaturized and a conventional near infrared reflectance (NIR) spectrometer for characterizing soil carbon and nitrogen. **Geoderma**, v. 338, n. February 2018, p. 422–429, mar. 2019.

BEĆ, K. B. et al. Handheld near-infrared spectrometers: Where are we heading? **NIR news**, v. 31, n. 3–4, p. 28–35, 2020.

DEMATTE, J. A. M. et al. The Brazilian Soil Spectral Library (BSSL): A general view, application and challenges. **Geoderma**, v. 354, n. August, p. 113793, 2019a.

DEMATTE, J. A. M. et al. Soil analytical quality control by traditional and spectroscopy techniques: Constructing the future of a hybrid laboratory for low environmental impact. **Geoderma**, v. 337, p. 111–121, mar. 2019b.

FANG, Q. et al. Visible and Near-Infrared Reflectance Spectroscopy for Investigating Soil Mineralogy: A Review. **Journal of Spectroscopy**, v. 2018, p. 1–14, 2018.

FUENTES, M. et al. NIR Spectroscopy: An Alternative for Soil Analysis. **Communications in Soil Science and Plant Analysis**, v. 43, n. 1–2, p. 346–356, jan. 2012.

GUERRERO, C.; ROSSEL, R. A. V.; MOUAZEN, A. M. Special issue “Diffuse reflectance spectroscopy in soil science and land resource assessment.” **Geoderma**, v. 158, n. 1–2, p. 1–2, 2010.

HARDEN, J. W. et al. Networking our science to characterize the state, vulnerabilities, and management opportunities of soil organic matter. **Global Change Biology**, v. 24, n. 2, p. e705–e718, 2018.

HONG, Y. et al. Transferability of Vis-NIR models for Soil Organic Carbon Estimation between Two Study Areas by using Spiking. **Soil Science Society of America Journal**, v. 82, n. 5, p. 1231–1242, 2018.

JACKSON, R. B. et al. The Ecology of Soil Carbon: Pools, Vulnerabilities, and Biotic and Abiotic Controls. **Annual Review of Ecology, Evolution, and Systematics**, v. 48, p. 419–445, 2017.

JACONI, A. et al. Log-ratio transformation is the key to determining soil organic carbon fractions with near-infrared spectroscopy. **European Journal of Soil Science**, v. 70, n. 1, p. 127–139, 2019.

JACONI, A.; DON, A.; FREIBAUER, A. Prediction of soil organic carbon at the country scale: stratification strategies for near-infrared data. **European Journal of Soil Science**, v. 68, n. 6, p. 919–929, nov. 2017.

KANG, H.; GAO, H.; YU, W. Evaluation of spectral pretreatments, spectral range, and regression methods for quantitative spectroscopic analysis of soil organic carbon composition. **Spectroscopy Letters**, v. 50, n. 3, p. 143–149, 16 mar. 2017.

KARYOTIS, K. et al. Evaluation of a Micro-Electro Mechanical Systems Spectral Sensor for Soil Properties Estimation. **Land**, v. 10, n. 1, p. 63, 13 jan. 2021.

LAVALLEE, J. M.; SOONG, J. L.; COTRUFO, M. F. Conceptualizing soil organic matter into particulate and mineral-associated forms to address global change in the 21st century. **Global Change Biology**, v. 26, n. 1, p. 261–273, 2 jan. 2019.

LI, S.; VISCARRA ROSSEL, R. A.; WEBSTER, R. The cost-effectiveness of reflectance spectroscopy for estimating soil organic carbon. **European Journal of Soil Science**, v. 73, n. 1, 30 jan. 2022.

LIU, S. et al. Estimating forest soil organic carbon content using vis-NIR spectroscopy: Implications for large-scale soil carbon spectroscopic assessment. **Geoderma**, v. 348, n. December 2018, p. 37–44, 2019.

MCBRIDE, M. B. Estimating soil chemical properties by diffuse reflectance spectroscopy: Promise versus reality. **European Journal of Soil Science**, v. 73, n. 1, p. 1–11, 11 jan. 2022.

MILTZ, J.; DON, A. Optimising sample preparation and near infrared spectra measurements of soil samples to calibrate organic carbon and total nitrogen content. **Journal of Near Infrared Spectroscopy**, v. 20, n. 6, p. 695–706, 2012.

MINASNY, B. et al. Soil carbon 4 per mille. **Geoderma**, v. 292, p. 59–86, 2017.

MOURA-BUENO, J. M. et al. Stratification of a local VIS-NIR-SWIR spectral library by homogeneity criteria yields more accurate soil organic carbon predictions. **Geoderma**, v. 337, n. October 2018, p. 565–581, 2019.

MURA, S. et al. Vibrational spectroscopy and Aquaphotomics holistic approach to determine chemical compounds related to sustainability in soil profiles. **Computers and Electronics in Agriculture**, v. 159, n. February, p. 92–96, 2019.

O'ROURKE, S. M.; HOLDEN, N. M. Optical sensing and chemometric analysis of soil organic carbon - a cost effective alternative to conventional laboratory methods? **Soil Use and Management**, v. 27, n. 2, p. 143–155, 2011.

PASQUINI, C. Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. **Analytica Chimica Acta**, v. 1026, p. 8–36, out. 2018.

PASQUINI, C.; HESPANHOL, M. C. A rotational-linear sample probing device to improve the performance of compact near-infrared spectrophotometers. **Microchemical Journal**, v. 170, p. 106747, nov. 2021.

PROUT, J. M. et al. What is a good level of soil organic matter? An index based on organic carbon to clay ratio. **European Journal of Soil Science**, v. 72, p. 2493–2503, 2020.

SANTANA, F. B.; SOUZA, A. M.; POPPI, R. J. Green methodology for soil organic matter analysis using a national near infrared spectral library in tandem with learning machine. **Science of the Total Environment**, v. 658, p. 895–900, 2019.

SANTOS, C. A. T. DOS et al. A Review on the Applications of Portable Near-Infrared Spectrometers in the Agro-Food Industry. **Applied Spectroscopy**, v. 67, n. 11, p. 1215–1233, 1 nov. 2013.

SHARIFIFAR, A. et al. Evaluating a low-cost portable NIR spectrometer for the prediction of soil organic and total carbon using different calibration models. **Soil Use and Management**, v. 35, n. 4, p. 607–616, dez. 2019.

SORIANO-DISLA, J. M. et al. Evaluation of the performance of portable visible-infrared instruments for the prediction of soil properties. **Biosystems Engineering**, v. 161, p. 24–36, 2017.

ST. LUCE, M. et al. Rapid determination of soil organic matter quality indicators using visible near infrared reflectance spectroscopy. **Geoderma**, v. 232–234, p. 449–458, nov. 2014.

TAMBURINI, E. et al. Effects of Moisture and Particle Size on Quantitative Determination of Total Organic Carbon (TOC) in Soils Using Near-Infrared Spectroscopy. **Sensors**, v. 17, n. 10, p. 2366, 17 out. 2017.

VISCARRA ROSSEL, R. A. et al. A global spectral library to characterize the world's soil. **Earth-Science Reviews**, v. 155, p. 198–230, 2016.

VISCARRA ROSSEL, R. A.; BEHRENS, T. Using data mining to model and interpret soil diffuse reflectance spectra. **Geoderma**, v. 158, n. 1–2, p. 46–54, 2010.

WALKLEY, A.; BLACK, I. A. An examination of the degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. **Soil Science**, v. 37, n. 1, p. 29–38, jan. 1934.

WINOWIECKI, L. A. et al. Landscape-scale assessments of stable carbon isotopes in soil under diverse vegetation classes in East Africa: application of near-infrared spectroscopy. **Plant and Soil**, v. 421, n. 1–2, p. 259–272, 16 dez. 2017.

Capítulo 1

Effect of the sample measurement representativeness on soil carbon determination using near-infrared compact spectrophotometers

Abstract

Compact or miniaturized near-infrared (NIR) spectrophotometers are inexpensive instruments that can boost the use of NIR spectroscopy (NIRS) for determining soil organic carbon (SOC). Nevertheless, these instruments are deficient concerning the probing capability due the small area of the sample assessed in spectra measurement. Protocols normally used during spectral data acquisition can be insufficient to obtain representative spectra, which can affect the performance of the models to estimate SOC. This paper aimed to understand how the representativeness of soil spectra is affected by increasing the area probed during spectra acquisition and determine its effect on the performance of multivariate regression models constructed using compact NIR spectrophotometers for SOC determination. Four measurement modes of the flask containing the soil sample were evaluated using a compact FT-NIR. Measurement modes I (full radial and axial scanning of sample flask) and II (full radial scanning in a fixed axial line of sample flask) were performed using a rotational-linear sample probing device to move the sample flask over the optical probing beam during spectral data acquisition. Measurement modes III and IV, in turn, were performed in static mode. Using the measurement mode III, the spectral data were acquired at variable points of the sample flask, and using the measurement mode IV, the spectral data were acquired at a single fixed point in the sample flask. The study was developed in two stages. Firstly, five spectra of three soil samples (selected according to their carbon content) were acquired using the measurement modes. The repeatability and representativeness of the spectral data were then evaluated using principal component analysis (PCA). Partial least square regression (PLSR) models were constructed to determine the quantitative effects of the representativeness of the four measurement modes used to acquire the spectral data set. The calibration and independent validation sets were composed of 49 and 20 soil samples, respectively. The increase in the sample area probed using the a rotational-linear sample probing device resulted in more representative and reproducible spectral data, which improves the performance of compact spectrophotometer in SOC determination. Significant enhancements in SOC determination ranged from 146 to 400 % were found, considering the most (mode I) over less representative measurement modes. These

results were verified by a comparative randomization test between models' prediction performances. In this context, using a rotational-linear sample probing device is recommended for SOC determinations using compact NIR spectrophotometers.

Keywords: Soil near-infrared spectra representativeness; rotational-linear sample probing device; compact near-infrared FT spectrophotometer; soil organic matter determination; Near-infrared (NIR) spectroscopy; NIRS.

1. Introduction

Increasing soil organic carbon (SOC) stocks have been considered a possible solution to mitigate global climate changes (MINASNY et al., 2017). Soil organic matter is responsible for the functionality of several ecosystem services and can act as a source or sink of carbon (C), which can impact the concentration of carbon dioxide in the atmosphere (BÜNEMANN et al., 2018; MINASNY et al., 2017; SANDERMAN; HENGL; FISKE, 2017). Managing SOC requires characterizing this soil attribute more extensively. The method predominantly used in soil analysis laboratories to determine SOC (dichromate oxidation) is laborious, time-consuming, and generates large amounts of hazardous wastes (DEMATTE et al., 2019b; SANTANA; SOUZA; POPPI, 2019). In contrast, near-infrared (NIR) spectroscopy (NIRS) is a fast and environmental-friendly analytical technique (PASQUINI, 2018) that has proven to be a valuable technique to determine SOC content (BELTRAME et al., 2016; DEMATTE et al., 2019b; SANTANA; SOUZA; POPPI, 2019). Nevertheless, purchasing a conventional NIR spectrophotometer remains rather expensive (BARTHÈS et al., 2019), mainly for small laboratories. Recently, several types of low-cost, compact/miniaturized NIR spectrophotometers were developed (BEĆ et al., 2020; BEĆ; GRABSKA; HUCK, 2021; CROCOMBE, 2018; PASQUINI, 2018). These compact instruments may spread the use of NIRS by soil researchers and in soil analysis laboratories. However, factors that can affect their performance in SOC determinations need to be better understood to effectively achieve this goal.

Near-infrared spectroscopy is an analytical technique based on the interaction between radiation and matter (PASQUINI, 2003, 2018). When NIRS is used to analyze soil samples spectral information is usually acquired through diffuse reflectance (GUERRERO; ROSSEL; MOUAZEN, 2010). Thus, the optical beam is impinged on the grounded sample and, depending

on its composition, the radiant energy can be selectively and partially absorbed in the NIR wavelength range by different vibrational modes of the chemical bonds and reemitted (reflected) towards the instrument detector (BURNS; CIURCZAK, 2007). Conventional NIR spectrophotometers probe relatively large areas of solid samples to produce a single spectrum that may represent their average composition (SILVA; PASQUINI, 2015). On the other hand, compact NIR spectrophotometers typically have a small optical window. Thus, they only probe a few millimeters squared of grounded solid samples to produce the final spectrum (PASQUINI; HESPANHOL, 2021; SANTOS et al., 2013). This intrinsic limitation of the compact instruments can affect the representativeness of the spectral data acquired, mainly when heterogeneous samples are analyzed (PASQUINI; HESPANHOL, 2021; REEVES, 1996).

Soil samples present high physical and chemical heterogeneity, so, even when spectrophotometers with higher probing capability are used, the spectral data acquired may not represent the average composition of the laboratory sample, which harms the performance of multivariate regression models usually employed to determine SOC (MILTZ; DON, 2012; TAMBURINI et al., 2017).

Two strategies usually are adopted to increase the representativeness of the final spectrum: (i) reduction of the heterogeneity of the sample by intensive grinding (MILTZ; DON, 2012; TAMBURINI et al., 2017), which is a laborious and time/energy-consuming process; or (ii) use of average spectra obtained from many scans collected over a more significant fraction of the heterogeneous sample (JACONI et al., 2019; LIN et al., 2017; SANTANA; SOUZA; POPPI, 2019; THOMSEN et al., 2009). To estimate soil properties using compact spectrophotometers it is usual to scan up to three points on the sample to obtain the final average spectra (Table 1). However, the use of this protocol can still result in the acquisition of poorly representative spectra, as demonstrated in this work. The acquisition of average spectra can be improved by rotating devices used to move the samples over the optical probing beam during scanning (PASQUINI; HESPANHOL, 2021). There were no rotation devices specifically designed to work with compact instruments; however, nowadays, a new rotational-linear sample probing device (RLSPD) capable of accommodating different types of compact NIR instruments is available (PASQUINI; HESPANHOL, 2021). This device can be built easily using low-cost components, as shown in Pasquini and Hespanhol (2021) where a detailed schematic of this device can be found. Considering that the RLSPD can probe the sample flask in an axial and radial direction, it is hypothesized that this device could increase the

representativeness of soil NIR spectra and improve the performance of compact NIR spectrophotometers in SOC determination.

Table 1. Spectral data acquisition protocols using portable spectrophotometers on soil samples

Soil granulometry	Spectral data acquisition protocol	Reference
0.2 mm	Measurement performed on a single point of the soil sample	(BARTHÈS et al., 2019)
2mm	Two measurements in different positions of the petri dishes with soil samples	(MURA et al., 2019)
2 mm	Three measurements in different positions of the petri dishes with soil samples	(AHMAD JANI et al., 2021; KARYOTIS et al., 2021; SHARIFIFAR et al., 2019b)
2 mm	Four measurements in different positions of the petri dishes with soil samples	(SORIANO-DISLA et al., 2017)
2 mm	Five measurements in different positions of the petri dishes with soil samples	(SUT et al., 2012)

The main objective of this work was to understand: (i) how the representativeness of soil NIR spectra is altered by increasing the sample area probed by a compact spectrophotometer during spectral data acquisition by different measurement protocols; and (ii) how the representativeness of soil spectra impacts the performance of a compact NIR spectrophotometer in SOC determination using multivariate models.

2. Materials and Methods

2.1. Soil samples

This study used two sets of topsoil samples, a calibration set, and a validation set, both obtained from a field experiment conducted in a post-mining area, in which different management practices were employed to increase SOC levels. This experiment was conducted over a total area comprising 2 km², and the soil is classified as a Clay Ferralsol (IUSS WORKING GROUP WRB, 2014). A complete description of the experiment can be found in (COSTA, 2020).

The calibration set comprised 49 soil samples (air-dried and sieved at an average grain size of 2 mm) selected according to their SOC content previously determined. Values of SOC in this sample set varied in the range 0.65 - 3.46 % (m/m). In a second moment, other 20 independent soil samples (validation set), air-dried and sieved at 2 mm, were randomly selected from the same field experiment, and used to validate the SOC multivariate regression models constructed based on the NIR spectral data. Values of SOC in the validation set were in the range 0.90 - 3.39 % (m/m) ensuring models interpolation.

2.2. Reference values of soil organic carbon

Subsamples of soil (< 2 mm) with 10 g were crushed in ball mill until their entire mass passed through a 0.2 mm sieve. Then, the reference values of SOC were determined by the dichromate oxidation method (YEOMANS; BREMNER, 1988).

2.3. NIR spectrophotometer

Spectral data were acquired using a NeoSpectra micro-FT instrument (Si-Ware, Egypt). This compact/miniaturized FT-NIR, manufactured by Si-Ware Systems, operates in the 1350-2500 nm wavelength range with a spectral resolution of 16 nm at 1550 nm. Spectra were composed by 257 points (wavelengths) after Fourier transformation of the mean interferogram. The integration time used per scan was 20 s. Although no specific information on the number of spectra acquired per second is available, the manufacturer mentions that within this integration time, thousands of spectra are scanned to calculate the final average spectrum.

2.4. Sample Measurement Device

The spectrophotometer was accommodated in the RLSPD. This device consists of an acrylic box with mechanical components and electronics employed to move the sample flask during spectrum acquisition by diffuse reflectance. Soil samples were transferred to cylindrical glass vials with 20 mL volume (Sigma-Aldrich, ref.DWK986541-500EA). A complete description of the RLSPD components and information on the device's operating mode can be found in Pasquini and Hespanhol (2021).

2.5. Data acquisition and data processing

Four spectral data acquisition modes were evaluated: (I) full radial and axial scanning (six revolutions and a simultaneous complete 4 cm long linear movement) using the RLSPD to

move the samples over the optical probing beam during a measurement cycle (20 s); (II) full radial scanning (six revolution in a fixed axial line of the flask) using RLSPD to move the samples over the optical probing beam during a measurement cycle (20 s); (III) scanning at variable random points (20 s each) of the sample flask, performed in static mode; (IV) scanning at a single fixed point (20 s) of the sample flask, also performed in static mode. All measurement modes produce average spectra resulted from the same number of scans, ensuring the same instrumental signal to noise ratio.

Each measurement mode was performed using the sample packing in two forms. In the first form (A), the total volume of the flasks was filled with soil (particle size < 2 mm), so the soil particles remained static inside of the flask when it was moved during the spectral acquisition. In the second form (B), only two-thirds of the flask volume was filled with soil (particle size < 2 mm), thus, when rotation is employed (modes I and II), the soil particles can move radially inside the flask during measurement.

Data sets were processed using the Unscrambler 11.0 software (CAMO Analytics, Norway). The randomization test (VOET, 1994), used to statistically compare the accuracy of the multivariate models, was processed using the Matlab 2000 rev.12 software, using 20,000 iterations to improve the robustness of the calculation of the p-value.

2.6. Evaluation of reproducibility and representativeness

A preliminary qualitative test was carried out using three soil samples selected from the calibration set (section 2.1), according to their SOC content (low SOC content - 0.65 %, medium SOC content - 1.77 % and high SOC content - 3.46 %) to assess the reproducibility and representativeness of the NIR spectra collected using the different proposed measurement modes (I, II, III, IV).

Firstly, twelve spectra (20 s integration time) of each soil sample were acquired using full radial and axial scanning (mode I) to obtain an average spectrum that best represents the soil sample composition (the expected best representative spectrum). Secondly, five individual spectra (20 s integration time) of each sample were acquired using the four measurement modes. This procedure was performed using a fully packed flask and a 2/3-packed flask. As a reference of 100 % reflectance, a flask fully packed with Spectralon[®] was scanned every five consecutive measurements.

The spectral data was pre-processed to minimize the effect of radiation scattering. The Second-derivative using the Savitzky-Golay algorithm with a window size of 15 data points

was used, 2nd-degree polynomial (SAVITZKY; GOLAY, 1964). In addition, the spectral regions between 1350-1365 nm and 2470-2500 nm, showing lower signal-to-noise ratio after transformation by second derivative, were removed for further processing.

Principal components analysis (PCA) was used to evaluate the reproducibility of the five NIR spectra collected from samples measured using modes I, II, III, and IV, using fully and 2/3 packed flasks. The PCA was also used to assess the similarity among NIR spectra, collected by the different measurement modes and the most representative spectrum of each soil sample.

2.7. Soil carbon determination by NIRS

The 49 calibration samples were used to construct the soil models to determine SOC from spectra collected using the proposed measurement modes (I, II, III, and IV). The 20 soil independent samples of the second dataset (validation set), in turn, were used to validate the models. The sample's measurement was conducted in a random order. The final spectra of each soil sample used for model development and validation were the average of three scans (20 s integration time each) to ensure better repeatability. The flask fully packed with Spectralon[®] was scanned every five samples (15 measurements) to produce a fresh reference, minimizing possible instrumental drift.

As in the reproducibility and representativeness test, the spectral data set was transformed by second-derivative using Savitzky-Golay as described before, and the noisy regions between 1350-1365 nm and 2470-2500 nm were removed from the transformed spectra. Partial least square regression (PLSR) was performed using the leave-one-out cross-validation technique to calibrate the models. Outliers in calibration and validation sets were identified using Hotelling statistics and spectral residuals. Just one spectral outlier was identified in calibration set and there were no outliers detected in validation set. The variables producing the more stable regression coefficients were identified by the Jack-knife algorithm of the Unscrambler (WESTAD; MARTENS, 2000). The root means square error (RMSE) of cross-validation (RMSECV) and prediction (RMSEP, using the validation set), the coefficient of determination (R^2), and bias were used to assess the models' performance (Table 2) (BELLON-MAUREL et al., 2010; OLIVIERI, 2018).

Table 2. Metrics used to assess the performance of the models

Metrics	Formula ^a
Root means square error (RMSE)	$RMSE (\%) = \sqrt{\frac{\sum_{i=1}^{i=n} (\hat{y}_i - y_i)^2}{n}}$
Coefficient of determination (R^2)	$R^2 = 1 - \frac{\sum_{i=1}^{i=n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{i=n} (y_i - \bar{y})^2}$
Bias	$Bias = \frac{1}{n} \sum_{i=1}^{i=n} (\hat{y}_i - y_i)$

^a \hat{y}_i = predicted values; y_i = reference values; \bar{y} = mean of reference values; n is the number of samples.

3. Results

3.1. NIR spectra reproducibility and representativeness

The five NIR spectra acquired using the measurement modes I (A and B), II (A and B), III (A and B), and IV (A and B), as well as their 2nd-derivative, showed similar qualitative features for soils with low, medium, and high SOC contents (Figures 1-6). The spectra acquired using the RLSPD (modes I and II) were similar, and only minor variations were observed among them. In contrast, when the RLSPD was not used (modes III and IV), visible differences among the features of the five acquired spectra can be observed. Variations among the five spectra acquired at different points on the sample flask (measurement mode III) were observed in several wavelengths of the NIR region. On the other hand, when mode IV (scanning at a fixed point of the sample flask) was used, the main difference among the spectra occurred close to 1940 nm.

The variability of the five NIR spectra collected using the different measurement modes (I, II, III, and IV) was evaluated using PCA (Figure 7). When considered together, the first two principal components capture at least 88% of the variance present in the 2nd derivative spectra sets. Regardless of the SOC content in the sample, the scores of the spectra collected using the measurement modes III and IV showed higher dispersion, which indicates a more deficient measurement reproducibility and/or representativeness. The score plot for the first two principal components also showed that the five spectra acquired using measurement modes I and II present a lower dispersion (Figure 7) compared to measurement modes II and IV. Although no

visible differences can be noted among the features of the spectra measured using full flask (A) and 2/3-filled flask (B), the PCA scores evidence a lower dispersion of the spectral data when full flasks are used (Figure 7).

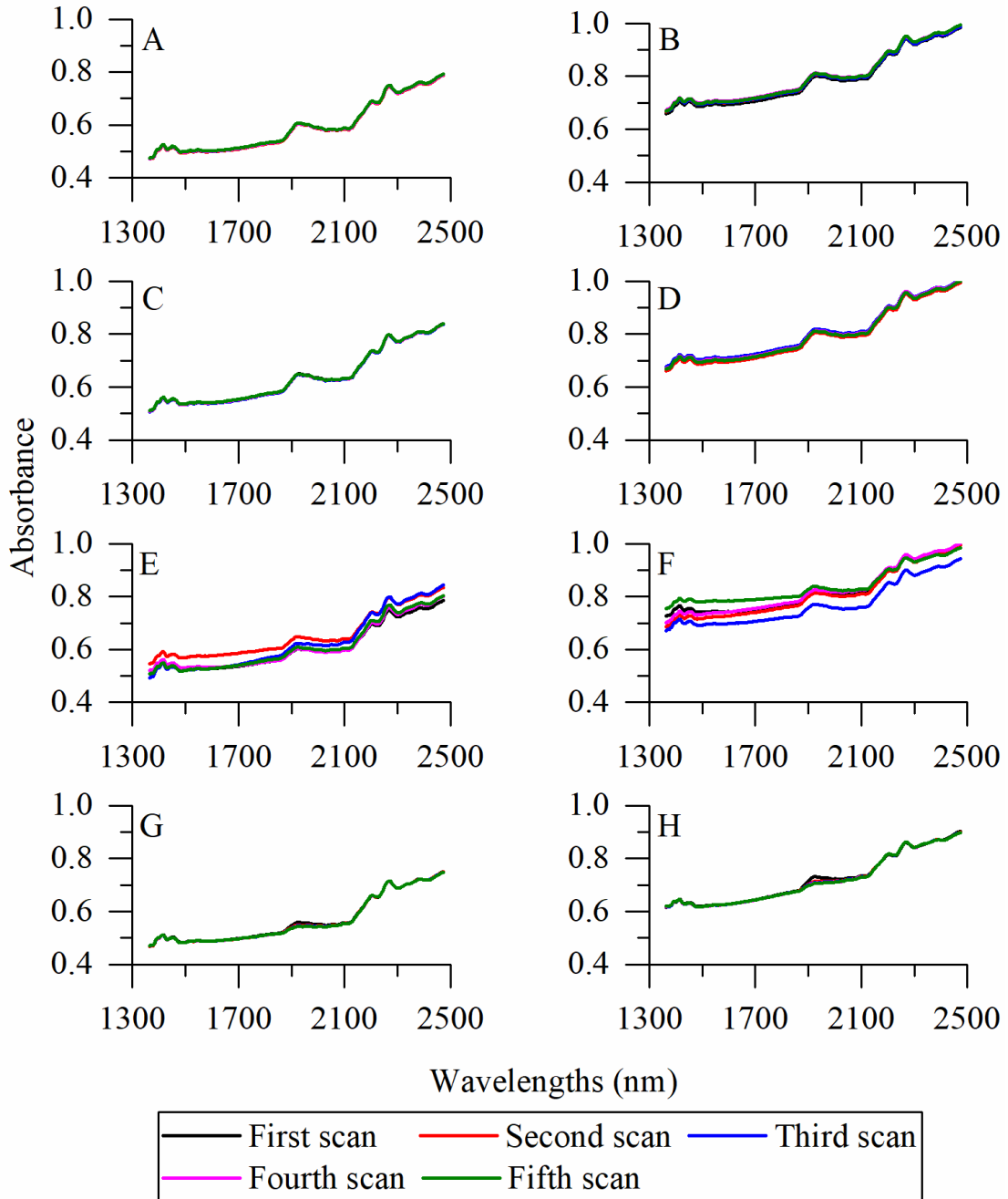


Figure 1. The raw absorbance spectra of the soil sample with low SOC content (0.65 %). The spectral data were acquired by: full radial and axial scanning (using A-full packing mode and B-2/3 packing mode); full radial scanning (using C-full packing mode and D-2/3 packing mode); scanning at variable points (using E-full packing mode and F-2/3 packing mode); scanning at a fixed point (using G-full packing mode and H-2/3 packing mode).

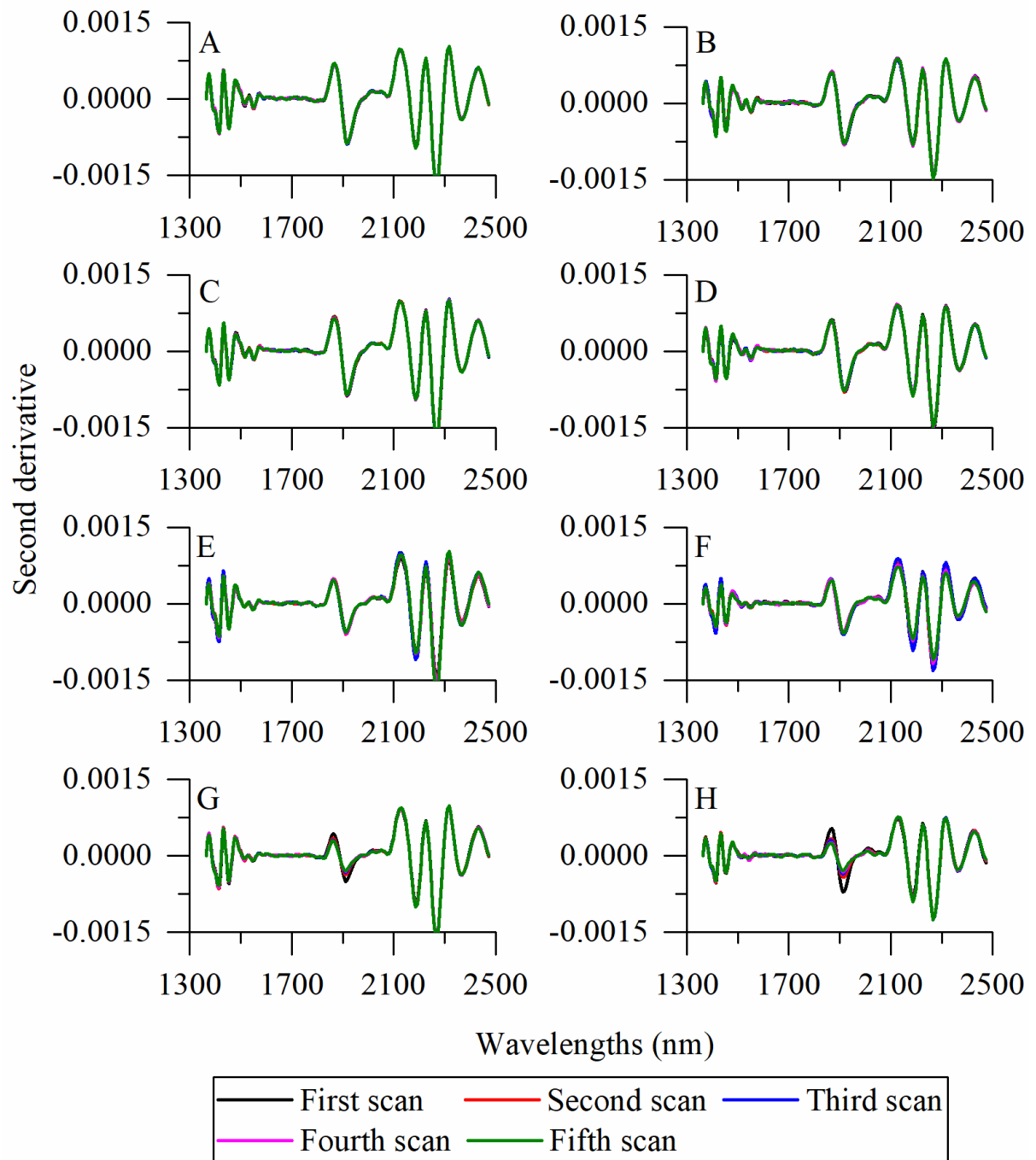


Figure 2. Second derivative spectra of the soil sample with low SOC content (0.65 %). The spectral data were acquired by: full radial and axial scanning (using A-full packing mode and B-2/3 packing mode); full radial scanning (using C-full packing mode and D-2/3 packing mode); scanning at variable points (using E-full packing mode and F-2/3 packing mode); scanning at a fixed point (using G-full packing mode and H-2/3 packing mode).

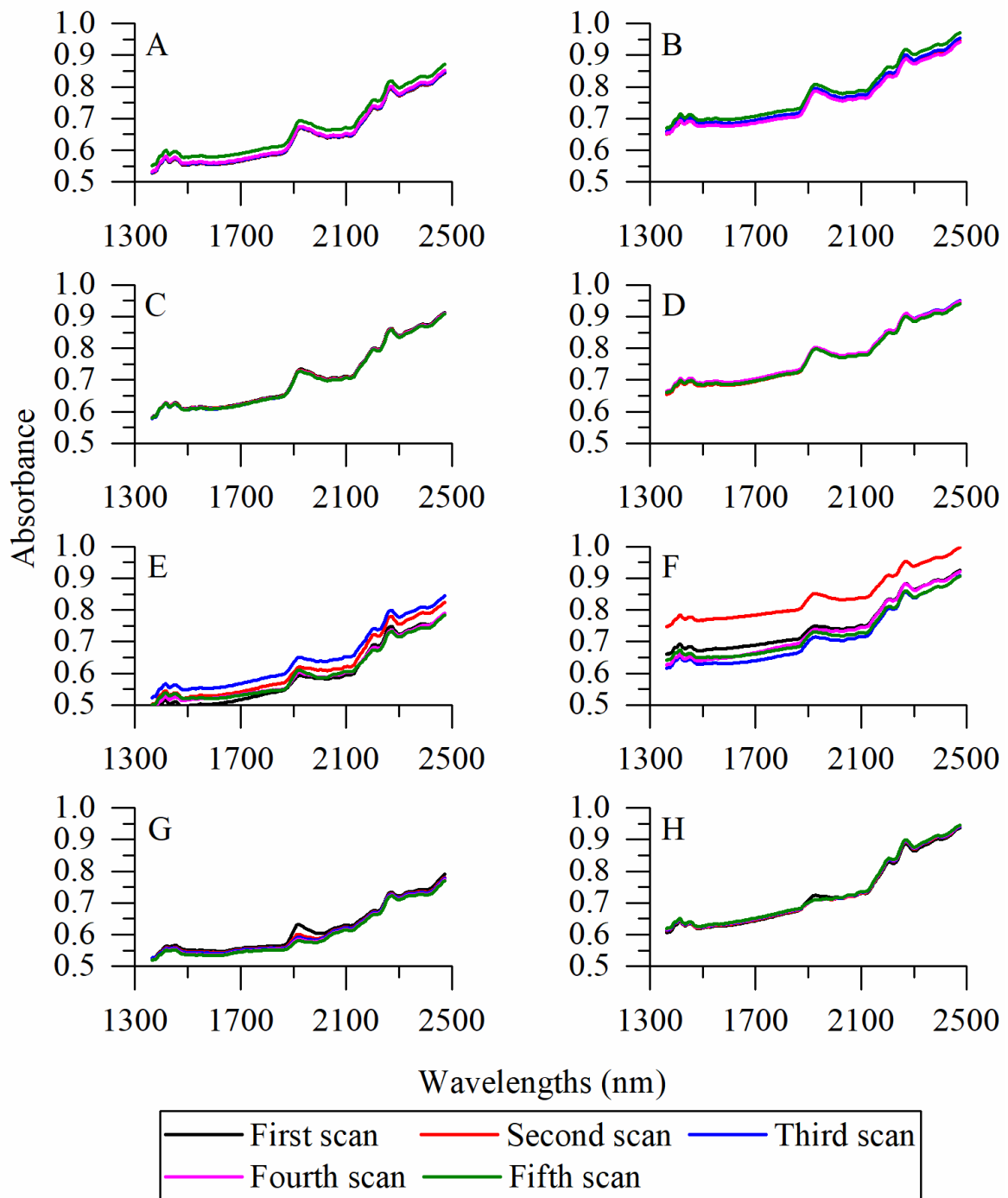


Figure 3. The raw absorbance spectra of the soil sample with medium SOC content (1.77 %). The spectral data were acquired by: full radial and axial scanning (using A-full packing mode and B-2/3 packing mode); full radial scanning (using C-full packing mode and D-2/3 packing mode); scanning at variable points (using E-full packing mode and F-2/3 packing mode); scanning at a fixed point (using G-full packing mode and H-2/3 packing mode).

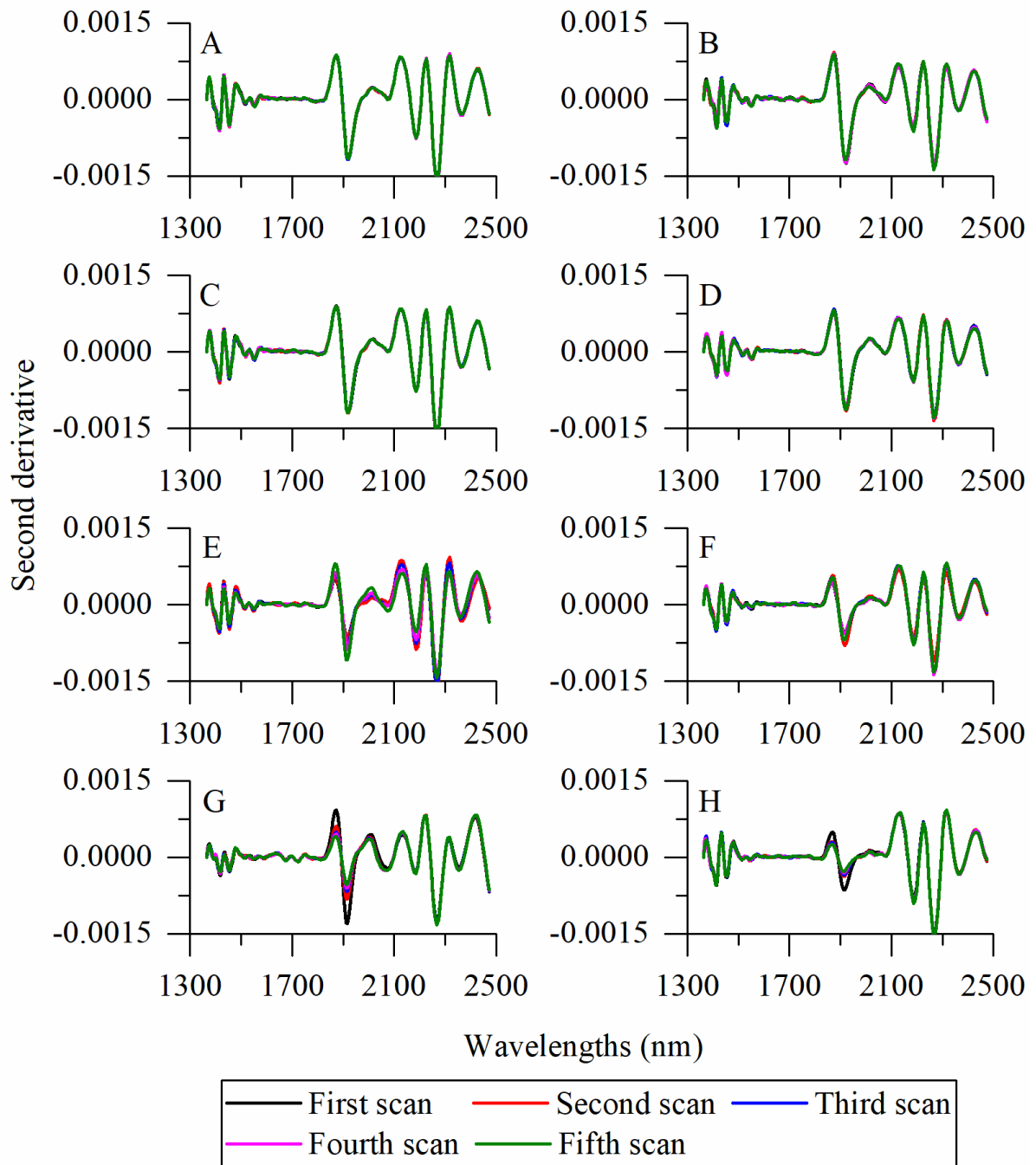


Figure 4. Second derivative spectra of the soil sample with medium SOC content (1.77 %). The spectral data were acquired by: full radial and axial scanning (using A-full packing mode and B-2/3 packing mode); full radial scanning (using C-full packing mode and D-2/3 packing mode); scanning at variable points (using E-full packing mode and F-2/3 packing mode); scanning at a fixed point (using G-full packing mode and H-2/3 packing mode).

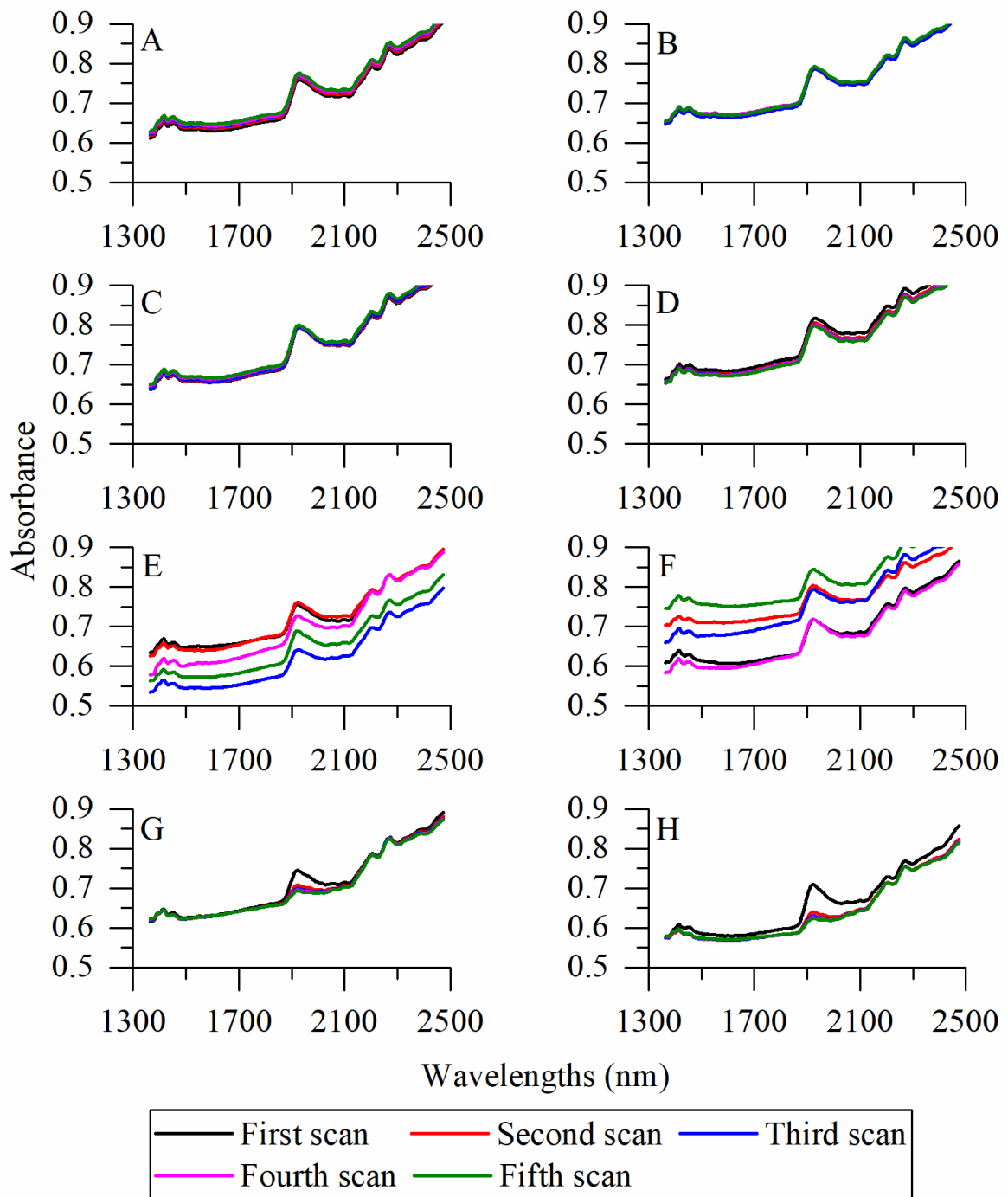


Figure 5. The raw absorbance spectra of the soil sample with high SOC content (3.46 %). The spectral data were acquired by: full radial and axial scanning (using A-full packing mode and B-2/3 packing mode); full radial scanning (using C-full packing mode and D-2/3 packing mode); scanning at variable points (using E-full packing mode and F-2/3 packing mode); scanning at a fixed point (using G-full packing mode and H-2/3 packing mode).

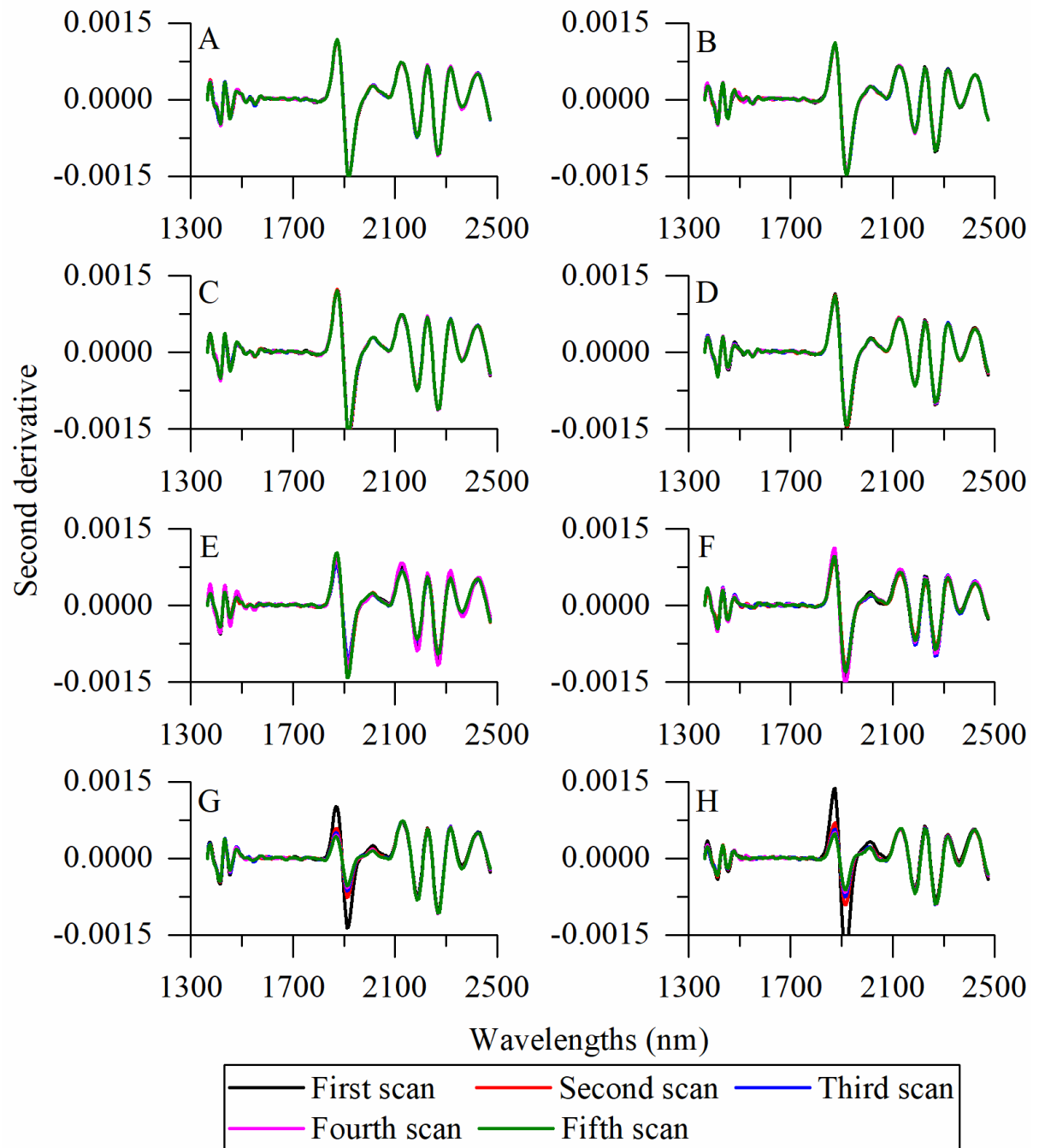


Figure 6. Second derivative spectra of the soil sample with high SOC content (3.46 %). The spectral data were acquired by: full radial and axial scanning (using A-full packing mode and B-2/3 packing mode); full radial scanning (using C-full packing mode and D-2/3 packing mode); scanning at variable points (using E-full packing mode and F-2/3 packing mode); scanning at a fixed point (using G-full packing mode and H-2/3 packing mode).

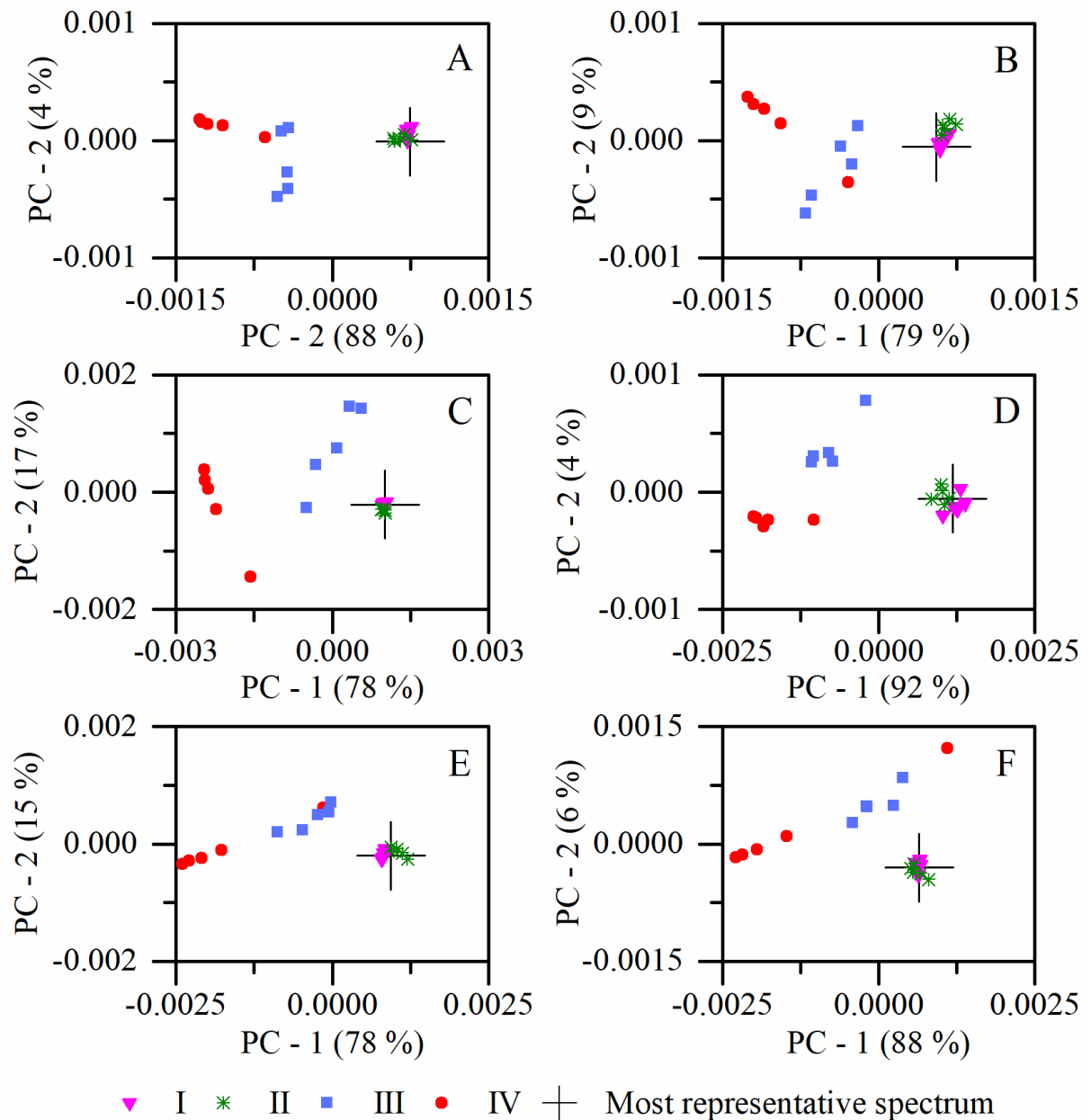


Figure 7. PCA score plots showing the distribution of the spectra acquired from soil samples with low (using A-full packing mode and B-2/3 packing), medium (using C-full packing and D-2/3 packing mode), and high (using E-full packing and F-2/3 packing mode) SOC content. These spectral data were obtained using different measurement modes: I-full radial and axial scanning, II-full radial scanning, III-scanning at variable points, IV-scanning at a fixed point. The most representative spectrum of the sample corresponds to the average of 12 spectra acquired using full radial and axial scanning. The cross-lines locate the most representative average.

The representativeness of the NIR spectra acquired using measurement modes I, II, III, and IV was also evaluated using the PCA scores (Figure 7). The spectral data acquired using the RLSPD (modes I and II) were clustered close to the most representative average spectrum of the samples (Figure 7), which indicated high similarity among them and, consequently, high representativeness. On the other hand, spectra collected using measurement modes III and IV formed two clusters distant from the most representative spectrum (Figure 7).

The PCA scores also evidenced that the sample packing form can affect the soil NIR spectra (Figure 7). According to the results shown in Figure 7, when the total volume of the flasks was filled with the soil sample (Figure 7A), the dispersion of spectra acquired using measurement modes I and II was lower (Figure 7).

3.2. Soil carbon determination by NIRS

The raw absorbance spectra of the calibration set are presented in Figure 8. After pre-processing the spectral data sets by 2nd-derivative, it was possible to visualize better the differences among the spectra acquired using measurement modes I, II, III, and IV (Figure 9). Spectra collected using the measurement mode IV (Figure 9A and 9B) presented wide fluctuations throughout the scanned NIR region (Figure 9G and 9H). The spectral variability is considerably reduced (Figure 9E and 9F) using the measurement mode III and when the RLSPD was used for data acquisition (Figure 9A, 9B, 9C and 9D).

The statistics obtained by the PLSR models constructed to determine the SOC content are presented in Table 3. The RLSPD improved the performance of NeoSpectra for SOC determination, which is yet more evidenced considering the external validation results. When measurement mode I (full axial and radial scanning) was used, the SOC prediction (RMSEP) improved by up to 146 and 185 % over the measurement modes IV and III, respectively (Table 3). Using the RLSPD programmed to move the sample flask in the radial direction at a fixed axial line (measurement mode II) also reduced the prediction error compared to the measurement modes III and IV; however, the RMSEP of this model was 54 and 21 % higher compared to measurement modes I-A and I-B, respectively (Table 3).

The models developed for SOC determination using spectral data acquired using the measurement mode III (obtained through three sample scans) presented low prediction error in cross-validation (RMSECV). However, when the models were used in independent samples, the RMSEP increased substantially (Table 3). The models developed using spectral data

acquired at a fixed point on the sample presented higher cross-validation and external validation errors (Table 3).

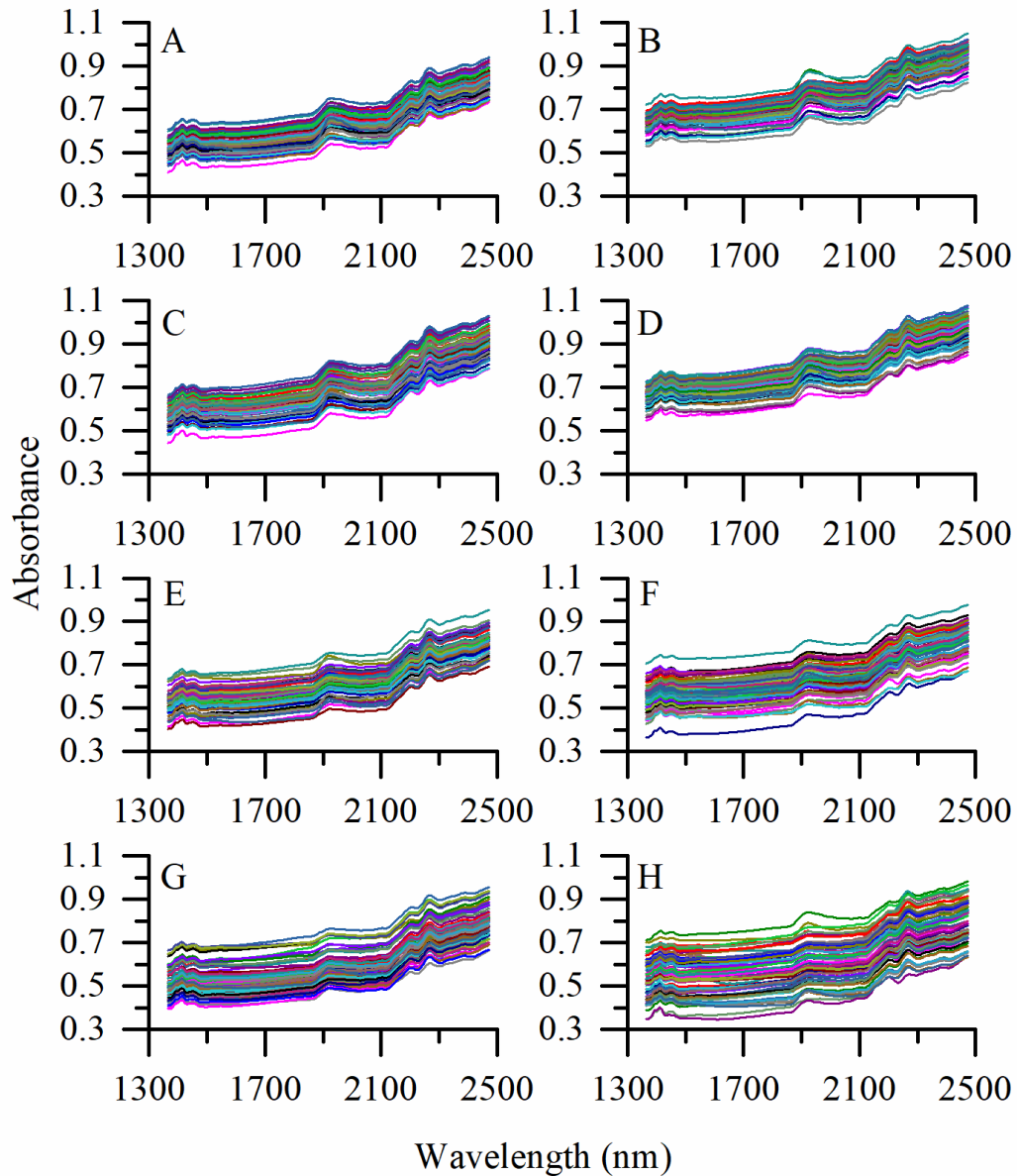


Figure 8. NIR spectra sets of soil samples after the spectral data transformation to absorbance. The spectral data were acquired by: full radial and axial scanning (using A-full packing mode and B-2/3 packing mode); full radial scanning (using C-full packing mode and D-2/3 packing mode); scanning at variable points (using E-full packing mode and F-2/3 packing mode); scanning at a fixed point (using G-full packing mode and H-2/3 packing mode).

Regardless of the measurement mode used, when the spectral data were acquired using full packing mode, a small number of latent variables were required for the PLSR models, and

the error during external validation (RMSEP) was smaller (Table 3). The models developed with data collected using the 2/3-filled flask presented acceptable RMSECV; however, when these models were used to determine SOC content in independent samples (validation set), their RMSEP increased considerably (Table 3). In general, the low predictive performance of the models generated using the 2/3-filled flask was associated with high bias values, which did not occur when the full packing was used (Table 3).

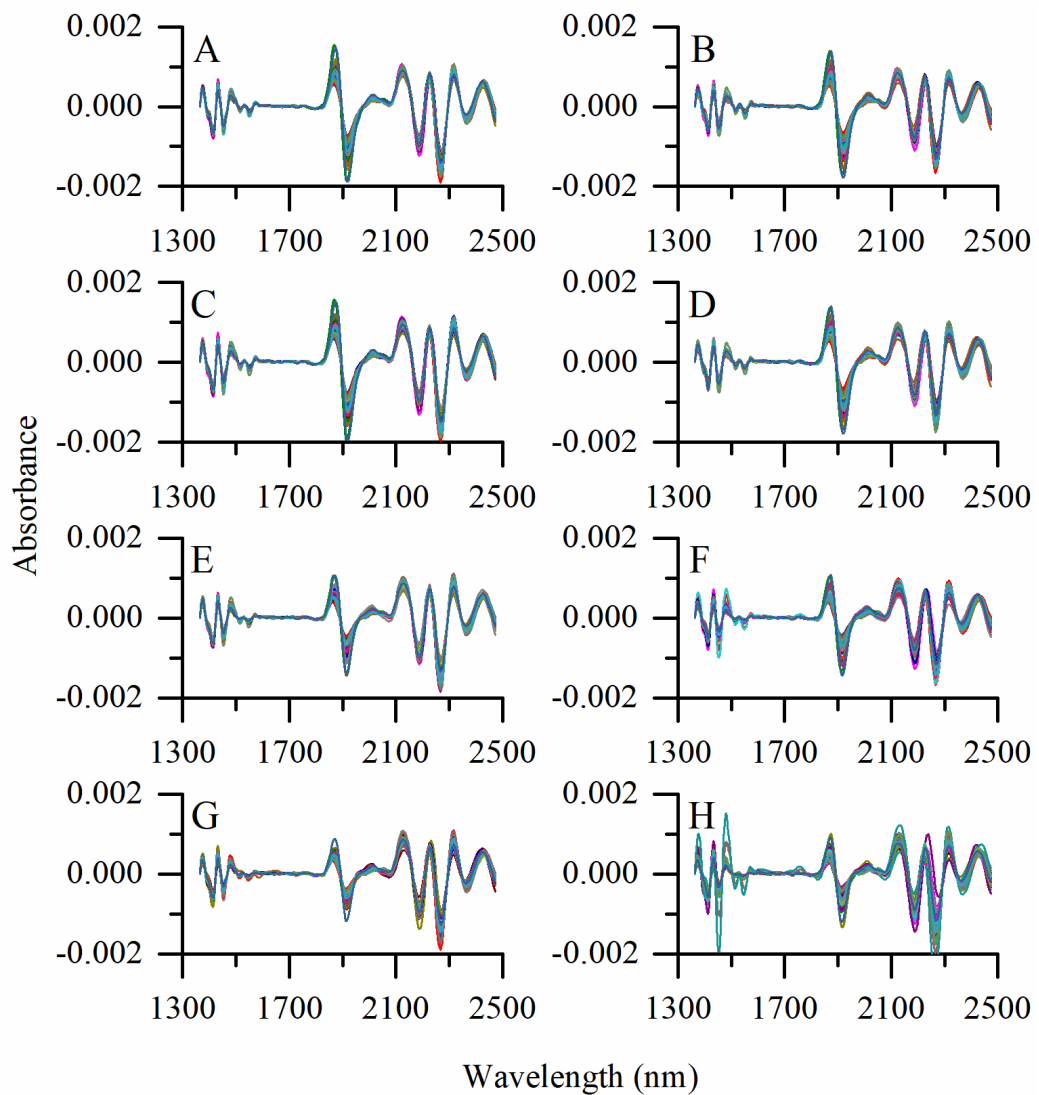


Figure 9. NIR spectra of soil samples after second derivative application using Savitzky-Golay algorithm with a window size of 15 data points and 2nd-degree polynomial. The spectral data were acquired by: full radial and axial scanning (using A-full packing mode and B-2/3 packing mode); full radial scanning (using C-full packing mode and D-2/3 packing mode); scanning at variable points (using E-full packing mode and F-2/3 packing mode); scanning at a fixed point (using G-full packing mode and H-2/3 packing mode).

Table 3- Indicators of model performance using cross-validation (calibration set) and independent validation (validation set)

Measurement mode ^a	Calibration set					Validation set			
	<i>n</i> ^b	LV ^c	Bias	RMSECV ^d	<i>R</i> ²	<i>n</i>	Bias	RMSEP ^e	<i>R</i> ²
Data acquisition using the full-packing mode									
I	48	2	0.01	0,22	0,921	20	0.03	0,13	0,947
II	48	1	-0.01	0,27	0,880	20	0.07	0,20	0,884
III	48	4	0.01	0,22	0,915	20	0.13	0,37	0,614
IV	48	3	-0.01	0,31	0,842	20	0.07	0,32	0,710
Data acquisition using the 2/3-packing mode									
I	48	5	-0.01	0,21	0,925	20	0.23	0,28	0,777
II	48	5	-0.01	0,23	0,912	20	0.32	0,34	0,663
III	48	6	-0.01	0,19	0,941	20	0.41	0,52	0,226
IV	48	5	0.06	0,46	0,645	20	-0.16	0,39	0,560

^aI, full radial and axial scanning; II, full radial scanning in a fixed axial line of the sample flask; III, measurement at three different points of the sample; IV, measurement at a fixed point of the sample. ^bOnly one outlier sample was detected and removed from the Calibration set. ^cLatent variables used by PLSR. ^dRoot means square error of cross-validation (% , m/m). ^eRoot means square error of prediction (% , m/m).

Table 4 summarizes the results obtained for the accuracy comparison between selected models using the randomization test. Models showing preferentially similar RMSEP values were compared. The p-values are low enough to reject the null hypothesis, which anticipates no significant difference among the results, with a confidence of at least 86%.

The regression coefficients of the models developed using the different measurement modes are presented in Figure 10. In general, the highest regression coefficients occurred at wavelengths near 1900 nm. This spectral region produced stable regression coefficients regardless of the measurement mode used (Figure 10).

Table 4. Results for accuracy comparison between selected models using the randomization test.

Measurement modes		RMSEP ^a		P-value
X ₁	X ₂	X ₁	X ₂	
I-A	II-A	0.13	0.20	0.145
I-A	III-A	0.13	0.37	0.050
I-A	IV-A	0.13	0.32	0.041
I-A	I-B	0.13	0.28	0.001
I-B	II-B	0.23	0.34	0.052

^aRoot mean square error of prediction (% , m/m).

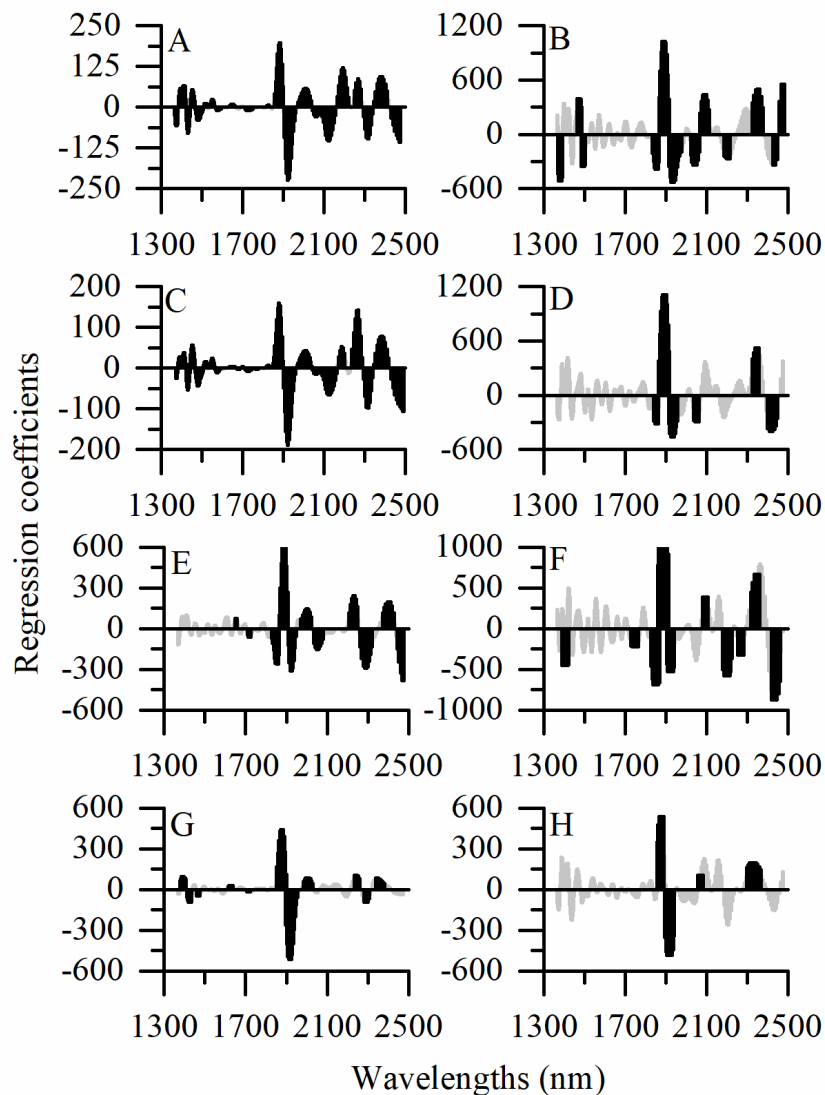


Figure 10. Regression coefficients of the PLSR models. The variables producing the more stable regression coefficients are marked in black. A-H labels as in figure 8.

The models generated using the measurement modes I-A and II-A presented very similar regression coefficients. Almost all wavelengths of the NIR spectrum provided information for SOC determination and presented high stability during the calibration process (Figures 5A and 5C). When these measurement modes were used with 2/3-packed flasks (modes I-B and II-B), however, the number of stable regression coefficients is significantly reduced (Figures 5B and 5D).

Regression coefficients of the models generated using spectral data acquired using measurement modes III and IV showed low stability during the calibration process (Figure 10). Again, the volume of soil in the flasks influenced the regression coefficients. When 2/3-packed flasks were used, more wavelengths were used to determine the SOC; however, these coefficients were not stable during the calibration process (Figure 10).

4. Discussion

The reproducibility and representativeness of the soil NIR spectra were influenced by the protocol used for sample measurement, which defines the amount of soil that interacts with NIR radiation in each measurement mode. The sample area probed by the spectrophotometer during data acquisition produces a single spectrum that represents the average composition of the particles illuminated by the NIR radiation (SILVA; PASQUINI, 2015). Thus, local variations in the probed sample area, caused, for example, by the origin, morphology, and size distribution of the particles, can affect the final spectrum (PASQUINI; HESPANHOL, 2021). This behavior is observed through the dispersion of the spectral data obtained using measurement mode III.

The quality of data employed to feed multivariate regression algorithms is relevant to increase models' robustness and accuracy (AMIGO, 2021). Frequently, works found in the literature omit or describe unsatisfactory measurement protocols unable to optimize the laboratory sample representativeness. Large, ideally global, spectral libraries are constituted by data collected under distinct measurement protocols resulting in the reduction of the quality of information. This lack of quality probably propagates influencing the performance of the regression models.

As soil is a complex and multi-component matrix with a great diversity of inorganic and organic compounds that can absorb NIR radiation (FANG et al., 2018; MILTZ; DON, 2012; STENBERG et al., 2010; VISCARRA ROSSEL; BEHRENS, 2010), there are significant variations among the spectral data measured at different points of the same soil sample. This

behavior is observed through the dispersion of the spectral data obtained using measurement mode III.

Variations in spectral features occurred even when data were acquired at a fixed point of the soil sample (measurement mode IV), but, in this case, the spectral variability was not related to sample heterogeneity. The main differences among the five spectra collected using this protocol occurred around 1940 nm, where combinations of vibrational modes of O-H groups present in water molecules occur (CLARK et al., 1990; FANG et al., 2018; HUNT, 1977; PASQUINI, 2018; TAMBURINI et al., 2017). After each measurement, the intensity of this absorption band was reduced (Figures 2, 4 and 6), which indicates a reduction in the local moisture of the soil samples (ROUDIER et al., 2017). Water molecules are removed by evaporation from the probed fixed amount of sample heated by exposition to the relative intense optical beam provided by the radiation source of the compact NIR spectrophotometer (REEVES, 1996).

Increasing the area probed during spectral data acquisition can mitigate the influence of sample heterogeneity and sample heating effects since a more extensive portion of the sample is briefly exposed to the optical beam (REEVES, 1996; SCHMUTZLER; HUCK, 2014). Thus, as expected, the use of RLSPD increased reproducibility and representativeness of the soil spectral data acquired using the compact spectrophotometer. Although it is difficult to estimate the portion of the sample illuminated by the spectrophotometer in each measurement mode, moving the samples over the optical probing beam during scanning can certainly and significantly increase the probed amount of the sample.

The surface area of the sample recipient probed when the RLSPD were used can be calculated considering the internal radius of these cylinder flasks (1 cm) and the effective height of the flasks that was exposed to the optical beam during the measurement cycle. Measurement mode I employed six revolutions and a linear simultaneous displacement of 4 cm (section 2.5). Considering the internal radius of the cylinder flasks and that the effective height of these flasks exposed to the optical beam during the linear displacements was 4 cm, the maximum surface area probed following this protocol was 25 cm². Using the full radial scanning in the absence of the linear movement (single rotation measurement modes II-A and II-B), the effective height of the flask exposed to the optical beam was considered as the diameter of the optical window of NeoSpectra (0.25 cm), so the surface area probed was 1.57 cm². When the spectral data were acquired using static modes (measurement mode III and IV), only 0.05 cm² (area of the

NeoSpectra's optical window) of the sample was probed in each measurement, equivalent to 0.20 and 3.19 % of the area probed by measurement modes I and II, respectively.

Regarding the effects of soil volume present in the flasks, the results were different than expected. Typically, the NIR radiation penetrates only a few millimeters down the surface of ground samples (BURNS; CIURCZAK, 2007). When fully-packed flasks (mode A) were used, only the most superficial soil particles were probed, once soil particles remained static inside of the flask. The radial mass transport inside the flasks provided by rotating 2/3-packed flasks (B) causes mixing of soil particles, which can increase the volume of sample probed by NIR radiation during spectra acquisition (PASQUINI; HESPANHOL, 2021). However, this measurement mode reduced the spectral reproducibility. Probably, the movement of the particles and their redistribution during the acquisition of the spectra could have affected the measurements.

Using measurement protocols that increase the representativeness of soil NIR spectra improved SOC determination employing the low-cost compact NIR spectrophotometer. PLSR models the relationship between a matrix of predictors and response variable vectors (LEONE et al., 2012; OLIVIERI, 2018). Therefore, when the predictor matrix is composed of unrepresentative spectral data, it is more difficult to establish its relationship with the SOC content (response variable).

As seen before, when spectral data were obtained from a single point of the sample flask, there is an increasing risk of including poorly representative spectra in the calibration set, which significantly affects the accuracy of the models. In addition, if the optical beam focuses, for example, on fragments of plant tissue or on an area of the sample flask in which there was an agglomeration of clay particles, the spectra features can be significantly modified (LACERDA et al., 2016; STENBERG et al., 2010; VISCARRA ROSSEL et al., 2009), which explain the occurrence of false outliers in the calibration set when the measurement mode IV was used. In this case, the outliers are not, in fact, outliers but a result of the poor representativeness of the spectral data set.

Measuring three points of the sample (mode III) to calibrate the models using average spectra improved the spectral representativeness and reproducibility. Consequently, the number of false outliers is considerably reduced. Models calibrated following this protocol presented good performance during the calibration. Nevertheless, the prediction error increased during the external validation process, which probably occurred due to the low stability of the regression coefficients. The vector of regression coefficients is formed when latent variables

are decompressed in the real space (OLIVIERI, 2018). Considering that latent variables reflect the information present in the predictor matrix, which is essential for modelling and determining the analyte (LEONE et al., 2012), the instability of the regression coefficients indicates that the spectral information related to SOC content varied during the leave-one-out cross-validation process, which makes the prediction of independent samples less accurate.

Table 3 shows that the number of PLSR factors required to optimize the models increases significantly and unnecessarily when lesser representative measurement modes are used. More complex models reflect lower robustness, as evidenced by the worse performance of the models constructed using data sets with reduced representativeness after validation.

The best results in SOC prediction were obtained using measurement protocols in which the sample was moved over the optical beam to increase the amount of sample probed, which is attributed to the use of more representative spectral data in the calibration and validation sets (REEVES, 1996; SCHMUTZLER; HUCK, 2014; SILVA; PASQUINI, 2015). The models developed using the RLSPD required few latent variables, and almost all regression coefficients showed high stability, indicating that the relationship between evaluated wavelengths and SOC content were very similar during the model training (LEONE et al., 2012; OLIVIERI, 2018). These improvements provided greater accuracy in SOC prediction when the models were applied to the independent samples of the validation set.

Although the scores for the first two components of PCA did not show significant differences between the representativeness of the soil spectra collected using measurement modes I and II, the accuracy in SOC prediction was higher when the full radial and axial scanning (mode I) was used (Table 2). This result was expected since a larger amount of sample is probed using this measurement mode. As has been extensively discussed throughout this paper, in addition to increasing the representativeness of the NIR spectra, increasing the amount of probed sample by the compact spectrophotometer reduces the occurrence of errors related, for example, to local density, particle size distribution, particles morphology, sample heterogeneity, and sample heating (PASQUINI; HESPANHOL, 2021; REEVES, 1996).

In the current literature, the RMSEP values of the models developed in order to estimate SOC content using compact NIR spectrophotometers range from 0.36 to 0.94 % (BARTHÈS et al., 2019; KARYOTIS et al., 2021; MURA et al., 2019; SHARIFIFAR et al., 2019b; SORIANO-DISLA et al., 2017). The soil samples remained static, and few points were probed to acquire the spectral data used in these models (Table 1), similarly to what was done using measurement modes III and IV in the present study. The comparison among the RMSEP values

obtained in the present work and those reported in the works in Table 1 is not straightforward because they refer to different sample sets. Nevertheless, based on the results of the present work, it is possible to anticipate the accuracy of these models would probably be improved if measurement protocols allowing the acquisition of more representative spectral data had been used, such as the measurement modes I and II.

5. Conclusion

Our results demonstrate that using a rotational-linear sample probing device is highly recommended for SOC determinations using compact NIR spectrophotometers. In short, when this device is used, a more significant portion of the soil sample is exposed to NIR radiation, which increases spectra representativeness and reproducibility, reducing the risk of false outliers caused by sample heterogeneity. These improvements reduce the error of PLSR models to predict SOC in independent samples. All these benefits were more significant when the device was programmed to simultaneously move the sample flasks in the radial and axial direction since this measurement mode increases the sample amount probed by the spectrophotometer.

Based on the results of this work, it is not recommended to use spectral data acquired at a single point of the sample to develop SOC determination models based on compact NIR spectrophotometers. Furthermore, it is necessary to be cautious even using average spectra calculated from measuring three or few distinct points of the soil samples. Although this protocol may reduce the cross-validation error, it produces more complex models increasing the error when used for the determination of SOC in independent samples.

References

AHMAD JANI, H. F. et al. Near infrared spectroscopy of plantation forest soil nutrients in Sabah, Malaysia, and the potential for microsite assessment. **Journal of Near Infrared Spectroscopy**, v. 29, n. 3, p. 148–157, 17 jun. 2021.

AMIGO, J. M. Data Mining, Machine Learning, Deep Learning, Chemometrics. Definitions, common points and Trends (Spoiler Alert: VALIDATE your models!). **Brazilian Journal of Analytical Chemistry**, v. 8, n. 32, p. 45–61, 14 maio 2021.

BARTHÈS, B. G. et al. Performance comparison between a miniaturized and a conventional near infrared reflectance (NIR) spectrometer for characterizing soil carbon and nitrogen. **Geoderma**, v. 338, n. February 2018, p. 422–429, mar. 2019.

BEĆ, K. B. et al. Handheld near-infrared spectrometers: Where are we heading? **NIR news**, v. 31, n. 3–4, p. 28–35, 2020.

BEĆ, K. B.; GRABSKA, J.; HUCK, C. W. Principles and Applications of Miniaturized Near-Infrared (NIR) Spectrometers. **Chemistry - A European Journal**, v. 27, n. 5, p. 1514–1532, 2021.

BELLON-MAUREL, V. et al. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. **TrAC - Trends in Analytical Chemistry**, v. 29, n. 9, p. 1073–1081, 2010.

BELTRAME, K. K. et al. Soil organic carbon determination using NIRS: Evaluation of dichromate oxidation and dry combustion analysis as reference methods in multivariate calibration. **Journal of the Brazilian Chemical Society**, v. 27, n. 9, p. 1527–1532, 2016.

BÜNEMANN, E. K. et al. Soil quality – A critical review. **Soil Biology and Biochemistry**, v. 120, n. September 2017, p. 105–125, 2018.

BURNS, D. A.; CIURCZAK, E. W. **Handbook of Near-Infrared Analysis**. 3rd Edition. Boca Raton: CRC Press, 2007.

CLARK, R. N. et al. High spectral resolution reflectance spectroscopy of minerals. **Journal of Geophysical Research**, v. 95, n. B8, p. 12,653-12,680, 1990.

COSTA, D. C. **Efeitos residuais da fertilização orgânica e mineral aplicadas na recuperação de solo de pastagem minerado para bauxita**. [s.l.] Universidade Federal de Viçosa, 2020.

CROCOMBE, R. A. Portable Spectroscopy. **Applied Spectroscopy**, v. 72, n. 12, p. 1701–1751, 2018.

DEMATTE, J. A. M. et al. Soil analytical quality control by traditional and spectroscopy techniques: Constructing the future of a hybrid laboratory for low environmental impact. **Geoderma**, v. 337, p. 111–121, mar. 2019.

FANG, Q. et al. Visible and Near-Infrared Reflectance Spectroscopy for Investigating Soil Mineralogy: A Review. **Journal of Spectroscopy**, v. 2018, p. 1–14, 2018.

GUERRERO, C.; ROSSEL, R. A. V.; MOUAZEN, A. M. Special issue “Diffuse reflectance spectroscopy in soil science and land resource assessment.” **Geoderma**, v. 158, n. 1–2, p. 1–2, 2010.

HUNT, G. R. Spectral signatures of particulate minerals in the visible and near infrared. **Geophysics**, v. 42, n. 3, p. 501–513, 1977.

IUSS WORKING GROUP WRB. World reference base for soil resources 2014. International soil classification system for naming soils and creating legends for soil maps. In: **World Soil Resources Reports No. 106**. [s.l: s.n.]. p. 192.

JACONI, A. et al. Log-ratio transformation is the key to determining soil organic carbon fractions with near-infrared spectroscopy. **European Journal of Soil Science**, v. 70, n. 1, p. 127–139, 2019.

KARYOTIS, K. et al. Evaluation of a Micro-Electro Mechanical Systems Spectral Sensor for Soil Properties Estimation. **Land**, v. 10, n. 1, p. 63, 13 jan. 2021.

LACERDA, M. P. C. et al. Tropical texture determination by Proximal Sensing using a regional spectral library and its relationship with soil classification. **Remote Sensing**, v. 8, n. 9, p. 1–20, 2016.

LEONE, A. P. et al. Prediction of Soil Properties with PLSR and vis-NIR Spectroscopy: Application to Mediterranean Soils from Southern Italy. **Current Analytical Chemistry**, v. 8, n. 2, p. 283–299, 1 abr. 2012.

LIN, Z. D. et al. Improvements of the Vis-NIRS Model in the Prediction of Soil Organic Matter Content Using Spectral Pretreatments, Sample Selection, and Wavelength Optimization. **Journal of Applied Spectroscopy**, v. 84, n. 3, p. 529–534, 2017.

MILTZ, J.; DON, A. Optimising sample preparation and near infrared spectra measurements of soil samples to calibrate organic carbon and total nitrogen content. **Journal of Near Infrared Spectroscopy**, v. 20, n. 6, p. 695–706, 2012.

MINASNY, B. et al. Soil carbon 4 per mille. **Geoderma**, v. 292, p. 59–86, 2017.

MURA, S. et al. Vibrational spectroscopy and Aquaphotomics holistic approach to determine chemical compounds related to sustainability in soil profiles. **Computers and Electronics in Agriculture**, v. 159, n. February, p. 92–96, 2019.

OLIVIERI, A. C. **Introduction to Multivariate Calibration**. Cham: Springer International Publishing, 2018.

PASQUINI, C. Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications. **Journal of the Brazilian Chemical Society**, v. 14, n. 2, p. 198–219, 2003.

PASQUINI, C. Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. **Analytica Chimica Acta**, v. 1026, p. 8–36, out. 2018.

PASQUINI, C.; HESPANHOL, M. C. A rotational-linear sample probing device to improve the performance of compact near-infrared spectrophotometers. **Microchemical Journal**, v. 170, p. 106747, nov. 2021.

REEVES, J. B. Improvement in Fourier Near- and Mid-Infrared Diffuse Reflectance Spectroscopic Calibrations through the Use of a Sample Transport Device. **Applied Spectroscopy**, v. 50, n. 8, p. 965–969, 31 ago. 1996.

ROUDIER, P. et al. Evaluation of two methods to eliminate the effect of water from soil vis–NIR spectra for predictions of organic carbon. **Geoderma**, v. 296, p. 98–107, 2017.

SANDERMAN, J.; HENGL, T.; FISKE, G. J. Soil carbon debt of 12,000 years of human land use. **Proceedings of the National Academy of Sciences of the United States of America**, v. 114, n. 36, p. 9575–9580, 2017.

SANTANA, F. B.; SOUZA, A. M.; POPPI, R. J. Green methodology for soil organic matter analysis using a national near infrared spectral library in tandem with learning machine. **Science of the Total Environment**, v. 658, p. 895–900, 2019.

SANTOS, C. A. T. DOS et al. A Review on the Applications of Portable Near-Infrared Spectrometers in the Agro-Food Industry. **Applied Spectroscopy**, v. 67, n. 11, p. 1215–1233, 1 nov. 2013.

SAVITZKY, A.; GOLAY, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. **Analytical Chemistry**, v. 36, n. 8, p. 1627–1639, jul. 1964.

SCHMUTZLER, M.; HUCK, C. W. Automatic sample rotation for simultaneous determination of geographical origin and quality characteristics of apples based on near infrared spectroscopy (NIRS). **Vibrational Spectroscopy**, v. 72, p. 97–104, 2014.

SHARIFIFAR, A. et al. Evaluating a low-cost portable NIR spectrometer for the prediction of soil organic and total carbon using different calibration models. **Soil Use and Management**, v. 35, n. 4, p. 607–616, dez. 2019.

SILVA, C. J.; PASQUINI, C. Comparing near-infrared conventional diffuse reflectance spectroscopy and hyperspectral imaging for determination of the bulk properties of solid samples by multivariate regression: determination of Mooney viscosity and plasticity indices of natural rubber. **The Analyst**, v. 140, n. 2, p. 512–522, 2015.

SORIANO-DISLA, J. M. et al. Evaluation of the performance of portable visible-infrared instruments for the prediction of soil properties. **Biosystems Engineering**, v. 161, p. 24–36, 2017.

STENBERG, B. et al. Visible and Near Infrared Spectroscopy in Soil Science. In: SPARKS, D. L. (Ed.). . **Advances in Agronomy**. Burlington: Academic Press, 2010. v. 107p. 163–215.

SUT, M. et al. Feasibility of Field Portable Near Infrared (NIR) Spectroscopy to Determine Cyanide Concentrations in Soil. **Water, Air, & Soil Pollution**, v. 223, n. 8, p. 5495–5504, 5 out. 2012.

TAMBURINI, E. et al. Effects of Moisture and Particle Size on Quantitative Determination of Total Organic Carbon (TOC) in Soils Using Near-Infrared Spectroscopy. **Sensors**, v. 17, n. 10, p. 2366, 17 out. 2017.

THOMSEN, I. K. et al. Assessing soil carbon lability by near infrared spectroscopy and NaOCl oxidation. **Soil Biology and Biochemistry**, v. 41, n. 10, p. 2170–2177, 2009.

VISCARRA ROSSEL, R. A. et al. In situ measurements of soil colour, mineral composition and clay content by vis-NIR spectroscopy. **Geoderma**, v. 150, n. 3–4, p. 253–266, 2009.

VISCARRA ROSSEL, R. A.; BEHRENS, T. Using data mining to model and interpret soil diffuse reflectance spectra. **Geoderma**, v. 158, n. 1–2, p. 46–54, 2010.

VOET, H. VAN DER. Comparing the predictive accuracy of models using a simple randomization test. **Chemometrics and Intelligent Laboratory Systems**, v. 25, n. 2, p. 313–323, nov. 1994.

WESTAD, F.; MARTENS, H. Variable Selection in near Infrared Spectroscopy Based on Significance Testing in Partial Least Squares Regression. **Journal of Near Infrared Spectroscopy**, v. 8, n. 2, p. 117–124, 2 mar. 2000.

YEOMANS, J. C.; BREMNER, J. M. A rapid and precise method for routine determination of organic carbon in soil. **Communications in Soil Science and Plant Analysis**, v. 19, n. 13, p. 1467–1476, 11 out. 1988.

Capítulo 2

Performance de espectrofotômetros NIR compactos na determinação de carbono orgânico do solo em escala regional: efeitos do pré-processamento das amostras e uso de modelos locais

Resumo

A espectroscopia na região do infravermelho-próximo – NIR (*near-infrared*) pode ser utilizada para quantificar os teores de carbono orgânico do solo (COS) de forma rápida, não destrutiva e sem gerar resíduos nocivos ao ambiente. Atualmente, diferentes modelos de espectrofotômetros NIR compactos e de baixo custo estão disponíveis, o que pode impulsionar o uso da NIRS nos laboratórios de análise de solo. Para isto, é necessário entender os aspectos metodológicos que envolvem o uso destes equipamentos. O objetivo deste estudo foi avaliar o desempenho de dois espectrofotômetros NIR compactos (NeoSpectra e NanoNIR) na determinação dos teores de COS em escala regional, assim como estabelecer o melhor protocolo para o uso destes equipamentos nesta condição. Foram utilizadas 559 amostras de solo. Os espectros foram adquiridos usando amostras nas seguintes condições: solo seco ao ar e peneirado a 2 mm (Modo de preparo I); solo seco ao ar e macerado em moinho de bola (Modo de preparo II); solo macerado e seco em estufa a 50°C por 48 horas (Modo de preparo III). A performance de modelos globais, desenvolvidos utilizando todo o conjunto de dados, e de modelos locais, desenvolvidos para cada amostra do conjunto de validação, também foi comparada. Os modelos foram desenvolvidos usando a regressão por mínimos quadrados parciais (PLSR) e sua performance avaliada pela raiz quadrada do erro quadrático médio (RMSEP). O NeoSpectra foi o espectrofotômetro compacto que apresentou melhor performance na determinação de COS. O RMSEP obtido com o uso deste equipamento variou entre 5,1 – 6,3 g kg⁻¹, acurácia semelhante à obtida com o uso dos espectrofotômetros convencionais. O uso de amostras maceradas proporcionou redução de até 8 % no RMSEP dos modelos desenvolvidos com espectrofotômetros compactos. A calibração local também melhorou a performance destes instrumentos. Os valores de RMSEP obtidos com o uso de modelos locais foram até 15 % menores do que com o uso de modelos globais. O melhor desempenho dos espectrofotômetros compactos foi obtido usando a calibração local e amostras finamente moídas e secas em estufa.

Palavras-chave: Espectroscopia no infravermelho próximo; matéria orgânica do solo; comparação da performance de espectrofotômetros NIR; comparação entre modelos globais e locais; PLSR-local; reflectância difusa.

1. Introdução

Maximizar a produção agrícola/florestal e, simultaneamente, reduzir as emissões de carbono (C) para a atmosfera é um dos principais desafios do século (MINASNY et al., 2017). Para isto, é necessário um amplo conhecimento a respeito da dinâmica espacial e temporal do carbono orgânico do solo (COS), o que, por sua vez, requer a caracterização extensiva deste atributo. Neste sentido, cresce a importância do desenvolvimento de métodos analíticos capazes de mensurar os teores de COS de forma rápida, barata, sem gerar resíduos úmidos e que possam ser utilizados em aplicações de larga escala, tais como em laboratórios de rotina e na agricultura de precisão.

Globalmente, mais de 75 % dos laboratórios de análise de solo utilizam a oxidação via úmida proposta por Walkley-Black para determinar os teores de COS (VISCARRA ROSSEL et al., 2016). Este método consiste na oxidação do COS presente na amostra por meio da adição de solução ácida contendo dicromato de potássio ($0,167 \text{ mol L}^{-1}$) e ácido sulfúrico concentrado (WALKLEY; BLACK, 1934). Embora não demande de equipamentos caros, esta determinação é laboriosa, demorada e gera resíduos ambientalmente tóxicos (SANTANA; SOUZA; POPPI, 2019). Considerando que aproximadamente 150 mL de resíduos úmidos são gerados a cada amostra de solo analisada e que mais de 600 milhões de amostras são analisadas no mundo todos os anos (DEMATTE et al., 2019b), $67.500.000 \text{ L ano}^{-1}$ de resíduos tóxicos são gerados pelos laboratórios de rotina para determinar os teores de COS. Deste total, $450.000 \text{ L ano}^{-1}$ são gerados apenas no Brasil (SANTANA; SOUZA; POPPI, 2019). Com a expansão do uso da agricultura de precisão, onde o monitoramento espacial e temporal dos atributos do solo é feito de maneira mais frequente, o volume de resíduos gerados por este método irá, provavelmente, aumentar nos próximos anos caso a oxidação via úmida continue sendo extensamente utilizada.

Nos últimos anos, técnicas espectroscópicas por reflectância difusa, principalmente na região do infravermelho-próximo – NIR (*near-infrared*) e do visível (Vis-NIR), têm demonstrado grande potencial para substituir/complementar o método de determinação de COS nos laboratórios de análise de solo (AHMADI et al., 2021; BELTRAME et al., 2016; DEMATTE et al., 2019a; MURA et al., 2019; STENBERG et al., 2010; VISCARRA ROSSEL

et al., 2016). Esta técnica espectroscópica é rápida, não destrutiva, altamente reprodutível, não gera resíduos nocivos ao ambiente, possui baixo custo operacional e tem apresentado bons resultados até mesmo quando utilizada em conjuntos de solos amplos e heterogêneos (DEMATTE et al., 2019a; JACONI; DON; FREIBAUER, 2017; VISCARRA ROSSEL et al., 2016). Contudo, o valor de aquisição dos espectrofotômetros Vis-NIR ainda é alto, o que pode limitar a adoção desta técnica em pequenos laboratórios de análise de solo ou grupos de pesquisa.

Com os avanços no desenvolvimento de instrumentos e computadores miniaturizados, diversos tipos de espectrofotômetros NIR compactos estão sendo produzidos e podem ser encontrados no mercado (BEĆ et al., 2020; PASQUINI, 2018). Estes instrumentos são extremamente pequenos e leves, mas a portabilidade não é a sua única vantagem (PASQUINI, 2018). O baixo valor de aquisição dos espectrofotômetros NIR compactos é um importante diferencial para a escolha destes equipamentos, ainda que para trabalhar em condições de laboratório (BARTHÈS et al., 2019). Na Ciência do Solo, o uso de espectrofotômetros NIR compactos ainda é recente, mas promissor. Embora estes instrumentos apresentem menor capacidade de sondar a amostra, menor resolução espectral e geralmente forneçam informações sobre uma faixa menor do espectro NIR (BEĆ et al., 2020; PASQUINI, 2018; SANTOS et al., 2013), a literatura atual demonstra que é possível utilizar estes espectrofotômetros para desenvolver modelos acurados para determinação dos teores de COS (AHMAD JANI et al., 2021; BARTHÈS et al., 2019; FONSECA et al., 2022; MURA et al., 2019). Contudo, até o presente momento, a performance dos espectrofotômetros NIR compactos na determinação de COS ainda não foi avaliada em conjuntos de dados amplos e com grande heterogeneidade de solos, tal como ocorre nos laboratórios de rotina.

Desenvolver modelos de determinação de COS em larga-escala, utilizando solos com diferentes texturas e mineralogias, formados a partir de diferentes materiais de origem e sob diferentes tipos de vegetação, pode ser extremamente difícil, dado que solo é uma matriz complexa e apresenta grande diversidade de compostos orgânicos e inorgânicos que podem absorver a radiação NIR (FANG et al., 2018; MILTZ; DON, 2012; STENBERG et al., 2010; VISCARRA ROSSEL et al., 2016; VISCARRA ROSSEL; BEHRENS, 2010). Estas variabilidades aumentam juntamente com a escala do conjunto de dados e tornam mais complexa a modelagem das informações espectrais relacionadas ao COS (LIU et al., 2019). Nestas condições, o pré-processamento das amostras, feito por meio da moagem e secagem do solo, pode melhorar o desempenho dos espectrofotômetros na predição de COS. Embora os

efeitos do preparo da amostra de solo, feito antes da aquisição dos espectros, no desempenho de modelos desenvolvidos com o uso de espectrofotômetros convencionais sejam divergentes (BARTHÈS et al., 2006; MILTZ; DON, 2012; TAMBURINI et al., 2017), os espectrofotômetros NIR compactos apresentam menor capacidade de sondar as amostras (FONSECA et al., 2022; PASQUINI; HESPANHOL, 2021) e menor resolução espectral (BÉC; GRABSKA; HUCK, 2021) do que espectrofotômetros convencionais, portanto, é plausível que a redução da heterogeneidade das partículas do solo e de variabilidades espectrais não informativas seja mais importante para a acurácia dos modelos desenvolvidos com estes instrumentos.

O uso de modelos locais, em substituição aos modelos “globais”, que são desenvolvidos utilizando todas as amostras do conjunto de calibração, é outra estratégia que pode melhorar o desempenho dos espectrofotômetros na determinação de COS em conjunto de dados amplos e heterogêneos. Os modelos locais são construídos utilizando apenas as amostras do conjunto de calibração consideradas semelhantes à amostra que se deseja avaliar, para isto, diferentes métodos podem ser utilizados, entre eles destacam-se: a estratificação em função de atributos do solo (JACONI; DON; FREIBAUER, 2017), estratificação em função do tipo de uso do solo (MOURA-BUENO et al., 2019) e estratificação em função das feições espectrais (LIU et al., 2019). Por não demandar informações analíticas ou de conhecimento prévio a respeito das amostras, a seleção por similaridade espectral pode ser o método mais prático a ser implementado nos laboratórios de análise de solo. No presente trabalho, ao invés da estratificação do conjunto de calibração, foram utilizados modelos de regressão locais, desenvolvidos para cada amostra do conjunto de validação utilizando apenas as amostras “vizinhas” (que atendiam ao critério de similaridade espectral utilizada) a ela.

Os objetivos do presente trabalho foram (i) avaliar o desempenho de dois espectrofotômetros NIR compactos, que trabalham com diferentes faixas do espectro NIR, na determinação dos teores de COS em escala regional; (ii) avaliar os efeitos do preparo das amostras de solo na performance dos modelos de determinação de COS desenvolvidos usando NIR compactos; e (iii) comparar a acurácia dos modelos regionais “globais” e dos modelos locais desenvolvidos usando os espectrofotômetros NIR compactos.

2. Materiais e métodos

2.1. Amostras de solo

O conjunto de amostras de solo avaliado neste estudo foi obtido no Banco de Solos de Minas Gerais (DE SOUZA et al., 2015). Ao todo, 559 amostras (0-20 cm) foram coletadas em todo o estado de Minas Gerais, abrangendo uma área aproximada de 589.000 km² e contemplando uma grande diversidade de classes de solo (Figura 1). Após terem sido coletadas, as amostras de solo foram secas ao ar, passadas em peneira de 2 mm e submetidas a caracterização física e química (Tabela 1).

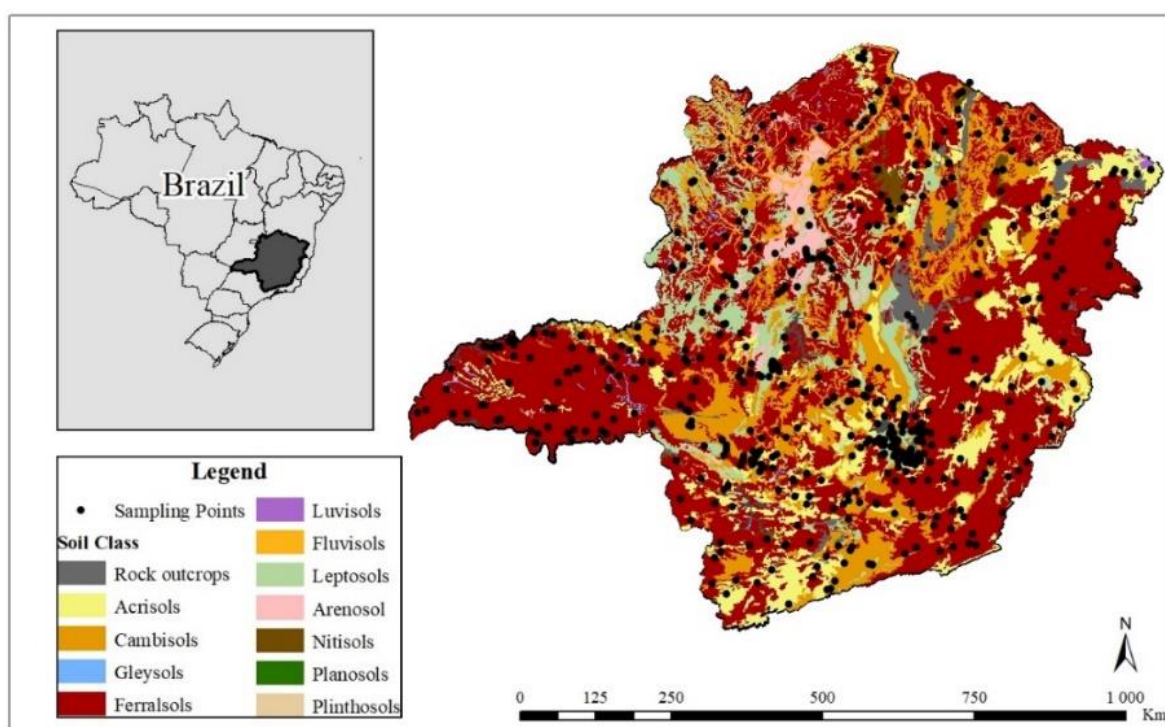


Figura 1. Localização da área em estudo, classes de solo^a e pontos de coleta das amostras de solo. ^a(UFV et al., 2010).

2.2. Determinação dos teores de carbono

Os teores de carbono foram determinados por oxidação via úmida (WALKLEY; BLACK, 1934), para isto, subamostras de solo (< 2 mm) foram maceradas em moinho de bola e passadas em peneira de 0.2 mm. Os resultados das análises laboratoriais foram disponibilizados pelo Banco de Solos de Minas Gerais (DE SOUZA et al., 2015). Os teores de carbono variaram de 0.6-195.9 g kg⁻¹ (Figura 2A).

Tabela 1. Atributos físicos e químicos das amostras de solo^a

	Mínimo	Média	Máximo	SD
Areia (g kg ⁻¹) ^b	10.0	394.8	950.0	243.9
Silte (g kg ⁻¹) ^b	0.0	204.5	770.0	153.0
Argila (g kg ⁻¹) ^b	10.0	400.7	870.0	206.2
pH H ₂ O (1:2,5) ^b	3.9	5.3	8.3	0.7
Fósforo (mg dm ⁻³) ^b	0.0	3.9	98.5	8.3
Potássio (mg dm ⁻³) ^b	0.0	74.4	602.0	66.6
Cálcio (mmolc dm ⁻³) ^b	0.0	17.1	219.0	27.0
Magnésio (mmolc dm ⁻³) ^b	0.0	6.3	87.0	8.7
CTC _{total} (mmolc dm ⁻³) ^b	1.5	85.1	343.0	39.7
CTC _{efetiva} (mmolc dm ⁻³) ^b	0.4	32.0	249.2	31.8
Ferro _{total} (mg kg ⁻¹) ^c	501.3	45571.6	210647.3	34375.0
Manganês _{total} (mg kg ⁻¹) ^c	1.6	433.3	15415.9	1025.5
Alumínio _{total} (mg kg ⁻¹) ^c	842.7	56083.5	248018.4	42287.3

^aValores cedidos pelo Banco de Solos de Minas Gerais (DE SOUZA et al., 2015).^bEmbrapa (1997); ^cUSEPA (2007).

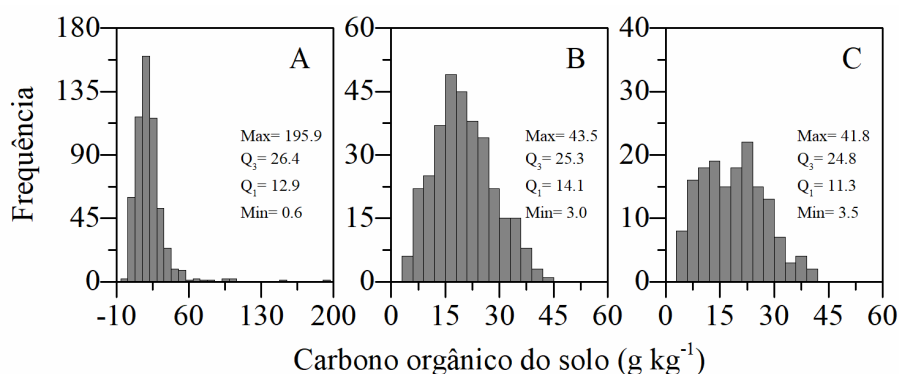


Figura 1. Teores de carbono orgânico do solo (COS) determinados por oxidação via úmida e sua frequência em todo o conjunto de dados (A) e nos conjuntos de calibração (B) e validação (C) separadamente.

2.3. Espectros NIR

2.3.1. Preparo das amostras de solo antes da aquisição dos espectros

Os espectros das 559 amostras foram adquiridos usando diferentes tipos de preparo da amostra de solo. O modo de preparo I consistiu na aquisição dos espectros usando amostras de solo secas ao ar e passadas em peneira de 2 mm. Após a coleta deste grupo de espectros, as

amostras foram maceradas em moinho de bola por 5 min e então os seus espectros foram novamente medidos (modo de preparo II). Por fim, as amostras maceradas foram secas em estufa de circulação forçada de ar (50 °C por 48 horas) e os seus espectros foram novamente medidos (modo de preparo III).

2.3.2. Instrumentação e escaneamento das amostras

Foram avaliados dois tipos de espectrofotômetros NIR compactos. Primeiramente, os espectros foram coletados com um instrumento Micro-FT NeoSpectra (NeoSpectra), produzido pela Si-Ware Systems (Si-Ware, Egypt). Este espectrofotômetro opera na região entre 1350-2500 nm com resolução de 16 nm (em 1550 nm). Os espectros finais eram compostos por 257 pontos (comprimentos de onda) e foram obtidos usando um tempo de integração de 20 s, o que proporciona a aquisição de milhares de espectros. Para aumentar a representatividade dos espectros finais, foi utilizado um sistema automatizado para rotacionar os frascos contendo as amostras de solo sobre o feixe de sondagem ótica durante a medição dos espectros (FONSECA et al., 2022; PASQUINI; HESPANHOL, 2021). Em um ciclo de medição completo (20 s), os frascos completamente preenchidos com amostras de solo sofreram seis rotações. Uma descrição completa do funcionamento do sistema automático de movimentação de amostras, e dos seus componentes, pode ser encontrada em PASQUINI; HESPANHOL (2021).

O segundo espectrofotômetro utilizado foi um NIRscan Nano (NanoNIR), produzido pela Texas Instruments (Texas Instruments, USA). Este instrumento opera entre 900-1700 nm, com resolução espectral de 10 nm. Os espectros finais foram calculados a partir de 99 scans (21 s). O NanoNIR também foi acoplado ao sistema de movimentação de amostras para aumentar a representatividade dos espectros adquiridos, sendo que a programação do dispositivo foi a mesma utilizada para o NeoSpectra.

Os espectros adquiridos no NeoSpectra e NanoNIR foram medidos em absorbância ($\log [1/\text{reflectância}]$). Como referência de 100 % de reflectância, a cada 10 amostras, um frasco completamente preenchido com Spectralon® foi escaneado. A performance dos espectrofotômetros NIR compactos na determinação de COS foi comparada com a performance de espectrofotômetros NIR convencionais. Para isto, também foram coletados espectros das amostras de solo utilizando espectrofotômetros convencionais de alta resolução (Tabela 2).

2.3.3. Pré-processamento dos espectros

Foram testados diferentes tipos de pré-tratamentos para remover fontes de variabilidade não informativas presentes nos espectros NIR, tais como espalhamento da radiação e deslocamentos na linha de base. O pré-tratamento que resultou em modelos com menor erro foi

a primeira derivada usando o filtro de Savitzky–Golay (SAVITZKY; GOLAY, 1964). Para os espectros obtidos usando o NeoSpectra e o NanoNIR, foram utilizadas janelas de 15 e 21 pontos, respectivamente. Para os espectros obtidos usando os instrumentos convencionais, foi utilizada uma janela de 31 pontos.

Tabela 2. Espectrofotômetros convencionais e protocolos utilizados na aquisição dos espectros

Grupo escaneado	Espectrofotômetro	Faixa espectral	Protocolo usado na aquisição dos espectros
Modo de preparo I	Bomem	1100-2500 nm	Para cada amostra, o espectro foi adquirido a partir de 32 escaneamentos com resolução espectral de 16 cm ⁻¹ . Para isto, foi utilizado um sistema automático de rotação de amostras.
Modo de preparo II	Antaris	1100-2500 nm	Para cada amostra, foram coletados cinco espectros, cada um deles adquiridos a partir de 32 escaneamentos com resolução espectral de 16 cm ⁻¹ . A amostra foi mantida estática durante a coleta dos espectros e cada medição foi feita em uma posição diferente da amostra.
Modo de preparo III	Antaris	1100-2500 nm	Para cada amostra, foram coletados cinco espectros, cada um deles adquiridos a partir de 32 escaneamentos com resolução espectral de 16 cm ⁻¹ . A amostra foi mantida estática durante a coleta dos espectros e cada medição foi feita em uma posição diferente da amostra.

As regiões dos espectros NIR que apresentaram baixa relação sinal/ruído foram removidas antes do desenvolvimento dos modelos. Desta forma, apenas os comprimentos de onda entre 1365-2475 e 953-1649 nm foram utilizados no desenvolvimento dos modelos de

predição de COS baseados nos espectros adquiridos usando o NeoSpectra e o NanoNIR, respectivamente. Dos espectros adquiridos usando os espectrofotômetros convencionais, foi removida a região espectral entre 1000-1100 nm, restando os comprimentos de onda entre 1100-2500 nm.

Todas as transformações e pré-processamentos dos espectros NIR foram realizados usando o software Unscrambler X 10.4 (Camo, Norway).

2.3.4. Remoção de outliers e definição dos conjuntos de calibração e validação

Possíveis outliers no conjunto de dados foram identificados com base no teste de Hotelling (T^2), nos resíduos espectrais e usando a função X-Y relation, implementada no Unscrambler X 10.4 (Camo, Norway). Ao todo, 79 amostras foram identificadas como possíveis outliers e removidas do conjunto de dados. Após a remoção destas amostras, as amostras de solo foram divididas em dois grupos (Figura 2): conjunto de calibração, composto por 320 amostras, e conjunto de validação, composto por 160 amostras independentes. Para garantir que o conjunto de calibração contemplasse todo o domínio espectral da população de solos em estudo, as amostras que o compunham foram selecionadas usando o algoritmo Kennard-Stone (KENNARD; STONE, 1969).

2.3.5. Desenvolvimento dos modelos de regressão globais

Todas as amostras do conjunto de calibração foram utilizadas para treinar os modelos globais de determinação de COS desenvolvidos a partir dos espectros NIR coletados usando os diferentes tipos de preparo das amostras de solo (modos I, II e III). O conjunto de validação foi utilizado para a validação externa de cada modelo.

A regressão por quadrados mínimos parciais - PLSR (*partial least square regression*) pelo método de validação cruzada completa (*leave-one-out*) (OLIVIERI, 2018), foi utilizada para calibrar os modelos de determinação de COS. A raiz do erro quadrático médio - RMSE (*root means square error*), da validação cruzada (RMSECV) e da predição (RMSEP), o coeficiente de determinação (R^2) e a relação entre o desempenho e a distância interquartil - RPIQ (*ratio of performance to interquartile range*) foram utilizados para avaliar a performance dos modelos (Tabela 4) (BELLON-MAUREL et al., 2010; OLIVIERI, 2018).

O Teste de Randomização (VOET, 1994) foi utilizado para comparar estatisticamente a acurácia dos modelos desenvolvidos a partir de espectros adquiridos com os diferentes tipos de

espectrofotômetros e também para avaliar os efeitos do preparo das amostras de solo no desempenho de cada espectrofotômetro.

2.3.6. Desenvolvimento dos modelos de regressão locais

A primeira etapa realizada para o desenvolvimento dos modelos de regressão local, foi avaliar a similaridade entre cada amostra do conjunto de validação e as 320 amostras do conjunto de calibração. Para isto, o número de variáveis foi reduzido através do PLSR global, desenvolvido usando as 320 amostras do conjunto de calibração e as 160 amostras do conjunto de validação foram projetadas na dimensão espacial gerada por este PLSR. Os escores das amostras de calibração e validação foram registrados. A similaridade entre cada amostra do conjunto de validação e as amostras do conjunto de calibração foi então avaliada pela Distância de Mahalanobis (MAHALANOBIS, 1936) entre os seus escores.

A partir da Distância de Mahalanobis entre uma amostra que se deseja analisar usando a NIRS e todas as amostras do conjunto de calibração, é possível identificar, no conjunto de calibração, as amostras que apresentam maior similaridade com a amostra em análise e, assim, os modelos de regressão local podem ser construídos. O nível de similaridade desejado para as amostras selecionadas foi imposto utilizando uma Distância de Mahalanobis máxima, pré-definida.

Foram testados diferentes critérios de similaridade (Distância de Mahalanobis máxima variando de 0.1 a 3.6) para o preparo dos modelos locais. Após a seleção do grupo de amostras do conjunto de calibração que atende ao critério de similaridade utilizado, os modelos locais foram construídos. O número ótimo de termos (variáveis latentes) utilizado em cada modelo local foi escolhido de forma a minimizar o valor de RMSECV. O desempenho dos modelos locais foi avaliado utilizando os valores de referência e os valores preditos das amostras de validação.

É importante destacar que, para a construção do modelo local utilizado para avaliar uma determinada amostra do conjunto de validação era necessário que, pelo menos três amostras do conjunto de calibração atendessem o critério de similaridade imposto pela Distância de Mahalanobis máxima. Caso contrário, se apenas 1, 2 ou nenhuma amostra do conjunto de calibração atendessem o critério definido, não era possível desenvolver o PLSR local, portanto, esta amostra era removida do conjunto de validação e o seu valor não foi determinado. O desempenho destes modelos locais não pôde ser comparado com o desempenho dos modelos globais, já que o Teste de Randomização requer necessariamente que todas as amostras preditas

sejam analisadas com os dois modelos comparados (VOET, 1994). Neste contexto, a definição final pela Distância de Mahalanobis máxima utilizada nos modelos locais a serem comparados com os modelos globais de cada grupo de espectros (tipo de espectrofotômetro e de preparo de solo) foi feita pela seleção do menor valor de RMSEP obtido quando todas as amostras do conjunto de validação atendiam ao critério de similaridade e puderam ser avaliadas pelo modelo.

Todos os processos que envolveram a construção dos modelos de regressão locais foram implementados na interface de desenvolvimento Jupyter Notebook com a linguagem de programação Python. As principais bibliotecas utilizadas foram NumPy (HARRIS et al., 2020), Pandas (MCKINNEY, 2010), SciPy (VIRTANEN et al., 2020) e Scikit-learn (PEDREGOSA et al., 2011).

3. Resultados

3.1. Teores de COS

Os teores de COS nas 559 amostras de solo variaram entre 0,6-195,9 g kg⁻¹ (Figura 2). Contudo, no conjunto de dados, aproximadamente 14 % das amostras foram identificadas como possíveis outliers. Após a remoção destas amostras, a amplitude dos teores de COS reduziu consideravelmente, dado que a maioria dos outliers identificados foram amostras com teores de COS maiores do que 40 g kg⁻¹ (Figura 2).

3.2. Dados espectrais

Os espectros das 559 amostras de solo, medidos usando os diferentes espectrofotômetros, são apresentados na Figura 3, e os espectros derivados médios, na Figura 4. Embora o NeoSpectra não forneça informações nos comprimentos de onda entre 1000-1350 nm, os espectros obtidos com o uso deste instrumento foram muito semelhantes aos obtidos com o uso dos espectrofotômetros de referência (Figura 4). As principais bandas de absorção nestes espectros foram observadas nas regiões próximas a 1400 e 1900 nm e entre 2050-2500 nm (Figura 4). Os espectros obtidos com o NanoNIR, por outro lado, fornecem informações em uma faixa espectral muito restrita. Desta forma, a principal banda observada nos espectros adquiridos com este instrumento ocorreu na região próxima a 1400 nm (Figura 3).

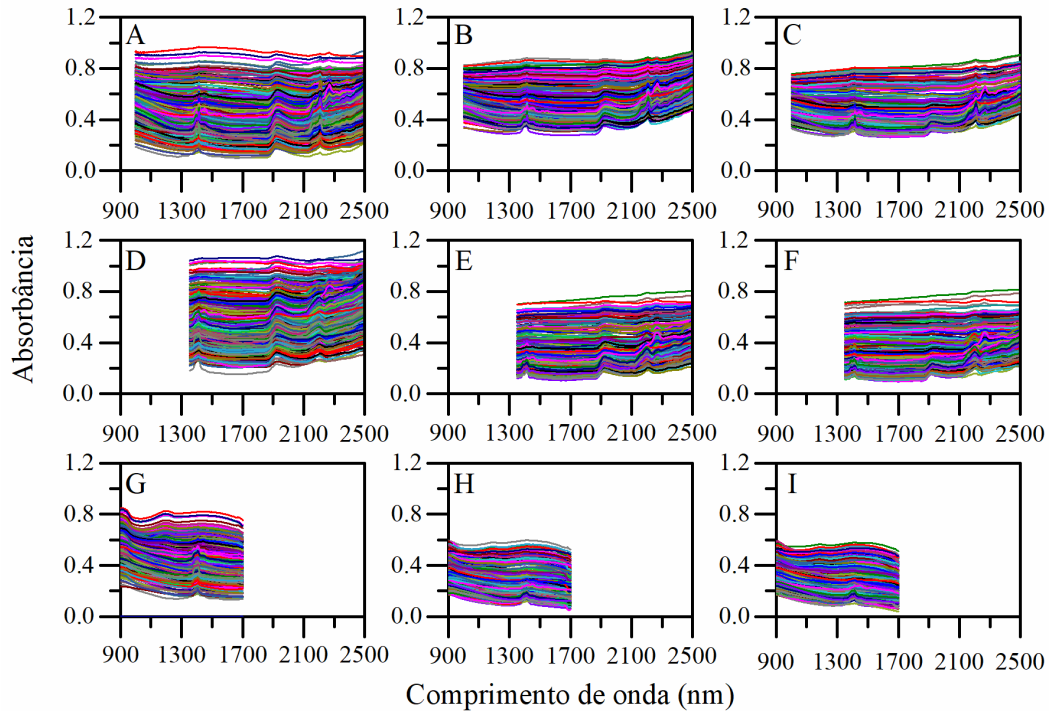


Figura 3. Espectros das 559 amostras de solo do conjunto de dados adquiridos usando: o modo de preparo I (amostras secas ao ar e peneiradas a 2 mm) e o espectrofotômetro convencional (A), o NeoSpectra (D) e o NanoNIR (G); o modo de preparo II (amostras secas ao ar e maceradas) e o espectrofotômetro convencional (B), o NeoSpectra (E) e o NanoNIR (H); e o modo de preparo III (amostras secas em estufa e maceradas) e o espectrofotômetro convencional (C), o NeoSpectra (F) e o NanoNIR (I).

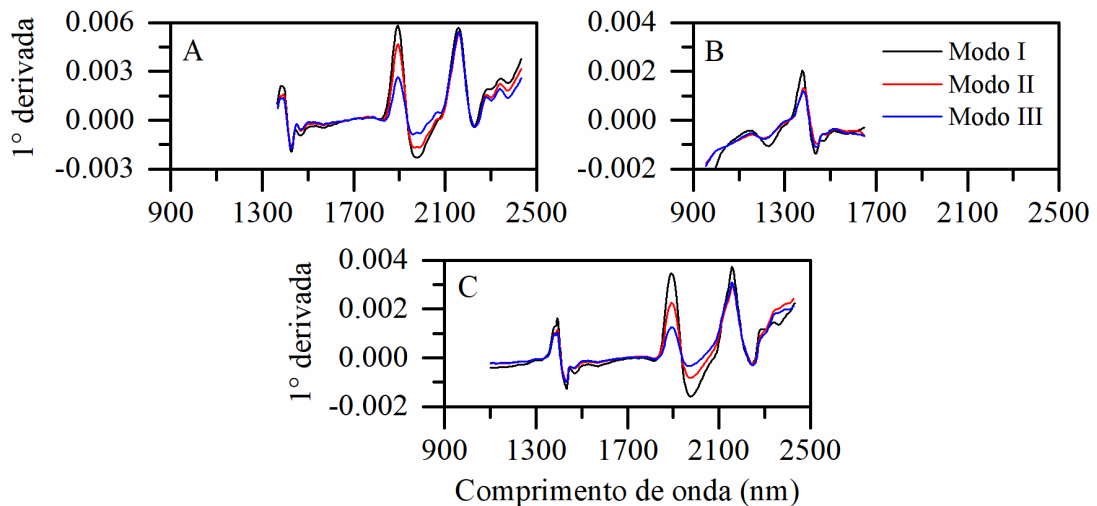


Figura 4. Espectros derivados médio das 559 amostras do conjunto de dados, adquiridos usando o NeoSpectra (A), o NanoNIR (B) e o espectrofotômetro convencional (C). Modo I, solo seco ao ar e peneirado a 2 mm; Modo II, solo seco ao ar e macerado; Modo III, solo seco em estufa e macerado.

O modo de preparo das amostras de solo afetou as características dos espectros NIR (Figuras 3 e 4). Independentemente do instrumento utilizado, os deslocamentos da linha de base dos espectros reduziram com o uso de amostras de solo maceradas (modos de preparo II e III) (Figura 3). Contudo, observando-se os espectros derivados, nota-se que as regiões espectrais próximas a 1400 e 1900 nm foram as mais afetadas pelo modo de preparo utilizado (Figura 4). A intensidade das absorções nestas regiões foi menor quando os espectros foram adquiridos usando amostras de solo maceradas do que usando amostras com granulometria mais grosseira (Figura 4). Assim como a moagem, a secagem do solo antes da aquisição dos espectros reduziu a intensidade destas bandas de absorção (Figura 4).

3.3. Determinação de COS usando modelos globais

Entre os espectrofotômetros NIR compactos, o NeoSpectra foi o que apresentou melhor desempenho na determinação de COS (Tabela 3). Os valores de RPIQ obtidos com o uso deste espectrofotômetro foram sempre maiores do que 2 (Figura 5) e, independentemente do modo de preparo utilizado, a performance dos modelos desenvolvidos utilizando o NeoSpectra foi semelhante ou superior a obtida com o uso de espectrofotômetros convencionais (Tabela 4).

Tabela 3. Indicadores da performance dos modelos globais desenvolvidos usando os diferentes modos de preparo da amostra de solo e diferentes espectrofotômetros

Modo de preparo da amostra de solo	Instrumento	Validação cruzada			Validação	
		VL	RSMECV	R ²	RSMEP	R ²
Modo de preparo I	NeoSpectra	10	6,2	0,44	6,3	0,47
	NanoNIR	3	7,7	0,14	8,0	0,16
	Convencional	10	6,1	0,46	6,4	0,47
Modo de preparo II	NeoSpectra	9	5,6	0,53	5,8	0,54
	NanoNIR	2	7,5	0,18	7,8	0,20
	Convencional	10	5,6	0,54	6,0	0,52
Modo de preparo III	NeoSpectra	10	5,7	0,53	6,0	0,52
	NanoNIR	3	7,4	0,20	7,6	0,24
	Convencional	13	5,5	0,56	5,8	0,56

^aModo I, solo seco ao ar e peneirado a 2 mm; II, solo seco ao ar e macerados; III, solo secos em estufa e macerado.

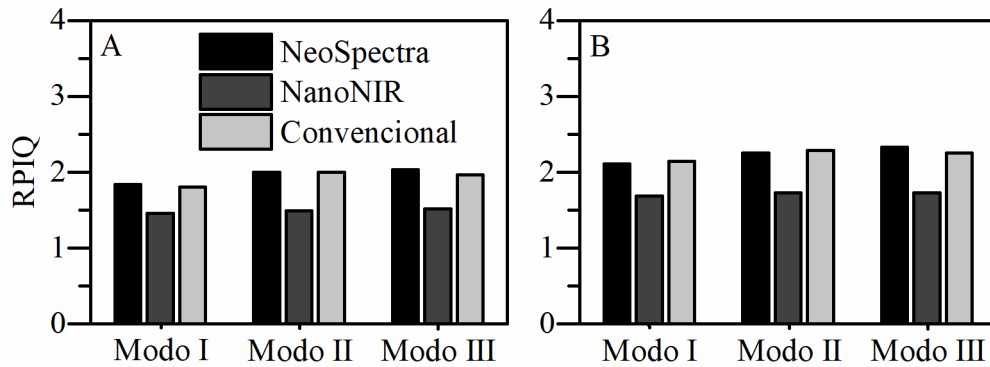


Figura 5. Relação entre o desempenho e a distância interquartil (RPIQ) na validação cruzada (A) e na validação externa (B) dos modelos. Modo I, solo seco ao ar e peneirado a 2 mm; Modo II, solo seco ao ar e macerado; Modo III, solo seco em estufa e macerado.

Os modelos desenvolvidos com os espectros adquiridos pelo NanoNIR foram os mais simples. Os modelos PLSR precisaram de 2 a 3 variáveis latentes para estimar os teores de COS (Tabela 3). Contudo, o uso deste instrumento resultou em modelos com altos valores de RMSECV e RMSEP e pequenos valores de R^2 e RPIQ, independentemente do modo de preparo das amostras de solo utilizado (Tabela 3 e Figura 2).

O modo de preparo das amostras de solo afetou a capacidade preditiva dos modelos desenvolvidos com o NeoSpectra (Tabela 3). Com o uso deste espectrofotômetro, o modelo desenvolvido utilizando o modo de preparo I (amostras secas ao ar e peneiradas a 2 mm) foi o que apresentou menor acurácia (Tabelas 3 e 4). Nos modelos desenvolvidos com o uso de amostras de solo maceradas (modos de preparo II e III), os valores de RMSECV e RMSEP até 10 e 8 %, respectivamente, e os valores de R^2 e RPIQ, na validação-cruzada e na validação externa, aumentaram (Tabelas 3 e 4 e Figura 5). Por outro lado, comparando os modelos desenvolvidos usando os modos de preparo II (amostras de solo maceradas e secas ao ar) e III (amostras de solo maceradas e secas em estufa), é possível observar que a secagem do solo em estufa não afeta significativamente o RMSEP dos modelos (Tabela 4).

Diferente do que ocorreu para o NeoSpectra, a acurácia dos modelos desenvolvidos com o uso do NanoNIR não foi afetada pelo modo de preparo das amostras de solo antes da aquisição de seus espectros (Tabela 3). Embora tenha sido observada redução nos valores de RMSECV e RMSEP quando foram utilizadas amostras de solo maceradas (Tabela 3), estes ganhos não foram significativos (Tabela 4).

Tabela 4. Comparação da acurácia dos modelos desenvolvidos utilizando os diferentes modos de preparo das amostras de solo e espectrofotômetros pelo Teste de Randomização

Instrumentos e modo de preparo das amostras de solo		RSMEP (g kg ⁻¹)		P-Valor ^b
A	B	A	B	
Modo de preparo I				
NeoSpectra	Convencional	6,3	6,4	ns
NanoNIR	Convencional	8,0	6,4	< 0,001
NanoNIR	NeoSpectra	8,0	6,3	< 0,001
Modo de preparo II				
NeoSpectra	Convencional	5,8	6,0	0,0457
NanoNIR	Convencional	7,8	6,0	< 0,001
NanoNIR	NeoSpectra	7,8	5,8	< 0,001
Modo de preparo III				
NeoSpectra	Convencional	6,0	5,8	ns
NanoNIR	Convencional	7,8	5,8	< 0,001
NanoNIR	NeoSpectra	7,8	6,0	< 0,001
NeoSpectra				
Modo de preparo I	Modo de preparo II	6,3	5,8	0,0288
Modo de preparo I	Modo de preparo III	6,3	6,0	0,0845
Modo de preparo II	Modo de preparo III	5,8	6,0	ns
NanoNIR				
Modo de preparo I	Modo de preparo II	8,0	7,8	ns
Modo de preparo I	Modo de preparo III	8,0	7,8	ns
Modo de preparo II	Modo de preparo III	7,8	7,8	ns

^aModo de preparo: I, amostras secas ao ar e peneiradas a 2 mm; II, amostras secas ao ar e maceradas; III, amostras secas em estufa e maceradas. ^bValores de $P > 0.2$ foram considerados não significativos (ns)

3.4. Determinação de COS usando modelos locais

Na Tabela 5 são apresentadas as Distâncias de Mahalanobis, utilizadas como limite máximo para a seleção das amostras do conjunto de calibração, o RMSEP e o RPIQ obtidos nos diferentes modelos locais. O critério de similaridade (Distâncias de Mahalanobis máxima)

que permitiu o desenvolvimento de modelos locais mais acurados variou em função do espectrofotômetro utilizado e do modo de preparo das amostras de solo (Tabela 5).

Novamente, o NeoSpectra foi o espectrofotômetro compacto com melhor desempenho (Tabela 5). O RMSEP dos modelos locais desenvolvidos com este equipamento variou entre 6,0 – 5,1 g kg⁻¹ e os valores de RPIQ foram os maiores entre os equipamentos analisados (Tabela 5). Em comparação com o desempenho dos modelos globais, o uso de modelos locais proporcionou redução nos valores de RMSEP, independentemente do modo de preparo utilizado (Tabela 6). O uso de modelos locais também beneficiou o desempenho do NanoNIR (Tabela 5). Os valores de RPIQ foram maiores do que os observados nos modelos globais e os valores de RMSEP reduziram significativamente, sendo que os ganhos na acurácia dos modelos desenvolvidos com este instrumento chegaram a 11 % (Tabela 5).

Tabela 5. Indicadores da performance dos modelos locais desenvolvidos usando os diferentes modos de preparo das amostras de solo e tipos de espectrofotômetros

	Instrumento	Validação		
		DM ^b	RMSEP (g kg ⁻¹)	RPIQ
Modo de preparo I	NeoSpectra	2,7	6,0	2,3
	NanoNIR	2,2	7,1	1,9
	Convencional	3,2	6,2	2,2
Modo de preparo II	NeoSpectra	3,5	5,3	2,6
	NanoNIR	3,5	7,3	1,9
	Convencional	3,4	5,7	2,4
Modo de preparo III	NeoSpectra	3,1	5,1	2,7
	NanoNIR	1,6	7,0	1,9
	Convencional	3,4	5,7	2,4

^aModo de preparo I, solo seco ao ar e peneirado a 2 mm; II, solo seco ao ar e macerado; III, solo seco em estufa e macerado. ^bDM, Distância de Mahalanobis máxima utilizada na seleção das amostras para calibração dos modelos de regressão local.

Tabela 6. Comparação da acurácia dos modelos desenvolvidos utilizando os diferentes espectrofotômetros e modos de preparo do solo pelo Teste de Randomização

Espectrofotômetros e modo de preparo das amostras de solo ^a			RSMEP (g kg ⁻¹)		P-Value ^b
	A	B	A	B	
NeoSpectra					
Modo de preparo I	Modelo global	Modelos local	6,3	6,0	0,1590
Modo de preparo II	Modelo global	Modelos local	5,9	5,3	0,0080
Modo de preparo III	Modelo global	Modelos local	6,0	5,1	< 0,0010
NanoNIR					
Modo de preparo I	Modelo global	Modelos local	8,0	7,1	< 0,0010
Modo de preparo II	Modelo global	Modelos local	7,8	7,3	0,0539
Modo de preparo III	Modelo global	Modelos local	7,8	7,0	< 0,0010

^aModo de preparo: I, amostras secas ao ar e peneiradas a 2 mm; II, amostras secas ao ar e maceradas; III, amostras secas em estufa e maceradas. ^bP-Valor > 0.2 foram considerados não significativos (ns).

4. Discussão

4.1. Outliers

No conjunto de dados, aproximadamente 14 % das amostras foram identificadas como possíveis outliers, o que pode ser atribuído as limitações analíticas impostas pelo método de referência. A taxa de recuperação de COS pela oxidação via úmida proposta por Walkley and Black varia entre 60 – 86 %. Para corrigir estes valores, os laboratórios de rotina aplicam um fator de correção nos resultados analíticos (BHATTACHARYYA et al., 2015). O índice utilizado nos laboratórios considera que 70 % do COS nas amostras é oxidado, contudo, a quantidade de carbono efetivamente oxidada pela solução ácida varia em função das condições do solo e do nível de proteção da matéria orgânica, entre outros fatores (BHATTACHARYYA et al., 2015; SATO et al., 2014), o que pode resultar em valores de COS superestimados ou subestimados. Além disto, o uso da oxidação via úmida em solos com altos teores de COS devem ser evitado, dado que a taxa de recuperação de carbono por este método, bem como a sua correlação com o método padrão de determinação de carbono (oxidação via seca), reduz com o aumento da concentração de COS (DE VOS et al., 2007; SATO et al., 2014).

4.2. Dados espectrais

As principais absorções nos espectros obtidos com os espectrofotômetros convencionais ocorreram nas regiões próximas a 1400 e 1900 nm e entre 2050-2500 nm. Estas feições são comuns para a maioria dos espectros NIR de amostras de solo e podem ser atribuídas a diferentes modos vibracionais de alta energia (sobretoms e combinação de vibrações) das ligações moleculares presentes nos compostos orgânicos e inorgânicos do solo (FANG et al., 2018; STENBERG et al., 2010; VISCARRA ROSSEL; BEHRENS, 2010). Brevemente, absorções próximas a 1400 nm estão principalmente associadas ao primeiro sobretom de ligações O-H e N-H e ao terceiro sobretom de ligações C=O (FANG et al., 2018; STENBERG et al., 2010; VISCARRA ROSSEL; BEHRENS, 2010). Já as absorções que ocorrem em comprimentos de onda próximos a 1900 nm estão geralmente relacionadas à combinações das vibrações O-H, presentes na água residual do solo, ao segundo sobretom das vibrações C=O e ao terceiro sobretom de ligações C-H (FANG et al., 2018; ROUDIER et al., 2017; STENBERG et al., 2010; VISCARRA ROSSEL; BEHRENS, 2010). Por fim, as combinações das vibrações O-H, C-H e N-H são as principais responsáveis pelas absorções observadas na região entre 2050-2500 nm (STENBERG et al., 2010; VISCARRA ROSSEL; BEHRENS, 2010).

Entre os espectrofotômetros compactos, o NeoSpectra foi o que abrangia a maior faixa espectral (1350-2500 nm). Os espectros adquiridos usando este instrumento, assim como os espectros obtidos com o uso de equipamentos convencionais, continham maior riqueza de informações sobre os constituintes do solo. Os espectros obtidos com o NanoNIR, por outro lado, apresentaram apenas uma fração destas informações.

O preparo das amostras de solo antes da aquisição dos espectros afeta as feições espectrais, independentemente do tipo de espectrofotômetro utilizado. Embora o aumento da homogeneidade na distribuição de partículas do solo tenha proporcionado uma redução nos deslocamentos da linha de base dos espectros, através dos espectros derivados é possível observar que as regiões próximas a 1400 e 1900 nm foram as mais afetadas pelo pré-processamento das amostras de solo. Nestas regiões, a radiação NIR pode ser absorvida por grupos OH presentes na água residual do solo (FANG et al., 2018; ROUDIER et al., 2017; VISCARRA ROSSEL; BEHRENS, 2010). Com a moagem das amostras, os microagregados do solo são quebrados e parte da água higroscópica pode ser perdida, reduzindo assim as absorções nestas regiões. Nas amostras maceradas e secas em estufa, a umidade do solo é ainda menor, com isto, a intensidade das absorções próximas a 1400 e 1900 nm nos espectros adquiridos usando o modo de preparo III, foi ainda menor.

4.3. Determinação de COS usando modelos globais

O NeoSpectra apresentou desempenho semelhante ao dos espectrofotômetros NIR convencionais e foi o instrumento compacto com maior acurácia na determinação dos teores de COS. Este resultado está relacionado, em grande parte, com a faixa espectral de trabalho destes instrumentos. Embora a região espectral abrangida pelo NanoNIR (900-1700 nm) contenha informações sobre diferentes compostos orgânicos do solo (STENBERG et al., 2010; VISCARRA ROSSEL; BEHRENS, 2010), os espectros obtidos com o NeoSpectra são mais ricos em informação, principalmente pelo fato de as ligações moleculares serem avaliadas em diferentes níveis vibracionais (combinações e sobretons) (STENBERG et al., 2010; VISCARRA ROSSEL; BEHRENS, 2010; WORKMAN; WEYER, 2008).

Os valores RPIQ dos modelos desenvolvidos com o NeoSpectra variaram entre 1,8 e 2,3. Este índice é baseado no RMSE do modelo e no intervalo interquartil da variável na população de solos em estudo. Valores entre 1,8-2,0 e 2,0-2,5 indicam modelos acurados e com capacidade preditiva boa e muito boa, respectivamente (BELLON-MAUREL et al., 2010; LEONE et al., 2012). Apesar do desempenho dos modelos desenvolvidos usando o modo de preparo I ter sido satisfatório, a acurácia na determinação dos teores de COS foi maior com o uso de amostras de solo maceradas (modos de preparo II e III). Este benefício proporcionado pelo preparo da amostra já foi relatado em trabalhos com espectrofotômetros convencionais e pode ser atribuído à quebra de microagregados e ao aumento da homogeneidade das amostras (BARTHÈS et al., 2006; MILTZ; DON, 2012; TAMBURINI et al., 2017), o que aumenta a representatividade dos espectros adquiridos e, conseqüentemente, melhora o desempenho deste espectrofotômetro compacto na determinação de COS (FONSECA et al., 2022). É importante destacar que, embora a acurácia dos modelos de determinação de COS melhore com o uso de amostras maceradas, e a moagem do solo seja um procedimento que já é realizado nos laboratórios de rotina antes da determinação de COS por oxidação via úmida, os ganhos na predição foram pequenos.

O RMSEP dos modelos obtidos com o uso do NeoSpectra variou entre 5,9 e 6,3 g kg⁻¹. Em uma revisão recente a respeito da performance de espectrofotômetros Vis-NIR, foi estimado que o RMSEP médio dos modelos de determinação de COS, desenvolvidos em escalas variadas (local, regional, nacional e global), é de 5 g kg⁻¹ (AHMADI et al., 2021), aproximadamente 1.0 g kg⁻¹ menor do que os valores de RMSEP obtidos com o NeoSpectra. Contudo, o RMSEP dos modelos aumenta com o aumento da escala do conjunto de dados (global > nacional > regional

> local) (VISCARRA ROSSEL et al., 2016). Considerando apenas os modelos desenvolvidos em escala regional, o RMSEP médio observado em um levantamento realizado por Viscarra-Russel et. al. foi de 9.0 g kg^{-1} (VISCARRA ROSSEL et al., 2016). Nota-se que, o RMSEP obtido com o uso do NeoSpectra, um espectrofotômetro NIR compacto de baixo custo, foi menor do que o RMSEP médio dos modelos desenvolvidos em escala regional com espectrofotômetros Vis-NIR, que possui um alto custo.

A principal limitação do NeoSpectra, assim como de outros instrumentos compactos, em relação aos instrumentos convencionais, é a sua menor capacidade de sondar a amostra, o que pode prejudicar a representatividade dos espectros adquiridos (FONSECA et al., 2022). Contudo, no presente trabalho, os espectrofotômetros compactos foram acoplados a um sistema de movimentação de amostras capaz de movimentar os frascos contendo solo sobre o feixe de sondagem óptica destes equipamentos. O uso deste sistema aumenta a área sondada da amostra durante a medição dos espectros (PASQUINI; HESPANHOL, 2021), proporcionando o aumento da representatividade dos espectros adquiridos e, conseqüentemente, a melhoria na performance de espectrofotômetros compactos na determinação de COS (FONSECA et al., 2022), ao ponto de o NeoSpectra apresentar performance semelhante ou até mesmo superior a de espectrofotômetros convencionais.

4.4. Determinação de COS usando modelos locais

A performance dos espectrofotômetros NIR compactos na determinação de COS melhorou com o uso de modelos locais. Estes ganhos estão relacionados com a redução das variabilidades espectrais, que não estão associadas ao COS, no conjunto de calibração. Nos espectros NIR de amostras de solo, as absorções associadas aos compostos inorgânicos são intensas e proeminentes (FANG et al., 2018), enquanto as bandas associadas aos compostos orgânicos são geralmente sutis, com menor intensidade (VISCARRA ROSSEL et al., 2011). Assim, nos modelos desenvolvidos em escala regional, que incluem diversos tipos de solo, a modelagem das informações espectrais relacionadas ao COS é dificultada, o que prejudica a acurácia dos modelos desenvolvidos (LIU et al., 2019). Os modelos locais, por outro lado, são calibrados utilizando apenas as amostras do conjunto de calibração que apresentam maior similaridade espectral com a amostra analisada, com isto a interferência de outros atributos do solo nos espectros é menor e a acurácia dos modelos aumenta (LIU et al., 2019; SHI et al., 2015).

É importante destacar que, com o uso de modelos locais, foi possível determinar os teores de COS em amostras minimamente processadas (modo de preparo I) com acurácia semelhante à obtida com o uso de modelos globais e amostras finamente moídas (modos de preparo II e III). Contudo, assim como ocorreu nos modelos globais, o aumento da homogeneidade na distribuição de partículas do solo e a redução das variabilidades espectrais relacionadas à água do solo, proporcionadas pela moagem e secagem das amostras (BARTHÈS et al., 2006; MILTZ; DON, 2012; TAMBURINI et al., 2017), melhorou o desempenho dos modelos locais, reduzindo os valores de RMSEP para até 5.1 g kg^{-1} . Como mencionado, os espectros de amostras processadas representam melhor a composição média do solo. Portanto, quando estes espectros são utilizados no desenvolvimento de modelos locais, a identificação das amostras com composição semelhante é facilitada, assim como o processo de predição. Nota-se que, para os modos de preparo II e III, os modelos locais desenvolvidos usando o NeoSpectra apresentaram melhor desempenho do que os modelos locais desenvolvidos usando os espectrofotômetros convencionais ($P\text{-valor} < 0.06$, em ambos os casos). Este resultado pode ser explicado justamente pela melhor representatividade dos espectros adquiridos usando o NeoSpectra. Embora os instrumentos convencionais possuam maior capacidade de sondar as amostras, a ausência de um sistema de movimentação provavelmente afetou a representatividade dos espectros adquiridos com este instrumento, piorando o seu desempenho.

É importante destacar que, mesmo com os ganhos proporcionados pelo pré-processamento das amostras de solo, os valores de RMSEP obtidos com o uso dos modelos locais foram maiores do que o esperado. Quando o NeoSpectra foi utilizado para determinar os teores de COS nas amostras coletadas em um único local (Capítulo 1), o RMSEP foi de até 0.1% (1 g kg^{-1}), mesmo com o uso de amostras minimamente processadas (equivalente ao modo de preparo I). Partindo da premissa que os modelos locais simulam esta situação por meio da seleção das amostras usadas na calibração, era esperado que, com o seu uso o RMSEP se aproximasse de 0.2% , mas não foi o que ocorreu. Provavelmente, o desempenho abaixo do esperado ocorreu devido ao conjunto de calibração possuir poucas amostras que atendessem os critérios de similaridade mais rigorosos. De forma que, para prever todo o conjunto de dados, foi necessário utilizar maiores distâncias de Mahalanobis como critério limite de similaridade.

O RMSEP dos modelos locais foi influenciado pelo critério de similaridade utilizado e, embora o uso de modelos locais tenha sido bem-sucedido, os valores de RMSEP poderiam ter sido ainda menores caso estes modelos tivessem sido construídos utilizando critérios de similaridade mais restritivos. A exemplo disto, na Figura 6 são apresentados os valores de

RMSEP, dos modelos desenvolvidos usando o NeoSpectra e o modo de preparo III (amostras de solo maceradas e secas em estufa), em função da distância de Mahalanobis máxima, imposta como limite para a seleção das amostras do conjunto de calibração utilizadas no desenvolvimento dos modelos locais. Quando foram utilizadas condições mais restritivas (menor valor distância de Mahalanobis máxima) na seleção destas amostras, o RMSEP dos modelos locais reduziu para até 1.6 g kg^{-1} (Figura 6). Contudo, este valor é baseado no erro médio de apenas quatro amostras do conjunto de validação, já que as outras 156 amostras não atenderam ao critério de similaridade imposto (possuir um mínimo de três amostras do conjunto de calibração com distância de Mahalanobis de até 1.6). Caso este mesmo método de desenvolvimento dos modelos locais fosse utilizado em conjuntos de dados mais amplos, formados, por exemplo, por 10,000-20,000 amostras, provavelmente seria possível utilizar as condições de similaridade mais restritivas e o RMSEP dos modelos seria reduzido.

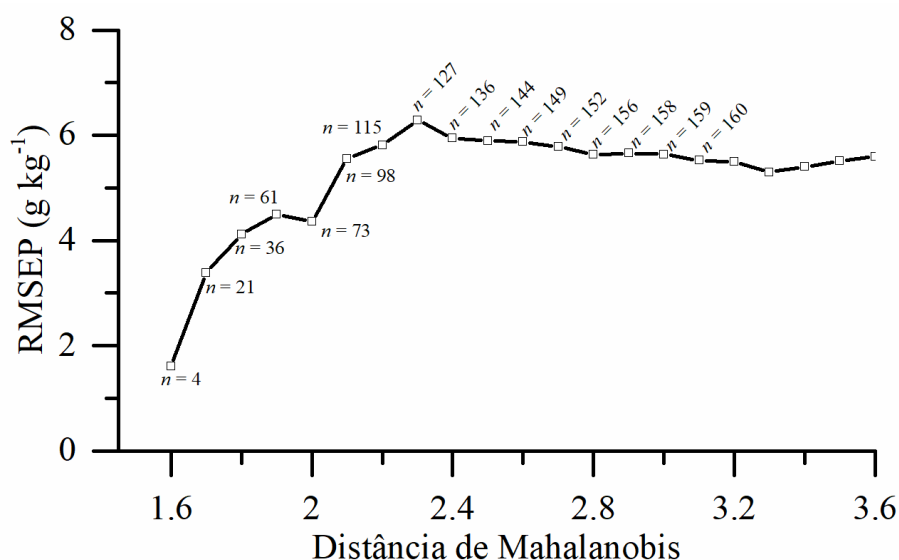


Figura 6. Valores de RMSEP, dos modelos locais desenvolvidos usando o NeoSpectra e o modo de preparo III, em função da Distância de Mahalanobis máxima imposta como limite de similaridade entre a amostra de validação e as amostras de calibração selecionadas. Os valores de n referem-se o número de amostras do conjunto de validação que atenderam os requisitos impostos e que, portanto, foram utilizadas no cálculo do RMSEP.

5. Conclusão

O tipo de instrumento compacto utilizado na aquisição dos espectros influencia diretamente a performance dos modelos de predição. No presente trabalho, o FT-NIR compacto, que abrange a faixa espectral entre 1350-2500 nm, foi espectrofotômetro com melhor

desempenho. Sua performance na predição de COS em larga escala foi semelhante, ou superior, em alguns casos, a performance dos modelos desenvolvidos usando espectrofotômetros convencionais.

O modo de preparo das amostras de solo antes da aquisição dos espectros afetou a acurácia dos modelos desenvolvidos com este espectrofotômetro, sendo os menores valores de RMSEP obtidos com o uso de amostras maceradas. No entanto, os ganhos deste processamento foram de no máximo 0.6 g kg^{-1} . Como o preparo da amostra é um processo laborioso e demorado, o custo-benefício do seu uso deve ser julgado pelo usuário. Por outro lado, por apresentarem melhor desempenho do que os modelos “globais” em todas as condições estudadas, o uso de modelos de regressão local é indicado para melhorar a acurácia dos espectrofotômetros NIR compactos na determinação de COS em escala regional.

Referências

AHMAD JANI, H. F. et al. Near infrared spectroscopy of plantation forest soil nutrients in Sabah, Malaysia, and the potential for microsite assessment. **Journal of Near Infrared Spectroscopy**, v. 29, n. 3, p. 148–157, 17 jun. 2021.

AHMADI, A. et al. Soil properties prediction for precision agriculture using visible and near-infrared spectroscopy: A systematic review and meta-analysis. **Agronomy**, v. 11, n. 3, 2021.

BARTHÈS, B. G. et al. Determination of total carbon and nitrogen content in a range of tropical soils using near infrared spectroscopy: Influence of replication and sample grinding and drying. **Journal of Near Infrared Spectroscopy**, v. 14, n. 5, p. 341–348, 2006.

BARTHÈS, B. G. et al. Performance comparison between a miniaturized and a conventional near infrared reflectance (NIR) spectrometer for characterizing soil carbon and nitrogen. **Geoderma**, v. 338, n. February 2018, p. 422–429, mar. 2019.

BEĆ, K. B. et al. Handheld near-infrared spectrometers: Where are we heading? **NIR news**, v. 31, n. 3–4, p. 28–35, 2020.

BEĆ, K. B.; GRABSKA, J.; HUCK, C. W. Principles and Applications of Miniaturized Near-Infrared (NIR) Spectrometers. **Chemistry - A European Journal**, v. 27, n. 5, p. 1514–1532, 2021.

BELLON-MAUREL, V. et al. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. **TrAC - Trends in Analytical Chemistry**, v. 29, n. 9, p. 1073–1081, 2010.

BELTRAME, K. K. et al. Soil organic carbon determination using NIRS: Evaluation of dichromate oxidation and dry combustion analysis as reference methods in multivariate calibration. **Journal of the Brazilian Chemical Society**, v. 27, n. 9, p. 1527–1532, 2016.

BHATTACHARYYA, T. et al. Walkley-Black Recovery Factor to Reassess Soil Organic Matter: Indo-Gangetic Plains and Black Soil Region of India Case Studies. **Communications in Soil Science and Plant Analysis**, v. 46, n. 20, p. 2628–2648, 2015.

DE SOUZA, J. J. L. L. et al. Geochemistry and spatial variability of metal(loid) concentrations in soils of the state of Minas Gerais, Brazil. **Science of the Total Environment**, v. 505, p. 338–349, 2015.

DE VOS, B. et al. Walkley-Black analysis of forest soil organic carbon: Recovery, limitations and uncertainty. **Soil Use and Management**, v. 23, n. 3, p. 221–229, 2007.

DEMATTE, J. A. M. et al. The Brazilian Soil Spectral Library (BSSL): A general view, application and challenges. **Geoderma**, v. 354, n. August, p. 113793, 2019a.

DEMATTE, J. A. M. et al. Soil analytical quality control by traditional and spectroscopy techniques: Constructing the future of a hybrid laboratory for low environmental impact. **Geoderma**, v. 337, p. 111–121, mar. 2019b.

FANG, Q. et al. Visible and Near-Infrared Reflectance Spectroscopy for Investigating Soil Mineralogy: A Review. **Journal of Spectroscopy**, v. 2018, p. 1–14, 2018.

FONSECA, A. A. DA et al. Effect of the sample measurement representativeness on soil carbon determination using near-infrared compact spectrophotometers. **Geoderma**, v. 409, n. September 2021, p. 115636, mar. 2022.

HARRIS, C. R. et al. Array programming with NumPy. **Nature**, v. 585, n. 7825, p. 357–362, 2020.

JACONI, A.; DON, A.; FREIBAUER, A. Prediction of soil organic carbon at the country scale: stratification strategies for near-infrared data. **European Journal of Soil Science**, v. 68, n. 6, p. 919–929, nov. 2017.

KENNARD, R. W.; STONE, L. A. Computer Aided Design of Experiments. **Technometrics**, v. 11, n. 1, p. 137–148, fev. 1969.

LEONE, A. P. et al. Prediction of Soil Properties with PLSR and vis-NIR Spectroscopy: Application to Mediterranean Soils from Southern Italy. **Current Analytical Chemistry**, v. 8, n. 2, p. 283–299, 1 abr. 2012.

LIU, S. et al. Estimating forest soil organic carbon content using vis-NIR spectroscopy: Implications for large-scale soil carbon spectroscopic assessment. **Geoderma**, v. 348, n. December 2018, p. 37–44, 2019.

MAHALANOBIS, P. C. On the Generalized Distance in Statistics. **Proceedings of the National Institute of Sciences of India**, v. 2, n. 1, p. 49–55, 1936.

MCKINNEY, W. Data Structures for Statistical Computing in Python. **Proceedings of the 9th Python in Science Conference**, v. 1, n. Scipy, p. 56–61, 2010.

MILTZ, J.; DON, A. Optimising sample preparation and near infrared spectra measurements of soil samples to calibrate organic carbon and total nitrogen content. **Journal of Near Infrared Spectroscopy**, v. 20, n. 6, p. 695–706, 2012.

MINASNY, B. et al. Soil carbon 4 per mille. **Geoderma**, v. 292, p. 59–86, 2017.

MOURA-BUENO, J. M. et al. Stratification of a local VIS-NIR-SWIR spectral library by homogeneity criteria yields more accurate soil organic carbon predictions. **Geoderma**, v. 337, n. October 2018, p. 565–581, 2019.

MURA, S. et al. Vibrational spectroscopy and Aquaphotomics holistic approach to determine chemical compounds related to sustainability in soil profiles. **Computers and Electronics in Agriculture**, v. 159, n. February, p. 92–96, 2019.

OLIVIERI, A. C. **Introduction to Multivariate Calibration**. Cham: Springer International Publishing, 2018.

PASQUINI, C. Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. **Analytica Chimica Acta**, v. 1026, p. 8–36, out. 2018.

PASQUINI, C.; HESPANHOL, M. C. A rotational-linear sample probing device to improve the performance of compact near-infrared spectrophotometers. **Microchemical Journal**, v. 170, p. 106747, nov. 2021.

PEDREGOSA, F. et al. SCIKIT-LEARN: MACHINE LEARNING IN PYTHON. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, set. 2011.

ROUDIER, P. et al. Evaluation of two methods to eliminate the effect of water from soil vis-NIR spectra for predictions of organic carbon. **Geoderma**, v. 296, p. 98–107, 2017.

SANTANA, F. B.; SOUZA, A. M.; POPPI, R. J. Green methodology for soil organic matter analysis using a national near infrared spectral library in tandem with learning machine. **Science of the Total Environment**, v. 658, p. 895–900, 2019.

SANTOS, C. A. T. DOS et al. A Review on the Applications of Portable Near-Infrared Spectrometers in the Agro-Food Industry. **Applied Spectroscopy**, v. 67, n. 11, p. 1215–1233, 1 nov. 2013.

SATO, J. H. et al. Methods of soil organic carbon determination in Brazilian savannah soils. **Scientia Agricola**, v. 71, n. 4, p. 302–308, ago. 2014.

SAVITZKY, A.; GOLAY, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. **Analytical Chemistry**, v. 36, n. 8, p. 1627–1639, jul. 1964.

SHI, Z. et al. Prediction of soil organic matter using a spatially constrained local partial least squares regression and the Chinese vis-NIR spectral library. **European Journal of Soil Science**, v. 66, n. 4, p. 679–687, 2015.

STENBERG, B. et al. Visible and Near Infrared Spectroscopy in Soil Science. In: SPARKS, D. L. (Ed.). . **Advances in Agronomy**. Burlington: Academic Press, 2010. v. 107p. 163–215.

TAMBURINI, E. et al. Effects of Moisture and Particle Size on Quantitative Determination of Total Organic Carbon (TOC) in Soils Using Near-Infrared Spectroscopy. **Sensors**, v. 17, n. 10, p. 2366, 17 out. 2017.

UFV et al. **Mapa de Solos do Estado de Minas Gerais**. Viçosa: [s.n.].

VIRTANEN, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. **Nature Methods**, v. 17, n. 3, p. 261–272, 2020.

VISCARRA ROSSEL, R. A. et al. On the soil information content of visible-near infrared reflectance spectra. **European Journal of Soil Science**, v. 62, n. 3, p. 442–453, 2011.

VISCARRA ROSSEL, R. A. et al. A global spectral library to characterize the world's soil. **Earth-Science Reviews**, v. 155, p. 198–230, 2016.

VISCARRA ROSSEL, R. A.; BEHRENS, T. Using data mining to model and interpret soil diffuse reflectance spectra. **Geoderma**, v. 158, n. 1–2, p. 46–54, 2010.

VOET, H. VAN DER. Comparing the predictive accuracy of models using a simple randomization test. **Chemometrics and Intelligent Laboratory Systems**, v. 25, n. 2, p. 313–323, nov. 1994.

WALKLEY, A.; BLACK, I. A. An examination of the degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. **Soil Science**, v. 37, n. 1, p. 29–38, jan. 1934.

WORKMAN, J.; WEYER, L. **Practical guide to interpretive near-infrared spectroscopy**. Boca Raton: Taylor & Francis Group, 2008.

Capítulo 3

Spectral information supporting near-infrared spectroscopy to estimate C, N, $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ in soil organic matter fractions

Abstract

Monitoring carbon (C) and nitrogen (N) content and their isotopic composition ($\delta^{13}\text{C}$ and $\delta^{15}\text{N}$) in soil organic matter (SOM) fractions is important to understand C and N cycling in soil and near-infrared (NIR) spectroscopy (NIRS) can be used to estimate these attributes quickly, cheaply and non-destructively. Our goal was to build regression models from NIR spectra to estimate C, N, $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ in mineral-associated organic matter (MAOM) and particulate organic matter (POM), and assess the spectral information supporting these models. The contents of C and N in bulk soil (TOC and TON, respectively) and in SOM fractions, as well as the $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ values, were measured using an isotopic ratio mass spectrometer. Based on reference values and NIR spectra of soil, we calibrated and validated regression models using partial least square regression (PLSR) and the leave-one-out cross-validation technique. The models performance was evaluated using the root mean square error (RMSE) and ratio of performance to interquartile (RPIQ). The PLSR models to estimate C and N content in SOM fractions presented small RMSE and R^2 close to one. These results, achieved by PLSR, are based on the information from several wavelengths distributed throughout the NIR region, but absorption between 1350-1600 nm were particularly important to estimate MAOM-C and MAOM-N and those between 2300-2500 nm were relevant to determine POM-C and POM-N. Our results demonstrate that PLSR-NIRS can be used to non-destructively characterize SOM. In addition to providing information on TOC and TON content in bulk soil, PLSR-NIRS allows to infer the contents of C, N, $\delta^{13}\text{C}$ in POM and MAOM fractions of SOM.

Keywords: Diffuse reflectance spectroscopy; mineral-associated organic matter, particulate organic matter, soil organic carbon determination, soil organic nitrogen determination, determination of stable isotopes of carbon and nitrogen.

1. Introduction

Soil organic matter (SOM) is a primary indicator of soil quality. Defined as a key factor to mitigate global climate change and to increase plant yield (HARDEN et al., 2018), SOM is the largest terrestrial carbon (C) pool, with approximately 3051 Pg of C (PAUL, 2016), and is also the main source of nitrogen (N) for plants, which is the most important nutrient to optimise plant yield (STRAWN; HOHN; O'CONNOR, 2020). Managing soil to increase C and N stocks requires well-substantial knowledge of SOM dynamics and monitoring only the amount of SOM may be insufficient (LAVALLEE; SOONG; COTRUFO, 2019). The SOM can be divided into components with contrasting behaviour (LAVALLEE; SOONG; COTRUFO, 2019). The particulate organic matter (POM) is mostly made up of plant and fungal fragments that are relatively undecomposed and that are rapidly cycled in soil (CAMBARDELLA; ELLIOTT, 1992; LAVALLEE; SOONG; COTRUFO, 2019). In contrast, mineral-associated organic matter (MAOM) is biochemically different from the plant tissue and its interaction with the mineral matrix of soil promotes a greater protection against mineralisation (COTRUFO et al., 2015; COTRUFO; WALLENSTEIN; BOOT, 2013; LAVALLEE; SOONG; COTRUFO, 2019). Monitoring the C and N contents in each SOM fraction can provide a more comprehensive understanding of C and N transformations in soil (LAVALLEE; SOONG; COTRUFO, 2019), especially when the isotopic composition of them is also determined (FERREIRA et al., 2021). Nevertheless, the methods currently used to determine these soil properties are destructive, laborious and time-consuming, which still limits their use for large-scale applications.

Near-infrared (NIR) spectroscopy (NIRS) is a rapidly, non-destructive, highly reproducible and environmentally friendly analytical technique that can be very useful for soil researchers, mainly in studies related to SOM (DEMATTE' et al., 2019b; FANG et al., 2018; GUERRERO; ROSSEL; MOUAZEN, 2010; VISCARRA ROSSEL et al., 2016). The NIRS mainly uses overtones and combinations of OH, CH, CO, NH and SH vibrations as a source of qualitative and quantitative information about a material under analysis (WORKMAN; WEYER, 2008). The development of NIRS-based analytical methods preferentially requires a direct correlation between spectral data and the presence/concentration of the species of interest in the sample (causal effect) (MCBRIDE, 2021; PASQUINI, 2018). Many studies have sought to explain NIR spectral information that can support the determination of total organic C (TOC) and total organic N (TON) in soils (LIN et al., 2017; STENBERG et al., 2010; VISCARRA

ROSSEL; BEHRENS, 2010). On the other hand, the potential of NIRS to provide specific information about C and N present in fractions of SOM is rarely studied (BARTHÈS et al., 2008; JACONI et al., 2019; ST. LUCE et al., 2014) and, to our knowledge, there are no studies reporting the wavelengths sourcing the spectral information supporting these models, which is necessary to demonstrate that these determination methods are not based on internal correlations between sample constituents (MCBRIDE, 2021).

Considering that different soil organic compounds, such as polysaccharides, carboxylic acids, amides and amine, absorb NIR radiation in different regions of the spectrum (STENBERG et al., 2010) and that broad chemical patterns of POM and MAOM are distinct (LAVALLEE; SOONG; COTRUFO, 2019), we hypothesize that NIRS can distinguish the spectral information related to each SOM fraction. Such information would be highly useful to non-destructively infer the chemical composition of SOM fractions and provide the fundamental basis to quantify the C and N content in MAOM and POM fractions using NIRS. Furthermore, changes in the mass of atoms involved in molecular vibrations that occur in the NIR region alter the energies of chemical bonds, which cause a shift in the wavelengths of the absorption bands (PASQUINI, 2018). This spectral information could be used by NIRS to assess the isotopic composition of C ($\delta^{13}\text{C}$) and N ($\delta^{15}\text{N}$) in SOM fractions. In this context, this work aimed to build regression models from NIR spectra to determine C and N content in SOM fractions with contrasting behaviour (POM and MAOM) in a fast and accurate way. Additionally, the potential of this analytical technique to estimate $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ in POM and MAOM fractions was evaluated. In addition to previous other studies, we attempt to identify the sources of the information supporting the regression models, which is necessary to establish a reliable NIRS-based method (MCBRIDE, 2021; PASQUINI, 2018).

2. Material and methods

2.1. Soil samples

The soil samples used in the present study were obtained from an incubation experiment that was carried out to assess the influence of soil mineralogy, carbon levels and type of Eucalyptus harvest residues on the formation of soil organic matter. Briefly, this incubation experiment was set up using four types of soil (Table 1): (i) Vetic Ferralsol with high carbon content; (ii) Vetic Ferralsol with low carbon content; (iii) Xanthic Ferralsol with high carbon content; (iiii) Xanthic Ferralsol with low carbon content. Each type of soil was air-dried, sieved

at 2 mm and placed into pots together with different types of Eucalyptus residues. The Eucalyptus harvest residues (leaves, bark, twigs, stem, and roots) were obtained from Eucalyptus seedlings enriched with ^{13}C . This experiment was carried out for one year. A complete description of the incubation experiment and its results can be found in Jesus (2012).

Table 1. Textural analyses and total organic carbon (TOC) content of soils

	Sampling Depth (cm)	Sand (g kg ⁻¹)	Silt (g kg ⁻¹)	Clay (g kg ⁻¹)	TOC (g kg ⁻¹)
Vetic Ferralsol with high carbon content	0-10	29.0	4.0	67.0	52.6
Vetic Ferralsol with low carbon content	60-100	29.0	6.0	65.0	18.6
Xanthic Ferralsol with high carbon content	0-10	37.0	5.0	58.0	17.7
Xanthic Ferralsol with low carbon content	60-100	24.0	6.0	70.0	7.1

As the transfer rate of C and N from plant residues to the MAOM and POM fractions, as well as the efficiency of SOM formation, varies according to the characteristics of the residues added into the soil (ALMEIDA et al., 2018; FERREIRA et al., 2021), a wide variation in TOC, TON, POM-C, MAOM-C, POM-N, MAOM-N, MAOM- $\delta^{13}\text{C}$, POM- $\delta^{13}\text{C}$, MAOM- $\delta^{15}\text{N}$ and POM- $\delta^{15}\text{N}$ contents was observed among the samples at the end of the incubation period (Table 2). This fact makes the samples gathered from this experiment ideal for evaluating the potential of NIRS to determine these soil attributes. Furthermore, few types of soils were used in this experiment, facilitating the identification of wavelengths regions that are preferentially related to each soil attribute evaluated, given that the NIR spectrum can be dominated by spectral features associated with soil mineralogy (VISCARRA ROSSEL et al., 2011).

2.2. Reference values for carbon and nitrogen

The physical fractionation of soil organic matter proposed by Cambardella and Elliott (1992) was used. In brief, the soil samples were air-dried and sieved at 2 mm. Subsamples of 10 g were placed into centrifuge tubes with 30 mL of sodium hexametaphosphate (5 g L⁻¹) and continuously agitated for 15 hours. The POM, fraction retained on the 53 μm sieve, and MAOM, fraction passing through a 53 μm sieve, were obtained by wet sieving.

The reference values of TOC and TON, and the C, $\delta^{13}\text{C}$, N, and $\delta^{15}\text{N}$ associated with each SOM fraction were obtained by dry combustion using an isotopic ratio mass spectrometer (Table 2). The correlation between the variables under analysis was also verified (Table 3).

Table 2. Descriptive statistics for soil attributes of the reference samples

	Calibration set ($n = 48$)						Validation set ($n = 24$)					
	Min.	Q1	Q3	Max.	Mean	SD	Min.	Q1	Q3	Max.	Mean	SD
TOC ¹	6.9	9.7	22.1	73.8	23.7	17.9	8.7	20.8	50.9	70.3	34.8	21.8
TON	0.6	0.8	1.7	4.7	1.6	1.1	0.6	1.2	3.5	4.6	2.3	1.4
MAOM-C	5.6	8.3	19.2	59.5	20.3	15.3	7.2	19.1	44.9	60.5	30.6	18.7
POM-C	0.3	1.7	3.4	17.0	3.4	3.5	1.0	1.6	6.3	10.9	4.2	3.3
MAOM-N	0.5	0.7	1.5	4.0	1.5	1.0	0.5	1.1	3.3	4.1	2.0	1.3
POM-N	0.0	0.1	0.2	2.3	0.2	0.3	0.1	0.1	0.3	0.5	0.2	0.1
MAOM- $\delta^{13}\text{C}$	-28.7	-16.8	33.0	120.2	17.5	40.6	-24.4	-11.6	26.1	87.4	12.7	27.2
POM- $\delta^{13}\text{C}$	-26.7	30.5	289.9	559.1	181.5	179.9	-25.2	49.6	221.7	486.9	177.2	151.6
MAOM- $\delta^{15}\text{N}$	5.1	17.1	28.0	37.7	22.5	7.8	-3.9	12.4	23.1	229.7	27.1	44.5
POM- $\delta^{15}\text{N}$	-31.7	14.9	36.0	171.4	30.0	37.7	-36.9	15.9	47.9	69.2	29.0	23.8

¹TOC, total organic carbon (g kg^{-1}); MAOM-C, carbon in mineral-associated organic matter (g kg^{-1}); POM-C, carbon in particulate organic matter (g kg^{-1}); TON, total nitrogen (g kg^{-1}); MAOM-N, nitrogen in mineral-associated organic matter (g kg^{-1}); POM-N, nitrogen in particulate organic matter (g kg^{-1}); MAOM- $\delta^{13}\text{C}$, ratio between stable isotopes of carbon in mineral-associated organic matter (‰); POM- $\delta^{13}\text{C}$, ratio between stable isotopes of carbon in particulate organic matter (‰); MAOM- $\delta^{15}\text{N}$, ratio between stable isotopes of nitrogen in mineral-associated organic matter (‰); POM- $\delta^{15}\text{N}$, ratio between stable isotopes of nitrogen in particulate organic matter (‰).

2.3. Near-infrared spectra

The NIR spectra of soil samples were obtained using an Antaris II FT-NIR Analyzer spectrometer (Thermo Scientific™, MA, USA) in the 1000 - 2500 nm range with 2 cm^{-1} resolution. We used air-dried, 2 mm sieved subsamples of bulk soil for spectra collection. For each sample, two spectra were collected, each one obtained from the average of 32 scans, and averaged to obtain the final spectrum. The final spectra, obtained in reflectance (R), were transformed in absorbance (A) by the Equation (1):

$$A = \text{Log}_{10}(1/R) \quad (1)$$

The spectral data were pre-processed using the second derivative by Savitzky-Golay filter with a window size of 41 data points (SAVITZKY; GOLAY, 1964) to minimize non-informative spectral variability (PASQUINI, 2018). In addition, the spectral region between

1000 and 1100 nm showed a low signal-to-noise ratio; therefore, it was removed after the application of the second derivative. After pre-processing the spectral data, an exploratory analysis was carried out through principal component analysis (PCA).

The samples were split into calibration (two thirds of dataset, n=48) and validation (one third of dataset, n=24) sets using the Kennard-Stone algorithm (KENNARD; STONE, 1969). Outliers were identified and removed from the calibration set using both the Hotelling statistics and spectral residuals (BRO; SMILDE, 2014). The calibrations with partial least square regression (PLSR) were performed using the leave-one-out cross-validation technique. The variables producing the more stable regression coefficients were identified by the Jack-knife algorithm of the Unscrambler (WESTAD; MARTENS, 2000). Aiming to create simpler models, the PLSR models were then rebuilt using only the stable variables identified by the Jack-knife algorithm.

The following statistical indices were used to evaluate model performance: the root mean square error (RMSE) of calibration (cross-validation) and prediction (external validation set), calculated according to Equation (2) (OLIVIERI, 2018); the ratio of performance to interquartile (RPIQ), given by Equation (3) (BELLON-MAUREL et al., 2010) and the coefficient of determination (R^2).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{i=n} (\hat{y}_i - y)^2}{n}} \quad (2)$$

$$RPIQ = \frac{IQ}{RMSE} \quad (3)$$

where \hat{y}_i is the predicted values, y are the reference values, n is the number of samples and IQ is the difference between the third and first quartiles (= $Q3 - Q1$) of the reference values.

Table 3 - Pearson's correlation between the variables¹ under study

	TOC	C-MAOM	C-POM	TON	N-MAOM	N-POM	$\delta^{13}\text{C-MAOM}$	$\delta^{13}\text{C-POM}$	$\delta^{15}\text{N-MAOM}$	$\delta^{15}\text{N-POM}$
TOC	1									
C-MAOM	0.99	1								
C-POM	0.83	0.76	1							
TON	0.90	0.89	0.78	1						
N-MAOM	0.98	0.98	0.79	0.91	1					
N-POM	0.44	0.35	0.81	0.51	0.38	1				
$\delta^{13}\text{C-MAOM}$	-0.56	-0.57	-0.43	-0.52	-0.55	-0.30	1			
$\delta^{13}\text{C-POM}$	-0.56	-0.57	-0.40	-0.58	-0.58	-0.32	0.76	1		
$\delta^{15}\text{N-MAOM}$	-0.21	-0.20	-0.22	-0.18	-0.19	-0.16	0.31	0.21	1	
$\delta^{15}\text{N-POM}$	0.06	0.07	0.02	0.11	0.11	0.02	-0.02	-0.04	-0.04	1

¹TOC, total organic carbon; C-MAOM, carbon in mineral-associated organic matter; C-POM, carbon in particulate organic matter; TON, total nitrogen; N-MAOM, nitrogen in mineral-associated organic matter; N-POM, nitrogen in particulate organic matter; $\delta^{13}\text{C-MAOM}$, ratio between stable isotopes of carbon in mineral-associated organic matter; $\delta^{13}\text{C-POM}$, ratio between stable isotopes of carbon in particulate organic matter; $\delta^{15}\text{N-MAOM}$, ratio between stable isotopes of nitrogen in mineral-associated organic matter; $\delta^{15}\text{N-POM}$, ratio between stable isotopes of nitrogen in particulate organic matter.

3. Results

3.1. NIR spectra of soil

The baseline-shift in the original spectra, caused mainly by the radiation scattering, makes it difficult to visualise the informative absorption bands about the composition of soil samples (Figure 1A and Figure 1B). However, it was possible to identify the main spectra features after the application of the second derivative using the Savitzky-Golay filter (Figure 1C). The main absorption bands highlighted occurred in the ranges of 1300-1650, 1800-1950 and 2050-2500 nm (Figure 1C).

While the NIR spectra of the different soils overlapped across many bands, many spectral differences between the Xanthic Ferralsol and Vetic Ferralsol were apparent (Figure 1C). The carbon levels did not change the NIR spectra main absorption bands; however, the peaks in the spectra of soil with low carbon content showed absorption intensity higher than those with high carbon content (Figure 1C). The type of harvest residues added into soil did not change the main absorption bands in soil NIR spectra.

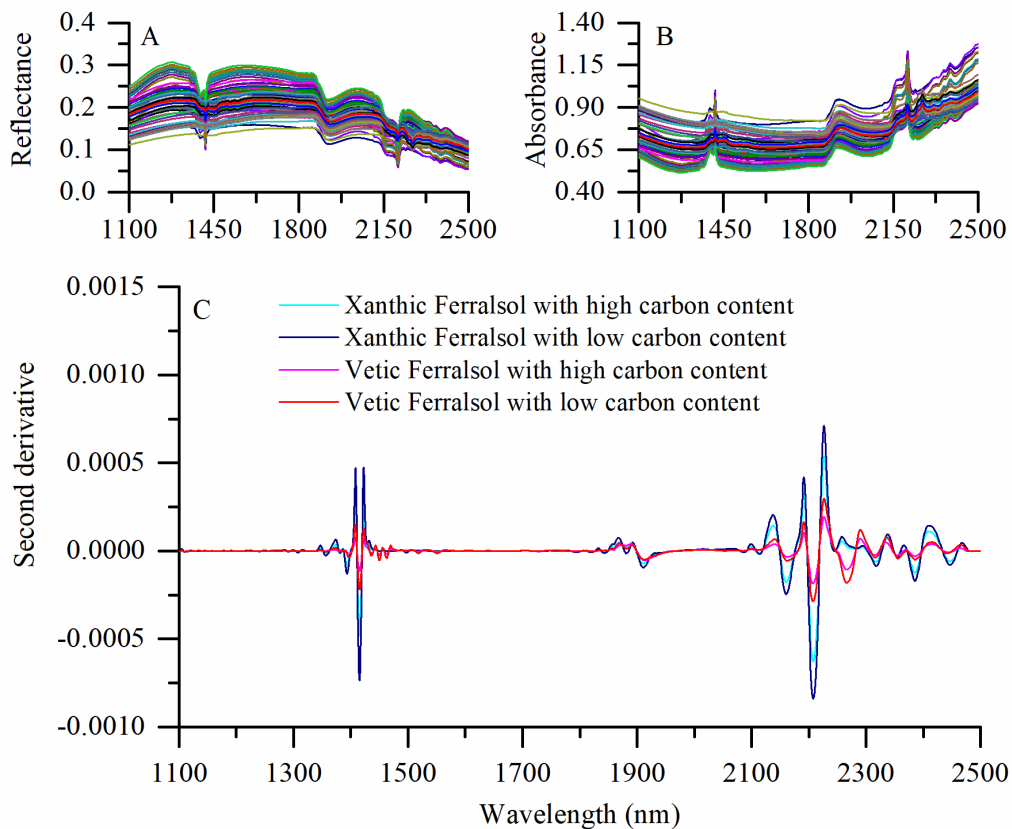


Figure 1. NIR spectra of soil samples measured in reflectance (A), after the spectral data transformation to absorbance (B) and after second derivative application using Savitzky-Golay filter (separated by the soil type) (C).

3.2. Exploratory analysis of soil spectra

The first two principal components explain 99 % of the total variance of the soil spectra, splitting the soil samples into four distinct groups (Figure 2). The first principal component (PCA-1) explained 96 % of the data variability and separated the two soil classes. The second principal component (PCA-2) explained 3 % of the data variability, but subdivides these groups according to the soil carbon content. In general, samples were randomly distributed within the clusters, so it was not possible to observe the grouping of the samples according to types of residues that were added to the soil.

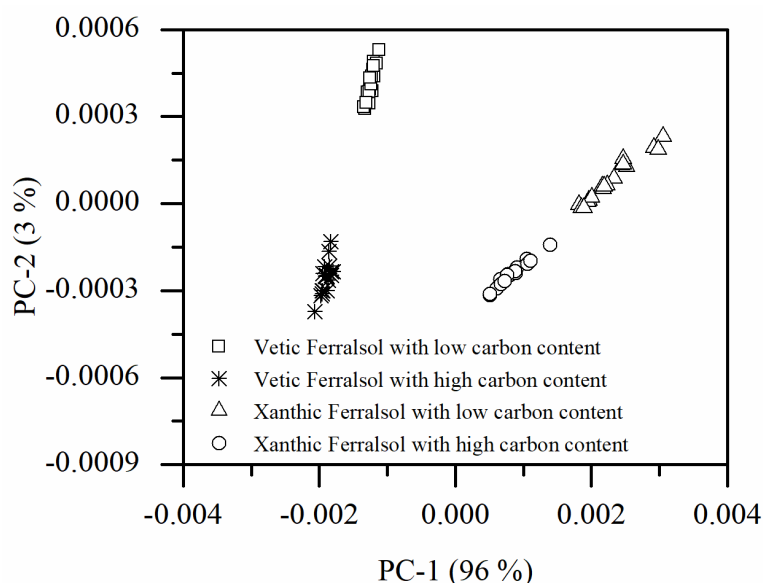


Figure 2. Score plots of the soil samples spectra in the principal components space.

3.3. Performance of PLSR models

3.3.1. Carbon

The model to determine TOC required five latent variables and presented good performance, with small RMSE (3.6 g kg^{-1}), determination coefficients close to one (0.960) and RPIQ > 2.5 (Table 4). The regression coefficients of the PLSR model for TOC determination are presented in Figure 3A. A large number of wavelengths distributed throughout the NIR region contributed to the determination of TOC and many of these variables yield stable regression coefficients during the calibration procedure. The stable regression coefficients positively correlated with TOC content occurred in the ranges of 1200-1420, 1850-1940, 2190-2250 and 2400-2500 nm (Figure 3). When only the stable regression coefficients were used, the PLSR models required less latent variables to estimate TOC, although no major changes in the performance of the models were observed (Table 4).

Table 4- Merit figures for the models using cross-validation (calibration set) and validation (external validation set)

	PLSR models							PLSR models developed with selected variables						
	Calibration set					Validation set ($n = 24$)		Calibration set					Validation set ($n = 24$)	
	n	LV ¹	RMSE	R^2	RPIQ	RMSE	R^2	n	LV ¹	RMSE	R^2	RPIQ	RMSE	R^2
TOC ²	44	5	3.6	0.957	3.5	4.3	0.960	44	4	3.8	0.951	3.3	4.6	0.954
C-MAOM	48	6	2.9	0.964	3.8	3.3	0.968	48	5	3.2	0.957	3.4	3.6	0.961
C-POM	42	5	0.8	0.716	2.2	1.3	0.852	42	4	0.7	0.759	2.4	1.3	0.836
TON	46	5	0.3	0.931	2.9	0.5	0.867	46	5	0.2	0.952	4.5	0.5	0.877
N-MAOM	41	6	0.2	0.956	4.0	0.2	0.970	41	4	0.2	0.958	4.0	0.2	0.971
N-POM	45	9	0.1	0.843	2.8	0.1	0.763	45	9	0.1	0.921	5.6	0.1	0.760
$\delta^{13}\text{C}$ -MAOM	39	10	14.5	0.802	3.4	15.4	0.667	39	7	11.8	0.869	4.2	17.0	0.592
$\delta^{13}\text{C}$ -POM	33	7	60.4	0.862	4.3	45.6	0.906	33	5	48.4	0.911	5.4	42.3	0.920
$\delta^{15}\text{N}$ -MAOM	35	3	4.7	0.451	2.3	44.4	0.000	35	4	4.5	0.500	2.4	44.2	0.000
$\delta^{15}\text{N}$ -POM	31	1	12.5	0.000	1.7	24.5	0.000	31	3	8.3	0.550	2.5	27.3	0.000

¹LV, Latent Variables; RMSE, root mean square error (g kg^{-1}); R^2 , coefficient of determination; bias (g kg^{-1}); RPIQ, ratio of performance to interquartile; ²Labels as in Table 2.

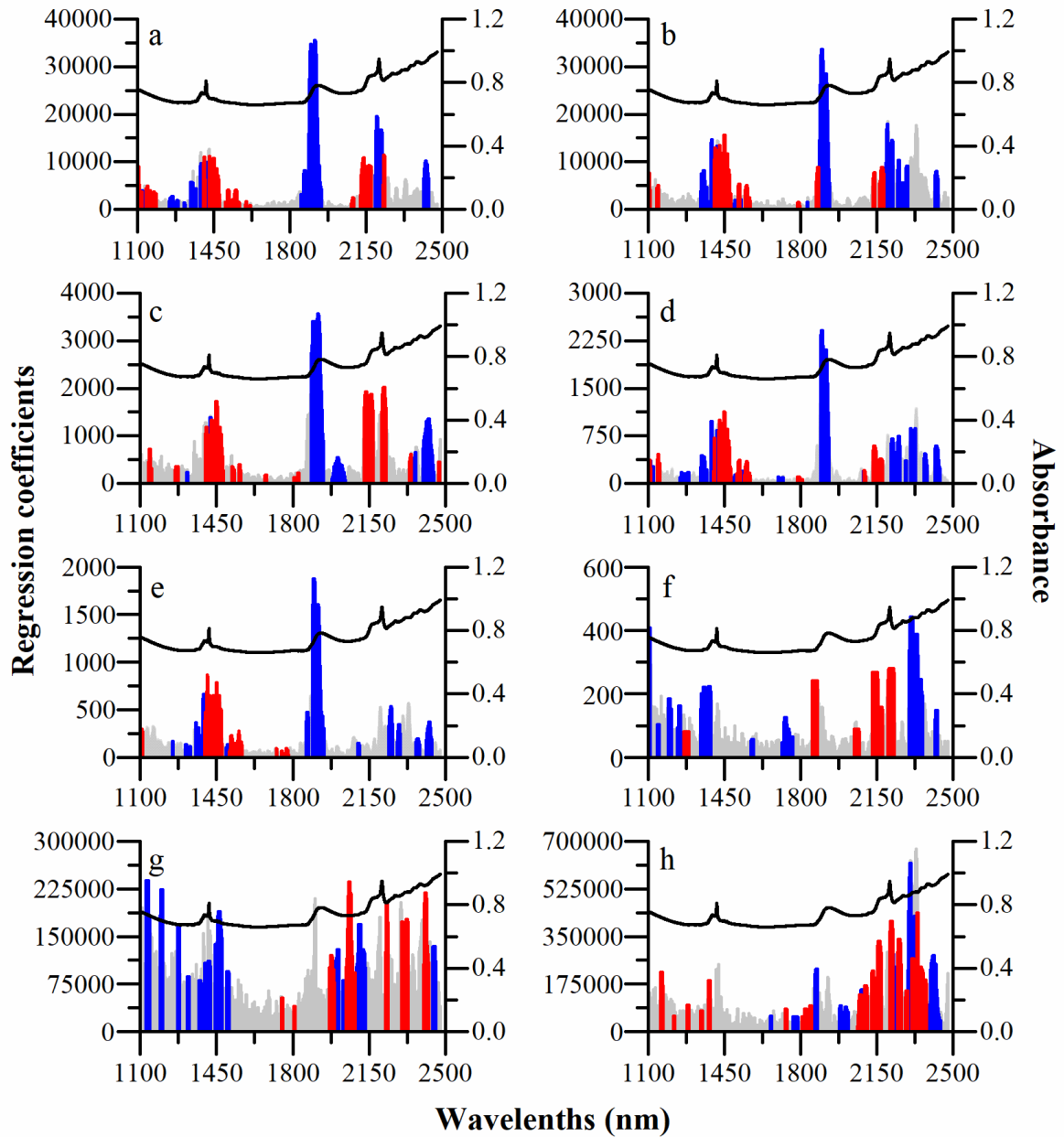


Figure 3. Absolute values of the regression coefficients of the PLSR models to predict: TOC (A), C in MAOM fraction (B), C in POM fraction (C), TON (D), N in MAOM fraction (E), N in POM fraction (F), $\delta^{13}\text{C}$ in MAOM fraction (G); $\delta^{13}\text{C}$ in POM fraction (H). The stable coefficients related to the 2nd derivative values of the spectral variables positively correlated with the soil propriety are marked in blue, and those negatively correlated with soil proprieties were marked in red. The average raw NIR spectrum of the soil samples set is shown as a black line.

The PLSR required one more latent variable to explain variations in MAOM-C content than to explain variations in TOC content, but the model to determine MAOM-C also presented

good performance, with small RMSE (2.9 g kg^{-1}), high R^2 (0.964) and RPIQ > 2.5 (Table 4). To determine MAOM-C fraction, the PLSR used information from several wavelengths throughout the NIR spectrum. The stable regression coefficients positively correlated with MAOM-C contents occurred mainly between 1350-1600, 1850-1940 and 2190-2250 (Figure 3). However, in the ranges of 1350-1600 and 2190-2250 nm wavelengths negatively correlated with the content of MAOM-C were also observed. The performance of the model to predict POM-C was also satisfactory (Table 4). The RMSE and R^2 values of this model were 0.8 g kg^{-1} and 0.716, respectively, and the RPIQ > 2.0 (Table 4). Several wavelengths distributed throughout the NIR region contain information on POM-C, but the stable PLSR regression coefficients positively correlated with this parameter were observed mainly between 1890-2050 and 2350-2500 nm (Figure 3). Wavelengths in the ranges of 1350-1600 and 2190-2250 nm also produced stable regression coefficients to estimate POM-C, nevertheless, most of them were negatively correlated with POM-C content. As in the model to determine TOC content, the PLSR models were simpler when only the variables selected by the Jack-knife algorithm were used to estimate the MAOM-C and POM-C contents, but although no major changes in the performance of the models were observed (Table 4).

3.3.2. Nitrogen

The PLSR model required five latent variables to estimate TON content (Table 4), with small RMSE (0.3 g kg^{-1}), determination coefficients close to one (0.930) and RPIQ > 2.5 (Table 4). In general, the regression coefficients of this model presented similar behaviour to those used by PLSR to determine TOC content (Figure 3). The highest regression coefficients of these models occurred also between 1350-1600, 1890-1940, 2190-2500 nm, nevertheless, the variables between 2250-2400 nm showed an increased relationship with TON content than with TOC content. (Figure 3).

Among the PLSR models to estimate N in SOM fractions, the highest values of RPIQ were obtained in MAOM fraction (Table 4). The PLSR required six latent variables to estimate MAOM-N and this model presented small RMSE (0.2 g kg^{-1}) and $R^2 > 0.95$ (Table 4). The stable regression coefficients produced by PLSR to estimate N in MAOM fraction occurred in the ranges of 1420-1600, 1890-1940 and 2190-2500 nm (Figure 3). Among them, the spectral variables between 1890-1940 nm and 2190-2500 nm presented a positively correlation with MAOM-N content. In the region between 1420-1600 nm, absorptions showing positive correlation with MAOM-N content were also observed, however, the intensity of absorption at

some wavelengths in this spectral region reduced with increasing MAOM-N contents. The PLSR model for predicting POM-N also presented good performance, with small RMSE (0.1 g kg^{-1}), R^2 close to one (0.843) and RPIQ > 2.0 (Table 4). The stable variables produced by PLSR to estimate N in POM fraction occurred in the ranges of 1100-1400, 1700-1760, 2050-2500 nm (Figure 3), but the variables with a positive correlation with POM-N content occurred predominantly between 1100-1400, 1700-1760, 2300-2500 nm (Figure 3).

The performance of the models to estimate the N content in bulk soil and in SOM fractions had small improvements when only the regression coefficients selected by the Jack-knife algorithm were used. Furthermore, the number of latent variables required by PLSR to estimate MAOM-N was reduced.

3.3.3. Ratio between stable isotopes of carbon and nitrogen

The models for predicting the ratio between stable isotopes of C in MAOM fraction ($\text{MAOM-}\delta^{13}\text{C}$) and in POM fraction ($\text{POM-}\delta^{13}\text{C}$) presented high values of R^2 and RPIQ, and the RMSE of these models were 14.5 and 60.4 ‰, respectively (Table 4). The stable regression coefficients supporting the use of NIRS-PLSR to estimate $\delta^{13}\text{C}$ in MAOM fraction occurred in the ranges of 1100-1470 and 1970-2500 nm (Figure 3). Among them, the variables between 1100-1470, 2000-2150 and 2440-2460 nm presented a positive correlation with $\text{MAOM-}\delta^{13}\text{C}$ (Figure 3). The most important wavelengths to estimate $\delta^{13}\text{C}$ in POM fraction were between 1650-1850 and 2050-2500 nm (Figure 3). In these regions, $\text{POM-}\delta^{13}\text{C}$ was positively correlated with some variables and negatively correlated with others. The PLSR models developed only with the variables selected by the Jack-knife algorithm required less latent variables to estimate $\text{MAOM-}\delta^{13}\text{C}$ and $\text{POM-}\delta^{13}\text{C}$. Furthermore, the performance of these models was improved.

The models to predict the ratio between stable isotopes of N in SOM fractions presented unsatisfactory results. These models showed high RMSE and their determination coefficients were extremely low, even when they were remade using only the regression coefficients selected by the Jack-knife algorithm (Table 4).

4. Discussion

4.1. Soil spectra and exploratory analysis

Absorptions in NIR region result from overtones and combinations of fundamental vibrations of bonded atoms present in soil organic and inorganic chemical species (CLARK et

al., 1990; FANG et al., 2018; HUNT, 1977; VISCARRA ROSSEL et al., 2009; VISCARRA ROSSEL; BEHRENS, 2010), but the exact place where the absorption features will appear depends upon to what the bonded atoms are directly attached to (FANG et al., 2018; HUNT, 1977). It is known that absorptions in the 1300-1650 nm range are due to the first overtone of O-H, C-H and N-H bonds and to the third overtone of C=O bond (FANG et al., 2018; STENBERG et al., 2010; VISCARRA ROSSEL; BEHRENS, 2010). The O-H groups are usually associated to minerals, such as Kaolinite (FANG et al., 2018), and to residual soil water (CLARK et al., 1990; HUNT, 1977; VISCARRA ROSSEL; BEHRENS, 2010), while C-H, N-H and C=O groups are related to functional groups of soil organic compounds, such as carboxylic acids, aromatics, amides (STENBERG et al., 2010; VISCARRA ROSSEL; BEHRENS, 2010). The prominent peak observed between 1800-1900 nm are associated to the combinations of vibrational modes of O-H groups of residual soil water (CLARK et al., 1990; HUNT, 1977; VISCARRA ROSSEL; BEHRENS, 2010), nevertheless, bands due to vibrational modes of C-H and C=O bonds present in organics, like carboxylic acids and methyl groups, can also occur in this region (STENBERG et al., 2010; VISCARRA ROSSEL; BEHRENS, 2010). The spectral features observed in 2050-2500 nm are due to combinations of vibrational modes of hydroxyl groups attached to metals, such as Fe-OH, Mg-OH and Al-OH (FANG et al., 2018; VISCARRA ROSSEL; BEHRENS, 2010), and combinations of vibrational modes of N-H, C-H and C-O bonds present in organic functional groups of soil, such as carbohydrates, polysaccharides, amine, methyl and aliphatic groups (STENBERG et al., 2010; VISCARRA ROSSEL; BEHRENS, 2010). Although the high-energy vibrational modes of CH, NH and CO groups present in organic compounds of soil occur in NIR region (STENBERG et al., 2010; VISCARRA ROSSEL; BEHRENS, 2010), most of these bands are weak. The spectral features related to inorganic compounds of soil, on the other hand, show high intensity (VISCARRA ROSSEL et al., 2011); therefore, they are mostly responsible for the major absorption bands observed in NIR spectra.

The PCA considers all the sources of variability present in NIR spectra to project the sample spectra in the principal component space. It was found that the first two principal components explain 99 % of the spectral data variability and made it possible to group the soil samples according to soil type and carbon saturation deficit. The first PC explains 96 % of the data variance and separates the soil samples into two groups with different mineralogy. This result confirms that the main spectral features in NIR spectra are related to soil mineralogy (VISCARRA ROSSEL et al., 2011). The second PC further subdivides each of these groups

according to the soil carbon saturation deficit. In each soil class, the main spectral differences between the soils with different carbon saturation deficits are related to the absorption intensity. Considering the similar texture of the soils along the profile (JESUS, 2012), the main source of variability considered by PC2 may be related to the difference in SOM concentration between these soils. In addition to absorbing radiation in different regions of the NIR spectrum (VISCARRA ROSSEL; BEHRENS, 2010), the SOM directly influences the formation of stable aggregates in the soil, which can modify the profile of NIR spectra (SHI et al., 2020). Regarding the influence of eucalyptus components, although variations in the type of the harvest residue added into the soil alter the chemical composition of SOM (ALMEIDA et al., 2018; FERREIRA et al., 2021), this information is obscured by weak absorption relative to the strong bands of inorganic compounds of soil (MCBRIDE, 2021).

4.2. PLSR models

4.2.1. Carbon

The RPIQ > 2.0 indicated that C content in bulk soil (TOC) and in each SOM fractions (MAOM-C and POM-C) were predicted accurately by the NIRS-PLSR (BELLON-MAUREL et al., 2010; LEONE et al., 2012). To this, the PLSR used information from several wavelengths across the NIR spectrum. The stable regression coefficients produced by PLSR to estimate POM-C were mainly observed in regions related to vibrational modes of: C=O bonds, present for example in carboxyl (VISCARRA ROSSEL; BEHRENS, 2010) and carbonyl groups (WORKMAN; WEYER, 2008) (near to 1950 and 1450 nm); C-O bonds, present in organics like phenolic groups (between 1950-2044 nm) (VISCARRA ROSSEL; BEHRENS, 2010); and C-H bonds, present in cellulosic materials (between 2350-2500 nm) (ST. LUCE et al., 2016). These organic compounds are important constituents of POM fraction of SOM (COTRUFO et al., 2015; LAVALLEE; SOONG; COTRUFO, 2019). Therefore, the spectral information generated through the direct absorption of NIR radiation by these compounds has greater importance for determining POM-C content. The stable regression coefficients produced by PLSR to estimate C in MAOM fraction, on the other hand, were mainly observed in the ranges of 1340-1600, 1890-1940 and 2190-2250 nm. In these regions can occur the vibrational modes of C=O bonds present in carboxylic acids (STENBERG et al., 2010; VISCARRA ROSSEL; BEHRENS, 2010), and O-H bonds present in inorganic compounds of soil, such as kaolinite, aluminium oxides and iron oxides (between 1340-1600 and 2190-2250 nm) (CLARK et al.,

1990; FANG et al., 2018; HUNT, 1977; VISCARRA ROSSEL et al., 2009; VISCARRA ROSSEL; BEHRENS, 2010). Carboxyl group is one of the main responsible for the attachment between SOM and mineral compounds of soil (DICK et al., 2009; SOKOL; SANDERMAN; BRADFORD, 2019). This association affect the energy required by the vibrational modes of O-H and C=O bonds (HUNT, 1977), which can reduce the intensity of absorption at some wavelengths and increase it at others. These spectral variations were used by PLSR to estimate MAOM-C content.

4.2.2. Nitrogen

The PLSR models to estimate TON, MAOM-N and POM-N presented RPIQ values greater than 2.5, which indicate models with excellent performance (BELLON-MAUREL et al., 2010; LEONE et al., 2012). The spectral information supporting the use of NIRS-PLSR to determine these soil attributes came from a large number of wavelengths distributed throughout the NIR region. Wavelengths between 2190-2500 nm were very informative on soil N. Absorptions in this region are usually associated to the combination of vibrational modes of N-H and C-H bonds present in compounds of soil containing N, such as proteins and amino sugars (ONO et al., 2007; ST. LUCE et al., 2016). These compounds can be found in partially decomposed plant-derived materials in the soil (ST. LUCE et al., 2016) and in the surface charges of soil minerals (SOKOL; SANDERMAN; BRADFORD, 2019), thus, the absorption of NIR radiation by them can provided information on N content in both fractions of SOM. To estimate POM-N, the PLSR also used information from regions associated to C-H bonds present in protein and polysaccharides (between 1100-1400 and 1700-1760 nm) (ONO et al., 2007; STENBERG et al., 2010). To determine MAOM-N, on the other hand, the main variables used by PLSR were observed in the ranges of 1340-1600 and 1890-1940 nm. The spectral information in these regions is generated through the direct absorption of NIR radiation by N-H and C=O bonds, present in amines and other proteins (around 1500 nm) (ONO et al., 2007; STENBERG et al., 2010; VISCARRA ROSSEL; BEHRENS, 2010), and by changes in binding energy and vibration anharmonicity of NH, CO and OH bonds caused by interaction between carboxyl groups, present in N compounds with low molecular weight, such as proteins, and the mineral fraction of the soil (DICK et al., 2009; LAVALLEE; SOONG; COTRUFO, 2019).

It is evident that PLSR-NIRS can use spectral information generated by the presence of compounds in the soil to estimate the contents of MAOM-N and POM-N. Nevertheless, many similarities were observed between the regression coefficients used to estimate C and N content. PLSR finds the relationships between the predictor matrix and the vector with the parameter

under analysis (LEONE et al., 2012), so, the built model can also use spectral variabilities that are not directly associated with the analyte, caused by internal correlations, for example, as a source of information (PASQUINI, 2018). Considering the strong correlation between C and N, in bulk soil and in SOM fractions (Table 3), and that N content was much lower than the C content into the samples, it is plausible that PLSR also used spectral information related to C compounds to estimate N, in bulk soil or in SOM fractions, and vice-versa.

4.2.3. Ratio between stable isotopes of carbon and nitrogen

Among the models to predict the isotopic composition of MAOM and POM, the best results were obtained in the prediction of the ratio between stable C isotopes (MAOM- $\delta^{13}\text{C}$ and POM- $\delta^{13}\text{C}$). The PLSR models for estimate MAOM- $\delta^{13}\text{C}$ and POM- $\delta^{13}\text{C}$ presented RPIQ values greater than 2.5, which indicates models with excellent predictive capacity (BELLON-MAUREL et al., 2010; LEONE et al., 2012). These models used information along the entire NIR spectrum. The stable regression coefficients positively correlated with MAOM- $\delta^{13}\text{C}$ occurred in regions usually associated with vibrational modes of C-H bonds (1100-1420 and 2440-2460 nm), C-O and C=O bonds (2000-2150 nm) and O-H bonds (1400-1470 nm) (ONO et al., 2007; STENBERG et al., 2010; VISCARRA ROSSEL; BEHRENS, 2010; WORKMAN; WEYER, 2008). The stable spectral variables positively correlated with POM- $\delta^{13}\text{C}$, on the other hand, were mainly observed in spectral regions associated to the high energy vibrational modes of C-H bonds (1650-1850 and 2200-2500 nm) (ONO et al., 2007; STENBERG et al., 2010; WORKMAN; WEYER, 2008). The mass of atoms involved in a chemical bond, as well as the mass of atoms close to the bond in chemical species, can change the natural frequency of vibration associated to a given chemical bond. Consequently, the energy (wavelength) associated to an absorption feature associated with that chemical bond also changes (PASQUINI, 2003, 2018). Increasing the ^{13}C content in the sample can shift the wavelength of the absorption bands. Although this shift is subtle, the PLSR can detect these variations in the spectra and correlate them with the concentration of ^{13}C (OLIVIERI, 2018). Nevertheless, it is important to stress that the residues employed in the present work were enriched with ^{13}C . Thus, the RMSE are low considering the population under study, but the magnitude of these errors limits the use of the models to trace the natural abundance of ^{13}C , for example. It was not possible to fit a model to predict MAOM-N and POM- $\delta^{15}\text{N}$ using PLSR. This may be attributed to the low content of ^{15}N in the samples, since the eucalyptus seedlings were not enriched with ^{15}N . Variations in the mass of N atoms present in the soil sample can influence the features of the NIR spectrum. However, when the concentration of the analyte is close to the detection

limit of the analytical technique, imposed by the signal-to-noise of the instrument, the information can be confused with noise, resulting in inaccurate models (PASQUINI, 2018).

5. Conclusion

Our results demonstrate that PLSR-NIRS can be used to characterize SOM rapidly and with good precision. In addition to providing information on TOC content in bulk soil, PLSR-NIRS can be used to estimate the contents of C in POM and MAOM fractions of SOM. The stable regression coefficients of these models confirmed that PLSR used direct analyte information to estimate TOC, MAOM-C, POM-C. It was also possible to evaluate the ratio between stable C isotopes in SOM fractions resulted from enriched materials, using NIRS. In this context, NIRS has proven to be a valuable tool that can be used to expand the knowledge on the dynamics of C in soil. Additional studies are recommended to evaluate in field experiments the use of conventional and compact NIR spectrophotometers for monitoring the dynamics of C in SOM fractions. The models for evaluating the N content in bulk soil and in SOM fractions are sufficiently accurate. Although the regression coefficients show slight differences in comparison to the models developed to measure C, there was a high correlation between N and C contents. Therefore, further studies are needed to assess the use of NIRS to measure the N present in both, soil and SOM fractions, preferably using soils that also received N from mineral fertilisation and with residues enriched with ^{15}N . Thus, it will become possible to verify whether the information accessed by the models is unequivocally related to the variation in the concentration of N in soil.

References

ALMEIDA, L. F. J. et al. Soil organic matter formation as affected by eucalypt litter biochemistry — Evidence from an incubation study. **Geoderma**, v. 312, p. 121–129, fev. 2018.

BARTHÈS, B. G. et al. Determining the distributions of soil carbon and nitrogen in particle size fractions using near-infrared reflectance spectrum of bulk soil samples. **Soil Biology and Biochemistry**, v. 40, n. 6, p. 1533–1537, 2008.

BELLON-MAUREL, V. et al. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. **TrAC - Trends in Analytical Chemistry**, v. 29, n. 9, p. 1073–1081, 2010.

BELTRAME, K. K. et al. Soil organic carbon determination using NIRS: Evaluation of dichromate oxidation and dry combustion analysis as reference methods in multivariate calibration. **Journal of the Brazilian Chemical Society**, v. 27, n. 9, p. 1527–1532, 2016.

BRO, R.; SMILDE, A. K. Principal component analysis. **Analytical Methods**, v. 6, n. 9, p. 2812–2831, 2014.

CAMBARDELLA, E. T.; ELLIOTT, C. A. Particulate Soil Organic-Matter Changes across a Grassland Cultivation Sequence. **Soil Science Society of America Journal**, v. 56, p. 777–783, 1992.

CARRA, J. B. et al. Near-Infrared Spectroscopy Coupled with Chemometrics Tools: A Rapid and Non-Destructive Alternative on Soil Evaluation. **Communications in Soil Science and Plant Analysis**, v. 50, n. 4, p. 421–434, 2019.

CLARK, R. N. et al. High spectral resolution reflectance spectroscopy of minerals. **Journal of Geophysical Research**, v. 95, n. B8, p. 12,653-12,680, 1990.

COTRUFO, M. F. et al. The Microbial Efficiency-Matrix Stabilization (MEMS) framework integrates plant litter decomposition with soil organic matter stabilization: do labile plant inputs form stable soil organic matter? **Global Change Biology**, v. 19, n. 4, p. 988–995, abr. 2013.

COTRUFO, M. F. et al. Formation of soil organic matter via biochemical and physical pathways of litter mass loss. **Nature Geoscience**, v. 8, n. 10, p. 776–779, 7 out. 2015.

DEMATTE, J. A. M. et al. Soil analytical quality control by traditional and spectroscopy techniques: Constructing the future of a hybrid laboratory for low environmental impact. **Geoderma**, v. 337, p. 111–121, mar. 2019.

DICK, D. P. et al. Química da Matéria Orgânica do Solo. In: MELO, V. DE F.; ALLEONI, L. R. F. (Eds.). . **Química e Mineralogia do Solo**. 1. ed. Viçosa: [s.n.]. p. 685.

FANG, Q. et al. Visible and Near-Infrared Reflectance Spectroscopy for Investigating Soil Mineralogy: A Review. **Journal of Spectroscopy**, v. 2018, p. 1–14, 2018.

FERREIRA, G. W. D. et al. Retaining eucalyptus harvest residues promotes different pathways for particulate and mineral-associated organic matter. **Ecosphere**, v. 12, n. 3, p. 1–18, 1 mar. 2021.

GUERRERO, C.; ROSSEL, R. A. V.; MOUAZEN, A. M. Special issue “Diffuse reflectance spectroscopy in soil science and land resource assessment.” **Geoderma**, v. 158, n. 1–2, p. 1–2, 2010.

HARDEN, J. W. et al. Networking our science to characterize the state, vulnerabilities, and management opportunities of soil organic matter. **Global Change Biology**, v. 24, n. 2, p. e705–e718, 2018.

HUNT, G. R. Spectral signatures of particulate minerals in the visible and near infrared. **Geophysics**, v. 42, n. 3, p. 501–513, 1977.

JACONI, A. et al. Log-ratio transformation is the key to determining soil organic carbon fractions with near-infrared spectroscopy. **European Journal of Soil Science**, v. 70, n. 1, p. 127–139, 2019.

JESUS, G. L. DE. **Compactação Do Solo, Espaçamento De Plantio, Resíduos De Colheita E Déficit De Saturação De Carbono Na Matéria Orgânica Do Solo Sob Povoamentos De Eucalipto**. [s.l.] Universidade Federal de Viçosa, 2012.

KENNARD, R. W.; STONE, L. A. Computer Aided Design of Experiments. **Technometrics**, v. 11, n. 1, p. 137–148, fev. 1969.

LAVALLEE, J. M.; SOONG, J. L.; COTRUFO, M. F. Conceptualizing soil organic matter into particulate and mineral-associated forms to address global change in the 21st century. **Global Change Biology**, v. 26, n. 1, p. 261–273, 2 jan. 2019.

LEONE, A. P. et al. Prediction of Soil Properties with PLSR and vis-NIR Spectroscopy: Application to Mediterranean Soils from Southern Italy. **Current Analytical Chemistry**, v. 8, n. 2, p. 283–299, 1 abr. 2012.

LIN, Z. D. et al. Improvements of the Vis-NIRS Model in the Prediction of Soil Organic Matter Content Using Spectral Pretreatments, Sample Selection, and Wavelength Optimization. **Journal of Applied Spectroscopy**, v. 84, n. 3, p. 529–534, 2017.

MCBRIDE, M. B. Estimating soil chemical properties by diffuse reflectance spectroscopy: Promise versus reality. **European Journal of Soil Science**, v. 73, n. 1, p. 1–11, 11 jan. 2022.

OLIVIERI, A. C. **Introduction to Multivariate Calibration**. Cham: Springer International Publishing, 2018.

ONO, K. et al. Spectrophotometrical characteristics in the near infrared region in beech (*Fagus crenata*) and pine (*Pinus densiflora*) litters at the various decomposing stages. **Journal of Forest Research**, v. 12, n. 4, p. 255–261, 20 ago. 2007.

PASQUINI, C. Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications. **Journal of the Brazilian Chemical Society**, v. 14, n. 2, p. 198–219, 2003.

PASQUINI, C. Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. **Analytica Chimica Acta**, v. 1026, p. 8–36, out. 2018.

PAUL, E. A. The nature and dynamics of soil organic matter: Plant inputs, microbial transformations, and organic matter stabilization. **Soil Biology and Biochemistry**, v. 98, p. 109–126, 2016.

SAVITZKY, A.; GOLAY, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. **Analytical Chemistry**, v. 36, n. 8, p. 1627–1639, jul. 1964.

SHI, P. et al. Vis-NIR spectroscopic assessment of soil aggregate stability and aggregate size distribution in the Belgian Loam Belt. **Geoderma**, v. 357, n. May 2019, p. 113958, jan. 2020.

SOKOL, N. W.; SANDERMAN, J.; BRADFORD, M. A. Pathways of mineral-associated soil organic matter formation: Integrating the role of plant carbon source, chemistry, and point of entry. **Global Change Biology**, v. 25, n. 1, p. 12–24, 2019.

ST. LUCE, M. et al. Rapid determination of soil organic matter quality indicators using visible near infrared reflectance spectroscopy. **Geoderma**, v. 232–234, p. 449–458, nov. 2014.

ST. LUCE, M. et al. Net nitrogen mineralization enhanced with the addition of nitrogen-rich particulate organic matter. **Geoderma**, v. 262, p. 112–118, 2016.

STENBERG, B. et al. Visible and Near Infrared Spectroscopy in Soil Science. In: SPARKS, D. L. (Ed.). . **Advances in Agronomy**. Burlington: Academic Press, 2010. v. 107p. 163–215.

STRAWN, D. G.; HOHN, H. L.; O'CONNOR, G. A. **Soil chemistry**. 5th Edition ed. [s.l.] Wiley Blackwell, 2020.

VISCARRA ROSSEL, R. A. et al. In situ measurements of soil colour, mineral composition and clay content by vis-NIR spectroscopy. **Geoderma**, v. 150, n. 3–4, p. 253–266, 2009.

VISCARRA ROSSEL, R. A. et al. On the soil information content of visible-near infrared reflectance spectra. **European Journal of Soil Science**, v. 62, n. 3, p. 442–453, 2011.

VISCARRA ROSSEL, R. A. et al. A global spectral library to characterize the world's soil. **Earth-Science Reviews**, v. 155, p. 198–230, 2016.

VISCARRA ROSSEL, R. A.; BEHRENS, T. Using data mining to model and interpret soil diffuse reflectance spectra. **Geoderma**, v. 158, n. 1–2, p. 46–54, 2010.

WESTAD, F.; MARTENS, H. Variable Selection in near Infrared Spectroscopy Based on Significance Testing in Partial Least Squares Regression. **Journal of Near Infrared Spectroscopy**, v. 8, n. 2, p. 117–124, 2 mar. 2000.

WORKMAN, J.; WEYER, L. **Practical guide to interpretive near-infrared spectroscopy**. Boca Raton: Taylor & Francis Group, 2008.

Conclusões finais

No presente estudo, foi demonstrado que os espectrofotômetros NIR compactos podem ser utilizados para determinar os teores de carbono orgânico do solo (COS) de forma acurada. Por aumentar a representatividade e a reprodutibilidade dos espectros NIR adquiridos com instrumentos compactos e proporcionar ganhos entre 146-400 % na predição de COS, o uso de um sistema de movimentação de amostras, capaz de movimentar o solo sobre o feixe de sondagem ótica durante a aquisição de espectros, é altamente recomendado.

Quando os espectrofotômetros NIR compactos foram utilizados para determinação de COS em escala regional, o modo de preparo das amostras de solo e as estratégias usadas no desenvolvimento dos modelos afetam diretamente a sua performance. O uso de amostras secas em estufa e finamente moídas resultou em modelos mais acurados, principalmente quando os espectros destas amostras foram utilizados no desenvolvimento de modelos de regressão local. O uso de modelos locais se mostrou uma boa alternativa para substituir a necessidade do processamento das amostras de solo antes da aquisição dos espectros. Para estudos futuros, recomenda-se a construção de bibliotecas espectrais, semelhantes às que são utilizadas com os equipamentos Vis-NIR, empregando o uso do FT-NIR compacto. Para a predição dos teores de COS usando estas bibliotecas, é recomendado o uso de modelos locais. Uma avaliação dos melhores critérios de similaridade e do número mínimo de amostras utilizadas nos modelos de regressão local, também é indicada.

Por fim, os resultados no presente estudo comprovam que, além de ser utilizada para mensurar os teores totais de C na matéria orgânica do solo (MOS), a espectroscopia NIR pode ser utilizada para quantificar o C e $\delta^{13}\text{C}$ em diferentes frações da MOS (matéria orgânica associada aos minerais e na matéria orgânica particulada) formada sob influência de resíduos enriquecidos com ^{13}C , uma vez que existem informações espectrais fundamentadas que sustentam o uso destes modelos. Estudos destinados a avaliar a performance de instrumentos NIR compactos para o monitoramento temporal do C nas frações da MOS, são recomendados.