

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE AGRONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DO SOLO

**AVALIAÇÃO DE MÉTODOS DE PRÉ-PROCESSAMENTO E MODELAGEM
MATEMÁTICA APLICADOS À ESPECTROSCOPIA DE INFRAVERMELHO
PRÓXIMO PARA ESTIMAR TEORES DE ARGILA E MATÉRIA ORGÂNICA
DO SOLO**

**Jacques Krticka Carvalho
(Dissertação de mestrado)**

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE AGRONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DO SOLO

**AVALIAÇÃO DE MÉTODOS DE PRÉ-PROCESSAMENTO E MODELAGEM
MATEMÁTICA APLICADOS À ESPECTROSCOPIA DE INFRAVERMELHO
PRÓXIMO PARA ESTIMAR TEORES DE ARGILA E MATÉRIA ORGÂNICA
DO SOLO**

Jacques Krticka Carvalho
Engenheiro Agrônomo (UFRGS)

Dissertação apresentada como um
dos requisitos à obtenção do Grau de
Mestre em Ciência do Solo

Porto Alegre (RS) Brasil
Março de 2019

CIP - Catalogação na Publicação

Carvalho, Jacques Krticka
AVALIAÇÃO DE MÉTODOS DE PRÉ-PROCESSAMENTO E
MODELAGEM MATEMÁTICA APLICADOS À ESPECTROSCOPIA DE
INFRAVERMELHO PRÓXIMO PARA ESTIMAR TEORES DE ARGILA E
MATÉRIA ORGÂNICA DO SOLO / Jacques Krticka Carvalho.
-- 2019.
65 f.
Orientador: Tales Tiecher.

Coorientadora: Amanda Posselt Martins.

Dissertação (Mestrado) -- Universidade Federal do
Rio Grande do Sul, Faculdade de Agronomia, Programa de
Pós-Graduação em Ciência do Solo, Porto Alegre, BR-RS,
2019.

1. Espectroscopia. 2. Carbono orgânico. 3.
Modelagem multivariada. 4. Quimiometria. 5.
Pre-processamento de dados espectrais. I. Tiecher,
Tales, orient. II. Martins, Amanda Posselt, coorient.
III. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os
dados fornecidos pelo(a) autor(a).

JACQUES KRTICKA CARVALHO
Engenheiro-Agrônomo (UFRGS)

DISSERTAÇÃO

submetida como parte dos requisitos para
a obtenção do grau de

MESTRE EM CIÊNCIA DO SOLO

Programa de Pós-Graduação em Ciência do Solo
Faculdade de Agronomia
Universidade Federal do Rio Grande do Sul
Porto Alegre (RS), Brasil

Homologado em: ____/____/____

Por

TALES TIECHER

Orientador-Departamento de
Solos/UFRGS

DEBORAH PINHEIRO DICK

Professora do Departamento
de Físico-Química/UFRGS

ELVIO GIASSON

Professor do Departamento de
Solos/UFRGS

JEAN MICHEL MOURA BUENO

Pós-Doutorando do Departamento
de Solos/UFPel

Aprovada em: ____/____/____

Pela Banca Examinadora

CARLOS GUSTAVO
TORNQUIST

Coordenador do PPGCS/UFRGS

CARLOS ALBERTO BISSANI

Diretor da Faculdade
Agronomia

"Tudo é ousado para quem nada se atreve."

Fernando Pessoa

AGRADECIMENTOS

Gostaria de agradecer a meus pais, João e Eliane, primeiramente pela vida, e também pelo apoio e suporte que me deram durante toda minha jornada, muito obrigado.

Ao meu orientador, Dr. Tales Tiecher, pelo apoio e tempo dispensado para compartilhar o conhecimento e auxiliar na elaboração deste trabalho, além da confiança e paciência.

Ainda, gostaria de agradecer ao laboratório de análises de solo da URI de Frederico Westphalen/RS, pela disponibilidade e pelas amostras fornecidas.

Gostaria de agradecer ao Daniel Fernando Hoppe, bolsista sem o qual este trabalho não seria possível, obrigado pela paciência e pelo tempo disponibilizado.

Aos professores e servidores da Faculdade de Agronomia, obrigado pelo conhecimento e suporte, o conhecimento é a base de tudo.

Gostaria de agradecer também aos meus colegas do programa de Pós-Graduação em Ciência do Solo da UFRGS, pela amizade e por compartilhar os momentos de alegria e tristeza.

Por fim, gostaria de agradecer a Deus, pela benção da vida e por tudo que me foi concebido durante esta longa jornada.

AVALIAÇÃO DE MÉTODOS DE PRÉ-PROCESSAMENTO E MODELAGEM MATEMÁTICA APLICADOS À ESPECTROSCOPIA DE INFRAVERMELHO PRÓXIMO PARA ESTIMAR TEORES DE ARGILA E MATÉRIA ORGÂNICA DO SOLO¹

Autor: Jacques Krticka Carvalho

Orientador: Prof. Tales Tiecher

Co-orientadora: Profa. Amanda Posselt Martins

RESUMO

A estimativa da matéria orgânica do solo (MOS) e argila por espectroscopia de infravermelho próximo (NIRS) é uma alternativa rápida e barata aos métodos de análise utilizados em laboratórios de rotina. Entretanto, a acurácia deste método é altamente afetada pelo uso adequado de técnicas de pré-processamento e métodos de análise multivariada. Ainda, os modelos são altamente específicos para diferentes locais, indicando a necessidade de se desenvolver calibrações para diferentes regiões. Esse estudo teve como objetivo avaliar o efeito dos métodos de pré-processamento e calibração na predição da MOS e argila utilizando a espectroscopia de infravermelho próximo. Um total de 2388 amostras de solo do norte do estado do Rio Grande do Sul e oeste de Santa Catarina, Brasil, foram analisadas em relação ao teor de argila e MOS, a refletância das amostras foi medida por meio de um espectrômetro de infravermelho próximo (1200-2400 nm). Sete técnicas de pré-processamento foram aplicadas ao espectro original, entre elas: derivada de Savitzki-Golay (SGD), *continuum removal* (CR), *detrend* (DET), *binning* (BIN), *smoothing* (SMO), e *standard normal variate* (SNV). *Multiple linear regression* (MLR), *partial least squares regression* (PLSR), *support vector machine* (SVM) e *gaussian process regression* (GPR) foram os modelos multivariados utilizados. A utilização de técnicas de pré-processamento melhorou a predição dos teores de MOS e argila. Ainda, quando comparada com os métodos de referência, a quantidade de acerto da espectroscopia foi de 85% para o teor de matéria orgânica e 78% no teor de argila, demonstrando performance satisfatória deste método analítico e com a vantagem de ser mais rápido, não destrutivo, barato e não gerar resíduos tóxicos, possibilitando uma análise com maior resolução espacial e temporal

Palavras-chave: Fertilidade do solo, carbono orgânico, pré-processamento espectral, espectroscopia do solo.

¹ Dissertação de Mestrado em Ciência do Solo. Programa de Pós-Graduação em Ciência do Solo, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul. Porto Alegre. (52p.) Março, 2019.

EVALUATION OF PRE-PROCESSING METHODS AND MATHEMATICAL MODELING APPLIED TO NEAR INFRARED SPECTROSCOPY TO ESTIMATE CLAY AND SOIL ORGANIC MATTER CONTENT ¹

Author: Jacques Krticka Carvalho

Advisor: Prof. Tales Tiecher

Co-advisor: Prof. Amanda Posselt Martins

ABSTRACT

The estimation of soil organic matter (MOS) and clay by near infrared spectroscopy (NIRS) is a fast and cost-effective alternative to conventional methods utilized in routine analysis. However, the accuracy of this method is highly affected by the adequate use of pre-processing techniques and multivariate methods. Furthermore, the models are highly specific for different locations, indicating the need to develop calibrations for different regions. This study aimed to evaluate the effect of preprocessing and multivariate calibration methods on the prediction of MOS and Clay using near infrared spectroscopy. A total of 2388 soil samples from the northern state of Rio Grande do Sul and western Santa Catarina, Brazil, were analyzed for clay content and SOM content, and the reflectance of the samples was measured by NIR-spectrometer (1200 - 2400 nm). Seven pre-processing techniques were applied to the original spectrum, including: Savitzki-Golay derivative (SGD), continuum removal (CR), detrend (DET), binning (BIN), smoothing (SMO), and standard normal variate (SNV). Multiple linear regression (MLR), partial least squares regression (PLSR), support vector machine (SVM) and gaussian process regression (GPR) were the multivariate models used. The use of pre-processing techniques improved the prediction of SOM and clay contents. As compared with the reference methods, the NIR spectroscopy classified correctly 85% for organic matter content and 78% for clay content, demonstrating satisfactory performance of this analytical method and with the advantage of being faster, non-destructive, cheaper and not generating toxic waste.

Keywords: Soil fertility, organic carbon, spectral preprocessing, soil spectroscopy.

¹Master's dissertation in Soil Science. Graduate Program in Soil Science, Faculty of Agronomy, Universidade Federal do Rio Grande do Sul. Porto Alegre. (52p.) March, 2019.

SUMÁRIO

1. INTRODUÇÃO.....	1
2. REVISÃO BIBLIOGRÁFICA	4
2.1 A radiação eletromagnética	4
2.2 Interação da radiação eletromagnética com a matéria	5
2.3 Espectroscopia de refletância no infravermelho próximo aplicada à análise do solo.....	6
2.4 Espectroscopia aplicada a análise do conteúdo matéria orgânica do solo	7
2.5 Espectroscopia aplicada a análise do conteúdo de argila do solo.....	9
2.6 Pré-processamento dos dados espectrais.....	11
3. HIPÓTESE.....	16
4. OBJETIVO.....	17
5. MATERIAL E MÉTODOS.....	18
5.1 Amostras de solo.....	18
5.2 Análise laboratorial e espectroscópica	19
5.3 Pré-processamento espectral.....	20
5.4 Modelos calibrados.....	21
5.5 Validação dos modelos e acurácia	22
6. RESULTADOS E DISCUSSÃO.....	25
6.1 Descrição estatística das amostras em relação ao teor de matéria orgânica e argila	25
6.2 Características das curvas espectrais	26
6.3 Conteúdo de matéria orgânica do solo	30
6.3.1 Performance das técnicas de pré-processamento nos modelos gerados	30
6.3.2 Performance dos modelos calibrados.....	33
6.3.3 Classificação dos teores preditos de MOS	36
6.4 Conteúdo de argila no solo.....	39
6.4.1 Performance das técnicas de pré-processamento nos modelos	39
6.4.2 Performance dos diferentes modelos de calibração.....	41
6.4.3 Classificação dos teores preditos de argila.....	43

7. CONCLUSÃO GERAL	46
8. CONSIDERAÇÕES FINAIS.....	47
9. REFERENCIAS	48

RELAÇÃO DE TABELAS

Tabela 1. Interpretação do teor de argila e de matéria orgânica.....	24
Tabela 2. Estatística descritiva dos teores de MOS e argila entre os grupos treinamento e validação.....	25
Tabela 3. Bandas correspondentes aos tons fundamentais no IR e seus sobretons e combinações no NIR	27
Tabela 4. Performance dos grupos de treinamento e validação em relação aos diferentes métodos de pré-processamento e calibração.....	30
Tabela 5. Porcentagem das amostras classificadas corretamente a partir do teor predito pela espectroscopia NIR em relação ao teor de MOS mensurado por digestão úmida.....	37
Tabela 6. Performance dos grupos de treinamento e validação em relação aos diferentes métodos de pré-processamento e calibração.....	40
Tabela 7. Porcentagem de classificação correta das amostras nas suas respectivas classes de Argila.....	44

RELAÇÃO DE FIGURAS

Figura 1. Número e distribuição das amostras no RS e SCI	19
Figura 2. Curvas espectrais das amostras de solo no comprimento de onda de 1200 até 2400.	26
Figura 3. Melhores performances da predição de MOS para os quatro modelos em relação a MOS determinada por digestão úmida	35
Figura 4. Melhores performances da predição de argila pelos quatro modelos e os respectivos métodos de pré-processamento em relação a argila medida pelo método do densímetro	43

RELAÇÃO DE ABREVIATURAS

MOS – Matéria orgânica do solo

NIR – Infravermelho próximo

NIRS – Espectroscopia de infravermelho próximo

SNV – *Standard Normal Variate*

BIN – *Binning*

ABS – Absorbância

DET – *Detrend*

CR – *Continuum removal*

SMO – Suavização

SGD – Derivada de Savitzky-Golay

SVM – *Support vector machine*

MLR – *Multiple linear regression*

PLSR – *Partial least squares regression*

GPR – *Gaussian process regression*

1. INTRODUÇÃO

A produção de alimentos é fundamental para manutenção da vida na terra. Estima-se que até 2050 a produção anual de alimentos no mundo precise aumentar em mais de 50% para que seja possível alimentar a população (ALEXANDRATOS; BRUINSMA, 2012). O solo é um recurso natural não renovável, do qual a humanidade depende para manutenção da segurança alimentar e da produção de fibra e energia, além de servir de fundação para a construção de habitações. Essencialmente, o solo é um corpo natural composto por matéria mineral e orgânica, água e ar (HILLEL, 2004). Se mal manejado, o solo como recurso natural pode ser degradado em um curto espaço de tempo, sendo necessárias medidas para sua preservação (HARTEMINK; MCBRATNEY, 2008). A urbanização e a degradação do solo também ameaçam os serviços ecossistêmicos por ele prestados, aumentando a necessidade de se monitorar e manejar o solo sustentavelmente.

A geração de dados sobre o solo, por um baixo custo e de rápida execução, é uma demanda da comunidade científica mundial, a fim de se monitorar os efeitos do aquecimento global, aumentar a produtividade de áreas agrícolas e a eficiência de uso dos insumos, e de se avaliar áreas contaminadas (VISCARRA ROSSEL *et al.*, 2011; NOCITA, 2016). A produção sustentável de alimentos depende do uso eficiente dos insumos, como fertilizantes e corretivos, e para isso também é necessária a geração de dados sobre o solo, com alta resolução espacial e de forma eficiente. A aplicação variável de nutrientes em uma área, por exemplo, requer uma grande quantidade de dados sobre o solo para que seja efetiva (STENBERG; JONSSON; BÖRJESSON, 2002). Além disso, o monitoramento das propriedades do solo a uma escala continental depende diretamente do desenvolvimento de novas técnicas de análise que

gerem resultados que possam ser comparados e armazenados de forma acessível (NOCITA, 2016), além de possibilitarem economicamente a análise de um grande número de amostras.

O conhecimento científico sobre as características físico-químicas e as funções do solo no ecossistema foi adquirido ao longo do tempo através de intensas pesquisas científicas e análises em laboratório (VISCARRA ROSSEL *et al.*, 2011). Os métodos convencionais de análise de matéria orgânica do solo e argila apresentam alta confiabilidade, porém requerem preparação das amostras, muitas vezes o uso de aparelhos complexos e geram uma série de resíduos químicos, o que dificulta a caracterização das propriedades do solo em áreas extensas (ARMENTA; DE LA GUARDIA, 2014; VISCARRA ROSSEL *et al.*, 2016). Apesar das desvantagens do uso dos métodos convencionais para a análise da MOS e argila, estes ainda continuam sendo utilizados como métodos de referência, o que evidencia a crescente necessidade de se desenvolver métodos analíticos alternativos, que possibilitem a obtenção de dados em grande quantidade, com qualidade e que possam ser reprodutíveis.

A espectroscopia de refletância de infravermelho próximo surge como uma alternativa rápida, de baixo custo e não destrutiva, que pode ser utilizada para a análise do solo. O crescimento do uso da espectroscopia pela comunidade científica pode ser demonstrado pelo aumento no número de artigos publicados sobre a utilização desta técnica analítica nas últimas décadas (GUERRERO; VISCARRA ROSSEL; MOUAZEN, 2010). A análise de solo por meio da espectroscopia NIR requer apenas a secagem e peneiramento das amostras, o que reduz drasticamente o tempo necessário para análise, permitindo processar uma maior quantidade de amostras (VISCARRA ROSSEL *et al.*, 2006).

A radiação eletromagnética no comprimento de onda do visível e no infravermelho pode ser utilizada efetivamente para caracterizar o solo, gerando informações qualitativas e quantitativas sobre o solo (VISCARRA ROSSEL *et al.*, 2016). O uso da espectroscopia baseia-se na detecção da radiação eletromagnética em comprimentos de onda específicos refletidos, emitidos ou espalhados pela matéria (CLARK, 1999). Isto se deve ao fato do comportamento espectral do solo refletir a combinação dos materiais mineral, orgânico e fluido que o compõe (STONER; BAUMGARDNER, 1981). Deste modo, com a

espectroscopia é possível se gerar um espectro do solo, que contém informações importantes sobre suas propriedades físicas, químicas e biológicas (MCBRATNEY; MINASNY; VISCARRA ROSSEL, 2006).

Apesar da espectroscopia ser uma técnica analítica promissora para análise do solo, este método analítico exige um tratamento matemático dos dados espectrais para a obtenção das informações desejadas. As técnicas de pré-processamento dos dados espectrais possuem a finalidade de remover qualquer informação indesejada, como ruídos, e realçar as características relacionadas com as propriedades estudadas.

2. REVISÃO BIBLIOGRÁFICA

2.1 A radiação eletromagnética

A radiação eletromagnética pode ser descrita por duas teorias: uma clássica, na qual a radiação eletromagnética é considerada um campo eletromagnético oscilante que se propaga na forma de ondas; e outra moderna, na qual as ondas eletromagnéticas são uma forma de energia dinâmica, composta por partículas microscópicas que carregam frações de energia (BENDOR, *et al.*, 1999; NIEMZ, 2007; BAUER, 2011).

As ondas eletromagnéticas podem ser classificadas pelo seu comprimento de onda, frequência ou número de ondas por unidade de distância. O comprimento de onda pode ser medido através da distância entre dois pontos equidistantes na onda e a frequência representa o número de ondas que passa por um ponto por unidade de tempo, medida em ciclos por segundo (Hertz) (NIEMZ, 2007). O espectro eletromagnético compreende todos os comprimentos de onda, das ondas menos energéticas, como as ondas de rádio, até as mais energéticas, como os raios gama (NIEMZ, 2007).

Além de ser descrita na forma de ondas, a radiação eletromagnética ainda pode ser considerada uma forma de energia dinâmica composta por partículas que carregam quantidades específicas de energia (BEN-DOR; IRONS; EPEMA, 1999). As partículas que carregam energia através das ondas eletromagnéticas são chamadas de fótons, sendo esta energia diretamente relacionada à frequência da radiação eletromagnética (NIEMZ, 2007). As ondas eletromagnéticas se propagam na velocidade da luz e sua energia pode ser calculada de acordo com a equação de Bohr, sendo $Energia = h \times \nu$, sendo h a constante de Planck (6.626×10^{-34} J s⁻¹) e ν a frequência (STUART, 2004).

Quanto maior a frequência e menor o comprimento de onda, maior a energia contida na onda eletromagnética.

2.2 Interação da radiação eletromagnética com a matéria

O fenômeno da interação da radiação eletromagnética com a matéria pode ser entendido através de uma série de efeitos físicos (LEROY, 2009). Esta interação depende da constituição da matéria e também da energia contida na radiação incidente (NIEMZ, 2007). As moléculas que formam a matéria são, por sua vez, formadas por átomos, que se ligam uns nos outros através da interação dos elétrons na camada de valência (NIEMZ, 2007).

A interação das ondas eletromagnéticas com a matéria pode ocorrer de diferentes formas. Dependendo da energia contida na radiação, a interação pode ser relativa à mudança de orbital energético dos elétrons ou aos processos vibracionais dos grupos funcionais presentes nas moléculas. Para a radiação eletromagnética promover um elétron de um orbital de energia para um orbital mais energético, é necessário que a radiação possua a quantidade de energia igual a diferença de energia entre os orbitais (STENBERG *et al.*, 2010). Ainda, a transição entre os orbitais deve ser acompanhada de uma redistribuição das cargas elétricas ou pela reorientação nuclear dos átomos (BEN-DOR; IRONS; EPEMA, 1999).

No caso de radiações menos energéticas, apesar de não ocorrer a transição eletrônica entre os orbitais, as ligações entre os átomos podem vibrar de diferentes formas, como no caso da radiação no comprimento de onda do infravermelho (NIEMZ, 2007). Isto ocorre porque os átomos, que compõem as moléculas, estão ligados quimicamente e em constante movimento, de modo que as moléculas se encontrem em uma frequência de energia vibracional fundamental, porém a amplitude desta vibração aumenta quando a molécula absorve energia radiante correspondente (PASQUINI, 2003).

2.3 Espectroscopia de refletância no infravermelho próximo aplicada à análise do solo

A espectroscopia de refletância no infravermelho próximo (NIR) é uma técnica analítica que se baseia no uso da radiação infravermelho próximo e sua interação com ligações das moléculas do solo. Após a radiação ser incidida na amostra de solo, esta radiação penetra na amostra, sofrendo múltiplas refrações e reflexões pelas partículas, resultando em uma radiação difusa, a qual é captada e pode prover informações sobre a composição do solo (ARMENTA; DE LA GUARDIA, 2014). Esta técnica analítica foi desenvolvida no início dos anos 1970, com a finalidade de analisar a umidade de grãos, entretanto se expandiu para outras áreas por sua eficiência e robustez (ZORNOZA; GUERRERO, 2008). O intervalo dos comprimentos de onda compreendidos pelo espectro eletromagnético do infravermelho próximo vai de 780 a 2500 nm (BLANCO; VILLARROYA, 2002).

Moléculas que absorvem energia na região do infravermelho próximo apresentam um momento dipolo elétrico associado a vibração ou rotação da molécula, o que serve como regra para selecionar os compostos que são infravermelho-ativos (STUART, 2004). A energia absorvida pela radiação incidente é resultado das ligações presentes nas moléculas do solo e também de fatores ambientais como o tipo de grupo funcional presente, moléculas adjacentes e presença de ligações de hidrogênio (ARMENTA; DE LA GUARDIA, 2014). A resposta espectral do solo para a radiação eletromagnética no comprimento de onda do NIR é resultado, principalmente, da interação da radiação com as moléculas H-O-H, metal-OH, O-H, C-H, N-H, S-H, C=O (NOCITA, 2016; ZORNOZA; GUERRERO, 2008).

Os modelos harmônicos e anarmônicos podem ser utilizados para descrever a vibração das moléculas (BLANCO; VILLARROYA, 2002). As bandas de absorção no infravermelho próximo podem ser classificadas como sobretons ou bandas de combinação. Os sobretons (780 a 2000 nm) são oriundos da transição entre níveis de energia vibracional não consecutivos e as bandas de combinação (1900 a 2500 nm) representam dois ou mais modos vibracionais atuando ao mesmo tempo (BLANCO; VILLARROYA, 2002).

A espectroscopia de refletância no infravermelho próximo, por exemplo, apresenta diversas vantagens em relação aos métodos convencionais de análise do solo, entre elas: (i) não há necessidade de preparação das amostras de solo, (ii) a possibilidade de se determinar diversos parâmetros do solo em alguns segundos, (iii) os menores custos para análise e (IV) não gera resíduos ambientais (ARMENTA; DE LA GUARDIA, 2014). Entretanto a espectroscopia NIR é um método indireto de análise do solo, o qual necessita de métodos de referência para calibração dos modelos utilizados para predição. O espectro do solo, obtido pela espectroscopia NIR é utilizado para estabelecer um modelo, no qual as informações significantes são concentradas em algumas variáveis, com a finalidade de se predizer, com melhor correlação possível, as propriedades do solo (ZORNOZA; GUERRERO, 2008).

2.4 Espectroscopia aplicada a análise do conteúdo matéria orgânica do solo

A matéria orgânica do solo (MOS) é um importante componente da matriz do solo, exercendo papel fundamental na fertilidade de solos agrícolas, armazenando e disponibilizando água e nutrientes para as plantas (LAL, 2014). Dada a sua importância, a MOS é frequentemente utilizada como um indicador de qualidade do solo (GLOBAL SOIL PARTNERSHIP, 2017), pois melhora uma série de propriedades químicas, físicas e biológicas, sendo um fator importante para a manutenção da produtividade do solo ao longo do tempo (JACKSON *et al.*, 2017; TIESSEN; CUEVAS; CHACON, 1994). Além disso, o sequestro de carbono por meio do acúmulo de MOS é uma estratégia, não somente para a manutenção da produtividade do solo e manutenção da segurança alimentar, mas também para mitigar o efeito estufa, evitando o aquecimento global (LAL, 2004). Como o conteúdo de MOS é diretamente afetado pelo cultivo e sistema de manejo empregado (JACKSON *et al.*, 2017), quando o solo é mal manejado, os estoques de carbono podem ser perdidos em um curto espaço de tempo (HARTEMINK; MCBRATNEY, 2008). Desta forma, o monitoramento do teor de MOS, com alta resolução espacial e temporal se torna indispensável (AMUNDSON *et al.*, 2015).

Além disso, por ser a principal fonte de nitrogênio no solo, a MOS é utilizada, no sul do Brasil (i.e. Rio Grande do Sul - RS e Santa Catarina - SC), como parâmetro para recomendação de adubação nitrogenada em diversas culturas. Nestas regiões, a avaliação da MOS é fundamental para a utilização eficiente dos fertilizantes nitrogenados na lavoura. Segundo o Manual de Adubação e Calagem dos estados do RS e SC (CRQFS/RS, 2016), após a análise do solo, o teor de MOS é classificado em alto ($>5,0\%$), médio ($2,6-5,0\%$) e baixo ($\leq 2,5\%$), sendo que quanto maior a quantidade de MOS menor a quantidade de fertilizante nitrogenado a ser utilizada. A geração de dados sobre a MOS, que representem a variabilidade espacial deste atributo, é de grande importância para o manejo sustentável do solo e utilização eficiente dos fertilizantes, particularmente para a agricultura de precisão, onde um grande volume de dados é necessário para aplicação variável de insumos (VISCARRA ROSSEL; MCBRATNEY, 1998). Entretanto, a avaliação do conteúdo de MOS por meio dos métodos convencionais (e.g. combustão seca e combustão úmida) é cara e trabalhosa, e ainda gera resíduos (VISCARRA ROSSEL *et al.*, 2006).

Nesse contexto, a demanda da comunidade científica mundial da Ciência do Solo, por métodos de análise que permitam a quantificação da MOS, de um grande número de amostras, motiva o desenvolvimento de técnicas analíticas alternativas. Neste sentido, a espectroscopia de infravermelho próximo (NIRS) surge como uma alternativa aos métodos convencionais, a qual permite a quantificação da MOS, e de várias propriedades do solo, de maneira rápida, não destrutiva, barata, sem uso de reagentes químicos e com pouca preparação de amostras (MCBRATNEY; MINASNY; VISCARRA ROSSEL, 2006; VISCARRA ROSSEL; BEHRENS, 2010; STENBERG *et al.*, 2010). A capacidade de modelar a MOS, por meio da espectroscopia NIR tem sido evidenciada em diversos estudos (GHOLIZADEH *et al.*, 2013; STENBERG *et al.*, 2010; VISCARRA ROSSEL *et al.*, 2006). A utilização desta técnica analítica se deve a interação da radiação NIR com as moléculas da MOS, uma vez que a energia contida em comprimentos de onda específicos da radiação NIR interagem com diferentes ligações dos átomos constituintes da matéria orgânica (STENBERG *et al.*, 2010). Dentro do espectro NIR, os comprimentos de onda entre 1100–2500 nm têm sido identificados como bandas particularmente importantes para a análise da MOS (BELLON-MAUREL *et al.*, 2010; STENBERG *et al.*, 2010).

Entretanto, apesar das vantagens comparativamente às metodologias convencionais utilizadas nos laboratórios de análise de solo de rotina, a espectroscopia é uma técnica analítica que requer o desenvolvimento de modelos matemáticos complexos para se estabelecer uma relação entre o espectro do solo e as propriedades do solo de interesse.

2.5 Espectroscopia aplicada a análise do conteúdo de argila do solo

O solo é um componente fundamental do planeta terra, o qual é responsável por diversos serviços ecossistêmicos, além de servir como meio para produção de alimentos, fibra e energia para a humanidade (STENBERG *et al.*, 2010). O solo é um corpo natural, caracterizado por ser um sistema trifásico, aberto, composto por uma fase sólida, uma líquida e uma fase gasosa. Sua fase sólida é composta por diferentes minerais e matéria orgânica do solo, sendo os minerais classificados em diferentes frações granulométricas. Dentre os tamanhos de partícula presentes no solo, a argila (0-2 μm de diâmetro) é a responsável por apresentar a maior área superficial específica e reatividade (HILLEL, 2004), sendo importante para diversos processos químicos e físicos do solo.

A argila é a fração granulométrica do solo que apresenta maior reatividade com o meio, isto se deve a sua maior área superficial específica, a qual confere a exposição de grupos funcionais de superfície nas arestas dos minerais, além das cargas permanentes oriundas da substituição isomórfica na estrutura dos minerais, as quais interagem com os íons presentes no solo (e.g. NH_4 , K^+ , Ca^{2+} , Mg^{2+} , etc.) (DIXON, 1991). A presença em alta quantidade desta fração confere ao solo características físicas como maior porosidade total, baixa permeabilidade e alta capacidade de retenção de água (KLEIN, 2014).

Em solos tropicais e subtropicais altamente intemperizados a maior quantidade de óxidos de Fe e Al na fração argila é associado a uma alta capacidade de adsorção de fósforo no solo. O teor de fósforo no solo, extraído por solução Mehlich-I, é interpretado conforme o teor de argila no solo, sendo importante para a recomendação de fertilizantes fosfatados para culturas de sequeiro (CQFS-RS/SC, 2016). Por isso, para dimensionar de forma correta a

adubação fosfatada em solos tropicais e subtropicais é imprescindível conhecer a variabilidade espacial do teor de argila do solo, principalmente para a agricultura de precisão, onde existe a demanda de um grande volume de dados sobre o solo para aplicação de insumos a taxa variável.

Entretanto, análise convencional do teor de argila, seja por hidrômetro ou por pipeta, demanda tempo para preparação de agentes e amostras, o que restringe o número de análises a algumas amostras ou amostras compostas, não representando a variabilidade espacial do teor de argila no solo. Nesse contexto, espectroscopia de infravermelho próximo surge como uma alternativa aos métodos de análise de rotina. Entretanto, como desvantagem, a espectroscopia necessita a geração de modelos matemáticos complexos que requerem a calibração com os métodos convencionais (SILVA, 2002).

A técnica de espectroscopia de refletância tem sido aplicada como alternativa para análises agronômicas rotineiras do solo desde o início dos anos 90 (DEMATTÊ *et al.*, 2018). Esta técnica se baseia na interação da energia contida na radiação eletromagnética com as ligações presentes nas moléculas do solo, e depende de fatores como o tipo de grupo funcional presente e natureza das moléculas adjacentes (ARMENTA; DE LA GUARDIA, 2014). Além disso, diversas propriedades físicas do solo podem afetar o espectro, tais como tamanho de partícula e estrutura do solo (ARMENTA; DE LA GUARDIA, 2014), por exemplo.

O espectro gerado pela espectroscopia NIR contém ruídos causados pelo espalhamento da luz, oriundos da difusão da radiação eletromagnética pelos diferentes tamanhos de partícula presentes no solo, apresentando variações sistemáticas (RINNAN, 2014), além de sobreposições de bandas de interesse. Devido à influência de tais não linearidades causadas pelo espalhamento da luz, as quais representam a maior parte da variação indesejada no espectro, a espectroscopia NIR é a técnica de análise que utiliza a maior diversidade de técnicas de pré-processamento espectral (RINNAN; BERG; ENGELSEN, 2009). Nesse contexto, a escolha adequada dos métodos de pré-processamento e calibração multivariada, pode melhorar a predição das propriedades estudadas pelo modelo gerado (GHOLIZADEH *et al.*, 2016).

O teor de argila do solo influencia as características das curvas espectrais devido ao aumento ou diminuição da absorção de energia relacionada

a ligações de hidroxilas e suas ligações com os minerais da fração Argila (STEVENS *et al.*, 2013), o que fundamenta quimicamente a sua análise quantitativa.

2.6 Pré-processamento dos dados espectrais

A espectroscopia de infravermelho requer uma série de tratamentos matemáticos com a finalidade de se extrair as informações desejadas. O espectro, além de apresentar informações importantes sobre a amostra, ainda pode apresentar variações sistemáticas que são causadas pelo espalhamento da radiação incidida. Estas variações são indesejadas e geralmente constituem a maior parte da variação da curva espectral (RINNAN; BERG; ENGELSEN, 2009).

A espectroscopia de infravermelho próximo é a técnica espectroscópica que utiliza a maior diversidade de técnicas de pré-processamento do espectro, devido à grande influência de não-linearidades no espectro gerado (RINNAN; BERG; ENGELSEN, 2009). A aplicação de métodos de pré-processamento adequados pode evitar que haja uma combinação entre as informações de interesse e dos ruídos presentes no modelo gerado (RINNAN, 2014). A escolha correta dos métodos de pré-processamento e análise multivariada pode resultar em predições mais confiáveis, devido a remoção de informações irrelevantes do espectro, realçando as características espectrais que representam as propriedades estudadas (GHOLIZADEH *et al.*, 2016). O pré-processamento dos espectros pode ser feito através de diversas técnicas, entre elas: Suavização, *Binning*, Absorbância, *Detrend*, Normalização, *continuum removal*, derivada de Savitzky-Golay e *standard normal variate*.

Suavização

Esta técnica de pré-processamento consiste de um filtro de média móvel dos dados espectrais que utiliza uma função de convolução (STEVENS; RAMIREZ LOPEZ, 2014). É uma das técnicas mais comuns utilizadas na espectroscopia, sendo utilizada para reduzir ruídos de fundo mantendo as características de interesse. Para a aplicação da suavização é necessário

escolher o número de pontos a serem suavizados. Este pré-processamento segue a seguinte equação:

$$S_i = \frac{1}{M} \sum_{j=0}^{M-1} x[i + j]$$

onde: x é o espectro original ($i = 1, 2, \dots$), S_i é o espectro após suavização e M o número de pontos usados na média móvel.

Binning

Esta técnica consiste na divisão dos dados espectrais em agrupamentos de mesmo tamanho, gerando novo espectro, representado pelos valores médios de cada bin (STEVENS; RAMIREZ LOPEZ, 2014). Para isso deve-se escolher o número de agrupamentos (BINS). Esta técnica é utilizada, principalmente, para grande volume de dados, como no caso de dados espectrais.

Absorbância

Esta técnica consiste na transformação dos dados de refletância para absorbância. A refletância pode ser transformada em absorbância através da seguinte equação:

$$A = \log_{10} \frac{1}{R}$$

onde: A é a absorbância e R a refletância.

Detrend

Este pré-processamento serve para remover ruídos presentes nos dados espectrais através da utilização da *standard normal variate* seguida de um modelo de regressão polinomial de segunda ordem para padronizar os dados (BARNES; DHANOA; LISTER, 1989).

Continuum removal

Esta técnica consiste na remoção contínua de feições indesejadas no espectro e possui a finalidade de diminuir o ruído no espectro isolando bandas de absorção específicas (CLARK; ROUSH, 1984). Esta técnica é aplicada através da identificação e seleção dos pontos máximos de refletância no espectro original, então estes pontos são interpolados linearmente para se obter o espectro transformado com a *continuum removal*.

Derivada de Savitzky-Golay

O algoritmo de Savitzky-Golay (SAVITZKY; GOLAY, 1964) é geralmente utilizado para derivação espectral. Derivadas são técnicas comuns de pré-processamento espectral utilizadas para remover ruídos não importantes através do cálculo da derivada das refletâncias em relação aos comprimentos de onda. Esta técnica requer a seleção do número de pontos a serem derivados, além da ordem do polinômio e da derivada. Esta técnica pode ser descrita pela seguinte equação:

$$x_i = \frac{1}{N} \sum_{-m}^m C_h x_{j+m}$$

onde: x_i é o espectro gerado, N é o coeficiente de normalização, m é o número de valores adjacentes em cada lado de j e C_h são coeficientes pré computados, dependendo da ordem polinomial e da derivada.

Standard normal variate

Esta técnica de pré-tratamento é geralmente utilizada para remoção da interferência por variação multiplicativa causada pelo espalhamento da radiação e também pela interferência no espectro devido ao tamanho de partícula (BARNES; DHANOA; LISTER, 1989). Pode ser obtida através da seguinte equação:

$$SNV_i = \frac{x_i - \bar{x}_i}{S_i}$$

onde: x_i representa a refletância original, \bar{x}_i é a média da refletância original, S_i é o desvio padrão da refletância original e SNV_i é o espectro resultante.

2.5 Calibração multivariada

A análise multivariada é uma técnica utilizada para análise no campo da quimiometria, que engloba o desenvolvimento de modelos matemáticos para a análise de dados complexos (CLEGG *et al.*, 2009). Diversos métodos são utilizados e tem como finalidade extrair a informação desejada de um grande número de dados espectrais (GHOLIZADEH *et al.*, 2016). A estimativa dos componentes do solo através da espectroscopia depende diretamente da seleção de métodos adequados de calibração multivariada (VOHLAND *et al.*, 2011). Algumas das técnicas de análise multivariada mais utilizadas na calibração multivariada são a *partial least squares regression* (PLSR), *multiple linear regression*, *support vector machine*, *random forest* e *Gaussian process regression* (CHEN; MORRIS; MARTIN, 2007; PASQUINI, 2003; VISCARRA ROSSEL; BEHRENS, 2010).

Partial least squares regression

Esta técnica de análise multivariada foi desenvolvida para o desenvolvimento de modelos matemáticos que apresentam relações complexas entre as variáveis resposta e preditora (WOLD; SJÖSTRÖM; ERIKSSON, 2001). *Partial least squares regression* é uma técnica amplamente utilizada na análise multivariada, a qual suporta a alta colinearidade, ruídos e a análise de inúmeras variáveis (GRINAND *et al.*, 2012; WOLD; SJÖSTRÖM; ERIKSSON, 2001), como no caso da análise de solo através da espectroscopia. Esta técnica de análise multivariada utiliza fatores ortogonais, também chamados de variáveis latentes, para prever as variáveis resposta, sendo as variáveis latentes combinações lineares das variáveis preditoras (NOCITA *et al.*, 2014).

Multiple linear regression

Este é o método mais simples de calibração multivariada. A *multiple linear regression* é uma técnica de análise estatística que utiliza várias variáveis para prever a variável resposta em um modelo linear (BALABIN; SAFIEVA; LOMAKINA, 2007). Este método assume que existe uma relação linear entre as variáveis, que não há colinearidades nos dados.

Support vector machine

Este método de análise multivariada se baseia no método de aprendizagem estatística “Kernel” (CORTES; VAPNIK, 1995). Utilizando esta técnica é possível derivar um plano espectral linear como uma função de decisão para problemas não lineares (VISCARRA ROSSEL; BEHRENS, 2010).

Gaussian process regression

Este é um método de modelagem não-linear, de aprendizagem de máquina, que utiliza regressões não-paramétricas através de processos gaussianos para gerar um modelo (CHEN; MORRIS; MARTIN, 2007). Este método funciona através da aplicação da função *kernel* para treinamento e predição, sendo a função *kernel* um algoritmo para análise padrão.

A composição do solo varia conforme os fatores de formação e os processos pedogenéticos atuantes. Assim varia também as características espectrais dos solos, as quais podem conter diferentes tipos de ruídos e interferências no sinal, sendo necessário escolher adequadamente as técnicas de pré-processamento para que seja possível se obter a melhor predição com os modelos matemáticos, evitando que informações irrelevantes se misturem com informações de interesse. O surgimento de novas técnicas de pré-processamento de dados espectrais, bem como novos métodos de modelagem matemática multivariada no campo da quimiometria, motivam novos trabalhos que visam melhorar a predição dos modelos gerados.

3. HIPÓTESE

A utilização de técnicas de pré-processamento e modelagem multivariada adequados pode produzir resultados com precisão adequada para a estimativa do teor de MOS e argila para laboratório de análise de solo de rotina.

4. OBJETIVO

Avaliar a espectroscopia de infravermelho próximo como técnica alternativa para análise dos teores de argila e matéria orgânica do solo, comparando diferentes métodos de modelagem multivariada e técnicas de pré-processamento espectral e comparar o resultado das predições com o método de referência em relação a classificação agronômica do conteúdo de argila e matéria orgânica do solo conforme o Manual de Adubação e Calagem do RS e SC.

5. MATERIAL E MÉTODOS

5.1 Amostras de solo

Um total de 2.388 amostras representativas da região noroeste do Rio Grande do Sul e Oeste de Santa Catarina foram analisadas, compreendendo 1.731 amostras distribuídas em 86 municípios do Estado do Rio Grande do Sul e 659 amostras de 13 municípios de Santa Catarina (Figura 1), sendo todas as amostras oriundas da camada superficial do solo (0-10 e 0-20 cm). As amostras foram coletadas por agricultores e técnicos e enviadas ao laboratório de análises de solo da Universidade Regional Integrada do Alto Uruguai e das Missões para fins de análise de rotina, sendo que foram selecionadas 2.388 amostras para este estudo. De acordo com a classificação climática de Köppen, na região predomina um clima subtropical úmido (Cfa) com chuvas bem distribuídas ao longo do ano e verões quentes. A região é caracterizada por sua diversidade no clima, relevo, vegetação e solos.

O local do estudo (Figura 1) apresenta várias classes de solos, predominantemente derivados de rochas de basalto e com a ocorrência de diferentes tipos de minerais, contendo quantidades consideráveis de Fe, Mg e Ca (TIECHER *et al.*, 2016). Os solos predominantes na região são os Neossolos, Chernossolos, Cambissolos, Latossolos e Nitossolos (TIECHER *et al.*, 2016). A maioria dos solos da região é utilizada para a agricultura, com diferentes tipos de manejo de solo e culturas.

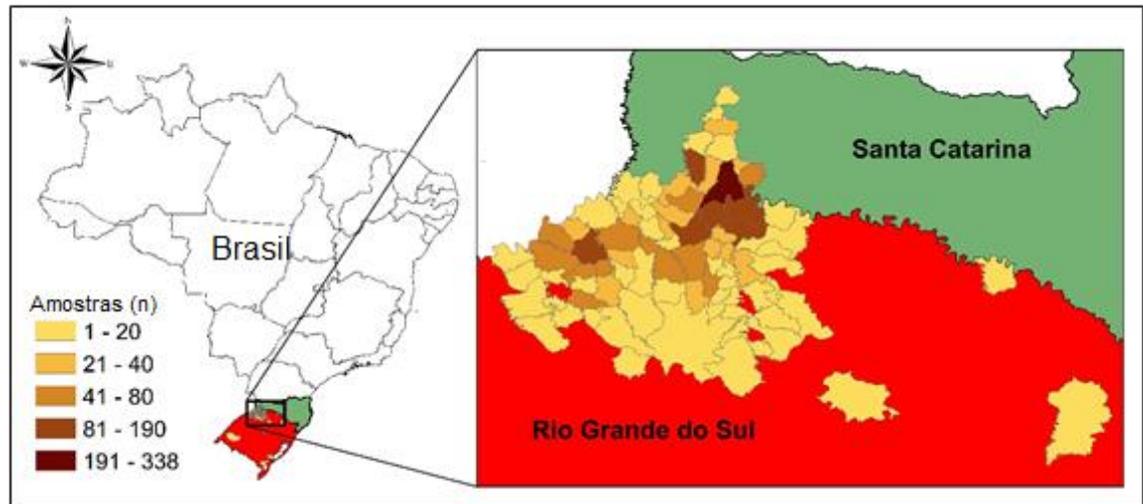


Figura 1. Distribuição das amostras nos Estados do Rio Grande do Sul e de Santa Catarina, Brasil.

5.2 Análise laboratorial e espectroscópica

As análises laboratorial e espectroscópica foram conduzidas no Laboratório de Análise de Solo da Universidade Regional Integrada do Alto Uruguai e das Missões (URI), sediada em Frederico Westphalen, RS, seguindo os métodos utilizados como na análise de rotina, sendo aceitos pela Sociedade Brasileira de Ciência do Solo (CQFS-RS/SC, 2016).

As amostras foram secas em estufa, com circulação forçada de ar, a 45-50°C. Após a secagem as amostras foram peneiradas a 2,0 mm. As amostras peneiradas foram submetidas, então, à análise do teor de carbono orgânico por combustão úmida pelo método Walkley-Black (WALKLEY; BLACK, 1934), sendo oxidadas com dicromato de potássio e sua posterior medida por colorimetria. Por fim, o teor de carbono orgânico foi transformado para matéria orgânica do solo pelo fator de conversão 1,72 (TEDESCO, 1995). Sendo o método mais utilizado em laboratórios da rede oficial de laboratórios de análise de solo no Rio Grande do Sul.

Para a análise do conteúdo de argila do solo, 50 g de cada amostra foram colocadas em frascos de vidro de 300 mL. Nos frascos foram colocados 250 mL de água destilada e 25 mL de hexametáfosfato de sódio (dispersante). Após foi realizada a dispersão por agitação horizontal durante 5 horas. A suspensão foi transferida para provetas de 1000 mL, passando-se pela peneira com malha de 0,053 mm para separação da fração areia. A solução peneirada

foi agitada na proveta por 1 minuto e, após decantação da fração silte, foi medida a densidade da solução (água + solo + dispersante) utilizando o densímetro de Bouyoucos.

Para a análise espectroscópica, amostras previamente secas ao ar e peneiradas foram colocadas em placas de petry e inseridas no espectrômetro NIR, localizado em sala escura. A refletância das amostras foi obtida com a utilização do espectrômetro NIR Unity Spectra Star 2400®. O aparelho apresenta uma resolução espectral de 1 nm e trabalha com o intervalo de comprimentos de onda de 1.200 a 2.400 nm. As amostras foram analisadas uma a uma e foi gerada uma biblioteca espectral composta de 2.388 assinaturas espectrais.

5.3 Pré-processamento espectral

O espectro NIR é altamente afetado pela granulometria e pelo espalhamento da radiação, sendo o espalhamento da radiação dependente de vários fatores, entre eles, a natureza das partículas, o tamanho e sua organização na amostra (BARNES; DHANOA; LISTER, 1989). Existe uma grande diversidade de técnicas de pré-processamento utilizadas para melhorar a predição dos modelos calibrados a partir de dados espectrais. Para o tratamento dos dados espectrais, neste estudo, foram utilizadas sete técnicas de pré-processamento, além da refletância original (RAW) sem qualquer pré-processamento, utilizada como controle. Os 8 diferentes espectros foram comparados conforme a sua performance nos diferentes modelos de calibração multivariada para predição da MOS e argila. As sete técnicas de pré-processamento utilizadas neste trabalho foram: Suavização (SMO), *Continuum removal* (CR), *Detrend* (DET), *Binning* (BIN), *Standard Normal Variate* (SNV), Absorbância (ABS) e derivada de Savitzky-Golay (SGD). Todas técnicas de pré-processamento foram calculadas no ambiente do software R (R CORE TEAM, 2017). Os pré-processamentos foram aplicados nas curvas espectrais das 2.388 amostras de solos no intervalo de comprimento de onda de 1.200 a 2.400 nm.

O pré-processamento SMO consiste em uma média móvel, a qual utiliza uma função de convolução para suavizar o espectro, sendo necessário definir o número de pontos a serem suavizados (STEVENS; RAMIREZ LOPEZ, 2014), neste estudo o SMO foi aplicado com 11 pontos de suavização, devido

ter o melhor resultado em relação a utilização de outros valores testados. CR é uma técnica de pré-processamento utilizada para remover as feições contínuas do espectro e realçar picos de absorção, minimizando o ruído, por meio da identificação de pontos máximos no espectro e então conectando-os por interpolação linear (CLARK; ROUSH, 1984). A função DET foi aplicada por meio da normalização do espectro pela aplicação de SNV ajustado a um modelo de regressão polinomial de segundo grau. O pré-processamento SGD foi aplicado aos dados espectrais ajustados para um polinômio de primeira ordem sobre uma janela móvel de 5 nm, após testes com outros tamanhos de janela. O SNV foi aplicado em cada espectro subtraindo a média da refletância do espectro e escalando com o desvio padrão do espectro. O pré-processamento BIN foi calculado pelo no agrupamento de intervalos de refletância e substituição por sua média no espectro. Neste estudo Binning foi aplicado com tamanhos de BIN igual a 2, para evitar a remoção de informações de interesse no espectro. O pré-processamento ABS consistiu da transformação da refletância (R) para absorbância por meio da expressão $\log(1/R)$. Todos os pré-processamentos foram aplicados por meio de do software R (R CORE TEAM, 2017)

5.4 Modelos calibrados

Para avaliar o efeito das diferentes técnicas de pré-processamento na performance preditiva da MOS e argila, foram utilizados quatro modelos matemáticos. Dois modelos lineares: PLSR e MLR e dois modelos baseados em algoritmos de aprendizagem de máquina: SVM e GPR. Cada um dos espectros preprocessados pelas sete técnicas foi utilizado em conjunto com cada um dos quatro métodos de calibração multivariada, resultando em 28 combinações. Além destes, foram gerados quatro modelos preditivos com os dados espectrais originais, os quais foram utilizados como controle. Todos os modelos foram gerados por meio do uso da interface gráfica Alrad spectra (DOTTO, 2017) no ambiente do software R (R CORE TEAM, 2017).

Entre os modelos lineares o MLR é um modelo estatístico mais simples, que utiliza diversas variáveis explanatórias para predizer os resultados de uma variável resposta, em um modelo linear. Este modelo assume que a

relação entre as variáveis independentes e dependentes são lineares e que não há multicolinearidades (VARMUZA & FILZMOSE, 2009).

Já o modelo PLSR é um modelo mais complexo o qual é utilizado para lidar com relações complexas entre as variáveis preditoras e as variáveis resposta, sendo geralmente utilizado para gerar modelos em que os dados apresentam um alto grau de colinearidade (WOLD; RUHE; DUNN, 1984). Este modelo tem sido amplamente utilizado para análise quantitativa de dados espectrais, sendo o modelo mais utilizado (DOTTO *et al.*, 2018).

SVM é um algoritmo de aprendizagem de máquina, o qual consiste em um método supervisionado, de análise estatística não paramétrica, que apresenta a capacidade de lidar com um grande número de dados e com os ruídos contidos no espectro, como é o caso dos dados oriundos da espectroscopia. Este modelo é baseado na teoria do aprendizado estatístico, a qual utiliza a função kernel para projetar os dados num novo hiperplano, onde padrões não lineares podem ser representados (CORTES; VAPNIK, 1995).

Assim como SVM, GPR é um modelo de aprendizagem de máquina, que utiliza uma função kernel para predição dos dados de interesse. Este algoritmo se baseia na perspectiva não paramétrica Bayesiana para a formulação do modelo de regressão, sendo capaz de gerar predições confiáveis tanto para dados não lineares quanto dados lineares (CHEN; MORRIS; MARTIN, 2007).

5.5 Validação dos modelos e acurácia

As 2.388 amostras foram divididas ao acaso em dois grupos, sendo treinamento [~70%] ($n = 1.672$), para calibração dos modelos, e validação [~30%] ($n = 716$), a qual foi utilizada para avaliar a performance preditiva dos modelos. A acurácia dos modelos em relação a predição da MOS e argila, bem como a performance dos diferentes métodos de pré-processamento dos dados espectrais, foi avaliada por meio do coeficiente de determinação (R^2), o qual representa porcentagem da variância explicada pelo modelo, e a raiz quadrada do erro médio (RMSE), que é comumente utilizada para medir a diferença entre os valores observados e os valores preditos pelos modelos. Ainda, foi calculada a razão da diferença interquartil pelo RMSE (RPIQ – *ratio of*

performance/prediction to interquartile), com a finalidade de melhor representar o espalhamento dos dados pela diferença interquartis, em relação ao RMSE (BELLON-MAUREL *et al.*, 2010), sendo quanto maior o seu valor, melhor será a capacidade preditiva do modelo. Ainda, foi realizada a descrição estatística dos conjuntos de dados de argila e MOS. Foi calculado o teste de normalidade dos dados de Shapiro Wilk, bem como o coeficiente de variância e o desvio padrão.

5.6 Classificação dos teores de matéria orgânica do solo e argila

Após a geração dos modelos, os valores preditos de MOS foram classificados conforme o manual de adubação e calagem do RS e SC (CQFS-RS/SC, 2016) em baixo ($\leq 2,5\%$), médio (2,6-5,0%) e alto ($>5,0\%$) (Tabela 1), para fins de recomendação de adubação nitrogenada, como no caso da análise de rotina. O nível de acerto dos valores de MOS preditos pelos modelos dentro das classes estabelecidas no manual foi comparada com a classificação obtida para os valores quantificados por meio de digestão úmida (WALKLEY; BLACK, 1934).

Para o conteúdo de argila, as predições obtidas foram utilizadas para interpretação das classes de argila conforme o manual de adubação e calagem do RS e SC (CQFS-RS/SC, 2016), conforme os teores estabelecidos para as diferentes classes (Tabela 1), para fins de recomendação de adubação fosfatada, como no caso da análise de rotina.

Tabela 1. Interpretação do teor de argila e de matéria orgânica do solo conforme o manual de adubação e calagem do RS e SC (CQFS-RS/SC, 2016).

Interpretação do teor de argila	Classe	Argila (%)
	1	>60
	2	41-60
	3	21-40
	4	≤20
Interpretação do teor de MOS	Classe	MOS (%)
	Alto	>5,0
	Médio	2,6-5,0
	Baixo	≤2,5

O resultado da interpretação dos teores de argila preditos pelos modelos, conforme o Manual de Adubação e Calagem do RS e SC (CQFS-RS/SC, 2016), foram comparados com o resultado da interpretação feita com os dados obtidos através do método do densímetro de Bouyoucos. O resultado da interpretação dos valores preditos por meio dos modelos multivariados em relação aos obtidos através da digestão úmida foi utilizado para averiguar o desempenho desta técnica para recomendação de adubação e calagem.

6. RESULTADOS E DISCUSSÃO

6.1 Descrição estatística das amostras em relação ao teor de matéria orgânica e argila

O conjunto das amostras contempla uma variação no teor de matéria orgânica (Tabela 2), sendo a média igual a 2,49%. As amostras apresentaram uma distribuição normal, sendo que mais de 90% das amostras apresentou um teor de MOS entre 0,4 e 5,0%. O conjunto de amostras apresentou um coeficiente de variação de 33%, e segundo o teste de normalidade de Shapiro Wilk os dados não apresentaram distribuição normal. Os conjuntos amostras de treinamento (70%) e validação (30%) apresentaram a média e desvio-padrão similares entre os conjuntos, indicando que possuem homogeneidade da variância.

Tabela 2. Estatística descritiva dos teores de MOS e argila entre os grupos treinamento e validação.

Amostras	Treinamento		Validação		Total	
	Argila (%)	MOS (%)	Argila (%)	MOS (%)	Argila (%)	MOS (%)
Número de amostras	1672	1672	716	716	2388	2388
Mínimo	7	0,4	8	0,4	7	0,4
Máximo	92	12	85	7,9	92	12
Média	36	2,49	36	2,49	36	2,49
Mediana	33	2,4	34	2,4	34	2,4
Desvio padrão	15,9	0,86	15,5	0,80	15,8	0,84

Já para o teor de argila, o conjunto das amostras apresentou grande variação no teor de argila (Tabela 2), apresentando a mesma média entre os conjuntos (36%), indicando similaridade entre os conjuntos de dados. Foi

realizado o teste de normalidade de Shapiro Wilk os dados não apresentaram distribuição normal.

6.2 Características das curvas espectrais

As assinaturas espectrais do conjunto de amostras com os espectros sem pré-processamentos (*Raw spectra*) e com aplicação dos sete processamentos espectrais são apresentadas na Figura 2.

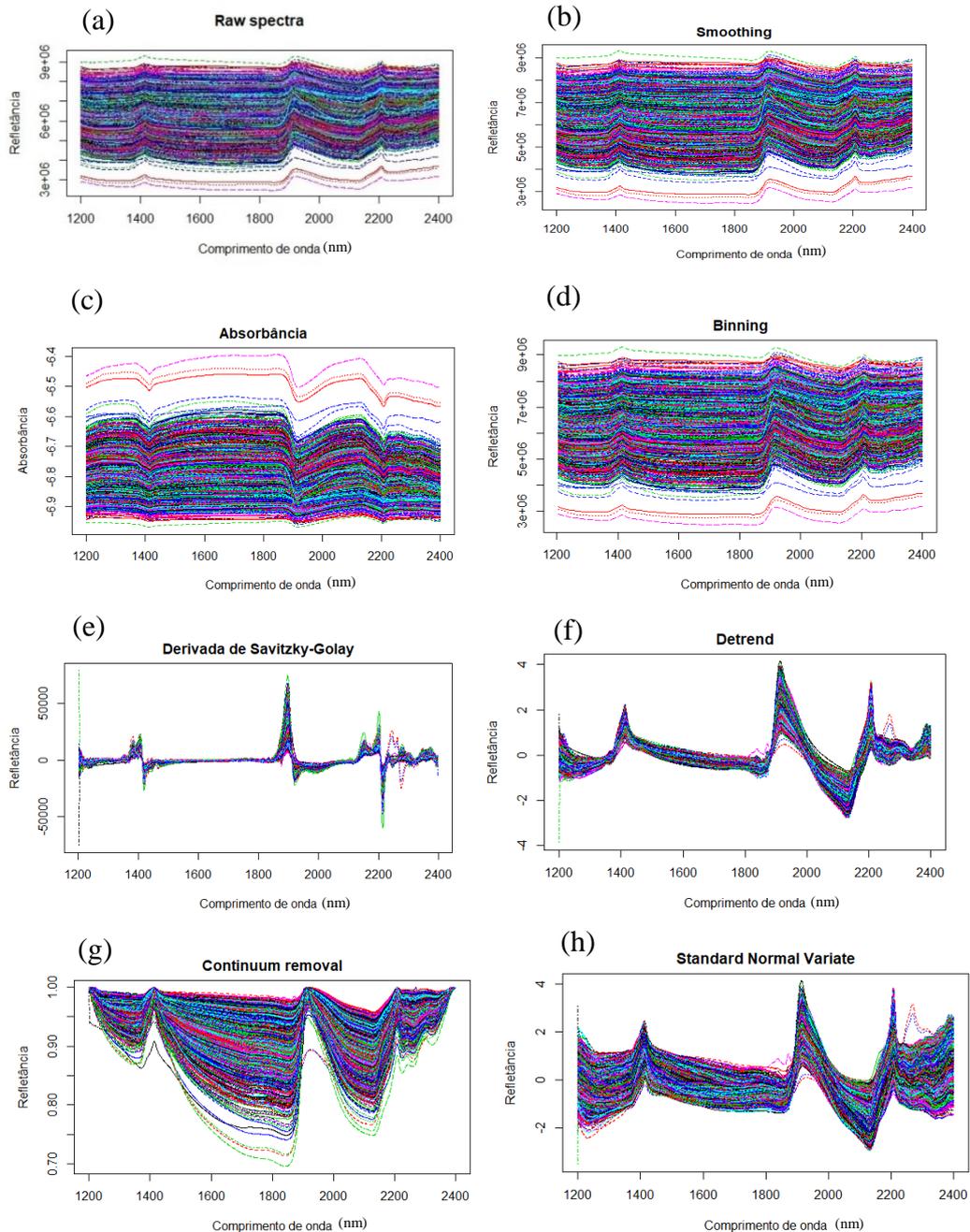


Figura 2. Curvas espectrais das amostras de solo no comprimento de onda de 1200 até 2400 nanômetros. a) Raw spectra. b) Suavização. c) Absorbância. d) Binning. e) Savitzky-Golay. f) Detrend; g) Continuum removal; h) Standard Normal Variate.

As curvas espectrais mostram uma alta variação na reflectância, indicando que existe uma variação composicional dos constituintes orgânicos e inorgânicos dos solos. As curvas de reflectância demonstram, de acordo com sua forma e feições, a relação dos constituintes do solo com as bandas da região do NIR. A forma da curva espectral é fortemente influenciada pela composição organomineral dos solos, principalmente pela textura, teor óxidos de ferro e de matéria orgânica (DALMOLIN *et al.*, 2005; STONER; BAUMGARDNER, 1981). Portanto, a avaliação qualitativa do espectro fornece informações relevantes sobre a composição das amostras, indicando possíveis bandas do espectro NIR correlacionadas com os constituintes organominerais e suas ligações relacionadas às moléculas da MOS. As assinaturas espectrais no comprimento de onda NIR são principalmente compostas por feições causadas por sobretons fracos e combinações de bandas de vibração fundamental para H-C, H-N e ligações O-H (VISCARRA ROSSEL *et al.*, 2006), bem como metal-OH, presente nos minerais do solo (Tabela 3).

Tabela 3. Bandas correspondentes a sobretons e combinações ocorrentes no NIR (VISCARRA ROSSEL; BEHRENS, 2010).

Constituinte do solo	Sobretons e combinações de bandas NIR (nm)	Modo vibracional NIR
Minerais de argila		
Água	1915	$\nu_2 + \nu_3$
	1455	$2\nu_2 + \nu_3$
	1380	$\nu_1 + \nu_3$
Hidroxilas	1400	$2\nu_1$
Caulinita	1395	$2\nu_{1a}$
	1415	$2\nu_{1b}$
	2160	$\nu_{1a} + \delta$
Esmectita	2208	$\nu_{1b} + \delta$
	2206	$\nu_1 + \delta_a$
	2230	$\nu_1 + \delta_b$
Ilita	2206	$\nu_1 + \delta$
	2340	
Carbonatos	2336	$3\nu_3$

(cont.)		
Grupos funcionais orgânicos		
Aromáticos	1650	$2\nu_1$
Amina	2060	$\nu_1 + \delta$
	1500	$2\nu_1$
Alquil	1706	$2\nu_3$
	1754	$2\nu_1$
Ácido carboxílico	1930	$3\nu_1$
	1449	$4\nu_1$
Amidas	2033	$3\nu_1$
	1524	$4\nu_1$
Alifáticos	2275	$3\nu_1$
	1706	$4\nu_1$
Metil	2307-2469	$3\nu_1$
	1730-1852	$4\nu_1$
Fenóis	1961	$4\nu_1$
Polissacarídeos	2137	$4\nu_1$
Carboidratos	2381	$4\nu_1$

As curvas espectrais que apresentam baixa reflectância (Figura 2a), em geral, indicam que são amostras com alto teor de MOS e óxidos de ferro (DALMOLIN *et al.*, 2005). As bandas de absorção presentes no comprimento de onda de 1400 e 1900 nm são principalmente devido a ligações O-H e metal-OH (Tabela 3), presente nos minerais (DALMOLIN *et al.*, 2005; VISCARRA ROSSEL *et al.*, 2006). Diversas bandas de absorção foram apresentadas Viscarra & Behrens (2010), para os constituintes do solo e seus respectivos modos de vibração (Tabela 3). Ainda, picos de absorção associados as ligações C-H são relacionados com os comprimentos de onda entre 2250 e 2450 nm, presentes na MOS (Tabela 3) (BEN-DOR; INBAR; CHEN, 1997). Variações nos teores de argila das amostras podem causar diferenças nas feições presentes no espectro, devido ao aumento ou diminuição da absorção de energia relacionada a presença de hidroxilas e suas ligações com os minerais da fração argila (STEVENS *et al.*, 2013). Nota-se que algumas dessas feições relacionadas aos constituintes do solo ficaram mais realçadas nos espectros submetidos aos pré-processamentos SGD, DET, CR e SNV, com maior destaque para a técnica SGD (Figura 2e), comparado aos dados sem qualquer pré-processamento (Figura 2a). Além disso, as curvas espectrais mostram que existe uma variação nos constituintes organominerais dos solos, pois se observa variação na reflectância nas regiões de 1400 e 1900 nm, 2200 nm e entre 2250 e 2450 nm (Figura 2a),

onde ocorreram os picos de menor absorbância e menor refletância. Os comprimentos de onda de próximos a 1400 nm (1395 e 1415 nm) são devidos a sobretons, enquanto aqueles perto de 2200 nm (2165 e 2207 nm) são devidos a Al–OH e combinação de O–H. Esmeclitas apresentam grupos funcionais fortes absorções próximas de 1400, 1900, e entre 2200 e 2400 nm (Tabela (STENBERG, 2010)). Essas bandas são semelhantes às aquelas relatadas por Madejova e Komadel (2001) para distinguir diferentes minerais argilosos, como montmorilonita e caulinita, por exemplo.

A alta presença de não-linearidades e multicolinearidades no espectro NIR, causadas pelo espalhamento da luz e diferentes tamanhos de partícula (RINNAN; BERG; ENGELSEN, 2009) torna a avaliação qualitativa do espectro NIR menos precisa devido a sobreposição de bandas (CHANGWEN *et al.*, 2015; VISCARRA ROSSEL *et al.*, 2006). No entanto, o espectro NIR das amostras da região estudada apresentou diferenças importantes em relação aos picos de absorção relacionados aos grupos funcionais presentes nos componentes do solo. Assim, a análise quantitativa com a combinação de diferentes técnicas de processamento espectral e métodos de calibração multivariada pode melhorar o estabelecimento de relações entre os constituintes do solo e sua assinatura espectral e, conseqüentemente, gerar modelos para a predição do conteúdo de MOS e de argila mais acurados.

6.3 Conteúdo de matéria orgânica do solo

6.3.1 Performance das técnicas de pré-processamento nos modelos gerados

Apesar dos dados espectrais apresentarem informações qualitativas importantes sobre cromóforos presentes no solo, a presença de sobreposição de bandas presentes nas curvas espectrais NIR influencia negativamente na análise espectral quantitativa (RINNAN; BERG; ENGELSEN, 2009). Porém, tais feições podem ser realçadas para que haja uma relação mais consistente com as propriedades de interesse. Isso pode ser feito por meio de métodos de pré-processamento, seja por realce das bandas relacionadas com as propriedades estudadas ou pela remoção de ruídos causados por tamanho de partícula e espalhamento da luz. Neste estudo foram avaliados sete métodos de pré-processamento em relação a performance de modelos de predição para o conteúdo de MOS. Os resultados demonstraram que as diferentes técnicas de pré-processamento apresentaram performances distintas em relação a predição dos teores de MOS (Tabela 4), melhorando ou piorando as predições dos modelos em relação aos dados espectrais sem aplicação de pré-processamento.

Tabela 4. Performance dos grupos de treinamento e validação em relação aos diferentes métodos de pré-processamento e calibração.

Modelo	Pré-processamento	Treinamento			Validação		
		R ²	RMSE	RPIQ	R ²	RMSE (%)	RPIQ
MLR	SNV	0,62	0,53	1,71	0,64	0,49	2,06
	Detrend (DET)	0,63	0,53	1,71	0,63	0,49	2,05
	Suavização (SMO)	0,62	0,53	1,70	0,63	0,49	2,05
	Raw data (RAW)	0,61	0,54	1,67	0,61	0,50	1,99
	Absorbância (ABS)	0,58	0,55	1,63	0,59	0,52	1,94
	Cont. removal (CR)	0,57	0,57	1,59	0,56	0,53	1,89
	1º derivada de Savitzky-Golay (SGD)	0,56	0,57	1,58	0,55	0,54	1,84
	Binning (BIN)	0,15	0,79	1,13	0,13	0,75	1,33
PLSR	Detrend (DET)	0,68	0,49	1,84	0,66	0,47	2,12
	SNV	0,68	0,48	1,86	0,66	0,47	2,12
	1º derivada de Savitzky-Golay (SGD)	0,71	0,46	1,94	0,65	0,48	2,10

(cont.)							
	Suavização (SMO)	0,67	0,49	1,83	0,65	0,48	2,10
	Raw data (RAW)	0,66	0,50	1,80	0,65	0,48	2,10
	Binning (BIN)	0,67	0,50	1,82	0,64	0,48	2,08
	Absorbância (ABS)	0,64	0,51	1,75	0,63	0,49	2,03
	Cont. removal (CR)	0,69	0,48	1,89	0,62	0,50	2,00
SVM	SNV	0,72	0,46	1,97	0,70	0,44	2,26
	Detrend (DET)	0,73	0,45	2,00	0,70	0,45	2,24
	Binning (BIN)	0,57	0,56	1,59	0,61	0,51	1,97
	Cont. removal (CR)	0,68	0,49	1,84	0,60	0,52	1,94
	Raw data (RAW)	0,57	0,57	1,59	0,59	0,52	1,92
	Suavização (SMO)	0,57	0,57	1,58	0,59	0,52	1,92
	Absorbância (ABS)	0,53	0,59	1,52	0,57	0,54	1,87
	1º derivada de Savitzky-Golay (SGD)	0,84	0,35	2,60	0,52	0,61	1,64
GPR	Detrend (DET)	0,75	0,43	2,09	0,67	0,46	2,16
	SNV	0,68	0,48	1,86	0,67	0,47	2,15
	Cont. removal (CR)	0,66	0,50	1,78	0,61	0,50	1,98
	Raw data (RAW)	0,57	0,56	1,60	0,57	0,53	1,89
	Suavização (SMO)	0,57	0,56	1,59	0,57	0,53	1,88
	Absorbância (ABS)	0,54	0,58	1,55	0,55	0,55	1,83
	1º derivada de Savitzky-Golay (SGD)	0,88	0,29	3,06	0,52	0,61	1,64
	Binning (BIN)		0,50	0,61	1,47	0,45	0,60 1,68

*MLR: Multiple linear regression. *PLSR: Partial least squares regression. *SVM: Support vector machine. *GPR: Gaussian Process Regression. *R²: Coeficiente de determinação *RMSE: Raiz quadrada do erro médio. *RPIQ: Relação De Desempenho para distância interquartis; Standard normal variate SNV

A melhor performance na validação entre os sete métodos de pré-processamento foi encontrada para o modelo SVM em conjunto com os dados espectrais pré-processados por SNV ($R^2 = 0,70$, $RMSE = 0,44$ e $RPIQ = 2,26$). Ainda pode-se ressaltar que quando foi utilizado o pré-processamento SGD em conjunto com SNV e GPR houve um sobreajuste dos modelos em relação ao conjunto de dados de treinamento, o que não se confirmou na validação. Quando o modelo linear MLR foi utilizado, somente SNV, SMO e DET melhoraram a predição da MOS em relação ao modelo gerado com os dados sem qualquer pré-processamento, sendo SNV o pré-processamento que apresentou maior acurácia ($R^2 = 0,64$, $RMSE = 0,49$ e $RPIQ = 2,06$). Para o modelo PLSR, a predição da MOS foi melhorada com o pré-processando DET e SNV, os quais apresentaram mesma acurácia ($R^2 = 0,66$, $RMSE = 0,47$ e $RPIQ = 2,12$), quando comparada com a predição utilizando os dados originais ($R^2 = 0,65$, $RMSE = 0,48$ e $RPIQ = 2,10$). Em contraste, as técnicas de pré-processamento BIN, CR,

ABS and SGD não melhoraram a predição da MOS para os dois modelos lineares, quando comparados com a utilização do espectro original.

Ainda, a performance utilizando SVM combinado com DET ($R^2 = 0,70$, $RMSE = 0,45$ e $RPIQ = 2,24$), melhorou consideravelmente a predição em relação a modelagem com os dados originais. Os pré-processamentos BIN e CR também melhoraram as predições do modelo SVM quando comparados com o modelo baseado nos dados originais. Para o modelo GPR, as melhores predições foram obtidas para o pré-processando os dados espectrais com DET, SNV e CR.

Em geral, as melhores predições observadas foram alcançadas modelando os dados preprocessados por SNV, isto pode ser devido a remoção dos efeitos de espalhamento da radiação presentes nos dados originais. Para os modelos gerados a partir dos dados preprocessados pela técnica DET, as predições foram similares as observadas utilizando SNV, a causa disso é a similaridade entre os dois tratamentos matemáticos dos dados espectrais (RINNAN; BERG; ENGELSEN, 2009) resultando em efeitos similares nos dados espectrais. Os modelos baseados nos espectros preprocessados por ABS e SGD não melhoraram a predição de nenhum dos modelos utilizados quando comparados aos dados sem qualquer pré-processamento.

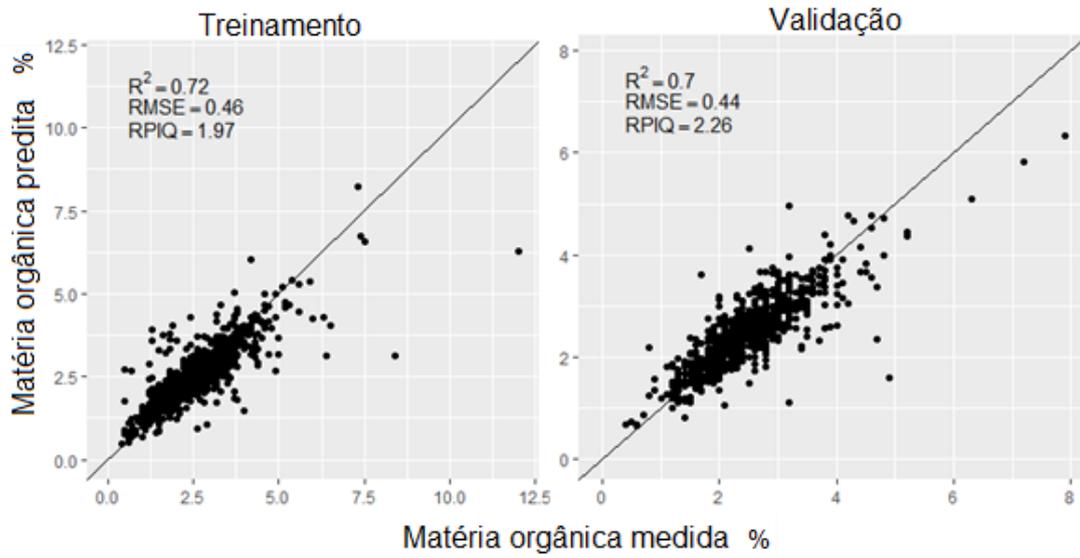
Essa variação na acurácia dos modelos de predição em relação aos tipos de pré-processamentos utilizados também foi observada em estudos de Vasquez *et al.* (2008), que compararam quatro modelos e 30 pré-processamentos na construção de modelos preditivos de carbono orgânico na Flórida, EUA, encontrando a melhor predição por meio da utilização de PLSR, entre outros modelos multivariados. Da mesma forma, Moura-Bueno *et al.* (2019) encontraram variação na performance dos modelos por PLSR, MLR, SVM e Random Forest ao comparar quatro pré-processamentos em uma biblioteca espectral na região sul do Brasil. Portanto, os resultados observados na literatura e encontrados no presente estudo, demonstram que existe uma variação nos resultados observados, independentemente de qual pré-processamento ou técnica multivariada é empregada. Porém, deve-se destacar que os modelos multivariados podem apresentar maior similaridade por alguma técnica de pré-

processamento específica, corroborando com o estudo de Moura-Bueno *et al.* (2019).

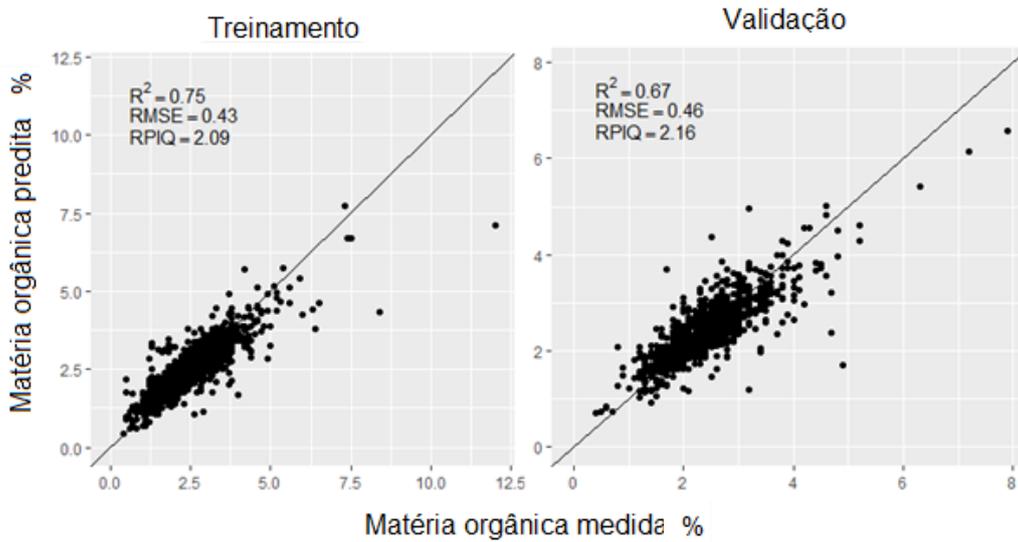
6.3.2 Performance dos modelos calibrados

Os modelos utilizados neste estudo apresentaram diferentes resultados para predição da MOS, sendo em ordem crescente de RMSE (%) do conjunto de validação: SVM>GPR>PLSR>MLR. Em geral, os modelos não lineares (SVM e GPR) apresentaram maior acurácia (Figura 3A e 3B). O modelo baseado em kernel function SVM apresentou a melhor performance entre os modelos utilizados, com maior acurácia ($R^2 = 0,70$, RMSE (%) = 0,44 e RPIQ = 2,26) obtida com o espectro SNV (Figura 4A). O método SVM utiliza uma abordagem de aprendizagem de máquina que extrai padrões dos dados, sendo amplamente utilizado por conseguir lidar com relações complexas observadas em conjuntos de dados espectrais (VISCARRA ROSSEL; BEHRENS, 2010). O modelo SVM foi utilizado para estimar o teor de carbono orgânico do solo por Stevens *et. al.* (2010) obtendo um coeficiente de determinação de 0,83 para calibrações regionais, sendo superior as calibrações globais ($R^2 = 0,73$), utilizando o mesmo modelo. Demonstrando que calibrações realizadas para regiões específicas, apresentaram melhor desempenho em relação a predição do carbono orgânico do solo.

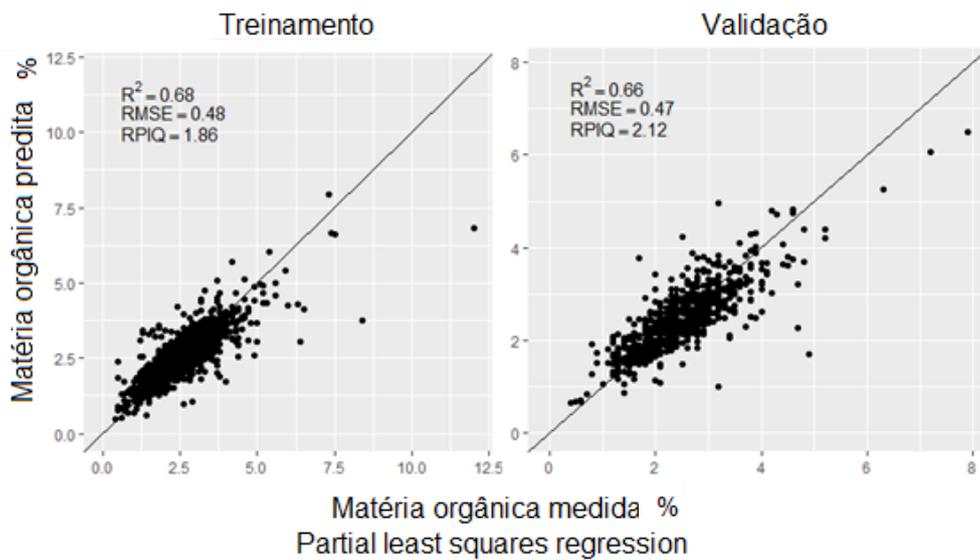
(a)



(b)



(c)



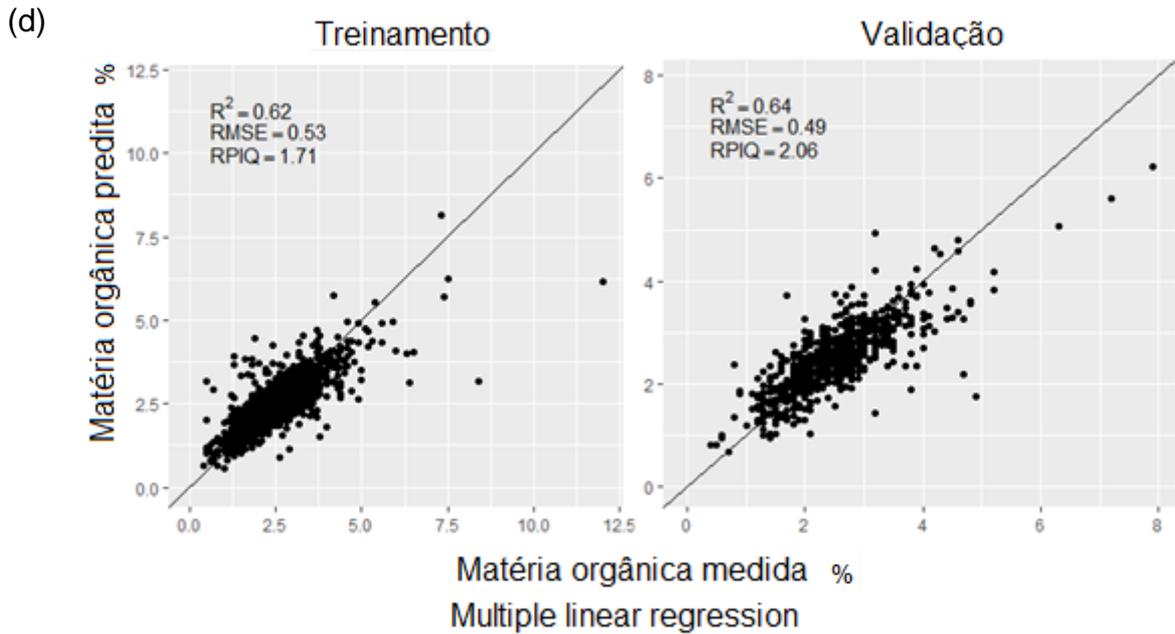


Figura 3. Melhores performances da predição de MOS (%) para os quatro modelos em relação a MOS (%) determinada por digestão úmida. A) SVM+SNV, B) GPR+DET, C) PLSR+SNV e D) MLR+SNV.

O método GPR, quando calibrado com dados espectrais DET, apresentou a segunda melhor performance para predição da MOS (Figura 3B). Em contraste com os resultados deste estudo, em estudo realizado por Dotto (2017) foram avaliados diferentes modelos multivariados para predição de carbono orgânico do solo, sendo que GPR, apresentou performance inferior ($R^2 = 0,79$) ao método linear PLSR ($R^2 = 0,81$). O que evidencia que diferentes modelos podem apresentar resultados diferentes conforme a variação das propriedades das amostras utilizadas.

O modelo PLSR apresentou performance inferior na predição de MOS (Figura 3C) em relação aos modelos não lineares. Quando aplicado no espectro preprocessados por SNV, apresentou a terceira melhor performance (Figura 3C) entre os modelos estudados. Entretanto, este modelo apresentou a menor variação na acurácia e performance satisfatória para todos os pré-processamentos ($R^2 > 0,62$ e $RMSE (\%) \leq 0,5$), mostrando ser um modelo robusto e justificando a sua utilização em grande parte dos trabalhos que estudam a espectroscopia para predição de propriedades do solo (DOTTO *et al.*, 2018). Apesar de ser um modelo linear, o modelo multivariado PLSR tem a capacidade de lidar com um grande volume de dados e com alta colinearidade

(WOLD; RUHE; DUNN, 1984), como é o caso de dados espectrais, por exemplo. Entretanto apresenta uma deficiência de compensar possíveis ruídos presentes no espectro (VOHLAND *et al.*, 2011), o que pode ter influenciado as predições por este modelo no presente estudo. Bellon-Maurel & Mcbratney, em 2011, compararam diferentes estudos utilizando spectroscopy para predição de matéria orgânica do solo, sendo que PLSR foi o modelo mais utilizado, entretanto apresentando resultados similares a outros modelos lineares e não lineares, variando para diferentes regiões e número de amostras utilizadas para calibração.

O modelo MLR, apresentou a menor acurácia na predição entre os modelos utilizados, sendo a sua melhor performance alcançada utilizando o espectro pré-processado pela técnica SNV (Figura 3D). Isto se deve ao fato deste modelo assumir que não haja colinearidades e não linearidades no espectro, e que as relações entre as variáveis dependentes e independentes são lineares. Entretanto, o espectro NIR apresenta alta não-linearidade e ruídos presentes no espectro, causados por pelo espalhamento da luz (RINNAN; BERG; ENGELSEN, 2009). Em geral, a performance inferior dos modelos lineares (MLR e PLSR) pode ser ligada ao comportamento espectral do solo, que é complexo e não linear, e que varia para diferentes locais, afetando a predição de modelos que possuem o pressuposto que as variáveis dependentes e independentes apresentem uma relação linear.

6.3.3 Classificação dos teores preditos de MOS

O teor de MOS predito pelos diferentes modelos, quando classificados nos intervalos baixo ($\leq 2,5\%$), médio (2,6-5,0%) e alto ($>5,0\%$) de MOS, apresentaram performance satisfatória na classificação dos teores de MOS mensurado por meio do método Walkley-Black (Tabela 5).

Tabela 5. Porcentagem das amostras classificadas corretamente a partir do teor predito pela espectroscopia NIR em relação ao teor de MOS por digestão úmida.

Pré-processamentos	Modelos	Teor de matéria orgânica			Todas amostras
		≤2,5	2,6 a 5,0	>5	
		-----%-----			
ABS	SVM	79	77	17	78
	GPR	73	82	17	76
	PLSR	76	82	35	78
	MLR	74	80	30	76
CR	SVM	83	80	35	82
	GPR	76	82	30	78
	PLSR	79	83	39	80
	MLR	72	79	30	74
DET	SVM	86	85	48	85
	GPR	83	83	43	83
	PLSR	80	81	43	80
	MLR	79	80	35	79
BIN	SVM	81	77	22	79
	GPR	69	81	13	74
	PLSR	78	82	39	79
	MLR	59	62	0	60
RAW	SVM	81	77	22	79
	GPR	74	82	17	76
	PLSR	77	80	39	78
	MLR	75	80	26	77
SNV	SVM	86	84	48	85
	GPR	81	81	39	80
	PLSR	81	81	39	80
	MLR	79	80	35	79
SMO	SVM	81	76	22	79
	GPR	74	82	17	77
	PLSR	77	82	35	78
	MLR	75	80	35	77
SGD	SVM	85	86	70	85
	GPR	82	84	74	83
	PLSR	78	82	43	79
	MLR	73	79	22	75
Número de amostras		1357	1008	23	2388

Em geral, todos os modelos apresentaram resultados satisfatórios, em ordem decrescente de concordância em relação aos dados de referência: SVM>GPR>PLSR>MLR. Os pré-processamentos SNV e SGD quando utilizados no modelo SNV apresentaram maior concordância com o método de análise de rotina, o que pode ser devido ao realce que estes pré-processamentos dão a

feições no espectro ligadas a matéria orgânica. Apesar do coeficiente de determinação máximo de 0,70 obtido entre os diferentes modelos, o modelo SVM classificou corretamente o teor de matéria orgânica em cerca de 85% das amostras em relação ao analisado por meio de digestão úmida. Entretanto, quando o teor das mostras foi maior que 5% os modelos apresentaram desempenho insatisfatório na predição da MOS. Isso pode ser devido ao pequeno número de amostras com teor maior do que 5% utilizadas na calibração dos modelos (Tabela 3). Esses resultados mostram que é necessário testar diferentes métodos de processamento espectral e de calibração multivariada, uma vez que, a acurácia dos modelos responde de forma distinta em relação a combinação desses métodos, não existindo um protocolo definido. Isso se deve as características composicionais e espectrais dos solos que compõem o conjunto de dados, devido à variabilidade pedológica que existe na pedosfera (ARAÚJO *et al.*, 2014; WIJEWARDANE *et al.*, 2016; MOURA-BUENO *et al.*, 2019), confirmando que modelos regionais gerados pela espectroscopia NIR são mais acurados.

A análise da MOS por meio da espectroscopia NIR tem como vantagens a rápida execução e a não geração de resíduos químicos, como no caso da análise por meio de digestão úmida, por exemplo, tornando a NIRS uma técnica analítica promissora para quantificação da MOS. O uso de novas técnicas analíticas, como a espectroscopia NIR é de fundamental importância para a quantificação de propriedades do solo, principalmente quando a demanda requer maior número de amostras, como é o caso de projetos de agricultura de precisão e monitoramento dos estoques de carbono no solo ao longo do tempo. Além disso, tais resultados demonstram a necessidade de novos estudos serem realizados com a finalidade de verificar a eficiência desta técnica analítica para outras propriedades do solo. A utilização de técnicas de pré-processamento espectral e calibração multivariada adequados melhorou a predição de MOS dos modelos. A performance nas predições variou conforme os modelos multivariados utilizados, sendo que SNV apresentou a melhor performance quando utilizado o modelo SVM. A classificação dos teores de MOS quantificados por NIRS apresentaram 85% de concordância quando comparado com os teores obtidos pelo método utilizado nos laboratórios de análise de solo de rotina. Sendo que os 15% de erro podem ser oriundos do baixo número de

amostras com menor teor de matéria orgânica, as quais obtiveram menor número de concordância.

6.4 Conteúdo de argila no solo

6.4.1 Performance das técnicas de pré-processamento nos modelos

Os sete métodos de pré-processamento em relação a performance de modelos de predição para argila são apresentados na tabela 6. Os resultados demonstraram que as diferentes técnicas de pré-processamento apresentaram performances distintas em relação à predição dos teores de argila no solo.

Após testados os diferentes métodos de pré-processamento, a melhor performance foi encontrada por meio da utilização do modelo SVM em conjunto com os dados espectrais preprocessados por SNV ($R^2 = 0,79$ e RMSE (%) = 7,22). Seguindo posteriormente em ordem decrescente pelo modelo GPR em conjunto com os dados espectrais preprocessados por SNV ($R^2 = 0,78$ e RMSE (%) = 7,27), modelo PLSR em conjunto com os dados espectrais preprocessados por Suavização e *Binning* ($R^2 = 0,78$ e RMSE (%) = 7,28) e modelo MLR em conjunto com os dados espectrais preprocessados por *Suavização* e *Raw data* ($R^2 = 0,76$ e RMSE (%) = 7,62), indicando o pior desempenho de MLR em relação aos outros modelos com resultados similares. Resultados similares foram obtidos por estudos de Viscarra-Rossel (2008), que relata desempenho satisfatório do pré-processamento SNV, devido a atenuação dos ruídos causados pelo espalhamento da radiação eletromagnética causado pelos diferentes tamanhos de partícula presentes no solo.

Tabela 6. Performance dos grupos de treinamento e validação em relação aos diferentes métodos de pré-processamento e calibração para argila.

Modelo	Pré-processamento	Treinamento			Validação		
		R ²	RMSE (%)	RPIQ	R ²	RMSE (%)	RPIQ
MLR	Suavização (SMO)	0,78	7,40	3,24	0,76	7,62	3,15
	Detrend (DET)	0,79	7,22	3,32	0,76	7,66	3,13
	Binning (BIN)	0,70	8,73	2,75	0,68	8,83	2,72
	1º derivada de Savitzky-Golay (SGD)	0,73	8,24	2,91	0,71	8,32	2,89
	Standard normal variate (SNV)	0,79	7,25	3,31	0,76	7,65	3,14
	Continuum removal (CR)	0,77	7,59	3,16	0,75	7,75	3,10
	Absorbância (ABS)	0,76	7,84	3,06	0,74	7,87	3,05
	Raw data (RAW)	0,78	7,42	3,23	0,76	7,62	3,15
PLSR	Suavização (SMO)	0,80	7,09	3,38	0,78	7,28	3,29
	Detrend (DET)	0,81	6,96	3,45	0,78	7,39	3,25
	Binning (BIN)	0,81	7,01	3,42	0,78	7,28	3,30
	1º derivada de Savitzky-Golay (SGD)	0,82	6,78	3,54	0,76	7,62	3,15
	Standard normal variate (SNV)	0,81	6,93	3,46	0,78	7,40	3,24
	Continuum removal (CR)	0,80	7,06	3,40	0,76	7,57	3,17
	Absorbância (ABS)	0,78	7,40	3,24	0,76	7,69	3,12
	Raw data (RAW)	0,81	6,98	3,44	0,77	7,43	3,23
SVM	Suavização (SMO)	0,78	7,43	3,23	0,77	7,47	3,21
	Detrend (DET)	0,85	6,14	3,91	0,78	7,36	3,26
	Binning (BIN)	0,78	7,40	3,24	0,77	7,42	2,23
	1º derivada de Savitzky-Golay (SGD)	0,89	5,36	4,48	0,67	9,36	2,56
	Standard normal variate (SNV)	0,82	6,75	3,56	0,79	7,22	3,32
	Continuum removal (CR)	0,80	7,19	3,34	0,77	7,42	3,23
	Absorbância (ABS)	0,78	7,40	3,24	0,77	7,42	3,23
	Raw data (RAW)	0,78	7,40	3,24	0,78	7,38	3,25
GPR	Suavização (SMO)	0,75	8,00	3,00	0,73	8,04	2,98
	Detrend (DET)	0,85	6,08	3,95	0,78	7,37	3,26
	Binning (BIN)	0,74	8,18	2,93	0,72	8,18	2,93
	1º derivada de Savitzky-Golay (SGD)	0,92	4,64	5,17	0,67	9,24	2,60
	Standard normal variate (SNV)	0,82	6,75	3,56	0,78	7,27	3,30

(cont.)

Cont. removal (CR)	0,80	7,11	3,38	0,76	7,57	3,17
Absorbância (ABS)	0,73	8,32	2,88	0,72	8,18	2,93
Raw data (RAW)	0,75	7,99	3,00	0,73	8,02	2,99

*MLR: Multiple Linear Regression. *PLSR: Partial Least Squares Regression. *SVM: Support Vector Machine. *GPR: Gaussian Process Regression.

6.4.2 Performance dos diferentes modelos de calibração

Os modelos utilizados neste estudo apresentaram diferentes resultados para predição da argila do solo, sendo em ordem decrescente de coeficiente de determinação (R^2) do conjunto de validação: SVM, GPR, PLSR e MLR. Em geral, os modelos não lineares apresentaram uma melhor capacidade de processar um grande volume de dados, que possam apresentar relações não lineares no espectro, como é o caso da espectroscopia NIR, onde existe uma série de ruídos no espectro, causados pelo espalhamento da radiação eletromagnética quando incidida no solo.

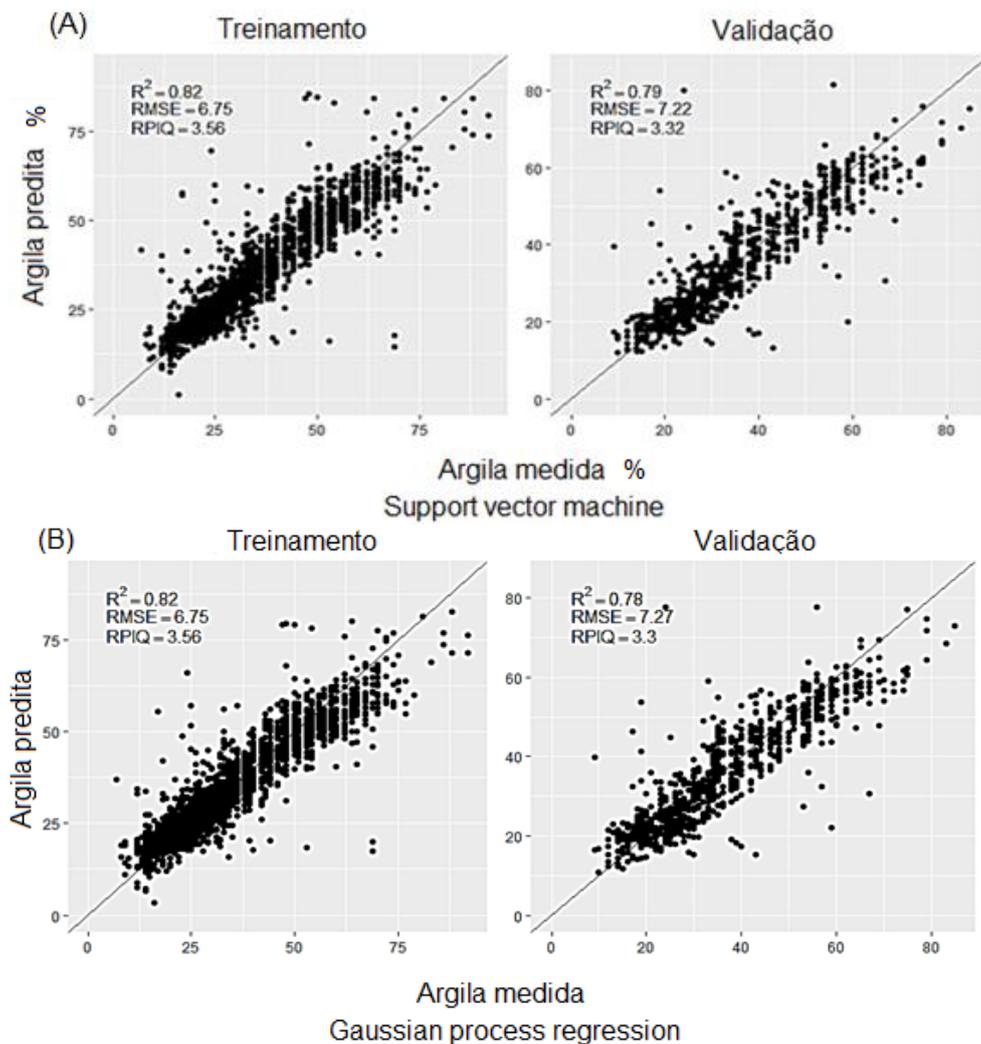
O modelo *Support Vector Machine* (SVM) utiliza uma abordagem de aprendizagem de máquina, extraíndo padrões dos dados. Este modelo é utilizado amplamente por conseguir lidar com relações complexas entre os conjuntos de dados (VISCARRA ROSSEL; BEHRENS, 2010). Este modelo, baseado em *kernel function*, apresentou a melhor performance entre os modelos utilizados, sendo a melhor predição ($R^2 = 0,79$ e RMSE (%) = 7,22) obtida com o espectro SNV (Figura 2A).

O modelo *Gaussian Process Regression* (GPR), também baseado em *kernel function*, apresentou a segunda melhor performance para predição do teor de argila no solo (Figura 4B). Já o modelo *Partial Least Squares Regression* (PLSR), o mais utilizado em espectroscopia NIR, apresentou performance inferior na predição de argila (Figura 4C) em relação aos modelos não lineares.

Apesar de ser um modelo linear, o modelo multivariado *Partial Least Squares Regression* (PLSR) tem a capacidade de lidar com um grande volume de dados e com alta colinearidade (WOLD; RUHE; DUNN, 1984), como é o caso de dados espectrais, por exemplo. Entretanto apresenta uma deficiência de compensar possíveis ruídos presentes no espectro no espectro (VOHLAND *et al.*, 2011), o que pode ter influenciado as predições por este modelo. A

performance inferior dos modelos lineares pode ser ligada ao comportamento espectral do solo, que é complexo e não linear, e que varia para diferentes locais, afetando a predição de modelos lineares.

O modelo *Multiple Linear Regression* (MLR), apresentou a pior predição entre os modelos utilizados, sendo a sua melhor performance quando combinado com o espectro pré-processado pela técnica SMO (Figura 4D). Isto se deve ao fato deste modelo assumir que as relações entre as variáveis dependentes e independentes são lineares. Entretanto, o espectro NIR apresenta alto grau e não-linearidade e ruídos presentes no espectro, causados por dispersão de luz (RINNAN; BERG; ENGELSEN, 2009).



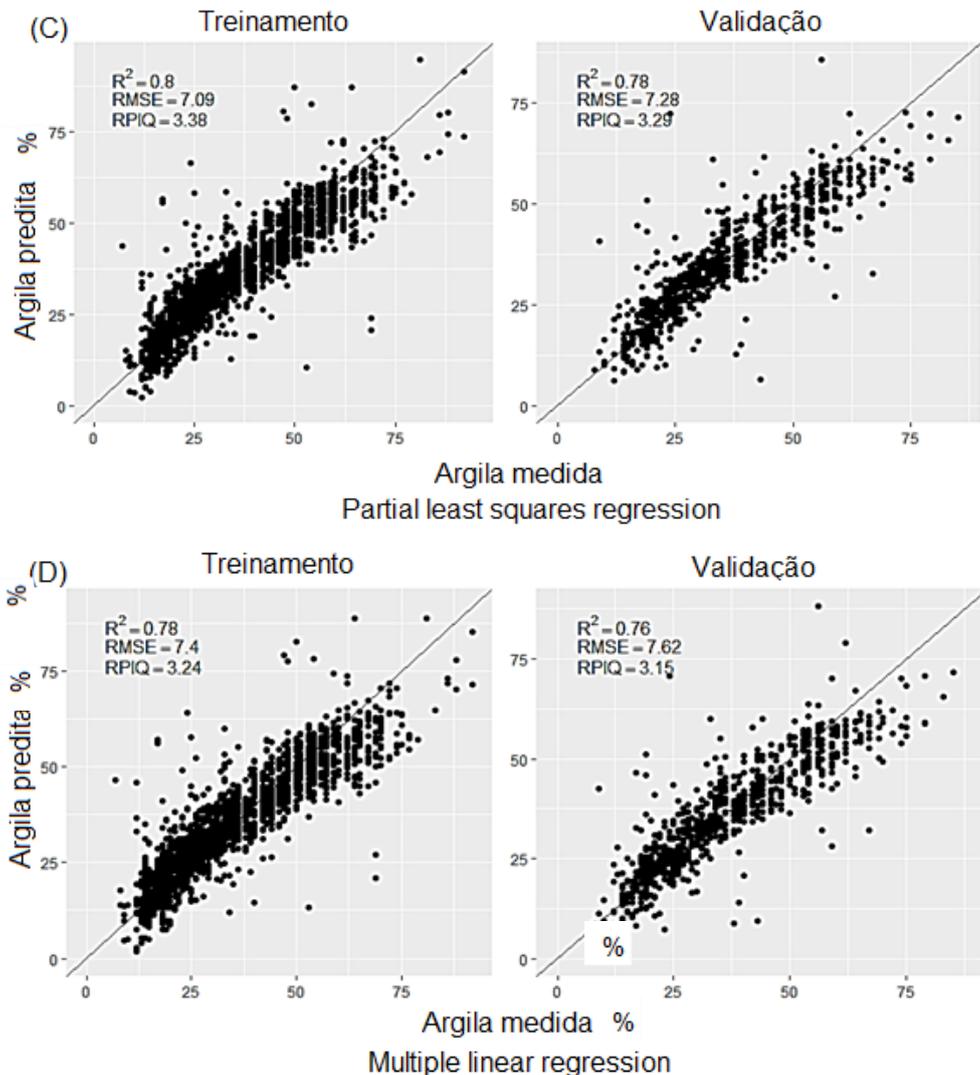


Figura 4. Melhores performances da predição de argila (%) pelos quatro modelos e os respectivos métodos de pré-processamento em relação a argila (%) medida pelo método do densímetro. A) SVM: SNV, B) GPR: SNV, C) PLSR: SMO D) MLR:SMO;

Ainda, em consonância com este estudo, Demattê e colaboradores (2018) demonstraram que a argila pode ser determinada pela técnica de espectroscopia com resultado de alta qualidade, em relação a determinação por métodos convencionais utilizados na rotina em laboratórios certificados de análise de solo.

6.4.3 Classificação dos teores preditos de argila

Os diferentes modelos de predição de argila do solo apresentaram classificação satisfatória do teor de argila em relação ao mensurado por meio do

método do densímetro conforme descrito (TEIXEIRA et al 2017) em podendo ser observado na tabela 7.

Tabela 7. Porcentagem de classificação correta das amostras nas suas respectivas classes de argila (<20, 20-40, 40-60 e >60%).

Pré-processamentos	Modelos			
	MLR	PLSR	SVM	GPR
	-----%-----			
Suavização	75	76	75	73
Detrend	72	73	78	76
Binning	71	45	63	71
Savitzky-Golay 1 derivative	66	69	65	65
Standard normal variate	72	73	75	74
Continuum removal	73	75	76	75
Absorbância	74	75	76	73
Espectro original	75	76	75	73

De forma geral, todos os modelos apresentaram resultados satisfatórios, em ordem decrescente de concordância em relação aos dados de referência: SVM, GPR, PLSR e MLR. Apesar do coeficiente de determinação máximo de 0,79, obtido entre os diferentes modelos, o modelo que apresentou a melhor performance, classificou corretamente cerca de 78% dos teores de argila das amostras, no mesmo nível em relação ao analisado por meio da metodologia convencional. Dotto e colaboradores (2016), por meio da utilização de modelo de regressão linear um coeficiente de determinação (R^2) de 0,75, com aproximadamente 81% de acurácia na classificação textural das amostras utilizadas. Além do nível de concordância aceitável da análise da argila por meio da espectroscopia NIR em relação a outros estudos, esta ainda permite a análise de um maior número de amostras por unidade de tempo, o que possibilita a análise do teor de argila no solo com uma maior resolução espacial e temporal, tornando a espectroscopia NIR uma técnica analítica alternativa promissora em relação aos métodos convencionais (VISCARRA ROSSEL; MCBRATNEY, 1998), utilizados nos laboratórios de rotina de análise de solos.

A NIRS apresentou resultados satisfatórios demonstrando ser uma técnica alternativa aos métodos convencionais de análise do teor de argila no solo, para fins de análise de rotina. Ainda, a performance nas predições variou conforme os modelos e pré-processamentos utilizados, sendo SNV o pré-

processamento que apresentou a melhor performance quando utilizado o modelo SVM. A classificação dos teores de argila quantificados por NIRS apresentou 78% de concordância quando comparado com os teores obtidos pelo método convencional.

7. CONCLUSÃO GERAL

Os resultados apresentados nesta dissertação demonstram a performance positiva da espectroscopia de infravermelho próximo em prever o teor de matéria orgânica e argila do solo. Tais propriedades são relacionadas diretamente com os grupos funcionais constituintes do solo, sendo possível sua quantificação por meio da utilização de bandas específicas do espectro eletromagnético. O pré-processamento dos dados espectrais das amostras contribuíram para uma melhor predição das propriedades estudadas. Sendo que o pré-processamento *standard normal variate*, apresentou predições superiores em relação às outras técnicas de pré-processamento espectral para predição de argila e matéria orgânica do solo. Em geral, com exceção do modelo MLR todos os modelos apresentaram performances satisfatórias para predição de argila e matéria orgânica do solo, sendo que os modelos não lineares apresentaram performance superior, com ênfase para *support vector machine* que apresentou predição superior aos demais modelos, tanto para argila quanto para MOS.

8. CONSIDERAÇÕES FINAIS

A espectroscopia de infravermelho próximo é uma alternativa aos métodos utilizados nos laboratórios de rotina para análise de solos, que possibilita uma análise de um maior número de amostras por tempo e a um menor custo. Para futuros trabalhos a estratificação dos conjuntos de dados formando modelos para diferentes conjuntos de solos seria uma opção ao baixo número de amostras com menor conteúdo de MOS e argila. Visando também a relação entre esses constituintes e sua influência na curva espectral do solo. Ainda, com o avanço na área de quimiometria, outros métodos de pré-processamento e modelagem podem ser testados com a finalidade de se melhorar as predições das propriedades do solo por meio da espectroscopia.

9. REFERENCIAS

ALEXANDRATOS, N. *et al.* **World agriculture towards 2030/2050: the 2012 revision.** [Rome]: FAO, Agricultural Development Economics Div., 2012. (ESA Working paper, 12-03).

AMUNDSON, R. *et al.* Soil and human security in the 21st century. **Science**, Washington, v. 348, n. 6235, p. 647, 2015.

ARMENTA, S.; DE LA GUARDIA, M. Vibrational spectroscopy in soil and sediment analysis. **Trends in Environmental Analytical Chemistry 2**, Amsterdam, v. 2, p. 43–52, 2014.

BALABIN, R. M.; SAFIEVA, R. Z.; LOMAKINA, E. I. Comparison of linear and nonlinear calibration models based on near infrared (NIR) spectroscopy data for gasoline properties prediction. **Chemometrics and Intelligent Laboratory Systems**, Amsterdam, v. 88, n. 2, p. 183–188, 2007.

BARNES, R. J.; DHANOA, M. S.; LISTER, S. J. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. **Applied Spectroscopy**, Baltimore, v. 43, n. 5, p. 772–777, 1989.

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**, Amsterdam, v. 29, n. 9, p. 1073–1081, 2010.

BELLON-MAUREL, V.; MCBRATNEY, A. Soil biology & biochemistry near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils e Critical review and research perspectives. **Soil Biology and Biochemistry**, Elmsford, v. 43, n. 7, p. 1398–1410, 2011.

BEN-DOR, E.; INBAR, Y.; CHEN, Y. The reflectance spectra of organic matter in the visible near-infrared and short-wave infrared region (400-2500 nm) during a controlled decomposition process. **Remote Sensing of Environment**, New York, v. 61, n. 1, p. 1–15, 1997.

BEN-DOR, E.; IRONS, J. R.; EPEMA, G. F. Soil reflectance. *In: Manual of remote sensing: remote sensing for earth sciences* edited by RENCZ, A. N. 3rd ed. New York: Wiley, 1999. v. 3, p. 111-188.

BLANCO, M.; VILLARROYA, I. NIR spectroscopy: a rapid-response analytical tool. **TrAC - Trends in Analytical Chemistry**, Amsterdam, v. 21, n. 4, p. 240–250, 2002.

CHEN, T.; MORRIS, J.; MARTIN, E. Gaussian process regression for multivariate spectroscopic calibration. **Chemometrics and Intelligent Laboratory Systems**, Amsterdam, v. 87, n. 1, p. 85–97, 2007.

CLARK, R. N. Spectroscopy of rocks and minerals, and principles of spectroscopy. *In: Manual of remote sensing: remote sensing for earth sciences* edited by RENCZ, A. N. 3rd ed. New York: John Wiley, 1999. v. 3, p. 3–58.

CLARK, R. N.; ROUSH, T. L. Reflectance spectroscopy: quantitative analysis techniques for remote sensing applications. **Journal of Geophysical Research: Solid Earth**, Hoboken, v. 89, n. B7, p. 6329–6340, 1984.

CLEGG, S. M. *et al.* Multivariate analysis of remote laser-induced breakdown spectroscopy spectra using partial least squares, principal component analysis, and related techniques. **Spectrochimica Acta - Part B Atomic Spectroscopy**, Amsterdam, v. 64, n. 1, p. 79–88, 2009.

CORTES, C.; VAPNIK, V. Support-Vector Networks. **Machine Learning**, New York, v. 20, n. 3, p. 273–297, 1995.

DALMOLIN, R. S. D. *et al.* Relação entre os constituintes do solo e seu comportamento espectral. **Ciência Rural**, Santa Maria, v. 35, n. 2, p. 481–489, 2005.

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**, Amsterdam, v. 337, p. 111–121, 2018.

DIXON, J. B. Roles of clays in soils. **Applied Clay Science**, Amsterdam, v. 5, n. 5–6, p. 489–503, 1991.

DOTTO, A. C. *et al.* A systematic study on the application of scatter-corrective and spectral- derivative preprocessing for multivariate prediction of soil organic carbon by Vis-NIR spectra. **Geoderma**, Amsterdam, v. 314, p. 262–274, 2018.

DOTTO, A. C. *et al.* Potential of spectroradiometry to classify soil clay content. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 40, p. 1-8, 2016.

DOTTO, A. C. **Soil VIS-NIR spectroscopy: predictive potential and the development of a graphical user interface in R.** 2017. 112 f. Tese (Doutorado) – Programa de Pós-Graduação em Ciência do Solo, Universidade Federal de Santa Maria, Santa Maria, 2009.

DU CHANGWEN, M. F.; YUZHEN, L.; JIANMIN, Z. Soil fertility assessed by infrared spectroscopy. *In: LAL, R.; STEWART, B. A. (ed.). Soil-specific farming: precision agriculture.* Boca Raton: CRC press, 2015. (Advances in soil science; v. 22). p. 155-177.

- GHOLIZADEH, A. *et al.* Visible, near-infrared, and mid-infrared spectroscopy applications for soil assessment with emphasis on soil organic matter content and quality: state-of-the-art and key issues. **Applied Spectroscopy**, Baltimore, v. 67, p. 1349–1362, 2013.
- GHOLIZADEH, A. *et al.* Comparing different data preprocessing methods for monitoring soil heavy metals based on soil spectral features. **Soil and Water Research**, Praga, v. 10, p. 218–227, 2016.
- GRINAND, C. *et al.* Prediction of soil organic and inorganic carbon contents at a national scale (France) using mid-infrared reflectance spectroscopy (MIRS). **European Journal of Soil Science**, Oxford, v. 63, p. 141–151, 2012.
- GUERRERO, C.; ROSSEL, R. A. V.; MOUAZEN, A. M. Diffuse reflectance spectroscopy in soil science and land resource assessment. **Geoderma**, Amsterdam, v. 158, n. 1–2, p. 1–2, 2010.
- HARTEMINK, A. E.; MCBRATNEY, A. A soil science renaissance. **Geoderma**, Amsterdam, v. 148, n. 2, p. 123–129, 2008.
- HILLEL, D. **Introduction to environmental soil physics**. Burlington: Elsevier Science, 2003.
- JACKSON, R. B. *et al.* The ecology of soil carbon: pools, vulnerabilities, and biotic and abiotic controls. **Annual Review of Ecology, Evolution, and Systematics**, Palo Alto, Calif., v. 48, p. 419–445, 2017.
- LAL, R. Soil Carbon sequestration impacts on global climate change and food security. **American Association for the Advancement of Science**, Washington, v. 304, n. 5677, p. 1623–7, 2004.
- LAL, R. Societal value of soil carbon. **Journal of Soil and Water Conservation**, Ankeny, v. 69, n. 6, p. 186–192, 2014.
- LEROY, C. Radiation interaction in matter and principles of detection. *In*: CONFERENCE, 1., 2007, Prague. **Proceedings** [...]. Melville, NY: AIP, 2007. v. 958. p. 85–91.
- MCBRATNEY, A. B.; MINASNY, B.; VISCARRA ROSSEL, R. Spectral soil analysis and inference systems: A powerful combination for solving the soil data crisis. **Geoderma**, Amsterdam, v. 136, n. 1–2, p. 272–278, 2006.
- 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**, Amsterdam, v. 337, p. 565–581, 2019.
- NIEMZ, M. H. Light and matter. *In*: HOUCK, M. M.; SIEGEL, J. A. **Fundamentals of forensic science**. 3rd ed. Amsterdam: Academic Press, 2015. p. 94–119.
- NOCITA, M. *et al.* Prediction of soil organic carbon content by diffuse reflectance spectroscopy using a local partial least square regression

approach. **Soil Biology and Biochemistry**, Elmsford, v. 68, p. 337–347, 2014.

NOCITA, M. *et al.* Soil Spectroscopy: an alternative to wet chemistry for soil monitoring. *In*: SPARKS, D. L. (ed.). **Advances in Agronomy**. Amsterdam: Elsevier, 2015. v. 132, p.139-159.

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

R CORE TEAM. **R**: a language and environment for statistical computing. Version 3.6.0. Vienna, Austria, 2017.

RINNAN, A. Pre-processing in vibrational spectroscopy: when, why and how. **Analytical Methods**, London, v. 6, n. 18, p. 7124–7129, 2014.

RINNAN, A.; VAN DEN BERG, F.; ENGELSEN, S. B. Review of the most common pre-processing techniques for near-infrared spectra. **TrAC - Trends in Analytical Chemistry**, Amsterdam, v. 28, n. 10, p. 1201–1222, 2009.

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

SAVITZKY, A.; GOLAY, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. **Analytical Chemistry**, Amsterdam, v. 36, n. 8, p. 1627–1639, 1964.

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. 107, p. 163-215.

STENBERG, B.; JONSSON, A.; BÖRJESSON, T. Near infrared technology for soil analysis with implications for precision agriculture. *In*: INTERNATIONAL CONFERENCE ON NEAR INFRARED SPECTROSCOPY, 10., 2001, Kyongju, Korea. **Proceedings** [...]. Chichester: NIR Publications, 2002. p. 279-284.

STEVENS, A. *et al.* Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy. **Geoderma**, Amsterdam, v. 158, n. 1–2, p. 32–45, 2010.

STEVENS, A. *et al.* Prediction of soil organic carbon at the European scale by visible and near infrared reflectance spectroscopy. **PLoS ONE**, San Francisco, v. 8, n. 6, 2013.

STEVENS, A.; RAMIREZ-LOPEZ, L. **An introduction to the prospectr package**. [S. l.], 2013. Disponível em: <https://cran.r-project.org/web/packages/prospectr>. Acesso em: 18 out. 2017.

STONER, E. R.; BAUMGARDNER, M. F. Characteristic variations in reflectance of surface soils. **Soil Science Society of America Journal**, Madison, v. 45, n. 6, p. 1161-1165, 1981.

STEWART, B. **Infrared spectroscopy: fundamentals and applications**. New York: Wiley, 2004.

TIECHER, T. *et al.* **Evolução e estado da fertilidade do solo no Norte do Rio Grande do Sul e Sudoeste de Santa Catarina**. Porto Alegre: UFRGS, 2016.

TIESSEN, H.; CUEVAS, E.; CHACON, P. The role of organic matter in sustaining soil fertility. **Nature**, London, v. 371, p. 783–785, 1994.

VASQUEZ, G. M.; GRUNWALD, S.; SICKMAN, J. O. Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. **Geoderma**, Amsterdam, v. 146, n. 1–2, p. 14–25, 2008.

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

VISCARRA ROSSEL, R. A. *et al.* Proximal soil sensing: an effective approach for soil measurements in space and time. *In*: SPARKS, D. L. (ed.). **Advances in agronomy**. Burlington: Academic Press, 2011. p. 243-291.

VISCARRA ROSSEL, R.; MCBRATNEY, A. B. Soil chemical analytical accuracy and costs: implications from precision agriculture. **Australian Journal of Experimental Agriculture**, Camberra, v. 38, n. 7, p. 765, 1998.

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

VOHLAND, M. *et al.* Comparing different multivariate calibration methods for the determination of soil organic carbon pools with visible to near infrared spectroscopy. **Geoderma**, Amsterdam, v. 166, n. 1, p. 198–205, 2011.

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**, Philadelphia, v. 37, n. 1, p. 29-38, 1934.

WOLD, S. *et al.* The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. **SIAM Journal on Scientific and Statistical Computing**, Philadelphia, v. 5, n. 3, p. 735-743, 1984.

WOLD, S.; SJÖSTRÖM, M.; ERIKSSON, L. PLS-regression: A basic tool of chemometrics. **Chemometrics and Intelligent Laboratory Systems**, Amsterdam, v. 58, n. 2, p. 109–130, 2001.

ZORNOZA, R.; GUERRERO, C. Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. **Soil Biological Biochemistry**, Amsterdam, v. 40, n. 7, p. 1923–1930, 2008.