

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

**ESTIMATIVA DE PROPRIEDADES DO SOLO POR ESPECTROSCOPIA Vis-  
NIR-SWIR E VARIÁVEIS AMBIENTAIS NO SUL DO BRASIL**

**José Janderson Ferreira Costa**  
**(Tese de Doutorado)**

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

**ESTIMATIVA DE PROPRIEDADES DO SOLO POR ESPECTROSCOPIA Vis-  
NIR-SWIR E VARIÁVEIS AMBIENTAIS NO SUL DO BRASIL**

JOSÉ JANDERSON FERREIRA COSTA  
Engenheiro-Agrônomo (UFRA)  
Mestre em Ciência do Solo (UFRGS)

Tese apresentada como um dos  
requisitos à obtenção do Grau de  
Doutor em Ciência do Solo

Porto Alegre (RS) Brasil  
Setembro de 2020

### CIP - Catalogação na Publicação

Costa, José Janderson Ferreira  
ESTIMATIVA DE PROPRIEDADES DO SOLO POR  
ESPECTROSCOPIA Vis-NIR-SWIR E VARIÁVEIS AMBIENTAIS NO  
SUL DO BRASIL / José Janderson Ferreira Costa. --  
2020.  
94 f.  
Orientador: Elvio Giasson.

Tese (Doutorado) -- Universidade Federal do Rio  
Grande do Sul, Faculdade de Agronomia, Programa de  
Pós-Graduação em Ciência do Solo, Porto Alegre, BR-RS,  
2020.

1. Sensoriamento proximal hiperespectral. 2.  
Carbono orgânico. 3. Textura do solo. 4. Cor do solo.  
5. Modelos multivariados. I. Giasson, Elvio, orient.  
II. Título.

JOSÉ JANDERSON FERREIRA COSTA  
Engenheiro Agrônomo - UFRA  
Mestre em Ciência do Solo - UFRGS

## TESE

Submetida como parte dos requisitos  
para obtenção do Grau de

### DOUTOR EM CIÊNCIA DO SOLO

Programa de Pós-Graduação em Ciência do Solo

Faculdade de Agronomia

Universidade Federal do Rio Grande do Sul

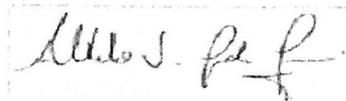
Porto Alegre (RS), Brasil

Aprovado em: 23.09.2020  
Pela Banca Examinadora

Homologado em: 27.12.2020  
Por



ELVIO GIASSON  
Orientador-PPG Ciência do Solo

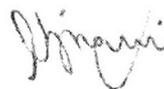


ALBERTO VASCONCELLOS INDA JUNIOR  
Coordenador do  
Programa de Pós-Graduação em  
Ciência do Solo

DANIEL DE BORTOLI TEIXEIRA  
Universidade de Marília  
(através do Mconf da UFRGS)

IVAN LUIZ ZILLI BACIC  
EPAGRI/SC  
(através do Mconf da UFRGS)

TALES TIECHER  
PPG Ciência do Solo/UFRGS  
(através do Mconf da UFRGS)



CARLOS ALBERTO BISSANI  
Diretor da Faculdade  
de Agronomia

*"Toda a nossa ciência comparada com a realidade, é primitiva e infantil – e, no entanto, é a coisa mais preciosa que temos..."*

*Albert Einstein*

*Dedico a minha mãe (Rosinete) e ao meu pai (José Joaquim).*

## AGRADECIMENTOS

À Deus por me permitir a conclusão deste trabalho.

Ao orientador, professor Elvio Giasson, pelos ensinamentos e confiança nos seis anos de orientação no Departamento de Solos da UFRGS.

À Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina, em nome da pesquisadora Elisângela Benedet da Silva, pelo fornecimento de parte do material utilizado nos estudos e que compõem a Biblioteca Espectral de Solos do Brasil.

Aos professores da banca Daniel de Bortoli Teixeira, Tales Tiecher e Ivan Luiz Zilli Bacic pelas contribuições e aprimoramento dos estudos.

Aos professores do PPGCS da UFRGS, pelas contribuições que proporcionaram o meu crescimento acadêmico.

Aos servidores da Faculdade de Agronomia, que contribuíram nos processos necessários para a conclusão deste trabalho e do curso.

Ao Departamento de Solos, em nome da Faculdade de Agronomia e da Universidade Federal do Rio Grande do Sul, pela estrutura e recursos que possibilitaram o desenvolvimento deste trabalho e da realização do curso.

Aos colegas e amigos que fiz no PPGCS, em especial aos do laboratório Aquário: Alcinei, Israel, Elisângela, Benito, Joelma, Fabrício, Antony, Ryshardson, João e Saulo. Agradeço pela parceria, alegrias e pelas contribuições na construção deste trabalho.

À minha família, pelo apoio e incentivo que me dispensaram durante todo o período de desenvolvimento e execução deste trabalho.

Aos meus avós paternos, Dorcelino e Rita (*in memoriam*), que infelizmente nos deixaram e agora restam boas lembranças.

À minha noiva Monique Nascimento, pela companhia, amor e apoio durante minha caminhada na pós-graduação.

Às minhas irmãs Janilce e Jacielle Ferreira por sempre estarem ao meu lado dividindo às alegrias e os sonhos.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão da bolsa de estudos.

# ESTIMATIVA DE PROPRIEDADES DO SOLO POR ESPECTROSCOPIA Vis-NIR-SWIR E VARIÁVEIS AMBIENTAIS NO SUL DO BRASIL<sup>1</sup>

Autor: José Janderson Ferreira Costa

Orientador: Prof. Elvio Giasson

## RESUMO

Esta tese apresenta o uso das técnicas de espectroscopia de reflectância (ER) e variáveis ambientais (VAs) para a predição da textura do solo e do carbono orgânico do solo (COS). A ER é uma alternativa às análises físico-químicas laboratoriais na quantificação da textura do solo e do teor de COS, no entanto, a capacidade preditiva dos modelos é influenciada pelas diferenças existentes entre os tipos de pré-processamento espectral, pelos métodos multivariados e pela variabilidade das amostras de solos, dentre outros. Assim, a presente tese teve como objetivo geral estudar os efeitos que a estratificação das amostras pela cor do solo provoca no desempenho dos modelos e estudar os efeitos da inclusão de uma segunda variável preditora (variáveis ambientais) à matriz espectral. A tese é composta por três estudos. O Estudo 1 avaliou o uso de parâmetros de cor baseados no espectro visível (Vis) combinados com estatística multivariada para estratificar amostras de solo de uma biblioteca espectral local e comparar os resultados de diferentes modelos de predição de COS, argila, areia e silte. As amostras de solos foram coletadas no Parque Nacional de São Joaquim, Santa Catarina, Brasil. Dados de reflectância do solo foram obtidos por espectroscopia Vis-NIR-SWIR. Os teores de COS foram determinados por combustão seca. As frações areia, silte e argila foram determinadas utilizando o método da Pipeta. Nesse estudo concluiu-se que o uso de parâmetros de cor aplicados na classificação de amostras de solos melhorou o desempenho dos modelos. O Estudo 2 teve como objetivos: (a) aplicar os parâmetros de cor baseados em Vis na estratificação de uma biblioteca espectral regional de solos; (b) avaliar o desempenho dos modelos preditivos gerados a partir da estratificação *Fuzzy K-means* da biblioteca espectral de solo e (c) explicar as possíveis mudanças na precisão da predição com base nos modelos estratificados. Para isso, foi utilizada uma biblioteca espectral regional de solos com 1.535 amostras de solo. Nesse estudo concluiu-se que a estratificação da biblioteca espectral regional é uma boa estratégia para melhorar as avaliações regionais dos recursos do solo. O Estudo 3 teve como objetivo combinar as VAs com às curvas espectrais Vis-NIR-SWIR para melhorar as estimativas do teor de COS e das frações areia, argila e silte. Nesse estudo concluiu-se que a inclusão de uma segunda variável preditora melhorou o desempenho dos modelos de predição. Assim, concluiu-se de modo geral que as metodologias testadas melhoraram o desempenho dos modelos na predição das propriedades do solo e que, do ponto de vista analítico e agrônomo, podem ser utilizados como métodos complementares às análises de solo.

**Palavras-chave:** sensoriamento proximal hiperespectral, carbono orgânico, textura do solo, cor do solo, modelos multivariados.

---

<sup>1</sup>Tese de Doutorado em Ciência do Solo. Programa de Pós-Graduação em Ciência do Solo, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul. Porto Alegre. (94 p.). Setembro, 2020.

## ESTIMATIVE OF SOIL PROPERTIES BY Vis-NIR-SWIR SPECTROSCOPY AND ENVIRONMENTAL VARIABLES IN SOUTHERN BRAZIL<sup>2</sup>

Author: José Janderson Ferreira Costa

Adviser: Prof. Elvio Giasson

### ABSTRACT

This thesis presents the use of reflectance spectroscopy (RS) techniques and environmental variables (EVs) for the prediction of soil texture and soil organic carbon (SOC). Reflectance spectroscopy is an alternative to physical-chemical laboratory analyzes in the quantification of soil texture and SOC content; however, the predictive capacity of the models is influenced by differences between the types of spectral pre-processing, multivariate methods, and variability of soil samples used to calibrate multivariate methods, among others. Thus, the present thesis aimed to study the effects that soil heterogeneity causes on the performance of the models, and study the effects of including a second predictor variable (environmental variables) to the spectral matrix. The thesis is composed of three studies. Study 1 evaluated the use Vis-based-color parameters combined with multivariate statistical techniques to group soil samples from a local spectral library and comparing the results of different SOC, clay, sand and silt prediction models. Soil sampling in the region of São Joaquim National Park, Santa Catarina State, Brazil. Soil reflectance data were obtained by Vis-NIR-SWIR spectroscopy. Soil organic carbon content was determined by dry combustion in an elemental analyzer. Sand, silt and clay fractions were determined using the pipette method. In this study, it was concluded that the use of color parameters applied in the classification of soil samples improved the performance of the models in predicting the properties of the soil studied. Study 2 aimed to (a) to apply Vis-based color parameters on the stratification of a regional soil spectral library; (b) to evaluate the performance of the predictive models generated from the Fuzzy K-means stratification of the spectral library and (c) to explain the possible changes in prediction accuracy based on the stratified models. Thus, a regional soil spectral library with 1,535 samples from the State of Santa Catarina, Brazil, was used. In this study it was concluded that the stratification of a regional spectral library is a good strategy to improve regional soil resource assessments, reducing the prediction error in qualitative soil properties determination. Study 3 aimed to combine the EVs with the Vis-NIR-SWIR spectral curves to improve the estimates of the SOC content and the fractions sand, clay, and silt. In this study, it is concluded that the inclusion of a second predictor variable improved the performance of the prediction models, with an increase in the value of  $R^2$  and a reduction of RMSE. Thus, it was generally concluded that the methodological procedures tested improved the performance of the models in predicting soil properties and that, from an analytical and agronomic point of view, soil analyses can be used as complementary methods.

**Keywords:** proximal hyperspectral sensing, organic carbon, soil texture, soil color, multivariate models.

---

<sup>2</sup>Doctoral thesis in Soil Science. Graduate Program in Soil Science, Faculty of Agronomy, Universidade Federal do Rio Grande do Sul. Porto Alegre. (94 p.). September, 2020.

## SUMÁRIO

Página

<b>1. INTRODUÇÃO GERAL .....</b>	<b>1</b>
<b>2. CAPÍTULO I - REVISÃO BIBLIOGRÁFICA.....</b>	<b>5</b>
2.1. Espectroscopia de reflectância do solo.....	5
2.2. Espectroscopia aplicada à análise da textura do solo .....	8
2.3. Espectroscopia aplicada a análise do teor de carbono orgânico do solo .	11
2.4. Parâmetros de cor aplicados na estratificação de amostras de solos .....	14
2.5. Estratificação de amostras de solo na espectroscopia .....	17
2.6. Modelos de calibração multivariados .....	20
<b>3. CAPÍTULO II – ESTUDO 1: EFEITOS DA ESTRATIFICAÇÃO DAS AMOSTRAS NO DESEMPENHO DOS MODELOS DE PREDIÇÃO DA TEXTURA DO SOLO E DO CARBONO ORGÂNICO DO SOLO<sup>2</sup> .....</b>	<b>25</b>
3.1. INTRODUÇÃO .....	25
3.2. MATERIAL E MÉTODOS.....	30
3.2.1. Caracterização da área de estudo .....	30
3.2.2. Medidas de reflectância espectral e análise laboratorial .....	32
3.2.3. Cálculo dos parâmetros de cor baseados em Vis.....	33
3.2.4. Análises estatísticas descritiva e multivariadas .....	35
3.2.5. Análise de componentes principais (ACP).....	36
3.2.6. Análise de clusterização das amostras.....	37
3.2.7. Modelagem preditiva e validação.....	37
3.3. RESULTADOS E DISCUSSÃO.....	39
3.3.1. Estatísticas descritivas.....	39
3.3.2. Análise de componentes principais (ACP).....	40
3.3.3. Análise de cluster.....	42
3.3.4. Desempenho de predição dos modelos espectrais .....	45
3.3.5. Predição das propriedades do solo utilizando os clusters e o MSE .....	48
3.4. CONCLUSÕES .....	52
<b>4. CAPÍTULO III – ESTUDO 2: PREDIÇÃO DA TEXTURA DO SOLO A PARTIR DA ESTRATIFICAÇÃO DE UMA BIBLIOTECA ESPECTRAL REGIONAL DE SOLOS .....</b>	<b>53</b>
4.1. INTRODUÇÃO .....	53
4.2. MATERIAL E MÉTODOS .....	56

4.2.1. Área de estudo e biblioteca espectral de solos.....	56
4.2.2. Medidas de reflectância espectral e pré-processamento.....	57
4.2.3. Cálculo dos parâmetros de cor baseados em Vis.....	58
4.2.4. Análise de componentes principais (ACP).....	58
4.2.5. Análise de clusterização das amostras.....	59
4.2.6. Pré-processamento e construção dos modelos.....	59
4.2.7. Derivadas de ordem fracionária.....	61
4.2.8. Precisão do modelo .....	61
<b>4.3. RESULTADOS E DISCUSSÃO.....</b>	<b>62</b>
4.3.1. Análise descritiva .....	62
4.3.2. Análise de componentes principais (ACP).....	63
4.3.4. Análise das curvas espectrais.....	65
4.3.5. Derivadas fracionárias dos conjuntos estratificados .....	66
4.3.6. Modelos iniciais (estratificados e não estratificados).....	68
4.3.7. Validação entre grupos .....	72
<b>4.4. CONCLUSÃO.....</b>	<b>74</b>
<b>5. CAPÍTULO IV – ESTUDO 3: PREDIÇÃO DE PROPRIEDADES DO SOLO USANDO ESPECTROSCOPIA Vis-NIR-SWIR E VARIÁVEIS AMBIENTAIS .....</b>	<b>76</b>
5.1. INTRODUÇÃO .....	76
5.2. MATERIAL E MÉTODOS.....	78
5.2.1. Área de estudo e coleta de dados .....	78
5.2.2. Medidas de reflectância espectral e análise laboratorial .....	80
5.2.3. Combinação das variáveis ambientais com as curvas espectrais .....	80
5.2.4. Modelagem preditiva e validação.....	82
<b>5.3. RESULTADOS E DISCUSSÃO.....</b>	<b>83</b>
5.3.1. Análise descritiva .....	83
5.3.2. Análise de correlação das variáveis ambientais .....	85
5.3.3. Desempenho dos modelos de predição.....	87
5.3.4. Importância das bandas na predição (IBP).....	88
<b>5.4. CONCLUSÕES .....</b>	<b>91</b>
<b>6. CAPÍTULO V – CONSIDERAÇÕES FINAIS.....</b>	<b>92</b>
<b>7. REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>94</b>

## RELAÇÃO DE TABELAS

Página

<b>Tabela 1.</b> Principais feições de absorções e a ocorrência de sobretons e combinações na região do Vis-NIR. ....	7
<b>Tabela 2.</b> Principais estratégia que já foram utilizadas para estratificar as amostras de solos na predição de propriedades por espectroscopia. ....	18
<b>Tabela 3.</b> Compilação de estudos que aplicaram diferentes tipos de pré-processamento e modelos multivariados na estimativa do teor de COS e das frações argila, areia e silte utilizando ER. ....	27
<b>Tabela 4.</b> Parâmetros de cores baseados em Vis derivados de diferentes modelos de espaço de cores calculados usando o software Munsell Conversion (WALLKILLCOLOR, 2019). ....	34
<b>Tabela 5.</b> Estatística descritiva das propriedades do solo e dos parâmetros de cor. ....	40
<b>Tabela 6.</b> Avaliações da clusterização fuzzy k-means e escolha do número ideal de clusters. ....	42
<b>Tabela 7.</b> Desempenho preditivo das propriedades do solo para o conjunto de validação considerando o MSE e a performance dos modelos após a análise de cluster baseado nos parâmetros de cor. ....	46
<b>Tabela 8.</b> Diferença estatística e redução do erro de predição nos resultados de predição dos métodos PLSR e SVM para cada propriedade do solo. ....	49
<b>Tabela 9.</b> Estatística das frações granulométricas do solo a partir dos clusters adquiridos pela análise de estratificação. ....	63
<b>Tabela 10.</b> Resultados de validação dos modelos estratificados e não estratificados para a textura do solo (%) usando diferentes tipos de pré-processamentos. ....	69
<b>Tabela 11.</b> Variáveis ambientais derivadas do modelo digital do terreno. ....	81
<b>Tabela 12.</b> Análises descritivas dos conjuntos de treinamento e validação das propriedades do solo utilizadas na modelagem. ....	84
<b>Tabela 13.</b> Resultado da validação dos modelos preditivos, considerando os três cenários de predição. ....	88

## RELAÇÃO DE FIGURAS

	Página
<b>Figura 1.</b> Módulo de pesquisa e localização das parcelas onde foram coletadas as amostras de solo no Parque Nacional de São Joaquim, no Estado de Santa Catarina. ....	31
<b>Figura 2.</b> Sequência metodológica. ....	35
<b>Figura 3.</b> Resultado da análise de componentes principais dos parâmetros de cor. ....	41
<b>Figura 4.</b> Distribuição estatística e variabilidade do teor de COS (a), argila (b), areia (c) e silte (d) para os modelos sem estratificação (MSE) e clusters. ....	43
<b>Figura 5.</b> Espectros médios de reflectância dos clusters com base nos parâmetros de cor (a) e a primeira derivada das curvas espectrais, destacando bandas espectrais relacionadas aos componentes do solo e as alterações na intensidade da reflectância (b). ....	45
<b>Figura 6.</b> Performance dos modelos de predição considerando os modelos sem estratificação (MSE) e a performance dos modelos após a análise de cluster com base nos parâmetros de cor. Valores de mediana, 1° e 2° quartil. ....	50
<b>Figura 7.</b> Localização dos pontos de coleta de solos no Estado de Santa Catarina, Brasil. ....	57
<b>Figura 8.</b> Resultado da análise de componentes principais dos parâmetros que melhor explicaram a variação das informações do solo com base na cor do solo (a, b). Contribuição (loadings) dos parâmetros nas duas primeiras CPs (c, d). ....	65
<b>Figura 9.</b> Espectros médios de reflectância dos clusters, destacando bandas espectrais relacionadas aos componentes do solo e as alterações na intensidade da reflectância. ....	66
<b>Figura 10.</b> Curvas espectrais dos diferentes clusters e derivadas fracionárias: C1 com derivadas fracionárias de 0.5, 1.5 e 2 (a); C2 com derivadas fracionárias de 0.5, 1.5 e 2 (b); C3 com derivadas fracionárias de 0.5, 1.5 e 2 (c); C4 com derivadas fracionárias de 0.5, 1.5 e 2 (d). As áreas em azul representam os desvios padrão das curvas espectrais. ....	68

<b>Figura 11.</b> Validação de dados medidos em laboratório em relação às predições Vis-NIR-SWIR para a argila (a, b, c, d) com os clusters C1, C2, C3 e C4, respectivamente, e para a areia (e, f, g, h) com os clusters C1, C2, C3 e C4, respectivamente. ....	72
<b>Figura 12.</b> Comparação do RMSE (a), $R^2$ (b) e RPIQ (c) dos modelos iniciais na predição da areia e argila para a validação de sua aplicabilidade entre os clusters com maior e menor DP. (Cc1 = calibração usando o cluster 1; Cc3 = calibração usando o cluster 3; Cc4 = calibração usando o cluster 4; Vc1 = predição usando o cluster 1; Vc3 = predição usando o cluster 3; Vc4 = predição usando o cluster 4. ....	73
<b>Figura 13.</b> Curva espectrais do solo Vis-NIR-SWIR (400–2500 nm) suavizadas com filtro Savitzky-Golay. ....	85
<b>Figura 14.</b> Correlação de Pearson entre as variáveis ambientais e as propriedades do solo (a, b). As cores em azul e vermelho simbolizam correlações positivas e negativas, respectivamente. ....	86
<b>Figura 15.</b> Importância das bandas na predição (IBP) e os b-coeficientes do modelo PLSR na predição do teor de COS e das frações argila e areia utilizando espectroscopia Vis-NIR-SWIR. (a) na predição do teor de COS, o desvio padrão foi de $\sigma = \pm 2,93$ , (b) na predição da fração argila, o desvio padrão foi de $\sigma = \pm 3,51$ e (c) na predição da fração areia, o desvio padrão foi de $\sigma = \pm 3,59$ . ....	90

## RELAÇÃO DE ABREVIATURAS

- ACP:** Análise de componentes principais
- BESB:** Biblioteca Espectral de Solos do Brasil
- CBT:** *Cubist*
- CIE:** *Commission Internationale de l'Eclairage*
- COS:** Carbono Orgânico do Solo
- CR:** *Continuum Removed*
- DOFs:** Derivadas de Ordem Fracionária
- EMBRAPA:** Empresa Brasileira de Pesquisa Agropecuária
- EPAGRI:** Empresa de Pesquisa Agropecuária e Extensão Rural de SC
- ER:** Espectroscopia de reflectância
- FDSG:** *Savitzky–Golay first derivative*
- FKM:** *Fuzzy K-means*
- GPR:** *Gaussian Process Regression*
- MDT:** Modelo Digital do Terreno
- MEs:** Modelos Estratificados
- MSC:** *Multiplicative Scatter-Correction*
- MSE:** Modelos sem estratificados
- NBR:** *Normalizations by Range*
- PLSR:** *Partial Least Square Regression*
- R<sup>2</sup>:** Coeficiente de determinação
- RMSE:** Raiz do Erro Quadrático Médio
- RPIQ:** Intervalo Interquartil
- RF:** *Random Forest*
- SDSG:** *Savitzky–Golay second derivative*
- SNV:** *Standard Normal Variate*
- SVM:** *Support Vector Machines*
- VAs:** Variáveis Ambientais

## 1. INTRODUÇÃO GERAL

Para conhecer melhor o solo, tem-se aumentado a densidade espacial de informações e buscado estratégias para melhorar a acurácia dos modelos de predição de propriedades do solo. Assim, para suprir essa demanda por informações que auxiliem no processo de tomada de decisão relacionada à expansão agrícola e em questões relacionadas com a preservação ambiental, é preciso desenvolver métodos efetivos para medir e monitorar esse recurso (VISCARRA ROSSEL et al., 2016).

A espectroscopia de reflectância (ER) é uma técnica utilizada para caracterizar e identificar as propriedades do solo, fornece uma análise rápida, com boa relação custo-benefício, não destrutiva, sem uso de reagentes químicos e, principalmente, uma técnica não poluente, sendo um apoio à gestão sustentável do meio ambiente (JI et al., 2016; NOCITA et al., 2015; STENBERG et al., 2010; VISCARRA ROSSEL; BEHRENS, 2010). Segundo STONER; BAUMGARDNER (1981), a reflectância do solo é uma propriedade cumulativa derivada do comportamento espectral inerente da combinação heterogênea de seus componentes, como tamanho de partícula, estrutura do solo, rugosidade superficial, umidade, teor de matéria orgânica (MO), teor de minerais carbonatos, presença ou não de quartzo e dos óxidos de ferro, entre outros. Essa relação dependente entre propriedade do solo e as bandas espectrais produz uma assinatura de reflectância espectral única que pode ser descrita e distinguida (BEN-DOR et al., 2008; CLARK, 1999; DEMATTÊ et al., 2004; STENBERG et al., 2010). As faixas espectrais comumente utilizadas na estimativa de propriedades do solo correspondem a 400-700 nm (visível-Vis), 700-1100 nm (infravermelho próximo-NIR) e 1100-2500 nm (infravermelho de ondas curtas - SWIR) (DEMATTE et al., 2019; VISCARRA ROSSEL; WEBSTER, 2012). Estas faixas estão sendo amplamente utilizadas para analisar os solos, porque correlacionam eficientemente os componentes químicos com suas características espectrais de absorções específicas (STENBERG et al., 2010;

VISCARRA ROSSEL et al., 2006a). As características de absorção na faixa do Vis são dominadas por excitações moleculares eletrônicas (sobretons) e as da faixa do NIR-SWIR contêm uma combinação de vibrações moleculares de alta qualidade.

Nesse contexto, o comportamento espectral do solo é diretamente dependente da sua composição química, biológica, física e mineralógica (SOUSA JUNIOR; DEMATTÊ; ARAÚJO, 2011) resultante da interação dos diversos fatores e processos que atuam na sua formação. Essa interação fornece resultados satisfatórios para a caracterização dos constituintes orgânicos e inorgânicos do solo (BEN-DOR; BANIN, 1995; VISCARRA ROSSEL et al., 2006b), permitindo a predição de várias propriedades do solo, como as frações areia, silte e argila (DEMATTÊ et al., 2019; DOTTO et al., 2017; JACONI; VOS; DON, 2019; MARGENOT et al., 2020; SILVA et al., 2019; VASAVA et al., 2019), o teor de COS (ALLO et al., 2020; ALLORY et al., 2019; HUTENGS et al., 2019; MARGENOT et al., 2020; MOURA-BUENO et al., 2019), a capacidade de troca de cátions e o pH (BROWN et al., 2006; VISCARRA ROSSEL et al., 2016), os óxidos de ferro (BAHIA et al., 2015; RAMOS et al., 2020) e o nitrogênio (AITKENHEAD et al., 2013a), entre outras.

Considerando que a acurácia dos modelos é influenciada pela composição dos solos que integram as bibliotecas espectrais e visando refinar a aplicação da técnica de ER em laboratórios de análise de solos, vê-se a necessidade de estudos que levem em consideração a heterogeneidade dos solos em relação as frações areia, silte e argila e do teor de COS. Além disso, a construção de modelos com conjuntos de dados mais homogêneos pode melhorar a capacidade preditiva dos modelos (ARAÚJO et al., 2014; GUERRERO et al., 2014; MOURA-BUENO et al., 2019).

As bibliotecas espectrais representam um banco de dados que integram as informações químicas e físicas dos solos, onde as informações qualitativas são representadas por curvas espectrais, possibilitando a construção de modelos de calibração mais robustos (BROWN, 2007; CAMBULE et al., 2012; VISCARRA ROSSEL; WEBSTER, 2012). Os dados que compõem as bibliotecas espectrais, geralmente são dados legados que fazem parte de acervos de centros de pesquisa, universidades, entre outros, e representam uma rica fonte de informações dos solos (VISCARRA ROSSEL et al., 2016). Assim, são

necessários estudos utilizando bibliotecas espectrais de diferentes escalas que possam representar a heterogeneidade dos solos.

No Estado de Santa Catarina (SC), mais de 50% do território encontra-se recoberto por rochas da formação Serra Geral, formada sobre uma geologia composta por rochas magmáticas extrusivas. Nesse contexto ambiental, os solos são formados a partir do basalto, diabásio, riolito e riolito, que são rochas de composição básica a ácida, que variam em sua composição quanto ao teor de óxidos de silício e quanto a teores de óxidos de ferro e manganês. O Estado também possui solos desenvolvidos de rochas sedimentares do Grupo Passa Dois (EMBRAPA, 2004). Assim, os diferentes materiais de origem e os fatores de formação dos solos de SC são responsáveis pelo desenvolvimento de solos com diferentes composições físico-química e mineralogia, compreendendo uma grande variabilidade de texturas e de composição, incluindo o teor de carbono orgânico do solo (COS).

Além do uso da ER utilizando um conjunto de dados de solos mais homogêneos, outra possibilidade para melhorar o desempenho dos modelos de predição é a utilização de outras variáveis preditoras associadas às curvas espectrais. Poucos estudos tentaram avaliar o uso de uma segunda variável preditora combinada com às curvas espectrais, como nos estudos propostos por BROWN et al. (2006); PENG et al. (2015); STEVENS et al. (2013) que incluíram uma segunda variável preditora nos modelos de predição com o objetivo de melhorar a capacidade preditiva dos modelos. Assim, uma possibilidade é a utilização de variáveis ambientais (VAs) derivadas dos modelos digitais do terreno (MDT), que podem ser empregadas para caracterizar o ambiente nas mais variadas escalas (BEHRENS et al., 2010; COELHO, 2010; GIASSON et al., 2013; TEN CATEN et al., 2012).

Diante do exposto, os objetivos gerais desta tese foi estudar os efeitos que a estratificação das amostras pela cor do solo provoca no desempenho dos modelos e estudar os efeitos da inclusão de uma segunda variável preditora (variáveis ambientais) à matriz espectral.

Os seguintes objetivos específicos foram definidos: (i) usar os parâmetros de cor baseados no espectro visível (Vis) combinados com estatística multivariada para estratificar amostras de solo de uma biblioteca espectral local e comparar os resultados de diferentes modelos de predição de

COS, argila, areia e silte, (ii) aplicar os parâmetros de cor baseados em Vis na estratificação de uma biblioteca espectral regional de solos; avaliar o desempenho dos modelos preditivos gerados a partir da estratificação da biblioteca espectral de solo e explicar as possíveis mudanças na precisão da predição com base nos modelos estratificados, (iii) combinar as variáveis ambientais com as curvas espectrais Vis-NIR-SWIR para aumentar o poder de predição e alcançar predições mais acuradas das propriedades teor de COS, areia, argila e silte. Assim, a tese está estruturada da seguinte forma:

Capítulo I: Revisão bibliográfica sobre a aplicação da espectroscopia de reflectância (ER) na análise de solo, uso da ER aplicada na predição da textura do solo e do teor de COS, estratificação de amostras de solos na espectroscopia, principais modelos de calibração utilizados para correlacionar os espectros de reflectância com os dados de propriedades do solo medidos em laboratório e os principais parâmetros de cor aplicados na estratificação de amostras de solos.

Capítulo II: Estudo 1 - Efeitos da estratificação das amostras no desempenho dos modelos de predição da textura do solo e do carbono orgânico do solo;

Capítulo III: Estudo 2 - Predição da textura do solo a partir da estratificação de uma biblioteca espectral regional de solos;

Capítulo IV: Estudo 3 - Predição de propriedades do solo usando espectroscopia Vis-NIR-SWIR e variáveis ambientais.

## **2. CAPÍTULO I - REVISÃO BIBLIOGRÁFICA**

### **2.1. Espectroscopia de reflectância do solo**

A espectroscopia de reflectância (ER) é a uma técnica que estuda a interação da radiação eletromagnética com a matéria. Essa interação ocorre a partir de uma série de efeitos físicos que dependem da energia contida na radiação incidente e da constituição da matéria (NIEMZ, 2015). A reflectância do solo pode ser quantificada medindo a porção de energia incidente que é refletida em cada comprimento de onda, que é chamada de reflectância espectral do solo (LILLESAND; KIEFER; CHIPMAN, 2015).

As características da reflectância espectral do solo dependem da combinação heterogênea de seus componentes. Da energia eletromagnética incidente sobre a superfície do solo, uma parte será absorvida e outra será refletida e/ou transmitida. Apenas uma porção da energia emitida pela fonte de radiação retorna ao sensor, sendo o restante da energia perdida ou absorvida pelos constituintes do solo. A absorção da energia pelo solo nos diversos comprimentos de onda ocorre em função dos diversos materiais componentes do solo, os quais individualmente absorvem ou refletem a energia de forma diferente nos diferentes comprimentos de onda (BAUMGARDNER et al., 1986; STONER; BAUMGARDNER, 1981). A energia eletromagnética refletida pode ser representada como a relação entre a radiação espectral refletida pela superfície da amostra de solo e a radiação espectral refletida pelo material de referência. Essa interação de absorção de energia pode ser relativa à mudança de orbital energético dos elétrons a um nível de energia mais elevado ou aos processos vibracionais dos grupos funcionais presentes nas moléculas. Esse processo de promoção de elétrons ocorre quando o átomo ou a molécula absorvem a quantidade exata de energia que corresponde à diferença de energia entre os orbitais (STENBERG et al., 2010).

Na região do Vis, a promoção de elétrons a partir da absorção da radiação eletromagnética ocorre entre os níveis de energia eletrônicos (sobretens), estimulando um processo interno no átomo ou nas moléculas de mudança de estado de energia. Quando um átomo ou molécula absorve energia, um elétron é promovido de um orbital ocupado para um orbital desocupado de maior energia potencial (HUNT; SALISBURY, 1970). Na região NIR-SWIR, as radiações são menos energéticas (menor frequência e maior comprimento de onda). Dessa forma, não existe energia suficiente para promover os elétrons para um orbital mais energético. Em vez disso, as ligações entre os átomos podem vibrar de diferentes formas, como no caso da radiação no comprimento de onda do NIR-SWIR (NIEMZ, 2015). Assim como ocorre nas interações Vis com a matéria, as absorções de NIR-SWIR também são quantificáveis. A energia de um fóton deve corresponder exatamente à energia adequada da vibração de uma das ligações da molécula. Cada ligação química diferente na molécula tem as suas próprias vibrações e cada ligação química C-H, O-H, N-H, C=O e Al-OH pode sofrer uma série de tipos diferentes de vibrações. No entanto, a interpretação dos espectros Vis-NIR-SWIR pode ser prejudicada quando o constituinte (elemento químico) está presente em pequenas quantidades no solo, podendo ter sua detecção comprometida. A Tabela 1 apresenta a compilação das principais feições de absorções e a ocorrência de sobretens e combinações na região do Vis-NIR usados na interpretação das curvas espectrais do solo. Assim, a partir da interpretação das curvas espectrais é possível identificar feições relacionadas a materiais orgânicos e inorgânicos presentes no solo.

Dada a influência da composição mineralogia e das propriedades físico-químicas no comportamento espectral do solo, alguns estudos utilizaram amostras e bibliotecas espectrais que representam a variabilidade dos solos em diferente escalas (ARAÚJO et al., 2014; DEMATTÊ et al., 2019; GHOLIZADEH et al., 2016; NOCITA et al., 2014; RAMIREZ-LOPEZ et al., 2013; STENBERG et al., 2010; VISCARRA ROSSEL et al., 2016; VISCARRA ROSSEL; CHEN, 2011). Essas bibliotecas espectrais geralmente integram um grande número de amostras o que pode aumentar a variabilidade total do conjunto dos solos, podendo comprometer a capacidade preditiva dos modelos.

Tabela 1. Principais feições de absorções e a ocorrência de sobretons e combinações na região do Vis-NIR.

Constituintes do solo	Região Vis-NIR (nm)	Sobretons e combinações de bandas Vis-NIR (nm)
Água	1915	$V_2+V_3$
	1455	$2V_2+V_3$
	1380	$V_1+V_3$
	1135	$V_1+V_2+V_3$
	940	$2V_1+V_3$
Hidroxisilas	1400	$2V_1$
	930	$3V_1$
	700	$4V_1$
<b>Óxidos de ferro</b>		
Goethita	434, 480	eletrônico
	650, 920	transições
Hematita	404, 444, 529	eletrônico
	650, 884	transições
<b>Minerais da fração argila</b>		
Caulinita	1395	$2V_{1a}$
	1415	$2V_{1b}$
	2160	$V_{1a}+\delta$
	2208	$V_{1b}+\delta$
Esmectita	2206	$V_1+\delta_a$
	2230	$V_1+\delta_b$
Ilita	2206	$V_1+\delta$
	2340	
Carbonatos	2450	indefinido
	2336	$3V_3$
<b>Compostos orgânicos</b>		
Aromáticos	1650	$2V_1$
	1100	$3V_1$
	825	$4V_1$
	2060	$V_1+\delta$
Aminas	1500	$2V_1$
	1000	$3V_1$
	751	$4V_1$
Ácido carboxílico	1930	$3V_1$
	1449	$4V_1$
Amida	2033	$3V_1$
	1524	$4V_1$
Alifático	2275	$3V_1$
	1706	$4V_1$
Metil	2307–2469	$3V_1$
	1730–1852	$4V_1$
Fenólicos	1961	$4V_1$
Polissacarídeos	2137	$4V_1$
Carboidratos	2381	$4V_1$

Fonte: Compilação adaptada do estudo de VISCARRA ROSSEL; BEHRENS (2010).

## 2.2. Espectroscopia aplicada à análise da textura do solo

A análise granulométrica permite classificar os componentes sólidos em classes, de acordo com os seus diâmetros. Dentre esses componentes estão as frações areia, silte e argila, que correspondem as dimensões de 2-0,05, 0,05-0,002 e <0,002 mm, respectivamente (KLEIN et al., 2010; TÜMSAVAŞ et al., 2019).

A areia é a fração granulométrica do solo que apresenta baixa reatividade com o meio, isto se deve a sua baixa superfície específica, possuindo pouca capacidade de reter água ou nutrientes. Devido à essas características, os solos arenosos possuem alta porosidade e baixa capacidade de reter água e nutrientes, sendo também inférteis e propensos à seca. A fração silte apresenta a forma e a composição mineral semelhantes às da fração areia, no entanto, apresentam partículas menores, o que lhes confere maior área superficial específica. Devido ao menor tamanho das partículas, os solos siltosos retém mais água e tem menor capacidade de drenagem do que os solos arenosos (MELO; ALLEONI, 2009). As partículas de argila que compõem a fração argila dos solos são menores, conferindo maior área superficial específica e maior reatividade com o meio, a qual confere a exposição de grupos funcionais de superfície nas arestas dos minerais. Solos com maiores teores de argila apresentam maior número de cargas elétricas, provenientes da substituição isomórfica na estrutura dos minerais, as quais interagem com os íons presentes no solo (e.g.  $\text{NH}_4^+$ ,  $\text{K}^+$ ,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ , etc.) (MELO; ALLEONI, 2009). A presença em alta quantidade desta fração confere ao solo características físicas como maior porosidades total, baixa permeabilidade e alta capacidade de retenção de água (KLEIN et al., 2010).

As técnicas laboratoriais tradicionais utilizadas na determinação dos teores das frações minerais do solo e, conseqüentemente, da textura do solo, empregam métodos convencionais de análise física do solo, incluindo, principalmente, os métodos do hidrômetro e da pipeta (DONAGEMMA et al., 2017). Embora esses métodos sejam relativamente precisos e amplamente aceitos, eles demandam a preparação das amostras e o uso de reagentes químicos, além de serem métodos demorados, dificultando a execução das análises quando estas envolvem um grande número de amostras.

Nesse contexto, a ER surge como uma alternativa as análises convencionais, apresentando como vantagem ser um método rápido, não destrutivo, sem uso de reagente químicos, com menor necessidade de preparação das amostras e com boa relação custo-benefício (ALLO et al., 2020; JACONI; VOS; DON, 2019; VISCARRA ROSSEL et al., 2016). A ER está se destacando na determinação de importantes propriedades do solo (JACONI; DON; FREIBAUER, 2017; NOCITA et al., 2014; VISCARRA ROSSEL; WEBSTER, 2012), incluindo a textura (ARAÚJO et al., 2014; DEMATTÊ et al., 2019; NANNI et al., 2017; PINHEIRO et al., 2017; SILVA et al., 2019; VASAVA et al., 2019; XU et al., 2018; ZHANG et al., 2017). Nesse contexto, as curvas espectrais contêm informações que permitem a estimativa das frações areia, silte e argila.

Na fração areia, o quartzo é o mineral dominante e não tem características de absorção na região Vis-NIR-SWIR, aumentando a reflectância em todo o espectro (DEMATTÊ et al., 2007; WIGHT; ASHWORTH; ALLEN, 2016). Em contrapartida, a fração argila influencia às características de absorção espectral, especialmente na faixa do Vis (480 e 550 nm) relacionadas aos óxidos de ferro (DALMOLIN et al., 2005) e na faixa SWIR (1400 e 1900 nm) devido às vibrações de estiramento de OH<sup>-</sup> e H<sub>2</sub>O mantida nas estruturas dos argilominerais (BEN-DOR et al., 2008; GHOLIZADEH et al., 2016). DEMATTÊ et al. (2014) observaram uma diminuição na reflectância do solo com um aumento do conteúdo da fração argila e, inversamente, a reflectância aumentou com o aumento do conteúdo da fração areia, atribuindo isso ao efeito que os minerais presentes nas frações texturais causam na reflectância do solo. Mais recentemente, XU et al. (2018) avaliaram a aplicabilidade do uso da espectroscopia Vis-NIR na estimativa da textura do solo e identificaram as bandas de maior importância na predição das propriedades do solo. De acordo com os autores, as bandas em aproximadamente 1400, 1900 e 2200 nm foram importantes na estimativa das frações areia, silte e argila. Além dessas regiões, a banda em torno de 480 nm foi significativa na predição das frações areia e argila. Da mesma forma, JACONI; VOS; DON (2019) avaliaram o desempenho de estimativas da textura do solo utilizado ER a partir de um conjunto de dados de solo em escala nacional e concluíram que as frações texturais obtiveram boa predição. No entanto, os autores identificaram que a estimativa da fração argila

superou as das frações areia e silte, o que reflete a influência direta dos minerais da fração argila nas propriedades de absorção do solo.

As curvas espectrais normalmente contêm feições de absorção a aproximadamente 2200 nm, que é uma característica de absorção dos minerais silicatados (CHABRILLAT et al., 2002). DUFRECHOU; GRANDJEAN; BOURGUIGNON (2015) ao analisarem os parâmetros geométricos (profundidade, posição, assimetria e largura) das bandas de absorção da fração argila, identificaram que as feições de absorção em torno de 1400, 1900 e 2200 nm são influenciadas pelos argilominerais, principalmente illita, montmorilonita e caulinita, respectivamente. Além disso, as frações de tamanho de partícula do solo também influenciam a linha de base das curvas espectrais (WORKMAN; SHENK, 2015).

Em geral, as pesquisas tem mostrado que as feições de absorção estão principalmente relacionadas aos minerais da fração argila, como no trabalho de HUNT; SALISBURY (1970) que relataram que o óxido de ferro mostrou feição de absorção próxima a 550 nm, relacionado, principalmente, à hematita ( $\alpha\text{-Fe}_2\text{O}_3$ ). REZENDE (1980) observou um padrão côncavo-convexo em torno de 440 a 470 nm, relacionado à goethita ( $\alpha\text{-FeOOH}$ ). STEVENS et al. (2013) relataram feições de absorção relacionadas a óxidos de ferro em torno de 540, 640 e 900 nm (goethita em 620 e 920 nm). DEMATTÊ et al. (2017) identificaram feições de absorção relacionadas a goethita em torno de 425, 480 e 600 nm e de hematita em torno de 750 e 1050 nm, e a característica de absorção do Fe cristalino em torno de 850 nm. FANG et al. (2018) ao realizarem uma revisão sobre os principais minerais da fração argila que apresentam feições de absorção na região do Vis, identificaram um pico positivo na região em torno de 523 nm, referentes à hematita. Mais recentemente, SAHWAN et al. (2020) relataram uma alta correlação de comportamentos espectrais e variações de cor com os óxidos de ferro. Os autores identificaram que a influência do teor de óxidos de ferro é claramente aparente na região do Vis. Além disso, eles relataram que a presença de  $\text{CaCO}_3$  aumentou a reflectância em todas as faixas espectrais.

DEMATTÊ et al. (2017) também identificaram que as feições de absorção em torno de 1400, 1900 e 200 nm estão relacionadas a grupos  $\text{OH}^-$  e vibrações da molécula de água, e a absorção em torno de 2200 nm indica a

presença de caulinita. As feições de absorção das moléculas de água em torno de 1400 e 1900 nm são características de espectros de solo e os comprimentos de onda em torno de 2200 nm são características de minerais de argila (CLARK, 1999). De acordo com THENKABAIL et al. (2011), existem três feições de maior absorção relacionadas a minerais de argila em geral, por volta de 1300-1400, 1800-1900 e 2200-2500 nm. Para caulinita, mineral 1:1, o sinal de absorção em torno de 1400 e 2200 nm é relativamente forte, enquanto o sinal em 1900 nm é muito fraco.

Essas bandas de absorção na região Vis-NIR-SWIR podem ser usadas para identificar os constituintes que compõem a textura do solo, o que fundamenta quimicamente a sua análise quantitativa.

### **2.3. Espectroscopia aplicada a análise do teor de carbono orgânico do solo**

O solo participa do ciclo global do carbono e, juntamente com as plantas, atua como um sumidouro de CO<sub>2</sub> atmosférico através da fotossíntese. Cerca de metade do carbono é devolvido a atmosfera na forma de CO<sub>2</sub> e a outra metade é devolvido ao solo na forma de dejetos ou tecidos orgânicos, sendo, portanto, integrados ao solo como matéria orgânica, que é composta de aproximadamente 50% a 58% de carbono (GOMEZ et al., 2020).

O teor de COS pode ter uma grande variabilidade em áreas com diferentes altitudes e declividades, assim, quando se efetuam amostragens para a caracterização de sua variabilidade espacial, é necessário aumentar a densidade espacial de coleta de dados para se obter uma amostragem mais representativa (BARTHOLOMEUS et al., 2011). No entanto, a análise de grandes quantidades de amostras de solo ainda não é uma prática economicamente viável devido aos altos custos associados às análises químicas laboratoriais (NANNI et al., 2011). Portanto, é importante desenvolver técnicas que sejam menos onerosas. A ER pode ser uma destas técnicas, desde que se disponha de modelos confiáveis e com boa relação custo-benefício para prever o teor de COS (ZERAATPISHEH et al., 2019).

A ER na região do visível (Vis, 400-700 nm), do infravermelho próximo (NIR, 700-1100 nm) e de ondas curtas (SWIR, 1100-2500 nm) é um conceito de correlação que explora a relação quantitativa entre os espectros de reflectância

com os dados de propriedades do solo medidos em laboratório e obtém medições de baixo custo, rápida e em larga escala a nível nacional e regional (ALLO et al., 2020; GUPTA; ISLAM; HASAN, 2018; PABÓN; SOUZA FILHO; OLIVEIRA, 2019; STEVENS et al., 2013). Muitos estudos destacaram o potencial da espectroscopia Vis-NIR-SWIR para prever várias propriedades do solo, incluindo o teor de COS (ALLO et al., 2020; ALLORY et al., 2019; HUTENGS et al., 2019; MARGENOT et al., 2020; VISCARRA ROSSEL et al., 2006b).

A maioria dos trabalhos sobre a quantificação de COS por espectroscopia Vis-NIR-SWIR testaram diferentes técnicas de pré-processamento e modelos de calibração com o objetivo de melhorar a eficiência das estimativas do teor de COS (GRUNWALD; VASQUES; RIVERO, 2015). Técnicas de pré-processamento espectral são tratamentos matemáticos utilizados com a finalidade de eliminar ruídos, enfatizar recursos e destacar às características de maior interesse do sinal espectral (GRUNWALD; VASQUES; RIVERO, 2015). ALLORY et al. (2019) observaram melhores estimativas do teor de COS utilizando uma combinação de técnicas de pré-processamento espectral (SG + SNV), com  $R^2$  de 0,83 e RMSE de 0,5%, mostrando melhor desempenho quando se utilizou apenas uma técnica de pré-processamento. Mais recentemente, ALLO et al. (2020) testaram diferentes técnicas de pré-processamento e o método de calibração PLSR em um conjunto de dados com teor de COS variando de 0,4 a 13% e observaram que a combinação das técnicas de pré-processamento (DET+FDSG) obtiveram melhor desempenho que as demais, com  $R^2$  de 0,94 e RMSE de 0,7%. Da mesma forma, HUTENGS et al. (2019), em escala local, com 90 amostras, aplicaram o modelo PLSR e o pré-processamento MSC, em um conjunto de dados com o teor de COS variando de 1,39 a 2,76% e encontraram  $R^2$  de 0,66 e RMSE de 0,2%.

Além disso, outros estudos também identificaram os principais grupos funcionais que possuem feições de absorções ao longo das curvas espectrais, como no trabalho de MARGENOT et al. (2020), que utilizaram a espectroscopia na estimativa do teor de COS e da textura do solo e identificaram os grupos funcionais associados a essas propriedades do solo. Segundo os autores, as feições de absorção estavam mais associadas a grupos funcionais inorgânicos (minerais), como os filossilicatos OH, SiO<sub>2</sub> (por exemplo o quartzo), enquanto as feições de absorção associadas a grupos funcionais orgânicos estavam mais

relacionadas a C-H aromático. Igualmente, NAWAR; MOUAZEN (2018) ao utilizarem a espectroscopia na predição do teor de COS, identificaram que as absorções em torno de 1035 e 1680 nm estavam associadas a C-H aromático e contribuíram para a predição do teor de COS.

Além dos estudos relacionados a avaliação das técnicas de pré-processamento, dos modelos multivariados e da identificação dos principais grupos funcionais, outros estudos também realizaram a determinação dos comprimentos de onda mais importantes na predição do teor de COS (BARTHOLOMEUS et al., 2008; LI et al., 2012; NAWAR; MOUAZEN, 2018; NOCITA et al., 2014; VISCARRA ROSSEL et al., 2006b; WANG et al., 2010). VISCARRA ROSSEL et al. (2006b) relataram que os comprimentos de onda em torno de 410, 570 e 660 nm foram importantes na estimativa do teor de COS. WANG et al. (2010) relataram cinco bandas de comprimento de onda mais importantes (440, 560, 625, 740 e 1336 nm) na predição do teor de COS em condições de laboratório. Igualmente, BARTHOLOMEUS et al. (2008) utilizaram a ER na estimativa do teor de COS e identificaram que as absorções em torno de 640 e 690 nm tiveram as melhores correlações na determinação do teor de COS. Além disso, LI et al. (2012) relataram feições de absorções mais importantes em torno de 2140-2240 nm relacionados ao teor de COS. Mais recentemente, NAWAR; MOUAZEN (2018), em escala local, com 148 amostras, identificaram que na região do Vis, em 410 e 680 nm, foram os comprimentos de onda mais significativos para a estimativa do teor de COS. KNOX et al. (2015), em escala regional, com 1014 amostras, modelou o teor de COS utilizando espectroscopia Vis-NIR-SWIR, além de aplicarem dez diferentes técnicas de pré-processamento espectral, resultando em um valor médio de  $R^2$  de 0,80 e RMSE de 0,48 g kg<sup>-1</sup> para o modelo PLSR. Igualmente, KUANG; TEKIN; MOUAZEN (2015) compararam a calibração da espectroscopia Vis-NIR para medições online do teor de COS e obtiveram desempenho de  $R^2$  semelhante para o modelo PLSR na validação cruzada e RMSE inferior ( $R^2$  de 0,81 e RMSE de 1,99%). Da mesma forma, DING et al. (2018), ao avaliarem o potencial da ER na estimativa do teor de COS, constataram que os comprimentos de onda entre 745 a 910 nm e entre 1911 a 2254 nm foram as regiões espectrais mais importantes na estimativa do teor de COS. No entanto, o COS também foi identificado por picos de absorção a aproximadamente 1910 nm (VISCARRA

ROSSEL et al., 2006b; VISCARRA ROSSEL; WEBSTER, 2012) e 2050 a 2150 nm (WORKMAN, JR.; WEYER, 2007). As diferenças nos comprimentos de onda relevantes na estimativa do teor de COS podem estar relacionadas aos diferentes tipos de solos (WANG et al., 2016). Além disso, os diferentes constituintes do solo podem gerar muita variação nos resultados de acurácia (VISCARRA ROSSEL et al., 2016), tornando-se essenciais estratégias para selecionar ou estratificar as amostras de solos utilizadas para calibrar os modelos multivariados.

#### **2.4. Parâmetros de cor aplicados na estratificação de amostras de solos**

A cor do solo é uma característica morfológica facilmente perceptível e que fornece informações sobre outras propriedades do solo (BRADY; WEIL, 2012). Devido a sua importância como indicador pedogenético e pedoambiental, a descrição precisa das cores é necessária para a classificação e a interpretação dos solos (RAMOS et al., 2020).

Os solos exibem uma ampla variedade de cores e pode exibir tons desde os mais escuros até os quase brancos, incluindo também os tons marrons, amarelos, vermelhos ou até mesmo verdes (BRADY; WEIL, 2012). Como a cor do solo é uma variável contínua que varia na x, y e z dimensões espaciais, ela pode variar em pequenas distâncias da paisagem, assim também como em pequenas profundidades (VISCARRA ROSSEL et al., 2006a). Essa variação na coloração dos solos está relacionada a vários fatores, principalmente ao teor de matéria orgânica, a drenagem do solo e a presença ou o estado de oxidação dos óxidos de ferro (BRADY; WEIL, 2012).

A matéria orgânica é um dos pigmentos mais importantes da coloração do solo, conferindo cores escuras aos horizontes superficiais e em alguns horizontes subsuperficiais (iluviação) (VODYANITSKII; SAVICHEV, 2017). A água influencia indiretamente a cor controlando os níveis de O<sub>2</sub> no solo, este, por sua vez, controla a taxa de acúmulo de MO. Quanto menor os níveis de O<sub>2</sub>, maior o acúmulo de MO, deixando o solo mais escuro. A água também afeta o estado de oxidação do ferro e manganês. Solos bem drenados e intemperizados conferem coloração vermelha (hematita) ou amarela (goethita), principalmente. Na ausência do O<sub>2</sub> (condições anaeróbicas), os óxidos de ferro

são reduzidos e lixiviados, dando surgimento a solos com coloração acinzentadas devido à condição de má drenagem (AITKENHEAD et al., 2013a, 2013b; BEN-DOR; INBAR; CHEN, 1997; DAVEY; RUSSELL; WILSON, 1975; MURTI; SATYANARAYANA, 1971; VISCARRA ROSSEL et al., 2009; VODYANITSKII; SAVICHEV, 2017). Outros pigmentos básicos do solo também são conferidos pelo quartzo ( $\alpha\text{SiO}_2$ ), calcita ( $\text{CaCO}_3$ ) e outros carbonatos (SAHWAN et al., 2020).

Na Ciência do Solo, a cor é geralmente descrita qualitativamente por comparação visual utilizando o sistema de classificação de cores de Munsell, que define a cor em termos do matiz-H (posição da cor no espectro Vis), croma-C (pureza da cor que vai do cinza à cor púrpura) e o valor-V (brilho da cor que vai do preto ao branco) (MELO; ALLEONI, 2009). No entanto, a cor também pode ser obtida a partir das curvas espectrais (VISCARRA ROSSEL et al., 2006a). Os sensores registram a energia refletida pelos solos, que absorve determinados comprimentos de onda e a combinação dos comprimentos de onda refletidos é que determina a cor do solo (SCHANDA, 2007; VISCARRA ROSSEL et al., 2009).

Segundo VISCARRA ROSSEL et al. (2006a), a cor é um fenômeno psicofísico tridimensional e pode ser representada em modelos de espaços de cor (ou parâmetros de cor) em que as cores individuais são especificadas por pontos nestes espaços. Assim, a *Commission Internationale de l'Eclairage* (CIE, 1931) propôs os modelos CIE para facilitar a visualização e padronizar os modelos de cores. Os modelos tridimensionais de espaços de cores XYZ (ou valores triestímulos) são calculadas com base nas funções de correspondência de cores definidas em 1931 pela CIE (CIE, 1931), onde Y representa o brilho e X e Z são componentes virtuais dos espectros primários. Em 1976, a CIE desenvolveu um novo modelo de cor, denominado sistema CIE  $L^*a^*b^*$ , que é um espaço de cor universal em coordenadas cartesianas (MELVILLE; ATKINSON, 1985). A partir dos valores XYZ derivados, também podem ser obtidos outros espaços de cor (por exemplo, CIE  $L^*u^*v^*$ , CIE xyY, RGB, CIE  $L^*h^*c^*$ , entre outros) (VISCARRA ROSSEL et al., 2006a). O modelo CMYK é uma abreviação do sistema formado pelas cores Ciano, Magenta, Amarelo (Yellow) e Preto (Black). Esses sistemas representam a cor em diferentes escalas e podem ser utilizados para caracterizar a cor sob análise espectroscópica, bem como para

avaliar a contribuição da cor no solo (VISCARRA ROSSEL et al., 2006a; VODYANITSKII et al., 2007).

Nesse contexto, os parâmetros de cor derivados das curvas espectrais têm sido frequentemente utilizados em estudos na Ciência do Solo, como no trabalho de MOUAZEN et al. (2007), em escala local, com 342 amostras, avaliaram a cor do solo utilizando a carta de Munsell para utilização na classificação do solo. Os autores utilizaram os parâmetros Matiz, Valor e Cromo para caracterizar as cores do solo e a ACP para discriminar as amostras. EVRARD et al. (2019) utilizaram os parâmetros de cor para identificar a dinâmica das fontes de sedimentos espacial e temporal em duas bacias hidrográficas. Os autores identificaram que nove parâmetros de cor ( $L^*$ ,  $a^*$ ,  $b^*$ ,  $v^*$ ,  $h$ ,  $a$ ,  $b$ ,  $u'$  e  $L$ ) classificaram corretamente 100% a origem das amostras. Além disso, outros estudos foram desenvolvidos e utilizaram os parâmetros de cor na pesquisa de impressões digitais de fontes de sedimentos, como nos trabalhos realizados por MARTÍNEZ-CARRERAS et al. (2010b, 2010a). SELLIER et al. (2020) utilizaram os parâmetros de cor para discriminar as fontes de sedimentos de duas sub-bacias mineradoras na Nova Caledônia, França. Os resultados desse estudo demonstraram que a inclusão dos parâmetros de cor melhorou a discriminação das fontes de sedimentos (ou seja, 92,6% da variância da fonte foi explicada). Outros estudos usaram os parâmetros de cor para estimar as proporções de minerais, como hematita e goethita (RAMOS et al., 2020; TIECHER et al., 2015), ou determinar o tipo de MO encontrada no sedimento com base na análise dos primeiros espectros de refletância derivada (DEBRET et al., 2011). Também foram aplicados em estudos para estimar a composição mineral e o teor de argila (AITKENHEAD et al., 2013a; DOMINGUEZ et al., 2010; VISCARRA ROSSEL et al., 2009) e na identificação do teor de COS (VODYANITSKII; SAVICHEV, 2017).

Assim, devido as relações estabelecidas entre os parâmetros de cor e as propriedades do solo e também aos vários componentes do solo que exibem resposta espectral na região do Vis (400 e 700), eles podem ser utilizados como critério para classificar as amostras de solo e auxiliar na modelagem, fornecendo grupos homogêneos e com menor variabilidade de matéria orgânica, quartzo, teor de COS, óxidos de ferro ou outros elementos.

## 2.5. Estratificação de amostras de solo na espectroscopia

Os tipos de solos e os diferentes constituintes provocam não só diferenças na identificação dos comprimentos de onda relevantes como também afetam a capacidade preditiva dos modelos de predição (GUERRERO et al., 2014; NAWAR; MOUAZEN, 2018). Assim, é essencial selecionar ou estratificar o conjunto de calibração dos modelos que pode resultar em melhores desempenhos de predição e menor custo de análise.

Foi relatado que a capacidade preditiva dos modelos é influenciada pela variabilidade das amostras de solos utilizadas para calibrar os modelos multivariados (LUCÀ et al., 2017), pois quanto menor a heterogeneidade do conjunto de calibração, mais preciso deverá ser o resultado dos modelos preditivos (GUERRERO et al., 2014). Para obter melhores resultados e confiabilidade na predição, um conjunto de dados representativo deve ser selecionado com base em propriedades analíticas (AÏCHI et al., 2009), características espectrais (GUERRERO et al., 2010) ou ambos (LUCÀ et al., 2017). No entanto, muitos estudos ainda propõem a inclusão de amostras locais nas bibliotecas espectrais existentes (local/global) para melhorar os modelos de predição (GUERRERO et al., 2014; NAWAR; MOUAZEN, 2018; WETTERLIND; STENBERG, 2010). Por exemplo, WETTERLIND; STENBERG (2010) incorporaram amostras locais em uma biblioteca espectral sueca e, alcançaram melhores resultados para predição do teor de COS com  $R^2$  e RMSE de 0,87 e 0,21% respectivamente, em comparação com a biblioteca nacional sozinha. Igualmente, GUERRERO et al. (2014) sugeriram a adaptação dos modelos conforme às características das amostras locais, inserindo amostras locais ou “spiking” em bibliotecas espectrais em escala nacional com o objetivo de melhorar as estimativas das propriedades do solo.

As amostras de solos também podem ser selecionadas ou estratificadas utilizando a profundidade do solo (Tabela 2), por exemplo, JIANG et al. (2017) coletaram amostras de solos em camadas superficiais (0 – 0,1 m) e subsuperficiais (0,1 – 0,3 m) na região central da China e identificaram diferenças na variabilidade dos teores de COS para cada profundidade do solo, o que afetou o desempenho dos modelos de predição. Igualmente, ALLORY et al. (2019) ao utilizarem a espectroscopia na predição do teor de COS, selecionaram as

amostras de solos de acordo com a profundidade e identificaram que os conjuntos de amostras mais homogêneas tiveram resultados de predição mais precisos.

Tabela 2. Principais estratégias utilizadas para estratificar as amostras de solos na predição de propriedades por espectroscopia.

Parâmetro utilizado	Conjunto de amostras	Nº de amostras	Propriedades do solo	Fonte
Profundidade do solo	Camada superficial	900	COS	(JIANG et al., 2017)
	Camada sub-superficial	900		
Profundidade do solo	-	137	COS	(ALLORY et al., 2019)
Mineralogia da fração argila	-	7172	Argila	(ARAÚJO et al., 2014)
Classes de solo; Usos e camadas do solo	-	841	COS	(MOURA-BUENO et al., 2019)
Teor de umidade	-	352	COS	(WIJEWARDANE; GE; MORGAN, 2016)
Uso do solo	-	20000	COS	(STEVENS et al., 2013)

No entanto, as amostras de solos também podem ser selecionadas de acordo com a mineralogia e propriedades físico-químicas do solo, como por exemplo, ARAÚJO et al. (2014), ao estratificarem os dados de uma biblioteca espectral de solos em grupos mais homogêneos em relação à mineralogia da fração argila, obtiveram redução de 21% do erro de predição dessa fração. MOURA-BUENO et al. (2019), ao estratificarem um conjunto de amostras de solo com base nas classes de solos, usos e camadas do solo, observaram que a separação das amostras em grupos mais homogêneos melhorou a precisão dos modelos multivariados na predição de COS. WIJEWARDANE; GE; MORGAN (2016) estratificaram um conjunto de amostras de uma biblioteca espectral nacional com o objetivo de formar grupos mais homogêneos em relação ao teor de COS e identificaram que a estratificação aumentou a precisão da predição do COS. Os autores também observaram que a estratificação da biblioteca

espectral baseada nos horizontes foi mais eficaz em relação à separação por região ou uso do solo.

A variação no desempenho dos modelos de predição utilizando Vis-NIR-SWIR foi atribuída à variabilidade das amostras de solos, conforme descrita por STEVENS et al. (2013), que construíram modelos de predição para uma biblioteca espectral contendo 20.000 amostras de 23 Estados da União Europeia, nas quais as amostras foram divididas em grupos de acordo com o uso da terra. Os autores observaram variações nos valores de  $R^2$  de 0,67 a 0,89 e RMSE de 4,0 a 50,6 g kg<sup>-1</sup>. As fontes de variação foram atribuídas as frações argila, areia e COS. Da mesma forma, DEMATTÊ et al. (2016) utilizaram duas bibliotecas espectrais (local e regional) para a predição de propriedades do solo no Brasil e observaram que a presença de solos com alta variação de componentes afetou a predição das propriedades químicas do solo.

No entanto, outra alternativa que ainda é pouca explorada, seria estratificar o conjunto de amostras utilizando a própria cor do solo e sua divisão em grupos homogêneos de acordo com às características apresentadas pelos solos. Na Ciência do Solo, a forma mais utilizada para obter a cor é por comparação visual utilizando a carta de Munsell. No entanto, a cor do solo também pode ser obtida a partir das curvas espectrais, utilizando a região do Vis, sendo, portanto, uma forma rápida de obtenção da cor e sem custos adicionais. Assim, estratégias para identificar o conjunto de calibração ideal são de grande importância para a espectroscopia do solo. É necessário e desejável escolher um conjunto de calibração ideal e representativo de um número de amostras para calibrar os modelos multivariados. Isso pode resultar em uma precisão de predição mais satisfatória, melhorando as estimativas das propriedades do solo, além de reduzir o tempo de análise nos laboratórios (SEIDEL et al., 2019). Muitos estudos avaliaram a estimativa do teor de COS por espectroscopia Vis-NIR-SWIR, conforme revisado por VISCARRA ROSSEL et al. (2006b), mas poucos investigaram a influência que a variabilidade dos solos causa no desempenho dos modelos de predição.

## 2.6. Modelos de calibração multivariados

Dentre os modelos de calibração mais utilizados para correlacionar os espectros de reflectância com os dados das propriedades do solo, estão os modelos de redes neurais artificiais (RNA) (MARGENOT et al., 2020; WERE et al., 2015), regressão de mínimos quadrados parciais (PLSR) (ARAÚJO et al., 2014; DOTTO et al., 2018; STENBERG et al., 2010), máquina de vetores de suporte (*support vector machines* - SVM) (LUCÀ et al., 2017; MORELLOS et al., 2016), floresta aleatória (*random forest* - RF) (SHAHRAYINI et al., 2020; WERE et al., 2015) e o modelo de regressão gaussiana (GPR) (DOTTO et al., 2018), entre outros.

A regressão de mínimos quadrados parciais (PLSR) (WOLD; SJÖSTRÖM; ERIKSSON, 2001) é o método de calibração multivariada mais comum, comumente empregado na modelagem estatística (quimiometria) para prever as propriedades do solo a partir dos espectros (STENBERG et al., 2010). Essa técnica multivariada define um conjunto de variáveis latentes através da projeção do processo e dos espaços de qualidade em novos subespaços ortogonais, maximizando a covariância entre os dois espaços. As variáveis latentes são definidas como combinações lineares das variáveis originais (ALLORY et al., 2019). Embora sejam calculadas muitas variáveis latentes como  $\min(N, M)$ , onde  $N$  é o tamanho da amostra e  $M$  é o número de variáveis do processo, é previsto que as variáveis latentes da ordem inferior sejam associadas ao ruído do processo e devem ser excluídas do modelo. Portanto, para remover o ruído, um critério é necessário para selecionar o número de variáveis latentes a serem incluídas no modelo PLSR (LI; MORRIS; MARTIN, 2002). O modelo PLSR tem sido frequentemente utilizado na estimativa de propriedades do solo, como no trabalho de VISCARRA ROSSEL; BEHRENS (2010), que aplicaram o método PLSR, entre outros, para prever o teor de COS com base na espectroscopia Vis-NIR usando uma biblioteca espectral com 1104 amostras de solo. DOTTO et al. (2018), com 595 amostras, utilizaram nove modelos de calibração e sete técnicas de pré-processamento em um conjunto de dados com teor de COS variando de 0,02 a 6,87% e observaram que o modelo PLSR obteve melhor desempenho que os demais, com  $R^2$  de 0,82 e RMSE de 0,48%. Utilizando um conjunto de dados com maior número de

amostras (n=1534), SILVA et al. (2019), utilizaram cinco modelos de calibração e seis técnicas de pré-processamento na predição da textura do solo e observaram que o modelo PLSR obteve melhor desempenho para a fração argila em relação a fração areia, com  $R^2$  de 0,75 e RMSE de 8,9%. VASQUES; GRUNWALD; SICKMAN (2008) compararam métodos multivariados para modelar o teor de COS e os modelos PLSR alcançaram um valor médio de  $R^2$  de 0,82 para 30 métodos de pré-processamento espectral.

A rede neural artificial (RNA) é um método de aprendizagem de máquina, não-linear e não-paramétrico, que tem obtido destaque nos últimos anos (LIU et al., 2017) e tem sido frequentemente utilizado na Ciência do Solo para mapeamento digital (BAGHERI BODAGHABADI et al., 2015; WERE et al., 2015) e na predição de propriedade do solo utilizando ER (MARGENOT et al., 2020; YANG et al., 2020a). A grande vantagem da RNA é o uso de relações não lineares entre as curvas espectrais e as propriedades de solos medidas em laboratório. Como uma RNA pode ser capaz de abordar respostas espectrais não lineares melhor do que o PLSR, ela foi aproveitada para melhorar a precisão das estimativas em várias aplicações, incluindo a estimativa de propriedades do solo a partir das curvas espectrais (ZHAO et al., 2006). A RNA tem sido usada de forma proeminente para predições de propriedades do solo e, especificamente, na estimativa do teor de COS, por exemplo em MARGENOT et al. (2020) e YANG et al. (2020a). Comparações em larga escala de RNA e PLSR para predição de propriedades do solo (por exemplo, argila) utilizando ER descobriram que a RNA supera o modelo PLSR usando um modelo global de conjuntos de dados relativamente grandes ( $n > 20.000$ ), mas o oposto foi encontrado para modelos locais e/ou conjuntos de dados menores (VISCARRA ROSSEL; BEHRENS, 2010; WIJEWARDANE; GE; MORGAN, 2016).

Além desses métodos, os métodos de mineração de dados não lineares, como máquinas de vetores de suporte (SVM) (CORTES; VAPNIK, 1995) e o método de aprendizagem de conjunto de floresta aleatória (RF) (BREIMAN, 2001) recentemente ganharam espaço como métodos multivariados para prever propriedades do solo.

O SVM é um algoritmo de aprendizagem de máquina que despertou atenção ao lidar com uma grande quantidade de dados (RAMIREZ-LOPEZ et al., 2013). Como um método não linear, o SVM mostra vantagens sobre o PLSR por

ser capaz de lidar eficientemente com valores de entrada de alta dimensão (GUPTA; ISLAM; HASAN, 2018). O PLSR tende a encolher as direções de baixa variância ou algumas direções de alta variância, tornando o PLSR um pouco instável (VISCARRA ROSSEL; BEHRENS, 2010). Além disso, o PLSR reduz os ruídos das curvas espectrais, mas não as excluem. Portanto, muito ruído nas curvas espectrais pode afetar a acurácia do modelo de predição de propriedades do solo, sendo necessário o pré-processamento dos espectros (BONA et al., 2017). Além disso, variáveis altamente correlacionadas tendem a ser escolhidas juntas, como resultado, pode haver muita redundância no conjunto de variáveis selecionadas. Isso pode indicar que o método PLSR é mais propenso a *overfitting* (sobre ajuste) do que o SVM. A forte base de treinamento do SVM, fornece alta capacidade de generalização e evita *overfitting*, que é quando os classificadores possuem um grande volume de dados de entrada e tendem a se adaptar a detalhes específicos da base de treinamento, ocasionando uma redução da taxa de acerto (ARGYRI et al., 2013). O *overfitting* é mais provável de ocorrer em dados de alta dimensão, e os espectros de infravermelho tipicamente mostram uma dimensão muito alta (LIU; YANG; DENG, 2015). Devido a essas características, o SVM tem sido utilizado na estimativa de propriedades do solo utilizado ER, como no trabalho de LUCÀ et al. (2017), que utilizaram três métodos de calibração na predição do teor de COS, incluindo a análise de componentes principais (ACP), o PLSR e o SVM, e identificaram que embora a ACP seja menos sensível ao número de amostras utilizadas na calibração dos modelos em relação aos métodos PLSR e SVM, o método SVM superou os demais métodos com  $R^2$  mais alto e RMSE mais baixo. Da mesma forma, DOTTO et al. (2017), em escala local, com 299 amostras, aplicaram o modelo SVM e o pré-processamento DET, em um conjunto de dados com a fração argila variando de 21 a 78% e a fração areia variando de 1 a 35%. Os autores encontram  $R^2$  de 0,62 e RMSE de 6,8% para a fração argila e  $R^2$  de 0,25 e RMSE de 6,4% para a fração areia. Mais recentemente, DEISS et al. (2020), avaliaram o desempenho preditivo dos modelos PLSR e SVM na estimativa de propriedades físicas e químicas do solo utilizando ER e identificaram que o modelo SVM superou o modelo PLSR na predição de todas as propriedades do solo estudadas.

O RF é um modelo de aprendizagem desenvolvido por BREIMAN (2001) para melhorar o método das árvores de decisão (*Classification and Regression Trees*). O RF é um modelo não paramétrico, que busca melhorar o desempenho de predição deste modelo. Trata-se de uma combinação de muitas árvores de decisão (floresta), em que cada árvore é gerada a partir de um vetor aleatório, amostrado de forma independente e com a mesma distribuição para todas as árvores na floresta. As subdivisões dentro de cada árvore são determinadas com base em um subconjunto de variáveis preditoras, escolhido aleatoriamente a partir de um conjunto de variáveis. A média dessas árvores resulta em grande ganho na variação reduzida. No caso de RF para regressão, o resultado final consiste da média dos resultados de todas as árvores (BREIMAN, 2001). VISCARRA ROSSEL; BEHRENS (2010) compararam diferentes modelos de calibração utilizando espectroscopia Vis-NIR-SWIR e obtiveram os menores resultados para estimativa do teor de COS com RF ( $R^2$  de 0,71 e RMSE de 1,23%), com a melhor predição encontrada para RNA. KNOX et al. (2015), em escala regional, com 1014 amostras, avaliaram o potencial da espectroscopia Vis-NIR-SWIR na estimativa do teor de COS e identificaram que o modelo RF produziu um  $R^2$  e RMSE de 0,63 a 0,88 e de 0,70 a 0,38  $\log g \text{ kg}^{-1}$ , respectivamente. Mais recentemente, SHAHRAYINI et al. (2020) avaliaram a capacidade da espectroscopia Vis-NIR na estimativa do teor de COS em diferentes profundidades, utilizando quatro modelos de calibração (PLSR, PCR, SVM e RF) e concluíram que o modelo RF superou os demais modelos de calibração e a maior precisão foi alcançada sem a utilização de pré-processamento espectral.

O GPR é um modelo de aprendizagem de máquina que foi desenvolvido de acordo com a teoria bayesiana, tendo como base a teoria de aprendizagem estatística. É um modelo aplicável a problemas complexos, como os de alta dimensão, de pequena amostra e os não lineares. Além disso, possui alto desempenho de generalização. Em comparação com outros métodos de aprendizado de máquina, como o SVM e RNA, possui a capacidade de operar com fácil realização, ser adaptável a muitos parâmetros e significado probabilístico de saídas. Devido ao seu bom desempenho, o GPR tem sido aplicado com sucesso na Ciência do Solo. CHEN; MORRIS; MARTIN (2007), aplicaram o GPR como um método alternativo para o desenvolvimento de um

modelo multivariado de calibração para curvas espectrais. GIJSBERTS; METTA (2013) confirmaram que o desempenho do GPR foi superior de acordo com a validação empírica em vários problemas de aprendizagem, enquanto a exigência computacional foi evidentemente mais baixa.

Nesse contexto, os diferentes modelos de calibração podem ser utilizados na predição das propriedades do solo como o teor de COS e das frações texturais. No entanto, a heterogeneidade dos solos prejudica a capacidade preditiva dos modelos, sendo necessários estabelecer critérios para estratificar o conjunto de amostras em grupos mais homogêneos.

### **3. CAPÍTULO II – ESTUDO 1: EFEITOS DA ESTRATIFICAÇÃO DAS AMOSTRAS NO DESEMPENHO DOS MODELOS DE PREDIÇÃO DA TEXTURA DO SOLO E DO CARBONO ORGÂNICO DO SOLO<sup>2</sup>**

#### **3.1. INTRODUÇÃO**

Compreender a variabilidade espacial das propriedades do solo permite melhorar as estratégias de manejo e conservação, bem como o uso consciente do solo para garantir seu uso econômico e ambientalmente sustentável (GHOLIZADEH et al., 2018). A fim de apoiar o desenvolvimento e a aplicação de estratégias de manejo, é necessário aumentar a densidade espacial de dados analíticos sobre os solos (NOCITA et al., 2014; VISCARRA ROSSEL et al., 2016). No entanto, essa ainda não é uma prática economicamente viável devido aos altos custos associados às análises químicas laboratoriais (NANNI et al., 2011).

Técnicas de sensoriamento remoto e proximal têm sido propostas como métodos eficientes no estudo das propriedades do solo (GUPTA; ISLAM; HASAN, 2018). Entre esses métodos, a espectroscopia de reflectância (ER) na região do visível (Vis), do infravermelho próximo (NIR) e de ondas curtas (SWIR) (350-2500 nm) está se destacando como uma técnica capaz de mensurar várias propriedades do solo a baixo custo (O'ROURKE; HOLDEN, 2011). Atualmente, a ER já é uma técnica muito utilizada em laboratórios de rotina de solo e, nos últimos anos, tem aumentado também a sua aplicação no campo. As assinaturas espectrais do solo permitem inferir informações sobre a composição físico-química e mineralógica do solo e a inferência é baseada em modelos de calibração multivariados desenvolvidos a partir de bibliotecas espectrais que correlacionam os espectros de reflectância com os dados de propriedades do

---

<sup>2</sup>Adaptado do artigo publicado na revista *Computers and Electronics in Agriculture*.  
<https://doi.org/10.1016/j.compag.2020.105710>

solo medidos em laboratório (GUPTA; ISLAM; HASAN, 2018; PABÓN; SOUZA FILHO; OLIVEIRA, 2019; STEVENS et al., 2013).

Como os solos podem variar muito em termos de sua composição mineralógica e propriedades físico-químicas, muitos estudos têm utilizado amostras e bibliotecas espectrais que representam a variabilidade dos solos em escala local (DOTTO et al., 2016; GHOLIZADEH et al., 2016; SILVA et al., 2016), regional (FRANCESCHINI et al., 2013; RAMIREZ-LOPEZ et al., 2013), nacional (ARAÚJO et al., 2014; SHI et al., 2014; VISCARRA ROSSEL; CHEN, 2011) continental (NOCITA et al., 2014; STEVENS; RAMIREZ-LOPEZ, 2013) e global (RAMIREZ-LOPEZ et al., 2013; VISCARRA ROSSEL et al., 2016). Outros estudos ainda propõem a adaptação dos modelos conforme as características das amostras locais, denominadas de “spiking” (GUERRERO et al., 2010, 2014; ZHAO et al., 2018), onde amostras locais representativas da área de estudo são inseridas em bibliotecas espectrais em escala regional. O desenvolvimento dessas bibliotecas com um grande número de amostras é realizado para garantir que as amostras locais se enquadrem no domínio do modelo (GRINAND et al., 2012; VISCARRA ROSSEL; WEBSTER, 2012). No entanto, isso não pode ser garantido porque os solos têm características muito variáveis, mesmo em escala regional, causando variabilidade na precisão de predição dos modelos (STENBERG et al., 2010; VISCARRA ROSSEL et al., 2016).

Para melhorar o desempenho dos modelos, muitos estudos têm utilizado diferentes tipos de técnicas de pré-processamento espectral e modelos multivariados (Tabela 2). ALLO et al. (2020), em escala local, com 95 amostras, aplicaram o modelo PLSR e três técnicas de pré-processamento em um conjunto de dados com teor de COS variando de 0,4 a 13% e observaram que a combinação das técnicas de pré-processamento (DET+FDSD) obtiveram melhor desempenho que as demais. ALLORY et al. (2019) observaram uma melhor capacidade de predição utilizando a combinação de técnicas de pré-processamento (SG + SNV) que mostraram melhor desempenho quando se utilizou apenas uma técnica de pré-processamento. Da mesma forma, DOTTO et al. (2017), em escala local, com 299 amostras, aplicaram o modelo SVM e o pré-processamento DET, em um conjunto de dados com a fração argila variando de 21 a 78% e a fração areia variando de 1 a 35%, encontraram  $R^2$  de 0,62 e

RMSE de 6,8% para a fração argila e  $R^2$  de 0,25 e RMSE de 6,4% para a fração areia.

No entanto, os resultados de predição obtidos por DEMATTÊ et al. (2016), em diferentes escalas, mostraram diferenças no desempenho dos modelos do teor de COS e das frações areia, argila e silte, que podem ter sido influenciados pela estrutura dos dados de calibração. Isso demonstra que tanto o tipo de pré-processamento quanto a estrutura dos dados de entrada são fatores que podem influenciar na capacidade preditiva dos modelos de predição.

Tabela 3. Compilação de estudos que aplicaram diferentes tipos de pré-processamento e modelos multivariados na estimativa do teor de COS e das frações argila, areia e silte utilizando ER.

bEscala	aModelo	Pré-processamento	Amostras	COS		Argila		Areia		Silte		Fonte
				$R^2_v$	RMSE <sub>v</sub> %							
Local	Cubist	-	197	-	-	0,89	5,1	0,81	6,5	0,06	4,1	(COBLINSKI et al., 2020)
Local	PLSR	PCA	144	0,84	0,1	0,78	3,9	0,79	3,4	0,65	2,6	(MARGENOT et al., 2020)
Local	PLSR	-	92	0,92	0,8	-	-	-	-	-	-	(ALLO et al., 2020)
Local	PLSR	DET + FDSG	92	0,94	0,7	-	-	-	-	-	-	(ALLO et al., 2020)
Local	PLSR	SNV + FDSG	94	0,92	0,8	-	-	-	-	-	-	(ALLO et al., 2020)
Local	PLSR	SG	186	0,40	1,0	-	-	-	-	-	-	(ALLORY et al., 2019)
Local	PLSR	SG + SNV	186	0,83	0,5	-	-	-	-	-	-	(ALLORY et al., 2019)
Local	PLSR	SG + FDSG	186	0,83	0,5	-	-	-	-	-	-	(ALLORY et al., 2019)
Local	PLSR	SG + SDSG	186	0,82	0,5	-	-	-	-	-	-	(ALLORY et al., 2019)
Nacional	Cubist	CR	39284	-	-	0,88	7,6	0,87	10,3	-	-	(DEMATTÊ et al., 2019)
Local	PLSR	MSC	90	0,66	0,2	-	-	-	-	-	-	(HUTENGS et al., 2019)
Regional	Cubist	MSC	1534	-	-	0,83	7,3	0,70	11,0	-	-	(SILVA et al., 2019)
Regional	SVM	MSC	1534	-	-	0,76	8,7	0,60	12,8	-	-	(SILVA et al., 2019)
Regional	SVM	DET	1534	-	-	0,73	9,4	0,58	13,0	-	-	(SILVA et al., 2019)
Regional	GPR	DET	1534	-	-	0,71	9,6	0,58	13,0	-	-	(SILVA et al., 2019)
Regional	RF	CRR	1534	-	-	0,68	10,2	0,61	12,9	-	-	(SILVA et al., 2019)
Regional	PLSR	SDSG	1534	-	-	0,55	12,4	0,54	13,7	-	-	(SILVA et al., 2019)
Regional	PLSR	SG	1013	-	-	0,73	5,4	0,72	9,2	0,59	5,9	(VASAVA et al., 2019)
Nacional	PLSR	SNV, SG	10802	-	-	0,97	1,9	0,96	6,6	0,94	5,6	(JACONI; VOS; DON, 2019)
Local	PLSR	SG	148	0,74	0,2	-	-	-	-	-	-	(NAWAR; MOUAZEN, 2018)
Regional	PLSR	SG	591	0,80	0,1	0,58	6,2	0,67	7,8	0,70	7,8	(XU et al., 2018)
Local	SVM	DET	299	0,86	0,4	0,62	6,8	0,25	6,4	0,50	6,2	(DOTTO et al., 2017)
Local	SVM	CRR	299	0,86	0,4	0,58	7,2	0,25	6,3	0,50	6,2	(DOTTO et al., 2017)
Local	PLSR	CRR	299	0,83	0,4	0,52	7,5	0,17	6,5	0,56	5,3	(DOTTO et al., 2017)
Local	PLSR	CR + RB	299	0,79	0,5	0,45	8,3	0,18	6,5	0,44	6,7	(DOTTO et al., 2017)
Local	PLSR	DET	299	0,72	0,5	0,40	8,8	0,19	6,5	0,44	6,7	(DOTTO et al., 2017)

Tabela 3. Continuação.

Local	SVR	FDSG	300	-	-	0,53	7,9	0,18	5,5	0,39	9,1	(DUDA et al., 2017)
Local	PCR	-	216	0,69	0,9	-	-	-	-	-	-	(LUCÀ et al., 2017)
Local	PLSR	SNV	216	0,79	0,7	-	-	-	-	-	-	(LUCÀ et al., 2017)
Local	SVM	-	216	0,82	0,7	-	-	-	-	-	-	(LUCÀ et al., 2017)
Local	PLSR	MSC, FDSG	200			0,78	6,9	0,82	6,7	0,42	2,1	(NANNI et al., 2017)
Local	PLSR	SG/SNV	434	0,71	0,6	0,78	6,2	0,62	11,5	0,36	9,5	(PINHEIRO et al., 2017)
Local	Cubist	SDSG/SNV	257			0,70	14,7	0,50	18,3	0,00	16,7	(ZHANG et al., 2017)
Regional	PLSR	PCA	7185	0,63	0,3	0,85	9,8	0,86	10,2	0,51	3,8	(DEMATTÊ et al., 2016)
Regional	PLSR	PCA	903	0,65	0,3	0,77	10,4	0,78	11,8	0,55	3,3	(DEMATTÊ et al., 2016)
Regional	PLSR	PCA	3093	0,61	0,2	0,67	3,2	0,64	4,2	0,30	2,6	(DEMATTÊ et al., 2016)
Local	PLSR	PCA	621	0,62	0,2	0,77	7,3	0,77	7,3	0,21	1,4	(DEMATTÊ et al., 2016)
Local	PLSR	PCA	563	0,54	0,3	0,74	11,3	0,77	12,7	0,71	3,1	(DEMATTÊ et al., 2016)
Local	PLSR	PCA	843	0,66	0,1	0,71	2,8	0,68	3,8	0,21	3,0	(DEMATTÊ et al., 2016)
Local	PLSR	PCA	541	0,61	0,4	0,55	7,6	0,45	6,7	0,50	3,8	(DEMATTÊ et al., 2016)
Local	SVR	FDSG	102	-	-	0,55	9,9	-	-	-	-	(NAWAR et al., 2016)
Local	PLSR	CR	102	-	-	0,72	7,8	-	-	-	-	(NAWAR et al., 2016)
Local	MARS	SDSG	102	-	-	0,73	8,2	-	-	-	-	(NAWAR et al., 2016)
Local	MARS	CR	102	-	-	0,79	7,6	-	-	-	-	(NAWAR et al., 2016)
Global	Cubist	CRR	23631	0,92	1,1	0,80	10,3	0,68	18,8	0,79	10,3	(VISCARRA ROSSEL et al., 2016)
Local	PLSR	SNV	306	-	-	0,78	3,1	-	-	-	-	(CAMARGO et al., 2016)
Regional	PLSR	SDSG	7172	0,60	0,6	0,82	10,9	-	-	-	-	(ARAÚJO et al., 2014)
Regional	SVM	SDSG	7172	0,69	0,5	0,89	8,9	-	-	-	-	(ARAÚJO et al., 2014)
Local	PLSR	SVN, SG, MSC	129	-	-	0,92	5,2	0,87	7,3	0,80	3,1	(FRANCESCHINI et al., 2013)
Local	PLSR	FDSG	120	-	-	0,86	17,7	0,87	22,0	0,83	46,5	(FERRARESI et al., 2012)
Local	PLSR	SNV	148	-	-	0,74	12,0	0,56	24,0	0,46	49,0	(VENDRAME et al., 2012)
Global	BTR	SMO, 1st DER	3793	0,82	0,9	0,73	9,5	-	-	-	-	(BROWN et al., 2006)

<sup>a</sup> PLSR: *partial least squares regression*; SVM: *support vector machine*; SVR: *support vector regression*; GPR: *Gaussian process regression*; RF: *random forest*; MARS: *multivariate adaptive regression splines*; BTR: *boosted regression trees*; RB: *band ratio*; CRR: *continuum removal*; MSC: *multiplicative scatter-correction*; DET: *detrend*; SDSG: *Savitzky–Golay second derivative*; FDSG: *Savitzky–Golay first derivative*; SNV: *standard normal variate*; PCA: *principal component analysis*; DER: *derivative*;  $R^2_v$ : *coefficient of determination*; RMSE<sub>v</sub>: *root mean square error of validation*. <sup>b</sup> O critério para diferenciar à escala da biblioteca espectral, geralmente é a cobertura espacial e o número de amostras de solos.

Atualmente, apenas algumas publicações abordam o efeito que a estrutura dos dados de entrada e a heterogeneidade dos solos provocam no desempenho dos modelos. JIANG et al. (2017) coletaram amostras de solos em camadas superficiais (0 – 0,1 m) e subsuperficiais (0,1 – 0,3 m) na região central da China e identificaram diferenças na variabilidade dos teores de COS para cada profundidade do solo, o que afetou o desempenho dos modelos de predição. ARAÚJO et al. (2014), ao dividirem os dados de uma biblioteca

espectral de solos em grupos mais homogêneos em relação a mineralogia da fração argila, obtiveram redução de 21% do erro de predição para a fração argila. MOURA-BUENO et al. (2019), ao estratificarem um conjunto de amostras de solo com base nas classes de solos, usos e camadas do solo, observaram que a separação das amostras em grupos mais homogêneos melhorou a precisão dos modelos multivariados na predição do teor de COS.

No entanto, uma alternativa que ainda pode ser explorada é o estabelecimento de um critério para classificar as amostras com base na cor dos solos e sua divisão em grupos homogêneos de acordo com as características apresentadas pelos solos. Existem muitas relações conhecidas entre a cor do solo e suas características mineralógicas e físico-químicas (AITKENHEAD et al., 2013a, 2013b; BEN-DOR; INBAR; CHEN, 1997; DAVEY; RUSSELL; WILSON, 1975; MURTI; SATYANARAYANA, 1971; VISCARRA ROSSEL et al., 2009; VODYANITSKII; SAVICHEV, 2017). Exemplos de variações na coloração do solo associadas as propriedades do solo incluem a coloração escura da MO, a coloração branca do quartzo, calcita e outros carbonatos, a coloração vermelha da hematita ( $\alpha\text{-Fe}_2\text{O}_3$ ) e a coloração amarela da goethita ( $\alpha\text{-FeOOH}$ ) (AITKENHEAD et al., 2013a, 2013b). Assim, a clusterização prévia dos solos pela cor pode auxiliar na modelagem, fornecendo aos modelos de predição subgrupos com menor variabilidade de matéria orgânica, quartzo, óxidos de ferro ou outros elementos.

A forma mais fácil e rápida de medir-se ou caracterizar-se a cor do solo é por meio do sistema *Munsell* de cores (MUNSELL SOIL COLOR CHARTS, 2000) que classifica as cores em três componentes: matiz, valor e croma, onde o matiz é a cor espectral dominante, o valor é a tonalidade da cor e o croma é a pureza da cor (SOIL SURVEY DIVISION STAFF, 2017). No entanto, existem abordagens mais recentes para a mensuração da cor e incluem o uso de computador acoplado a sensores, como os colorímetros, espectrofotômetros e espectrorradiômetros (LEVIN; BEN-DOR; SINGER, 2005; MORENO-RAMÓN; MARQUÉS-MATEU; IBÁÑEZ-ASENSIO, 2014; VISCARRA ROSSEL; FOUAD; WALTER, 2008). Esses equipamentos registram a energia refletida pelos solos, que absorve determinados comprimentos de onda e a combinação dos comprimentos de onda refletidos é que determina a cor do solo (SCHANDA, 2007; VISCARRA ROSSEL et al., 2009). Dessa forma, a cor do solo pode ser

obtida a partir das próprias curvas espectrais sem custos adicionais e ser representada por modelos de espaço de cores.

Embora os parâmetros de cor tenham sido aplicados em estudos para estimar a cor do solo, a composição mineral e a fração argila (AITKENHEAD et al., 2013a; DOMINGUEZ et al., 2010; VISCARRA ROSSEL et al., 2009), na identificação do teor de COS (VODYANITSKII; SAVICHEV, 2017), e na discriminação da fonte de sedimentos (MARTÍNEZ-CARRERAS et al., 2010a, 2010b; TIECHER et al., 2015), desconhece-se estudos que tenham utilizado os parâmetros de cor para classificar amostras de solo e agrupá-las em grupos homogêneos para predição de propriedades do solo.

Neste sentido, os objetivos deste estudo foram usar os parâmetros de cor baseados no espectro visível (Vis) combinados com estatística multivariada para estratificar amostras de solo de uma biblioteca espectral local e comparar os resultados de diferentes modelos de predição de COS, argila, areia e silte.

## **3.2. MATERIAL E MÉTODOS**

### **3.2.1. Caracterização da área de estudo**

A área de estudo está localizada no Parque Nacional de São Joaquim (PNSJ), região do planalto sul do Estado de Santa Catarina, Brasil (Figura 1). Essa Unidade de Conservação (UC) apresenta uma área aproximada de 49.300 ha, com relevo bastante acidentado e altitudes que variam entre 300 a 1.826 m. O parque situa-se em uma região fria, com a temperatura média anual variando entre 12 °C a 14 °C (VIANNA et al., 2015). Destaca-se por ser o único no Brasil voltado para a conservação da araucária (*Araucária angustifolia*) e cobre quatro tipos principais de fitofisionomia (KLEIN, 1978): os campos gerais ou de altitude, as matas de araucárias, a floresta pluvial subtropical e as matas nebulares das encostas de altitudes. Os solos são formados predominantemente a partir de rochas vulcânicas e sedimentares da Bacia do Paraná do Grupo Serra Geral (Formação Serra Geral, Palmas, Paranapanema, Gramado e Botucatu) e do Grupo Passa Dois (Formação Rio do Rasto e Teresina) (EMBRAPA, 2004; WILDNER et al., 2014). As rochas vulcânicas da Formação Serra Geral são dominantes e caracterizadas por basaltos e andesitos (básicas) e riolitos e riodacitos (ácidas). Os solos dessa região apresentam, em geral, horizonte A húmico ou

proeminente, com teores muito altos de Al trocável e alta acidez potencial, baixa soma de bases e saturação por base, relevo forte ondulado e ondulado, um tanto rasos pertencendo às classes dos Neossolos Litólicos e Regolíticos e dos Cambissolos Húmicos e Háplicos (DALMOLIN et al., 2017; EMBRAPA, 2004).

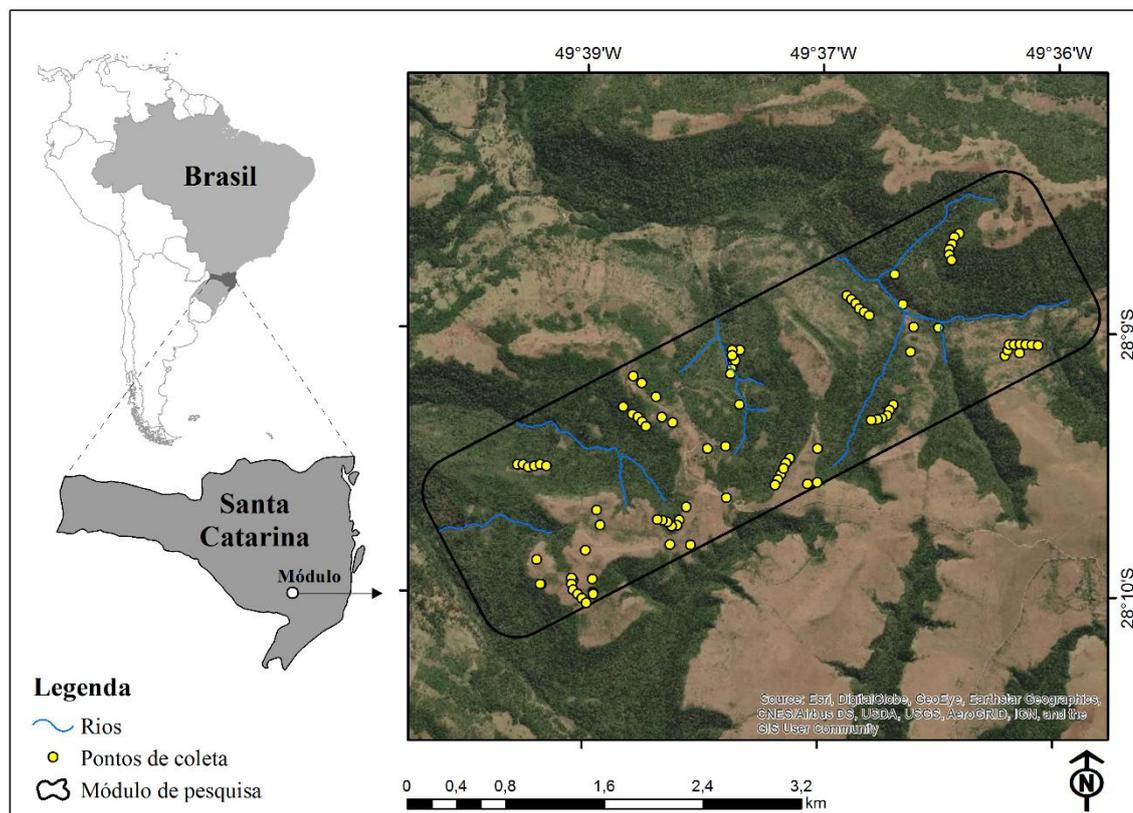


Figura 1. Módulo de pesquisa e localização das parcelas onde foram coletadas as amostras de solo no Parque Nacional de São Joaquim, no Estado de Santa Catarina.

Nessa área, a Epagri instalou módulos de pesquisa do Programa de Pesquisa em Biodiversidade, Mata Atlântica, Santa Catarina (PPBio-MA-SC). A área do módulo tem 500 ha (1 × 5 km) e é composta por dois acessos no sentido Leste-Oeste. No interior do módulo 1 foram instaladas 20 parcelas: 10 terrestres (Figura 1) e 10 ripárias. As parcelas são corredores de 250 m que seguem às curvas de nível e são demarcadas por 6 piquetes, um a cada 50 m (0, 50, 100, 150, 200 e 250 m). Um total de 260 amostras de solo foram coletadas em novembro de 2016, usando dois sistemas de amostragem. Com o primeiro sistema de amostragem foram coletadas 145 amostras de solos (nas parcelas terrestres), usando o protocolo do PPBio, com adaptação na profundidade de coleta segundo o GlobalSoilMap.Net (0-10, 10-20, 20-30 e 30-60 cm de

profundidade). Com o segundo sistema de amostragem foram coletadas 115 amostras de solos (fora das parcelas, em diferentes locais do módulo), usando uma abordagem do *conditioned Latin Hypercube Sampling* (cLHS) (MINASNY; MCBRATNEY, 2006). A amostragem por cLHS utilizou um conjunto de variáveis ambientais contínuas ou categóricas que representam o "espaço preditor" no qual foram selecionados os locais mais representativos da paisagem para a coleta das amostras.

### **3.2.2. Medidas de reflectância espectral e análise laboratorial**

Dados de reflectância do solo foram obtidos por espectroscopia Vis-NIR-SWIR, com base no espectro de absorção que é produzido após a leitura da amostra de solo com um sensor espectral. Para obtenção da reflectância espectral, foi utilizado um espectraloradiômetro *FieldSpec 3 (Analytical Spectral Devices, Boulder, EUA)* com o conjunto ASD "Contact Probe®" (CP) que recobre a faixa espectral entre 350 a 2500 nm e resolução de 1 nm. Inicialmente as amostras foram secas, moídas e peneiradas (malha 2mm) para obtenção da terra fina seca ao ar (TFSA). Para realizar as medidas espectrais, as amostras de TFSA foram distribuídas homogeneamente em um porta amostras. O sensor espectral utilizado capturou a luz através de um cabo de fibra ótica acoplado no CP, que ficou em contato com a superfície da amostra. O sensor foi calibrado no início das medições espectrais e a cada 20 amostras de solo, usando uma placa branca *Spectralon (Labsphere, North Sutton, NH, USA)* com mais de 99% de reflectância. Para cada amostra foram obtidas três repetições, sendo utilizada a curva espectral média. Os dados espectrais do solo foram pré-processados em uma etapa prévia à modelagem. As curvas espectrais foram levemente suavizadas com o filtro *Savitzky-Golay* (com ajuste poligonal de segunda ordem e janela móvel de 11 nm). A parte ruidosa do espectro foi removida previamente a modelagem.

Para determinar os teores de COS foram coletadas sub-amostras com aproximadamente 3 g de TFSA, que foram moídas e peneiradas para uma fração de partícula < 0,25 mm. As sub-amostras foram armazenadas em *Eppendorfs* para serem feitas as análises por combustão seca em um analisador elementar, modelo *Flash 2000 Organic Elemental Analyzer* (ZOBECK et al., 2013). Para

determinar as frações areia, silte e argila nas amostras de TFSA foi utilizado o método da pipeta, usando dispersante NaOH (DONAGEMMA et al., 2017).

### 3.2.3. Cálculo dos parâmetros de cor baseados em Vis

A sequência metodológica aplicada neste estudo é mostrada no fluxograma da Figura 2. Vinte e dois (22) componentes foram derivados do espectro Vis usando vários modelos de colorimétrica descritos em detalhes por VISCARRA ROSSEL et al. (2006). A *Commission Internationale de l'Eclairage* (CIE, 1996) propôs os modelos CIE para facilitar a visualização e padronizar os modelos de cores. Foram calculados os valores triestímulos XYZ com base nas funções de correspondência de cores definidas em 1931 pela CIE (CIE, 1996), onde Y representa o brilho e X e Z são componentes virtuais dos espectros primários. Os valores XYZ derivados foram, em seguida, transformados em dezenove outros modelos de espaço de cor (RGB, coordenadas cromáticas de Munsell HVC, de CIE xyY, de CIE L\*a\*b\*, de CIE L\*u\*v\*, de CIE L\*c\*h\* e de CMYK) usando o *software Munsell Conversion* (WALLKILLCOLOR, [s.d.]). No sistema CIE xyY, o Y representa a luminosidade e x e y representam variações de cor de azul para vermelho e de azul para verde, respectivamente. Nos sistemas CIE L\*a\*b\* e CIE L\*u\*v\*, o L\* representa o brilho ou a luminosidade, e a\* e b\* e u\* e v\* representam coordenadas cromáticas, vermelhas-verdes e azuis-amarelas, respectivamente. O modelo CIE L\*c\*h\* representa uma transformação do espaço de cor esférico CIE L\*a\*b\* em coordenadas cilíndricas, resultando em valores de matiz (h\*) e de croma (c\*). O sistema RGB forma um cubo que compreende eixos ortogonais vermelho (R), verde (G) e azul (B), de onde cada cor pode ser produzida por uma mistura dessas três cores primárias. O modelo CMYK é uma abreviação do sistema formado pelas cores Ciano, Magenta, Amarelo (Yellow) e Preto (Black). O sistema Munsell HVC usado na Ciência do Solo, descreve a cor do solo pelo uso de matiz (H), valor (V) e croma (C). A Tabela 3 apresenta um resumo dos modelos de espaço de cores usados neste estudo e as abreviações para todos os 22 parâmetros de cor calculados.

Tabela 4. Parâmetros de cores baseados em Vis derivados de diferentes modelos de espaço de cores calculados usando o *software Munsell Conversion* (WALLKILLCOLOR, 2019).

Modelos de espaços de cor	Parâmetros de cor	Abreviação
RGB	Red (vermelho)	R
	Green (verde)	G
	Blue (azul)	B
CIE xyY	Coordenadas cromáticas x	x
	Coordenadas cromáticas y	y
	Luminosidade	Y
CIE XYZ	Componente virtual X	X
	Componente virtual Z	Z
CIE Luv	Brilho ou Luminosidade	L
	Coordenadas cromáticas vermelhas-verdes	u*
	Coordenadas cromáticas azul-amarela	v*
CIE Lab	Coordenadas cromáticas vermelhas-verdes	a*
	Coordenadas cromáticas azul-amarela	b*
CIE Lch	CIE croma	c*
	CIE hue (matiz)	h*
CMYK	Cyan (ciano)	C
	Magenta	M
	Yellow (amarela)	Ye
	Black (preta)	K
Munsell HVC	Hue (matiz)	Matiz
	Value (valor)	Valor
	Chroma (croma)	Croma

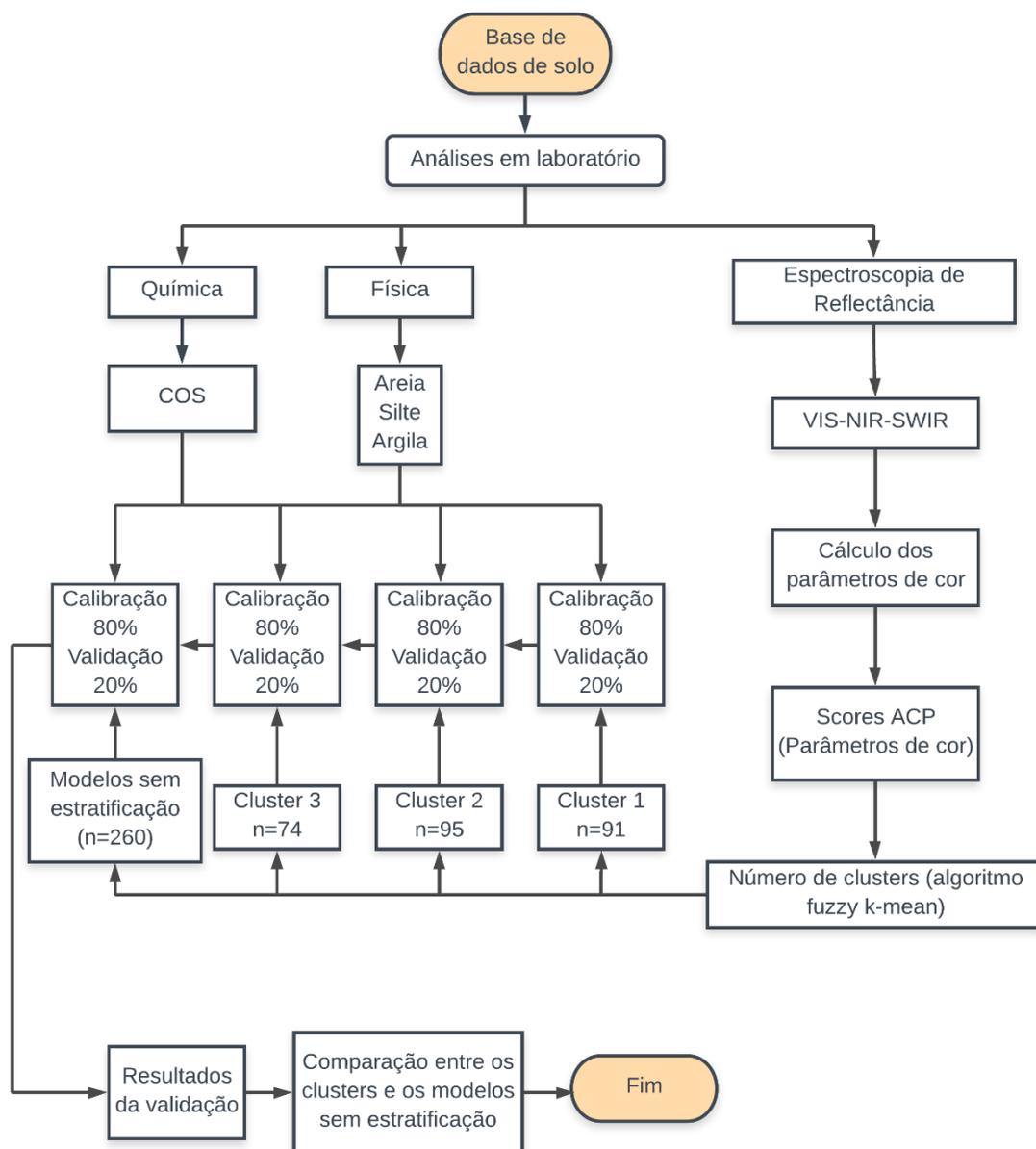


Figura 2. Sequência metodológica.

### 3.2.4. Análises estatísticas descritiva e multivariadas

A estatística descritiva foi realizada nos teores de COS e das frações areia, silte e argila. Com base nessas informações, verificou-se que existia uma variabilidade dessas propriedades, principalmente para o teor de COS e das frações areia e argila (Tabela 4). Dessa forma, foi necessário estabelecer critérios para dividir o grupo de amostras em subgrupos. Para tal, foram aplicadas aos valores de parâmetros de cor duas análises estatísticas multivariadas: Análise de Componentes Principais (ACP) e Análise de

clusterização das amostras (ou *cluster*), ambas realizadas na plataforma computacional R (R DEVELOPMENT CORE TEAM, 2017).

### 3.2.5. Análise de componentes principais (ACP)

A ACP foi aplicada aos valores de parâmetros de cor para reduzir sua dimensionalidade no espaço multivariado e visualizar suas estruturas e padrões de distribuição que permitem identificar *clusters* e *outliers* (GALVÃO; VITORELLO; PARADELLA, 1995). Cada componente principal (CP) é representada por um eixo ortogonal aos demais eixos o qual é formado por uma combinação linear das variáveis originais. Nessa técnica, a distância preservada entre os descritores foi a distância euclidiana e as relações identificadas são lineares (BORCARD; GILLET; LEGENDRE, 2011). Os descritores utilizados nesse estudo foram os parâmetros de cor: XYZ, RGB, coordenadas cromáticas de Munsell HVC, CIE xyY, CIE L\*a\*b\*, CIE L\*u\*v\*, CIE L\*c\*h\* e CMYK. Como essas variáveis apresentam unidades de medida distintas, foi necessário transformar os valores para uma mesma escala, com o objetivo de equalizar a importância estatística de todas as variáveis simplificando as relações matemáticas, como sugerido por LEGENDRE; LEGENDRE (1998). Para isso, as variáveis foram linearmente transformadas por um processo de translação e expansão dos dados subtraindo-se cada valor da variável por uma constante (média) e, em seguida, dividindo-os por outra constante (desvio padrão). Essa transformação é denominada de standardização ou padronização, "z scores". Para realizar essa transformação, foi utilizada a função *scale* do pacote *base*. A ACP foi aplicada usando a função *princomp* do pacote *stats*, nos valores dos descritores para extrair as informações dos 22 parâmetros de cor em todo o conjunto de amostras, em única matriz de 260 x 22. Foram utilizados os scores das CPs para determinar o número ideal de *clusters* na etapa seguinte. Uma implementação geral da ACP é dada pela Eq. (1).

$$\begin{aligned}
 CP_1 &= (B_1 \cdot e_{1,1}) + (B_2 \cdot e_{1,2}) + \dots + (B_n \cdot e_{1,n}) \\
 CP_2 &= (B_1 \cdot e_{2,1}) + (B_2 \cdot e_{2,2}) + \dots + (B_n \cdot e_{2,n}) \\
 &\dots \\
 CP_n &= (B_1 \cdot e_{n,1}) + (B_2 \cdot e_{n,2}) + \dots + (B_n \cdot e_{n,n})
 \end{aligned} \tag{1}$$

onde: CP é o componente principal (*scores*), B é o dado original (parâmetros de cor), e é o autovetor (contribuição de cada dado original para o *score*) e n é o número de parâmetros.

### 3.2.6. Análise de clusterização das amostras

A fim de estratificar o conjunto de amostras de solo, a clusterização *Fuzzy K-means* (FKM) foi aplicada aos *scores* da ACP para discriminar diferentes amostras de solos com base nos valores de parâmetros de cor, descritos em detalhes por TERRA; DEMATTÊ; VISCARRA ROSSEL (2018). O número mais apropriado de *clusters* e as pontuações da ACP para a clusterização foram estabelecidos pelos seguintes índices: o coeficiente de partição (PC) (Eq. 2), a entropia de partição (PE) (Eq. 3) e o coeficiente de partição modificado (MPC) (Eq. 4) (FERRARO; GIORDANI, 2015; WU; YANG, 2005). Foram testados de 2 a 4 *scores* da ACP e de 2 a 8 *clusters*. De acordo com a melhor pontuação de PC e MPC, foi definido que o número ideal de *clusters* seria de três. Portanto, o grupo de 260 amostras de solo foi dividido em três subgrupos, sendo o primeiro com 91 amostras, o segundo com 95 amostras e o terceiro com 74 amostras. Para facilitar o entendimento, os subgrupos foram considerados como *clusters* (C1, C2 e C3) e o grupo das propriedades do solo contendo todas as amostras (sem clusterização) foi considerado como modelo sem estratificação (MSE).

$$PC(k) = \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^k u_{ij}^2 \quad (2)$$

onde  $\frac{1}{k} \leq PC(k) \leq 1$ .

$$PE(k) = -\frac{1}{n} \sum_{j=1}^n \sum_{i=1}^k u_{ij} \log_2 u_{ij} \quad (3)$$

onde  $0 \leq PE(k) \leq \log_2 k$ .

$$MPC(k) = 1 - \frac{k}{k-1} (1 - PC(k)) \quad (4)$$

onde  $0 \leq MPC(k) \leq 1$ .

### 3.2.7. Modelagem preditiva e validação

Modelos preditivos para o teor de COS e das frações areia, silte e argila foram calibrados para cada *cluster* e também para o MSE. Os dados foram

divididos em dois conjuntos com 80% das amostras para calibração (selecionadas aleatoriamente) e 20% para validação. Dessa forma, os conjuntos de calibração dos *clusters* 1, 2 e 3 continham 73, 76 e 59 amostras, respectivamente (ou seja, 80% das amostras de cada *cluster*). Os conjuntos de validação dos *clusters* 1, 2 e 3 continham 18, 19 e 15 amostras, respectivamente (ou seja, 20% do *cluster*). No total, foram utilizadas 208 amostras para calibrar os modelos e, no total, foram preditas 52 amostras. Para construir o MSE, as mesmas 208 amostras dos *clusters* foram utilizadas para calibrar o MSE e prever as 52 amostras usadas na validação. Assim, os grupos estratificados e não estratificados são equivalentes e permitiram uma comparação justa dos índices de acurácia. Para avaliar a performance dos modelos preditivos, dois índices de acurácia foram utilizados: o coeficiente de determinação ( $R^2$ ), que varia entre 0 e 1 e fornece a porcentagem da variação que é explicada pelo modelo e a raiz do erro quadrático médio (RMSE), que mede a acurácia geral do modelo de predição.

Nos modelos de predição, a cada vez que o modelo é executado pode retornar um valor diferente de  $R^2$  e RMSE, uma vez que é realizada a amostragem aleatória no conjunto amostral (ao acaso). Para verificar a variação de predição dos modelos foram realizadas as seguintes simulações: 1º) Para cada *cluster* os modelos de predição foram executados 18 vezes (3 *clusters* × 18 repetições = 54 modelos de predição para cada propriedade do solo); 2º) o mesmo foi realizado para o MSE (1 grupo × 18 repetições para cada propriedade do solo). As repetições foram definidas de acordo com o poder do teste estatístico (95%) para verificar a variabilidade de predição dos modelos. O poder do teste foi de 95% e é a probabilidade de que a hipótese nula fosse rejeitada a favor da hipótese alternativa (UTTLEY, 2019). O poder do teste estatístico foi realizado pelo pacote *stats* R (R DEVELOPMENT CORE TEAM, 2017).

Dessa forma, as estimativas foram obtidas considerando a média de  $R^2$  e RMSE, levando em considerando a variação apresentada pelos modelos. Os modelos foram calibrados utilizando a combinação das curvas espectrais VIS-NIR-SWIR correspondente a cada solo. Dois métodos multivariados foram implementados para avaliar o desempenho preditivo, sendo eles: *partial-least-squares regression* (PLSR) no pacote *pls* (MEVIK; WEHRENS; LILAND, 2016) e *support vector machines* (SVM) no pacote *e1071* (MEYER et al., 2020). Para

comparar o desempenho da modelagem, o teste t de *Student* foi aplicado aos valores médios de RMSE para verificar as diferenças entre as médias dos *clusters* e do MSE. O teste t de *Student* foi realizado pelo pacote *stats* R (R DEVELOPMENT CORE TEAM, 2017).

### 3.3. RESULTADOS E DISCUSSÃO

#### 3.3.1. Estatísticas descritivas

O teor de COS e a fração areia apresentam valores de coeficiente de variação (CV) superior a 70%, caracterizando a heterogeneidade dessas propriedades físico-químicas do solo (Tabela 4). A fração silte foi a propriedade que apresentou o menor CV, seguido da fração argila, 32 e 48%, respectivamente. A variabilidade do teor de COS e da fração areia está associada a característica húmica dos solos conferida pela condição climática, pelo relevo e pela formação geológica da área de estudo (DALMOLIN et al., 2017). A presença dos campos de altitude com acúmulo de COS nas camadas superficiais do solo, nas áreas de relevo plano (solos rasos) e a formação de matas de araucárias nas vertentes (solos mais profundos) também são responsáveis pela variabilidade granulométrica e dos teores de COS das amostras. Essa variabilidade é refletida tanto na distribuição das propriedades do solo, bem como nas medições quantificadas da cor do solo (Tabela 4). A matiz dos solos variou de 7,69YR a 10,00YR, com o valor variando de 1,64 a 5,15 unidades e o croma de 0,73 a 5,24 unidades. Como a medição da cor do solo de Munsell é subjetiva (POST et al., 1994), a obtenção da cor baseada em Vis pode fornecer medições quantitativas rápidas da cor do solo (VISCARRA ROSSEL; WALTER; FOUAD, 2003) e auxiliar na escolha da cor predominante.

Tabela 5. Estatística descritiva das propriedades do solo e dos parâmetros de cor.

Propied. do solo e parâmetros de cor	Média	DP	Mínimo	Máximo	CV %
COS (%)	11,07	7,70	0,64	34,84	70
Argila (%)	34,19	16,53	8,83	65,72	48
Areia (%)	27,30	19,95	2,77	79,74	73
Silte (%)	38,50	12,39	6,74	65,53	32
Matiz	9,04YR	0,48YR	7,69YR	10,00YR	5
Valor	3,14	0,74	1,64	5,15	23
Croma	2,47	1,07	0,73	5,24	43
X	8,22	4,08	2,37	22,58	50
Y	7,75	3,72	2,31	21,16	48
Z	4,92	1,52	1,89	13,55	31
x	0,38	0,02	0,34	0,43	6
y	0,37	0,02	0,34	0,39	4
L	32,27	7,64	17,01	53,12	24
a*	5,13	2,15	1,89	10,84	42
b*	14,63	6,78	3,99	31,23	46
u*	14,08	7,05	3,41	33,04	50
v*	16,02	7,98	3,55	35,78	50
c*	15,52	7,10	4,41	32,82	46
h*	70,19	2,01	64,57	74,82	3
R	92,02	23,21	50,00	157,00	25
G	73,78	15,29	44,00	120,00	21
B	55,64	6,71	37,00	89,00	12
C	19,03	0,16	19,00	20,00	1
M	26,65	3,23	22,00	35,00	12
Ye	33,76	6,94	23,00	52,00	21
K	43,20	9,11	18,00	60,00	21

DP: Desvio padrão; CV: Coeficiente de variação.

### 3.3.2. Análise de componentes principais (ACP)

Os resultados da análise de componentes principais (ACs) (Figura 3) mostraram que na primeira componente principal (CP1), responsável por 84% da variação total, praticamente todos os parâmetros apresentaram valores significativos de contribuição (0,92, em média), exceto o h\* (0,54) e o C (0,41) que apresentaram os menores valores de contribuição positiva. A segunda componente principal (CP2), responsável por 7%, está associada a uma forte influência da Matiz (0,86). Os parâmetros de cor que melhor determinaram o comportamento do conjunto de amostras de solos foram o valor, o croma, o CIE L\*b\*, o CIE L\*v\*c\*, o RG e o MYK (0,92, em média) no primeiro eixo e a Matiz (0,86) no segundo eixo. Esses resultados estão de acordo com VISCARRA ROSSEL; WALTER; FOUAD (2003) que ao estudarem os modelos de cor mais

adequados para descrição quantitativa da cor solo e sua relação com o teor de COS, observaram que os parâmetros de luminosidade (L), R dos modelos RGB e  $v^*$  dos modelos CIE  $L^*u^*v^*$  obtiveram as maiores correlações com o teor de COS. Da mesma forma, BAUMANN et al. (2016) ao utilizarem a cor do solo e a espectroscopia, observaram que a luminosidade (L) apresenta boa correlação com os teores de COS. Os parâmetros de luminosidade (L) variam conforme a região geográfica e o uso da terra e estão relacionados ao teor de COS e à composição química da MO (SPIELVOGEL; KNICKER; KÖGEL-KNABNER, 2004). O valor  $a^*$  dos modelos CIE  $L^*a^*b^*$  está relacionado a cor vermelha dos solos que pode estar sendo influenciado pelos óxidos de Fe, principalmente a hematita. O valor  $b^*$  dos modelos CIE  $L^*a^*b^*$ , está relacionado a cor amarela dos solos, refletindo diferenças nas condições geológicas e pedogenéticas (FISCHER et al., 2010) e pode ser uma estimativa do conteúdo de goethita, mostrando boa contribuição (0,98) na CP1. Dessa forma, os parâmetros de cor baseados em Vis podem ser aplicados para determinar a cor, a composição mineral, o conteúdo de argila do solo (AITKENHEAD et al., 2013a; DOMINGUEZ et al., 2010; VISCARRA ROSSEL et al., 2009), e na identificação do teor de COS (VODYANITSKII; SAVICHEV, 2017).

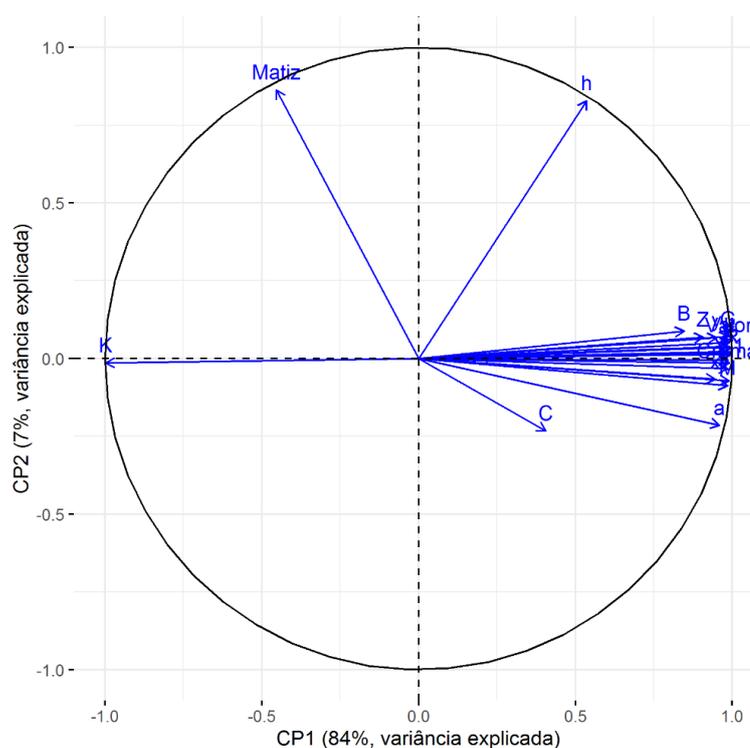


Figura 3. Resultado da análise de componentes principais dos parâmetros de cor.

### 3.3.3. Análise de cluster

A clusterização *fuzzy k-means* (FKM) foi aplicada para particionar o conjunto de amostras de solos em grupos (*clusters*) mais homogêneos, maximizando a heterogeneidade entre eles. Com base nas avaliações dos *clusters* (Tabela 5) e levando em consideração os valores de parâmetros de cor, as amostras de solos foram divididas em 3 *clusters* usando os dois primeiros *scores* da ACP que explicaram 91% da variabilidade dos dados (Tabela 5). A escolha do número ideal de *clusters* levou em consideração valores mais altos de PC (0,99) e MPC (0,99), que implicam em uma boa partição e também menores valores de PE (0,02), que implicam em uma partição mais nítida ("*crisp class*"), conforme sugerido por BEZDEK; EHRLICH; FULL (1984) e VISCARRA ROSSEL et al. (2016).

Tabela 6. Avaliações da clusterização *fuzzy k-means* e escolha do número ideal de *clusters*.

Número de pontuações	Variação acumulada %	Índices	Números de <i>clusters</i>						
			8	7	6	5	4	3	2
2	91	PC	0,93	0,95	0,95	0,98	0,97	0,99	0,98
		PE	0,13	0,09	0,10	0,04	0,05	0,02	0,04
		MPC	0,92	0,94	0,93	0,97	0,97	0,99	0,95
3	96	PC	0,97	0,98	0,98	0,96	0,95	0,92	0,94
		PE	0,97	0,05	0,06	0,09	0,10	0,15	0,10
		MPC	0,97	0,98	0,97	0,95	0,94	0,88	0,89
4	99	PC	0,97	0,99	0,99	0,96	0,93	0,90	0,92
		PE	0,06	0,03	0,04	0,08	0,14	0,18	0,14
		MPC	0,97	0,99	0,98	0,96	0,91	0,86	0,84

PC: coeficiente de partição; PE: entropia de partição; MPC: coeficiente de partição modificado.

Na classificação por FKM, as amostras de solos foram divididas de acordo com suas características. Os espectros médios de todas as amostras em cada *cluster* mostraram características de pico de absorção espectral semelhantes (Figura 5a). No entanto, o valor médio de reflectância entre os *clusters* foi significativamente diferente, principalmente por causa da diferença na média do teor de COS e a tendência da reflectância diminuir com o aumento do teor de COS. Por exemplo, o *cluster* 1 apresentou os maiores valores de

reflectância média e os menores teores de COS (4%, em média) (Figura 4a e 5a). Em contraste, o *cluster 2* apresentou a menor reflectância média e os maiores teores de COS (18%, em média). Esses resultados estão de acordo com o trabalho de LIU et al. (2019), que ao aplicarem a análise de *clusters* em amostras de solos, constataram que a variação do teor de COS influenciou na clusterização dos solos. Comportamentos espectrais semelhantes foram relatados por STENBERG et al. (2010) e NOCITA et al. (2014) que encontraram diferenças na reflectância para solos orgânicos na faixa de comprimento de onda NIR. Outro aspecto importante na análise espectral é que o *cluster 2* apresentou os solos com os maiores teores de MO e areia (Figura 4a e 4c). No entanto, a MO tem baixa reflectância e quando está acompanhada de material com maior reflectância (como o quartzo presente na fração areia), mascara a resposta espectral diminuindo a reflectância e, conseqüentemente, reduz o contraste das feições de absorção do material (DEMATTÊ; EPIPHANIO; FORMAGGIO, 2003). A absorção 2207 nm relacionada a argilominerais foi menos evidente para o *cluster 2*, já que agrupou amostras com os menores conteúdos de argila. Em contrapartida, no *cluster 1* a absorção foi mais perceptível, pois agrupou amostras com maior conteúdo de argila (Figura 4b e 5a).

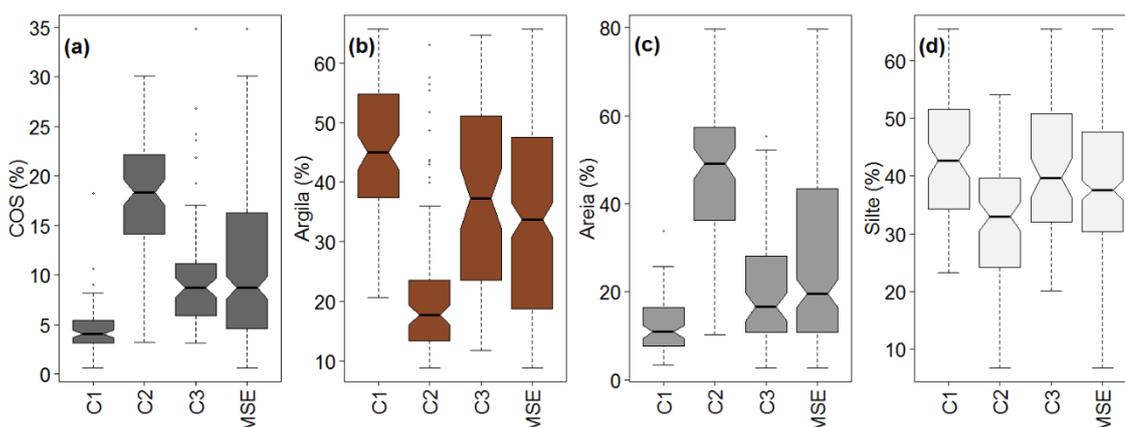


Figura 4. Distribuição estatística e variabilidade do teor de COS (a), argila (b), areia (c) e silte (d) para os modelos sem estratificação (MSE) e *clusters*.

As amostras do *cluster 1* mostraram características bem definidas para a absorção de óxidos de ferro ( $Fe^{III}$ ) nas bandas 480, 550 e 850 nm (Figura 5b), identificando a presença de goethita e hematita (DALMOLIN et al., 2005; MOURA-BUENO et al., 2019). Nos comprimentos de onda 1400 e 2200 nm

foram identificadas bandas de absorção devido às vibrações moleculares dos grupos OH e Al-OH presentes nos argilominerais, sendo principalmente relacionadas a presença de caulinita e/ou montmorilonita. No *cluster* 1, em determinados comprimentos de onda foram observados maior absorção de energia devido as amostras apresentarem textura argilosa e maior teor de óxidos de ferro ( $\text{Fe}^{\text{III}}$ ), enquanto as amostras dos *clusters* 2 e 3, com textura menos argilosa e menor teor de óxidos, esse comportamento não foi observado (Figura 5a e 5b).

O COS influenciou na forma e no albedo ao longo de toda a curva espectral (Figura 5a e 5b), sendo que, na literatura, essa propriedade do solo pode influenciar principalmente na região entre 400 e 1000 nm. Essa característica foi observada em amostras com alto teor de COS, que apresentaram reduções no albedo em toda a curva espectral, principalmente nas amostras dos *clusters* 2 e 3, mas com maior modificação de características entre 400 e 1000 nm. Essa redução no albedo devido ao maior teor de COS foi observada por LIU et al. (2019) e MOURA-BUENO et al. (2019) que constataram que a variabilidade do teor de COS influenciou no comportamento das curvas espectrais.

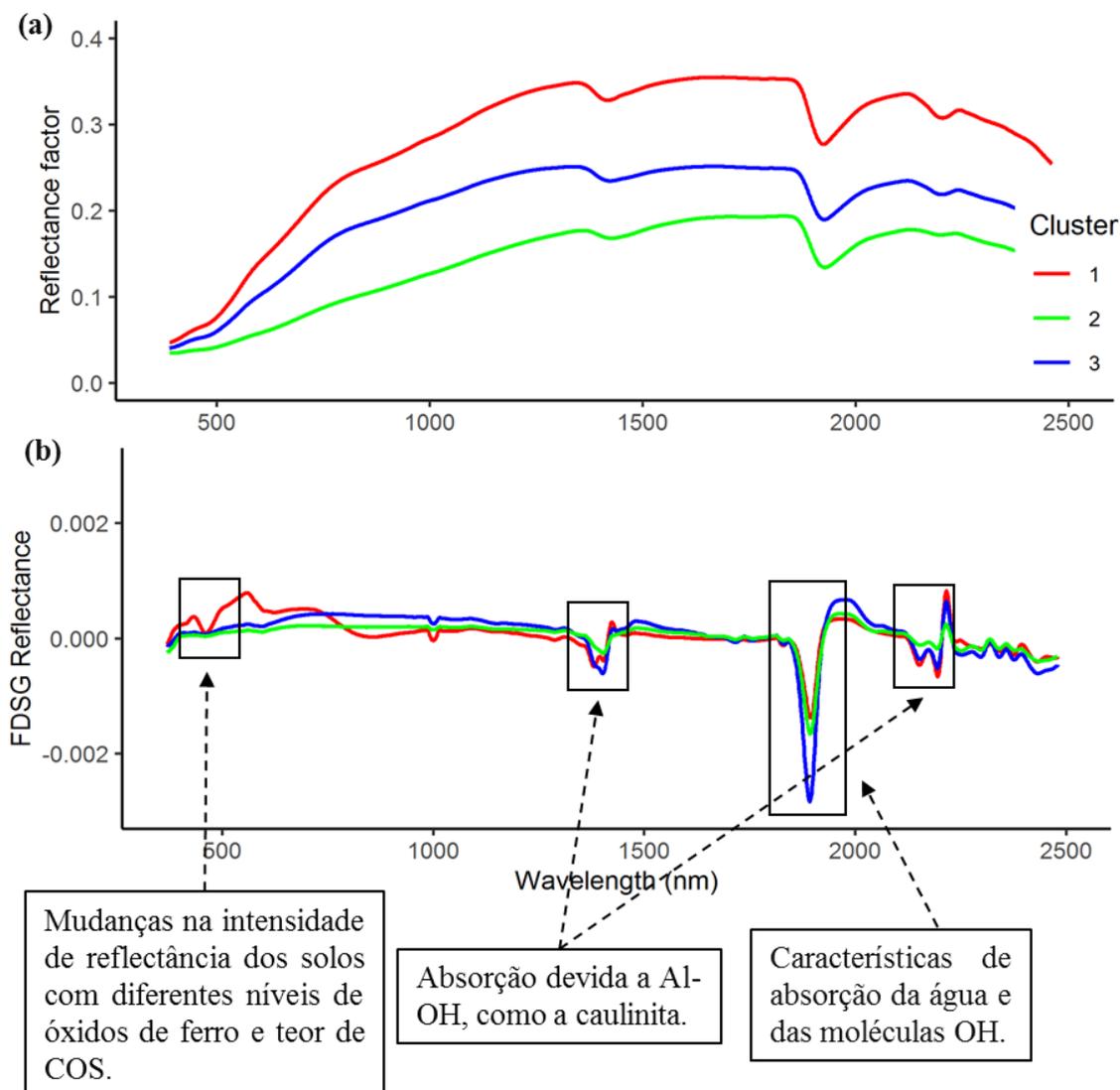


Figura 5. Espectros médios de reflectância dos clusters com base nos parâmetros de cor (a) e a primeira derivada das curvas espectrais, destacando bandas espectrais relacionadas aos componentes do solo e as alterações na intensidade da reflectância (b).

### 3.3.4. Desempenho de predição dos modelos espectrais

O desempenho dos modelos foi avaliado considerando as 18 repetições para cada modelo de predição (*clusters* e MSE), dessa forma consideramos a média de  $R^2$  e RMSE para avaliar o desempenho dos mesmos. De acordo com os resultados (Tabela 6), observa-se que existe variação na predição das propriedades do solo e as performances de predição dos modelos PLSR, em média, foram melhores que as dos modelos SVM. Na modelagem para o teor de COS, com o método PLSR, o *cluster 3* obteve os melhores índices

de acurácia, com valores médios de  $R^2$  e RMSE de 0,88 e 2,3%, respectivamente. Enquanto que o *cluster 1* obteve os menores índices com valores médios de  $R^2$  e RMSE de 0,63 e 1,4%, respectivamente. Para o método SVM, o *cluster 3* obteve os melhores índices com valores médios de  $R^2$  e RMSE de 0,82 e 2,7%, respectivamente. Enquanto que o *cluster 2* obteve o maior valor médio de RMSE de 2,9%. Na predição da fração argila, com o método PLSR, o *cluster 2* obteve os melhores índices com valores médios de  $R^2$  e RMSE de 0,72 e 7%, respectivamente. Enquanto que o *cluster 3* obteve o maior valor médio de RMSE de 9%. Para o método SVM, o *cluster 2* obteve valores médios de  $R^2$  e RMSE de 0,59 e 8,3%, respectivamente. Enquanto que o *cluster 1* obteve o menor desempenho com valores médios de  $R^2$  e RMSE de 0,30 e 9,6%, respectivamente. Com os resultados do desvio padrão, observa-se que houve maior variabilidade absoluta das validações dos *clusters* (Tabela 6), principalmente para o *cluster* contendo amostras de solos com menor teor de COS. Ao realizar as repetições, é possível encontrar valores mais precisos de predição, já que se considera as variações existentes em cada modelo, sem depender da escolha do melhor ou pior modelo de predição.

Tabela 7. Desempenho preditivo das propriedades do solo para o conjunto de validação considerando o MSE e a performance dos modelos após a análise de cluster baseado nos parâmetros de cor.

	Cluster 1 (n - 91)				Cluster 2 (n - 95)				Cluster 3 (n - 74)				MSE (n - 260)			
	SVM		PLSR		SVM		PLSR		SVM		PLSR		SVM		PLSR	
	$R^2$	RMSE	$R^2$	RMSE	$R^2$	RMSE	$R^2$	RMSE	$R^2$	RMSE	$R^2$	RMSE	$R^2$	RMSE	$R^2$	RMSE
<b>COS</b>		%		%		%		%		%		%		%		%
1	0,91	0,8	0,86	1,5	0,86	2,9	0,85	2,3	0,82	3,3	0,86	2,5	0,93	2,2	0,91	2,4
2	0,82	0,7	0,72	1,1	0,80	2,6	0,87	2,7	0,85	3,1	0,87	3,0	0,91	2,5	0,92	2,3
3	0,66	0,8	0,64	1,0	0,78	2,3	0,82	2,2	0,73	4,9	0,82	3,9	0,89	3,2	0,91	2,7
4	0,76	1,2	0,74	1,3	0,69	2,5	0,76	2,3	0,82	1,8	0,87	1,4	0,91	2,1	0,90	2,3
5	0,81	0,9	0,82	1,0	0,70	2,9	0,73	2,6	0,82	1,6	0,86	1,7	0,87	2,5	0,86	2,7
6	0,72	1,3	0,74	1,2	0,84	3,1	0,86	2,5	0,91	1,8	0,97	1,3	0,88	2,3	0,88	2,3
7	0,79	0,8	0,58	1,4	0,70	3,5	0,78	3,1	0,72	3,5	0,85	2,3	0,91	2,5	0,92	2,3
8	0,11	3,6	0,26	3,3	0,86	2,6	0,88	2,6	0,82	4,4	0,90	3,4	0,85	3,6	0,88	3,3
9	0,94	0,9	0,90	1,1	0,74	3,2	0,75	3,1	0,82	1,9	0,88	1,6	0,83	2,9	0,83	3,0
10	0,66	1,1	0,56	1,5	0,59	4,1	0,70	3,0	0,83	2,6	0,89	2,1	0,89	2,3	0,88	2,6
11	0,66	1,0	0,66	1,5	0,73	3,5	0,86	2,2	0,94	0,9	0,91	1,7	0,90	2,3	0,90	2,4
12	0,68	1,0	0,56	1,2	0,81	2,8	0,82	2,9	0,87	1,4	0,87	1,6	0,91	2,3	0,91	2,3
13	0,83	0,7	0,71	0,9	0,80	2,5	0,89	2,0	0,66	3,5	0,84	3,3	0,92	2,3	0,92	2,4
14	0,12	3,6	0,27	3,2	0,85	2,1	0,85	2,3	0,87	2,1	0,94	2,3	0,86	2,7	0,86	2,7

Tabela 7. Continuação.

15	0,60	1,0	0,56	1,1	0,75	2,7	0,83	2,5	0,89	4,1	0,90	3,5	0,86	3,1	0,90	2,7
16	0,67	0,9	0,44	1,1	0,78	2,7	0,88	2,0	0,70	3,2	0,81	2,7	0,90	2,3	0,87	2,7
17	0,69	1,0	0,54	1,2	0,76	3,4	0,82	2,8	0,81	2,0	0,82	1,9	0,87	2,7	0,86	2,8
18	0,83	0,6	0,73	0,9	0,70	2,9	0,74	2,9	0,96	2,4	0,96	1,5	0,88	2,9	0,88	2,8
Média	0,68	1,2	0,63	1,4	0,76	2,9	0,82	2,6	0,82	2,7	0,88	2,3	0,89	2,6	0,89	2,6
DP	0,23	0,9	0,18	0,7	0,07	0,5	0,06	0,4	0,08	1,1	0,05	0,8	0,03	0,4	0,03	0,3
Max.	0,94	3,6	0,90	3,3	0,86	4,1	0,89	3,1	0,96	4,9	0,97	3,9	0,93	3,6	0,92	3,3
Min.	0,11	0,6	0,26	0,9	0,59	2,1	0,70	2,0	0,66	0,9	0,81	1,3	0,83	2,1	0,83	2,3
<b>Argila</b>																
1	0,63	6,4	0,67	6,0	0,82	8,1	0,90	5,4	0,76	7,2	0,88	5,4	0,76	8,2	0,77	7,9
2	0,19	10,5	0,64	6,4	0,80	5,5	0,88	4,3	0,77	8,9	0,90	6,3	0,77	8,3	0,76	8,4
3	0,56	8,0	0,67	6,7	0,71	7,8	0,86	5,4	0,90	6,9	0,88	6,8	0,71	9,2	0,67	9,7
4	0,34	9,5	0,50	6,9	0,80	7,1	0,81	7,3	0,84	7,4	0,82	7,5	0,76	8,4	0,81	7,5
5	0,47	8,5	0,61	6,9	0,59	6,4	0,79	5,2	0,65	9,4	0,70	8,0	0,70	8,7	0,73	8,1
6	0,46	8,2	0,59	6,9	0,81	5,6	0,79	5,9	0,62	9,2	0,75	8,8	0,76	7,7	0,72	8,3
7	0,24	9,5	0,51	7,1	0,23	11,1	0,78	6,0	0,62	9,8	0,66	8,8	0,64	9,8	0,65	9,3
8	0,43	8,0	0,54	7,2	0,69	8,4	0,77	6,8	0,75	7,9	0,82	8,9	0,71	8,7	0,74	8,3
9	0,16	10,3	0,48	7,2	0,63	10,2	0,76	7,9	0,70	9,8	0,74	9,1	0,65	10,2	0,67	9,9
10	0,27	7,8	0,27	7,4	0,60	8,1	0,68	8,7	0,66	9,3	0,68	9,5	0,79	7,8	0,72	8,9
11	0,20	9,2	0,39	7,6	0,52	7,8	0,67	6,4	0,47	11,3	0,62	9,7	0,74	8,9	0,73	9,1
12	0,30	9,4	0,52	7,7	0,63	6,6	0,66	6,3	0,69	9,9	0,68	9,7	0,73	8,4	0,67	9,2
13	0,02	13,4	0,48	7,8	0,69	7,0	0,66	6,9	0,76	9,3	0,79	9,8	0,81	7,4	0,78	7,9
14	0,44	9,3	0,56	7,8	0,50	8,5	0,65	8,0	0,52	11,8	0,61	10,4	0,59	10,0	0,63	9,6
15	0,38	9,6	0,64	7,9	0,48	10,1	0,61	9,6	0,51	12,3	0,57	10,6	0,71	9,1	0,72	8,9
16	0,18	11,9	0,60	7,9	0,34	13,0	0,60	10,1	0,63	12,0	0,74	10,9	0,75	8,3	0,71	8,9
17	0,17	10,4	0,64	8,0	0,39	11,7	0,54	10,0	0,66	10,7	0,80	11,2	0,72	8,7	0,74	8,4
18	0,00	12,9	0,50	8,0	0,44	7,0	0,54	6,3	0,48	13,3	0,64	11,3	0,69	10,1	0,71	9,7
Média	0,30	9,6	0,54	7,3	0,59	8,3	0,72	7,0	0,67	9,8	0,74	9,0	0,72	8,8	0,72	8,8
DP	0,17	1,8	0,10	0,6	0,17	2,1	0,11	1,7	0,12	1,8	0,10	1,7	0,05	0,8	0,05	0,7
Max.	0,63	13,4	0,67	8,0	0,82	13,0	0,90	10,1	0,90	13,3	0,90	11,3	0,81	10,2	0,81	9,9
Min.	0,00	6,4	0,27	6,0	0,23	5,5	0,54	4,3	0,47	6,9	0,57	5,4	0,59	7,4	0,63	7,5

MSE: Modelos sem estratificação

Em geral, o erro do modelo PLSR (*clusters* e MSE) foi menor do que os observados para o modelo SVM (Tabela 6). O erro médio dos modelos foi frequentemente relacionado ao aumento do teor de COS (Tabela 6 e Figura 4a). Por exemplo, para o *cluster* 1 o teor médio de COS foi de 4% e os valores de RMSE, em média, foram de 1,2 e 1,4%, para os modelos SVM e PLSR, respectivamente. Enquanto que o teor médio de COS do *cluster* 2 foi de 18% e os valores de RMSE, em média, foram de 2,9 e 2,6%, para os modelos SVM e PLSR, respectivamente. Segundo NOCITA et al. (2014) os erros de predição tendem a aumentar com o aumento do teor de COS. Esse resultado foi

confirmado pela baixa reflectância do *cluster* 2 (Figura 5a), devido ao aumento do teor de MO, que possui baixo nível de reflectância e mascara a reflectância das amostras (NOCITA et al., 2014; STENBERG et al., 2010).

### 3.3.5. Predição das propriedades do solo utilizando os clusters e o MSE

A análise de *cluster* baseada na cor dos solos, influenciou na predição do teor de COS (Tabela 7 e Figura 6). Em geral, o erro obtido dos *clusters* foi menor que o erro obtido no MSE. De acordo com os resultados, o teor de COS foi a propriedade do solo onde a redução do erro foi mais significativa e as performances de predição dos modelos PLSR, em média, foram melhores que as dos modelos SVM. Na modelagem para o teor de COS, com o método PLSR, os valores médios de RMSE foram de  $2,6 \pm 0,3$  para o MSE e  $2,1 \pm 0,8$  para os *clusters*, representando uma redução de 19% do erro na predição (Tabela 7 e Figura 6). Para o método SVM, os valores médios de RMSE foram de  $2,6 \pm 0,4$  para o MSE e  $2,3 \pm 1,1$  para os *clusters*, não sendo verificada redução ( $p > 0,05$ ). Esses valores estão próximos aos observados em um estudo de predição do teor de COS por JACONI; DON; FREIBAUER (2017), que aplicaram métodos de *clusters* baseados em dados de profundidade do solo, uso da terra, pH e textura do solo para desenvolver modelos espectrais individuais para cada subconjunto. Segundo os autores, a estratificação do conjunto de dados reduziu em 22% o erro dos modelos agrupados comparado ao MSE. Da mesma forma, NOCITA et al. (2014) ao predizerem o teor de COS, selecionaram amostras de solos com base na distância espectral e na combinação espectral com a fração areia (variável preditora). Segundo os autores, solos com maior teor de COS apresentaram maior erro de predição.

Para a predição da fração argila, com o método PLSR, os valores médios de RMSE foram de  $8,8 \pm 0,7$  para o MSE e  $7,8 \pm 1,6$  para os *clusters*, representando uma redução significativa de 11% do erro na predição (Tabela 7 e Figura 6). Para a fração areia, com o método PLSR, a redução do erro também foi significativa, com os valores médios de RMSE de  $10,9 \pm 0,7$  para o MSE e  $9,1 \pm 3,5$  para os *clusters*, representando uma redução de 17% do erro na predição. Para a fração silte, com o método PLSR, não houve redução do erro. Para o método SVM, os valores médios de RMSE dos *clusters* foram superiores aos do

MSE, mostrando baixo desempenho dessa metodologia na performance de predição para a fração silte.

Na comparação entre os métodos de classificação das amostras (Tabela 7), o método que considerou apenas a região do Vis obteve a maior diferença entre os valores médios de RMSE (para os *clusters* e MSE), com reduções significativas do erro para o teor de COS (19%, método PLSR), para a fração argila (11%, método PLSR) e para a fração areia (17%, método PLSR). Enquanto o método que considerou o VIS-NIR-SWIR não apresentou reduções significativas entre os valores médios de RMSE.

Tabela 8. Diferença estatística e redução do erro de predição nos resultados de predição dos métodos PLSR e SVM para cada propriedade do solo.

Propriedades do solo	Modelo	Processo	Média dos valores RMSE (%)*	Redução do erro (%)*	Média dos valores RMSE (%)**	Redução do erro (%)**
COS	PLSR	<i>Clusters</i>	2,1 ± 0,8 a	19	2,7 ± 0,6 a	ns
		MSE	2,6 ± 0,3 b		2,7 ± 0,2 a	
	SVM	<i>Clusters</i>	2,3 ± 1,1 a	ns	2,7 ± 0,6 a	ns
		MSE	2,6 ± 0,4 a		2,7 ± 0,4 a	
Argila	PLSR	<i>Clusters</i>	7,8 ± 1,6 a	11	10,0 ± 2,5 a	ns
		MSE	8,8 ± 0,7 b		9,3 ± 1,1 a	
	SVM	<i>Clusters</i>	9,2 ± 2,0 a	ns	10,0 ± 2,4 a	ns
		MSE	8,8 ± 0,8 a		9,3 ± 1,2 a	
Areia	PLSR	<i>Clusters</i>	9,1 ± 3,5 a	17	10,5 ± 2,5 a	ns
		MSE	10,9 ± 0,7 b		10,8 ± 1,0 a	
	SVM	<i>Clusters</i>	11,1 ± 4,6 a	ns	11,0 ± 2,5 a	ns
		MSE	11,4 ± 0,6 a		10,6 ± 1,0 a	
Silte	PLSR	<i>Clusters</i>	9,2 ± 1,6 a	ns	10,4 ± 1,8 a	ns
		MSE	9,7 ± 0,8 a		9,7 ± 0,7 a	
	SVM	<i>Clusters</i>	10,4 ± 2,0 a	ns	11,0 ± 2,1 a	12
		MSE	9,9 ± 1,0 a		9,8 ± 0,7 b	

\*Considerando apenas a região Vis para classificar as amostras de solo; \*\*Considerando a região Vis-NIR-SWIR para classificar as amostras de solo; Letras iguais indicam que não há diferença significativa e letras diferentes indicam que há diferença significativa entre os valores médios de RMSE para os *clusters* e MSE, segundo o teste t *student* em  $p < 0.05$ . ns, não significativo.

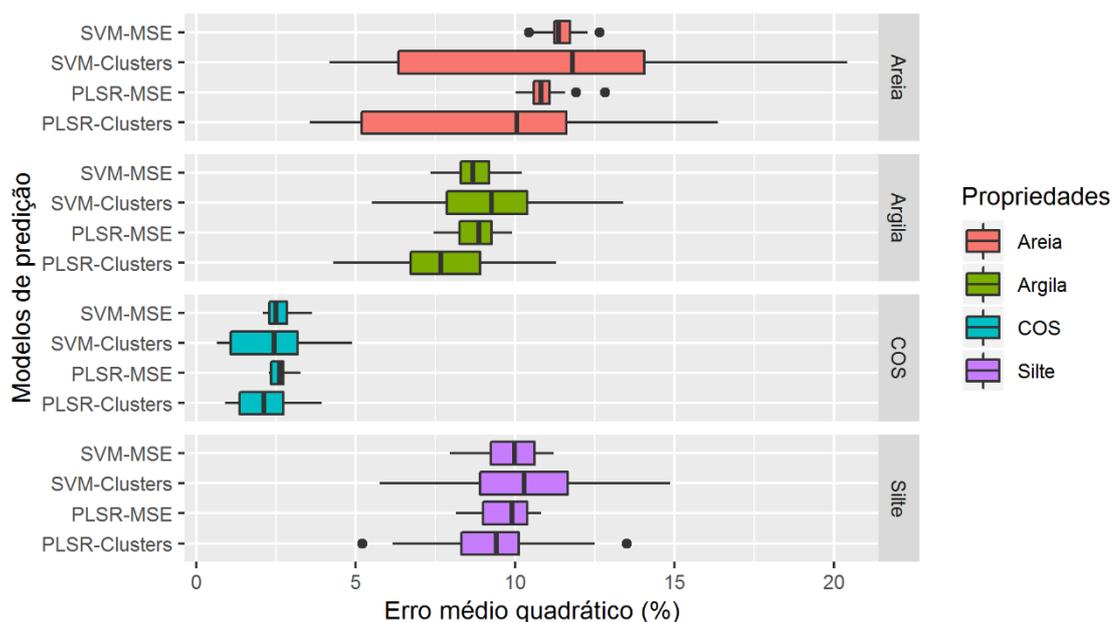


Figura 6. Performance dos modelos de predição considerando os modelos sem estratificação (MSE) e a performance dos modelos após a análise de *cluster* com base nos parâmetros de cor. Valores de mediana, 1º e 2º quartil.

A aplicação da análise de *cluster* não apenas melhorou a predição dos modelos PLSR e SVM, mas também reduziu o erro de ambos. Como os solos deste estudo continham informações abrangentes de várias propriedades do solo (Tabela 4 e Figura 4), a análise multivariada classificou as amostras de solo com base nos parâmetros de cor e agrupou o conjunto de dados complexos em *clusters* com características espectrais semelhantes, reduzindo assim a interferência que a variação dessas propriedades poderia representar na predição (GUERRERO et al., 2010; SHI et al., 2015). Ao contrário do MSE, que apresentou amostras de solos como maior variabilidade de dados, resultando em maior erro de predição. Segundo GUERRERO et al. (2014), se as amostras de solos forem semelhantes ao conjunto geral de predição, o modelo deverá fornecer resultados mais precisos.

Com a aplicação desta metodologia, é necessário entender porque os parâmetros de cor conseguiram classificar as amostras com base nas relações das características do solo e da cor. Para algumas propriedades do solo a explicação de uma relação com a cor do solo pode ser mais clara do que para outras, por exemplo: a) a matéria orgânica é um constituinte primário da coloração do solo, conferindo cores escuras aos horizontes superficiais e em

alguns horizontes subsuperficiais (iluviação); b) o quartzo presente na fração areia confere coloração branca, aumentando a intensidade da reflectância do solo; e c) os óxidos de ferro ( $\text{Fe}^{+3}$ ) presentes em solos bem drenados e intemperizados, conferem coloração vermelha (hematita) ou amarela (goethita), principalmente (AITKENHEAD et al., 2013a). Cada um desses fatores influencia a cor do solo, permitindo, dessa forma, que técnicas de clusterização baseadas em parâmetros de cor sejam bem estabelecidas.

Este estudo, diferente de outras tentativas de melhorar os modelos de predição, buscou alternativas mais simples e sem custos adicionais para aumentar o potencial de ganho na predição das propriedades do solo. Outras abordagens usaram métodos de estratificação baseada na profundidade, pH e textura do solo para dividir o grupo de dados em subgrupos (JACONI; DON; FREIBAUER, 2017). As estratégias de clusterizar as amostras de solos, geralmente são baseadas nas propriedades do solo de acordo com a profundidade e também com os tipos de uso e cobertura da terra, ou suas combinações (STEVENS et al., 2013; VASQUES; GRUNWALD; HARRIS, 2010). STEVENS et al. (2013) adicionaram novas preditoras à matriz espectral na tentativa de melhorar os modelos de predição do teor de COS, estes autores testaram as frações areia e argila como preditoras auxiliares. No entanto, assim como o teor de COS, essas propriedades físicas do solo também necessitam da sua determinação em laboratório e usá-las como variáveis preditoras encarece a predição. Em contraste, a clusterização das amostras pelas cores, não gera maiores custos já que a cor pode ser obtida a partir dos parâmetros de cor derivados das curvas espectrais (Vis), sendo um método promissor, podendo ser utilizado nos laboratórios para obter resultados mais precisos na determinação de propriedades do solo.

O potencial de discriminação da análise de estatística multivariada aplicada aos parâmetros de cor foi possível devido à variabilidade das características dos solos para os teores de COS e textura do solo. Isso garantiu que na classificação das amostras fosse possível dividir 3 *clusters* com padrões bem definidos. Em um conjunto de solos com maior variabilidade nas características dos solos, o potencial desta metodologia poderia ser ainda maior.

### 3.4. CONCLUSÕES

Este estudo utilizou os parâmetros de cores derivados do espectro visível para classificar as amostras de solos na predição de propriedades físico-químicas do solo. Além disso, comparamos os diferentes modelos de predição, tanto os dos conjuntos de amostras agrupadas com base nos parâmetros de cores, quanto do MSE (sem estratificar o conjunto de amostras). A análise de estatística multivariada aplicada aos parâmetros de cor foi capaz de clusterizar (ou agrupar) amostras de solos com características semelhantes, reduzindo a amplitude dos dados e melhorando a precisão das predições das propriedades do solo.

A redução do erro com a estratificação foi significativa para o teor de COS, seguido das frações areia e argila. Os erros de predição para o teor de COS usando o modelo PLSR foram em média  $2,6 \pm 0,3$  para o MSE e  $2,1 \pm 0,8$  para os *clusters*, o que representou uma redução de 19% no erro de predição. Para a fração argila, os erros de predição usando o PLSR foram em média  $8,8 \pm 0,7$  para o MSE e  $7,8 \pm 1,6$  para os *clusters*, o que representou uma redução de 11% no erro de predição. No geral, o modelo PLSR apresentou melhor desempenho que o modelo SVM, o que se confirmou pela diferença estatística entre os resultados do RMSE.

Portanto, o uso de parâmetros de cores baseados em Vis para classificar amostras de solos pode ser uma maneira de aumentar o potencial de ganho na predição de propriedades físico-químicas do solo, podendo ser utilizado nos laboratórios sem custos adicionais.

## **4. CAPÍTULO III – ESTUDO 2: PREDIÇÃO DA TEXTURA DO SOLO A PARTIR DA ESTRATIFICAÇÃO DE UMA BIBLIOTECA ESPECTRAL REGIONAL DE SOLOS<sup>3</sup>**

### **4.1. INTRODUÇÃO**

A textura do solo é uma importante propriedade que influencia na capacidade de infiltração e disponibilidade de água, na absorção de nutrientes pelas plantas, na resistência à penetração radicular, na atividade microbiana, na suscetibilidade à erosão e compactação, entre outras (JACONI; VOS; DON, 2019; PHOGAT et al., 2014; TÜMSAVAŞ et al., 2019). Assim, pela importância que desempenha na produção agrícola, são necessários métodos quantitativos rápidos, precisos e com boa relação custo-benefício que possibilitem predizer essa propriedade em diferentes escalas (JACONI; VOS; DON, 2019) e subsidiar tomadas de decisão no contexto agrícola e ambiental.

Nas últimas décadas, pesquisadores tem utilizado a espectroscopia de reflectância (ER) na determinação de importantes propriedades do solo (JACONI; DON; FREIBAUER, 2017; NOCITA et al., 2014; VISCARRA ROSSEL; WEBSTER, 2012), incluindo a textura (SILVA et al., 2019; XU et al., 2018). As faixas espectrais comumente utilizadas na predição correspondem a 400-700 nm (visível-Vis), 700-1100 nm (infravermelho próximo-NIR) e 1100-2500 nm (infravermelho de ondas curtas - SWIR) (DEMATTE et al., 2019; VISCARRA ROSSEL; WEBSTER, 2012). Essa técnica aumentou sua utilidade por ser capaz de mensurar várias propriedades do solo em apenas uma leitura espectral, podendo analisar grandes quantidades de amostras de maneira rápida, com boa relação custo-benefício, sem uso de reagentes químicos e produção de resíduos (DEMATTE et al., 2019; JI et al., 2016; MINASNY; MCBRATNEY, 2008; VISCARRA ROSSEL; BEHRENS, 2010).

---

<sup>3</sup>Adaptado do artigo aceito na revista *Pedosphere* em junho de 2020.

Para obter melhor desempenho dos modelos preditivos, muitos estudos testaram diferentes técnicas de pré-processamento espectral (tratamentos) e métodos multivariados (ARAÚJO et al., 2014; DEMATTÊ et al., 2019; JACONI; VOS; DON, 2019; SILVA et al., 2019). A Tabela 2 (Estudo 1 – Capítulo II) apresenta estudos realizados na estimativa da textura do solo utilizando ER, com diferentes tipos de pré-processamento espectral e métodos multivariados. Esses estudos mostraram a existência de uma variação nos índices de acurácia (coeficiente de determinação =  $R^2$  e raiz do erro quadrático médio = RMSE), que pode ser atribuída às diferenças existentes entre os tipos de pré-processamento e entre os métodos multivariados aplicados aos dados espectrais (ARAÚJO et al., 2014; DEMATTÊ et al., 2016; DOTTO et al., 2017; NAWAR et al., 2016; SILVA et al., 2019). PINHEIRO et al. (2017) utilizaram uma biblioteca espectral de solos e o modelo multivariado *partial least square regression* (PLSR) na predição da textura do solo, relatando capacidade preditiva moderada de areia ( $R^2 = 0,62$  e RMSE = 11,5%), baixa capacidade preditiva do silte ( $R^2 = 0,36$  e RMSE = 9,5%) e maior desempenho de predição para a argila ( $R^2 = 0,78$  e RMSE = 6,2%). NAWAR et al. (2016) testaram três modelos multivariados e sete técnicas de pré-processamento na predição da argila, onde o modelo *multivariate adaptive regression splines* (MARS) em combinação com o pré-processamento contínuo removível (CR) obtiveram os melhores resultados de validação ( $R^2 = 0,79$  e RMSE = 7,6%). SILVA et al. (2019) obtiveram melhores resultados preditivos usando o modelo *cubist* em comparação com *support vector machines* (SVM), *gaussian process regression* (GPR), *random forest* (RF) e PLSR. Nos estudos de ARAÚJO et al. (2014), o modelo SVM apresentou melhores índices de acurácia em comparação ao PLSR nas estimativas da argila. Enquanto que para DOTTO et al. (2017), ao utilizarem os métodos PLSR e SVM, a estimativa para os teores de argila e silte apresentaram desempenho moderado em oposição ao teor de areia, que apresentou baixo desempenho.

A variabilidade preditiva dos modelos também pode ser atribuída às características muito variáveis dos solos que integram as bibliotecas espectrais, como o teor de argila e mineralogia, o conteúdo de matéria orgânica, a umidade e os óxidos de ferro, entre outros (DALMOLIN et al., 2005; STENBERG et al., 2010; VISCARRA ROSSEL et al., 2016). Para contornar esse problema, uma

alternativa seria a estratificação das bibliotecas espectrais em conjuntos de amostras mais homogêneos. ARAÚJO et al. (2014), ao dividirem os dados de uma biblioteca espectral na tentativa de melhorar as precisões do teor de argila, obtiveram redução de 21% do erro de predição para a argila. MOURA-BUENO et al. (2019), ao estratificarem um conjunto de amostras de solo com base nas classes de solos, usos e camadas do solo, observaram que a separação das amostras em grupos mais homogêneos melhorou a precisão dos modelos multivariados na predição do teor de COS.

Muitas propriedades e condições do solo exercem influência na sua cor, como a matéria orgânica que confere uma coloração escura; a coloração branca do quartzo, calcita e outros carbonatos; a coloração vermelha da hematita ( $\alpha\text{-Fe}_2\text{O}_3$ ) e a coloração amarela da goethita ( $\alpha\text{-FeOOH}$ ) (AITKENHEAD et al., 2013a, 2013b). Pelas relações conhecidas entre a cor do solo e suas características mineralógicas e físico-químicas (AITKENHEAD et al., 2013a, 2013b; BEN-DOR; INBAR; CHEN, 1997; VISCARRA ROSSEL et al., 2009), a cor também pode ser usada como critério na estratificação das amostras de solos, podendo ainda fornecer aos modelos de predição conjuntos de amostras com características mais homogêneas. Além disso, a cor do solo pode ser obtida a partir das próprias curvas espectrais e ser representada por modelos tridimensionais de espaço de cores (coordenadas XYZ). Os valores XYZ podem ser transformados em outros modelos de espaço de cor (por exemplo, RGB, CIE  $L^*a^*b^*$ , CIE  $L^*u^*v^*$  e CIE xyY, entre outros) com a finalidade de melhorar a representação das cores (SIMON et al., 2020; VISCARRA ROSSEL et al., 2006a).

Motivado pela hipótese de que a estratificação de uma biblioteca espectral regional de solos pode melhorar a precisão da predição da textura do solo, uma vez que geralmente apresenta alta variabilidade espacial, neste estudo é proposta uma estratégia para estratificar o conjunto de amostras com base na cor dos solos. Assim, os objetivos deste estudo foram (a) aplicar os parâmetros de cor baseados em Vis na estratificação de uma biblioteca espectral regional de solos; (b) avaliar o desempenho dos modelos preditivos gerados a partir da estratificação *Fuzzy K-means* da biblioteca espectral de solo e (c) explicar as possíveis mudanças na precisão da predição da textura do solo com base nos modelos estratificados.

## 4.2. MATERIAL E MÉTODOS

### 4.2.1. Área de estudo e biblioteca espectral de solos

O estudo foi realizado com 1.535 amostras do banco de dados de solos da Epagri – Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina e que compõem a Biblioteca Espectral de Solos do Brasil – BESB (DEMATTÊ et al., 2019). A distribuição espacial abrangeu 260 municípios (cerca de 90%) do Estado de Santa Catarina (SC), Brasil (Figura 7). De acordo com o sistema de classificação Köppen, SC apresenta clima mesotérmico úmido - Cf, incluindo os subtipos Cfa e Cfb. O clima subtropical Cfa é caracterizado por temperatura média no mês mais frio inferior a 18°C (super úmido e mesotérmico) e temperatura média no mês mais quente acima de 22°C (PANDOLFO et al., 2002). O clima temperado Cfb é caracterizado por temperatura média no mês mais frio abaixo de 18°C (bastante úmido e mesotérmico) e temperatura média no mês mais quente abaixo de 22°C (PANDOLFO et al., 2002). De acordo com o Sistema Brasileiro de Classificação de Solos (SANTOS et al., 2013), as classes de solos predominantes são Cambissolos (*Inceptisols*), Neossolos (*Entisols*), Nitossolos (*Ultisols*, *Oxisols*), Argissolos (*Ultisols*), Latossolos (*Oxisols*) e Gleissolos (*Aquents*).

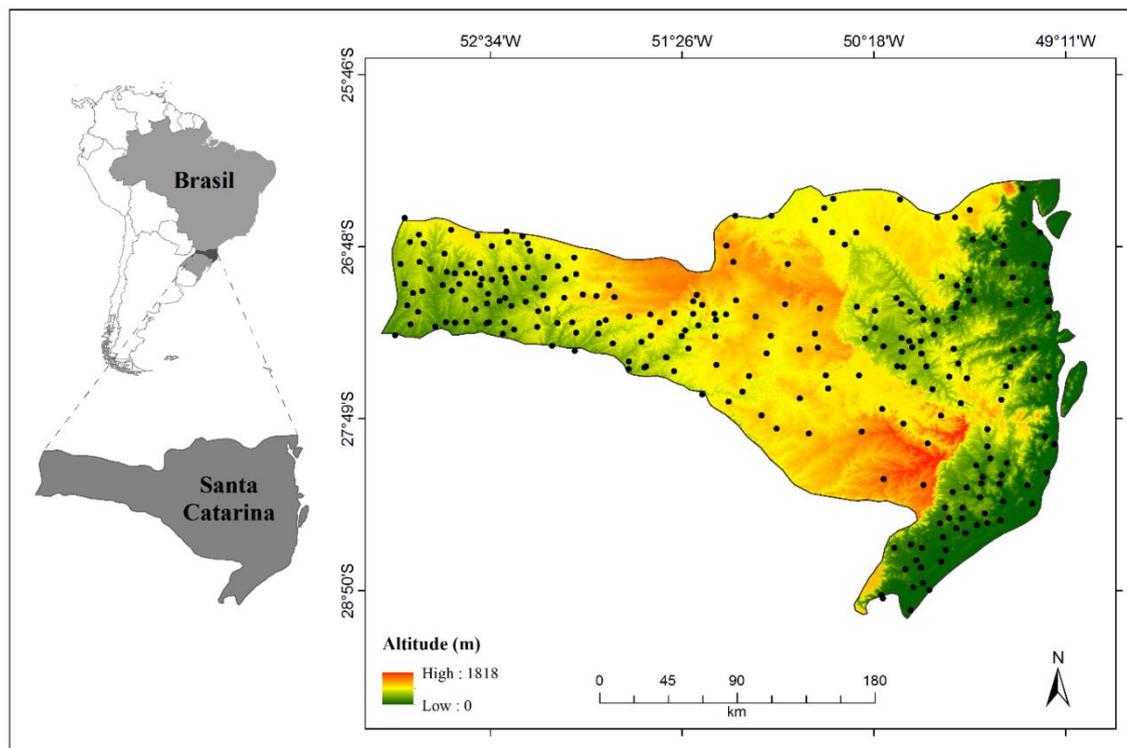


Figura 7. Localização dos pontos de coleta de solos no Estado de Santa Catarina, Brasil.

As amostras de solos foram coletadas pela Epagri em áreas agrícolas, em propriedades rurais. A profundidade de coleta foi de 0,0 – 0,5 m e as frações areia, silte e argila foram determinadas pelo método da Pipeta (DONAGEMMA et al., 2017), conforme descrito por VEIGA et al. (2012).

#### 4.2.2. Medidas de reflectância espectral e pré-processamento

Para obtenção da reflectância espectral, foi utilizado um espectrorradiômetro modelo FieldSpec 3 (Analytical Spectral Devices, Boulder, EUA) que recobre a faixa espectral entre 350 a 2500 nm, com resolução de 1 nm. A geometria da leitura espectral seguiu a metodologia da BESB e está descrita em detalhes em ROMERO et al. (2018). O sensor foi calibrado no início das medições espectrais e a cada 20 leituras, usando uma placa branca *Spectralon*. Para cada amostra foram obtidas três repetições, sendo utilizada a curva espectral média, conforme descrito por SILVA et al. (2019).

### 4.2.3. Cálculo dos parâmetros de cor baseados em Vis

A cor do solo pode ser representada por modelos tridimensionais de espaços de cores (coordenadas XYZ) e ser transformada para outros modelos de cores usando modelos de colorimetria. Esses modelos foram propostos pela *Commission Internationale de l'Eclairage* (CIE, 1996) para facilitar a visualização e padronizar os modelos de cores. A região do visível do espectro (400 a 700 nm) e o iluminante C foram usados para calcular os valores triestímulos XYZ. Nesse cálculo foram utilizados os mesmos critérios e parâmetros de cor utilizados no Estudo 1 (Capítulo II). Assim, foram obtidos 22 parâmetros de cor derivados do espectro Vis usando vários modelos de colorimetria descritos em detalhes por VISCARRA ROSSEL et al. (2006). A Tabela 3 (Estudo 1 – Capítulo II) apresenta um resumo dos modelos de espaço de cores usados neste estudo e as abreviações.

### 4.2.4. Análise de componentes principais (ACP)

A ACP foi aplicada aos valores de parâmetros de cor para reduzir sua dimensionalidade no espaço multivariado e melhorar a eficiência computacional (GALVÃO; VITORELLO; PARADELLA, 1995). Nessa técnica, a distância preservada entre os descritores foi a distância euclidiana e as relações identificadas são lineares (BORCARD; GILLET; LEGENDRE, 2011). Os descritores utilizados nesse estudo foram os parâmetros de cor: XYZ, RGB, coordenadas cromáticas de Munsell HVC, CIE xyY, CIE L\*a\*b\*, CIE L\*u\*v\*, CIE L\*c\*h\* e CMYK. Como essas variáveis apresentam unidades de medida distintas, elas foram linearmente transformadas por um processo de translação e expansão dos dados subtraindo-se cada valor da variável por uma constante (média) e, em seguida, dividindo-os por outra constante (desvio padrão). Essa transformação é denominada de standardização ou padronização (*z scores*). A ACP foi aplicada usando a função *princomp* do pacote *stats*, nos valores dos descritores para extrair as informações dos parâmetros de cor em todo o conjunto de amostras, em única matriz de 1535 x 22. Foram utilizados os *scores* e autovetores da ACP para determinar o número ideal de *clusters* na etapa seguinte.

#### 4.2.5. Análise de clusterização das amostras

A fim de estratificar a biblioteca espectral de solos e representá-la em padrões bem definidos, a clusterização *Fuzzy K-means* (FKM) foi aplicada aos *scores* da ACP para discriminar diferentes amostras de solos com base nos valores de parâmetros de cor, descritos em detalhes por TERRA; DEMATTÊ; VISCARRA ROSSEL (2018). Para determinar o número ideal de *clusters* necessários para representar o conjunto de amostras, foi utilizado os quatro primeiros *scores* da ACP. A técnica FKM atribui um grau de associação *fuzzy* a cada amostra com base na sua distância para o centro do *cluster*. Os graus de associações *fuzzy* são contínuos e variam de 0 a 1. Cada amostra pode ter maior associação a determinado *cluster*, onde a proximidade do valor 1 indica um alto grau de similaridade entre a amostra e um *cluster* e quando mais próxima ao valor 0 indica uma baixa similaridade (BEZDEK; EHRLICH; FULL, 1984). O número mais apropriado de *clusters* e as pontuações da ACP para a clusterização foram estabelecidos pelos seguintes índices: o coeficiente de partição (PC), a entropia de partição (PE) e o coeficiente de partição modificado (MPC) (FERRARO; GIORDANI, 2015; WU; YANG, 2005). Foram testados de 2 a 4 *scores* da ACP e de 2 a 8 *clusters*. A escolha do número ideal de *clusters* foi definida de acordo com a melhor pontuação de PC e MPC. Portanto, o grupo de 1535 amostras de solo foi dividido em quatro *clusters*, sendo o *cluster* 1 (C1) com 400 amostras, o *cluster* 2 (C2) com 265 amostras, o *cluster* 3 (C3) com 447 amostras e o *cluster* 4 (C4) com 423 amostras. Todas as análises e procedimentos estatísticos descritos acima foram realizados no *software* R (R DEVELOPMENT CORE TEAM, 2017). A clusterização FKM foi realizada usando o pacote *ppclust* (CEBECI et al., 2020). Para facilitar o entendimento, os modelos estratificados (*clusters*) foram considerados como (MEs) e o grupo contendo todas as amostras (sem clusterização) foi considerado como modelo sem estratificação (MSE).

#### 4.2.6. Pré-processamento e construção dos modelos

Os dados espectrais do solo foram levemente suavizados (SMO) com o filtro Savitzky-Golay (com ajuste poligonal de segunda ordem e janela móvel de 11 nm). O SMO realizou uma leve suavização nas curvas espectrais, sendo

dessa forma considerado como um método de pré-processamento. Em seguida, foram aplicadas três técnicas de pré-processamento espectral frequentemente utilizadas em estudos de espectroscopia (BA et al., 2020; DOTTO et al., 2017; SILVA et al., 2019; YANG et al., 2020b) com o objetivo de testar sua capacidade de predição da textura do solo a partir dos MEs: (i) *multiplicative scatter correction* (MSC); (ii) *normalizations by range* (NBR) e (iii) *standard normal variates* (SNV). Essas técnicas de pré-processamento foram aplicadas às curvas espectrais com o objetivo de eliminar ruídos provocados pela dispersão da luz e destacar às características de maior interesse do sinal espectral. Todas as técnicas de pré-processamento foram realizadas no software R (R DEVELOPMENT CORE TEAM, 2017).

Para verificar o desempenho dos modelos estratificados (MEs) e não estratificados (MSE) na predição da textura do solo, foi utilizado o modelo *Cubist* (CBT) no pacote *cubist* (KUHN et al., 2020). O *Cubist* é um modelo que constrói árvores de decisão usando a abordagem CART. Nesse algoritmo de regressão é criada uma árvore de decisão modelo, sendo posteriormente reduzida a uma série de regras (KUHN et al., 2020). A partir dessas regras, as amostras de solos são realocadas de acordo com seus espectros e um modelo linear é aplicado para prever a variável de destino. O *Cubist* exige a configuração de dois parâmetros, que é o *committees* e o *neighbors*. O *committees* é um impulsor que utiliza dados vizinhos do conjunto de calibração (*neighbours*) e desenvolve uma série de árvores sequencialmente com pesos ajustados. Esses parâmetros foram ajustados em 1 e 100 para o impulsor *committees* e 0 e 9 para o *neighbours*.

Para testar a hipótese deste estudo, dois tipos de modelos foram gerados.

1º *Modelos iniciais*: Para estudar o efeito dos dados de entrada e da heterogeneidade no desempenho dos modelos, foram desenvolvidos dois tipos de modelos iniciais: modelos estratificados (MEs) e modelos sem estratificação (MSE). Para os MEs, 75% das amostras foram utilizadas para desenvolver o modelo de calibração e 25% foram utilizadas para validação. Dessa forma, os conjuntos de calibração dos *clusters* 1, 2, 3 e 4 continham 300, 199, 335 e 317 amostras, respectivamente (ou seja, 75% das amostras de cada *cluster*). Os conjuntos de validação dos *clusters* 1, 2, 3, 4 continham 100, 66, 112 e 106

amostras, respectivamente (ou seja, 25% do *cluster*). No total, foram utilizadas 1151 amostras para calibrar os modelos e, no total, foram preditas 384 amostras. Para construir o modelo MSE, as mesmas 1151 amostras dos MEs foram utilizadas para calibrá-lo e prever as 384 amostras dos MEs. Assim, os grupos estratificados e não estratificados são equivalentes e permitiram uma comparação justa dos índices de acurácia.

2º *Validação entre os clusters*: Para verificar o efeito da estratificação das amostras no desempenho dos modelos de calibração e validação, os modelos calibrados dos clusters com menor variabilidade de dados e melhor desempenho nos modelos iniciais foram validados com o cluster que apresentou a maior variabilidade de dados e menor desempenho nos modelos iniciais. Repetimos esse procedimento de forma inversa.

#### **4.2.7. Derivadas de ordem fracionária**

As derivadas de ordem fracionária (DOFs), assim como as derivadas de primeira e segunda ordem, são utilizadas para realçar os recursos espectrais de interesse e otimizar a extração de informações espectrais úteis. Além disso, possuem a vantagem de detectar as mais leves alterações geradas pelos grupos funcionais que compõem as propriedades do solo, pois variam em pequenos intervalos (HONG et al., 2019; ZHANG et al., 2016). Para explicar as possíveis mudanças no desempenho dos MEs, foram utilizadas as derivadas de ordem fracionária, que são comumente utilizadas em estudos de ER (HONG et al., 2019; WANG et al., 2018). Neste estudo, as derivadas fracionárias foram obtidas utilizando o pacote *prospectr* (STEVENS; RAMIREZ-LOPEZ, 2013) no software (R DEVELOPMENT CORE TEAM, 2017), onde o  $m$  variou de 0.5 a 2 com incrementos de 1 e 0.5.

#### **4.2.8. Precisão do modelo**

Para avaliar a performance dos modelos preditivos, foram calculados vários índices de acurácia que medem a precisão do modelo. Esses parâmetros incluem o coeficiente de determinação ( $R^2$ ), que varia entre 0 e 1 e fornece a porcentagem da variação que é explicada pelo modelo; a raiz do erro quadrático médio (RMSE), que mede a acurácia geral do modelo de predição; e a razão de

desempenho para o intervalo interquartil (RPIQ). Geralmente, obtém-se bom desempenho do modelo quando o RMSE é baixo com  $R^2$  e RPIQ altos.

### 4.3. RESULTADOS E DISCUSSÃO

#### 4.3.1. Análise descritiva

Para todo o conjunto de 1.535 amostras, o conteúdo de areia variou de 1 a 99%, indicando um amplo intervalo de valores, assim como a fração argila, que variou de 0 a 77%. Para a fração silte, a distribuição dos teores ficou mais próxima da média, com desvio padrão de 11%. Com a estratificação do conjunto de dados, a distribuição estatística diferiu entre os grupos. Para a areia e argila, uma menor variabilidade foi observada nos clusters C1, C2 e C3 (Tabela 8), enquanto que para o silte a menor variabilidade foi observada nos clusters C2 e C3. Para a argila, a assimetria negativa no C2 (-0,8) indica uma distribuição mais próxima ao teor máximo dessa fração, enquanto no C3, a distribuição levemente positiva (0,3) indica valores próximos ao teor mínimo dessa fração. Para a areia e o silte, a inclinação foi positiva para ambas as frações (Tabela 8). A variabilidade da textura do solo pode ser atribuída à grande distribuição espacial do conjunto de amostras, aos diversos processos de formação do solo e aos diferentes materiais de origem dos solos do Estado de Santa Catarina (EMBRAPA, 2004).

A reflectância na região do visível revelou que no conjunto de dados a cor do solo variou de 0,1 a 10 YR, sendo que o valor médio no C1 foi de 9,2 YR (amarelo-vermelho), no C2 foi de 6,5 YR (amarelo-vermelho), no C3 foi de 7,4 YR (amarelo-vermelho) e no C4 foi de 9,3 YR (amarelo-vermelho) (Tabela 8). O valor e o croma variaram de 2,3 a 6,4 e de 1,7 a 6,9, com valores médios de 5,2 e 4,6 no C1, 3,4 e 4,8 no C2, 3,3 e 4,0 no C3 e 3,9 e 3,5 no C4, respectivamente. No geral, a matiz foi baixa no C2 e C3 e o valor e o croma apresentaram variação de acordo com o *cluster*. Ou seja, a cor do solo era mais avermelhada e mais clara no C2 e C3. Resumidamente, esses quatro *clusters* forneceram conjuntos de amostras com padrões bem definidos de uma área com ampla cobertura espacial.

Tabela 9. Estatística das frações granulométricas do solo a partir dos clusters adquiridos pela análise de estratificação.

Clusters	Nº amost.	Média ± DP	Máx.	Min.	Ske.	Cor (média)			(Hm/Hm+Gt)	Classif. óxidos ferro
						Matiz	Valor	Croma		
Argila (%)										
C1	400	29,2 ± 12,0	64,0	0,0	0,2	9,2YR	5,2	4,6	8	Gt
C2	265	50,2 ± 14,9	77,0	2,0	-0,8	6,5YR	3,4	4,8	51	HG
C3	447	42,4 ± 16,0	77,0	6,7	0,3	7,4YR	3,3	4,0	45	GH
C4	423	35,0 ± 18,5	74,0	0,0	0,0	9,3YR	3,9	3,5	9	Gt
Todas amostras	1535	38,3 ± 17,3	77,0	0,0	0,1	8,1YR	3,9	4,1	27	GH
Areia (%)										
C1	400	36,2 ± 18,7	99,0	1,0	0,3	9,2YR	5,2	4,6	8	Gt
C2	265	18,6 ± 12,2	86,0	2,0	1,6	6,5YR	3,4	4,8	51	HG
C3	447	22,1 ± 11,8	67,0	3,0	0,6	7,4YR	3,3	4,0	45	GH
C4	423	35,4 ± 24,5	97,0	2,0	0,8	9,3YR	3,9	3,5	9	Gt
Todas amostras	1535	28,8 ± 19,5	99,0	1,0	1,1	8,1YR	3,9	4,1	27	GH
Silte (%)										
C1	400	34,4 ± 13,6	83,0	0,0	0,5	9,2YR	5,2	4,6	8	Gt
C2	265	31,1 ± 6,7	59,0	12,0	0,6	6,5YR	3,4	4,8	51	HG
C3	447	35,4 ± 8,4	67,0	15,0	0,3	7,4YR	3,3	4,0	45	GH
C4	423	29,5 ± 12,9	78,0	1,0	0,3	9,3YR	3,9	3,5	9	Gt
Todas amostras	1535	32,8 ± 11,3	83,0	0,0	0,3	8,1YR	3,9	4,1	27	GH

Relação Hematita/Goethita foi calculada a partir das equações e bandas de absorção propostas por SHERMAN; WAITE (1985) e BAHIA et al. (2014). \*Gt: solos goethíticos = 0-25%; GH: Goethíticos-Hematíticos = >25-50%; HG: Hematíticos-Goethíticos = >50-75%; Hm: Hematíticos = >75-100%.

#### 4.3.2. Análise de componentes principais (ACP)

De acordo com os resultados do ACP, no primeiro componente principal (CP1), que representa 59% da variação total, os parâmetros de cor Valor, X, Y, L, R, G, K e  $v^*$  apresentaram as porcentagens mais elevadas de contribuição para este eixo, com cada parâmetro explicando 7%, em média (Figura 8a). Os parâmetros com maior contribuição positiva foram o Valor, X, Y, L, R, G e  $v$  (com 0,9), e contribuição negativa (sentido inverso), o parâmetro K (-0,9) (Figura 8c). O Valor no Sistema Munsell, assim como os parâmetros de luminosidade (L) e o  $v^*$  do CIE Luv, dizem respeito à quantidade de luz que uma determinada superfície pode emitir ou refletir. Solos com predomínio de minerais de quartzo (mais claros) e material orgânico (mais escuro) podem ser detectados por seu comportamento natural de refletir e absorver a luz, respectivamente

(VISCARRA ROSSEL et al., 2006a). A segunda componente principal (CP2) explicou 31% da variação total e está associada a uma forte influência dos parâmetros Croma, CIE  $a^*$ ,  $x^*$ ,  $u^*$  e M, onde cada parâmetro explicou, em média, 12% da variabilidade das amostras (Figura 8b). Os parâmetros com maior contribuição positiva foram o Croma,  $a$ ,  $x$  e  $u$  (com 0,9), e a contribuição negativa, o parâmetro  $h$  (-0,6) (Figura 8d). O valor  $a^*$  dos modelos CIE Lab está relacionado a cor vermelha dos solos, que pode estar sendo influenciado pelos óxidos de Fe e indicando predominância de hematita. Dessa forma, os parâmetros de cor baseados em Vis podem ser aplicados para determinar a cor, a mineralogia e auxiliar na escolha da cor predominante do solo (DOMINGUEZ et al., 2010; VISCARRA ROSSEL; WALTER; FOUAD, 2003). Para SIMON et al. (2020), a predição da cor do solo de forma rápida e com melhor relação custo-benefício é uma importante ferramenta para o planejamento de práticas de manejo do solo, bem como, para a prospecção e abertura de novas áreas.

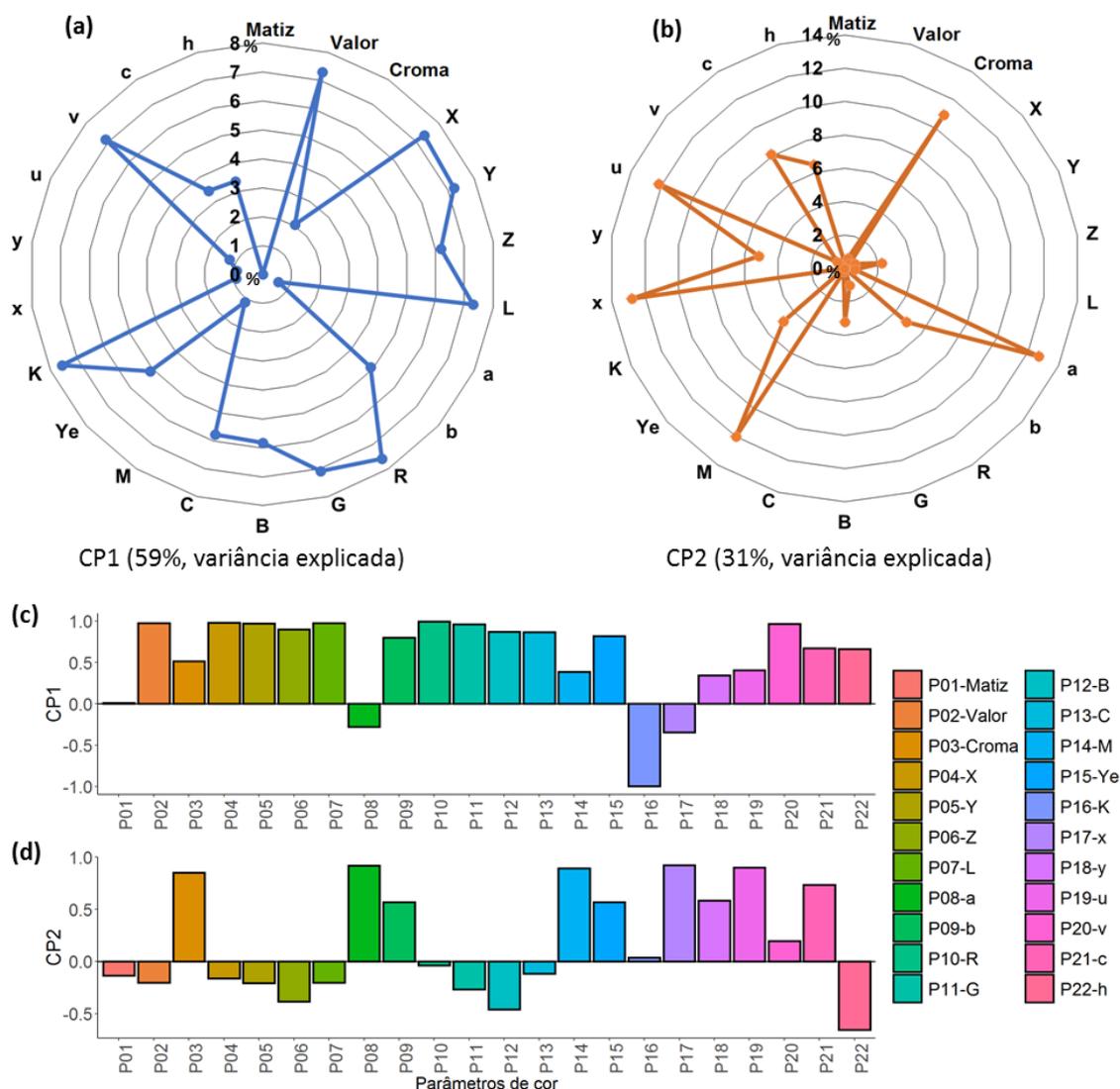


Figura 8. Resultado da análise de componentes principais dos parâmetros que melhor explicaram a variação das informações do solo com base na cor do solo (a, b). Contribuição (*loadings*) dos parâmetros nas duas primeiras CPs (c, d).

#### 4.3.4. Análise das curvas espectrais

Na interpretação da curva espectral é preciso observar a intensidade de reflectância, a forma da curva ao longo do espectro (por exemplo, ascendente, descendente ou plana) e às características de absorção, que geralmente são influenciadas pela composição físico-química e mineralógica dos solos (Figura 9). Os *clusters* 1 e 4 apresentaram curvas ascendente com alta intensidade de reflectância, esse comportamento ascendente é explicado pelo maior conteúdo de areia. Na fração areia, o quartzo é o mineral dominante e não tem características espectrais de absorção na região do VIS-NIR-SWIR,

aumentando a reflectância em todo o espectro (DEMATTÊ et al., 2007; WIGHT; ASHWORTH; ALLEN, 2016). Enquanto que o C2 e o C3 apresentaram curvas planas com intensidades médias próximas de 0.2 que, em geral, indicam que são amostras com alto teor de óxidos de ferro e argila (DALMOLIN et al., 2005). Além disso, nas regiões entre 480, 550, 850 e 900 nm observou-se uma concavidade indicativa da presença de óxidos de ferro (DALMOLIN et al., 2005; DEMATTÊ et al., 2014; MOURA-BUENO et al., 2019). Assim, a partir da interpretação das curvas espectrais é possível inferir sobre algumas características dos solos e indicar a análise quantitativa mais apropriada, reduzindo custos e tempo nos laboratórios.

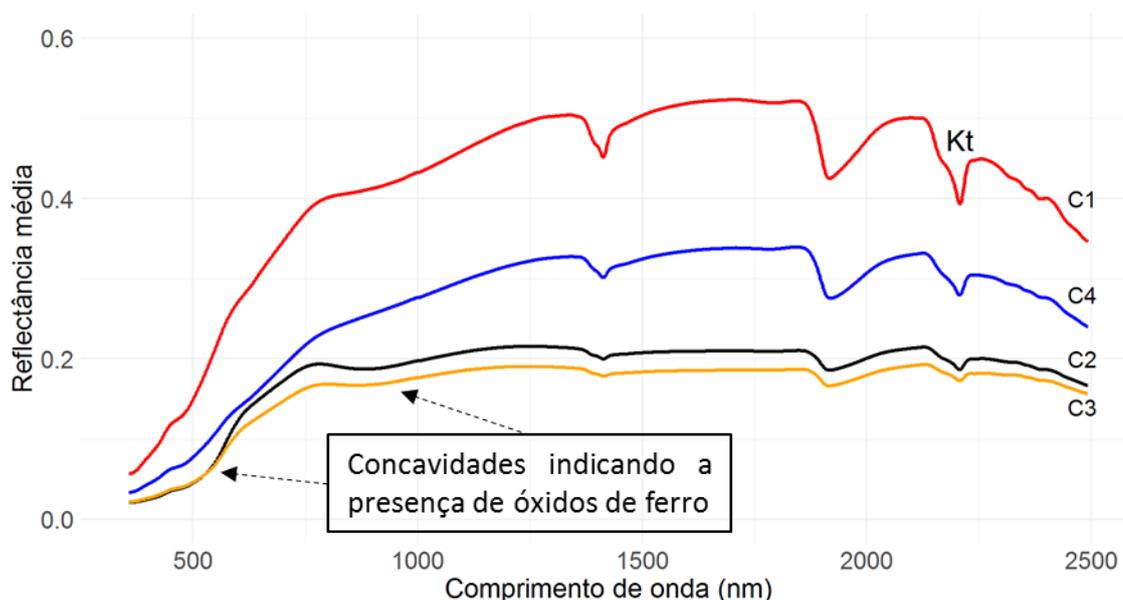


Figura 9. Espectros médios de reflectância dos *clusters*, destacando bandas espectrais relacionadas aos componentes do solo e as alterações na intensidade da reflectância.

#### 4.3.5. Derivadas fracionárias dos conjuntos estratificados

Os espectros de reflectância de diferentes ordens fracionárias são mostrados na Figura 10. A ordem = 0.5 representa o espectro com pouca alteração em relação ao original. Na ordem = 0.5, para o C1 e o C4, há três feições de absorção bem definidas em ~1400, ~1900 e ~2200. A primeira e a segunda feição referem-se às vibrações de estiramento dos grupos OH<sup>-</sup> e H<sub>2</sub>O dos minerais de argila, que são visualmente identificados nessas regiões

(DOTTO et al., 2018; GHOLIZADEH et al., 2016; NOCITA et al., 2014). Na terceira feição, às características de absorção são atribuídas aos argilominerais, principalmente caulinita e illita (CAMARGO et al., 2018; ZHAO et al., 2018). Quando a ordem fracionária aumentou de 0.5 para 1.5, as três características de absorção citadas anteriormente ficaram mais evidentes. Além disso, as características de absorção na região do visível (bandas 480 nm e 550 nm) ficaram bem definidas, devido aos óxidos de ferro (Figura 10), identificando a presença de goethita e hematita (DALMOLIN et al., 2005; MOURA-BUENO et al., 2019). Para o C1 e o C4 a absorção na região do 480 nm é mais evidente, indicando a presença de maior conteúdo de goethita, enquanto o C2 e o C3 indicam maior conteúdo de hematita (conforme mostrado na Tabela 8), com absorção mais evidente nas bandas de 550 nm. Na região de 1400 nm houve mudanças na absorção e a feição espectral transformou-se suavemente para uma forma bipolar (um pico positivo e um pico negativo) (ZHANG et al., 2020), observados com maior intensidade no C1 e C4.

Quando a ordem aumentou de 1.5 para 2, na região do visível, todos os *clusters* apresentam dois picos positivos em 420 e 490 nm (causados por goethita e hematita, respectivamente), no entanto, o C2 e C3 apresentam um pico positivo na região de 523 nm confirmando que esses *clusters* apresentam solos com maior relação de hematita/goethita (Tabela 8) que os demais (DOTTO et al., 2018; FANG et al., 2018).

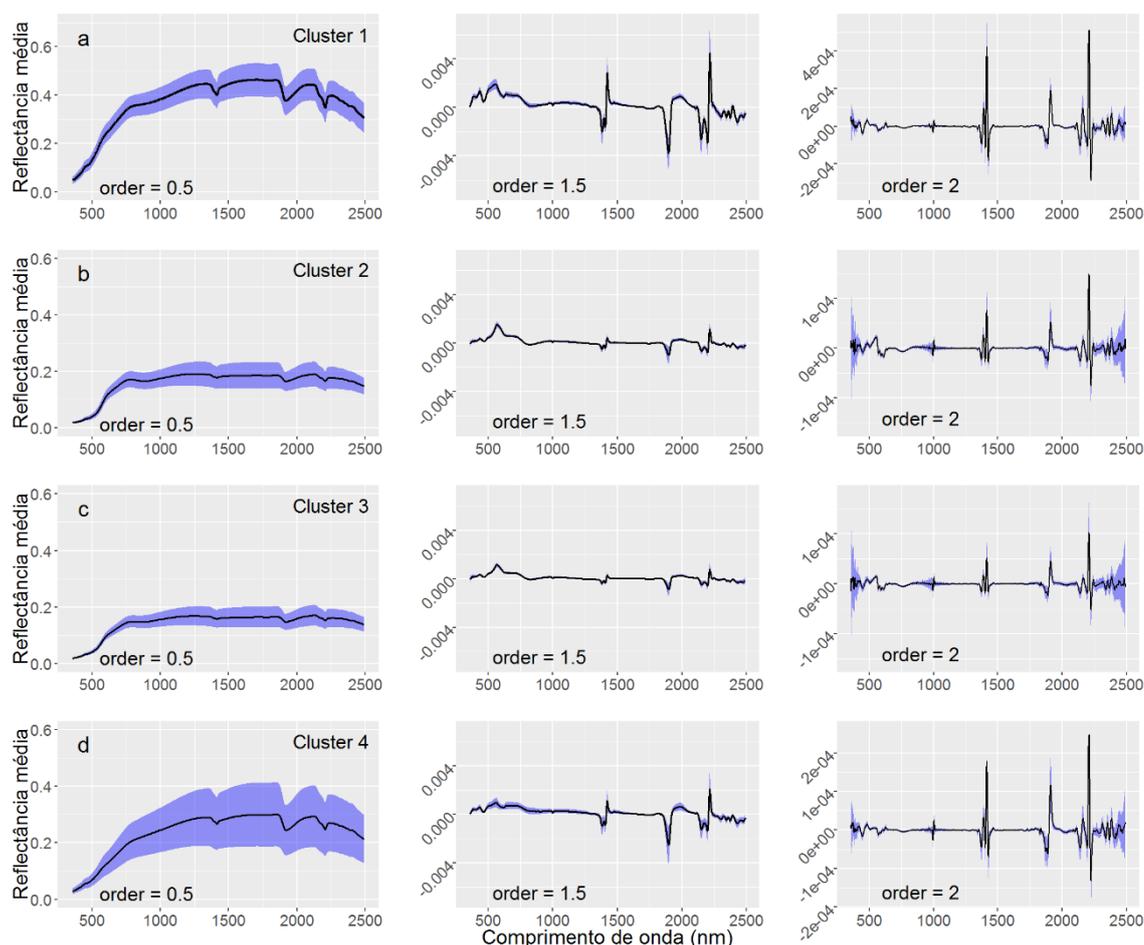


Figura 10. Curvas espectrais dos diferentes clusters e derivadas fracionárias: C1 com derivadas fracionárias de 0.5, 1.5 e 2 (a); C2 com derivadas fracionárias de 0.5, 1.5 e 2 (b); C3 com derivadas fracionárias de 0.5, 1.5 e 2 (c); C4 com derivadas fracionárias de 0.5, 1.5 e 2 (d). As áreas em azul representam os desvios padrão das curvas espectrais.

#### 4.3.6. Modelos iniciais (estratificados e não estratificados)

Com base nos parâmetros de cor, a biblioteca espectral foi estratificada em diferentes clusters (conforme mostrado na Tabela 8). Os resultados de validação do modelo *Cubist* indicaram que a estrutura dos dados de entrada gerou diferenças no desempenho da predição das frações de argila, areia e silte (Tabela 9). Em relação aos MEs para a argila, os resultados mais precisos foram obtidos para o C2 ( $R^2 = 0,84$  e  $RMSE = 5,7\%$ ) e C3 ( $R^2 = 0,84$  e  $RMSE = 6,1\%$ ). O melhor desempenho do C2 e C3 na predição do teor de argila está associado ao maior conteúdo dessa fração nestes *clusters*. Na fração argila, às características de absorção espectral, especialmente na faixa Vis (480 e 550 nm) estão relacionadas aos óxidos de ferro (BAHIA et al., 2015; RAMOS et al.,

2020) e na faixa SWIR (1400 e 1900 nm) devido às vibrações de estiramento de OH<sup>-</sup> e H<sub>2</sub>O mantida nas estruturas dos argilominerais (BEN-DOR et al., 2008; GHOLIZADEH et al., 2016; STENBERG et al., 2010). Por outro lado, o C2 e o C3 apresentaram a menor variabilidade do teor de areia, o que pode ter contribuído para o melhor desempenho dos modelos. Em termos dos valores de RMSE, os resultados para a fração areia indicaram que os modelos do C2 (4,9%) e C3 (4,5%) obtiveram melhor capacidade preditiva em comparação aos demais. Esses *clusters* apresentaram a menor variabilidade e o menor conteúdo da respectiva fração (Tabela 8). Para o silte, os resultados do C3 foram semelhantes para os quatro pré-processamentos ( $R^2 = 0,58$ , RMSE = 5,7 e o RPIQ = 19), com os demais modelos apresentando baixo desempenho, tanto para os MEs quando para o MSE, seguindo a mesma performance obtida em outros estudos (DOTTO et al., 2017; PINHEIRO et al., 2017; VASAVA et al., 2019; ZHANG et al., 2017).

Em relação ao tipo de pré-processamento, muitos estudos obtiveram melhor desempenho dos modelos preditivos quando utilizaram algum pré-processamento nos dados espectrais na estimativa da textura do solo (DEMATTE et al., 2019; DOTTO et al., 2017; JACONI; VOS; DON, 2019), enquanto COBLINSKI et al. (2020) obtiveram melhores resultados com dados espectrais não transformados. No presente estudo, os melhores resultados de desempenho dos modelos para a argila foram obtidos com os pré-processamentos SMO e MSC (Tabela 9). Para a areia, foram os pré-processamentos SMO, SNV e MSC que alcançaram os melhores desempenhos.

Tabela 10. Resultados de validação dos modelos estratificados e não estratificados para a textura do solo (%) usando diferentes tipos de pré-processamentos.

Clusters	Pré-Tratamento	Validação			Validação			Validação		
		R <sup>2</sup>	RMSE (%)	RPIQ	R <sup>2</sup>	RMSE (%)	RPIQ	R <sup>2</sup>	RMSE (%)	RPIQ
		Argila			Areia			Silte		
Cluster 1	SMO	0,70	7,0	2,45	0,76	11,0	2,83	0,51	8,9	2,17
	NBR	0,67	7,1	2,29	0,70	11,5	2,34	0,47	9,2	2,10
	MSC	0,67	7,1	2,29	0,71	11,5	2,25	0,42	9,6	2,00
	SNV	0,67	7,1	2,28	0,67	11,6	2,24	0,51	8,9	2,16
Cluster 2	SMO	0,84	5,7	2,85	0,78	4,9	2,20	0,35	5,3	1,90
	NBR	0,80	6,6	2,43	0,85	5,3	3,11	0,28	5,4	1,84

Tabela 10. Continuação.

	MSC	0,76	6,6	2,41	0,87	5,1	3,26	0,37	5,2	1,93
	SNV	0,77	6,6	2,48	0,88	4,9	3,40	0,36	5,2	1,93
Cluster 3	SMO	0,84	6,6	3,55	0,80	5,4	3,04	0,58	5,8	1,95
	NBR	0,83	6,3	3,09	0,80	5,4	3,07	0,58	5,7	1,96
	MSC	0,84	6,1	3,17	0,86	4,5	3,63	0,58	5,7	1,96
	SNV	0,83	6,2	3,13	0,85	4,7	3,51	0,58	5,7	1,96
Cluster 4	SMO	0,83	7,5	3,73	0,78	12,1	2,86	0,42	8,7	1,47
	NBR	0,81	8,3	3,21	0,75	12,9	2,28	0,36	9,3	1,37
	MSC	0,79	8,2	3,40	0,74	13,1	2,25	0,43	8,6	1,49
	SNV	0,80	8,0	3,48	0,74	12,9	2,29	0,29	9,6	1,33
Todas	SMO	0,82	7,4	3,66	0,72	10,7	2,30	0,37	9,3	1,50
	NBR	0,82	7,2	3,47	0,69	11,3	2,00	0,29	10,1	1,38
	MSC	0,83	7,1	3,52	0,71	11,1	2,10	0,32	9,8	1,42
	SNV	0,80	7,7	3,26	0,71	11,0	2,53	0,28	10,2	1,37

<sup>a</sup>SMO: *smoothing*; NBR: *normalization by range*; MSC: *multiplicative scatter-correction*; SNV: *standard normal variate*.

Os resultados obtidos para os MEs (Tabela 9) mostraram que às características do solo tiveram maior efeito sobre a capacidade preditiva dos modelos do que o tipo de pré-processamento espectral, confirmando o potencial da estratégia da estratificação em fornecer conjuntos mais homogêneos na tentativa de obter bons modelos de predição. De acordo com os resultados, as precisões dos MEs (com valor médio de RMSE de 6,9% para a argila e de 8,6% para a areia) foram comparativamente melhores que a maioria dos estudos publicados anteriormente, onde os valores de RMSE variaram de 1,9 a 18% para argila e 3,4 a 24% para a areia (Tabela 2 – Capítulo II). ZHANG et al. (2017), em escala local, com 257 amostras, aplicaram o algoritmo *cubist* em um conjunto de dados com teor de argila variando de 3 a 88% e obtiveram  $R^2$  de 0,70 e RMSE de 14,7% para validação. Da mesma forma, DOTTO et al. (2017), em escala local, com 299 amostras de solos e teor de argila variando de 21 a 78% e areia variando de 1 a 35%, encontram  $R^2$  de 0,62 e RMSE de 6,8% para argila e  $R^2$  de 0,25 e RMSE de 6,4% para areia. Mais recentemente, DEMATTÊ et al. (2019), em escala nacional, utilizaram 39284 amostras na predição de propriedades do solo representativas de todos os Estados do Brasil. Nesse estudo o teor de argila variou de 0 a 98,7% e a areia variou de 0 a 99%. Os autores encontram  $R^2$  de 0,88 e RMSE de 7,6% para argila e  $R^2$  de 0,87 e RMSE de 10,3% para areia. No entanto, os resultados de predição obtidos por

COBLINSKI et al. (2020), em escala local, com 197 amostras e com  $R^2$  de 0,89 e RMSE de 5,1% para argila e  $R^2$  de 0,81 e RMSE de 6,5% para a areia, foram superiores aos encontrados no presente estudo. Uma explicação para o melhor desempenho dos modelos nesse estudo pode ser devido ao menor número de amostras utilizadas em comparação ao presente estudo e a homogeneidade da base de dados (biblioteca espectral local).

Em relação ao desempenho do C4, observou-se valores mais altos de RMSE em relação aos demais (7,5% a 8,3% para a argila e de 12,1% a 13,1% para a areia) (Tabela 9). Isso pode ser atribuído à amplitude dos dados em C4, mesmo após a estratificação. Esses resultados são consistentes com os descritos na literatura, nos quais conjuntos mais homogêneos de solos geralmente fornecem melhores desempenhos dos modelos de precisão (ARAÚJO et al., 2014; JIANG et al., 2017; MOURA-BUENO et al., 2019). JIANG et al. (2017) coletaram amostras de solos em camadas superficiais (0 – 0,1 m) e subsuperficiais (0,1 – 0,3 m) na região central da China e identificaram diferenças na variabilidade dos teores de COS para cada profundidade do solo, o que afetou o desempenho dos modelos de predição. ARAÚJO et al. (2014) separaram os dados de uma biblioteca espectral nacional na tentativa de melhorar as precisões na predição da fração argila. Os autores concluíram que a divisão da biblioteca em subconjuntos melhorou as estimativas com redução de 21% do erro de predição para essa fração. A Figura 11 mostra os resultados da validação para os modelos de areia e argila que apresentaram o melhor desempenho.

Com o agrupamento prévio das amostras de solos, houve redução da variabilidade espectral e composicional das amostras, principalmente para o C2 e C3, que foram os grupos que mostraram menor variação nos teores das frações texturais e nos dados espectrais. Além disso, é importante observar que o C1 também apresentou baixa variabilidade para o teor de argila, que pode ser confirmada pelo DP (Tabela 8). No entanto, o desempenho dos modelos para esse grupo foi inferior aos demais, podendo ser explicado pelo menor conteúdo de argila para esse *cluster*, conforme esclarecido anteriormente. Portanto, dois fatores devem ser considerados na análise de desempenho dos modelos quando construídos a partir da estratificação de bibliotecas espectrais, a variabilidade dos dados e o conteúdo das frações que integram essas bibliotecas.

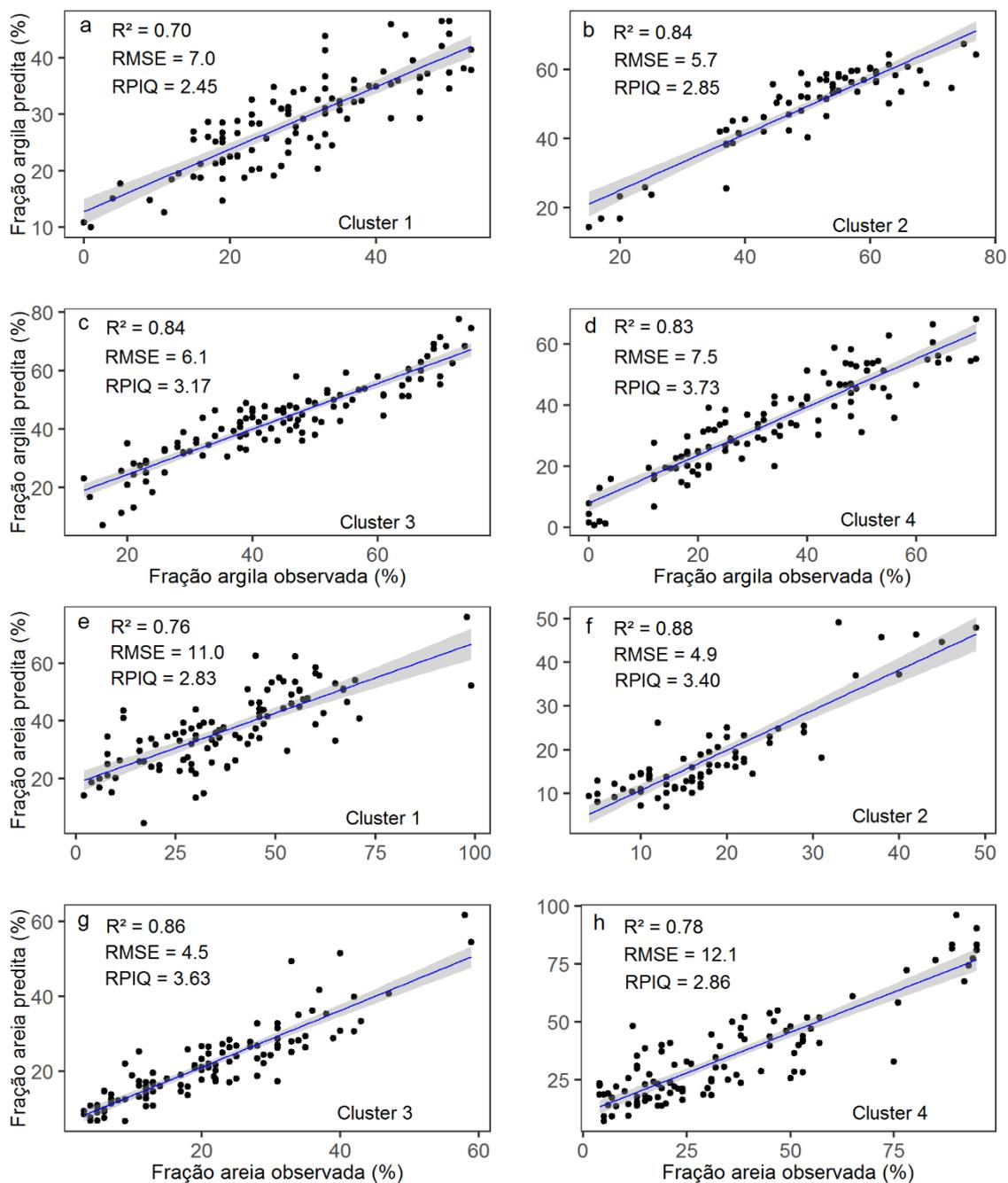


Figura 11. Validação de dados medidos em laboratório em relação às predições Vis-NIR-SWIR para a argila (a, b, c, d) com os *clusters* C1, C2, C3 e C4, respectivamente, e para a areia (e, f, g, h) com os *clusters* C1, C2, C3 e C4, respectivamente.

#### 4.3.7. Validação entre grupos

Para verificar a aplicabilidade dos modelos iniciais, os *clusters* com menor desvio padrão foram utilizados para estimar a areia e a argila dos *clusters* com maior desvio padrão. As frações argila e areia foram consideradas nesta

análise em função do melhor desempenho dos modelos obtidos na etapa anterior. Para tanto, os modelos não foram calibrados novamente, utilizou-se o mesmo conjunto de calibração e validação da etapa anterior. Para a fração areia, foram utilizados os *clusters* C3 e C4, que apresentaram o menor e o maior DP, respectivamente. Para a fração argila, foram utilizados os *clusters* C1 e C4, que apresentaram o menor e o maior DP, respectivamente. Quando comparado com os resultados de validação dos modelos iniciais estratificados, os modelos de calibração mostraram baixa qualidade de predição das frações areia e argila para os diferentes *clusters* (Figura 12). Resultado semelhante foi obtido no C3, onde o RMSE aumentou de 5,4% para 10,7%, o  $R^2$  diminuiu 0,80 para 0,26 e o RPIQ diminuiu de 3,04 para 2,01 para a predição da fração areia. Além disso, a predição do C4 quando calibrado com o C3 revelou baixo desempenho dos modelos (o RMSE aumentou de 12,1% para 14,8% e o  $R^2$  diminuiu de 0,78 para 0,20). Para a fração argila, os resultados da comparação da validação entre os *clusters* revelaram baixa precisão dos modelos, sendo observado, principalmente, em relação ao C1 que teve um aumento no valor de RMSE, passando de 7,0% para 8,4%, assim como uma redução no valor de  $R^2$  de 0,70 para 0,57.

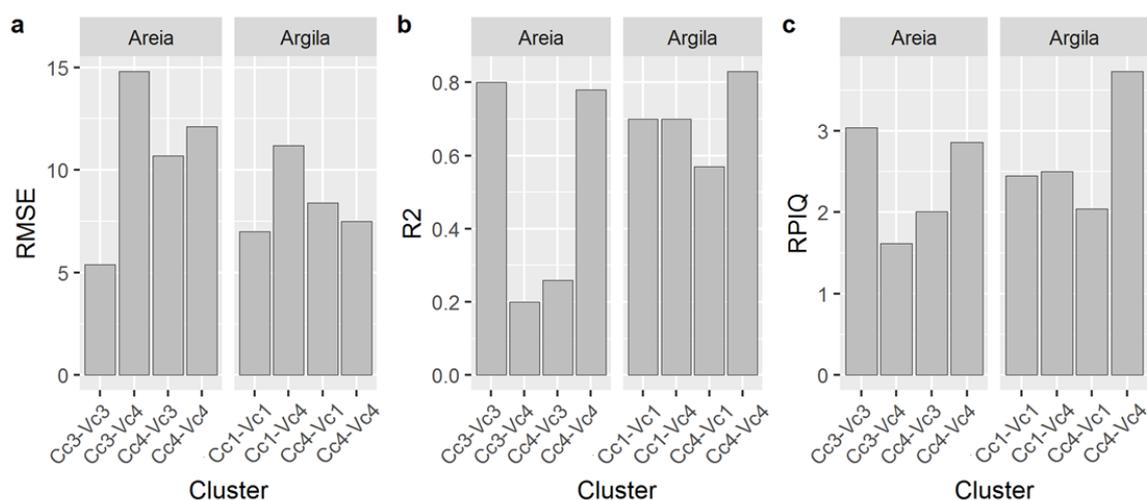


Figura 12. Comparação do RMSE (a),  $R^2$  (b) e RPIQ (c) dos modelos iniciais na predição da areia e argila para a validação de sua aplicabilidade entre os clusters com maior e menor DP. (Cc1 = calibração usando o cluster 1; Cc3 = calibração usando o cluster 3; Cc4 = calibração usando o cluster 4; Vc1 = predição usando o cluster 1; Vc3 = predição usando o cluster 3; Vc4 = predição usando o cluster 4).

Essa etapa adicional de calibração entre os diferentes *clusters* no processo de validação foi necessária para confirmar que a estratificação é de fato capaz de melhorar a predição na estimativa da textura do solo. Além disso, os modelos têm baixo desempenho quando utilizados em um grupo de amostras mais heterogêneo ou com maior DP. Esses resultados corroboram com JIANG et al. (2017), que observaram que os modelos de validação cruzada apresentaram baixo desempenho quando amostras da superfície com maior DP foram utilizadas na estimativa do teor de COS da subsuperfície.

Do ponto de vista analítico e agronômico, a estratificação da biblioteca espectral pela cor do solo pode ser usada nos laboratórios comerciais de análise de solo como um método complementar, direcionando às análises, poupando tempo e dinheiro.

#### 4.4. CONCLUSÃO

A estratificação da biblioteca espectral de escala maior melhorou a capacidade preditiva dos modelos na estimativa dos teores das frações texturais do solo, dada a influência da variabilidade dos constituintes dos solos que compõem essas bibliotecas e que interferem no comportamento espectral. Portanto, a estrutura dos dados de entrada deve ser levada em consideração na construção de modelos de predição mais precisos.

Considerando a predição a partir dos modelos estratificados, dois fatores tiveram maior influência nos resultados, a variabilidade dos solos e o teor das frações argila, areia e silte. O menor desvio padrão e o maior conteúdo de argila apresentados no C2 e C3 levaram a uma boa estimativa dessa fração, com redução do erro de 5%, em média, em relação ao modelo sem estratificação. Enquanto para a areia, a menor variabilidade no C2 e C3 levaram a uma melhor capacidade de predição dessa fração, com redução do erro de 22%, em média.

Os modelos estratificados com maior conteúdo de argila e óxidos de ferro obtiveram os melhores índices de acurácia e as bandas espectrais de maior importância foram observadas, principalmente, nas faixas de 480-550, 850-900 e 1400, 1900 e 2200 nm.

A validação entre *clusters* revelou um baixo desempenho quando os modelos iniciais foram calibrados e validados com conjuntos de dados

apresentando maior desvio padrão, confirmando a hipótese inicial e elucidando a influência de dados muito heterogêneos na precisão dos modelos. Portanto, a estratificação das bibliotecas espectrais de escala maior é uma boa estratégia para melhorar as avaliações regionais dos recursos do solo, reduzindo o erro de predição na determinação qualitativa das propriedades do solo.

## 5. CAPÍTULO IV – ESTUDO 3: PREDIÇÃO DE PROPRIEDADES DO SOLO USANDO ESPECTROSCOPIA Vis-NIR-SWIR E VARIÁVEIS AMBIENTAIS

### 5.1. INTRODUÇÃO

A espectroscopia Vis-NIR-SWIR provou gradualmente seu potencial para estimar várias propriedades do solo (ALLO et al., 2020; DEMATTÊ et al., 2019; HUTENGS et al., 2019; VISCARRA ROSSEL et al., 2016). Essa técnica requer uma pequena quantidade de amostras para a medição, apresentando boa relação custo-benefício, sem uso de reagentes químicos e a possibilidade de reaproveitar as amostras após a análise, sendo, portanto, uma alternativa às análises químicas laboratoriais (BA et al., 2020; NOCITA et al., 2015; VISCARRA ROSSEL et al., 2014; YANG et al., 2020b; ZHANG et al., 2020).

A textura do solo é umas das propriedades do solo que pode ser predita pela espectroscopia (DUDA et al., 2017; JACONI; VOS; DON, 2019; MARGENOT et al., 2020; SILVA et al., 2019; VASAVA et al., 2019). O estudo de métodos alternativos para a análise da textura é justificado pela importância que essa propriedade desempenha no uso e manejo dos solos, como na capacidade de infiltração e disponibilidade de água, na absorção de nutrientes e desenvolvimento das plantas, além de influenciar outras propriedades físicas do solo, como a densidade, a estabilidade de agregados e a porosidade, entre outros (KLEIN et al., 2010; TÜMSAVAŞ et al., 2019). Outra importante propriedade do solo que também pode ser predita pela espectroscopia é o teor de COS, que tem grande importância no monitoramento ambiental, além de ser um parâmetro essencial no controle da dinâmica de vários agroquímicos no solo (GHOLIZADEH et al., 2018; STEVENS et al., 2013).

Na determinação qualitativa dessas propriedades do solo a partir das assinaturas espectrais, são utilizados modelos robustos de calibração, como a regressão por mínimos quadrados parciais (*partial-least-squares regression* -

PLSR) (WOLD; SJÖSTRÖM; ERIKSSON, 2001), a máquina de vetores de suporte (*support vector machines* - SVM) (CORTES; VAPNIK, 1995), a floresta aleatória (*random forest* - RF) (BREIMAN, 2001) e o modelo de regressão gaussiana (*gaussian process regression* - GPR) (WILLIAMS; BARBER, 1998), entre outros. O modelo PLSR tem sido amplamente utilizado nesse tipo de estudo (ALLO et al., 2020; HUTENGS et al., 2019), onde captura a relação linear entre determinada propriedade do solo (variável dependente) e os espectros de reflectância correspondentes a essa propriedade (variável independente), diferindo de outros modelos como SMV, RF, GPR, que capturam as relações não lineares (GUPTA; ISLAM; HASAN, 2018).

Em relação aos dados de entrada dos modelos, têm sido poucos os trabalhos que utilizam os espectros de reflectância combinados com outra variável preditora na predição das propriedades do solo. No entanto, os modelos de calibração suportam a utilização de outras variáveis que podem ser incorporadas à regressão multivariada e trazer melhores resultados a modelagem. BROWN et al. (2006) aplicaram a espectroscopia de reflectância (ER) junto com variáveis auxiliares, incluindo a fração areia e o pH do solo como preditores. Ao combinar essas duas variáveis com as assinaturas espectrais, eles obtiveram melhores resultados de predição de propriedades do solo como a fração argila, o teor de óxidos de ferro, a capacidade de troca de cátions, etc.

STEVENS et al. (2013) adicionaram novas preditoras à matriz espectral na tentativa de melhorar os modelos de predição do teor de COS, testando as frações areia e argila como preditoras auxiliares, que segundo eles são parâmetros prontamente disponíveis e que sofrem poucas mudanças ao longo tempo e, portanto, podem ser diretamente explorados em modelos preditivos. Nesse estudo, a utilização da fração areia integrada à matriz espectral melhorou os modelos de solos de pastagens e florestas em comparação com os modelos usando apenas os dados espectrais. PENG et al. (2015) combinaram os espectros Vis-NIR-SWIR, os índices espectrais derivados de imagens de satélite e variáveis ambientais auxiliares (por exemplo a elevação, o mapa de solos, a geologia, o uso da terra, etc.) na predição do teor de COS e, empregando o método multivariado *Cubist*, melhoraram a acurácia de predição dos modelos para essas propriedades do solo.

Aparentemente, os poucos estudos que tentaram melhorar a precisão dos modelos espectroscópicos tratam da inclusão de outras propriedades do solo (por exemplo, as frações areia, argila e pH) à matriz espectral. No entanto, assim como teor de COS, essas propriedades também necessitam da sua determinação em laboratório e usá-las como variáveis preditoras encarecem a predição. Dessa forma, uma alternativa é a utilização de variáveis que já possam estar disponíveis, como as variáveis ambientais (VAs) obtidas dos modelos digitais do terreno (MDT), que podem ser empregadas para caracterizar o ambiente nas mais variadas escalas (COELHO, 2010; GIASSON et al., 2013; TEN CATEN et al., 2012).

A partir dos MDTs podem ser derivadas VAs que caracterizam o relevo e podem auxiliar na predição das propriedades do solo. O emprego dessas VAs nos modelos espectroscópicos pode representar determinações mais seguras e confiáveis, já que pesquisas demonstraram forte correlação entre atributos do solo e atributos do relevo. Por exemplo, a textura do solo apresentou forte correlação com a elevação, a curvatura planar e a curvatura do perfil (COX et al., 2003; KOKULAN et al., 2018; LIESS; GLASER; HUWE, 2012). Outros estudos encontraram relação moderada entre o teor de COS e a elevação (ADHIKARI et al., 2014; KOKULAN et al., 2018), a declividade e o índice de umidade topográfica (ADHIKARI et al., 2014; FLORINSKY et al., 2002).

O objetivo deste estudo foi combinar as VAs com às curvas espectrais Vis-NIR-SWIR para melhorar as estimativas do teor de COS e das frações areia, argila e silte. Para esse propósito, foram testados três cenários diferentes de predição: a) a predição das propriedades do solo utilizando apenas as assinaturas espectrais na região do Vis-NIR-SWIR; b) a predição das propriedades utilizando apenas as VAs; c) a predição das propriedades do solo utilizando a combinação das leituras espectrais na região do Vis-NIR-SWIR e as VAs.

## **5.2. MATERIAL E MÉTODOS**

### **5.2.1. Área de estudo e coleta de dados**

A área de estudo está localizada no Parque Nacional de São Joaquim (PNSJ), região do planalto sul do Estado de Santa Catarina, Brasil (Figura 1 –

Capítulo II). Essa Unidade de Conservação (UC) apresenta uma área aproximada de 49.300 ha, com relevo bastante acidentado e altitudes que variam entre 300 a 1.826 m. Os solos são formados predominantemente a partir de rochas vulcânicas e sedimentares da Bacia do Paraná do Grupo Serra Geral (Formação Serra Geral, Palmas, Paranapanema, Gramado e Botucatu) e do Grupo Passa Dois (Formação Rio do Rastro e Teresina) (EMBRAPA, 2004; WILDNER et al., 2014). As rochas vulcânicas da Formação Serra Geral são dominantes e caracterizadas por basaltos e andesitos (básicas) e riolitos e riodacitos (ácidas). Os solos dessa região apresentam, em geral, horizonte A húmico ou proeminente, com teores muito altos de Al trocável e alta acidez potencial, baixa soma de bases e saturação por base, relevo forte ondulado e ondulado, um tanto rasos pertencendo às classes dos Neossolos Litólicos e Regolíticos e dos Cambissolos Húmicos e Háplicos (DALMOLIN et al., 2017; EMBRAPA, 2004).

Nessa área, foram instalados pela Epagri módulos de pesquisa do Programa de Pesquisa em Biodiversidade, Mata Atlântica, Santa Catarina (PPBio-MA-SC). A área do módulo tem 500 ha (1 × 5 km) e é composta por dois acessos no sentido Leste-Oeste. No interior do módulo 1 foram instaladas 20 parcelas: 10 terrestres (Fig. 1) e 10 ripárias. As parcelas são corredores de 250 m que seguem as curvas de nível e são demarcadas por 6 piquetes, um a cada 50 m (0, 50, 100, 150, 200 e 250 m). Um total de 260 amostras de solo foram coletadas em novembro de 2016, usando dois sistemas de amostragem. Com o primeiro sistema de amostragem foram coletadas 145 amostras de solos (nas parcelas terrestres), usando o protocolo do PPBio, com adaptação na profundidade de coleta segundo o GlobalSoilMap.Net (0-10, 10-20, 20-30 e 30-60 cm de profundidade). Com o segundo sistema de amostragem foram coletadas 115 amostras de solos (fora das parcelas, em diferentes locais do módulo), usando uma abordagem do *conditioned Latin Hypercube Sampling* (cLHS) (MINASNY; MCBRATNEY, 2006). A amostragem por cLHS utilizou um conjunto de VAs contínuas ou categóricas que representam o "espaço preditor" no qual foram selecionados os locais mais representativos da paisagem para a coleta das amostras.

### 5.2.2. Medidas de reflectância espectral e análise laboratorial

Os dados de reflectância do solo foram obtidos por espectroscopia Vis-NIR-SWIR, utilizando um espectroradiômetro *FieldSpec 3 (Analytical Spectral Devices, Boulder, EUA)* com o conjunto *ASD Contact Probe*, que abrange a faixa espectral entre 350 a 2500 nm e resolução de 1 nm. Inicialmente as amostras foram secas, moídas e peneiradas (malha 2 mm) para obtenção da terra fina seca ao ar (TFSA). Para realizar as medidas espectrais, as amostras de TFSA foram distribuídas homogeneamente em um porta amostras. O sensor espectral utilizado capturou a luz através de um cabo de fibra ótica acoplado no *Contact Probe*, que ficou em contato com a superfície da amostra. O sensor foi calibrado no início das medições espectrais e a cada 20 amostras de solo, usando uma placa branca *Spectralon (Labsphere, North Sutton, NH, USA)* com mais de 99% de reflectância. Para cada amostra foram obtidas três repetições, sendo utilizada a curva espectral média.

Para determinar os teores de COS foram coletadas sub-amostras com aproximadamente 3 g de TFSA, que foram moídas e peneiradas para uma fração de partícula < 0,25 mm. As sub-amostras foram armazenadas em *Eppendorfs* para serem feitas as análises por combustão seca em um analisador elementar, modelo *Flash 2000 Organic Elemental Analyzer (ZOBECK et al., 2013)*. Para determinar as frações areia, silte e argila nas amostras de TFSA foi utilizado o método da pipeta, usando dispersante NaOH (DONAGEMMA et al., 2017).

### 5.2.3. Combinação das variáveis ambientais com as curvas espectrais

Foram utilizados os métodos de regressão *Stepwise* e os coeficientes de correlação para selecionar as VAs mais representativas para a modelagem das propriedades do solo. As VAs foram derivadas do modelo digital do terreno (MDT) com resolução espacial de 1 m. O MDT foi obtido pela Secretaria de Estado do Desenvolvimento Econômico e Sustentável (SANTA CATARINA, 2013) a partir do Levantamento Aerofotogramétrico do Estado de SC, pelo método de reamostragem utilizando técnicas de interpolação bilinear e imagens no formato *Geotiff*. Dezoito VAs foram derivadas a partir do MDT (Tabela 10), utilizando-se o pacote RSAGA, versão 2.2.2, integrado ao *software R* versão 3.5.0 (R DEVELOPMENT CORE TEAM, 2017).

Tabela 11. Variáveis ambientais derivadas do modelo digital do terreno.

Área de contribuição (ACT)	Fator LS (FATLS)
Convexidade (CONV)	Índice de posição topográfico (IPT)
Curvatura de perfil (CURVPERF)	Índice de umidade topográfica (IUT)
Curvatura total (CURVTOT)	Índice de rugosidade do terreno (IRT)
Curvatura planar (CURVPLAN)	Insolação difusa (INSOLDIF)
Declividade (DECL)	Insolação direta (ISOLDIR)
Distância da rede de drenagem (DRD)	Insolação total (INSOLTOT)
Elevação (MDT)	Orientação de vertentes (ASPECT)
Escoamento acumulado (ESCACUM)	Profundidade do vale (PROFVALE)

Na sequência, para diminuir a colinearidade das variáveis nos modelos, todas as VAs foram submetidas a um procedimento de seleção automatizado pelo método de regressão *stepwise*. Nesse método, as variáveis são selecionadas (*forward*) ou removidas (*backward*) do conjunto de VAs de acordo com a significância de 5% do critério de Akaike (AIC). A seleção *forward* envolveu a definição de um modelo nulo (sem nenhuma VA definida) e por iteração foram adicionadas as variáveis do conjunto de variáveis candidatas. Cada VA adicionada foi computando a distância entre o modelo verdadeiro e o modelo elegível (AIC). Quando o AIC diminuiu com a adição de uma nova variável, esta foi mantida no modelo e o algoritmo prosseguiu com a próxima variável. Se o AIC não diminuiu, a respectiva variável foi excluída do conjunto de CPs. Em contraste, a seleção *backward* iniciou com todas as VAs candidatas, e por iteração foi feita a retirada das variáveis. Quando o AIC diminuiu com a retirada da variável, essa VA foi retirada do conjunto de variáveis e, como oposto, se o AIC não diminuiu, a VA foi devolvida ao conjunto. Esse procedimento foi realizado com a função *stepAIC* do pacote MASS do software R (VENABLES; RIPLEY, 2002). Ao final da seleção, a regressão *Stepwise* selecionou um conjunto de VAs para cada propriedade do solo, sendo nove para o teor de COS, 12 para a fração areia, 10 para a fração argila e 11 variáveis para a fração silte.

Essas VAs foram as mais representativas para cada propriedade e foram integradas à matriz Vis-NIR-SWIR para que fossem feitas as previsões das propriedades do solo utilizando a combinação das VAs com às assinaturas espectrais (Vis-NIR-SWIR + MDT).

#### 5.2.4. Modelagem preditiva e validação

Quatro métodos multivariados foram implementados para avaliar o desempenho preditivo da inclusão das VAs combinadas com as curvas espectrais. Para tal, testamos três cenários diferentes de previsão: 1º) os modelos preditivos foram calibrados utilizando somente as curvas espectrais; 2º) os modelos foram calibrados utilizando apenas as informações das VAs como variável independente e, finalmente 3º) os modelos foram calibrados utilizando a combinação das curvas espectrais Vis-NIR-SWIR e as VAs (Vis-NIR-SWIR + MDT).

Cada modelo tem parâmetros específicos para a entrada de dados e esse procedimento foi realizado manualmente para obter o melhor ajuste possível. A seguir estão os modelos multivariados com os pacotes do *software* R que foram utilizados para a execução desse estudo: *partial-least-squares regression* (PLSR) no pacote *pls* (MEVIK; WEHRENS; LILAND, 2016), *random forest* (RF) no pacote *randomForest* (LIAW; WIENER, 2002), *support vector machines* (SVM) no pacote *e1071* (MEYER et al., 2020) e *gaussian process regression* (GPR) no pacote *kernelab* (KARATZOGLOU et al., 2004).

Para o modelo PLSR, por ser o mais robusto, foi realizado a determinação dos comprimentos de onda mais importantes na previsão, também conhecido como importância das bandas espectrais na previsão (IBP), descrito em detalhes por CHONG; JUN (2005), e também a determinação dos coeficientes PLSR (chamados de b-coeficientes) (HAALAND; THOMAS, 1988). O vetor dos b-coeficientes possui dimensões de cada espectro individual. O limite para o IBP foi definido como 1 (CHONG; JUN, 2005) e o limite para os b-coeficientes foi baseado em seus desvios padrão ( $\sigma$ ) (VISCARRA ROSSEL; FOUAD; WALTER, 2008). Se as pontuações IBP e os b-coeficientes atingirem esses dois limites, o comprimento de onda espectral é considerado importante (GOMEZ; LAGACHERIE; COULOUMA, 2008).

Das 260 amostras de solo usadas para esse estudo, 70% foram aleatoriamente selecionadas para calibração e 30% para a validação do modelo. Esse procedimento de divisão foi adotado em outros estudos de ER (DEMATTÊ et al., 2017; ROMERO et al., 2017). O teste de Levene (LEVENE, 1960) foi aplicado para avaliar a homogeneidade de variâncias (homocedasticidade) do conjunto de treinamento e validação e para confirmar a confiabilidade da divisão do conjunto. Para avaliar a precisão dos modelos preditivos, três medidas preditivas foram utilizadas: o coeficiente de determinação ( $R^2$ ), que varia entre 0 e 1 e fornece a porcentagem da variação que é explicada pelo modelo; a raiz do erro quadrático médio (RMSE), que mede a acurácia geral do modelo de predição; e a razão de desempenho para o intervalo interquartil (RPIQ).

### **5.3. RESULTADOS E DISCUSSÃO**

#### **5.3.1. Análise descritiva**

Os resultados da estatística descritiva para os conjuntos de treinamento e validação, mostraram que o teor de COS e a fração areia apresentam valores de coeficiente de variação (CV) superior a 68% em ambos os casos, caracterizando uma alta heterogeneidade (Tabela 11). A fração silte apresentou o menor CV, seguida da fração argila. O teor de COS apresentou altos valores em ambos os conjuntos, característico do tipo de clima e solo da região. A conservação das matas de araucária associada a característica húmica dos solos conferida pelas altas altitudes e baixas temperaturas, promovem um maior acúmulo do teor de COS nessa região, esses resultados estão de acordo com DALMOLIN et al. (2017). O teste de Levene (LEVENE, 1960) indicou a homogeneidade de variâncias (homocedasticidade) entre os conjuntos de treinamento e validação para o teor de COS (p-valor = 0,555), fração argila (p-valor = 0,482), fração areia (p-valor = 0,861) e fração silte (p-valor = 0,491). A um nível de significância de  $\alpha = 0,05$ , os conjuntos de treinamento e validação mostraram-se estatisticamente semelhantes.

Tabela 12. Análises descritivas dos conjuntos de treinamento e validação das propriedades do solo utilizadas na modelagem.

	Conjunto de treinamento (%)				Conjunto de validação (%)			
	COS	Argila	Areia	Silte	COS	Argila	Areia	Silte
Amostras	182	182	182	182	78	78	78	78
Mínimo	0,64	8,83	2,77	6,74	1,14	11,35	3,11	6,74
Máximo	34,84	65,72	79,74	65,53	30,10	62,49	79,74	60,41
1º quartil	4,85	18,35	10,96	29,53	4,27	19,04	9,54	32,13
3º quartil	17,35	47,22	43,52	47,35	15,40	47,66	40,33	49,16
Média	11,44	34,07	27,60	38,33	10,20	34,48	26,60	38,92
Mediana	9,05	32,10	19,27	37,91	7,24	35,51	20,29	37,31
Média DP	0,58	1,25	1,48	0,93	0,85	1,81	2,26	1,38
Assimetria	0,71	0,22	0,76	0,00	0,87	0,10	0,86	-0,34
Curtose	-0,55	-1,30	-0,66	-0,58	-0,44	-1,36	-0,30	-0,18
CV (%)	67,87	49,36	72,45	32,59	73,54	46,24	75,03	31,42
Levene (p-valor)	0,555	0,482	0,861	0,491	0,555	0,482	0,861	0,491

Todas as 260 curvas espectrais Vis-NIR-SWIR diferiram e apresentaram intensidades e formas características das assinaturas espectrais de amostras de solo (Figura 13). A reflectância foi menor na região Vis (400-700 nm) e maior na região do infravermelho próximo e de ondas curtas (NIR-SWIR) (700-2500 nm). As amostras de solo foram bem distintas quanto à reflectância espectral que estão associadas a diferentes formas e bandas de absorção. Esta distinção deve-se, principalmente, ao teor de COS e aos óxidos de ferro presentes na fração argila. A presença dos óxidos pode ser aferida pelo comportamento das curvas espectrais nos comprimentos de ondas nas faixas de 420 a 850 nm, que são características de hematita e goethita. Existem características típicas de absorção que são devidas às vibrações das moléculas de água e dos grupos OH<sup>-</sup> e essas feições de absorção são perceptíveis entre as faixas 1400 e 1900 nm (DOTTO et al., 2018; GHOLIZADEH et al., 2016; NOCITA et al., 2014) (Figura 13). As curvas espectrais também apresentaram uma pequena característica de absorção em torno de 2200 nm relacionada aos argilominerais, principalmente caulinita e illita (CAMARGO et al., 2018; NOCITA et al., 2014; ZHAO et al., 2018). Na região do NIR, devido as radiações menos

energéticas, não ocorre transições eletrônicas entre os orbitais, no entanto, as ligações entre os átomos podem vibrar de diferentes formas (NIEMZ, 2015), sendo possível a quantificação das absorções nessa região.

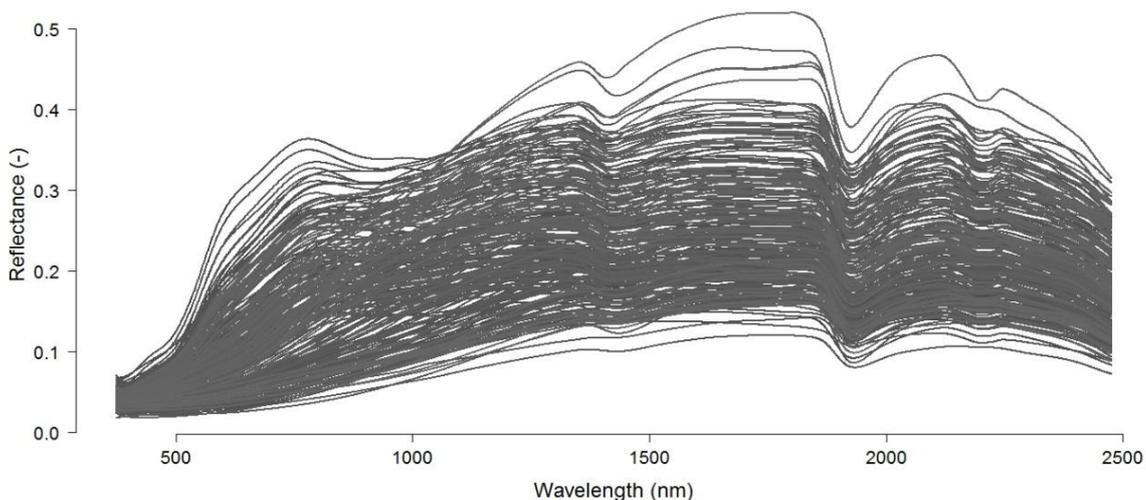


Figura 13. Curva espectrais do solo Vis-NIR-SWIR (400–2500 nm) suavizadas com filtro Savitzky-Golay.

### 5.3.2. Análise de correlação das variáveis ambientais

O teor de COS apresentou forte correlação com as variáveis elevação (0,74), profundidade do vale (-0,47), insolação difusa (-0,22), índice de posição topográfica (0,22) e o índice de umidade topográfica (0,20) (Figura 15). Para a fração argila, as variáveis elevação (-0,71), profundidade do vale (0,41) e insolação difusa (0,16) apresentaram as correlações mais significativas. Para a fração areia, quatro VAs apresentaram correlações significativas, sendo elas a elevação (0,79), a profundidade do vale (-0,53), o índice de umidade topográfica (-0,26) e o índice de posição topográfica (0,25).

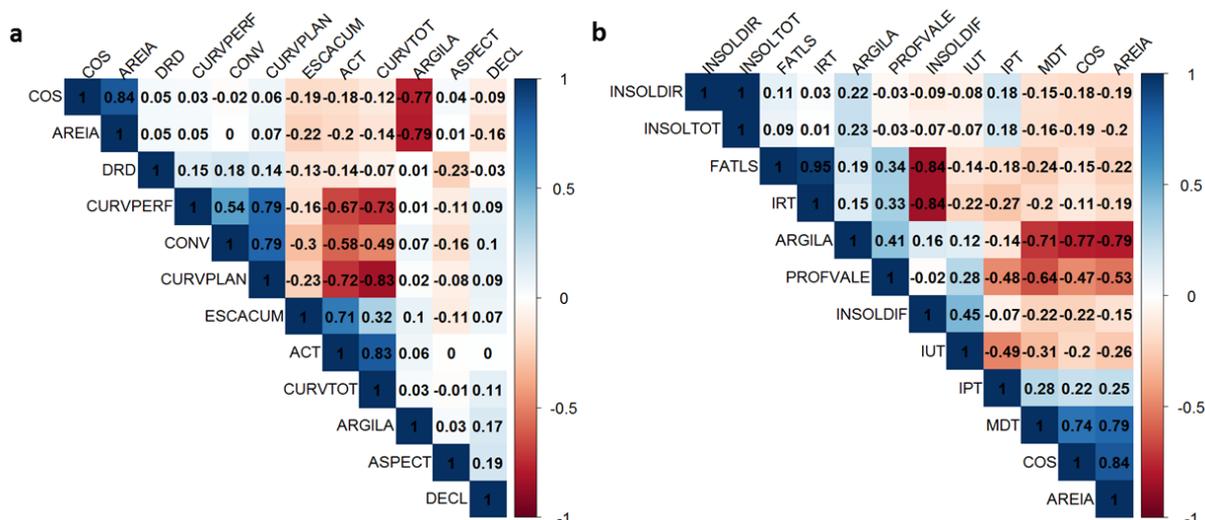


Figura 14. Correlação de Pearson entre as variáveis ambientais e as propriedades do solo (a, b). As cores em azul e vermelho simbolizam correlações positivas e negativas, respectivamente.

A elevação (MDT), o índice de posição topográfica (IPT) e a declividade (DECL) controlam a dinâmica da água do relevo, que influencia a intensidade dos processos de erosão e redistribuição das partículas do solo (LIU et al., 2020). Além disso, a elevação apresentou boa correlação e foi comum a todas as propriedades do solo estudadas, isso pode estar relacionado ao importante papel que essa variável desempenha no desenvolvimento de microclimas (GRIFFITHS; MADRITCH; SWANSON, 2009), que por sua vez influencia os processos do solo (LOZANO-GARCÍA; PARRAS-ALCÁNTARA; BREVIK, 2016). Locais de maior altitude normalmente apresentam temperaturas mais baixas, o que interfere na velocidade do intemperismo dos solos e das rochas, influenciando na degradação da matéria orgânica. Esses resultados estão de acordo com HINGE; SURAMPALLI; GOYAL (2018), que ao realizarem a predição de propriedades do solo, verificaram que a elevação foi a variável com melhor resposta preditiva. A variável ambiental IUT também obteve boa correlação, comportamento explicado pela sua ligação com os gradientes de umidade do solo (SIEWERT, 2018). A segunda variável com melhor correlação e comum a todas as propriedades do solo estudadas foi o índice de profundidade de vale (PROFVALE). Essa variável é calculada como a distância vertical para rede de canais, a qual representa um nível de base (SCHAETZL; MILLER, 2016).

As variáveis curvaturas planar e do perfil (CURVPLAN e CURVPERF) e a orientação de vertentes (ASPECT) obtiveram uma correlação relativamente baixa, embora essas variáveis normalmente controlem a umidade local, as condições térmicas e a redistribuição de massa a curto alcance sobre as paisagens (MCBRATNEY; SANTOS; MINASNY, 2003).

### 5.3.3. Desempenho dos modelos de predição

Os resultados de predição das propriedades do solo diferiram conforme o conjunto dos dados utilizados para treinamento (Tabela 12). No geral, o valor de  $R^2$  foi menor nos modelos que tiveram como variável independente somente as VAs e, notavelmente, mais alto quando se utilizou apenas as curvas espectrais ou a combinação VIS-NIR-SWIR+MDT. No entanto, na predição da fração argila com o modelo *Random Forest* e as VAs, o valor de  $R^2 = 0,81$  e  $RMSE = 7\%$ , foi superior ao conjunto VIS-NIR-SWIR e a combinação VIS-NIR-SWIR+MDT, mostrando bom desempenho desse modelo.

Ao utilizar somente as curvas espectrais na predição do teor de COS, o  $R^2$  variou de 0,77 a 0,91 para os modelos RF e SVM, respectivamente. Na predição da fração argila, o  $R^2$  variou de 0,67 a 0,78 para os modelos RF e PLSR, respectivamente, enquanto que na predição da fração areia, o  $R^2$  variou de 0,68 a 0,73 para os modelos GPR e PLSR, respectivamente. Para a fração silte, os modelos tiveram baixa eficiência de predição (com  $R^2$  variando de 0,23 a 0,48) quando comparados as outras propriedades do solo preditas. Resultados semelhantes foram observados por DOTTO et al. (2017), que ao realizarem a predição de propriedades do solo por espectroscopia, obtiveram  $R^2$  variando de 0,28 a 0,57 para modelos preditivos da fração silte e  $R^2$  variando de 0,13 a 0,33 para os modelos preditivos da fração areia.

Tabela 13. Resultado da validação dos modelos preditivos, considerando os três cenários de predição.

Propriedades do solo	Método	Vis-NIR-SWIR			MDT			Vis-NIR-SWIR + MDT		
		R <sub>val</sub>	RMSE <sub>val</sub>	RPIQ	R <sub>val</sub>	RMSE <sub>val</sub>	RPIQ	R <sub>val</sub>	RMSE <sub>val</sub>	RPIQ
COS	SVM	0,91	2,25	5,29	0,63	4,57	2,44	0,93	2,10	5,68
	GPR	0,90	2,43	4,91	0,64	4,52	2,46	0,92	2,21	5,40
	PLSR	0,88	2,60	4,59	0,66	4,45	2,50	0,92	2,24	5,34
	RF	0,77	3,67	3,25	0,75	3,79	2,94	0,79	3,49	3,42
Argila	PLSR	0,78	7,59	3,57	0,58	10,28	2,78	0,81	6,95	3,89
	SVM	0,74	8,54	3,17	0,57	10,73	2,67	0,77	7,94	3,41
	GPR	0,75	8,14	3,32	0,58	10,41	2,75	0,77	7,72	7,72
	RF	0,67	9,13	2,96	0,81	7,00	4,09	0,71	8,54	3,17
Areia	PLSR	0,73	11,01	2,80	0,62	12,30	2,50	0,77	9,85	3,13
	RF	0,71	10,74	2,87	0,69	11,31	2,72	0,75	10,03	3,07
	SVM	0,72	10,63	2,90	0,62	12,38	2,49	0,79	9,29	3,32
	GPR	0,68	11,50	2,68	0,62	12,68	2,43	0,76	10,27	3,00
Silte	RF	0,23	9,34	1,61	0,48	9,33	1,83	0,28	8,84	1,70
	SVM	0,48	7,56	1,99	0,08	10,04	1,50	0,40	8,09	1,86
	GPR	0,45	7,68	1,96	0,13	11,96	1,42	0,42	7,95	1,89
	PLSR	0,42	7,78	1,93	0,15	11,38	1,50	0,41	9,15	1,64

Ao combinar os espectros Vis-NIR-SWIR com as VAs, os modelos preditivos obtiveram melhor desempenho, principalmente para a fração argila e o teor de COS. No desempenho preditivo da fração argila utilizando os modelos RF e PLSR, o  $R^2$  e RMSE variaram de 0,71 e 8,54% a 0,81 e 6,95%, respectivamente. Na análise comparativa entre os resultados obtidos usando somente as curvas espectrais e a combinação Vis-NIR-SWIR+MDT, com o modelo PLSR, o  $R^2$  e RMSE passaram de 0,78 e 7,59% para 0,81 e 6,95%, configurando um desempenho de 4 e 8% melhor, respectivamente. Para o teor de COS, o desempenho do modelo PLSR apresentou melhor performance quando utilizou-se a combinação Vis-NIR-SWIR+MDT, onde o  $R^2$  e RMSE passaram de 0,88 e 2,6% (utilizando Vis-NIR-SWIR) para 0,92 e 2,24% (utilizando Vis-NIR-SWIR+MDT), configurando um desempenho de 4 e 14% melhor, respectivamente.

#### 5.3.4. Importância das bandas na predição (IBP)

As bandas de maior importância na predição e os b-coeficientes do modelo PLSR são mostrados na Figura 14. De acordo com esses resultados, os

comprimentos de onda próximos a 420, 520, 1400, 1900, 2000, 2200 e 2300 nm foram identificados como os mais importantes e comuns para a estimativa do teor de COS (Figura 14a). A seleção das bandas de maior importância para predição dessa propriedade é muito variável e depende do teor de COS presente nas amostras. Geralmente, amostras com elevado teor de COS apresentam um decréscimo significativo na reflectância, enquanto que em amostras com baixo teor esse comportamento não é observado (DEMATTÊ et al., 2019; MOURA-BUENO et al., 2019; VISCARRA ROSSEL et al., 2006b). De acordo com os resultados, na predição da fração argila, os comprimentos de onda próximos a 420, 480, 520, 1400, 1900, 2200 e 2300 nm são faixas espectrais importantes na estimativa dessa propriedade do solo (Figura 14b). Para a fração areia, os comprimentos de onda mais importantes foram identificados perto de 520, 1000, 1900, 2200 e 2300 nm. (Figura 14c). Da perspectiva de comprimentos de onda importantes, muitos foram identificados como sendo comuns para prever o teor de COS e as frações argila e areia. As curvas espectrais comuns foram observadas principalmente nas regiões de 420 a 800 e 1900 a 2300 nm, bem como na região em torno de 1000 nm. As assinaturas espectrais na região do NIR-SWIR são compostas, principalmente, por feições causadas por sobretons fracos e combinações de bandas de vibração fundamental para C-H, N-H e ligações O-H (VISCARRA ROSSEL et al., 2006b), bem como Al-OH, presente nos argilominerais. Esses resultados são consistentes com outros estudos, nos quais os comprimentos de onda importantes para a predição do teor de COS são encontrados em 520, 570, 620 e 1000 nm (DING et al., 2018; JIANG et al., 2017; MOURA-BUENO et al., 2019) e para a textura do solo em torno de 420, 480, 520 e 1900 (DALMOLIN et al., 2005; FANG et al., 2018).

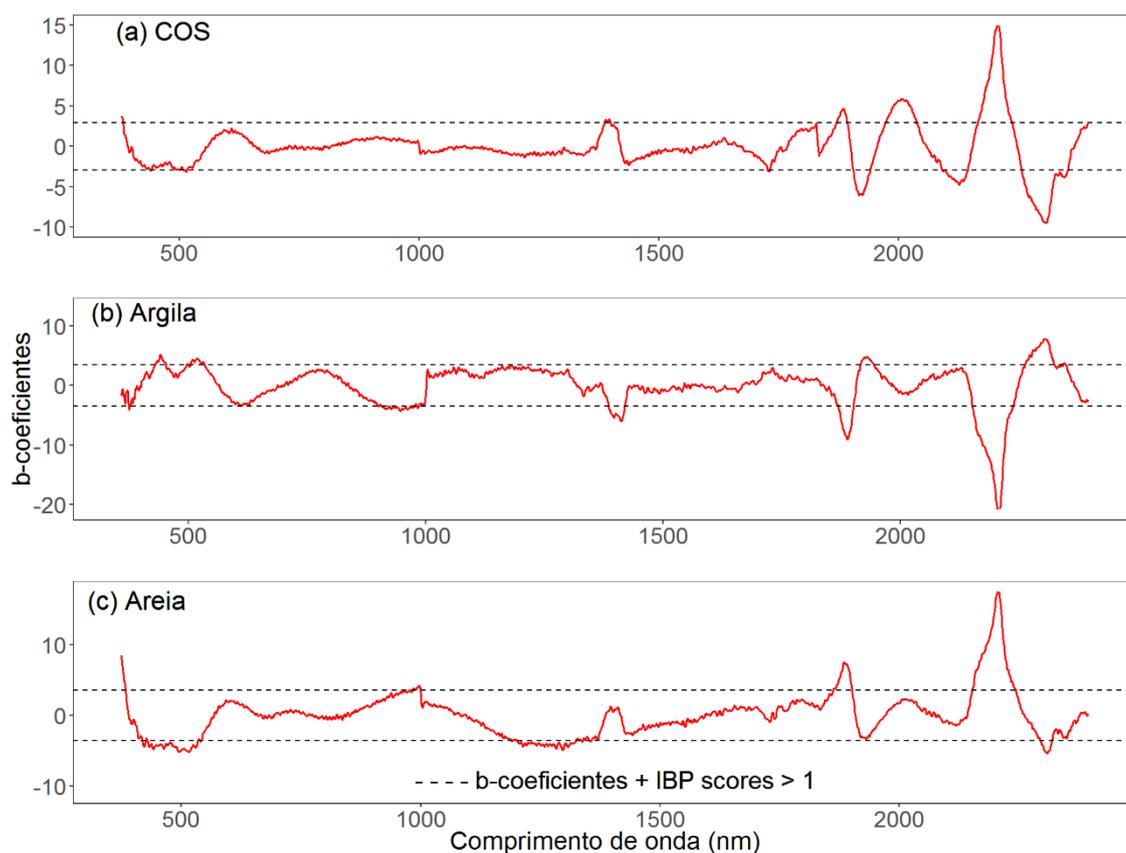


Figura 15. Importância das bandas na predição (IBP) e os b-coeficientes do modelo PLSR na predição do teor de COS e das frações argila e areia utilizando espectroscopia Vis-NIR-SWIR. (a) na predição do teor de COS, o desvio padrão foi de  $\sigma = \pm 2,93$ , (b) na predição da fração argila, o desvio padrão foi de  $\sigma = \pm 3,51$  e (c) na predição da fração areia, o desvio padrão foi de  $\sigma = \pm 3,59$ .

Para as frações texturais do solo, às características de absorção espectral estão relacionadas a presença de minerais de óxidos de ferro, sendo a goethita ( $\alpha\text{-FeOOH}$ ) e a hematita ( $\alpha\text{-Fe}_2\text{O}_3$ ) os mais comuns. As bandas de absorção em torno de 420, 480, 600 e 1700 indicam a presença de goethita, enquanto as absorções relacionadas a presença de hematita são identificadas em torno de 520, 650 e 850 nm (BAHIA et al., 2015; RAMOS et al., 2020). As bandas de absorção na região Vis estão relacionadas às transições eletrônicas dos óxidos de ferro presentes nas amostras de solo (DOTTO et al., 2018; HONG et al., 2019). Por outro lado, na região SWIR, as bandas de 1680 e 1750 nm estão relacionadas à vibração de estiramento de C-H (VISCARRA ROSSEL; BEHRENS, 2010), e as bandas de 1400, 1900 nm estão relacionadas à adsorção de água na superfície do solo e na estrutura dos argilominerais (DOTTO et al.,

2018; NOCITA et al., 2014). A análise da importância das bandas espectrais na predição, mostrou que existem padrões de absorção bem específicos que estão associados às características dos solos e diferem quanto a região espectral (NOCITA et al., 2014). Com a heterogeneidade das amostras, a faixa flutuante das curvas do b-coeficiente foi diferente para cada propriedade do solo, apresentando algumas feições de absorção em comum.

#### **5.4. CONCLUSÕES**

Neste estudo, foram construídos modelos de predição para as estimativas do teor de COS e das frações texturais do solo utilizando as curvas espectrais individuais e combinadas com variáveis ambientais. Concluiu-se que os espectros de reflectância (350-2500 nm) podem ser usados na estimativa dessas propriedades do solo. No entanto, os resultados mostraram que a inclusão de uma segunda variável preditora integrada à matriz espectral melhorou o desempenho dos modelos de predição.

As variáveis ambientais que apresentaram maior correlação com as propriedades do solo foram a elevação, seguida dos índices de profundidade do vale e umidade topográfica e da insolação difusa. Essas variáveis desempenham importantes processos no solo que estão relacionados, principalmente, a dinâmica da água do relevo e da umidade do solo, o que influencia a intensidade dos processos de erosão e deslocamento das partículas do solo.

Os comprimentos de onda mais importantes e comuns na predição do teor de COS e das frações texturais foram observados nas regiões de 400-800 e 2000-2300 nm, bem como na região de 1900 nm, foram importantes na estimativa dessas propriedades do solo.

## 6. CAPÍTULO V – CONSIDERAÇÕES FINAIS

Os estudos que compõem esta tese buscaram refinar a aplicação da técnica de espectroscopia de reflectância (ER) na predição de propriedades do solo. O Estudo 1 destacou o uso de parâmetros de cor derivados do espectro visível (Vis) aplicados na classificação de amostras de solos utilizadas na predição do teor de COS e das frações texturais, contribuindo para a ER no aprimoramento de metodologias que podem ser aplicadas em laboratório de análise de solo. No entanto, o agrupamento de conjunto de amostras com alto teor de COS prejudicou o desempenho dos modelos, uma vez que o COS está associado a matéria orgânica do solo, que tem baixa reflectância e mascara a resposta espectral, atentando que é preciso adotar outras estratégias ao trabalhar com conjuntos de amostras com alto teor de COS e MO.

O Estudo 1, também destacou que a classificação das amostras com base na cor do solo é uma boa estratégia para alcançar melhores resultados de predição do teor de COS e das frações areia e argila, uma vez que, separou e/ou agrupou conjuntos de amostras mais homogêneos em relação aos constituintes do solo. No entanto, a metodologia não foi eficaz para melhorar a capacidade preditiva dos modelos de predição da fração silte, sendo necessário desenvolver outras estratégias para melhorar os modelos de predição para essa propriedade do solo. Assim, a aplicação dessa metodologia alcançou resultados satisfatórios mesmo utilizando uma biblioteca espectral local, criando a possibilidade de aplicá-la em bibliotecas espectrais regionais, que integram solos com características mais heterogêneas.

O Estudo 2, destacou que a estratificação de bibliotecas espectrais regionais é uma estratégia para melhorar as avaliações regionais dos recursos do solo e obter resultados mais precisos da textura do solo. Com a separação e/ou agrupamento das amostras foi possível identificar os grupos funcionais que compõem as propriedades do solo e que interferem nas absorções da energia, principalmente na região do Vis-SWIR. O Estudo 2 também revelou que a estrutura dos dados de entrada é um fator que tem maior efeito sobre a

capacidade preditiva dos modelos do que o tipo de pré-processamento espectral, confirmando a hipótese do presente estudo e elucidando a influência de dados muito heterogêneos na precisão dos modelos.

O Estudo 3 buscou explorar o potencial dos modelos de predição ao integrar outras variáveis preditoras à matriz espectral. As variáveis ambientais (VAs) possuem potencial no MDS ao serem utilizadas na predição de classes de solos. No entanto, viu-se a possibilidade de explorar a aplicação dessas variáveis na predição de propriedades do solo, pois essas variáveis caracterizam o ambiente nas mais variadas escalas, auxiliando na modelagem. Ao integrar as VAs à matriz espectral, os modelos ficaram mais robustos e com maior possibilidade de alcançar resultados mais satisfatórios. A aplicação dessa metodologia possui a vantagem de alcançar resultados mais precisos e como desvantagem a exigência computacional e o tempo de processamento dos dados.

Esses resultados reiteram que a utilização de bibliotecas espectrais de diferentes escalas na predição de propriedades físico-químicas do solo necessita de análise prévia da heterogeneidade das amostras que compõem essas bibliotecas, principalmente quando se tratam de amostras de solos oriundas de áreas que apresentem diferentes materiais de origem. Assim, os resultados obtidos na presente tese criam a possibilidade de melhor entendimento dos fatores que podem influenciar nos modelos de calibração. Além disso, os laboratórios de rotina de solos podem adotar a presente metodologia, estratificando as amostras de solos pela cor em uma etapa prévia à análise espectroscópica.

Trabalhos futuros podem ser realizados para medir e relatar a abundância relativa de minerais no solo (como os óxidos de ferro, caulinita, gibbsita e também os minerais 2:1) a partir das curvas espectrais do solo. Outros estudos também podem explorar o potencial da presente metodologia ao aplicá-la em bibliotecas espectrais a nível nacional ou global, visando obter melhores estimativas de propriedades de solos.

## 7. REFERÊNCIAS BIBLIOGRÁFICAS

- ADHIKARI, K. *et al.* Digital mapping of soil organic carbon contents and stocks in Denmark. **PLoS ONE**, San Francisco, v. 9, n. 8, [art.] e105519, 2014.
- AÏCHI, H. *et al.* Regional predictions of soil organic carbon content from spectral reflectance measurements. **Biosystems Engineering**, London, v. 104, n. 3, p. 442-446, 2009.
- AITKENHEAD, M. *et al.* Prediction of soil characteristics and colour using data from the National Soils Inventory of Scotland. **Geoderma**, Amsterdam, v. 200/201, p. 99–107, 2013a.
- AITKENHEAD, M. *et al.* E-smart: environmental sensing for monitoring and advising in real-time. **IFIP Advances in Information and Communication Technology**, Boston, v. 413, p. 129-142, 2013b.
- ALLO, M. *et al.* Prediction of tropical volcanic soil organic carbon stocks by visible-near- and mid-infrared spectroscopy. **Catena**, Cremlingen, v. 189, [art.] 104452, 2020.
- ALLORY, V. *et al.* Quantification of soil organic carbon stock in urban soils using visible and near infrared reflectance spectroscopy (VNIRS) in situ or in laboratory conditions. **Science of the Total Environment**, Amsterdam, v. 686, p. 764–773, 2019.
- ARAÚJO, S. R. *et al.* Improving the prediction performance of a large tropical vis-NIR spectroscopic soil library from Brazil by clustering into smaller subsets or use of data mining calibration techniques. **European Journal of Soil Science**, Oxford, v. 65, n. 5, p. 718–729, 2014.
- ARGYRI, A. A. *et al.* A comparison of Raman and FT-IR spectroscopy for the prediction of meat spoilage. **Food Control**, Kidlington, v. 29, n. 2, p. 461-470, 2013.
- BA, Y. *et al.* Application of Vis-NIR spectroscopy for determination the content of organic matter in saline-alkali soils. **Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy**, Amsterdam, v. 229, [art.] 117863, 2020.
- BAGHERI BODAGHABADI, M. *et al.* Digital soil mapping using artificial neural networks and terrain-related attributes. **Pedosphere**, Beijing, v. 25, n. 4, p. 580-591, 2015.
- BAHIA, A. S. R. S. *et al.* Iron oxides as proxies for characterizing anisotropy in soil CO<sub>2</sub> emission in sugarcane areas under green harvest. **Agriculture, Ecosystems and Environment**, Amsterdam, v. 192, p. 152–162, 2014.
- BAHIA, A. S. R. S. *et al.* Field-scale spatial correlation between contents of iron

oxides and CO<sub>2</sub> emission in an Oxisol cultivated with sugarcane. **Scientia Agricola**, Piracicaba, v. 72, n. 2, p. 157–166, 2015.

BARTHOLOMEUS, H. *et al.* Soil organic carbon mapping of partially vegetated agricultural fields with imaging spectroscopy. **International Journal of Applied Earth Observation and Geoinformation**, Washington, DC, v. 13, n. 1, p. 81–99, 2011.

BARTHOLOMEUS, H. M. *et al.* Spectral reflectance based indices for soil organic carbon quantification. **Geoderma**, Amsterdam, v. 145, n. 1/2, p. 28–36, 2008.

BAUMANN, K. *et al.* Rapid assessment of soil organic matter: soil color analysis and Fourier transform infrared spectroscopy. **Geoderma**, Amsterdam, v. 278, p. 49–57, Sept. 2016.

BAUMGARDNER, M. F. *et al.* Reflectance properties of soils. **Advances in Agronomy**, San Diego, v. 38, p. 1–44, 1986.

BEHRENS, T. *et al.* Multi-scale digital terrain analysis and feature selection for digital soil mapping. **Geoderma**, Amsterdam, v. 155, n. 3/4, p. 175–185, 2010.

BEN-DOR, E. *et al.* Imaging spectrometry for soil applications. **Advances in Agronomy**, San Diego, v. 97, p. 321–392, 2008.

BEN-DOR, E.; BANIN, A. Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. **Soil Science Society of America Journal**, Madison, v. 59, n. 2, p. 364–372, 1995.

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.

BEZDEK, J. C.; EHRLICH, R.; FULL, W. FCM: the fuzzy c-means clustering algorithm. **Computers and Geosciences**, New York, v. 10, n. 2/3, p. 191–203, 1984.

BONA, E. *et al.* Support vector machines in tandem with infrared spectroscopy for geographical classification of green arabica coffee. **LWT - Food Science and Technology**, London, v. 76, p. 330–336, 2017.

BORCARD, D.; GILLET, F.; LEGENDRE, P. **Numerical ecology with R**. New York: Springer, 2011. 306 p.

BREIMAN, L. Random forests. **Machine Learning**, Boston, v. 45, p. 5–32, 2001.

BROWN, D. J. *et al.* Global soil characterization with VNIR diffuse reflectance spectroscopy. **Geoderma**, Amsterdam, v. 132, n. 3/4, p. 273–290, 2006.

- BROWN, D. J. Using a global VNIR soil-spectral library for local soil characterization and landscape modeling in a 2nd-order Uganda watershed. **Geoderma**, Amsterdam, v. 140, n. 4, p. 444-453, 2007.
- CAMARGO, L. A. *et al.* Pedotransfer functions to assess adsorbed phosphate using iron oxide content and magnetic susceptibility in an Oxisol. **Soil Use and Management**, Oxford, v. 32, n. 2, p. 172–182, 2016.
- CAMARGO, L. A. *et al.* Predicting potentially toxic elements in tropical soils from iron oxides, magnetic susceptibility and diffuse reflectance spectra. **Catena**, Cremlingen, v. 165, p. 503–515, 2018.
- CAMBULE, A. H. *et al.* Building a near infrared spectral library for soil organic carbon estimation in the Limpopo National Park, Mozambique. **Geoderma**, Amsterdam, v. 183/184, p. 41–48, 2012.
- CEBECI, Z. *et al.* **Pclust**: probabilistic and possibilistic cluster analysis. [S.l.]: R Package Documentation, 2020.
- CHABRILLAT, S. *et al.* Field and Imaging spectrometry for identification and mapping of expansive soils. *In*: VAN DER MEER, F. D.; JONG, S. M. (ed.). **Imaging spectrometry: basic principles and prospective applications**. Dordrecht: Springer, 2002. cap. 4, p. 87-109.
- CHEN, T.; MORRIS, J.; MARTIN, E. Gaussian process regression for multivariate spectroscopic calibration. **Chemometrics and Intelligent Laboratory Systems**, Amsterdam, v. 87, n. 1, p. 59-71, 2007.
- CHONG, I. G.; JUN, C. H. Performance of some variable selection methods when multicollinearity is present. **Chemometrics and Intelligent Laboratory Systems**, Amsterdam, v. 78, n. 1/2, p. 103-112, 2005.
- COMMISSION INTERNATIONALE DE L'ECLAIRAGE, 1931, Cambridge, UK. **Proceedings** [...]. Cambridge: University Cambridge, 1932.
- CLARK, R. N. Spectroscopy of rocks and minerals, and principles of spectroscopy. *In*: RENCZ, A. N. (ed.). **Remote sensing for the earth sciences: manual of remote sensing**. New York: John Wiley, 1999. cap. 1, p. 1-50.
- COBLINSKI, J. A. *et al.* Prediction of soil texture classes through different wavelength regions of reflectance spectroscopy at various soil depths. **Catena**, Cremlingen, v. 189, [art.] 104485, 2020.
- COELHO, F. F. **Comparação de métodos de mapeamento digital de solos através de variáveis geomorfológicas e sistemas de informações geográficas**. 2010. Dissertação (Mestrado) - Programa de Pós-Graduação em Sensoriamento Remoto, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2010.

CORTES, C.; VAPNIK, V. Support-vector networks. **Machine Learning**, Boston, v. 20, n. 3, p. 273–297, 1995.

COX, M. S. *et al.* Variability of selected soil properties and their relationships with soybean yield. **Soil Science Society of America Journal**, Madison, v. 67, n. 4, p. 1296-1302, 2003.

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.

DALMOLIN, R. S. D. *et al.* Solos do planalto das Araucárias. *In*: CURTI, N. *et al.* (ed.). **Pedologia: solos dos biomas brasileiros**. Viçosa, MG: Sociedade Brasileira de Ciência do Solo, 2017. cap. 8, p. 353-406.

DAVEY, B. G.; RUSSELL, J. D.; WILSON, M. J. Iron oxide and clay minerals and their relation to colours of red and yellow podzolic soils near Sydney, Australia. **Geoderma**, Amsterdam, v. 14, n. 2, p. 125–138, 1975.

DEISS, L. *et al.* Tuning support vector machines regression models improves prediction accuracy of soil properties in MIR spectroscopy. **Geoderma**, Amsterdam, v. 365, [art.] 114227, 2020.

DEMATTÊ, J. A. M. *et al.* Visible-NIR reflectance: a new approach on soil evaluation. **Geoderma**, Amsterdam, v. 121, n. 1/2, p. 95–112, 2004.

DEMATTÊ, J. A. M. *et al.* Spectral reflectance for the mineralogical evaluation of Brazilian low clay activity soils. **International Journal of Remote Sensing**, London, v. 28, n. 20, p. 4537-4559, 2007.

DEMATTÊ, J. A. M. *et al.* Morphological Interpretation of Reflectance Spectrum (MIRS) using libraries looking towards soil classification. **Scientia Agricola**, Piracicaba, v. 71, n. 6, p. 509–520, 2014.

DEMATTÊ, J. A. M. *et al.* Spectral regionalization of tropical soils in the estimation of soil attributes. **Revista Ciência Agronômica**, Fortaleza, v. 47, n. 4, p. 589–598, 2016.

DEMATTÊ, J. A. M. *et al.* Chemometric soil analysis on the determination of specific bands for the detection of magnesium and potassium by spectroscopy. **Geoderma**, Amsterdam, v. 288, p. 8–22, 2017.

DEMATTÊ, J. A. M. *et al.* The Brazilian Soil Spectral Library (BSSL): a general view, application and challenges. **Geoderma**, Amsterdam, v. 354, [art.] 113793, 2019.

DEMATTÊ, J. A. M.; EPIPHANIO, J. C. N.; FORMAGGIO, A. R. Influência da matéria orgânica e de formas de ferro na reflectância de solos tropicais. **Bragantia**, Campinas, v. 62, n. 3, p. 451–464, 2003.

- DING, J. *et al.* Machine-learning-based quantitative estimation of soil organic carbon content by VIS/NIR spectroscopy. **PeerJ**, Corte Madera, CA, v. 6, [art.] e5714, 2018.
- DOMINGUEZ, J. S. *et al.* Sistema de Notación Munsell y CIELab como herramienta para evaluación de color en suelos. **Revista Mexicana de Ciencias Agrícolas**, Mexico, DF, v. 3, n. 1, p. 141–155, 2010.
- DONAGEMMA, G. K. *et al.* **Padronização de métodos para análise granulométrica no Brasil**. Rio de Janeiro: Embrapa Solos, 2017. v. 3, 573 p.
- DOTTO, A. C. *et al.* Potential of spectroradiometry to classify soil clay content. **Revista Brasileira de Ciência do Solo**, Viçosa, MG, v. 40, p. 1–8, 2016.
- DOTTO, A. C. *et al.* Two preprocessing techniques to reduce model covariables in soil property predictions by Vis-NIR spectroscopy. **Soil and Tillage Research**, Amsterdam, v. 172, p. 59–68, 2017.
- 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.
- DUDA, B. M. *et al.* Soil characterization across catenas via advanced proximal sensors. **Geoderma**, Amsterdam, v. 298, p. 78–91, 2017.
- DUFRECHOU, G.; GRANDJEAN, G.; BOURGUIGNON, A. Geometrical analysis of laboratory soil spectra in the short-wave infrared domain: clay composition and estimation of the swelling potential. **Geoderma**, Amsterdam, v. 243/244, p. 92-107, 2015.
- EMBRAPA. **Solos do Estado de Santa Catarina**. Rio de Janeiro: Embrapa Solos, 2004. (Boletim de Pesquisa e Desenvolvimento, n. 46).
- FANG, Q. *et al.* Visible and near-infrared reflectance spectroscopy for investigating soil mineralogy: a review. **Journal of Spectroscopy**, New York, v. 2018, [art.] ID 3168974, 2018.
- FERRARESI, T. M. *et al.* Espectroscopia de infravermelho na determinação da textura do solo TT - Infrared spectroscopy in determination of soil texture. **Revista Brasileira de Ciência do Solo**, Viçosa, MG, v. 36, n. 1, p. 1769–1777, 2012.
- FERRARO, M. B.; GIORDANI, P. A toolbox for fuzzy clustering using the R programming language. **Fuzzy Sets and Systems**, Amsterdam, v. 279, p. 1–16, 2015.
- FISCHER, M. *et al.* Implementing large-scale and long-term functional biodiversity research: the biodiversity exploratories. **Basic and Applied Ecology**, Jena, v. 11, n. 6, p. 473–485, 2010.

FLORINSKY, I. V. *et al.* Prediction of soil properties by digital terrain modelling. **Environmental Modelling and Software**, Amsterdam, v. 17, n. 3, p. 295-311, 2002.

FRANCESCHINI, M. H. D. *et al.* Abordagens semiquantitativa e quantitativa na avaliação da textura do solo por espectroscopia de reflectância bidirecional no VIS-NIR-SWIR. **Pesquisa Agropecuária Brasileira**, Brasília, DF, v. 48, n. 12, p. 1569–1582, 2013.

GALVÃO, L. S.; VITORELLO, Í.; PARADELLA, W. R. Spectroradiometric discrimination of laterites with principal components analysis and additive modeling. **Remote Sensing of Environment**, New York, v. 53, n. 2, p. 70–75, 1995.

GHOLIZADEH, A. *et al.* A memory-based learning approach as compared to other data mining algorithms for the prediction of soil texture using diffuse reflectance spectra. **Remote Sensing**, Basel, v. 8, n. 4, [art.] 341, 2016.

GHOLIZADEH, A. *et al.* Soil organic carbon and texture retrieving and mapping using proximal, airborne and sentinel-2 spectral imaging. **Remote Sensing of Environment**, New York, v. 218, p. 89–103, 2018.

GIASSON, E. *et al.* Avaliação de cinco algoritmos de árvores de decisão e três tipos de modelos digitais de elevação para mapeamento digital de solos a nível semidetalhado na Bacia do Lageado Grande, RS, Brasil. **Ciência Rural**, Santa Maria, v. 43, n. 11, p. 1967-1973, 2013.

GIJSBERTS, A.; METTA, G. Real-time model learning using incremental sparse spectrum gaussian process regression. **Neural Networks**, New York, v. 41, p. 59-69, 2013.

GOMEZ, C. *et al.* Prediction of soil organic and inorganic carbon concentrations in Tunisian samples by mid-infrared reflectance spectroscopy using a French national library. **Geoderma**, Amsterdam, v. 375, [art.] 114469, 2020.

GOMEZ, C.; LAGACHERIE, P.; COULOUMA, G. Continuum removal versus PLSR method for clay and calcium carbonate content estimation from laboratory and airborne hyperspectral measurements. **Geoderma**, Amsterdam, v. 148, n. 2, p. 141-148, 2008.

GRIFFITHS, R. P.; MADRITCH, M. D.; SWANSON, A. K. The effects of topography on forest soil characteristics in the Oregon Cascade Mountains (USA): implications for the effects of climate change on soil properties. **Forest Ecology and Management**, Amsterdam, v. 257, n. 1, p. 1-7, 2009.

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, n. 2, p. 141–151, 2012.

GRUNWALD, S.; VASQUES, G. M.; RIVERO, R. G. Fusion of soil and remote

sensing data to model soil properties. **Advances in Agronomy**, San Diego, v. 131, p. 1-109, 2015.

GUERRERO, C. *et al.* Spiking of NIR regional models using samples from target sites: effect of model size on prediction accuracy. **Geoderma**, Amsterdam, v. 158, n. 1/2, p. 66–77, 2010.

GUERRERO, C. *et al.* Assessment of soil organic carbon at local scale with spiked NIR calibrations: effects of selection and extra-weighting on the spiking subset. **European Journal of Soil Science**, Oxford, v. 65, n. 2, p. 248–263, 2014.

GUPTA, S.; ISLAM, S.; HASAN, M. M. Analysis of impervious land-cover expansion using remote sensing and GIS: a case study of Sylhet sadar upazila. **Applied Geography**, Sevenoaks, v. 98, p. 156–165, June 2018.

HAALAND, D. M.; THOMAS, E. V. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. **Analytical Chemistry**, Washington, DC, v. 60, n. 11, p. 1193–1202, 1988.

HINGE, G.; SURAMPALLI, R. Y.; GOYAL, M. K. Prediction of soil organic carbon stock using digital mapping approach in humid India. **Environmental Earth Sciences**, Heidelberg, v. 77, n. 5, [art.] 172, 2018.

HONG, Y. *et al.* Combination of fractional order derivative and memory-based learning algorithm to improve the estimation accuracy of soil organic matter by visible and near-infrared spectroscopy. **Catena**, Cremlingen, v. 174, p. 104–116, 2019.

HUNT, G. R.; SALISBURY, J. W. Visible and near-infrared spectra of minerals and rocks: 1. Silicate minerals. **Modern Geology**, London, v. 1, p. 283–300, 1970.

HUTENGS, C. *et al.* In situ and laboratory soil spectroscopy with portable visible-to-near-infrared and mid-infrared instruments for the assessment of organic carbon in soils. **Geoderma**, Amsterdam, v. 355, [art.] 113900, 2019.

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

JACONI, A.; VOS, C.; DON, A. Near infrared spectroscopy as an easy and precise method to estimate soil texture. **Geoderma**, Amsterdam, v. 337, p. 906–913, 2019.

JI, W. *et al.* Prediction of soil attributes using the chinese soil spectral library and standardized spectra recorded at field conditions. **Soil and Tillage Research**, Amsterdam, v. 155, p. 492–500, 2016.

JIANG, Q. *et al.* Estimation of soil organic carbon and total nitrogen in different soil layers using VNIR spectroscopy: effects of spiking on model applicability. **Geoderma**, Amsterdam, v. 293, p. 54–63, 2017.

KARATZOGLOU, A. *et al.* Kernlab - An S4 package for kernel methods in R. **Journal of Statistical Software**, Los Angeles, v. 11, n. 9, p. 1-20, 2004.

KLEIN, R. M. Mapa fitogeográfico do estado de Santa Catarina. *In*: REITZ, R. (ed.). **Flora ilustrada catarinense**. Itajaí: Herbário Barbosa Rodrigues, 1978. p. 24.

KLEIN, V. A. *et al.* Textura do solo e a estimativa do teor de água no ponto de murcha permanente com psicrômetro. **Ciência Rural**, Santa Maria, v. 40, n. 7, p. 1550-1556, 2010.

KNOX, N. M. *et al.* Modelling soil carbon fractions with visible near-infrared (VNIR) and mid-infrared (MIR) spectroscopy. **Geoderma**, Amsterdam, v. 239/240, p. 229-239, 2015.

KOKULAN, V. *et al.* Importance of terrain attributes in relation to the spatial distribution of soil properties at the micro scale: a case study. **Canadian Journal of Soil Science**, Ottawa, v. 98, n. 2, p. 292–305, 2018.

KUANG, B.; TEKIN, Y.; MOUAZEN, A. M. Comparison between artificial neural network and partial least squares for on-line visible and near infrared spectroscopy measurement of soil organic carbon, pH and clay content. **Soil and Tillage Research**, Amsterdam, v. 146, pt. B, p. 243-252, 2015.

KUHN, M. *et al.* **Package ‘Cubist’. Rule- and instance-based regression modeling: version 0.2.3.** [S.l.: Packages R CRAN], 10 Jan. 2020. Disponível em: <https://cran.r-project.org/web/packages/Cubist/Cubist.pdf>. Acesso em: 27 nov. 2020.

LEGENDRE, P.; LEGENDRE, L. **Numerical ecology**. 2nd ed. New York: Elsevier, 1998. (Developments in Environmental Modelling, v. 24).

LEVENE, H. Robust tests for equality of variances. **Journal of the American Statistical Association**, New York, v. 69, n. 346, p. 364-367, 1960.

LEVIN, N.; BEN-DOR, E.; SINGER, A. A digital camera as a tool to measure colour indices and related properties of sandy soils in semi-arid environments. **International Journal of Remote Sensing**, London, v. 26, n. 24, p. 5475–5492, 2005.

LI, B.; MORRIS, J.; MARTIN, E. B. Model selection for partial least squares regression. **Chemometrics and Intelligent Laboratory Systems**, Amsterdam, v. 64, n. 1, p. 79-89, 2002.

LI, D. *et al.* Prediction of soil organic matter content in a litchi orchard of South China using spectral indices. **Soil and Tillage Research**, Amsterdam, v. 123,

p. 78-86, 2012.

LIAW, A.; WIENER, M. Classification and regression by random forest. **R News**, Wien, v. 2, n. 3, p. 18–22, 2002.

LISS, M.; GLASER, B.; HUWE, B. Uncertainty in the spatial prediction of soil texture. Comparison of regression tree and random forest models. **Geoderma**, Amsterdam, v. 170, p. 70-79, 2012.

LILLESAND, T.; KIEFER, R.; CHIPMAN, J. **Remote sensing and image interpretation**. 7th ed. Hoboken: Wiley, 2015.

LIU, C.; YANG, S. X.; DENG, L. A comparative study for least angle regression on NIR spectra analysis to determine internal qualities of navel oranges. **Expert Systems with Applications**, Oxford, v. 42, n. 22, p. 8497-8503, 2015.

LIU, F. *et al.* High-resolution and three-dimensional mapping of soil texture of China. **Geoderma**, Amsterdam, v. 361, [art.] 114061, 2020.

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

LIU, W. *et al.* A survey of deep neural network architectures and their applications. **Neurocomputing**, Amsterdam, v. 234, p. 11-26, 2017.

LOZANO-GARCÍA, B.; PARRAS-ALCÁNTARA, L.; BREVIK, E. C. Impact of topographic aspect and vegetation (native and reforested areas) on soil organic carbon and nitrogen budgets in Mediterranean natural areas. **Science of the Total Environment**, Amsterdam, v. 544, p. 963-970, 2016.

LUCÀ, F. *et al.* Effect of calibration set size on prediction at local scale of soil carbon by Vis-NIR spectroscopy. **Geoderma**, Amsterdam, v. 288, p. 175–183, 2017.

MARGENOT, A. *et al.* Predicting soil permanganate oxidizable carbon (POXC) by coupling DRIFT spectroscopy and artificial neural networks (ANN). **Computers and Electronics in Agriculture**, Amsterdam, v. 168, [art.]. 105098, 2020.

MARTÍNEZ-CARRERAS, N. *et al.* A rapid spectral-reflectance-based fingerprinting approach for documenting suspended sediment sources during storm runoff events. **Journal of Soils and Sediments**, Landsberg, v. 10, n. 3, p. 400–413, 2010a.

MARTÍNEZ-CARRERAS, N. *et al.* The use of sediment colour measured by diffuse reflectance spectrometry to determine sediment sources: application to the Attert River catchment (Luxembourg). **Journal of Hydrology**, Amsterdam, v. 382, n. 1/4, p. 49–63, 2010b.

MCBRATNEY, A. B.; SANTOS, M. L. M.; MINASNY, B. On digital soil mapping. **Geoderma**, Amsterdam, v. 117, n. 1/2, p. 3-52, 2003.

MELO, V. F.; ALLEONI, L. R. F. **Química e mineralogia do solo**. Viçosa, MG: SBCS, 2019.

MEVIK, B.-H.; WEHRENS, R.; LILAND, K. H. **Partial least squares and principal component regression**. [S.l.]: Packages R CRAN, 2016.

MEYER, D. SVM - Support Vector Machines. *In*: MEYER, D. *et al.* **Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071)**: Package 'e1071': version 1.7-4. Wien: TU Wien, 14 Oct. 2020. p. 49-53. Disponível em: <https://cran.r-project.org/web/packages/e1071/e1071.pdf>. Acesso em: 26 nov. 2020.

MINASNY, B.; MCBRATNEY, A. B. A conditioned Latin hypercube method for sampling in the presence of ancillary information. **Computers and Geosciences**, New York, v. 32, n. 9, p. 1378–1388, 2006.

MINASNY, B.; MCBRATNEY, A. B. Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. **Chemometrics and Intelligent Laboratory Systems**, Amsterdam, v. 94, n. 1, p. 72–79, 2008.

MORELLOS, A. *et al.* Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. **Biosystems Engineering**, London, v. 152, p. 104-116, 2016.

MORENO-RAMÓN, H.; MARQUÉS-MATEU, Á.; IBÁÑEZ-ASENSIO, S. Significance of soil lightness versus physicochemical soil properties in semiarid areas. **Arid Land Research and Management**, London, v. 28, n. 4, p. 371–382, 2014.

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.

MUNSELL COLLOR. **Munsell soil color charts**. Baltimore: Munsell Color, 2000. Disponível em: <http://wallkillcolor.com/Munsell20/CMC%20Instruments.htm>. Acesso em: 3 fev. 2020.

MURTI, G. S. R. K.; SATYANARAYANA, K. V. S. Influence of chemical characteristics in the development of soil colour. **Geoderma**, Amsterdam, v. 5, n. 3, p. 243–248, 1971.

NANNI, M. R. *et al.* Optimum size in grid soil sampling for variable rate application in site-specific management. **Scientia Agricola**, Piracicaba, v. 68, n. 3, p. 386–392, 2011.

NANNI, M. R. *et al.* Partial least squares regression (PLSR) associated with

spectral response to predict soil attributes in transitional lithologies. **Archives of Agronomy and Soil Science**, Berlin, v. 64, n. 5, p. 682-695, 2018. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/03650340.2017.1373185?journalCode=gags20>. Acesso em: 5 fev. 2020.

NAWAR, S. *et al.* Estimating the soil clay content and organic matter by means of different calibration methods of vis-NIR diffuse reflectance spectroscopy. **Soil and Tillage Research**, Amsterdam v. 155, p. 510–522, 2016.

NAWAR, S.; MOUAZEN, A. M. Optimal sample selection for measurement of soil organic carbon using on-line vis-NIR spectroscopy. **Computers and Electronics in Agriculture**, Amsterdam, v. 151, p. 469–477, 2018.

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**, Oxford, v. 68, p. 337–347, 2014.

NOCITA, M. *et al.* Soil spectroscopy: an opportunity to be seized. **Global Change Biology**, Oxford, v. 21, n. 1, p. 10–11, 2015.

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

PABÓN, R. E. C.; SOUZA FILHO, C. R.; OLIVEIRA, W. J. Reflectance and imaging spectroscopy applied to detection of petroleum hydrocarbon pollution in bare soils. **Science of the Total Environment**, Amsterdam, v. 649, p. 1224–1236, 2019.

PANDOLFO, C. *et al.* **Atlas climatológico do Estado de Santa Catarina**. Florianópolis: Epagri, 2002. CD-Rom, v. 1, 13 p.

PENG, Y. *et al.* Modeling soil organic carbon at regional scale by combining multi-spectral images with laboratory spectra. **PLoS ONE**, São Francisco, v. 10, n. 11, [art.] e0142295, 2015.

PHOGAT, V. *et al.* Seasonal simulation of water, salinity and nitrate dynamics under drip irrigated mandarin (*Citrus reticulata*) and assessing management options for drainage and nitrate leaching. **Journal of Hydrology**, Amsterdam, v. 513, p. 504–516, 2014.

PINHEIRO, É. *et al.* Prediction of soil physical and chemical properties by visible and near-infrared diffuse reflectance spectroscopy in the central amazon. **Remote Sensing**, Basel, v. 9, n. 4, [art.] 293, 2017.

POST, D. F. *et al.* Relations between soil color and landsat reflectance on

semiarid rangelands. **Soil Science Society of America Journal**, Madison, v. 58, n. 6, p. 1809-1816, 1994.

R DEVELOPMENT CORE TEAM. **A language and environment for statistical computing**. [S.l.]: R Foundation for Statistical Computing, 2017. v. 2. Disponível em: <https://www.R-project.org>. Acesso em: 10 fev. 2020.

RAMIREZ-LOPEZ, L. *et al.* The spectrum-based learner: a new local approach for modeling soil vis-NIR spectra of complex datasets. **Geoderma**, Amsterdam, v. 195/196, p. 268–279, 2013.

RAMOS, P. V. *et al.* Color in subtropical brazilian soils as determined with a Munsell chart and by diffuse reflectance spectroscopy. **Catena**, Cremlingen, v. 193, [art.] 104609, 2020.

REZENDE, S. B. **Geomorphology, mineralogy and genesis of four soils on gneiss in southeastern Brazil**. Ann Arbor: University Microfilms International, 1980.

ROMERO, D. J. *et al.* Diffuse reflectance spectroscopy (Vis-Nir-Swir) as a promising tool for blue carbon quantification in mangrove soils: a case of study in tropical semiarid climatic conditions. **Soil Science Society of America Journal**, Madison, v. 81, n. 6, p. 1661-1667, 2017.

ROMERO, D. J. *et al.* Internal soil standard method for the Brazilian soil spectral library: performance and proximate analysis. **Geoderma**, Amsterdam, v. 312, p. 95–103, 2018.

SAHWAN, W. *et al.* Relationships between spectral features, iron oxides and colours of surface soils in northern Jordan. **European Journal of Soil Science**, Oxford, p. 1-18, 2020. Disponível em: <https://onlinelibrary.wiley.com/doi/epdf/10.1111/ejss.12986>. Acesso em: 12 fev. 2020.

SANTA CATARINA (Estado). Secretaria de Estado do Desenvolvimento Econômico Sustentável (SDS). **Levantamento aerofotogramétrico**. Florianópolis: SDS, 2013. Disponível em: <https://www.sc.gov.br/governo/acoesdegoverno/saneamento-e-recursos-hidricos/levantamento-aerofotogrametrico>. Acesso em: 14 de fevereiro de 2020.

SANTOS, H. G. *et al.* **Sistema brasileiro de classificação de solos**. 3. ed. rev. e ampl. Brasília, DF: Embrapa, 2013. 353 p.

SCHAETZL, R.; MILLER, B. A. Use of soil maps and surveys to interpret soil-landform assemblages and soil-landscape evolution. *In*: ZINCK, J. A. **Geopedology: an integration of geomorphology and pedology for soil and landscape studies**. Cham: Springer, 2016. cap. 15, p. 251-264.

SCHANDA, J. **Colorimetry: understanding the CIE system**. Hoboken: John Wiley, 2007.

SEIDEL, M. *et al.* Strategies for the efficient estimation of soil organic carbon at the field scale with vis-NIR spectroscopy: spectral libraries and spiking vs. local calibrations. **Geoderma**, Amsterdam, v. 354, [art.] 113856, 2019.

SHAHRAYINI, E. *et al.* Multiple-depth modeling of soil organic carbon using visible–near infrared spectroscopy. **Geocarto International**, Abingdon, [p. 1-18], May 2020. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/10106049.2020.1765887>. Acesso em: 25 maio 2020.

SHERMAN, D. M.; WAITE, T. D. Electronic spectra of Fe<sup>3+</sup> oxides and oxide hydroxides in the near IR to near UV. **American Mineralogist**, Berlin, v. 70, n. 11/12, p. 1262–1269, 1985.

SHI, Z. *et al.* Development of a national VNIR soil-spectral library for soil classification and prediction of organic matter concentrations. **Science China Earth Sciences**, Heidelberg, v. 57, n. 7, p. 1671–1680, 2014.

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

SIEWERT, M. B. High-resolution digital mapping of soil organic carbon in permafrost terrain using machine learning: a case study in a sub-Arctic peatland environment. **Biogeosciences**, Katlenburg-Lindau, v. 15, n. 6, p. 1663–1682, 2018.

SILVA, E. B. *et al.* Estimating soil texture from a limited region of the Visible/Near-Infrared Spectrum. *In*: HARTEMINK, A. E.; MINASNY, B. (ed.). **Digital soil morphometrics**. Cham: Springer International Publishing, 2016. (Progress in Soil Science). p. 73–87.

SILVA, E. B. *et al.* A regional legacy soil dataset for prediction of sand and clay content with VIS-NIR-SWIR, in southern Brazil. **Revista Brasileira de Ciência do Solo**, Viçosa, MG, v. 43, [art.] e0180174, 2019.

SIMON, T. *et al.* Predicting the color of sandy soils from Wisconsin, USA. **Geoderma**, Amsterdam, v. 361, [art.] 114039, 2020.

SOIL SURVEY DIVISION STAFF. Examination and description of soil profiles. *In*: SOIL SURVEY DIVISION STAFF. **Soil survey manual**. 4th ed. Washington, DC: United States Department of Agriculture, 2017. (United States Department of Agriculture. Handbook, no. 18). cap. 3, p. 83-233.

SOUSA JUNIOR, J. G.; DEMATTÊ, J. A. M.; ARAÚJO, S. R. Modelos espectrais terrestres e orbitais na determinação de teores de atributos dos solos: potencial e custos. **Bragantia**, Campinas, v. 70, n. 3, p. 610-621, 2011.

SPIELVOGEL, S.; KNICKER, H.; KÖGEL-KNABNER, I. Soil organic matter

composition and soil lightness. **Journal of Plant Nutrition and Soil Science**, Weinheim, v. 167, n. 5, p. 545–555, 2004.

STENBERG, B. *et al.* Visible and near infrared spectroscopy in soil science. **Advances in Agronomy**, San Diego, v. 107, p. 163-215, 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, [art.] e66409, 2013.

STEVENS, A.; RAMIREZ-LOPEZ, L. **An introduction to the prospectr package**. [S.l.: R Package Vignette], 14 Aug. 2013. 22 p. Disponível em: [https://www.researchgate.net/publication/255941339\\_An\\_introduction\\_to\\_the\\_prospectr\\_package](https://www.researchgate.net/publication/255941339_An_introduction_to_the_prospectr_package). Acesso em: 8 dez. 2020.

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.

TEN CATEN, A. *et al.* Mapeamento digital de classes de solos: características da abordagem brasileira. **Ciência Rural**, Santa Maria, v. 42, n. 11, p. 1989-1997, 2012.

TERRA, F. S.; DEMATTÊ, J. A. M.; VISCARRA ROSSEL, R. A. Proximal spectral sensing in pedological assessments: vis–NIR spectra for soil classification based on weathering and pedogenesis. **Geoderma**, Amsterdam, v. 318, p. 123–136, 2018.

THENKABAIL, A. *et al.* Characterization of soil properties using reflectance spectroscopy. *In*: THENKABAIL, P. S. **Hyperspectral remote sensing of vegetation**. Boca Raton: Taylor & Francis, 2011. p. 513–558.

TIECHER, T. *et al.* Combining visible-based-color parameters and geochemical tracers to improve sediment source discrimination and apportionment. **Science of the Total Environment**, Amsterdam, v. 527–528, p. 135–149, 2015.

TÜMSAVAŞ, Z. *et al.* Prediction and mapping of soil clay and sand contents using visible and near-infrared spectroscopy. **Biosystems Engineering**, London, v. 177, p. 90–100, 2019.

UTTLEY, J. Power analysis, sample size, and assessment of statistical assumptions—Improving the evidential value of lighting research. **LEUKOS: the Journal of Illuminating Engineering Society of North America**, New York, v. 15, n. 2/3, p. 143-162, 2019.

VASAVA, H. B. *et al.* Assessment of soil texture from spectral reflectance data of bulk soil samples and their dry-sieved aggregate size fractions. **Geoderma**, Amsterdam, v. 337, p. 914–926, 2019.

VASQUES, G. M.; GRUNWALD, S.; HARRIS, W. G. Spectroscopic models of

soil organic carbon in Florida, USA. **Journal of Environmental Quality**, Madison, v. 39, n. 3, p. 923–934, 2010.

VASQUES, 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.

VEIGA, M. *et al.* Distribuição espacial dos teores de argila, silte e areia na camada superficial do solo em Santa Catarina. **Agropecuária Catarinense**, Florianópolis, v. 25, n. 1, p. 63–68, 2012.

VENABLES, W. N.; RIPLEY, B. D. **Modern applied statistics with S**. 4th ed. New York: Springer, 2002. v. 53.

VENDRAME, P. R. S. *et al.* The potential of NIR spectroscopy to predict soil texture and mineralogy in Cerrado Latosols. **European Journal of Soil Science**, Oxford, v. 63, n. 5, p. 743–753, 2012.

VIANNA, L. F. N. *et al.* Aplicação de descritores de heterogeneidade ambiental na seleção de áreas para sistemas de parcelas amostrais: um estudo de caso para a determinação de rotspots potenciais de biodiversidade. **Geografia**, Rio Claro, v. 40, n. 2, p. 211–239, 2015.

VISCARRA ROSSEL, R. A. *et al.* Colour space models for soil science. **Geoderma**, Amsterdam, v. 133, n. 3/4, p. 320–337, 2006a.

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, 2006b.

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

VISCARRA ROSSEL, R. A. *et al.* Baseline map of organic carbon in Australian soil to support national carbon accounting and monitoring under climate change. **Global Change Biology**, Oxford, v. 20, n. 9, p. 2953–2970, 2014.

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

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

VISCARRA ROSSEL, R. A.; CHEN, C. Digitally mapping the information content of visible-near infrared spectra of surficial Australian soils. **Remote Sensing of Environment**, New York, v. 115, n. 6, p. 1443–1455, 2011.

VISCARRA ROSSEL, R. A.; FOUAD, Y.; WALTER, C. Using a digital camera to measure soil organic carbon and iron contents. **Biosystems Engineering**, London, v. 100, n. 2, p. 149–159, 2008.

VISCARRA ROSSEL, R. A.; WALTER, C.; FOUAD, Y. Assessment of two reflectance techniques for the quantification of the within-field spatial variability of soil organic carbon. *In*: EUROPEAN CONFERENCE ON PRECISION AGRICULTURE, 4., 2003, Berlin, Germany. **Proceedings** [...]. Wageningen: Wageningen Academic, 2003. p. 697–703.

VISCARRA ROSSEL, R. A.; WEBSTER, R. Predicting soil properties from the Australian soil visible-near infrared spectroscopic database. **European Journal of Soil Science**, Oxford, v. 63, n. 6, p. 848–860, 2012.

VODYANITSKII, Y. N.; SAVICHEV, A. T. The influence of organic matter on soil color using the regression equations of optical parameters in the system CIE-L\*a\*b\*. **Annals of Agrarian Science**, Tbilisi, v. 15, n. 3, p. 380–385, 2017.

WALLKILL COLOR. **Munsell conversion software**. [S.l.]: Wallkill Color, [2019]. Disponível em: <http://wallkillcolor.com/>. Acesso em: 18 out. 2019.

WANG, J. *et al.* Mapping soil organic matter based on land degradation spectral response units using Hyperion images. **International Journal of Applied Earth Observation and Geoinformation**, Washington, DC, v. 12, p. S171-S180, 2010. Supl. 2.

WANG, P. *et al.* Spectral exploration of calcium accumulation in organic matter in gray desert soil from northwest China. **PLoS ONE**, San Francisco, v. 11, n. 1, [art.] e0145054, 2016.

WANG, X. *et al.* Estimation of soil salt content (SSC) in the Ebinur Lake Wetland National Nature Reserve (ELWNNR), Northwest China, based on a Bootstrap-BP neural network model and optimal spectral indices. **Science of the Total Environment**, Amsterdam, v. 615, p. 918–930, 2018.

WERE, K. *et al.* A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afrotropical landscape. **Ecological Indicators**, New York, v. 52, p. 394-403, 2015.

WETTERLIND, J.; STENBERG, B. Near-infrared spectroscopy for within-field soil characterization: small local calibrations compared with national libraries spiked with local samples. **European Journal of Soil Science**, Oxford, v. 61, n. 6, p. 823-843, 2010.

WIGHT, J. P.; ASHWORTH, A. J.; ALLEN, F. L. Organic substrate, clay type, texture, and water influence on NIR carbon measurements. **Geoderma**, Amsterdam, v. 261, p. 36-43, 2016.

WIJEWARDANE, N. K.; GE, Y.; MORGAN, C. L. S. Moisture insensitive

prediction of soil properties from VNIR reflectance spectra based on external parameter orthogonalization. **Geoderma**, Amsterdam, v. 267, p. 92-101, 2016.

WILDNER, W. *et al.* **Mapa geológico do Estado de Santa Catarina**. Porto Alegre: CPRM, 2014. 1 mapa. Escala 1:500.000. Disponível em: <http://geobank.cprm.gov.br/>. Acesso em: 6 jul. 2020.

WILLIAMS, C. K. I.; BARBER, D. Bayesian classification with gaussian processes. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, New York, v. 20, n. 12, p. 1342 -1351, 1998.

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.

WORKMAN, J. Jr.; WEYER, L. **Practical guide to interpretive near-infrared spectroscopy**. Boca Raton: CRC Press, 2007.

WORKMAN, J.; SHENK, J. Understanding and using the near-infrared spectrum as an analytical method. *In*: CRAIG, A. R.; WORKMAN, J. Jr.; REEVES, J. B. (ed.). **Near-infrared spectroscopy in agriculture**. Madison: American Society of Agronomy, 2004. (Agronomy, 44). cap. 1, p. 1-10.

WU, K. L.; YANG, M. S. A cluster validity index for fuzzy clustering. **Pattern Recognition Letters**, Amsterdam, v. 26, n. 9, p. 1275–1291, 2005.

XU, D. *et al.* Assessment of important soil properties related to chinese soil taxonomy based on vis–NIR reflectance spectroscopy. **Computers and Electronics in Agriculture**, Amsterdam, v. 144, p. 1–8, 2018.

YANG, J. *et al.* Combination of convolutional neural networks and recurrent neural networks for predicting soil properties using Vis–NIR spectroscopy. **Geoderma**, Amsterdam, v. 380, [art.] 114616, 2020a.

YANG, Z. *et al.* Fast determination of oxides content in cement raw meal using NIR spectroscopy combined with synergy interval partial least square and different preprocessing methods. **Measurement: Journal of the International Measurement Confederation**, Amsterdam, v. 149, [art.] 106990, 2020b.

ZERAATPISHEH, M. *et al.* Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. **Geoderma**, Amsterdam, v. 338, p. 445-452, 2019.

ZHANG, D. *et al.* Quantitative estimating salt content of saline soil using laboratory hyperspectral data treated by fractional derivative. **Journal of Spectroscopy**, New York, v. 2016, [art.] ID 1081674, 2016.

ZHANG, Y. *et al.* Depth-specific prediction of soil properties in situ using vis-nir spectroscopy. **Soil Science Society of America Journal**, Madison, v. 81, n. 5, p. 993–1004, 2017.

ZHANG, Z. *et al.* Prediction of soil organic matter in northwestern China using fractional-order derivative spectroscopy and modified normalized difference indices. **Catena**, Cremlingen, v. 185, [art.] 104257, 2020.

ZHAO, D. *et al.* A Vis-NIR spectral library to predict clay in Australian cotton growing soil. **Soil Science Society of America Journal**, Madison, v. 82, n. 6, p. 1347–1357, 2018.

ZHAO, S. J. *et al.* Nonlinear projection to latent structures method and its applications. **Industrial and Engineering Chemistry Research**, Washington, DC, v. 45, n. 11, p. 3843–3852, 2006.

ZINCK, J. A. *et al.* (ed.). **Geopedology**: an integration of geomorphology and pedology for soil and landscape studies. Cham: Springer, 2016.

ZOBECK, T. M. *et al.* Soil property effects on wind erosion of organic soils. **Aeolian Research**, Amsterdam, v. 10, p. 43–51, 2013.