UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE CIÊNCIAS RURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DO SOLO

Jean Michel Moura-Bueno

USO DE BIBLIOTECAS ESPECTRAIS PARA A PREDIÇÃO DO CARBONO ORGÂNICO DO SOLO

Santa Maria, RS, Brasil
2018

Jean Michel Moura-Bueno

USO DE BIBLIOTECAS ESPECTRAIS PARA A PREDIÇÃO DO CARBONO ORGÂNICO DO SOLO

Tese apresentada ao Curso de Doutorado do Programa de Pós-Graduação em Ciência do Solo da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Doutor em Ciência do Solo.**

Orientador: Prof. Ricardo Simão Diniz Dalmolin

Santa Maria, RS, Brasil
2018

Moura-Bueno, Jean Michel
USO DE BIBLIOTECAS ESPECTRAIS PARA A PREDIÇÃO DO CARBONO ORGÂNICO DO SOLO / Jean Michel Moura-Bueno.- 2018.
201 p.; 30 cm

Orientador: Ricardo Simão Diniz Damolin
Coorientador: Alexandre ten Caten
Tese (doutorado) - Universidade Federal de Santa Maria, Centro de Ciências Rurais, Programa de Pós Graduação em Ciência do Solo, RS, 2018

1. sensoriamento remoto hiperespectral 2. predição de carbono orgânico 3. processamento espectral 4. modelos de aprendizado 5. estratificação de dados espectrais I. Damolin, Ricardo Simão Diniz II. ten Caten, Alexandre III. Título.

Sistema de geração automática de ficha catalográfica da UFSM. Dados fornecidos pelo autor(a). Sob supervisão da Direção da Divisão de Processos Técnicos da Biblioteca Central. Bibliotecária responsável Paula Schoenfeldt Patta CRB 10/1728.

© 2018
Todos os direitos autorais reservados a Jean Michel Moura-Bueno. A reprodução de partes ou do todo deste trabalho só poderá ser feito mediante a citação da fonte.
E-mail: bueno.jean1@gmail.com

Jean Michel Moura-Bueno

USO DE BIBLIOTECAS ESPECTRAIS PARA A PREDIÇÃO DO CARBONO ORGÂNICO DO SOLO

Tese apresentada ao Curso de Doutorado do Programa de Pós-Graduação em Ciência do Solo da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Doutor em Ciência do Solo.**

Aprovada em 21 de maio de 2018:

_____

Ricardo Simão Diniz Dalmolin, Dr. (UFSM)
(Presidente/Orientador)

_____

Fabrício de Araújo Pedron, Dr. (UFSM)

_____

Gustavo de Mattos Vasques, Dr. (Embrapa Solos)

_____

Pablo Miguel, Dr.(UFPel)

_____

Ricardo Bergamo Schenato, Dr. (UFSM)

Santa Maria, RS, Brasil
2018

AGRADECIMENTOS

A Deus, que sempre esteve em minha vida, guiando meus passos e meu caminho, e me dando forças para que chegasse até o fim.

Aos meus pais Adair Sidnei Bueno e Heidi Selent, por me ensinarem que o estudo é sempre o melhor caminho. Obrigado por acreditarem em mim.

Ao meu irmão Charles Bueno por sempre estar ao meu lado e dividir as alegrias e sonhos.

A minha esposa, amiga e companheira Fernanda Pascotini, pela parceria, pela paciência, pelo apoio e por toda ajuda durante minha caminha na pós-graduação.

Aos meus sogros, Luis Henrique e Lilian, por sempre torcerem e acreditarem em mim.

Ao meu orientador e amigo, professor Ricardo Dalmolin, pela confiança depositada durante os 11 anos de convivência no Setor de Pedologia do Departamento de Solos da UFSM. Tenha certeza que você contribuiu muito para que eu chegasse até aqui.

Ao meu co-orientador, Prof. Alexandre ten Caten, pelos momentos de discussão e pelo auxílio, sempre que precisei.

Aos membros da banca, Fabrício Pedron, Gustavo Vasques, Pablo Miguel e Ricardo Schenato pelas contribuições sugeridas no trabalho.

Aos colegas e amigos do grupo de Pedologia, Luciano Cancian, André Dotto, Estelita Penteado (*in memoriam*), Nícolas Rosin, João Pedro Flores, Diego Gris, Luíza Kreimeier, Jéssica da Costa, Ismael Backes e Mario Wolski pela convivência e ajuda, e em especial, a colega Taciara Horst pelas contribuições e ajuda na construção do trabalho.

Ao pesquisador e amigo Alessandro Samuel-Rosa pelos ensinamentos, troca de ideias e conselhos.

Ao pesquisador Carlos Alberto Flores pelos ensinamentos, troca de ideias e conselhos.

Ao proprietário de área de estudo de Giruá, Darcy Urhy, pela disponibilidade da área.

Aos amigos da turma 80 da Agronomia da UFSM por estarem ao meu lado nos momentos de descontração.

À UFSM e ao Programa de Pós-Graduação em Ciência do Solo pela oportunidade de realização do doutorado.

Ao CNPq pela concessão da bolsa de estudos.

E agradeço a todas as pessoas que, de uma maneira ou de outra, contribuíram para a realização desse trabalho. **MUITO OBRIGADO**!

RESUMO

USO DE BIBLIOTECAS ESPECTRAIS PARA A PREDIÇÃO DE CARBONO ORGÂNICO DO SOLO

Autor: Jean Michel Moura-Bueno
Orientador: Ricardo Simão Diniz Dalmolin

Este estudo apresenta o uso da técnica de espectroscopia de reflectância difusa (ERD) aplicada à predição de carbono orgânico do solo (COS). O escopo do trabalho tem relação com o aumento da demanda por informações de solo para auxiliar no monitoramento ambiental, produção agrícola e abastecer modelos de simulação de cenários futuros. A ERD associada a bibliotecas espectrais de solo (BES) é uma alternativa para a quantificação do conteúdo de COS. No entanto, a capacidade preditiva dos modelos está atrelada às características dos dados espectrais do solo. Assim, torna-se essencial avaliar a construção de modelos quando a BES é composta por amostras com grande variação nas características físicas, químicas e mineralógicas, como é o caso dos solos da região sul do Brasil. Pois a acurácia dos modelos é definida pela complexidade e abrangência das características pedológicas dos solos representados na BES. Assim, usando uma BES da região sul do Brasil, os objetivos deste estudo foram: i) avaliar o efeito de técnicas de pré-processamentos, métodos multivariados e estratificação amostral de BES na acurácia dos modelos; ii) identificar a influência da complexidade espectral das amostras na qualidade das predições; iii) definir as características mínimas da BES que possam impactar em seu poder preditivo. No ESTUDO 1, foi utilizada uma BES de escala local, composta por 841 amostras. A partir dela testou-se a influência de técnicas de pré-processamentos, métodos multivariados e estratificação da BES com base na variância espectral, classe de solo, uso da terra e profundidade das amostras de solo na predição de COS. No ESTUDO 2, foi utilizada uma BES de escala regional, composta por 2599 amostras. A partir dela, foi avaliado o efeito da estratificação da BES com base em características ambientais regionais, classe textural do solo, usos da terra e classes espectrais na acurácia dos modelos. Os espectros de solo apresentaram alta variação espectral, decorrente da variação pedológica e ambiental da região sul do Brasil. Modelos mais acurados foram alcançados com a técnica de processamento espectral *Savitzky-Golay* - 1ª derivada associada ao método calibração de regressão por mínimos quadrados parciais e com o método *Cubist* com processamento remoção do contínuo. A estratificação de BES com base em características do solo e do ambiente (regiões fisiográficas) revelou que o agrupamento de amostras mais homogêneas, principalmente em relação a regiões fisiográficas e usos da terra aumentou a acurácia das predições. A redução no número de amostras devido à estratificação afetou negativamente o desempenho dos modelos, principalmente para grupos com alta variação pedológica e espectral. As predições de COS apresentaram menor acurácia para amostras com textura grossa (areia > 15% e argila < 35%). Os resultados reiteram que o uso de BES para a predição de COS exige um estudo prévio sobre a variância dos dados.

Palavras-chave: sensoriamento remoto hiperespectral, predição de carbono orgânico, processamento espectral, modelos de aprendizado, estratificação de dados espectrais.

ABSTRACT

USE OF SPECTRAL LIBRARIES FOR SOIL ORGANIC CARBON PREDICTION

AUTHOR: Jean Michel Moura-Bueno
ADVISOR: Ricardo Simão Diniz Dalmolin

This study demonstrates the use of diffuse reflectance spectroscopy (DRS) applied to soil organic carbon (SOC) prediction. The scope of the study is related to the increase in the demand for soil information to support environmental monitoring, agricultural production and to feed models for simulation of future scenarios. DRS associated to soil spectral libraries (SSL) is an alternative for quantification of SOC content. However, the predictive ability of the models is linked to the characteristics of soil spectral data. Therefore, it becomes essential to evaluate the construction of models when the SSL is composed by samples with high variability in physical, chemical and morphological characteristics, which is the case for soils in the south of Brazil. Considering that the accuracy of the models is defined by the complexity and coverage of pedological features of soils represented in the SSL, associated to techniques of spectral processing and multivariate methods, the objectives of this study were: i) evaluate the effect of preprocessing techniques, multivariate methods and sample stratification of SSL on the accuracy of the models; ii) identify the influence of the spectral complexity of the samples on the quality of predictions; iii) define the minimum features of the SSL that might impact their predictive ability. In STUDY 1, a local SSL composed of 841 samples was used. With this SSL, the influence of preprocessing techniques, multivariate methods and SSL stratification based on spectral variance, soil class, land use and soil sample depth on SOC prediction was assessed. In STUDY 2, a regional-scale SSL composed of 2,599 samples was used. With this SSL, the effect of stratification of the SSL based on regional environmental features, soil texture, land use and spectral class on the accuracy of models was evaluated. Soil spectra presented high spectral variation, due to the pedological and environmental variation of the southern region of Brazil. More accurate models were obtained with the Savitzky-Golay spectral processing technique - 1st derivative associated with the partial least squares regression calibration method, and with the Cubist method with continuous removal processing. The stratification of the SSL based on soil and environmental (physiographic regions) characteristics showed that grouping more homogeneous samples, especially in relation to physiographic regions and land uses, increased the accuracy of the predictions. The reduced number of samples due to stratification negatively affected the performance of the models, especially for groups with high pedological and spectral variation. SOC predictions presented lower accuracy for samples with coarse texture (sand > 15 % and clay < 35 %). The results confirm that the use of SSL for SOC prediction requires a previous study on the data variance.

Keywords: hyperspectral remote sensing, prediction of organic carbon, spectral processing, learning model, stratification of spectral dataset.

LISTA DE FIGURAS

REVISÃO BIBLIOGRÁFICA

Figura 1 – Diagrama mostrando a interdisciplinaridade da pedometria. 36
Figura 2 – Representação do modelo *scorpan* com os planos de informação espacial. 37
Figura 3 – Representação das etapas envolvidas na integração da técnica de ERD e MDS para obtenção de informações de solos. 38
Figura 4 – Ilustração mostrando uso da técnica de ERD em solos na região do Vis-NIR (350-2500 nm) do espectro eletromagnético. 42
Figura 5 – Representação das etapas envolvidas na quantificação de propriedade do solo pela técnica de ERD. 44
Figura 6 – Representação das diferentes formas de interação organomineral do solo conforme a estrutura dos argilominerais. 46
Figura 7 – Curva espectral de uma amostra de Latossolo – horizonte Bw (curva vermelha) e Argissolo – horizonte Bt (curva azul) indicando a principais regiões do espectro Vis-NIR influenciadas pelos constituintes do solo. Em que: MO – matéria orgânica; Gt – goethita; Hm – hematita; Gb – gibbsita. 49
Figura 8 – Curvas espectrais mostrando o comportamento espectral do solo de amostras mais argilosas e menos argilosas da região sul do Brasil. 50
Figura 9 – Curvas espectrais mostrando a influencia da mineralogia no comportamento espectral do solo em amostras da região sul do Brasil. Hematita (Hm); Goethita (Gt); Caulinita (K). 52
Figura 10 – Curvas espectrais mostrando a influencia da umidade no comportamento espectral do solo em amostras da região sul do Brasil. 53
Figura 11 – Curvas espectrais mostrando a influencia da matéria orgânica (MO) no comportamento espectral do solo em amostras da região sul do Brasil. 54

ESTUDO 1

Figura 12 – Localização da área de estudo no município de Giruá, Estado do Rio Grande do Sul, Brasil (a) e área ampliada com imagem Google Earth (ano de 2016) com os pontos de amostragem (b). 69
Figura 13 – Diagrama esquemático das etapas executadas na metodologia. CTA: conjunto total de amostras; SMO: smoothed utilizando uma janela móvel de 9 nm; CRR: continuum removed reflectance; SNV:standard normal variate; DET: Detrend; SGD Savitzky-Golay com 1ª derivada usando um polinômio de 1ªordem com janela de busca de 9 nm; PLSR: regressão por mínimos quadrados parciais; MLR: regressão linear múltipla; SVM: máquina de vetor de suporte; RF: floresta aleatória. 76
Figura 14 – Triângulo textural indicando a classe textural dos Latossolos (pontos em vermelho) e Gleissolos (pontos em preto). 77
Figura 15 – Estatística descritiva e histogramas dos teores de COS do CTA, Latossolo Vermelho – LV, Gleissolo Háplico – GX, Lavoura – LA, Campo nativo – CA, Floresta nativa – FL, camada de 0,0 – 0,05 m (C1); 0,05-0,15 m (C2); 0,15-0,30 m (C3); 0,30-0,60 m (C4); 0,60-1,00 m. 79
Figura 16 – Curvas espectrais representativas da BEL, amostras de Latossolo Vermelho – LV e Gleissolo Háplico – GX, nos usos da terra Lavoura – LA, Campo nativo – CA e Floresta nativa – FL, com identificação das características das amostras. 81

Figura 17 – Análise de componentes principais (a) CTA e grupos: (b) classes de solo Latossolo Vermelho – LV e Gleissolo Háplico – GX, (c) usos da terra Lavoura – LA, Campo nativo – CA e Floresta nativa – FL e (d) camadas amostradas C1 (0,0 – 0,05 m); C2 (0,05-0,15 m); C3 (0,15-0,30 m); C4 (0,30-0,60 m); C5 (0,60-1,00 m). 83
Figura 18 – Espectros da BEL após a aplicação dos pré-processamentos. (a) SMO: suavização utilizando uma janela móvel de 9 nm; (b) CRR: remoção do contínuo; (c) SNV: padrão normal variado; (d) DET: Detrend; (e) SGD: Savitzky-Golay com 1ª derivada usando um polinômio 1ªordem com janela de busca de 9 nm. 86
Figura 19 – Desempenho dos modelos de predição em relação aos valores de RMSEv para os quatro modelos multivariados e seis técnicas de pré-processamento. 87
Figura 20 – Melhores desempenhos dos modelos de predição para o COS com o CTA. PLSR: regressão por mínimos quadrados parciais; MLR: regressão linear múltipla; SVM: máquina de vetor de suporte; RF: floresta aleatória. 89
Figura 21 – Importância das faixas espectrais na predição do COS (a) e pesos (*loadings*) das componentes PLS1 e PLS2 obtidas pelo modelo PLSR (b). Ressaltado com setas e em cinza as faixas espectrais consideradas importantes na predição do COS pelo modelo. 91

ESTUDO 2

Figura 22 – Localização na região sul do Brasil dos bancos de dados da BE e mapa de regiões (a), clima (b), material parental (c), uso da terra (d), classes de solo (f) e classes texturais (e). 102
Figura 23 – Estatística descritiva e histogramas dos teores de COS, argila, silte e areia e triângulo textural do BD1. 103
Figura 24 – Estatística descritiva e histogramas dos teores de COS, argila, silte e areia e triângulo textural do BD2. 105
Figura 25 – Estatística descritiva e histogramas dos teores de COS, argila, silte e areia e triângulo textural do BD3. 106
Figura 26 – Estatística descritiva e histogramas dos teores de COS, argila, silte e areia e triângulo textural do BD4. 108
Figura 27 – Estatística descritiva e histogramas dos teores de COS, argila, silte e areia e triângulo textural do BD5. 109
Figura 28 – Estatística descritiva e histogramas dos teores de COS, argila, silte e areia triângulo textural do BD6. 111
Figura 29 – Estatística descritiva e histogramas dos teores de COS, argila, silte e areia e triângulo textural do BD7. 112
Figura 30 – Diagrama esquemático das etapas executadas na metodologia. 116
Figura 31 – Estatística descritiva, histograma e boxplots dos teores de COS da BEG (a), conjunto externo de validação – CEV (b) e sub-bibliotecas por regiões fisiográficas (c), classes texturais (d), usos da terra (e), classes espectrais (f). 120
Figura 32 – Triângulo textural com a identificação das amostras das bibliotecas espectrais agrupadas por região fisiográficas (a), classe textural (b), uso da terra (c) e classe espectral (d). 125
Figura 33 – *Scores* do CP1, CP2 e CP3 derivados da ACP dos dados espectrais da BEG. 126
Figura 34 – *Loadings* derivados da ACP os quais representam as principais variações espectrais causadas pelas variações na composição do solo nas diferentes bandas. 127

Figura 35 – Análise de componentes principais e classificação de *fuzzy-c-means*: (a) três primeiros componentes (CP) da ACP, (b, i – iii) pontuações dos CP para coloridas para as quatro classes espectrais geradas pela *fuzzy-c-means*, (c, e, g, i) média dos espectros e desvios padrão para as quatro classes espectrais com cores correspondentes às classes mostradas em (b), (d, f, h, j, l, n, i-iii) são as funções de pertinência para cada classe espectral. 130
Figura 36 – *Scores* do CP1 e CP2 derivados da ACP dos dados espectrais com identificação das amostras (a) regiões fisiográficas, (b) classe textural e (c) usos da terra. 131
Figura 37 – Curvas espectrais da BEG de solos, com destaque para curvas representativas das três regiões com diferentes materiais parentais e composição organomineral..... 133
Figura 38 – Espectros médios e seus desvios-padrão do conjunto global de solos, regiões fisiográficas, classes texturais e usos da terra..... 135
Figura 39 – Espectros da BEG após a aplicação dos seis pré-processamentos. (a) SMO: suavização utilizando uma janela móvel de 11 nm; (b) CRR: remoção do contínuo; (c) SNV: correção pelo padrão normal variado; (d) MSC: *multiplicative scatter correction* - correção de dispersão; (e) SGD: Savitzky-Golay com 1ª derivada usando um polinômio de 1ªordem com janela de busca de 11 nm; (f) DET: Detrend - redução de tendência..... 140
Figura 40 – Importância das variáveis explicativas (VARIMP) na predição do COS para o modelo global Cubist-CRR. 142
Figura 41 – *Scatter plots* da calibração e validação do teor de COS mensurado em laboratório versus o predito pelo modelo global Cubist-CRR com identificação dos solos em (a) regiões fisiográficas, (b) classes texturais (c) usos da terra. 144
Figura 42 – Importância das variáveis explicativas na predição do COS dos modelos regionais (a) modelo região R1 – MR1, (b) modelo região R2 – MR2 e (c) modelo região R3 – MR3..... 149
Figura 43 – *Scatter plots* da validação do teor de COS medido em laboratório versus o predito com método Cubist-CRR com identificação das classes texturais e usos da terra para os modelos regionais (a) modelo região R1 – MR1, (b) modelo região R2 – MR2 e (c) modelo região R3 – 151
Figura 44 – Importância das variáveis explicativas na predição do COS dos modelos texturais (a) modelo textura muito argilosa – MTMA, (b) modelo textura argilosa – MTA, (c) modelo textura média – MM e (d) modelo textura arenosa – MTAR..... 154
Figura 45 – *Scatter plots* da validação do teor de COS medido em laboratório versus o predito com método Cubist-CRR com identificação das regiões fisiográficas e usos da terra para os modelos texturais (a) modelo textura muito argilosa – MTMA, (b) modelo textura argilosa – MTA, (c) modelo textura média – MTM e (d) modelo textura arenosa – MTAR..... 155
Figura 46 – Importância das variáveis explicativas na predição do COS dos modelos usos da terra (a) modelo lavoura – MLA, (b) modelo floresta – MFL e (c) modelo campo – MCA..... 159
Figura 47 – *Scatter plots* da validação do teor de COS medido em laboratório versus o predito com método Cubist-CRR com identificação das regiões fisiográficas e classes texturais para os modelos uso da terra (a) modelo lavoura – MLA, (b) modelo floresta – MFL e (c) modelo campo – MCA..... 160
Figura 48 – Importância das variáveis explicativas na predição do COS dos modelos espectrais (a) modelo espectral classe 1 – MC1, (b) modelo espectral classe 2 – MC2, (c) modelo espectral classe 3 – MC3 e (d) modelo espectral classe 4 – MC4..... 163
Figura 49 – *Scatter plots* da validação do teor de COS medido em laboratório versus o predito com método Cubist-CRR com identificação das regiões fisiográficas, classes texturais e usos da terra para os modelos espectrais (a) modelo espectral classe 1 – MC1, (b) modelo espectral classe 2 – MC2, (c) modelo espectral classe 3 – MC3 e (d) modelo espectral classe 4 – MC4..... 165

Figura 50 – *Scatter plots* da validação (n = 738) mostrando o teor de COS medido em laboratório versus o predito (a) modelo global – MG, (b) modelo regional – MR, (c) modelos texturais – MT, (d) modelos usos da terra – MUT, (e) modelos espectrais – ME. 167

LISTA DE TABELAS

REVISÃO BIBLIOGRÁFICA
Tabela 1 – Constituintes do solo, grupos químicos e atribuições das absorções fundamentais na região do infravermelho médio (MIR) com as sobreposições de absorções e suas combinações na região Vis-NIR. Compilação adaptada de Viscarra Rossel e Behrens (2010). 48

ESTUDO 1
Tabela 2 – Compilação de trabalhos de predição de COS com diferentes métodos multivariados, técnicas de pré-processamento e escalas de bibliotecas espectrais. 67
Tabela 3 – Resumo estatístico dos modelos geradas para os quatros métodos multivariados e seis pré-processamentos para o CTA (n = 810)..... 85
Tabela 4 – Grupos de amostras formados após a estratificação da BEL e descrição das amostras que compõem cada grupo. 95
Tabela 5 – Resumo estatístico dos modelos gerados para cada grupo pelo modelo PLSR e pré-processamento SGD. 95

ESTUDO 2
Tabela 6 – Matriz correlação entre os *scores* da ACP e o conteúdo de COS, argila, silte e areia da BEG (n = 2461). 127
Tabela 7 – Número de classes e índices de coeficiente de partição e entropia de partição..... 128
Tabela 8 – Resultados da predição do COS com o modelo global (MG) gerado com a combinação de diferentes métodos multivariados e técnicas de pré-processamento espectral. 137
Tabela 9 – Resultados de calibração e validação para predição do COS com os modelos regionais (MR) gerados com diferentes abordagens de métodos..... 146
Tabela 10 – Resultados de calibração e validação para predição do COS com os modelos texturais (MT) gerados com diferentes abordagens de métodos. 153
Tabela 11 – Resultados de calibração e validação para predição do COS com os modelos usos da terra (MUT) gerados com diferentes abordagens de métodos..... 157
Tabela 12 – Resultados de calibração e validação para predição do COS com os modelos espectrais (ME) gerados com diferentes abordagens de métodos. 162
Tabela 13 – Resumo estatístico da acurácia da predição do COS para o conjunto externo de validação (CEV) pelo modelo global (MG), modelos regionais (MR) e modelos usos da terra (MUT)..... 169

SUMÁRIO

INTRODUÇÃO GERAL 27
1. JUSTIFICATICA E RELEVÂNCIA 29
2. REVISÃO BIBLIGRÁFICA 33
2.1. PEDOMETRIA 33
2.2. MÉTODOS DE DETERMINAÇÃO DO CONTEÚDO DE COS 39
2.2.1. Oxidação por via úmida 40
2.2.2. Oxidação por via seca 41
2.3. ESPECTROSCOPIA DE REFLECTÂNCIA DIFUSA DO SOLO 42
2.4. COMPORTAMENTO ESPECTRAL DO SOLO 45
2.4.1. Efeito da granulometria 49
2.4.2. Efeito da mineralogia 50
2.4.3. Efeito da umidade 52
2.4.4. Efeito da matéria orgânica 53
2.5. PREDIÇÃO DO CARBONO ORGÂNICO DO SOLO POR ERD 55
3. HIPÓTESES 61
4. OBJETIVOS 63
4.1. GERAL 63
4.2. ESPECÍFICOS 63
5. ESTUDO 1 – ESTRATIFICAÇÃO DE UMA BIBLIOTECA ESPECTRAL LOCAL POR CRITÉRIOS DE HOMOGENEIDADE PRODUZ PREDIÇÕES DE CARBONO ORGÂNICO DO SOLO MAIS PRECISAS 65
5.1. INTRODUÇÃO 65
5.2. MATERIAL E MÉTODOS 69
5.2.1. Descrição da área de estudo 69
5.2.2. Amostragem de campo e análises de solo 70
5.2.3. Medições e processamento espectral 70
5.2.5. Calibração e validação dos modelos 74
5.3. RESULTADOS E DISCUSSÃO 77
5.3.1. Caracterização dos solos e análises estatísticas dos dados de COS 77
5.3.2. Caracterização dos dados espectrais vis-NIR 80
5.3.3. Análises de componentes principais dos dados espectrais 82
5.3.4. Performance dos modelos de predição da BEL 84
5.3.5. Estratificação da BEL e desempenho dos modelos de predição 94

5.4. CONCLUSÕES 98
6. ESTUDO 2 – A ESTRATIFICAÇÃO DE UMA BIBLIOTECA ESPECTRAL DE SOLOS SUBTROPICAIS MELHORA AS PREDIÇÕES DE CARBONO ORGÂNICO NO SOLO 99
6.1. INTRODUÇÃO 99
6.2. MATERIAL E MÉTODOS 101
6.2.1. Descrição dos bancos de dados 101
6.2.1.1. Banco de dados 1 – BD1 103
6.2.1.2. Banco de dados 2 – BD2 104
6.2.1.3. Banco de dados 3 – BD3 106
6.2.1.4. Banco de dados 4 – BD4 107
6.2.1.5. Banco de dados 5 – BD5 109
6.2.1.6. Banco de dados 6 – BD6 110
6.2.1.7. Banco de dados 7 – BD7 112
6.2.2. Análises de solo e espectrais 113
6.2.3. Organização da biblioteca espectral – BE 114
6.2.4. Desenvolvimento dos modelos espectroscópicos 117
6.3. RESULTADOS E DISCUSSÃO 119
6.3.1. Caracterização do solo e análise estatística das bibliotecas espectrais 119
6.3.2. Análise exploratória quantitativa dos espectros 125
6.3.3. Análise exploratória qualitativa dos espectros 132
6.3.4. Desempenho do modelo espectroscópico global - MG 136
6.3.5. Modelos espectroscópicos a partir da estratificação da BEG 145
6.3.5.1. Modelos preditivos de base regional (MR) 146
6.3.5.2. Modelos preditivos de base textural (MT) 152
6.3.5.3. Modelos preditivos de base uso da terra (MUT) 156
6.3.5.4. Modelos preditivos de base espectrais (ME) 161
6.4. CONCLUSÃO 170
7. CONSIDERAÇÕES FINAIS 173
8. REFERÊNCIAS 175
9. APÊNDICES 191

INTRODUÇÃO GERAL

Mudanças nos diferentes ecossistemas terrestres têm comprometido a sustentabilidade do recurso natural solo em escala local, regional, nacional e global. Esse cenário tem refletido no aumento da demanda por informações espaciais de propriedades do solo, com destaque para o Carbono Orgânico do Solo (COS), em resoluções espaciais finas, com o objetivo de buscar o uso sustentável das terras frente às diferentes formas de exploração agrícola e fornecer dados para modelagem e simulação de cenários futuros.

Pesquisas mostram que a organização de dados solos em banco de dados é passo essencial para desenvolver pesquisas que resultem em avanço para a ciência do solo, e gerar resultados que atendam as demandas atuais por informações de solo. As Bibliotecas Espectrais de Solo (BES) são banco de dados que reúnem informações espectrais do solo de um determinado local. As BES tiveram início na década de 70 com Condit e 80 com Stoner e Baumgadner, que reuniram dados espectrais de solos dos Estados Unidos. Em seguida, vários trabalhos científicos foram publicados sobre o tema. No Brasil, a primeira BES foi compilada em 1992 pelos pesquisadores do Instituto Nacional de Pesquisas Espaciais (INPE). Em nível de instituição de ensino, a primeira iniciativa de organização de uma BES foi realizada pelo Departamento de Ciência do Solo da Escola Superior de Agricultura "Luiz de Queiroz" (ESALQ), em 1995. Atualmente, essa BES é denominada de Biblioteca Espectral de Solos do Brasil (BESB) e abrange solos de todo o território nacional (http://bibliotecaespectral.wixsite.com/esalq).

O uso da técnica de Espectroscopia de Reflectância Difusa (ERD) para obter BES e integrar os dados com métodos estatísticos multivariados tem demonstrado potencial para construir modelos de predição do conteúdo de COS. Porém, são necessários mais estudos regionalizados para aumentar a base de conhecimento e desenvolver estratégias para construção de modelos de predição de COS e avaliar a possibilidade de aplicação destes para outras regiões.

O grupo de pesquisa em Mapeamento Digital de Solos coordenado pelo professor Dr. Ricardo Simão Diniz Dalmolin, vinculado ao Programa de Pós-Graduação em Ciência do Solo da Universidade Federal de Santa Maria e com contribuição do professor Dr. Alexandre ten Caten da Universidade Federal de Santa Catarina (Campus Curitibanos), vem trabalhando na organização de BES compostas por solos amostrados na região sul do Brasil (Estados do Rio Grande do Sul – RS e Santa Catarina – SC). As pesquisas do grupo têm utilizado a técnica de ERD associada a métodos estatísticos para desenvolver modelos de predição de propriedades do

solo (argila, silte, areia e COS) para bacias hidrográficas dessa região. Os resultados encontrados pelo grupo até o presente estudo desta tese sugerem que existem fatores, provavelmente, relacionados à variação pedológica das amostras das BES, influenciando na acurácia dos modelos de predição.

Considerando que a acurácia dos modelos é definida pela variação das características pedológicas dos solos representados na BES e visando a implementação da técnica de ERD em laboratórios de análises de solo, vê-se a necessidade de pesquisas em regiões onde está presente a grande diversidade pedológica associada com a heterogeneidade em relação ao teor de argila, areia e COS. Tais características são observadas na região sul do Brasil, onde no RS, na região do Planalto predominam solos com textura argilosa e muito argilosa, enquanto na região da Depressão Central estão presentes distintos usos da terra (campo nativo, lavouras, florestas nativas, florestas plantadas, entre outros) e solos com textura arenosa nos horizontes superficiais e argila nos subsuperficiais. Já na região Serrana do Estado de SC, predominam solos de textura média e argilosa e uso da terra com florestas plantadas de pinus.

Diante do exposto, para dar continuidade nas pesquisas do grupo Mapeamento Digital de Solos da UFSM, objetivo geral desta tese foi avaliar a influência de processamentos espectrais, métodos estatísticos multivariados e estratificação amostral de BES na acurácia dos modelos de predição para o teor de COS. Para isso, foram realizados dois estudos: ESTUDO 1, testou-se a influência de técnicas de pré-processamentos, métodos multivariados e estratificação da BES com base na variância espectral, classe de solo, uso da terra e profundidade do solo amostrada na predição de COS, utilizando uma BES de escala local; ESTUDO 2, avaliou o efeito da estratificação da BES com base em características ambientais regionais, textura do solo, usos da terra e classes espectrais na acurácia dos modelos a partir de uma BES de escala regional.

1. JUSTIFICATICA E RELEVÂNCIA

Em virtude do acelerado crescimento populacional, em 2050, a Terra será habitada por cerca de 9 bilhões de pessoas, segundo projeções da Organização das Nações Unidas (ONU). Diante desse cenário, a produção de alimentos deverá crescer entre 60 e 90%, o que poderá aumentar a pressão sob os recursos naturais, em especial o solo, se não forem tomadas medidas adequadas de uso e conservação desse recurso. Em função disso, nas últimas décadas, o homem tem buscado soluções mais consistentes para problemas de escala global que decorrem destes fatos tais como mudanças climáticas, segurança alimentar e energética, preocupação com as formas de uso da terra, conteúdo e sequestro de carbono. E é na ciência do solo que estão encontrando grande parte destas soluções.

O solo é o principal responsável pela produção de alimentos, pelos ciclos biogeoquímicos do Carbono (C) e Nitrogênio (N) e pela regulação climática, por isso, manter suas funções é uma questão vital. Infelizmente, o relatório lançado recentemente pela Organização das Nações Unidas para a Alimentação e Agricultura (FAO) revelou que 33% dos solos do mundo já estão degradados pela erosão, salinização, compactação, acidificação e contaminação (FAO; ITPS. 2015). O relatório mostra que práticas agrícolas inadequadas e a pressão do ser humano sobre o solo estão a intensificar a degradação e reforça a importância desse recurso natural para a subsistência humana.

Esse cenário tem refletido no aumento da demanda por informações espaciais de propriedades do solo em escalas detalhadas e resoluções espaciais finas para grandes áreas, com o objetivo de buscar o uso sustentável das terras frente às diferentes formas de exploração agrícola (AMUNDSON et al., 2015). Por isso, as pesquisas na ciência do solo têm focado no desenvolvimento de métodos que permitam a obtenção de dados de solo de forma rápida, com menor custo e maior acurácia. Nesse âmbito, a Pedometria e o Mapeamento Digital de Solo (MDS) são técnicas que atendem essa demanda por métodos mais sustentáveis para obter dados solos, possibilitando construção de modelos de predição de um vasto conjunto de propriedades do solo (McBRATNEY; MENDONÇA-SANTOS; MINASNY, 2003; DALMOLIN; TEN CATEN, 2015).

O COS é uma dessas propriedades do solo, a qual desempenha diversas funções no ambiente, sendo considerado um indicador de qualidade do solo (GLOBAL SOIL

PARTNERSHIP, 2017). Portanto, conhecer o conteúdo e a distribuição do COS na paisagem é fundamental, tendo em vista a importância deste no cenário global (STOCKMANN et al., 2015; JACKSON et al., 2017). Assim, esforços têm sido realizados para quantificar e mapear com precisão o conteúdo de COS em nível global (KÖCHY et al., 2017; HENGL et al., 2014) e nacional (VASQUES et al., 2017). No entanto, realizar mapeamento do conteúdo de COS nas escalas mais detalhadas, por exemplo, em nível regional, municipal e de propriedade rural é geralmente dispendioso e demanda tempo devido ao elevado número de pontos que precisam ser amostrados na paisagem, em especial para o COS (VASQUES et al., 2010). A necessidade de uma amostragem densa para caracterização do COS está relacionada com o número de fatores ambientais e antrópicos que influenciam na sua distribuição na pedosfera. Então, é necessário o uso de novas técnicas de amostragem, processamento de dados e a construção de modelos de predição que representem de maneira adequada as relações complexas entre o solo e o ambiente em que o mesmo ocorre.

Historicamente, sabe-se que todo o conhecimento a respeito dos solos e suas características foram obtidos a partir de análises de rotina em laboratório (VISCARRA ROSSEL et al., 2006), no entanto, a caracterização do solo baseada em métodos analíticos padrão está sendo discutida, já que as metodologias utilizadas apresentam alto custo por análise, geram resíduos químicos e demoram, limitações essas que poderiam impossibilitar o seu uso dado o grande volume de amostras necessárias para o conhecimento e planejamento do uso do solo em escalas detalhadas no Brasil (POLIDORO et al., 2016).

Um dos desafios é propor uma técnica com potencial para o monitoramento do solo que permita uma leitura rápida, barata e não destrutiva das amostras (VISCARRA ROSSEL et al., 2006; BELLINASO; DEMATTÊ; ROMEIRO, 2010). A espectroscopia de refletância difusa (ERD) tem despontado como uma ferramenta eficiente para a predição de várias propriedades do solo (McBRATNEY, MINASNY, VISCARRA ROSSEL, 2006; BELLINASO et al., 2010; VISCARRA ROSSEL et al., 2016; DOTTO et al., 2018; HONG et al., 2018; TERRA, DEMATTÊ, VISCARRA ROSSEL, 2018; VAUDOUR et al., 2018). Com base nessa técnica, são construídos modelos de predição a partir de variáveis espectrais que apresentam correlação com as características do solo. As variáveis utilizadas na construção desses modelos são as bandas espectrais obtidas pela técnica de ERD, as quais contém a assinatura espectral do solo nas regiões do visível ao infravermelho-próximo (Vis-NIR) (VISCARRA ROSSEL et al., 2006; VASQUES

et al., 2010; KNOX et al., 2015; VISCARRA ROSSEL et al., 2016; WIGHT; ASHWORTH; ALLEN, 2016). Essa técnica permite a predição de várias propriedades do solo somente com uma única amostra espectral, além de apresentar uma série de vantagens em relação às metodologias tradicionais de análise do solo, como rapidez, maior eficiência quando um grande número de amostras se fizer necessário, baixa produção de resíduos químicos prejudiciais ao ambiente e menor custo (BEN-DOR et al., 2009; VISCARRA ROSSEL; BEHRENS, 2010).

Porém, as predições ainda contém muita variação nos resultados de acurácia, sendo que a capacidade preditiva dos modelos está atrelada às características dos dados espectrais e, consequentemente, às características do solo (VISCARRA-ROSSEL et al., 2016). Assim, avaliar a construção de modelos com uso de BES torna-se essencial, principalmente quando esta é composta por amostras de solo com grande variação nas características físicas, químicas e mineralógicas, como é o caso dos solos da região sul do Brasil. Conforme Grunwald et al. (2015), não há como transferir os modelos desenvolvidos de um local (região) para outro, visto a diversidade pedológica, ou seja, os modelos devem ser regionais e considerar amostras locais e suas características. Somado a isso, não há protocolos de uso da ERD para desenvolver modelos de predição para o COS mais acurados quando se utiliza uma BES compostas por amostras de solos com grande variação composicional e espectral.

2. REVISÃO BIBLIGRÁFICA

2.1. PEDOMETRIA

Estudos pioneiros a respeito do solo como corpos naturais organizados e a sua gênese, foram difundidos pelo russo Dokuchaev, em 1883, e posteriormente por seus discípulos, na metade do século 19. Nesse período varias equações matemáticas foram apresentadas com o objetivo de representar a lei básica da ciência do solo apresentada por Dokuchaev, a qual diz que "o solo é sempre e em qualquer lugar função dos seguintes fatores de formação: (1) natureza da rocha parental; (2) clima do terreno; (3) massa e carácter da vegetação; (4) a idade do terreno; e (5) a topografia do terreno" (FLORINSKY, 2012). A equação mais aceita na época e aceita até os dias atuais para representar estas relações foi a apresenta pelo russo Sergey Zakharov em 1927. Com base nesses estudos, Hans Jenny, em 1941, formalizou as bases conceituais sobre a gênese e a distribuição dos solos na paisagem. Jenny propôs um modelo clássico abordando os fatores relacionados com a formação do solo. Para ele, o solo (S) é função do clima (cl), dos organismos (o) e do relevo (r), agindo sobre o material parental (p) durante um período de tempo (t) e fatores desconhecidos (...), conforme representado pela equação 1:

$$\boldsymbol{S = f\ (cl, o, r, p, t, \ldots)} \qquad (\mathbf{1})$$

O modelo determinístico de Jenny representou uma maneira complexa de predizer os solos, em função das interações entre os fatores de formação. Porém, se o solo é função dos fatores de formação e, se a distribuição espacial desses fatores for conhecida, o solo e suas propriedades podem ser inferidos a partir da interpretação de informações de campo e construção de um modelo conceitual (relação solo-paisagem) (JENNY, 1941). Esse modelo busca compreender as relações entre a paisagem e demais variáveis ambientais que caracterizam o solo, sendo esse aplicado para inferir a distribuição espacial dos solos e suas propriedades em áreas análogas (MENDONÇA-SANTOS et al., 2007).

A partir da introdução do modelo de Jenny na ciência do solo, mais especificamente na área da Pedologia, vários levantamentos e mapeamentos de solos foram elaborados em escala mundial, conforme as necessidades e demandas da sociedade em geral de conhecer os tipos de solo e sua distribuição na superfície terrestre. Inicialmente o levantamento pedológico era

realizado com descrições dos solos, classificação, determinação de limites entre as classes e, com a realização de inferências sobre o comportamento dos solos perante o uso e manejo (MENDONÇA-SANTOS; SANTOS, 2003). No mapeamento do solo utilizando métodos convencionais, o pedólogo adquire conhecimentos tácitos a respeito da ocorrência das classes de solos na paisagem, sendo de fundamental importância, a experiência adquirida no decorrer dos muitos trabalhos de campo. Contudo, existem poucos pedólogos treinados e capazes de descrever um modelo mental generalizado que atenda as diversas situações que ocorrem no ambiente.

Nos últimos anos, a demanda por informações de qualidade sobre o solo tem aumentado devido à estreita relação desse recurso natural com os ciclos hidrológicos, dinâmica do clima em escala global, ciclagem de nutrientes e produção de alimentos e energia (HARTEMINK, 2008; SANCHEZ; AHAMED; CARRE, 2009; GRUNWALD et al., 2012). A gestão e o planejamento para fins específicos, com alto nível de detalhe, requerem mapas funcionais de solos (ZHU; LIN; DOOLITTLE, 2013), como por exemplo, mapa de aptidão agrícola da terra e área de risco de erosão (VAN ZIJL et al., 2014).

Nesse contexto, uma das limitações é a indisponibilidade de mapas de solo em escalas adequadas para cada propósito. Segundo Sanchez, Ahamed e Carre (2009), apenas 109 países possuem mapas coropléticos de solos nas escalas 1:1.000.000 ou maiores, representando somente 31% da superfície da Terra. O restante, dispõe apenas do mapa de solos na escala 1:5.000.000, elaborado pela FAO, em escala incompatível para planejamento de uso do solo necessário para atender as demandas atuais.

Ainda, os levantamentos e mapeamentos de solos realizados com base no “modelo mental”, desenvolvido pelo pedólogo experiente, estabelecem limites abruptos arbitrários entre as classes de solo (ten CATEN et al., 2011) e as informações não ficam documentadas, dificultando a reprodução e atualização desses mapas no futuro. Esses são alguns dos motivos pelos quais o mapeamento convencional vem recebendo críticas a respeito de suas limitações para atender a sociedade no sentido de informações quantitativas e com precisão sobre a distribuição espacial do solo e suas propriedades (HARTEMINK e McBRATNEY, 2008; ARRUDA et al., 2013; CARVALHO JÚNIOR et al., 2014).

Frente a isso, precisa-se compreender melhor os processos que ocorrem na pedosfera em conjunto e quantificar precisamente a distribuição espacial do solo e suas propriedades físicas, químicas e biológicas, as quais exercem funções no ambiente. Todavia, essa é uma tarefa

complexa, uma vez que, o solo é um sistema heterogêneo e contínuo na paisagem, onde seus processos de formação e dinâmica de funcionamento são de difícil compreensão. Além disso, as coletas de solos para formação de bancos de dados apresentam um custo elevado e demanda de tempo (OLIVEIRA et al., 2012).

O desenvolvimento de novas técnicas e o uso frequente pelos pesquisadores de tecnologias digitais, tem contribuído para reduzir esses problemas, como é o caso da técnica de MDS (McBRATNEY; MENDONÇA-SANTOS; MINASNY, 2003). O MDS foi defino por Lagacherie e McBratney (2007) como "a criação e população de sistemas de informação espacial de solos através do uso de métodos observacionais de campo e laboratório, acoplados a sistemas de inferência espacial e não-espacial de solos". No MDS as informações extraídas de amostras de solo e associação com variáveis ambientais são empregadas em modelos de predição, que possibilitam gerar informação espacial a respeito do solo com as incertezas e erros associados (MINASNY; MCBRATNEY; SALVADOR-BLANES, 2008), além de permitirem análises e interpretações em qualquer época. Para isso, são utilizadas tecnologias como: sensoriamento remoto, informática, gerenciamento de dados espaciais em Sistema de Informações Geográficas (SIG), visualização e integração em conjunto do solo e as variáveis ambientais correlacionadas, métodos quantitativos para predizer os processos que ocorrem no solo e suas propriedades (McBRATNEY; MENDONÇA-SANTOS; MINASNY, 2003; MINASNY; MCBRATNEY; SALVADOR-BLANES, 2008; GRUNWALD, 2009; ten CATEN et al., 2012).

O MDS engloba uma área de conhecimento da Ciência do Solo chamada de *Pedometria*. O termo Pedometria possui significado em um neologismo derivado do grego *pedos* (solo) e *metron* (medida). A pedometria visa solucionar os problemas relacionados com as incertezas, inerentes ao método convencional de mapeamento de solos, com investigações sobre novas abordagens de modelagem quantitativa, com a finalidade de aprofundar o conhecimento da variabilidade espacial dos solos, da qualidade e precisão da informação (HEUVELINK, 2003). Para atender esse objetivo, a pedometria lança mão do uso algumas ferramentas (Figura 1) como a geoinformação, as informações do solo e a estatística (modelos lineares generalizados, redes neurais, árvores de decisão, regressão, componentes principais). Portanto, é um campo interdisciplinar entre a Ciência do Solo, a estatística e a matemática aplicada à ciência da geoinformação, que permite compreender o funcionamento dos processos do solo (HENGL, 2003), eliminando algumas incertezas associadas ao mapeamento convencional. Segundo ten

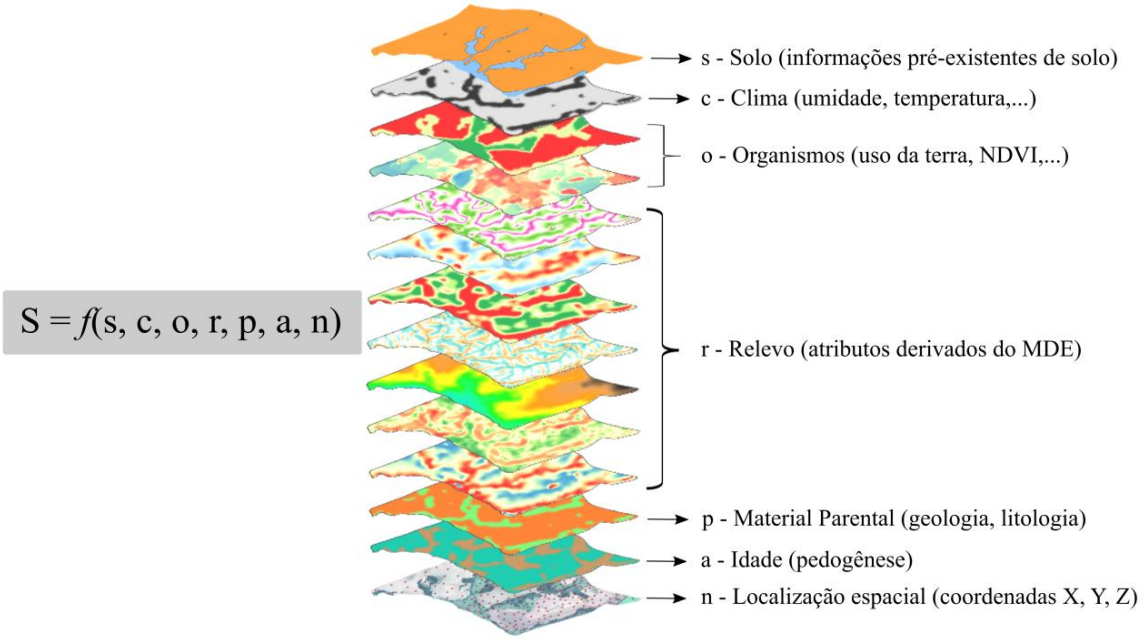
Caten et al. (2012), o avanço dos estudos em pedometria é também resultado de avanços tecnológicos, técnicas de sensoriamento remoto, posicionamento global de satélites e ao acesso à informação através da rede mundial de computadores.

Figura 1 – Diagrama mostrando a interdisciplinaridade da pedometria

Fonte: Adaptado de HENGL (2003).

O MDS utiliza os fundamentos básicos da pedometria através do uso de modelos de predição espacial de solos. Tais modelos constituem um método de ajuste empírico das relações quantitativas existentes entre o solo e o ambiente em que o mesmo ocorre (McBRATNEY; MENDONÇA-SANTOS; MINASNY, 2003). A equação que representa os modelos de predição constitui o modelo *S=f(scorpan*) (Figura 2), nas quais o S representa a classe ou a propriedade do solo a ser predita em função de s = informação prévia do solo, c = clima, o = organismos, r = relevo, p = material de origem, a = período decorrido, e n = posição espacial (McBRATNEY; MENDONÇA-SANTOS; MINASNY, 2003). Este modelo apresenta fatores semelhantes ao de Jenny (1941). Porém, o modelo *scorpan* acrescenta em sua composição os fatores *s* e *n*, que representam a relação solo e ambiente em determinado espaço geográfico, possibilitando inferir a variabilidade espacial. Assim, o MDS se torna útil à medida que aumenta a velocidade dos levantamentos e facilita a obtenção de dados, diminuindo os custos de obtenção e análise dos mesmos.

Figura 2 – Representação do modelo *scorpan* com os planos de informação espacial



Fonte: Autor.

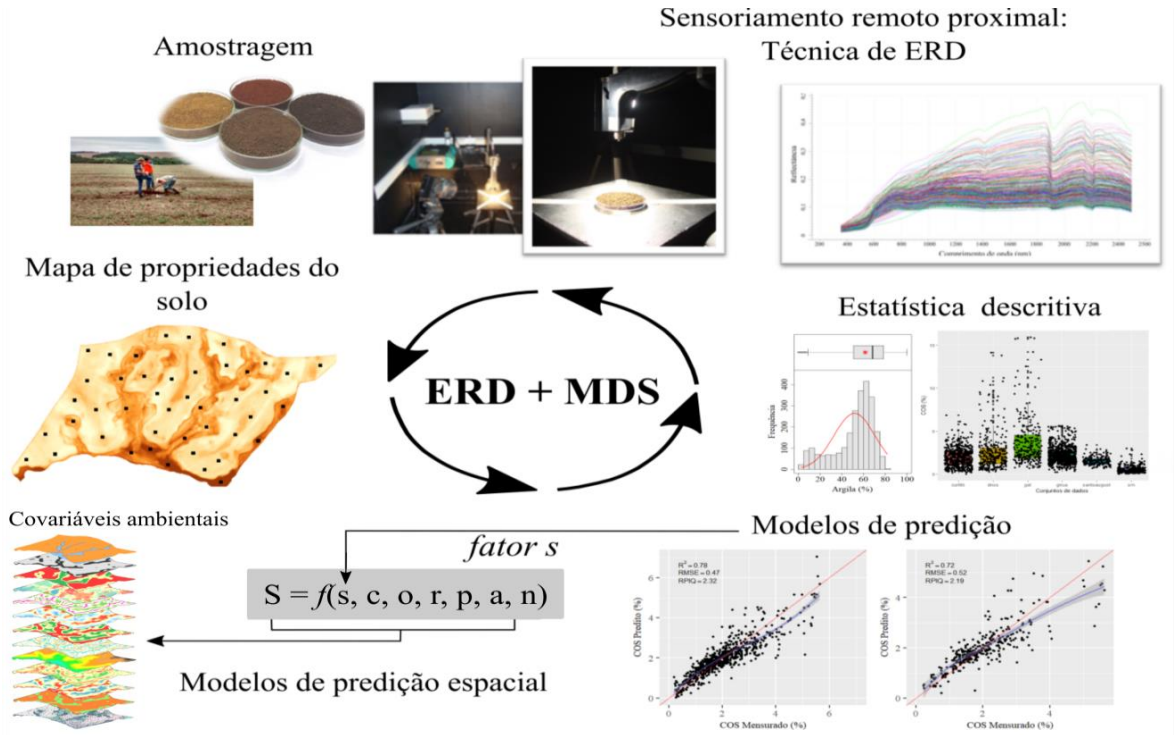
O reconhecimento do MDS pela comunidade científica internacional como uma ferramenta promissora para fornecer informações de qualidade e precisão a respeito do solo, se deu a partir da publicação do trabalho de McBratney, Mendonça-Santos e Minasny, (2003), que abordou os fundamentos básicos do MDS. No Brasil, as pesquisas em MDS são mais recentes, com destaque para o mapeamento de classes de solos, como os trabalhos de Giasson et al. (2006) e ten Caten et al. (2011). Segundo ten Caten et al. (2012), somente após o 2° Workshop Global em MDS, realizado no Rio de Janeiro, em 2006, os pesquisadores brasileiros da área da Ciência do Solo deram importância para os métodos empregados no MDS. Esses trabalhos utilizaram como variáveis preditoras os atributos do terreno derivados de um Modelo Digital de Elevação (MDE), representado o fator relevo do modelo *scorpan*.

O MDS pode atender as necessidades globais por informações de solos mencionadas anteriormente, através do Grupo de Trabalho de MDS da *International Union of Soil Sciences* e do consórcio GlobalSoilMap.net. O objetivo do consórcio é gerar um mapa digital em escala global das propriedades funcionais do solo, como por exemplo, a composição granulométrica (areia, silte e argila), teor de COS, pH, CTC, CE e densidade do solo. Esse grupo estabeleceu

especificações que regem alguns aspectos que devem ser estudados e seguidos nos trabalhos de MDS, apresentados no livro *GlobalSoilMap: Basis of the global spatial soil information system* (ARROUAYS et al., 2014). Este novo mapa global do solo visa auxiliar a tomada de decisões em uma série de questões, tais como a produção de alimentos e a erradicação da fome, mudanças climáticas e degradação ambiental.

Assim, estudos têm mostrado a importância de desenvolver métodos para caracterizar de forma rápida e precisa todo o perfil de solo (BEN-DOR; HELLER; CHUDNOVSKY, 2008; VISCARRA ROSSEL et al., 2011; VISCARRA ROSSEL et al., 2016), visando quantificar as propriedades do solo (*fator s*) para serem incorporadas no modelo *scorpan* por técnicas de MDS (Figura 3) (LAGACHERIE; GOMEZ, 2018). Neste contexto, a Espectroscopia de Reflectância Difusa e a organização das informações espectrais do solo em bibliotecas espectrais têm se mostrado ferramentas com potencial em capturar a variabilidade pedológica e quantificar múltiplas propriedades do solo (STENBERG; VISCARRA ROSSEL, 2010; NOCITA et al., 2015; SOUZA et al, 2016; VISCARRA ROSSEL et al., 2016; VAUDOUR et al., 2018).

Figura 3 – Representação das etapas envolvidas na integração da técnica de ERD e MDS para obtenção de informações de solos



Fonte: Autor.

2.2. MÉTODOS DE DETERMINAÇÃO DO CONTEÚDO DE COS

O carbono (C) no solo pode estar presente de três formas básicas: C elementar, C inorgânico e C orgânico (NELSON; SOMMERS, 1996). Algumas das formas elementares de C incluem fuligem, carvão e grafite. As formas inorgânicas compreendem os minerais carbonatos como a calcita e dolomita. E as formas orgânicas incluem os restos de plantas, animais e microrganismos em diferentes estágios de decomposição, mas consiste principalmente de materiais fortemente alterado (humificado). Em regiões úmidas, onde o perfil de solo está sujeito à lixiviação, o C ocorre predominantemente ou inteiramente como C orgânico, enquanto que em regiões áridas a quantidade de C inorgânico frequentemente excede o teor de C orgânico. Embora as formas elementares de C não serem consideradas C orgânico do solo, elas são determinadas pelos métodos de combustão por via úmida e seca, utilizados para quantificar o COS (NELSON; SOMMERS, 1996).

O C orgânico é o elemento chefe presente na matéria orgânica do solo, compreendendo de 48 a 58 % do peso total. Portanto, determinações de COS são frequentemente usadas como base para estimativas de matéria orgânica através da multiplicação do valor de C orgânico por um fator. O fator mais utilizado é o de “Van Bemmelen” (f = 1,724), baseado na suposição que a matéria orgânica contém 58 % de C orgânico (PRIBYL, 2010). No entanto, estudos tem mostrado que a proporção de C na matéria orgânica do solo é altamente variável para uma variedade de solos e não há fator apropriado para todos os solos. Porém, o fator de “Van Bemmelen” ainda é o mais utilizado (PRIBYL, 2010).

Em geral, o conteúdo de COS é determinado nos laboratórios de rotina de análise de solo por: i) oxidação por via seca por meio de analisador elementar e ii) oxidação com dicromato de potássio por via úmida (WALKLEY; BLACK, 1934). O segundo método conhecido como “Walkley-Black” é utilizado pela maioria dos laboratórios do Brasil e do mundo. No entanto, essa metodologia apresenta como principais desvantagens o alto custo, geração de resíduos químicos e tempo excessivo de análise (O’ROURKE e HOLDEN, 2011). Por essa razão, pesquisas vêm sendo desenvolvidas para propor metodologias mais eficientes, apresentado baixo custo, possibilitando medições do conteúdo de COS sem o uso de reagentes químicos e de forma mais rápida.

2.2.1. Oxidação por via úmida

Estudo de Schollenberger (1927) propôs que a matéria orgânica do solo pode ser oxidada por uma mistura quente de $K_2Cr_2O_7$ e H_2SO_4 de acordo com a equação abaixo.

$$2Cr_2O_7^{2-} + 3C^0 + 16H^+ \rightleftharpoons 4Cr^{3+} + 3CO_2 + 8H_2O$$

Após a reação, em amostras que requerem mais precisão, a determinação da quantidade de íons Cr^{3+} reduzidos é realizada por titulação dos íons Cr^{3+} em excesso pelos íons de Fe^{2+} da solução padronizada de sulfato ferroso amoniacal $[Fe(NH_4)_2(SO_4)_2]$, conforme mostra a equação abaixo.

$$Cr^{2+}O_7^{2-} + 6Fe^{2+} + 14H^+ \rightleftharpoons 2Cr^{3+} + 3CO_2 + 8H_2O$$

Outra forma de determinar a quantidade dos íons Cr^{3+} em excesso, é o método de colorimetria, medindo-se a intensidade da cor esverdeada produzida por esses íons em solução. Esse método é normalmente empregado em laboratórios de rotina, onde se requer mais rapidez na obtenção dos resultados analíticos. Porém, esse método necessita da montagem de curva-padrão de calibração obtida a partir de uma série de amostras, nas quais o conteúdo de COS é determinado por titulação por apresentar maior precisão. Todos os métodos baseados na determinação de $Cr_2O_7^{2-}$ remanescente ou Cr^{3+} formado assumem que o C orgânico presente na matéria orgânica tem uma valência zero (C°), ou seja, estado de oxidação zero, o que não é verdadeiro (WALKLEY; BLACK, 1934).

O método original de Walkley-Black não utiliza aquecimento externo, contato somente com o calor gerado pela diluição do ácido sulfúrico. Assim, a oxidação do C orgânico não é completa, gerando inconsistências nos resultados obtidos (WALKLEY; BLACK, 1934). Para contornar esse problema, vários estudos foram realizados com objetivo de testar formas e tempo de aquecimento no processo de oxidação do C orgânico. Nelson e Sommers (1975) propuseram que o C orgânico poderia ser determinado aquecendo a mistura solo-ácido-dicromato em tubos de digestão de 50 mL, com refluxo, colocados em uma chapa quente. O tempo de aquecimento e temperatura de digestão recomendada foi de 30 min e 150 °C, respectivamente. O método mais

aceito e empregado até hoje, foi o proposto por Yeomans e Bremner (1988), que substituíram a chapa de aquecimento pelo bloco de digestão (Mebius no bloco), no intuito de uniformizar a temperatura durante o processo de digestão e aumentando o número de amostras manipuladas por bateria. Este último método é descrito a seguir, em detalhes.

Amostras de terra fina seca ao ar (TFSA), de 0,20 a 0,50 g, são transferidas para tudo de digestão de 2,5 cm de diâmetro e 25,0 cm de altura. Adiciona-se em seguida, 10 mL da solução sulfocrômica ($K_2Cr_2O_7$ 0,067 mol L^{-1} (0,4N) + H_2SO_4 concentrado). Funis de refluxo são colocados em todos os tudo para servirem de condensadores. Em cada bateria se utiliza três provas em branco. Após o pré-aquecimento do bloco a 150 °C, os tubos são colocados no bloco, e deixados sob aquecimento por 30 minutos. Posteriormente, os tubos são retirados do bloco e deixados em repouso para esfriarem. Em seguida, o conteúdo é transferido para erlenmeyers de 250 mL já contendo 20 mL de água destilada e 2 mL de ácido ortofosfórico (H_3PO_4). Em cada tubo de digestão são adicionados 20 mL de água destilada, com posterior agitação do tubo e transferência para o erlenmeyer, e, em seguida é adicionado mais 40 mL água destilada e novamente o tubo é agitado e o conteúdo transferido para o erlenmeyer. Ao final desse processo, no erlenmeyer irá conter aproximadamente 80 mL de solução. A titulação do excesso de íons Cr^{3+} na amostra é realizada pela solução de sulfato ferroso amoniacal [$Fe(NH_4)_2(SO_4)_2\ 6H_2O$] 0,2 mol L^{-1} na presença de 5 gotas da solução indicadora ferroína (O-fenantrolina), até o aparecimento da cor violeta-escuro. A concentração real do sulfato ferroso amoniacal é determinada por titulação das provas em branco. Por fim, são realizados os cálculos para obter o conteúdo de C orgânico na amostra expresso em grama por quilograma de solo.

2.2.2. Oxidação por via seca

No método de oxidação por via seca, a amostra de solo previamente moída em almofariz (a 100 mesh) é oxidada em presença de $Cr_2O_7^{2-}$ e CO_3O_4/Ag em forno, com temperatura entre 1.020–1.500°C, com circulação de oxigênio ultrapuro sob pressão de 30 kPa. As moléculas orgânicas são oxidadas na forma N_xO_y e $CO^{2-}H_2O$ + O_2, que são carregadas nas formas de gás para um forno de redução sob fluxo de hélio. O forno de redução contém íons Cu, à temperatura de 750°C, de onde o N sai na forma de N_2. Em seguida, o N_2 e o CO_2 são separados em uma coluna cromatográfica e determinados em detector apropriado. Toda a operação é realizada em

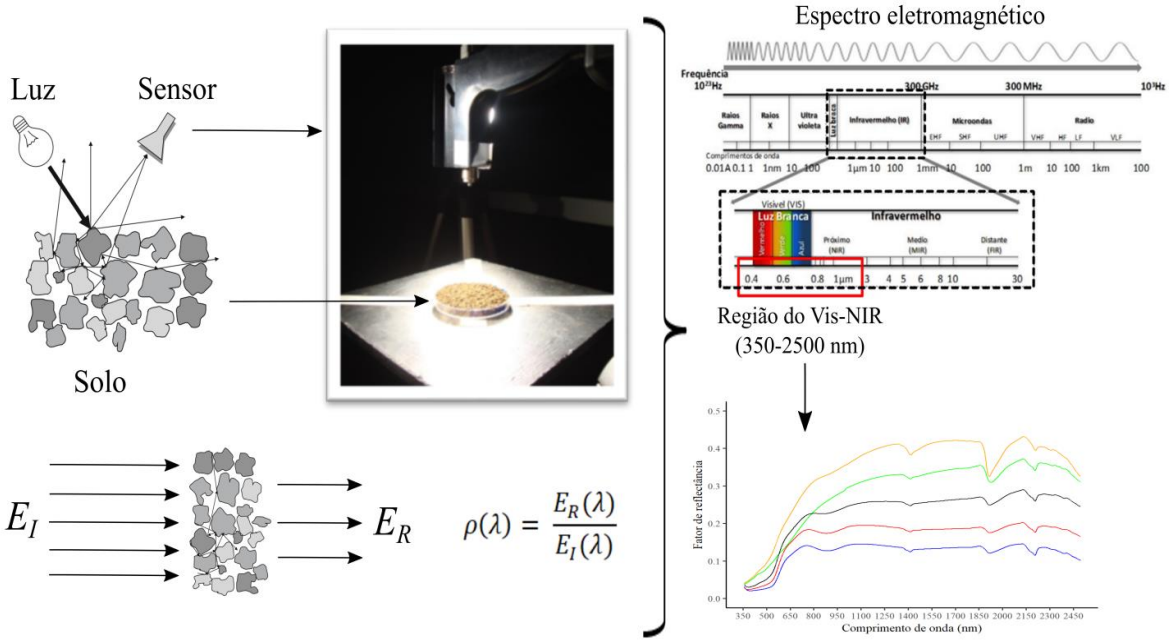
analisador elementar. O método da combustão a seco é considerado padrão, devido a sua alta precisão e exatidão nos resultados (NELSON; SOMMERS, 1996).

2.3. ESPECTROSCOPIA DE REFLECTÂNCIA DIFUSA DO SOLO

A espectroscopia é a ciência que estuda a interação da radiação eletromagnética com a matéria com base nos fundamentos da Lei de Beer. Essa lei também é conhecida como lei de Beer-Lambert-Bouguer por ser a união dos estudos de Pierre Bouguer em 1729, Johann Heinrich Lambert em 1760 e August Beer em 1852. A lei estabelece uma relação entre a absorção de energia de um material, por exemplo, o solo e suas propriedades, quando este é atravessado por uma radiação luminosa monocromática colimada (raios luminosos paralelos) (Figura 4).

A reflectância espectral [ρ (λ)] é função do comprimento de onda (λ), sendo definido matematicamente como a razão entre a energia do comprimento de onda (λ) refletida pelo objeto [E_R (λ)] e a energia do comprimento de onda (λ) incidente sobre o objeto [E_I (λ)] (LILLESAND; KIEFER, 1994), como mostra a equação na figura 4.

Figura 4 – Ilustração mostrando uso da técnica de ERD em solos na região do Vis-NIR (350-2500 nm) do espectro eletromagnético



Fonte: Autor.

A técnica de ERD do solo é um conjunto de técnicas utilizadas para medir quantitativamente a energia radiante do solo (reflectância espectral), sendo que os sistemas utilizados para sua medida são os espectroradiômetros (STONER; BAUMGARDNER, 1981). A denominação da técnica de “ERD” é devido à ocorrência de reflectância difusa (espalhada) de parte da energia eletromagnética proveniente da fonte, por exemplo, de uma lâmpada ao atingir a superfície do solo irá interagir por meio de trocas de energia, o que resulta em uma absorção, reflexão e/ou transmissão de parte da energia. Nesse processo, apenas uma fração da energia refletida alcança a sensor receptor, sendo o restante da energia absorvida ou perdida. A quantidade de energia que será refletida ou absorvida será dependente das propriedades físicas, químicas e mineralógicas do solo (STONET et al., 1980; BAUMGARDNER et al., 1985; STONER; BAUMGARDNER, 1986; BEN-DOR; EPEMA, 1999).

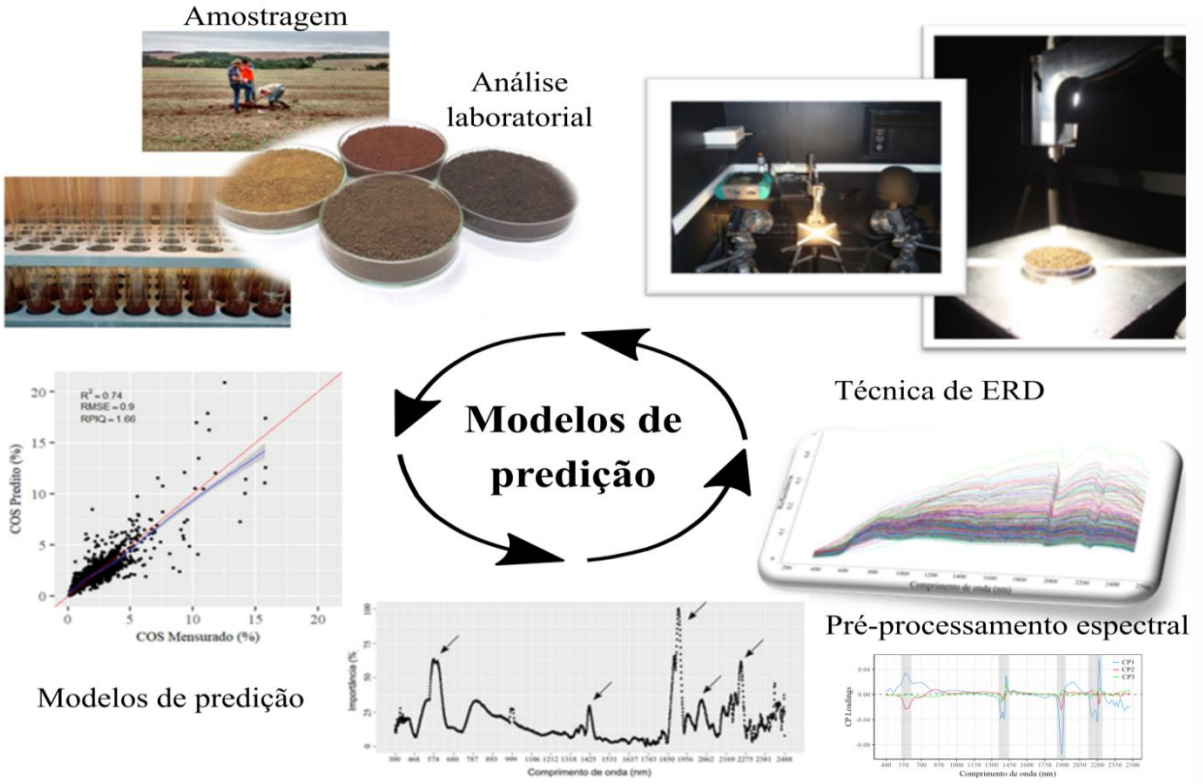
Os primeiros trabalhos publicados mostrando o efeito da água do solo, tamanho de partícula e composição química do solo nos espectros Vis-NIR foram: Angström (1925), O’Neal (1927), Carter (1931), Obukhov e Orlov (1964), Bowers e Hanks (1965), Hunt e Salisbury (1970), Al-Abbas et al. (1972), Hunt (1977), Stoner e Baumgardner (1981) e Baumgardner et al. (1985). Após esses estudos, várias pesquisas foram desenvolvidas, como objetivo de gerar modelos matemáticos para estimar algumas propriedades do solo, como por exemplo, o trabalho de Dalal e Henry (1986) que avaliou o uso do método estatístico de regressão linear múltipla para predizer o conteúdo de COS a partir de dados espectrais na região do NIR.

Em geral, as pesquisas tem mostrado que a aplicação da ERD em solos pode reduzir os custos de amostragem e melhorar ou substituir os métodos tradicionais de laboratórios de análise de solo (BEN-DOR et al., 1995; VISCARRA ROSSEL; MCBRATNEY, 2008; VISCARRA ROSSEL et al., 2016; LAGACHERIE; GOMEZ, 2018), visto que a ERD tem algumas vantagens sobre a determinação das propriedades do solo por análises tradicionais, como por exemplo, aquisição dos dados de maneira mais rápida e menos onerosa, com a mesma eficiência, especialmente quando um grande número de análises é necessário (VISCARRA ROSSEL; BEHRENS, 2010). Nessa técnica não há demora no preparo da amostra e utilização de reagentes químicos prejudiciais para o ambiente (VISCARRA ROSSEL et al., 2006). Além disso, a obtenção dos dados espectrais de solo poder ser realizada *in situ* (BEN-DOR; GRANOT; NOTESCO, 2017; KÜHNEL; BOGNER, 2017), em que o número de amostras a serem coletadas

e processadas para envio ao laboratório pode ser reduzido, diminuindo os custos para obtenção de dados de solo.

A técnica de ERD já vem sendo utilizada em laboratório de análise de solos, como por exemplo, o *Commonwealth Scientific and Industrial Research Organisation* (*CSIRO*), em Sidney, Austrália e o *Africa Soil Information Service (AFSIS) Dry Spectroscopy LAB*, na África. O *CSIRO* disponibiliza esse tipo de serviço em escala comercial e com um custo aproximadamente 83 % menor em relação à análise tradicional oferecida nos laboratórios de rotina. Após a obtenção da assinatura espectral do solo em laboratório, o mesmo espectro pode ser utilizado para a construção de modelos de predição para várias propriedades do solo (Figura 5), como o teor de areia, silte, argila, COS, capacidade de trocas de cátions, pH, ferro, nitrogênio e fósforo (STENBERG et al., 2010, NOCITA et al., 2014; RAMÍREZ-LÓPEZ et al., 2014, DOTTO et al., 2018; KNOX et al., 2015, VISCARRA ROSSEL et al., 2016; DOTTO et al., 2018; VAUDOUR et al., 2018).

Figura 5 – Representação das etapas envolvidas na quantificação de propriedade do solo pela técnica de ERD



Fonte: Autor

2.4. COMPORTAMENTO ESPECTRAL DO SOLO

O comportamento espectral do solo pode ser medido através da leitura das amostras em espectroradiômetros, obtendo-se medidas espectroradiométricas em diferentes comprimentos de onda da energia eletromagnética refletida da superfície da amostra de solo. Essa é representada na forma de gráficos denominados de curva de reflectância espectral do solo ou assinatura espectral do solo.

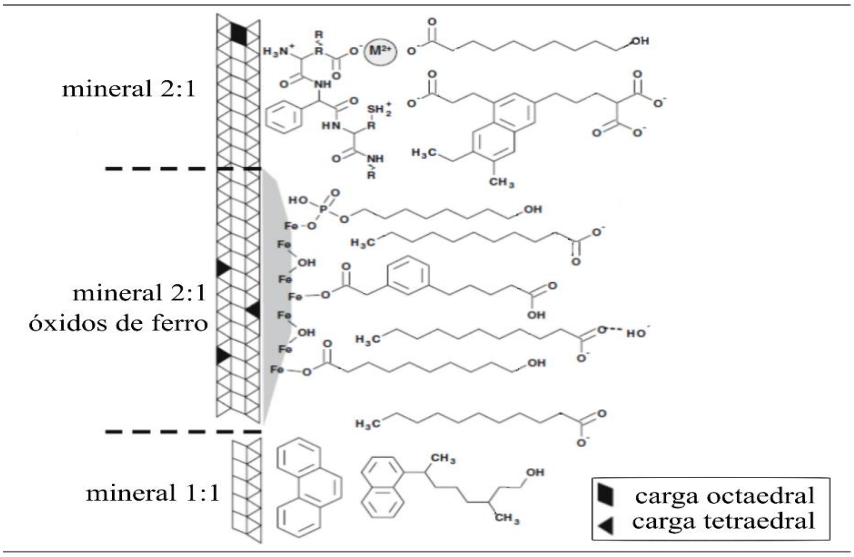
Os fatores que controlam a reflectância são geridos por dois processos de interação da radiação eletromagnética com a superfície do solo, sendo as interações macroscópicas e microscópicas. As *interações macroscópicas* estão relacionadas com as propriedades ópticas do material e envolvem efeitos como refração, difração e espalhamento conforme as características de superfície do material a ser analisado (STONER; BAUMGARDNER, 1986). As *interações microscópicas* ocorrem devido à penetração da radiação eletromagnética na superfície da amostra de solo. Essa penetração ocorre em uma camada muito fina, aproximadamente 10 a 15 micrômetros, o que resulta na interação da radiação incidente com os elétrons e moléculas do solo. Essa interação é responsável por detectar feições de absorção espectral da radiação relacionadas com o tipo de estrutura interna dos constituintes do solo, tamanho dos seus raios iônicos e tipo e força das ligações químicas (HUNT; SALISBURG; LENHOF, 1971). Tais interações são possíveis, em função do modelo quântico da radiação eletromagnética, no qual os materiais se diferenciam por constituírem sistemas com diferentes arranjos de níveis de energia, sendo a reflectância uma resposta indireta das transições ocorridas nesses níveis de energia. As transições ocorrem em nível eletrônico (níveis de energia dos elétrons unidos aos íons na rede cristalina) e molecular (níveis de energia entre as ligações que unem os átomos) (HUNT; SALISBURG; LENHOF, 1971).

Os processos de transição eletrônicas e moleculares ocorrem quando a radiação incidente é absorvida, estimulando um processo interno no átomo ou nas moléculas de mudança de estados de energia. A transição de estado de energia é acompanhada por uma oscilação eletromagnética, sendo a energia absorvida em um dado comprimento de onda exatamente suficiente para provocar uma transição nos níveis de energia das moléculas ou átomos (HUNT; SALISBURG; LENHOF, 1971). Como a distribuição dos níveis de energia é diferente em cada constituinte do solo, ou seja, é função da sua composição, a intensidade na energia absorvida em determinados

comprimentos de onda também difere. Logo, a composição do solo irá resultar em diferentes intensidades de energia absorvida ao longo do espectro eletromagnético.

A composição química dos argilominerais do solo (Figura 6) irá determinar a quantidade e a intensidade da manifestação de cargas na superfície dos argilominerais (grupos funcionais), suas condições físico-química e sua relação com o conteúdo orgânico (KLEBER; JOHNSON, 2010; HAN et al., 2015). Dessa forma, a medida de reflectância está relacionada com a composição organomineral do solo e suas as interações (STONER; BAUMGARDNER, 1986; STENBERG et al., 2010).

Figura 6 – Representação das diferentes formas de interação organomineral do solo conforme a estrutura dos argilominerais



Fonte: Adaptado de KLEBER; SOLLINS; SUTTON (2007).

O espectro Vis-NIR apresenta bandas largas, provenientes da sobreposição de absorções, devido principalmente a sobretons (transições de níveis de energia) e combinações de vibrações de ligações químicas C – H, O – H, N – H, C =O e Al – OH. Isso é um problema no momento de identificação dos constituintes relacionados com cada feição de absorção, principalmente em amostras de solo, onde as interações de ligações químicas são complexas (Figura 6). No entanto, a vibrações moleculares fundamentais ocorrem na região do infravermelho médio (MIR), enquanto seus sobretons (transições de níveis de energia) e combinações ocorrem no Vis-NIR, onde o principal processo pelo qual moléculas absorver energia é transições eletrônicas em

átomos do solo para maior estado de energia. Portanto, interpretações do espectro Vis-NIR do solo podem ser difíceis, particularmente quando o constituinte (espécie química) está presente apenas em pequenas quantidades no solo ou quando a propriedade do solo sendo estudada depende de correlações com a espécie química. A Tabela 1 apresenta um resumo de importantes absorções na região do MIR e a ocorrência de seus sobretons e combinações na região do Vis-NIR, que podem ser usadas para ajudar interpretação de dados espectrais de solo. Portanto, a região do Vis-NIR fornece informações úteis sobre o efeito de materiais orgânicos e inorgânicos presentes na fase sólida do solo, sendo possível identificar regiões do espectro Vis-NIR que mostram absorção devido à água, material mineral e orgânica no solo.

Estudos pioneiros buscaram investigar quais as características físicas, químicas e mineralógicas que influenciam no comportamento espectral do solo (HUNT; SALISBURG; LENHOF, 1971; MATHEWS et al., 1973; STONER et al., 1980; BAUMGARDNER et al., 1985; STONER; BAUMGARDNER, 1986; COLEMAN et al., 1991). No Brasil, os estudos do comportamento espectral dos solos brasileiros foi iniciando com os trabalhos de Formaggio (1983), Madeira Netto (1991), Stoner et al. (1991), Epiphanio et al. (1992), Demattê e Garcia (1999) na região sudeste e na região sul por Dalmolin (2002) e Dalmolin et al. (2005).

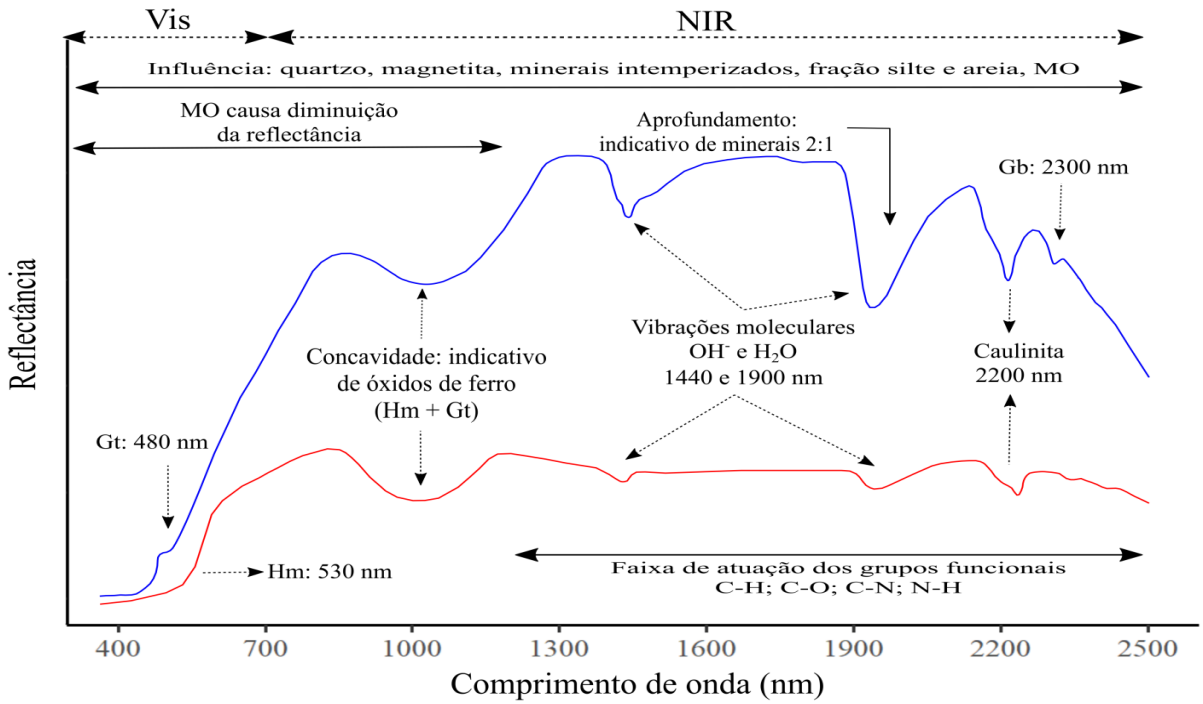
As principais regiões do Vis-NIR em que ocorrem feições de absorção relacionadas aos constituintes do solo são observadas na figura 7, onde é apresentada como exemplo, a curva espectral de um Latossolo – horizonte Bw e Argissolo – horizonte Bt, representativos da região sul do Brasil. A influência da granulometria, mineralogia, umidade e matéria orgânica do solo, na região do Vis-NIR, são apresentadas nos próximos itens.

Tabela 1 – Constituintes do solo, grupos químicos e atribuições das absorções fundamentais na região do infravermelho médio (MIR) com as sobreposições de absorções e suas combinações na região Vis-NIR

Constituintes do solo	Grupos químicos e estados fundamentais no MIR	Região Vis-NIR (nm)	Sobreposições e combinações na região Vis-NIR
Água	v_1 O – H v_2 H – O – H v_3 O – H	1915 1455 1380 1135 940	$v_2 + v_3$ $2v_2 + v_3$ $v_1 + v_3$ $v_1 + v_2 + v_3$ $2v_1 + v_3$
Hidroxilas	v_1 O – H	1400 930 700	$2v_1$ $3v_1$ $4v_1$
Oxidos de ferro e alumínio			
Goethita		434, 480 650, 920	eletrônico transições
Hematita		404, 444, 530 650, 884	eletrônico transições
Gibbsita		2300	eletrônico
Minerais da fração argila			
Caulinita	v_{1a} O – H v_{1b} O – H δ Al – OH	1395 1415 2160 2208	$2v_{1a}$ $2v_{1b}$ $v_{1a} + \delta$ $v_{1b} + \delta$
Esmectita	v_1 O – H δ_a Al – OH δ_b AlFe – OH	2206 2230	$v_1 + \delta_a$ $v_1 + \delta_b$
Ilita	v_1 O – H	2206 2340 2450	$v_1 + \delta$ indefinido
Carbonatos	v_3 CO_3^{2-}	2336	$3v_3$
Compostos orgânicos			
Aromáticos	v_1 C – H	1650 1100 825	$2v_1$ $3v_1$ $4v_1$
Amina	δ N – H v_1 N – H	2060 1500 1000 751	$v_1 + \delta$ $2v_1$ $3v_1$ $4v_1$
Alquila assimétrica Alquila simétrica	v_3 C – H v_1 C – H	1706 1754 1138 1170 853 877	$2v_3$ $2v_1$ $3v_3$ $3v_1$ $4v_3$ $4v_1$
Ácido carboxílico	v_1 C = O	1930 1449	$3v_1$ $4v_1$
Amida	v_1 C = O	2033 1524	$3v_1$ $4v_1$
Alifático	v_1 C – O	2275 1706	$3v_1$ $4v_1$
Metil	v_1 C – H	2307-2469 1730-1852	$3v_1$ $4v_1$
Fenólicos	v_1 C – OH	1961	$4v_1$
Polissacarídeos	v_1 C – O	2127	$4v_1$
Carboidratos	v_1 C – O	2381	$4v_1$

Fonte: Compilação adaptada de Viscarra Rossel e Behrens (2010).

Figura 7 – Curva espectral de uma amostra de Latossolo – horizonte Bw (curva vermelha) e Argissolo – horizonte Bt (curva azul) indicando a principais regiões do espectro Vis-NIR influenciadas pelos constituintes do solo. Em que: MO – matéria orgânica; Gt – goethita; Hm – hematita; Gb – gibbsita

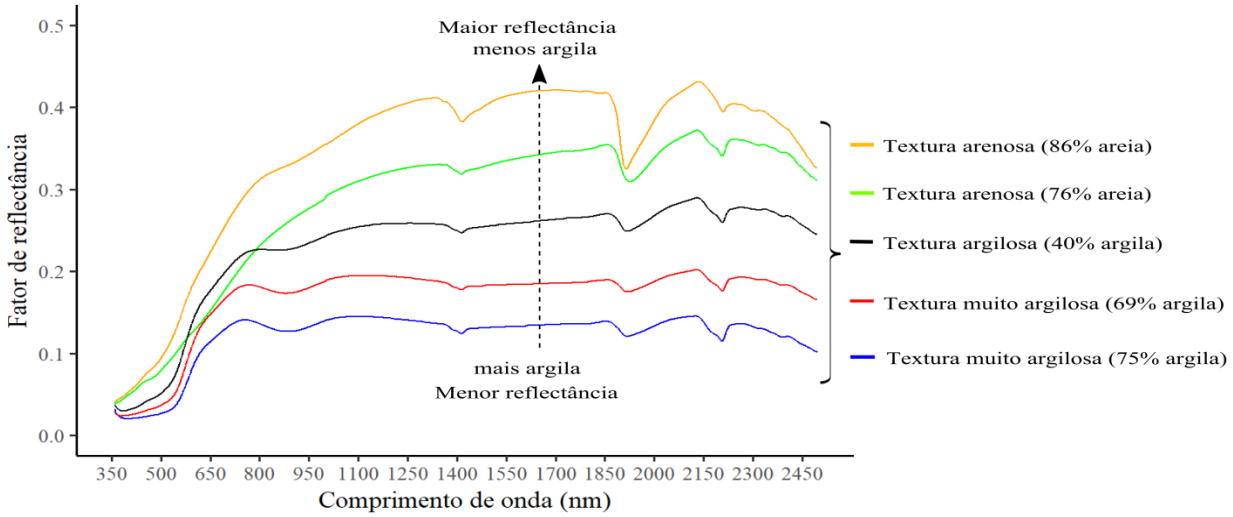


Fonte: Autor.

2.4.1. Efeito da granulometria

O solo é considerado uma superfície rugosa, por isso ocorre à reflectância difusa (energia incidente é refletida uniformemente em todas as direções), sendo essa maior quando o tamanho das partículas é maior. Existe um efeito inverso entre o tamanho das partículas e a reflectância do solo, sendo a absorção maior quanto menor forem as partículas, reduzindo a reflectância (Figura 8). Os solos com altos teores de areia tendem a ter maior reflectância devido à sua composição mineralógica rica em quartzo. Além disso, estes são solos frágeis em termos agrícolas e apresentam baixos teores de matéria orgânica, conteúdo de água e argilominerais. Stoner e Baumgardner (1981) observaram aumento na reflectância de solos associado ao aumento das frações areia fina e muito fina e um decréscimo na reflectância na faixa de 2080 a 2320 nm para solos argilosos.

Figura 8 – Curvas espectrais mostrando o comportamento espectral do solo de amostras mais argilosas e menos argilosas da região sul do Brasil



Fonte: Autor

Em amostras da biblioteca espectral de solos do Brasil analisadas por Sato (2015), foi observada maior absorção de energia em amostras com textura argilosa ou mais fina ao longo do espectro Vis-NIR. As regiões do espectro com feições de absorção significativas foram às regiões entre 520 e 600 nm, 1150 a 1300 e 2030 a 2035 nm. Coleman et al. (1991), que, estudaram o comportamento espectral de solos, no norte do Alabama, EUA, e utilizaram modelos estatísticos para procuraram quantificar propriedades do solo, observaram que essas regiões apresentam correlação com os teores de argila, sendo indicadas como regiões ativas para estimar esse atributo do solo.

2.4.2. Efeito da mineralogia

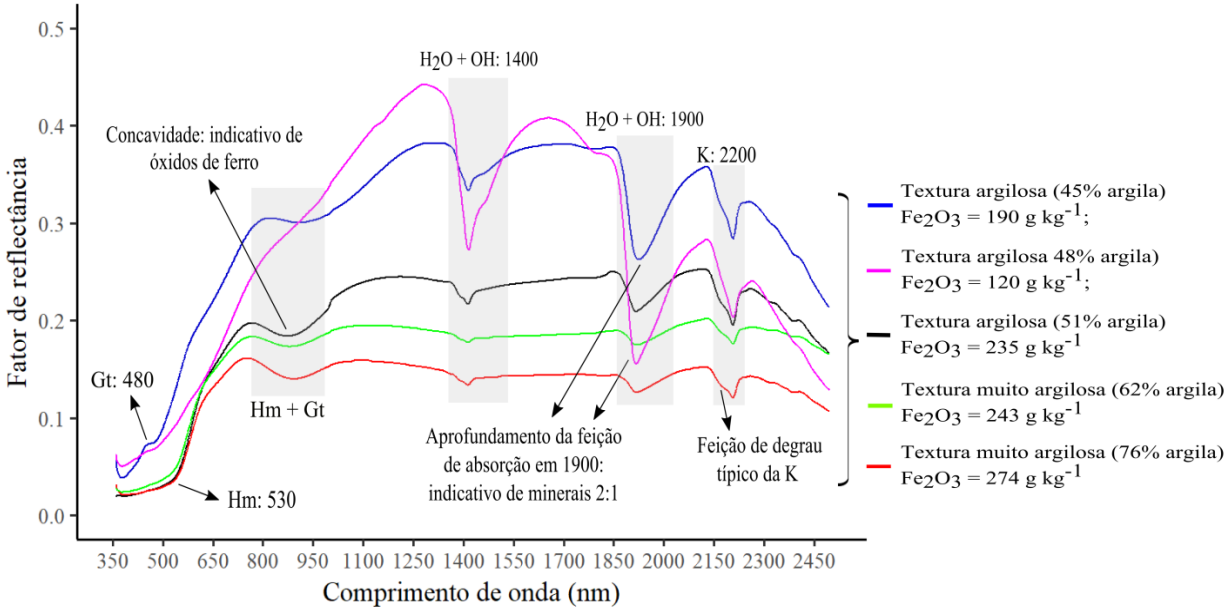
Considerando que o comportamento espectral de cada mineral depende da sua constituição básica, solos com diferentes mineralogias podem apresentar comportamento espectral distintos, baseados em suas unidades constituintes. Conforme o solo é intemperizado a relação entre oxigênio e silício e oxigênio alumínio é alterada, causando modificações nas vibrações eletrônicas. Os argilominerais frequentemente encontrados em amostras de solos são a

caulinita (1:1) e montmorilonita (2:1), em que os espectros de reflectância apresentam feições de absorção centrados em 1400 e 2200 nm (Figura 9), reflexos das vibrações dos grupos OH e ligações Al-OH presentes na estrutura desses argilominerais. A caulinita pode ser observada na região de 2200 nm (Figura 9), resultado da influência de OH (MATHEWS et al., 1973; STONER; BAUMGARDNER, 1986). Essa feição característica da caulinita também foi observada por Sato (2015) na maior parte das amostras da biblioteca espectral de solos do Brasil. Já feições relacionadas ao argilomineral ilita pode ser observada na região de 1700 nm, como observado em solos da Pensilvânia, EUA por MATHEWS et al., 1973.

Os óxidos de ferro hematita e goethita influenciam no comportamento espectral do solo, apresentando feições características na região do Vis-NIR entre 400 a 1100 nm principalmente devido a excitações eletrônicas associadas com esses minerais (STONER; BAUMGARDNER, 1986). Nesse intervalo ocorrem as transições eletrônicas que correspondem à presença de ferro nas formas de Fe^{+2} (íon ferroso) e Fe^{+3} (íon férrico), sendo essas vibrações observadas pela absorção de energia nesses comprimentos de onda (Figura 9). Essas formas de ferro estão presentes nos óxidos de ferro hematita e goethita, conferindo aos espectros acentuada queda de reflectância em direção à região dos menores comprimentos de onda, sendo observada forte absorção próxima a 400 nm. A hematita apresenta feição centrada em 530 nm e a goethita em 480 nm relacionadas a transições eletrônicas (Figura 9). Além disso, se observa forte absorção para hematita em torno de 859 nm e para goethita em 650 e 900-940 nm devido às transições no campo cristalinos desses óxidos (Figura 9). Em relação aos óxidos de alumínio, a gibbsita apresenta feições de absorção devido à vibração das moléculas do grupo OH, sendo observado em 2300 nm.

Feições de absorção de energia relacionadas aos óxidos de ferro na região entre 400 e 550, 650, 850 e 900 nm foram observadas por Stoner et al. (1991), Dalmolin (2002) e Sato (2015) em Latossolos brasileiros. Stoner et al. (1991) relata em seu estudo que teores de ferro maiores que 4 % podem mascarar o efeito do conteúdo de matéria orgânica. Em Latossolos da região do Planalto do Estado do RS, feições de absorção na região de 450-480 nm para goethita e 800-1100 nm para hematita + goethita, foram observadas em estudos de Dalmolin (2002) e Dalmolin et al. (2005). Demattê e Garcia (1999) relataram que formas de ferro amorfo também influenciam na reflectância ao longo de todo o espectro (400 a 2500 nm) e que os óxidos de ferro apresentaram feições típicas nas bandas 480 e 850 nm.

Figura 9 – Curvas espectrais mostrando a influencia da mineralogia no comportamento espectral do solo em amostras da região sul do Brasil. Hematita (Hm); Goethita (Gt); Caulinita (K)



Fonte: Autor.

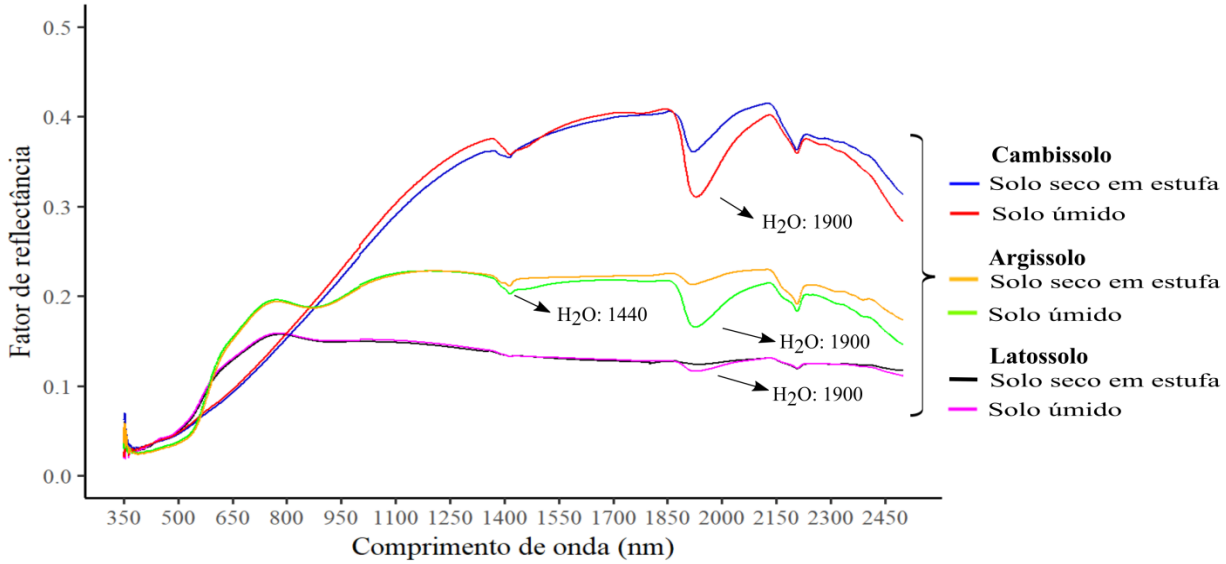
2.4.3. Efeito da umidade

A umidade também é um fator que influencia na resposta espectral do solo, em que solos mais úmidos podem se apresentar mais escuros e opacos, causando diminuição na reflectância (STONER et al., 1991). O efeito da água na reflectância do solo pode causar decréscimo da reflectância em todos os comprimentos de onda, sendo observadas bandas de absorção centradas em 760, 970, 1190, 1440 e 1940 nm (STONER et al., 1991).

A presença de moléculas de água na amostra resulta em duas feições típicas de absorção sendo em 1440 e 1900 nm (Figuras 10). As feições no espectro das vibrações das moléculas podem aparecer bem agudas, relacionadas a moléculas bem definidas, ou feições muito largas relacionadas a moléculas desordenadas. Quando apresenta apenas a absorção em 1400 nm, indica a presença de hidroxilas em moléculas de água na estrutura de minerais 2:1. Isso pode ser observado nos comprimentos de onda de 2000 a 2500 nm, onde ocorrem feições de absorção relacionadas com ligações metal-hidroxilas (M-OH), onde o metal pode ser Fe, Mg e Al. Em

geral, a banda de absorção próxima a 2200 nm está relacionada a presença de Al-OH, em 2300 nm a Mg-OH e 2290 nm por Fe-OH.

Figura 10 – Curvas espectrais mostrando a influencia da umidade no comportamento espectral do solo em amostras da região sul do Brasil



Fonte: Autor.

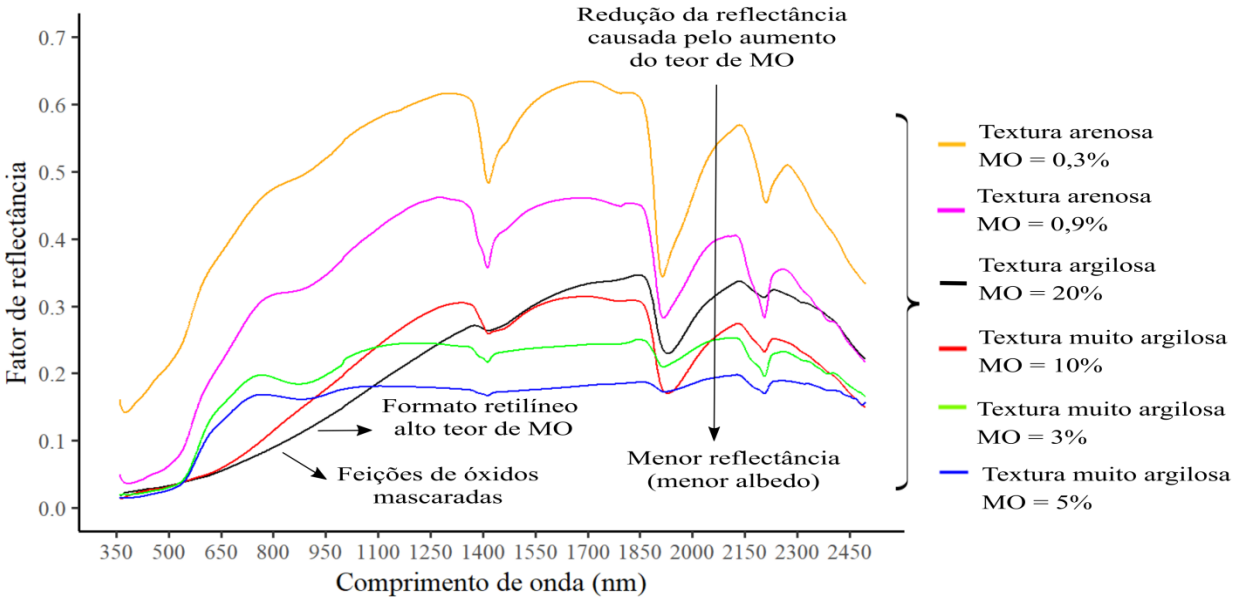
Porém, os efeitos da umidade são dependentes do comprimento de onda e da posição e intensidade das feições espectrais do solo em estudo. Os Latossolos da região sul do Brasil com altos teores de ferro, por exemplo, apresentam fortes bandas de absorção na região do visível, relacionadas aos óxidos, apresentando alta opacidade, em que os diferentes regimes de umidade não causaram mudança significativa na resposta espectral do solo (Dalmolin, 2002) (Figura 10).

2.4.4. Efeito da matéria orgânica

A matéria orgânica (MO), em uma determinada porcentagem, reduz a reflectância do solo e modifica o formato da curva espectral (Figura 11) e pode mascarar as feições espectrais diagnósticas de outros componentes presentes no solo, por exemplo, região relacionada aos óxidos de ferro (500-1000 nm) (Figura 11).

A MO tem varias regiões de absorção em toda a faixa espectral do Vis-NIR (BEN-DOR; BANIN, 1995), sendo essas regiões associadas aos grupos OH, SO_4, CO_3 e combinações de H_2O e CO_2 presentes na composição da MO (CLARK, 1999). No Vis-NIR, a combinação de bandas e sobretons resultam do alongamento e flexão de grupos N–H, C–H, C–O, C–OH e C=O, conforme apresentado em revisão por Viscarra Rossel e Behrens (2010). Feições de absorção em torno de 1400 nm, que estão relacionadas à molécula de água na estrutura de algolominerais, mas também são associados à água ligada á celulose em compostos orgânicos. A MO também pode apresentar feições de absorção fracas em torno de 2300 nm (VISCARRA ROSSEL; BEHRENS, 2010), que às vezes pode ser confundida com absorções de hidroxila e carbonato em argilominerais (CLARK et al., 1999).

Figura 11 – Curvas espectrais mostrando a influencia da matéria orgânica (MO) no comportamento espectral do solo em amostras da região sul do Brasil



Fonte: Autor.

Trabalhos pioneiros a respeito da influência da MO no comportamento espectral do solo forma realizados por Stoner e Baugardner (1981), que observaram a diferença no comportamento espectral do solo conforme o conteúdo de MO. Foi observado, que solos com textura média e com alto teor de MO apresentam baixa refletância e feições côncavas de 500 a 1300 nm. Já nos

solos com baixo teor de MOS foi observado alta reflectância de 500 a 1300 nm. Em amostras com altos teores de óxidos de Fe e argila, observou-se que a curva espectral foi influenciada pelos óxidos, apresentando albedos muito baixos e valores crescentes de absorção nos comprimentos de onda maiores que 800 nm.

Madeira Netto (1991) ao estudar três Latossolos com diferentes teores e tipos de óxidos de ferro, observou comportamento diferenciado na reflectância dos mesmos quando a MO foi removida. Nos solos com baixos teores de ferro, houve aumento da reflectância ao longo do espectro, e solos com teores mais elevados de ferro tiveram aumento na reflectância até 600 nm e decréscimo até 2500 nm. Estudo de Ben-Dor et al. (1997) relatou que diferentes estágios de decomposição da MO conferem diferente comportamento espectral na região do Vis-NIR. Em solos da região Sudeste do Brasil, Demattê e Garcia (1999) observaram aumento na reflectância das amostras de solo com o decréscimo do conteúdo de MO, sendo que a feição relacionada aos óxidos de Fe (900 nm) foi evidenciada nas amostras com menor teor de MO. Galvão e Vitorello (1998) notaram que a presença de 1,7 % de MO foi capaz de mascarar as feições espectrais do ferro presentes nas amostras.

Dalmolin (2002) e Dalmolin et al. (2005) observaram a influência da MO na forma e no albedo da curva espectral ao longo de todo o espectro Vis-NIR para amostras de Latossolo no Estado do RS. Os autores observaram que o aumento do teor de MO no solo implica no escurecimento do mesmo, causando uma maior absorção de radiação eletromagnética e, consequentemente, uma reflectância menor, sendo que a presença da MO reduz o albedo ao longo da região do Vis-NIR (Figura 11). Bellinaso (2009) observou diminuição na reflectância na região dos 350 a 1350 nm devido à influência da MO nos horizontes superficiais de amostra de Latossolos e Nitossolos, além de uma suavização das feições côncavas dos óxidos de ferro na região de 800 a 900 nm. Essa alteração no formato da curva e redução no albedo em função da MO também foi observada em trabalho de Sato (2015), que analisou amostras da biblioteca espectral do Brasil.

2.5. PREDIÇÃO DO CARBONO ORGÂNICO DO SOLO POR ERD

A predição do COS é realizada por meio de ajuste (calibração) de modelos matemáticos entre dados de referência variável de solo (variável dependente) e variáveis preditoras (variáveis

independentes), neste caso, dados espectrais obtidos pela técnica de ERD. Existem vários métodos estatísticos multivariados que podem ser utilizados para construir esses modelos, sendo o mais simples a regressão linear múltipla (*multiple linear regression – MLR)*, representa pela equação 1.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \quad (1)$$

onde Y_i é o valor da variável dependente na *i*-ésima observação, β_0, β_1, β_2 ... β_{p-1} são os coeficientes, X_{i1}, X_{i2} ... $X_{i,p-1}$ são constantes conhecidas, ou seja, o valor das variáveis preditoras na *i*-ésima observação, ε_i é o erro aleatório na *i*-ésima observação e $i = 1, \ldots, n$ observações (VARMUZA; FILZMOSER, 2009).

As pesquisas com uso da técnica de ERD para a predição do COS, ganharam destaque em escala internacional, com a trabalhos publicados por Al-Abbas et al. (1972), Krishnan, Butler e Hummel (1980), Dalal e Henry, (1986) e Ben-Dor et al. (1995) que utilizaram como método multivariado a MLR e suas variações. No Brasil, os estudos são mais recentes, sendo o primeiro trabalho realizado por Demattê e Garcia (1999), que utilizaram informações das assinaturas espectrais do solo para estimar por MLR o teor de COS e as formas de ferro para um grupo de solo do Estado do Paraná. Os resultados mostraram que a ERD é uma técnica eficiente na estimativa dessa propriedade do solo

Com o avanço na última década em relação à capacidade computacional, métodos estatísticos mais robustos e banco de dados com maior número de amostras têm sido empregados na construção dos modelos de predição do teor de COS. Os métodos mais utilizados incluem a regressão linear múltipla stepwise (*stepwise multiple linear regression – SMLR*), a regressão dos mínimos quadrados parciais (*partial least squares regression – PLSR*), as redes neurais artificiais (*artificial neural networks – ANN*), as máquinas de vetor de suporte (*support vector machines – SVM*), árvores de regressão (*regression tree – RT*), floresta aleatória (*random forest – RF*) e *Cubist*. No entanto, a análise multivariada mais utilizada para a obtenção de modelos de predição a partir de dados espectrais é a PLSR, como mostra compilação de dados realizada por Vasques et al (2008), Stenberg et al. (2010) e Dotto et al. (2018).

O método *partial least squares (PLS)* original foi desenvolvido por volta de 1975 pelo estatístico Herman Wold para um tratamento de cadeias de matrizes e aplicações em econometria (VARMUZA; FILZMOSER, 2009). O uso combinado das técnicas espectroscópicas e a regressão por PLS (*partial least squares regression – PLSR)* teve início com a publicação do trabalho de Haaland e Thomas (1988), que quantificou diferentes compostos químicos. Para a quantificação de propriedades solo, em especial para o COS, pode-se citar os trabalhos pioneiros de Janik e Skjemstad (1995), Masserschmidt et al. (1999) e Dunn et al. (2002). O método PLSR pode manipular conjuntos de dados contendo mais variáveis independente do que dependentes e aceita variáveis altamente correlacionadas, como é o caso de dados espectrais Vis-NIR. A análise de mínimos quadrados parciais (*partial least squares)* reduz grandes conjuntos de dados em um pequeno número de fatores ortogonais, não correlacionados, de modo a minimizar a soma dos quadrados dos erros entre os valores a serem preditos (VARMUZA; FILZMOSER, 2009). A estimativa da variável dependente, neste caso o conteúdo de COS, pode ser realizada pela equação 2.

$$Y_i = \beta_0 + \sum_{i=1}^{n} (X_i * \beta_i) + \varepsilon_i \qquad (2)$$

onde Y_i é o valor da variável dependente na *i*-ésima observação, β_0 é o fator de correção, β_i são os coeficientes de regressão para cada região do espectro, X_i são constantes conhecidas, ou seja, o valor das variáveis preditoras na *i*-ésima observação, ε_i é o erro aleatório na *i*-ésima observação e $i = 1, \ldots, n$ observações (VARMUZA; FILZMOSER, 2009).

Para avaliar a incerteza (erro da predição), que sempre está associada aos modelos preditivos, são utilizadas algumas medidas de erro, como por exemplo, erro médio (EM) ou diferença sistemática entre os valores preditos e medidos (BIAS), a raiz do erro médio quadrático (RMSE), desvio padrão do erro (SDE) e a proporção de desempenho no intervalo interquartil (RPIQ), que explica a disseminação do conjunto de dados com base na distância interquartil, que fornece uma informação de acurácia mais realista das predições (BELLON-MAUREL et al., 2010). A medida de erro mais frequentemente utilizada na literatura para calcular as diferenças

entre os valores preditos pelo modelo e os valores efetivamente observados é RMSE (VASQUES et al., 2008; STENBERG et al., 2010).

Trabalhos realizados no Brasil têm mostrado que existe uma variação na acurácia dos modelos de predição para o teor de COS para solos brasileiros, sendo os estudos das causas de tal fato ainda incipientes. Fiorio e Demattê (2009) desenvolveram modelos para solos, oriundos do sudeste do Estado de São Paulo, a partir de técnicas ERD com sensor de laboratório para a predição do teor de matéria orgânica, onde o R^2 foi de 0,30. Os autores observaram que os dados espectrais obtidos em laboratório foram mais eficazes quando comparados com os dados orbitais, devido a uma melhor resolução espectral em laboratório e interferência atmosférica em dados orbitais. Sousa Junior, Demattê e Araújo (2011) geraram modelos a partir de dados espectrais e, como modelo, utilizaram a MLR para predição do COS, no qual o modelo apresentou R^2 de 0,38. Dotto et al. (2014) gerou modelos de predição para o COS em uma bacia hidrográfica de 1.892 ha, com relevo complexo e distintos tipos de solos e uso da terra, localizada na região central do Estado do RS, e alcançou R^2 de 0,62 e logRMSE de 0,20 g kg^{-1}.

Estudos têm apontado que a confiabilidade e precisão dos modelos são dependentes das dos pré-processamentos aplicados nos dados espectrais e do uso de modelos matemáticos apropriados (VASQUES et al., 2008; KNOX et al., 2015; DOTTO et al., 2017). Em relação aos pré-processamentos aplicados na curva espectral, pode- se citar a suavização com variações no número de janelas de filtro, primeira derivada com filtro Savitzky-Golay, remoção do contínuo e normalização por faixa. Esses processamentos e suas variações foram utilizados em trabalho realizado por Vasques et al. (2008) e Dotto et al. (2017). Os autores concluíram que a escolha do pré-processamento depende da técnica multivariada a ser utilizada para construção de modelos preditivos. Além disso, os processamentos auxiliam na interpretação e extração de informações dos espectros de reflectância e ressaltam feições espectrais relacionadas com as propriedades do solo.

Dotto et al. (2017), que avaliou a performance dos modelos PLSR e SVM e quatro técnicas de pré-processamento na capacidade preditiva do COS em escala local, alcançou RMSE de 0,45 % - PLSR e 0,44 % - SVM. Nesse mesmo estudo, os autores também observaram alta variação no desempenho dos modelos preditos em relação a cada pré-processamento espectral, sendo aponta a grande variação das características do solo na paisagem como as causas da baixa acurácia das predições. Araújo et al. (2014), na tentativa de melhorar o desempenho de previsão

dos modelos, trabalho de realizado em solos tropicais do Brasil, alcançou um R^2 de 0,60 e RMSE de 0,55 % na predição do COS e quando o conjunto de dados foi particionada em grupos mais uniformes em relação à mineralogia, foi observado aumento na acurácia dos modelos para R^2 = 0,71 e RMSE = 0,47 %. Da mesma forma, Demattê et al. (2016) abordou a regionalização espectral em escala local e regional para a predição de propriedades do solo no Brasil, em que observou que a presença de solos com alta variação de seus constituintes prejudica a predição de propriedades químicas do solo.

Como citado anteriormente, a composição organomineral do solo influencia a absorção da radiação eletromagnética, causando variação da reflectância espectral (DALMOLIN et al., 2005; STENBERG et al., 2010; DEMATTÊ; TERRA, 2014). Estudo de Viscarra Rossel et al. (2006) aponta que distribuição da variável dependente, a composição do solo e a origem dos solos da BES são as principais causas da variação observada na acurácia das predições do COS. Além, o COS apresenta feições fracas, comparadas as feições do argilominerais, que são tipicamente mais significativas (STENBERG; VISCARRA ROSSEL, 2010). Tal fato pode marcar as feições ativas para predição do COS, especialmente quando características espectrais de argilominerais e COS coincidirem. Logo, se a mineralogia e a textura diferirem substancialmente entre as amostras que compõem a BES, as feições do COS são mais difíceis de separar na calibração dos modelos preditivos. Isso foi observado em estudo de Stenberg, Jonsson e Börjesson (2002), que utilizaram um conjunto de amostras de solos suecos, em que os autores encontraram baixa acurácia nas predições de COS, mas que foi significativamente aumentada quando se excluiu os solos com menor teor de argila. Resultados semelhantes foram observados por Stenberg (2010), que estudou o efeito do areia na predição de COS em solos suecos, onde observou que as predições eram mais imprecisas para solos com alto teor de areia.

O efeito da variação das características do solo no comportamento espectral e, consequentemente, nos modelos de predição de propriedades do solo foi relatado por Viscarra Rossel et al. (2016), que estudaram a biblioteca espectral global, composta por 23.631 amostras. Os autores mostraram que variações nos dados espectrais podem ser associadas com a variabilidade pedológica, ambiental e com sua distribuição geográfica global. Também observaram que modelos regionais nem sempre foram mais precisos que o global, com exceção do modelo de predição para o COS, onde o agrupando dos espectros em classes espectrais mais homogêneas, independentemente da localização geográfica, melhorou a acurácia das predições.

Estudo de Wight, Ashworth e Allen (2016), mostrou que os constituintes da matéria orgânica podem afetar a capacidade preditiva dos modelos pela técnica de ERD na região do Vis-NIR. Os autores recomendam que amostras de solo de diferentes usos da terra (por exemplo, floresta nativa, campo nativo e lavouras) possam ser separadas em grupos para realizar a predição do COS separadamente, com a intenção de melhorar a acurácia das predições. Também foi observado efeito da textura do solo, em que as amostras com maior teor de argila apresentaram a maior porcentagem de erros. Nesse caso, os autores sugerem que, se for conhecida previamente a textura do conjunto de amostras, esse deve ser separada em grupos para realizar a predição do COS, com base na classe textural. Porém, essa separação em grupos, ou estratificação com base em critérios pré-definidos reduz o número de amostras para calibração dos modelos. Tal fato, também tem efeito sobre o desempenho e acurácia das predições, principalmente quando se trata de conjuntos pequenos de amostras (RAMIREZ-LOPEZ et al., 2014).

Modelos de calibração mais precisos também podem ser alcançados pela seleção de variáveis espectrais mais informativas, ao invés de utilizar o espectro completo (XIAOBO et al., 2010). Nesse trabalho, os autores apresentam a importância da seleção de variáveis espectrais e uma revisão dos diferentes de algoritmos utilizados para esse propósito. Essa abordagem foi utilizada em trabalho recente de Vohland et al. (2014), visando a melhoria da predição de propriedades do solo. A técnica utilizada foi a simulação de Monte Carlo, com a finalidade de realizar amostragens aleatórias que selecionem um conjunto mais adequado de alguns comprimentos de onda que mais se correlacionam com a propriedade do solo de interesse.

Neste contexto, a complexidade das interações organominerais que ocorrem no solo influenciam na acurácia dos modelos de predição (VISCARRA ROSSEL et al., 2006; BELLON-MAUREL; MCBRATNEY, 2011). Essa hipótese se sustenta à medida que aumenta o número de amostras que compõem uma BES e, principalmente, quando a extensão geográfica de cobertura dessas amostras cresce, por exemplo, escala local > regional > nacional > global, pois cresce a hetorogeneidade das características do solo. Esse efeito na variação do solo em diferentes escalas é observado na região sul do Brasil, onde o solo e suas propriedades são reflexos da heterogeneidade dos fatores de formação clima, organismos, relevo e material parental (ALMEIDA, 2017; DALMOLIN et al., 2017; STRECK et al.., 2018).

3. HIPÓTESES

i) Processamentos espectrais e métodos estatísticos multivariados influenciam na acurácia dos modelos de predição do teor de carbono orgânico do solo.

ii) A estratificação de bibliotecas espectrais solo heterogêneas em sub-bibliotecas mais homogêneos aumenta a acurácia dos modelos para a predição do teor de carbono orgânico do solo.

iii) A predição do teor de carbono orgânico do solo é mais acurada em solos com textura argilosa.

4. OBJETIVOS

4.1. GERAL

Avaliar a influência de processamentos espectrais, métodos estatísticos multivariados e estratificação amostral de bibliotecas espectrais na acurácia dos modelos de predição para o teor de carbono orgânico do solo.

4.2. ESPECÍFICOS

i) Avaliar o efeito de técnicas de pré-processamentos e métodos multivariados na acurácia dos modelos de predição;

ii) Avaliar o efeito da estratificação amostral de BES na acurácia dos modelos de predição;

iii) Identificar a influência composicional do solo na variação espectral das amostras;

iv) Avaliar a influência da variância do teor de carbono orgânico do solo, composição granulométrica e espectral na acurácia das predições;

v) Definir as características da BES que possam impactar na acurácia dos modelos de predição;

5. ESTUDO 1 – ESTRATIFICAÇÃO DE UMA BIBLIOTECA ESPECTRAL LOCAL POR CRITÉRIOS DE HOMOGENEIDADE PRODUZ PREDIÇÕES DE CARBONO ORGÂNICO DO SOLO MAIS PRECISAS

5.1. INTRODUÇÃO

Intensas modificações no uso da terra, alterações climáticas e a demanda em simular cenários futuros têm refletido no aumento do interesse pelas informações espaciais em solos (MCBRATNEY; FIELD; KOCH, 2014; AMUNDSON et al., 2015). Como resposta, pesquisas em ciência do solo têm focado no desenvolvimento de métodos que permitam a predição em escala local, regional, nacional e global de um vasto conjunto de propriedades do solo (GRUNWALD; VASQUES; RIVERO, 2015). O COS responsável por uma série de funções no ambiente, e chave no ciclo global do carbono (STOCKMANN et al., 2015). Em razão disso, surge a necessidade de gerar informações confiáveis sobre o COS. No entanto, o mapeamento desta propriedade do solo demanda tempo e torna-se dispendioso em decorrência da elevada intensidade amostral requerida (VASQUES et al, 2010), devido a diversidade de fatores ambientais e antrópicos que influenciam sua distribuição na pedosfera.

Assim, são necessários métodos mais rápidos e econômicos para quantificação do COS. A técnica ERD na região Vis-NIR e a organização das informações espectrais do solo em BES são ferramentas com potencial de capturar a variabilidade pedológica e quantificar propriedades do solo (BEN-DOR; BANIN, 1995; NOCITA et al., 2015; GUERRERO et al., 2016; VISCARRA ROSSEL et al., 2016; DEMATTÊ et al., 2017; DOTTO et al., 2017; PINHEIRO et al., 2017; VASÁT et al., 2017). Essas ferramentas apresentam uma série de vantagens em relação às metodologias tradicionais de determinação de COS, como rapidez das análises, não geram resíduos químicos (cromo – Cr) e apresentam baixo custo. Tais dados foram confirmados em estudo de O'Rourke e Holden (2011), que avaliaram o custo e tempo envolvido na análise do COS entre os métodos Walkley-Black e a ERD Vis-NIR, e concluíram que a ERD apresenta um custo 10 vezes menor, associado à rapidez das análises, principalmente quando se deseja avaliação do solo em grandes áreas, onde o número de amostras é elevado. Da mesma forma, Schwartz, Ben-Dor e Eshel (2012) encontraram resultados mais precisos, a um custo menor,

utilizando a ERD para monitorar a contaminação por petróleo em solos, comparado aos métodos tradicionais de química analítica.

A quantificação de propriedades do solo, em especial o COS, a partir dessa metodologia se dá pela construção de modelos de predição, utilizando-se diferentes métodos multivariados e técnicas de pré-processamento de dados espectrais a partir de BES de escala local, regional, nacional e global (Tabela 2). Nesses estudos, se observa que a definição da escala está relacionada com a abrangência geográfica das amostras de solo coletadas. Em geral, BES locais se limitam a fazendas, bacias hidrográficas e munícipios. Já as regionais abrangem, como por exemplo, regiões fitográficas de um Estado e/ou as cinco regiões do Brasil. As nacionais contemplam os países e a global considera todo o globo terrestre. O número de amostras em cada escala de BES é bastante variável, sendo isso uma característica particular de cada estudo.

Outro aspecto verificado na tabela 2 é que a acurácia das predições de COS apresenta variação entre os trabalhos, sendo isso já destacado na literatura (VISCARRA ROSSEL et al.,2006; STENBERG et al 2010). Há várias explicações possíveis para essa variação na acurácia dos modelos preditivos. Uma delas é o tipo de pré-processamento aplicado nos dados espectrais e os métodos multivariados de predição utilizados no ajuste dos modelos. Trabalhos têm mostrado que diferentes pré-processamentos aplicados aos dados espectrais resultam em distintos desempenhos modelos de predição para o COS (DUNN et al., 2002; VASQUES et al. 2008; KNOX et al., 2015; DOTTO et al., 2017). Em relação aos métodos multivariados, Vasques et al. (2008) encontraram melhores resultados para regressão por mínimos quadrados parciais - *Partial Least Square Regression (PLSR)* comparado com a regressão linear múltipla - *Multiple Linear Regression (MLR*), sendo o inverso observado por Vasát et al. (2017). Já estudos de Sequeira et al. (2014) e Knox et al. (2015), com a técnica de floresta aleatória - *Randon Forest (RF)* apresentaram vantagens em relação a PLSR quando os valores de COS não apresentam distribuição normal. Viscarra Rossel e Behrens (2010) constataram que o modelo máquina de vetor de suporte - *Support Vector Machines (SVM)* gerou os modelos mais acurados, quando comparado com PLSR e RF. Lucà et al. (2017) observaram maior capacidade preditiva de SVM em relação às PLSR e *principal components regression* (PCR), além de diferenças conforme o número de amostras na calibração dos modelos. Já Dotto et al. (2017) não observaram diferenças significativas entre métodos PLSR e SVM na predição de COS.

Tabela 2 – Compilação de trabalhos de predição de COS com diferentes métodos multivariados, técnicas de pré-processamento e escalas de bibliotecas espectrais

Fonte	Escala da BES	Método multivariado[(1)]	Processamento espectral[(1)]	Amostras	R^2	$RMSE_v$[(1)]
Dotto et al., 2017	Local	PLSR	CR	299	0,90	0,32 %
		SVM			0,87	0,35 %
		PLSR	DET		0,86	0,36 %
		PLSR	RB + CR		0,79	0,46 %
		SVM	RB + CR		0,69	0,53 %
		PLSR	RB + DET		0,72	0,50 %
Jiang et al., 2017	Local	PLSR	PCA			
Lucà et al., 2017	Local	PCR	SNV	216	0,69	0,88%
		PLSR			0,79	0,71%
		SVM			0,82	0,68%
Pinheiro et al., 2017	Local	PLSR	SNV	434	0,71	5,69 g kg^{-1}
Vasát et al., 2017	Local	MLR	MW-CR	107	0,81	0,16 %
		RF			0,75	0,18 %
		PLSR			0,78	0,17 %
		SVM			0,72	0,19 %
Demattê et al., 2016	Regional	PLSR	PCA	7185	0,63	2,73 mg kg^{-1}
		PLSR		3093	0,61	1,73 mg kg^{-1}
		PLSR		903	0,65	2,54 mg kg^{-1}
	Local	PLSR		843	0,66	1,31 mg kg^{-1}
		PLSR		563	0,54	3,30 mg kg^{-1}
		PLSR		621	0,62	2,49 mg kg^{-1}
		PLSR		541	0,61	3,52 mg kg^{-1}
Wijewardane et al. (2016)	Nacional	PLSR	-	19.891	0,83	7,38%
		ANN			0,96	3,61%
Guerrero et al., 2016	Nacional	PLSR	1st DER	3482	0,91	9,93 g kg^{-1}
		PLSR		1096	0,82	25,6 g kg^{-1}
		PLSR		362	0,80	25,4 g kg^{-1}
	Regional	PLSR	1st DER	147	0,21	45,8 g kg^{-1}
		PLSR		73	0,46	31,5 g kg^{-1}
		PLSR-SM			0,96	6,21 g kg^{-1}
	Local	PLSR-SM	1st DER	123	0,94	7,65 g kg^{-1}
		PLSR			0,39	92,5 g kg^{-1}
Viscarra Rossel et al., 2016	Global	Cubist	CRR	23.631	0,92	1,11%
Knox et al., 2015	Nacional	PLSR	SGD 1st DER	1014	0,87	0,40 log g kg^{-1}
		RF		1014	0,88	0,38 log g kg^{-1}
Terra et al., 2015	Regional	SVM	PCA	1259	0,72	0,14 log g kg^{-1}
Araújo et al., 2014	Regional	PLSR		7172	0,60	0,55
		SVM			0,69	0,48
Dotto et al., 2014	Local	MLR	-	303	0,62	0,20 log g kg^{-1}
Gogé et al., 2014	Local	PLSR	SGD 1st DER	141	0,58	3,2 g kg^{-1}
	Local/ Nacional	PLSR	SGD 1st DER	94/2126	0,40	5,8 g kg^{-1}
		FFT			0,56	3,4 g kg^{-1}
Ramírez-López et al., 2013	Global	SVM	-	4438	0,57	0,93 %
		PLSR	-		0,50	1,08 %
	Regional	SVM	-	4200	0,54	0,27 %
		PLSR	-		0,48	0,28 %
Stevens et al., 2013	Global	SVM	SGD 1st DER	250.000	0,67	4,9 g kg^{-1}
Cambule et al., 2012	Regional	PLSR	RAW	129	0,65	0,37%
			RAW - 1st DER		0,85	0,32%
			SGD 1st DER		0,81	0,32%
			MSC		0,75	0,34%
			MSC - SGD		0,77	0,33%
			SNV		0,73	0,35%
			MSC - 1st DER		0,78	0,34%
			MSC – SGD – 1st DER		0,84	0,32%
Vasques et al., 2008	Regional	SMLR	SNV	554	0,85	0,18 log mg kg^{-1}
		PLSR	LOG		0,86	0,18 log mg kg^{-1}
		PLSR	SGF-1-9		0,85	0,18 log mg kg^{-1}
		PCR	NRA		0,84	0,18 log mg kg^{-1}
		RT	NGD-5		0,86	0,19 log mg kg^{-1}
		CT	NGD-7		0,86	0,17 log mg kg^{-1}
Brown et al., 2006	Global	BTR	SMO, 1st DER	3793	0,82	9,0 g kg^{-1}
Viscarra Rossel et al., 2006	Local	PLSR	LOG	118	0,60	1,8 g kg^{-1}
Brown et al., 2005	Regional	PLSR	SMO, 1st DER	283	0,86	1,08 g kg^{-1}
Chang; Laird, 2002	Regional	PLSR	-	108	0,89	6,2 g kg^{-1}
Dunn et al., 2002	Regional	PLSR	LOG, 1st and 2nd DER	360	0,66	-
Palacios-Orueta; Ustin, 1998	Local	PCR	-	74	0,51	-
Dalal; Henry, 1986	Regional	MLR	LOG	72	0,86	2,2 g kg^{-1}

Fonte: Autor. BES = biblioteca espectral de solo; PLRS = partial least-squares regression; SVM = Support vector machine; RF = random forest; MLR = multiple linear regression; FFT = fast fourier transform; PCR = principal components regression; RT = regression tree; CT = committee trees; BTR = boosted regression trees; MW-CR = moving window continuum removal; CR = continuum removal; ; MSC: multiplicative scatter correction; DET = detrend; RB = band ratio; SNV = standard normal variate transformation; PCA = principal components analyses; DER = derivative; LOG = Log (1/reflectance); SGF-1-9 = Savitzky–Golay 1st derivative using a 1st-order polynomial 9 windows; NRA = normalization by the range; NGD-5 and NGD-7 = norris gap derivative 5 and 7 windows; SMO = smoothing; RMSEv = root mean square error validation.

Outra explicação para as variações na acurácia dos modelos está relacionada com a complexidade intrínseca ao recurso natural solo. A mineralogia, granulometria, umidade e o teor de COS interferem na absorção da radiação eletromagnética, causando variação na reflectância (STEVENS et al., 2013; WIGHT; ASHWORTH; ALLEN, 2016). Assim, a heterogeneidade do conjunto amostral pode levar a um viés nas estimativas das propriedades do solo, em estudos que utilizam a técnica da espectroscopia (BELLON-MAUREL; MCBRATNEY, 2011). Em revisão apresentada por Stenberg et al. (2010), foi mencionada a existência de correlação positiva entre o erro das predições e o desvio padrão da propriedade do solo sob investigação, corroborando com Demattê et al. (2016), que utilizaram BES em escala regional e local. Os autores reportaram que em escala local, a presença de solos com alta variação de seus constituintes prejudica a predição de propriedades químicas do solo. A fim de contornar esse problema, pesquisas têm se dedicado a desenvolver estratégias para melhorar a acurácia das predições, como a combinação de informações de dados espectrais locais e nacionais (GOGÉ et al., 2014).

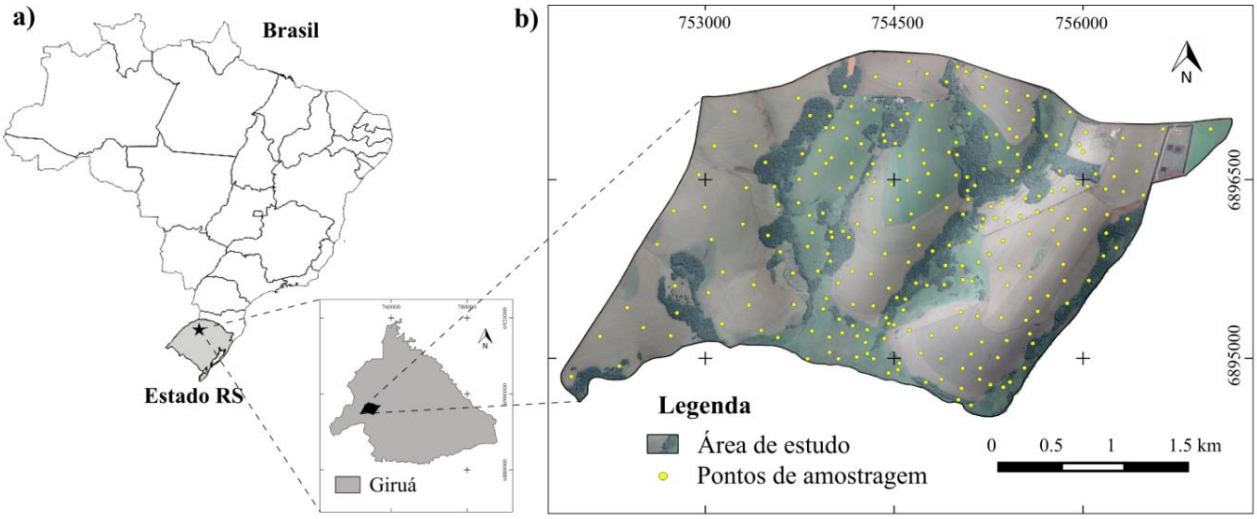
Percebe-se uma significativa variabilidade encontrada nas predições de COS na literatura e uma necessidade de mais estudos utilizando dados espectrais de escala local, com a intenção de, no futuro, serem incorporadas em BES de maior escala. Porém, há uma lacuna no conhecimento científico, no que diz respeito ao tipo de pré-processamento espectral, métodos multivariados e estratégias para construção de modelos de predição para o COS em BES locais de solos subtropicais do Brasil. Considerando que a capacidade preditiva dos modelos esteja atrelada às características dos dados espectrais, o uso de diferentes técnicas de pré-processamento e métodos multivariados, acredita-se que a estratificação de uma biblioteca espectral em grupos é uma estratégia para melhorar a acurácia das predições. Assim, os objetivos deste trabalho foram: i) comparar técnicas de pré-processamento de dados espectrais e métodos multivariados para predição do COS e, ii) avaliar a acurácia das predições do COS a partir de modelos gerados pela estratificação de uma biblioteca espectral local (BEL), com base nas características pedológicas e espectrais do solo.

5.2. MATERIAL E MÉTODOS

5.2.1. Descrição da área de estudo

O estudo foi realizado em uma área de 940 ha, localizada no município de Giruá, no Estado Rio Grande do Sul (RS), Brasil (Figura 12a). A área pertence à região fisiográfica do Planalto do RS, com altitudes que variam entre 339 e 426 m. O clima predominante é do tipo Cfa conforme a classificação de Köppen, caracterizado como clima subtropical úmido sem período de estiagem definido (ALVARES et al., 2013).

Figura 12 – Localização da área de estudo no município de Giruá, Estado do Rio Grande do Sul, Brasil (a) e área ampliada com imagem Google Earth (ano de 2016) com os pontos de amostragem (b)



Fonte: Autor.

O relevo é suave ondulado a ondulado, com declividades mais acentuadas próximo das drenagens, sendo comuns colinas com declividades de 3 a 10 %. As classes de solos que ocorrem na área são Latossolos Vermelhos Distroférricos típicos, aproximadamente 89 % da área, e Gleissolos Háplicos Tb Distrófico típicos em 11% da área. O material de origem desses solos são rochas vulcânicas do tipo basalto derivadas da Formação Serra Geral. As coberturas da terra dominantes na área são: floresta nativa (19 %), campo nativo (11 %) e lavouras anuais em sistema plantio direto (70 %).

5.2.2. Amostragem de campo e análises de solo

Foram amostrados 261 pontos (Figura 12b) nas camadas 0 – 0,05; 0,05 – 0,15 e 0,15 – 0,30 m (261 * 3 = 783), e 29 pontos nas camadas 0,30 – 0,60 e 0,60 – 1,0 m (29 * 2 = 58), conforme especificação do consórcio GlobalSoilMap (ARROUAYS et al., 2014), totalizando 841 amostras de solo. Em cada ponto amostral foram coletadas três amostras em cada profundidade para compor uma única amostra composta utilizada para as análises laboratoriais. A definição dos pontos de coleta no campo foi realizada de duas formas: (1) foi confeccionada uma malha de amostragem irregular, contendo 400 pontos, com o objetivo de representar toda a variabilidade ambiental existente na área; (2) a partir dos 400 pontos foram selecionados 261 pontos com base no conhecimento tácito do pedólogo em relação ao uso da terra, classe de solo e unidades geomorfológicas. O objetivo dessa metodologia de coleta foi capturar toda a variabilidade da área e incorporar o conhecimento tácito do pedólogo.

Em laboratório todas as amostras foram secas ao ar, destorroadas, passadas em peneira com malha de 2 mm e submetidas à análises químicas em triplicata. O teor de COS foi determinado via combustão úmida com aquecimento externo (0,5 g de solo + 10 ml $K_2Cr_2O_7$ 0,067 mol L^{-1}) e a titulação dos extratos do COS foi realizada com $Fe(NH_4)_2(SO_4)_2$ $6H_2O$ 0,5 mol L^{-1}, segundo Yeomans e Bremner (1988).

5.2.3. Medições e processamento espectral

A leitura da reflectância espectral das amostras de solo foi realizada com o uso do espectroradiômetro FildSpec 3 (Analytical Spectral Devices, Boulder, USA), com capacidade de amostragem no intervalo de 350 à 2500 nm e resolução espectral de 1 nm. Para realização da leitura foram utilizados aproximadamente 15 cm^3 de solo acondicionado em placa de petri. Como fonte de energia foi utilizada uma lâmpada de halogênio 50 W com o feixe colimado não-alvo para o plano, posicionado a 35 cm da amostra num ângulo de 30º zênite. Uma placa branca Spectralon foi utilizada como padrão de referência de 100 % de reflectância para a calibração do equipamento a cada 20 minutos. Os espectros reflectância foram obtidos para cada amostra, sendo realizadas duas leituras por amostra, com um giro de 180° das placas entre as leituras, onde cada leitura foi calculada a média a partir de 100 leituras em 10 segundos.

Para reduzir as interferências nos espectros causadas pela combinação solo-aparelho-fonte de luz-ambiente, as curvas espectrais originais foram suavizadas - *smoothed* (SMO). O SMO consiste na aplicação de uma média móvel a partir de uma função de convolação (STEVENS; RAMIREZ-LOPEZ, 2013) (Equação 3). Neste estudo foi utilizada suavização de 9 pontos, ou seja, uma janela de busca de 9 nm. A partir das curvas espectrais SMO, consideradas neste estudo como curvas testemunhas foram aplicadas cinco técnicas de pré-processamentos com frequente aplicação em estudos de espectroscopia (RINNAN; BERG; ENGELSEN, 2009; PINHEIRO et al., 2017; DOTTO et al., 2018), com objetivo de testar suas habilidades na modelagem e na previsão do COS: (i) remoção do contínuo – *continuum removed reflectance* (CRR) (Equação 4), utilizado para remover as feições contínuas do espectro por meio da identificação de pontos máximos no espectro e, assim realçar os picos de absorção (CLARK; ROUSH, 1984); ii) suavização *Savitzky-Golay* (SAVITZKY; GOLAY, 1964) (Equação 5), com 1ª derivada usando um polinômio de 1ª ordem, com janela de busca de 9 nm, após testes preliminares (SGD). A 1ª derivada calcula a taxa de mudança da reflectância em taxa de variação em relação ao comprimento de onda. Essa técnica é largamente utilizada para remover mudanças de linha de base (*baseline shifts*) e realçar as feições espectrais de interesse; (iii) padrão normal variado – *standard normal variate* (SNV) (Equação 6), consiste na subtração da média e divisão pelo desvio padrão da refletância de cada espectro (BERNES; DHANOA; LISTER, 1989) e tem como finalidade remover as variações indesejadas no espectro causada por ruídos e diferentes tamanhos de partículas; (iv) *Detrend* (DET), consiste na normalização do espectro pela aplicação do *standard normal variate* – SNV (BERNES; DHANOA; LISTER, 1989), seguido de uma função polinomial de segundo grau, a qual é utilizada para remover variações de baixa frequência (*offset* constantes) presentes em dados espectrais, realçando a variações de maior frequência; (v) transformação para absorção (Equação 7) e aplicação de Savitzky-Golay com 1ª derivada usando um polinômio 1ªordem com janela de busca de 9 nm (ASG).

$$S_i = \frac{1}{M} \sum_{j=0}^{M-1} x\,[i+j] \tag{3}$$

onde, S_i é o valor de reflectância suavizado, x é o valor original da refletância (i = 1, 2,…, n) e M é o número de pontos usados na média móvel.

$$\boldsymbol{CR_i = \frac{X_i}{C_i}; i = [1, \dots p]} \quad (4)$$

onde, X_i são os valores de refletância originais e C_i são os valores de refletância do contínuo no iésimo comprimento de onda de um conjunto de p comprimentos de onda, e CR_i é o valor da refletância final após a remoção do contínuo.

$$\boldsymbol{D_s = \frac{dS}{d\lambda_i} = \frac{S(\lambda_i) - S(\lambda_j)}{m} * \beta} \quad (5)$$

onde, D_s é o novo valor de refletância derivado, S é a reflectância, λ_i é a reflectância no menor comprimento de onda da janela de busca, λj é a reflectância no maior comprimento de onda da janela de busca, m é a diferença entre os comprimentos de onda máximo e mínimo da janela de busca e β é o coeficiente dependente da ordem da derivada e do polinômio.

$$\boldsymbol{SNV = \frac{x_i - \overline{x}_i}{s_i}} \quad (6)$$

onde, x_i é a refletância original, $\overline{x}_i$ é a média da refletância original, s_i é o desvio padrão da refletância original.

$$\boldsymbol{ABS = log_{10} \frac{1}{R}} \quad (7)$$

onde, *ABS* é a absorbância, *log10* é o logaritmo na base 10, e R é o reflectância.

Esses pré-processamentos foram aplicados nas curvas de reflectância de solo na faixa de 350-2500 nm para remover a variabilidade física devido à dispersão da luz e enfatizar as características de interesse ao longo do espectro. Todos os pré-processamentos foram realizados utilizando o pacote *prospectr* (STEVENS; RAMIREZ-LOPEZ, 2013) no software R (R CORE TEAM, 2018).

5.2.4. Análises estatísticas, métodos multivariados e estratificação da BEL

A união dos dados analíticos de COS, reflectância espectral e informações do local de amostragem (classe de solo e uso da terra) resultou na BEL. Foi realizada análise estatística descritiva, gerados histogramas de frequência e testada a normalidade dos dados de COS pelo teste Shapiro-Wilk (SHAPIRO; WILK, 1965) (stats R package) com nível de significância de $\alpha = 0,05$. Antes da modelagem, os dados espectrais foram submetidos à análise de componentes principais (ACP) para identificar similaridades e diferenças entre os espectros das classes de solo, usos da terra e camadas das amostras da biblioteca espectral, visando identificar grupos de amostras com espectros homogêneos. Além disso, a ACP auxiliou na detecção de espectros atípicos (outliers), uma vez que, podem prejudicar a construção dos modelos de predição (TERRA et al., 2015). Esses espectros foram identificados pelo cálculo da distância de Mahalanobis, em que amostras que apresentaram valores de *H* superiores a 3, foram consideradas outliers e removidas da biblioteca espectral antes das análises estatísticas, de acordo com Viscarra Rossel et al. (2016).

A fim de comparar a capacidade preditiva de diferentes métodos multivariados na construção dos modelos de predição para o COS, quatro métodos estatísticos multivariados foram utilizados: (i) PLSR (pacote *pls* do R (MEVIK et al., 2016)), modelo linear baseado em uma projeção das variáveis preditoras e a variável resposta em um conjunto de variáveis latentes e escores correspondentes, minimizando a dimensionalidade dos dados, ao mesmo tempo em que maximiza a covariância entre variáveis, sendo o modelo mais amplamente utilizado na predição de propriedades do solo a partir de dados espectrais; (ii) RF (pacote *randomForest* do R (LIAW; WIENER, 2002)) técnica não paramétrica, que consiste de uma combinação de muitas árvores preditoras, 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 (BREIMAN (2001); (iii) SMLR (pacote *caret* do R (KUHN, 2016)) técnica paramétrica, que descreve as relações lineares entre duas ou mais variáveis preditoras em relação a uma variável resposta, que utiliza para a seleção das variáveis preditoras a função *stepwise*, onde o modelo é ajustado seguindo o Critério de Informação de Akaike (AIC), em que o método seleciona o melhor preditor para a variável dependente e as demais variáveis preditoras são incluídas no modelo à medida que essas aumentam o poder explicativo do modelo, resultando na maior minimização do valor de AIC; (iv)

SVM (pacote *e1071* do R (MEYER et al., 2017)), técnica não paramétrica, baseada na teoria da aprendizagem estatística, em que através das funções do *kernel*, o modelo busca identificar uma função de interpolação, ajustando os dados de treinamento para um tubo com raio ε usando amostras de limites (representando os vetores de suporte) perto da margem definida por ε, em que o melhor modelo é obtido usando uma função de custo com o objetivo de minimizar simultaneamente o tamanho dos coeficientes e os erros de predição, em que conjuntos de dados com padrões não lineares podem ser representados mais facilmente (IVANCIUC, 2007).

A geração dos modelos foi realizada em duas etapas, sendo a primeira a construção de um modelo a partir do total de amostras da BEL, após a remoção dos outliers, denominado de conjunto total de amostras (CTA). Nessa etapa, foi avaliado qual pré-processamento dos dados espectrais e método multivariado alcançou melhor desempenho na predição. O teste Scott Knott (pacote *ScottKnott* do R (JELIHOVSCHI et al., 2014)), com nível de significância de $\alpha = 0{,}05$ foi utilizado para comparar a performance dos modelos. Em uma segunda etapa, foi realizada a estratificação da BEL em grupos com base na análise qualitativa das curvas espectrais dos diferentes usos da terra, classes de solo e camadas amostradas, e análise qualitativa da distribuição da variância dos dados espectrais obtida pela ACP. A partir disso, foram gerados modelos para cada grupo utilizando o pré-processamento e método multivariado selecionado na primeira etapa.

5.2.5. Calibração e validação dos modelos

Para a construção dos modelos de predição, cada grupo de amostras foi particionado em dois subconjuntos de forma aleatória, sendo 70 % das amostras utilizadas para construção dos modelos com validação cruzada "*leave one out*" (BRUS; KEMPEN; HUVELINK, 2011) e 30 % das amostras para validação. O teste de Levene (LEVENE, 1960) (pacote *car* do R (FOX; WEISBERG, 2011)) foi aplicado para verificar a homocedasticidade entre dois conjuntos. Para comparar os desempenhos dos modelos foram calculados os seguintes parâmetros: coeficiente de determinação (R^2) (Equação 3), raiz quadrada média do erro de predição (RMSE) (Equação 4), e proporção de desempenho no intervalo interquartil (RPIQ) (Equação 5):

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{8}$$

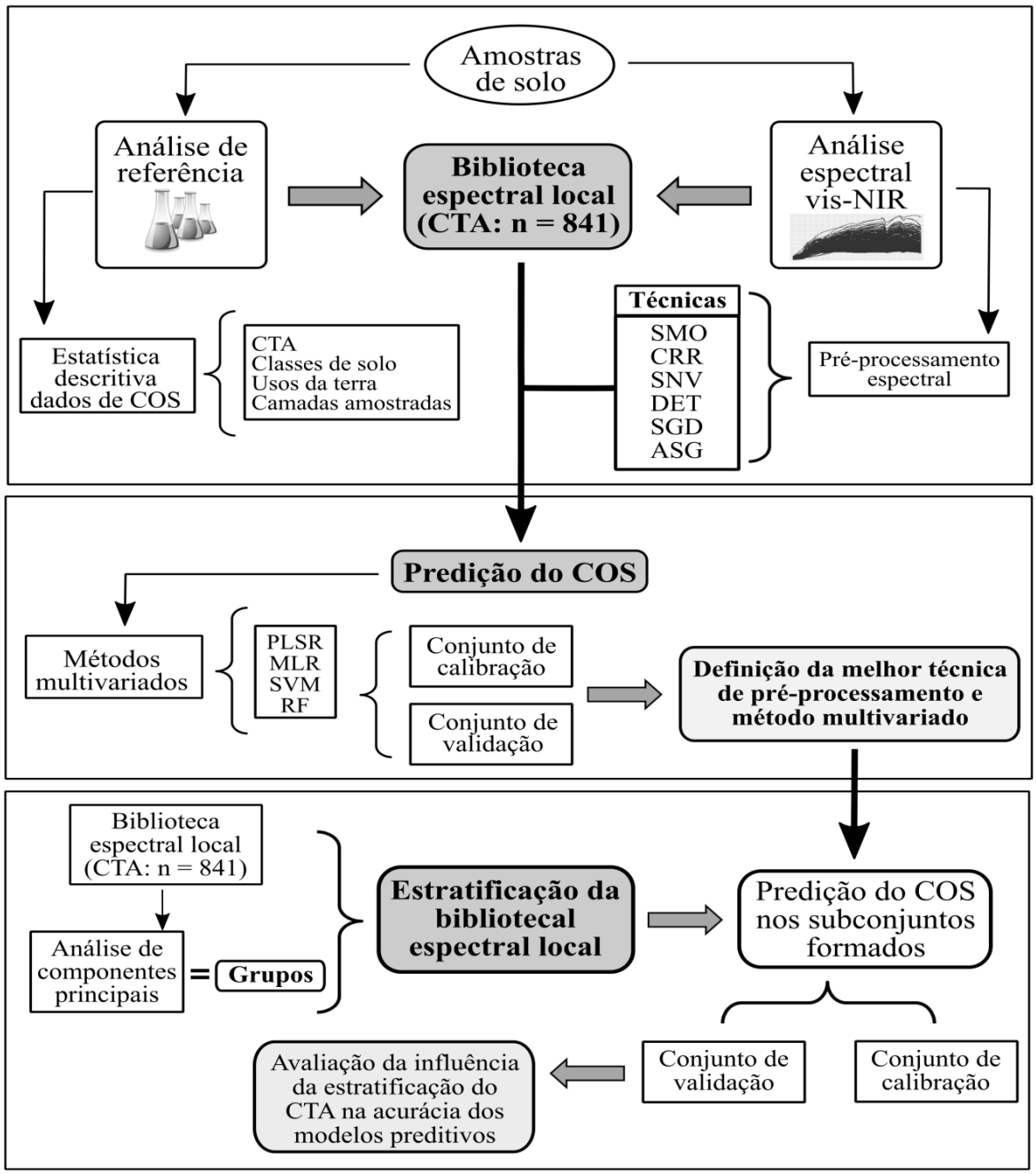
$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{9}$$

$$RPIQ = \frac{(Q3 - Q1)}{RMSE} \tag{10}$$

Em que: $\hat{y}$ = valor predito; $\bar{y}$ = valor médio observado; y = valores observados; n = número de amostras com *i* = 1, 2,… , n; Q1 = 25 % das amostras; Q3 = 75 % das amostras; distância interquartil = Q3 – Q1, representa o intervalo responsável por 50 % da população em torno da mediana.

A figura 13 apresenta o resumo das etapas executadas na metodologia do trabalho.

Figura 13 – Diagrama esquemático das etapas executadas na metodologia. CTA: conjunto total de amostras; SMO: smoothed utilizando uma janela móvel de 9 nm; CRR: continuum removed reflectance; SNV:standard normal variate; DET: Detrend; SGD Savitzky-Golay com 1ª derivada usando um polinômio de 1ªordem com janela de busca de 9 nm; PLSR: regressão por mínimos quadrados parciais; MLR: regressão linear múltipla; SVM: máquina de vetor de suporte; RF: floresta aleatória.



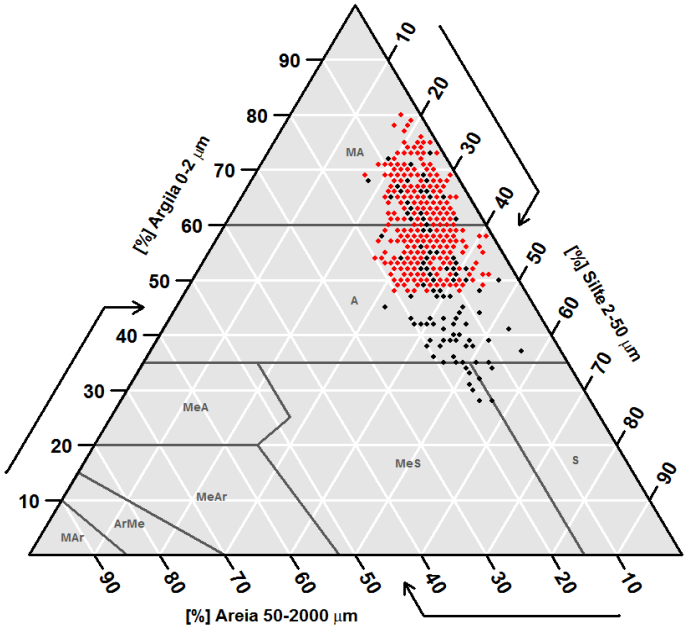
Fonte: Autor.

5.3. RESULTADOS E DISCUSSÃO

5.3.1. Caracterização dos solos e análises estatísticas dos dados de COS

O banco de dados utilizado neste estudo, embora seja de escala local, é composto por um grande número amostras (n = 810, após remoção de 31 *outliers*), sendo superior aos apresentados na Tabela 2 para escala local. Na área de estudo predominam as classes de solos LV (80 % da área) em posições mais elevadas e bem drenadas da paisagem, apresentando textura muito argilosa a argilosa (Figura 14) e altos teores de óxidos de ferro (~ 234 g kg^{-1} de Fe_3O_2). Os GX ocorrem em 20 % da área, em posições menos elevadas e mal drenadas da paisagem, com textura predominantemente argilosa e argilossiltosa (Figura 14) e baixos teores de óxidos de ferro (~ 130 g kg^{-1} de Fe_3O_2). Essas diferenças nas características físicas e químicas estão relacionadas aos fatores pedogenéticos intrínsecos a cada classe de solo.

Figura 14 – Triângulo textural indicando a classe textural dos Latossolos (pontos em vermelho) e Gleissolos (pontos em preto)



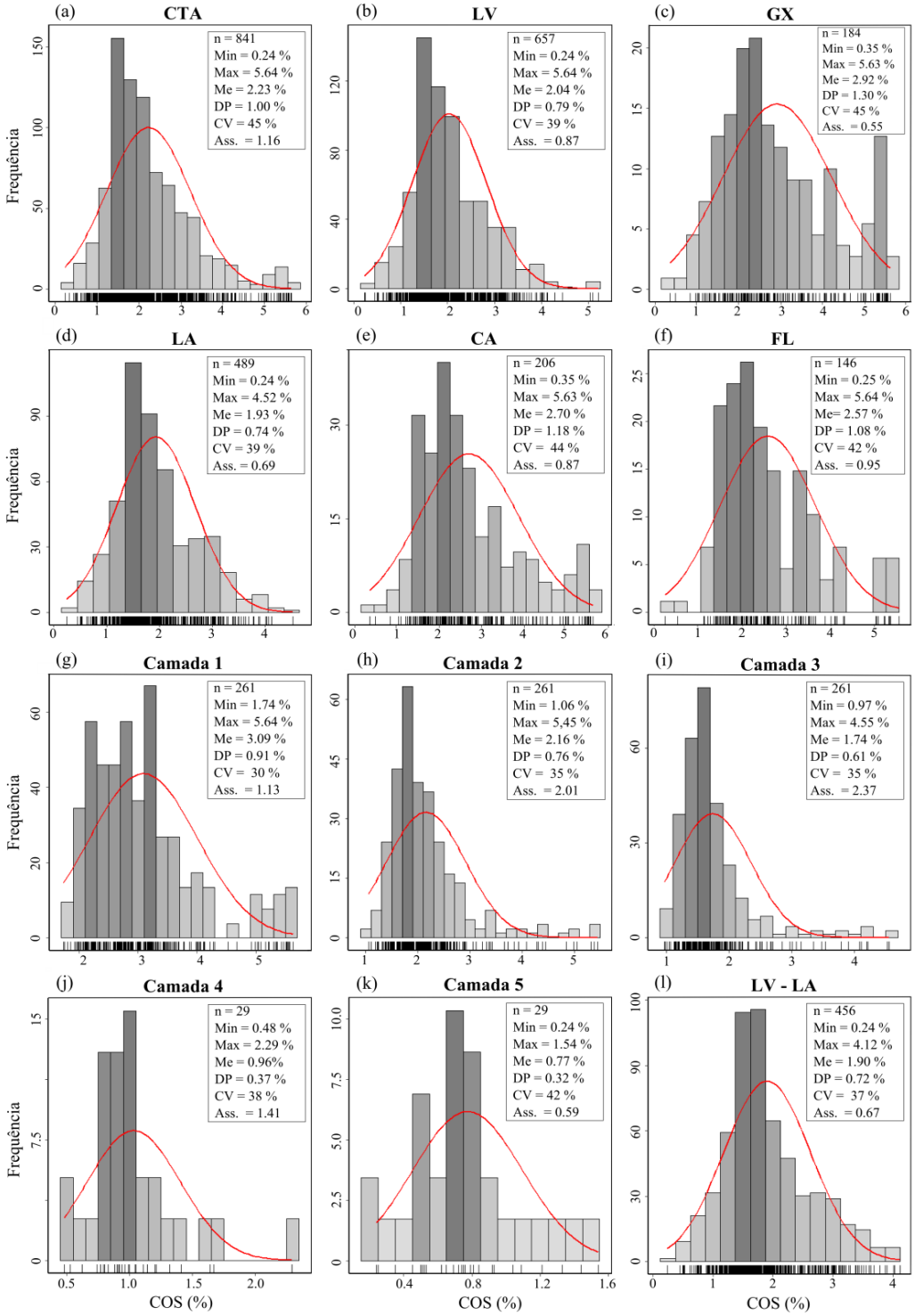
Fonte: Autor.

O número de amostras, teores mínimos, máximos, médios, desvio-padrão (DP), coeficiente de variação (CV), assimetria para o CTA e grupos separados com base na classe de solo, uso da terra e camadas das amostras são apresentados na Figura 15. Os teores de COS do CTA variaram de 0,24 a 5,64 % e teor médio de 2,23 % (Figura 16a). O valor de DP e CV do CTA foram altos, mostrando uma dispersão de 45 % dos teores de COS em relação à média.

Os teores mais elevados de COS ocorreram nas áreas de FL e CA na classe de solo GX, atribuindo-se esse dado ao maior aporte de resíduos vegetais e ausência de revolvimento nesses usos e posição na paisagem que ocorre esse solo. Houve pouca variação nos teores mínimos e máximos de COS nas classes de solo LV e GX (Figura 15b e 15c), porém o teor médio para amostras de GX foi maior. Nas amostras de LV, os maiores teores de COS estão associados ao uso da terra CA na camada 0,0 – 0,05 m. Os teores de COS nos usos da terra LA, CA e FL (Figura 15d, 15e e 15f) e nas cinco camadas de amostragem (Figura 15g, 15h, 15i, 15j e 15k) apresentaram diferenças expressivas em relação aos teores mínimos, máximos, médios, DP e CV e assimetria. Os maiores valores de assimetria foram observados na C3, C2, C4, CTA e C1, já os menores na camada C5 e nos grupos GX, LA, LV, CA e FL.

Os valores de assimetria para todos os grupos indicam que os dados não apresentam uma distribuição normal. A variação nos teores de COS em todos os grupos é consequência da variabilidade das características do terreno, camadas de amostragem, classes de solos e usos da terra da área de estudo, os quais são fatores que interferem na variabilidade do COS na paisagem (BODDEY et al., 2010).

Figura 15 – Estatística descritiva e histogramas dos teores de COS do CTA, Latossolo Vermelho – LV, Gleissolo Háplico – GX, Lavoura – LA, Campo nativo – CA, Floresta nativa – FL, camada de 0,0 – 0,05 m (C1); 0,05-0,15 m (C2); 0,15-0,30 m (C3); 0,30-0,60 m (C4); 0,60-1,00 m



Fonte: Autor.

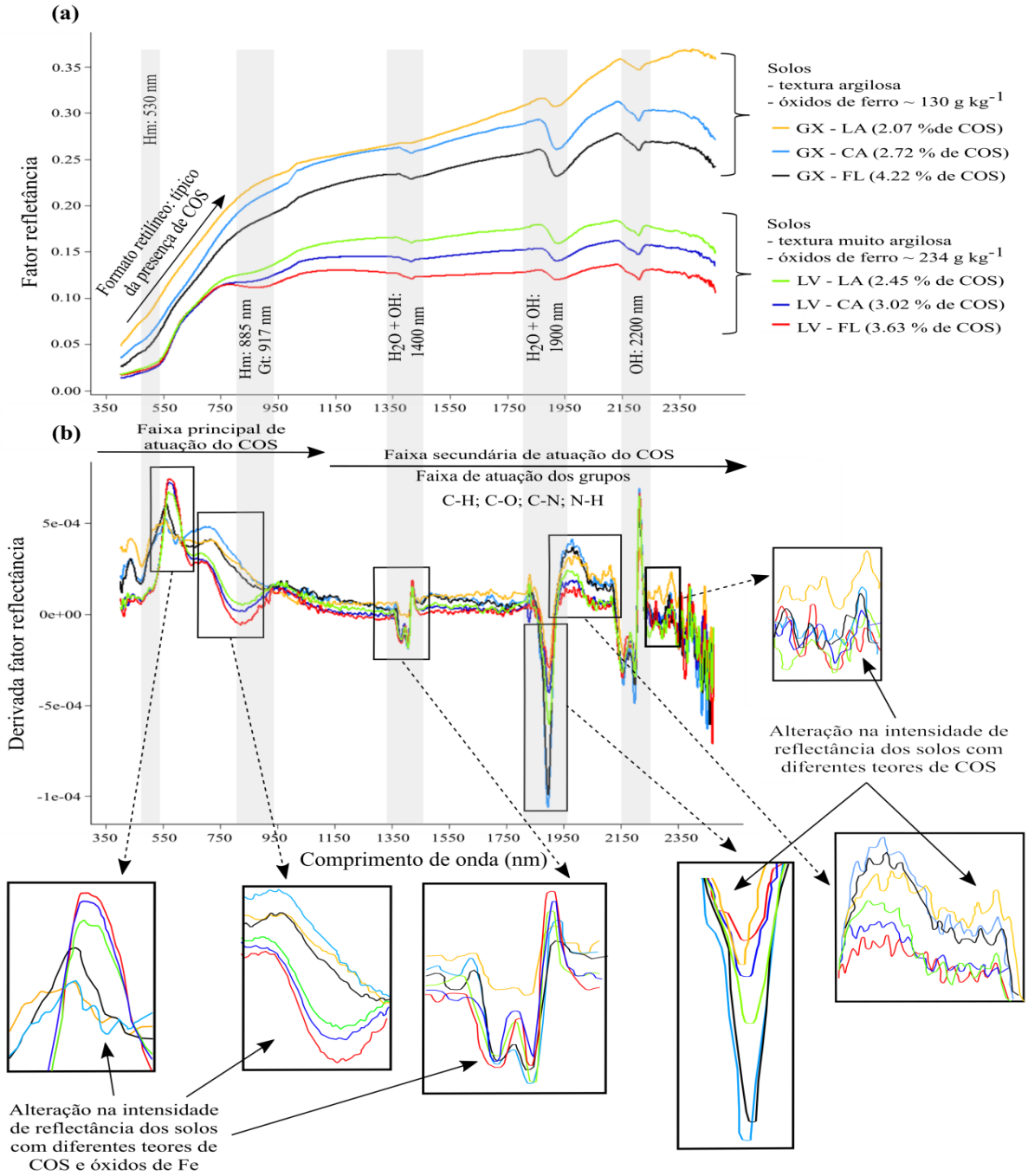
5.3.2. Caracterização dos dados espectrais vis-NIR

Os espectros de reflectância Vis-NIR representativos da BEL são apresentados na Figura 16. As curvas espectrais dos LV apresentam menor intensidade quando comparadas com as amostras de GX. Em relação ao uso da terra, amostras com maior conteúdo de COS em cada classe de solo apresentaram menor intensidade (Figura 16a).

Amostras de LV apresentaram características bem definidas de absorção de óxidos de ferro nas bandas 435, 550 e 850 nm, identificando a presença de goethita e hematita (DALMOLIN et al., 2005; DEMATTÊ; GARCIA, 1999). Também são observadas características de absorção bem definidas na região de 1900 e 2200 nm as quais, por sua vez, estão relacionadas às moléculas de hidroxilas (OH) presentes na estrutura de argilominerais. Características de absorção em 2200 nm estão relacionadas à presença de caulinita (DEMATTÊ; GARCIA, 1999). Também foram notadas feições de absorção entre 1350 a 2300 nm produzidas por vibrações moleculares relacionadas aos grupos C – H, C – O, C = O, C – N e N – H (Figura 16b), sendo esses associadas ao COS (VISCARRA ROSSEL; BEHRENS, 2010).

Em geral, os espectros de reflectância Vis-NIR das amostras de cada classe de solo e uso da terra mostraram forma e localização semelhantes das características de absorção, mas com considerável variabilidade nas suas intensidades de absorção. Portanto, as regiões do espectro associadas ao COS podem ser mascaradas ou distorcidas por óxidos de ferro, que são comumente encontradas nestes solos (DALMOLIN et al., 2005). Isso mostra uma questão crítica da predição de propriedades do solo por espectroscopia Vis-NIR, uma vez que, a composição da fração argila podem criar diferentes feições de absorção e interferir na predição de uma determinada propriedade de interesse, por exemplo, o COS, indicando a necessidade de se criar estratégias para minimizar esse problema e melhorar o desempenho dos modelos de predição para o COS.

Figura 16 – Curvas espectrais representativas da BEL, amostras de Latossolo Vermelho – LV e Gleissolo Háplico – GX, nos usos da terra Lavoura – LA, Campo nativo – CA e Floresta nativa – FL, com identificação das características das amostras



Fonte: Autor.

5.3.3. Análises de componentes principais dos dados espectrais

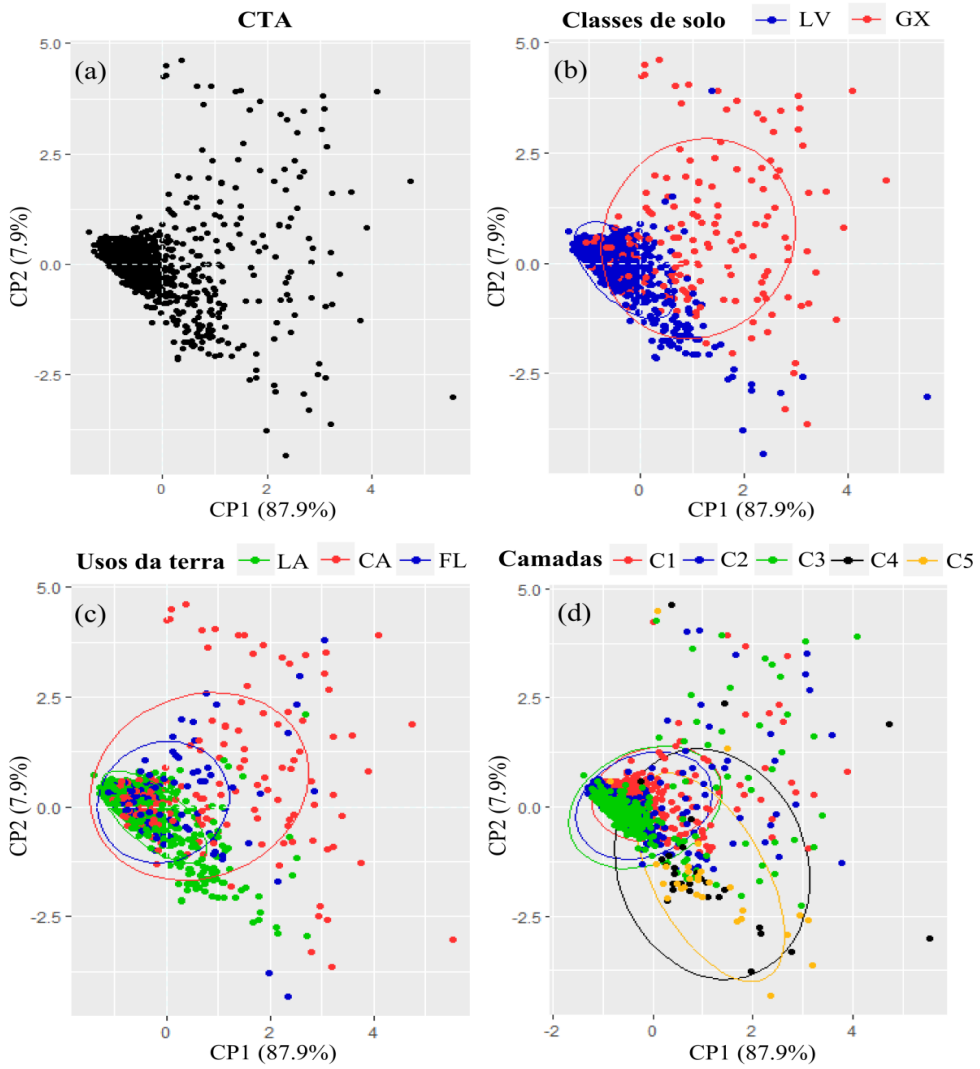
A análise das variáveis latentes da ACP para o CTA, classes de solo, usos da terra e camadas amostradas indica que as componentes principais CP1 e CP2 contemplam 95,8 % da variância (Figura 17). A ACP permitiu identificar o comportamento da variância dos dados espectrais, com maior dispersão daquelas observadas nos quartis superior e inferior direito (Figura 17a), com algumas amostras que destoam das demais. O cálculo da distância de Mahalanobis mostrou 31 amostras com valores de H superiores a 3. Após a verificação do local de amostragem e comportamento espectral heterogêneo destas amostras, as mesmas foram removidas da biblioteca espectral por serem consideradas *outliers*, assim o CTA passou a ter 810 amostras. Procedimento semelhante foi utilizado em estudo de Terra et al. (2015) e Viscarra Rossel et al. (2016).

Nos dados espectrais das amostras de LV e GX, maior dispersão é notada nas amostras de GX (Figura 17b). Nos espectros obtidos de amostras de diferentes usos da terra, a maior dispersão é observada nos usos CA e FL (Figura 17c), com destaque para as amostras coletadas na CA, principalmente nos quartis superior e inferior direto. As três camadas apresentaram espectros distintas, sendo a maior dispersão notada nas camadas C1 e C2 (Figura 17d).

As amostras com maior dispersão estão relacionadas à classe de solo GX (Figura 17b) e uso da terra CA (Figura 17c), consequência das reações de oxi-redução presentes nesses solos mal drenados, que influenciam nas formas e teores de óxidos, somada a variabilidade observada nos teores de COS (Figura 17c e 17e). Portanto, o ambiente de ocorrência das classes de solo e forma de uso da terra foram os principais fatores responsáveis pela variação na composição das amostras e, consequentemente, na alta dispersão dos dados espectrais observada na Figura 18. Comportamento semelhante foi observado por Demattê et al. (2016), que estudaram a regionalização espectral de solos tropicais para a predição de propriedades do solo. Os autores observaram que bibliotecas espectrais regionais mostraram maior dispersão em relação a bibliotecas locais, sendo a causa dessa variação atribuída a solos formados por diferentes materiais parentais, que refletiram em distintos teores de óxidos de ferro, argila e COS nas amostras de solo. Em estudo realizado por Jiang et al. (2017) na região central da China, com amostras coletadas na camada superficial (0,0 - 0,10 m) e subsuperficial (0,10 – 0,30 m), foi

possível identificar diferenças nos espectros pela ACP e assim construir modelos preditivos de COS para cada camada.

Figura 17 – Análise de componentes principais (a) CTA e grupos: (b) classes de solo Latossolo Vermelho – LV e Gleissolo Háplico – GX, (c) usos da terra Lavoura – LA, Campo nativo – CA e Floresta nativa – FL e (d) camadas amostradas C1 (0,0 – 0,05 m); C2 (0,05-0,15 m); C3 (0,15-0,30 m); C4 (0,30-0,60 m); C5 (0,60-1,00 m)



Fonte: Autor.

Na análise conjunta das Figuras 16 e 17 fica evidente a relação entre os espectros Vis-NIR e a composição das amostras de solo, em que diferentes conteúdos de argila, óxidos de ferro e COS influenciaram no comportamento espectral, uma vez que esses constituintes são os

principais componentes espectralmente ativos do solo, tendo influência direta nos espectros (GALVÃO; VITORELLO, 1998; DALMOLIN et al., 2005; VISCARRA ROSSEL; BEHRENS, 2010). Portanto, para o uso de bibliotecas espectrais em ambientes complexos como o solo, deve-se conhecer a variabilidade do conjunto amostral como uma etapa importante no desenvolvimento de modelos preditivos mais acurados.

5.3.4. Performance dos modelos de predição da BEL

As estatísticas do desempenho de cada modelo preditivo associada à técnica de pré-processamento aplicada nos dados espectrais para o CTA da BEL é apresentada na Tabela 3. Ao comparar o valor médio de $RMSE_v$ entre os quatro os modelos multivariados, observa-se que não há diferença entre os modelo (Tabela 3). Comportamento semelhante foi observado por Dotto et al. (2017), que avaliaram a capacidade preditiva do COS em escala local com os modelos PLSR e SVM, e quatro técnicas de pré-processamento, onde alcançaram RMSE de 0,45 % - PLSR e 0,44 % - SVM. Nesse mesmo estudo, os autores observaram alta variação no desempenho dos modelos preditos em relação a cada pré-processamento espectral, sendo isso apontado como a causa de não se observar um padrão no desempenho dos modelos. Esse mesmo comportamento é notado no presente estudo, ao observar a amplitude entre os valores mínimos e máximos de $RMSE_v$ dos quatros modelos avaliados (Tabela 3).

A variação no desempenho dos modelos multivariados para distintos tipos de pré-processamento espectral foi observada por Vasques et al. (2008), que compararam quatro modelos e 30 pré-processamentos na construção de modelos preditivos de COS escala regional na Flórida. Os autores encontraram valores de RMSE variando de 0,18 - 0,27 log (mg kg^{-1}) na modelagem por MLR e de 0,18 - 0,26 por PLSR. Desempenho semelhante foi observado em trabalho de Knox et al. (2015), que encontraram variação na performance dos modelos por PLSR e RF ao se comparar dez pré-processamentos em uma biblioteca espectral de escala nacional da Flórida, EUA. Da mesma forma, Cambule et al. (2012) avaliaram o efeito de nove técnicas na predição de COS por PLSR, em Moçambique - África, e encontraram valores de R^2 entre 0,65 e 0,85 e RMSE entre 0,32 % e 0,37 %.

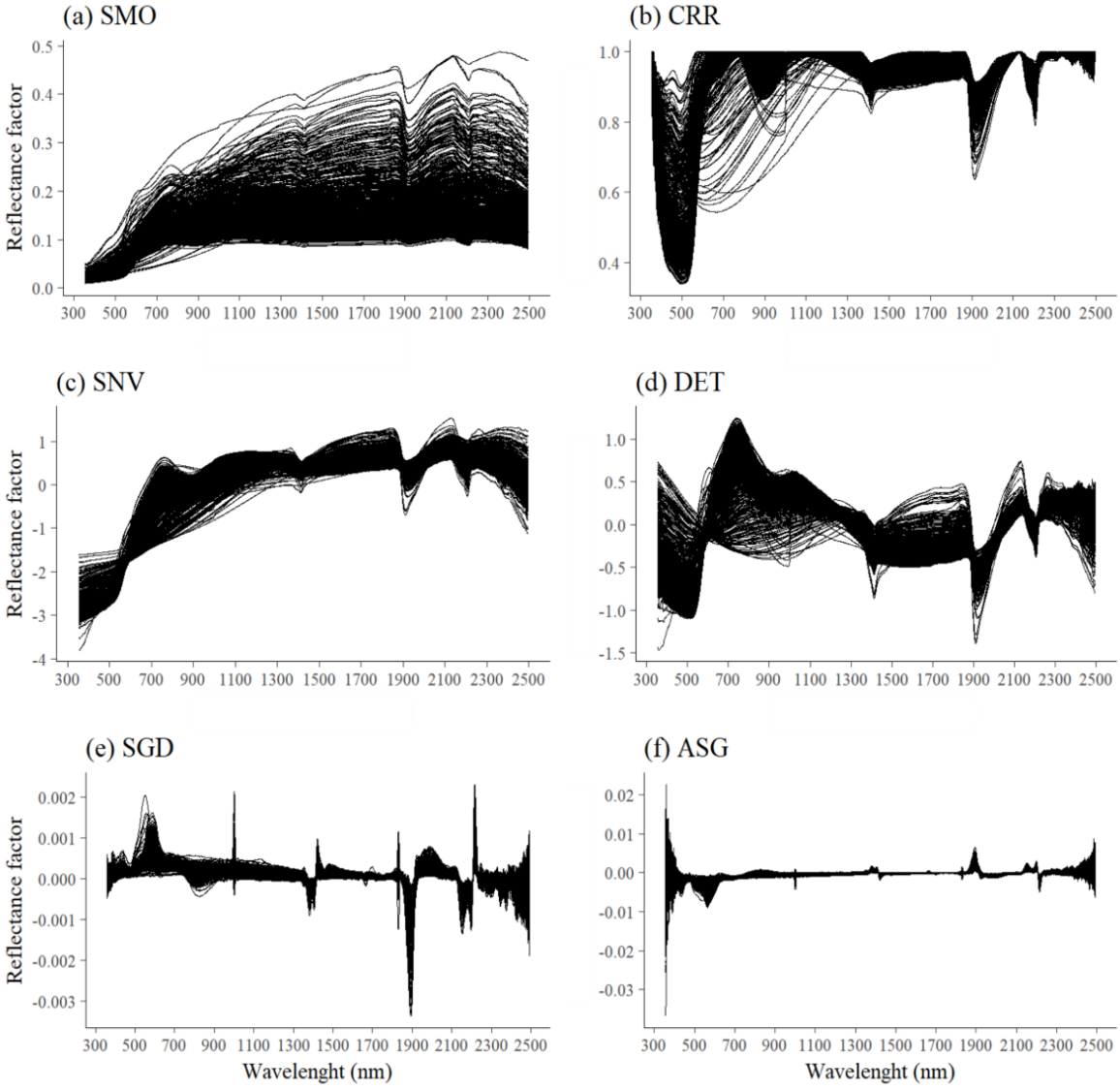
Tabela 3 – Resumo estatístico dos modelos geradas para os quatros métodos multivariados e seis pré-processamentos para o CTA (n = 810)

	Calibração (n = 567)			**Validação (n = 243)**		
Pré-processamentos	R^2	RMSE	RPIQ	R^2	RMSE	RPIQ
			PLSR			
SMO*	0,68	0,57	2,03	0,72	0,56	2,14
CRR	0,67	0,56	2,05	0,63	0,66	1,82
SGD	0,68	0,54	2,11	**0,74**	**0,52**	**2,23**
SNV	0,69	0,54	2,13	0,71	0,59	2,04
ASG	0,70	0,53	2,17	0,64	0,65	1,86
DET	0,69	0,53	2,15	0,69	0,60	1,98
Mínimo	0,67	0,53	2,03	0,63	0,52	1,82
Máximo	0,70	0,57	2,17	0,74	0,66	2,23
Média	0,69	0,55	2,11	0,69	0,60 a**	2,01
			MLR			
SMO*	0,76	0,47	2,42	0,71	0,58	2,08
CRR	0,74	0,45	2,40	0,70	0,60	2,01
SGD	0,76	0,47	2,41	**0,72**	**0,57**	**2,09**
SNV	0,74	0,49	2,32	0,70	0,59	2,02
ASG	0,69	0,54	2,12	0,62	0,66	1,82
DET	0,74	0,49	2,31	0,70	0,59	2,04
Mínimo	0,69	0,45	2,12	0,62	0,57	1,82
Máximo	0,76	0,54	2,42	0,72	0,66	2,09
Média	0,74	0,49	2,32	0,69	0,60 a	2,01
			SVM			
SMO*	0,75	0,50	2,20	0,70	0,58	2,02
CRR	0,88	0,35	3,30	0,66	0,57	2,03
SGD	0,76	0,50	2,27	**0,72**	**0,55**	**2,10**
SNV	0,72	0,53	2,25	0,55	0,58	1,68
ASG	0,92	0,27	4,39	0,59	0,56	1,74
DET	0,91	0,30	3,84	0,56	0,70	1,57
Mínimo	0,72	0,27	2,25	0,55	0,55	1,57
Máximo	0,92	0,53	4,39	0,72	0,70	2,10
Média	0,82	0,41	3,04	0,63	0,59 a	1,86
			RF			
SMO*	0,53	0,69	1,67	0,57	0,64	1,82
CRR	0,71	0,54	2,14	0,67	0,57	2,06
SGD	0,65	0,57	1,98	**0,72**	**0,56**	**2,09**
SNV	0,59	0,64	1,88	0,52	0,59	1,63
ASG	0,67	0,58	2,07	0,59	0,59	1,77
DET	0,60	0,61	1,88	0,67	0,62	1,92
Mínimo	0,53	0,54	1,67	0,52	0,56	1,63
Máximo	0,71	0,69	2,14	0,72	0,64	2,09
Média	0,63	0,61	1,94	0,62	0,59 a	1,88

Fonte: Autor.*Pré-processamento considerado testemunha. SMO: suavização utilizando uma janela móvel de 9 nm; CRR: remoção do contínuo; SNV: padrão normal variado; DET: Detrend; SGD: Savitzky-Golay com 1ª derivada usando um polinômio de 1ªordem com janela de busca de 9 nm. ASG: transformação para absorção e aplicação de Savitzky-Golay com 1ª derivada usando um polinômio 1ªordem com janela de busca de 9 nm; PLSR: Regressão por mínimos quadrados parciais; RLM: regressão linear múltipla; SVM: Support vector machine; RF: Random forest; R^2: coeficiente de determinação; RMSE: raiz do erro quadrático médio em %; RPIQ: Proporção do erro de predição no intervalo interquartil; c: calibração; v: validação. Destacado em negrito representa o melhor resultado para o modelo em relação a cada pré-processamento. **Valores com a mesma letra não diferem entre si pelo teste de Scott Knott ($\alpha = 0,05$).

Uma explicação para a variação na acurácia do mesmo método multivariado, porém com pré-processamento espectral diferente, são as alterações no valor de reflectância causadas por cada tipo de processamento espectral (Figura 18).

Figura 18 – Espectros da BEL após a aplicação dos pré-processamentos. (a) SMO: suavização utilizando uma janela móvel de 9 nm; (b) CRR: remoção do contínuo; (c) SNV: padrão normal variado; (d) DET: Detrend; (e) SGD: Savitzky-Golay com 1ª derivada usando um polinômio 1ªordem com janela de busca de 9 nm

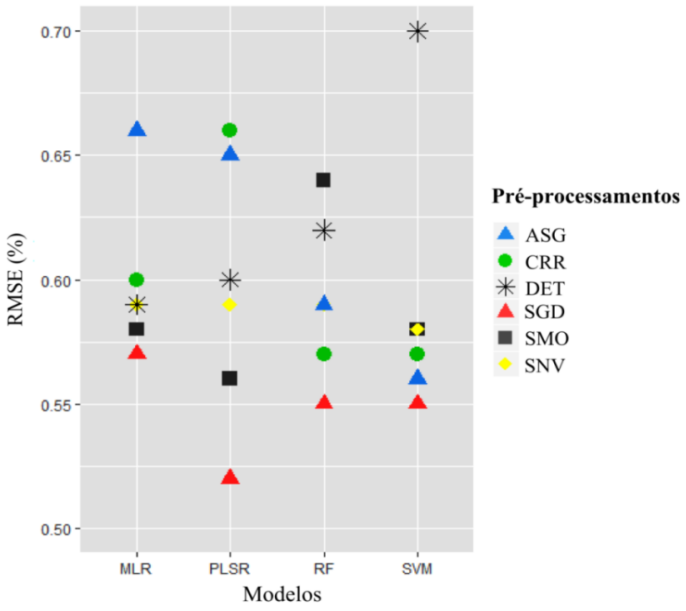


Fonte: Autor.

O uso destas técnicas tem por objetivo minimizar a influência de efeitos físicos (tamanho de partícula e rugosidade) e variações óticas, os quais introduzem alterações nos espectros (SHEPHERD; WALSH, 2002) e realçam feições de absorção de energia. Na Figura 18 nota-se que algumas feições relacionadas com os constituintes do solo mencionadas na Figura 17 ficaram mais realçadas nas técnicas CRR, SNV, DET, ASG e com maior destaque para a técnica SGD, comparado ao pré-processamento testemunha SMO (Figura 18a). Estudos têm mostrado melhorias na acurácia dos modelos preditivos de COS por Vis-NIR usando pré-processamento dos dados espectrais, tais como 1ª e 2ª derivadas com filtro Savitzky-Golay (CAMBULE et al., 2012; KNOX et al., 2015; GUERRERO et al., 2016), standard normal variate (LUCÀ et al., 2017) e remoção do contínuo (CLAIROTTE et al., 2016; DOTTO et al., 2017), enquanto Kooistra et al. (2001) encontraram melhores resultados com dados espectrais não transformados.

O efeito dos pré-processamentos na capacidade preditiva dos modelos é observado quando é considerado o valor de $RMSE_v$ das técnicas de pré-processamento em cada método multivariado separadamente (Figura 19).

Figura 19 – Desempenho dos modelos de predição em relação aos valores de RMSEv para os quatro modelos multivariados e seis técnicas de pré-processamento



Fonte: Autor. SMO: suavização utilizando uma janela móvel de 9 nm; CRR: remoção do contínuo; SNV: padrão normal variado; DET: Detrend; SGD: Savitzky-Golay com 1ª derivada usando um polinômio de 1ªordem com janela de busca de 9 nm. ASG: transformação para absorção e aplicação de Savitzky-Golay com 1ª derivada usando um polinômio 1ªordem com janela de busca de 9 nm.

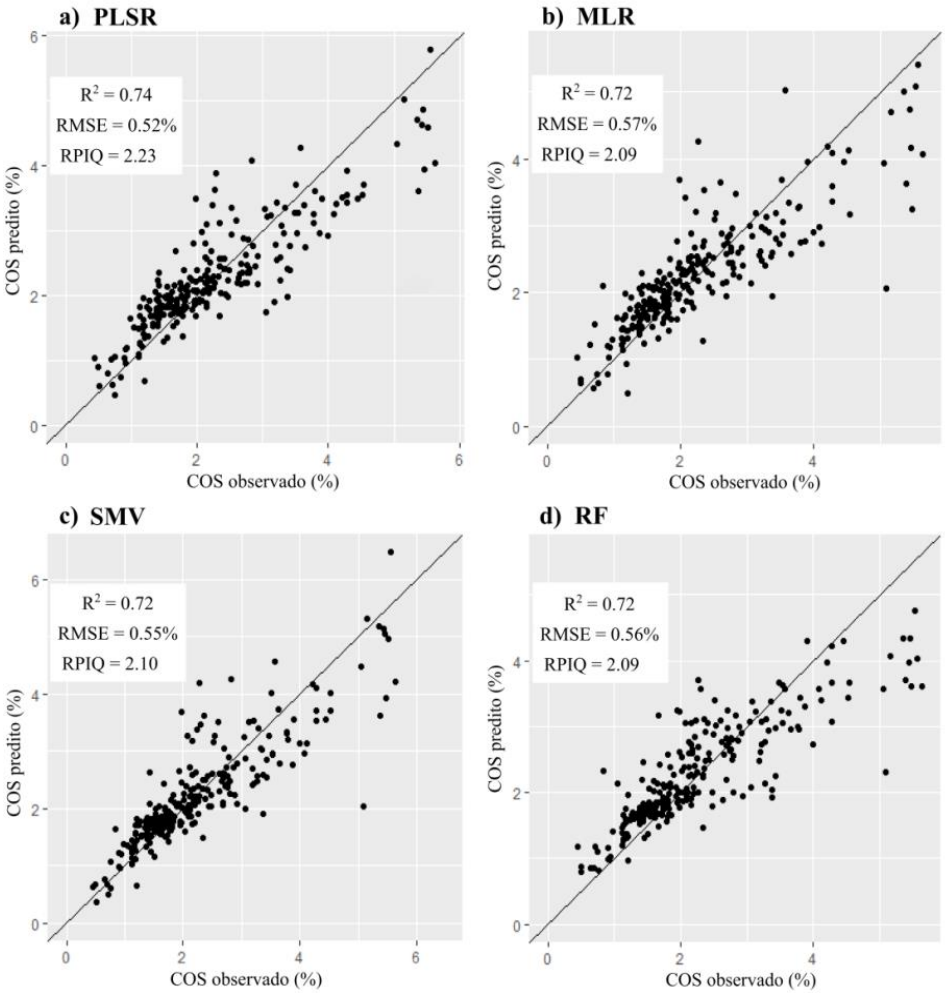
O pré-processamento SMO considerado testemunha, para efeito de comparação entre as demais técnicas, apresentou RMSE ligeiramente maior em relação ao pré-processamentos SGD e menor comparado aos demais, com destaque para o ASG no modelo por MLR (Figura 19). No modelo PLSR, foi observada maior amplitude dos valores de RMSE entre os pré-processamentos, onde o SMO apresentou diferença maior de RMSE quando comparado ao SDG, SVN, DET, CRR ASG (Figura 19). No entanto, para o modelo PLSR o processamento SMO apresentou o segundo menor valor de RMSE. Já nos modelos RF o SMO apresentou o maior valor de RMSE, mostrando que uma suavização mínima dos espectros não foi adequada. Isso pode ser um indicativo que o modelo RF requer valores de reflectância que passaram por uma transformação mais rigorosa para melhorar o desempenho preditivo, com destaque para os pré-processamentos SGD e CRR.

O modelo SVM apresentou baixa amplitude nos valores de RMSE nas técnicas SGD, ASG CRR, SMO e SNV, exceto para a técnica DET, em que apresentou o maior valor de RMSE. Para esse mesmo modelo, o segundo maior valor de RMSE foi observado para o pré-processamento testemunha SMO. Trabalho de Dotto et al. (2017) encontrou o menor desempenho na predição de COS como o modelo SVM com a técnica DET (RMSE = 0,42 %), sendo os menores valores de RMSE (0,35 %) observados com a técnica de CRR. Essa baixa amplitude observada para o modelo SVM no presente estudo, pode ter relação com a função matemática de kernel que o modelo emprega para estabelecer relações com o valor de reflectância e teor de COS, em que o modelo busca identificar uma função de interpolação entre as variáveis e cria vetores de suporte, sendo um procedimento robusto em modelos de aprendizagem estatística (IVANCIUC, 2007).

O pré-processamento SNV apresentou, em geral, baixa acurácia no presente estudo, contrariando os resultados observados por Vasques et al. (2008), que encontraram valores iguais de RMSE (0,18 log g kg^{-1}) para os pré-processamentos SNV e SGD nos modelos PLSR e MLR, gerados com dados espectrais de escala regional na Flórida, EUA. Por outro lado, Knox et al. (2015) encontraram maiores valores de RMSE com a técnica SNV (0,68 log g kg^{-1} para o modelo RF e 0,51 log g kg^{-1} para PLSR) comparado aos modelos obtidos com os espectros sem pré-processamento (0,41 log g kg^{-1} para RF e 0,42 log g kg^{-1} para PLSR) em uma biblioteca espectral em escala nacional da Flórida, EUA.

No presente estudo, é notável que o pré-processamento CRR apresentou melhor desempenho na predição de COS nos modelos não paramétricos RF e SVM e, para os modelos paramétricos MLR e PLSR, a técnica SGD mostrou mais capacidade preditiva. Comportamento semelhante foi observado por Vasques et al. (2008), em que os MLR e PLSR tiveram melhor desempenho com SGD e para modelos baseados em árvores, a transformação *norris gap derivative* (NGD). Na presente pesquisa, em geral, o melhor desempenho na predição do COS foi obtido para o modelo gerado pelo método de PLSR (R_v^2 = 0,74; $RMSE_v$ = 0,52 %) com o pré-processamento SGD (Figura 20a).

Figura 20 – Melhores desempenhos dos modelos de predição para o COS com o CTA. PLSR: regressão por mínimos quadrados parciais; MLR: regressão linear múltipla; SVM: máquina de vetor de suporte; RF: floresta aleatória



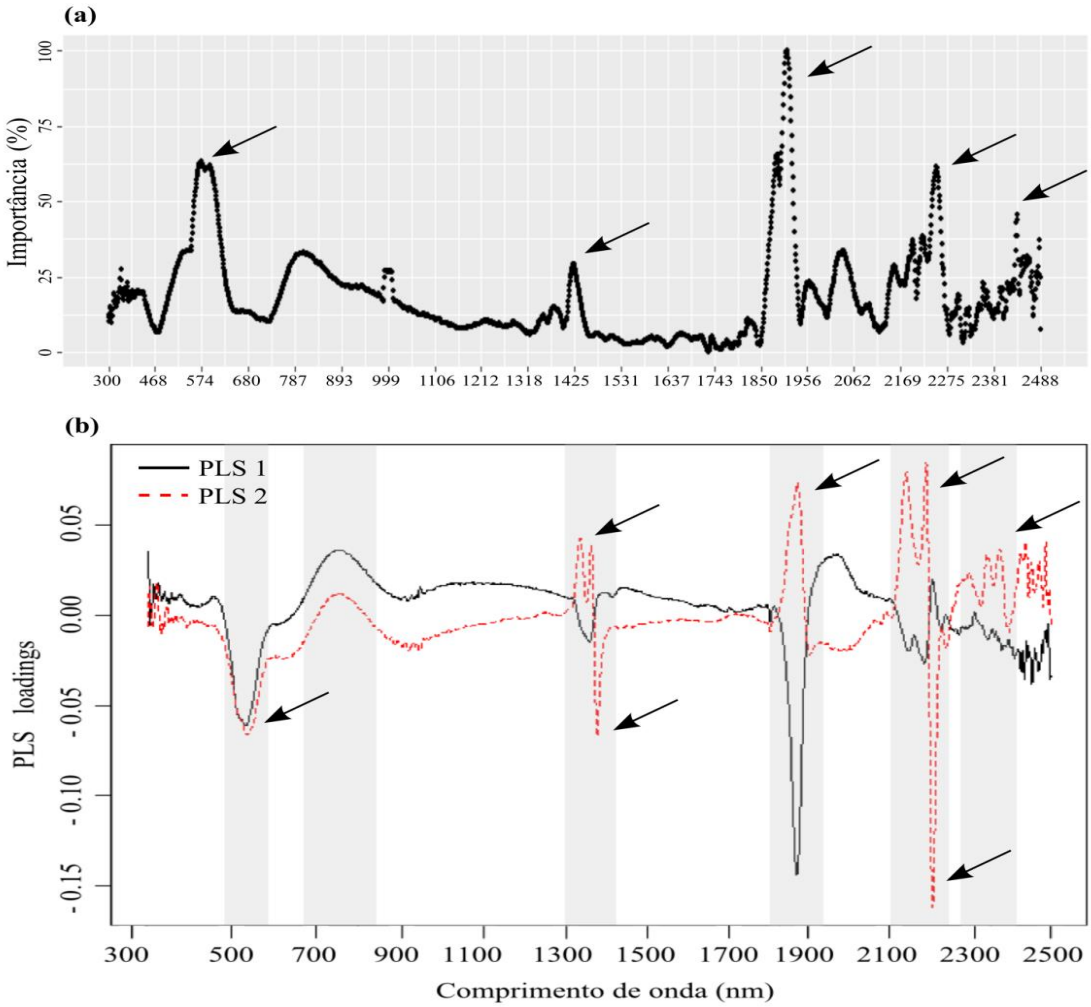
Fonte: Autor.

Na sequência, o melhor desempenho foi observado nos modelos SVM, MLR e RF com o mesmo pré-processamento (Figura 20c, 20b e 20d, respectivamente). Trabalhos que utilizaram o modelo PLSR associado à pré-processamentos derivativos dos dados espectrais (VASQUES et al., 2008; CAMBULE et al., 2012; KNOX et al., 2015) também obtiveram bom desempenho dos modelos por PLSR, como R^2 e RMSE de 0,85 e 0,18 log mg kg^{-1}; 0,81 e 0,32 % e 0,87 e 0,40 log g kg^{-1}, respectivamente. Diante dos resultados observados na literatura e do presente trabalho, fica evidente que existe uma variação nos resultados, independentemente de qual pré-processamento ou técnica multivariada foi empregada. No entanto, nota-se que os modelos multivariados podem apresentar maior similaridade por alguma técnica de pré-processamento específica.

Diante do melhor desempenho na predição do COS ter sido alcançado para o modelo PLSR com pré-processamento SGD, o grau de importância das faixas espectrais (Figura 21a) e as cargas (*loadings*) dos componentes um e dois do modelo PLSR (Figura 21b), foram plotadas para visualizar a relação dos picos com as faixas espectrais relacionados com COS.

No início da faixa espectral ocorrem picos longos na região de 500 a 850 nm (Figura 22b) que coincidem com bandas relacionadas com combinações O – H e C – H correlacionadas com o COS, conforme observado em estudo de Viscarra Rossel et al. (2006). Picos mais estreitos ocorrem nas faixas de 1300, 1900 e 2200 nm, sendo associados à característica de absorção de água e moléculas de OH. Nas regiões de 2200 a 2400 nm ocorrem picos bem definidos relacionados com a zona de alongamento da combinação C – H. As regiões dos espectros consideradas de maior importância como variáveis preditoras do COS no modelo PLSR coincidem com as regiões relacionadas com o COS discutidas anteriormente na Figura 5 e observadas em estudo de Vasques et al. (2008) e Viscarra Rossel e Behrens (2010). Em estudo de escala local, Jiang et al. (2017) também observaram as regiões entre 400-800, 1900, e 2000-2350 como importantes na predição do COS por Vis-NIR, confirmando o potencial da técnica de ERD como metodologia para geração de informações do teor de COS em escala local.

Figura 21 – Importância das faixas espectrais na predição do COS (a) e pesos (*loadings*) das componentes PLS1 e PLS2 obtidas pelo modelo PLSR (b). Ressaltado com setas e em cinza as faixas espectrais consideradas importantes na predição do COS pelo modelo



Fonte: Autor.

Os resultados encontrados no presente estudo apresentaram desempenho ligeiramente maior quando comparados a outros estudos que avaliaram a capacidade preditiva para o COS com bibliotecas espectrais em escala local. Na Austrália, em uma área agrícola de 17,5 ha, Viscarra Rossel et al. (2006) obteve R^2 de 0,60 e RMSE de 1,8 g kg^{-1} método multivariado PLSR com 118 amostras. Na região sul do Brasil, em uma bacia hidrográfica de 1.892 ha com distintos tipos de solos e uso da terra, Dotto et al. (2014) utilizando 303 amostras e modelo de MLR alcançaram R^2 de 0,62 e RMSE de 0,20 log g kg^{-1}, sendo a grande variação das características do

solo na paisagem as causas apontadas pelos autores pela baixa acurácia das predições. Em estudo realizado na Itália, Lucà et al. (2017), utilizando 216 amostras distribuídas em 33 ha, alcançaram R^2 de 0,79 e RMSE de 0,71 % com o modelo PLSR e R^2 de 0,82 e RMSE de 0,68 % com SVM. Os autores relataram a necessidade de mais pesquisas para definir as melhores técnicas de calibração, especialmente quando os solos são altamente heterogêneos. Em uma área da Amazônia Central no Brasil, Pinheiro et al. (2017) aplicando o modelo PLSR com pré-processamento espectral SNV e 434 amostras, obtiveram R^2 de 0,71 e RMSE de 5,69 g kg^{-1}. Os autores observaram que a heterogeneidade dos solos e a abrangência geográfica da biblioteca espectral são fatores determinantes no desempenho dos modelos e na sua aplicabilidade.

Resultados de predição do COS em escala local, com maior acurácia, comparado aos encontrados no presente estudo, foram obtidos por Dotto et al. (2017) na região central do Estado de Santa Catarina, Brasil. Os autores avaliaram a combinação de pré-processamentos espectrais, seleção de regiões do espectro e modelos para a predição de COS em uma biblioteca espectral de 299 amostras especializadas em 170,000 ha. A melhor acurácia obtida com o modelo PLSR, pré-processamentos CRR e utilizando todo o espectro Vis-NIR (R^2 de 0,90 e RMSE de 0,32 %). Em estudo de Vašát et al. (2017) em uma área de 100 ha na República Tcheca, com 107 amostras, os autores alcançaram R^2 de 0,81 e RMSE de 0,16 % por MLR, R^2 de 0,78 e RMSE de 0,17 % por PLSR, R^2 de 0,75 e RMSE de 0, 18% por RF e R^2 de 0,71 e RMSE de 0,19 % por SVM. Uma explicação para o desempenho superior desses trabalhos pode ser as características dos solos, uma vez que, não apresentaram valores elevados de argila e de óxidos de ferro comparados aos do presente trabalho (argila ~ 600 g kg^{-1} e óxidos de ferro ~ 234 g kg^{-1}), os quais são constituintes que influenciaram na assinatura espectral do solo dos Latossolos Vermelhos (DALMOLIN et al., 2005) e mascararam feições entre 400 e 700 nm correlacionadas com a predição de COS (VISCARRA ROSSEL et al., 2006). Isso também foi observado por Stevens et al. (2013), em que amostras com maior conteúdo de argila (~ 600 g kg^{-1}) mascararam regiões de absorção importantes para a predição do COS.

Com relação a trabalhos realizados com bibliotecas espectrais em escala regional, resultados superiores aos do presente estudo foram observados por Vasques et al. (2008) (R_v^2= 0,86 e RMSE = 0,18 log mg kg^{-1}) com 554 amostras e modelo PLSR na região da Flórida, EUA. Na África, Cambule et al. (2012) também encontrou bom desempenho dos modelos, com R^2 alcançando 0,85 e RMSE 0, 32%. Já resultados inferiores são encontrados em trabalho de Terra

et al. (2015), com o uso do modelo SVM (R^2= 0,72 e RMSE = 0,14 log g kg^{-1}). Porém, tais autores se utilizaram de um banco de 1259 amostras, representando oito regiões de três Estados diferentes do Brasil, em que a variabilidade de solos é alta. Da mesma forma, Araújo et al. (2014) obtiveram R^2 de 0,60 e RMSE de 0,55 % na predição para uma grande biblioteca espectral (n = 7172) de solos tropicais do Brasil. Desempenhos menores também são observados em trabalho de Ramirez-Lopez et al. (2013), realizado com 4200 amostras do Estado de São Paulo - Brasil, no qual obteve R^2 de 0,48 e RMSE de 0, 28%.

O uso de bibliotecas de escala nacional para construção de modelos preditos é apresentado em estudo de Knox et al. (2015), que alcançou R^2 de 0,87 e RMSE de 0,40 utilizando PLSR e pré-processamento SGD, com um conjunto de 1014 amostras coletadas em todo o Estado da Flórida, EUA. Na mesma linha, Viscarra Rossel e Behrens (2010) compararam diferentes modelos para a predição do COS por Vis-NIR com um conjunto de dados de 1104 amostras coletadas na Austrália, obtendo bons resultados (R^2 variando de 0,81 a 0,84). Wijewardane et al. (2016) usando 19.891 amostras distribuídas pelos EUA encontraram bom desempenho para PLSR (R^2 = 0,83 e RMSE = 7,38 %) e Rede Neural Artificial - RNA (R^2 = 0,96 e RMSE = 3,61 %). Já trabalho realizado na Espanha por Guerreiro et al. (2016) mostrou variação no desempenho dos modelos por PLSR conforme o número de amostras. A maior acurácia foi observada com 3482 amostras (R^2 = 0,91 e RMSE = 9,93 g kg^{-1}), intermediária com 1096 amostras (R^2 = 0,82 e RMSE = 25,6 g kg^{-1}) e menor com 362 amostras (R^2 = 0,80 e RMSE = 25,4 g kg^{-1}).

Em nível de escala global, estudo abordando a predição de COS foi apresentado primeiramente por Brown et al. (2006), onde os autores alcançaram desempenho de R^2 = 0,82 e RMSE = 9,0 g kg^{-1} com 3793 amostras e modelo *boosted regression trees*. Resultados semelhantes de RMSE foram obtidos por Ramirez-Lopez et al. (2013), utilizando-se 4438 amostras com modelo SVM (R^2 = 0,57 e RMSE = 9,3 g kg^{-1}), porém, com o modelo PLSR a performance reduziu (R^2 = 0,50 e RMSE = 10,8 g kg^{-1}). Em trabalho recente apresentado por Viscarra Rossel et al. (2016), em que foi compilado 23,631 amostras em nível global, a capacidade preditiva para o COS com o modelo cubist foi de R^2 = 0,92 e RMSE = 1,11 %. Os valores de RMSE destes estudos são inferiores aos obtidos no presente trabalho (RMSE = 0,52 %), sendo isso reflexo da escala das bibliotecas espectrais, uma vez que, a variabilidade dos fatores ambientais que influenciam no conteúdo de COS em escala global é grande (Guerreiro et

al., 2016). Nesse sentido, trabalhos realizados em escala regional, nacional e global mostram que a predição do COS por técnicas espectroscópicas apresenta potencial, no entanto, maior capacidade preditiva dos modelos pode ser alcançada com a inclusão de informações de bibliotecas espectrais locais (GOGÉ et al., 2014) e as características da composição do solo devem ser consideradas durante o processo de modelagem.

No entanto, tem se observado variações no desempenho dos modelos preditos de COS em escala local na literatura (DEMATTÊ et al., 2016; DOTTO et., 2017; LUCÀ et al., 2017; VASÁT et al., 2017), em que se observa distintos números de amostras por área, pré-processamento e modelos multivariados. Isso abre espaço para estudos que desenvolvam estratégias a serem utilizadas em dados espectrais de escala local, visando aumentar a capacidade preditiva dos modelos e buscar respostas para tal variação. No presente trabalho testamos a estratificação do CTA (n = 810) em grupos, conforme apresentado no próximo item.

5.3.5. Estratificação da BEL e desempenho dos modelos de predição

A partir da análise da variância dos dados espectrais (Figura 17), a BEL foi estratificada em grupos conforme mostra Tabela 4. Esses grupos foram utilizados na predição do COS para avaliar a estratégia de estratificação do CTA da BEL.

O valor de R_v^2 variou de 0,42 a 0,82; $RMSE_v$ de 0,29 a 0,70 % e $RPIQ_v$ de 1,99 a 2,60 (Tabela 5). Os modelos preditivos ajustados para os três primeiros grupos apresentaram desempenho superior comparado com o CTA (R_v^2 = 0,74; $RMSE_v$ = 0,55% e $RPIQ_v$ = 2,16). O modelo com melhor desempenho foi atingido no grupo LV+LA (redução de 45 % das amostras), com R^2_v = 0,82; $RMSE_v$ = 0,29 % e $RPIQ_v$ = 2,60. O segundo melhor desempenho foi observado para o grupo com amostras da camada C3 (redução de 68 % das amostras), com R^2_v = 0,77; $RMSE_v$ = 0,30 % e $RPIQ_v$ = 2,25. O modelo preditivo ajustado para o grupo LA (redução de 41 % das amostras) apresentou o terceiro melhor desempenho, com R^2_v = 0,76; $RMSE_v$ = 0,39 % e $RPIQ_v$ = 2,20. Esses modelos apresentam bom desempenho, com baixos valores e erro nas predições comparado aos obtidos por Lucà et al. (2017) (RMSE = 0,71 %) e Jiang et al. (2017) (RMSE = 0,88% e 0,54 %), que utilizaram o modelo PLSR em escala local.

Tabela 4 – Grupos de amostras formados após a estratificação da BEL e descrição das amostras que compõem cada grupo

Grupos	Descrição das amostras
LV + LA	amostras da classe de solo LV em áreas de LA
C3	amostras na camada de 0,15 – 0,30 m nas classes de solo LV e GX e usos da terra LA, CA, FL
LA	amostras de áreas de LA nas classes de solo LV e GX
CTA	conjunto total de amostras
LV	amostras da classe de solo LV nos usos da terra LA, CA e FL
CA	amostras de áreas de CA e nas classes de solo LV e GX
CA + FL + LV + GX	amostras de áreas de CA e FL nas classes de solos LV e GX
FL	amostras de áreas de FL e nas classes de solo LV e GX
C2	amostras na camada de 0,05 – 0,15 m nas classes de solo LV e GX e usos da terra LA, CA, FL
C1	amostras na camada de 0,00 – 0,05 m nas classes de solo LV e GX e usos da terra LA, CA, FL
CA + FL + GX	amostras de áreas de CA e FL na classe de solo GX
GX	amostras da classe de solo GX nos usos da terra LA, CA e FL
CA + FL + LV	amostras de áreas de CA e FL na classe de solo LV

Fonte: Autor. LV: Latossolo Vermelho; GX: Gleissolo Háplico; LA: lavoura; CA: campo nativo; FL: floresta nativa; CTA: conjunto total de amostras.

Tabela 5 – Resumo estatístico dos modelos gerados para cada grupo pelo modelo PLSR e pré-processamento SGD

Grupos	n_t	n	R^2	RMSE	RPIQ	n	R^2	RMSE	RPIQ
		calibração				validação			
LV + LA	445	312	0,75	0,37	2,41	134	0,82	0,29	2,60
C3	256	179	0,82	0,26	1,81	77	0,77	0,30	2,25
LA	478	335	0,76	0,36	2,54	143	0,76	0,39	2,20
CTA*	**810**	**567**	**0,68**	**0,54**	**2,11**	**243**	**0,74**	**0,55**	**2,16**
LV	646	452	0,55	0,47	1,84	194	0,74	0,56	2,60
CA	187	131	0,67	0,70	2,21	56	0,69	0,61	2,09
CA + FL + LV + GX	321	225	0,61	0,70	2,11	96	0,65	0,71	1,50
FL	143	100	0,55	0,70	2,11	43	0,61	0,78	1,64
C2	255	179	0,70	0,42	1,74	77	0,60	0,47	1,46
C1	256	179	0,50	0,66	1,67	77	0,55	0,79	1,17
CA + FL + GX	136	95	0,71	0,74	2,93	41	0,54	0,86	1,48
GX	164	115	0,71	0,71	2,60	49	0,51	0,85	1,76
CA + FL + LV	196	137	0,58	0,54	1,70	59	0,42	0,70	1,99

Fonte: Autor. n_t: número de amostras total; n: número de amostras; R^2: coeficiente de determinação; RMSE: raiz do erro quadrático médio em %; RPIQ: Proporção do erro de predição no intervalo interquartil; c: calibração; v: validação; LV: Latossolo Vermelho; GX: Gleissolo Háplico; LA: lavoura; CA: campo nativo; FL: floresta nativa; CTA: conjunto total de amostras. * modelo gerado com o conjunto total de amostras (CTA) da biblioteca espectral local (BEL).

Na sequência, o grupo formado apenas por amostras de LV (redução de 20 % das amostras) apresentou acurácia muito próxima ao CTA, com R^2_v = 0,74; $RMSE_v$ = 0,55 % e $RPIQ_v$ = 2,60. O grupo CA (n = 187) apresentou desempenho melhor comparado com os modelos ajustados para os grupos CA+FL+LV+GX (n = 321), camadas C1 (n = 256) e C2 (n = 255), que apresentam maior número de amostras (Tabela 5). O modelo FL (n = 143), com menos da metade das amostras do modelo C2 (n = 255), alcançou desempenho semelhante (Tabela 5). Esse comportamento também é observado nos modelos C1 e CA+FL+GX, em que o modelo CA+FL+GX (n = 136) apresentou R_v^2 = 0,54; $RMSE_v$ = 0,86 % e $RPIQ_v$ = 1,48. Os modelos ajustados para os grupos GX e CA+FL+LV apresentaram os menores desempenhos na predição do COS.

Os resultados encontrados para os três primeiros grupos mostram que o agrupamento de amostras com características pedológicas mais homogêneas teve efeito positivo no aumento da acurácia destes modelos e, a redução do número de amostras causado pela estratificação do CTA não teve efeito negativo na redução da acurácia para estes modelos (Tabela 5). Esses achados corroboram com os observados em trabalho de Araújo et al. (2014), para uma biblioteca espectral regional de solos tropicais do Brasil. Os autores alcançaram um R^2 de 0,60 e RMSE de 0,55 % na predição do COS. Na tentativa de melhorar o desempenho de previsão dos modelos, a biblioteca foi particionada em grupos mais uniformes em relação à mineralogia e observou-se aumento na acurácia dos modelos para R^2 = 0,71 e RMSE = 0,47 %. Na China, Jiang et al. (2017) avaliando o potencial de uma biblioteca espectral local na predição do COS, alcançaram melhores desempenhos dos modelos quando agruparam as amostras em camada superficial (RMSE = 0,88 %) e subsuperficial (RMSE = 0,54 %) comparado ao modelo gerado para o conjunto total (RMSE = 0,86 %). Em um estudo em escala nacional, o agrupamento de amostras com o objetivo de formar grupos mais homogêneos (espectralmente ou composicionalmente) aumentou a precisão das previsões de SOC, como observado em um estudo de Wijewardane et al. (2016). Os autores observaram que a estratificação da biblioteca espectral baseada na textura e horizontes de solo apresentaram maior desempenho na predição de SOC, sendo mais efetivos em relação à separação baseada na região ou no uso da terra.

A variação no desempenho de modelos de predição por Vis-NIR atribuída a variação das características químicas e físicas dos solos foi mencionada por Stenvens et al. (2013), que construiu modelos preditivos para uma biblioteca espectral contento 250.000 amostras de 23

estados na União Europeia, em que as amostras foram separadas em grupos conforme uso da terra. Os autores observaram variação nos valores de R^2 de 0,67 a 0,89 e RMSE de 4,0 a 50,6 g kg^{-1}, sendo as causas da variação atribuída ao teor de argila, areia e COS. A influência da granulometria do solo na predição de COS foi observada em estudo de Wight, Ashworth e Allen (2016), onde amostras com maior teor de argila apresentaram a maior porcentagem de erros na predição do COS. Nesse caso, os autores sugerem que, se for conhecida previamente a textura do conjunto de amostras, está deve ser separada em grupos para realizar a predição do COS com base na classe textural. Da mesma forma, Demattê et al. (2016) abordou a regionalização espectral em escala local e regional para a predição de propriedades do solo no Brasil, em que observou que a presença de solos com alta variação de seus constituintes prejudica a predição de propriedades químicas do solo.

A variação no comportamento espectral das amostras observada e discutida anteriormente nas Figuras 17 e 18, em que amostras de GX e CA apresentaram a maior variação, influenciou na capacidade preditiva. Modelos gerados com grupos em que essas amostras estão presentes apresentaram menor acurácia (Tabela 5). Tal fato está relacionado à variação das características desses solos, tornando o grupo de amostras muito heterogêneo. Quando os modelos são gerados a partir de grupos de amostras mais homogêneas, por exemplo, grupos formados por amostras de LV+LA, camada C3 e LA a acurácia das predições aumentou (Tabela 5).

No entanto, verifica-se que a redução de amostras também influenciou na acurácia dos modelos, sendo observada nos grupos formados por amostras de CA e FL (Tabela 5). Isso é relatado em estudo de Ramirez-Lopez et al. (2014), devendo-se tomar cuidado quando se trabalha com conjuntos pequenos de amostras, pois pequenos conjuntos de calibração podem levar a uma baixa cobertura da variação dos dados e refletir em maiores erros nas predições. Os autores também observaram que erros nas predições de conjunto de dados em escala regional são maiores que os observados em escala local. Isto ocorre, provavelmente, devido a uma maior complexidade de conjunto de dados de escala regional comparada com dados de escala local.

A influência do número de amostras na calibração de modelos preditivos de COS em escala local foi estudada por Lucà et al. (2017), em que observaram redução nos valores de R^2 e RMSE ao diminuir amostras, sendo o melhor resultado obtido com 114 amostras (R^2 = 0,79 e RMSE = 0,71 % por PLSR e R^2 = 0,82 e RMSE = 0,68 % por SVM). Os autores concluíram que

a capacidade preditiva dos modelos pode depender não apenas do tamanho da amostra, mas também da localização e variabilidade da área de estudo.

No presente estudo, apesar da biblioteca espectral ser local, os espectros apresentaram elevada variabilidade e a separação das amostras em grupos, com base nas características pedológicas e espectrais, mostrou ser eficiente para aumentar a acurácia de alguns modelos. Essa estratégia deve ser considerada na tomada de decisão para compor grupos de amostras em bibliotecas espectrais de solo para predição do COS em escala local, servindo de base para a agricultura de precisão.

5.4. CONCLUSÕES

A análise qualitativa das curvas espectrais apresentou características típicas que indicaram diferenças na composição física e química das amostras em relação às classes de solos, usos da terra e camadas de amostragem. O teor de óxido de ferro, argila e COS influenciaram as características espectrais das amostras, principalmente nos comprimentos de onda observados nas regiões de 400-800, 1400, 1900 e 2000-2350 nm.

A técnica de pré-processamento do SGD demonstrou os maiores valores de R^2_v e $RMSE_v$ para todos os modelos. Entre os métodos multivariados, o PLSR apresentou o melhor desempenho para predição do SOC, seguido pelos modelos SVM> MLR> e RF. As regiões de 400-800, 1900 e 2000-2350 nm foram consideradas as mais importantes para a predição de COS neste estudo.

A estratificação da biblioteca espectral local revelou que a separação e / ou agrupamento de amostras mais homogêneas em relação às características dos espectros, classes de solo, usos da terra e camadas do solo é uma estratégia para alcançar melhor precisão para as previsões de COS em comparação ao uso de todas as amostras juntas na calibração do modelo.

Este estudo destacou o potencial da aplicação da espectroscopia Vis-NIR como uma ferramenta confiável e econômica para quantificar concentrações de COS em solos subtropicais com altos níveis de óxidos de ferro e argila em escala local. Mais pesquisas são necessárias para determinar a eficácia da estratificação de bibliotecas espectrais para obter modelos de predição propriedades do solo mais acurados.

6. ESTUDO 2 – A ESTRATIFICAÇÃO DE UMA BIBLIOTECA ESPECTRAL DE SOLOS SUBTROPICAIS MELHORA AS PREDIÇÕES DE CARBONO ORGÂNICO NO SOLO?

6.1. INTRODUÇÃO

O interesse crescente sobre a captura de carbono levantou questões sobre qual o potencial e quanto de carbono orgânico está armazenado no solo (LAL et al., 2018). Essa quantificação é essencial para identificarmos áreas onde os solos se tornarão uma fonte ou sumidouros de carbono orgânico do solo (COS) frente aos cenários futuros de mudança climática global (BODDEY et al., 2010; JACKSON et al., 2017; SORENSEN et al., 2018). A crescente necessidade de dados do solo, especialmente de COS, tem sido referida na literatura como uma crise de dados do solo (MCBRATNEY et al., 2006; AMUNDSON et al., 2015). Para atender essa demanda métodos inovadores e mais sustentáveis para quantificar o COS são cada vez mais necessários. Dentre eles, a técnica de espectroscopia na região do Vis-NIR possibilita a obtenção de grande volume de dados de solo, com base no seu comportamento frente à energia eletromagnética.

Com o avanço das pesquisas em espectroscopia Vis-NIR na ciência do solo, os dados da análise espectral do solo passaram a ser organizados em bibliotecas espectrais (BE), para a construção de modelos de predição para quantificação do teor de carbono orgânico do solo (COS) em diferentes escalas geográficas (VASQUES et al., 2010; STEVENS et al., 2013; TERRA et al., 2015; ARAÚJO et al., 2014; NOCITA et al, 2014; DEMATTÊ et al., 2016; GUERREIRO et al., 2016; VICARRA ROSSEL et al., 2016; WIJEWARDANE et al., 2016; JIANG et al., 2017; DOTTO et al., 2018; MOURA-BUENO et al., 2019). O objetivo é obter o teor de COS de uma amostra de forma rápida, barata, não destrutiva e livre de resíduos químicos, ao contrário das metodologias tradicionais (VISCARRA ROSSEL et al., 2008; O'ROURKE; HOLDEN, 2011; SCHWARTZ et al., 2012; NOCITA et al., 2015). A definição da escala da BE está relacionada à cobertura territorial das amostras de solo coletadas, de forma que, à medida que aumenta o número de amostras que compõem uma BE, e principalmente, quando a extensão geográfica de abrangência dessas amostras cresce, aumenta a heterogeneidade das características do solo e a

variação espectral (ARAÚJO et al., 2014; VICARRA ROSSEL et al., 2016). A qualidade dos produtos gerados a partir da BE, portanto, é limitada pelas características do banco de dados utilizado e pela modelagem de dados a partir dos quais ele deriva (VICARRA ROSSEL et al., 2016).

O teor de COS pode ser predito a partir de BE em escala regional (DEMATTÊ et al., 2016; GUERRERO et al., 2016; WIJEWARDANE et al., 2016), continental (NOCITA et al., 2014; Stevens et al., 2013) e global (VISCARRA ROSSEL et al., 2016) com precisão razoável. No entanto, desenvolver modelos espectroscópicos precisos continua sendo uma tarefa desafiadora. Esta dificuldade é agravada quando as BE são compostas por amostras de diversas origens geográficas, material parental, protocolos de medição e condições ambientais (GUERRERO et al., 2010). A complexidade intrínseca do recurso natural solo devido à variabilidade dos fatores de formação implica em variância dos componentes do solo e de sua respectiva variância espectral (STENBERG et al., 2010; DEMATTÊ et al., 2014; VISCARRA ROSSEL et al., 2016) que influenciam diretamente no desempenho dos modelos de predição (VISCARRA ROSSEL et al., 2006; BELLON-MAUREL; MCBRATNEY, 2011; MOURA-BUENO et al., 2019) como, a composição granulométrica (STENBERG et al., 2002; STENBERG et al., 2010; WIJEWARDANE et al., 2016), mineralogia (ARAÚJO et al., 2014), classes de solos (VASQUES; GRUNWALD; HARRIS, 2010) e matéria orgânica (WIGHT; ASHWORTH; ALLEN, 2016).

Estudos mostram que a adição de covariáveis ambientais, tais como clima, geologia e uso da terra (RIAL et al., 2016) e coordenadas geográficas (TZIOLAS et al., 2019) podem ser úteis na modelagem espectroscópica. Apesar disso, a maior parte das pesquisas que se comprometem em melhorar a acurácia das predições se limita a avaliar estratégias de pré-processamento espectral e métodos multivariados (VASQUES et al. 2008; RAMIREZ-LOPEZ et al., 2013; CAMBULE et al., 2012; PINHEIRO et al., 2017; DOTTO et al., 2017; VASÁT et al., 2017; DOTTO et al., 2018). Sendo a variância do solo reflexo das características ambientais, a utilização de dados homogêneos pode ser uma estratégia valiosa para alcançar maior acurácia dos modelos, principalmente quando o objetivo é gerar dados para fomentar projetos de conservação do solo em bacias hidrográficas e agricultura de precisão que normalmente demandam maior acurácia.

A redução da variância de uma BE pode ser alcançada pela estratificação desta em sub-bibliotecas. Os critérios de estratificação podem ser ambientais, baseados no local de origem da amostragem, ou estatísticos, baseados na amplitude de variância espectral. Porém, ao passo que a estratificação pode melhorar acurácia das predições de COS pela redução variância da BE (VASQUES et al., 2010; ARAÚJO et al., 2014; VISCARRA-ROSSEL et al., 2016; WIJEWARDANE et al., 2016; MOURA-BUENO et al., 2019), em alguns casos pode não sortir efeito (DEMATTÊ et al., 2016), ou, ainda, prejudicar devido a perda da capacidade amostral em representar adequadamente a população que se deu origem (RAMIREZ-LOPEZ et al., 2014). Não há, portanto, um consenso na literatura que indique os critérios de estratificação amostral de BE adequados, especialmente para solos tropicais (VISCARRA ROSSEL et al., 2016). Entender o desempenho dos modelos espectrais nesse cenário é decisivo para direcionar estratégias futuras de modelagem espectroscópica em diferentes escalas geográficas e ambientes.

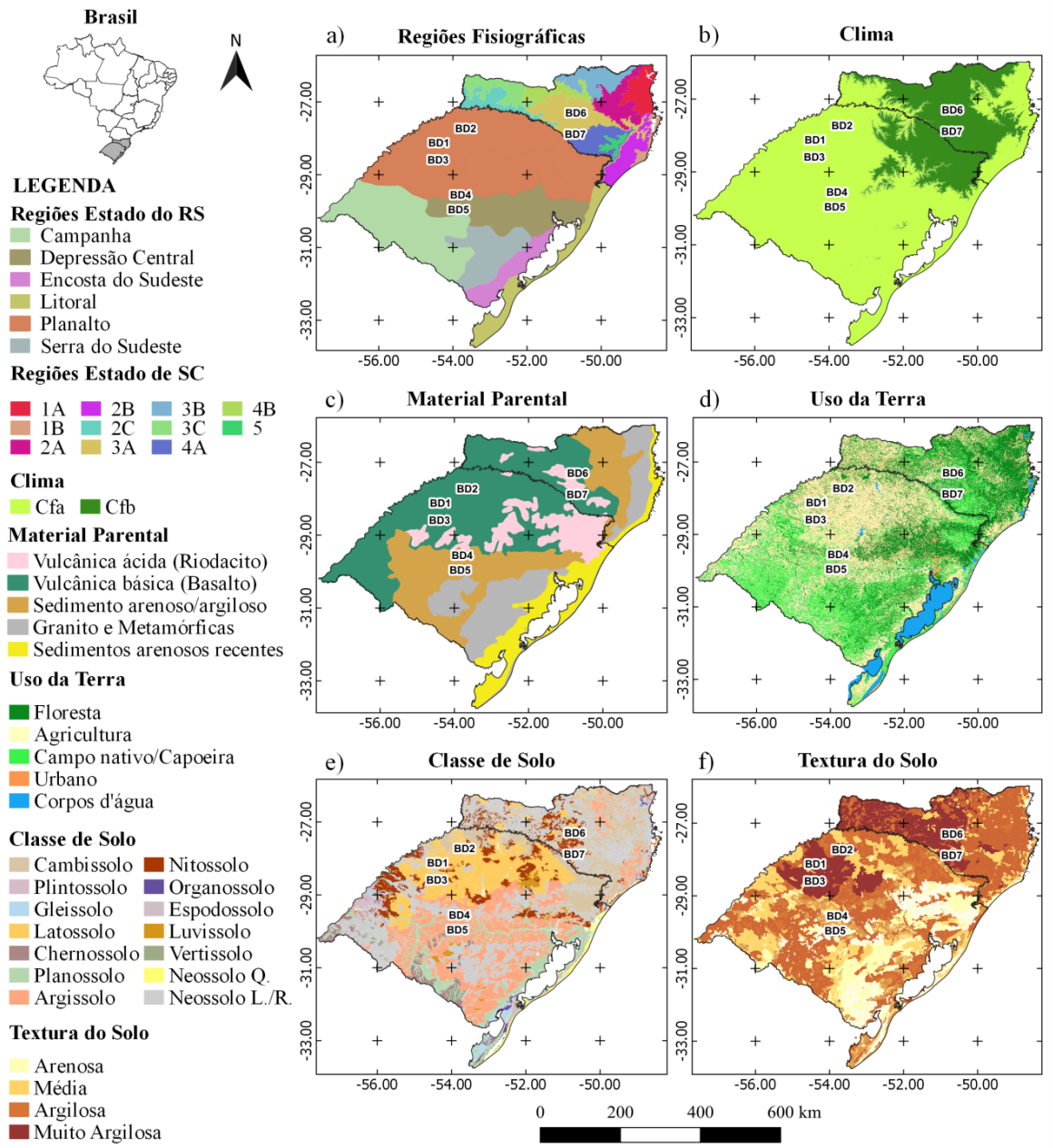
Esse cenário nos sugere que, a capacidade preditiva dos modelos está atrelada a composição das amostras e suas variâncias, a estratificação de uma BE aumenta a acurácia dos modelos espectroscópicos de COS, mas não está claro em que circunstância funciona. Para comprovar se essa hipótese é verdadeira, o objetivo principal do presente estudo foi avaliar o desempenho dos modelos de predição do COS a partir da estratificação amostral com base em quatro critérios, três deles baseados em características ambientais das amostras: i) região fisiográfica; ii) classe textural do solo, iii) uso da terra e, um critério estatístico, baseado na variância espectral: iv) classe espectral. Somado a isso exploramos qualitativamente os espectros, identificando as características que influenciam o comportamento espectral dos solos subtropicais, e, testamos o efeito de seis pré-processamentos dos espectros e cinco métodos de calibração multivariada na modelagem espectroscópica.

6.2. MATERIAL E MÉTODOS

6.2.1. Descrição dos bancos de dados

Foram utilizados sete bancos de dados (BD), oriundos de diferentes locais dos Estados do Rio Grande do Sul – RS e Santa Catarina – SC (Figura 22). A seguir, é apresentada uma descrição de cada BD, separadamente.

Figura 22 – Localização no sul do Brasil dos bancos de dados da BE e mapa de regiões (a), clima (b), material parental (c), uso da terra (d), classes de solo (f) e classes texturais (e)

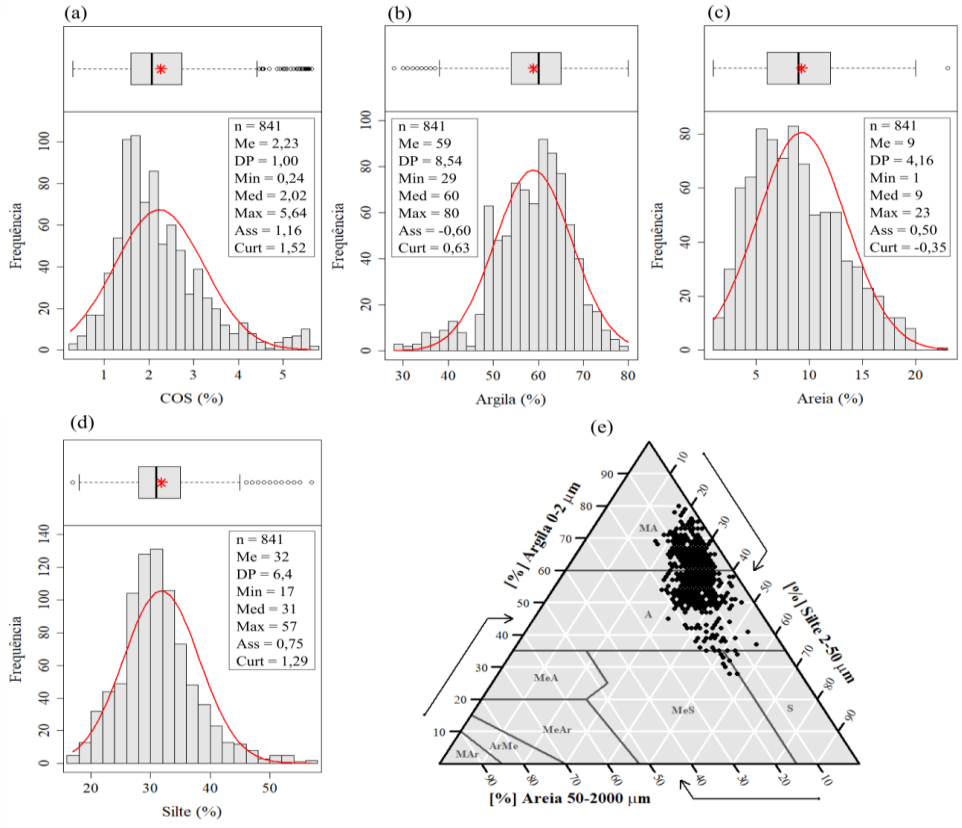


Fonte: Autor. Banco de dados 1; BD2: Banco de dados 2; BD3: Banco de dados 3; BD4: Banco de dados 4; BD5: Banco de dados 5; BD6: Banco de dados 6; BD7: Banco de dados 7; Regiões Fisiográficas de SC segundo Thomé et al. (1999), em que 1A:Litoral Norte, Vales dos Rios Itajaí e Tijucas; 1B: Litoral de Florianópolis e Laguna; 2A: Alto Vale do Rio Itajaí; 2B:Carbonífera, Extremo Sul e Colonial Serrana; 2C: Vale do Rio Uruguai; 3A: Planalto Central; 3B: Planalto Norte Catarinense; 3C: Noroeste Catarinense; 4A: Campos de Lages; 4B: Alto Vale do Rio do Peixe e Alto Iran: 5: Planalto Serrano de São Joaquim; C: Regiões Fisiográficas do RS segundo Fortes (1959); Clima do RS e SC segundo a classificação de Köppen's (ALVARES et al., 2013); Material Parental do RS (WILDNER et al., 2008) e SC (WILDNER et al., 2014); Uso da Terra segundo projeto MapBiomas - ano 2017 (http://mapbiomas.org/); Mapa de Classe de Solos do RS (BRASIL, 1973) e SC (POTTER et al., 2004), em que: Neossolo L./R. – Neossolo Litólico e Neossolo Regolítico e Neossolo Q. – Neossolo Quartzarênico; Textura do Solo: derivada do mapa de solos do RS e SC.

6.2.1.1. Banco de dados 1 – BD1

O BD1 compreende dados de solo de 261 observações, nas camadas 0 – 0,05; 0,05 – 0,15 e 0,15 – 0,30 m (261 * 3 = 783), e 29 pontos nas camadas 0,30 – 0,60 e 0,60 – 1,0 m (29 * 2 = 58), conforme especificação do consórcio GlobalSoilMap (ARROUAYS et al., 2014), totalizando 841 amostras de solo. Essas amostras foram utilizadas anteriormente no ESTUDO 1. A amostragem foi realizada entre os anos de 2012 e 2015, em uma área de aproximadamente 940 ha, localizada no município de Giruá, região fisiográfica do Planalto do RS, Brasil (Figura 22).

Figura 23 – Estatística descritiva e histogramas dos teores de COS, argila, silte e areia e triângulo textural do BD1



Fonte: Autor. n: número de amostras, Me: média, DP: desvio-padrão, Min: mínimo, Med: mediana, Max: máximo, Ass: assimetria, Curt: curtose, MA: muito argilosa, A: argilosa, S: siltosa, MeS: média siltosa, MeA: média argilosa, MeAr: média arenosa, ArMe: arenosa média, Mar: muito arenosa.

A área de abrangência das amostras varia entre 339 e 426m de altitude e com geologia composta por rochas ígneas básicas (basalto) da Formação Serra Geral. As classes de solos

predominante na área são os Latossolos Vermelhos Distroférricos típicos e Gleissolos Háplicos Tb Distrófico típicos. O clima da região, segundo a classificação de Köppen, é do tipo Cfa – subtropical (ALVARES et al., 2013). O relevo predominante é o suave ondulado a ondulado. As coberturas da terra predominantes na área são: mata nativa, campo nativo e lavouras anuais em sistema plantio direto (SPD).

Neste banco de dados predominam amostras com textura argilosa e muito argilosa (Figura 23e) e com teores de carbono entre 0,24 e 5,64 % (Figura 23a) e mineralogia predominante na fração argila, composta por caulinita e óxidos de ferro (hematita) (DALMOLIN et al., 2017).

Parte das amostras deste BD foi utilizada em projeto de pesquisa de mapeamento digital de solo (MOURA-BUENO, 2014).

6.2.1.2. Banco de dados 2 – BD2

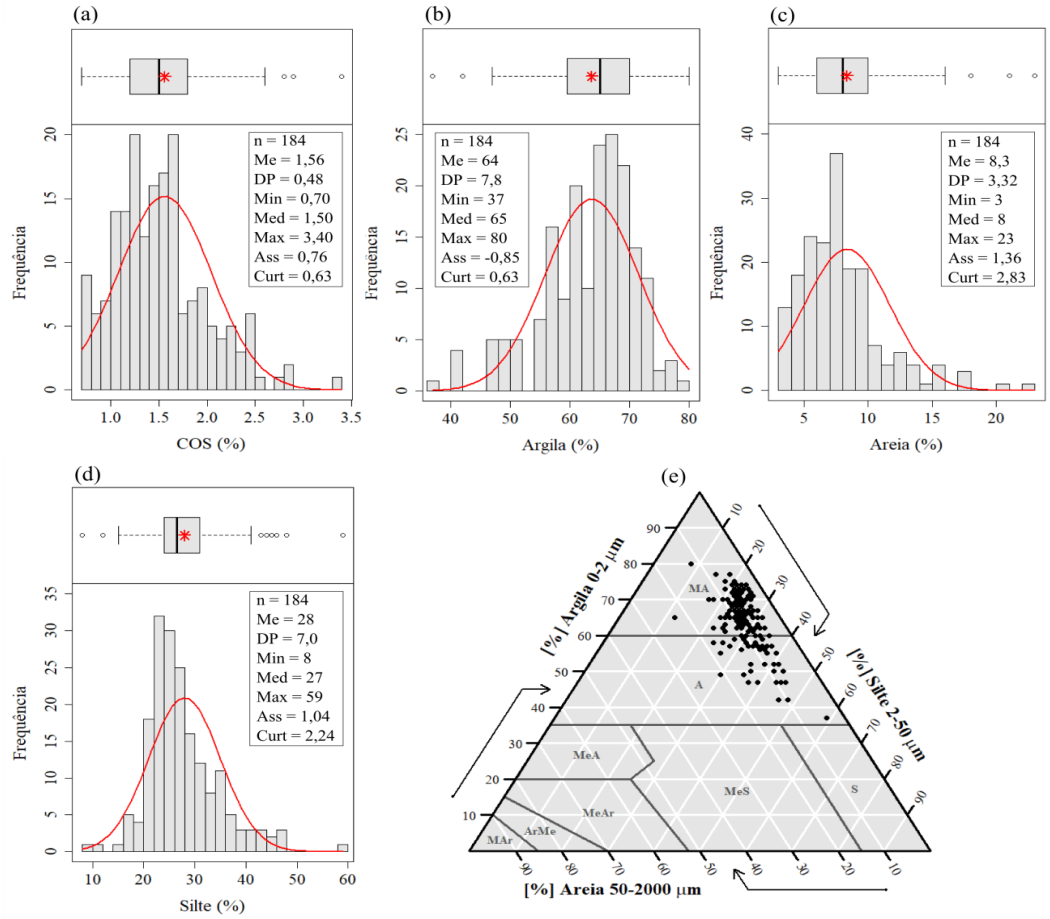
O BD2 compreende 87 observações de solo feitas em 2014, em uma área agrícola de aproximadamente 22 ha, no município de Santo Augusto, no Estado do RS, localizado na região fisiográfica do Planalto do RS, Brasil (Figura 22). A região tem geologia composta por rochas ígneas básicas (basalto) da Formação Serra Geral. A altitude média é de 495 metros, com predomínio de Latossolos Vermelhos Distroférricos típicos.

O clima da região, segundo a classificação de Köppen, é do tipo Cfa – subtropical do ponto de vista térmico e temperado chuvoso, com precipitação média anual elevada e bem distribuída ao longo do ano, apresentando média anual de pluviosidade de 1833 mm. A temperatura média da região é em torno de 19°C e o relevo predominante é o suave ondulado e ondulado. A cobertura da terra predominante na área é lavoura.

Os locais de observação foram selecionados intencionalmente a partir do estabelecimento de três transecções a partir do topo para o sentido do caimento mais suave do declive, distanciadas 50 metros uma da outra. Em cada transecção foram definidos 29 pontos com distância regular de 50 metros. Em cada ponto de observação coletou-se amostras de solo na profundidade de 0,00-0,05, 0,05-0,15, 0,15-0,30 e 0,30-0,60 m, conforme proposição do consórcio GlobalSoilMap.net, totalizando 348 amostras. No presente estudo, foram utilizadas apenas 179 amostras devido à falta de dados espectrais das demais amostras.

Neste banco de dados, predominam amostras com textura argilosa e muito argilosa (Figura 24e) e com teores de carbono entre 0,7 e 3,4 % (Figura 24a) e mineralogia predominante na fração argila, composta por caulinita e óxidos de ferro (hematita) (DALMOLIN et al., 2017). Estes dados do solo foram produzidos durante o desenvolvimento do projeto de pesquisa que visava utilizar a suscetibilidade magnética na estimativa de atributos do solo (RAMOS et al, 2017).

Figura 24 – Estatística descritiva e histogramas dos teores de COS, argila, silte e areia e triângulo textural do BD2

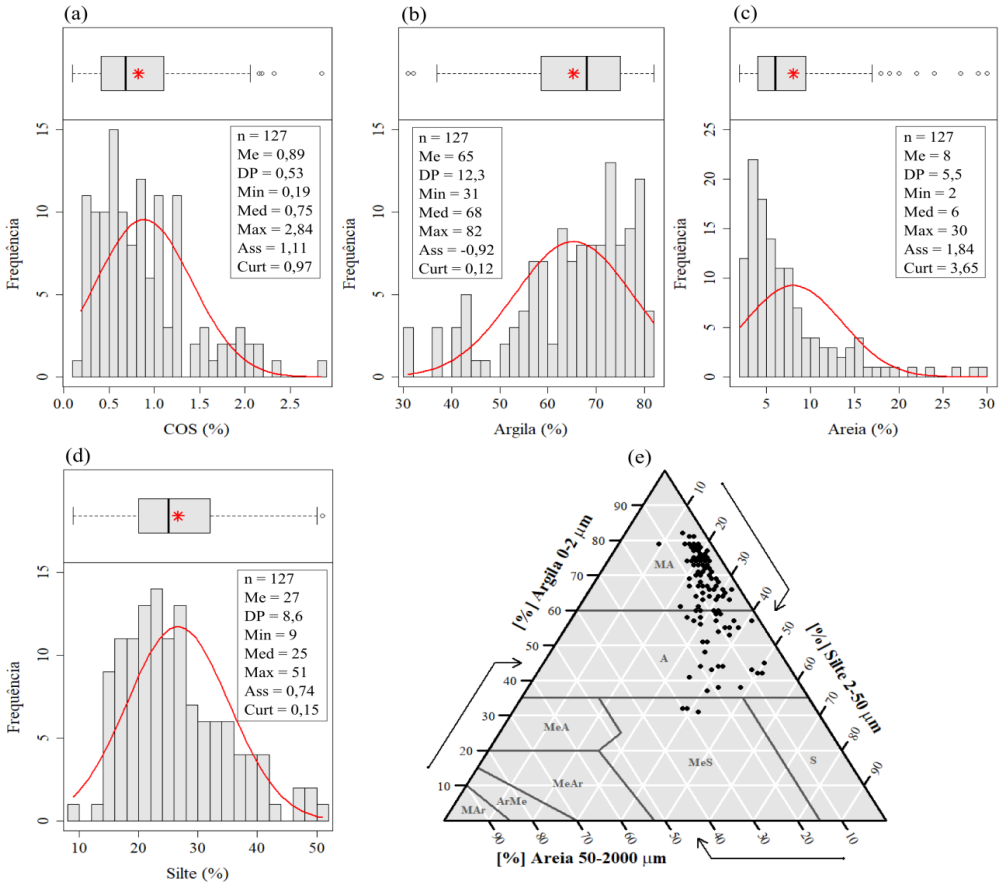


Fonte: Autor. n: número de amostras, Me: média, DP: desvio-padrão, Min: mínimo, Med: mediana, Max: máximo, Ass: assimetria, Curt: curtose, MA: muito argilosa, A: argilosa, S: siltosa, MeS: média siltosa, MeA: média argilosa, MeAr: média arenosa, ArMe: arenosa média, Mar: muito arenosa.

6.2.1.3. Banco de dados 3 – BD3

O BD3 compreende dados de 34 observações (perfis) de solo, feitas em 2014 em uma área agrícola de aproximadamente 1000 ha, no município de São Miguel das Missões, localizado na região fisiográfica do Planalto do RS, Brasil (Figura 22).

Figura 25 – Estatística descritiva e histogramas dos teores de COS, argila, silte e areia e triângulo textural do BD3



Fonte: Autor. n: número de amostras, Me: média, DP: desvio-padrão, Min: mínimo, Med: mediana, Max: máximo, Ass: assimetria, Curt: curtose, MA: muito argilosa, A: argilosa, S: siltosa, MeS: média siltosa, MeA: média argilosa, MeAr: média arenosa, ArMe: arenosa média, Mar: muito arenosa.

A área apresenta geologia composta por rochas ígneas básicas (basalto) da Formação Serra Geral. O relevo local é pouco movimentado, com predominância de plano à suave ondulado e com altitudes que variam de 220 a 320 metros, aproximadamente. A área caracteriza-se pela predominância de solos profundos com a ocorrência do Latossolo Vermelho Distroférrico típico.

O clima predominante da região é do tipo Cfa, de acordo com o sistema de classificação de Köppen (ALVARES et al., 2013) e cobertura da terra predominante na área é lavoura.

Todos os locais de observação foram selecionados intencionalmente e, em cada um, foram coletadas amostras em cada horizonte pedogênico descrito, totalizando 127 amostras. Neste banco de dados predominam amostras com textura argilosa e muito argilosa (Figura 25e) e com teores de carbono entre 0,19 e 2,84 % (Figura 25a) e mineralogia predominante na fração argila composta por caulinita e óxidos de ferro (hematita) (DALMOLIN et al., 2017).

Estes dados do solo foram produzidos durante o desenvolvimento do projeto de pesquisa que visava o mapeamento digital de solos (WOLSKI et al, 2017).

6.2.1.4. Banco de dados 4 – BD4

O BD4 compreende dados de 303 observações de solo feitas em 2011 na bacia hidrográfica do Rio Vacacaí-Mirim, de aproximadamente 2000 ha entre os municípios de Santa Maria e Itaara, localizado entre as regiões fisiográficas da Depressão Central e do Planalto do Estado do RS, Brasil (Figura 22).

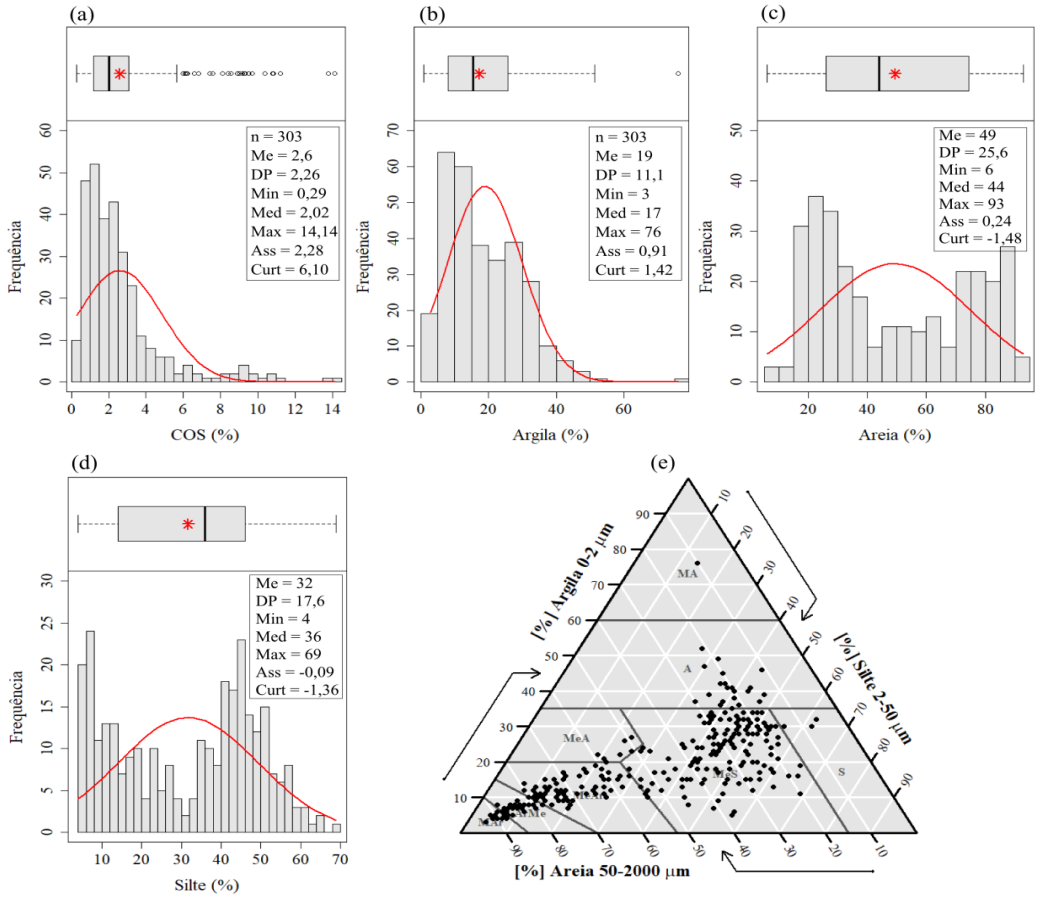
A geologia local consiste em rochas ígneas básicas, intermediárias e ácidas (riolito-riodacito e andesito-basalto) do período Cretáceo, rochas sedimentares consolidadas (arenitos eólicos e fluviais) e não consolidados (depósitos fluviais e coluviais) do Quaternário (MACIEL FILHO, 1990; SARTORI, 2009).

A bacia hidrográfica se caracteriza por diferentes usos e tipos de solo. Mais da metade da área tem ocupação do solo com Floresta Nativa seguida de áreas de Campo Nativo, Capoeira, Lavouras anuais, em menor expressão Silvicultura (Eucalipto) e área Urbanizada (SAMUEL-ROSA et al., 2011). Em mais de 50% da área são encontrados solos pouco desenvolvidos, Neossolos Litólicos, geralmente em locais de maior declividade. Também são encontrados solos mais profundos, como Argissolos Vermelhos, em uma proporção de 12%, e em locais com características de hifromorfismo, Planossolos Háplicos, pouco expressivos, com participação de 1,6% da área (MIGUEL et al., 2011). O clima predominante da região é do tipo Cfa, de acordo com o sistema de classificação de Köppen (ALVARES et al., 2013).

Todos os locais de observação foram selecionados intencionalmente e, em cada um, foram coletadas amostras na profundidade 0 – 20 cm, totalizando 303 amostras. Neste banco de dados,

predominam amostras com textura arenosa, média arenosa e média siltosa (Fgura 26e) e, com teores de carbono entre 0,29 e 14,1 % (Figura 26a). A mineralogia predominante na fração argila é composta por caulinita, óxidos de ferro e 2:1 com hidróxi-Al entre camadas (SCHMITT, 2015).

Figura 26 – Estatística descritiva e histogramas dos teores de COS, argila, silte e areia e triângulo textural do BD4



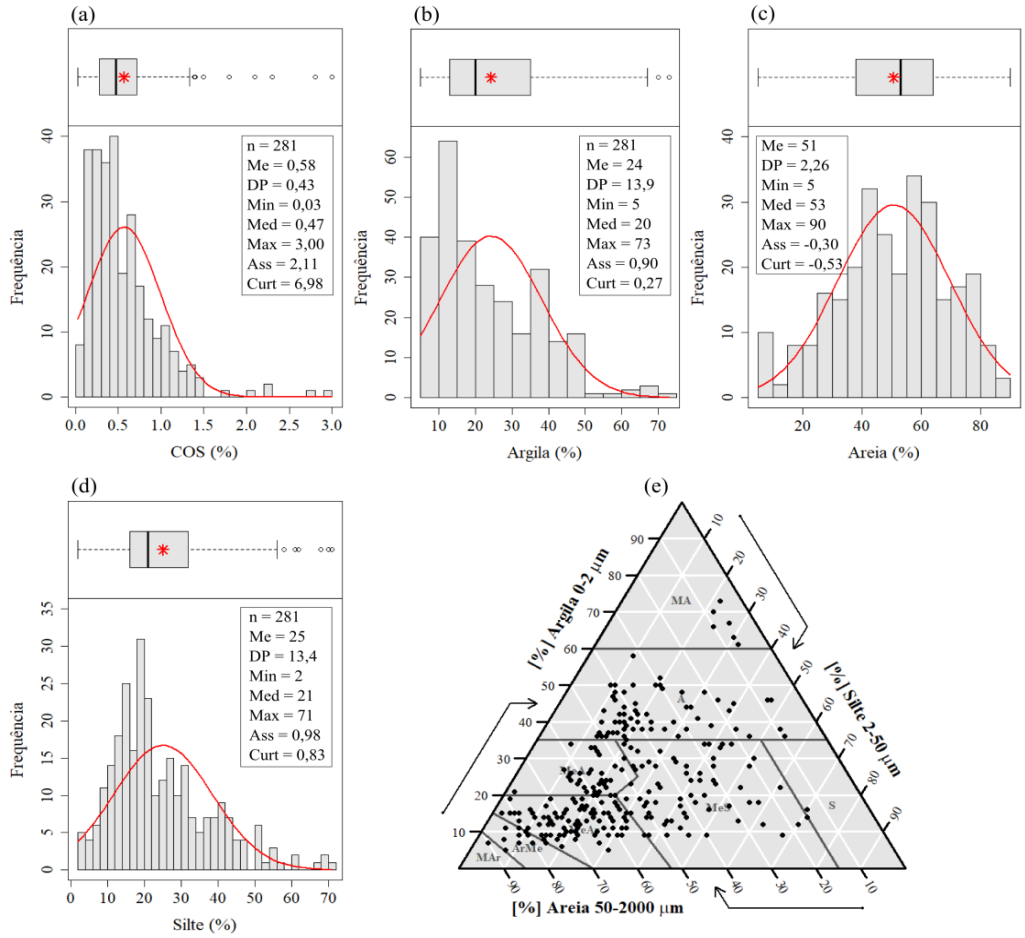
Fonte: Autor. n: número de amostras, Me: média, DP: desvio-padrão, Min: mínimo, Med: mediana, Max: máximo, Ass: assimetria, Curt: curtose, MA: muito argilosa, A: argilosa, S: siltosa, MeS: média siltosa, MeA: média argilosa, MeAr: média arenosa, ArMe: arenosa média, Mar: muito arenosa.

Estes dados do solo foram produzidos durante o desenvolvimento do projeto de pesquisa que visavam produzir informações de solo para essa bacia hidrográfica (SAMUEL-ROSA et al., 2011; MIGUEL et al., 2011; DOTTO et al., 2014, SAMUEL-ROSA et al., 2015; MOURA-BUENO et al., 2018).

6.2.1.5. Banco de dados 5 – BD5

O BD5 compreende dados de 72 observações de solo (perfis), totalizando 281 amostras, localizados no município de Santa Maria, na região fisiográfica da Depressão Central do Rio Grande do Sul, Brasil (Figura 22).

Figura 27 – Estatística descritiva e histogramas dos teores de COS, argila, silte e areia e triângulo textural do BD5



Fonte: Autor. n: número de amostras, Me: média, DP: desvio-padrão, Min: mínimo, Med: mediana, Max: máximo, Ass: assimetria, Curt: curtose, MA: muito argilosa, A: argilosa, S: siltosa, MeS: média siltosa, MeA: média argilosa, MeAr: média arenosa, ArMe: arenosa média, Mar: muito arenosa.

A área apresenta geologia composta por rochas ígneas básicas, intermediárias e ácidas (riolito-riodacito e andesito-basalto) do período Cretáceo, rochas sedimentares consolidadas (arenitos eólicos e fluviais) e não consolidados (depósitos fluviais e coluviais) (MACIEL FILHO,

1990; SARTORI, 2009). A altitude varia entre 80 e 400 m. As classes de solos predominantes na área são: Argissolos, Planossolos, Neossolos e Cambissolos (FLORES et al., 2015). O clima predominante é do tipo Cfa, conforme a classificação de Köppen (ALVARES et al., 2013). As coberturas da terra predominantes na área são lavouras e campo nativos e floresta nativa (FLORES et al., 2015).

Todos os locais de observação foram selecionados intencionalmente e por conveniência e em cada perfil de solo foram coletadas amostras nos horizontes pedogenético descritos, totalizando 286 amostras. Neste banco de dados, predominam amostras com textura que varia de muito arenosa até muito argilosa (Figura 27e) e com teores de carbono entre 0,03 e 3 % (Figura 27a). A mineralogia predominante na fração argila é composta por caulinita, óxidos de ferro e 2:1 com hidróxi-Al entrecamadas (SCHMITT, 2015).

Estes dados do solo foram produzidos durante o desenvolvimento do projeto de zoneamento edáfico para culturas no município de Santa Maria - RS (FLORES et al., 2015).

6.2.1.6. Banco de dados 6 – BD6

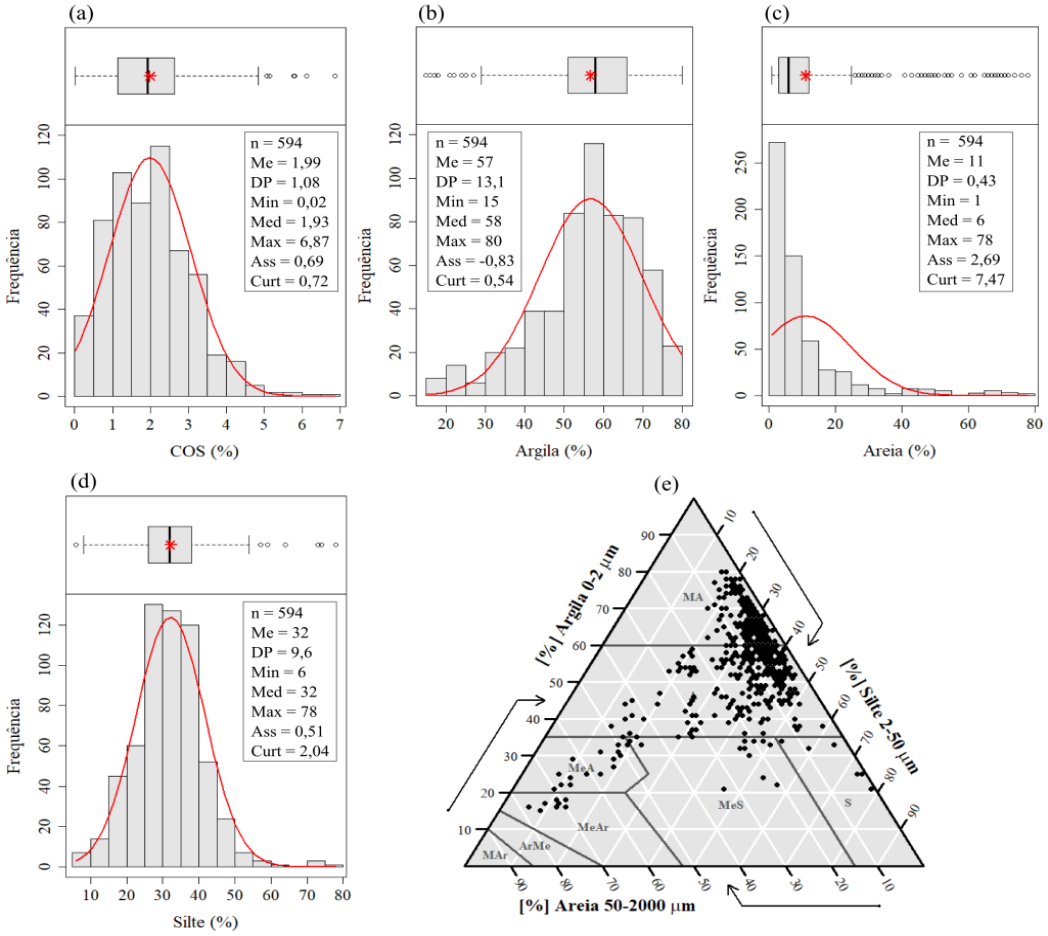
O BD6 compreende dados de 12 perfis de solo e 104 pontos de observações de solo realizadas em 2014, em parte da bacia hidrográfica do Rio Marombas, área de aproximadamente 180.000 ha (1800 km^2), localizada entre os municípios de Curitibanos, Brunópolis, Frei Rogério, Ponte Alta do Norte e São Cristovão do Sul, na região do Planalto Central do Estado de SC, Brasil (Figura 22).

A geologia da região é composta por rochas ígneas básicas, intermediárias e ácidas rochas (riolito-riodacito e andesito-basalto) da Formação Serra Geral. As altitudes variam entre 800 e 1300 m. As classes de solos predominantes na área são Latossolos Vermelhos, Nitossolos Vermelhos, Cambissolos e Neossolos. As coberturas da terra predominantes na área são: campo nativo, floresta nativa, floresta planta de pinus e lavouras. Todos os locais de observação foram selecionados pelo algoritmo Hipercubo Latino Condicionado (*Conditioned Latin Hipercube Sampling* - cLHS) (MINASNY; MCBRATNEY, 2006). Em cada ponto foram coletadas amostras seguindo o método recomendado pelo Consórcio GlobalSoilMap.net, nas 5 camadas: 0,0 – 0,05 m; 0,05 – 0,15 m; 0,15 – 0,30 m; 0,30 - 0,60 m e 0,60 – 1,00 m, além de amostras nos horizontes pedogenéticos de cada perfil descrito, totalizando em 595 amostras. Neste banco de dados,

predominam amostras com textura que varia de muito média arenosa até muito argilosa (Figura 28e) e com teores de carbono entre 0,02 e 6,87 % (Figura 28a) e mineralogia predominante na fração argila composta por caulinita e óxidos de ferro (hematita e goethita) (ALMEIDA et al., 2000; DALMOLIN et al., 2017).

Estes dados do solo foram produzidos durante o desenvolvimento do projeto de pesquisa que visava realizar predições de carbono orgânico do solo (DOTTO et al., 2018).

Figura 28 – Estatística descritiva e histogramas dos teores de COS, argila, silte e areia triângulo textural do BD6

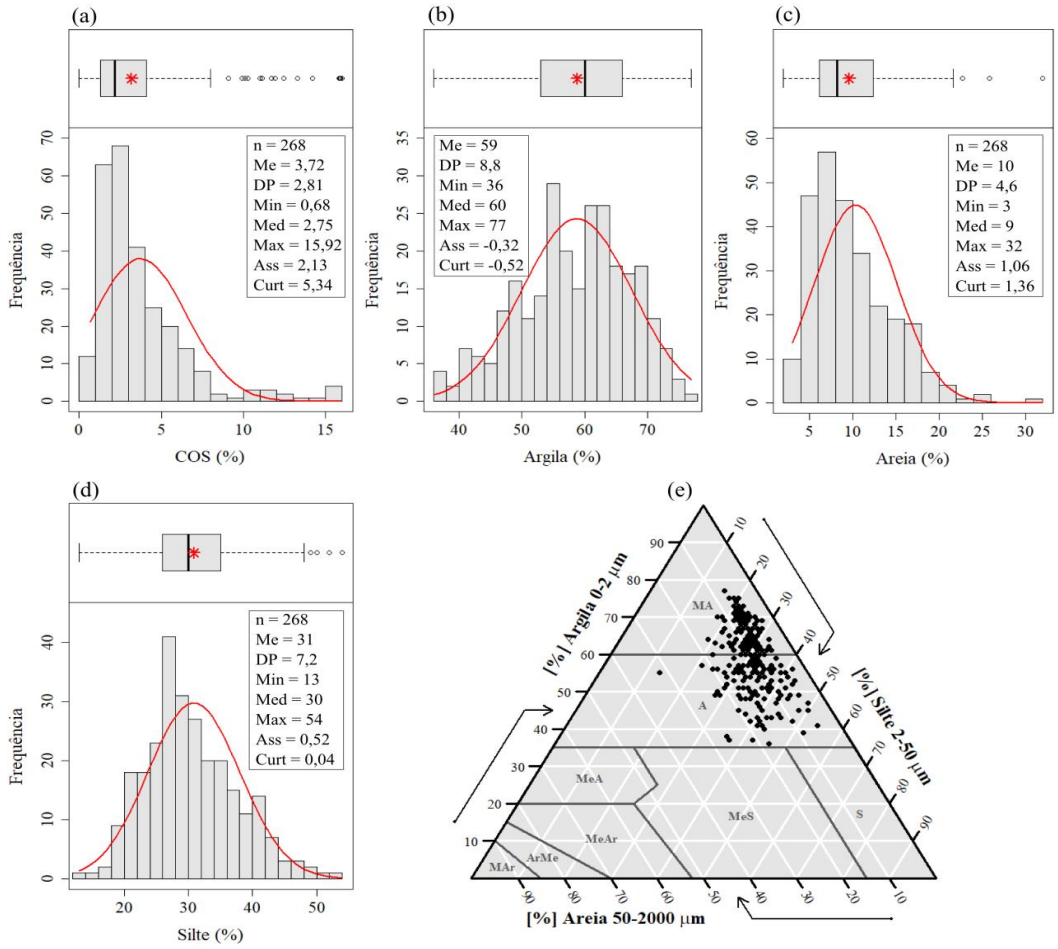


Fonte: Autor. n: número de amostras, Me: média, DP: desvio-padrão, Min: mínimo, Med: mediana, Max: máximo, Ass: assimetria, Curt: curtose, MA: muito argilosa, A: argilosa, S: siltosa, MeS: média siltosa, MeA: média argilosa, MeAr: média arenosa, ArMe: arenosa média, Mar: muito arenosa.

6.2.1.7. Banco de dados 7 – BD7

O BD7 compreende dados de 107 observações de solo feitas em 2016, em uma floresta plantada de *Pinus taeda*, com aproximadamente 108 ha, no município de Campo Belo do Sul, localizado na região dos Campos de Lages do Estado de SC, Brasil (Figura 22).

Figura 29 – Estatística descritiva e histogramas dos teores de COS, argila, silte e areia e triângulo textural do BD7



Fonte: Autor. n: número de amostras, Me: média, DP: desvio-padrão, Min: mínimo, Med: mediana, Max: máximo, Ass: assimetria, Curt: curtose, MA: muito argilosa, A: argilosa, S: siltosa, MeS: média siltosa, MeA: média argilosa, MeAr: média arenosa, ArMe: arenosa média, Mar: muito arenosa.

A geologia regional é composta por rochas ígneas básicas, intermediárias e ácidas rochas (riolito-riodacito e andesito-basalto) da Formação Serra Geral. A altitude varia entre 865 e 935 m.

As classes de solos predominante na área são os Neossolos Litólicos, Neossolos Regolíticos, Cambissolos Húmicos, Latossolos Vermelho-Amarelos, Gleissolos Melânicos e Organossolos.

Todos os locais de observação foram selecionados pelo algoritmo Hipercubo Latino Condicionado (*Conditioned Latin Hipercube Sampling* - cLHS) (MINASNY; MCBRATNEY, 2006) e, em cada um, foram coletadas amostras em 4 profundidades: 0 – 20, 20 – 40, 40 – 60, 60 – 100 cm, totalizando 261 amostras. A classe textural das amostras contidas neste banco de dados varia de muito arenosa até muito argilosa (Figura 29e). O teor de COS varia entre 0,68 e 15,92 % (Figura 29a) e mineralogia predominante na fração argila composta por caulinita e óxidos de ferro (DALMOLIN et al., 2017).

Estes dados do solo foram produzidos durante o desenvolvimento do projeto de pesquisa que visava produzir mapas digitais de propriedades do solo para utilização na silvicultura (HORST, 2017).

6.2.2. Análises de solo e espectrais

A determinação do COS em amostras derivadas de solos com teor elevado de COS foi realizada pelo método de combustão seca (CHNS/O) em analisador elementar (modelo Flash EA 1112, Thermo Finigan). Nas demais amostras o COS foi determinado via combustão úmida com aquecimento externo (bloco de digestão), de acordo com Yeomans e Bremner (1988) com posterior multiplicação por 1,14 para equivalência entre os métodos analíticos (RHEINHEIMER et al., 2008). Em seguida, os valores de COS foram utilizados como dados de referência para calibração e validação dos modelos de predição. A determinação da granulometria (fração argila, silte e areia) foi realizada pelo método da pipeta, utilizando NaOH 1 mol L^{-1} como agente dispersante, conforme Teixeira et al. (2017).

A reflectância espectral das amostras de solo foi obtida pelo espectroradiômetro FildSpec 3 (Analytical Spectral Devices, Boulder, USA), com capacidade de amostragem no intervalo de 350 à 2500 nm e resolução espectral de 1 nm. Para realização da leitura, foi utilizado aproximadamente 15 cm^3 de solo acondicionado em placa de petri. Como fonte de energia, foi utilizada uma lâmpada de halogênio 50 W com o feixe colimado não-alvo para o plano, posicionado a 35 cm da amostra num ângulo de 30º zênite. Uma placa branca *Spectralon* foi utilizada como padrão de referência de 100% de reflectância para a calibração, a cada 20 minutos

do espectroradiômetro. Foram obtidos os espectros de reflectância para cada amostra, realizando-se duas leituras por amostra, com um giro de 180° entre as leituras, onde em cada leitura, foi calculada a média a partir de 100 leituras em 10 segundos.

Para remover a variabilidade física devido à dispersão da luz, enfatizar as características de interesse ao longo do espectro e avaliar o efeito de diferentes técnicas de pré-processamento espectral, foi aplicado nos espectros brutos (RAW), seis técnicas de pré-processamento comumente empregadas em espectroscopia (RINNAN; BERG; ENGELSEN, 2009): (i) suavização utilizando uma janela móvel de 11 nm, após teste inicial para definir a melhor janela de busca (SMO); (ii) remoção do contínuo – *continuum removed reflectance* (CRR); (iii) correção por padrão normal variado – *standard normal variate* (SNV); (iv) correção de dispersão multiplicativa – *multiplicative scatter correction* (MSC); (v) *Savitzky-Golay* com 1ª derivada usando um polinômio de primeira ordem (SGD), com janela de busca de 11 nm, após teste inicial para definir a melhor janela de busca; (vi) redução de tendência – *Detrend* (DET); Todos os pré-processamentos foram realizados na linguagem de programação R (R CORE TEAM, 2018), utilizando o *pacote prospectr* (STEVENS; RAMIREZ-LOPEZ, 2013).

6.2.3. Organização da biblioteca espectral – BE

A união dos dados analíticos de COS, classe textural, reflectância espectral e informações do local de amostragem (classe de solo, uso da terra, clima, material parental e região fisiográfica) das amostras dos sete BD resultou na BE (n = 2599).

Foram identificadas 10 amostras da BE consideradas *outliers* pelo cálculo da distância de Mahalanobis. Essas amostras apresentaram dados de reflectância atípicos em relação ao conjunto total e foram removidas, para evitar possível prejuízo as análises estatísticas (TERRA et al., 2015; VISCARRA ROSSEL et al., 2016).

Foram retiradas aleatoriamente 128 amostras da BE para composição de um conjunto externa de validação (CEV) (Figura 30). Esse número corresponde a 5 % de cada BD que deu origem a BE e foi utilizado para garantir que todos os conjuntos tivessem a mesma probabilidade de ser amostrados. O restante das amostras (n = 2461), doravante denominada de biblioteca espectral global (BEG) foi então estratificado de acordo com critérios ambientais: i) fisiográficos; ii) texturais do solo, iii) de uso da terra e puramente estatístico iv) de classes espectrais.

Para avaliar o efeito das características ambientais relacionadas com cada região fisiográfica, a BEG foi estratificada considerando a região fisiográfica de origem das amostras (Figura 22a), resultando em três sub-bibliotecas, a saber: BE da região do Planalto do RS, n = 1095 (BER1); BE da região da Depressão Central do RS, n = 546 (BER2); BE da região do Planalto de Lages de SC, n = 820 (BER3) (Figura 30).

Para avaliar o efeito da textura do solo, a BEG foi estratificada com base na classe textural. Foram utilizadas as classes: textura arenosa (teor de areia menos teor de argila > 70%), argilosa (argila 35 - 60%), muito argilosa (argila > 60%), média (< 35% de argila e > 15% areia) e siltosa (< 35% argila e < 15% areia), conforme Santos et al. (2018). Cada classe de textura deu origem a uma sub-biblioteca: BE de textura arenosa, n = 148 (BETAR); BE de textura média, n = 343 (BETM); BE de textura argilosa, n = 1000 (BETA); BE de textura muito argilosa, n = 951 (BETMA) (Figura 30). Para a classe de textura siltosa, foram observadas apenas 19 amostras que, portanto foram desconsideradas.

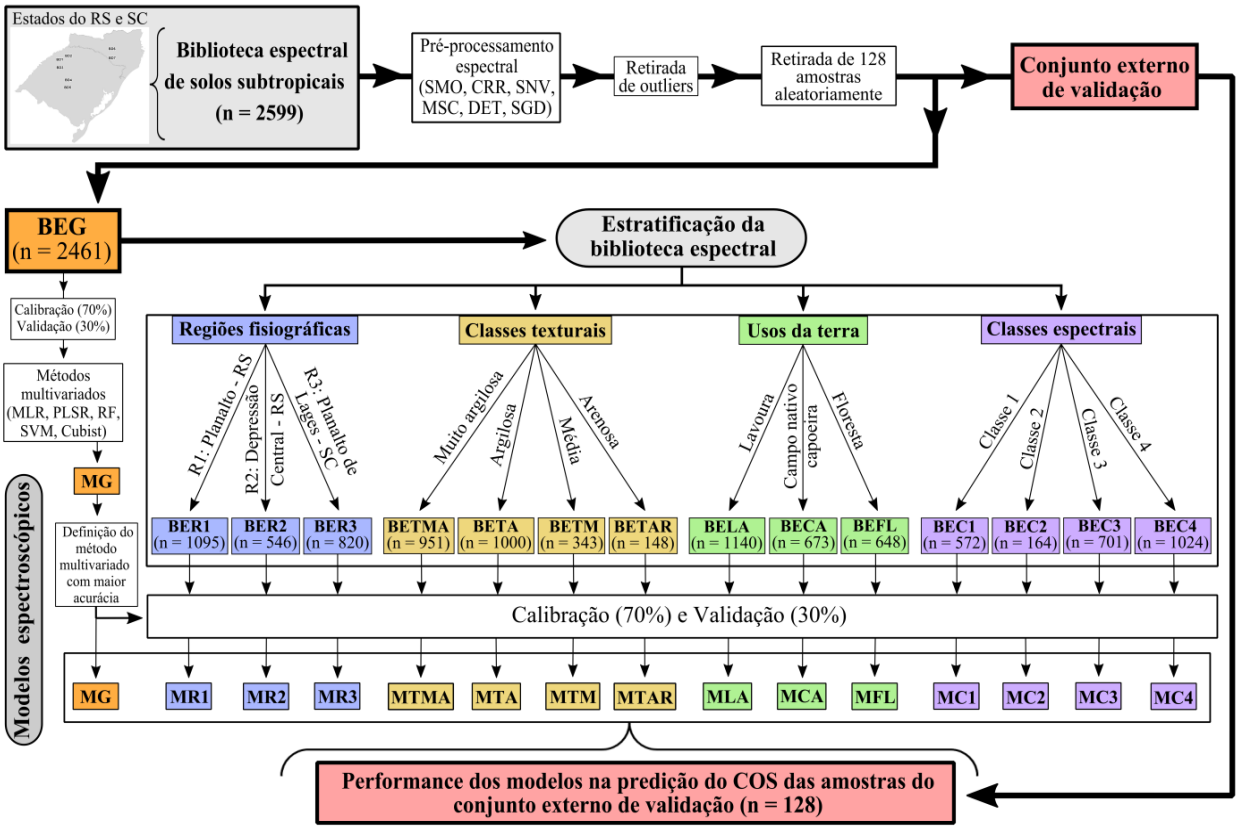
Para avaliar o efeito do uso da terra, a BEG foi estratificada de acordo com o uso da terra de origem das amostras em sub-bibliotecas: BE lavoura, n = 1140 (BELA), BE floresta, n = 648 (BEFL) e BE campo nativo/capoeira, n = 673 (BECA) (Figura 30).

Para avaliar o efeito da variância espectral foi utilizado o método de classificação pelo algoritmo *fuzzy c-means* (BEZDEK et al., 1984) para estratificar a BEG. Esse algoritmo subdividiu a BEG de acordo com as pontuações da análise de componentes principais, em *c*-classes de modo que a variância dentro da classe fosse minimizada. O número de classes foi definido pela relação entre os índices de coeficiente de partição (Cp) e de entropia de partição (Ep) (BEZDEK et al., 1984), resultando em quatro sub-bibliotecas: BE classe espectral 1 (BEC1), BE classe espectral 2 (BEC2), BE classe espectral 3 (BEC3), e BE classe espectral 4 (BEC4) (Figura 30).

Os dados da BEG e das sub-bibliotecas foram descritos estatisticamente. Os dados de COS não significativos pelos coeficientes de assimetria e teste de normalidade de *Kolmogorov-Smirnov* ($\alpha = 0,05$) foram transformados por Box-Cox (BOX; COX, 1964) e a transformação de volta foi realizada pelo inverso da função antes da validação dos modelos. Para avaliar a distribuição granulométrica das amostras que compõem a BEG e as sub-bibliotecas, foi gerado o triângulo textural utilizando o pacote *soiltexture* no software R (R CORE TEAM, 2018).

A avaliação qualitativa dos espectros do solo foi realiza para os dados da BEG biblioteca e das sub-bibliotecas, conforme sugerido por Demattê et al. (2014). Para explorar a variância espectral da BEG e das sub-bibliotecas, foi utilizada a análise de componentes principais (ACP). Para compreender a relação entre as projeções latentes dos componentes principais e os dados de COS, argila, silte e areia foram realizadas a análise de correlação de Pearson ($\alpha = 0.05$).

Figura 30 – Diagrama esquemático das etapas executadas na metodologia



Fonte: Autor. SMO: suavização utilizando uma janela móvel de 11 nm; CRR: remoção do contínuo; SNV: padrão normal variado; DET: *detrend*; MSC: *multiplicative scatter correction*; SGD: *Savitzky-Golay* com 1ª derivada usando um polinômio de 1ªordem com janela de busca de 11 nm. PLSR: Regressão por mínimos quadrados parciais; MLR: regressão linear múltipla; SVM: *Support vector machine*; RF: *Random forest*; BER1: biblioteca espectral região 1; BER2: biblioteca espectral região 2; BER3: biblioteca espectral região 3; BETMA: biblioteca espectral textura muito argilosa; BETA: biblioteca espectral textura argilosa; BETM: biblioteca espectral textura média; BETAR: biblioteca espectral textura arenosa; BELA: biblioteca espectral lavoura; BECA: biblioteca espectral campo: BEFL: biblioteca espectral floresta; BEC1: biblioteca espectral classe 1; BEC2: biblioteca espectral classe 2; BEC3: biblioteca espectral classe 3; BEC4: biblioteca espectral classe 4; MG: modelo global; MR1: modelo região 1; MR2: modelo região 2; MR3: modelo região 3; MTMA: modelo textura muito argilosa; MTA: modelo textura argilosa; MTM: modelo textura média; MTAR: modelo textura arenosa; MLA: modelo lavoura; MCA: modelo campo; MFL: modelo floresta; MC1: modelo classe 1; MC2: modelo classe 2; MC3: modelo classe 3MC4: modelo classe 4

6.2.4. Desenvolvimento dos modelos espectroscópicos

A BEG e cada sub-biblioteca foram particionadas aleatoriamente em conjuntos calibração (70% das amostras) e validação (30% das amostras) com variância homogênea conforme teste de *Levene* (α = 0,05). Modelos foram calibrados para os conjuntos de calibração com da BEG, denominado de modelo global (MG); e das sub-bibliotecas originadas das estratificações, denominados de modelo região R1 (MR1), modelo região R2 (MR2) e modelo região R3 (MR3); modelo textura muito argilosa (MTMA), modelo textura argilosa (MTA), modelo textura média (MTM) e modelo textura arenosa (MTAR); classe espectral 1 (MEC1), classe espectral 2 (MEC2), classe espectral 3 (MEC3), e classe espectral 4 (MEC4) (Figura 30).

Cinco métodos de calibração multivariada foram testados: (i) PLSR (pacote *pls* do R (MEVIK et al., 2016)) técnica paramétrica amplamente utilizada na modelagem espectroscópica (Dotto et al., 2018) e tem boa capacidade de estimar o COS a partir de espectros de solos tropicais (Moura-Bueno et al., 2019). No processamento o número de fatores dos mínimos quadrados parciais foi escolhido para minimizar o erro médio quadrático na validação cruzada; (ii) RF (pacote *randomForest* do R (LIAW; WIENER, 2002)) técnica não paramétrica, que consiste de uma combinação de muitas árvores preditoras, 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 (BREIMAN (2001); (iii) MLR (pacote *caret* do R (KUHN, 2016)) técnica paramétrica, que descreve as relações lineares entre duas ou mais variáveis preditoras em relação a uma variável resposta, que utiliza para a seleção das variáveis preditoras a função *stepwise*, onde o modelo é ajustado seguindo o Critério de Informação de Akaike (AIC), em que o método seleciona o melhor preditor para a variável dependente e as demais variáveis preditoras são incluídas no modelo à medida que essas aumentam o poder explicativo do modelo, resultando na maior minimização do valor de AIC; (iv) SVM (pacote *e1071* do R (MEYER et al., 2017)), técnica não paramétrica, baseada na teoria da aprendizagem estatística, em que através das funções do *kernel*, o modelo busca identificar uma função de interpolação, ajustando os dados de treinamento para um tubo com raio ε usando amostras de limites (representando os vetores de suporte) perto da margem definida por ε, em que o melhor modelo é obtido usando uma função de custo com o objetivo de minimizar simultaneamente o tamanho dos coeficientes e os erros de predição, em que conjuntos de dados com padrões não lineares podem ser representados mais

facilmente (IVANCIUC, 2007); (v) *Cubist*, método baseado na construção de árvores de regressão, onde o conjunto da variável resposta é dividido em subconjuntos, nos quais suas características são similares em relação aos seus espectros (QUINLAN, 1992). Primeiramente, é construída uma árvore pela definição de regras, que particiona os dados em grupos razoavelmente homogêneos para a propriedade do solo de interesse em relação aos comprimentos de onda específicos das observações. Em seguida, dentro de cada grupo, a propriedade é predita a partir das absorções nesses comprimentos de onda por regressão de mínimos quadrados ordinários. Nesse processo, são geradas várias regras *if–then–else* com base em vários comprimentos ondas e organizadas de forma hierárquica. É possível que as observações e seus preditores satisfaçam mais de um conjunto de regras, sendo nesse caso, a média das previsões é tomada como a predição geral do modelo. A modelagem foi realizada no software R, pacote *Cubist* (KUHN et al., 2011). Esse modelo tem apresentado bom desempenho para conceber modelos espectroscópicos para BES com alta variação espectral (VISCARRA ROSSEL et al., 2016).

Após a calibração, cada modelo foi validado com o respectivo conjunto de validação. Para avaliar a influência das variáveis explicativas nas predições, foi calculada a importância de cada banda na predição (VARIMP), pelo pacote *caret* do software R (KUHN et al., 2016).

O desempenho das predições foi avaliado usando: coeficiente de determinação (R^2) (Equação 11), erro médio (EM) (Equação12), raiz quadrada média do erro de predição (RMSE) (Equação 13), desvio padrão do erro (SDE) (Equação 14) e proporção de desempenho no intervalo interquartil (RPIQ) (Equação 15).

$$\mathrm{R}^2 = \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}}(\hat{\mathrm{y}}_\mathrm{i} - \bar{\mathrm{y}}_\mathrm{i})^2}{\sum_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{y}_\mathrm{i} - \bar{\mathrm{y}}_\mathrm{i})^2} \tag{11}$$

$$\mathrm{EM} = \frac{1}{N}\sum_{i=1}^{N}(\hat{\mathrm{y}}_\mathrm{i} - \bar{\mathrm{y}}_\mathrm{i}) \tag{12}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}(\hat{\mathrm{y}}_\mathrm{i} - \mathrm{y}_\mathrm{i})^2} \tag{13}$$

$$\mathrm{SDE} = \sqrt{\sum_{i=1}^{N} \frac{(\hat{y}_i - \bar{y}_i)^2}{(N-1)}} \tag{14}$$

$$\mathrm{RPIQ} = \frac{(\mathrm{Q3} - \mathrm{Q1})}{\mathrm{RMSE}} \tag{15}$$

Em que: $\hat{y}$ = valor predito; $\bar{y}$ = valor médio observado; y = valores observados; N = número de amostras com *i* = 1, 2,…, n; Q1 = 25% das amostras; Q3 = 75% das amostras; Q3 – Q1 = distância interquartil, representa o intervalo responsável por 50% da população em torno da mediana.

6.2.5. Desempenho dos modelos na predição do COS do conjunto de validação externa

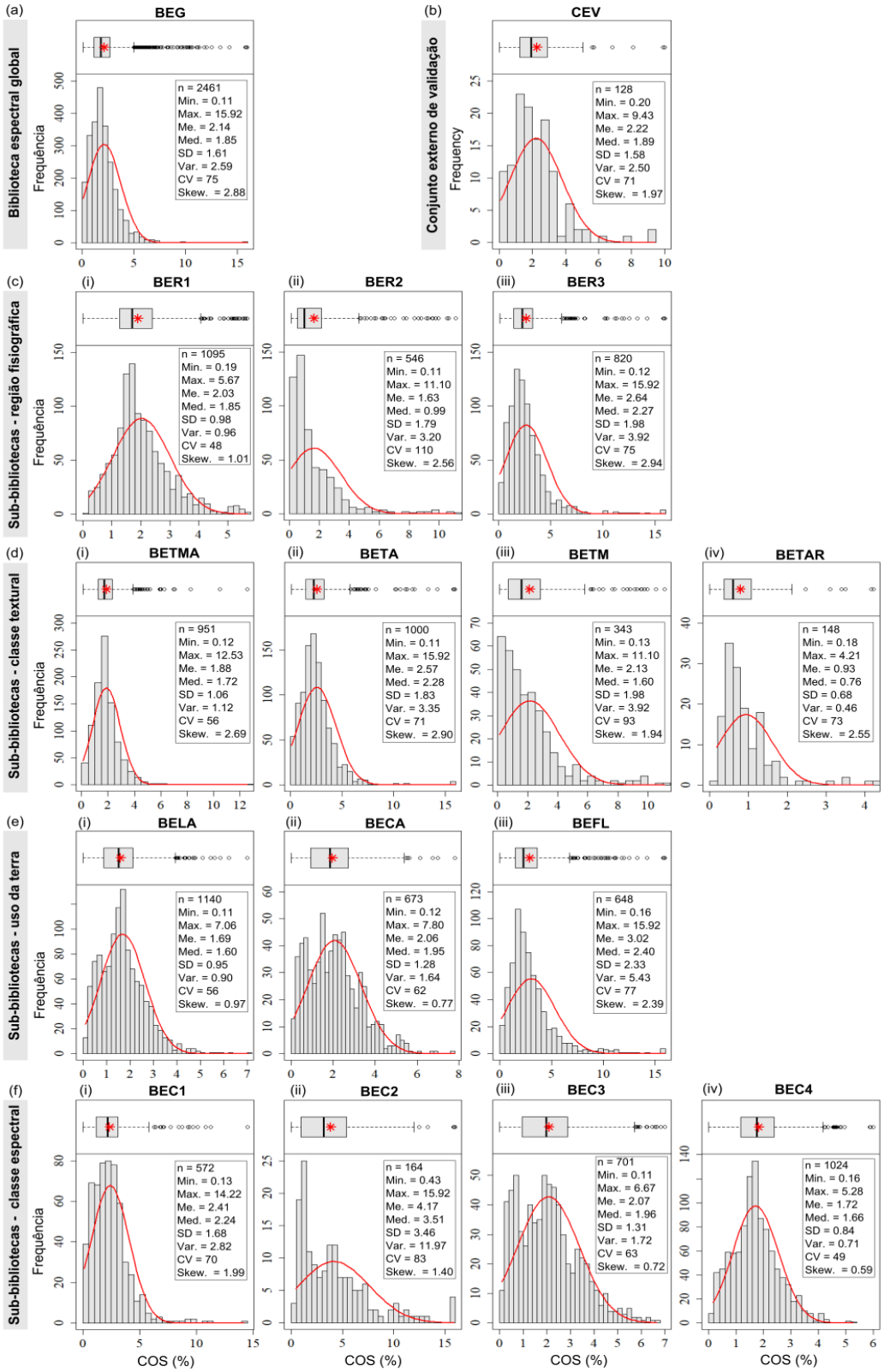
Para entender o comportamento de cada modelo gerado e identificar as circunstâncias em que cada critério de estratificação é eficaz, os modelos foram validados com o CEV (n = 128). Além disso, o CEV foi estratificado baseado nos mesmos critérios da BEG com o objetivo de verificar a aplicabilidade de cada modelo a dados externos com diferentes características ambientais. O desempenho das predições foi avaliado usando: EM, RMSE e SDE.

6.3. RESULTADOS E DISCUSSÃO

6.3.1. Caracterização do solo e análise estatística das bibliotecas espectrais

O conteúdo médio de COS da BEG é 2,14 % com amplitude de 0,11 a 15,92 % (Figura 3a). A maior parte das amostras estão em um intervalo de frequência entre os valores de 1 e 4 % de COS (Figura 31a), valores comumente encontrados nos solos subtropicais do sul do Brasil (BODDEY et al., 2010). Os valores DP, Var e CV do conteúdo de COS são elevados, com 75 % de desvio em torno da média. Solos orgânicos, com teor de COS acima de 8 %, são de baixa ocorrência na região sul do Brasil (DALMOLIN et al., 2017; STRECK et a., 2018), em função disso, há baixa frequência de amostras nesse intervalo.

Figura 31 – Estatística descritiva, histograma e boxplots dos teores de COS da BEG (a), conjunto externo de validação – CEV (b) e sub-bibliotecas por regiões fisiográficas (c), classes texturais (d), usos da terra (e), classes espectrais (f)



Fonte: Autor. BER1: biblioteca espectral região 1; BER2: biblioteca espectral região 2; BER3: biblioteca espectral região 3; BETMA: biblioteca espectral textura muito argilosa; BETA: biblioteca espectral textura argilosa; BETM: biblioteca espectral textura média; BETAR: biblioteca espectral textura arenosa; BELA: biblioteca espectral lavoura; BECA: biblioteca espectral campo: BEFL: biblioteca espectral floresta; BEC1: biblioteca espectral classe 1; BEC2: biblioteca espectral classe 2; BEC3: biblioteca espectral classe 3; BEC4: biblioteca espectral classe 4; n: número de amostra, Me: média, SD: desvio-padrão, Var: variância, Min: mínimo, Med: mediana, Max: máximo, CV: coeficiente de variação, Skew: assimetria.

O conteúdo de COS do CEV apresenta teor médio de 2,22 %, DP de 1,58 %, Var de 2,50 e CV de 73 % (Figura 31b), próximos à medida de tendência central e de dispersão observada na BEG.

A estratificação da BEG em sub-bibliotecas nem sempre resultou em redução da variância do conteúdo de COS. Em todos os critérios de estratificação houve pelo menos uma sub-biblioteca cuja variância aumentou consideravelmente. Na estratificação por região, a variância do COS reduziu apenas na sub-biblioteca BER1, passando de 2,59 na BEG para 0,96 (Figura 31c-i). Isso se deve a redução na amplitude do COS da BER1, relacionada principalmente ao uso da terra na região, onde predomina o uso lavoura e os teores de COS dificilmente ultrapassam 5 %, resultando ainda em um baixo CV (48 %). Além disso, os solos da região apresentam baixa variação na composição granulométrica por serem formados a partir rochas ígneas básicas (basalto) da Formação Serra Geral (Figura 22c). Na região R1 predomina a classe dos Latossolos com elevado teor de argila (Figura 22a, e, f) e mineralogia composta por caulinita e óxidos de ferro hematita (DALMOLIN et al., 2017). A BER3 guarda semelhanças em relação ao material parental da BER1, sendo rochas ígneas básicas e ácidas (riolito-riodacito e andesito-basalto) da Formação Serra Geral (Figura 22c). Em contrapartida, existe uma variação maior de grupos de solos, com predomínio de Latossolos, Nitossolos, Cambissolos e Neossolos com mineralogia predominante na fração argila composta por caulinita, óxidos de ferro (hematita e goethita) e argilominerais com hidróxi-Al entre camadas (KER; RESENDE, 1990; ALMEIDA et al., 2000). Somado a isso, predomina o clima mais frio e úmido (Figura 22b) e uso da terra predominante de florestas (Figura 22d). Essa variação nas características ambientais resultou na maior amplitude de valores de COS e maior variância (Figura 3c-iii), tanto em relação a BER1 e BER2 (Figura 31c-i, 31c-ii), quanto em relação à BEG (Figura 3a).

Apesar do conteúdo médio de COS da BER2 ser baixo (Figura 31c-ii) em relação as demais sub-bibliotecas regionais, a variância do COS aumentou, passando de 2,59 na BEG para 3,20 na BER2 (Figura 31c-ii). Resultado da elevada amplitude do conteúdo de COS (0,11 a 11,10 %) e baixa frequência de valores acima de 4 %, o CV foi alto (110 %). Essa dispersão nos teores

de COS é relacionada com o uso da terra na região que, quando não é floresta, normalmente é campo nativo degradado e lavoura com baixo teor de COS (SAMUEL-ROSA et al., 2011; MOURA-BUENO et al., 2018). Além disso, na região R2 predomina as classes de solos Argissolos e Planossolos (Figura 22e), originados de material parental sedimentar de arenito, arenito-argilito e argilito-siltito (SARTORI, 2009) (Figura 22c). A mineralogia da fração argila desses solos também é mais variada em relação as regiões R1 e R3, sendo composta por caulinita, óxidos de ferro e 2:1 com hidróxi-Al entre camadas (BORTOLUZZI et al., 2005; SCHMITT, 2015). Esses solos apresentam predominantemente, horizontes superficiais com textura média ou mais grossa e subsuperficiais de textura média e argilosa (STRECK et al., 2018), refletindo em solos com alta variação na composição granulométrica (Figura 22e, f), com efeitos diretos no potencial de armazenamento de COS (SIX et al., 2000; KLEBER et al., 2007).

Em relação ao conteúdo de COS entre as bibliotecas agrupadas por classe textural (Figura 31d), a BETA apresentou o maior teor médio (2,57 %), seguido das BETM (2,13 %), BETM (1,88 %) e BETAR (0,93 %). Isso porque solos argilosos tendem a armazenar mais carbono em função da maior área superficial específica das partículas e maior quantidade de grupos funcionais manifestados em superfície, o que possibilita a formação de complexos organominerais, conferindo estabilidade ao COS (KLEBER et al., 2007). Esse fenômeno proporciona um efeito de proteção físicoquímica do COS condicionado pela agregação das partículas e menor taxa de mineralização devido a inacessibilidade aos microorganismos (SIX et al., 2000). Assim, solos argilosos conseguem armazenem maior teor de COS. No entanto, esta condição é dependente da quantidade de COS no sistema, com influência direta do uso da terra, sistema de manejo e clima. Isso explica porque o valor médio de COS nas BETMA (Figura 31d-i) é menor em relação a BETA e BETM (Figuras 31d-ii, 31d-iii), uma vez que, a maior parte das amostras da BETMA provém da região R1, onde uso da terra predominante é lavoura.

Na estratificação por classe textural houve redução da variância do conteúdo de COS nas sub-bibliotecas BETMA e BETAR, passando de 2,59 na BEG para 1,12 e 0,46, respectivamente (Figura 31d). Entretanto, o CV do COS na BETAR continuou muito alto com 73 % de desvio em torno da média. Isso se deve ao número baixo de amostras e maior frequência de valores expressivos acima da média em relação ao BETMA. A redução do CV no conteúdo de COS observado na BETMA pode ser consequência das amostras dessa classe textural de amostras com textura muita argilosa serem derivadas da BER1 e BER3 onde a variabilidade ambiental é menor

em relação à região R2, com destaque para o fator material parental e uso da terra. O aumento da variância de COS nas sub-bibliotecas BETA e BETM (Figura 31d) pode estar associado à origem das amostras, principalmente as da BETM, grande parte delas provindas da região R2, onde a variabilidade ambiental é maior (Figura 22). Esses resultados sugerem que em condições de alta variabilidade ambiental, a estratificação por granulometria nem sempre reduz a variância do conteúdo de COS.

A estratificação da BEG pelo uso da terra diminuiu a variância do conteúdo de COS nas sub-bibliotecas BEFA e BECNS, passando de 2,59 na BEG para 0,90 e 1,64, respectivamente (Figuras 31e). A redução na amplitude do COS (0,11 e 7,06 %) e um baixo CV (56 %) na BEFA, se deve características típicas de solos subtropicais sob o uso agrícola na região sul do Brasil.

O expressivo aumento da variância na BEFO em relação à BEG é consequência da presença de valores acima de 5 % de COS que contribuíram para dispersão dos dados em relação à média amostral, CV (77 %). Além de ser característico de solos florestais da região sul do Brasil, o teor elevado de COS tem relação com variação da composição granulométrica observada nessa biblioteca espectral (Figura 32c). Os valores acima de 8 % de COS (Figura 31e-iii) são atribuídos a solos de áreas de florestas de pinus com mais de 30 anos de implantação na região R3 e a permanência do COS no sistema devido ao clima mais úmido e frio (Figura 22b), associado à formação de complexos organominerais devido à textura argilosa (Figura 32c).

O critério de estratificação baseado na reflectância espectral reduziu a variância do COS nas BEC1, BEC3 e BEC4, sendo a maior redução observada na BEC4 (Figura 31f-iv). Na BEC2 houve aumento significativo da variância (11,97) e CV (83 %), sendo isso consequência da alta amplitude do COS (0,43 a 15,92 %) e redução significativa do número de amostras (Figura 31f-ii). Somado a isso, as amostras que compõem a BEC2 são derivadas das três regiões fisiográficas, impondo assim maior variação nas características do solo, com destaque para a composição granulométrica (Figura 32d), devido a variabilidade do material parental entre as regiões.

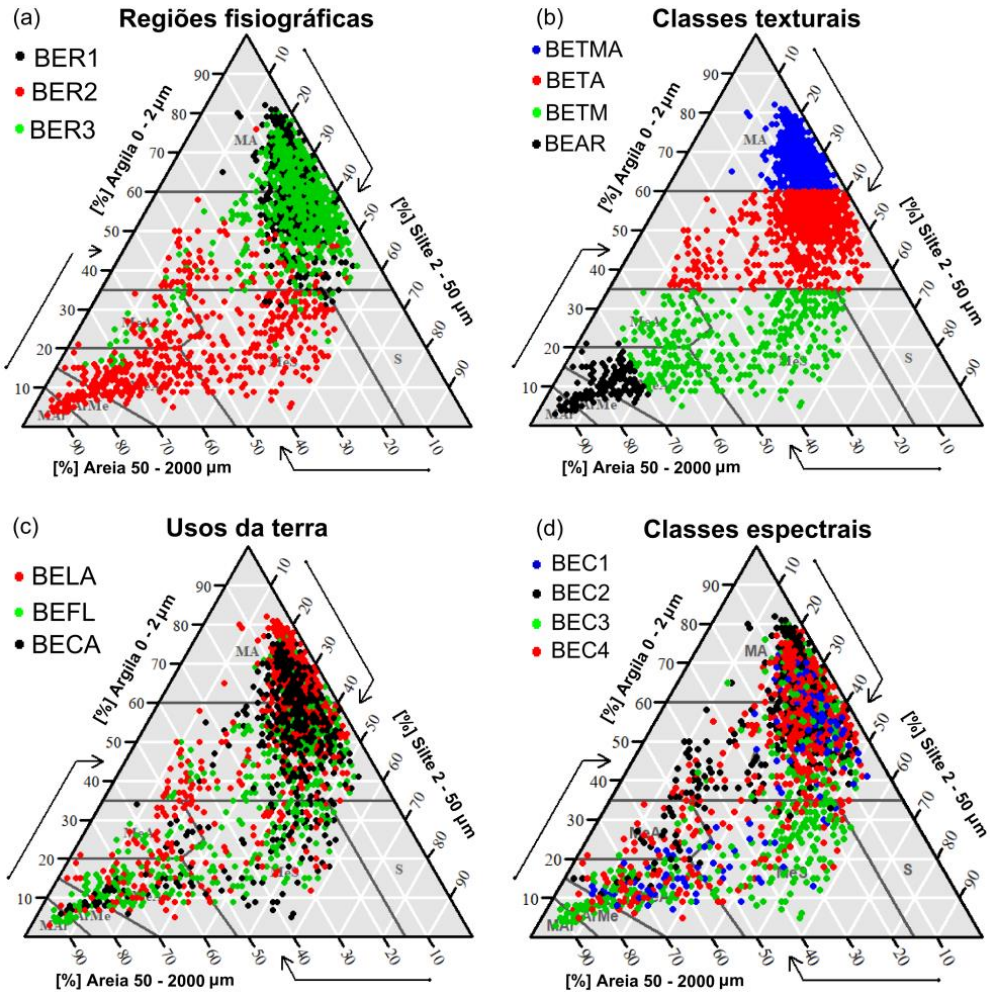
A redução da variância observada na BEG de 2,59 para 0,71 na BEC4 é resultado da menor amplitude do conteúdo de COS (0,16 a 5,28 %) e distribuição de frequência tendendo a normal, resultado em um baixo CV (49 %) (Figura 31f-iv). Além disso, a maior parte dos solos da BEC4 tem como origem as regiões R1 e R3, sendo isso a causa da menor variação na composição granulométrica da BEC4 (Figura 32d). Ainda, os grupos de solos e sua composição mineralógica para ambas a regiões são semelhantes, resultado em uma reflectância espectral

peculiar. As sub-bibliotecas BEC1 e BEC3 (Figuras 31f-i, 31f-iii, respetivamente) apresentaram redução da variância do COS mais tímida em relação à BEG. Na BEC1, tal fato esta relacionado com a menor frequência de valores de COS acima de 5 %, resultando em uma distribuição de frequência assimétrica, que elevou a variância e o CV (Figura 31f-i). Já na BEC3, a explicação é o maior número de amostras com textura média ou mais grossa (Figura 32d), provenientes da região R2, onde a variação de uso terra é maior. Isso resultou na maior dispersão dos teores de COS, sendo confirmado pelo maior intervalo interquartil (Figura 31f-ii).

A dispersão do conteúdo de COS e composição granulométrica entre as bibliotecas espectrais é consequência da interação dos diferentes fatores de formação do solo atuantes nos locais de origem das amostras. Essa variação é esperada em bibliotecas espectrais compostas por solos de diferentes regiões, principalmente quando existe uma variabilidade dos fatores de formação (BELLINASO et al., 2010; STEVENS et al., 2013; SATO, 2015; VISCARRA-ROSSEL et al., 2016), como é o caso da região sul do Brasil.

Da mesma forma, é de conhecimento que essa variação vai refletir na variância espectral do solo (STENBERG et al., 2002; ARAÚJO et al., 2014; VISCARRA ROSSEL et al., 2016; WIGHT et al., 2016; WIJEWARDANE et al., 2016). Porém, pouco se sabe sobre o desempenho de modelos de predição para COS calibrados a partir de BE de solos subtropicais com essas características. Além disso, nesse cenário, a estratégia de estratificação da BE e os critérios mais adequados visando o aumento da acurácia das predições ainda é uma lacuna científica.

Figura 32 – Triângulo textural com a identificação das amostras das bibliotecas espectrais agrupadas por região fisiográficas (a), classe textural (b), uso da terra (c) e classe espectral (d)



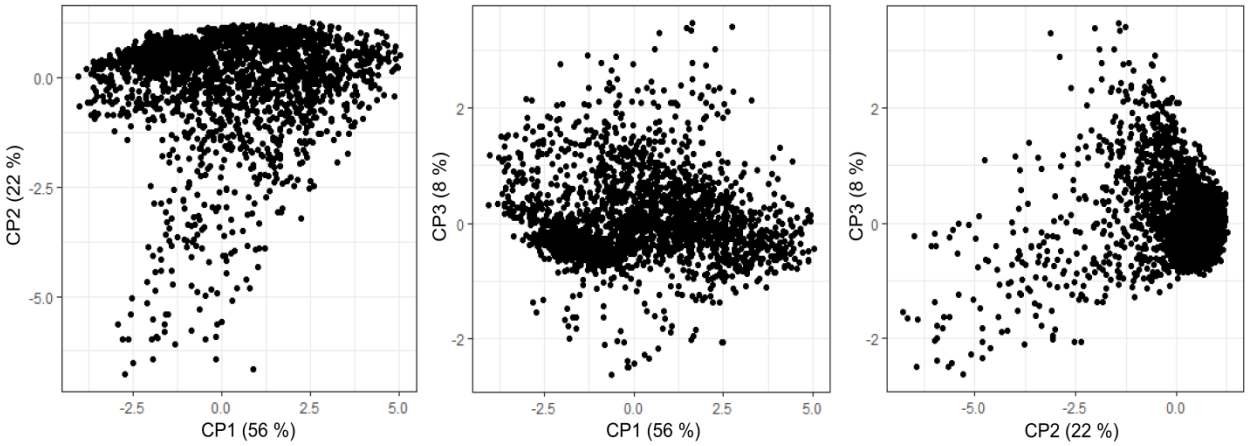
Fonte: Autor. BER1: biblioteca espectral região 1; BER2: biblioteca espectral região 2; BER3: biblioteca espectral região 3; BETMA: biblioteca espectral textura muito argilosa; BETA: biblioteca espectral textura argilosa; BETM: biblioteca espectral textura média; BETAR: biblioteca espectral textura arenosa; BELA: biblioteca espectral lavoura; BECA: biblioteca espectral campo: BEFL: biblioteca espectral floresta; BEC1: biblioteca espectral classe 1; BEC2: biblioteca espectral classe 2; BEC3: biblioteca espectral classe 3; BEC4: biblioteca espectral classe 4.

6.3.2. Análise exploratória quantitativa dos espectros

A análise das variáveis latentes da ACP mostra que as componentes principais CP1+ CP2+CP3 resumem 86 % da variância espectral das amostras da BEG (Figura 33). Os componentes restantes representam apenas pequenas proporções da variância dos espectros e assim eles não foram usados nas análises posteriores. A ACP mostra que maior parte da variância

se concentra no CP1 (56 %), seguida da CP2 (22%) e CP3 (8 %), com maior dispersão observada no quartil inferior da projeção CP1-CP2 (Figura 33).

Figura 33 – *Scores* do CP1, CP2 e CP3 derivados da ACP dos dados espectrais da BEG



Fonte: Autor.

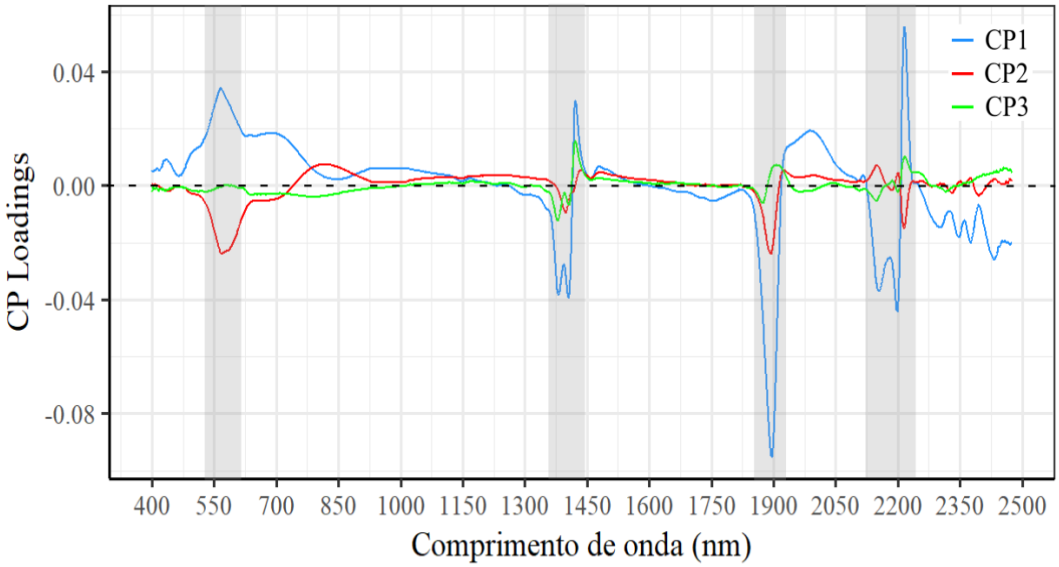
A relação entre a composição organomineral do solo e a variação espectral foi confirmada pela análise de correlação entre os *scores* das CP e os teores de COS, argila, silte e areia (Tabela 6). A CP1 apresentou uma alta correlação inversa com o os teores de COS (r = -0,72) e teores de areia (r = -0,49). A CP2 teve uma alta correlação direta com os teores de argila (r = 0,57). As correlações altas observadas entre os componentes minerais (fração argila) e orgânicos (matéria orgânica) indicam que esses constituintes do solo são componentes espectralmente ativos do solo, tendo influência direta nos espectros da BEG. Isso se confirma ao analisar os *loadings* derivados da ACP, os quais representam as regiões do espectro em que ocorrem variações causadas pela composição organomineral do solo (Figura 34). As bandas com os maiores picos são observadas nas feições relacionadas com variações nos teores de óxidos de ferro, argila e tipo de argilomineral (BEN-DOR; HELLER; CHUDNOVSKY, 2008).

Tabela 6 – Matriz correlação entre os *scores* da ACP e o conteúdo de COS, argila, silte e areia da BEG (n = 2461)

	COS	Argila	Silte	Areia	CP1	CP2	CP3
COS	1						
Argila	0,51	1					
Silte	0,39	-0,07	1				
Areia	-0,20	-0,88	-0,44	1			
CP1	**-0,72**	0,46	-0,14	**-0,49**	1		
CP2	0,38	**0,57**	-0,22	-0,46	0,62	1	
CP3	-0,34	0,24	-0,14	-0,14	-0,43	0,16	1

Fonte: Autor.

Figura 34 – *Loadings* derivados da ACP os quais representam as principais variações espectrais causadas pelas variações na composição do solo nas diferentes bandas



Fonte: Autor.

A estratificação da BEG em sub-bibliotecas com base na reflectância espectral a partir do algoritmo *fuzzy c-means* resultou em quatro sub-bibliotecas. Foram definidas quatro classes devido o coeficiente de partição (Cp) ser maximizado e a entropia de partição (Ep), ser minimizada nessa partição (Tabela 7), que foi então tomada para estratificar os dados da BEG.

Tabela 7 – Número de classes e índices de coeficiente de partição e entropia de partição

Número de classes	Coeficiente de partição (Cp)	Entropia de partição (Ep)
2	0,78	0,36
3	0,65	0,62
4	0,62	0,72
5	**0,56**	**0,88**
6	0,51	1,01
7	0,50	1,05
8	0,49	1,06

Fonte: Autor.

Os autovetores dos três primeiros componentes principais são apresentados na figura 35a, onde se observa que o primeiro componente CP1 é dominado por cargas positivas altas ao redor de comprimentos de onda característicos de absorções dos óxidos de ferro hematita e goethita (480-680 nm), comuns nos solos do sul do Brasil (DALMOLIN et al., 2005). Portanto, o CP1 descreve variações relacionadas aos óxidos de ferro das amostras da BEG dos solos da região sul do Brasil. O CP2 tem cargas positivas altas em trono dos comprimentos de onda relacionados aos grupos funcionais da matéria orgânica (800-900 e 1850 nm) (VISCARRA ROSSEL; BEHRENS, 2010). Já o autovetor do CP3 tem grandes cargas negativas próximas a 400 nm, que é atribuído a óxidos de ferro e também nas regiões de 1400 e 1900 nm, associadas ligações Al-OH da caulinita (VISCARRA ROSSEL; BEHRENS, 2010). Comportamento semelhante foi observado por Viscarra Rossel et al. (2016) ao classificar a biblioteca espectral global com base nos espectros e também em estudo de Araújo et al. (2014) para uma biblioteca de solos tropicais do Brasil.

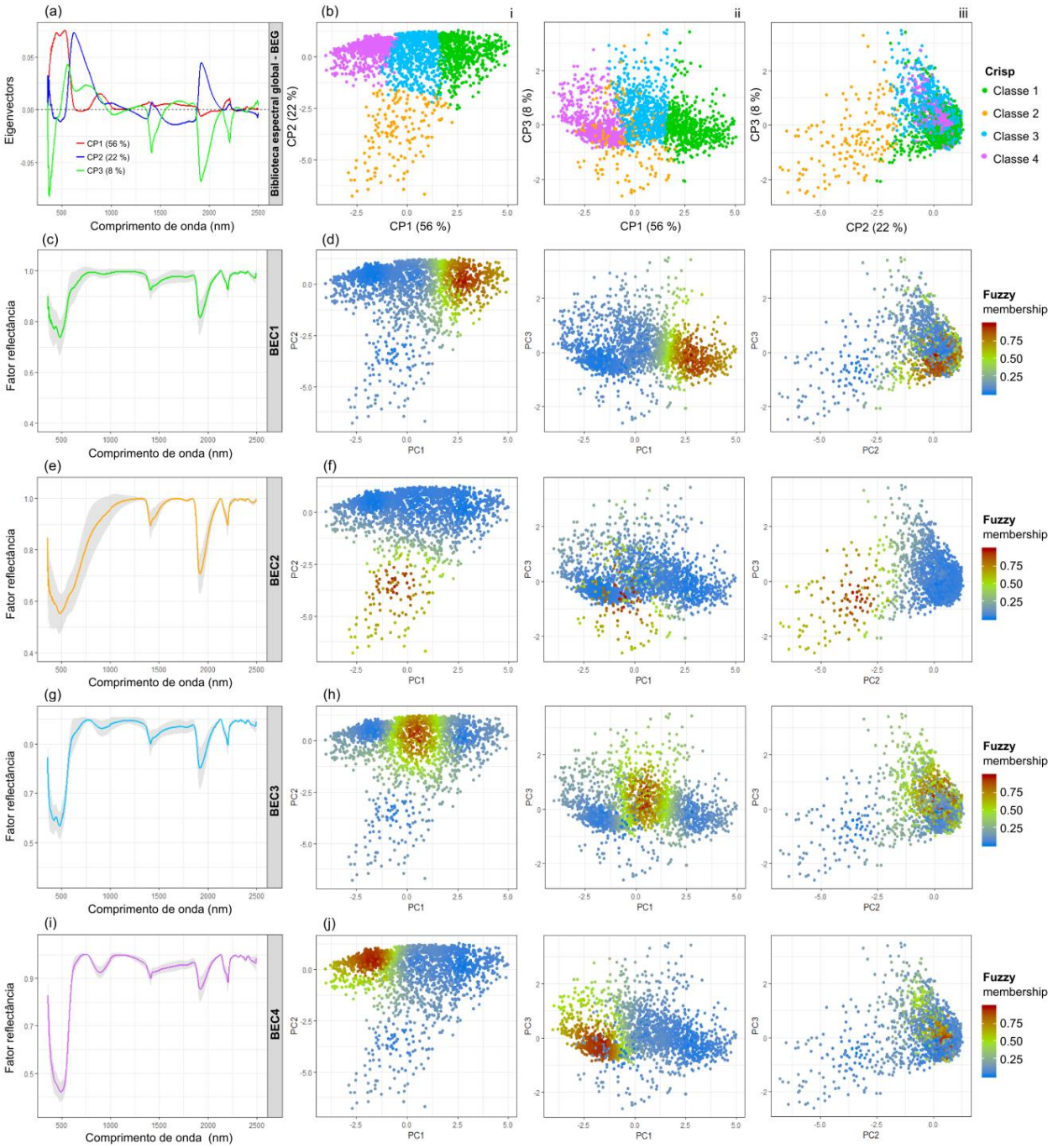
Os diagramas de dispersão das pontuações dos três primeiros componentes e a identificação das quatro classes obtidas na classificação *fuzzy-c-means* são apresentados na figura 35b-i-ii-iii. Nota-se que os espectros da BEC1 se concentraram no quadrante superior direito (Figura 35d-i-ii-iii) e o espectro médio apresenta variação espectral alto ao longo de todo espectro. O espectro é caracterizado por absorções que descrevem solos com quantidade moderada de óxidos de ferro, mas com alta variação espectral nessas feições e também nas feições de absorção dos grupos OH e Al-OH dos argilominerais em torno de 1400 e 1900 nm (Figura 35c). As amostras que compõem essa biblioteca espectral apresentam, em geral, textura média e argilosa, associadas ao uso da terra floresta e campo nativo e são provindas da região R2,

onde os solos apresentam alta variação na mineralógica (BORTOLUZZI et al., 2005; SCHMITT, 2015).

A BEC2 apresenta suas amostras dispersas ao longo do quadrante inferior esquerdo (Figura 35f-i-ii-iii) mostrando alta variação espectral, assim como observado no espectro médio (Figura 35e). A fonte de variância se deve ao fato das amostras que compõem essa classe espectral serem derivadas da região R2, com textura média e arenosa, e região R3, com textura argila. Além disso, nas duas regiões as amostras estão associadas aos usos da terra FL e CA, em que apresentam altos teores de COS, o qual refletiu na variação observada na região de 400 a 1000 nm (Figura 35e). Esse comportamento está relacionado à presença de complexos organominerais entre os grupos funcionais da matéria orgânica e óxidos de ferro (VISCARRA ROSSEL; BEHRENS, 2010; VISCARRA ROSSEL; HICKS, 2015).

A BEC3 teve as amostras concentradas no quadrante superior, próximas do eixo central dos CP (Figura 35h-i-ii-iii) e seu espectro médio indica forte influência no comportamento espectral na região dos óxidos de ferro (500 nm) e grupos funcionais da matéria orgânica em 750 nm (Figura 35g) (VISCARRA ROSSEL; HICKS, 2015). Além disso, a ampla absorção entre 500 nm, que é característico de solos escuros contendo maiores quantidades de matéria orgânica do solo (DALMOLIN et al., 2005). Isso é corroborado devido a grande parte das amostras da BEC3 serem derivadas da região R3, sob o uso da terra floresta, onde se observa valores elevados de COS. Contudo, os espectros da BEC3 são caracterizados por apresentar variação moderada comparada às demais bibliotecas. A BEC4 apresenta as amostras concentradas no quadrante superior esquerdo (Figura 35j-i-ii-iii) e seu espectro médio indica forte absorção na região dos óxidos de ferro e caulinita (Figura 35i). Em geral, a BEC4 mostra comportamento similar a BEC3, mas suas absorções são mais intensas em torno de 500 nm e a variação espectral da BEC4 é a menor observada. Os solos que compõem esse conjunto de dados são derivados da região R1, com predomínio de textura muito argilosa e argilosa, e uso da terra lavoura. Além disso, a composição mineralógica é mais homogênea, caracterizando espectros com menor variação. Por fim, as associações *fuzzy* (*memberships fuzzy*) dos espectros para cada classe espectral mostram as transições e a sobreposição entre as classes, demostrando a natureza contínua, complexa e diversificada da informação contidas nos espectros dos solos da região sul do Brasil. Isso também foi observado na caracterização da biblioteca espectral global por Viscarra Rossel et al. (2016) e por Demattê et al. (2016) para 7185 espectros das regiões sudeste, centro oeste e norte do Brasil.

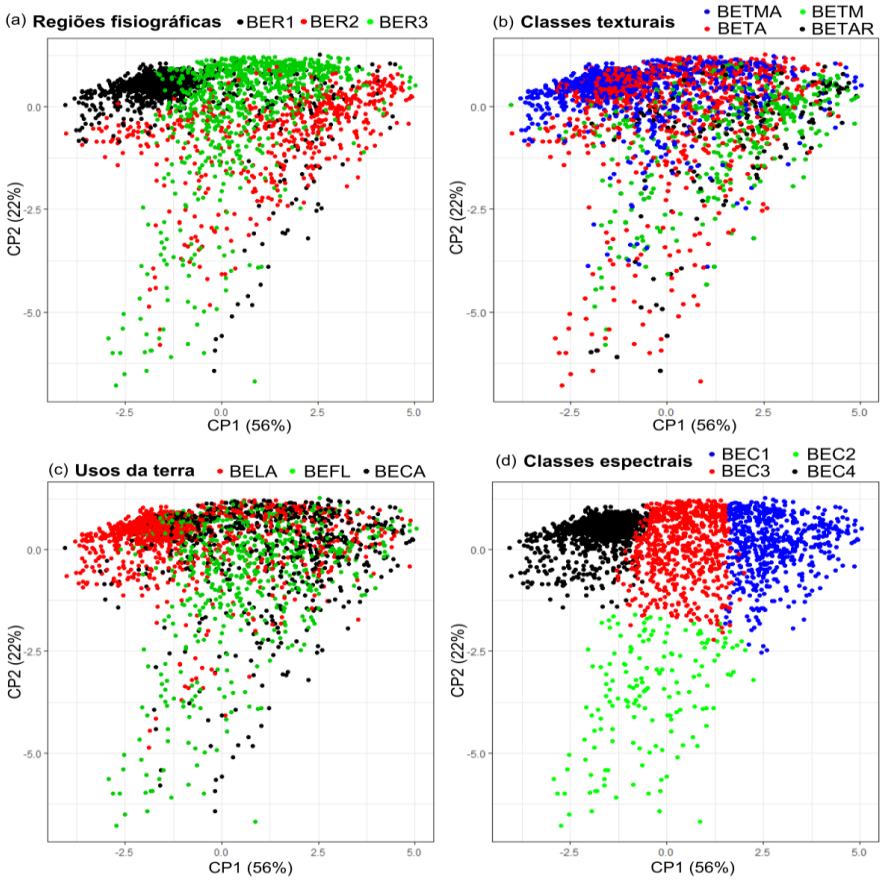
Figura 35 – Análise de componentes principais e classificação de *fuzzy-c-means*: (a) três primeiros componentes (CP) da ACP, (b, i – iii) pontuações dos CP para coloridas para as quatro classes espectrais geradas pela *fuzzy-c-means*, (c, e, g, i) média dos espectros e desvios padrão para as quatro classes espectrais com cores correspondentes às classes mostradas em (b), (d, f, h, j, l, n, i-iii) são as funções de pertinência para cada classe espectral



Fonte: Autor. BEC1: biblioteca espectral classe 1; BEC2: biblioteca espectral classe 2; BEC3: biblioteca espectral classe 3; BEC4: biblioteca espectral classe 4.

Através dos *scores* dos CP é possível identificar relações entre o comportamento espectral do solo das regiões fisiográficas, classes texturais, usos da terra e classes espectrais. Nota-se que a variação espectral é diferente entre as três regiões (Figura 36a), porém com ligeira similaridade entre as regiões R2 e R3, devido os solos dessas regiões apresentarem características pedológicas e usos da terra em comum. A menor variância é observada na região R1 e a maior na região R2, seguida da R3 (Figura 36a). Nas regiões R2 e R3 nota-se aumento significativo dos valores de *scores* em direção ao quartil inferior. Tal fato está relacionado com a variação espectral presente em alguns solos, em geral, com classe textural argilosa, média e arenosa (Figura 36b), e sob usos da terra floresta e campo nativo (Figura 36c).

Figura 36 – *Scores* do CP1 e CP2 derivados da ACP dos dados espectrais com identificação das amostras (a) regiões fisiográficas, (b) classe textural e (c) usos da terra



Fonte: Autor. BER1: biblioteca espectral região 1; BER2: biblioteca espectral região 2; BER3: biblioteca espectral região 3; BETMA: biblioteca espectral textura muito argilosa; BETA: biblioteca espectral textura argilosa; BETM: biblioteca espectral textura média; BETAR: biblioteca espectral textura arenosa; BELA: biblioteca espectral lavoura; BECA: biblioteca espectral campo: BEFL: biblioteca espectral floresta; BEC1: biblioteca espectral classe 1; BEC2: biblioteca espectral classe 2; BEC3: biblioteca espectral classe 3; BEC4: biblioteca espectral classe 4.

Como se observa, a variação no comportamento espectral dos solos da BEG e das sub-bibliotecas está vinculada com a variação nas características químicas, físicas e mineralógicas, ou seja, na composição organomineral dos solos que compõem cada biblioteca espectral. Além disso, a distribuição da variação espectral dos solos agrupados com base nas regiões fisiográficas indica que os espetros regionais guardam relação com a variação ambiental da região sul do Brasil. Comportamento semelhante foi observado por Demattê et al. (2016), que estudaram a regionalização espectral de solos tropicais do Brasil para a predição de propriedades do solo. Os autores observaram que bibliotecas espectrais regionais mostraram maior dispersão em detrimento a bibliotecas locais, sendo a causa dessa variação atribuída a solos formados por diferentes materiais parentais, que refletiram em distintos teores de óxidos de ferro, argila e COS nas amostras de solo. Em escala global, Viscarra-Rossel et al. (2016) observaram comportamento semelhante na caracterização da biblioteca espectral global, em que observaram distinção na variação espectral entre os solos dos sete continentes, entre países, grupos de solos e usos da terra. Os autores relataram que as informações codificadas nos espectros podem ser associadas com a cobertura do solo e sua distribuição geográfica global.

Em função disso, a estratificação de amostras com base nas regiões fisiográficas, classes texturais, usos da terra e classes espectrais pode ser uma estratégia para melhorar a acurácia das predições dos modelos espectroscópicos de COS, uma vez que, leva em consideração os quatro fatores citados anteriormente. Essa hipótese é oportuna e corroborada em estudo de Viscarra Rossel et al. (2016) a partir da biblioteca espectral global.

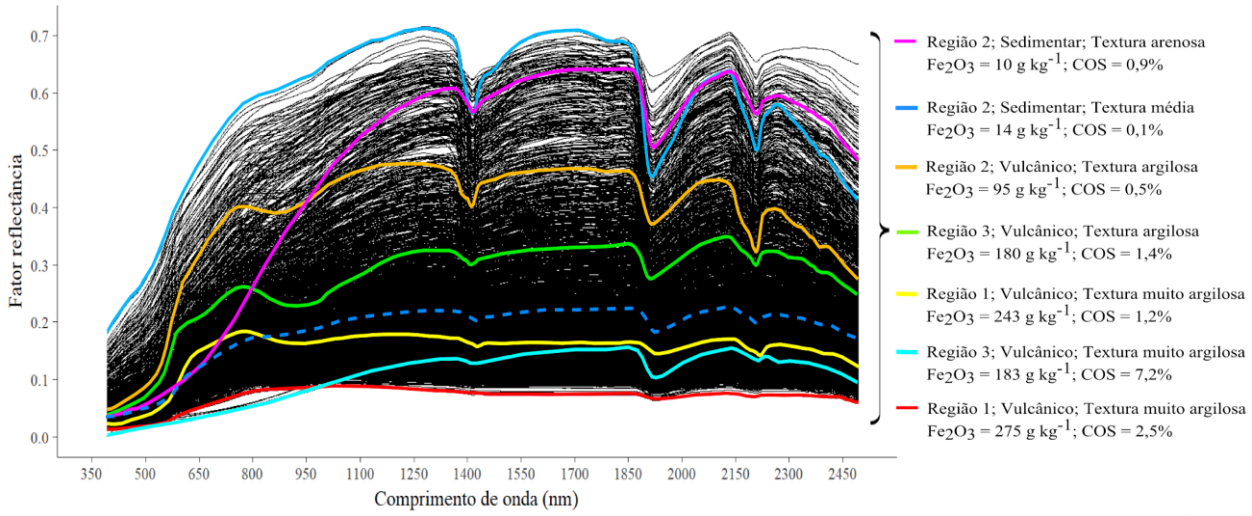
6.3.3. Análise exploratória qualitativa dos espectros

A presença de solos na BEG com diferentes teores de COS, argila, areia e composição mineralógica da fração argila, resulta em inúmeras combinações nas proporções da fase organomineral e suas interações físico-químicas. Todas estas interações se refletiram no comportamento espectral do solo, em que foi observada uma variação grande na intensidade de reflectância das amostras da BEG, conforme mostrado em destaque para algumas amostras da BEG (Figura 37).

Em geral, solos com menor conteúdo de COS apresentaram fortes feições de absorção em 1400, 1900 e 2200 nm (Figura 37). Já amostras com maior conteúdo de COS apresentaram

redução do albedo ao longo de todo o espectro eletromagnético, com atenuação significativa nas principais feições de absorção dos óxidos de ferro e argilominerais (Figura 37) e modificando o formato da curva espectral, deixando-a mais côncava e retilínea. O aumento do teor de argila também teve influência principalmente na redução do albedo ao longo de toda curva (Figura 37). Por fim, a variação no conteúdo de Fe_2O_3 refletiu na mudança nas feições de absorção relacionadas aos óxidos de ferro, conduzindo a um formato mais côncavo das curvas nessas feições à medida que aumenta o teor de Fe_2O_3 e reduzindo o albedo (Figura 37).

Figura 37 – Curvas espectrais da BEG de solos, com destaque para curvas representativas das três regiões com diferentes materiais parentais e composição organomineral



Fonte: Autor.

As curvas inferiores da figura 38 mostram o espectro médio de refletância e seu desvio padrão para os espectros sem pré-processamento e curvas superiores mostram os espectros após a aplicação do pré-processamento espectral remoção do contínuo, onde é possível identificar regiões onde ocorrem absorções relacionadas com a variação da composição organomineral da BEG, conjunto de validação externa e das sub-bibliotecas formadas com base nas regiões fisiográficas, classes texturais, uso da terra e classes espectrais. Todos os espectros têm uma forma geral semelhante, com aumento da reflectância na faixa do visível (400 a 700 nm) em direção a região do infravermelho (700 - 2500 nm), dentro da qual se observa feições de absorção acentuadas na relacionadas com variações nos teores de matéria orgânica, óxidos de ferro, argila

e tipo de argilomineral presente no solo (BEN-DOR et al., 1997; VISCARRA ROSSEL; BEHRENS, 2010). São observadas mudanças na reflectância nas faixas espectrais específicas de absorção de óxidos (bandas 435, 550 e 850 nm), sugerindo à presença de hematita e goethita (DALMOLIN et al., 2005; DEMATTÊ et al., 2014), matéria orgânica, em torno de 800-900 nm, associados ao grupo funcional "*aryl*" derivados de anéis aromáticos (VISCARRA ROSSEL; HICKS, 2015) e caulinita em 1900 nm (VISCARRA ROSSEL; BEHRENS, 2010).

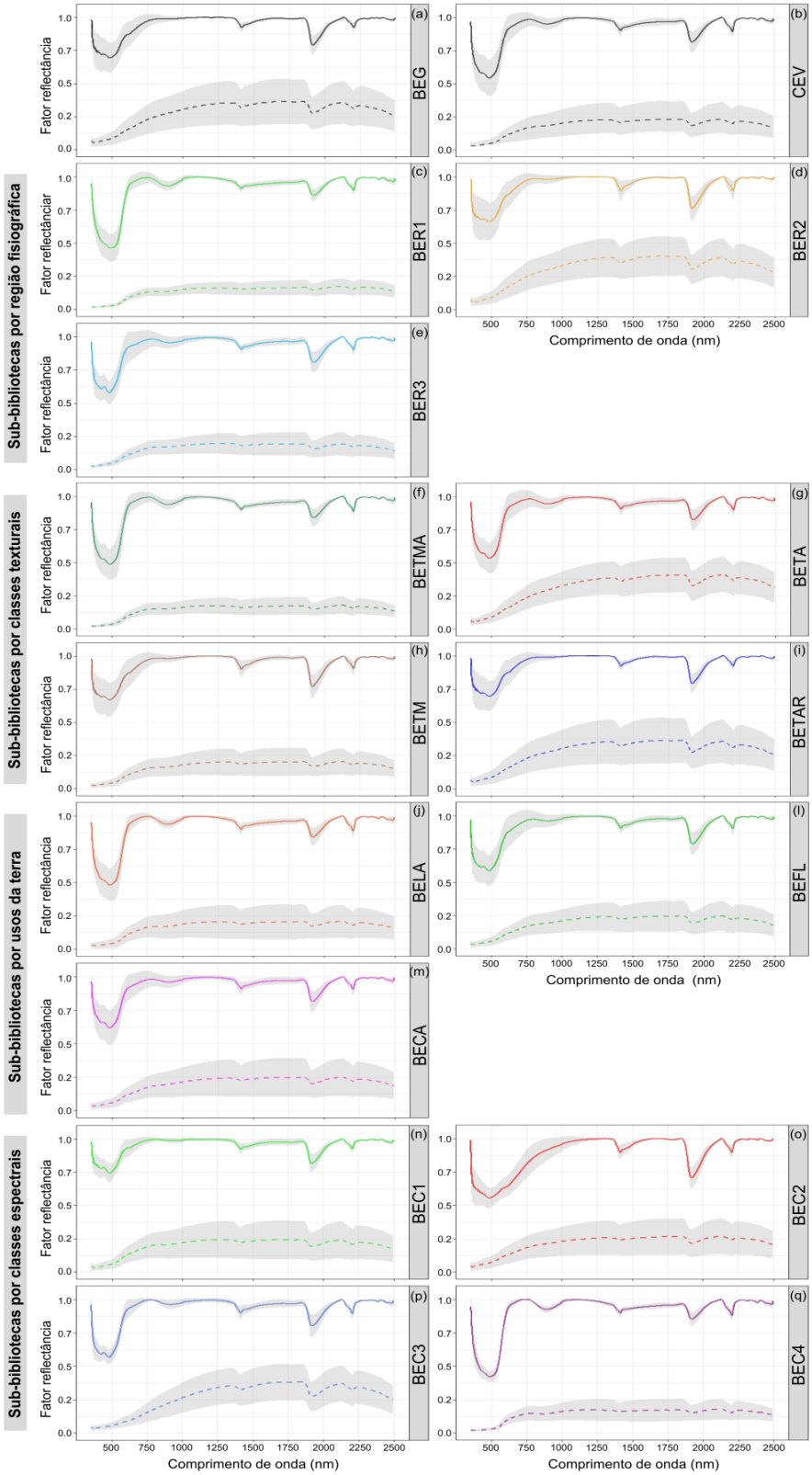
A BEG apresenta grande variação na intensidade de reflectância das amostras (Figura 38a) resultado da variação organomineral. O espectro médio dos solos do CEV (Figura 38b) apresentou regiões de absorção semelhantes, indicando que ambos os conjuntos são homogêneos.

Os espectros com menor intensidade de reflectância e menor variação espectral são observados para solos da BER1 (Figura 38c), BETMA (Figura 38f) e BEC4 (Figura 38q). Tal fato é consequência dos altos teores de argila e óxidos de ferro presentes nos solos dessas sub-bibliotecas, que causam aumento da quantidade de energia absorvida ao longo de todo espectro, refletindo na redução do albedo, característica típica de solos derivados de rochas vulcânicas em solos da região sul do Brasil, principalmente para a classe dos Latossolos (DALMOLIN et al., 2005). Já o menor desvio-padrão da reflectância observado para a BER1 e BETMA é devido a menor variação na composição organomineral dos solos dessas sub-bibliotecas.

Nas sub-bibliotecas BER2, BETAR, BECA e BEC1 são observadas as maiores intensidades de reflectância e variação espectral (Figura 38d, i, m, n, respectivamente), onde a variação é crescente ao longo de todo espectro. Isso está associada com a maior variância do teor de COS e predomínio da fração areia nos solos dessas sub-bibliotecas. Em geral, solos com textura arenosa apresentaram maior intensidade de reflectância em relação aos argilosos, devido ao maior conteúdo de areia, causado pela presença de quartzo nesta fração que reflete grande quantidade de energia (STONER; BAUMGARDNER, 1981; BEN-DOR et al., 1997).

Na BER3 a maior variação espectral é observada na região de absorção de óxidos (bandas 435, 550 e 850 nm) e matéria orgânica em torno de 1800 nm (Figura 38d). Tal comportamento está relacionado aos altos teores de óxidos presentes nos solo da R3, mas também pelo efeito de atenuação nas feições de absorção dos óxidos causado pelo alto teor de matéria orgânica (DALMOLIN et al., 2005; DEMATTÊ et al., 2014) observados nessa região. O aumento do teor de argila também teve influência principalmente na redução do albedo ao longo de todo espectro (Figura 38c) (DALMOLIN et al., 2005).

Figura 38 – Espectros médios e seus desvios-padrão do conjunto global de solos, regiões fisiográficas, classes texturais e usos da terra



Fonte: Autor. BEG: biblioteca espectral global; CEV: conjunto externo de validação; BER1: biblioteca espectral região 1; BER2: biblioteca espectral região 2; BER3: biblioteca espectral região 3; BETMA: biblioteca espectral textura muito argilosa; BETA: biblioteca espectral textura argilosa; BETM: biblioteca espectral textura média; BETAR: biblioteca espectral textura arenosa; BELA: biblioteca espectral lavoura; BECA: biblioteca espectral campo: BEFL: biblioteca espectral floresta; BEC1: biblioteca espectral classe 1; BEC2: biblioteca espectral classe 2; BEC3: biblioteca espectral classe 3; BEC4: biblioteca espectral classe 4.

Em relação ao comportamento espectral das amostras agrupadas com base na classe textural, solos com textura muito argilosa apresentaram a menor reflectância e variação espectral (Figura 38f) que solos com textura arenosa (Figura 38i). Isso esta relacionado ao comportamento espectral intrínseco dessas frações granulométricas do solo (STONER; BAUMGARDNER, 1981; BEN-DOR et al., 1997). Solos com textura argilosa (Figura 38g) e média (Figura 38h) apresentaram variações intermediárias, com ligeiro aumento da variação para os solos de textura argilosa. Tal fato está relacionado com a variação pedológica dos solos agrupados na classe textural argila, os quais são derivados das três regiões fisiográficas.

O comportamento espectral observado para os solos agrupados com base nos usos da terra, em geral, não mostraram diferenças significativas na reflectância e variação espectral (Figura 38j, l, m). Porém, são observadas variações na região de absorção de óxidos (bandas 435, 550 e 850 nm), sendo isso relacionado ao efeito de mascara causado pela matéria orgânica (DALMOLIN et al., 2005; DEMATTÊ et al., 2014). Neste caso, as maiores variações são observadas nos usos florestes (Figura 38l) e campo nativo (Figura 38m), em o conteúdo de matéria orgânica são maiores. Segundo Galvão e Vitorello (1998), teores acima de 1,7% em solos tropicais do Brasil já causam efeito nas feições de absorção dos óxidos de ferro e reduzem o albedo ao longo de todo o espectro eletromagnético. Isso ocorre pela atenuação significativa nas principais feições de absorção dos óxidos de ferro e argilominerais, modificando o formato da curva espectral (DALMOLIN et al., 2005) e deixando-a mais côncava e retilínea (DEMATTÊ et al., 2014).

6.3.4. Desempenho do modelo espectroscópico global - MG

Os resultados de calibração e validação do modelo global (MG) obtido com os cinco métodos multivariados a partir dos espectros sem pré-processamento (RAW) e com os seis processamentos espectrais, são resumidos na tabela 8. Em geral, se observa uma variação na acurácia do mesmo método multivariado, porém com pré-processamentos espectrais diferentes

(Tabela 8). Desempenho semelhante foi observado em trabalho de Vasques et al. (2008), Cambule et al. 2012, Knox et al. (2015) e Dotto et al. (2018).

Tabela 8 – Resultados da predição do COS com o modelo global (MG) gerado com a combinação de diferentes métodos multivariados e técnicas de pré-processamento espectral

Modelo Global	R^2	RMSE(%)	RPIQ	R^2	RMSE(%)	RPIQ
	Conjunto de calibração (n = 1723)			Conjunto de validação (n = 738)		
			MLR			
MLR-SGD	**0,64**	**1,02**	**1,51**	**0,64**	**0,97**	**1,60**
MLR-RAW*	0,61	1,04	1,45	0,57	0,99	1,50
MLR-SNV	0,58	1,06	1,44	0,59	1,00	1,56
MLR-CRR	0,62	1,00	1,61	0,58	1,03	1,32
MLR-DET	0,58	1,06	1,44	0,52	1,08	1,42
MLR-SMO	0,61	0,98	1,54	0,59	1,09	1,43
MLR-MSC	0,58	1,01	1,53	0,59	1,09	1,33
Média	0,60	1,02	1,50	0,58	1,04 b**	1,45
			PLSR			
PLSR-SGD	**0,72**	**0,97**	**1,59**	**0,71**	**0,90**	**1,65**
PLSR-SMO	0,69	1,02	1,52	0,67	0,99	1,50
PLSR-RAW*	0,68	0,92	1,68	0,65	0,93	1,59
PLSR-CRR	0,75	0,81	1,89	0,61	0,99	1,50
PLSR-SNV	0,74	0,94	1,66	0,59	1,05	1,42
PLSR-DET	0,71	1,07	1,39	0,58	1,07	1,39
PLSR-MSC	0,68	0,86	1,75	0,62	1,09	1,41
Média	0,71	0,94	1,64	0,63	1,00 b	1,49
			RF			
RF-SGD	**0,72**	**0,87**	**1,76**	**0,70**	**0,92**	**1,77**
RF-CRR	0,68	0,95	1,60	0,64	0,99	1,58
RF-SMO	0,47	1,15	1,30	0,59	1,10	1,48
RF-MSC	0,50	1,17	1,29	0,57	1,02	1,54
RF-SNV	0,53	1,16	1,34	0,56	0,98	1,50
RF-DET	0,64	0,99	1,50	0,55	1,11	1,48
RF-RAW*	0,52	1,11	1,38	0,49	1,17	1,29
Média	0,58	1,06	1,45	0,59	1,04 b	1,52
			SVM			
SVM-SMO	**0,70**	**0,92**	**1,59**	**0,66**	**0,92**	**1,77**
SVM-RAW*	0,65	0,94	1,58	0,68	0,97	1,66
SVM-CRR	0,86	0,61	2,48	0,61	1,11	1,41
SVM-MSC	0,74	0,84	1,76	0,59	1,13	1,45
SVM-SNV	0,71	0,89	1,72	0,58	1,19	1,42
SVM-DET	0,75	0,85	1,84	0,54	1,07	1,37
SVM-SGD	0,88	0,58	2,69	0,33	1,49	0,99
Média	0,76	0,80	1,95	0,57	1,13 c	1,44
			Cubist			
Cubist-CRR	**0,90**	**0,54**	**2,85**	**0,76**	**0,78**	1,90
Cubist-MSC	0,87	0,61	2,54	0,75	0,80	1,86
Cubist-SMO	0,84	0,68	2,25	0,75	0,81	1,83
Cubist-DET	0,89	0,57	2,70	0,74	0,81	1,82
Cubist-SNV	0,90	0,54	2,84	0,70	0,88	1,76
Cubist-RAW	0,86	0,65	2,35	0,68	0,91	1,69
Cubist-SGD	0,74	0,90	1,70	0,61	1,02	1,51
Média	0,86	0,64	2,46	0,71	0,86 a	1,77

Fonte autor. *RAW: espectros sem pré-processamento (testemunha). SMO: suavização utilizando uma janela móvel de 11 nm; CRR: remoção do contínuo; SNV: padrão normal variado; DET: Detrend; MSC: *multiplicative scatter correction*; SGD: Savitzky-Golay com 1ª derivada usando um polinômio de 1ªordem com janela de busca de 11 nm. PLSR: Regressão por mínimos quadrados parciais; RLM: regressão linear múltipla; SVM: Support vector machine; RF: Random forest; R^2: coeficiente de determinação; RMSE: raiz do erro quadrático médio em %; RPIQ: Proporção do erro de predição no intervalo interquartil;. Destacado em negrito representa o melhor resultado para o modelo em relação a cada pré-processamento. **Valores com a mesma letra não diferem entre si pelo teste de Scott Knott (α = 0,05).

Para os métodos PLSR, RF e MLR, os modelos gerados com os espectros SGD apresentaram maior desempenho (RMSE de 0,90 %; 0,92 % e 0,97 %, respectivamente) em relação aos demais pré-processamentos (Tabela 8). Comportamento semelhante foi observado por Cambule et al. (2012) para uma biblioteca espectral local de Moçambique, África, encontraram melhor performance (R^2 = 0,85 e RMSE = 0,32 %) para modelos PLSR com espectros submetidos a 1ª derivada. Nocita et al. (2014) para uma biblioteca espectral de solos dos países da União Européia, contento 19.969 amostras, alcançaram melhor desempenho para modelo PLSR com o espectros submetidos a 1ª derivada. Resultados apresentados por Vasques et al. (2008), mostraram maior acurácia para o pré-processamento de SNV e SGD (RMSE = 0,18 log [mg kg^{-1}]) com o método PLSR a partir de dados em escala regional na Flórida, EUA.

Em contraste, estudos de Dotto et al. (2017) e Dotto et al. (2018), realizado na região sul do Brasil, mostrou melhor desempenho do modelo PLSR com os espectros submetidos ao processamento CRR (RMSE de 0,32 % e 0,49 %, respectivamente) . Já Pinheiro et al. (2017) testaram 41 pré-processamentos e alcançaram melhor acurácia (R^2 = 0,71 e RMSE = 0,57 %) para a predição do COS para solos da Amazônia Central, Brasil, com os espectros processados por SNV. Knox et al. (2015) também encontraram maiores valores de RMSE com a técnica SNV (0,68 log g kg^{-1} para o modelo RF e 0,51 log g kg^{-1} para PLSR) comparado aos modelos obtidos com os espectros sem pré-processamento (0,41 log g kg^{-1} para RF e 0,42 log g kg^{-1} para PLSR) em uma biblioteca espectral em escala nacional da Flórida, EUA. Por outro lado, no presente estudo os modelos PLSR-SNV e RF-SNV apresentaram baixa acurácia (RMSE = 1,05 % e RSME = 0,98 %, respectivamente) (Tabela 8).

Os modelos obtidos pelos métodos SVM e Cubist com os espectros SGD apresentaram a menor acurácia (RMSE de 1,49 % e 1,02 %, respectivamente), sendo esses valores de RMSE inferiores aos observados para os modelos ajustados com os espectros sem pré-processamento (Tabela 8). Esse comportamento é o inverso do observado no estudo de Dotto et al. (2018), que alcançaram melhor desempenho para o método SVM (RMSE = 0,52 %) ao ajustar os modelos com espectros submetidos ao processamento espectral SGD. Para o presente estudo, tais

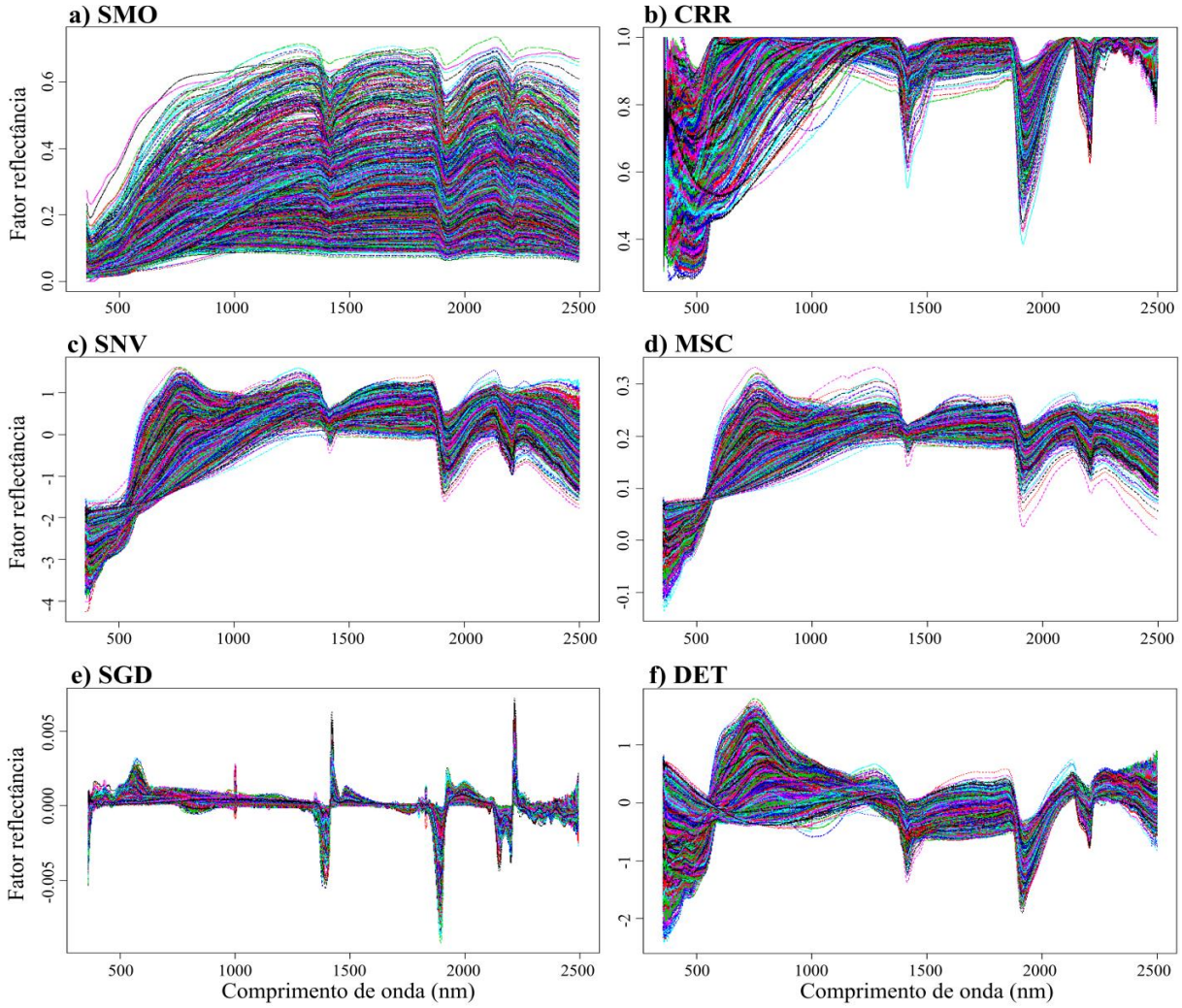
resultados sugerem que o processamento SGD não é adequado para uso com o método SVM e Cubist. Tal fato pode estar relacionado o processo matemático empregado nestes métodos multivariados, sugerindo que esses modelos são mais sensíveis a transformações rigorosas dos espectros, como as realizadas pelas técnicas que empregam as derivadas (SAVITZKY; GOLAY, 1964; RINNAN et al., 2009). Neste caso, o pré-processamento baseado na derivada da reflectância pode estar eliminando feições (variáveis preditoras) importantes para o ajuste destes modelos matemáticas.

A explicação para essa variação são as alterações no valor de reflectância causadas por cada tipo de processamento espectral (Figura 39). O uso destas técnicas tem por objetivo minimizar a influência de efeitos físicos (tamanho de partícula e rugosidade) e variações óticas, os quais introduzem alterações nos espectros (SHEPHERD; WALSH, 2002) e realçam feições de absorção de energia. Na Figura 39 nota-se que algumas feições relacionadas com os constituintes do solo ficaram mais realçadas nas técnicas CRR, SNV, MSC, SGD e DET, com maior destaque para a técnica SGD, comparado ao pré-processamento SMO (Figura 39a).

Estudos mostraram, em geral, uma melhora na acurácia dos modelos preditivos para COS por meio do VIS-NIR utilizando o pré-processamento dos dados espectrais, tais como 1º e 2º derivadas com filtro de Savitzky-Golay (VASQUES et al., 2008; CAMBULE et al., 2013; MOURA-BUENO et al., 2019), padrão variação normal (LUCÀ et al., 2017; PINHEIRO et al., 2017) e remoção do contínuo (DOTTO et al., 2018), enquanto Kooistra et al. (2001) obtiveram melhores resultados com dados espectrais não transformados.

Diante dos resultados observados na literatura e do presente trabalho, fica evidente que existe uma variação nos resultados, independentemente de qual pré-processamento ou método multivariado foi empregado na modelagem. No entanto, nota-se que os métodos multivariados podem apresentar maior similaridade por alguma técnica de pré-processamento específica. Portanto, não há uma maneira priori para determinar as técnicas de pré-processamento que fornecem a melhor precisão de modelos preditivos (BEN-DOR; BANIN, 1995), indicando que é sempre necessário realizar um teste preliminar, principalmente para solos da região sul do Brasil com altos teores de matéria orgânica, argila e óxidos de ferro, devido à forte influência desses componentes no comportamento espectral do solo (DALMOLIN et al., 2005; DEMATTÊ et al., 2014; MOURA-BUENO et al., 2019).

Figura 39 – Espectros da BEG após a aplicação dos seis pré-processamentos. (a) SMO: suavização utilizando uma janela móvel de 11 nm; (b) CRR: remoção do contínuo; (c) SNV: correção pelo padrão normal variado; (d) MSC: *multiplicative scatter correction* - correção de dispersão; (e) SGD: Savitzky-Golay com 1ª derivada usando um polinômio de 1ªordem com janela de busca de 11 nm; (f) DET: Detrend - redução de tendência



Fonte: Autor.

Em relação os métodos multivariados, no presente estudo, o melhor desempenho na predição foi observado para o modelo Cubist-CRR (R^2 = 0,76 e RMSE = 0,78 %), seguido no modelo PLSR-SGD (R^2 = 0,71 e RMSE = 0,90 %) (Tabela 8). Já os modelos MLR, RF e SVM apresentam baixa acurácia, com valores de RMSE iguais estatisticamente. O resultado encontrado para o modelo Cubist-CRR apresentou boa previsão do COS, com baixo erro nas estimativas

(RMSE de 0,78 %) (Tabela 8), considerando o alto intervalo de predição de COS (0,11 – 15,92 %) e a alta variação organomineral e espectral observada na BEG. O modelo subestimou os valores altos e superestimou os valores baixos, os quais são teoricamente de difícil quantificação por modelos (VISCARRA ROSSEL; BEHRENS, 2010) e também por métodos analíticos de laboratório (NELSON; SOMMERS, 1975). O modelo Cubist-CRR foi utilizado para caracterizar a biblioteca espectral global (VISCARRA ROSSEL et al., 2016), onde estimativas para o COS pelo modelo Cubist-CRR apresentaram R^2 = 0,92 e RMSE = 1,11 %. Esse resultado mostrar um R^2 ligeiramente maior para a biblioteca espectral global comparado ao R^2 observado no presente estudo. Por outro lado, o erro de previsão foi menor para a biblioteca espectral regional de solos do sul do Brasil (RMSE = 0,78 %), o que pode ser atribuída a menor variação do COS (0,11 – 15,92 %) dessa biblioteca espectral em relação a biblioteca espectral global (0,01 – 55,70 %).

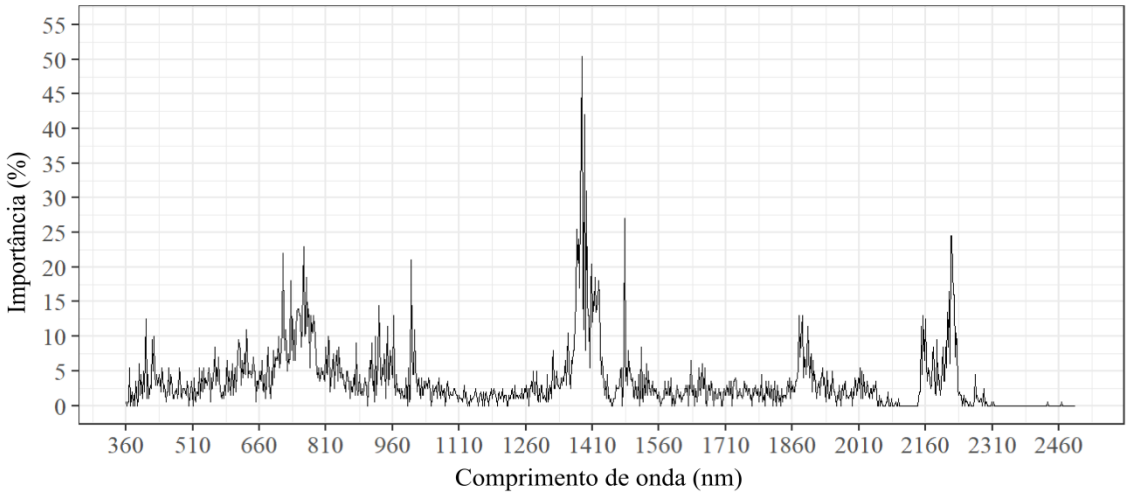
Em relação aos estudos realizados com bibliotecas espectrais em escala regional, melhores resultados do que o presente estudo foram observados por Vasques et al. (2008) (R^2 = 0,86 e RMSE = 0,18 log mg kg^{-1}) para 554 amostras com o modelo PLSR na Flórida, EUA. Na África, Cambule et al. (2012) também obtiveram modelos com bom desempenho, com R^2 de 0.85 e RMSE de 0,32 %. Terra et al. (2015) encontraram valores de R^2= 0,65 e RMSE = 0,16 % utilizando uma biblioteca espectral (n = 1259) com amostras de oito regiões tropicais do Brasil, local em que a variabilidade dos solos é alta, com intervalo de predição de 0,2 – 7,0 % de COS. Pinheiro et al. (2017), utilizando um banco de dados (n = 434) da Amazônia Central, Brasil, encontraram R^2 = 0,71; RMSE = 0,57 % na predição de teor de COS, com intervalor de predição de 0,01 – 10,56 % de COS. Araújo et al. (2014) obtiveram um R^2 de 0.60 e um RMSE de 0,55% para predições com uma grande biblioteca espectral (n = 7172) de solos tropicais no Brasil. Nessa mesma região do Brasil, Ramirez-Lopez et al. (2013) utilizando 4200 amostras, obtiveram R^2 de 0,48 e RMSE de 0,28 %. Em resumo, os autores desses estudos realizados no Brasil verificaram que as características e a variabilidade dos solos foram os fatores que afetaram a previsão do COS. Isso vai de encontro ao observado no presente estudo, onde está presente um conjunto de solos com alta variação nos componentes organominerais, resultado em uma biblioteca espectral heterogênea, que implica em menor acurácia dos modelos espectroscópicos para o COS.

Em relação aos estudos realizados com bibliotecas em escala nacional, estudo realizado por Knox et al. (2015), obtiveram um R^2 de 0,87 e um RMSE de 0,40 usando o pré-processamento de PLSR e SGD, com um conjunto de 1014 amostras coletadas em todo o Estado

da Flórida, EUA. Em contraste, resultados mais pobres foram obtidos em um estudo de Wijewardane et al. (2016), utilizando 19.891 amostras distribuídas nos EUA, encontraram bom desempenho para PLSR (R^2 = 0,83 e RMSE = 7,38 %) e Rede Neural Artificial - RNA (R^2 = 0,96 e RMSE = 3,61 %). Além disso, um estudo realizado na Espanha por Guerrero et al. (2016) demonstraram variações no desempenho dos modelos de PLSR de acordo com o número de amostras. Os autores encontraram a maior precisão no modelo com 3482 amostras (R^2 = 0,91 e RMSE = 9,93 g kg^{-1}), uma precisão média com 1096 amostras (R^2 = 0,82 e RMSE = 25,6 g kg^{-1}) e a menor precisão com 362 amostras (R^2 = 0,80 e RMSE = 25,4 g kg^{-1}). Em geral, se observa uma variação nos valores de RMSE desses estudos em escala regional, nacional e global. Isso se deve à maior escala dessas bibliotecas espectrais, onde a maior variabilidade de fatores ambientais que influenciam as características do solo é observada, especialmente para o conteúdo de COS (NOCITA et al., 2014; RAMIREZ-LOPEZ et al., 2013; DEMATTÊ et al., 2016; GUERRERO et al., 2016; VISCARRA ROSSEL et al., 2016).

As variações nos valores de importância das variáveis preditoras (VARIMP) (Figura 40) destaca que as relações entre o COS e a composição do solo (mineralógica, química e física) na BEG são complexas.

Figura 40 – Importância das variáveis explicativas (VARIMP) na predição do COS para o modelo global Cubist-CRR



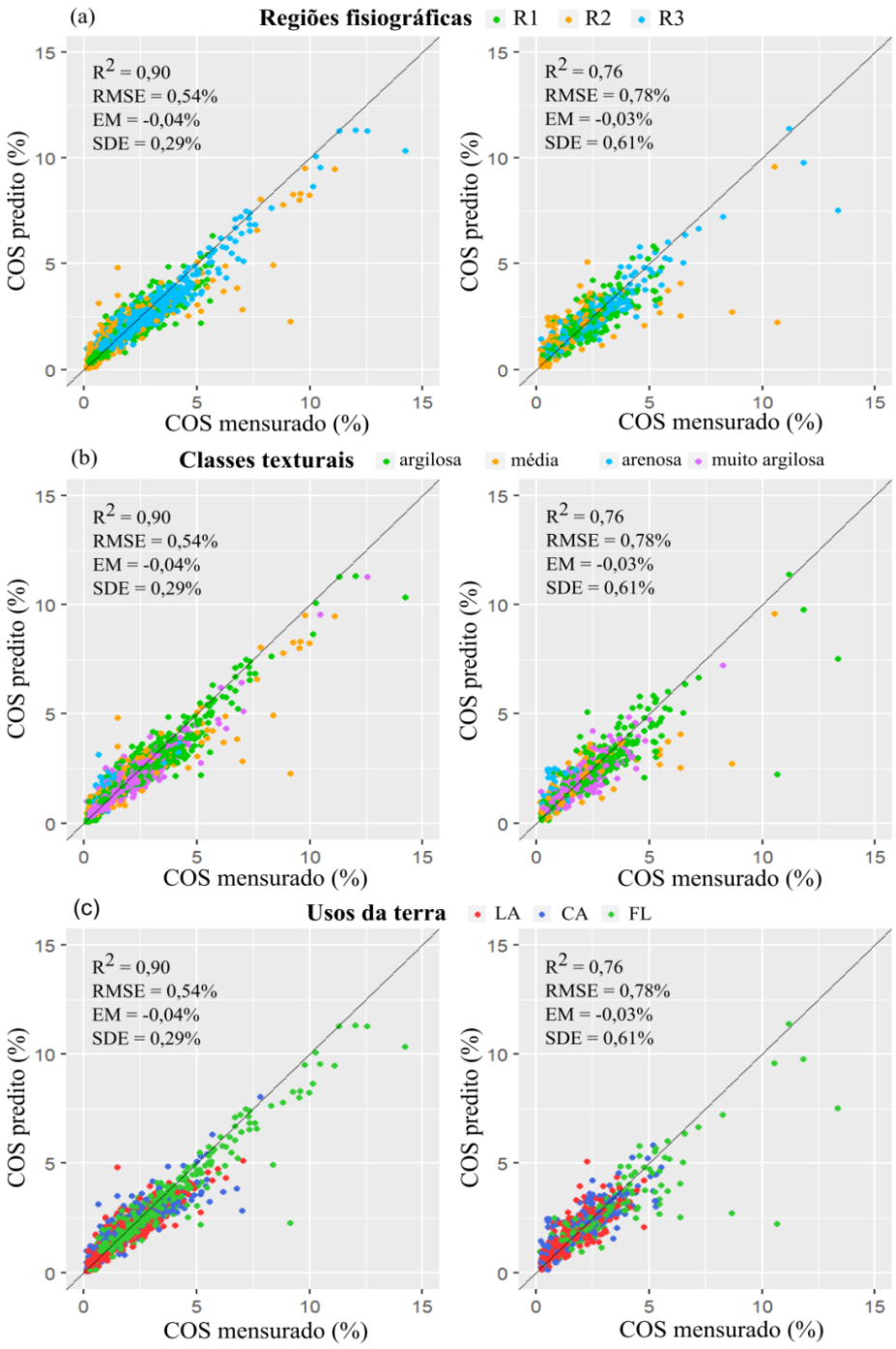
Fonte autor.

O pico VARIMP em torno de 480 e 550 nm está ligado à hematita e os picos por volta de 1370-1400 e 1875-1935 estão relacionados com a influência de grupos OH de minerais e ácidos carboxílicos. Já os picos em 2140-2280 nm têm relação com os grupos Al-OH dos argilominerais e com vibrações moleculares dos grupos funcionais C – H, C – O, C = O, C – N (BEN-DOR et al., 1997; VISCARRA ROSSEL; BEHRENS, 2010). Pinheiro et al. (2017) observaram maior importância da predição nessas mesmas regiões para o COS de solos da Amazônia Central, Brasil. Mesmo comportamento também foi observado por Santana, Souza e Poppi (2018) estudando 3854 espectros derivados de diversas regiões do Brasil.

Diante do melhor desempenho na predição do COS ter sido alcançado para o modelo Cubist-CRR, foram plotadas gráficos de dispersão para visualizar a correlação entre o teor de COS mensurado em laboratório versus o predito pelo modelo Cubist-CRR, com a indicação dos solos de cada região fisiográficas, classes texturais e usos da terra (Figura 41). Observa-se que os maiores erros de predição do COS estão relacionados a solos das regiões R2 e R3 (Figura 41a), em que esses solos apresentam textura argilosa e média (Figura 41b) e estão associados ao uso da terra floresta (Figura 41c). Isso sugere que modelos gerados separadamente para cada região ou uso terra, alcancem maior acurácia nas predições de COS, com destaque para a região R1, onde são observados os menores erros.

Em revisão sobre as perspectivas da espectroscopia, Bellon-Maurel e McBratney (2011) mostraram predições de COS por PLSR com valor de erro médio variando de -0,45 a 1,6 e sua relação com amplitude do intervalo de predição do conteúdo de COS. Da mesma forma, trabalho de Stenberg et al. (2010) menciona a existência de correlação positiva entre o erro das predições e o desvio padrão da propriedade do solo sob investigação. No entanto, a alta variação de características do solo da BEG já discutidas, que mostraram ter influência no comportamento espectral e na variância das variáveis preditoras (bandas espectrais), foram as principais causas pelas quais a acurácia do modelo de predição MG não tenha atingido valor superior ao observado. Em estudos de escala local (DEMATTÊ et al., 2016; MOURA-BUENO et al., 2019), regional (ARAÚJO et al., 2014), nacional (WIJEWARDANE et al., 2016) e global (Stevens et al., 2013; VISCARRA ROSSEL et al., 2016) observaram que a presença de solos com alta variação de seus constituintes prejudicou a predição de propriedades químicas do solo.

Figura 41 – *Scatter plots* da calibração e validação do teor de COS mensurado em laboratório versus o predito pelo modelo global Cubist-CRR com identificação dos solos em (a) regiões fisiográficas, (b) classes texturais (c) usos da terra



Fonte: Autor. LA – lavoura; CA – campo nativo; FL – floresta; R^2: coeficiente de determinação; RMSE: raiz do erro quadrático médio em %; RPIQ: Proporção do erro de predição no intervalo interquartil.

Nesse sentido, estudos realizados em escala regional mostraram que modelos com maior acurácia nas predições podem ser alcançados incluindo informações de bibliotecas espectrais locais, com objetivo de aumentar o número de amostras e reduzir a heterogeneidade do conjunto de dados (GOGÉ et al., 2014). Além disso, considerar as características pedológicas e ambientais durante o processo de modelagem pode contribuir para a obtenção de modelos mais precisos (STEVENS et al., 2013; ARAÚJO et al., 2014; WIGHT; ASHWORTH; ALLEN, 2016; WIJEWARDANE et al., 2016).

Na tentativa de melhorar o desempenho de previsão dos modelos, Araújo et al. (2014) particionaram uma BE em subconjuntos menores, dividindo os dados em grupos mais uniformes em mineralogia, observando-se aumento na acurácia dos modelos, com redução de 15 % no valor de RMSE. Já a influência do conteúdo de COS na acurácia dos modelos foi relatada em estudo de Morgan et al. (2009), onde a presença de material orgânico parcialmente degradado no solo pode aumentar a incerteza das predições. Stevens et al. (2013) estudando uma biblioteca espectral contendo cerca de 20.000 amostras da União Europeia, atribuíram as diferenças na acurácia dos modelos ao uso da terra e variação no conteúdo de argila, areia e COS.

Diante disso, há necessidade de estudos para desenvolver estratégias que possam ser usadas em bibliotecas espectrais regionais heterogêneas de solos subtropicais do sul do Brasil, com o objetivo de aumentar capacidade preditiva dos modelos e buscar respostas para tais variações. Assim, com base nos resultados da análise exploratória dos dados de COS, composição granulométrica e reflectância espectral discutidos anteriormente, nos próximos itens, é apresentado os modelos de predição gerados a partir da estratificação da BEG pelos critérios de regiões fisiográficas, classes texturais, usos da terra e classes espectrais.

6.3.5. Modelos espectroscópicos a partir da estratificação da BEG

Diante do melhor desempenho na predição do COS pelo modelo global ter sido alcançado para o modelo Cubist-CRR, seguido do PLSR-SGD e somado ao bom desempenho desse modelo na predição do COS de solos subtropicais (MOURA-BUENO et al., 2019) e também pelo amplo uso na literatura (DOTTO et al., 2018), nas próximas etapas de modelagem do presente estudo foram testados apenas os modelos PLSR-SGD e Cubist-CRR.

6.3.5.1. Modelos preditivos de base regional (MR)

Os resultados de calibração e validação dos modelos regionais obtidos com os métodos multivariados PLSR e Cubist, a partir dos espectros processamentos com SGD e CRR, são resumidos na tabela 9. No geral, os modelos gerados pelo método Cubist-CRR apresentam maior acurácia (Tabela 9). Para o modelo MR1, o método PLSR com o espectro SGD apresentou desempenho ligeiramente maior em relação ao modelo PLSR-CRR (Tabela 9). Já para o método Cubist, o modelo Cubist-CRR apresentou acurácia significativamente maior comparado ao modelo Cubist-SGD. Esse comportamento em relação às técnicas de pré-processamento e métodos multivariados são similares ao observado na modelagem espectroscópica do MG.

Tabela 9 – Resultados de calibração e validação para predição do COS com os modelos regionais (MR) gerados com diferentes abordagens de métodos

Modelos	R^2	RMSE	ME	SDE	RPIQ	R^2	RMSE	ME	SDE	RPIQ
MG*	0,90	0,54	-0,04	0,29	2,85	0,76	0,78	-0,03	0,61	1,90
MR1	Calibração (n = 767)					Validação (n =328)				
PLSR-SGD	0,74	0,50	0,01	0,25	2,13	0,69	0,55	-0,06	0,30	2,02
PLSR-CRR	0,74	0,50	0,01	0,26	2,10	0,68	0,56	-0,05	0,31	1,98
Cubist-SGD	0,73	0,55	-0,11	0,29	1,92	0,66	0,62	-0,17	0,36	1,77
Cubist-CRR	**0,85**	**0,38**	**-0,02**	**0,14**	**2,70**	**0,74**	**0,50**	**-0,08**	**0,26**	**2,28**
MR2	Calibração (n = 382)					Validação (n =164)				
PLSR-SGD	0,68	1,06	-0,19	1,09	1,52	0,47	1,35	-0,35	1,72	1,21
PLSR-CRR	0,65	1,06	-0,01	1,13	1,52	0,46	1,31	-0,16	1,69	1,25
Cubist-SGD	0,87	0,73	-0,13	0,52	2,21	0,53	1,29	-0,29	1,60	1,26
Cubist -CRR	**0,94**	**0,52**	**-0,07**	**0,27**	**3,08**	**0,62**	**1,15**	**-0,09**	**1,32**	**1,49**
MR3	Calibração (n = 574)					Validação (n = 246)				
PLSR-SGD	0,95	0,48	0,01	0,24	3,80	0,80	0,81	0,03	0,66	2,14
PLSR-CRR	0,91	0,65	0,01	0,42	2,97	0,79	0,72	-0,02	0,52	2,35
Cubist-SGD	0,92	0,71	-0,09	0,50	2,59	0,84	0,73	-0,06	0,53	2,40
Cubist-CRR	**0,97**	**0,37**	**-0,02**	**0,13**	**5,26**	**0,85**	**0,63**	**-0,07**	**0,40**	**2,68**

Fonte: Autor. MR1: modelo região R1; MR2: modelo região R2; MR3: modelo região R3; R^2: coeficiente de determinação; RMSE: raiz do erro quadrático médio em %; ME: erro médio; SDE: desvio padrão do erro; RPIQ: Proporção do erro de predição no intervalo interquartil; *Modelo global (MG) gerado pelo método Cubist-CRR com todas as amostras da biblioteca espectral global (BEG).

Em geral, o modelo Cubist-CRR apresentou R^2 de 0,74, sendo esse valor inferior ao observado para o modelo MG. Porém, os valores de RMSE (0,50 %) e SDE (0,26 %) do modelo MR1 reduziram significativamente (Tabela 9). Esse resultado está relacionado com a redução no intervalo de predição do teor de COS, mas principalmente pela redução da variação no teor e composição dos constituintes do solo (matéria orgânica, argila e mineralogia) (ARAÚJO et al.,

2014; NOCITA et al., 2014; WIJEWARDANE et al., 2016). Além disso, o uso da terra, na grande maioria das amostras da região R1, é lavoura (68 %). Portanto, a união de amostras com essas características resultou na menor variação das interações organomineral e, consequentemente, menor variação espectral do conjunto de dados. Nessas condições, a calibração e validação do MR1 apresentaram maior acurácia em detrimento do MG, confirmando que o agrupamento de amostras em bibliotecas espectrais mais homogêneas é uma estratégia para melhorar a acurácia das predições de COS, pois possibilita a identificação de preditores mais correlacionados com o COS pelos modelos espectroscópicos.

Os modelos gerados para a região MR2, em geral, apresentaram a menor acurácia em relação aos modelos MR1 e MR3 e também ao MG (Tabela 9). Esse resultado tem relação com as particularidades dos solos dessa região, sendo esses formados sob influência de material parental sedimentar (arenitos, argilitos e siltitos), associado à geomorfologia complexa, em que a perda, transporte e deposição de material na paisagem conferem ao solo características alta variação nos componentes organominerais. As amostras desse conjunto de dados apresentam como principais características o baixo teor de óxidos de ferro (~ 50 g kg^{-1}) e alta variação no teor de argila (3 – 76 %, média = 22 %) e areia (5 – 93 %, média = 50 %), devido a amostragem em diferentes profundidades. Esses fatores resultaram em alta variância na variável explicativa, conforme já discutido, que por sua vez, refletem na dificuldade do modelo estabelecer relações entre a reflectância e teor de COS. Além disso, o menor número de amostras (n = 546) também é um fator a ser considerado nesse conjunto de dados, uma vez que, pequenos conjuntos de amostras na modelagem espectroscópica podem levar à captura insuficiente de variações nos dados, refletindo em maiores erros de predição (RAMIREZ-LOPEZ et al., 2014). Os autores também observaram que os erros de previsão eram maiores para conjuntos de dados em escala regional do que em escala local, devido à maior complexidade dos conjuntos de dados em escala regional em comparação com os dados em escala local. Isso é corroborado pelo estudo de Lucà et al. (2017), que observaram diminuição nos valores de R^2 e RMSE com menor número de amostras e concluíram que a capacidade preditiva dos modelos pode dependem não apenas do tamanho da amostra, mas também da localização e variabilidade dos solos da área de estudo. Portanto, as características do conjunto de amostras do MR2 explicam porque a estratificação da BEG não aumentou a acurácia das predições em relação ao modelo MG.

Com relação ao pré-processamento, a técnica de CRR mostrou melhor desempenho para os métodos PLSR e Cubist para o MR2. Esse comportamento foi contrário ao observado para os modelos MG e MR1. Isso confirma os resultados observados para o modelo MG, que não há como determinar a técnica de pré-processamento que fornece a melhor precisão na modelagem espectroscópica (BEN-DOR; BANIN, 1995). Da mesma maneira, cada pré-processamento vai ter comportamento distinto (RINNAN et al., 2009), sendo esse condicionado a variação espectral do conjunto de dados.

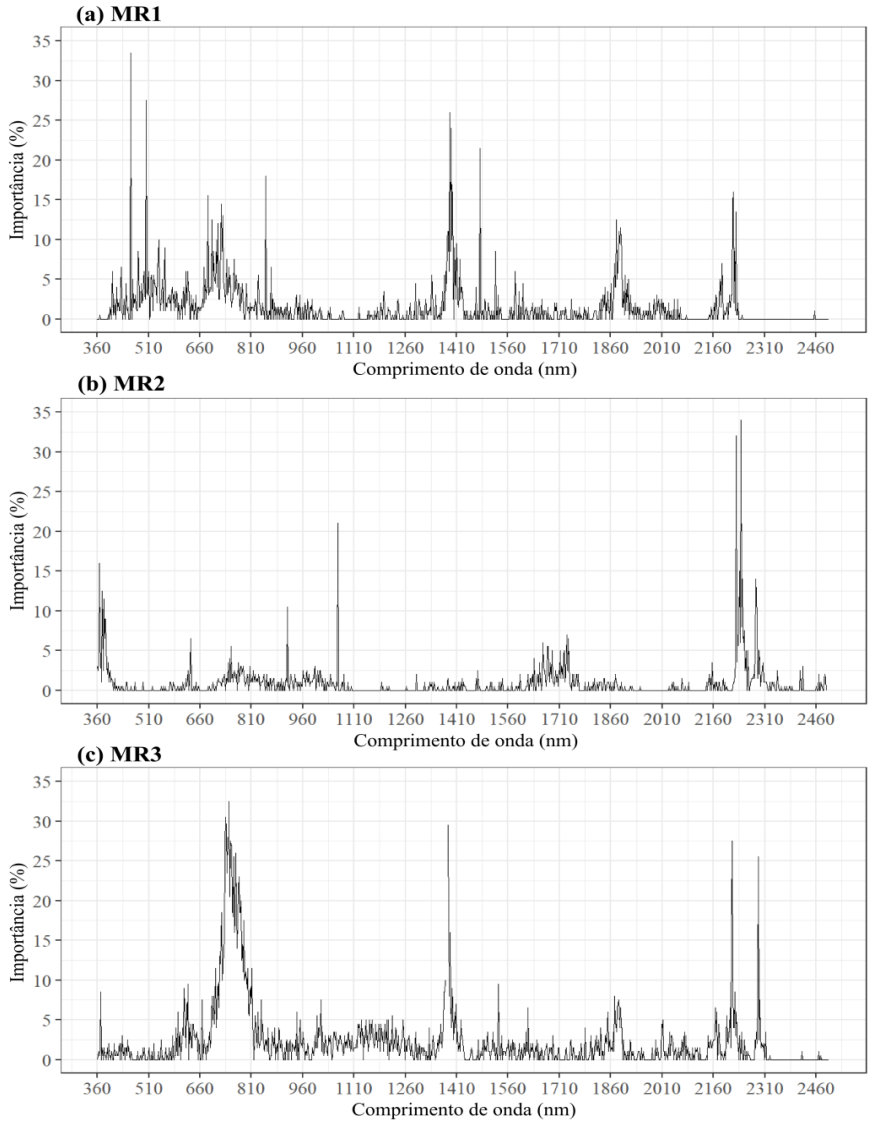
A acurácia do modelo MR3 aumentou de forma significativa (Tabela 9) em relação ao MG e MR2, mas com RMSE ligeiramente maior comparado ao MR1. O melhor desempenho observado para o MR3 tem relação com a maior homogeneidade das características mineralógicas do solo e baixa variação de argila (média = 57 % e desvio-padrão = 12 %), que tem efeito na variação espectral. Associado a isso, o alto teor de COS acumulado no solo pela floresta de pinus durante 30 anos e sua composição também são fatores a serem considerados, já que a degradação das acículas de pinus disponibilizam para o sistema solo, compostos orgânicos altamente lignificados, implicando em uma mineralização lenta e da permanência de grande parte dos compostos orgânicos no sistema, relacionando-se com a matriz mineral do solo.

Essas interações provavelmente estão relacionadas ao compartilhamento de elétrons do COS com óxidos de ferro e argilominerais. O alto teor de compostos orgânicos no sistema faz com que os grupos funcionais e a superfície das partículas sejam saturadas e envolvidas por moléculas orgânicas. Esse recobrimento mineral passa então a influenciar no comportamento espectral. Além disso, em função da matriz mineral estar saturada, compostos orgânicos permanecem sorvidos sem ligação química com os colóides (KLEBER; JOHNSON, 2010), assim, o comportamento espectral do solo passa a ter correlação maior com o teor de COS em outras regiões do espectro. Tal fato foi observado pelo aumento do valor do VARIMP na região em torno de 700-800 nm e 2300 nm (Figura 42c). Portanto, estratificar a BEG em amostras com essas características apresentou vantagem devido ao aumento na acurácia das predições.

As variações nos valores VARIMP entre os modelos regionais destaca que as relações entre a composição organomineral e a reflectância do solo para cada conjunto são distintas (Figura 42), refletindo na seleção de diferentes variáveis espectrais na modelagem espectroscópica. Para o modelo MR1 (Figura 42a) os maiores picos das VARIMP são observadas nas regiões relacionadas aos óxidos de ferro (480 a 680 nm) e grupos funcionais da matéria

orgânica (700 a 800 nm) (VISCARRA ROSSEL; BEHRENS, 2010), sendo esses picos maiores quando comparados aos observados no MG. Tal fato indica uma interação maior dos COS com os óxidos de ferro (hematita), indicando que os óxidos da fração argila podem estar atuando como um preditor para o COS (CLARK et al., 1990). Notam-se picos menores em torno de 1400 nm para o MR1, indicando a fraca correlação da variação espectral com o COS nessa feição, discordando com o observado no MG.

Figura 42 – Importância das variáveis explicativas na predição do COS dos modelos regionais (a) modelo região R1 – MR1, (b) modelo região R2 – MR2 e (c) modelo região R3 – MR3



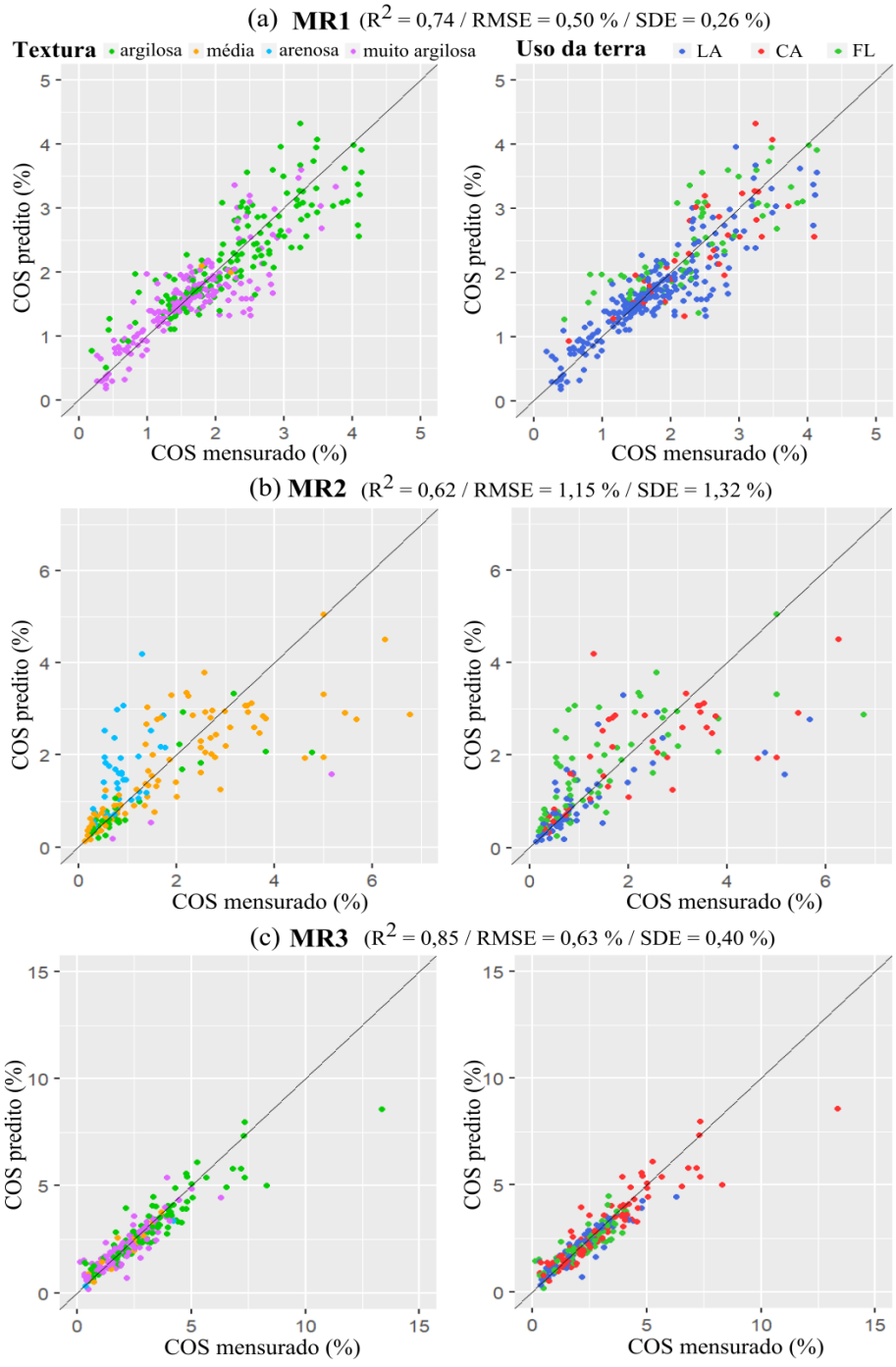
Fonte: Autor.

Esse resultado provavelmente está relacionado à baixa adsorção de grupos funcionais da matéria orgânica com grupos OH da caulinita, em detrimento ao elevado teor de óxidos de ferro nos solos da região R1. Os picos em 2200-2300 nm têm relação com os grupos Al-OH dos argilominerais e com vibrações moleculares dos grupos funcionais C – H, C – O, C = O, C – N (BEN-DOR et al., 1997; SORIANO-DISLA et al., 2014). Pinheiro et al. (2017) também observou importância maior maiores nessas mesmas regiões, na predição de teor de COS em amostras da Amazônia Central, Brasil. Essas mesmas regiões espectrais tiveram importância na modelagem espectroscópica para o COS em escala global (VISCARRA ROSSEL et al., 2016).

O modelo MR2, em geral, apresentou baixos valores de VARIMP, com redução acentuada para as variáveis espectrais relacionadas aos óxidos de ferro (Figura 42b). Isso é devido ao menor conteúdo de óxidos de ferro nos solos da região R2. A região espectral com maior VARIMP é observa em 2200nm, sendo essa característica de feições de absorção grupos OH e Al-OH dos argilominerais, indicando maior interação do COS com esses compostos do solo na região R2. O modelo MR3 apresentou comportamento diferenciado em relação aos demais modelos. Os maiores de VARIMP são observados na região de 660 a 800 nm (Figura 42c), sendo essas relacionadas a vibrações moleculares relacionadas aos grupos funcionais de compostos orgânicos (VISCARRA ROSSEL; BEHRENS, 2010; SORIANO-DISLA et al., 2014), indicando que essa região espectral é importante para predição do COS. O pico VARIMP observado em torno de 1400 e 2200 nm esta relacionado com a influência de grupos OH e em 2300 nm com grupos funcionais da matéria orgânica (VISCARRA ROSSE; BEHRENS, 2010).

Em relação ao comportamento dos modelos regionais de predição do COS, observa-se que os modelos MR1 (Figura 43a) e MR3 (Figura 43c) tendem a subestimar e/ou superestimar o teor de COS em amostras com maior teor de COS, sendo essas amostras associadas ao uso da terra campo nativo e floreta. Já o modelo MR2 (Figura 43b) apresentou tendência a superestimar valores baixos de COS, associados a solos com textura arenosa e, subestimar valores altos relacionados a solos com textura média, sendo que em ambos os solos, o uso terra floresta. Isso é atribuído à dificuldade dos modelos em quantificar valores muito baixos e altos de COS, associado também a menor proporção de amostras com essas características para a calibração do modelo (VISCARRA ROSSEL; BEHRENS, 2010). A presença de material orgânico parcialmente degradado no solo também é um fator que causa aumento na incerteza dos modelos (MORGAN et al., 2009).

Figura 43 – *Scatter plots* da validação do teor de COS medido em laboratório versus o predito com método Cubist-CRR com identificação das classes texturais e usos da terra para os modelos regionais (a) modelo região R1 – MR1, (b) modelo região R2 – MR2 e (c) modelo região R3 – MR3



Fonte: Autor. LA: lavoura; CA: campo nativo; FL: floresta.

6.3.5.2. Modelos preditivos de base textural (MT)

Os resultados de calibração e validação dos modelos texturais obtidos com os métodos multivariados PLSR e Cubist, a partir dos espectros processamentos com SGD e CRR, são resumidos na tabela 10. Em resumo, os modelos gerados pelo método Cubist-CRR apresentam o maior desempenho para os quatro MT (Tabela 10), sendo a maior acurácia observada para o MTMA (R^2 = 0,75; RMSE = 0,50 %; SDE = 0,24 %), seguido dos MTA (R^2 = 0,82; RMSE = 0,88 %; SDE = 0,78 %), MTM (R^2 = 0,58; RMSE = 1,50 %; SDE = 2,17 %) e a menor para o MTAR (R^2 = 0,38; RMSE = 0,40 %; SDE = 0,15 %). A maior acurácia alcançada para o modelo MTMA está relacionada com a menor variação do teor de COS (CV = 56 %, Figura 31) e dos dado espectrais (Figuras 36 e 38). Além disso, grande parte das amostras com textura muito argilosa são solos da região da R1, onde a mineralogia do solo predominante é caulinita e óxidos, conferido menor variação mineralógica. Esses resultado vão em encontro dos observados por Araújo et al. (2014), que obtiveram modelos mais precisos quando as amostras foram agrupadas com base na mineralogia.

Para os MTA e MTM, ambos os conjuntos de amostras apresentam elevada variação nos teores de COS (CV de 71 e 93 %, respectivamente, Figura 31), associado à alta variação espectral (Figuras 36 e 38). Esse comportamento é em função da maior parte das amostras que compõem as sub-bibliotecas de textura argilosa e média serem derivadas da região R2, onde está presente solos com composição mineralógica mais heterogêneas (BORTOLUZZI et al., 2005; SCHMITT, 2015), refletindo na maior variação do comportamento espectral (VISCARRA ROSSEL; BEHRENS, 2010; ARAÚJO et al., 2014). Estudo de Lucà et al. (2017), mostrou que a capacidade preditiva dos modelos pode depender não apenas do tamanho do conjunto de dados, mas também da variabilidade dos solos, em que menores valores de RMSE são obtidos para grupos de amostras mais homogêneos. Para o modelo TM-T4, a menor acurácia está relacionada com a variação do COS (CV = 73 %, Figura 31), mas principalmente pela maior variação espectral observada neste conjunto de dados (Figura 36 e 38). Além disso, o menor número de amostras (n = 148) também é um fator a ser considerado, uma vez que, pequenos conjuntos de dados podem levar à captura insuficiente de variações nos dados, refletindo em maiores erros na modelagem espectroscópica (RAMIREZ-LOPEZ et al., 2014).

Tabela 10 – Resultados de calibração e validação para predição do COS com os modelos texturais (MT) gerados com diferentes abordagens de métodos

Modelos	R^2	RMSE	ME	SDE	RPIQ	R^2	RMSE	ME	SDE	RPIQ
MG*	0,90	0,54	-0,04	0,29	2,85	0,76	0,78	-0,03	0,61	1,90
MTMA	Calibração (n = 665)					Validação (n = 286)				
PLSR-SGD	0,78	0,48	0,01	0,23	2,23	0,72	0,59	0,01	0,35	1,75
PLSR-CRR	0,74	0,52	0,01	0,28	1,97	0,73	0,57	-0,01	0,33	1,91
Cubist-SGD	0,68	0,64	-0,09	0,40	1,67	0,59	0,80	-013	0,62	1,30
Cubist-CRR	**0,86**	**0,41**	**-0,03**	**0,17**	**2,49**	**0,75**	**0,50**	**-0,09**	**0,24**	**2,14**
MTA	Calibração (n = 700)					Validação (n = 300)				
PLSR-SGD	0,76	0,85	0,01	0,73	2,15	0,74	1,02	-0,04	1,04	1,52
PLSR-CRR	0,79	0,80	0,01	0,64	2,30	0,70	1,10	-0,06	1,21	1,41
Cubist-SGD	0,79	0,90	-0,08	0,81	2,04	0,66	1,29	-0,15	1,65	1,20
Cubist-CRR	**0,92**	**0,54**	**-0,02**	**0,29**	**3,40**	**0,82**	**0,88**	**-0,06**	**0,78**	**1,76**
MTM	Calibração (n = 240)					Validação (n = 103)				
PLSR-SGD	0,68	1,11	-0,21	1,18	1,73	0,53	1,66	-0,52	2,52	1,49
PLSR-CRR	0,67	1,06	0,01	1,14	1,80	0,52	1,57	-0,24	2,42	1,59
Cubist-SGD	0,82	0,85	-0,12	0,71	2,26	0,55	1,57	-0,43	2,30	1,58
Cubist-CRR	**0,94**	**0,53**	**-0,08**	**0,28**	**3,58**	**0,58**	**1,50**	**-0,31**	**2,17**	**1,66**
MTAR	Calibração (n = 104)					Validação (n = 44)				
PLSR-SGD	0,53	0,55	-0,12	0,29	1,27	0,28	0,43	-0,01	0,19	1,12
PLSR-CRR	0,77	0,36	0,01	0,13	1,97	0,34	0,52	0,11	0,26	0,92
Cubist-SGD	0,87	0,34	0,08	0,11	2,08	0,29	0,47	0,08	0,22	1,01
Cubist-CRR	**0,94**	**0,24**	**-0,03**	**0,06**	**2,96**	**0,38**	**0,40**	**0,12**	**0,15**	**1,19**

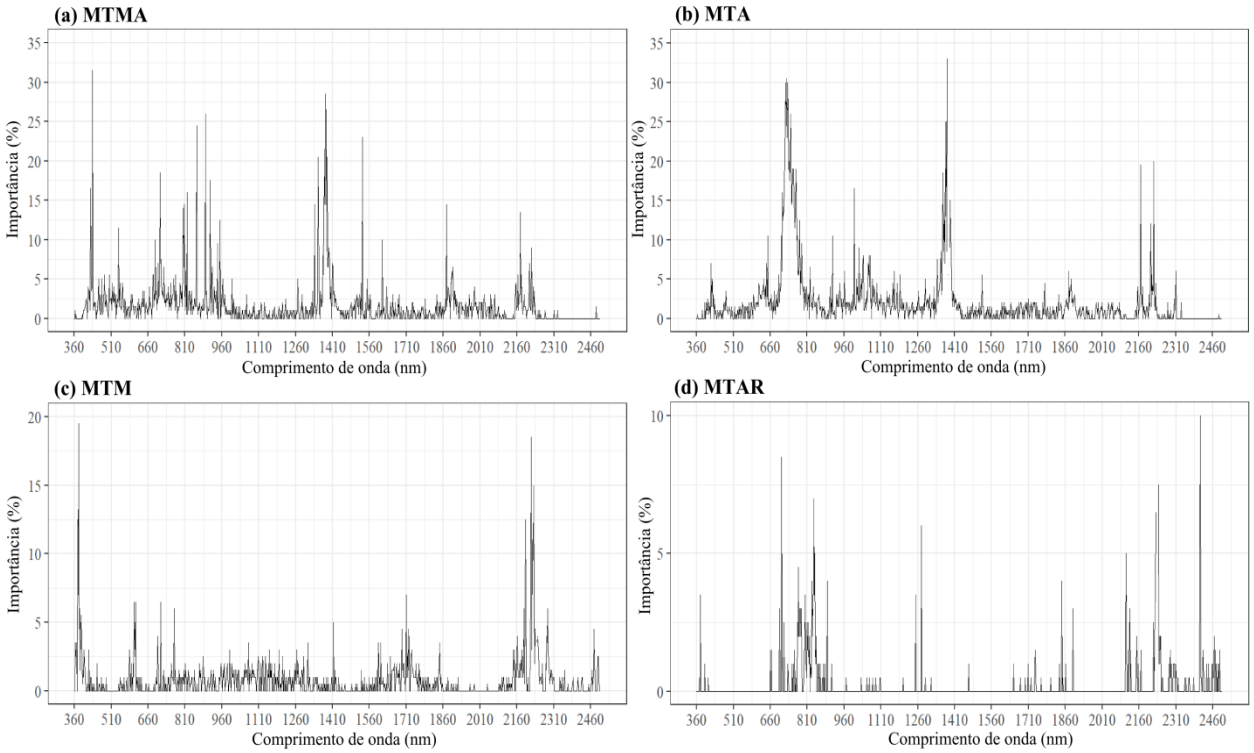
Fonte: Autor. MTMA: modelo textura muito argilosa; MTA: modelo textura argilosa; MTM: modelo textura média; MTAR: modelo textura arenosa; R^2: coeficiente de determinação; RMSE: raiz do erro quadrático médio em %; ME: erro médio; SDE: desvio padrão do erro; RPIQ: Proporção do erro de predição no intervalo interquartil; *Modelo global (MG) gerado pelo método Cubist-CRR com todas as amostras da biblioteca espectral global (BEG).

Com relação à importância das regiões do espectro Vis-NIR na predição do COS pelos modelos texturais, os picos de VARIMP na predição apresentaram diferença marcante para os modelos MTMA e MTA comparados aos MTM e MTAR (Figura 44). As principais mudanças ocorrem na região em torno de 550 e 850 nm, com valores de VARIMP maiores para os modelos MTMA e MTA (Figura 44a, 44b). Isso tem relação com o maior teor de óxidos de ferro nos solos dessas sub-bibliotecas espectrais, em que os picos em torno de 550 nm estão ligados à hematita e em 850 nm hematita e/ou goethita (DALMOLIN et al., 2005), indicando que os óxidos influenciaram na predição do COS devido sua interação com os grupos funcionais da matéria orgânica. Valores altos de VARIMP em torno de 1400 nm também são observados para os modelos MTMA e MTA, sendo essa feição relacionada com grupos funcionais Al-OH da caulinita (VISCARRA ROSSEL; BEHRENS, 2010).

O modelo MTM apresentou os maiores valores de VARIMP na região espectral de 2200 nm (Figura 44c), sendo essa região característica de feições de absorção de grupos OH e Al-OH

dos argilominerais, indicando maior interação do COS com esses compostos do solo. Já o modelo MTAR apresentou o maior valor de VARIMP na região de 700 nm e 2400 nm (Figura 44d). Esses valores altos, provavelmente, são devido a menor influência dos picos de óxidos de ferro e argilominerias, uma vez que, os solos dessa sub-biblioteca espectral apresentam textura arenosa. Assim, esses picos provavelmente estão associados aos grupos funcionais da matéria orgânica (VISCARRA ROSSEL; BEHRENS, 2010).

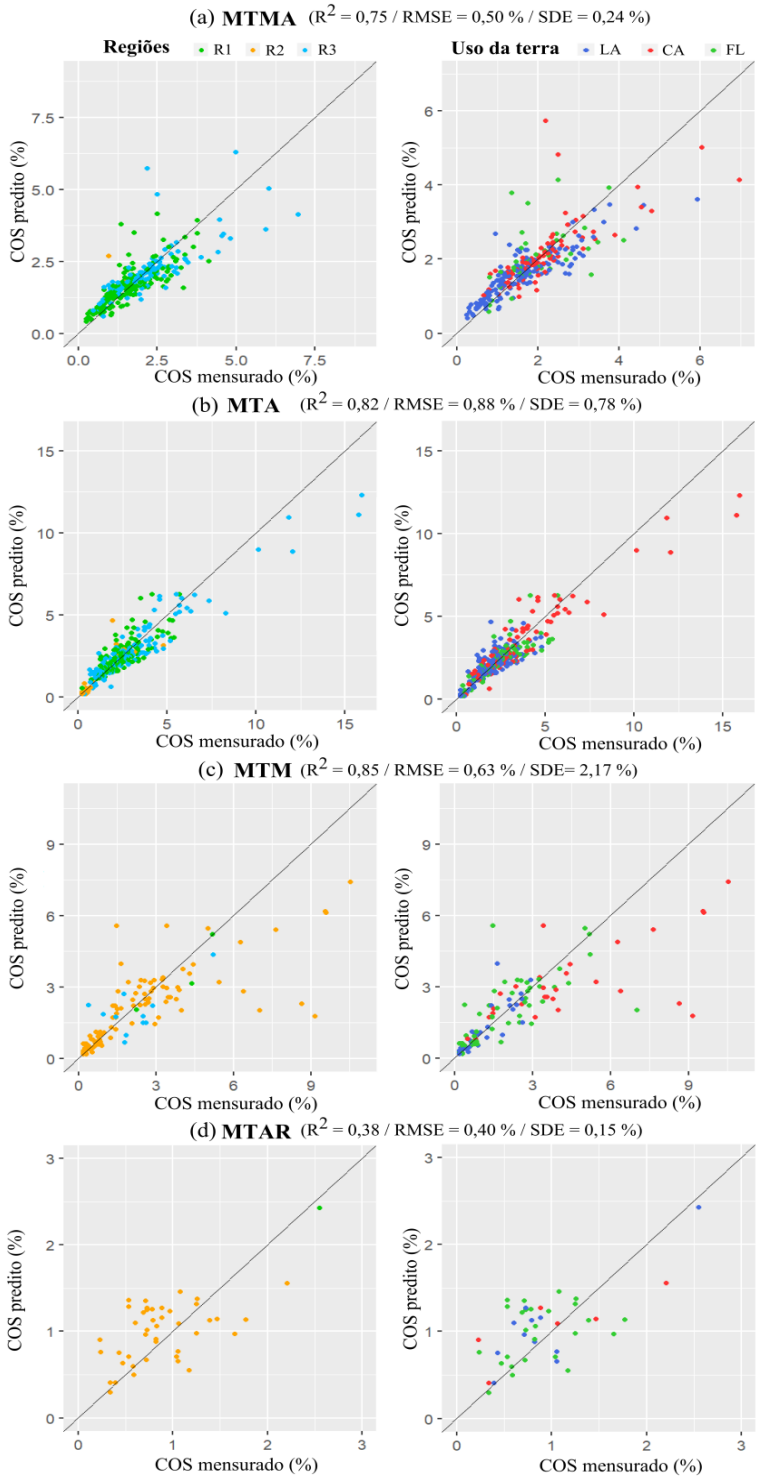
Figura 44 – Importância das variáveis explicativas na predição do COS dos modelos texturais (a) modelo textura muito argilosa – MTMA, (b) modelo textura argilosa – MTA, (c) modelo textura média – MM e (d) modelo textura arenosa – MTAR



Fonte: Autor.

A identificação das amostras de cada região fisiográficas e uso da terra para o conjunto de validação dos quatro modelos texturais são apresentadas na figura 45. Observa-se que os modelos MTMA tende a subestimar e/ou superestimar o teor de COS em amostras com alto teor de COS, sendo essas amostras procedidas da região R3 e associadas ao uso da terra campo nativo e floresta (Figura 45a). Já o modelo MTA apresentou tendência a subestimar amostras com elevado COS da região R3 derivadas do uso da terra PS (Figura 45b).

Figura 45 – *Scatter plots* da validação do teor de COS medido em laboratório versus o predito com método Cubist-CRR com identificação das regiões fisiográficas e usos da terra para os modelos texturais (a) modelo textura muito argilosa – MTMA, (b) modelo textura argilosa – MTA, (c) modelo textura média – MTM e (d) modelo textura arenosa – MTAR



Fonte: Autor. R1: região 1. R2: região 2; R3: região 3; LA: lavoura; CA: campo nativo; FL: floresta.

O modelo MTM apresentou esse mesmo comportamento, porém para amostras do uso da terra campo nativo da região R2 (Figura 45c). Esse comportamento observado nas predições é atribuído à dificuldade dos modelos em quantificar valores altos de COS e a menor proporção de amostras com essas características para a calibração do modelo (VISCARRA ROSSEL; BEHRENS, 2010). A presença de material orgânico parcialmente degradado no solo é um fator que causa aumento na incerteza dos modelos (MORGAN et al., 2009), com destaque para as amostras de floresta e campo nativo. O modelo MTAR, em geral, superestimou os valores de COS, principalmente em amostras com teores baixos, sendo essas amostras da região R2 (Figura 45d). Tal comportamento é comum na modelagem espectroscópica (VISCARRA ROSSEL; BEHRENS, 2010), principalmente quando se tem um conjunto pequeno de amostras para calibração do modelo, como é o caso do conjunto do MTAR (n =148). Contudo, métodos analíticos de laboratório por oxidação úmida com solução sulfocrômica também subestimam o COS em amostras com baixos teores (NELSON; SOMMERS, 1975).

O efeito do tamanho das partículas do solo na predição do COS foi observado estudo de Wight et al. (2016), em que amostras com maior teor de argila apresentaram a maior porcentagem de erros para a previsão do COS. Nesse caso, os autores sugeriram que se a textura do conjunto de amostras fosse previamente conhecida, previsões mais acuradas poderiam ser alcançadas. Isso foi abordado em um estudo de Wijewardane et al. (2016), onde os autores observaram maior RMSE (3,07 %) para solos textura arenosa e menor RMSE (0,48 %) para textura argilosa. No presente estudo, essa mesma tendência foi observada, onde os modelos obtidos para grupos de amostras com textura mais argilosa (MTMA e MTA) foram mais precisos em relação aos modelos de textura mais grossa (MTM e MTAR) (Figura 45). Em resumo, os resultados obtidos para os modelos com base na textura do solo, mostram que estratificar bibliotecas espectrais de solos tropicais, compostas por amostras com características semelhantes às da região R2 (Depressão Central do Estado do RS, Brasil) pelo critério granulometria, pode apresentar desvantagem, quando se busca aumentar a acurácia das predições de COS.

6.3.5.3. Modelos preditivos de base uso da terra (MUT)

Os resultados de calibração e validação dos modelos de uso da terra obtidos com os métodos multivariados PLSR e Cubist, a partir dos espectros processamentos com SGD e CRR,

são resumidos na tabela 11. No geral, os modelos gerados pelo método Cubist-CRR apresentam maior acurácia (Tabela 11), seguindo os resultados observados para o modelo global, modelo regionais e modelos texturais. Em relação aos processamentos espectrais SGD e CRR, o método Cubist-CRR apresentou a maior acurácia nos três MUT. Já o método PLSR mostrou desempenho distinto entre os conjuntos de amostras. Para modelo MLA e MCA, o método PLSR-SGD apresentou mais acurácia em relação ao PLSR-CRR. Já para o conjunto de amostras do uso da terra floresta, o método PLSR-SGD mostrou melhor desempenho (Tabela 11). Isso tem relação com a maior variação nas características químicas, físicas e espectrais observada nos solos da floresta já discutidas (Figuras 31, 32, 36 e 38, respectivamente). Assim, sugerindo que o método PLSR exige uma transformação mais rigorosa de espectros com alta variação espectral. Nesse sentido, o processo matemático de derivação utilizado pela técnica SGD enfatiza com mais rigor as feições espectrais de maior absorção (SAVITZKY; GOLAY, 1964; RINNAN et al., 2009).

Tabela 11 – Resultados de calibração e validação para predição do COS com os modelos usos da terra (MUT) gerados com diferentes abordagens de métodos

Modelos	R^2	RMSE	ME	SDE	RPIQ	R^2	RMSE	ME	SDE	RPIQ
MG*	0,90	0,54	-0,04	0,29	2,85	0,76	0,78	-0,03	0,61	1,90
MLA	Calibração (n = 798)					Validação (n = 342)				
PLSR-SGD	0,68	0,52	0,01	0,27	2,29	0,48	0,73	0,03	0,53	1,70
PLSR-CRR	0,66	0,54	0,01	0,29	2,14	0,62	0,61	0,04	0,37	2,09
Cubist-SGD	0,76	0,49	-0,07	0,24	2,42	0,65	0,61	-0,06	0,37	2,03
Cubist-CRR	**0,85**	**0,38**	**-0,03**	**0,14**	**3,06**	**0,70**	**0,55**	**0,01**	**0,30**	**2,32**
MFL	Calibração (n = 471)					Validação (n = 202)				
PLSR-SGD	0,84	0,97	0,01	0,94	2,23	0,66	1,31	0,25	1,67	1,36
PLSR-CRR	0,83	1,01	0,01	1,02	2,14	0,71	1,17	0,06	1,38	1,52
Cubist-SGD	0,86	1,10	-0,16	1,18	1,97	0,76	1,12	-0,05	1,25	1,59
Cubist -CRR	**0,94**	**0,62**	**-0,02**	**0,39**	**3,48**	**0,79**	**0,99**	**0,01**	**0,99**	**1,79**
MCA	Calibração (n = 454)					Validação (n = 194)				
PLSR-SGD	0,67	0,75	0,01	0,57	2,31	0,43	0,99	0,14	0,88	1,94
PLSR-CRR	0,77	0,62	0,01	0,38	2,82	0,48	0,95	0,08	0,98	1,85
Cubist-SGD	0,74	0,75	-0,13	0,55	2,33	0,55	0,83	0,01	0,69	2,22
Cubist-CRR	**0,89**	**0,47**	**-0,03**	**0,22**	**2,76**	**0,54**	**0,83**	**0,07**	**0,69**	**2,21**

Fonte: Autor. MLA: modelo lavoura; MFL: modelo floresta; MCA: modelo campo nativo; R^2: coeficiente de determinação; RMSE: raiz do erro quadrático médio em %; ME: erro médio; SDE: desvio padrão do erro; RPIQ: Proporção do erro de predição no intervalo interquartil; *Modelo global (MG) gerado pelo método Cubist-CRR com todas as amostras da biblioteca espectral global (BEG).

O modelo MLA teve a maior acurácia (R^2 = 0,70; RMSE = 0,55 %; SDE = 0,30 %), seguido do MCA (R^2 = 0,54; RMSE = 0,83 %; SDE = 0,69 %) e o MFL com a menor (R^2 = 0,79; RMSE = 0,99 %; SDE = 0,99 %) (Tabela 11). Esse resultado está relacionado com a redução no

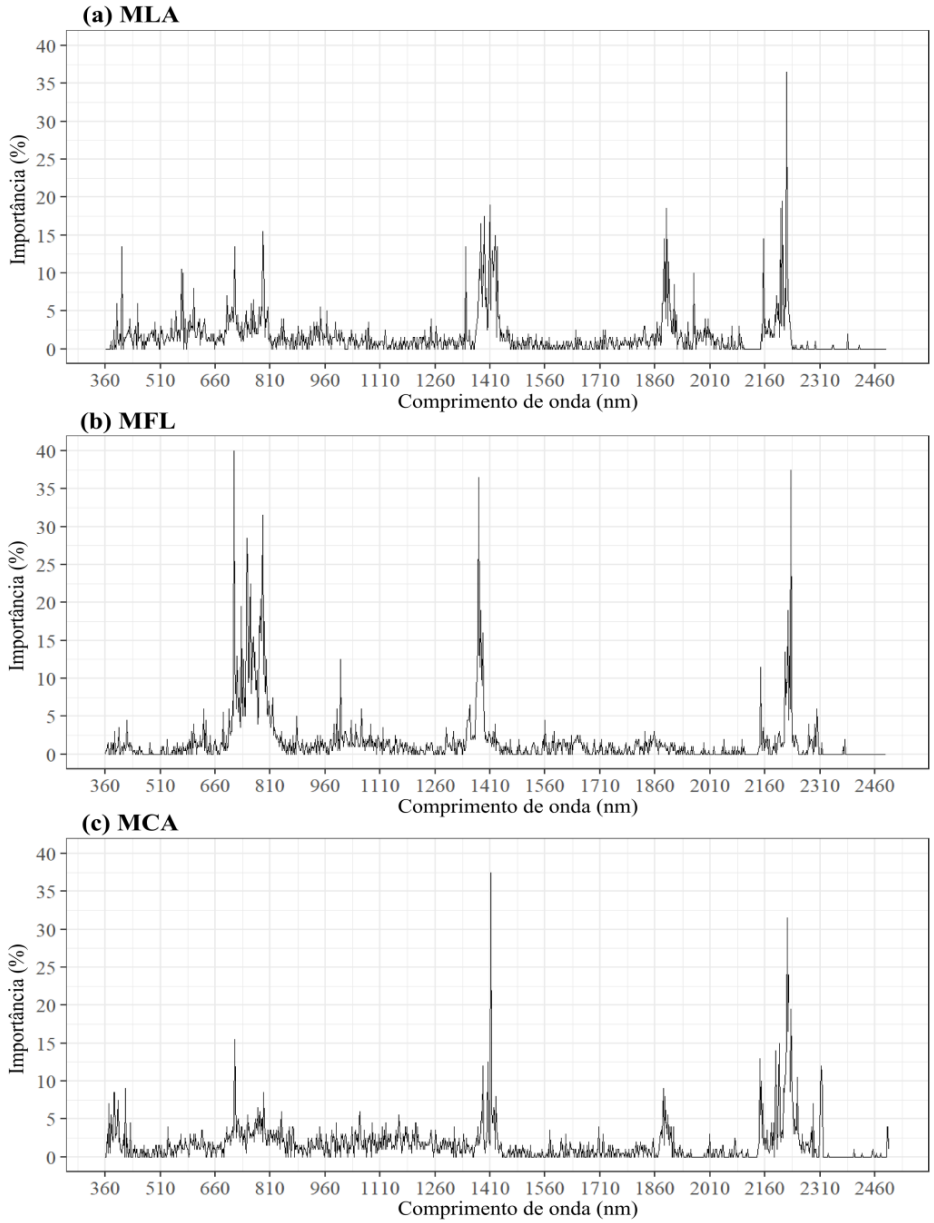
intervalo de predição para o conjunto de amostra de FA, mas principalmente pela redução da variação dos componentes organominerais (ARAÚJO et al., 2014; NOCITA et al.,2014; WIJEWARDANE et al., 2016). Dessa maneira, a união de amostras com essas características resultou na menor variação espectral do conjunto de dados. Comportamento semelhante foi observado para o modelo regional MR1. Nessas condições, a calibração e validação do MLA apresentaram maior acurácia em detrimento do modelo global, confirmando que o agrupamento de amostras com base no uso da terra para compor bibliotecas espectrais mais homogêneas é uma estratégia para melhorar a acurácia das predições de COS (STEVENS et al., 2013). Contudo, os modelos MFL e MCA apresentam desempenho reduzido em relação ao modelo global. Tal fato está vinculado a redução do número de amostras, mas principalmente ao maior intervalo de predição do teor de COS, alta variação organomineral e espectral observada nessas sub-bibliotecas (Figuras 31, 32, 36 e 38).

A importância das regiões do espectro Vis-NIR na predição do COS pelos modelos MUT apresentaram diferenças marcantes entre modelos (Figura 46), sendo as principais mudanças observadas nas regiões relacionadas aos óxidos de ferro (480 e 850 nm) e grupos funcionais OH e Al-OH dos argilominerais (1400 e 1900 nm) (VISCARRA ROSSEL; BEHRENS, 2010). No modelo MLA nota-se picos de VARIMP maiores em torno de 480 nm (Figura 46a) comparado aos demais modelos. Isso está associado ao maior teor de óxidos de ferro nos solos dessa sub-biblioteca espectral, os quais são derivados na sua grande maioria de Latossolos com altos teores de hematita na fração argila (KER; RESENDE, 1990; DALMOLIN et al., 2005).

No modelo MFL, maiores valores de VARIMP são observados na região em trono de 660-800, 1400 e 2200 nm (Figura 46b), sendo esses relacionados às vibrações moleculares dos grupos funcionais de compostos orgânicos e também ligações metal-OH (BEN-DOR et al., 1997; VISCARRA ROSSEL; BEHRENS, 2010). Tal fato era esperado devido o maior teor de COS presente no uso da terra floresta, resultando na maior influência desses componentes na modelagem espectroscópica. Na modelagem espectroscópica do COS para 3854 amostras de diversas regiões do Brasil fo1 observada maior importância na predição nessas mesmas regiões espectrais (SANTANA; SOUZA; POPPI, 2018), confirmando a importância dessas variáveis para quantificar o COS. Esse mesmo comportamento era esperado para o modelo MCA, onde o teor de COS também é elevado, no entanto, os maiores valores de VARIMP são observados em 1400 e 2200 nm (Figura 46c). Essas regiões são característica de feições de absorção de grupos

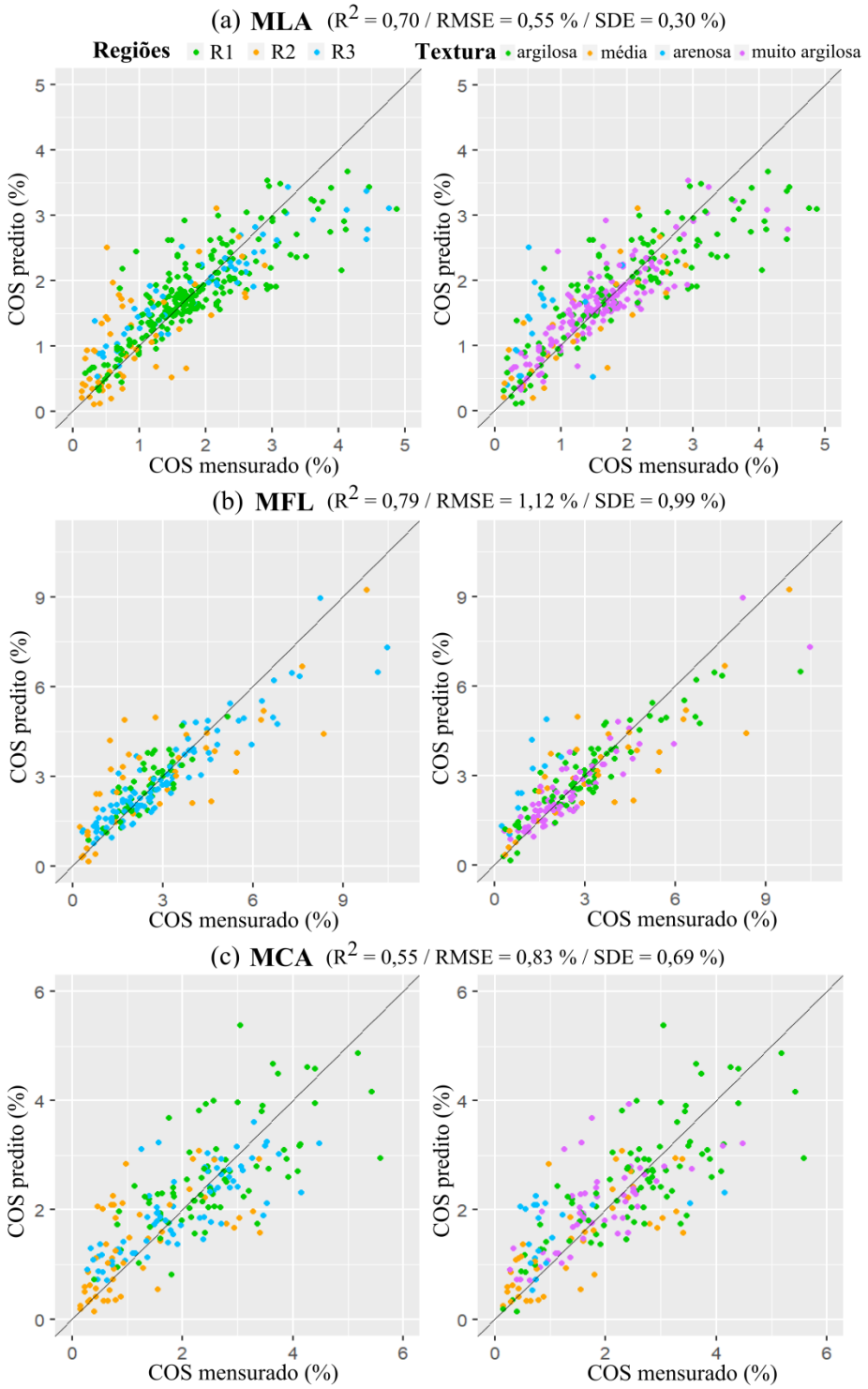
OH e Al-OH dos argilominerais (VISCARRA ROSSEL; BEHRENS, 2010), indicando maior interação do COS com esses compostos do solo no uso da terra campo nativo. Nota-se ainda aumento do VARIMP em torno de 2200 e 2310 nm, sendo essas feições associadas com grupos funcionais da matéria orgânica (SORIANO-DISLA et al., 2014).

Figura 46 – Importância das variáveis explicativas na predição do COS dos modelos usos da terra (a) modelo lavoura – MLA, (b) modelo floresta – MFL e (c) modelo campo – MCA



Fonte: Autor.

Figura 47 – *Scatter plots* da validação do teor de COS medido em laboratório versus o predito com método Cubist-CRR com identificação das regiões fisiográficas e classes texturais para os modelos uso da terra (a) modelo lavoura – MLA, (b) modelo floresta – MFL e (c) modelo campo – MCA



Fonte: Autor. R1: região 1. R2: região 2; R3: região 3; LA: lavoura; CA: campo nativo; FL: floresta.

A identificação das amostras de cada região fisiográficas e classe textural para o conjunto de validação dos três modelos gerados com base nos usos da terra são apresentadas na figura 47. Observa-se que o modelo MLA tende a subestimar o COS em amostras com alto teor de COS (Figura 47a), sendo essas amostras procedidas da região R3 e com textura argilosa Já os modelos MFL e MCA tende a superestimar o COS em amostras com baixo teor e com textura arenosa e, subestimar em amostras com alto teor e texturas mais finas (Figuras 47b e 47c). A grande maioria dessas amostras é provinda da região R2. Esse comportamento é associado com a menor representatividade de amostras com essas particulares para a calibração do modelo (VISCARRA ROSSEL; BEHRENS, 2010). Contudo, métodos analíticos de laboratório por oxidação úmida com solução sulfocrômica também podem superestimar o COS em amostras com baixos teores e subestimar em amostras com alto COS (NELSON; SOMMERS, 1975).

6.3.5.4. Modelos preditivos de base espectrais (ME)

Os resultados de calibração e validação dos modelos espectrais obtidos com os métodos multivariados PLSR e Cubist, a partir dos espectros processamentos com SGD e CRR, são resumidos na tabela 12. Em resumo, os modelos gerados pelo método Cubist-CRR apresentam o maior desempenho para os quatro modelos espectrais (Tabela 12). A maior acurácia observada para o MC4 (R^2 = 0.70; RMSE = 0,44 %; SDE = 0,20 %), seguido dos MC3 (R^2 = 0,80; RMSE = 0,58 %; SDE = 0.33%), MC1 (R^2 = 0,34; RMSE = 1,45 %; SDE = 2,12 %) e a menor para o MC2 (R^2 = 0,87; RMSE = 1,54 %; SDE = 2,39 %). A maior acurácia alcançada para o modelo MC4 está relacionada com a menor variância do teor de COS (SD = 0,84 %; CV = 49 %; Var. = 0,71, Figura 31f) e dos dados espectrais (Figuras 38q). Além disso, grande parte das amostras é derivada da região da R1, onde os solos apresentam menor variação mineralógica. Esse resultado vai de encontro com os observados por Araújo et al. (2014) para solos da região sudeste e centro oeste do Brasil, que obtiveram modelos mais precisos quando as amostras foram agrupadas com base na mineralogia.

Em relação aos processamentos espectrais SGD e CRR, o método Cubist-CRR apresentou a maior acurácia nos quatro modelos espectrais. Já o método PLSR mostrou desempenho distinto (Tabela 12), em que os modelos MC1 e MC2 apresentam maior acurácia para o método PLSR-SGD. Já para os modelos MC3 e MC4 o maior desempenho foi obtido para o método PLSR-

CRR. Isso tem relação com a maior variação nas características químicas, físicas e espectrais observada nos solos das sub-bibliotecas espectrais classe 1 e 2, já discutidas nas figuras 31, 32 e 36. Isso indica que o método PLSR exige uma transformação mais rigorosa de espectros com alta variação espectral. Esse mesmo comportamento foi observado para os modelos baseados nos usos terra do presente estudo.

Tabela 12 – Resultados de calibração e validação para predição do COS com os modelos espectrais (ME) gerados com diferentes abordagens de métodos

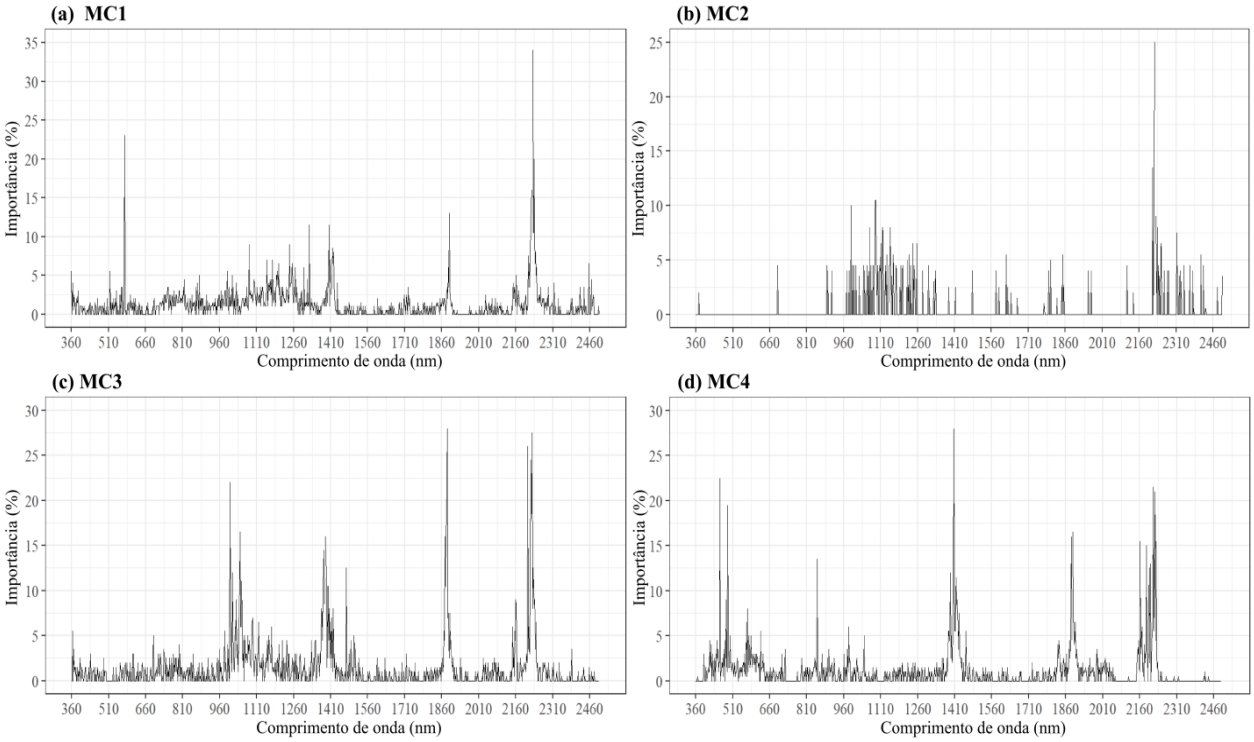
Modelos	R^2	RMSE	ME	SDE	RPIQ	R^2	RMSE	ME	SDE	RPIQ
MG*	0,90	0,54	-0,04	0,29	2,85	0,76	0,78	-0,03	0,61	1,90
MC1	Calibração (n = 400)					Validação (n = 172)				
PLSR-SGD	0,51	1,15	0,09	1,32	1,65	0,39	1,39	0,10	1,93	1,27
PLSR-CRR	0,59	1,04	0,01	1,09	1,81	0,33	1,51	0,07	2,29	1,17
Cubist-SGD	0,66	1,06	-0,11	1,12	1,78	0,31	1,49	-0,11	2,21	1,19
Cubist-CRR	**0,85**	**0,71**	**-0,06**	**0,51**	**2,65**	**0,34**	**1,45**	**0,08**	**2,12**	**1,21**
MC2	Calibração (n = 115)					Validação (n = 49)				
PLSR-SGD	0,99	0,27	0,00	0,07	14,65	0,87	1,55	-0,26	2,37	3,09
PLSR-CRR	0,98	0,46	0,00	0,21	8,62	0,63	2,64	-0,25	1,59	1,81
Cubist-SGD	0,86	1,39	-0,21	1,92	2,85	0,87	2,00	-0,46	3,85	2,39
Cubist-CRR	**0,96**	**0,60**	**-0,01**	**0,37**	**6,59**	**0,87**	**1,54**	**-0,20**	**2,39**	**3,09**
MC3	Calibração (n = 491)					Validação (n = 210)				
PLSR-SGD	0,79	0,60	0,00	0,36	3,03	0,57	0,85	-0,04	0,72	2,04
PLSR-CRR	0,77	0,63	0,00	0,39	2,89	0,69	0,71	-0,06	0,51	2,42
Cubist-SGD	0,81	0,65	-0,10	0,41	2,79	0,73	0,72	-0,12	0,51	2,39
Cubist-CRR	**0,93**	**0,37**	**-0,01**	**0,14**	**4,91**	**0,80**	**0,58**	**-0,05**	**0,33**	**2,99**
MC4	Calibração (n = 717)					Validação (n = 307)				
PLSR-SGD	0,72	0,45	0,00	0,20	2,36	0,63	0,53	0,01	0,28	1,90
PLSR-CRR	0,70	0,47	0,00	0,22	2,26	0,69	0,46	0,03	0,21	2,19
Cubist-SGD	0,77	0,44	-0,06	0,19	2,41	0,67	0,48	0,00	0,23	2,09
Cubist-CRR	**0,84**	**0,34**	**-0,03**	**0,12**	**3,05**	**0,70**	**0,44**	**0,01**	**0,20**	**2,20**

Fonte: Autor. MC1: modelo espectral classe 1; MC2: modelo espectral classe 2; MC3: modelo espectral classe 3; MC4: modelo espectral classe 4; R^2: coeficiente de determinação; RMSE: raiz do erro quadrático médio em %; ME: erro médio; SDE: desvio padrão do erro; RPIQ: Proporção do erro de predição no intervalo interquartil; *Modelo global (MG) gerado pelo método Cubist-CRR com todas as amostras da biblioteca espectral global (BEG).

A importância das regiões do espectro Vis-NIR na predição do COS pelos modelos espectrais apresentaram diferenças significativas entre modelos (Figura 48), sendo as principais mudanças observadas nas regiões dos óxidos de ferro (480 e 850 nm) e grupos funcionais OH e Al-OH dos argilominerais (1400 e 1900 nm) (VISCARRA ROSSEL; BEHRENS, 2010; SORIANO-DISLA et al., 2014). Essa variação está associada com as diferenças na composição organomineral dos solos de cada conjunjo de dados e a resposta espectral perante aos componentes espectralmente ativos presentes (VISCARRA ROSSEL; BEHRENS, 2010;

VISCARRA ROSSEL et al., 2016). Em geral os picos de VARIMP na região dos óxidos de ferro são observados para o MC4 (Figura 48d). Isso está associado ao maior teor de óxidos de ferro nos solos dessa sub-biblioteca espectral. Nas regiões em torno de 1900 e 2200 nm, os maiores picos de VARIMP são observados no modelo MC3 (Figura 48c), sendo esses relacionados aos grupos OH e Al-OH dos argilominerais (VISCARRA ROSSEL; BEHRENS, 2010; SORIANO-DISLA et al., 2014). Já os modelos MC1 e MC2 apresentaram picos altos de VARIMP apenas na região em torno de 2200 nm (Figuras 48a e 48b). Essa região tem relação com os grupos Al-OH dos argilominerais e com vibrações moleculares dos grupos funcionais C – H, C – O, C = O, C – N (VISCARRA ROSSEL; HICKS, 2015). Isso indica que essa região espectral é a variável preditora mais importante para estabelecer relações com o teor de COS para solos com composição organomineral heterogênea, conforme observado nessas sub-bibliotecas.

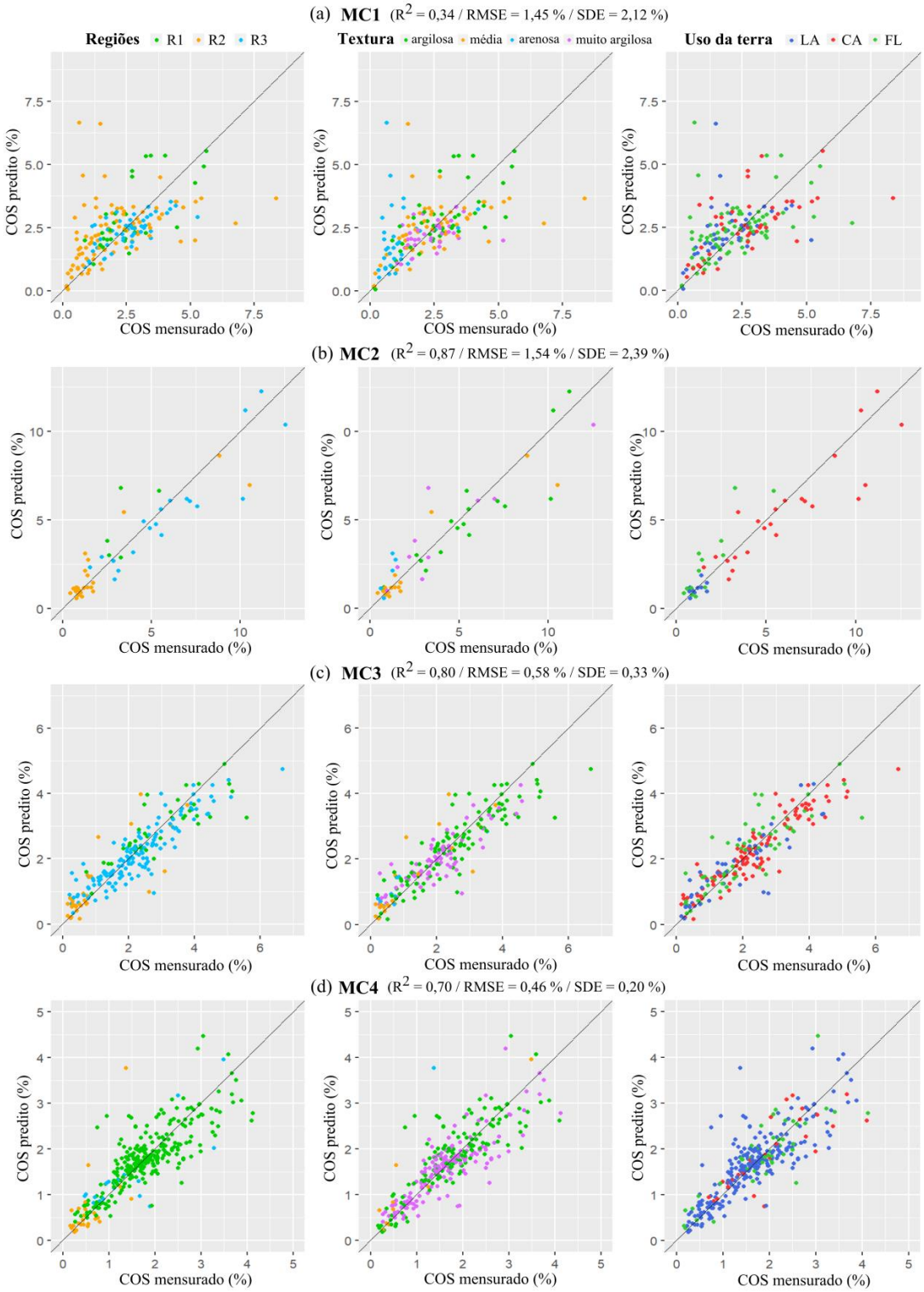
Figura 48 – Importância das variáveis explicativas na predição do COS dos modelos espectrais (a) modelo espectral classe 1 – MC1, (b) modelo espectral classe 2 – MC2, (c) modelo espectral classe 3 – MC3 e (d) modelo espectral classe 4 – MC4



Fonte: Autor.

A identificação das amostras de cada região fisiográficas, classe textural e uso da terra para o conjunto de validação dos quatro modelos gerados com base nas classes espectrais é apresentada na figura 49. Observa-se que os modelos MC1 e MC2 tendem a superestimar o COS em amostras com baixo teor e com textura arenosa e, subestimar em amostras com alto teor de COS e texturas mais finas (Figura 49a e 49b, respectivamente), sendo essas amostras pertencentes à região R2 e R3. Esse comportamento é associado com a menor representatividade de amostras com essas particulares para a calibração do modelo (VISCARRA ROSSEL; BEHRENS, 2010), resultando nos altos valores de RMSE e SDE encontrados. Já os modelos MC3 e MC4 apresentam menor dispersão dos valores de COS medidos versus preditos, consequentemente os menores valores de RMSE e SDE (Figura 49c e 49d, respectivamente). Esse resultado está relacionado com a menor variação organomineral e espectral para esses conjuntos de dados, com destaque para a sub-biblioteca espectral classe 4 (BEC4), confirmando que modelos espectroscópicos mais acurados para o COS podem ser obtidos com bibliotecas espectrais regionais heterogêneas mediante o agrupamento de amostras com base na variação espectral. Isso foi observado por Araújo et al. (2014) para uma biblioteca espectral de solos da região sudeste e centro oeste do Brasil, que obtiveram modelos mais precisos para o teor de matéria orgânica (RMSE = 0,47 %) quando as amostras foram agrupadas em cinco classes espectrais. Da mesma forma, em escala global Viscarra Rossel et al. (2016) agruparam as amostras em seis classes espectrais para desenvolver modelos de predição de COS, sendo observado redução do valor de RMSE de 1,45 % para 1,11 %. Nesse sentido, os resultados encontrados mostram que modelos espectroscópicos gerados com base nas características espectrais pode ser uma estratégia para alcançar maiores acurácia nas estimativas do COS em escala regional no sul do Brasil.

Figura 49 – *Scatter plots* da validação do teor de COS medido em laboratório versus o predito com método Cubist-CRR com identificação das regiões fisiográficas, classes texturais e usos da terra para os modelos espectrais (a) modelo espectral classe 1 – MC1, (b) modelo espectral classe 2 – MC2, (c) modelo espectral classe 3 – MC3 e (d) modelo espectral classe 4 – MC4



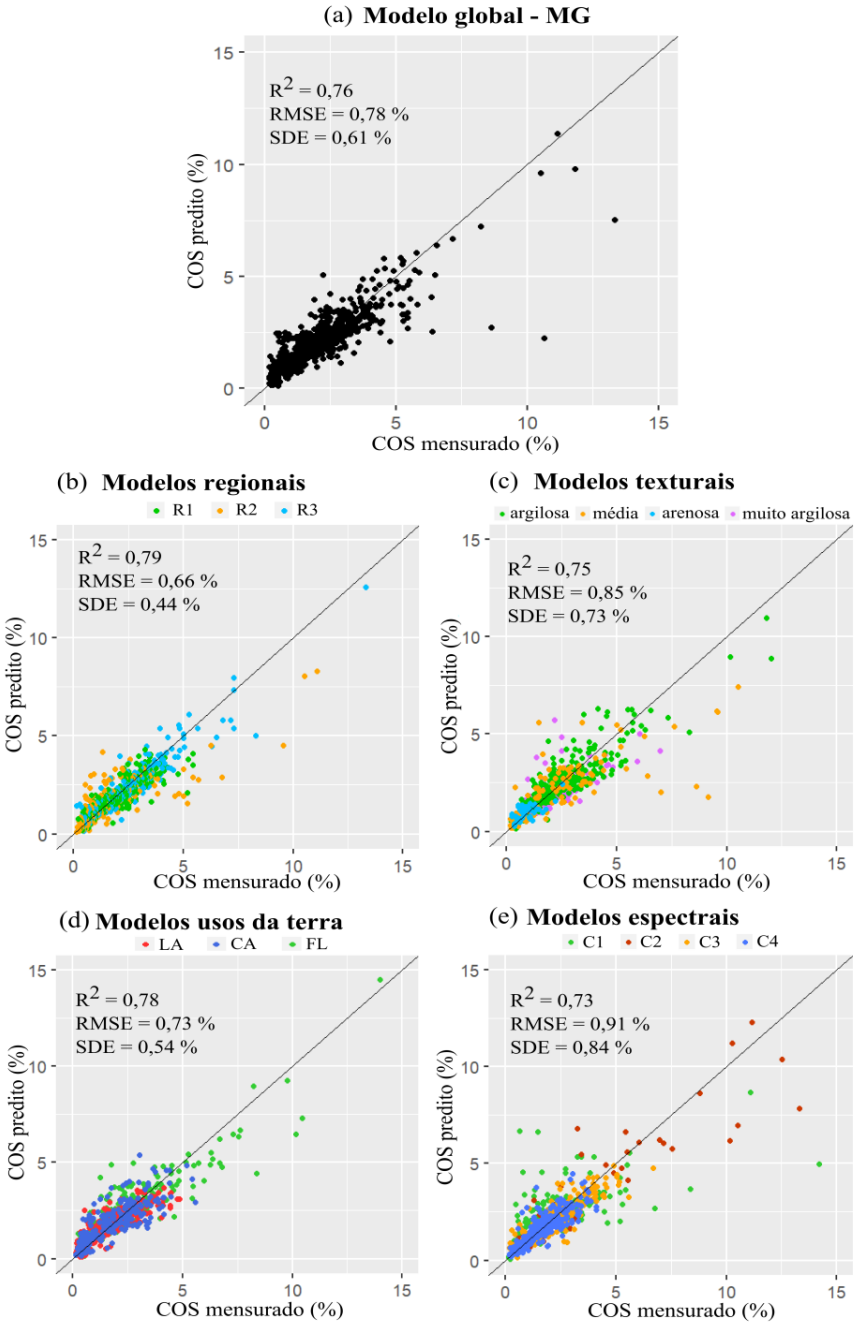
Fonte: Autor.

6.3.6. Validação externa dos modelos

A partir do modelo global e dos modelos obtidos pela estratégia de estratificação da BEG em sub-bibliotecas mais homogêneas com base nos critérios regiões fisiográficas, classes texturais, usos da terra e classes espectrais, foram calculados o R^2, RMSE e SDE para o conjunto de validação da BEG (n = 738) com base no valor predito por cada um dos modelos (Figura 50). As predições realizadas pelos modelos regionais (Figura 50b) apresentaram os menores erros em relação o modelo global e demais (Figura 50). Ainda, se observa na modelagem regional, que os maiores erros estão associados ao modelo da região R2, seguido do modelo R3, em que valores de COS são subestimados. Em segundo lugar no ranking de acurácia estão os modelos baseados nos usos da terra (Figura 50d), que também exibiram melhor desempenho em relação ao modelo global. Em terceiro lugar, se observa os modelos baseados na classe textural (Figura 50c) e por fim os modelos de classes espectrais (Figura 50e), onde ambos apresentaram acurácia inferior em relação ao modelo global.

Diante da maior acurácia alcançada para os modelos regionais e de usos da terra (Figura 50), esses foram utilizados na predição do COS das amostras do CEV com o objetivo entender o comportamento de cada modelo gerado e verificar a aplicabilidade de cada modelo a dados externos com diferentes características ambientais. Os resultados da predição para as 128 amostras do CEV são apresentados na tabela 13. As estatísticas do erro para o MG apresentaram EM = -0,08 %; RMSE = 0,67 % e SDE = 0,45 %, quando considerado o conjunto total de amostras (n = 128) (Tabela 13). Porém, quando considerado apenas as amostras das regiões R1 e R3, os erros de predição reduziram (Tabela 13). Já as predições do COS para as região R2 pelo modelo GM apresentaram maior incerteza, com EM = -0,15 %; RMSE = 0,90 %; SDE = 0,81 %. Esses resultados podem ser explicados devido às características destas regiões estarem mais representadas na BEG, com as quais o MG foi calibrado. Enquanto a R1 tem 45 % (n = 767), a R2 e R3 têm 22 % (n = 382) e 33 % (n = 574), respectivamente.

Figura 50 – *Scatter plots* da validação (n = 738) mostrando o teor de COS medido em laboratório versus o predito (a) modelo global – MG, (b) modelo regional – MR, (c) modelos texturais – MT, (d) modelos usos da terra – MUT, (e) modelos espectrais – ME



Fonte: Autor.

Os modelos regionais MR1 e MR3 apresentaram aumento da acurácia nas estimativas do COS para as regiões R1 e R3, comparada a acurácia do MG (Tabela 13). O MR1 alcançou EM = -0,02 %; RSME = 0,49 %; SDE = 0,28 % e o MR3 teve EM = 0,01 %; RMSE = 0,50 %; SDE =

0,30 %. Em contrapartida, o MR2 apresentou menor acurácia em relação ao MG (EM = -0,24 %; RMSE = 1,18 % e SDE = 1,39 %) (Tabela 13), com tendência a subestimar os teores de COS. Tais resultados estão relacionados com a variação composicional e espectral de cada região, sendo essa maior da R2, mas também é preciso considerar o efeito no número de amostras utilizadas na calibração dos modelos, em que o menor se observa no MR2. Portanto, conjunto de dados com essas particularidades necessitam de um número maior de amostras, visando representar melhor a variação, caso contrário, maiores acurácias são obtidas com modelos globais (ARAÚJO et al., 2014; VISCARRA ROSSEL et al. 2016).

Comportamento semelhante foi observado para os modelos de usos da terra, onde MLA apresentou maior acurácia na predição das amostras de lavoura (EM= -0,04 %; RMSE = 0,51 %; SDE = 0,26 %) em relação ao MG (EM = -0,08 %; RMSE = 0,67 % e SDE = 0,45 %). Porém, os MFL e MCA não apresentaram redução na incerteza das estimadas para as amostras derivadas de floresta e campo nativo (Tabela 13). Isso tem relação com a alta variação composicional e espectral dos solos de floresta e campo nativo, conforme já discutido, mas também devido o reduzido número de amostras utilizadas na calibração dos modelos.

Com relação à aplicação dos modelos baseados em regiões e usos da terra para a predição do COS do CEV, maiores erros comparado ao erro obtido no modelo global (Tabela 13). Esse comportamento era esperado, uma vez que, os modelos regionais e de usos da terra são específicos para solos com as características para os quais foram calibrados. No entanto, os resultados de acurácia observados na tabela 13 permitem compreender a semelhança entre os modelos e sua aplicabilidade em outras regiões e usos da terra. Por exemplo, o MR1 apresentou boa acurácia na predição do COS da região R3 (EM = 0,01 %; RMSE = 0,60 %; SDE = 0,37 %), sendo esses valores superiores aos observados para o modelo global para o total de amostras (n = 128) (EM = -0,08 %; RMSE = 0,67 % e SDE = 0,45 %) e semelhante quando considerado apenas amostras da R3 (n = 42) (EM = -0,03 %; RMSE = 0,58 % e SDE = 0,35 %) (Tabela 13). O mesmo comportamento é observado para o modelo MLA na predição do COS das amostras da região R1, que alcançou EM = 0,11 %, RMSE = 0,68 % e SDE = 0,45 %, sendo esses valores similares aos observados para o modelo global. No primeiro exemplo, a explicação é a similaridade na composição organomineral e variação espectral dos solos das regiões R1 e R3, já no segundo, é a maior representatividade de amostras provindas do uso da terra lavoura na sub-bibliotecas da região R1.

Tabela 13 – Resumo estatístico da acurácia da predição do COS para o conjunto externo de validação (CEV) pelo modelo global (MG), modelos regionais (MR) e modelos usos da terra (MUT)

Conjunto de dados	Número de amostras	EM	RMSE	SDE
		Modelo global (MG)		
Conjunto total	128	-0,08	0,67	0,45
Conjunto região R1	57	-0,08	0,59	0,34
Conjunto região R2	29	-0,15	0,90	0,81
Conjunto região R3	42	-0,03	0,58	0,35
		Modelo região R1 (MR1)		
Conjunto total	128	-0,05	1,28	1,66
Conjunto região R1	57	-0,02	0,49	0,28
Conjunto região R2	29	-0,24	1,18	1,39
Conjunto região R3	42	0,01	0,60	0,37
		Modelo região R2 (MR2)		
Conjunto total	128	-0,27	1,08	1,09
Conjunto região R1	57	-0,08	0,56	0,32
Conjunto região R2	29	-0,24	1,18	1,39
Conjunto região R3	42	0,01	0,60	0,37
		Modelo região R3 (MR3)		
Conjunto total	128	0,40	1,50	2,11
Conjunto região R1	57	0,02	0,86	0,75
Conjunto região R2	29	1,70	2,82	5,23
Conjunto região R3	42	0,01	0,50	0,30
		Modelo lavoura (MLA)		
Conjunto total	128	-0,48	1,43	1,83
Conjunto região R1	57	-0,19	0,69	0,44
Conjunto região R2	29	-0,54	1,62	2,43
Conjunto região R3	42	0,83	1,95	3,17
Conjunto lavoura	60	-0,04	0,51	0,26
Conjunto floresta	38	-1,34	2,43	4,24
Conjunto campo nativo	30	-0,29	0,87	0,70
		Modelo floresta (MFL)		
Conjunto total	128	0,16	0,87	0,74
Conjunto região R1	57	0,11	0,68	0,45
Conjunto região R2	29	0,34	1,22	1,43
Conjunto região R3	42	0,12	0,82	0,67
Conjunto lavoura	60	0,25	0,72	0,46
Conjunto floresta	38	-0,03	1,00	1,10
Conjunto campo nativo	30	0,22	0,98	0,94
		Modelo campo nativo (MCA)		
Conjunto total	128	-0,27	1,15	1,25
Conjunto região R1	57	0,03	0,70	0,49
Conjunto região R2	29	-0,50	1,48	2,00
Conjunto região R3	42	-0,52	1,36	1,61
Conjunto lavoura	60	0,05	0,56	0,31
Conjunto floresta	38	-0,88	1,84	2,66
Conjunto campo nativo	30	-0,14	0,85	0,73

Fonte: Autor.

Em geral, se observa que os maiores erros de predição dos modelos espectroscópicos estão associados a bibliotecas espectrais com maior variação ambiental e espectral, conforme amplamente discutidos nos itens 6.31, 6.3.2 e 6.3.3. Modelos mais acurados foram obtidos a partir da estratificação de BE em grupos de amostras mais homogêneas com relação à região fisiográficas, usos da terra, classes texturais e classes espectrais, confirmando nossa hipótese. Da mesma forma, Grunwald et al. (2015) comenta que não há como transferir os modelos desenvolvidos de um local (região) para outro, visto a diversidade pedológica, ou seja, os modelos devem ser regionais e considerar amostras locais e suas características. No entanto, a estratificação ganha mais importância quando as predições obtidas serão utilizadas em projetos de escala mais detalhada, por exemplo, projetos de agricultura de precisão e em laboratórios de análises de solo para fins de recomendação. Quando o objetivo de uso das predições COS for menos detalhista, modelos globais podem atender a acurácia requerida.

Por fim, o desenvolvimento de um modelo espectroscópico robusto para predição do teor de COS para solos tropicais, depende das características pedológicas e espectrais das amostras utilizadas na calibração dos modelos. Os fatores que tiveram maior efeito no desempenho dos modelos foram a composição granulométrica e mineralogia do solo, sendo isso reflexo dos fatores de formação do solo atuantes em cada uma das três regiões do sul do Brasil estudas. Nesse sentido, os resultados encontrados mostraram as principais características pedológicas que influenciaram no comportamento espectral, consequentemente na modelagem espectroscópica, podendo esses fatores auxiliar na tomada de decisão para compor grupos de amostras em bibliotecas espectrais e também orientar posteriores designer amostral para predição do COS em nível regional e nacional para solos tropicais.

6.4. CONCLUSÃO

A análise dos espectros de solo indicou presença de alta variação espectral, decorrente da variação pedológica e ambiental da região sul do Brasil. As principais variações nos espectros ocorrem nas regiões de 400-800, 1400, 1900 e 2000-2350 nm.

A técnica de pré-processamento CRR associada ao método de calibração Cubist tiveram o melhor desempenho para todos os modelos, exceto para alguns conjuntos de dados com alta

variação espectral, em que a técnica SGD associada ao método PLSR alcançou maiores acurácias.

A estratificação BEG revelou que agrupar amostras mais homogêneas em relação às regiões fisiográficas, classes texturais, usos da terra e classes espectrais é uma estratégia para alcançar melhor acurácia para as predições de COS em comparação ao modelo global.

O modelo global apresentou RMSE = 0,76 %, SDE = 0,61 % e RPIQ = 1,90. A maior acurácia foi observada nos modelos regionais (RMSE = 0,66 %; SDE = 0,44 % e RPIQ = 2,65) e usos da terra (RMSE = 0,73 %; SDE = 0,54 % e RPIQ = 2,14).

A redução no número de amostras devido à estratificação foi um fator que afetou negativamente no desempenho dos modelos, principalmente para grupos com alta variação pedológica e espectral, portanto, recomendamos um maior número de amostras para obter predições mais acuradas para COS para solos com essas características.

A representatividade de amostras de uma região em detrimento de outra influenciou na acurácia dos modelos para cada região individualmente. E como já esperado, modelos calibrados com amostras de uma determinada região não se aplicam para predizer o COS em outras regiões, onde os fatores de formação do solo atuantes são muito distintos, confirmando que os modelos de predição devem ser regionais.

7. CONSIDERAÇÕES FINAIS

Os resultados encontrados nesse trabalho mostram que a técnica de espectroscopia de reflectância difusa na região Vis-NIR tem potencial para quantificar concentrações de COS em solos subtropicais com altos níveis de óxidos de ferro, argila e COS, em escala local e regional.

A estratificação das bibliotecas espectrais com base em características do solo, do ambiente (regiões fisiográficas) e características espectrais (reflectância espectral do solo) revelou que a separação e/ou agrupamento de amostras mais homogêneas, especialmente em relação às regiões e uso da terra, é uma estratégia para alcançar melhores predições de COS em comparação ao uso de todas as amostras juntas na calibração do modelo.

A presença de amostras com textura grossa prejudicou as predições do COS, onde predições mais acuradas são alcançadas em solos de textura argilosa ou mais fina. No entanto, para alguns modelos, a redução no número de amostras em função da estratificação reduziu o desempenho dos modelos, atentando que é preciso ter cuidado ao trabalhar com conjuntos pequenos de amostras.

A estratificação ganha mais importância quando aos resultados da predição serão utilizadas em projetos de escala mais detalhada, por exemplo, projetos de agricultura de precisão e em laboratórios de análises de solo para fins de recomendação. Por outro lado, se uso das predições COS for menos detalhista, modelos globais podem atender a acurácia requerida.

Modelos calibrados com amostras de uma determinada região não se aplicam para predizer o COS em outras regiões, nas quais os fatores de formação do solo atuantes são muito distintos daqueles que atuam nas amostras utilizadas na calibração do modelo, confirmando que os modelos de predição devem ser regionais nessas circunstâncias.

Esses resultados reiteram que o uso de bibliotecas espectrais para a previsão do COS necessita de estudo prévio sobre o agrupamento do conjunto de amostras. Tal agrupamento pode ser realizado por análises de seus próprios espectros, que refletem o conjunto de características físicas, químicas e mineralógicas do solo. Além disso, o estudo do comportamento espectral do solo pode auxiliar no planejamento da coleta de espectros em escala de campo, buscando otimizar a amostragem, para potencializar a calibração dos modelos preditivos.

8. REFERÊNCIAS

AL-ABBAS, A.H.; SWAIN, P.H.; BAUMGARDNER, M.F. Relating organic matter and clay content to the multispectral radiance of soils. **Soil Science**, v.114, p.477–485, 1972.

ALMEIDA, J.A. Solos das pradarias mistas do sul do Brasil. In: CURI, N. et al. **Pedologia**: Solos dos Biomas Brasileiros.1.ed. Viçosa, MG: Sociedade Brasileira de Ciência do Solo, 2017. cap.9, p.353-406.

ALMEIDA, J.A.; MAÇANEIRO, K.C.; KLAMT, E. Mineralogia da fração argila de solos vermelhos com horizontes superficiais brunados do Planalto de Lages (SC). **Revista Brasileira de Ciência do Solo**, v.24, p.815-828, 2000.

ALVARES, C. A. et al. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v.22, n.6, 711-728, 2013.

AMUNDSON, R.et al. Soil and human security in the 21st century. **Science**, v.348, p.647-654, 2015.

ÅNGSTRÖM, A. The albedo of various surfaces of ground. **Geografiska Annaler**, v.7, p.323-342, 1925.

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**, v.65, n.5, p.718-729, 2014.

ARROUAYS, D. et al. **GlobalSoilMap**: Basis of the Global Spatial Soil Information System. CRC Press/Balkema, 2014.

BARNES, R.J; DHANOA, M.S.; LISTER, S.J. Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra. **Applied Spectroscopy**, v.43, p.772–777, 1989.

BAUMGARDNER, M.F. et al. Reflectance properties of soils. **Advances in Agronomy**, v.38, p.2–44, 1985.

BELLINASO, H. **Biblioteca espectral de solos e sua aplicação na quantificação de atributos e classificação**. 2009. 264p. Dissertação (Mestrado em Agronomia, Área de Concentração em Solos e Nutrição de Plantas). Universidade de São Paulo, Piracicaba, SP, 2009.

BELLINASO, H.; DEMATTÊ, J. A.M.; ROMEIRO, S.A. Soil spectral library and its use in soil classification. **Revista Brasileira de Ciência do Solo**, v.34, n.3, p.861-870, 2010.

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

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

BEN-DOR, E.; GRANOT, A.; NOTESCO, G. A simple apparatus to measure soil spectral information in the field under stable conditions. **Geoderma**, v.306, p.73-80, 2017.

BEN-DOR, E. et al. Using imaging spectroscopy to study soil properties. **Remote Sensing of Environment**, v.113, p.538-555, 2009.
BEN-DOR, E.; HELLER, D.; CHUDNOVSKY, A. A novel method of classifying soil profiles in the field using optical means. **Soil Science Society American Journal**, v.72, p.1113–1123, 2008.

BEN-DOR, E. et al. **Soil reflectance**. In: RENCZ, A.N. Remote sensing for the earth sciences. New York : J. Wiley & Sons, 1999. p.111-188.

BEN-DOR, E. et al. The reflectance spectra of organic matter in the visible near-infrared and short wave infrared region (400-2500) during a controlled decomposition process. **Remote Sensing of Environment**, v.61, p.1-15, 1997.

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

BISHOP, J.L.; LANE, M.D.; DYAR, M.D.; BROWN, A.J. Reflectance and emission spectroscopy study of four groups of phyllosilicates: smectites, kaolinite– serpentines, chlorites and micas. **Clay Minerals**, v.43, p.35– 54, 1994.

BODDEY, R.M. et al. Carbon accumulation at depth in Ferralsols under zero-till subtropical agriculture. **Global Change Biology**, v.16, n.2, p.784-795, 2010.

BORTOLUZZI, E.C.; RHEINHEIMER,D.S.; KAMINSKI, J.; GATIBONI, L.C.; TESSIER, D. Alterações na mineralogia de um argissolo do Rio Grande do Sul submetido à fertilização potássica. **Revista Brasileira de Ciência do Solo**, v.29, n.3, p.327-335, 2005.

BOX, G.E.P.; COX, D.R. An Analysis of Transformations. **Journal of the Royal Statistical Society**, v.26, p.211-252, 1964.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Secretaria de Política Agrícola. **Instrução Normativa nº. 2**, de 9 de out. de 2008. Diário Oficial da União, n.193, Seção 1, p.5, out. de 2008. [Acessado em: 15 de ago. 2015]

BRASIL. Ministério da Agricultura. **Levantamento de reconhecimento dos solos do estado do Rio Grande do Sul**. Recife: MA/DPP-AS/DRNR, 1973. 431p. (Boletim técnico, 30)
BREIMAN, L. Random forests. **Machine Learning**, v.45, p.5–32, 2001.

BOWERS, S.; HANKS, R. Reflection of radiant energy from soils. **Soil Science**, v.100(2), p.130 –138, 1965.

BROWN, D. J.; BRICKLEMYER, R. S.; MILLER, P. R. Validation requirements for diffuse reflectance soil characterization models with a case study of vis–NIR soil C prediction in Montana. **Geoderma**, n.129, v.251–267, 2005.

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

BRUS, D.J., KEMPEN, B., HEUVELINK, G.B.M. Sampling for validation of digital soil maps. **European Journal of Soil Science**, v.62, p.394–407, 2011.

CAMBULE, A.H. et al. Building a near infrared spectral library for soil organic carbon estimation in the Limpopo National Park, Mozambique. **Geoderma**, v.183, p.41-48, 2012.

CARTER, W.T. Color analysis of soils with spectrophotometer. **American Soil Survey Association Bull**, B12, 169 –170, 1931.

CARVALHO JÚNIOR, W. et al. Método do hipercubo latino condicionado para a amostragem de solos na presença de covariáveis ambientais visando o mapeamento digital de solos. **Revista Brasileira de Ciência do Solo**, v.38, p.386-396, 2014.

CARVALHO, J.L.N. et al. Potencial de sequestro de carbono em diferentes biomas do Brasil. **Revista Brasileira de Ciência do Solo**, v.34, p.277-289, 2010.

CHANG, C. W.; LAIRD, D.A. Near-infrared reflectance spectroscopic analysis of soil C and N. **Soil Science**. v.167, p.110–116, 2002.

CHONG, I.G.; JUN, C.H. Performance of some variable selection methods when multicollinearity is present. **Chemometrics and Intelligent Laboratory Systems**, v.78, p.103–112, 2005.

CLAIROTTE, M. et al. National calibration of soil organic carbon concentration using diffuse infrared reflectance spectroscopy. **Geoderma**, v.276, p.41 –52, 2016.

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

CLARK, R.N. et al. High spectral resolution reflectance spectroscopy of minerals. **Journal of Geophysical Research**, v.95, p.12653–12680, 1990.

CLARK, R.N. Spectroscopy of rocks and minerals, and principles of spectroscopy. In: Rencz N (ed) **Remote sensing for the earth sciences: manual of remote sensing**. Wiley, New York, NY, 1999. p.3–52.

COLEMAN, T.L. et al. Spectral band selection for quantifying selected properties in highly weathered soils. **Soil Science**, v.151, p.355-361, 1991.

DALAL, R. C.; HENRY, R. J. Simultaneous determination of moisture, organic carbon, and total nitrogen by near infrared reflectance spectrophotometry. **Soil Science Society American Journal**, v.50, p.120–123, 1986.

DALMOLIN, R.S.D. et al. Solos do Planalto das Araucárias. In: CURI, N. et al. **Pedologia**: Solos dos Biomas Brasileiros.1.ed. Viçosa, MG: Sociedade Brasileira de Ciência do Solo, 2017. cap.8, p.353-406.

DALMOLIN, RSD; TEN CATEN, A. Mapeamento Digital: nova abordagem em levantamento de solos. **Investigación Agraria**, 17(2):77-86, 2015.

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

DALMOLIN, R. S. D. et al. Organic matter characteristics and distribution in Ferralsol profiles of a climosequence in southern Brazil. **European Journal of Soil Science**, v. 57, p. 644- 654, 2006.

DALMOLIN, R.S.D. **Matéria orgânica e características físicas, químicas, mineralógicas e espectrais de Latossolos de diferentes ambientes**. 2002. 151f. Tese (Doutorado em Ciência do Solo) - Programa de Pós-graduação em Ciência do Solo, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 2002.

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**, v.288, p.8–22, 2017.

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

DEMATTÊ, J.A.M.; TERRA, F.S. Spectral pedology: A new perspective on evaluation of soils along pedogenetic alterations. **Geoderma**, v.217–218, p.190–200, 2014.

DEMATTÊ, A.M.; BELLINASO, H., ROMERO, D.J.; FONGARO, C.T. Morphological Interpretation of Reflectance Spectrum (MIRS) using libraries looking towards soil classification. **Scientia Agricola**, v.71, n.6, p.509-520, 2014.

DEMATTÊ, J.A.M.; GARCIA, G.J. Alteration of soil properties through a weathering sequence as evaluated by spectral reflectance. **Soil Science Society America Journal**, v.63, n.2, p.327-342, 1999.

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**, v.314, p.262-274, 2018.

DOTTO, A.C. et al. Two preprocessing techniques to reduce model covariables in soil property predictions by Vis-NIR spectroscopy. **Soil & Tillage Research**, v.172, p.59–68, 2017.

DOTTO, A.C. et al. Mapeamento digital de atributos: granulometria e matéria orgânica do solo utilizando espectroscopia de reflectância difusa. **Revista Brasileira de Ciência do Solo**, v.38, p.1663-1671, 2014.

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

EPIPHANIO, J.C.N. et al. **Comportamento espectral de solos do Estado de São Paulo**. São José dos Campos, Instituto Nacional de Pesquisas Espaciais, 1992. 131p.

FAO; ITPS. 2015. **Status of the World's Soil Resources (SWSR) – Main Report**. Food and Agriculture Organization of the United Nations and Intergovernmental Technical Panel on Soils, Rome, Italy.

FIORIO, P.R.; DEMATTÊ, J.L.M. Orbital and laboratory spectral data to optimize soil analysis. **Scientia Agricola**, v.66, n.2, p.250-257, 2009.

FLORINSKY, I. V. The Dokuchaev hypothesis as a basis for predictive digital soil mapping (on the 125th anniversary of its publication). **Eurasian Soil Science**, v.45, p.445-451, 2012.

FLORES, C. A. et al. **Levantamento semidetalhado de solos, Município de Santa Maria, Rio Grande do Sul.** Zoneamento edáfico de culturas para o município de santa maria/rs. 1ed. Brasília: Embrapa, 2015, v. 1, p. 11-79.

FORMAGGIO, A.R. **Comportamento espectral de quatro solos do Estado de São Paulo nos níveis orbital, de campo e de laboratório**. São José dos Campos, INPE, 1983. 90p. (Tese de Mestrado).

FORTES, A.B. Zonas fisiográficas. 1959. In: **Geografia física do Rio Grande do Sul**. Porto Alegre, Globo. p.133.

FOX, J.; WEISBERG, S. **An R Companion to Applied Regression**, 2011. Sage Publication.

GALVÃO, L.S.; VITORELLO, I. Role of organic matter in obliterating the effects of iron on spectral reflectance and colour of Brazilian tropical soils**. International Journal Remote Sensing**, v.19, n10, p.1969-1979, 1998.

GIASSON, E. et al. Digital soil mapping using multiple logistic regression on terrain parameters in Southern Brazil**. Scientia Agricola**, v.63, n.3, p.262-268, 2006.

GLOBAL SOIL PARTNERSHIP. Global Soil Organic Carbon Map – Leaflet. Roma: FAO, 2017. 5 p. disponível em: <http://www.fao.org/3/i8195en/I8195EN.pdf>. Acesso em: 25 de abril de 2018.

GOGÉ, F. et al. Which strategy is best to predict soil properties of a local site from a national Vis–NIR database? **Geoderma**, v.213, p.1–9, 2014.

GRUNWALD, S. et al. Digital Soil Mapping in a Changing World. In: MINASNY, B.; MALONE, B.P.; MCBRATNEY, A.B. (Eds). **Digital Soil Assessments and Beyond**: Proceedings of the 5th Global Workshop on Digital Soil Mapping. CRC Press, Sydney, Australia. 2012.

GRUNWALD, S.; VASQUES, G.M.; RIVERO, R.G. Fusion of soil and remote sensing data to model soil properties. In: Sparks, D.L. (Ed.), **Advances in Agronomy**, v.131, p.1–109, 2015.

GRUWALD, S. Multi-criteria characterization of recent digital soil mapping and modeling approaches. **Geoderma**, v.152, p.195-207, 2009.

GUERRERO, C. et al. Do we really need large spectral libraries for local scale SOC assessment with NIR spectroscopy? **Soil & Tillage Research**, v.155, p.501–509, 2016.

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,** v.60, p.1193-1202, 1988.

HAN, L. et al. Some concepts of soil organic carbon characteristics and mineral interaction from a review of literature. **Soil Biology and Biochemistry**, v.94, p.107-121, 2016.

HARTEMINK, A.E. Soils are back on the global agenda. **Soil Use and Management**, v.24, p.327–330, 2008.

HENGL, T. et al. SoilGrids250m: Global gridded soil information based on machine learning. **PLoS One**, 12, e0169748, 2017.

HENGL, T. **Pedometric mapping**. PhD-thesis, ITC, Enschede, 2003.

HEUVELINK, G. The definition of pedometrics. **Pedometron,** v.15, p.11-12, 2003.

HONG, Y. et al. Prediction of Soil Organic Matter by VIS–NIR Spectroscopy Using Normalized Soil Moisture Index as a Proxy of Soil Moisture. **Remote Sensing**, v.10, n.28, 2018.

HORST, T.Z. **Variação de parâmetros dendrométricos de pinus taeda l. e a distribuição espacial de atributos do solo por técnicas de mapeamento digital de solos**. 2017. 122p. Dissertação (Mestrado em Ciência do Solo). Universidade Federal de Santa Maria, Santa Maria, RS, 2017.

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

HUNT, G.R.; SALISBURY, J.W., 1970. Visible and near-infrared spectra of minerals and rocks. I. Silicate minerals. **Modern Geology**, v1, p.283 –300, 1970.

HUNT, G.R.; SALISBURG, J.W; LENHOFF, C.J. Visible and near-infrared spectra of minerals and rocks: III Oxides and hidroxides. **Modern Geology**, v.1, p.195-205, 1971.

IVANCIUC, O. **Applications of Support Vector Machines in Chemistry**. In: Reviews in Computational Chemistry, Volume 23, Eds.: K. B. Lipkowitz and T. R. Cundari. Wiley-VCH, Weinheim, p. 291–400, 2007.

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

JANIK, L.J.; SKJEMSTAD, J. Characterization and analysis of soils using midinfrared partial least-squeres. .2. Correlations with some laboratory data. **Australian Journal of Soil Research**, v.33(4), p.637 – 650, 1995.

JELIHOVSCHI, E.G. et al. ScottKnott: a package for performing the scott-knott clustering algorithm in R**. Trends in Applied and Computational Mathematics**, v.15, n.1, p.3-17, 2014.

JENNY, H. **Factors of soil formation - a system of quantitative pedology**. New York: Dover Publications, 1941.

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**, v.293, p.54–63, 2017.

KER, J.C.; RESENDE, M. Caracterização química e mineralógica de solos brunos subtropicais do Brasil. **Revista Brasileira de Ciência do Solo**, v.14, p.215-225, 1990.

KLEBER, M.; SOLLINS, P.; SUTTON, R. A conceptual model of organo-mineral interactions in soils: self-assembly of organic molecular fragments into zonal structures on mineral surfaces. **Biogeochemistry**, v.85, p.9-24, 2007.

KLEBER, M.; JOHNSON, M.G. Advances in understanding the molecular structure of soil organic matter: implicattions for interactions in the environment. In: Sparks, D.L. (Ed.), **Advances in Agronomy**, v.106, cap.3, p.77-142, 2010.

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

KÖCHY, M., HIEDERER, R., FREIBAUER, A. Global distribution of soil organic carbon, based on the Harmonized World Soil Database – Part 1: Masses and frequency distribution of SOC stocks for the tropics, permafrost regions, wetlands, and the world. **SOIL Discussions**, v.1, p.327–362, 2014.

KOOISTRA, L. et al. Possibilities of visible-near-infrared spectroscopy for the assessment of soil contamination in river floodplains. **Analytica Chimica Acta**, v.446, p.97–105, 2001.

KRISHNAN, P. et al. Reflectance technique for predicting soil organic matter. **Soil Science Society American Journal**, v.44, p.1282–1285, 1980.

KUHN, M. **Caret:** Classification and Regression Training. R package version 6.0-73, 2016. URLhttps://CRAN.R-project.org/package=caret.

KÜHNEL, A.; BOGNER, C. In-situ prediction of soil organic carbon by vis–NIR spectroscopy: an effiient use of limited fild data. **European Journal Soil Science**, v.68, p.689-702, 2017.

LAGACHERIE, P.; GOMEZ, C. Vis-NIR-SWIR Remote Sensing Products as New Soil Data for Digital Soil Mapping. In: MCBRATNEY, A.; MINASNY, B.; STOCKMANN, U. (Eds): **Pedometrics: Progress in Soil Science**. Springer, Cham, Switzercap.13, p.77-415-436, 2018.

LAGACHERIE, P.; MCBRATNEY, A.B. Spatial soil information systems and spatial soil inference systems: Perspectives for digital soil mapping. In: LAGACHERIE, P.; MCBRATNEY, A.; VOLTZ, M., (Eds): **Digital soil mapping: An introductory perspective**. Amsterdam, Elsevier. p.3-22, 2007.

LAL, R; SMITH, P.; JUNGKUNST, H.F. et al. The carbon sequestration potential of terrestrial ecosystems. **Journal of Soil and Water Conservation**, v.73(6), p.145A-152A, 2018.

LEVENE, H. Robust Test for Equality of Variances, in I. O. et al., ed., '**Contributions to Probability and Statistics**: Essays in Honor of Harold Hotteling', Stanford University Press, California, United States, p. 278–292, 1960.

LILLESAND, T.M.; KIEFER, R.W. **Remote Sensing and Image Interpretation**. 3. ed. Edition, John Wiley and Sons, Inc., Hoboken, 1994, 176p.

LOSS, A. et al. Frações granulométricas e oxidáveis de matéria orgânica sob diferentes sistemas de uso do solo, no Paraná, Brasil. **Bioscience Journal**, v.30, n.1, p.43-54, 2014.

LUCÀ, F.; CONFORTI, M.; CASTRIGNANÒ, A.; MATTEUCCI, G.; BUTTAFUOCO, G. Effect of calibration set size on prediction at local scale of soil carbon by Vis NIR spectroscopy. **Geoderma**, v.288, p.175-183, 2017.

MACIEL FILHO, C.L.**Carta geotécnica de Santa Maria**. Santa Maria: UFSM, 1990.

MADEIRA NETTO, J.S. **Etude quantitative des relations constituants minéralogiques - reflectance diffuse des latosols brésiliens. Applications à l'utilisation pedologique des données satellitaires TM (région de Brasilia)**. 1991. 238f. Thèse (Doctorat en Science du Sol) - Université Pierre et Marie Curie, UFR des Sciences de la Terre, Paris.

MapBiomas - ano 2017 (http://mapbiomas.org/)

MASSERSCHMIDT, I.; CUELBAS, C.J.; POPPI, R.J.; ANDRADE, J.C.; ABREU, C.A.; DAVANZO, C.U. Determination of organic matter in soils by FTIR/diffuse reflectance and multivariate calibration. **Journal of Chemometrics**, v.13, p.265–273, 1999.

MATHEWS, H.L. et al. Spectral reflectance of selected Pennsylvania soils. **Soil Science Society of America Proceedings**, v.37, p.421-424, 1973.

McBRATNEY, A.; FIELD, D.J.; KOCH, A. The dimensions of soil security. **Geoderma**, v.213, p.203–213, 2014.

MCBRATNEY, A.B.; MENDONÇA-SANTOS, M.; MINASNY, B. On digital soil mapping. **Geoderma**, v.117, p.3-52, 2003.

MCBRATNEY, A.B.; MINASNY, B.; VISCARRA ROSSEL, R. Spectral soil analysis and inference systems: A powerful combination for solving the soil data crisis, **Geoderma**, v.136, p.272-278, 2006.

McCARTY, G.W. et al. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. **Soil Science Society American Journal**, v.66, p.640–646, 2002.

MCDOWELL, M.L. et al. Soil total carbon analysis in Hawaiian soils with visible, near-infrared and mid-infrared diffuse reTflectance spectroscopy. **Geoderma**, v.189–190, p.312–320, 2012.

MCKENZIE, N. et al. The role of terrain analysis in soil mapping. In: WILSON, J.; GALLANT, J. (Eds.). **Terrain analysis**: principles and applications, New York: John Wiley and Sons, p.245-266, 2000.

MENDONÇA-SANTOS, M.L. et al. Modelagem e Mapeamento Digital de Estoque de Carbono Orgânico na Camada Superficial dos Solos (0-10 cm) do Estado do Rio de Janeiro. **Embrapa**, Boletim de pesquisa e desenvolvimento, p.23, 2007.

MENDONÇA-SANTOS, M.L.; SANTOS, H.G. Mapeamento Digital de Classes e Atributos de Solos. **Embrapa**, p.17, 2003.

MEVIK, B.H. **VIP.R:** Implementation of VIP (Variable Importance in Projection) for the 'pls' Package. Available online: http://mevik.net/work/software/VIP.R (accessed on 26 October 2016).
MEYER, D. et al. **e1071:** Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.6-8, 2017. https://CRAN.R-project.org/package=e1071

MIGUEL, P. et al. Solo e dinâmica de ocupação das terras em áreas do rebordo do planalto do Rio Grande do Sul. **Revista Brasileira de Agrociência** (UFPEL), v.17, n.4, p.447-455, 2011.

MINASNY, B.; McBRATNEY, A.B. A conditioned Latin hypercube method for sampling in the presence of ancillary information. **Computers & Geosciences**, v.32, p.1378-1388, 2006.

MINASNY, B.; MCBRATNEY, A.B.; SALVADOR-BLANES, S. Quantitative models for pedogenesis — a review. **Geoderma**, v.144, p.140–157, 2008.

MORGAN, C.L.S. et al. Simulated in situ characterization of soil organic and inorganic carbon with visible near-infrared diffuse reflectance spectroscopy. **Geoderma**, v.151, p.249–256, 2009.

MOURA-BUENO, J.M. 2014. **Modelos digitais de elevação e predição do carbono orgânico do solo no Planalto do Estado do Rio Grande do Sul**. 2014. 108p. Dissertação (Mestrado em Ciência do Solo). Universidade Federal de Santa Maria, Santa Maria, RS, 2014.

MOURA-BUENO, J.M. et al. Erosão em áreas de encosta com solos frágeis e sua relação com a cobertura do solo. **Revista Scientia Agrária**, v.19, n.1, p.102-112, 2018.

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

NELSON, D.W.; SOMMERS, L.E. A rapid and accurate procedure for estimation of organic carbon in soil. **Proceedings of Indiana Academy of Science**, v.84, p.456-462, 1975.

NELSON, D.W.; SOMMERS, L.E. Total carbon, organic carbon and organic matter. In: SPARKS, D.L.; PAGA, A.L.; HELMKE, P.A.; LOEPPERT, R.H.; SOLTANPOUR, P.N.; TABATABAI, M.A.; JOHNSTON, C.T. & SUMMER, M.E., eds. **Methods of soil analysis: Chemical methods. Part 3**. Madison, Soil Science Society of America, 1996. p.961-1010.

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**, v.68, p.337-347, 2014.

NOCITA, M. et al. Soil spectroscopy: An alternative to wet chemistry for soil monitoring. **Advances in Agronomy**, v.132, p.139-159, 2015.

OBUKHOV, A.I.; ORLOV, D.S. Spectral reflectivity of the major soil groups and possibility of using diffuse re flection in soil investigations. **Soviet Soil Science**, v.2, p.174 –184, 1964.

O'NEAL, A.M. The effect of moisture on the color of certain Iowa soils. **American Soil Survey Association Bull**, B.8, p.158 –174, 1927.

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

OLIVEIRA, A.H. et al. Caracterização ambiental e predição dos teores de matéria orgânica do solo na sub-Bacia do Salto, Extrema, MG. **Ciências Agrárias**, v.33, n.1, p.143-154, 2012.

PALACIOS-ORUETA, A., USTIN, S.L. Remote sensing of soil properties in the Santa Monica Mountains I. Spectral analysis. **Remote Sensing of Environment**, v.65, p.170–183, 1998.

PINHEIRO, E.F.M. et al. Prediction of Soil Physical and Chemical Properties by Visible and Near-Infrared Diffuse Reflectance Spectroscopy in the Central Amazon. **Remote Sensing**, v.9, n.4, p.293, 2017.

POLIDORO, J.C. et al. Programa Nacional de Solos do Brasil (PronaSolos). Dados eletrônicos. – Rio de Janeiro: Embrapa Solos, 2016. 53p.

POTTER, R.O.; CARVALHO, A.P.; FLORES, C.A.; BOGNOLA, I. **Solos do Estado de Santa Catarina**. Embrapa Solos, 2004. Embrapa Solos. Boletim de Pesquisa e Desenvolvimento; n. 46. 1 CD-ROM.

PRIBYL, D.W. "A critical review of the conventional SOC to SOM conversion factor", **Geoderma**, v.156, p.75, 2010.

QUINLAN, J. Learning with continuous classes. In: Adams, A., Sterling, L. (Eds.), Proceedings AI'92, 5th Australian Conference on Artificial Intelligence. **World Scientific**, Singapore, pp. 343–348, 1992.

R CORE TEAM (2018). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

RAMÍREZ-LÓPEZ, L. et al. Sampling optimal calibration sets in soil infrared spectroscopy, **Geoderma**, v.226–227, p.140–150, 2014.

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

RAMOS, P.V. et al. Magnetic Susceptibility of Soil to Differentiate Soil Environments in Southern Brazil. **Revista Brasileira de Ciência do Solo**, v. 41, p.e0160189, 2017.

RHEINHEIMER, D.S.; CAMPOS, B.C.; GIACOMINI, S.J.; CONCEIÇÃO, P.C.; BORTOLUZZI, E.C. Comparação de métodos de determinação de carbono orgânico total no solo. **Revista Brasileira de Ciência do Solo**, v.32, p.435-440, 2008.
RINNAN, A.; BERG, F.V.; ENGELSEN, S.B. Review of the most common pre-processing techniques for near-infrared spectra. **Trends in Analytical Chemistry**, v.28, p.1201–1222, 2009.

SAMUEL-ROSA, A. et al. Uso da terra no rebordo do Planalto do Rio Grande do Sul. **Revista Ciência e Natura**, v.33, p.161-173, 2011.

SANCHEZ, P.A.; AHAMED, S.; CARRE, F. et al. Digital soil map of the world. **Science**, v.325, p.680-681, 2009.

SANTANA, B.F.; SOUZA, A.M.; POPPI, R.J. Visible and near infrared spectroscopy coupled to random forest to quantify some soil quality parameters. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v.191, p.454 –462, 2018.

SANTOS, H.G. et al. **Sistema Brasileiro de Classificação de Solos**. 5. ed., rev. e ampl. – Brasília, DF: Embrapa, 2018. 356p.

SANTOS, R.D. et al. **Manual de descrição e coleta de solo no campo**. 7. ed. Viçosa: Sociedade Brasileira de Ciência do Solo, 2015. 100p.

SARTORI, P. Geologia e geomorfologia de Santa Maria.**Ciência e Ambiente**, v.39, p.19-42, 2009.

SATO, M.V. **Primeira aproximação da biblioteca espectral de solos do Brasil: caracterização de espectros de solos e quantificação de atributos**. 2015. 102f. Dissertação (Mestrado em Agronomia, Área de Solos e Nutrição de Plantas). Universidade de São Paulo, Piracicaba, SP, 2015.

SAVITZKY, A., GOLAY, M.J. Smooting and differentiation of data by simplified least squares procedures. **Analitical chemistry**, v.36, n.8, p.1627-1639, 1964.

SCHMITT, C. **Gênese de solos desenvolvidos de rochas sedimentares na depressão central do Rio Grande do Sul**. 2015. 172p. Tese (Doutorado em Agronomia do Programa de Pós-graduação em Ciência do Solo). Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 2015.

SCHOLLENBERGER, C.J. A rapid approximate method for determining soil organic matter. **Soil Science**, v.24, p.65-68, 1927.

SCHWARTZ, G., BEN-DOR, E., ESHEL, G. Quantitative analysis of total petroleum hydro-carbons in soils: comparison between reflectance spectroscopy and solvent extraction by 3 certified laboratories. **Applied Environmental Soil Science**, p.1-11, 2012.

SEQUEIRA, C.H.; WILLS, S.A.; GRUNWALD, S.; FERGUSON, R.R.; BENHAM, E.C.; WEST, L.T. Development and Update Process of VNIR-Based Models Built to Predict Soil Organic Carbon. **Soil Science Society of America Journal**, v.78, p.903–913, 2014.

SHEPHERD, K.D.; WALSH, M.G. Development of refl ectance spectral libraries for characterization of soil properties. **Soil Science Society of America Journal**, v.66, n.3, p.988–998, 2002.

SIX, J. et al. Soil Structure and Organic Matter: I. Distribution of Aggregate-Size Classes and AggregateAssociated Carbon. **Soil Science Society of America Journal**, v.64, n.2, p.681 -689, 2000.

SORENSEN, M.V. et al. Draining the Pool? Carbon Storage and Fluxes in Three Alpine Plant Communities.**Ecosystems,** v.21, n.2, p.316–330, 2018.

SORIANO-DISLA, J.M. et al. The Performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. **Journal of Applied Spectroscopy**, v.49, p.139-186, 2014.

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**, v.70, p.610-621, 2011.

SOUZA, A.M.; FIGUEIRAS, P. R.; COELHO, M.R.; FONTANA, A.; WINKLER, T.C.B.; VALDERRAMA, P.; POPPI, RONEI JESUS. Validation of the near infrared spectroscopy method for determining soil organic carbon by employing a proficiency essay for fertility laboratories. **Journal of Near Infrared Spectroscopy**, v.24, p.293-303, 2016.

STENBERG, B.; VISCARRA ROSSEL, R.A.; MOUAZEN, A.M.; WETTERLIND, J. Visible and Near Infrared Spectroscopy in Soil Science. In Donald L. Sparks, editor: **Advances in Agronomy**, v. 107, chapter 5, Burlington: Academic Press, p.163-215, 2010.

STENBERG, B.; VISCARRA ROSSEL, R.A., 2010. Diffuse Reflectance Spectroscopy for High-Resolution Soil Sensing. In Viscarra Rossel, R.A., McBratney, A.B., Minasny, B. (Eds): **Proximal Soil Sensing, Progress in Soil Science**, v1, chapter 3, Springer, p.29-47, 2010.

STENBERG, B.; JONSSON, A.; BÖRJESSON, T. 2002. Near infrared technology for soil analysis with implications for precision agriculture. In: Davies, A.; Cho, R. (Eds**.): Near Infrared Spectroscopy: Proceedings of the 10th International Conference**, NIR Publications, Chichester, UK, p.279-284, 2002.

STEVENS, A.; NOCITA, M.; TÓTH, G.; MONTANARELLA, L.; WESEMAEL, B. van. Prediction of Soil Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy. **PLoS ONE**, v.8, n.6, e66409, 2013.

STOCKMANN, U. et al. Global soil organic carbon assessment. **Global Food Security**, v.6, p.9–16, 2015.

STONER, E.R. et al. **Atlas of soil reflectance properties**. West Lafayette : Purdue University, 1980. 75p.
STONER, E.R.; BAUMGARDNER, M.F. Characteristic variations in reflectance of surface soils. **Soil Science Society America Journal**, v.45, p.1161-1165, 1981.

STONER, E.R.; BAUMGARDNER, M.F. Data acquisition through remote sensing. In: BEATTY, M.T.; PETERSEN, G.M.; SWINDALE, L.D. (Eds.). Planning the uses and management of land. Madison: Soil Science Society of America, 1986, p.159-185.

STONER, E. et al. Discriminação espectral de Latossolos do Planalto Central brasileiro. **Pesquisa Agropecuária Brasileira**, Brasília, v.26, p.1599-1606, 1991.

STRECK, E.V. et al. **Solos do Rio Grande do Sul**. 3.ed. rev. e ampl. Porto Alegre: Emater/RS, 2018.

TEIXEIRA, P.C. et al. **Manual de métodos de análise de solos.** 3ª. ed. rev. e ampl. – Brasília, DF: Embrapa, 2017. 573 p.

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

TEN CATEN, A. et al. Regressões logísticas múltiplas: fatores que influenciam sua aplicação na predição de classes de solos. **Revista Brasileira de Ciência do Solo**, v.35, p.53-62, 2011.

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

TERRA, F.S.; DEMATTÊ, J.Á.; VISCARRA-ROSSEL, R.A. Spectral libraries for quantitative analyses of tropical Brazilian soils: Comparing vis–NIR and mid-IR reflectance data. **Geoderma**, v.255, p.81-93, 2015.

THOMÉ, V.M.R. et al. **Zoneamento Agroecológico e Socioeconômico de Santa Catarina**. Florianópolis: Epagri, 1999, v.1000. p.1000. CD-ROOM.

VALERIANO, M. M.; ROSSETTI, D. F. Topodata: Brazilian full coverage refinement of SRTM data. **Applied Geography**, v.32, n.2, p.300-309, 2011.

VAN ZIJL, G.M. et al. Functional digital soil mapping: A case study from Namarroi, Mozambique. **Geoderma**, v.219–220, p.155–161, 2014.

VARMUZA, K.; FILZMOSER, P. Introduction to multivariate statistical analysis in chemometrics. CRC Press (Taylor & Francis), Boca Raton, USA. 2009. 321p.

VAUDOUR, E. et al. Predicting Key Agronomic Soil Properties with UV-Vis Fluorescence Measurements Combined with Vis-NIR-SWIR Reflectance Spectroscopy: A Farm-Scale Study in a Mediterranean Viticultural Agroecosystem. **Sensors**, v.18, p.1157, 2018.

VASÁT, R. et al. Combining reflectance spectroscopy and the digital elevation model for soil oxidizable carbon estimation. **Geoderma**, v.303, p.133–142, 2017.

VASQUES, G.M. et al. Mapa de estoque de carbono orgânico do solo (COS) a 0-30 cm do Brasil. Embrapa Solos, RJ. 2017. URL http://geoinfo.cnps.embrapa.br/documents/1115/metadata_read

VASQUES, G.M. et al. Regional modelling of soil carbon at multiple depths within a subtropical watershed. **Geoderma**, v.156, p.326-336, 2010.

VASQUES, G.M.; GRUNWALD, S.; HARRIS, W.G. Spectroscopic Models of Soil Organic Carbon in Florida, USA. **Journal of Environmental Quality**, v.39, 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**, v.146(1-2), p.14–25, 2008.

VASQUES, G.M.; GRUNWALD, S.; SICKMAN, J.O. Visible/near-infrared spectroscopy modeling of dynamic soil carbon fractions. **Soil Science Society of American Journal**, v.73, p.176–184, 2009.

VISCARRA ROSSEL, R.A; HICKS, W.S. Soil organic carbon and its fractions estimated by visible – near infrared transfer functions. **European Journal of Soil Science**, v.66, p.438–450, 2015.

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

VISCARRA ROSSEL, R.A. et al. Proximal soil sensing: an effective approach for soil measurements in space and time. **Advances in Agronomy**, v.113, p.243–291, 2011.

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

VISCARRA ROSSEL, R.A.; MCBRATNEY, A.B. Diffuse reflectance spectroscopy as a tool for digital soil mapping. In: HARTEMINK, A. A.M.L.M.-S. (Ed.), **Digital Soil Mapping with Limited Data**. Springer, p.165-172, 2008.

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

VITORELLO, I.; GALVÃO, L.S. Spectral properties of geologic materials in the 400 to 2.500 nm range: Review of applications to mineral exploration and lithologic mapping. **Photo Interpretation**, v.34, p.77-99, 1996.

VOHLAND, M. et al. Determination of soil properties with visible to near- and mid-infrared spectroscopy: Effects of spectral variable selection. **Geoderma**, v.223–225, p.88–96, 2014.

WALKLEY, A.; BLACK, I.A. An examination of Degtjareff method for determining soil organic matter, and proposed modification of the chromic acid tritation method. **Soil Science**, v.37,p.29-38, 1934.

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

WIJEWARDANE, N.; GE, Y.; WILLS, S.; LOECKE, T. Prediction of Soil Carbon in the Conterminous United States: Visible and Near Infrared Reflctance Spectroscopy Analysis of the Rapid Carbon Assessment Project. **Soil Science Society of America Journal**, v.80, p.973–982, 2016.

WILDNER, W.; CAMOZZATO, E.; TONIOLO, J.A.; BINOTTO, R.B. IGLESIAS, C.M.F.; LAUX, J.H. **Mapa Geológico do Estado de Santa Catarina**. Porto Alegre: CPRM –Serviço Geológico do Brasil, 2014. Escala 1:500.000.

WILDNER, W.; RAMGRAB, G.E.; LOPES, R.C.; IGLESIAS, C.M.F. **Mapa Geológico do Estado do Rio Grande do Sul**. Porto Alegre: CPRM - Serviço Geológico do Brasil, 2008. Escala 1:750.000.

WOLSKI, M. S. et al. Digital soil mapping and its implications in the extrapolation of soil-landscape relationships in detailed scale. **Pesquisa Agropecuária Brasileira**, v. 52, p.633-642, 2017.

WOLSKI, M. S. **Modelagem do terreno e mapeamento digital de solos por extrapolação das relações solo-paisagem**. 2016. 255p. Tese. (Doutorado em Ciência do Solo). Universidade Federal de Santa Maria, Santa Maria, RS, 2016.

XIAOBO, Z. et al. Variables selection methods in near-infrared spectroscopy. **Analytica Chimica Acta**, v.667, p.14–32, 2010.

XIONG, X. et al. Assessing uncertainty in soil organic carbon modeling across a highly heterogeneous landscape. **Geoderma**, v.251–252, p.105–116, 2015.

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

ZHU, Q.; LIN, H.S.; DOOLITTLE, J.A. Functional soil mapping for site specific soil moisture and crop yield management. **Geoderma**, v.200–201, p.45–54, 2013.

9. APÊNDICES

APÊNDICE A – DESCRIÇÃO MORFOLÓGICA DO PERFIL 1 – ÁREA DO ESTUDO 1

DESCRIÇÃO GERAL – Perfil 1 do ESTUDO 1

PERFIL Nº – 1
DATA – 02/11/2017
CLASSIFICAÇÃO – Latossolo Vermelho Distroférrico típico
LOCALIZAÇÃO, MUNICÍPIO, ESTADO E COORDENADAS – Região Campos dos Carvalhos, estrada sentido municípios Giruá - Senador Salgado Filho, aproximadamente 4,2 km da RS 344. Giruá, Rio Grande do Sul, 28º 01’ 23,01” S; 54º 24’ 28,78” W – Sirgas 2000.
SITUAÇÃO, DECLIVE E COBERTURA VEGETAL SOBRE O PERFIL – Descrito e coleta em barranco de estrada no terço superior de coxilha, com declive de 6%, sob lavoura com cultura anual de soja.
ELEVAÇÃO –.415 metros.
LITOLOGIA – Basalto
FORMAÇÃO GEOLÓGICA – Formação Serra Geral, Grupo São Bento.
PERÍODO – Jurássico-Cretácio, era Mesozóico
MATERIAL ORIGINÁRIO – Produto de alteração de rochas do tipo basalto.
PEDREGOSIDADE – Ausente
ROCHOSIDADE – Ausente
RELEVO LOCAL – Suave ondulado
RELEVO REGIONAL – Ondulado e suave ondulado
EROSÃO – Laminar
DRENAGEM – Bem drenado.
VEGETAÇÃO PRIMÁRIA – Campo subtropical
USO ATUAL – Lavoura com cultura anual (soja).
CLIMA – Cfa 2 da classificação de Köppen.
DESCRITO E COLETADO POR – Jean Michel Moura Bueno e Charles Nicholas Moura Bueno.

DESCRIÇÃO MORFOLÓGICA – Perfil 1 do ESTUDO 1

A_p 0 – 15 cm; bruno-avermelhado-escuro (5 YR 3/4, úmido), bruno-avermelhado-escuro (5 YR 3/4, seco); argilosa; moderada, pequena e média, granular e blocos subangulares; poroso; friável, ligeiramente plástico e ligeiramente pegajoso; transição plana gradual; raízes muitas.

AB 15 – 40 cm; vermelho-escuro (2,5 YR 3/4, úmido), vermelho-escuro (2,5 YR 3/6, seco); muito argilosa; moderada, pequena e média, blocos subangulares; poroso; friável, ligeiramente plástico e ligeiramente pegajoso; transição plana difusa; raízes poucas.

BA 40 – 80 cm; vermelho-escuro (2,5 YR 3/6, úmido), vermelho-escuro (2,5 YR 3/4, seco); muito argilosa; moderada, pequena e média, blocos subangulares; poroso; friável, plástico e pegajoso, transição plana difusa; raízes raras.

Bw1 80 – 140 cm; vermelho-escuro (10 R 3/6, úmido), vermelho-escuro (10 R 4/4, seco); muito argilosa; moderada, pequena e média, blocos subangulares; poroso; friável, plástico e pegajoso; transição plana difusa; raízes raras.

Bw2 140 – 200 cm; vermelho-escuro (10 R 3/6, úmido), vermelho (10 R 4/6, seco); muito argilosa; moderada, pequena e média, blocos subangulares; poroso; friável, plástico e pegajoso; transição plana difusa; raízes raras.

APÊNDICE B – CARACTERIZAÇÃO ANALÍTICA – Perfil 01 do ESTUDO 1

Análises físicas								
Horizonte		Composição Granulométrica da Terra Fina (g kg^{-1})			Argila dispersa em água (g kg^{-1})	Grau de floculação (%)	Relação Silte/Argila	Ds (g cm^{-3})
Símbolo	Profundidade (cm)	Areia	Silte	Argila				
Ap	0 -15	64	258	678	267	61	0,4	1,28
AB	15 - 40	40	213	746	287	62	0,3	1,32
BA	40 - 80	24	166	810	13	98	0,2	1,15
Bw1	80 - 140	20	126	853	12	99	0,1	108
Bw2	140 - 200	19	119	862	12	99	0,1	1,05

Análises químicas																
Hz.	pH H_2O	Complexo sortivo ($cmol_c$ kg^{-1})									V (%)	Al (%)	P (mg kg^{-1})	MO (%)	CO (g kg^{-1})	Fe_2O_3 (g kg^{-1})
		Ca^{2+}	Mg^{2+}	K^+	Na^+	S	Al^{3+}	H+Al	CTC_{ef}	CTC_{pH7}						
Ap	5,2	4,2	2,1	0,2	0,0	6,5	0,5	4,2	7,0	10,7	61	7	4,6	3,1	18,9	231
AB	5,0	1,6	0,7	0,1	0,0	2,4	0,8	5,8	3,2	8,2	29	25	2,5	2,4	12,7	236
BA	4,5	1,0	0,5	0,05	0,0	1,6	1,5	6,1	3,0	7,7	21	50	1,5	1,3	6,7	268
Bw1	4,5	1,0	0,5	0,04	0,0	1,5	1,7	6,6	3,2	8,1	18	53	1,2	0,9	5,1	264
Bw2	4,4	0,8	0,3	0,02	0,0	1,1	1,6	6,8	2,7	7,9	14	59	0,7	0,7	3,8	257

APÊNDICE C – DESCRIÇÃO MORFOLÓGICA DO PERFIL 2 – ÁREA DO ESTUDO 1

DECRIÇÃO GERAL – Perfil 2 do ESTUDO 1

PERFIL Nº – 2
DATA – 02/11/2017
CLASSIFICAÇÃO – Gleissolos Háplicos Tb Distrófico típicos
LOCALIZAÇÃO, MUNICÍPIO, ESTADO E COORDENADAS – Região Campos dos Carvalhos, estrada sentido município Giruá - Senador Salgado Filho, aproximadamente 2,7 km entrada à esquerda e seguindo a estrada até a várzea. Giruá, Rio Grande do Sul, 28º 02’ 25,77” S; 54º 24’ 28,77” W – Sirgas 2000.
SITUAÇÃO, DECLIVE E COBERTURA VEGETAL SOBRE O PERFIL – Descrito e coletado em trincheira no terço inferior – área de várzea, com declive de 3%, sob campo nativo (grama forquilha).
ELEVAÇÃO – 331 metros.
LITOLOGIA – Basalto
FORMAÇÃO GEOLÓGICA – Formação Serra Geral, Grupo São Bento.
PERÍODO – Jurássico-Cretácio, era Mesozóico
MATERIAL ORIGINÁRIO – Produto de alteração de rochas do tipo basalto e material depositado pela erosão.
PEDREGOSIDADE – Ausente
ROCHOSIDADE – Ausente
RELEVO LOCAL – Plano
RELEVO REGIONAL – Ondulado e suave ondulado
EROSÃO – Ausente
DRENAGEM – Mal drenado.
VEGETAÇÃO PRIMÁRIA – Campo subtropical
USO ATUAL – Campo nativo
CLIMA – Cfa 2 da classificação de Köppen.
DESCRITO E COLETADO POR – Jean Michel Moura Bueno e Charles Nicholas Moura Bueno.

DESCRIÇÃO MORFOLÓGICA – Perfil 02 do ESTUDO 1

A 0 – 23 cm; preto-azulado (GLEY 2 2,5/10 B, úmido), preto-azulado (GLEY 2 2,5/5 B, seca); argilosa, moderada, pequena e média, granular e blocos subangulares; firme, ligeiramente plástico e ligeiramente pegajoso; transição clara e plana; raízes muitas.

Cg1 23 – 52 cm; cinza-escuro (2,5 Y 4/1, úmido), cinza-escuro (2,5 4/1, seco); muito argilosa; forte, média e grande, prismática; pouco poroso; macio, friável, plástico e pegajoso, transição clara e plana; raízes poucas.

Cg2 52 – 90+ cm; cinzento muito escuro (10 YR 3/1, úmido), cinzento muito escuro (10 YR 3/1, seca); muito argilosa; forte, média e grande, macia que se desfaz em prismática; pouco poroso; macio, muito plástico e muito pegajoso; raízes raras.

OBSERVAÇÕES: Presença de slickensides no horizonte Cg_2.

APÊNDICE D – CARACTERIZAÇÃO ANALÍTICA – Perfil 02 do ESTUDO 1

Análises físicas								
Horizonte		Composição Granulométrica da Terra Fina (g kg^{-1})			Argila dispersa em água (g kg^{-1})	Grau de floculação (%)	Relação Silte/Argila	Ds (g cm^{-3})
Símbolo	Profundidade (cm)	Areia	Silte	Argila				
A	0 - 23	132	442	426	199	53	1,0	0,90
Cg1	23 - 52	85	386	529	302	43	0,7	0,85
Cg2	52 - 90	6	222	772	429	44	0,3	0,84

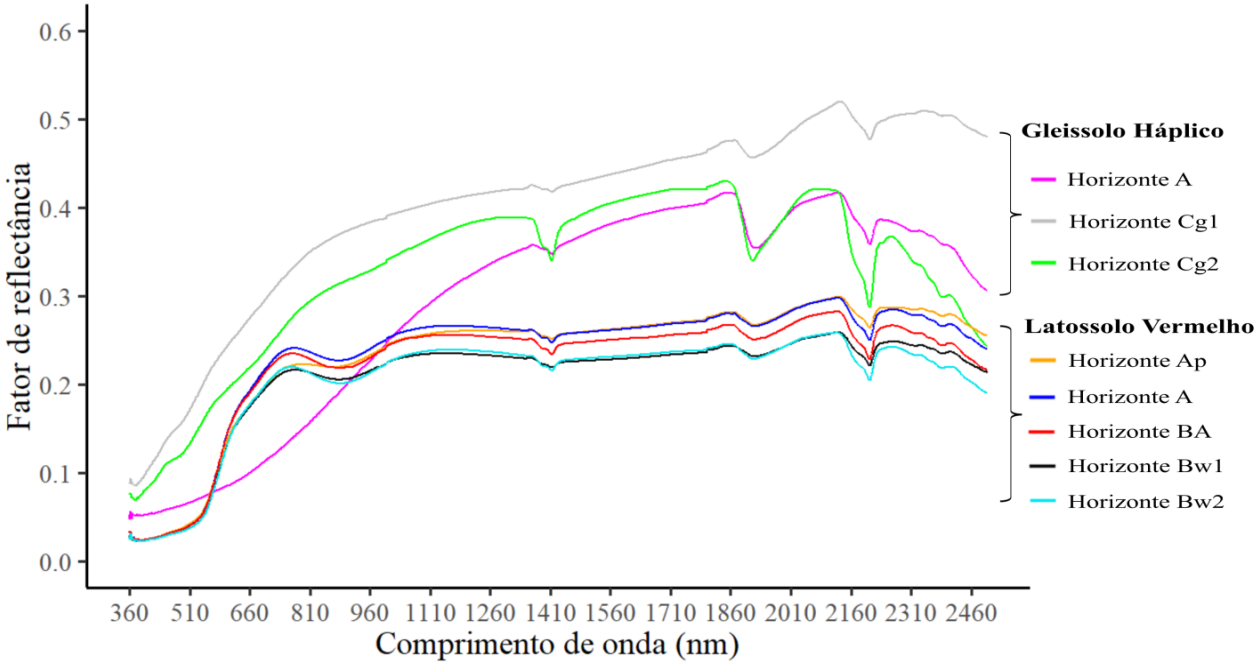
Análises químicas																
Hz.	pH	Complexo sortivo ($cmol_c$ kg^{-1})									V	Al	P	MO	CO	Fe_2O_3
	H_2O	Ca^{2+}	Mg^{2+}	K^+	Na^+	S	Al^{3+}	H+Al	CTC_{ef}	CTC_{ph7}	(%)	(%)	(mg kg^{-1})	(%)	(g kg^{-1})	(g kg^{-1})
A	4,8	11,4	3,0	0,38	0,00	14,7	0,5	9,5	15,2	24,2	61	3	9,2	11,8	61,7	131
Cg1	4,5	5,4	2,6	0,04	0,00	8,0	1,5	10,8	9,5	18,8	43	16	4,9	3,8	22,3	133
Cg2	4,0	4,1	1,5	0,03	0,00	5,6	2,1	11,6	7,7	17,2	33	27	3,8	1,3	7,1	160

APÊNDICE E– FOTOS DOS PERFIS DESCRITOS E DA PAISAGEM DE OCORRÊNCIA NA ÁREA DE ESTUDO 1

Latossolo Vermelho Distroférrico típico

Gleissolos Háplicos Tb Distrófico típicos

APÊNDICE F – CURVAS ESPECTRAIS DOS HORIZONTES DOS PERFIS DESCRITOS NA ÁREA DE ESTUDO 1



APÊNDICE G – COLETA DE AMOSTRAS NOS DIFERENTES USOS DA TERRA DA ÁREA DO ESTUDO 1

APÊNDICE H – ANÁLISES ESPECTRAIS DO SOLO