**UNIVERSIDADE FEDERAL DE SANTA MARIA**

**CENTRO DE CIÊNCIAS RURAIS**

**CURSO DE GRADUAÇÃO EM AGRONOMIA**

**DEPARTAMENTO DE SOLOS**

**Guilherme Reis**

**ÍNDICES ESPECTRAIS COMO ALTERNATIVA PARA PREDIÇÃO DE ARGILA E CARBONO ORGÂNICO EM SOLOS SUBTROPICAIS DO SUL DO BRASIL**

Santa Maria, RS2021

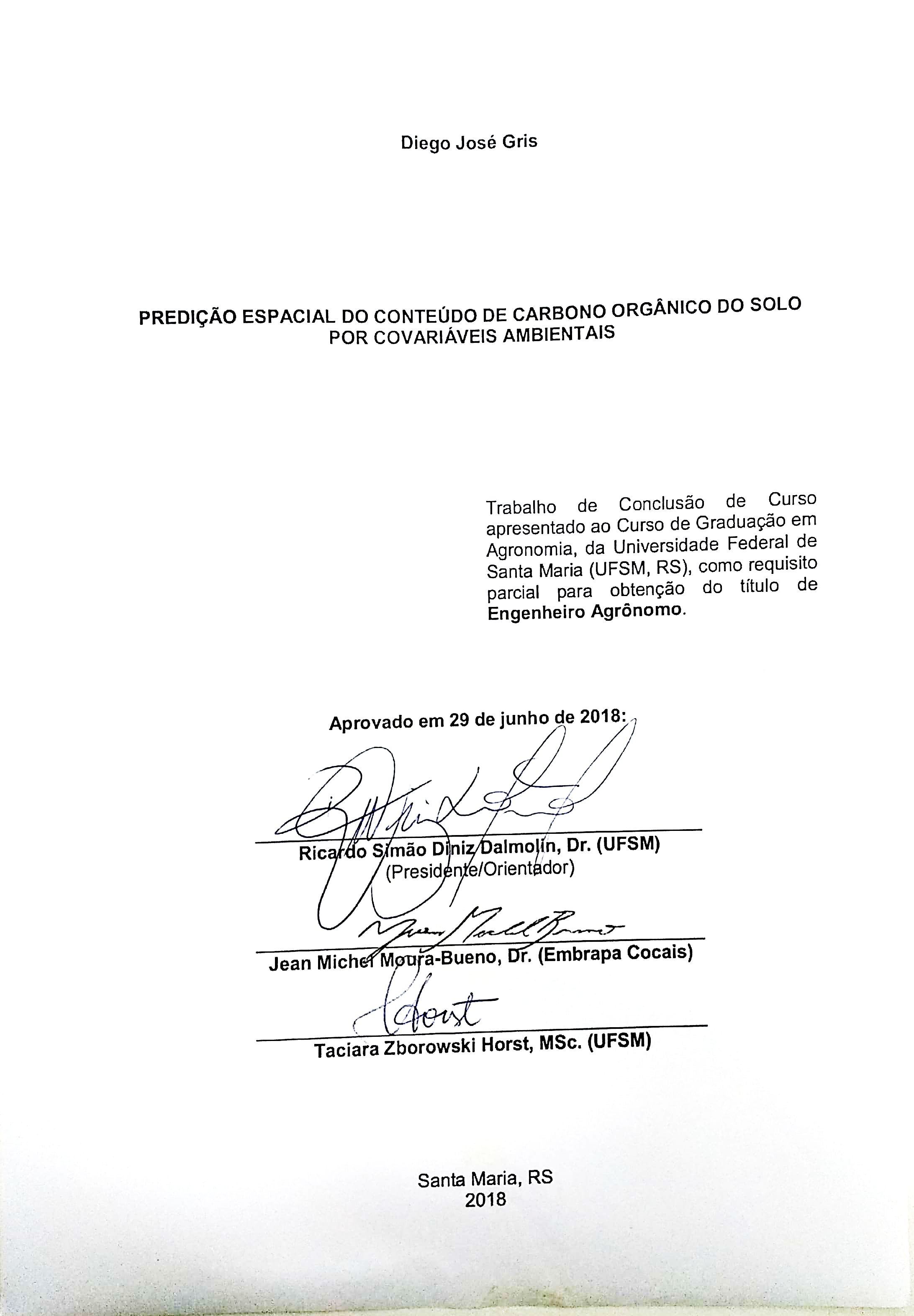
**Guilherme Reis**

**ÍNDICES ESPECTRAIS COMO ALTERNATIVA PARA PREDIÇÃO DE ARGILA E CARBONO ORGÂNICO EM SOLOS SUBTROPICAIS DO SUL DO BRASIL**

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Agronomia, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do título de **Engenheiro Agrônomo**.

Orientador: Prof. Dr. Ricardo Simão Diniz Dalmolin

Santa Maria, RS2021

**AGRADECIMENTOS**

A Deus, por ser a fonte e estar presente em tudo o que é.

Aos meus pais e ao meu irmão, por me apoiarem durante toda a jornada acadêmica.

Ao meu grande amigo Lucas, por estar presente em vários momentos importantes, pelas longas e profundas conversas e por todo o apoio fornecido.

Ao meu orientador, Professor Ricardo Damolin, por me receber em seu grupo ao longo destes anos no Laboratório de Pedologia e pela disposição em orientar a realização deste trabalho.

A todos os colegas com quem convivi no Laboratório de Pedologia da UFSM, pela companhia nas horas de trabalho e lazer, e por todos os bons momentos proporcionados.

Agradecimento especial ao Jean Bueno, por dar a ideia inicial para este trabalho e por disponibilizar os dados para a sua realização.

Agradecimentos especiais à Taciara e ao Nicolas, pelas enormes contribuições ao trabalho.

Aos colegas da Turma 93 da Agronomia, por me acolherem e proporcionarem ótimos momentos ao longo destes anos na UFSM. Grandes amizades e boas memórias serão levadas deste período.

A todos os amigos que conheci em Santa Maria, por tornarem os dias mais alegres e proveitosos.

**RESUMO**

**ÍNDICES ESPECTRAIS COMO ALTERNATIVA PARA PREDIÇÃO DE ARGILA E CARBONO ORGÂNICO EM SOLOS SUBTROPICAIS DO SUL DO BRASIL**

AUTOR: Guilherme Reis

ORIENTADOR: Ricardo Simão Diniz Dalmolin

O solo é um dos reservatórios de carbono (C) mais importantes no ambiente. Dependendo do seu manejo, ele pode atuar como dreno ou fonte de C para a atmosfera. Deste modo, o mapeamento do carbono orgânico do solo (COS) torna-se fundamental para compreender a distribuição do COS na paisagem e auxiliar no desenvolvimento de práticas conservacionistas, para que o solo atue como dreno de C. Portanto, o objetivo deste trabalho foi avaliar o efeito de diferentes conjuntos de covariáveis ambientais no desempenho dos modelos de predição espacial do COS. Os dados de COS foram provenientes de amostras coletadas de 0-5, 5-15 e 15-30 cm em 261 pontos de amostragem, em uma área de 940 ha na região do Planalto do Rio Grande do Sul. Um total de 20 covariáveis ambientais foram utilizadas como preditoras do COS, divididas em três grupos: i) variáveis topográficas, incluindo elevação, declividade, orientação, curvatura horizontal, curvatura vertical, sombreamento do relevo, índice de umidade, fator LS, insolação direta e difusa, derivadas do modelo digital de elevação Topodata com 30 m de resolução espacial; ii) cobertura do solo, incluindo as bandas B4, B5, B6, B7, razão B5/B7 e índice de vegetação ajustado ao solo (SAVI), derivadas de uma imagem do Landsat 8 com 30 m de resolução espacial, além de um mapa de uso da terra; iii) granulometria do solo, composto por mapas do conteúdo de argila, silte e areia das camadas amostradas. Três combinações de conjuntos de covariáveis foram testadas: C1) somente variáveis topográficas; C2) variáveis topográficas e de cobertura do solo; C3) variáveis topográficas, de cobertura do solo e de granulometria. O conteúdo de COS foi predito utilizando o modelo *Boosted Regression Trees* (BRT) no ambiente R, ajustando-se os parâmetros do modelo para obter-se o melhor desempenho. A acurácia dos modelos foi avaliada por validação cruzada. O modelo com maior acurácia foi obtido para a camada 0-5 cm quando foram utilizadas todas as covariáveis (C3) (R2 = 0,86; RMSE = 0,45%). Para as camadas 5-15 cm e 15-30 cm, os melhores modelos foram obtidos utilizando apenas as covariáveis topográficas (C1) (R2 = 0,74; RMSE = 0,46% e R2 = 0,76; RMSE = 0,34%, respectivamente). Características do terreno foram preditores eficientes do conteúdo de COS, com a granulometria e a cobertura do solo tendo relativamente pouco poder explicativo para a variação do COS abaixo de 5 cm. Elevação, índice de umidade e insolação foram bons preditores topográficos e demonstraram que a importância relativa dos atributos do terreno na predição do COS aumenta com a profundidade do solo. Quando maior acurácia é requerida, recomenda-se a inclusão de covariáveis de cobertura e granulometria para a camada superficial (0-5 cm).

**Palavras-chave:** Mapeamento Digital de Solos. Escala de Fazenda. Boosted Regression Trees. Seleção de Covariáveis.

**ABSTRACT**

**SPATIAL PREDICTION OF SOIL ORGANIC CARBON CONTENT BY ENVIRONMENTAL COVARIATES**

AUTHOR: Guilherme Reis

ADVISOR: Ricardo Simão Diniz Dalmolin

Soil is one of the most important carbon (C) reservoirs in the environment. Depending on how it is managed, it can act as a sink or source of C to the atmosphere. Hence, mapping soil organic carbon (SOC) becomes fundamental to understanding SOC distribution throughout the landscape and to help develop conservation practices, so that soil acts as a sink for C. Therefore, the objective of this study was to assess the effect of different combinations of environmental covariates on the performance of models for SOC spatial prediction. SOC data consisted on samples collected from 0-5, 5-15 and 15-30 cm at 261 sampling locations in an area of 940 ha in the Planalto region of Rio Grande do Sul, Brazil. A total of 20 environmental covariates were used as SOC predictors, divided in three groups: i) topographic variables, including elevation, slope, aspect, horizontal curvature, vertical curvature, hillshade, wetness index, LS factor, direct and diffuse insolation, derived from a 30-m Topodata digital elevation model; ii) land cover, including bands B4, B5, B6, B7, B5/B7 ratio and soil adjusted vegetation index (SAVI), derived from a 30-m Landsat 8 image, and a land use map; iii) soil particle-size distribution, consisting of maps for clay, silt and sand contents of the sampled layers. Three different combinations of environmental covariates for SOC spatial prediction were tested: C1) only topographic variables; C2) topographic and land cover variables; C3) topographic, land cover and particle-size distribution variables. SOC content was predicted for each layer using a Boosted Regression Trees (BRT) model in R environment, adjusting model parameters for the best performance. Model accuracy was evaluated using cross validation. The best accuracy was obtained for the 0-5 cm layer when using all covariates (C3) (R2 = 0.86, RMSE = 0.45%). For the 5-15 and 15-30 cm layers, the best models were obtained when only topographic variables were used (R2 = 0.74, RMSE = 0.46% and R2 = 0.76, RMSE = 0.34%, respectively). Terrain attributes were efficient predictors for SOC content, while land cover and particle-size distribution had relatively small importance for explaining SOC variation below 5 cm. Elevation, wetness index and insolation were good predictors among topographic variables and showed that the relative importance of terrain attributes for SOC spatial prediction increases with soil depth. When better accuracy is required, including land cover and particle-size distribution covariates is recommended for the topmost layer (0-5 cm).

**Keywords:** Digital Soil Mapping. Farm Scale. Boosted Regression Trees. Covariates Selection.

**SUMÁRIO**

[1 INTRODUÇÃO 7](#_Toc518373189)

[2 REVISÃO BIBLIOGRÁFICA 10](#_Toc518373190)

[2.1 VARIABILIDADE ESPACIAL DO SOLO 10](#_Toc518373191)

[2.2 MAPEAMENTO DIGITAL DO COS 11](#_Toc518373192)

[2.3 *BOOSTED REGRESSION TREES* 15](#_Toc518373193)

[3 MATERIAL E MÉTODOS 17](#_Toc518373194)

[3.1 ÁREA DE ESTUDO 17](#_Toc518373195)

[3.2 AMOSTRAGEM DO SOLO 17](#_Toc518373196)

[3.3 ANÁLISES LABORATORIAIS 18](#_Toc518373197)

[3.4 COVARIÁVEIS AMBIENTAIS 19](#_Toc518373198)

[3.4.1 Variáveis topográficas 20](#_Toc518373199)

[3.4.2 Cobertura do solo 22](#_Toc518373200)

[3.4.3 Granulometria do solo 23](#_Toc518373201)

[3.5 EXTRAÇÃO DAS COVARIÁVEIS PARA OS PONTOS DE AMOSTRAGEM 26](#_Toc518373202)

[3.6 PREDIÇÃO ESPACIAL DO COS 27](#_Toc518373203)

[4 RESULTADOS E DISCUSSÃO 29](#_Toc518373204)

[4.1 DISTRIBUIÇÃO DOS VALORES DE COS NAS AMOSTRAS DE SOLO 29](#_Toc518373205)

[4.2 CORRELAÇÃO ENTRE O COS E AS COVARIÁVEIS AMBIENTAIS 30](#_Toc518373206)

[4.3 DESEMPENHO DOS MODELOS NA PREDIÇÃO DO COS 33](#_Toc518373207)

[4.4 IMPORTÂNCIA RELATIVA DAS COVARIÁVEIS 36](#_Toc518373208)

[4.5 PREDIÇÃO ESPACIAL DO COS 38](#_Toc518373209)

[5 CONCLUSÃO 41](#_Toc518373210)

[REFERÊNCIAS 42](#_Toc518373211)

# 1 INTRODUÇÃO

Os solos figuram como um dos recursos naturais vulneráveis do mundo, impulsionado pelas mudanças climáticas, degradação das terras e perdas de biodiversidade. Diante disso, tem-se aumentado a demanda por informações das propriedades do solo em escalas maiores e com resoluções mais finas, visando explorar de forma sustentável os diferentes tipos de exploração agrícola dos solos, prover dados para a compreensão da dinâmica dos ecossistemas e previsão de cenários futuros. A argila e carbono orgânico do solo (COS) são as propriedades do solo mais estudadas pois ambos governam a maior parte dos processos e propriedades biológicas, físicas, hidrológicas e químicos do solo (OSMAN, 2013), além de serem fundamentais para a produção sustentável de alimentos e combate ao aquecimento global (LAL et al., 2018; SORENSEN et al., 2018)

Tradicionalmente a determinação dos teores de COS e argila baseia-se em coletas a campo e posterior analises convencionais em laboratório. Tais análises estabelecidas como padrão em laboratórios de rotina, apesar de serem largamente utilizadas em todo o mundo, caracterizam-se pela geração de contaminantes, destruição do solo e risco para saúde humana (COSTA e KLEIN, 2006), morosidade e inexequível para as demandas atuais. Diante disso, diversas técnicas têm sido propostas para substituir as metodologias tradicionais que devem possibilitar uma análise rápida, de melhor relação custo-benefício e não destrutiva das propriedades do solo (VISCARRA ROSSEL et al., 2006). A Espectroscopia de Reflectância Difusa (ERD) tem-se destacado como método alternativo para quantificação e análise de diversas propriedades do solo (VISCARRA ROSSEL et al., 2006; VISCARRA ROSSEL e MCBRATNEY, 2008; PINHEIRO et al., 2017; LAZZAR et al., 2020). A predição das propriedades do solo por ERD ocorre pelo uso de métodos estatísticos multivariados (modelos de predição), que correlacionam os padrões complexos de absorção (variáveis espectrais) do solo com as propriedades a serem preditas (STENBERG et al., 2010). Essa técnica possibilita a predição de várias propriedades biológicas, físicas, químicas e mineralógicas do solo (SORIANO-DISLA et al., 2014; FANG et al., 2018), superando as limitações impostas pelas análises tradicionais (PINHEIRO et al., 2017) e sendo inofensiva ao meio ambiente utilizando apenas uma única leitura espectral combinada a calibrações multivariadas (VISCARRA-ROSSEL et al., 2009; OLIVEIRA et al., 2013).

# Os dados espectrais do solo referem-se aos valores de reflectância espectral medidos por sensores denominados espectroradiômetros, que desagregam a radiação incidente e refletida em diferentes comprimentos de ondas. Os dados espectrais caracterizam-se pela ausência de especificidade em virtude da sobreposição de bandas de absorção dos constituintes do solo, estes por sua vez são variados e inter-relacionados. Com isso, as intensidades das feições diagnósticas dos constituintes do solo tornam-se fracas e amplas, prejudicando o desempenho de modelos preditivos (VASÁT et al., 2017).

# Uma alternativa para contornar esses problemas é a utilização de índices espectrais, medidas adimensionais de reflectância espectral obtida por operações algébricas, como razão, diferença ou diferença normalizada (JACKSON; HUETE, 1991). Os índices espectrais podem promover o refinamento das relações entre os teores de COS e argila com as feições espectrais (HONG et al., 2018; XIA et al., 2020). Além disso, reduzem a complexidade e dimensionalidade dos dados espectrais (ANGELOPOULOU et al., 2020), extraem informações valiosas ocultas entre as bandas (INOUE et al.,2012) e aceleram o processamento computacional quando se manipulam grandes bancos de dados espectrais (SAWUT et al., 2018). Diversos estudos demonstram a otimização da predição de argila e COS com uso de índices espectrais, usados tanto para a construção de modelos de regressão ou como potenciais covariáveis de métodos estatísticos multivariados (BARTHOLOMEUS et al., 2008; JIN et al., 2016; JIN et al., 2017; BAO et al., 2017, HONG et al., 2018; XIA et al., 2020;).

Atualmente, 33 % dos solos mundiais estão degradados e até 2050 poderá atingir o patamar de 90 % (FAO; ITPS, 2015). Isso implicaria em uma redução de 50% na produtividade das culturas agrícolas, levando em conta que 95% dos alimentos são direta ou indiretamente produzidos no solo (FAO, 2015). Em face desse panorama preocupante, aumentou-se a demanda por informações das propriedades de solo, nesse caso de COS e argila, em altas resoluções (NOCITA et al., 2015) que auxiliarão no enfrentamento dos desafios impostos pela segurança alimentar, mudança climática, degradação ambiental, escassez hídrica e perda de biodiversidade (SÁNCHEZ et al., 2009).

Portanto, o conhecimento do conteúdo, variabilidade e distribuição de COS e argila é imprescindível para prover uma atividade agrícola sustentável e ao mesmo tempo fornecer dados para a análise de cenários globais, sobretudo nos Estados do Rio Grande do Sul e Santa Catarina na região sul do Brasil. Além disso, não há estudos focados na aplicação de índices espectrais na predição de COS e argila em solos subtropicais brasileiros. Diante desse cenário, o objetivo desse trabalho foi avaliar o uso dos índices espectrais simulados brutos e pré-processados como covariáveis preditoras em modelos preditivos dos teores de COS e argila em solos subtropicais do Sul do Brasil.

# 

# 2 REVISÃO BIBLIOGRÁFICA

## 2.1 CARBONO ORGÂNICO DO SOLO

A matéria orgânica do solo (MOS) integra a fração sólida orgânica do sistema trifásico e é constituída por resíduos vegetais e animais em vários estágios de decomposição, variando desde materiais frescos não decompostos, parcialmente decompostos, produtos de decomposição até totalmente decomposto, caracterizado pelo húmus. Os resíduos animais e vegetais contém açucares, celulose, hemicelulose, proteínas, lipídios que quando incorporados ao solo são rapidamente decompostos pelos microrganismos enquanto a lignina é degradada lentamente. Com isso, acumula-se no solo produtos de decomposição lenta e substâncias ressintetizadas que formam associações complexas com os minerais do solo, sobretudo a argila. O húmus ou matéria orgânica completamente decomposta é construída ao longo de séculos ou até milênios em virtude da contínua decomposição, ressíntese e complexação (OSMAN, 2013).

A matéria orgânica do solo é uma combinação complexa de diversos materiais orgânicos, contendo vários grupos funcionais de carbono (LEHMANN e KLEBER, 2015). Em função disso, o carbono orgânico do solo (COS) é um indicador robusto da SOM (ROPER et al., 2019), na qual o valor de COS é multiplicado por um fator que varia entre 1,724 a 2,000, para obter o teor de MOS, conhecido como fator de Van Bemmelen. Esse fator de conversão resulta do fato que a quantidade de COS presente na MOS varia entre 50 – 58 %. O COS é determinado em laboratório com diferentes metodologias, de modo que todo carbono orgânico e até o inorgânico seja convertido em CO2. É conveniente utilizar somente o teor de COS, em virtude da variabilidade do teor de COS na MOS.

A MOS é um componente do solo primordial que imprime influência no estado e funcionamento dos outros componentes do solo, apesar de estar presente na maioria dos solos em teores inferiores a 5%. A MOS desempenha funções químicas no solo ao atuar na troca de íons, poder tampão, intemperismo pedoquímico, quelação e translocação de nutrientes dentro do solo (OSMAN, 2013). Os coloides de húmus possuem expressiva área superficial onde localizam-se grupos químicos de cargas negativas que incrementam a capacidade de troca de cátions efetiva e a pH 7 do solo em até 44 e 21%, respectivamente. (CIOTTA et al., 2003). Também é capaz de complexar metais (quelação) como Al, Fe, Mn, Zn e Cu, metais pesados como Cd e Pb e resíduos de agrotóxicos.

# As funções físicas promovidas pela MOS envolvem o aperfeiçoamento da agregação, aeração, movimento de água, redução da evaporação e condutividade térmica (HERNÁNDEZ-SORIANO,2012). Solos com elevados teores de SOM, exibem coloração mais escura e diminuição do albedo (DALMOLIN et al., 2005). A MOS atua como um agente cimentante contribuindo na estruturação e porosidade do solo, o que implica em maior taxa de infiltração e capacidade de armazenamento de água, reduzindo a erosão e escoamento superficial. Independente da classe textural do solo, à medida que aumenta o teor de MOS aumenta a capacidade de água disponível no solo (HUDSON,1994).

# A MOS também se destaca como um exímio promotor da atividade biológica do solo, sendo fonte de energia e carbono para a fauna do solo. Os resíduos vegetais e animais adicionados ao solo são degradados sobretudo pela microfauna do solo, microrganismos heterotróficos e detritívoros representados por diversos gêneros de bactérias e fungos. A partir da decomposição microbiana da MOS, ocorre a reciclagem e mineralização de macro e micronutrientes, que paulatinamente são liberados para a solução do solo, ficando disponíveis para a absorção das plantas e com isso melhoram a fertilidade do solo. Também favorece o desenvolvimento de microrganismos de ação desintoxicante de metais pesados e resíduos de agrotóxicos, além de estimular associação entre fungos micorrízicos e plantas, protegendo contra o ataque de patógenos do solo e ampliando a capacidade de absorção de água e nutrientes (FREY et al., 2019).

## 2.2 FRAÇÃO ARGILA DO SOLO

A granulometria refere-se à distribuição das partículas que compõem o solo, de natureza inorgânica ou mineral, em classes de tamanho ou frações granulométricas. De modo geral, há três classes de tamanho baseados no diâmetro das partículas: areia (0,05 – 2 mm), silte (0,002 – 0,05 mm) e argila (< 0,002 mm). As frações areia e silte são constituídos por minerais primários como quartzo, feldspatos, micas, etc, enquanto a fração argila integra os mineriais secundários como caolinita, esmectita, vermiculita, ilita, cloria e óxidos de ferro e alumínio.

A granulometria do solo figura-se como um dos mais importantes e estáveis atributos do solo, determinando e influenciando a grande maioria das propriedades do solo, com destaque para a fração argila. Em virtude de sua elevada área superficial especifica, frequentemente carregada negativamente, atua na retenção de nutrientes indispensáveis ao crescimento e desenvolvimento das plantas, tais como Ca+2, Mg+2, K+, NH4+, bem como atenua a contaminação de águas subterrâneas por agrotóxicos, metais pesados, esgoto da lixiviação e resiste as alterações extremas do pH (NEWMAN, 1984).

As propriedades físicas do solo sofrem grande influência pela quantidade de argila presente no solo. Ela atua em conjunto com a matéria orgânica na formação e consolidação de agregados (peds) e consequente estruturação do solo (TISDALL e OADES, 1982). Esta por sua vez opera o crescimento vegetal através da configuração da aeração, movimento da água, condução do calor, compactação e resistência a erosão. Além disso, impacta na consistência do solo mantendo a coesão e adesão entre partículas o que reduz a erodibilidade e aumenta a capacidade de retenção de água no solo (CHURCHMAN, 2018).

A fração argila é capaz de armazenar a maior parte do COS e protegê-lo da mineralização (SIX et al., 2002) comparativamente as frações areia e silte, favorecendo a sequestro do C atmosférico, melhoria da estrutura do solo, mitigação da emissão de gases de efeito estufa e promoção da atividade biológica do solo (ZHONG et al.,2018). A associação íntima entre argila e COS decorre que a argila dispõe simultaneamente de cargas superficiais variáveis e permanentes em conjunto com diferentes áreas superficiais especificas que são fundamentais para a formação de complexos organo-minerais e estabelecimento de agregados (SCHIMDT et al., 2011; SARKAR et al., 2018). Os organismos do solo são beneficiados por essa associação, já que extraem a energia necessária para a sobrevivência ao empreenderem a ciclagem de macro e micronutrientes, promovem o crescimento vegetal pela liberação de hormônios (auxinas, giberelinas e citocininas), mobilizam nutrientes para as plantas e inibição de fitopatógenos.

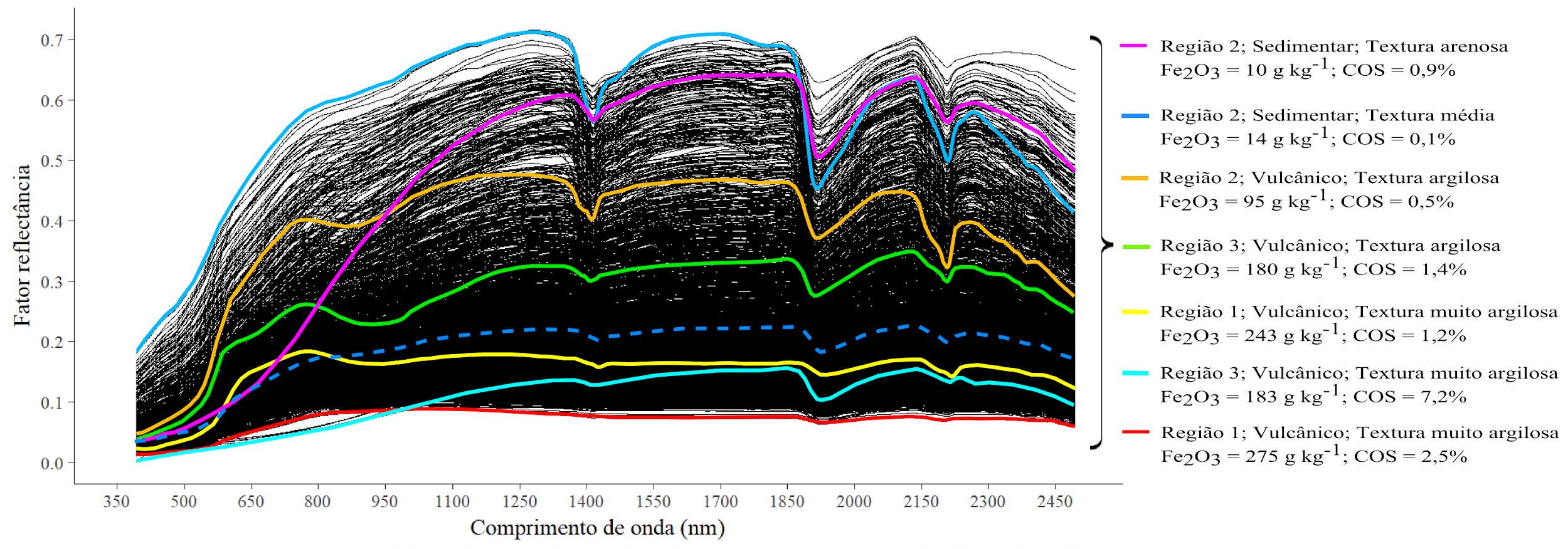
## 2.3 ESPECTROSCOPIA DE REFLECTÂNCIA DIFUSA (ERD)

A Espectroscopia é a ciência que estuda a luz em função do comprimento de onda refletido, emitido ou espalhado por um material gasoso, líquido ou sólido (CLARK,1999). A quantidade de fótons absorvidos (absorbância) ou refletidos (reflectância) por algum material, nesse caso o solo, podem ser quantificados e interconvertidos pela seguinte fórmula:

A reflectância, mais estritamente, a reflectância espectral representa a razão entre a radiância (radiação refletida) e irradiância (radiação incidente) em determinado comprimento de onda sobre o solo por exemplo, sendo captados por sensores denominados espectroradiomêtros (DALMOLIN et al., 2005, STONER e BAUMGARDNER,1981). O processo refletivo é denominado espalhamento, permitindo que os fótons espalhados possam ser detectados e medidos, compondo assim os dados de reflectância espectral. As bandas de absorção presentes na assinatura espectral do solo são o resultado da ocorrência de três processos físicos: transições eletrônicas, vibracionais e rotacionais (FANG et al., 2018).

A assinatura espectral de um solo é composta por diversos sinais complexos em virtude das absorções sobrepostas dos constituintes do solo (HONG et al., 2018; STENBERG et al., 2010), podendo ser representada graficamente através de curvas espectrais (Figura 1). O padrão complexo de absorção do solo é fortemente influenciado por suas propriedades biológicas, físicas, químicas e mineralógicas, destacando-se a matéria orgânica e óxidos de ferro, também associado com os efeitos da mineralogia da fração argila, textura, rugosidade e umidade do solo (DALMOLIN et al.,2005).

Figura 1 – Curvas espectrais de um Biblioteca Espectral de Solos (n = 2598), com destaque para curvas representativas de três regiões com diferentes materiais de origem e composição organo-mineral.



Fonte: MOURA-BUENO (2018)

A COS exerce influência expressiva na coloração (SCHULZE et al., 1993) e reflectância do solo (DALMOLIN et al., 2005). Solos com elevado teores de COS apresentam baixos valores de reflectância na região espectral do visível (DEMATTÊ et al, 2019, The Brazilian Soil Spectral Library (BSSL). Em virtude, da sensibilidade espectral da COS ao longo da região do vis-NIR, diversos estudos buscaram selecionar bandas, feições ou comprimento de ondas mais correlacionadas com a COS (KAWAMURA et al., 2020). Viscarra Rossel e Behrens (2010) encontraram que as melhores predições COS foram aquelas que possuíram forte correlação com os comprimentos de ondas associados aos compostos C – O, C = O e N – H. Laamrani et al. (2019) identificaram 7 bandas espectrais (433-435, 711-725, 727, 986-998, 2198-2206, 2365-2373 e 2481-2500 nm) que fornecerem informações relevantes para a predição satisfatória de SOC em uma lavoura conduzida em rotação soja-milho-trigo na província canadense de Ontário. A argila é espectralmente ativa ao longo da região espectral 350-2500 nm, em virtude das bandas de absorção de seus minerais que a compõem. As bandas de maior importância para a predição de argila estão situadas na região do visível (525, 620 e 680 nm), infravermelho próximo (875 nm) e infravermelho de ondas curtas (1410, 1920, 2200 e 2480 nm). (COBLINSKI et al., 2020).

Historicamente, a compreensão do sistema solo e suas propriedades ocorre através de análises tradicionais de laboratório (VISCARRA-ROSSEL et al., 2006), caracterizadas por serem morosas e dispendiosas. Diversas técnicas têm sido propostas substituir as metodologias tradicionais, enfocadas sobretudo no sensoriamento remoto e sistemas de informações geográficas. Nas últimas três décadas, a Espectroscopia de Reflectância Difusa emergiu como umas das técnicas mais promissoras do Sensoriamento Remoto com grande aplicabilidade na Ciência do Solo. O sucesso da ERD é derivado do fato de ser uma alternativa rápida e de melhor relação custo-benefício em comparação aos métodos analíticos para a estimativa de diversos constituintes do solo, majoritariamente de COS (MOURA-BUENO et al., 2019, MOURA-BUENO et al., 2020, ROSIN et al., 2021) e argila (Dotto et al.,2014, Dotto et al., 2016, Dotto et al., 2017) .

A quantificação de propriedades do solo, em especial a argila e COS, através da técnica da ERD, lastra-se na construção de modelos predições que utilizam dados espectrais de solo. Esses modelos de predição são oriundos da combinação de diferentes métodos estatísticos multivariados e técnicas de pré-processamento de dados espectrais, estes extraídos de bibliotecas espectrais de diferentes escalas geográficas, que podem ser locais, regionais, nacionais, continentais ou globais. (Tabela 1). Em cada um desses estudos, a definição da escala empregada está vinculada com o domínio territorial das amostras de solo coletadas. Além disso, tem ocorrido variabilidade da acurácia preditiva dos teores de argila (Tabela 1) e COS (Tabela 2) entre os trabalhos.

Tabela 1: Compilação de trabalhos de predição de argila com diferentes métodos multivariados, técnicas de pré-processamento, escalas de bibliotecas espectrais e regiões espectrais

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Fonte** | **Escala** | **Método**  **multivariado** | **Processamento**  **espectral**. | **Amostras** | **Região Espectral** | **R2** | **RMSEval** |
| Davari et al., 2021 | Local | PLSR | ABS + SG-FD | 220 | VIS-NIR-SWIR | 0,77 | 6,24 |
| Laazar et al., 2021 | Local | PLSR | SG-FD | 100 | VIS-NIR-SWIR | 00,53 | 2,0 % |
| Coblinsky et al., 2020 | Local | Cubist | Spline correction | 197 | VIS-NIR-SWIR | 0,84 | 5,1 % |
| MIR | 00,94 | 3,7% |
| VIS-NIR-SWIR-MIR | 00,94 | 3,6 % |
| Levi;Karnieli;Paz-Kagan, 2020 | Local | PLSR | Autoscale + GLSW | 121 | VIS-NIR-SWIR | 00,92 | 3,52% |
| Ramirez-Lopez et al., 2019 | Local | MBL | ABS  SG-FD  SNV | 910 | VIS-NIR-SWIR | 00,87 | 6,31 % |
| Jaconi; Vos; Don, 2019 | Local | MBL | SNV + SG-FD | 4080 | NIR-SWIR | 00,97 | 18,5 g kg-1 |
| Campbell et al., 2018 | Local | PLSR  ANN  RF  GPR linear  GPR radial  SVM linear | CR  ABS  SG-FD  SNV  MSC | 184 | MIR | 00,08 -0,50 | 5,16 – 9,77 % |
| VIS-NIR-SWIR | 0,19 – 0,69 | 4,38 – 8,93 % |
| VIS-NIR-SWIR-MIR | 0,04 – 0,54 | 5,41 – 18,42 % |
| Dotto et al., 2017 | Local | PLSR  SVM | CR  DET  CR + BR  DET + BR | 299 | VIS-NIR-SWIR | 0,35 – 0,62 | 6,84 - 8, |
| Dotto et al., 2016 | Regional | LRM | RAW | 412 | VIS-NIR-SWIR | 00,75 | 8,51 % |
| Terra, Demattê, Viscarra Rossel, 2015 | Regional | SVM | RAW | 1259 | VIS-NIR-SWIR | 0  00,86 | 95,34 g kg-1 |
| MIR | 00,88 | 84,48 g kg-1 |

Abreviações: PLRS = partial least-squares regression; SVM = Support vector machine; RF = random forest; ANN = Artificial neural network; GPR = Gaussian process regression; LRM = Linear regression model; MBL = Memory-based; CR = continuum removal; DET = Detrend; BR = band ratio; SNV = standard normal variate transformation; ABS = Absorbance; SG-FD = Savitzky-Golay 1st derivative; MSC = Multiplicative scatter correction; RAW = no preprocessing treatment; GLSW = Generalized least squares weighting; SMO = smoothing; VIS = Visible; NIR = Near-infrared; SWIR = Shortwave infrared, MIR = Mid-infrared;; R2val = Coefficient of determination (validation) RMSEval = Root mean square error (validation)

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

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Fonte** | **Escala** | **Método**  **multivariado** | **Processamento**  **espectral** | **Amostras** | **Região Espectral** | **R2val** | **RMSEval** |
| Reda et al., 2021 | Local | PLSR | ABS + SG-FD | 650 | NIR-SWIR | 0,90 | 0,38 |
| MBL | 0,88 | 0,41 |
| BPNN | 0,90 | 0,43 |
| Calreg | 0,92 | 0,29 |
| Rosin et al., 2021 | Regional | PLSR | SMO  SNV  SG-FD | 2471 | VIS-NIR-SWIR | 0,14 -0,24 | 0,77 – 0,90 % |
| 1887 | 0,41 – 0,62 | 0,50 – 0,89 % |
| 1025 | 0,65 – 0,78 | 0,38 – 0, 39  % |
| Allo et al., 2020 | Local | PLSR | RAW  DET  SMO + SNV | 95 | VIS-NIR-SWIR | 00,92 – 0,94 | 6,96 – 8,04 g kg-1 |
| RAW  SMO + SNV + BAS  SMO + SNV + BAS + SGD | MIR | 00,88 – 0,95 | 6,67 – 10,03 g kg-1 |
| Ghosh; Das; Reddy; 2020 | Local | PLSR | RAW  ABS  MSC  SNV  SG – FD  SG - SD | 280 | VIS-NIR-SWIR | 00,30 – 0,73 | 0,07 – 0,19 % |
| RF | 00,31 – 0,69 | 0,07 – 0,38 % |
| SVR | 00,20 – 0,57 | 0,08 – 0,59 % |
| MARS | 00,32 – 0,63 | 0,10 – 0,66 % |
| Moura-Bueno et al., 2019 | Local | PLSR | SMO  CR  DET  SG-FD  ABS + SG-FD | 810 | VIS-NIR-SWIR | 00,63 – 0,74 | 0,52 – 0,66 % |
| MLR | 00,62 – 0,72 | 0,57 – 0,66 % |
| SVM | 00,55 – 0,72 | 0,55 – 0,70 % |
| RF | 00,52 – 0,72 | 0,56 – 0,64 % |
| Johnson et al., 2019 | Continental | PLSR | ABS + SG-FD | 285 | SWIR | 00,78 | 0,55 % |
| MIR | 00,80 | 0,52 % |
| SWIR + MIR | 00,82 | 0,49 % |
| Xu et al., 2018 | Local | PCR | SG-FD | 585 | VIS-NIR-SWIR | 0,83 | 6,21 g kg-1 |
| PLSR | 0,86 | 5,59 g kg-1 |
| BPNN | 0,89 | 5,07 g kg-1 |
| SVMR | 0,90 | 4,92 g kg-1 |
| Dotto et al., 2018 | Local | PLSR | CR  NBR  SNV  MSC  ABS + SG-FD  SG-FD  SMO | 595 | VIS-NIR-SWIR | 00,67 – 0,80 | 0,49 – 0,67 % |
| PCR | 00,66 – 0,80 | 0,51 – 0,66 % |
| MLR | 00,69 – 0,79 | 0,52 – 0,64 % |
| SVM | 00,74 – 0,80 | 0,52 – 0,59 % |
| RF | 00,47 – 0,77 | 0,55 – 0,84 % |
| BMA | 00,68 – 0,80 | 0,51 – 0,65 % |
| WAPLS | 00,48 – 0,82 | 0,48 – 0,74 % |
| GPR | 00,65 – 0,79 | 0,52 – 0,69 % |
| ANN | 00,64 – 0,80 | 0,51 – 0,69 % |
| Dotto et al., 2017 | Local | PLSR | DET  CR  CR + BR  DET + BR | 299 | VIS-NIR-SWIR | 00,68 – 0,90 | 0,32 – 0,54 % |
| SVM | 00,68 – 0,87 | 0,35 – 0,56 % |
| Pinheiro et al., 2017 | Local | PLSR | SNV | 434 | VIS-NIR-SWIR | 0  00,71  0 | 5,69 g kg-1 |

Abreviações: PLRS = partial least-squares regression; SVM = Support vector machine; RF = random forest; MLR = multiple linear regression; ANN = Artificial neural network; GPR = Gaussian process regression; MBL = Memory-based learning; BPNN = Backpropagation neural network; Calreg = Combined classification and regression method; SVR = Support vector regression; MARS = Multivariate adaptive regression splines; SVMR = Support vector machine regression; PCR = Principal component regression; BMA = Bayesian model averaging; WAPLS = Weighted average partial least-squares; CR = continuum removal; DET = Detrend; BR = band ratio; NBR = Normalization by range; BAS = Baseline; SNV = standard normal variate transformation; ABS = Absorbance; SG-FD = Savitzky-Golay 1st derivative; SG-SD = Savitzky-Golay 2st derivative; MSC = Multiplicative scatter correction; SMO = smoothing RAW = no preprocessing treatment; VIS = Visible; NIR = Near-infrared; SWIR = Shortwave infrared, MIR = Mid-infrared; R2val = Coefficient of determination (validation) RMSEval = Root mean square error (validation)

## 2.4 MÉTODOS ESTATÍSTICOS MULTIVARIADOS E TÉCNICAS DE PRE-PROCESSAMENTO ESPECTRAL

Os dados espectrais do solo são constituídos por diversos sinais complexos com diferentes informações sobrepostas decorrente da composição heterogênea dos solos. Para relacionar os dados espectrais (variáveis preditoras) com a propriedades do solo (variáveis respostas), é necessário a utilização de modelos estatísticos. Além disso, os instrumentos utilizados na ERD fornecem milhares de valores de reflectância espectral, resultando em maior quantidade de variáveis preditoras do que variáveis respostas. Por isso, modelos matemáticos mais flexíveis normalmente são preferidos em detrimento de modelos mais simples.Os métodos estatísticos multivariados mais utilizados na literatura são o *Partial least squares regression* (PLSR), *Random forest* (RF), *Support vector machine* (SVM) (DOTTO et al., 2018) e mais recentemente o *Cubist (*MOURA*-*BUENOet al., 2020), em função da elevada habilidade em relacionar os dados espectrais com os teores de argila (SILVA et al., 2019) e COS (MOURA-BUENO et al., 2018).

O PLSR é um dos métodos lineares mais empregados, já que propicia a redução a dimensionalidade dos dados espectrais e problemas de colinearidade, através da compressão das variáveis preditoras iniciais em poucos componentes principais, com o mínimo de perda de informação (VASQUES et al., 2008, CAMPBELL et al., 2018), funcionamento similar à Análise de Componentes Principais (TEN CATEN et al., 2011). O RF é um algoritmo de regressão e classificação não linear e não paramétrico desenvolvido por Breiman (2001), fundamentado em um tipo de aprendizado de máquina (ensemble learning), que produz vários classificadores e agregados em seus resultados (NAWAR e MOUAZEN, 2019). O RF não exige nenhum pré-tratamento de dados, possui um tempo de processamento rápido (DOTTO et al., 2018) e apresenta uma potencial habilidade preditiva de COS e argila (VISCARRA-ROSSEL e BEHRENS, 2010; MOURA-BUENO et al., 2019).

Santana, Souza e Poppi (2018) alcançaram elevada acurácia preditiva na validação para argila (R2 = 0,80 e RMSE = 76,15 g/kg) e matéria orgânica do solo (R2 = 0,80 e RMSE = 5,49 g/dm3) com o uso de RF, evidenciando seu alto desempenho preditivo, sendo inclusive superior ao PLSR.O SVM é uma técnica de aprendizado de máquina baseado em funções *kernel* (CORTES; VAPNIK 1995), sobressaindo em relações aos modelos lineares por ser capaz de analisar as relações naturalmente não lineares entre dados espectrais e propriedades do solo (DOTTO et al., 2017) com alto desempenho preditivo (LI et al., 2015; XU et al., 2018; HONG et al., 2018). O *Cubist* é método de mineração de dados que opera através da construção de árvores de decisão, na qual a predição é baseada em modelos de regressão linear ao invés de valores discretos (MINASNY e MCBRATNEY, 2010). Silva et al., (2019) obtiveram com o uso do *Cubist* excelente desempenho na predição de argila (R2 = 0,83, RMSE = 7,29 % e RPIQ = 3,70) em 1534 amostras de solo. Resultados semelhantes foram encontrados por Moura-Bueno et al, (2020), na predição do teor de COS.

Diversos estudos compararam o desempenho entre modelos preditivos de COS e argila construídos principalmente a partir dos métodos PLSR, SVM, (VISCARRA –ROSSEL, BEHRENS, 2010; NAWAR et al., 2016; DOTTO et al., 2017; CAMPBELL et al., 2018, MOURA-BUENO et al., 2019; SILVA et al., 2019; LAZZARETTI et al., 2020), RF (VISCARRA-ROSSEL, BEHRENS,2010, CAMPBELL et al., 2018; MOURA-BUENO et al., 2019; SILVA et al., 2019) e *Cubist* (SILVA et al.,2019), com grande variabilidade do melhor método estatístico multivariado entre os trabalhos. Diante disso, é relevante testar diferentes métodos multivariados, pois cada um apresenta um funcionamento especifico para relacionar os dados espectrais com as propriedades do solo, permitindo assim encontrar o mais adequado e robusto para conjunto de dados espectrais e propriedades do solo em estudo.

Anteriormente a aplicação dos dados espectrais como variáveis preditoras nos modelos estatísticos multivariados, estes são submetidos a várias técnicas de pré-processamento espectral. O seu uso é justificado pelo fato de promover a redução da dimensionalidade de dados espectrais, remoção de ruídos de origens diversas e/ou realçar feições diagnósticas e, consequentemente, aperfeiçoar o modelo preditivo (NAWAR et al., 2016; DOTTO et al., 2018). De modo geral, as técnicas de pré-processamento mais empregadas são classificadas em duas categorias: métodos de correção de dispersão e derivados espectrais, maiores detalhes podem ser encontrados em Rinnan et al., (2009). Dentre todas as técnicas de pré-processamento disponíveis, a técnica de Savitzky- Golay 1ª derivada tem sido demonstrado por aumentar o desempenho de predições de COS quando associada aos métodos PLSR (MOURA-BUENO et al., 2019, MENDES et al., 2021), SVM e RF (DOTTO et al., 2018) e para argila quando combinada ao método RF (SILVA et al., 2019), PLSR (PINHEIRO et al., 2017) e Cubist (HAGHI; PÉREZ-FERNÁNDEZ; ROBERSTON, 2021).

A técnica de ERD mede o valor de reflectância do solo na região do visível (400-700 nm), infravermelho-próximo (700-1100 nm), infravermelho de ondas curtas (1100-2500 nm) e infravermelho médio (2500 – 25000 nm) do espectro eletromagnético. As feições de absorção no visível decorrem das excitações eletrônicas das moléculas enquanto no infravermelho próximo resultam da combinação de overtones das vibrações moleculares fundamentais (ADELINE et al., 2017). Comparativamente à região do infravermelho médio, os constituintes do solo apresentam ao longo da região do infravermelho próximo uma absorção fraca e ampla, resultando em poucas feições diagnósticas de difícil interpretação (STENBERG et al., 2010). Entretanto faixas espectrais sobretudo ao longo de 350-2500 nm tem sido utilizadas com sucesso como covariáveis de modelos preditivos de COS (NOCITA et al., 2014; DOTTO et al., 2018; MOURA-BUENO et al., 2019; LIU et al., 2019; ROSIN et al., 2021) e argila (CURCIO et al., 2013; DOTTO et al., 2016; NAWAR et al., 2016; ADELINE et al., 2017; SILVA et al., 2019).

2.5 INDÍCES ESPECTRAIS

As faixas do espectro eletromagnético mais correlacionadas com atributos do solo podem ser realçadas por transformações, dando origem a índices espectrais que podem ser criados com base em bandas de satélites (JIN et al., 2017) e posteriormente empregados como potenciais covariáveis preditoras de argila e COS.. Os índices espectrais são medidas adimensionais de reflectância espectral obtida por operações algébricas, como razão, diferença ou diferença normalizada (JACKSON; HUETE 1991). As bandas espectrais presentes nos sensores a bordo de satélites ou em espectroradiômetros referem-se a intervalos específicos do espectro eletromagnético com larguras estreitas de comprimentos de onda. Geralmente, o nome da banda espectral refere-se a região do espectro onde ela está situada, como verde, azul, infravermelho próximo e infravermelho termal.

As bandas comumente empregadas nos satélites são compatíveis com as principais feições diagnósticas ou bandas de absorção dos constituintes do solo, nesse caso argila e COS. As bandas 5 (1550 – 1750 nm) e 7 (2080 – 2350 nm) do Landsat 7 e banda 7 (773 – 783 nm) do Sentinel-2 são covariávies altamente úteis para a quantificação de argila (DEMATTÊ et al., 2018; GHOLIZADEH et al., 2018), pois estão próximas as bandas diagnósticas da hematita (~880 nm), goetita (~920 nm) , overtones do grupo O-H (~1400 nm), overtones da molécula da água (~1900 nm) e combinação de tons de Al-OH em 2200 nm (FANG et al., 2018). O teor de COS é espectralmente sensível nas bandas do visível (B4 E B5), infravermelho próximo (B8a) e infravermelho de ondas curtas (B11 e B12) do satélite Sentinel-2, podendo ser predito a partir delas com acurácia aceitável (GHOLIZADEH et al., 2018; VANDOUR et al., 2019). A bandas de absorção da MOS no visível são amplas e ocorrem em virtude dos cromóforos e coloração escura do solo, enquanto no infravermelho próximo e de ondas curtas resultam de overtones e combinação de bandas dos grupos C-H, O-H, N-H, C-O, C=O (STENBERG et al 2010, VISCARRA ROSSEL; BEHRENS 2010; DICK; SANTOS; FERRANTI, 2003).

# Diversos estudos buscaram desenvolver e avaliar índices espectrais para predizer os teores de COS e argila (XIA et al., 2020; HONG et al., 2018; JIN et al., 2017; JIN et al., 2016; BAO et al., 2017; BARTHOLOMEUS et al., 2008). A principal justificativa do uso de índices espectrais é a sua capacidade em extrair bandas espectrais ou feições de forte correlação e sensibilidade com as propriedades do solo (HONG et al., 2018). Além disso, os índices espectrais promovem o refinamento da relação entre os teores de MOS e argila com as feições espectrais (HONG et al., 2018; XIA et al., 2020) reduzem a complexidade e dimensionalidade dos dados espectrais (ANGELOPOULOU et al., 2020), extraem informações valiosas ocultas entre as bandas espectrais (INOUE et al., 2012) e aceleram o tempo de processamento dos métodos estatísticos multivariados (SAWUT et al., 2018).

JIN et al., (2016) estimaram o teor de SOM em uma área agrícola situada na Planície de Sanjiang com grande sucesso através da combinação de algoritmo ótimo de banda, da analise relacional Grey (GRA) e o método multivariado rede neural artificial (ANN). O algoritmo ótimo de banda foi utilizado para analisar a correlação entre o índice de diferença (ID), índice de vegetação da diferença normalizada (NDVI) e índice de vegetação da razão (RVI) com o teor de MOS. O ID apresentou maior correlação com a MOS, sensível à diferença entre os valores espectrais das bandas 364 e 1001, que posteriormente serviu como variável preditora em um modelo de regressão, alcançando um desempenho preditivo promissor (R2 = 0,63 e RMSE = 1,43%). O GRA foi utilizado para selecionar 18 parâmetros dentre os 40 presentes no artigo, que apresentaram maior grau de correlação grey (GCD) e sensibilidade com o MOS. Em seguida, foram introduzidas como covariáveis no ANN para estimar o teor de MOS. A combinação de GRA-ANN resultou em valores de acurácia semelhantes (R2 = 0,90 e RMSE = 0,88%) , comparativamente ao ANN usando todos os parâmetros espectrais ( R2 = 0,91 e RMSE = 0,87%), evidenciando a relevância do GRA em refinar a habilidade de modelos preditivos de SOM.

JIN et al., (2017) obtiveram excelentes resultados na predição de SOM na Planície de Sanjiang através da combinação de diferentes dados de satélites, do método estatístico SVM e do algoritmo de otimização por inteligência de enxames (PSO). Foram utilizados 8 satélites, na qual houve a convolução dos valores espectrais obtidos do espectroradiômetro de acordo com a resolução espectral de cada satélite (dados simulados). O algoritmo de bandas ótimo identificou o índice de vegetação aprimorado (EVI) com o melhor índice espectral acurácia quando associado com os dados simulados do satélite WorldView-2 (R2 = 0,43 e RMSE = 2,62%). O algoritmo PSO foi utilizado para encontrar os valores ótimos de dois parâmetros (C e γ), o qual melhorou a acurácia preditiva do modelo SVM. A estimação de SOM usando o algoritmo PSO-SVM teve desempenho superior ao obtido pelo algoritmo de banda ótimo, pois aumentou a sensibilidade das bandas espectrais as mudanças de SOM ao considerar diversas bandas. Além disso, a acurácia da estimação de SOM com o algoritmo PSO-SVM foi superior quando foi utilizado como variáveis preditoras a combinação de todas as bandas dos satélites e seus respectivos índices espectrais (R2= 0,82 e RMSE= 1,41%) comparativamente ao obtido somente com as bandas e índices espectrais de cada satélite (R2 =0,77 e RMSE= 1,66%).HONG et al., (2018) encontraram que o coeficiente de correlação de 2 dimensões (2D) foi eficiente em extrair informações relevantes para a predição de SOM, em relação ao coeficiente de correlação de 1 dimensão (OCC). Construiu-se o mapa de contorno de 2 dimensões para a representar o coeficiente de determinação (R2) dos índices espectrais testados no estudo com o teor de SOM, o que permitiu a compreensão da combinação entre duas diferentes bandas. Os índices de razão (RI), de diferença (DI) e diferença normalizada (NDI) foram os índices espectrais calculados no artigo para selecionar as bandas espectrais mais sensíveis à MOS. Os 3 índices espectrais e OCC posteriormente foram estabelecidos com os métodos de SVM e ELM (Extreme Machine Learning) para predizer a MOS. Os índices espectrais produziram predições mais acuradas em relação ao OCC, com maiores valores de R2 e RPIQ e menores de RMSE, independentemente do método testado. Em relação ao método, o ELM teve melhor performance ao SVM, independente da variável preditora usada. O melhor desempenho preditivo foi atingido com o método ELM e índice espectral RI (R2 = 0,83; RMSE = 0,45% e RPIQ = 3,49), revelando capacidade de índices espectrais associado com métodos matemáticos adequados em predizer satisfatoriamente a MOS.

XIA et al., 2020 desenvolveram uma predição rápida e acurada da textura do solo em uma área agrícola na Província de Hebei pela combinação de índices espectrais com o algoritmo de projeções sucessivas (SPA). Os índices de razão (RI), de diferença normalizada (NDI), de diferença (DI) e de solo exposto (BSI) foram os índices espectrais empregados para identificar bandas espectrais sensíveis ao teores de argila,silte e areia e analisar o efeito do pre-processamento dos dados espectrais. O SPA destinou-se a selecionar as melhores feições e bandas espectrais com o mínimo de RMSE para serem posteriormente introduzidas no cômputo dos índices espectrais. Em seguida, foram inseridas como covariáveis nos modelos de PLSR e Random Forest (RF) para estimar a textura do solo. A adoção do SPA aumentou a acurácia dos modelos RF e PLSR em comparação ao não uso do SPA, ressaltando a capacidade do algoritmo em extrair informações valiosas de um grande conjunto de dados espectrais. O modelo PLSR estabelecido pelo índice espectral BIS e SPA produziu a melhor predição para argila (R2 = 0,85 e RMSE= 0,57%), resultado semelhante para as frações areia e silte. Os autores observaram o refinamento da correlação entre os índices espectrais e textura do solo com o uso da técnica de pré-processamento MSC.

# 3 HIPÓTESE

1. A redução da multidimensionalidade de dados espectrais com o uso de índices espectrais não prejudica a acurácia preditiva de modelos preditivos de argila e COS;
2. A adição de índices espectrais como covariaveis de modelos preditivos de argila e COS melhoram a sua acurácia preditiva;
3. A técnica de pré-processamento Savitzky-Golay 1ª derivada proporciona modelos de predição de argila e COS mais acurados e robustos quando construídos a partir de métodos multivariados não paramétricos;
4. Os índices espectrais impactam de forma distinta no desempenho preditivo dos métodos estatísticos multivariados testados.

**4 OBJETIVOS**

## 4.1. GERAL

Avaliar o uso dos índices espectrais como covariáveis preditoras em modelos preditivos dos teores de COS e argila em solos subtropicais do Sul do Brasil.

## 4.2. ESPECÍFICOS

1. Testar o efeito da redução da dimensionalidade dos dados espectrais na acurácia dos modelos preditivos de argila e COS;
2. Avaliar o desempenho preditivo de modelos de argila e COS com a adição de índices espectrais;
3. Avaliar o impacto da técnica de pré-processamento Savitzky-Golay 1ª no desempenho preditivo de modelos de argila e COS desenvolvidos a partir métodos multivariados não paramétricos;
4. Avaliar o comportamento dos índices espectrais nos diferentes métodos estatísticos multivariados empregados.

# 

# 5 MATERIAL E MÉTODOS

## 5.1 BIBLIOTECA ESPECTRAL

Este estudo foi realizado nos Estados Rio Grande do Sul (RS) e Santa Catarina (SC), no sul do Brasil. O clima é predominantemente subtropical e temperado (classes de clima Cfa e Cfb de Köppen, respectivamente), segundo ALVARES et al. (2013). Os principais usos da terra incluem lavouras de culturas anuais, florestas plantadas do gênero *Pinus*, florestas nativas e campo nativo com exploração pecuária (MOURA-BUENO et al., 2020; ROSIN et al., 2021).

Uma biblioteca espectral (BES) contendo 2672 amostras foi utilizada (Figura 1). Essas amostras foram coletadas entre 2008 e 2017 em projetos independentes (SAMUEL-ROSA et al., 2013; FLORES et al., 2015; RAMOS et al., 2017; DOTTO et al., 2018; HORST et al., 2017; MOURA-BUENO et al., 2019; FALTA BD DA DANY). As amostras que compõem a BES são originadas de diferentes regiões fisiográficas, mais especificamente das regiões da Depressão Central e Planalto, ambas situadas no RS e a região Serrana de SC. A BES inclui amostras representativas do Sul do Brasil, incluindo amostras de Argissolos, Cambissolos Háplicos, Cambissolos Húmicos, Gleissolos Háplicos, Gleissolos Melânicos, Latossolos Vermelho-Amarelos, Latossolos Vermelhos, Neossolos Litólicos, Neossolos Regolíticos, Nitossolos Vermelhos, Organossolos e Planossolos. O material de origem varia entre basalto, riolito, riodacito, andesito-basalto, arenitos eólicos e fluviais e depósitos fluviais e coluviais em altitudes entre 200 a 1300 metros.

No laboratório todas as amostras foram secas ao ar, destorroadas, passadas em peneira com malha de 2 mm e submetidas à análises químicas. O teor de COS foi determinado via combustão úmida com aquecimento externo (0,5 g de solo + 10 ml K2Cr2O7 0,067 mol L-1) e a titulação dos extratos do COS foi realizada com Fe(NH4)2(SO4)2 6H2O 0,5 mol L-1, na presença de fenantrolina como indicador do fim da titulação, segundo YEOMANS e BREMNER (1988).A composição granulométrica dos solos da BES exibe uma elevada variabilidade, variando de muita arenosa a muito argilosa, conforme demonstrado nas figuras 2 e 4. Amostras da Depressão Central são provenientes de se solos formados a partir de rochas sedimentares (arenitos e argilitos). Nesse caso, tais materiais de origem produziram solos com horizontes superficiais com textura arenosa e subsuperficias de textura média e argilosa, de mineralogia da fração argila prevalecida por caulinita, óxidos de ferro e argilominerais 2:1 com hidróxi-Al entre camadas (SCHMITT, 2015).

Amostras das regiões do Planalto e Serrana são derivadas de solos de textura argilosa e muito argilosa, desenvolvidos a partir de rochas vulcânicas básicas (basalto) e ácidas (riodacito/riolito). Apesar de ambas as regiões compartilharem o mesmo material de origem, diferem em relação a mineralogia da fração argila. Na região do Planalto predominam caulinita e hematita, enquanto na região Serrana além da predominância de caulinita e hematita, também ocorre argilo-minerais 2:1 com hidróxi-entrecamadas, goethita e gibsita (KÄMPF; KLAMT, 1978). A granulometria do solo e, consequentemente o teor de argila foram determinados pelo método da pipeta, utilizando uma solução de NaOH 1 mol L-1 como agente dispersante, conforme descrito por TEIXEIRA et al. (2017).

Os valores de COS apresentaram grande variação, conforme exibido no histograma (Figura 3), fortemente influenciados pela granulometria do solo, uso da terra, condições climáticas e profundidade de amostragem de solo empregada (MOURA-BUENO et al., 2018).

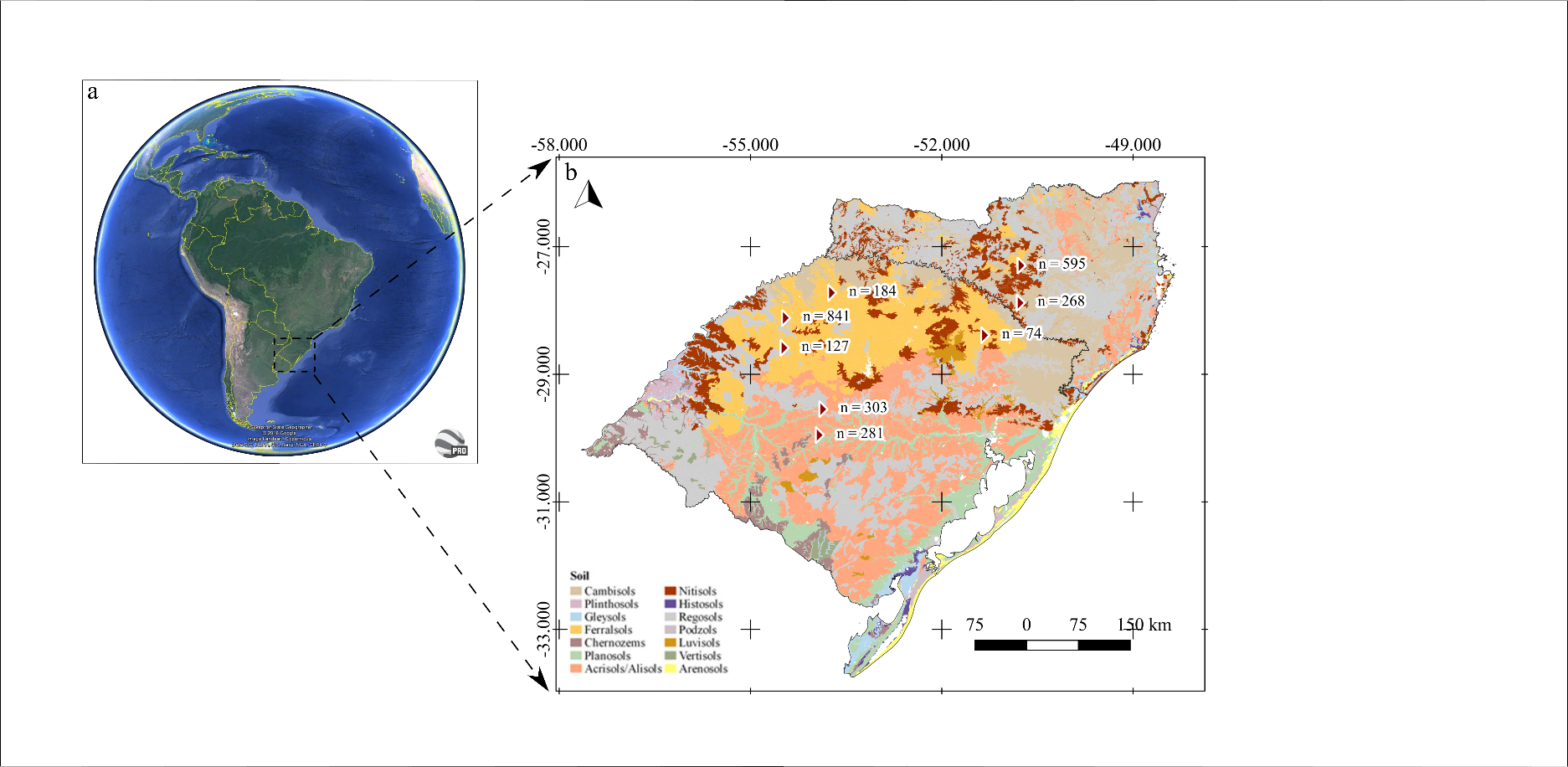
****

Figura 1: Mapa de localização da área de estudo com os pontos.

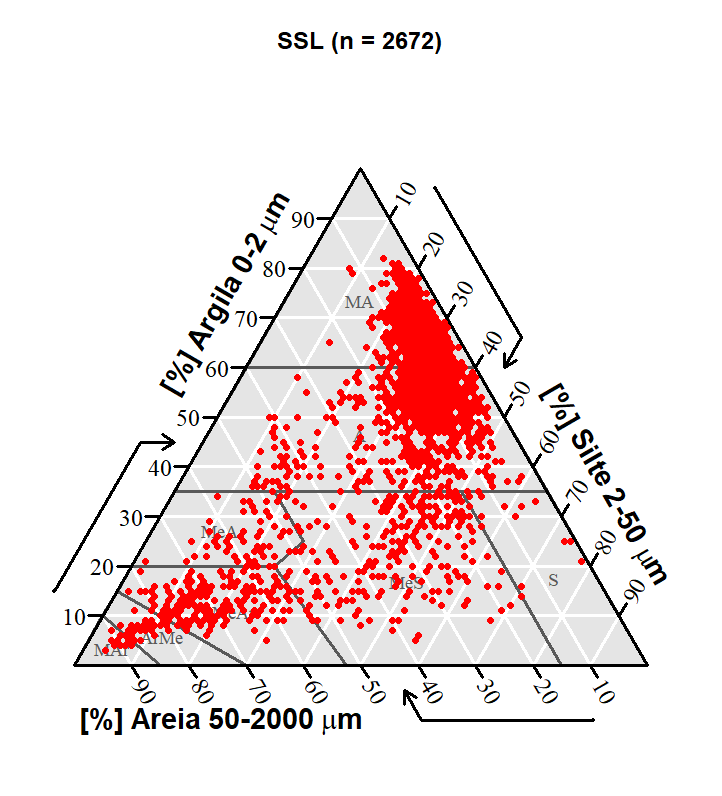
****

Figura 2: Triângulo textural das amostras

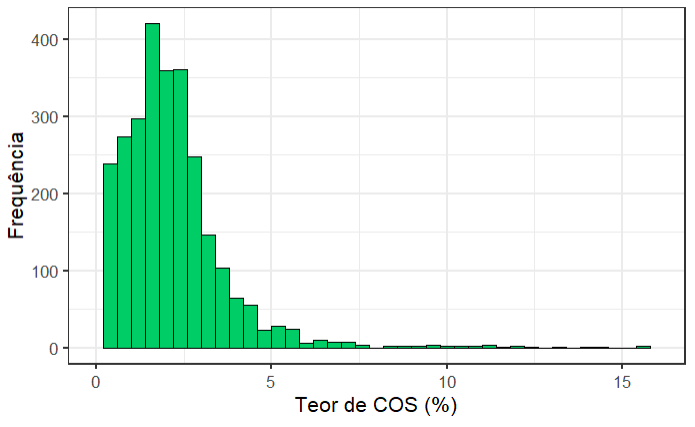


Figura 3: Histograma do teor de COS (%)

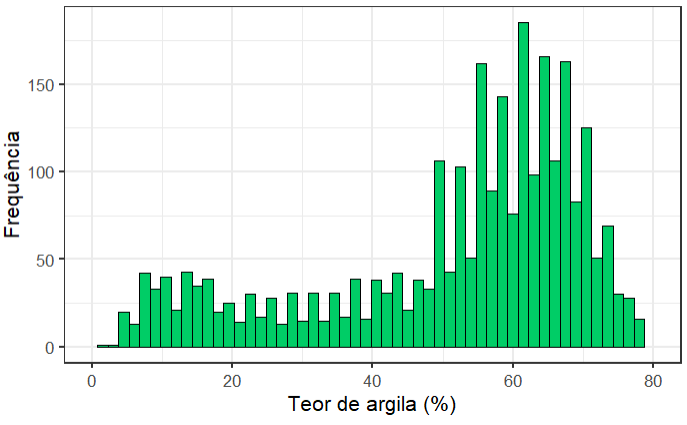


Figura 4: Histograma do teor de argila(%)

## 5.2 LEITURA ESPECTRAL

A leitura da reflectância espectral das amostras de solo foi realizada com o uso do espectroradiômetro FieldSpec 3 (Analytical Spectral Devices, Boulder, USA), com capacidade de amostragem no intervalo de 355 a 2500 nm e resolução espectral de 1 nm. Para realização da leitura foi utilizado aproximadamente 15 cm³ 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, sendo realizando-se duas leituras por amostra, com um giro de 180° entre as leituras, em que cada leitura foi calculada a média a partir de 100 leituras em 10 segundos.

## 5.3 OBTENÇÃO DAS COVARIÁVEIS

Os índices espectrais simulados foram calculados a partir dos dados RAW com base no intervalo de bandas definidos por três diferentes satélites. Sentinel-2 de bandas Costal Blue (443 nm), Blue (490 nm), Green (560 nm), Red (665 nm), Red edge1 (705 nm), Red edge2 (740 nm), Red edge3 (783 nm), Nir (842 nm), Water Vapor (945 nm), Swir Cirrus (1375 nm), Swir1 (1610 nm), Swir2 (2190 nm).

WorldView-3 com as bandas Green (510:580 nm), Yellow (585-625 nm), Red (630-690 nm), Red edge (705-745 nm), Nir1 (770-895 nm), Nir2 (860-1040 nm), Swir1 (1195-1225 nm), Swir2 (1550-1590 nm), Swir3 (1640-1680 nm), Swir4 (1710-1750 nm), Swir5 (2145-2185 nm), Swir6 (2185-2225 nm), Swir7 (2235-2285 nm), Swir8 (2295-2365 nm).

As bandas utilizadas do satélite Landsat foram Costal Blue (430-450 nm), Blue (450-510 nm), Green (530-590 nm), Red (640-690 nm), Nir (850-880 nm), Swir1 (1570-1650 nm), Swir2 (2110-2290 nm).

Os índices espectrais calculados foram Ratio Index (RI) (Equação 5), Difference Index (DI) (Equação 3), Normalized Difference Index (NDI) (Equação 2) HONG et al., (2018), Normalized Difference Vegetation Index (NDVI) (Equação 1), Ratio Index (SRI) (Equação 6) e Difference Index (SDI) (Equação 4). Onde e representam o valor de reflectância nos comprimentos de onda e , respectivamente.

(1)

(2) (3)

(4)

(5)

(6)

5.4 PRÉ-PROCESSAMENTO DE DADOS ESPECTRAIS

A partir das curvas espectrais brutas (RAW), foram aplicadas três técnicas de pré-processamentos comumente empregados em espectroscopia (VASQUES et al., 2008; KNOX et al., 2015; CLAIROTTE et al., 2016): suavização Savitzky-Golay com 1ª derivada usando um polinômio de primeira ordem, com janela de busca de 9 nm, após teste inicial (SGD); padrão normal variado – standard normal variate (SNV); suavização - smoothed utilizando uma janela móvel de 9 nm (SMO). Esses pré-processamentos foram aplicados nas curvas de reflectância de solo na faixa de 355-2500 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 na linguagem de programação R (R Core Team, 2021), utilizando o pacote prospectr (STEVENS e RAMIREZ-LOPEZ, 2013).

## 5.5 PREDIÇÃO DE COS E ARGILA

Os dados do solo foram divididos aleatoriamente em dois conjuntos de dados; um deles foi utilizado no processo de treinamento (70%, n = 1185) e outro no processo de validação (30%, n = 507). A predição dos teores de argila e COS foi realizada pelos modelos supporte vector machine (SVM), regressão linear múltipla (MLR), partial least squares regression (PLSR), considerando como covariáveis: i) dados brutos RAW (555 a 2500 nm); ii) índices espectrais brutos e pré-processados e, iii) RAW + índices espectrais brutos e pré-processados.

## 5.6 AVALIAÇÃO DOS MODELOS DE PREDIÇÃO

O desempenho do modelo foi avaliado aplicando o modelo ajustado aos dados de validação (n = 507). Os parâmetros utilizados para a avaliação foram o coeficiente de determinação (R2) (Equação 7), raiz quadrada do erro quadrático médio (RMSE) (Equação 8) e proporção de desempenho no intervalo interquartil (RPIQ) (Equação 9).

(7)

(8)

(9)

Em que: ŷ = valor predito; ȳ = valor médio observado; y = valor observado; 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.



Graphical user interface, application, Excel

Description automatically generated

Figura 4: Fluxograma da metodologia.

Todo o processamento dos dados de COS e covariáveis ambientais deste trabalho foi realizado no ambiente de estatística computacional R (R CORE TEAM, 2021).

# 6 RESULTADOS E DISCUSSÃO

## 6.1 CORRELAÇÃO ENTRE O COS E OS ÍNDICES E ESPECTROS

## 6.2 CORRELAÇÃO ENTRE A ARGILA E OS ÍNDICES E ESPECTROS

## 6.3 DESEMPENHO DOS MODELOS NA PREDIÇÃO

# 5 CONCLUSÃO

# REFERÊNCIAS

ANGELOPOULOU, T. et al.From Laboratory to Proximal Sensing Spectroscopy for Soil Organic Carbon Estimation—A Review. **Sustentability**, v. 12, n. 2, p. 443, 2020.

# ALLO, M. et al. Prediction of tropical volcanic soil organic carbon stocks by visible-near- and mid-infrared spectroscopy. Catena, v. 189, p. 104452, 2020.

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

# BAO, N. et al. Assessing soil organic matter of reclaimed soil from a large surface coalmine using a field spectroradiometer in laboratory. Geoderma, v. 288, p. 47 – 55, 2017.

# BARTHOLOMEUS, H.M et al. Spectral reflectance based indices for soil organic carbon quantification. Geoderma, v. 145, p. 28 – 36, 2008.

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

CAMPBELL, Patrícia Morais da Matta et al. Digital Soil Mapping of Soil Properties in the “Mar de Morros” Environment Using Spectral Data. **Revista Brasileira de Ciência do Solo**, v. 42, e0170413, 2018.

CHURCHMAN, G. J. Game Changer in Soil Science Functional Role of clay minerals in soil. **Journal of Plant Nutrition and Soil Science**, v. 181 (1), p. 99 – 103, 2018.

CIOTTA, M. N. et al. Matéria orgânica e aumento da capacidade de troca de cátions em solo com argila de atividade baixa sob plantio direto. **Ciência Rural**, v. 33, n.6, p.1161-1164, 2003.

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. Spectroscopy of Rocks and Minerals, and Principles of Spectroscopy. **Remote Sensing for the Earth Science: Manual of Remote Sensing**, Wiley, v. 3, n. 3, p. 3 – 52, 1999.

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

# COBLINSKY, J. A. et al. Prediction of soil texture classes through different wavelength regions of reflectance spectroscopy at various soil depths. Catena, v. 189, p. 104485, 2020.

CORTES, C., VAPNIK, V. Support-vector networks. **Mach Learn**, v.20**,** p. 273–297 1995.

COSTA, M.; KLEIN, C.B. Toxicity and carcinogenicity of chromium compounds in humans. **Critical Reviews in Toxicology**, v. 36, n. 2, p. 155-163, 2006.

# CURCIO, D. et al. Prediction of soil texture distributions using VNIR-SWIR reflectance spectroscopy. Procedia Environmental Sciences, v. 19, p. 494-503, 2013.

DALMOLIN, R. S. D.; TEN CATEN, A.; DOTTO, A. C. Pedometria: uma breve contextualização nacional e mundial. **A Pedometria na Ciência do Solo.** Sociedade Brasileira de Ciência do Solo, Viçosa, v. 43, n. 3, p. 18 – 21, 2017.

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. 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.

DAVANI. M. et al. Simultaneous prediction of several soil properties related to engineering uses based on laboratory Vis-NIR reflectance spectroscopy. **Catena**, v. 197, p. 104987, 2021.

DEMATTÊ, J. A. M. et al. Satellite Spectral Data on the Quantification of Soil Particle Size from Different Geographic Regions. **Revista Brasileira de Ciência do Solo**, v. 42, e0170392, 2018.

DICK, D. P.; SANTOS, J. H. Z.; FERRANTI, E. M.Chemical characterization and infrared spectroscopy of soil organic matter from two southern brazilian soils. . **Revista Brasileira de Ciência do Solo**, v. 27 (1), p. 29-39, 2003.

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. 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.

DOTTO, A.C. et al. Potential of Spectroradiometry to Classify Soil Clay Content. **Revista Brasileira de Ciência do Solo**, v.40, e0151105, 2016.

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.

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

FAO. Healthy soils are the basis for healthy food production. **FAO in action: 2015 International Year of Soils**, p. 2, 2015.

FAO; ITPS. 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**, p.650, 2015.

# 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.

FREI, R. W. Diffuse Reflectance Spectroscopy: Applications, Standards, and Calibration (With Special Reference to Chromatography). **Journal of research of the National Bureau of Standards, Section A: Physics and Chemistry**, v.80 (4), p. 551-565, 1976.

# FREY, D. S. Mycorrhizal Fungi as Mediators of Soil Organic Matter Dynamics. Annual Review of Ecology, Evolution, and Systematics, v. 50, p. 237 – 259, 2019.

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

HERNÁNDEZ-SORIANO, M. C. The Role of Aluminum-Organo Complexes in Soil Organic Matter Dynamics. In: ­\_\_\_\_\_\_. (Org.). **Soil Health and Land Use Management**. 1. ed. [S.1.]: IntechOpen, 2012, cap. 2, p. 17-32.

HEUVELINK, G. The definition of Pedometrics. In: GRUNWALD, S. (Ed.). **Pedometron** - Issue 15. Gainesville: International Working Group on Pedometrics – Provisional Commission on Pedometrics of the International Union of Soil Sciences (IUSS), 2003. p. 11-12. Disponível em: <http://www.pedometrics.org/Pedometron/ pedometron15.pdf>. Acesso em: 30 ago. 2020.

# HAGHI, R. K; PÉREZ-FERNÁNDEZ, E; ROBERSTON, A. H. J. Prediction of various soil properties for a national spatial dataset of Scottish soils based on four different chemometric approaches: A comparison of near-infrared and mid-infrared spectroscopy. Geoderma, v. 396, 115071, 2021.

HONG, Y. et al.. Rapid identification of soil organic matter level via visible and near-infrared spectroscopy: Effects of two-dimensional correlation coefficient and extreme learning machine**. Sci Total Environ**, v. 644, p. 1232 – 1243, 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.

# INOUE, Y et al. Diagnostic mapping of canopy nitrogen content in rice based on hyperspectral measurements. Remote Sensing of Environment, v. 126, p. 210 – 221, 2012.

# JACKSON, R.D; HUETE, A.R. Interpreting vegetation indices. Preventive Veterinary Medicine, v. 11 (3 – 4), p. 185 – 200, 1991.

# JACONI, A.; CORA, V.; DON, A. Near infrared spectroscopy as an easy and precise method to estimate soil texture. Geoderma, v. 337, p. 906-913, 2019.

# JIN, X. et al. Comparison of different satellite bands and vegetation indices for estimation of soil organic matter based on simulated spectral configuration. Agricultural and Forest Meteorology, v. 244 – 245, p. 57 – 61, 2017.

# JIN, X. et al. Remote estimation of soil organic matter content in the Sanjiang Plain, Northest China: The optimal band algorithm versus the GRA-ANN model. Agricultural and Forest Meteorology, v. 218 – 219, p. 250 – 260, 2016.

JOHNSON, J. M. et al. Near-infrared, mid-infrared or combined diffuse reflectance spectroscopy for assessing soil fertility in rice fields in sub-Saharan Africa. **Geoderma**, v. 354, p. 113840, 2019.

KÄMPF, N.; KLAMT, E. Mineralogia e Gênese de Latossolos (Oxisols) e Solos Podzólicos da Região Nordeste do Planalto Sul-Riograndense. **Revista Brasileira de Ciência do Solo**, v.2, p.68-73, 1978.

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.

# LAAMRANI, A. et al. Ensemble Identification of Spectral Bands Related to Soil Organic Carbon Levels over an Agricultural Field in Southern Ontario, Canada. Remote Sensing, v. 11, n. 11, p. 1298, 2019.

# LAAZAR, A. et al. The manifestation of VIS-NIRS spectroscopy data to predict and mapping soil texture in the Triffa plain (Morocco). Kuwait Journal of Science, v. 48, n. 1, p. 127-137, 2021.

LAL. R et al. The carbon sequestration potential of terrestrial ecosystems. **Journal of Soil and Water Conservation**, v. 73, n. 6, p. 145-152, 2018.

LAZAAR, A. et al. The application of proximal visible and near-infrared spectroscopy to estimate soil organic matter on the Triffa Plain of Morocco**. International Soil and Water Conservation Research**, v. 8, n. 2, p. 195 – 204, 2020.

LAZZARETTI, B. P. et al. Prediction of soil organic matter and clay contents by near-infrared spectroscopy - NIRS.**Ciência Rural**, Santa, v. 50, n.1, e20190506, 2020.

LEHMANN, J.; KLEBER, M. The contentious nature of soil organic matter. **Nature**, v. 528**,** p. 60–68, 2015.

# LEVI, N.; KARNIELI, A.; PAZ-KAZAN, T. Using reflectance spectroscopy for detecting land-use effects on soil quality in drylands. Soil and Tillage Research, v. 199, p. 104571, 2020.

LI, S. et al. In Situ Measurements of Organic Carbon in Soil Profiles Using vis-NIR Spectroscopy on the Qinghai−Tibet Plateau. **Environ. Sci. Technol.** v. 49 (8), p. 4980—4987, 2015.

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

# MENDES, W. S. et al. Soil spectral library of Piauí using machine learning for laboratory analysis in Northeastern Brazil. Revista Brasileira de Ciência do Solo, v. 45, e0200115, 2021.

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

MORRELOS, A. et al. Machine learning based of soil total nitrogen, organic carbon and moisture content by use vis-NIR spectroscopy. **Biosystemseng**, v. 152, p. 104 – 116, 2016.

# 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.

# MOURA-BUENO, J. M. et al. When does stratification of a subtropical soil spectral library improve predictions of soil organic carbon content? Science of The Total Environment, v. 737, p. 139895, 2020.

MOURA-BUENO, J. M. **Uso de Bibliotecas Espectrais para a predição do carbono orgânico do solo**. 2018. 201p. Tese. (Doutorado em Ciência do Solo). Universidade Federal de Santa Maria, Santa Maria, RS, 2018.

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

# NAWAR, S.; MOUAZEN, A. M. On-line vis-NIR spectroscopy prediction of soil organic carbon using machine learning. Soil and Tillage Research, v. 190, p. 120 – 127, 2019.

NEWMAN, A.C.D. The significance of clays in agriculture and soils. **Philosophical Transactions of the Royal Society of London, Series A - Mathematical Physical and Engineering Sciences**, v. 311 (1517), p. 375 – 389, 1984.

# NOCITA, M. et al. Chapter Four - Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring. Advances in Agronomy, v. 132, p. 139 - 159, 2015.

OLIVEIRA, J. F. et al. Soil discrimination using diffuse reflectance Vis-NIR spectroscopy in a local toposequence. **Comptes Rendus Geoscience**, v. 345, n. 11 – 12, p. 446 – 453, 2013.

OSMAN, K.T. Soils: **Principles, Properties and Management**. 1. ed. New York: Springer, 2012, p. 274, 2012.

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.

RAMIREZ-LOPEZ, L. et al. Robust soil mapping at the farm scale with vis–NIR spectroscopy. **European Journal of Soil Science**, v. 70, n. 2, p. 378-393, 2019.

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, e0160189, 2017.

REDA, R. et al. Comparing CalReg performance with other multivariate methods for estimating selected soil properties from Moroccan agricultural regions using NIR spectroscopy. **Chemometrics and Intelligent Laboratory Systems**, v. 211, p.104277, 2021.

ROGGO, Y. et al. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. **Journal of Pharmaceutical and Biomedical Analysis**, v. 44, n. 3, p. 683-700, 2007.

# ROPER, W. R. et al. Comparing Four Methods of Measuring Soil Organic Matter in North Carolina Soils. Soil Science Society of America Journal, v. 83 (2), p. 466-474, 2019.

# ROSIN, N. A. et al. Diffuse reflectance spectroscopy for estimating soil organic carbon and make nitrogen recommendations. Scientia Agricola, v. 78, n. 5, e20190246, 2021.

SAMUEL-ROSA, A.; DALMOLIN, R. S. D.; MIGUEL, P. Building predictive models of soil particle-size distribution**. Revista Brasileira Ciência do Solo**, v. 37, n. 2, p. 422-430, 2013.

SANCHEZ P. A. et al. Digital Soil Map of the World. **Science**, v. 325, n. 5941, p. 680 – 681, 2009.

SARKAR, B. et al. Clay Minerals – Organic Matter Interactions in Relation to Carbon Stabilization in Soils. In: GARCIA, C.; NANNIPIERI, P.; HERNANDEZ, T. **The future of Soil Carbon:** Its Conservation and Formation. 1. ed. London: Academic Press, 2018, cap. 3, p. 71 – 86.

# SAWUT, R. et al. Possibility of optimized indices for the assessment of heavy metal contents. International Journal of Applied Earth Observation and Geoinformation, v. 73, p. 14 – 25, 2018.

SCHMIDT, M. W. I. et al. Persistence of soil organic matter as an ecosystem property. **Nature**, v. 478, p. 49-56, 2011.

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.

SCHULZE, D.G. et al. Significance of organic matter in determining soils colors. In: BIGHAM, J.M.; CIOLKOSZ, E.J. (Ed.). **Soil color**. Madison: Soil Science Society of America, 1993. p. 71-90.

# SEEMA, A. K.; GHOSH, B. S. D.; REDDY, N. Application of VIS-NIR spectroscopy for estimation of soil organic carbon using different spectral preprocessing techniques and multivariate methods in the middle Indo-Gangetic plains of India. Geoderma Regional, v. 23, p. e00349, 2020.

SILVA, E. B. et al. 2019 A Regional Legacy Soil Dataset for Prediction of Sand and Clay Content with Vis-Nir-Swir, in Southern Brazil.**Revista Brasileira de Ciência do Solo**, v. 43, e0180174, 2019.

# SIX, J. et al. Stabilization mechanisms of soil organic matter: Implications for C-saturation of soils. Plant and Soil, v. 241, p. 155–176, 2002.

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.

STENBERG, B. et al. 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.

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

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. Estatística multivariada aplicada à diminuição do número de preditores no mapeamento digital do solo. **Pesquisa agropecuária brasileira**, v. 46, n. 5, p. 553-561, 2011.

# TERRA, F. S.; DEMATTÊ, J. M. A.; VISCARRA ROSSEL, R. A. Spectral libraries for quantitative analyses of tropical Brazilian soils: Comparing vis–NIR and mid-IR reflectance data. Geoderma, v. 255-256, p. 81-93, 2015.

TISDALL, J. M.; OADES, J. M. Organic matter and water-stable aggregates in soils. **European Journal of Soil Science**, v. 33, n. 2, p. 141-163, 1982.

VANDOUR, E. et al. Sentinel – 2 image capacities to predict common topsoil properties of temperate and Mediterranean agroecosystems. **Remote Sensing of Environment**, v. 223, p. 21-33, 2019.

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.; 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.

VISCARRA ROSSEL, R. A. et al. Using a legacy soil sample to develop a mid-IR spectral library. **Australian Journal of Soil Research.** CSIRO Publishing, v. 46, p. 1 – 16, 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.

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.

# 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.

# XIA, K. et al. Optimization of a Soil Particle Content Prediction Model Based on a Combined Spectral Index and Successive Projections Algorithm Using Vis-NIR Spectroscopy. Spectroscopy, v. 35 (12), p. 24 – 34, 2020.

XU, S. et al. Comparison of multivariate methods for estimating selected soil properties from intact soil cores of paddy fields by Vis–NIR spectroscopy. **Geoderma**, v. 310, p. 29-43, 2018.

YEOMANS, J. C.; BREMNER, J. M. A rapid and precise method for routine determination of organic carbon in soil. **Communications in soil science and plant analysis**, New York, v. 19, n. 13, p. 1467-1476, 1988.

ZHONG, Z. et al. Relationship between Soil Organic Carbon Stocks and Clay Content under Different Climatic Conditions in Central China. **Forests**, v. 9 (10), p. 598, 2018.