UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

FACULDADE DE AGRONOMIA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DO SOLO

ESPECTROSCOPIA DE REFLECTÂNCIA PARA ANÁLISE TEXTURAL DE AMOSTRAS DE SOLO LEGADAS DO ESTADO DE SANTA CATARINA

Elisângela Benedet da Silva

(Tese)

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

FACULDADE DE AGRONOMIA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DO SOLO

ESPECTROSCOPIA DE REFLECTÂNCIA PARA ANÁLISE TEXTURAL DE AMOSTRAS DE SOLO LEGADAS DO ESTADO DE SANTA CATARINA

ELISÂNGELA BENEDET DA SILVA

Engenheira Agrônoma (UFSC)

Mestre em Engenharia Civil (UFSC)

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

Porto Alegre (RS) Brasil

Maio de 2018

ELISÂNGELA BENEDET DA SILVA
Engenheira Agrônoma – UFSC
Mestre em Engenharia Civil – UFSC

TESE

Submetida como parte dos requisitos

para obtenção do Grau de

DOUTORA EM CIÊNCIA DO SOLO

Programa de Pós-Graduação em Ciência do Solo

Faculdade de Agronomia

Universidade Federal do Rio Grande do Sul

Porto Alegre (RS), Brasil

Aprovado em: 25.05.2018
Pela Banca Examinadora

Homologado em: 27.07.2018
Por

ELVIO GIASSON
Orientador-PPG Ciência do Solo
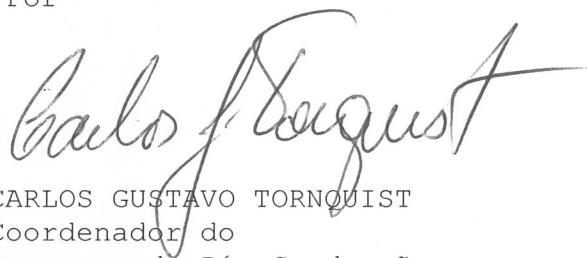
CARLOS GUSTAVO TORNQUIST
Coordenador do
Programa de Pós-Graduação em
Ciência do Solo

TALES TIECHER
Departamento de Solos/UFRGS

via Skype

ALEXANDRE TEN CATEN
UFSC

ANDRÉ CARNIELETTO DOTTO
ESALQ/USP
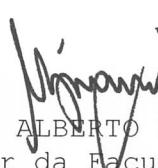
CARLOS ALBERTO BISSANI
Diretor da Faculdade
de Agronomia

"[...] a felicidade está menos nos resultados obtidos do que no processo de busca, seguindo a dialética do desejo. O desejo tem a tessitura da busca permanente, pois, uma vez satisfeito, deixa de ser desejo e consegue sempre menos do que queria. Todo resultado é decepcionante, porque é mais extenso do que intenso. [...] Para sermos felizes temos de inventar grandes causas, ou pelo menos cercar a rotina com alguma surpresa. Os resultados apenas se deterioram, ao passo que os processos transformam."

Pedro Demo

AGRADECIMENTOS

Gostaria de agradecer ao meu orientador prof. Élvio Giasson por ter me recebido com atenção e amizade, pelo apoio, confiança e paciência durante todos os momentos do curso; aos meus co-orientadores Ivan Bacic pela amizade e confiança, supervisão na Epagri, e incentivo desde o meu mestrado, sem ele eu não teria chegado aqui e ao prof. Alexandre Demattê pela oportunidade e apoio técnico.

Ao prof. Alexandre ten Caten pela amizade, longas conversas e sugestões técnicas, essenciais para o desenvolvimento deste trabalho; ao amigo André Dotto por ter me mostrado os caminhos do R e pela pronta ajuda nos momentos de dúvida. Ao prof. Milton da Veiga pelo incentivo através da concessão das amostras de solo.

Agradeço a Epagri por fomentar a formação de pesquisadores doutores através do seu Programa de Pós-Graduação, pelo apoio financeiro e confiança; ao Departamento de Solo da UFRGS e ao Programa de Pós-Graduação em Ciência do solo, seus professores e ao secretário Jader por todo suporte durante o curso; ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e a Coordenação de Aperfeiçoamento de Nível Superior (CAPES), pela concessão da bolsa de doutorado no Brasil.

Agradeço aos amigos do grupo de Gênese e Classificação do solo, Janderson Costa, Alcinei Campos, Benito Bonfatti, Israel Machado, Fabrício Coelho, João Coblinski e ao amigo Paulo Ricardo da Luz pela amizade, apoio nas saídas de campo e pelos memoráveis momentos compartilhados; aos amigos da pós-graduação em Ciência do Solo (PPGCS), em Sensoriamento Remoto (PPGSR) e em Engenharia de Minas, Metalurgia e Materiais (PPGE3M) pelo agradável convívio; ao pessoal do Grupo de Geotecnologia (GeoCis/USP/ESALQ) pelo apoio durante a realização das análises espectrais e amizade; ao pessoal do grupo de pesquisa em Ciência do Solo da UFSC-Curitibanos pelo apoio técnico e as boas conversas.

Agradeço aos meu pais, Agenor e Matilde, pelo amor incondicional e incentivo aos estudos desde o jardim de infância; aos meus tios, Lídio e Luzia que me receberam em Porto Alegre e me apoiaram com amor e paciência; a toda a minha FAMÍLIA e AMIGOS que me deram suporte nessa caminhada.

Agradeço, especialmente ao meu marido André e ao meu filho Gustavo, pelo amor sincero e por terem aceitado a minha ausência, mesmo quando eu estava presente.

Agradeço a Deus e ao Plano Espiritual pelos desafios e bênçãos recebidos durante minha vida.

E certa de que "uma tese não se faz sozinha", meu muito obrigada.

ESPECTROSCOPIA DE REFLECTÂNCIA PARA ANÁLISE TEXTURAL DE AMOSTRAS DE SOLO LEGADAS DO ESTADO DE SANTA CATARINA[1]

Autora: Elisângela Benedet da Silva
Orientador: Elvio Giasson

RESUMO

As demandas globais por informação de solos e para a construção de grandes bases de dados têm impulsionado a aplicação de abordagens quantitativas para a estimativa dos atributos do solo. Nesse contexto, a espectroscopia de reflectância (ER) surge como uma técnica rápida, não destrutiva, acurada e de melhor relação custo-benefício que as técnicas de análise convencionais. Entretanto, no Estado de Santa Catarina, os estudos de solos em que a ER foi aplicada são pouco abrangentes. Nessa pesquisa, a ER na região do VIS-NIR-SWIR foi empregada para a predição do conteúdo de areia e argila a partir de um conjunto de amostras de solos pré-existentes coletadas em áreas de lavoura, abrangendo cerca de 90% dos municípios catarinenses. No capítulo 1 foi realizada uma revisão de literatura sobre a produção científica da ER no âmbito da ciência do solo, destacando as pesquisas nas regiões do VIS-NIR-SWIR e MIR. A abordagem espectroscópica foi aplicada e avaliada em dois estudos distintos: (i) na comparação do desempenho de 5 modelos estatísticos multivariados e de 6 técnicas de pré-processamento dos espectros na predição dos teores de areia e argila do solo (Capítulo 2) e, (ii) na avaliação do uso dos teores de argila preditos para a alocação das amostras de solos em classes texturais distintas para fins de enquadramento nos programas de seguro agrícola (Capítulo 3). Os resultados alcançados confirmaram o grande potencial da ER associada as técnicas quimiométricas na predição da textura do solo usando um conjunto legado de amostras de solos. Os modelos de predição do conteúdo de argila foram mais precisos que os modelos de predição do conteúdo de areia. Em geral, as técnicas de pré-processamento por correção da dispersão apresentaram melhor desempenho que as técnicas com aplicação de derivadas. Os espectros não pré-processados apresentaram bons resultados, principalmente na predição do conteúdo de areia do solo. Entre os modelos avaliados, o Cubist apresentou o melhor desempenho na predição do conteúdo de areia (Cubist>RF>SVM>PLS>GPR) e argila (Cubist>SVM>GPR>RF>PLS). A ER mostrou-se útil na alocação das classes texturais indicando o grande potencial da técnica para uso nos laboratórios de rotina. Além disso, este estudo contribui para avançar na caracterização do comportamento espectral dos solos de Santa Catarina.

SOIL SPECTROSCOPY FOR TEXTURAL ANALYSIS OF LEGACY SOILS SAMPLES OF SANTA CATARINA STATE[1]

Author: Elisângela Benedet da Silva
Adviser: Elvio Giasson

ABSTRACT

Global demands for soil information and for the development of large databases has led to the application of the quantitative approaches for estimating soil attributes. In this context, soil spectroscopy appears as a fast, non-destructive, accurate and cost-effective technique to predict soil attributes. However, in the Santa Catarina State (SC), Brazil, soil studies in which soil spectroscopy was applied are not wide coverage limited to local scale.. In this research, VIS-NIR-SWIR reflectance spectroscopy was used to predict sand and clay contents from a set of legacy soils samples collected in farming areas, covering approximately 90% of the municipalities of SC. The spectroscopic approach was evaluated in two different studies: (i) to compare the performance of six preprocessing techniques and five multivariate methods in the prediction of sand and clay content (Chapter 2) and (ii) to use the predicted clay content values to classify soil into textural classes for agricultural insurance programs (Chapter 3). In addition, it was carried out a review of soil spectroscopy in the VIS-NIR-SWIR and MIR region based on scientific literature in soil science (Chapter 1). The results confirmed the potential of soil spectroscopy approach associated to chemometrics methods to predict soil attributes using heterogeneous soil samples. Models for predicting clay content were better than models to predict sand. In general, the scatter-corrective preprocessing techniques performed better than derivatives techniques. RAW spectra presented competitive results, mainly in the prediction of sand content. Among methods applied, the Cubist presented the best performance in the prediction of the two attributes. Reflectance spectroscopy was useful in the allocation of textural classes indicating its great potential for use in routine laboratories. In addition, this study contributes to improve the knowledge of the spectral behavior of the soils of Brazilian soils.

SUMÁRIO

RELAÇÃO DE TABELAS

RELAÇÃO DE FIGURAS

1. INTRODUÇÃO GERAL

O solo é um sistema complexo e heterogêneo que exerce papel fundamental nos ecossistemas terrestres. É reconhecido como um dos principais recursos naturais que fornece serviços ecossistêmicos necessários ao desenvolvimento da humanidade, tais como produção de alimentos, fonte de matéria prima (água e minerais), regulação e mitigação climática, sequestro de carbono, armazenamento, filtragem e transformação de nutrientes e água, biodiversidade, ambiente físico para a construção das habitações e herança cultural (BOUMA; McBRATNEY, 2013; HEMPEL et al., 2014). Entretanto, a degradação dos solos tem resultado no declínio da sua capacidade de produzir bens econômicos e fornecer serviços ecossistêmicos (LAL, 2010). A preservação e o manejo sustentável desse recurso natural não renovável são cruciais para evitar as principais ameaças ao solo e que põem em risco a manutenção da vida na terra, como a segurança alimentar, as mudanças climáticas, a degradação ambiental, a escassez de água e a biodiversidade (SANCHEZ et al., 2009).

Essas importantes questões sociais e ambientais têm aumentado a demanda global por grandes quantidades de dados de solo (locais amostrados e atributos analisados) de boa qualidade, consistência e relevância (HARTEMINK; McBRATNEY, 2008; HEMPEL et al., 2014). Para atender a essa crescente demanda, grandes bases de dados precisam ser desenvolvidas, mas o uso de métodos analíticos convencionais pode tornar esse processo lento e dispendioso. O uso do sensoriamento próximo aplicado ao solo e os contínuos avanços tecnológicos têm modernizado a pedologia, assim como os métodos de análises de solos (HARTEMINK; MINASNY, 2014). Há um amplo interesse na aplicação da espectroscopia de reflectância (ER, ou espectroradiometria) na

região do visível (VIS), infravermelho-próximo (NIR), -de ondas curtas (SWIR) (350-2500 nm) e o -médio (MIR) (2500-25000 nm) do espectro eletromagnético para geração dessas grandes bases e armazenamento dos dados. Esse interesse está fundamentado no uso de sensores proximais, instalados em laboratórios (ROMERO et al., 2018) ou diretamente no campo (POGGIO; BROWN; BRICKLEMYER, 2017; VISCARRA ROSSEL et al., 2017) que captam parte da radiação eletromagnética (REM) que é refletida, emitida ou espalhada pelo objeto (CLARK, 1999). Além disso, a ER é considerada uma técnica rápida e de fácil reprodução, não destrutiva, acurada e de melhor relação custo-benefício que as técnicas de análise convencional (Viscarra Rossel et al., 2006; Viscarra Rossel et al., 2009).

As informações de ER do solo têm sido empregadas no desenvolvimento de bibliotecas espectrais em escalas locais, regionais, nacionais, continentais e globais (ADAMCHUK et al., 2004; BELLINASO; DEMATTÊ; ROMEIRO, 2010; BROWN et al., 2006; CLAIROTTE et al., 2016; DOTTO et al., 2018; GOGÉ et al., 2012; LUCÀ et al., 2017; NOCITA et al., 2014b; RAMIREZ-LOPEZ et al., 2013; SATO, 2015; SHEPHERD; WALSH, 2002; STEVENS et al., 2013; VISCARRA ROSSEL et al., 2016; VISCARRA ROSSEL; WEBSTER, 2012).Entretanto, os custos de criação de grandes bases de dados são também altos, de modo que, na grande maioria, essas bibliotecas são desenvolvidas a partir de amostras de solo históricas e bases de dados legadas disponibilizadas pelas universidades, centros de pesquisa e agências governamentais, entre outras (VISCARRA ROSSEL; WEBSTER, 2012). Essas bibliotecas representam uma rica e valiosa fonte de informação sobre as características do solo no espaço e no tempo (BAXTER; CRAWFORD, 2008), tornaram-se ferramentas poderosas, ajudando a analisar e armazenar grandes quantidades de informações de maneira eficiente (McBRATNEY; MENDONÇA-SANTOS; MINASNY, 2003; VISCARRA ROSSEL et al., 2016).

O estado de Santa Catarina não possui uma biblioteca espectral de solos. As pesquisas na área da ER são incipientes no Estado e aplicadas apenas em escala local (DOTTO et al., 2017, 2018). A necessidade de preencher a lacuna sobre o comportamento espectral dos solos catarinenses

em escala regional e contribuir para o avanço dessa área na ciência do solo brasileira fundamentam o desenvolvimento dessa pesquisa.

O objetivo geral dessa pesquisa foi aplicar e avaliar espectroscopia de reflectância na região do VIS-NIR-SWIR (350-2500 nm) utilizando diferentes modelos estatísticos multivariados para a predição dos teores de areia e argila no solo em escala regional utilizando amostras de solos legadas.

Os seguintes objetivos específicos foram definidos: (i) estimar o conteúdo de areia e argila para gerar informação sobre o solo usando uma biblioteca espectral na região do VIS-NIR-SWIR (350-2500 nm), (ii) comparar o desempenho de diferentes técnicas de pré-processamento espectrais e modelos estatísticos multivariados na predição do conteúdo de areia e argila do solo, e (iii) investigar o potencial da ER para a classificação dos teores de argila preditos em classes texturais para fins de seguro agrícola.

No primeiro capítulo é apresentada uma revisão de literatura sobre a produção científica no âmbito da ciência do solo, onde a ER, na região do VIS-NIR-SWIR e MIR, é o aspecto principal. Além disso, foram associados os resultados de pesquisa aos pesquisadores, e aos conhecimentos e inovações que eles produziram através de uma avaliação específica das contribuições mais relevantes durante o período analisado.

No segundo capítulo, é apresentado um estudo que compara modelos baseados na ER para a predição dos teores de argila e areia em escala regional. Seis técnicas de pré-processamentos dos espectros foram aplicadas para diminuir os ruídos e não linearidades entre os espectros e o atributo avaliado e cinco modelos estatísticos multivariados foram desenvolvidos e comparados.

No terceiro capítulo, é apresentado um estudo sobre a aplicabilidade do uso da abordagem ER na região do VIS-NIR-SWIR, associada a um modelo multivariado não linear (SVM), na predição dos teores de argila e alocação do conteúdo de argila predito das amostras de solos em classes texturais distintas. A Normativa nº 02/2008 do Ministério da Agricultura, Pecuária e Abastecimento (MAPA) foi utilizada como referência para o enquadramento das classes texturais exigido pelos programas de seguro agrícola brasileiros.

2.CAPÍTULO I - REVISÃO BIBLIOGRÁFICA
ANÁLISE BIBLIOMÉTRICA DA ESPECTROSCOPIA
DE REFLECTÂNCIA NA CIÊNCIA DO SOLO

2.1. Introdução

A espectroscopia de reflectância (ER) é uma técnica de sensoriamento próximo e que vem sendo usada para estudar a relação empírica entre as curvas espectrais dos solos e seus atributos físico-químicos, mineralógicos e biológicos. Isso é possível devido à interação dos componentes do solo com a energia eletromagnética refletida. Essa técnica teve grandes contribuições na ciência do solo com o trabalho do pesquisador Condit (1970) que utilizou a análise espectral para descrever e determinar padrões de curvas espectrais nos solos americanos. Nesse estudo, ele definiu três padrões de curvas espectrais. No mesmo ano, Hunt; Salisbury (1970) mediram sistematicamente espectros de minerais e rochas e relacionaram a energia refletida por eles com suas frações químicas e mineralógicas encontradas nas amostras. Na década de 1980, Stoner; Baumgardner (1981) complementaram o trabalho de Condit (1970), adicionando mais dois tipos, definindo cinco tipos de curvas espectrais médias para solos de acordo com o conteúdo de matéria orgânica, óxidos de ferro e a composição mineralógica. Esses estudos pioneiros focaram principalmente em interpretação e classificação das curvas espectrais dos solos.

Nas décadas seguintes, as pesquisas sobre o uso da ER na região do VIS-NIR-SWIR e MIR aumentaram rapidamente em número de publicações e com foco direcionado para uma abordagem mais quantitativa com a predição de vários atributos do solo (BENDOR; BANIN, 1995; BROWN et al., 2006; CONFORTI et al., 2014; DEMATTÊ; TERRA, 2014; NAWAR et al., 2016; RAMIREZ-LOPEZ et al., 2013; SANTANA; SOUZA; POPPI, 2018; SHEPHERD; WALSH, 2002; STENBERG, 2010; VASQUES; GRUNWALD; SICKMAN, 2008; VISCARRA ROSSEL et al., 2006, 2014), utilizando sensores em laboratório (BENDOR; ONG; LAU, 2015; ROMERO et al., 2018), e a campo (*in situ* e *on-line*) (KUANG et al., 2012). O mesmo ocorreu no desenvolvimento de bibliotecas espectrais formadas a partir do armazenamento digital dos dados espectrais associados aos respectivos resultados analíticos, em diversas escalas (ARAÚJO et al., 2014; CAMBULE et al., 2012; FRANCESCHINI et al., 2013; GHOLIZADEH et al., 2016; KUANG; MOUAZEN, 2013; NAWAR et al., 2016; NOCITA et al., 2014a; RAMIREZ-LOPEZ et al., 2013; STENBERG, 2010; VISCARRA ROSSEL et al., 2016; WARURU et al., 2014; ZHOU et al., 2014)

Todos esses estudos levam a um melhor entendimento do comportamento espectral dos solos, da capacidade de predição para os principais atributos de interesse agrícola (NOCITA et al., 2015) e ambientais (SHI et al., 2014; TIECHER et al., 2017), bem como das limitações técnicas que ainda precisam ser trabalhadas (SORIANO-DISLA et al., 2014). Além disso, essa abordagem mostrou ser uma boa alternativa para complementar ou substituir os métodos convencionais de análise de solo, quando é necessário um grande número de amostras, respostas rápidas para o processo de tomada de decisão ou para expandir o conhecimento sobre os solos em áreas que atualmente são excluídas dos levantamentos por limitações econômicas (BENDOR; BANIN, 1995; SUMMERS et al., 2011; VISCARRA ROSSEL et al., 2006).

Nesse contexto, é oportuno avaliar o estado atual da pesquisa científica na área da ER na ciência do solo. Uma forma de avaliar o impacto de uma disciplina científica é através do número de publicações produzidas ao

longo do tempo, que é um indicador da produtividade (HARTEMINK; McBRATNEY, 2008). Embora as publicações científicas sejam apenas uma parte do processo de pesquisa científica, elas fornecem informações sobre a natureza e a evolução dessa pesquisa (BELTER; SEIDEL, 2013). O objetivo desta análise bibliométrica foi apresentar uma visão geral da produção científica no âmbito da ciência do solo, destacando as pesquisas relacionadas a ER nas regiões do VIS-NIR-SWIR e MIR. Além disso, foram associados os resultados de pesquisa aos pesquisadores, e aos conhecimentos e inovações que eles produziram através de uma visão específica das contribuições mais relevantes do período de 1973 a 2018.

2.2. Materiais e métodos

2.2.1. Análise bibliométrica da produção científica

A revisão envolveu cinco etapas, adaptadas de Zupic; Cater (2015). Na primeira etapa, foi elaborada a pergunta de pesquisa e definido o método apropriado para respondê-la. A pergunta de pesquisa foi descrita como objetivo desse estudo. O método escolhido foi a análise bibliométrica realizada com o pacote *bibliometrix* (ARIA; CUCCURULLO, 2017) implementado no R (R CORE TEAM, 2017). O pacote *bibliometrix* fornece uma série de ferramentas para análise bibliométrica e cientométrica. Ele permite a extração de dados de duas das principais bases de dados disponíveis online, denominadas *Clarivate Analytics Web of Science* (WoS at http://www.webofknowledge.com) e *Scopus* (http://www.scopus.com) (CUCCURULLO; ARIA; SARTO, 2016).

Na segunda etapa, foram compilados os dados bibliométricos. Essa etapa tratou da recuperação baseada na pesquisa de tópicos usando coringas (asteriscos) e o operador *AND* para as seguintes palavras-chave "*soil*" *AND* "*diffuse reflectance spect**" encontradas nos títulos, palavras-chave e resumos nas duas bases de dados bibliográficos: *WoS* e *Scopus*. Os seguintes filtros de busca foram aplicados para composição da matriz de referências: (i) intervalo de

tempo: todos os anos (1945-2018 para *WoS* e 1960-2018 para *Scopus*); (ii) tipos de documentos: artigos e artigos em impressão; (iii) apenas artigos escritos em inglês e português. A definição do período se fez necessária, pois a análise bibliométrica é realizada em um momento específico do tempo que representa uma imagem estática da área temática naquele momento (ZUPIC; CATER, 2015).

A conversão para o formato *BibTex* e o descarregamento dos dados bibliográficos selecionados constituíram a terceira etapa. Os arquivos exportados foram lidos com a função *readFiles*. Esses arquivos foram convertidos em data frames contendo vários elementos bibliográficos usando a função *convert2d*. Esses elementos constituíram os metadados para as análises (ARIA; CUCCURULLO, 2017). Na quarta etapa foi realizada a junção das bases em uma única matriz contento todas as referências, e a limpeza dos dados. Para tanto, utilizamos a função *mergeDbSources* com *remove.duplicate=TRUE*. Em seguida, verificamos artigo por artigo na matriz para verificar possíveis erros. Foram retiradas 22 duplicatas não identificadas no R e corrigidos erros ortográficos no campo "autor", manualmente. A etapa cinco constituiu a análise bibliométrica propriamente dita.

Foram identificados 548 artigos na base da *WoS* entre 1996 a 2018 (fevereiro) e, 298 na base da *Scopus* entre 1973 a 2018 (fevereiro), totalizando 846 documentos exportados entre as duas bases. Após a junção das bases no R, foram retirados 276 documentos, sendo 239 artigos duplicados e 37 publicados como resumos de congresso e capítulos de livros. Foram analisados um total de 570 artigos no período de 1973 a 2018 oriundos de 172 fontes bibliográficas, com um total de 1194 palavras chave. A palavra-chave "*spectroscopy*" foi a mais listadas pelos autores, com 51 ocorrências.

2.3. Resultados e discussão

No período analisado, observou-se um crescimento no número de publicações, passando de 1 publicação no ano de 1973 para 71 publicações em 2017, com taxa de crescimento média anual de 10.4% (Figura 1).
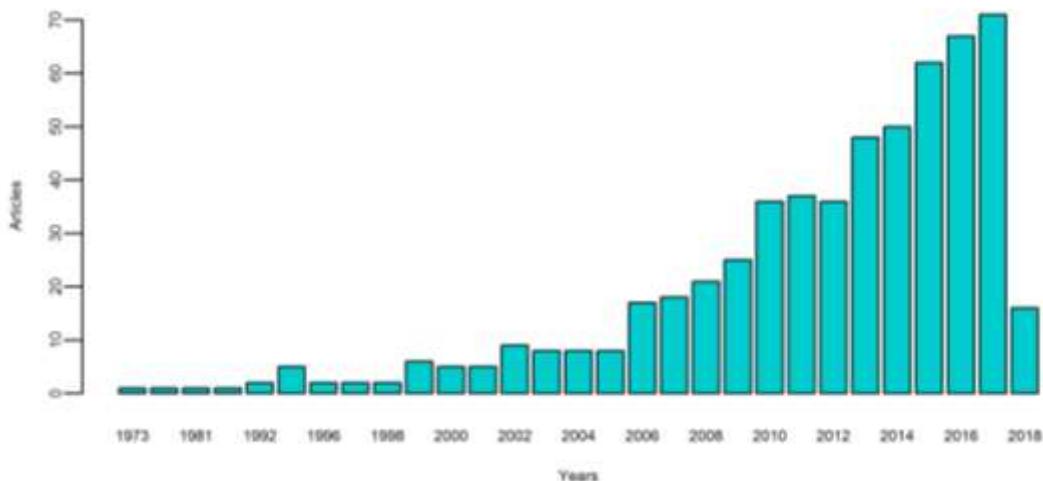
Figura 1. Número de publicações por ano.

2.3.1. Fontes bibliográficas

Esses artigos foram publicados em 172 fontes bibliográficas diferentes. O elevado número de fontes bibliográficas nacionais e internacionais ressalta a importância da análise bibliométrica para compilar e avaliar as pesquisas desenvolvidas nessa área. A Tabela 1 apresenta as fontes com maior frequência de publicações. A maioria dos artigos foram publicados em periódicos com alto fator de impacto na área da ciência do solo e afins, o que revela a importância e qualidade das pesquisas em ER nessas áreas.

Tabela 1. Número absoluto e percentual (%) de artigos publicados nas principais fontes bibliográficas.

Fonte bibliográfica	Número de artigos	Artigos (%)
Geoderma	98	17,2
Soil Science Society of America Journal	46	8,1
European Journal of Soil Science	23	4,0

Remote Sensing	14	2,5
Revista Brasileira de Ciência do Solo	13	2,3
Soil Science	13	2,3
Journal of Near Infrared Spectroscopy	12	2,1
Soil and Tillage Research	11	1,9
Applied Spectroscopy	10	1,7
Catena	10	1,7

Das fontes brasileiras, a Revista Brasileira de Ciência do Solo tem publicado o maior número de artigos nessa área, seguida pela Pesquisa Agropecuária Brasileira (5 artigos), Scientia Agrícola (2 artigos), Química Nova, Revista Ciência Agronômica e Semina-Ciências Agrárias com 1 artigo cada. A maioria dos estudos publicados tiveram como foco principal o uso da ER na região VIS-NIR-SWIR (350-2500nm) para identificação e predição dos teores de óxidos de ferro (hematita e goethita) e cor do solo (ALMEIDA; TORRENT; BARRÓN, 2003; CARMO et al., 2016; CORREA et al., 2008; DANTAS et al., 2014; FERNANDES et al., 2017, 2004; INDA et al., 2013; PELUCO et al., 2015; POPPIEL et al., 2018; RESENDE et al., 2014; RIZZO; DEMATTÊ; TERRA, 2014), predição do conteúdo de areia, silte e argila (BELLINASO; DEMATTÊ; ROMEIRO, 2010; CEZAR et al., 2012; DOTTO et al., 2014, 2016; FERRARESI et al., 2012), matéria orgânica (DOTTO et al., 2014; TEKIN; TUMSAVAS; MOUAZEN, 2014), carbono orgânico, P e N total (FELIX et al., 2016) e para análise qualitativa dos espectros para classificação do solo (BELLINASO; DEMATTÊ; ROMEIRO, 2010; POPPIEL et al., 2018).

2.3.2. Produção científica

Os dez países mais produtivos são apresentados na Tabela 2. Os EUA lideraram o ranking, seguidos da China, Brasil e Austrália. Os EUA e a China são os países com a maior produção de artigos científicos. A liderança dos EUA não é incomum, uma vez que esse país investe de forma continuada e permanente no desenvolvimento científico e tecnológico gerando pesquisa de alta qualidade e nível de impacto (TOLLEFSON, 2018) e, é o maior produtor de

artigos na maioria das áreas de pesquisa (ARIA; CUCCURULLO, 2017; CUCCURULLO; ARIA; SARTO, 2016). Entretanto o crescente investimento em ciência, tecnologia e inovação realizados pelo governo chinês nos últimos 10 anos tem impulsionado a produção de artigos científicos, que atingiu a marca de 426 mil artigos em 2016 (TOLLEFSON, 2018).

Tabela 2. Dez países mais produtivos na publicação de artigos entre 1973 a 2018 (com base na afiliação do primeiro autor).

País	Artigos	% Artigos
EUA	101	20.0
China	77	15.3
Brasil	47	9.4
Austrália	45	8.9
Alemanha	33	6.6
Itália	19	3.8
França	18	3.6
República Checa	15	3.0
Espanha	15	3.0
Inglaterra	14	2.8

2.3.3. Colaboração científica

Observou-se forte colaboração científica entre os países (Figura 2), com destaque para as redes de colaboração entre EUA e China, China e Austrália, Austrália com a Inglaterra, França, China e Espanha e Espanha com Brasil e Austrália. O Brasil aparece na terceira posição com 47 artigos (9.4%) e expressiva rede de colaboração científica com diversos países, com destaque para Espanha (Universidade de Córdoba, Universidade de *Rovira i Virgili*), França (*Internacional Development Research Centre* - IRD), EUA (várias instituições), Alemanha (Universidade de *Tübingen*) e Austrália (*CSIRO Land and Water*). O número de autores durante o período analisado foi de 1483 com média de 0.38 artigos por autor.
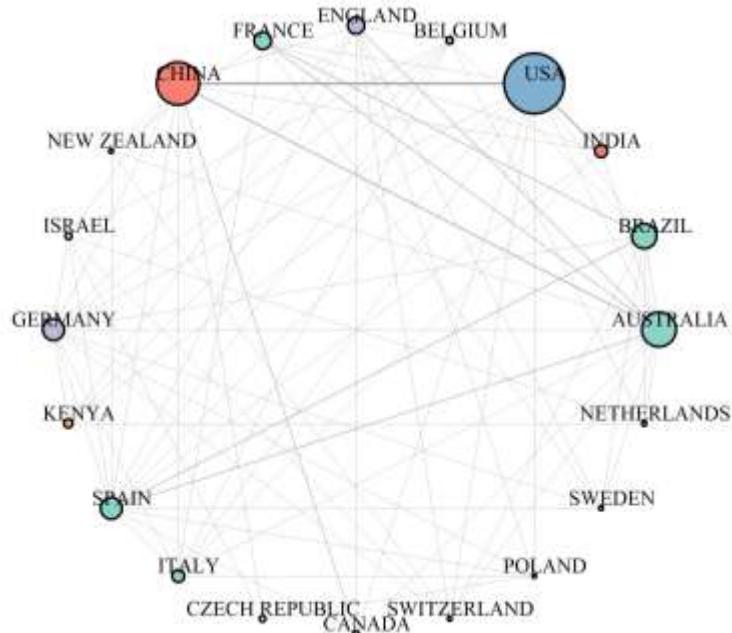
Figura 2. Rede de colaboração científica entre países

Na Figura 3, o número de citações refere-se ao número de vezes que cada um dos 20 autores mais citados incluídos na base de dados analisada (eixo y) foi citado por autores também incluídos na referida base. Entre os autores mais citados, destaca-se a presença de dois pesquisadores brasileiros: o Professor Dr. José Alexandre Mello Demattê da Universidade de São Paulo/Escola Superior de Agricultura "Luiz de Queiroz" e o pesquisador Dr. Gustavo Vasques da Embrapa Solos.
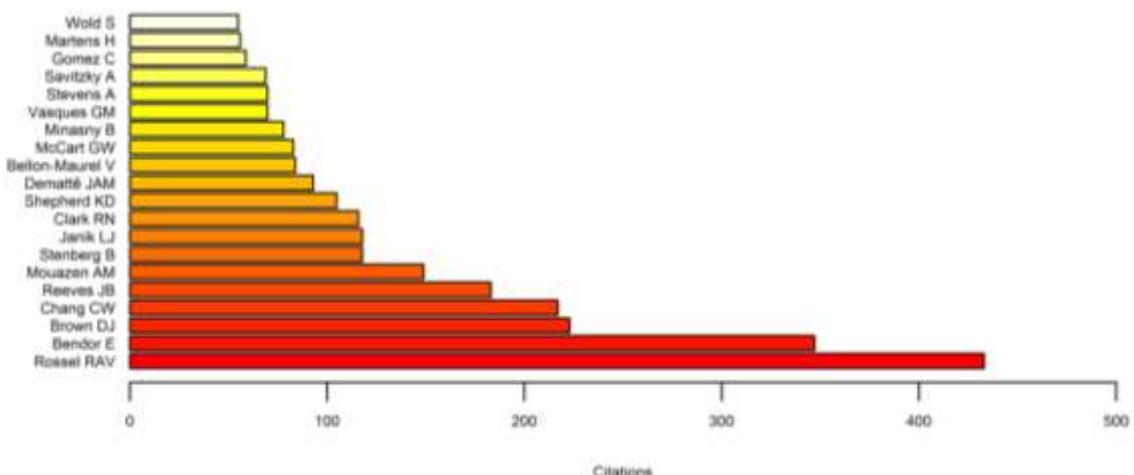
Figura 3. Vinte autores mais citados na base de dados selecionada.

A coautoria é uma das formas mais tangíveis e bem documentadas de colaboração científica (GLANZEL; SCHUBERT, 2005). Essa métrica foi usada para construir a rede de colaboração científica entre autores (ARIA; CUCCURULLO, 2017) apresentada na Figura 4. O trabalho científico colaborativo é um dos principais fatores que contribuíram para o aumento contínuo do número de artigos publicados (SOLLA PRICE, 1976). Tal tendência colaborativa foi observada nos artigos publicados nas bases analisadas. Dos 1483 autores, apenas 10 publicaram artigos de autoria única. Por outro lado, 1473 autores publicaram artigos de autoria múltipla, com média de 2,6 autores por artigo e índice de colaboração de 2,7.
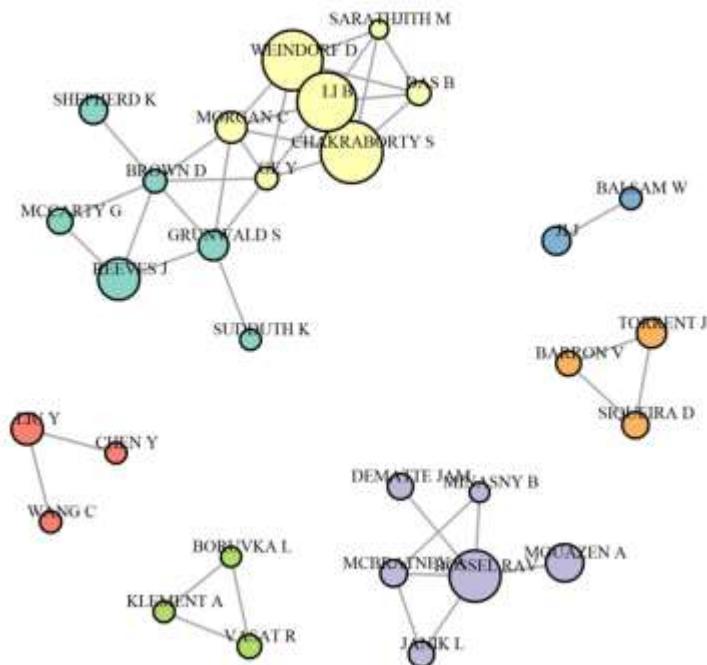
Figura 4. Rede de colaboração entre 30 autores. Os nós representam os autores e os links os coautores.

Observou-se claramente a formação de sete grupos colaborativos, com destaque para a rede de colaboração entre pesquisadores dos EUA, China e Índia. O maior agrupamento mostra que 13 dos 30 autores estão densamente interconectados, com autores colaborando em mais de um grupo. Nos agrupamentos menores, a colaboração ocorre, mas é limitada a poucos autores e/ou grupos de pesquisa. À medida que os autores publicam com diferentes pesquisadores, espera-se que a rede de colaboração de coautoria entre eles aumente (DE HAAN, 1997). A taxa de autor/publicação de 2,7 mostra a produtividade dos autores e a interligação com outros pesquisadores através da participação em coautoria. A taxa diminui à medida que a rede de colaboração e/ou a produtividade dos autores aumenta (DE HAAN, 1997).

Se considerarmos a rede de colaboração internacional, o número de artigos publicados com participação (autoria e coautoria) de pesquisadores brasileiros sobe de 47 (Tabela 1) para 61 artigos na base de dados. A maioria das publicações resultaram da colaboração com a Espanha (8 artigos), França (7 artigos), EUA (7 artigos) e Austrália (3 artigos). Um possível viés na análise

da autoria pode ter ocorrido, pois alguns autores podem ter usado nomes diferentes ao longo do tempo em seus artigos ou trocado sua afiliação (HO, 2007).

2.3.4. Contribuições relevantes: os dez artigos mais citados

A Tabela 3 apresenta os dez artigos mais citados na análise bibliométrica (item 2.2.1). O mais recente entre eles foi publicado no ano de 2010 e o mais antigo em 1995. Os dois artigos mais citados foram publicados em 2006 e 2002. Diferentes abordagens metodológicas para a predição dos atributos ou classificação interpretativa dos espectros foram empregadas nesses dez artigos e serão apresentadas e discutidas a seguir. Os aspectos teóricos e práticos do uso da ER podem ser encontrados com amplo detalhe nos trabalhos de Bendor; Banin (1995) e Viscarra Rossel et al. (2006).

Tabela 3. Os dez artigos mais citados.

Referência citada	Nº citações	
	Total	Freq.
Viscarra Rossel R.A. et al. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various properties. Geoderma, 131, 59-75. 2006.	591	20,2%
Shepherd K.D.; Walsh M.G. Development of reflectance spectral libraries for characterization of soil properties. Soil Sci. Soc. Am. J, 66, 988–998. 2002.	468	16,0%
Bendor E; Banin A. Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. Soil Sci. Soc. Am. J, 59, 364-372. 1995.	374	12,8%
Brown D.J. et al. Global soil characterization with VNIR diffuse reflectance spectroscopy. Geoderma, 132, 273-290. 2006.	348	11,9%
McCarty G.W. et al. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. Soil Sci. Soc. Am. J, 66, 640-646. 2002.	256	8,8%
Viscarra Rossel R.A.; Behrens T. Using data mining to model and interpret soil diffuse reflectance spectra. Geoderma, 158, 46-54. 2010.	253	8.6%
Scheinost A.C. et al. Use and limitations of second derivative diffuse reflectance spectroscopy in the visible to near infrared range to identify and quantify Fe oxide. Clays and clay minerals, 46, 528-536. 1998.	188	6.4%

Viscarra Rossel R.A. et al. In situ measurements of soil colour mineral composition and clay content by vis-NIR spectroscopy. Geoderma, 150, 253-266. 2009.	153	5.2%
Viscarra Rossel R.A.; McGlynn R.; McBratney A. Determining the composition of mineral organic mixes using UV-vis-NIR diffuse reflectance spectroscopy. Geoderma, 137, 70-82. 2006.	152	5.2%
Vasques, G.; Grunwald, S.; Sickman J. Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. Geoderma, 146, 14-25. 2008.	142	4.9%

2.3.4.1. Amostras de solos e bibliotecas espectrais

Dos dez artigos avaliados, cinco utilizaram amostras de solos históricas (legadas) coletadas para outras pesquisas (BROWN et al., 2006; McCARTY et al., 2002; SCHEINOST et al., 1998; SHEPHERD; WALSH, 2002; VISCARRA ROSSEL; BEHRENS, 2010). Nos estudos de Brown et al. (2006) e Shepherd; Walsh (2002) foram criadas bibliotecas espectrais dos solos em escala global e continental, respectivamente. Brown et al. (2006), utilizaram cerca de 46.000 amostras de solos legadas do *National Soil Survey Center-Soil Survey Laboratory* (EUA) coletadas em 50 estados americanos e em 36 países do continente Africano e Asiático. Shepherd e Walsh (2002) trabalharam com 1000 amostras de solos legadas coletadas no continente Africano. Scheinost et al. (1998) compilaram diversos estudos publicados e trabalharam com 176 amostras sintéticas e naturais com teores variáveis de hematita-goethita, lepidocrocita-goethita, óxidos de ferro, oxihidróxidos e oxihidroxisulfatos coletadas em vários países da União Europeia, África do Sul e Brasil. Viscarra Rossel; Behrens (2010), utilizaram 1104 amostras históricas de solos coletadas na Austrália por várias pessoas para atender as demandas de outras pesquisas e propósitos comerciais agronômicos. McCarty et al. (2002) trabalharam com 273 amostras de solos arquivadas de 14 regiões geograficamente distintas do Centro-Oeste dos Estados Unidos da América.

2.3.4.2. Feições espectrais

Todos os dez estudos utilizaram a técnica da ER na região do VIS-NIR-SWIR e MIR. As feições de absorção nessas regiões são causadas pelas interações entre a radiação eletromagnética incidente que é capaz de penetrar na superfície dos materiais (grupos funcionais dos solos) causando interações em nível microscópico com os elétrons e moléculas desses materiais. Essas interações são responsáveis pelas feições espectrais (depressões ou picos negativos) cujas intensidade e posição que ocupam na curva dependem das características de cada material (MENESES; MADEIRA NETO, 2001). A posição na banda espectral está tipicamente associada à composição química ou mineral do solo e a profundidade de absorção é proporcional à sua abundância (VAN DER MEER, 2018). As feições de absorção devido as transições eletrônicas ocorrem na região do UV ao VIS-NIR e estão associadas aos minerais que contêm Fe (ex., hematita, goethita, biotita e olivina) (BROWN et al., 2006) e aos metais de transição. A ER na região do visível foi usada para caracterizar a cor do solo a campo (VISCARRA ROSSEL et al., 2009), um importante atributo para a classificação do solo. A maioria dos minerais que apresentam silício, alumínio, potássio, cálcio ou radicais como o OH^-, não apresentam feições espectrais nessa região (MENESES; MADEIRA NETO, 2001).

Por outro lado, esses minerais apresentam feições de absorção muito mais intensas, definidas e pronunciadas a partir dos 1100 nm. Os picos em torno dos 1400 nm e 1900 nm são fortemente afetados pela água do solo. No nível de interação molecular, os processos vibracionais dos componentes do solo apresentam feições fortes e bem definidas na região do MIR (vibrações fundamentais), enquanto apenas os *overtones* e os *combination* tones (vibrações não fundamentais) são detectados na região do NIR-SWIR (VISCARRA ROSSEL et al., 2006). A matéria orgânica apresenta feições espectrais bem distintas na região do NIR-SWIR devido às vibrações não fundamentais relativas a grupos funcionais, tais como C-H, C-O, O-H e N-H, normalmente presentes nos componentes orgânicos (BENDOR; BANIN, 1995). As bandas de absorção na região do VIS-NIR-SWIR e MIR para a maioria dos

componentes do solo podem ser encontradas em Soriano-Disla et al. (2014) e o comportamento espectral de algumas classes de solos em Demattê et al. (2014) e Demattê et al. (2017).

2.3.4.3. Pré-processamento dos espectros de reflectância

A maioria dos dez artigos usaram técnicas de transformação dos espectros antes do processo de predição. Dentre elas, a absorbância (MCCARTY et al., 2002; VISCARRA ROSSEL et al., 2006; VISCARRA ROSSEL; MCGLYNN; McBRATNEY, 2006) e as derivadas (BENDOR; BANIN, 1995; SCHEINOST et al., 1998; SHEPHERD; WALSH, 2002; VASQUES; GRUNWALD; SICKMAN, 2008; VISCARRA ROSSEL; MCGLYNN; McBRATNEY, 2006) foram as mais usadas. Vasques; Grunwald; Sickman, (2008) compararam trinta técnicas de tratamento dos espectros de reflectância incluindo derivadas por *Savitzky-Golay* e *Norris-Williams*, transformações por *Kubelka-Munk*, absorbância, padronizações, normalizações e *baseline offset* para predição do carbono do solo em uma bacia hidrográfica na Flórida, EUA. Os autores concluíram que a escolha da técnica de transformação dos espectros depende do método multivariado usado. Uma revisão detalhada sobre os principais tratamentos dos espectros pode ser encontrada nos trabalhos de Buddenbaum; Steffens (2012) e Rinnan; van den Berg; Engelsen (2009).

2.3.4.4. Modelagem

Para modelar a relação entre os espectros de reflectância e os diversos atributos do solo, os estudos aplicaram diferentes métodos multivariados. Dos dez estudos analisados, sete usaram o método de regressão por mínimos quadrados parciais (*partial least square regression* - PLS) para predição dos atributos do solo a partir dos espectros. Entre os demais métodos usados estão: *multivariate adaptative regression splines* - MARS (FRIEDMAN, 1991), *stepwise multiple linear regression* -SMLR, *boosted trees* - BT

(FRIEDMAN, 2002), *support vector machines* - SVM (VAPNIK, 1995), *random forest* (BRIEMAN, 2001), *regression trees* e boosted regression trees (BREIMAN, 1994), *multiple linear regressions* - MLR, *committed trees* - CT (BREIMAN, 1994), *principal component regression* - PCR (MARTENS; NAES, 1992) e *artificial neural networks* - ANN (GALLANT, 1993). A PLS é o método estatístico multivariado mais comumente usados na predição de atributos do solo a partir dos espectros de reflectância. Viscarra Rossel et al. (2006) empregaram a PLS para predição de vários atributos dos solos. Esse método relaciona uma ou mais variáveis independentes (atributos do solo) a um conjunto de preditores (dados espectrais) para modelar uma relação linear provável entre os espectros e os atributos do solo, com boa capacidade de lidar com multicolinearidade dimensional elevada dos preditores (WOLD et al., 2001).

Os dez artigos apresentaram resultados promissores na predição de vários atributos (Tabela 4), tais como, matéria orgânica, área superficial específica, $CaCO_3$ (BENDOR; BANIN, 1995), carbono inorgânico carbono orgânico (BROWN et al., 2006; MCCARTY et al., 2002; SHEPHERD; WALSH, 2002; VASQUES; GRUNWALD; SICKMAN, 2008; VISCARRA ROSSEL et al., 2006; VISCARRA ROSSEL; BEHRENS, 2010) e inorgânico (BROWN et al., 2006; MCCARTY et al., 2002), conteúdo de areia, silte e argila com espectros medidos em laboratório e a campo (VISCARRA ROSSEL et al., 2009); algumas propriedades associadas a fertilidade do solo tais como bases trocáveis Ca e Mg, capacidade de troca de cátions, pH, Fe extraível (VISCARRA ROSSEL; BEHRENS, 2010, VISCARRA ROSSEL et al., 2006, SHEPHERD; WALSH, 2002, BENDOR; BANIN, 1995). Alguns estudos caracterizaram os minerais (e.g., goethita, hematita, esmectita, ilita, gibbsita, caulinita, quartzo, entre outros presentes na composição dos solos de forma quantitativa (BROWN et al., 2006; VISCARRA ROSSEL et al., 2009; VISCARRA ROSSEL; MCGLYNN; McBRATNEY, 2006).

Tabela 4. Descrição da abordagem por ER adotada nos 10 artigos mais relevantes.

Atributo	Compr. de onda	Prof.	n trein./n valid.	Pré-processamento	Método	Melhor R^2	Melhor RMSE	Área	Autores
pH_{Ca}	400-2500 nm e 2500-25.000 nm	0-20 cm	118	Absorbância	PLS	0.86	0.10	Austrália	Viscarra Rossel et al., (2006)
pH_w	400-2500 nm e 2500-25.000 nm	0-20 cm	118	Absorbância	PLS	0.75	0.13	Austrália	Viscarra Rossel et al., (2006)
pH_w	350-2500 nm	0-20 cm	758/378	Suavização por Savitzky-Golay	MARS	0.70	0.43	África	Shepherd; Walsh (2002)
pH_w	350-2500 nm	0-10 cm; 10-20 cm; 30-60 cm; 70-80 cm	1104	Reflectância	MLR; PLS; MARS; SVM; RF; BT; ANN;	0.81	0.53	Austrália	Viscarra Rossel; Behrens (2010)
Necessidade de calagem; Mg ha⁻¹	400-2500 nm e 2500-25.000 nm	0-20 cm	118	Absorbância	PLS	0.73	0.61	Austrália	Viscarra Rossel et al., (2006)
Carbono orgânico; dag kg⁻¹	400-2500 nm e 2500-25.000 nm	0-20 cm	118	Absorbância	PLS	0.76	0.15	Austrália	Viscarra Rossel et al., (2006)
Carbono orgânico; g kg⁻¹	350-2500 nm	0-20 cm	674/337	Suavização por Savitzky-Golay	MARS	0.80	3.10	África	Shepherd; Walsh (2002)
Carbono orgânico; g kg⁻¹	350-2500 nm		3793	Primeira derivada	BRT; PLS	0.87	7.9	EUA, África, Asia, America e Europa	Brown et al., (2006)
Carbono orgânico; g kg⁻¹	1000-2500 nm e 2500-25.000 nm	0-5 cm; 5-10 cm; 10-25 cm	177/61	Absorbância	PLS	0.94	3.20	EUA	McCarty et al., (2002)

Carbono orgânico; %	350-2500 nm	0-10 cm; 10-20 cm; 30-60 cm; 70-80 cm	1104	Reflectância	MLR; PLS; MARS; SVM; RF; BT; ANN;	0.89	0.75	Austrália	Viscarra Rossel; Behrens (2010)
Carbono orgânico; log mg kg^{-1}	350-2500 nm	0-30 cm; 30-60 cm; 60-120 cm; 120-180 cm	400/154	30 tipos diferentes	SMLR; PCR; PLS; RT; CT	0.82	0.193	EUA	Vasques, Grunwald, Sickman (2008)
Matéria orgânica; %	1000-2500 nm	0-5 cm	22/47	Reflectância	LR	0.55	nd	Israel	Bendor; Bannin (1995)

Continuação...

Carbono inorgânico; g kg^{-1}	350-2500 nm		4184	Primeira derivada	BRT; PLS	0.86	5.60	EUA, África, Ásia, Americas e Europa	Brown et al., (2006)
Carbono inorgânico; g kg^{-1}	1000-2500 nm e 2500-25.000 nm	0-5 cm; 5-10 cm; 10-25 cm	177/62	Absorbância	PLS	0.98	3.40	EUA	McCarty et al., (2002)
Carbono Total; g kg^{-1}	1000-2500 nm e 2500-25.000 nm		177/60	Absorbância	PLS	0.95	3.40	EUA	McCarty et al., (2002)
Argila; dag kg^{-1}	400-2500 nm e 2500-25.000 nm	0-20 cm	116	Absorbância	PLS	0.67	1.73	Austrália	Viscarra Rossel et al., (2006)
Argila; g kg^{-1}	350-2500 nm	0-20 cm	457/225	Suavização por Savitzky-Golay	MARS	0.78	75	África	Shepherd; Walsh (2002)
Argila; %	1000-2500 nm	0-5 cm	35/56	Derivada da absorbância	LR	0.56	nd	Israel	Bendor; Bannin (1995)
Argila; g kg^{-1}	350-2500 nm		4184	Primeira derivada	BRT; PLS	0.91	54	EUA, África, Ásia,	Brown et al., (2006)

								America e Europa	
Argila; %	350-2500 nm	0-10 cm; 10-20 cm; 30-60 cm; 70-80 cm	1104	Reflectância	MLR; PLS; MARS; SVM; RF; BT; ANN;	0.88	6.42	Austrália	Viscarra Rossel; Behrens (2010)
Argila; %	350-2500 nm	até 1.2 m	892/469	Suavização e primeira derivada por Savitzky-Golay	PLS	0.85	7.50	Austrália	Viscarra Rossel et al., (2009)
Silte; dag kg^{-1}	400-2500 nm e 2500-25.000 nm	0-20 cm	116	Absorbância	PLS	0.52	2.11	Austrália	Viscarra Rossel et al., (2006)

Continuação...

Silte; g kg^{-1}	350-2500 nm	0-20 cm	457/225	Suavização por Savitzky-Golay	MARS	0.67	49	África	Shepherd; Walsh (2002)
Areia; dag kg^{-1}	400-2500 nm e 2500-25.000 nm	0-20 cm	116	Absorbância	PLS	0.75	2.57	Austrália	Viscarra Rossel et al., (2006)
Areia; g kg^{-1}	350-2500 nm	0-20 cm	457/225	Suavização por Savitzky-Golay	MARS	0.76	108	África	Shepherd; Walsh (2002)
Capacidade de troca de cátions (CTC); mmol(+) kg^{-1}	400-2500 nm e 2500-25.000 nm	0-20 cm	49	Absorbância	PLS	0.09	10.77	Austrália	Viscarra Rossel et al., (2006)
Capacidade de troca de cátions (CTC); cmolc kg^{-1}	350-2500 nm	0-20 cm	740/369	Suavização por Savitzky-Golay	MARS	0.88	3.80	África	Shepherd; Walsh (2002)
Capacidade de troca de cátions (CTC); cmolc kg^{-1}	1000-2500 nm	0-5 cm	35/56	Derivada da absorbância	LR	0.64	nd	Israel	Bendor; Bannin (1995)

Capacidade de troca de cátions (CTC); cmolc kg^{-1}	350-2500 nm		4183	Primeira derivada	BRT; PLS	0.83	5.50	EUA, África, Ásia, America e Europa	Brown et al., (2006)
Cálcio trocável (Ca); mmol(+) kg^{-1}	400-2500 nm e 2500-25.000 nm	0-20 cm	49	Absorbância	PLS	0.37	6.60	Austrália	Viscarra Rossel et al., (2006)
Cálcio trocável (Ca); cmolc kg^{-1}	350-2500 nm	0-20 cm	740/369	Suavização por Savitzky-Golay	MARS	0.88	2.80	África	Shepherd; Walsh (2002)
Alumínio trocável (Al); mmol(+) kg^{-1}	400-2500 nm e 2500-25.000 nm	0-20 cm	49	Absorbância	PLS	0.37	1.06	Austrália	Viscarra Rossel et al., (2006)
Nitrato-nitrogênio (NO$_3$-N); mg kg^{-1}	400-2500 nm e 2500-25.000 nm	0-20 cm	49	Absorbância	PLS	-0.02	3.68	Austrália	Viscarra Rossel et al., (2006)
Fósforo disponível (P); mg kg^{-1}	400-2500 nm e 2500-25.000 nm	0-20 cm	49	Absorbância	PLS	0.07	5.21	Austrália	Viscarra Rossel et al., (2006)
Potássio trocável (K); mmol(+) kg^{-1}	400-2500 nm e 2500-25.000 nm	0-20 cm	49	Absorbância	PLS	0.46	1.83	Austrália	Viscarra Rossel et al., (2006)

Continuação...

Condutividade elétrica (CE); dS/m	400-2500 nm e 2500-25.000 nm	0-20 cm	49	Absorbância	PLS	0.29	0.01	Austrália	Viscarra Rossel et al., (2006)
Magnésio trocável (Mg); cmolc kg^{-1}	350-2500 nm	0-20 cm	739/369	Suavização por Savitzky-Golay	MARS	0.81	1.10	África	Shepherd; Walsh (2002)
Área superficial específica; m^2 g^{-1}	1000-2500 nm	0-5 cm	35/56	Derivada da absorbância	LR	0.70	nd	Israel	Bendor; Bannin (1995)
Umidade higroscopica; %	1000-2500 nm	0-5 cm	35/56	Derivada da reflectância	LR	0.62	nd	Israel	Bendor; Bannin (1995)

Carbonato (CaCO$_3$); %	1000-2500 nm	0-5 cm	39/52	Derivada da absorbância	LR	0.69	nd	Israel	Bendor; Bannin (1995)
Fed; g kg^{-1}	350-2500 nm	nd	2909	Primeira derivada	BRT; PLS	0.77	8.90	EUA, África, Ásia, America e Europa	Brown et al., (2006)
Caolinita	280-2500 nm	nd	nd	Diversos	PLS	0.94	0.036	Austrália	Viscarra Rossel, McGlynn, McBratney (2006)
Ilita	280-2500 nm	nd	nd	Diversos	PLS	0.96	0.034	Austrália	Viscarra Rossel, McGlynn, McBratney (2006)
Esmectita	280-2500 nm	nd	nd	Diversos	PLS	0.92	0.034	Austrália	Viscarra Rossel, McGlynn, McBratney (2006)
Goethita	280-2500 nm	nd	nd	Diversos	PLS	0.55	0.036	Austrália	Viscarra Rossel, McGlynn, McBratney (2006)
H-F	280-2500 nm	nd	nd	Diversos	PLS	0.60	0.028	Austrália	Viscarra Rossel, McGlynn, McBratney (2006)
Quartzo	280-2500 nm	nd	nd	Diversos	PLS	0.16	0.142	Austrália	Viscarra Rossel, McGlynn, McBratney (2006)

nd = não descrito.

Optou-se por mostrar os valores do coeficiente de determinação (R^2) dos conjuntos de validação dos dez artigos, como medida da performance das predições, pois nem todos os trabalhos avaliados apresentavam os valores da raiz quadrada do erro médio (RMSE) e/ou do desvio padrão do erro (RPD). Observou-se que a performance dos modelos variou bastante nos 10 estudos analisados. Para Soriano-Disla et al. (2014), essa variação ocorre devido às características espectrais intrínsecas de cada atributo avaliado, ao método multivariado usado, ao tamanho do conjunto de calibração e a especificidade da área de estudo. Portanto, os valores de performance da abordagem ER devem ser avaliados considerando que: (i) a utilidade de uma predição particular depende da natureza do atributo analisado e do uso potencial do valor predito (por exemplo, teste rotineiro do solo, mapeamento e/ou entrada nas funções de pedotransferência), bem como sobre o erro e robustez da predição e relevância das amostras utilizadas para o desenvolvimento do modelo em relação às amostras desconhecidas (SORIANO-DISLA et al., 2014) e (ii) o sucesso ou não dessa abordagem está associado aos valores de referência medidos em laboratório pelos métodos convencionais (KUANG et al., 2012) e a variabilidade dos resultados entre os laboratórios (CANTARELLA et al., 2006). Para Soriano-Disla et al. (2014), o objetivo de cada estudo é que deve determinar o grau de precisão adequado das predições.

2.4. Conclusões

O estudo da produção científica sobre a ER na ciência do solo foi analisado durante o período de 1973 a 2018. Observou-se um aumento acentuado no número de artigos publicados ao longo do período observado. Os três principais países de origem desses artigos foram EUA, China e Brasil. Há uma comunidade científica ativa pesquisando a relação entre os espectros de reflectância e a composição física, química e mineralógica do solo, usando sensores em laboratório e a campo. A colaboração científica por meio da coautoria foi a estratégia preferida para publicação, configurada em pequenos e grandes grupos. Apenas 3% dos artigos foram publicados com autoria única. A medida que aumentam as pesquisas em ER com artigos publicados em um número diversificado de periódicos, os métodos bibliométricos serão cada vez

mais úteis para mostrar os avanços, revelar os gargalos e indicar os passos futuros da ER na ciência do solo e áreas afins. Os dez artigos mais citados revelaram que os principais países envolvidos foram os EUA e a Austrália. Os principais resultados alcançados pelos autores nesses artigos mostram que (i) a ER é uma técnica que permite a estimativa de vários atributos do solo simultaneamente. Foi considerada uma técnica viável para a análise de um grande número de amostras, mostrando-se promissora para a aplicação prática na agricultura e no monitoramento ambiental como uma técnica que pode superar as limitações dos métodos convencionais de laboratório; (ii) o conhecimento da incerteza da predição torna o processo de tomada de decisão melhor informado sobre o *trade-off* do o custo da medição e o risco associado ao uso da predição; (iii) a capacidade preditiva do modelo usado e do nível de acurácia esperado dependem do número e da diversidade das amostras de solo utilizadas ; (iv) PLS é o método que apresentou melhor performance para vários atributos do solo e quando comparado a modelos não lineares apresenta predições competitivas; (v) predições na região do MIR são geralmente melhores que na região do VIS-NIR-SWIR; (vi) A ER na região do VIS-NIR-SWIR e MIR associada a métodos estatísticos robustos e os constantes avanços tecnológicos (global position systems - GPS, o desenvolvimento contínuo dos computadores e de plataformas computacionais livres, sensores hiperespectrais remotos e sensores portáteis de campo ou laboratório, etc.) tem modernizado a pedologia e os métodos de análise do solo.

3. CAPÍTULO II - REGIONAL LEGACY SOIL DATASET FOR ESTIMATION OF SAND AND CLAY CONTENT FROM VIS-NIR-SWIR REFLECTANCE SPECTROSCOPY, IN SOUTHERN BRAZIL

3.1. Introduction

Soil texture is one of the most important soil attributes as it controls many physical, chemical, biological and hydrological features and processes in the soil. It is the expression of soil particle size distribution, that is, the relative proportion of sand, silt and clay content. Soil texture has been used in pedotransfer functions to estimate bulk density and in estimating more complex, time-consuming soil attributes, such as hydrological characteristics (KARUP et al., 2016; KELLER; HÅKANSSON, 2010; MARTÍN; REYES; TAGUAS, 2017; POEPLAU; ERIKSSON; KÄTTERER, 2015; SCHJØNNING et al., 2017) and soil organic carbon (BLANCO-MOURE et al., 2016; JOHANNES et al., 2017), as well as in climatic, environmental and agricultural risk assessment models (JENSEN et al., 2015; LEVINE et al., 2016; PENNINGTON et al., 2017; VALVERDE-ARIAS et al., 2018b). Knowledge of the proportions of different-sized soil particles is critical for understanding soil behavior and management (BRADY; WEIL, 2007).

Reflectance spectroscopy in the visible, near and shortwave infrared (VIS-NIR-SWIR) regions has been proposed as a rapid, accurate and cost-effective method to predict chemical, physical and mineralogical attributes using

laboratory, field and airborne hyperspectral sensors (DEMATTÊ et al., 2016a; DOTTO et al., 2018; NOURI et al., 2017; POGGIO; BROWN; BRICKLEMYER, 2017; VASQUES; GRUNWALD; SICKMAN, 2008; VISCARRA ROSSEL et al., 2017). Soil VIS-NIR-SWIR spectra are non-specific and include weak, wide and overlapping absorption bands directly linked to soil composition, whereby moisture, particle size, organic matter and mineralogy of the clay fraction and iron oxides influence spectral behavior (STENBERG et al., 2010).

Different preprocessing techniques have been applied to transform soil spectra, removing noise from the multiple scattering effect, highlighting specific features of the spectra, eliminating redundant information and preparing the soil spectra for spectral modelling (RINNAN; VAN DEN BERG; ENGELSEN, 2009). As reported by Buddenbaum and Steffens (2012), these techniques represent an important step in the multivariate approach and include several algorithms such as, smoothing, normalization, scatter-correction, continuum removal and derivatives. Some studies have reported improvements in the performance of prediction models (DOTTO et al., 2017; NAWAR et al., 2016; VASQUES; GRUNWALD; SICKMAN, 2008), while others found similar or better results with no spectral preprocessing (SAWUT et al., 2014, VISCARRA ROSSEL; WEBSTER, 2012). The type and amount of required preprocessing techniques are site-specific (STENBERG et al., 2010) and with large datasets the effects of preprocessing steps are not clear (ENGEL et al., 2013).

Several multivariate methods based on VIS-NIR-SWIR have been applied to processing soil spectra in order to mathematically extract meaningful information from individual spectrum to accurately predict chemical and physical soil attributes, such as organic carbon/matter, pH, total nitrogen, soil moisture and cationic exchange capacity, among others (ARAÚJO et al., 2015; CONFORTI; MATTEUCCI; BUTTAFUOCO, 2017; DEMATTÊ et al., 2017; DOTTO et al., 2017, 2018; GHOLIZADEH et al., 2016; MORELLOS et al., 2016; NAWAR; MOUAZEN, 2017a; VENDRAME et al., 2012; VISCARRA ROSSEL; BEHRENS, 2010; XU et al., 2018). The capacity to predict sand, silt and clay has also been demonstrated in previous studies (DEMATTÊ et al., 2016b; DOTTO et al., 2017; LACERDA et al., 2016; NAWAR et al., 2016; SANTANA; SOUZA; POPPI, 2018; VENDRAME et al., 2012). The literature has shown the potential of VIS-NIR-SWIR spectroscopy and multivariate methods for

prediction of sand and clay content. Among the multivariate methods, the partial least square regression (PLS) is the most common multivariate method used (DOTTO et al., 2018), given its simplicity and robustness (LACERDA et al., 2016; VASQUES; GRUNWALD; SICKMAN, 2008; VISCARRA ROSSEL et al., 2006). However, other studies have established that nonlinear data-mining methods such as support vector machines (SVM), gaussian process regression (GPR) and random forest (RF) can outperform PLS when used to build predictive models from reflectance spectra (DOTTO et al., 2017; NAWAR et al., 2016; SANTANA; SOUZA; POPPI, 2018; TERRA; DEMATTÊ; VISCARRA ROSSEL, 2015). In addition to these methods, another data-mining tool based on Cubist regression-rules has been introduced into the spectroscopy approach to predicting soil attributes (MINASNY; McBRATNEY, 2008; MORELLOS et al., 2016; SORENSON; QUIDEAU; RIVARD, 2018; VISCARRA ROSSEL et al., 2016; VISCARRA ROSSEL; WEBSTER, 2012; ZENG et al., 2017). Minasny; McBratney (2008), Morellos et al. (2016) and Sorenson; Quideau; Rivard (2018) used Cubist to build predictive models of soil attributes, including clay content, total carbon, total nitrogen, moisture content and cation exchange capacity, and reported that Cubist provided better results than those provided by PLS.

The success of soil prediction by VIS-NIR-SWIR has led to considerable investment in large soil spectral libraries (BROWN et al., 2006; SHEPHERD; WALSH, 2002; VISCARRA ROSSEL; WEBSTER, 2012). Soil information stored by universities, research centers and government agencies, among others, could provide an opportunity to enlarge spectral libraries (NOCITA et al., 2015; VISCARRA ROSSEL et al., 2016). While several studies have been done to build predictive models for soil attributes based on local, regional and national spectral libraries in Brazil (ARAÚJO et al., 2014; BELLINASO; DEMATTÊ; ROMEIRO, 2010; DEMATTÊ et al., 2016b; LACERDA et al., 2016; RAMIREZ-LOPES et al., 2013), the application of soil reflectance spectroscopy has not been reported on a regional scale in South Brazil, especially in the state of Santa Catarina (SC). In this state, existing studies using soil spectral libraries are limited to small areas (DOTTO et al., 2017, 2018). This research aims to fill this gap and to further the use of reflectance spectroscopy for assessing sand and clay content in subtropical soils based on a regional spectral library.

The main objectives of this study were: (1) to develop a soil VIS-NIR-SWIR spectroscopy approach using legacy soil samples to improve spectral soil information in a regional scale; (2) to compare six spectral preprocessing techniques in the development of sand and clay content models; and (3) to compare the performance of linear and non-linear multivariate methods for prediction of sand and clay content. Given the high variability of legacy soil samples, we expected that non-linear multivariate methods would perform better than linear methods to predict sand and clay content.

3.2. Material and methods

3.2.1. Study area

The study area is the entire state of SC, which is located in Southern Brazil between longitudes 25°57'41" and 29°23'55"S and latitudes 48°19'37" and 53°50'00"W and has an area of approximately 95,738 km^2 (Figure 1). The state is characterized by its diversity in climate, vegetation, geology, relief and soil. SC has two climate types according to the Köeppen classification (Figure 1); super humid and mesothermal (Cfa) and quite humid and mesothermal (Cfb). Cfa with mean annual temperature of 18.5 °C and annual precipitation of 1,800 mm occurs in the Atlantic coastal area and in the Western region, while Cfb, with an average maximum temperature ranging from 20 to 26 °C and minimum temperatures from 9 to 13 °C, with annual precipitation of 1,600 mm, occurs in the Plateau region (PANDOLFO et al., 2002). The remaining original vegetation includes areas of Rain Forest type and four major subtypes, these being Dense Rain Forest, Araucaria Forest, Alpine Grassland, Deciduous Forest, and Coastal Vegetation (KLEIN, 1978). The geology consists of granitoids, charnockitics, gnaisses and granites in Eastern SC, the Gondwana Plateau and a basalt plateau in Western SC, with a predominance of basic volcanic rocks and quartz sandstones, with siliceous and argillaceous intercalations (SANTA CATARINA, 1986). The relief ranges from below sea level to about 1800 m a.s.l., with approximately 70% of the state with at least 300 m above sea level (SANTA CATARINA, 1986). Soils are diverse and represent several groups, according to the IUSS Working Group WRB (2014) classification, including Acrisols, Lixisols, Alisols, Ambisols, Chernozems,

Podzols, Gleysols, Solonchaks, Ferralsols, Fluvisols, Leptosolss, Arenosols, Regosols, Nitosols and Histosols (EMBRAPA, 2004).
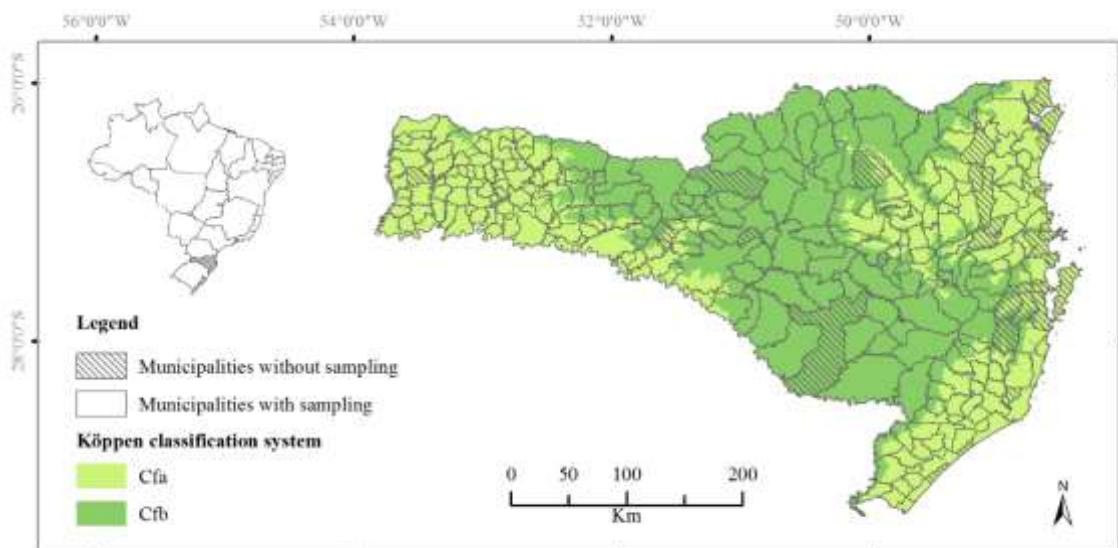
Figure 5. Sampling of soil data from municipalities of Santa Catarina state.

3.2.2. Soil spectral library

The soil information provided by the Agricultural Research and Rural Extension Corporation of Santa Catarina ("EPAGRI - Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina") was used to develop the soil spectral library. A total of 1,534 samples were collected from agricultural areas in 2009 on a regional scale, covering 260 municipalities (about 90%) of SC (Municipalities with sampling, Figure 5). Each sample was a composite sample of five points per area from the topsoil layer (0-50 cm). Samples were air-dried, ground and sieved to 2 mm. The sand (0.05-2 mm) and clay (<0.002 mm) content were determined according to the pipette standard method, using NaOH dispersant (TEIXEIRA et al., 2017).

3.2.3. Spectral measurements

The spectral reflectance of soil samples was obtained using a FieldSpec 3 spectroradiometer (Analytical Spectral Devices, Boulder, USA) in the Vis-NIR-SWIR range (350-2500 nm), following standard laboratory procedure of the Brazilian Soil Spectral Library (ROMERO et al., 2018). The

samples were placed in a petri dish and shaken to ensure a smooth surface for spectrum acquisition. The light source was two halogen (50W) bulbs with the beam non-collimated to the target plane, positioned at a distance of 35 cm from the sample with a Zenithal angle of 30⁰. The spectral sensor captured the light through a fiber-optic cable connected to the sensor, placed vertically within 8 cm of the sample, where the reflected light in an area of approximately 2 cm^2 at the center of the sample was measured. As a reference standard, a white Spectralon® was used at the beginning of the measurements and after every 20 readings. Each spectrum measurement was the result of the average of 50 sensor readings. A total of three scans were collected from each sample, rotating the petri dish by 90⁰ for each scan, and these were then averaged in order to obtain a representative spectrum.

3.2.4. Spectral preprocessing

The VIS-NIR-SWIR spectra are affected by particle size and light scatter, so spectral preprocessing using mathematical functions is applied to reduce the spectral noise and enhance the feature being sought (RINNAN; VAN DEN BERG; ENGELSEN, 2009), which can often improve performance of the models. In this study six spectral preprocessing techniques were applied and compared in the development of sand and clay prediction models. The techniques were divided into groups of scatter-correction and spectral-derivatives. The first group included: (i) multiplicative scatter correction (MSC), which removes additive and/or multiplicative signal effects (MARTENS; NAES, 1992); (ii) detrending (DET), to reduce the effect of particle size and additionally remove the linear, or curvilinear, trend of each spectrum (BARNES; DHANOA; LISTER, 1989); and (iii) normalizations by range (NBR), to get all data to approximately the same scale; (iv) continuum removed reflectance (CRR), which removes the continuous features of the spectra and isolates specific absorption features present in the spectrum to minimize noise (CLARK; ROUSH, 1984). The second group is spectral derivatives, represented by (v) Savitzky–Golay first derivative using a first order polynomial with a search window of 11 nm (FDSG) and (vi) Savitzky–Golay second derivative using a second order polynomial with a search window of 11 nm (SDSG). Derivatives

are often used to remove baseline shifts and enhance possible spectral features correlated with soil attributes (SAVITZKY; GOLAY, 1964).

The preprocessing techniques were selected because they have been shown to improve the performance of spectroscopy models and have different effects on model prediction. They were applied to soil reflectance curves in the range of 350–2500 nm. Reflectance spectra (RAW) with no spectral preprocessing applied was used as "control treatment". All preprocessing techniques were carried out using R (R CORE TEAM, 2017) by applying *prospectr* (RAMIREZ-LOPES et al., 2013), *pls* (MEVIK; WEHRENS; LILAND, 2016) and *clusterSim* (DUDEK, 2017) packages. The spectral preprocessing techniques presented here are discussed in more detail in Rinnan; van den Berg; Engelsen (2009) and Buddenbaum; Steffens (2012).

The Scott–Knott (SK) test (SCOTT; KNOTT, 1974) was used to compare the average values of R^2 and RMSE between different methods and preprocessing techniques to verify significant differences between them. The SK performs a hierarchical cluster analysis approach used to partition treatment into distinct homogeneous groups by minimizing variation within groups and maximizing variation between groups (SCOTT; KNOTT, 1974). The cluster procedure begins with the whole group of observed mean effects and then divides and keeps dividing subgroups in such a way that the intersection of any of the two formed groups remains empty (JELIHOVSCHI; FARIA; ALLAMAN, 2014). It was applied using the *ScottKnott* package (JELIHOVSCHI; FARIA; ALLAMAN, 2014).

3.2.5. Multivariate methods

In order to compare the performance of the proposed multivariate regression models using a regional spectral library, we compared to two supervised learning algorithms with linear kernel function, Support Vector Machines (SVM) and Gaussian Process Regression (GPR), two tree-based models, Cubist and Random Forest (RF), and one of the most common linear methods used in the spectroscopy approach, Partial Least Square Regression (PLS). The regression process was implemented based on the measured reflectance spectra (RAW and six spectral preprocessing techniques) and the

measured values of sand and clay content using the training set. The predictive models were assessed for each soil attribute using the independent assessment set. The modelling was performed using several packages in R (R CORE TEAM, 2017) and the parameters of each model were manually optimized to generate the best possible fit between the variables and outputs.

The regression-rules model (HOLMES; HALL; PRANK, 1999), known as Cubist (MINASNY; McBRATNEY, 2008), was used as tree-based model. Cubist is based on a decision tree that builds multivariate linear regression at the terminal leaves of a tree (HOLMES; HALL; PRANK, 1999), instead of using discrete values at each leaf of a decision tree. First, the response data is partitioned into subsets within which their characteristics are similar with respect to their spectra. A series of rules derived using *if-then-else* defines the partitions and a hierarchy is arranged. A condition may be simply based on one wavelength or, more often, it consists of several wavelengths, and then the next step is the prediction of the soil attribute by ordinary least square regression from the wavelengths in the partition (VISCARRA ROSSEL et al., 2016; VISCARRA ROSSEL; WEBSTER, 2012). It was implemented using *Cubist* package (KUHN, 2015) and some tuning parameters were adjusted, such as the ensemble of rule-based models called "committees", to improve predictability and stability of the models, and the number of neighboring observations, or "neighbors", to build composite models.

RF is an ensemble learning method based on decision of classification or regressions trees (BREIMAN, 2001). The response data for each tree is obtained through bootstrap sampling (with replacement) of the original observations of the training set (CUTLER; CUTLER; STEVENS, 2012). RF is considered resistant to overfitting, with relative robustness in respect to noise and irrelevant features, and is relatively insensitive to value range (BREIMAN, 2001). It was implemented using the *randomForest* package with the setting of basic tuning parameters as follows: mtry=100, the number of input variables that are randomly selected for each bootstrap; ntree=500, the number of trees to be constructed; and nodesize=5, the minimum size of nodes in which no split will be attempted.

SVM is a machine technique based on statistical learning theory that uses a kernel trick to project data onto a new hyperplane where high-

dimensional nonlinear patterns can be easily represented (VAPNIK, 1995). The epsilon SVR with basic "linear" kernel function or dot product was used in order to keep the models as simple as possible. SVM was performed using the *e1071* package (MEYER et al., 2017).

GPR is a probabilistic, non-parametric Bayesian approach that uses linear "*kernel tricks*" to insert the original feature space into a high dimensional space with linear or nearly linear structure (RASMUSSEN; WILLIAMS, 2006; WILLIAMS; BARBER, 1998). In this study, a regression type with the linear "*vanilladot*" kernel function was implemented using the *kernlab* package (KARATZOGLOU et al., 2004). Linear function was chosen for SVM and GPR to better compare the different performances with the reference method (PLS) and because use of the simplest methods may be preferred.

PLSR is a widely used method applied to spectral calibration and prediction of soil attributes (VISCARRA ROSSEL et al., 2006). In PLSR, the orthogonal basis of latent factors is constructed one by one and oriented in the direction of maximum covariance between predictor variables (X-variables - spectra) and response variables (Y-variables separate sand and clay content) (WOLD; SJÖSTRÖM; ERIKSSON, 2001). In the regression model, this relationship between variables X and Y is simplified to the smallest number of possible latent factors. RMSE was used in this study to identify the number of latent factors and *leave-one-out cross-validation* was used to verify prediction performance for the model training set (BOOS, 2003) and to prevent over or under-fitting the data. The PLSR was performed using the *pls* package (MEVIK; WEHRENS; LILAND, 2016).

3.2.6. Model training and assessment

From the total dataset (n = 1,534), seventy-five percent were separated at random into the training set (n = 1,151), to create the regression models, while the remaining 25% (n = 383, assessment set) were used to independently validate the models. In order to check the reliability of splitting of each subset, the Levene test (LEVENE, 1960) and Student's t-test were applied to verify the equality of variances and means, respectively.

3.2.7. Model accuracy

The coefficient of determination (R²), root mean square error (RMSE) and the ratio of performance to the interquartile range (RPIQ) were used to assess the performance of sand and clay content prediction models, using Equations (1), (2) and (3), respectively. R² provides the proportion of the variance explained by the model. RMSE provides the accuracy of predictions, giving the standard deviation of the model prediction error in the same units as the attributes. RPIQ was used instead of ratio of performance deviation (RPD), because it is based on quartile range and better represents the spread of the population for skewed distributions (BELLON-MAUREL et al., 2010). The highest R² and RPIQ values and the lowest RMSE value of the assessment set were used to determine selection of the best spectral preprocessing and multivariate regression models.

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

where $\hat{y}_i$ is the value predicted by the model; y_i is the measured value; $\bar{y}_i$ is the average value; and N is the number of samples.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \qquad \text{Equation (2)}$$

where N is the number of samples used in the prediction; and $\hat{y}_i$ and y_i are the values of predicted and measured soil properties, respectively.

$$RPIQ = \frac{IQ}{RMSE} \qquad \text{Equation (3)}$$

where IQ is the interquartile distance (IQ = Q3-Q1) of the observed values, which accounts for 50% of the population around the median.

3.3. Results and discussion

3.3.1. Descriptive analysis of soil data

The results of the descriptive statistical analyses for dataset, training and assessment sets are shown in Table 1. The samples under study presented a widespread variation with sand content ranging between 1.0 and

99.0%, with a mean value of 28.8%, and clay content ranging from 0.0 to 77.0%, with a mean value of 38.2%, indicating great variability in terms of particle size distribution (Table 5). Sand content showed the highest coefficient of variation (CV) and skewness, indicating high dispersion of the data in relation to the mean. The model prediction was influenced by this variation in the data, mainly in sand models. In all sets, the distribution of the properties is positively skewed.

Table 5. Descriptive statistics of soil properties for dataset, training set and assessment set.

	Dataset (%)		Training set (%)		Assessment set (%)	
	Sand	Clay	Sand	Clay	Sand	Clay
Observations	1534	1534	1151	1151	383	383
Minimum	1.00	0.00	1.00	0.00	3.00	0.00
Maximum	99.00	77.00	98.00	76.00	99.00	77.00
1st quartile	14.00	25.00	13.00	25.00	14.00	24.50
3rd quartile	40.00	52.00	39.00	52.00	43.00	52.00
Mean	28.85	38.25	28.30	38.45	30.53	37.63
Median	25.00	37.00	25.00	38.00	27.00	36.00
Std error of mean	0.50	0.40	0.55	0.50	1.09	0.91
Skewness	1.10	0.11	1.08	0.11	1.04	0.05
Kustosis	1.10	-0.80	1.17	-0.75	0.78	-0.83
CV (%)	67	45	67	44	70	47

Considering regional soil spectral libraries in Brazil, Ramirez-Lopez et al. (2013), working with a soil spectral library from the state of São Paulo, showed a wide range of variation for clay content (1% to 81%), with a mean value around 33.0%, considering a total of 4200 soil samples. Araújo et al. (2014) reported a mean value of clay content around 39.0% for 7172 samples representing a large soil spectral library of four states (Goiás, Minas Gerais, Mato Grosso do Sul and São Paulo). A spectral library comprising 1515 soil samples collected in the state of São Paulo was tested by Ackerson; Demattê; Morgan (2015). This dataset contained highly weathered soils with a mean value of clay content around 31.2%. Lacerda et al. (2016) worked with 3750 samples of a regional soil spectral library representing the state of São Paulo that showed high variation in particle size classes, varying from sandy to very clayey texture. Demattê et al. (2016b) worked with 7185 soil samples and general, regional and local spectral libraries representing several Brazilian states (São Paulo, Minas Gerais, Goiás, Amapá and Mato Grosso do Sul)

containing different ranges of data. They found great soil variability even in local spectral libraries.

Compared to published studies, the SC spectral library comprised soil samples with great variability of sand and clay content within a small territorial area. The variation in sand and clay content found in this spectral library was mainly caused by the great geological diversity of the parent material covering the state, which consists of granitoids, charnockitics, gnaisses, granites and marine sedimentary plains in the coastal region and volcanic rocks and quartz sandstones in the plateau region (COELHO; ROSSI; MATTOS, 2017; MAACK, 2001). However, samples with clay content equal to or greater than 35% comprised 85% of all dataset samples, indicating clay texture predominance among the soil samples, which is as expected, since the plateau region covers two-thirds of the surface of SC (MAACK, 2001) with finer-textured clay-rich soils.

Levene's test indicated a *p-value* of 0.609 and *p-value* of 0.175 for the equality of variances test between the training and assessment sets for sand and clay content values, respectively. No significant difference was observed between the variances, since the *p-values* were greater than the significance level of $\alpha = 0.05$. Comparison between the mean values for the training and assessment sets did not show a significant difference, according to Student's t test for sand (*p-value* = 0.179) and clay (*p-value* = 0.435), at a 5% significance level. Thus, statistical similarity (Table 5) between the sets indicates that the randomly split sample groups are representative of the population under study.

3.3.2. VIS-NIR-SWIR soil spectra characteristics

The spectral signature of soil samples carries information about its physical characteristics and chemical and mineralogical compositions, which are controlled by pedogenetic processes. Important absorption features of VIS-NIR-SWIR spectra have been presented in Table 6 and reported in a broad sense, for more detail see the relevant literature (e.g. BENDOR; IRONS; EPEMA, 1999; CLARK, 1999; VISCARRA ROSSEL; BEHRENS, 2010).

Table 6. Assignments for absorption features of soil constituents.

Soil constituents	VIS-NIR-SWIR wavelengths (nm)	References
Goethite	479 to 499, , 930, 950, 920	Scheinost et al. (1999), Sherman; Waite (1985), Viscarra Rossel; Behrens (2010)
Haematite	521, 565, 650	Scheinost et al. (1998), Viscarra Rossel; Behrens (2010), Demattê et al., 2010.
OH feature of hidration, higroscopi e free water	950, 1200; 1430, 1930, 2220	Clark et al. (1999)
Carbonate	1800, 2350, 2360, 2335	Bendor; Banin (1990b), Viscarra Rossel ; 104
Montimorillonite	2160 to 2170, 2204, 2220, 2280, 2350	Clark et al., 1999; Viscarra Rossel and Behrens, 2010
Kaolinite	1400, 2160, 2200, 2204, 2220, 2350, 2165	Hunt; Salisbury (1970), Clark et al. (1999), Viscarra Rossel; Behrens (2010), Demattê et al., 2010.
Ilita	2345, 2220, 2345, 2350, 2380	Post; Nobel (1993), Clark et al. (1999), Viscarra Rossel; Behrens (2010)
Organic compounds	410, 570, 660, 1100; 1600; 1700; 1800; 2000; 2200; 2316, 2382, 2400	Viscarra Rossel et al. (2006), Viscarra Rossel; Behrens (2010),

In order to illustrate soil variability, the reflectance spectral curves of all soil samples are shown in Figure 6. As reported by Waiser et al. (2006), a large variation in the shape and magnitude of the reflectance spectral curves was observed, this probably being due to the heterogeneity of soil textures and geological parent materials present in the dataset. The reflectance spectral curves exhibited the typical pattern of soil spectra, with the reflectance generally lower in the VIS range (350–700 nm) and higher in the NIR (700–1100 nm) and SWIR (1100-2500 nm) ranges.

Reflectance increased as iron content decreased and the soil texture changed from clayey to sandy (DEMATTÊ et al., 2007). Thus, soils with high sand content (highest reflectance spectra, Figure 6) present higher albedo compared to soils with high clay content (lowest reflectance spectra, Figure 6), which present high absorption energy. This is due to the high presence of quartz in the sand fraction increasing the intensity of spectral reflectance (albedo) (WHITE et al., 1997). Although sand content is one of the most important factors controlling soil spectral signatures, the VIS-NIR-SWIR region is less sensitive to SiO_2 related components such as quartz (EISELE et al., 2012; HUNT; SALISBURY, 1970) and has no distinct spectral features in this range (VISCARRA ROSSEL et al., 2006). In contrast, soil carbon and Fe-oxides (mainly hematite and goethite) absorb energy and decrease reflectance from the initial wavelengths by up to 1100 nm (DEMATTÊ et al., 2007). Soil spectra showed mineral absorption features in the VIS-NIR-SWIR range that are mainly associated with the presence of iron oxides. Demattê et al. (2017) found that goethite is defined at 425 nm, 480 nm, and 600 nm and hematite at 750 nm and 1050 nm, and the absorption feature of crystalline Fe is at 850 nm. Distinct absorption features were observed at around 1400, 1900 and 2400 nm in the SWIR range, due to OH- groups and water molecule vibrations (DEMATTÊ et al., 2017), while clay lattice OH features were also observed at 1400 and 2200 nm (CLARK, 1999). Small absorption bands near 850 nm, 1200 nm, 1400 nm and 1900 nm may be due to vibrational combinations and overtones of molecular water (HUNT; SALISBURY, 1970).

The high mineralogical variability of samples of the soil spectral library of SC, in terms of geological parent material, may be reflected in the variation of reflectance intensity and other features among soil spectra. This preliminary analysis of the soil spectral library of SC demonstrates that soil spectra are closely linked to key soil attributes, which supports the development of spectral prediction models.

3.3.3. Effects of the preprocessing techniques on modelling

The best preprocessing techniques were those that provided the best performance model for sand and clay content, that is, with the lowest value of

RMSE and the highest values of R^2 and RPIQ for assessment sets (Table 7 and 8; Figure 6).
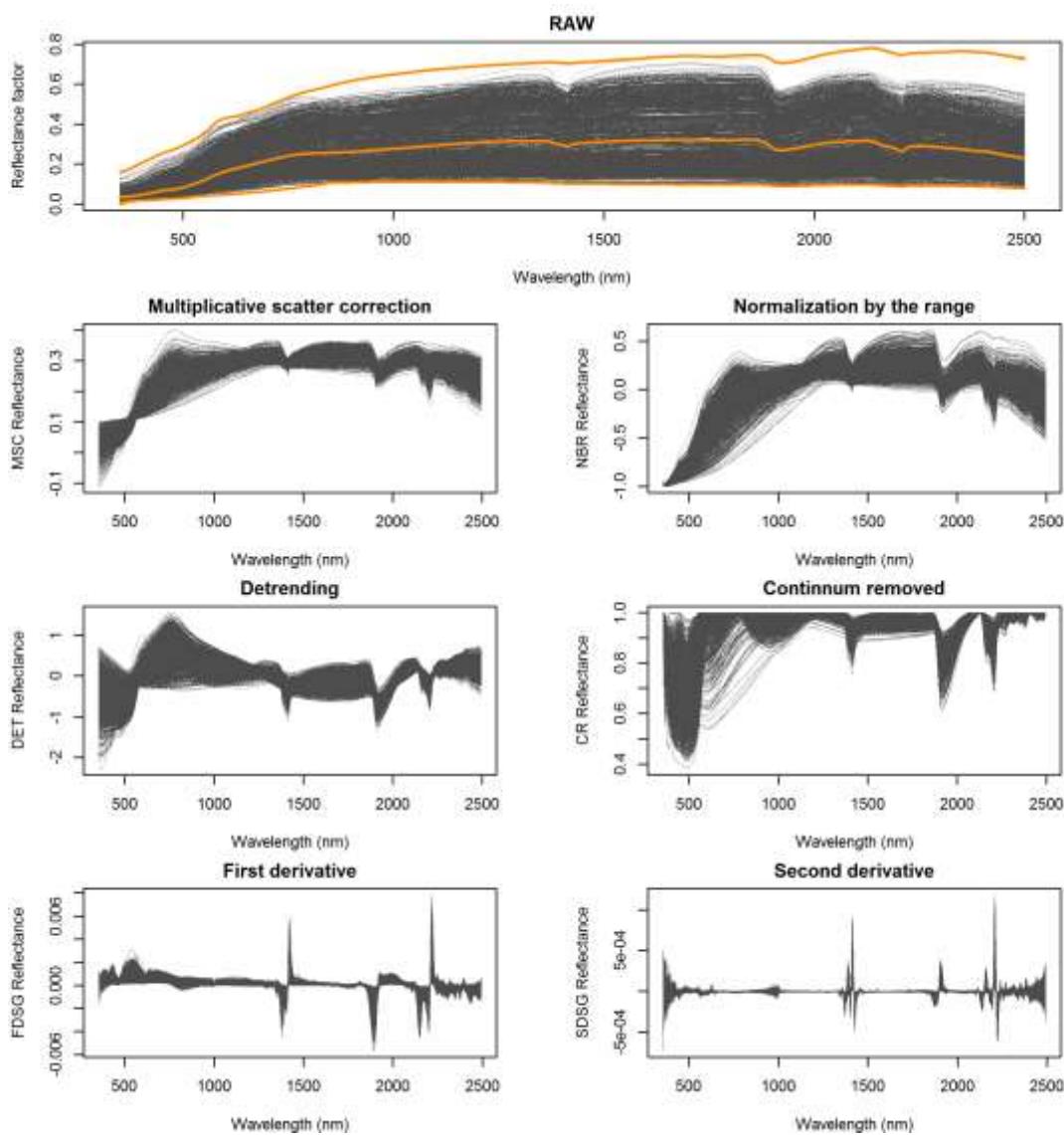
Figure 6. VIS-NIR-SWIR reflectance spectra and the preprocessing techniques of the spectral curves for all soil samples. The highlighted raw spectra correspond to the mean, the lowest and the highest reflectance spectra of the soil samples.

The best performance was found for the MSC and NBR scatter-correction groups and FDSG spectral derivatives in terms of performance models using preprocessing to predict sand and clay content. In general, model performances decreased as the noise resulting from preprocessing increased (from RAW to SDSG). The derivatives emphasize noise in the data more distinctly than the other methods (BUDDENBAUM; STEFFENS, 2012). Even though extremely flat structures can be evaluated with spectral derivatives

(FDSG and SDSG), this also tends to increase spectral noise (GARCÍA-SÁNCHEZ et al., 2017), especially by second spectral derivatives (BUDDENBAUM; STEFFENS, 2012). However, FDSG and SDSG still provide the best prediction of sand content when combined with an RF method. RF did not appear to be sensitive to enhanced noise in the spectrum (BREIMAN, 2001), revealing its capability to better handle derivative transformation (DOTTO et al., 2018). However, SDSG decreased model performance of the remaining multivariate methods for the two soil attributes studied. These results with spectral derivatives and ensemble-learning algorithms (RF) are in agreement with Dotto et al. (2018), Pinheiro et al. (2017), Santana; Souza; Poppi (2018), Vasques; Grunwald; Sickman (2008).

The RAW spectra achieved the best performance using Cubist, SVM and PLS methods for sand content (Table 7). Preprocessing the spectra prior to regression analysis did not improve sand prediction, with the exception of RF models. Therefore, spectral reflectance values without preprocessing (RAW spectra) were sufficient to obtain highly accurate models. There is no need to perform any preprocessing technique on the spectra to generate better prediction models for sand The results achieved in the current study are divergent from Franceschini et al. (2013), which found high performance of the sand model (R^2=0.87) with applying spectral preprocessing. Considering the GPR method, only a slight benefit was found using preprocessing on the spectra (NBR). This finding is in agreement with Duda et al. (2017), who reported no significant improvement in results with first derivative preprocessing compared with RAW spectra using the SVM method. They combined two proximal sensor approaches relative to a single sensor, to compare the efficacy in determining soil attributes, sand content and clay content, among others, in catena scale in Eastern Europe. Sawut et al. (2014) reported a small influence of preprocessing on spectral analysis for the prediction of sand content in a thermal infrared region.

Table 7. Performance of sand predictive models from five multivariate methods with the corresponding spectral preprocessing techniques.

Soil attribute	Method	Preprocessing	Assessment set		
			R^2	RMSE (%)	RPIQ
Sand	Cubist	RAW	0.73	10.6	2.4
		MSC	0.70	11.0	2.2
		NBR	0.67	11.7	2.1
		DET	0.65	12.0	2.1
		FDSG	0.64	12.0	2.1
		CRR	0.63	12.4	2.0
		SDSG	0.62	12.4	2.0
	RF	FDSG	0.68	11.6	2.2
		SDSG	0.63	12.8	2.0
		DET	0.61	12.8	2.0
		CRR	0.61	12.9	1.9
		MSC	0.61	12.8	1.9
		NBR	0.60	13.1	1.9
		RAW	0.57	13.4	1.9
	SVM	RAW	0.67	11.6	2.2
		NBR	0.64	12.1	2.1
		CRR	0.61	12.5	2.0
		MSC	0.60	12.8	2.0
		DET	0.58	13.0	1.9
		FDSG	0.58	13.4	1.9
		SDSG	0.19	27.9	0.9
	PLS	RAW	0.67	11.6	2.1
		FDSG	0.65	11.9	2.1
		NBR	0.63	12.2	2.0
		MSC	0.60	12.7	2.0
		CRR	0.60	12.7	2.0
		DET	0.59	12.9	1.9
		SDSG	0.54	13.7	1.8
	GPR	NBR	0.65	12.0	2.1
		RAW	0.64	12.1	2.1
		CRR	0.62	12.5	2.0
		MSC	0.61	12.5	2.0
		FDSG	0.61	12.8	1.9
		DET	0.58	13.0	1.9
		SDSG	0.30	20.8	1.2

Table 8. Performance of clay predictive models from five multivariate methods with the corresponding spectral preprocessing techniques.

Soil attribute	Method	Preprocessing	Assessment set		
			R^2	RMSE (%)	RPIQ
Clay	Cubist	MSC	0.83	7.3	3.7
		RAW	0.83	7.3	3.7
		DET	0.83	7.4	3.7
		NBR	0.82	7.5	3.6
		CRR	0.79	8.1	3.3
		FDSG	0.74	9.1	3.00
		SDSG	0.69	9.8	2.8
	SVM	NBR	0.77	8.7	3.1
		MSC	0.76	8.7	3.1
		DET	0.73	9.4	2.9
		RAW	0.69	9.9	2.7
		FDSG	0.63	11.6	2.3
		CRR	0.45	13.5	2.0
		SDSG	0.28	23.1	1.2
	GPR	NBR	0.76	8.8	3.1
		MSC	0.75	9.0	3.0
		DET	0.71	9.6	2.8
		RAW	0.67	10.3	2.6
		FDSG	0.67	10.5	2.6
		CRR	0.57	11.9	2.3
		SDSG	0.42	17.3	1.6
	RF	FDSG	0.76	8.8	3.0
		DET	0.70	10.0	2.7
		CRR	0.68	10.2	2.6
		MSC	0.68	10.3	2.6
		SDSG	0.69	10.3	2.6
		NBR	0.65	10.6	2.5
		RAW	0.63	10.9	2.5
	PLS	NBR	0.75	8.9	3.0
		MSC	0.73	9.3	2.9
		CRR	0.72	9.5	2.8
		DET	0.70	9.8	2.7
		RAW	0.70	10.1	2.7
		FDSG	0.67	10.5	2.6
		SDSG	0.55	12.4	2.2

Good performances were achieved with FDSG-RF (R^2 = 0.68, RMSE = 11.56%, RPIQ = 2.16,) for sand, but only small predictive accuracy was gained in relation to RAW spectra with the SVM method (R^2 = 0.67, RMSE = 11.60%, RPIQ = 2.16) and RAW spectra with the PLS method (R^2 = 0.67,

RMSE = 11.62%, RPIQ = 2.15) (Table 7). Preprocessing results for FDSG derivatives are in agreement with those found by Bilgili et al. (2010), Pinheiro et al. (2017) and Duda et al. (2017) , who reported different R^2 values 0.84, 0.62 and 0,25, respectively. Spectral preprocessing may emphasize the feature sought in the spectra and several authors have noted its benefit (DOTTO et al., 2018; NAWAR et al., 2016; VASQUES; GRUNWALD; SICKMAN, 2008). Sand fractions, tends to have quartz as the dominant mineral (DEMATTÊ et al., 2007) which has no diagnostic spectral features in the VIS-NIR-SWIR ranges, high values of reflectance intensity (albedo) and its reflectance spectrum is largely unvarying (CLARK, 1999; HUNT; SALISBURY, 1970; RAMIREZ-LOPES et al., 2013; WIGHT; ASHWORTH; ALLEN, 2016). Preprocessing techniques are designed for baseline corrections, so their effect is minimal when the baseline variance is small (BUDDENBAUM; STEFFENS, 2012). In general, sandy soils exhibits similar spectral behavior of quartz and this may provide an explanation as to why no preprocessing technique was very helpful in improving sand model performance in the present study.

For clay content the best performance was found with scatter-correction preprocessing, MSC (R^2 = 0.83, RMSE = 7.29%, RPIQ = 3.70), followed by DET (R^2 = 0.83, RMSE = 7.38%, RPIQ = 3.66) and NBR (R^2 = 0.82, RMSE = 7.50%, RPIQ = 3.60), using Cubist method (Table 8). MSC and NBR achieved the best performance with the SVM, GPR and PLS methods. These preprocessing techniques are normalization procedures commonly used to compensate for baseline shift and multiplicative effects in the spectral data, which are induced by physical effects such as particle size (GARCÍA-SÁNCHEZ et al., 2017; MARTENS; NAES, 1992; RINNAN; VAN DEN BERG; ENGELSEN, 2009). MSC, which attempts to eliminate the effects of the spectrum by linearizing each spectrum by the average spectrum of the sample, is the most popular normalization technique (MARTENS; NAES, 1992). In NBR, each spectrum is divided by the range. In this study MSC exhibited a slightly better performance than NBR, when combined with Cubist method, albeit with the same performance as RAW spectra (control treatment). NBR produced the best model result for the SVM, GPR and PLS methods.

CRR presented inferior performance than expected, given that this technique is commonly reported by other studies as an effective VIS-NIR-SWIR

data preprocessing technique (DOTTO et al., 2017, 2018; LAGACHERIE et al., 2008; NAWAR et al., 2016). CRR was introduced by Clark; Roush (1984) as a method to emphasize mineral energy absorption features (STEVENS; RAMIREZ-LOPEZ, 2013) and is commonly used for baseline corrections, especially for differences in illumination condition and in viewing geometry (BUDDENBAUM; STEFFENS, 2012). In our case, laboratory conditions of illumination and geometry were maintained constant, so it might explain the small improvement from using CRR.

In general, the scatter-corrective group gives similar, or better, performance than spectral-derivatives for sand and clay models. These results are in agreement with Dotto et al. (2017), who also reported better performance of the models with scatter-corrective preprocessing techniques compared to spectral-derivatives to predict soil organic carbon using a local spectral library from SC. All multivariate methods applied in the present study achieved different performance with two spectral preprocessing groups. For clay content, accuracy of the prediction models appears to depend on spectral preprocessing. For sand, this was not so evident. SVM seems to be more sensitive to spectral preprocessing applications, since R^2 dropped from 0.67 to 0.19 and from 0.77 to 0.28 for sand and clay content, respectively. The RMSE showed the inverse trend.

The effect of six preprocessing techniques fluctuated between models (Tables 7 and 8), so it is difficult to reach a clear conclusion as to whether the differences between the average values of R^2 and RMSE among preprocessing techniques are significant. In order to better compare these preprocessing techniques, including RAW spectra, the SK test of R^2 and RMSE mean values of the assessment set was carried out. It was observed that there was no statistical difference between the mean values of R^2 and RMSE for RAW spectra, NBR, MSC, DET, CRR and FDSG for sand and clay prediction models (letter a, Figure 3). In other words, on average, RAW spectra, NBR, MSC, DET, CRR, FDSG preprocessing techniques had equal effect on model performance to quantify sand and clay content and they clearly perform better than SDSG preprocessing. Thus, RAW spectra and MSC preprocessing are the best strategies for sand and clay content, as they consistently simplify the models. The worst results for both R^2 and RMSE were found with SDSG preprocessing

(letter b, Figure 7). There is significant variability across SDSG results. The poorest result was achieved by SDSG-SVM, which presented the lowest R^2 and the highest RMSE (0.19 and 27.94% for sand and 0.28 and 23.13% for clay).
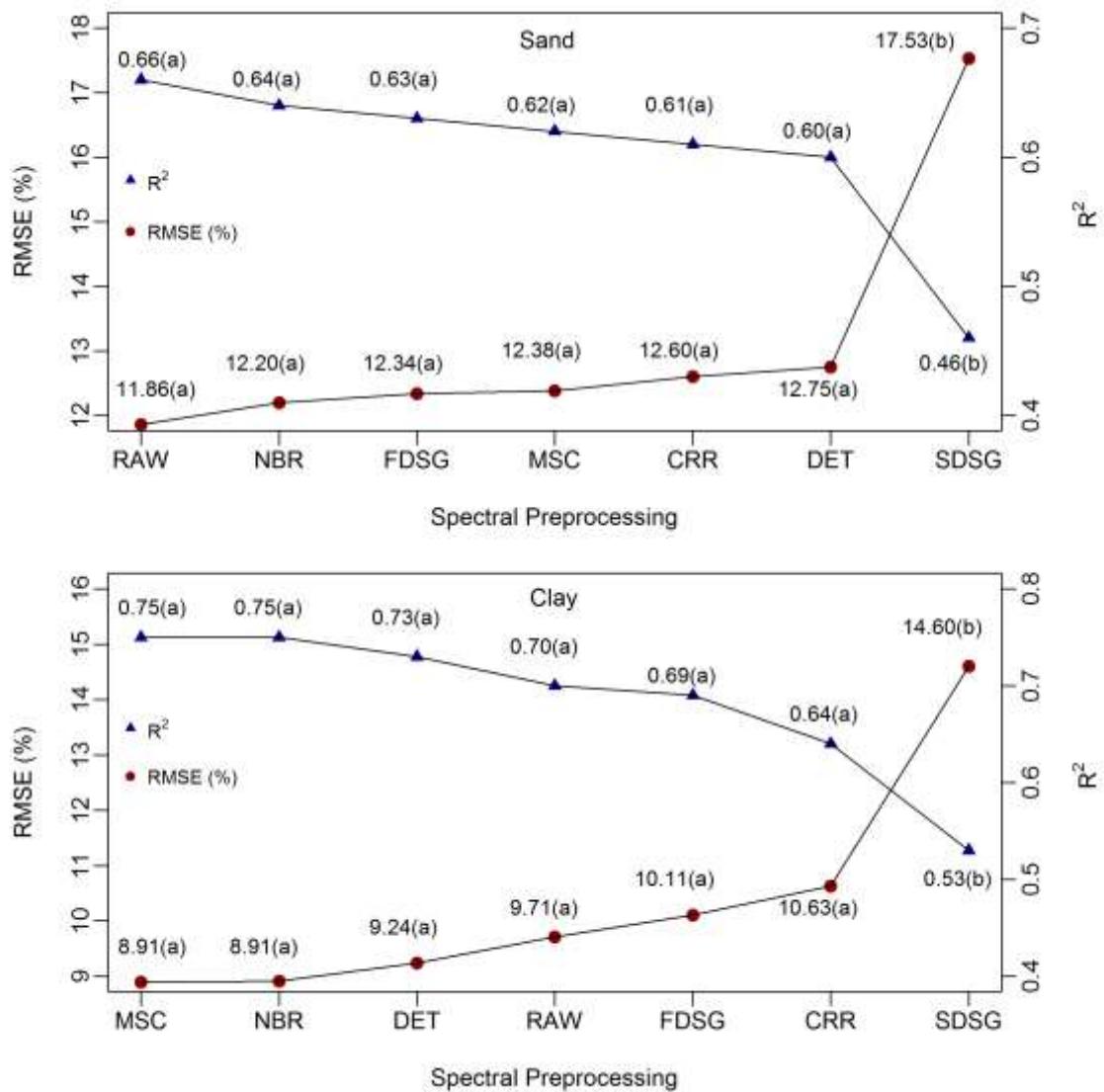
Figure 7. The mean values of RMSE and R^2 for each preprocessing technique for sand and clay content. The letters in parentheses represent the results of the Scott-Knott' test (significance level α=0.05).

3.3.4. Effects of the multivariate methods

The model approaches tested in this study are often used with highly dimensional and strongly correlated data. The predictive performance of sand content presented R^2, RMSE and RPIQ values ranging from 0.19 to 0.73, 27.94% to 10.60% and 0.89 to 2.36, respectively (Table 7). The predictive

performance for clay content presented R^2, RMSE and RPIQ values ranging from 0.28 to 0.83, 23.13% to 7.29% and 1.17 to 3.70, respectively. Between them the best R^2 values of this study were plotted in dark red color. Lacerda et al. (2016) developed a PLS prediction model to quantify soil texture from 3750 soil samples using the topossequence method from three areas of São Paulo state, Brazil. These authors found high-quality predictions in the validation set for sand (R^2 = 0.96, RMSE = 137.98 g kg^{-1}) and clay (R^2 = 0.93, RMSE = 82.50 g kg^{-1}) content. The R^2 and RMSE values are higher than those found in this study. The lowest R^2 found with the SC spectral library can be explained in terms of the high heterogeneity of soil samples collected throughout the state. Higher performance was achieved by Terra; Demattê; Viscarra Rossel (2015) in the VIS-NIR-SWIR region for clay (R^2 = 0.85, RMSE = 96.50 g kg^{-1}) and sand content (R^2 = 0.85, RMSE = 25.22 g kg^{-1}), using 1259 soil samples from four Brazilian states. In Nawar et al. (2016) the R^2 and RMSE for clay content in the validation set ranged from 0.52 to 0.79 and 11.35 to 7.75%, respectively, using different multivariate methods and preprocessing techniques based on a spectral library with limited soil samples (n = 102) from Northern Sinai, Egypt.

For sand, good predictions were obtained using all multivariate methods (Table 7). The best results were obtained by Cubist based on RAW spectra (R^2 = 0.73, RMSE = 10.60%, RPIQ = 2.36) and MSC preprocessing techniques (R^2 = 0.70, RMSE = 11.09%, RPIQ = 2.25), followed by lower performances using RF with FDSG (R^2 = 0.68, RMSE = 11.56%, RPIQ = 2.16), SVM-RAW (R^2 = 0.67, RMSE = 11.60%, RPIQ = 2.16) and PLS-RAW (R^2 = 0.67, RMSE = 11.62%, RPIQ = 2.15). RF outperformed the SVM, PLS and GPR models only with FDSG, whereas with all remaining spectral preprocessing showed similar performance (Table 7). Compared to published results using Cubist method (R^2 = 0.50) (e.g., ZENG et al., 2017), RF (R^2 = 0.82) (e.g., SANTANA; SOUZA; POPPI, 2018), SVM (R^2 = 0.25-0.90) (e.g., DOTTO et al., 2017; TERRA; DEMATTÊ; VISCARRA ROSSEL, 2015; VISCARRA ROSSEL; BEHRENS, 2010) and PLSR (R^2 = 0.33-0.96) (e.g., CONFORTI; MATTEUCCI; BUTTAFUOCO, 2017; DOTTO et al., 2017; LACERDA et al., 2016; PINHEIRO et al., 2017; SAWUT et al., 2014; TERRA; DEMATTÊ; VISCARRA ROSSEL, 2015; WETTERLIND; STENBERG; JONSSON, 2008) the results of the present study showed good quantitative predictions. However, all these studies used

different soils, different methodologies, multivariate methods and preprocessing techniques, with different population sizes, and VIS-NIR-SWIR or MID-IF in some situations.

The scatter plot (Figure 8) of laboratory-measured versus VIS-NIR-SWIR-predicted values of sand and clay content based on Cubist model using the assessment set showed quite low dispersion, with most of the values distributed close to the 1:1 line (red line), with small slope and intercept values, indicating good fit (VISCARRA ROSSEL; WEBSTER, 2012).
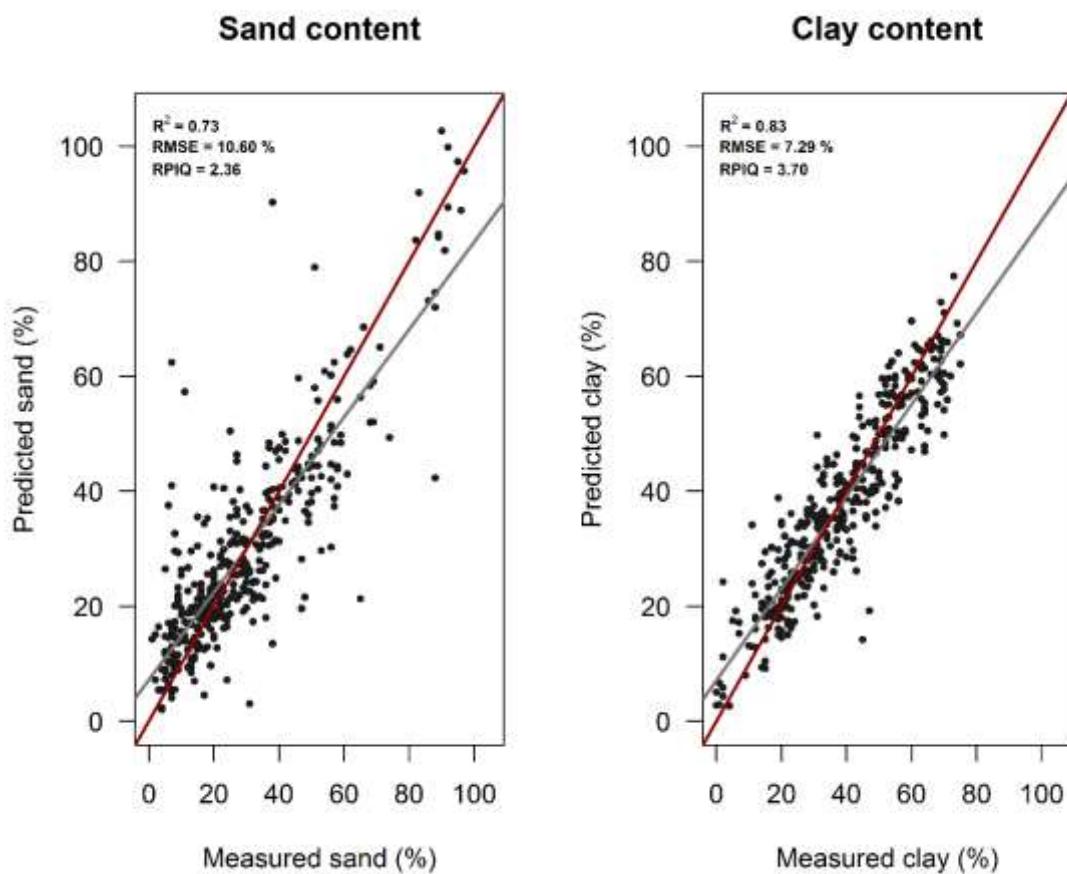
Figure 8. Assessment set of laboratory-measured versus VIS-NIR-SWIR-predicted values using Cubist (with RAW spectra) for sand and (with MSC spectra) for clay. Red line indicates 1:1 line.

The Cubist method showed the best performance in comparison with the other models for the prediction of clay content (Table 8). The R^2 values of the assessment set ranged from 0.69 to 0.83, whereas RMSE ranged from 9.79 to 7.29%. This confirms the superior performance of Cubist in predicting soil properties, as stated in the literature (MINASNY; McBRATNEY, 2008; MORELLOS et al., 2016; STEVENS et al., 2013; VISCARRA ROSSEL et al.,

2016; VISCARRA ROSSEL; WEBSTER, 2012). This result is in line with results reported by Viscarra Rossel; Webster (2012), who used Cubist to predict 24 soil attributes, including sand, silt and clay content, using a large soil dataset (N = 21,493) from all the states in Australia and found RMSE = 12.00% and RMSE = 8.49% for sand and clay content, respectively. They concluded that the rule-based method predicts sand and clay content well, working effectively with large and diverse datasets. A study presented by Minasny; McBratney (2008) showed that Cubist produced the best fit model (R^2 = 0.92) and lowest error (RMSE = 7.18%) when compared to PLS and another data-mining method, Treenet, using mid-infrared (2500-25000 nm) spectra of soil samples from Australia.

Scatter-correction (MSC and NBR) spectral preprocessing performed well and provided the best results for Cubist-MSC (R^2 = 0.83, RMSE = 7.29%, RPIQ = 3.70), SVM-NBR (R^2 = 0.77, RMSE = 8.69%, RPIQ = 3.11), GPR-NBR (R^2 = 0.76, RMSE = 8.80%, RPIQ = 3.07) and PLS-NBR (R^2 = 0.75, RMSE = 8.94%, RPIQ = 3.02), respectively. These results were consistent with previous studies estimating clay content based on VIS-NIR-SWIR regions (R^2 = 0.62-0.93) (ARAÚJO et al., 2014; DOTTO et al., 2017; LACERDA et al., 2016; LUCÀ et al., 2017; MINASNY; McBRATNEY, 2008; PINHEIRO et al., 2017; RAMIREZ-LOPEZ et al., 2013; SANTANA; SOUZA; POPPI, 2018; TERRA; DEMATTÊ; VISCARRA ROSSEL, 2015; VISCARRA ROSSEL; BEHRENS, 2010; VISCARRA ROSSEL; WEBSTER, 2012). Similar performance was achieved with RF-FDSG (R^2 = 0.76, RMSE = 8.85%, RPIQ = 3.05). Machine learning algorithms (Cubist, SVM, GPR and RF) outperformed the PLS approach (Table 8). This better performance may be explained by the inclusion of non-linear relationship between clay content and spectra, interaction effects of the regression task, as well as linear combinations of variables (GOMEZ et al., 2016; KOVAČEVIĆ; BAJAT; GAJIĆ, 2010). In the study presented by Dotto et al. (2017), SVM and PLS methods were applied to the prediction of soil organic carbon (SOC), sand, silt and clay content using VIS-NIR-SWIR ranges. A total of 299 soil samples from the central region of SC were used and statistical differences between the RMSE mean values of the SVM (RMSE = 7.68%) and PLS (RMSE = 8.58%) methods were found, whereby SVM produced the best fitting model (R^2_{val} = 0.62) and the lowest error ($RMSE_{val}$ = 6.84%) for clay content estimation. For SOC, sand and silt, they did not find any

statistical differences between the two multivariate methods. On the other hand, Santana; Souza; Poppi (2018) compared RF and PLS to assess sand and clay content and found a significant difference between the results of the two methods, RF (RMSE$_{val}$ = 7.61%) and PLS (RMSE$_{val}$ = 8.82%) for clay content, with RF proving to be the better approach. These authors used 641 soil samples from several regions of Brazil.

PLS is the most linear common multivariate method for quantitative spectroscopy analysis in soil. This method is based on decomposition of spectral data into latent variables that capture most of the variance existing in the spectrum, and linear models are then created using the scores of the most correlated features (MORELLOS et al., 2016). However, in PLS, non-linear relationships can only be modeled in a limited way and the model is a linear function of all wavenumbers, whereas in regression-rule models like Cubist these non-linearities can be efficiently modeled using a set of comprehensible linear equations (MINASNY; McBRATNEY, 2008).

The SK test presented the means comparison test for five multivariate methods. For sand, the SK test showed that there was no statistical difference between R^2 and RMSE mean values of the five methods applied, using a 5% significant level. In the SK test, Cubist (R^2 =0.66, RMSE=11.73%), RF (R^2=0.62, RMSE=12.76%), PLS (R^2=0.61, RMSE=12.54%), GPR (R^2=0.57, RMSE=13.68%) and SVM (R^2=0.55, RMSE=14.77%) presented the same performance in sand prediction (Figure 9). However, when comparing R^2 and RMSE performance for each method (Figure 9), it was clear that Cubist achieved the best performance, followed by regular performance by the PLS, RF and GPS methods, with the poorest performance being found for SVM, although it remained acceptable for sand prediction.

For clay, the SK test showed different results for R^2 and RMSE between the five methods. When comparing R^2 mean values, clay content was the only soil attribute where the performances between Cubist and the remaining multivariate methods presented a statistical difference (Figure 9). The results showed that the methods were divided into two groups (blue and dark red color groups), of which Cubist presented the best performance. However, for the mean RMSE values, there was no statistical difference in the SK test (α = 10%).The prediction quality decreased from Cubist to SVM, with mean RMSE

values increasing from 8.06 to 12.13%, suggesting different performances between the methods. These results again demonstrated the better performance of the Cubist method in comparison to the most common algorithms used in spectroscopy analysis.
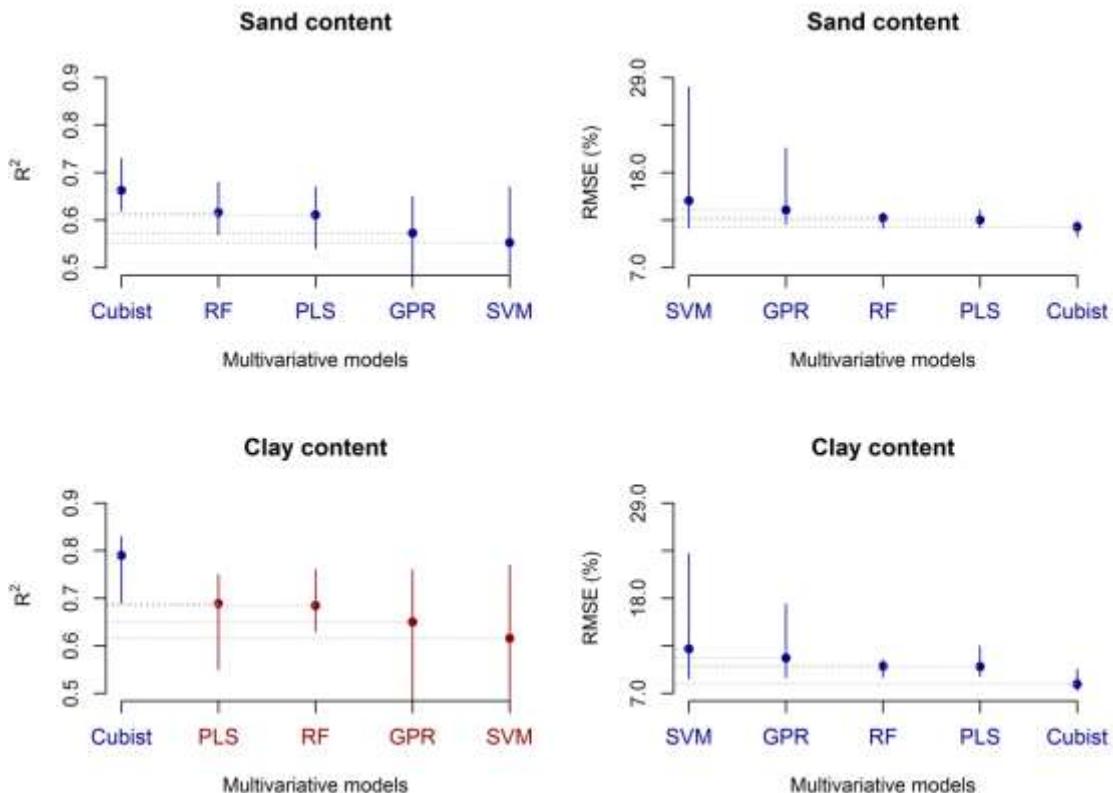
Figure 9. Statistical difference between multivariate methods resulted by Scott Knott test (significance level of 10%). The mean, maximum and minimum values of R^2 and RMSE for each method applied for sand and clay content.

3.4. Conclusions

This study evaluated a systematic spectroscopy approach to predict sand and clay content using different multivariate methods and preprocessing techniques based on a regional soil spectral library. The scatter-corrective preprocessing groups (MSC, NBR, DET and CRR) produced similar or better performance than spectral-derivatives (FDSG and SDSG). In addition, preprocessing spectra prior to regression analysis does not improve sand prediction, since RAW spectra achieved the best performance using Cubist, SVM and PLS methods. Spectral-derivatives only showed better results with RF

methods for both attributes. Considering R^2 and RMSE mean values, there were no statistical differences between preprocessing techniques applied for sand and clay content, except for the second derivative which produced the worst prediction.

In general, clay content presented better prediction accuracy than sand content. The best multivariate method to predict sand and clay content from soil VIS-NIR-SWIR spectra was Cubist. The best Cubist performance was achieved combined with RAW spectra (R^2 = 0.73; RMSE = 10.60%; RPIQ = 2.36) and MSC (R^2 = 0.83; RMSE = 7.29%; RPIQ = 3.70) for sand and clay content, respectively. Considering the mean RMSE values of the assessment set, the predictive ability of the multivariate methods decreased in the following order: Cubist>PLS>RF>GPR>SVM for both attributes. However, when comparing multivariate methods used for clay content, only R^2 mean values revealed a statistical difference between Cubist and the other four algorithms (PLS, RF, SVM and GPR), confirming the better performance of the rule-based method.

There is still no best preprocessing and multivariate method for prediction of all soil attributes from spectral data. However, the predictive ability of VIS-NIR-SWIR reflectance spectroscopy achieved in this study for sand and clay content using legacy data and heterogeneous soil samples confirmed the high potential of the spectroscopy approach. Furthermore, this study contributes towards understanding of the behavior of VIS-NIR-SWIR soil spectra in the Brazilian subtropical region.

These legacy soil data produced reliable information and it can be used to populate a soil database as input to soil assessment and monitoring, to guide new soil researches and field sampling strategy, as well as, a primary reference to establish standards of spectral behavior of SC soils. The use of legacy data did not required time and costs in sampling and laboratory analysis of soil attributes. For future work, new soil sampling may be required to populate the spectral library with more soil samples and soil profiles to adequately describe the SC soil variability and improve reliable soil information, considering its application in Brazil's National Soils Program (Pronasolos).

4.CAPÍTULO III - PRACTICAL APPLICATION OF SOIL SPECTROSCOPY TO CLASSIFY TEXTURE FOR AGRICULTURALINSURANCEC PURPOSE

4.1. Introduction

Agriculture is an economic activity characterized by its vulnerability to adverse events, since such adversities may generate significant losses and therefore represent a significant risk to production (LOYOLA; MOREIRA; DA VEIGA, 2016). To minimize these losses one of the most effective risk response mechanisms is crop insurance, which provides coverage against a drop in farmers' income because of low productivity, caused by critical weather events, among others (LOYOLA; MOREIRA; DA VEIGA, 2016). In this context, soil data, especially soil texture, plays a key role in the management of weather-related risk in crop and index insurance programs (BRYANT; O'CONNOR, 2017; VALVERDE-ARIAS et al., 2018a, 2018b; WOODARD, 2016a; WOODARD; VERTERAMO-CHIU, 2017). The impact of not integrating widely-available high resolution soil data into established insurance guarantees or rates in federal crop insurance programs (FCIP-USDA) was evaluated by Woodard;Verteramo-Chiu (2017). It was found that when ignoring soil data, it is not uncommon to have cases in which expected loss costs over program rates more than double. On the other hand, incorporating soil quality information into insurance rating procedures, based only on historical losses, increases pricing efficiency and improves soil health (WOODARD, 2016b).

The advent of big data represents a great potential source of information for the exploration of opportunities to employ soil databases on a

large scale, in order to establish feasible agricultural insurance schemes in terms of time and effort. The VIS-NIR-SWIR soil spectral libraries have become powerful tools for storing large amounts of soil information (McBRATNEY; MINASNY; VISCARRA ROSSEL, 2006), enabling spatio-temporal monitoring of soils in many regions worldwide (RIZZO et al., 2016) on local, regional, national and global scales (VISCARRA ROSSEL et al., 2016). VIS-NIR-SWIR reflectance spectroscopy (350-2500 nm) has been developed greatly over the last 30 years (VISCARRA ROSSEL et al., 2016), receiving attention as a complement to traditional laboratory methods. The technique stands out for its simplicity of sample preparation, the lack of necessity to use hazardous chemicals, its speed of measurement, the fact that various soil attributes can be analyzed from a single spectrum, and that the technique can be used both in the laboratory and *in situ* (VISCARRA ROSSEL et al., 2016).

Several studies in this area have shown good results in the prediction of soil attributes such as soil organic carbon/matter (R^2_{cal}=0.33-0.98; R^2_{val}=0.50-0.90) (ARAÚJO et al., 2014; CONFORTI; MATTEUCCI; BUTTAFUOCO, 2017; DOTTO et al., 2017; MORELLOS et al., 2016; NAWAR et al., 2016; NAWAR; MOUAZEN, 2017b; VENDRAME et al., 2012; XU et al., 2018), pH (R^2_{cal}=0.27-0.73; R^2_{val}=0.52-0.86) (CONFORTI; MATTEUCCI; BUTTAFUOCO, 2017; VENDRAME et al., 2012; XU et al., 2018), total nitrogen (R^2_{cal}=0.35-0.97; R^2_{val}=0.73-0.82) (CONFORTI; MATTEUCCI; BUTTAFUOCO, 2017; MORELLOS et al., 2016; NAWAR et al., 2016; NAWAR; MOUAZEN, 2017b; XU et al., 2018) and soil moisture (R^2_{cal}=0.48-0.92; R^2_{val}=0.80-0.81) (MORELLOS et al., 2016; NAWAR; MOUAZEN, 2017a), among others, using different chemometric approaches. The predictive capability for soil texture has been reported in recent studies with excellent performance for sand content (R^2_{cal}=0.68-0.84; R^2_{val}=0.13-0.96), silt content (R^2_{cal}=0.71-0.81; R^2_{val}=0.28-0.70) and clay content (R^2_{cal}=0.53-0.93; R^2_{val}=0.42-0.93) (ARAÚJO et al., 2014; CONFORTI; MATTEUCCI; BUTTAFUOCO, 2017; DOTTO et al., 2017; GHOLIZADEH et al., 2016; NAWAR et al., 2016; VENDRAME et al., 2012; XU et al., 2018). In addition, the characterization of sand and clay contents in textural classes using spectroscopy at VIS-NIR-SWIR has been observed and reported by various authors, such as Lacerda et al. (2016); Demattê et al., (2016); Dotto et al., (2016) and Bilgili et al., (2010).

While many studies apply VIS-NIR-SWIR spectra to predict soil attributes for different agricultural purposes, the applicability of such predictive models to classify soils for agricultural insurance programs has not been well documented. For this purpose, soils can be classified within the required range or threshold, rather than at a rigid numerical value. In addition, studies in this area should focus on the development of methods on a regional scale, or in limited geographic areas, since insurance coverage is applied by crop regions. In Brazil the Agro-Climatic Risk Zoning (ACRZ; or "ZARC - *Zoneamento Agrícola de Risco Climático*") is a technical-scientific tool applied to crop regions to minimize risks in agricultural production (ASSAD, 2008). In this tool, soil data is characterized by soil particle size distribution following textural classes set by Normative n.2 of the Ministry of Agriculture, Livestock, and Food Supply (or "MAPA - *Ministério de Agricultura, Pecuária e Abastecimento*") (BRASIL, 2008). ACRZ is the reference for the Farm Activity Guarantee Program for Smallholders (FAGPS; or "Proagro Mais - *Programa de Garantia da Atividade Agropecuária da Agricultura Familiar*"), which is one of the largest Brazilian public programs using agricultural insurance as a risk management mechanism.

In this context, it is hypothesized that VIS-NIR-SWIR spectroscopy could be used to classify soil samples into the required soil textural range for agricultural insurance purposes in the state of Santa Catarina (SC). Studies applying soil spectroscopy are incipient in SC, and to the best of our knowledge, this is the first involving the soil spectral library in the entire region. Thus, the objective of this study it to determine to what extent the soil clay content predicted by VIS-NIR-SWIR reflectance spectroscopy can be used to correctly classify soils into textural classes, considering MAPA regulations (BRASIL, 2008).

4.2. Material and Methods

4.2.1. Soil spectral library

In this study, soil samples were collected from 260 municipalities out of a total of 295 in SC. This soil spectral library is well representative of the state of SC, having samples distributed throughout its territory, varying from 1 to

20 samples per municipality and covering the entire state. The number of samples is considered large when compared to other regional libraries in the world (BELLINASO; DEMATTÊ; ROMEIRO, 2010; RAMIREZ-LOPEZ et al., 2013; VISCARRA ROSSEL et al., 2016). The climate of SC is classified into two types according to the Koeppen classification; super humid and mesothermal (Cfa) and quite humid and mesothermal (Cfb). Cfa with mean annual temperature of 18.5 ºC and annual precipitation of 1,800 mm occurs in Atlantic coastal areas and in the western region. Cfb with average maximum temperatures ranging from 20 to 26 °C, average minimum temperatures from 9 to 13 °C and annual precipitation of 1,600 mm occurs in the Plateau region (PANDOLFO et al., 2002). Soils are diverse and represent several groups, according to the IUSS Working Group WRB (2014) classification, including Acrisols, Lixisols, Alisols, Ambisols, Chernozems, Podzols, Gleysols, Solonchaks, Ferralsols, Fluvisols, Leptosolss, Arenosols, Regosols, Nitosols and Histosols (EMBRAPA, 2004). The remaining original vegetation includes areas of Rain Forest and four major subtypes - Dense Rain Forest, Araucaria Forest, Alpine Grassland and Deciduous Forest, besides Coastal Vegetation (KLEIN, 1978).

The legacy soil samples were carried out in Reference Units between April and September 2009. These Reference Units are smallholder farms and were implemented to monitor Family Agriculture Insurance for the main annual and perennial crops financed for state farmers. A total of 1,525 samples were collected from the upper soil layer (0-50 cm), by technical personal of the official technical assistance and rural extension, from different crops in smallholder farms, with a composite sample of five points per area. The locations of the farms were not geographically referenced since no allocation system (e.g. GNSS) was available at the time, hence the best spatial approximation of sample areas is at the level of municipality.

Nevertheless, a standard analytical method was applied to all samples. Prior to physical analysis, the soil samples were air-dried, ground and passed through a 2 mm sieve. Particle size distribution in percentage of clay (< 0.002 mm), silt (0.002 to 0.05 mm) and sand (0.05 to 2 mm) was analyzed using the Pipette method, with NaOH dispersant according to the standard soil analysis method (TEIXEIRA et al., 2017).

4.2.2. VIS-NIR-SWIR measurement of soil samples

Spectral reflectance of soil samples was obtained using a FieldSpec 3 spectroradiometer (Analytical Spectral Devices, Boulder, USA) in the VIS-NIR-SWIR (350-2500 nm) range following the routine protocol developed for the Brazilian Soil Spectral Library (ROMERO et al., 2018). For each sample, approximately 15 cm^3 of soil was placed in a petri dish and shaken to ensure a smooth surface for spectrum acquisition. The spectral sensor captured the light through a fiber-optic cable connected to the sensor and placed vertically within 8 cm of the sample, where reflected light in an area of approximately 2 cm^2 in the center of the sample was measured. The light source was a 50W halogen bulb with the beam non-collimated to the target plane, positioned 35 cm from the sample with a zenithal angle of 30º. All spectral measurements were carried out in a dark room to avoid interference from stray light. As a reference standard, a white Spectralon® plate was used before scanning and each 20 reads. Each spectrum was averaged from 50 reads. A total of three scans were collected from each sample and averaged into one spectrum.

4.2.3. Statistical analysis

Descriptive statistics were used to summarize the dataset. Normalization by range (NBR) was used as a preprocessing techniques of the soil spectral data to improve training model. Normalization was applied as a scatter-corrective preprocessing method on all the reflectance spectra to reduce physical variability in the samples, in order to enhance the feature sought in the spectra (RINNAN; VAN DEN BERG; ENGELSEN, 2009). All statistical and model development analyses were performed using R v.3.4.0 software (R CORE TEAM, 2017).

Prior to development of the model, the total number of soil samples (1,525) was randomly split into a training set [~ 75 %] (n = 1,143) and an assessment set [~ 25 %] (n = 382). Levene's test (LEVENE, 1960) was used to verify the assumption that variances are equal across training and assessment groups, with significance level of 5%, and the independent t-test was used to

determine whether a statistically significant difference exists between the means in the two unrelated groups.

As suggested by KOVAČEVIĆ; BAJAT; GAJIĆ (2010) to handle the calibration of large heterogeneous sample populations, a support vector machine regression (SVM) was used to produce predictive models for clay content. SVM is a supervised, nonparametric kernel-based learning method (VAPNIK, 1995). Kernel is a function that quantifies the similarity of two observations (JAMES et al., 2013). In the present study the linear kernel function was used. The fit and predictive accuracy assessment of the model used the parameters of coefficient of determination (R^2) between observed and predicted values and root mean square error of prediction (RMSE), calculated using Equations (1) and (2).

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

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

where $\hat{y}_i$ is the value predicted by the model; y_i is the measured value; $\bar{y}_i$ is the average value; and N is the number of samples.

To classify the soil samples into textural classes, laboratory clay content (%) was used as reference data and compared with the predicted value of the spectroscopy model. The soil spectral library was divided into four classes based on MAPA regulations on textural classes (BRASIL, 2008), mainly considering clay content, as follows: "class 0" - clay content less than 10% (soil class not included in the financial support); "class 1" – clay content ranging from 10% to 15% (soil class with restrictions for financial support in agricultural land with more than 20% of drought risk); "class 2" - clay content ranging from 15% to 35% (soil class with no restrictions for financial support), and "class 3" – clay content equal to or greater than 35% (soil class with no restrictions for financial support). The comparison between soil classes was performed using the confusion matrix. The confusion matrix describes the co-occurrence of soil textural classes between two clay content matrixes, by calculating the agreement (A) and overall agreement (OA) (COHEN, 1960).

4.3. Results and discussion

4.3.1. Soil attributes

The ranges of sand and clay content were wide, as were their coefficients of variation (CV), indicating that all textural classes are presented with great variability among the samples (Figure 10). This is due to the spatial distribution of soil-forming factors and processes over time (BOCKHEIM et al., 2014) resulting in high diversity of soil classes across the state. These two soil textural fractions had a close to normal distribution, as indicated by their mean and weak positive value for skewness. The samples represent a wide range of clay content with an overall variation of the whole dataset from 0.0 to 77.0%, with a mean value of 38.2% (Table 9), indicating large variability in terms of particle size distribution and textural classes. Most samples were classified into class 3 (838), followed by 589 samples into class 2, 50 samples into class 1 and 48 samples into class 0 (Figure 10). Sand content had the highest standard deviation (19.5%) (Table 9). Descriptive statistics for the training and assessment sets were comparable to those of the whole dataset. The CV was similar for the whole dataset and the training set but quite different for the assessment set, showing the greatest dispersion of the mean.

Table 9. Descriptive statistics of the entire dataset, training and assessment sets of measured sand, silt, and clay content.

|	Sand (%)			Clay (%)		
	Whole	Train	Asses.	Whole	Train	Asses.
Mean	28.80	28.30	30.50	38.20	38.40	37.60
Std error of mean	0.50	0.50	1.10	0.40	0.50	0.90
Std deviation	19.50	18.80	21.20	17.30	17.10	17.90
Kustosis	1.10	1.20	0.80	-0.80	-0.70	-0.80
Skewness	1.10	1.10	1.00	0.10	0.10	0.00
Minimum	1.00	1.00	3.00	0.00	0.00	0.00
Maximum	99.00	98.00	99.00	77.00	76.00	77.00
Coefficient of variation	67	67	70	45	44	47
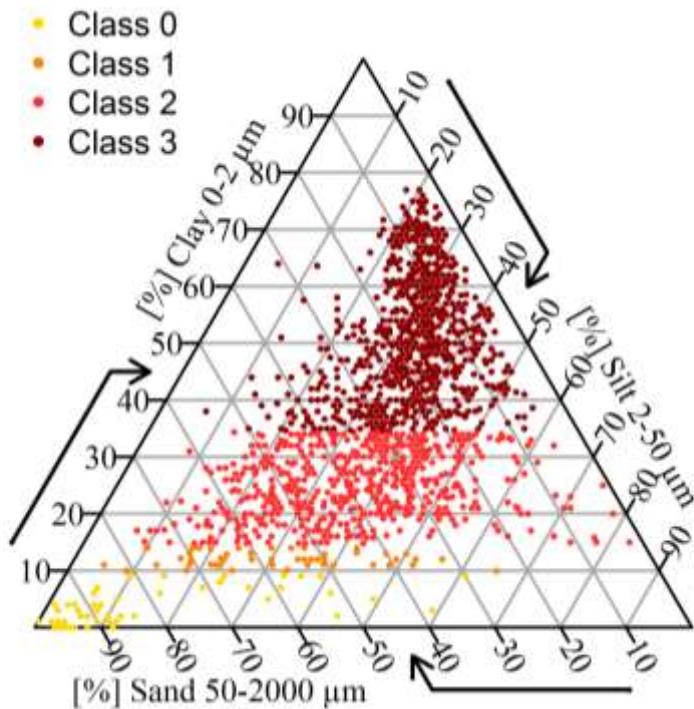
Figure 10. Soil particle size distribution of 1,525 samples measured by pipette method into the texture diagram. Class 0 is represented by the yellow points with clay content less than 10%; class 1 by orange points with clay content ranging from 10% to 15%; class 2 by red points with clay content ranging from 15% to 35%, and class 3 by dark red points with clay content greater than or equal to 35.

4.3.2. Characterization of soil spectra

In accordance with Demattê; Terra, (2014), three main topics of descriptive information can be observed on the spectral curves, as following: reflectance intensity variation (albedo), absorption features (depth and amplitude) and the overall spectral signature shape. A large variation in the features of mean spectral curves was observed. Soil reflectance intensity was higher for class 0 and 1 and decreased in class 3 samples throughout the entire range when the mean spectrum of each textural class was considered (Figure 11). High variability in terms of parent material and soil textural classes in SC can be reflected in variations in reflectance intensity and absorption features among soil spectra. In addition, a higher albedo response and angular alteration, ascending feature (Figure 11), can be observed on the spectral

signature from class 3 (high clay content) to class 0 (low clay content), due to quartz influence on the SWIR, in agreement with previous observations by Demattê (2002) in Brazilian soils. The ascendant aspect shows an increase in intensity and shape of spectral curves in SWIR as the sample contains less clay and silt, while in the VIS-NIR-SWIR range reflectance remains low, due to the presence of organic matter and Fe oxides that absorb energy in this range (DEMATTÊ, 2002; LACERDA et al., 2016) (Figure 2). According to Stevens et al. (2013), variations in clay content induce large differences in spectral shape with increased depth of absorptions related to OH and metal-OH in the mineral crystal lattice and OH in water, with increased clay content. As indicated by Demattê et al. (2016a), it is important to evaluate the descriptive information of spectra to see if it is coherent, before advancing to the statistical stage. A logical series of increased reflectance from class 3 < class 2 < class 1 < class 0 showed (Figure 11) a close relationship between soil reflectance spectral features and soil attributes, enabling separation of soil samples into textural classes as observed in previous studies by Bilgili et al., 2010; Dotto et al., 2016; Franceschini et al., 2013 and Lacerda et al., 2016.
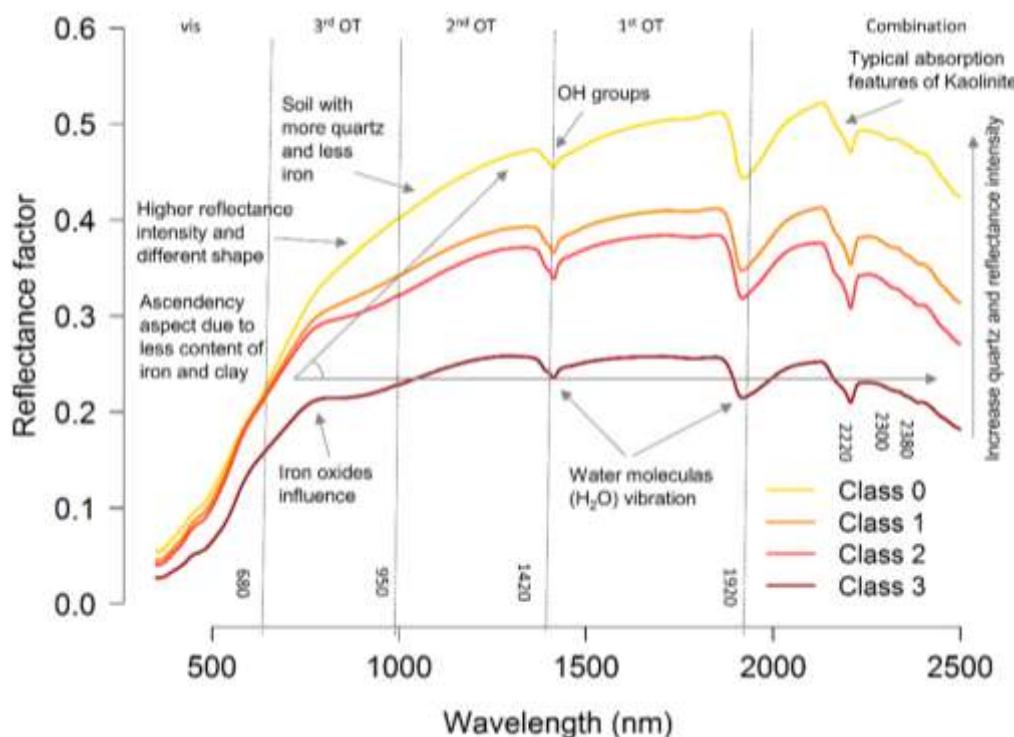

Figure 11. Mean reflectance soil spectra into four classes based on MAPA regulation of textural classes.

4.3.3. Performance of SVM for estimating clay content

Levene's test indicated variance equality for clay content values between the training and assessment sets (p-value = 0.175). The Student's "t" test demonstrated that clay content values for the training and assessment sets did not show a significant difference (p-value = 0.435), at a α = 0.05 significance level. The performance of spectroscopy to predict clay content is also demonstrated by descriptive statistics presented in Table 10. The average clay contents in the training and assessment sets for all classes are very similar. All the statistical parameters are very close for measured and predicted values in both datasets.

The results of the present study showed good predictive performance of clay content for the training (R^2_{train}=0.79 and $RMSE_{cal}$ = 7.75%) and assessment sets (R^2_{val}=0.77 and $RMSE_{val}$=8.69%) (Figure 12). These findings are consistent with previously reported results based on reflectance spectroscopy (ARAÚJO et al., 2014; DEISS; FRANZLUEBBERS; DE MORAES, 2017; GHOLIZADEH et al., 2016; TERRA; DEMATTÊ; VISCARRA ROSSEL, 2015; VISCARRA ROSSEL; BEHRENS, 2010; WANG et al., 2016). In the study presented by Dotto et al. (2017), located in the central region of SC, SVM produced the best-fitting model (R^2_{val}=0.62) and lowest error ($RMSE_{val}$=6.84%) for clay content estimation with a dataset of 299 samples and clay content ranging from 20.94 to 78.48%. Nawar et al. (2016) worked with a local soil spectral library (102 samples) with a wide range of clay contents (0 to 54.3%) on the Sinai Peninsula, Egypt. Despite the small dataset, they achieved high-quality results by applying SVM to predict clay content (R^2_{val}=0.64 and $RMSE_{val}$=8.84%). As suggested by Kuang and Mouazen (2011), the good prediction accuracy achieved in the present study occurred as a result of high soil heterogeneity and a wide variation in clay content, as shown in Table 9. On the other hand, this large variation increased the value of $RMSE_{val}$ (KUANG; MOUAZEN, 2011).
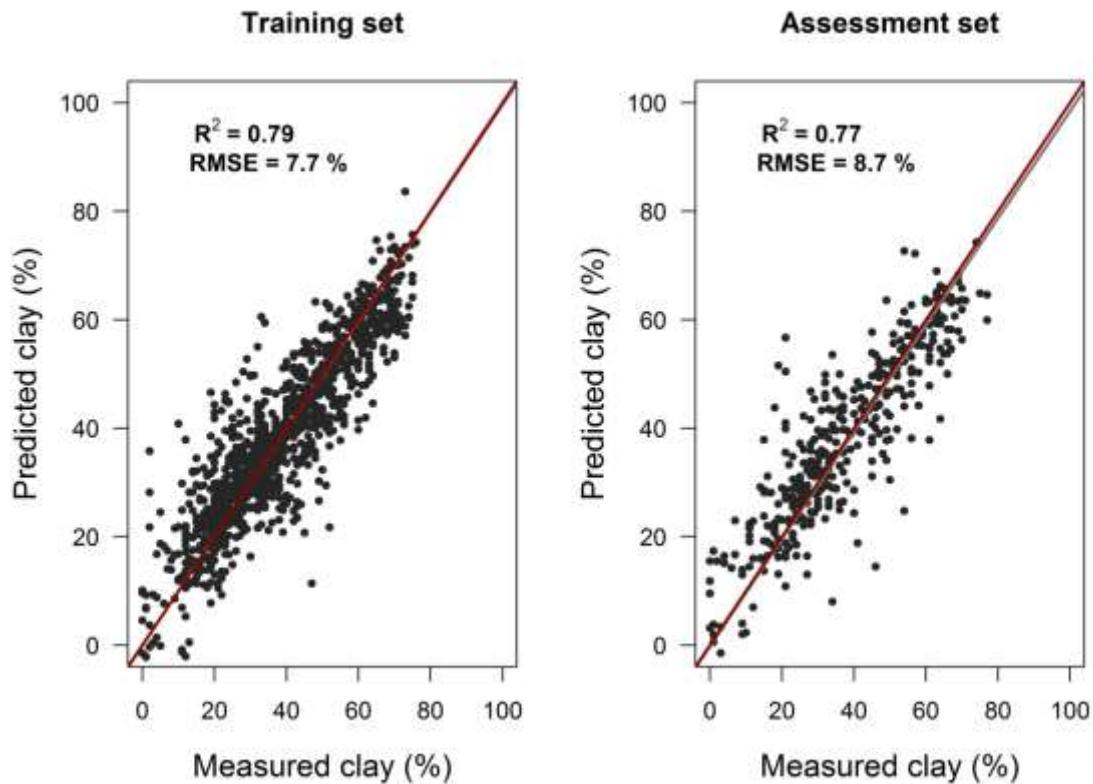
Figure 12. Scatter-plots of predicted versus measured values of clay content using SVM-NBR model for the training and assessment sets. Dark red indicates 1:1 line.

Descriptive statistics analyses were used to assess the performance of SVM for measured and predicted clay content (Table 10). It can be observed that the measured and predicted values of clay content are very close, with very low standard error of the mean. Cantarella et al. (2006) determined a tolerance interval of 20% (minimum and maximum values) as acceptable for discrepancies in results when they assessed the performance of different commercial laboratories using traditional methods to estimate discrepancies between results. If the 20% criterion is applied in the present study (Table 10), with the average measured clay content value (e.g. class 3) being 520 g.kg^{-1}, the minimum and maximum acceptable spectral model predicted values would be 416 g.kg^{-1} and 624 g.kg^{-1}, respectively. The average predicted clay content value (e.g. class 3) was 498 g.kg^{-1}, thus within the 20% confidence interval. The average predicted clay content value for all textural classes (0, 1, 2 and 3) lies within the tolerance interval established by laboratories, so it was considerate acceptable. These findings are corroborated by Demattê et al. (2010) and

Table 10. Descriptive statistical analysis for clay content measured by reference method and predicted by SVM model in assessment sets.

	Measured values (%)				Predicted values (%)			
	Classes				Classes			
	0	1	2	3	0	1	2	3
Mean	3.8	11.7	25.4	52.0	3.6	12.7	25.5	49.8
Std error of mean	0.8	0.3	0.5	0.7	0.5	0.2	0.3	0.4
Median	3.0	11.0	26.0	52.0	3.0	13.0	26.0	50.0
Standard deviation	3.5	1.2	5.7	10.6	2.3	1.4	5.6	9.4
Kustosis	-1.6	-0.5	-1.1	-0.9	-0.8	-1.1	-0.3	-0.8
Skewness	0.3	0.8	-0.2	0.2	0.8	-0.6	-1.0	0.3
Minimum	0.0	10.0	15.0	35.0	1.0	10.0	15.0	35.0
Maximum	9.0	14.0	34.0	77.0	8.0	14.0	34.0	74.0
Coefficient of variation	92.3	10.5	22.2	20.4	42.0	11.5	21.5	19.2

Lacerda et al., (2016), indicating that the spectroscopy model achieved results very close to those of routine laboratory analysis.

4.3.4. Comparisons between soil textural classes

The confusion matrix determined the number of correctly classified soil samples in the four textural classes using reference data and predicted clay values determined by soil spectroscopy (Table 11). The majority of samples were correctly classified (72.2, 28.1, 75.9, 85.2 and 70.0, 0.0, 75.2; Table 11) in both the training and assessment sets, as observed and reported by other authors, such as Bilgili et al. (2010); Dotto et al. (2016) and Franceschini et al. (2013). The accuracy assessment for the training set compared with the assessment set showed slightly higher classification performance. For the training set, class 3 had the highest agreement rate with reference data (85.3%), followed by classes 2, 0 and 1, with 76.0%, 72.2% and 28.1%, respectively (Table 11). The OA rate was 80.0%. For the assessment set, the classification followed the same trend for agreement rate in classes 3 (82.1%), 2 (75.2%), 0 (70.0%) and 1(0.0%) with OA=77.2% (Table 11). These findings are in line with Dotto et al. (2016), who used the VIS-NIR-SWIR to predict clay

content and grouped soil into three textural classes for insurance purposes in a watershed in SC state, Brazil. The best performance was found for soil class 3, followed by those of class 2, with poor performance for class 1. It was suggested that this sequence may be explained by the constituents of the clay fraction which have absorption regions characteristic of the visible and near infrared electromagnetic spectrum.

Table 11. Confusion matrix generated from the similarity between soil textural classes 0, 1, 2, and 3.

Predicted soil	Measured soil							
	Training set (n = 1,143)				Assessment set (n = 382)			
	Agreement (%)							
Classes	0	1	2	3	0	1	2	3
0	**72.2**	16.7	11.1	0.0	**70.0**	20.0	10.0	0.0
1	15.6	**28.1**	53.1	3.1	55.6	**0.0**	44.4	0.0
2	2.4	5.7	**75.9**	16.1	4.8	6.9	**75.2**	13.1
3	0.1	0.3	14.3	**85.2**	0.0	0.0	17.9	**82.1**
	OA = 80.0%				OA = 77.2%			

Some confusion was observed for class 1. Poor performance was found for both the training (28.1%) and assessment (0.0 %) sets, which were mainly mixed with class 2 (53.1% and 44.4%). This misclassification is detrimental, since class 1 is dominated by sand fractions and class 2 are dominated by medium texture, so, the behaviors of these two soils differ considerably and the management is very different. In addition, class 1 is only supported with restrictions by financial support due its limitations to maintain water for a certain period of time and vulnerability to the risk of crop losses. Class 1 also showed lower accuracy than class 0, according to agreement rate (Table 11). This poor performance was not expected, since the clay content is greater in class 1 than in class 0. A possible explanation would be that the prediction error of the clay content model ($RMSE_{train}$ = 7.75% and $RMSE_{val}$ = 8.69%) is larger than the clay content range of class 1 (≥ 10% and < 15%) but, on the other hand, lower than the clay content range of Class 0 (< 10%). Zeng et al. (2017), stated that when working with classification systems, soil

allocation accuracy depends not only on the prediction accuracy of the target property but the required limits (range) of the classes. This suggests that misclassification may be caused by arbitration in establishment of the small range of class 1 by MAPA regulations. For example, class 1 was overestimated as class 2 of 53.1% and 44.4% in training and assessment sets, respectively (Table 11). For insurance applications, this may benefit crop areas with restrictions for financial support, in which the behavior of soils differs considerably from class 2, increasing the agricultural risk.

Class 0 showed good agreement, with 72.2% and 70.0% for the training and assessment sets, respectively (Table 11). This result is significant because samples above this limit can be excluded from crop insurance, since crops established in soils with clay content less than 10% are not eligible for agricultural funding, in accordance with MAPA regulations. Considering the training and assessment sets, class 0 was confounded with class 1 and class 2 in less than 20% of the dataset (Table 11). The allocation mistake of class 0 in class 2 (11.1% and 10.0% in training and assessment sets, respectively) is a significant error if we consider that soils with these two different texture would have different management. The mainly misallocations are due to overestimated of class 0 and 1 and underestimated of class 1. These two classes are the less representative of the described classes (48 and 50 samples, respectively). They are very specific and occurred only in a restricted area of the Southern coast of the state.

Class 2 was mainly confounded with class 3 in the training (16%) and assessment (13%) sets. As mentioned by Vasques et al. (2014), clay content is an attribute that only indirectly influences the VIS-NIR-SWIR soil spectra, and thus small differences in clay content were probably not captured by the SVM model, causing misclassification between these two classes. Class 3 was discriminated from soils of class 0 and 1 in 100% of the classification in the assessment set. It was confounded with class 2 in less than 20% of both datasets. This finding is in line with Bilgili et al. (2010), who obtained good results using spectroscopy models for clay content estimation in 512 soil samples collected from a single field in Northern Turkey, reporting agreement between categorical classification by laboratory and VIS-NIR-SWIR estimations of more than 50% for clay class. Franceschini et al., 2013, achieved better

results between laboratory and VIS-NIR-SWIR estimation for classes with high clay content (AO = 69.8%), using a small dataset of 129 soil samples collected from different locations in the central-west region of Brazil, albeit with poor discrimination for classes with low clay content.

The VIS-NIR-SWIR spectroscopy used as a reference method for clay quantification was able to correctly classify classes 2 and 3, which together represent more than 90% (589 and 838 samples, respectively) of the whole dataset. Soils in class 2 predominate on the coast and in the extreme West regions of the state, while class 3 predominates on the Plateau and in the West of the state. These two soil textural classes are suitable soils for cultivation of most crops when considering the climatic risks associated with agricultural insurance purposes.

4.4. Conclusions

This study confirms the usefulness of soil analysis based on VIS-NIR-SWIR spectroscopy associated with chemometric techniques to predict clay content for agricultural insurance purposes, using a regional soil spectral library. Furthermore, this spectral library could be used advantageously in support of soil data in other areas of SC State. Spectroscopy predictions of clay content using the regional soil spectral library in a wide variability of soil texture reached R^2_{cal} = 0.79 with low $RMSE_{train}$ (7.75%) and R_{val}^2 = 0.77 with low $RMSE_{val}$ (8.69%). The prediction accuracy of clay content by spectra technique when compared with laboratory measurements lies within the required range.

Classification of the predicted clay content values into four defined classes achieved high levels of agreement. This led to clear allocation, with a high degree of accuracy for classes 0, 2 and 3, which together represent more than 95% of the whole dataset. However, the worst classification was observed for class 1. VIS-NIR-SWIR spectroscopy technique presented close relationship with traditional laboratory method, which could be recommended its use for soil attributes quantification especially when a large number of soil samples are analyzed. Results such as these, represent an additional step towards its application in Brazilian soil laboratory routines and indicate the importance of correctly allocating soils into textural classes to estimate the availability of soil water content for crops, considering the management of agricultural risk

promoted by insurance programs, such as *Proagro Mais*. The MAPA regulation is an important tool related to soil information in Brazilian agricultural insurance policies to mitigate the vulnerability of smallholder farmers to climate risk. Thus, further work needs to be done to include more soil samples of classes 0 and 1 (sandy soils) from different crops to make the regional soil spectral library more soil-representative and improve prediction accuracy and allocation of textural classes.

5. CONSIDERAÇÕES FINAIS

Essa tese foi elaborada na forma de artigos científicos que constituem os capítulos II e III. O objetivo geral dessa pesquisa foi aplicar e avaliar espectroscopia de reflectância (ER) na região do VIS-NIR-SWIR (350-2500 nm) utilizando diferentes modelos estatísticos multivariados para a predição dos teores de areia e argila no solo em escala regional utilizando amostras de solos legadas. Além disso foi avaliado o potencial dessa abordagem na alocação das amostras de solos em classes texturais distintas para fins de seguro agrícola.

No primeiro capítulo foi apresentada uma revisão de literatura sobre a aplicação da ER (VIS-NIR-SWIR e MIR) na ciência do solo. As bases de dados selecionadas foram WoS e Scopus no período de 1945 a 2018. Observou-se o aumento dramático no número de artigos publicados durante o período e a evolução tecnológica da espectroscopia (sensores de laboratório e campo) e da quimiometria (número de modelos empregados), bem como no número de atributos caracterizados e estimados com sucesso pela técnica. Além disso, foi possível verificar o nível de colaboração científica entre os principais autores das pesquisas e o quanto essa colaboração foi significativa para a evolução da ER no mundo. Verificou-se que os países com maior número de artigos publicados (EUA e China) são os países com os maiores investimentos em desenvolvimento científico e tecnológico através de programas permanentes que fomentam pesquisa de alta qualidade e sua divulgação em revistas com alto fator de impacto. Apesar desse cenário ser bem diferente no Brasil, onde tem havido cortes de verbas para CT&I de forma sistemática nos últimos anos, os pesquisadores brasileiros, através de

parcerias nacionais e internacionais, conseguiram realizar pesquisas na área da ER e publicar seus resultados para comunidade científica internacional, mas não se sabe até quando.

No segundo capítulo, foi apresentado um estudo que compara modelos baseados na ER para a predição dos teores de argila e areia em escala regional. Seis técnicas de pré-processamento dos espectros foram aplicadas para diminuir os ruídos e não linearidades entre os espectros e o atributo avaliado e cinco modelos estatísticos multivariados foram desenvolvidos e comparados. O efeito dos pré-processamentos aplicados aos espectros de reflectância variou em função do atributo predito e/ou do modelo multivariado usado. Em geral, os pré-processamentos baseados na correção da dispersão (MSC, NBR, DET e CRR) melhoraram o desempenho dos modelos de predição quando comparados com as técnicas de aplicação de derivadas (FDSG e SDSG). Na predição do conteúdo de areia, esse efeito não foi expressivo, visto que, os melhores modelos foram obtidos sem o pré-processamento (RAW) dos espectros. Considerando os valores médios de R^2 e RMSE dos seis pré-processamentos usados mais a reflectância (teste SK), a segunda derivada por SG (SDSG) foi o único que apresentou diferença estatística significativa. Entre os modelos multivariados usados, o Cubist apresentou o melhor desempenho na predição do conteúdo de areia e argila. Porém houve diferença estatística significativa entre o Cubist e os demais métodos usados apenas quando se avaliou os valores de R^2 médio dos modelos na predição do conteúdo de argila. Como sugestão para pesquisas futuras poderia se avaliar a retiradas das amostras com altos teores de areia da base de dados para aumentar a capacidade preditiva do teor de argila dos modelos, uma vez que o quartzo diminui as feições espectrais na composição do solo; avaliar o potencial da espectroscopia na predição do conteúdo de silte, visto que SC apresenta solos com altos teores dessa fração; e avaliar o potencial dos modelos gerados na predição dos teores de areia e argila em amostras desconhecidas e em pequena escala.

No terceiro capítulo, foi apresentado um estudo sobre a aplicabilidade do uso da abordagem ER na região do VIS-NIR-SWIR para alocação do conteúdo de argila predito das amostras de solos em classes texturais distintas, como referência aos programas de seguro agrícola

brasileiros. As classes texturais definidas para fins de enquadramento dos solos nos programas de seguro agrícola foram discriminadas a partir dos valores de argila preditos pela ER (SVMR-NBR), apresentando taxa de concordância superior a 70% para três das quatros classes avaliadas. A ER foi capaz de discriminar as classes texturais dos solos, mesmo em um conjunto heterogêneo de amostras, mostrando o potencial de uso da técnica nos laboratórios de rotina, como método complementar ou alternativo, principalmente quando um grande número de amostras é necessário. Para pesquisas futuras sugere-se trabalhar com outras faixas de classificação da textura do solo além daquelas previstas pelo MAPA; utilizar a região do MIR além do VIS-NIR-SWIR para leitura do espectros já que o quartzo apresenta resposta ativa no MIR e aumentar o número de amostras de solo com teores de areia entre 10 e 15% e argila acima de 80% na base de dados para aumentar a representatividade e a capacidade de predição dos modelos.

Os resultados alcançados com os estudos confirmaram que a espectroscopia de reflectância pode ser aplicada na prática e em alguns países como, a Austrália e o Continente Africano, ela está na fase operacional. Disso surge o questionamento sobre o que falta para que essa abordagem seja implementada nos laboratórios de rotina no Brasil. Todos esses questionamentos surgiram durante o doutorado e referem-se as indagações sobre as pesquisas futuras a serem desenvolvidas na Epagri e sobre a modernização e o avanço da ciência do solo em Santa Catarina. A partir disso, seguem algumas afirmações: (i) a espectroscopia de reflectância é uma técnica que apresenta acurácia similar ou superior aos métodos tradicionais de análise de rotina; (ii) a biblioteca espectral de solo precisa ser ampliada, adicionando amostras de solos e resultados analíticos (legados ou não) de outros atributos, bem como ampliando o número de parceiros para aumentar a representatividade dos solos em escala local e regional; (iii) a espectroscopia de reflectância deve ser incluída no protocolo do MAPA como método de enquadramento das classes texturais para fins de seguro agrícola; (iv) a rede de colaboração técnico-científica entre os pesquisadores precisa ser incentivada pelas instituições para que o processo de inovação da ciência aconteça, diante da escassez de recursos, e gere benefícios para a agricultura catarinense; e (v) em um mundo globalizado, a formação de um pesquisador

depende muito da rede (network) de colaboradores que ele formou ou ampliou, uma vez que, além possibilitar maior crescimento de suas atividades científicas, atualmente, a rede de colaboração já vem sendo um dos indicadores usados pelas agências de fomento para o financiamento e planejamento em C&TI.

Para desenvolver os estudos optamos por utilizar um conjunto já existente de amostras físicas de solos e seus respectivos resultados analíticos. As 1.534 amostras de solos foram coletadas por técnicos da Assistência Técnica e Extensão Rural (ATER) em áreas de lavoura na profundidade de 0-50 cm, entre abril e setembro de 2009 em propriedades agrícolas familiares do Estado de Santa Catarina e analisadas no Laboratório de Física do Solo da Epagri-Campos Novos sob a coordenação do pesquisador Dr. Milton da Veiga. Esses dados são denominados de dados "legados" e no contexto da pedometria e do mapeamento digital de solos apresentam grande potencial para gerar novas informações. No sensoriamento próximo os dados legados assumiram papel fundamental na criação das bibliotecas espectrais de solos (Austrália, Europa, Estados Unidos da América, Brasil e Mundo - *Soil Spectroscopy Group*) em diversas escalas. Nesse contexto, surgiram algumas questões, tais como, quanta informação sobre o solo existe hoje no Brasil e quanto dessa informação está disponível, qual o potencial desses dados na geração de novas informações, e, finalmente, qual o nível de colaboração entre as instituições e pesquisadores. No Brasil, a escassez de informações sobre o solo em escala compatível com a demanda de uso da terra é histórica e ocorre em todos os estados brasileiros. Os dados de solos legados são essenciais para o avanço das pesquisas em espectroscopia e as universidades (vide o projeto da Biblioteca Espectral de Solos do Brasil - [http://bibliotecaespectral.wix.com/esalq]) e centros de pesquisa públicos assumem papel fundamental para que esse avanço ocorra. Entretanto, é necessário que haja um esforço entre essas instituições para que esses dados (mapas de solos no formato analógico e digital, descrição de perfis, dados analíticos e relatórios) sejam capturados, compilados, harmonizados e disponibilizados em uma base única para que possamos conhecer e explorar o seu potencial para então definir quais são as escalas e as informações de solos (atributos ou classes) que precisam ser levantadas para orientar a tomada de

decisão nas diferentes áreas (agronomia, engenharias, hidrologia e hidrogeologia, avaliação e monitoramento ambiental, entre outras) nas diversas regiões do Brasil. A escassez de pedólogos e/ou cientistas do solo especializados em novas tecnologias (mapeamento digital de solos, espectroscopia, sistema de informação geográfica, etc.) também configura um dos gargalos para a viabilização de levantamentos detalhados de solos no Brasil em grande escala. Além disso, a pedologia no Brasil se configurou, dada a escassez de projetos nacionais, em grupos regionalizados e especializados em levantamentos utilitários de solos, direcionados para as necessidades de cada região na qual o grupo se insere. Todas essas constatações foram levantadas pelo Tribunal de Contas da União (TCU) na conferência sobre a Governança de Solos, realizada em março de 2015 em Brasília e que resultou na criação do Programa Nacional de Levantamento e Interpretação de Solos do Brasil (PronaSolos), instituído esse ano pelo Decreto Nº. 9.414, que prevê a execução de levantamentos de solos em escalas geográficas iguais ou mais detalhadas que 1:100.000. O inciso V do Art. 2º prevê a implementação das inovações em levantamento de solos e temas correlatos. A evolução tecnológica associada ao sensoriamento remoto (espectroscopia de reflectância) e as técnicas de mapeamento digital de solos têm modernizado a pedologia e correspondem as inovações de que trata o referido inciso.

Nessa tese optou-se pelo uso do ambiente R (https://www.r-project.org/). Embora existam softwares comerciais (Unscrambler®, MATLAB®, etc) que podem ser usados para análise quimiométrica com base na espectroscopia de reflectância, atualmente há na Epagri, uma política de incentivo de uso de plataformas computacionais livres, sendo a Epagri a instituição onde trabalho e onde desenvolverei minhas atividades futuras de pesquisa. O R é uma linguagem de programação de código aberto, gratuita, intuitiva, interativa e popular na ciência de análise de dados, modelagem estatística e exibição gráfica. O processo de aprendizagem de programação no R exigiu esforço e dedicação, no entanto, há um grande número de usuários do R (maior comunidade de usuários *on-line* no mundo) desenvolvendo pacotes e códigos específicos, em diversas áreas do conhecimento, o que tornou o processo de aprendizagem mais rápido.

6.REFERÊNCIAS

ACKERSON, J. P.; DEMATTÊ, J. A. M. M.; MORGAN, C. L. S. S. Predicting clay content on field-moist intact tropical soils using a dried, ground VisNIR library with external parameter orthogonalization. **Geoderma**, Amsterdam, v. 259–260, p. 196–204, 2015.

ADAMCHUK, V. I. et al. On-the-go soil sensors for precision agriculture. **Computers and Electronics in Agriculture**, New York, v. 44, p. 71–91, 2004.

ALMEIDA, J. A.; TORRENT, J.; BARRÓN, V. Cor de solo, formas do fósforo e adsorção de fosfatos em Latossolos desenvolvidos de basalto do extremo-sul do Brasil. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 27, n. 2, p. 985–1002, 2003.

ARAÚJO, S. R. et al. Improving the prediction performance of a large tropical vis-NIR spectroscopic soil library from Brazil by clustering into smaller subsets or use of data mining calibration techniques. **European Journal of Soil Science**, Oxford, v. 65, p. 718–729, 2014. Disponível em: <http://doi.wiley.com/10.1111/ejss.12165>. Acesso em: 16 mar. 2014.

ARAÚJO, S. R. et al. Determining soil properties in Amazonian Dark Earths by reflectance spectroscopy. **Geoderma**, Amsterdam, v. 237–238, p. 308–317, 2015. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0016706114003462>. Acesso em: 25 ago. 2015.

ARIA, M.; CUCCURULLO, C. Bibliometrix : an R-tool for comprehensive science mapping analysis. **Journal of Informetrics**, Amsterdam, v. 11, n. 4, p. 959–975, 2017.

ASSAD, E.D. et al. Zoneamento agrícola de risco climáticos do Brasil: base teórica, pesquisa e desenvolvimento. **Informe Agropecuário**, Belo Horizonte, v. 29, n. 246, p. 47-60, 2008.

BARNES, R. J.; DHANOA, M. S.; LISTER, S. J. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra.

Applied Spectroscopy, Baltimore, v. 43, n. 5, p. 772–777, 1989.

BAXTER, S. J.; CRAWFORD, D. M. Incorporating legacy soil pH databases into digital soil mapping. In: HARTEMINK, A. E.; MCBRATNEY, A.; MENDONÇA-SANTOS, M. L. (Ed.). **Digital soil mapping with limited data**. Wageningen: Springer Science+Business Media, 2008. p. 311–318.

BELLINASO, H.; DEMATTÊ, J. A. M.; ROMEIRO, S. A. Soil spectral library and its use in soil classification. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 34, n. 1, 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. **TrAC - Trends in Analytical Chemistry**, Amsterdam, v. 29, n. 9, p. 1073–1081, 2010. Disponível em: <http://dx.doi.org/10.1016/j.trac.2010.05.006>. Acesso em: 27 ago. 2015.

BELTER, C. W.; SEIDEL, D. J. A bibliometric analysis of climate engineering research. **WIREs Climate Change**, Hoboken, v. 4, n. 5, p. 417–427, 2013.

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

BENDOR, E.; IRONS, J. R.; EPEMA, G. F. Soil Reflectance. In: RENCZ, A.N.; RYERSON, R.A. **Manual of remote sensing for the earth sciences**. New York: Wiley, 1999. v.3, p. 111–188.

BENDOR, E.; ONG, C.; LAU, I. C. I. C. Reflectance measurements of soils in the laboratory: standards and protocols. **Geoderma**, Amsterdam, v. 245–246, p. 112–124, 2015. Disponível em: <http://dx.doi.org/10.1016/j.geoderma.2015.01.002>. Acesso em: 24 abr. 2015.

BILGILI, V. A. et al. Visible-near infrared reflectance spectroscopy for assessment of soil properties in a semi-arid area of Turkey. **Journal of Arid Environments**, London, v. 74, n. 2, p. 229–238, 2010. Disponível em: <http://dx.doi.org/10.1016/j.jaridenv.2009.08.011>. Acesso em: 22 mar. 2016.

BLANCO-MOURE, N. et al. Soil organic matter fractions as affected by tillage and soil texture under semiarid Mediterranean conditions. **Soil and Tillage Research**, Amsterdam, v. 155, p. 381–389, 2016.

BOCKHEIM, J. G. et al. Soil-forming factors and soil taxonomy. **Geoderma**, Amsterdam, v. 226–227, p. 231–237, 2014. Disponível em: <http://dx.doi.org/10.1016/j.geoderma.2014.02.016>. Acesso em: 07 maio 2015.

BOOS, D.D. Introduction to the bootstrao world. **Statistical Science**, Hayward, v. 18, n. 2, p. 168-174, 2003. Disponível em: <http://projecteuclid.org/euclid.ss/1063994971>. Acesso em: 22 abr. 2018.

BOUMA, J.; MCBRATNEY, A. Framing soils as an actor when dealing with wicked environmental problems. **Geoderma**, Amsterdam, v. 201, p. 130–139, 2013.

BRADY, N. C.; WEIL, R. R. **The nature and properties of soils**. 14th ed. Upper Saddle River: Prentice Hall, 2007.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Secretaria Executiva da Comissão Especial de Recursos. Instrução Normativa n.2, de 9 de outubro de 2008. **Diário Oficial [da] República Federativa da União**, Brasília, DF, 13 out. 2008. Seção 1, p. 71.

BREIMAN, L. Bagging predictors. **Machine Learning**, Boston, v. 24, n. 2, p. 123–140, 1994. Disponível em: <http://link.springer.com/10.1007/BF00058655>. Acesso em: 19 mar. 2017.

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

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

BRYANT, L.; O'CONNOR, C. Creating incentives to improve soil health trough the federal crop insurance program. In: FIELD, D. J.; MORGAN, C. L. S.; MCBRATNEY, A. B. (Ed.). **Global soil security**. New York: Springer, 2017.

BUDDENBAUM, H.; STEFFENS, M. The effects of spectral pretreatments on chemometric analyses of soil profiles using laboratory imaging spectroscopy. **Applied and Environmental Soil Science**, New York, v. 2012, art.[274903], 2012.

CAMBULE, A. H. et al. Building a near infrared spectral library for soil organic carbon estimation in the Limpopo National Park, Mozambique. **Geoderma**, Amsterdam, v. 183–184, p. 41–48, 2012. Disponível em: <http://dx.doi.org/10.1016/j.geoderma.2012.03.011>. Acesso em: 08 mar. 2016.

CANTARELLA, H. et al. Variability of soil analysis in commercial laboratories: Implications for lime and fertilizer recommendations. **Communications in Soil Science and Plant Analysis**, New York, v. 37, p. 2213–2225, 2006.

CARMO, D. A. B. et al. Cor do solo na identificação de áreas com diferentes potenciais produtivos e qualidade de café. **Pesquisa Agropecuária Brasileira**, Brasília, v. 51, n. 9, p. 1261–1271, 2016.

CEZAR, E. et al. Avaliação e quantificação das frações silte, areia e argila por meio de suas respectivas relfectâncias. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 36, n. 1, p. 1157–1165, 2012.

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

CLARK, R. N. Spectroscopy of rocks and minerals, and principles of spectroscopy. In: RENCZ, A. (Ed.). **Remote sensing for the earth sciences:** manual of remote sensing. 3. ed. New York: John Wiley, 1999. v. 3, p. 3–58.

CLARK, R. N.; ROUSH, T. L. Reflectance spectroscopy: quantitative analysis techniques for remote sensing applications. **Journal of Geophysical Research: Solid Earth**, Hoboken, v. 89, n. B7, p. 6329–6340, 1984. Disponível em: <http://doi.wiley.com/10.1029/JB089iB07p06329>. Acesso em: 22 abr. 2016.

COELHO, M. R.; ROSSI, M.; MATTOS, I. F. A. Solos da Mata Atlântica. In: CURI, N.; KER, J. C.; VIDAL-TORRADO, P. (Ed.). **Pedologia:** solos dos biomas brasileiros. Viçosa, MG: Sociedade Brasileira de Ciência do Solo, 2017. p. 597.

COHEN, J. A Coefficient of agreement for nominal scales. **Educational and Psychological Measurement**, Durham, v. 20, n. 1, p. 37–46, 1960. Disponível em: <http://journals.sagepub.com/doi/10.1177/001316446002000104>. Acesso em: 20 nov. 2017.

CONDIT, H. R. The spectral reflectance of American soils. **Photogrammetric Engineering**, Falls Church, v. 36, p. 955–966, 1970. Disponível em: <http://trid.trb.org/view.aspx?id=93590>. Acesso em: 19 mar. 2017.

CONFORTI, M. et al. Laboratory-based Vis–NIR spectroscopy and partial least square regression with spatially correlated errors for predicting spatial variation of soil organic matter content. **Catena**, Cremlingen, v. 124, p. 60–67, 2014. Disponível em: <http://dx.doi.org/10.1016/j.catena.2014.09.004>. Acesso em: 18 out. 2017.

CONFORTI, M.; MATTEUCCI, G.; BUTTAFUOCO, G. Using laboratory Vis-NIR spectroscopy for monitoring some forest soil properties. **Journal of Soils and Sediments**, Landsberg, v. 18, n.3, p. 1009-1019, 2018. Disponível em: <https://link.springer.com/article/10.1007/s11368-017-1766-5>. Acesso em: 30 out. 2017.

CORREA, M. M. et al. Caracterização de óxidos de ferro de solos do ambiente tabulerios costeiros. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 32, n. 1, p. 1017–1031, 2008.

CUCCURULLO, C.; ARIA, M.; SARTO, F. Foundations and trends in performance management and public administration domains. **Scientometrics**, Amsterdam, v. 108, p. 595–611, 2016.

CUTLER, A.; CUTLER, R. D.; STEVENS, J. R. Random Forest. In: ZHNAG, C.; YUNQIAN, M. (Ed.). **Ensemble machine learning**. New York: Springer Science+Business Media, 2012. p. 329.

DANTAS, J. S. et al. Coeficiente de erodibilidade em sulcos e entressulcos de Argissolos coesos estimado pela cor do solo coefficient of rill and interrill

erodibility of cohesive. **Pesquisa Agropecuária Brasileira**, Brasília, v. 49, n. 9, p. 700–707, 2014.

DE HAAN, J. Authorship patterns in Dutch sociology. **Scientometrics**, Amsterdam, v. 39, n. 2, p. 197–208, 1997.

DEISS, L.; FRANZLUEBBERS, A. J.; DE MORAES, A. Soil texture and organic carbon fractions predicted from near-infrared spectroscopy and geostatistics. **Soil Science Society of America Journal**, Madison, v. 81, n. 5, p. 1222–1234, 2017. Disponível em: <https://dl.sciencesocieties.org/publications/sssaj/abstracts/0/0/sssaj2016.10.0326>. Acesso em: 16 mar. 2017.

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

DEMATTÊ, J. A. M. Characterization and discrimination of soils by their reflected electromagnetic energy. **Pesquisa Agropecuaria Brasileira**, Brasília, v. 37, n. 10, p. 1445–1458, 2002.

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

DEMATTÊ, J. A. M. et al. Is it possible to classify topsoil texture using a sensor located 800 km away from the surface? **Revista Brasileira de Ciencia do Solo**, Viçosa, v. 40, p. 1–13, 2016a.

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

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

DEMATTÊ, J. A. M.; FIORIO, P. R.; ARAÚJO, S. R. Variation of routine soil analysis when compared with hyperspectral narrow band sensing method. **Remote Sensing**, Basel, v. 2, n. 8, p. 1998–2016, 2010.

DEMATTÊ, J. A. M.; TERRA, F. S. Spectral pedology: a new perspective on evaluation of soils along pedogenetic alterations. **Geoderma**, Amsterdam, v. 217–218, p. 190–200, 2014. Disponível em: <http://dx.doi.org/10.1016/j.geoderma.2013.11.012>. Acesso em: 24 abr. 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**, Viçosa, v. 38, p. 1663–1671, 2014. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-06832014000600001&lng=en&nrm=iso&tlng=pt>. Acesso em: 3 fev. 2015.

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

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

DOTTO, A. C. et al. A systematic study on the application of scatter-corrective and spectral- derivative preprocessing for multivariate prediction of soil organic carbon by Vis-NIR spectra. **Geoderma**, Amsterdam, v. 314, p. 262–274, 2018. Disponível em: <https://doi.org/10.1016/j.geoderma.2017.11.006>. Acesso em: 16 mar. 2018.

DUDA, B. M. et al. Soil characterization across catenas via advanced proximal sensors. **Geoderma**, Amsterdam, v. 298, p. 78–91, 2017.

DUDEK, M. W. **ClusterSim:** searching for optimal clustering procedure for a data set. 2017. Disponível em: <http://keii.ue.wroc.pl/clusterSim>. Acesso em: 01 fev. 2017.

EISELE, A. et al. Applicability of the thermal infrared spectral region for the prediction of soil properties across semi-arid agricultural landscapes. **Remote Sensing**, Basel, v. 4, n. 11, p. 3265–3286, 2012.

ENGEL, J. et al. Breaking with trends in pre-processing? **TrAC - Trends in Analytical Chemistry**, Amsterdam, v. 50, p. 96–106, 2013.

EMBRAPA. Centro Nacional de Pesquisa de Solos. **Solos do Estado de Santa Catarina**. Rio de Janeiro: Embrapa/CNPS, 2004. (Boletim de Pesquisa e Desenvolvimento, 46).

FERNANDES, K. L. et al. Magnetic and spectral signatures of cerrado soils in the state of Goiás, Brazil. **Pesquisa Agropecuária Brasileira**, Brasília, v. 52, n. 10, p. 923–932, 2017.

FERNANDES, R. B. A. et al. Quantificação de óxidos de ferro de Latossolos brasileiros por espectroscopia de relfectância difusa. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 28, n. 1, p. 245–257, 2004.

FERRARESI, T. et al. Espectroscopia de infravermelho na determinação da textura do solo. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 36, p. 1769–1777, 2012.

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

FRIEDMAN, J. H. Multivariate adaptative regression splines. **The Annals of Statistics**, Cleveland, v. 19, n. 1, p. 1–141, 1991.

FRIEDMAN, J. H. Stochastic gradient boosting. **Computational Statistics & Data Analysis**, Amsterdam, v. 38, p. 367–378, 2002. Disponível em: <http://onlinelibrary.wiley.com/doi/10.1002/cbdv.200490137/abstract%5Cnhttp:// www.sciencedirect.com/science/article/pii/S0167947301000652>. Acesso em: 19 mar. 2018.

GALLANT, S. I. **Neural network learning and expert systems**. Massachusetts: The MIT Press, 1993.

GARCÍA-SÁNCHEZ, F. et al. Using Near-Infrared Spectroscopy in Agricultural Systems. In: García-Sánchez, F. et al. Developments in near-infrared spectroscopy. London: Intechopen, 2017. cap. 5, p. 97-127. Disponível em: <http://www.intechopen.com/books/developments-in-near-infrared-spectroscopy/using-near-infrared-spectroscopy-in-agricultural-systems>. Acesso em: 30 out. 2017.

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

GLANZEL, W.; SCHUBERT, A. Analysing scientific networks through co-authorship. In: MOED, H. F.; GLÄNZEL, W.; SCHMOCH, U. (Ed.). **Handbook of quantitative science and technology research**. Dordrecht: Kluwer Academic, 2005. p. 257–276.

GOGÉ, F. et al. Optimization criteria in sample selection step of local regression for quantitative analysis of large soil NIRS database. **Chemometrics and Intelligent Laboratory Systems**, Amsterdam, v. 110, n. 1, p. 168–176, 2012.

GOMEZ, C. et al. Using legacy data for correction of soil surface clay content predicted from VNIR / SWIR hyperspectral airborne images. **Geoderma**, Amsterdam, v. 276, p. 84–92, 2016.

HARTEMINK, A. E.; MCBRATNEY, A. A soil science renaissance. **Geoderma**, Amsterdam, v. 148, n. 2, p. 123–129, 2008. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0016706108002802>. Acesso em: 26 out. 2014.

HARTEMINK, A. E.; MINASNY, B. Towards digital soil morphometrics. **Geoderma**, Amsterdam, v. 230–231, p. 305–317, 2014. Disponível em: <http://dx.doi.org/10.1016/j.geoderma.2014.03.008>. Acesso em: 14 out. 2014.

HEMPEL, J. W. et al. GlobalSoilMap project history. In: ARROUAYS, D. et al. (Ed.). **GlobalSoilMap:** basis of the global spatial soil information system. Leiden: CRC Press/Balkema, 2014. p. 478.

HOLMES, G.; HALL, M.; PRANK, E. Generating rule sets from model trees. In: AUSTRALIAN JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, 12., 1999, Sydney. **Proceedings …** Berlin: Springer, 1999. p. 1–12. Disponível em: <https://link.springer.com/content/pdf/10.1007%2F3-540-46695-9.pdf>. Acesso

em: 14 dez. 2015.

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

INDA, A. V. et al. Iron oxides dynamics in a subtropical Brazilian Paleudult under long-term no-tillage management. **Scientia Agricola**, Piracicaba, v. 70, n. p. 48–54, 2013.

IUSS WORKING GROUP WRB. **International soil classification system for naming soils and creating legends for soil maps**. Rome, 2014.

JAMES, G. et al. **An introduction to statistical learning with applications in R**. New York: Springer Science+Business Media, 2013.

JELIHOVSCHI, E.; FARIA, J. C.; ALLAMAN, I. B. ScottKnott: a package for performing the Scott-Knott clustering algorithm in R. **TEMA (São Carlos)**, São Carlos, v. 15, n. 1, p. 3-17, 2014. Disponível em: <http://tema.sbmac.org.br/tema/article/view/646>. Acesso em: 23 out. 2017.

JENSEN, D. K. et al. A new two-stage approach to predicting the soil water characteristic from saturation to oven-dryness. **Journal of Hydrology**, Amsterdam, v. 521, p. 498–507, 2015.

JOHANNES, A. et al. Optimal organic carbon values for soil structure quality of arable soils. Does clay content matter? **Geoderma**, Amsterdam, v. 302, p. 14-21, 2017.

KARATZOGLOU, A. et al. Kernlab: an S4 package for Kernel methods in R. **Journal of Statistical Software**, Los Angeles, v. 11, n. 9, p. 1–20, 2004. Disponível em: <http://statistik.wu-wien.ac.at/%5Cnhttp://www.jstatsoft.org/v11/i09/paper>. Acesso em: 08 mar. 2018.

KARUP, D. et al. Water and solute transport in agricultural soils predicted by volumetric clay and silt contents. **Journal of Contaminant Hydrology**, Amsterdam, v. 192, p. 194–202, 2016.

KELLER, T.; HÅKANSSON, I. Estimation of reference bulk density from soil particle size distribution and soil organic matter content. **Geoderma**, Amsterdam, v. 154, n. 3-4, p. 398–406, 2010.

KLEIN, R. M. **Mapa fitogeográfico do Estado de Santa Catarina**. Itajaí: Flora Ilustrada Catarinense, 1978.

KOVAČEVIĆ, M.; BAJAT, B.; GAJIĆ, B. Soil type classification and estimation of soil properties using support vector machines. **Geoderma**, Amsterdam, v. 154, n. 3–4, p. 340–347, 2010.

KUANG, B. et al. Sensing soil properties in the laboratory, in situ, and on-line. A

review. In: SPARKS, D. (Ed.). **Advances in Agronomy**. Amsterdam: Elsevier, 2012. v. 114, p. 155–223.

KUANG, B.; MOUAZEN, A. M. Calibration of visible and near infrared spectroscopy for soil analysis at the field scale on three European farms. **European Journal of Soil Science**, Oxford, v. 62, n. 4, p. 629–636, 2011.

KUANG, B.; MOUAZEN, A. M. Non-biased prediction of soil organic carbon and total nitrogen with vis-NIR spectroscopy, as affected by soil moisture content and texture. **Biosystems Engineering**, London, v. 114, n. 3, p. 249–258, 2013. Disponível em: <http://dx.doi.org/10.1016/j.biosystemseng.2013.01.005>. Acesso em: 27 ago. 2015.

KUHN, M. A Short introduction to the caret package. **R Foundation for Statistical Computing**, [s. l.], p. 1–10, 2015. Disponível em: < https://cran.r-project.org/web/packages/caret/vignettes/caret.html >. Acesso em: 03 abr. 2018.

LACERDA, M. et al. Tropical texture determination by proximal sensing using a regional spectral library and its relationship with soil classification. **Remote Sensing**, Basel, v. 8, n. 9, 2016.

LAGACHERIE, P. et al. Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements. **Remote Sensing of Environment**, New York, v. 112, n. 3, p. 825–835, 2008.

LAL, R. Managing soils and ecosystems for mitigating anthropogenic carbon emissions and advancing global food security. **BioScience**, Oxford, v. 60, n. 9, p. 708–721, 2010.

LEVENE, H. Robust tests for equality of variances. In: OLKIN, I. et al. (Ed.). **Contributions to probability and statistics:** essays in honor of Harold Hotelling. Palo Alto: Stanford University Press, 1960. p. 278–292.

LEVINE, N. M. et al. Ecosystem heterogeneity determines the ecological resilience of the Amazon to climate change. **Proceedings of the National Academy of Sciences**, Washington, v. 113, n. 3, p. 793–797, 2016. Disponível em: <http://www.pnas.org/lookup/doi/10.1073/pnas.1511344112>. Acesso em: 19 abr. 2018.

LOYOLA, P.; MOREIRA, V. R.; DA VEIGA, C. P. Analysis of the brazilian program of subsidies for rural insurance premium: evolution from 2005 to 2014. **Modern Applied Science**, Toronto, v. 10, n. 7, p. 87, 2016. Disponível em: <http://www.ccsenet.org/journal/index.php/mas/article/view/55796>. Acesso em: 18 out. 2017.

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

MAACK, R. Breves Notícias Sobre a Geologia dos Estados do Paraná e Santa

Catarina. **Brazilian Archives of Biology and Technology**, Curitiba, v. 1946-2001, p. 169–288, 2001.

MARTENS, H.; NAES, T. Multivariate Calibration. In: KOWALSKI, B.R. (Ed.). **Chemometrics:** mathematics and statistics in chemistry. Dordrecht: Springer, 1992. p. 419.

MARTÍN, M. Á.; REYES, M.; TAGUAS, F. J. Estimating soil bulk density with information metrics of soil texture. **Geoderma**, Amsterdam, v. 287, p. 66–70, 2017.

MCBRATNEY, A. B.; MENDONÇA-SANTOS, M. L.; MINASNY, B. On digital soil mapping. **Geoderma**, Amsterdam, v. 117, n. 1–2, p. 3–52, 2003. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0016706103002234>. Acesso em: 17 jul. 2014.

MCBRATNEY, A. B.; MINASNY, B.; VISCARRA ROSSEL, R. Spectral soil analysis and inference systems: A powerful combination for solving the soil data crisis. **Geoderma**, Amsterdam, v. 136, n. 1–2, p. 272–278, 2006. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0016706106001066>. Acesso em: 31 out. 2014.

MCCARTY, G. W. et al. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. **Soil Science Society of America Journal**, Madison, v. 66, n. 2, p. 640-649, 2002.

MENESES, P. R.; MADEIRA NETO, J. S. M. **Sensoriamento remoto:** reflectância dos alvos naturais. Brasília: UnB; Embrapa Cerrados, 2001.

MEVIK, B.-H.; WEHRENS, R.; LILAND, K. H. **PLS:** partial least squares and principal component regression. 2016. Disponível em: <https://cran.r-project.org/package=pls>. Acesso em: 14 jun. 2016.

MEYER, D. et al. **e1071**: Misc functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. 2017. Disponível em: <https://cran.r-project.org/package=e1071>. Acesso em: 22 abr. 2016.

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

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

NAWAR, S. et al. Estimating the soil clay content and organic matter by means of different calibration methods of vis-NIR diffuse reflectance spectroscopy. **Soil and Tillage Research**, Amsterdam, v. 155, p. 510–522, 2016. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0167198715001786>. Acesso em: 31 jan. 2017.

NAWAR, S.; MOUAZEN, A. M. Predictive performance of mobile vis-near infrared spectroscopy for key soil properties at different geographical scales by using spiking and data mining techniques. **Catena**, Cremlingen, v. 151, p. 118–129, 2017a.

NAWAR, S.; MOUAZEN, A. M. Comparison between random forests, artificial neural networks and gradient boosted machines methods of on-line Vis-NIR spectroscopy measurements of soil total nitrogen and total carbon. **Sensors**, Basel, v. 17, n. 10, p. 1–22, 2017b.

NOCITA, M. et al. Prediction of soil organic carbon content by diffuse reflectance spectroscopy using a local partial least square regression approach. **Soil Biology and Biochemistry**, Elmsford, v. 68, p. 337–347, 2014a. Disponível em: <http://dx.doi.org/10.1016/j.soilbio.2013.10.022>. Acesso em: 13 maio 2015.

NOCITA, M. et al. Soil spectroscopy: an opportunity to be seized. **Global Change Biology**, Oxford, v. 21, p. 10–11, 2014b.

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

NOURI, M. et al. Clay content mapping from airborne hyperspectral Vis-NIR data by transferring a laboratory regression model. **Geoderma**, Amsterdam, v. 298, p. 54-66, 2017.

PANDOLFO, C. et al. **Atlas climatológico do Estado de Santa Catarina**. Florianópolis: Epagri, 2002.

PELUCO, R. G. et al. Mapeamento do fósforo adsorvido por meio da cor e da suscetibilidade magnética do solo. **Pesquisa Agropecuária Brasileira**, Brasília, v. 50, n. 3, p. 259–266, 2015.

PENNINGTON, V. E. et al. Climate and soil texture influence patterns of forb species richness and composition in big sagebrush plant communities across their spatial extent in the western U.S. **Plant Ecology**, Dordrecht, v. 218, n. 8, p. 957–970, 2017.

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**, Basel, v. 9, n. 4, art.[293], 2017. Disponível em: <http://www.mdpi.com/2072-4292/9/4/293>. Acesso em: 16 mar. 2018.

POEPLAU, C.; ERIKSSON, J.; KÄTTERER, T. Estimating residual water content in air-dried soil from organic carbon and clay content. **Soil and Tillage Research**, Amsterdam, v. 145, p. 181–183, 2015.

POGGIO, M.; BROWN, D. J.; BRICKLEMYER, R. S. Comparison of Vis – NIR on in situ, intact core and dried, sieved soil to estimate clay content at field to regional scales. **European Journal of Soil Science**, Oxford, v. 68, n. 4, p.

434–448, 2017.

POPPIEL, R. R. et al. Surface spectroscopy of oxisols , entisols and inceptisol and relationships with selected soil properties. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 42, p. 1–26, 2018.

R CORE TEAM. **R:** a language and environment for statistical computing. R Foundation for Statistical Computing. 2017. Disponível em: <https://www.r-project.org/>. Acesso em: 19 mar. 2017.

RAMIREZ-LOPES, L. et al. The spectrum-based learner: a new local approach for modeling soil vis-NIR spectra of complex datasets. **Geoderma**, Amsterdam, v. 195–196, p. 268–279, 2013. Disponível em: <http://dx.doi.org/10.1016/j.geoderma.2012.12.014>. Acesso em: 24 abr. 2015.

RASMUSSEN, C. E.; WILLIAMS, C. K. I. Regression. In: RASMUSSEN, C. E.; WILLIAMS, C. K. I. (Ed.). **Gaussian processes for machine learning**. Massachusetts: MIT Press, 2006.

RESENDE, J. M. A. et al. Variabilidade espacial de atributos de solos coesos do leste Maranhense. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 38, p. 1077–1090, 2014.

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

RIZZO, R. et al. Digital soil mapping at local scale using a multi-depth Vis–NIR spectral library and terrain attributes. **Geoderma**, Amsterdam, v. 274, p. 18-27, 2016.

RIZZO, R.; DEMATTÊ, J. A. M.; TERRA, F. S. Using numerical classification of profiles based on Vis-NIR spectra to distinguish soils from the Piracicaba Region, Brazil. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 38, n. 2, p. 372–385, 2014. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-06832014000200002&lng=en&nrm=iso&tlng=en>. Acesso em: 3 fev. 2015.

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

SANCHEZ, P. A. et al. Digital soil map of the world. **Science**, Washington, v. 325, n. 5941, p. 680–681, 2009.

SANTA CATARINA. Secretaria de Ciência e Tecnologia, Minas e Energia. **Mapa geológico do Estado de Santa Catarina escala 1:500.000**: texto explicativo e mapa. Florianópolis: CRM, 1987.

SANTANA, F. B.; 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, Amsterdam, v. 191, p. 454–462, 2018.

SATO, M. V. **Primeira aproximação da biblioteca espectral de solos do Brasil:** caracterização de espectros de solos e quantificação de atributos. 2015. 102 f. Dissertação (Mestrado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2015.

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

SAWUT, M. et al. Estimating soil sand content using thermal infrared spectra in arid lands. **International Journal of Applied Earth Observation and Geoinformation**, Washington, v. 33, p. 203–210, 2014. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0303243414001305>. Acesso em: 19 fev. 2016.

SCHEINOST, A. C. et al. Use and limitations of second-derivative diffuse reflectance spectroscopy in the visible to near-infrared range to identify and quantify Fe oxide minerals in soils. **Clays and Clay Minerals**, Oxford, v. 46, n. 5, p. 528–536, 1998.

SCHJØNNING, P. et al. Predicting soil particle density from clay and soil organic matter contents. **Geoderma**, Amsterdam, v. 286, p. 83–87, 2017. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S001670611630636X>. Acesso em: 17 abr. 2018.

SCOTT, A. A. J.; KNOTT, M. A Cluster analysis method for grouping means in the analysis of variance. **Biometrics**, Washington, v. 30, n. 3, p. 507–512, 1974. Disponível em: <http://www.jstor.org/stable/2529204>. Acesso em: 18 dez. 2017.

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

SHI, T. et al. Visible and near-infrared reflectance spectroscopy - an alternative for monitoring soil contamination by heavy metals: a review. **Journal of Hazardous Materials**, Amsterdam, v. 265, p. 166–176, 2014. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S030438941300914X>. Acesso em: 17 mar. 2016.

SOLLA PRICE, D. J. A general theory of bibiiometric and other cumulative advantage processes. **Journal of the American Society for Information Science**, New York, v. 27, n. 5, p. 292–306, 1976.

SORENSON, P. T.; QUIDEAU, S. A.; RIVARD, B. High resolution measurement of soil organic carbon and total nitrogen with laboratory imaging spectroscopy. **Geoderma**, Amsterdam, v. 315, p. 170–177, 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. **Applied Spectroscopy Reviews**, New York, v. 49, p. 139–186, 2014.

SOUZA BAHIA, A. S. R.; MARQUES, J. J.; SIQUEIRA, D. S. Procedures using diffuse reflectance spectroscopy for estimating hematite and goethite in oxisols of São Paulo, Brazil. **Geoderma Regional**, Amsterdam, v. 5, p. 150–156, 2015.

STENBERG, B. Effects of soil sample pretreatments and standardised rewetting as interacted with sand classes on Vis-NIR predictions of clay and soil organic carbon. **Geoderma**, Amsterdam, v. 158, n. 1–2, p. 15–22, 2010. Disponível em: <http://dx.doi.org/10.1016/j.geoderma.2010.04.008>. Acesso em: 21 maio 2015.

STENBERG, B. et al. Visible and near infrared spectroscopy in soil science. In: SPARKS, D. L. (Ed.). **Advances in agronomy**. Burlington: Academic Press, 2010. v. 107, p. 163-215.

STEVENS, A. et al. Prediction of soil organic carbon at the european scale by visible and Near InfraRed Reflectance Spectroscopy. **PLoS ONE**, San Francisco, v. 8, n. 6, 2013.

STEVENS, A.; RAMIREZ-LOPEZ, L. An introduction to the prospectr package. 2013. Disponível em: <http://cran.r-project.org/web/packages/prospectr/vignettes/prospectr-intro.pdf>. Acesso em: 18 out. 2015.

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

SUMMERS, D. et al. Visible near-infrared reflectance spectroscopy as a predictive indicator of soil properties. **Ecological Indicators**, [Amsterdam], v. 11, p. 123–131, 2011.

TEIXEIRA, P. C. et al. **Manual de métodos de análise de solo**. 3. ed. rev. Brasília: Embrapa Solos, 2017. Disponível em: < http://ainfo.cnptia.embrapa.br/digital/bitstream/item/171907/1/Manual-de-Metodos-de-Analise-de-Solo-2017.pdf >. Acesso em: 03 fev. 2018.

TEKIN, Y.; TUMSAVAS, Z.; MOUAZEN, A. M. Comparing the artificial neural network with parcial least squares for prediction organic carbon and pH at different moisture content levels using visible and near-infrared. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 38, n. 3, p. 1794–1804, 2014.

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

TIECHER, T. et al. Tracing sediment sources using mid-infrared spectroscopy in Arvorizinha catchament, southern Brazil. **Land Degradation & Development**, Chichester, v. 1614, p. 1603–1614, 2017.

TOLLEFSON, J. China declared largest source of research articles. **Nature**, London, v. 553, 2018.

VALVERDE-ARIAS, O. et al. Using geographical information system to generate a drought risk map for rice cultivation: case study in Babahoyo canton (Ecuador). **Biosystems Engineering**, London, v. 168, p. 26–41, 2018a.

VALVERDE-ARIAS, O. et al. Homogenisation of a soil properties map by principal component analysis to define index agricultural insurance policies. **Geoderma**, Amsterdam, v. 311, p. 149–158, 2018b. Disponível em: <http://dx.doi.org/10.1016/j.geoderma.2017.01.018>. Acesso em: 31 jan. 2018.

VAN DER MEER, F. Near-infrared laboratory spectroscopy of mineral chemistry: a review. **International Journal of Applied Earth Observation and Geoinformation**, Washington, v. 65, p. 71–78, 2018. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0303243417302210>. Acesso em: 02 dez. 2017.

VAPNIK, V. N. **The nature of statistical learning theory**. 2th ed. New York: Springer, 1995.

VASQUES, G. M. et al. Soil classification using visible/near-infrared diffuse reflectance spectra from multiple depths. **Geoderma**, Amsterdam, v. 223–225, n. 1, p. 73–78, 2014. Disponível em: <http://dx.doi.org/10.1016/j.geoderma.2014.01.019>. Acesso em: 18 out. 2016.

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

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

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

VISCARRA ROSSEL, R. A. et al. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. **Geoderma**, Amsterdam, v. 131, n. 1–2, p. 59–75, 2006. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0016706105000728>. Acesso em: 13 jul. 2014.

VISCARRA ROSSEL, R. A. et al. Baseline map of organic carbon in Australian

soil to support national carbon accounting and monitoring under climate change. **Global Change Biology**, Oxford, v. 20, n. 9, p. 2953–2970, 2014.

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

VISCARRA ROSSEL, R. A. et al. Novel proximal sensing for monitoring soil organic C stocks and condition. **Environmental Science and Technology**, Easton, v. 51, n. 10, p. 5630–5641, 2017.

VISCARRA ROSSEL, R. A.; BEHRENS, T. Using data mining to model and interpret soil diffuse reflectance spectra. **Geoderma**, Amsterdam, v. 158, n. 1–2, p. 46–54, 2010. Disponível em: <http://dx.doi.org/10.1016/j.geoderma.2009.12.025>. Acesso em: 21 maio 2015.

VISCARRA ROSSEL, R. A.; MCGLYNN, R. N.; MCBRATNEY, A. B. Determining the composition of mineral-organic mixes using UV-vis-NIR diffuse reflectance spectroscopy. **Geoderma**, Amsterdam, v. 137, p. 70–82, 2006.

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

WAISER, T. H. et al. In situ characterization of soil clay content with visible near-infrared diffuse refelctance spectroscopy. **Soil Science Society of America Journal**, Madison, v. 71, n. 2, p. 389-396, 2006. Disponível em: < https://dl.sciencesocieties.org/publications/sssaj/abstracts/71/2/389?access=0& view=pdf>. Acesso em: 10 jul. 2018.

WANG, D.-C. et al. The prediction of soil texture from Visible–Near-Infrared Spectra under varying moisture conditions. **Soil Science Society of America Journal**, Madison, v. 80, n. 2, p. 420-427, 2016. Disponível em: <https://dl.sciencesocieties.org/publications/sssaj/abstracts/0/0/sssaj2015.10.03 79>. Acesso em: 31 out. 2017.

WARURU, B. K. et al. Rapid estimation of soil engineering properties using diffuse reflectance near infrared spectroscopy. **Biosystems Engineering**, London, v. 121, p. 177–185, 2014. Disponível em: <http://www.scopus.com/inward/record.url?eid=2-s2.0-84897431748&partnerID=40&md5=21be84ee688d14c88a2a60d3550cc300>. Acesso em: 17 out. 2016.

WETTERLIND, J.; STENBERG, B.; JONSSON, A. Near infrared reflectance spectroscopy compared with soil clay and organic matter content for estimating within-field variation in N uptake in cereals. **Plant and Soil**, Hague, v. 302, n. 1–2, p. 317–327, 2008.

WHITE, K. et al. Mapping the iron oxide content of dune sands, Namib Sand Sea, Namibia, using Landsat Thematic Mapper data. **Remote Sensing of Environment**, New York, v. 62, n. 1, p. 30–39, 1997.

WIGHT, J. P.; ASHWORTH, A. J.; ALLEN, F. L. Organic substrate, clay type, texture, and water influence on NIR carbon measurements. **Geoderma**, Amsterdam, v. 261, p. 36–43, 2016. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0016706115001974>. Acesso em: 02 dez. 2017.

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

WOLD, S. et al. Some recent developments in PLS modeling. **Chemometrics and Intelligent Laboratory Systems**, Amsterdam, v. 58, n. 2, p. 131–150, 2001.

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

WOODARD, J. D. Integrating high resolution soil data into federal crop insurance policy: implications for policy and conservation. **Environmental Science and Policy**, Exeter, v. 66, p. 93–100, 2016a. Disponível em: <http://dx.doi.org/10.1016/j.envsci.2016.08.011>. Acesso em: 07 nov. 2017.

WOODARD, J. D. Big data and Ag-Analytics. **Agricultural Finance Review**, Washington, v. 76, n. 1, p. 15–26, 2016b. Disponível em: <http://www.emeraldinsight.com/doi/10.1108/AFR-03-2016-0018>. Acesso em: 19 jan. 2017.

WOODARD, J. D.; VERTERAMO-CHIU, L. J. Efficiency impacts of utilizing soil data in the pricing of the federal crop insurance program. **American Journal of Agricultural Economics**, Cary, v. 99, n. 3, p. 757–772, 2017.

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**, Amsterdam, v. 310, p. 29–43, 2018.

ZENG, R. et al. How accurately can soil classes be allocated based on spectrally predicted physio-chemical properties? **Geoderma**, Amsterdam, v. 303, p. 78–84, 2017.

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

ZUPIC, I.; CATER, T. Bibliometric methods in management and organization. **Organizational Research Methods**, Thousand Oaks, v. 18, p. 429–472, 2015.

7.RESUMO BIOGRÁFICO

Elisângela Benedet da Silva nasceu em 18 de janeiro de 1975, na cidade de Criciúma - SC, Brasil. Filha de Agenor J. B. da Silva e Matilde Benedet da Silva. Graduou-se em Agronomia pela Universidade Federal de Santa Catarina - UFSC em 2004. Na mesma universidade concluiu em 2007 o curso de Mestrado em Engenharia Civil na área de Cadastro Técnico Multifinalitário e Gestão Territorial, sob orientação da professora Dra. Ruth Nogueira. Entre 2006 e 2007 prestou consultoria na área de pedologia para o Instituto de Pesquisas Ambientais da Universidade do Extremo Sul Catarinense - IPAT/UNESC nos seguintes projetos: Estudo Complementar do EIA/RIMA da Usina Termoelétrica de Santa Catarina - USITESC e na elaboração do mapa de solos do Plano Diretor do Município de Criciúma. Desde 2008 é Agente de Formação Superior III na Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina - Epagri exercendo a função de pesquisadora no Centro de Informações de Recursos Ambientais e de Hidrometeorologia de Santa Catarina - Epagri/Ciram. Em 2014 com liberação concedida pela Portaria DEX - 641/2014, referente ao processo de Pós-Graduação 2014/2015 da Epagri, iniciou o curso de doutorado no Programa de Pós-Graduação em Ciência do Solo da Universidade Federal do Rio Grande do Sul - UFRGS, sob orientação do professor Dr. Élvio Giasson na área de Pedometria e Mapeamento Digital de Solos. Desde 2012 é mãe de um menino.