

Pré-processamento dos dados de espectroscopia na região do Vis-NIR melhoram a predição do carbono orgânico do solo?

Diego José Gris¹
Ricardo Simão Diniz Dalmolin¹
Jean Michel Moura-Bueno¹
Nicolas Augusto Rosin¹
João Pedro Moro Flores¹

¹ Universidade Federal de Santa Maria - UFSM/Departamento de Solos
Avenida Roraima, 1000 - 97105-900 - Santa Maria - RS, Brasil
diegojgris@gmail.com, dalmolin@ufsm.br, bueno.jean1@gmail.com,
nicolasaugustorosin@gmail.com, m.fjoaopedro@gmail.com

Abstract. Soil is one of the most important carbon (C) reservoirs in the environment, acting as a sink or source of C to the atmosphere, depending on its management. Therefore, data on soil organic carbon (SOC) is important for managing C in the ecosystem, but traditional methods for mapping SOC are highly demanding in time and resources. In this scenario, diffuse reflectance spectroscopy (DRS) emerges as a more efficient technique for SOC assessment. This study evaluates preprocessing techniques for soil spectra and models for SOC prediction. The study was conducted at Giruá, State of Rio Grande do Sul (RS), Brazil, where 841 soil samples were collected at 261 sites, in an area of 940 ha. SOC was determined analytically through wet combustion. Soil spectra were obtained using a FieldSpec Pro spectroradiometer. Preprocessing techniques included: smoothing (SMO), continuum removal (CRR), standard normal variate (SNV), detrend (DET), Savitzky-Golay first derivative (SGD), and Savitzky-Golay first derivative after transformation to absorbance (ASG). The samples were divided in training (70%) and validation (30%) sets. Two models were tested for SOC prediction: partial least squares regression (PLSR) and random forest (RF). SMO was the best preprocessing technique for PLSR, while SGD performed best for RF. PLSR had the best performance for SOC prediction ($R_v^2 = 0.72$; $RMSE_v = 0.52\%$ and $RPIQ_v = 2.23$) when compared to RF ($R_v^2 = 0.71$; $RMSE_v = 0.53\%$ and $RPIQ_v = 2.20$).

Palavras-chave: remote sensing, digital soil mapping, partial least squares regression, random forest, sensoriamento remoto, mapeamento digital de solos, regressão por mínimos quadrados parciais, floresta aleatória.

1. Introdução

A demanda por informações espaciais de propriedades do solo tem aumentado devido à necessidade de incremento na produção de alimentos e energia junto com a preocupação com o uso sustentável da terra, sequestro de carbono (C) no solo e atual cenário de mudanças climáticas. O objetivo destas informações sobre o solo são tornar o seu uso mais adequado frente às suas diversas formas de exploração e simular cenários futuros (Hartemink e McBratney, 2008). As pesquisas em ciência do solo, portanto, devem focar no desenvolvimento de métodos que permitam a predição de um vasto conjunto de propriedades do solo em escala regional, nacional e global (Grunwald et al., 2015).

O carbono orgânico do solo (COS) desempenha uma série de funções no ambiente, dentre elas, a regulação do nível de dióxido de carbono (CO₂) atmosférico (Batjes e Sombroek, 1997). O solo é um dos principais reservatórios de C no ecossistema terrestre, podendo atuar como fonte ou dreno de C para a atmosfera, dependendo do uso e manejo a que for submetido (Carvalho et al., 2010). Por este motivo, o conhecimento da distribuição do COS é importante para o manejo do C no ecossistema. Entretanto, mapear o COS é uma atividade que demanda muito tempo e trabalho, devido ao elevado número de pontos que precisam ser amostrados (Vasques et al, 2010), consequência do número de fatores que influenciam a distribuição do COS na pedosfera. Além disso, os métodos analíticos demandam tempo para o processamento das amostras e geram resíduos químicos prejudiciais ao ambiente (Viscarra Rossel e Behrens, 2010; Sousa Junior et al., 2011; Dotto et al, 2014). O cientista do solo, portanto, tem o desafio de propor uma técnica com potencial para predição de propriedades do solo através de uma

análise rápida, barata e não destrutiva das amostras (Viscarra Rossel et al., 2006). Neste contexto, a espectroscopia de reflectância difusa (ERD) tem surgido como uma ferramenta com potencial para a predição de diversas propriedades do solo (Knox et al., 2015; Clairotte et al., 2016; Demattê et al., 2016; Dotto et al., 2016; Viscarra Rossel et al., 2016).

A ERD apresenta uma série de vantagens em relação às metodologias tradicionais de análise do solo, como maior rapidez e eficiência quando se analisa um grande número de amostras, baixa produção de resíduos químicos e menor custo (Viscarra Rossel e Behrens, 2010). Viscarra Rossel et al. (2016) mostrou que é possível predizer várias propriedades do solo a partir de dados espectrais, entre elas, textura, COS, pH e capacidade de troca de cátions (CTC). No entanto, a acurácia das funções de predição de solo (FPS) tem apresentado grande variabilidade (Vasques et al., 2008; Bayer et al., 2012; Knox et al., 2015; Wight et al., 2016). Uma possível explicação para isto é a variação na composição mineralógica, textura, umidade e teor de COS, que podem influenciar a absorção da radiação eletromagnética, causando variações na reflectância e, conseqüentemente, na acurácia das FPS (Wight et al., 2016). O aumento do teor de COS torna o solo mais escuro, causando uma maior absorção de radiação eletromagnética e, conseqüentemente, uma reflectância menor, devido à diminuição do albedo ao longo da região do Vis-NIR (Dalmolin et al., 2005). Estudo de Bellon-Maurel e McBratney (2011) mostrou que a heterogeneidade do conjunto amostral de solos (por exemplo, amostras provenientes de diferentes áreas) pode levar a um viés nas estimativas das propriedades do solo em estudos que utilizam a técnica de ERD. Isto foi observado em trabalho de Ramirez-Lopez et al. (2013), em que o desempenho das FPS para a argila apresentou R^2 de 0,91 em escala detalhada, 0,75 em escala regional e 0,50 em escala global.

Outro fator que influencia a capacidade preditiva das FPS é o tipo de pré-processamento aplicado aos dados espectrais e as técnicas utilizadas no ajuste dos modelos (Vasques et al., 2008; Knox et al., 2015). Segundo Vasques et al. (2008), os pré-processamentos auxiliam na derivação de informações dos espectros e ressaltam feições espectrais relacionadas com as propriedades do solo. Estes autores observaram que o pré-processamento Savitzky–Golay com primeira derivada apresentou o melhor desempenho na predição do COS. Na construção das FPS, a técnica de regressão por mínimos quadrados parciais (PLSR) tem apresentado melhores resultados em comparação à regressão linear múltipla (MLR) e árvore de regressão (RT) (Vasques et al., 2008). Por outro lado, Knox et al. (2015) mostrou que a técnica de floresta aleatória (RF) apresenta vantagens em relação à PLSR quando os valores de COS não apresentam distribuição normal.

Neste contexto, o uso de diferentes técnicas de pré-processamento e modelagem pode ser uma estratégia para melhorar a acurácia das predições de propriedades do solo em escala detalhada. Desta forma, o objetivo deste trabalho foi avaliar o desempenho de modelos de predição para o COS a partir de dados espectrais utilizando técnicas de pré-processamento.

2. Metodologia de Trabalho

2.1 Descrição da área de estudo

O estudo foi realizado em uma área de 940 ha, localizada no município de Giruá, no Estado do Rio Grande do Sul (RS), Brasil (Figura 1). A área pertence à região fisiográfica do Planalto do RS, com altitudes que variam entre 339 e 426 m. O clima predominante é do tipo Cfa, pela classificação de Köppen, caracterizado como clima subtropical úmido sem período seco definido (Alvares et al., 2013). O relevo é suave ondulado a ondulado, com declividades mais acentuadas próximo às drenagens, sendo comuns colinas com declividades de 3 a 10 %. A geologia da área é derivada da Formação Serra Geral. As classes de solos predominante na área são os Latossolos Vermelhos Distroféricos típicos e Gleissolos Háplicos Tb Distrófico

típicos. Os usos do solo predominantes na área são: floresta nativa, campo nativo e lavouras anuais em sistema de plantio direto (SPD) (Moura-Bueno et al., 2016).

Para a caracterização do solo e suas propriedades, foram coletadas amostras de solo em 261 pontos (Figura 1). Destes, 29 pontos foram amostrados nas cinco camadas especificadas pelo consórcio GlobalSoilMap: 0,00-0,05 m; 0,05-0,15 m; 0,15-0,30 m; 0,30-0,60 m e 0,60-1,0 m. No restante dos pontos, as amostras foram coletadas apenas nas três primeiras camadas: 0,00-0,05 m; 0,05-0,15 m e 0,15-0,30 m. A definição dos pontos de coleta no campo foi realizada de duas formas: (1) foi confeccionada uma malha de amostragem irregular, no software SAGA GIS (Conrad et al., 2015), contendo 400 pontos, com o objetivo de representar toda a variabilidade ambiental existente na área; (2) a partir dos 400 pontos foram selecionados 261 pontos com base no conhecimento do pedólogo a respeito da área de estudo, uso da terra, classe de solo e unidade geomorfológica. Foram coletadas três subamostras para compor cada amostra única que foi utilizada para as análises laboratoriais. Os pontos amostrais foram georreferenciados no campo utilizando um aparelho de GPS. Também foram coletadas informações ambientais sobre a condição de drenagem, presença de erosão, pedregosidade e demais informações pertinentes, para auxiliar na discussão dos resultados.

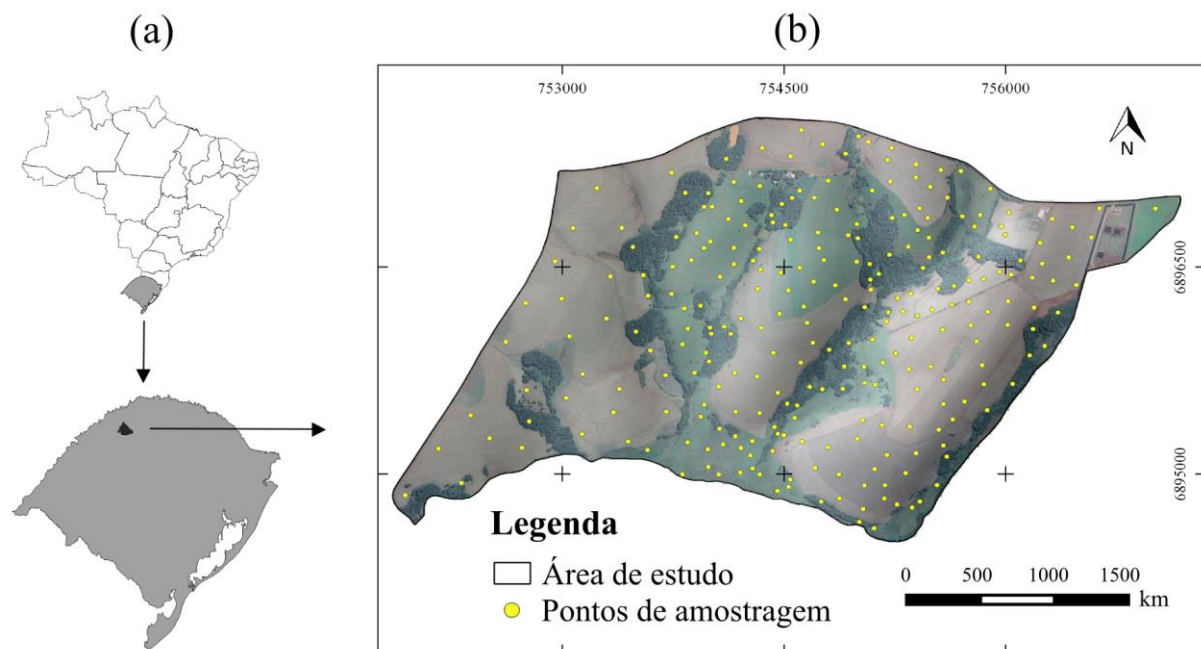


Figura 1. Localização do município de Giruá, Estado do Rio Grande do Sul, Brasil (a) e área de estudo ampliada com imagem do Google Earth e pontos de amostragem (n=261) (b).

2.2 Análises de solo

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

2.3 Obtenção dos dados de espectroscopia de reflectância difusa (ERD)

Os espectros de reflectância difusa das amostras de solo foram obtidos com o espectroradiômetro FieldSpec Pro (ASD Inc.), que possui capacidade de leitura da reflectância no intervalo de 350 a 2500 nm. Para a leitura no espectroradiômetro, foram utilizados aproximadamente 15 cm^3 de solo acondicionado em placa de Petri. Como fonte de

energia, foi utilizada uma lâmpada de halogênio de 50 W com o feixe colimado não-alvo para o plano, posicionada a 35 cm da amostra num ângulo de 30° zênite. Uma placa de sulfato de bário foi utilizada como padrão de referência de 100% de reflectância para a calibração do espectroradiômetro. Foram obtidos os espectros de reflectância do solo realizando-se duas leituras por amostra de solo, com um giro de 180° entre as leituras. Todas as medições espectrais foram realizadas em uma caixa escura para evitar interferência luminosa externa.

2.4 Pré-processamentos aplicados aos dados espectrais

Em primeiro lugar, para reduzir o ruído aleatório, as curvas de reflectância originais foram (a) suavizadas - *smoothing* (SMO) - sobre uma janela móvel de 9 nm. Os dados suavizados foram utilizados como testemunha. Os pré-processamentos aplicados a partir das curvas suavizadas incluíram: (b) *continuum removed reflectance* (CRR); (c) *standard normal variate* (SNV); (d) *detrend* (DET); (e) Savitzky-Golay com 1ª derivada usando um polinômio de primeira ordem, com uma janela de busca de 9 nm (SGD); (f) transformação para absorbância e aplicação do algoritmo Savitzky-Golay com 1ª derivada usando um polinômio de primeira ordem e janela de busca de 5 nm (ASG). Estes pré-processamentos foram aplicados nas curvas espectrais para remover a variabilidade física devido à dispersão da luz e realçar as características de interesse ao longo do espectro. Todos os pré-processamentos foram aplicados no ambiente de programação R (R Core Team, 2016).

2.5 Modelos utilizados para construir as FPS

Para gerar as FPS, as amostras foram separadas em conjuntos de treinamento (70% das amostras) e validação (30% das amostras). Foi realizada a análise descritiva dos conjuntos amostrais e o teste de Bartlett foi aplicado para verificar a existência de homogeneidade das variâncias entre os conjuntos de treinamento e validação.

A fim de comparar a capacidade preditiva de diferentes métodos na construção das FPS para o COS, dois modelos foram utilizados: PLSR e RF. A qualidade das FPS foi avaliada pelos seguintes parâmetros: coeficiente de determinação (R^2), raiz quadrada do erro quadrático médio (RMSE), e proporção de desempenho no intervalo interquartil (RPIQ). Os testes estatísticos e modelagem das FPS foram implementados no ambiente de programação R (R Core Team, 2016).

3. Resultados e Discussão

3.1 Análise descritiva

Os principais resultados da estatística descritiva dos conjuntos amostrais são apresentados na Tabela 1. O valor médio de COS foi de 2,23%, sendo este semelhante no conjunto de treinamento e ligeiramente superior no conjunto de validação. O desvio padrão seguiu a mesma tendência nos três conjuntos. Observando os valores mínimos e máximos dos conjuntos, verifica-se uma grande amplitude entre os valores de COS, o que se atribui às diversas profundidades amostradas e usos do solo na área.

Tabela 1. Estatística descritiva do conjunto total de amostras e dos conjuntos de treinamento e validação.

Conjunto	Número	COS (%)			
		Mínimo	Máximo	Média	DP
Total	841	0,24	5,64	2,23	1,00
Treinamento	589	0,24	5,63	2,21	1,00
Validação	252	0,45	5,64	2,27	0,98

DP: desvio padrão.

3.2 Pré-processamento dos dados espectrais

Considerando as seis técnicas de pré-processamento aplicadas nos espectros, o melhor pré-processamento para o modelo PLSR foi o SMO ($RMSE_v = 0,52\%$). Por outro lado, para o modelo RF, o melhor pré-processamento foi o SGD ($RMSE_v = 0,53\%$).

3.3 Modelos para predição do COS

Os resultados para as FPS ajustadas com os dois modelos e seis técnicas de pré-processamento são apresentados na Tabela 2. As FPS obtidas por PLSR apresentaram R_v^2 variando de 0,63 a 0,72; $RMSE_v$ de 0,52 a 0,59% e $RPIQ_v$ de 1,97 a 2,23. O número de componentes selecionados na PLSR apresentou grande variação (8 a 17). As FPS geradas por RF alcançaram R_v^2 variando de 0,57 a 0,71; $RMSE_v$ de 0,53% a 0,64% e $RPIQ_v$ de 1,82 a 2,20. Foram utilizadas 1.000 árvores para gerar os modelos RF.

No geral, o melhor desempenho na predição do COS foi obtido para as FPS geradas pelo método PLSR ($R_v^2 = 0,72$; $RMSE_v = 0,52\%$ e $RPIQ_v = 2,23$) com o pré-processamento SMO. O modelo RF com pré-processamento SGD ($R_v^2 = 0,71$; $RMSE_v = 0,53\%$ e $RPIQ_v = 2,20$) teve desempenho ligeiramente inferior ao PLSR.

Vasques et al. (2008), comparando métodos multivariados para predição de COS e 30 pré-processamentos aplicados aos dados espectrais, alcançaram R_v^2 variando de 0,66 a 0,85 na modelagem por MLR; R_v^2 de 0,68 a 0,86 por PLSR e R_v^2 de 0,51 a 0,76 por árvore de regressão. Este resultado é semelhante ao encontrado por Bayer et al. (2012), utilizando MLR para predição do COS na África do Sul, alcançando R_v^2 de 0,74 e $RMSE_v$ de 0,36%. Dotto et al. (2014) obteve $R^2 = 0,68$ com o modelo MLR em uma bacia hidrográfica de 1.892 ha e 303 pontos de amostragem, com distintos tipos de solos e uso da terra, localizada na região central do Estado do RS. Sousa Junior et al. (2011) obtiveram valor de $R^2 = 0,38$ utilizando MLR em uma área de 9.000 ha no Estado de São Paulo, com 319 pontos de amostragem.

Tabela 2. Resumo estatístico das FPS geradas por dois modelos e seis pré-processamentos.

Pré-processamentos	Treinamento (n = 589)			Validação (n = 252)		
	R_t^2	$RMSE_t$	$RPIQ_t$	R_v^2	$RMSE_v$	$RPIQ_v$
PLSR						
SMO	0,68	0,57	2,03	0,72	0,52	2,23
CRR	0,67	0,58	2,01	0,67	0,56	2,07
SNV	0,66	0,59	1,98	0,70	0,54	2,17
DET	0,67	0,58	2,01	0,67	0,56	2,08
SGD	0,67	0,58	2,02	0,69	0,55	2,13
ASG	0,65	0,60	1,94	0,63	0,59	1,97
Média do PLSR	0,67	0,58	2,00	0,68	0,55	2,11
RF						
SMO	0,53	0,69	1,68	0,57	0,64	1,82
CRR	0,71	0,54	2,14	0,68	0,56	2,10
SNV	0,62	0,62	1,86	0,64	0,59	1,99
DET	0,63	0,61	1,89	0,64	0,59	1,98
SGD	0,72	0,54	2,16	0,71	0,53	2,20
ASG	0,71	0,54	2,15	0,69	0,55	2,13
Média do RF	0,65	0,59	1,98	0,65	0,57	2,04

PLSR: regressão por mínimos quadrados parciais; RF: floresta aleatória; R^2 : coeficiente de determinação; RMSE: raiz do erro quadrático médio, em %; RPIQ: proporção do erro de predição no intervalo interquartil; t: treinamento; v: validação. Destaque em negrito representa o melhor resultado para cada modelo, baseado no RMSE da validação.

Os resultados obtidos no presente estudo são inferiores aos encontrados por Knox et al. (2015), que alcançou R^2 de 0,87 e RMSE de 0,40 utilizando PLSR e pré-processamento SGD na predição de COS com espectroscopia Vis-NIR, com um conjunto de 1.014 amostras coletadas em todo o estado da Flórida, EUA. Viscarra Rossel e Behrens (2010) compararam diferentes algoritmos de mineração de dados para modelagem do COS por Vis-NIR com um conjunto de dados de 1.104 amostras na Austrália. Estes autores produziram modelos de predição de COS com R_v^2 variando de 0,81 a 0,84, alcançando resultados melhores em comparação ao nosso estudo.

Estudo realizado por Ramirez-Lopez et al. (2014) mostrou que o número de amostras no conjunto de treinamento das FPS tem efeito sobre o desempenho e acurácia das predições, devendo-se tomar mais cuidado quando se trabalha com conjuntos pequenos de amostras. No presente estudo, este fator pode ter influenciado no desempenho das FPS.

4. Conclusões

O modelo PLSR apresentou o melhor desempenho na predição do COS ($R_v^2 = 0,72$; $RMSE_v = 0,52\%$ e $RPIQ_v = 2,23$), seguido do modelo RF ($R_v^2 = 0,71$; $RMSE_v = 0,53\%$ e $RPIQ_v = 2,20$).

A melhor técnica de pré-processamento para o modelo PLSR foi SMO e, para o modelo RF, SGD.

Agradecimentos

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela concessão de bolsa de Iniciação Científica ao primeiro autor e bolsa PQ ao segundo autor.

Referências Bibliográficas

- Alvares, C.A.; Stape, J.L.; Sentelhas, P.C.; Gonçalves, J.L.M.; Sparovek, G. Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift*, v.22, n.6, 711-728, 2013.
- Batjes, N.H.; Sombroek, W.G. Possibilities for carbon sequestration in tropical and subtropical soils. *Glob. Chang. Biol.* 3, 161-173, 1997. <http://dx.doi.org/10.1046/j.1365-2486.1997.00062.x>.
- Bayer, A.; Bachmann, M.; Müller, A.; Kaufmann, H. A comparison of feature-based MLR and PLS regression techniques for the prediction of three soil constituents in a degraded South African ecosystem. *Applied and Environmental Soil Science*, v. 2012, p. 20, 2012.
- Bellon-Maurel, V.E.; McBratney, A. Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils e Critical review and research perspectives. *Soil Biology & Biochemistry*, v.43, p.1398-1410, 2011.
- Carvalho, J.L.N.; Avanzi, J.C.; Silva, M.L.N.; Mello, C.R.; Cerri, C.E.P. Potencial de sequestro de carbono em diferentes biomas do Brasil. *Rev. Bras. Ciênc. Solo*, v.34, p.277-289, 2010.
- Clairotte, M; Grinand, C.; Kouakoua, E.; Thébault, A.; Saby, N.P.A.; Bernoux, M.; Barthès, B.G. National calibration of soil organic carbon concentration using diffuse infrared reflectance spectroscopy. *Geoderma*, v.276, p.41-52, 2016.
- Conrad, O.; Bechtel, B.; Bock, M.; Dietrich, H.; Fischer, E.; Gerlitz, L.; Wehberg, J.; Wichmann, V.; Böhner, J. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4, *Geosci. Model Dev.*, 8, 1991-2007, doi:10.5194/gmd-8-1991-2015, 2015.
- Dalmolin, R.S.D.; Goncalves, C.N.; Klamt, E.; Dick, D.P. Relação entre os constituintes do solo e seu comportamento espectral. *Ciência Rural*, v. 35, p. 481-489, 2005.

Demattê, J.A.M.; Morgan, C.L.S.; Chabrilat, S.; Rizzo, R.; Franceschini, M.H.D.; Terra, F.S.; Vasques, G.M.; Wetterlind, J. Spectral sensing from ground to space in soil science: state of the art, applications, potential, and perspectives. In: Thenkabail, P.S. (Ed.). **Land Resources Monitoring, Modeling, and Mapping with Remote Sensing**. Remote Sensing Handbook Series 2. Boca Raton: CRC Press, 2016. p. 661-732.

Dotto, A.C.; Dalmolin, R.S.D.; Pedron, F.A.; Ten Caten, A.; Ruiz, L.F.C. Mapeamento digital de atributos: granulometria e matéria orgânica do solo utilizando espectroscopia de reflectância difusa. **Rev. Bras. Ciênc. Solo**, v.38, p.1663-1671, 2014.

Dotto, A.C.; Dalmolin, R.S.D.; Ten Caten, A.; Moura-Bueno, J.M. Potential of Spectroradiometry to Classify Soil Clay Content. **Rev. Bras. Ciênc. Solo**, Viçosa, v. 40, e0151105, 2016.

Grunwald S.; Vasques, G.M.; Rivero, R.G. Fusion of soil and remote sensing data to model soil properties. In: Sparks, D.L. (Ed.). **Advances in Agronomy**. Academic Press, 2015. v. 131, cap. 1, p. 1-109.

Hartemink, A. E.; McBratney, A.B. A soil science renaissance. **Geoderma**, v.148, p.123-129, 2008.

Knox, N.M.; Grunwald, S.; McDowell, M.L.; Bruland, G.L.; Myers, D.B.; Harris, W.G. Modelling soil carbon fractions with visible near-infrared (VNIR) and mid-infrared (MIR) spectroscopy. **Geoderma**, v. 239–240, p. 229–239, 2015.

Moura-Bueno, J.M.; Dalmolin, R.S.D.; Ten Caten, A.; Ruiz, L.F.C.; Ramos, P.V.; Dotto, A.C. Assessment of Digital Elevation Model for Digital Soil Mapping in a Watershed with Gently Undulating Topography. **Rev. Bras. Ciênc. Solo**, v.40, 2016.

R Core Team (2016). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

Ramirez-Lopez, L.; Stevens, A. **Memory based learning methods and tools**: towards efficient modelling, predicting and managing tasks in large scale soil spectral libraries. 2013.

Ramirez-Lopez, L.; Schmidt, K.; Behrens, T.; van Wesemael, B.; Demattê, J.A.M.; Scholten, T. Sampling optimal calibration sets in soil infrared spectroscopy, **Geoderma**, v.226–227, p.140–150, 2014.

Sousa Junior, J.G.; Demattê, J.A.M.; Araújo, S.R. Modelos espectrais terrestres e orbitais na determinação de teores de atributos dos solos: potencial e custos. **Bragantia**, v.70, p.610-621, 2011.

Stenberg, B.; Viscarra Rossel, R.A.; Mouazen, A.M.; Wetterlind J. Visible and Near Infrared Spectroscopy in Soil Science. In: Sparks, D.L. (Ed.). **Advances in Agronomy**. Academic Press, 2010. v. 107, cap. 5, p. 163-215.

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

Vasques, G.M.; Grunwald, S.; Comerford, N.B.; Sickman, J.O. Regional modelling of soil carbon at multiple depths within a subtropical watershed. **Geoderma**, v.156, p.326-336, 2010.

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

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

Viscarra Rossel, R.A.; Behrens, T.; Ben-Dor, E.; Brown, D.J.; Demattê, J.A.M.; Shepherd, K.D.; Shi, Z.; Stenberg, B.; Stevens, A.; Adamchuk, V.; Aichi, H.; Barthès, B.G.; Bartholomeus, H.M.; Bayer, A.D.; Bernoux, M.; Böttcher, K.; Brodský, L.; Du, C.W.; Chappell, A.; Fouad, Y.; Genot, V.; Gomez, C.; Grunwald, S.; Gubler, A.; Guerrero, C.; Hedley, C.B.; Knadel, M.; Morrás, H.J.M.; Nocita, M.; Ramirez-Lopez, L.; Roudier, P.; Rufasto Campos, E.M.; Sanborn, P.; Sellitto, V.M.; Sudduth, K.A.; Rawlins, B.G.; Walter, C.; Winowiecki, L.A.; Hong, S.Y.; Ji, W. A global spectral library to characterize the world's soil. **Earth-Science Reviews**, v.155, p.198–230, 2016.

Viscarra Rossel, R.A.; Walvoort, D.J.J.; McBratney, A.B.; Janik, L. J.; Skjemstad, J.O. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. **Geoderma**, v.131, p.59-75, 2006.

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

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