



XXXIII Congresso Brasileiro de Ciência do Solo

Solos nos biomas brasileiros: sustentabilidade e mudanças climáticas
31 de julho à 05 de agosto - Center Convention - Uberlândia/Minas Gerais

DETERMINAÇÃO DO CARBONO ORGÂNICO DO SOLO POR ESPECTROSCOPIA DE INFRAVERMELHO PRÓXIMO E REGRESSÃO POR QUADRADOS MÍNIMOS PARCIAIS

André Marcelo de Souza⁽¹⁾; Maurício Rizzato Coelho⁽²⁾; Etelvino Henrique Novotny⁽²⁾; Ronei Jesus Poppi⁽³⁾; Ricardo de Oliveira Dart⁽¹⁾; Maria de Lourdes Mendonça Santos⁽²⁾; Ricardo Luis Louro Berbara⁽⁴⁾

⁽¹⁾Analista; Embrapa Solos, Rua Jardim Botânico, 1.024 - Jardim Botânico Rio de Janeiro, RJ - CEP 22460-000, Andre.souza@cnpq.embrapa.br; ⁽²⁾ Pesquisador; Embrapa Solos, Rua Jardim Botânico, 1.024 - Jardim Botânico Rio de Janeiro, RJ - CEP 22460-000; ⁽³⁾ Pesquisador; Instituto de Química, Universidade Estadual de Campinas, Caixa Postal 6154, Campinas, SP, CEP 13083-970; ⁽⁴⁾ Professor; Curso de Agronomia - UFRRJ, BR 465, km 7, Seropédica, RJ, CEP 23890-000.

Resumo – Uma abordagem alternativa para a determinação do carbono orgânico do solo (Corg), usando Espectroscopia de Infravermelho Próximo (*Near-infrared Spectroscopy*, NIR) e Regressão por Mínimos Quadrados Parciais (*Partial Least Squares*, PLS) é aplicada neste trabalho. Para tal, foram selecionadas 56 amostras de perfis de solos descritos e coletados no Parque Estadual da Mata Seca, Manga (MG), com o objetivo original de mapeamento dos solos. Este trabalho é uma etapa preliminar de um projeto maior que visa a utilização de tecnologias associadas ao mapeamento digital de solos. Dentre os resultados alcançados, destacam-se: os dados obtidos com as amostras sem pré-tratamento indicaram que as variações observadas nos espectros estão relacionadas ao Corg, tal como pode ser observado pelo coeficiente de correlação (R^2) de 0.901 (calibração do modelo). Os seguintes procedimentos serão futuramente utilizados a fim de elevar a capacidade preditiva do modelo: preprocessar as amostras (secagem e moagem) e aumentar seu número.

Palavras-Chave: calibração multivariada; mapeamento digital de solos.

INTRODUÇÃO

A Sociedade Americana de Testes e Materiais (*American Society of Testing and Material*, ASTM) define a região NIR (*Near Infrared*) como sendo aquela compreendida entre 4.000 a 10.000 cm^{-1} do espectro eletromagnético. Nesta região, a interação da radiação com a matéria provém de sobretons e bandas de combinação de vibrações de grupos -CH, -NH, -OH e -SH (Skoog *et al.*, 2002).

Dentre as principais vantagens da espectroscopia NIR pode se citar a possibilidade de desenvolvimento de métodos analíticos rápidos (com pouco ou nenhum preparo da amostra), não invasivos e não destrutivos, limpos (sem geração de resíduo) e que forneçam a melhor relação custo/benefício possível (Skoog *et al.*, 2002). Devido a essas vantagens, a espectroscopia NIR tem sido empregada por diversos autores para a caracterização química e física do solo (Reeves *et al.*,

2001; McCarty *et al.*, 2002; Calderon *et al.*, 2004; Fidencio *et al.*, 2002; Madari *et al.*, 2005; Madari *et al.*, 2006; Reeves *et al.*, 2006a; Sarkhot *et al.*, 2007; Reeves *et al.*, 2008; Reeves and Smith, 2009; Pimstein *et al.*, 2011). Entretanto, a presença de bandas largas e sobrepostas dos espectros NIR (Fidencio *et al.*, 2002), faz com que seu uso para os mais variados fins, requeram um método multivariado de tratamento de dados a fim possibilitar sua utilização como uma ferramenta de análise.

O método de regressão multivariada utilizado neste estudo foi o método de regressão por Quadrados Mínimos Parciais (*Partial Least Squares*, PLS). O PLS é, em geral, um dos métodos de regressão mais utilizados o tratamento de dados multivariados e no tratamento de dados de espectroscopia de amostras de solos (Wold *et al.*, 1984; Sjoblom *et al.*, 1998; Wold *et al.*, 2001; Reeves *et al.*, 2006b). Como qualquer outro método de regressão, tem como objetivo encontrar uma relação entre a matriz (\mathbf{X}) contendo os espectros das amostras e o vetor que armazena as respectivas propriedades de interesse (\mathbf{y}). O modelo pode ser representado pela equação 1:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{E} \quad (1)$$

onde \mathbf{E} representa os erros do modelo e \mathbf{b} é o vetor de regressão. O método PLS é especialmente indicado, quando \mathbf{X} contém variáveis altamente correlacionadas (como os dados de espectroscopia descritos neste trabalho). Outra vantagem é que este método pode ser usado mesmo quando as amostras contêm interferentes (que estarão presentes no conjunto de calibração). No modelo PLS, a matriz \mathbf{X} é decomposta em escores, \mathbf{t} , e pesos, \mathbf{w} , i.e. $\mathbf{X}\mathbf{W}=\mathbf{T}$, onde $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k)$ é escolhido de maneira que $\mathbf{T} = (\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_k)$ apresente covariância máxima com \mathbf{y} (k é o número de variáveis latentes). Assim, as informações espectrais e as concentrações são usadas ao mesmo tempo na fase de calibração. Fatores importantes na construção de um modelo PLS são: o pré-processamento dos espectros de modo a retirar informações físicas não relevantes, a escolha do número de variáveis latentes k (VL) a serem incluídas no modelo e a repetibilidade do método de referência, o qual fornece os valores da propriedade na etapa de calibração (Wold *et al.*, 2001).

Este trabalho é a etapa inicial de um projeto maior que visa aplicar e validar técnicas de mapeamento digital de classes e atributos de solos em escalas mais detalhadas que 1:50.000. Para tal, qualquer técnica que promova a economia de tempo e recursos financeiros na elaboração de mapas de solos (por meio de mapeamento digital ou convencional), como a aplicação da espectroscopia NIR na determinação de atributos de solo, são recomendados. Dentre esses atributos, cita-se o carbono orgânico do solo (Corg). Este trabalho objetiva testar um método de determinação do conteúdo de Corg por meio de espectroscopia NIR e PLS.

MATERIAL E MÉTODOS

Meio físico

As amostras estudadas, em número de 56, foram obtidas de perfis de solos descritos e coletados para fins de estudos de mapeamento digital de solos no Parque Estadual da Mata Seca (PEMS), município de Manga (MG) (Figura 1). Com o clima semi-árido, vegetação de Caatinga e litologia relacionada a depósitos quarternários do Rio São Francisco e terciários da Formação Bambuí (calcáreos e arenitos), os solos do PEMS apresentam grande variabilidade de seus atributos, incluindo-se o carbono orgânico do solo (Corg), o que é facilmente observado nos trabalhos de mapeamento de solos da área. Latossolos, Cambissolos, Vertissolos, Plintossolos, Neossolos Flúvicos, Gleissolos e Chernossolos, são os solos predominantes do PEMS (Dart et al., 2010).

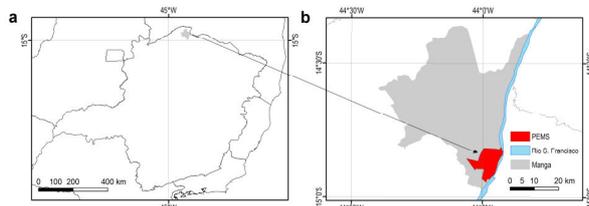


Figura 1. Mapa mostrando a localização (a) do município de Manga no Estado de Minas Gerais; e (b) da área de estudo (Parque Estadual da Mata Seca – PEMS) no município.

Amostras

Para a construção dos modelos de regressão foram utilizadas 56 amostras de solo, destorroadas, secas à temperatura ambiente e passadas em peneira de 2 mm - Terra Fina Seca ao Ar (Embrapa, 1997).

Método de referência

Para determinação do Corg utilizou-se o procedimento preconizado por Embrapa (1997). Para tal, a matéria orgânica foi oxidada com dicromato de potássio em meio sulfúrico, empregando-se como fonte de energia o calor desprendido do ácido sulfúrico. O excesso de dicromato após a oxidação foi titulado com solução padrão de sulfato ferroso amoniacal. Como este trabalho utilizou os resultados analíticos direcionados ao mapeamento de solos do PEMS, não houve réplicas na obtenção dos conteúdos de Corg.

Espectroscopia no infravermelho próximo

Os espectros foram obtidos em um espectrômetro Perkin Elmer *Spectrum 100N* equipado com acessório

de reflectância *SPIN* do Laboratório de Quimiometria em Química Analítica (LQQA) do Instituto de Química da Universidade Estadual de Campinas (UNICAMP).

As amostras (Terra Fina Seca ao Ar) foram colocadas sobre uma cela de vidro e os espectros obtidos na faixa de 4000 a 10000 cm^{-1} , com resolução de 2 cm^{-1} e 128 *scans* por espectro, no modo de reflectância (% R).

Análise multivariada

O conjunto total de espectros foi dividido em dois subconjuntos, sendo que dois terços foi utilizado para a construção do modelo de calibração e, um terço, para a validação interna. Para se determinar o número de Variáveis Latentes a serem utilizadas no modelo, utilizou-se o procedimento de validação cruzada no conjunto de calibração: uma amostra do conjunto de calibração é excluída, o modelo é construído. O processo é repetido até que todas as amostras sejam previstas para 1, 2,...n Variáveis Latentes. O número de Variáveis Latentes foi selecionado com base no gráfico de porcentagem de variância explicada na matriz X e no vetor y . Para calibração, os espectros NIR foram dispostos na forma de uma matriz 36 x 1020, e na etapa de validação, em uma matriz 20 x 1020. Para tornar a matriz X melhor condicionada para a análise e aumentar diferenças espectrais, os dados foram centrados na média.

A capacidade do modelo de calibração para estimar (ou prever) o Corg baseado nos espectros NIR foi avaliada através da correlação dos valores previstos e valores medidos pelo método de referência. Os dados foram exportados no formato ASCII e processados empregando Matlab 7.0 (Mathworks Co). Os espectros foram convertidos para log (1/R), transformados para segunda derivada utilizando o algoritmo de Savitzky-Golay com nove pontos e polinômio de terceira ordem. Os dados foram centrados na média antes da regressão por PLS (Fidêncio et al., 2002).

RESULTADOS E DISCUSSÃO

A Figura 2 apresenta os 56 espectros obtidos após a transformação para log 1/R. Como se pode observar, existem deslocamentos na linha base, os quais podem ser originados tanto por fatores instrumentais, quanto por diferenças naturais de tamanho de partícula entre as diferentes amostras utilizadas neste trabalho (Brereton, 2007). Para corrigir este efeito, utilizou-se o filtro de segunda derivada de Savitzky-Golay (Brereton, 2007). Os espectros, após esta transformação, são mostrados na Figura 3.

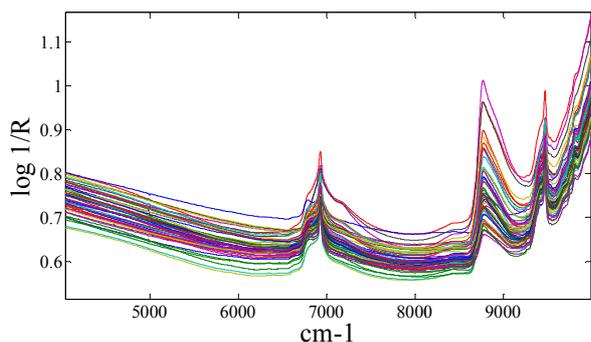


Figura 2: Espectros NIR após a transformação para $\log 1/R$

As regiões contendo apenas informação de linha base (10000-8500 cm^{-1} e 6500-5500 cm^{-1}) foram eliminadas, pois não possuem relevância do ponto de vista químico.

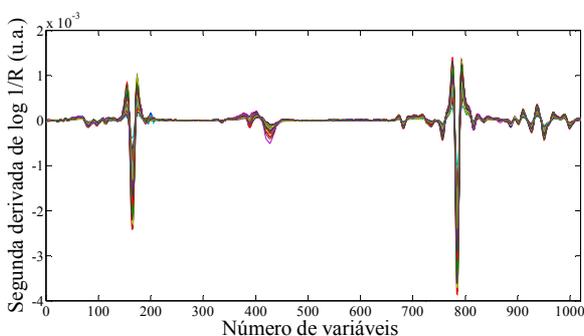


Figura 3: Espectros NIR após a transformação para segunda derivada de $\log 1/R$, enfatizando a eliminação de flutuações de linha base.

Como é possível observar na Figura 4, quatro Variáveis Latentes são suficientes para descrever a totalidade da variância nos espectros. Entretanto, são requeridas oito Variáveis Latentes a fim de explicar a grande variabilidade nos valores de concentração.

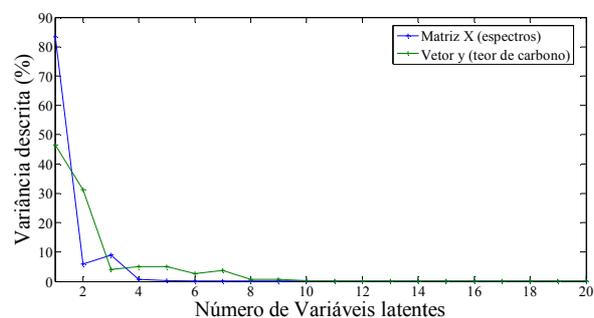


Figura 4: Gráfico de porcentagem de variância descrita vs o número de Variáveis Latentes utilizado para seleção do número de Variáveis Latentes a ser utilizado no modelo.

Para avaliar se as variações espectrais observadas podem ser relacionadas com o teor de carbono, bem como a concordância entre os valores de Corg medidos

peelo método de referência e previstos pelo modelo PLS, avaliou-se o gráfico de valores medidos pelo método de referência vs valores previstos (Figura 5). Pelo exame da Figura 5 verifica-se que as variações espectrais observadas estão relacionadas com o conteúdo de Corg e que existe uma boa concordância entre os valores de referência e os valores previstos, pois o coeficiente de correlação é de 0,901. Entretanto, a grande dispersão de alguns pontos em torno da reta, sugere a existência de um elevado erro de previsão para algumas amostras, o que será estudado em etapas posteriores. Tal variação provavelmente deve-se à variabilidade encontrada no método de referência. Na Figura 5 também é mostrada a reta 1:1 ideal, representando uma correlação perfeita entre o método de referência e o método proposto, bem como a reta de regressão obtida.

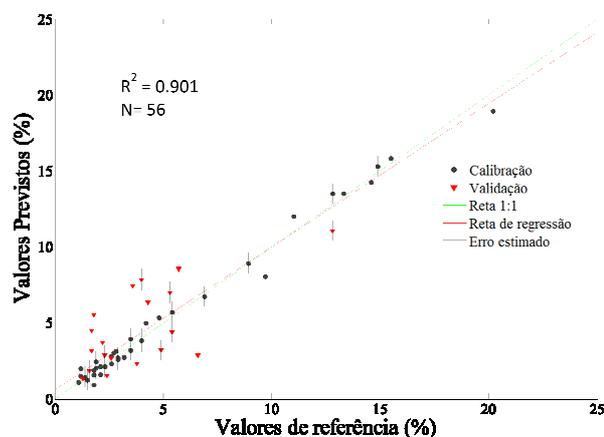


Figura 5: Gráfico de valores de referência vs valores previstos pelo modelo PLS.

CONCLUSÕES

1. Apesar do pequeno número de amostras utilizadas na calibração do modelo ($n=56$), e sem tratamento (Terra Fina Seca ao Ar), os espectros NIR foram relacionados ao conteúdo de Corg ($R^2=0,901$) dos solos estudados através do método PLS, corroborando os resultados de outros autores;

2. O aumento do número de amostras e seu pré-tratamento (moagem e secagem), bem como a utilização de algoritmos quimiométricos (Redes Neurais, *Iterative Partial Least Squares - iPLS*) para a seleção de variáveis, serão futuramente testados a fim de melhorar a capacidade preditiva do modelo.

REFERÊNCIAS

- CALDERON, F.J., MCCARTY, G.W., VAN KESSEL, J.A.S., REEVES, J.B. Carbon and nitrogen dynamics during incubation of manured soil. *Soil Sci. Soc. Am. J.*, 1592-1599, 2004.
- DART R. O.; MENDONÇA-SANTOS, M. L.; COELHO, M. R.; BERBARA, R. L. L.; PARES, J. G. Digital Soil Mapping at Parque Estadual da Mata Seca, MG, Brazil: Applying Regression Tree to predict soil classes. 4th International Workshop on Digital Soil Mapping: From Digital Soil Mapping to Digital Soil Assessment: identifying key gaps from fields to continents. Rome, Italy, 24-26 May, 2010.

- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. CENTRO NACIONAL DE PESQUISA EM SOLOS. MANUAL DE MÉTODOS DE ANÁLISE DE SOLO. 2. ED. RIO DE JANEIRO: EMBRAPA, CENTRO NACIONAL DE PESQUISA DE SOLOS. RIO DE JANEIRO, 1997. 212P.
- FIDENCIO, P.H., POPPI, R.J., DE ANDRADE, J.C., CANTARELLA, H. Determination of organic matter in soil using near-infrared spectroscopy and partial least squares regression. *Commun. Soil Sci. Plant Anal.*, 1607-1615, 2002b.
- MADARI, B.E., REEVES, J.B., COELHO, M.R., MACHADO, P., DE-POLLI, H. Mid- and near-infrared spectroscopic determination of carbon in a diverse set of soils from the Brazilian National Soil Collection. *Spectrosc. Lett.*, 721-740, 2005.
- MADARI, B.E., REEVES, J.B., MACHADO, P., GUIMARAES, C.M., TORRES, E., MCCARTY, G.W. Mid- and near-infrared spectroscopic assessment of soil compositional parameters and structural indices in two Ferralsols. *Geoderma*, 245-259, 2006.
- MCCARTY, G.W., REEVES, J.B., REEVES, V.B., FOLLETT, R.F., KIMBLE, J.M. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. *Soil Sci. Soc. Am. J.*, 640-646, 2002.
- PIMSTEIN, A., NOTESCO, G., BEN-DOR, E. Performance of Three Identical Spectrometers in Retrieving Soil Reflectance under Laboratory Conditions. *Soil Sci. Soc. Am. J.*, 746-759, 2011.
- REEVES, J.B., DELWICHE, S.R., REEVES, V.B. Least squares means multiple comparison testing of reference versus predicted residuals for evaluation of partial least squares spectral calibrations. *J. Near Infrared Spectrosc.*, 371-377, 2006b.
- REEVES, J.B., FOLLETT, R.F., MCCARTY, G.W., KIMBLE, J.M. Can near or mid-infrared diffuse reflectance spectroscopy be used to determine soil carbon pools? *Commun. Soil Sci. Plant Anal.*, 2307-2325, 2006a.
- REEVES, J.B., MCCARTY, G.W., REEVES, V.B. Mid-infrared diffuse reflectance spectroscopy for the quantitative analysis of agricultural soils. *J. Agric. Food Chem.*, 766-772, 2001.
- REEVES, J.B., MCCARTY, G.W., RUTHERFORD, D.W., WERSHAW, R.L. Mid-infrared diffuse reflectance spectroscopic examination of charred pine wood, bark, cellulose, and lignin: Implications for the quantitative determination of charcoal in soils. *Appl. Spectrosc.*, 182-189, 2008.
- REEVES, J.B., SMITH, D.B. The potential of mid- and near-infrared diffuse reflectance spectroscopy for determining major- and trace-element concentrations in soils from a geochemical survey of North America. *Appl. Geochem.*, 1472-1481, 2009.
- SARKHOT, D.V., COMERFORD, N.B., JOKELA, E.J., REEVES, J.B. Effects of forest management intensity on carbon and nitrogen content in different soil size fractions of a North Florida Spodosol. *Plant Soil*, 291-303, 2007.
- SJOBLOM, J., SVENSSON, O., JOSEFSON, M., KULLBERG, H., WOLD, S. An evaluation of orthogonal signal correction applied to calibration transfer of near infrared spectra. *Chemom. Intell. Lab. Syst.*, 229-244, 1998.
- SKOOG, DA; HOLLER, F.J; NEIMAN, T.A. *Análise Instrumental*. 5.ed. São Paulo. Bookman. 2002. 836p.
- WOLD, S., RUHE, A., WOLD, H., DUNN, W.J. The collinearity problem in linear-regression - the partial least-squares (pls) approach to generalized inverses. *Siam Journal on Scientific and Statistical Computing*, 735-743, 1984.
- WOLD, S., TRYGG, J., BERGLUND, A., ANTTI, H. Some recent developments in PLS modeling. *Chemom. Intell. Lab. Syst.*, 131-150, 2001.
- BRERETON, R.G. *Applied Chemometrics for Scientists*. 1ed. Chichester. John Wiley & Sons. 2007. 356p.