

Determinação atributos físicos do solo por espectroscopia no infravermelho próximo e regressão por quadrados mínimos parciais ⁽¹⁾.

André Marcelo de Souza ⁽²⁾; Maurício Rizzato Coelho ⁽³⁾; Paulo Roberto Figueiras ⁽⁴⁾; Lorena Alves Amaro da Costa ⁽⁵⁾; Etelvino Henrique Novotny ⁽³⁾; Ronei Jesus Poppi ⁽⁶⁾;

⁽¹⁾ Trabalho executado com recursos do Projeto Embrapa - MP3 n°. 03.10.06.013.00.00.

⁽²⁾ Analista; Embrapa Solos, Rua Jardim Botânico, 1.024 - Jardim Botânico Rio de Janeiro, RJ - CEP 22460-000, Andre.souza@cnpq.embrapa.br; ⁽³⁾ Pesquisador; Embrapa Solos; ⁽⁴⁾ Estudante de Doutorado; Instituto de Química, Universidade Estadual de Campinas, Caixa Postal 6154, Campinas, SP, CEP 13083-970; ⁽⁵⁾ Estagiária Embrapa Solos; ⁽⁶⁾ Professor, Instituto de Química, Universidade Estadual de Campinas, Caixa Postal 6154, Campinas, SP, CEP 13083-970.

RESUMO: Uma abordagem alternativa para a determinação da granulometria do solo, empregando espectroscopia na região do infravermelho próximo e Regressão por Mínimos Quadrados Parciais é aplicada neste trabalho. Para tal, foram selecionadas 148 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 à utilização de tecnologias associadas ao mapeamento digital de solos. Dentre os resultados alcançados, destacam-se: os dados obtidos com as amostras na condição de terra fina seca ao ar indicaram que as variações observadas nos espectros estão relacionadas aos teores de areia grossa, areia fina, areia total, argila natural e silte. Entretanto, por se tratar de um estudo preliminar, os resultados apresentados estão em fase de otimização em que diversos algoritmos para calibração multivariada e pré-processamento dos espectros estão sendo avaliados.

Termos de indexação: calibração multivariada; mapeamento digital de solos.

INTRODUÇÃO

A espectroscopia na região do infravermelho próximo (*near infrared spectroscopy*, NIR) vem encontrando cada vez aplicações na área de solos devido a vantagens como rapidez, baixo custo e, principalmente, por ser uma técnica que não necessita de reagentes químicos e, portanto, não gera resíduos agressivos ao meio ambiente. Além disso, recentes desenvolvimentos de acessórios de reflectância difusa que permitem a análise de 100-400 gramas de solos de uma só vez (pratos giratórios ou *spinning*) têm levado a um aumento da representatividade das amostras, contribuindo para a construção de grandes bancos de dados espectrais (Gogé et al., 2012). Diversos trabalhos têm sido publicados com o objetivo de avaliar o potencial da espectroscopia NIR em substituição aos métodos analíticos tradicionalmente empregados em análises de rotina de diversos

atributos do solo. A determinação de atributos relacionados com sua composição física, como areia grossa, areia fina, areia total, silte, argila e argila dispersa em água (McBratney et al., 2006; Viscarra Rossel et al., 2006, 2009) tem sido constantemente relatada na literatura.

Diversos autores afirmam que a espectroscopia NIR e a quimiometria convivem em simbiose, uma vez que a espectroscopia NIR possui bandas largas e sobrepostas, esta se beneficia da quimiometria para tornar-se uma ferramenta cada vez mais robusta para identificação e quantificação de diversos parâmetros em diferentes matrizes, e a quimiometria se beneficia do tipo e da elevada quantidade de informação espectral gerada para o desenvolvimento de novos métodos (McClure, 1994; Pasquini, 2003). A calibração multivariada é um dos métodos quimiométricos mais utilizado para a extração de informações quantitativas dos diversos atributos do solo por espectroscopia NIR. O método de calibração multivariada mais comumente utilizado em espectroscopia NIR é a regressão por quadrados mínimos parciais (*partial least squares*, PLS) (Feudale et al., 2002; Pasquini, 2003). O PLS é utilizado para construir modelos de calibração que descrevem a relação entre os dados espectrais (matriz **X**) e as variáveis dependentes de interesse (matriz **Y**) medidas pelos métodos de referência (Igne et al., 2009).

Este trabalho é a etapa inicial de um projeto maior que visa aplicar e validar a espectroscopia NIR para a determinação de diversos atributos do solo, dentre eles a granulometria. Nesta etapa, foram construídos modelos de calibração multivariada empregando PLS para previsão dos teores de areia grossa, areia fina, areia total, argila natural e silte. Os modelos apresentados se encontram em fase de otimização e seus resultados são preliminares.

MATERIAL E MÉTODOS

Amostras de solo

As amostras estudadas, em número de 148, 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). Com o clima semi-árido, vegetação de Caatinga e litologia relacionada a depósitos quaternários do Rio São Francisco e terciários da Formação Bambuí (calcários e arenitos), os solos do PEMS apresentam grande variabilidade de seus atributos, incluindo-se os teores de areia grossa, areia fina, areia total, argila natural e silte, 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).

Preparo das 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

Espectroscopia no infravermelho próximo

Os espectros foram obtidos em um espectrômetro Perkin Elmer Spectrum 100N equipado com acessório de reflectância difusa NIRA do Laboratório de Análise de Solos e Plantas (LASP) da Embrapa Solos. 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 4 cm^{-1} e 64 scans por espectro, no modo de refletância (% R).

Análise multivariada

O conjunto total de espectros foi dividido em dois subconjuntos utilizando-se o método de Kennard-Stone com o auxílio de um *toolbox* gratuito, desenvolvido por Wen Wu and Sijmen de Jong, do FABI – Vrije Universiteit Brussel, 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 100 x 6001, e na etapa de validação, em uma matriz 48 x 6001. A capacidade do modelo de calibração para estimar (ou prever) os teores de areia grossa, areia fina, areia total, argila natural e silte, baseada nos espectros NIR, foi avaliada através dos valores da raiz quadrada do erro médio de calibração (*root*

mean squares error of calibration, RMSEC), de previsão (*root mean squares error of prediction*, RMSEP) e dos coeficientes de determinação para os valores previstos e valores observados (medidos) pelo método de referência. Os dados foram exportados no formato ASCII e processados empregando Matlab R2011b (The Math-Works, Natick, MA) e PLS_toolbox 4.2.1 (EigenvectorResearch, Wenatchee, WA). Os espectros foram convertidos para $\log(1/R)$, e preprocessados pelos métodos de variação normal padrão (*standard normal variate*, SNV) seguido de primeira derivada utilizando o algoritmo de Savitzky-Golay com quinze pontos e polinômio de segunda ordem (Xu et al., 2008).

RESULTADOS E DISCUSSÃO

A Figura 1 apresenta os 148 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. Para corrigir este efeito, utilizou-se SNV e o filtro de primeira derivada de Savitzky-Golay. Os espectros após preprocessamento por SNV e SNV seguido pela primeira derivada são mostrados na Figura 2 e 3 respectivamente.

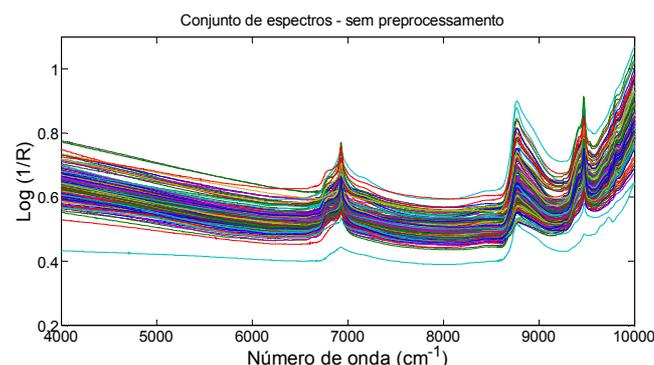


Figura 1: Espectros NIR após a transformação para $\log 1/R$.

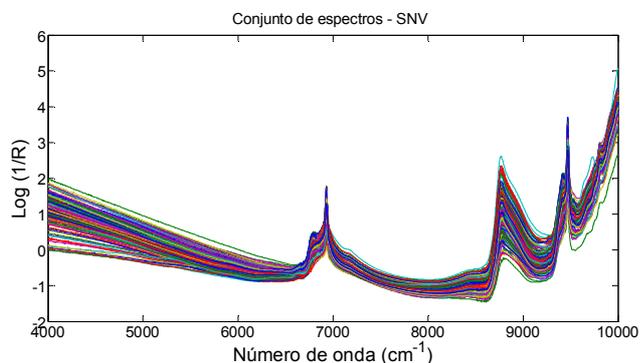


Figura 2: Espectros NIR após pré-processamento por SNV.

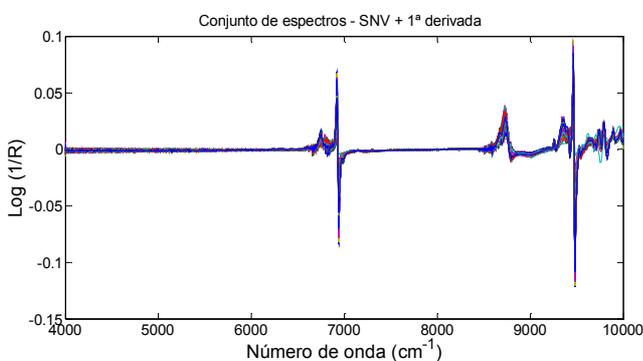


Figura 3: Espectros NIR após pré-processamento por SNV e 1ª derivada, enfatizando a eliminação de flutuações de linha base.

Para avaliar se as variações espectrais observadas podem ser relacionadas (de forma indireta) com os teores de areia grossa, areia fina, areia total, argila natural e silte, bem como a concordância entre os valores obtidos pelo método de referência e previstos pelo modelo PLS, avaliou-se o gráfico de valores previstos pelo método de referência vs valores observados para cada um dos atributos estudados (Figuras 4, 5, 6, 7 e 8). De maneira geral, pode ser observado que as variações espectrais observadas estão relacionadas com o conteúdo de dos atributos medidos. Entretanto, a grande dispersão de alguns pontos em torno da reta, sugere a existência de elevado erro de previsão para algumas amostras. Tal fato será estudado em etapas posteriores.

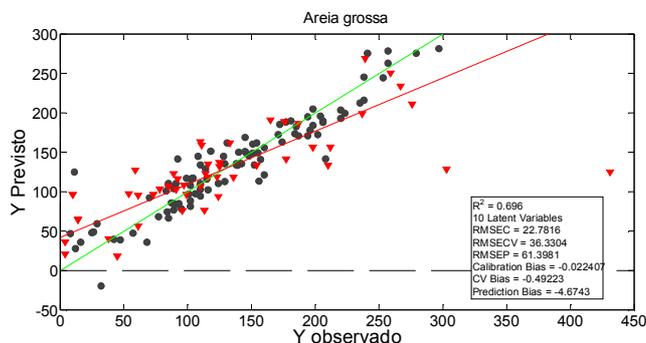


Figura 4: Gráfico de valores de referência vs previstos pelo modelo PLS para o teor de areia grossa.

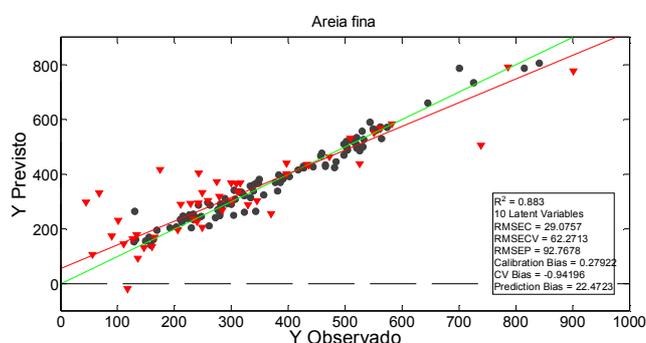


Figura 5: Gráfico de valores de referência vs previstos pelo modelo PLS para o teor de areia fina.

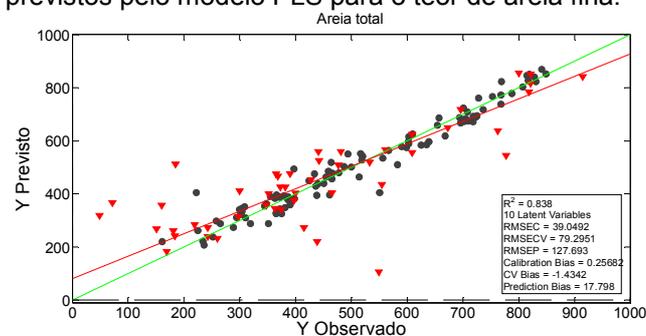


Figura 6: Gráfico de valores de referência vs previstos pelo modelo PLS para o teor de areia total.

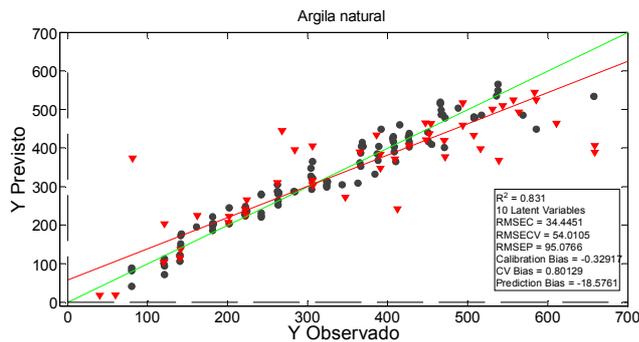


Figura 7: Gráfico de valores de referência vs previstos pelo modelo PLS para o teor de argila natural.

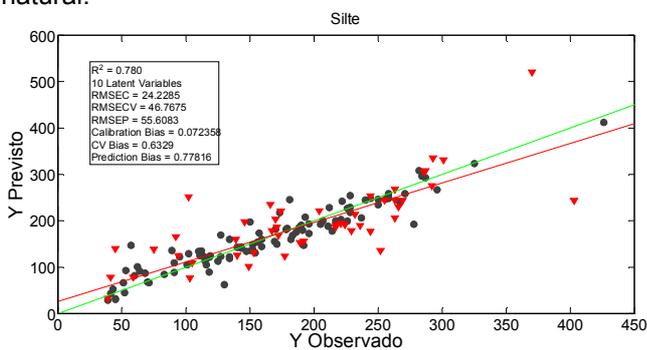


Figura 8: Gráfico de valores de referência vs previstos pelo modelo PLS para o teor de silte.

CONCLUSÕES

1. Apesar do pequeno número de amostras na condição de TFSA utilizadas na calibração do modelo ($n=100$), os espectros NIR foram relacionados ao conteúdo de areia grossa, areia fina, areia total, argila natural e silte dos solos estudados através do método PLS, corroborando os resultados de outros autores;

2. O aumento do número de amostras, bem como a otimização dos modelos de calibração multivariada através da utilização de outros

métodos de pré-processamento do conjunto e de seleção de variáveis, serão futuramente testados a fim de melhorar a capacidade preditiva do modelo.

REFERÊNCIAS

- Barboza, F.D. & Poppi, R.J. Determination of alcohol content in beverages using short-wave near-infrared spectroscopy and temperature correction by transfer calibration procedures. *Analytical and bioanalytical chemistry* 377:695–701, 2003
- Feudale, R., Woody, N., Tan, H. Transfer of multivariate calibration models: a review. *Chemometrics and Intelligent Laboratory Systems* 64:181–192, 2002
- Gogé, F., Joffre, R., Jolivet, C., Ross, I., Ranjard, L. Optimization criteria in sample selection step of local regression for quantitative analysis of large soil NIRS database. *Chemometrics and Intelligent Laboratory Systems* 110:168–176, 2012
- Igné, B., Roger, J.-M., Roussel, S., Bellon-Maurel, V., Hurburgh, C.R. Improving the transfer of near infrared prediction models by orthogonal methods. *Chemometrics and Intelligent Laboratory Systems* 99: 57–65, 2009.
- McBratney, A.B., Minasny, B., Viscarra Rossel, R. Spectral soil analysis and inference systems: A powerful combination for solving the soil data crisis. *Geoderma* 136:272–278, 2006.
- Pasquini, C. Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. *Journal of the Brazilian Chemical Society* 14:198–219, 2003.
- Viscarra Rossel, R.A., Cattle, S.R., Ortega, A., Fouad, Y. In situ measurements of soil colour, mineral composition and clay content by vis-NIR spectroscopy. *Geoderma* 150:253–266, 2009.
- Xu, L., Zhou, Y.-P., Tang, L.-J., Wu, H.-L., Jiang, J.-H., Shen, G.-L., Yu, R.-Q. Ensemble preprocessing of near-infrared (NIR) spectra for multivariate calibration. *Analytica chimica acta* 616:138–43, 2008