

## Modelos de quantificação de atributos do solo a partir de bibliotecas espectrais

Suzana Romeiro Araújo<sup>1</sup>  
José Alexandre Melo Demattê<sup>1</sup>  
Rodnei Rizzo<sup>1</sup>  
Henrique Bellinaso<sup>1</sup>  
Simone Vicente<sup>1</sup>

<sup>1</sup> Universidade de São Paulo - USP/ESALQ  
Caixa Postal 09 - 13416-000 - Piracicaba - SP, Brasil  
sraraujo@esalq.usp.br, jamdemat@esalq.usp.br, rizzo@esalq.usp.br,  
henriquebellinaso@yahoo.com.br, vicente\_simone@yahoo.com.br

**Abstract.** Soil science has sought to develop better techniques for the soil evaluation, one of which is the use of remote sensing applications. The use of ground sensors to obtain soil spectral data has enabled the characterization of these data and the advancement of techniques for the quantification of soil attributes. In order to do this, the creation of a soil spectral library is necessary. A spectral library should be representative of the variability of the soils in a region. The objective of this study was to (i) assess the quality of models to predict soil attributes, clay and iron (Fe<sub>2</sub>O<sub>3</sub>), obtained by spectral libraries of distinct soils from several agricultural regions of Brazil., (ii) illustrate the use of a spectral library of soils in the quantification of its attributes, (iii) evaluate the performance of a benchmark for when the data set is subdivided according to their spectral characteristic. Spectral data were collected (using a Fieldspec sensor, 350–2,500 nm) for the horizons of 223 soil profiles from the several regions of Brazil. A Principal Component Analysis (PCA) of the data was then performed. Using a K-means clustering and the PCA scores, samples were divided in 3 subgroups. A PLSR was applied in each subgroup, algorithms to predict clay and iron based on soil spectra were created. A discriminant analysis was performed, we aimed to create a model which indicates if samples from validation set are similar to subgroup 1, 2 or 3. According to this classification, validation dataset was divided. These validation sets were used to evaluate PLSR models performance.

**Palavras-chave:** sensoriamento remoto laboratorial, teor de argila, ferro total, espectrorradiometria

### 1. Introdução

A análise de rotina é o método mais utilizado para avaliar os atributos do solo e determinar a necessidade de nutrientes para as plantas. Além disso, auxilia na sua classificação e na adoção do manejo agrícola mais adequado para uma região. Os métodos de análise utilizados em laboratório de solos levaram bastante tempo para atingir o grau de confiabilidade que possuem. Trata-se de métodos em que grande quantidade de reagentes é gasto e resíduos produzidos, além do elevado custo e tempo para obtenção de resultados. No Brasil podem ser citadas as metodologias adotadas por Camargo et al. (1986), Rajj et al. (2001) e Embrapa (1997).

O sensoriamento remoto próximo surge como uma técnica alternativa eficaz na obtenção de informações sobre os atributos químicos, físicos e mineralógicos de solos (Cohen et al., 2007). Trata-se de análises não destrutivas de amostras, em que as interações entre a luz incidente e a superfície de um “corpo” são estudadas (Stark et al., 1986). Segundo Viscarra Rossel et al. (2006) a avaliação dos atributos do solo por meio da sua resposta espectral pode resultar em dados mais precisos que aqueles obtidos por métodos convencionais de análise, os quais alteram o equilíbrio entre as fases do solo.

Os modelos de predição são construídos através de métodos quimiométricos (Minasny, 2008) e são constantemente calibrados em função da região de estudo. Portanto, o uso adequado desta técnica depende da construção de um banco de dados espectrais ou *biblioteca espectral*, que apresente grande variabilidade/representatividade de solos (Viscarra-Rossel et al., 2008).

Nesse sentido, este trabalho objetiva: (i) avaliar a qualidade de modelos de predição de atributos de solo, argila e ferro ( $\text{Fe}_2\text{O}_3$ ), gerados a partir de bibliotecas espectrais (dados obtidos com sensor FieldSpec em laboratório, na região do visível e infravermelho próximo) com abrangência regional e grande quantidade de dados, (ii) exemplificar a utilização de uma biblioteca espectral de solos na quantificação de seus atributos, (iii) avaliar o desempenho de um modelo de quantificação quando o conjunto de dados é subdividido de acordo com sua característica espectral.

## 2. Metodologia de trabalho

Foram utilizados os dados espectrorradiométricos de 1428 amostras de terra, obtidas de 233 perfis de solos, representativos de diversas regiões do país (Mato Grosso do Sul, Minas Gerais, São Paulo e Goiás) para avaliar o teor de argila. Já as análises dos teores de ferro total ( $\text{Fe}_2\text{O}_3$ ) foram realizadas em 248 amostras deste total.

Os dados espectrais foram obtidos em laboratório com o espectrorradiômetro FieldSpec Pro (Analytical Spectral Devices, Boulder, Colo.) (Hatchell, 1999). Trata-se de um sensor cuja resolução espectral é de 1 nm nos comprimentos de onda de 350 a 1100 nm e de 2 nm nos comprimentos que vão de 1100 a 2500 nm. Para a coleta dos dados de reflectância, as amostras secas em estufa à 45 °C durante 24 horas (Henderson et. al., 1992), em seguida moídas e peneiradas (malha de 2 mm). Cada amostra corresponde à aproximadamente um volume de 15 cm<sup>3</sup> de solo. Essa foi alocada em placa de petri para leitura no sensor. A reflectância de cada amostra foi dada pela média de 100 leituras realizadas pelo sensor. A ponta de captação luminosa do equipamento (entrada do cabo de fibra ótica) foi colocada em posição vertical a 8 cm de distância da amostra, onde foi medida a luz refletida numa área aproximada de 2 cm<sup>2</sup> no centro da amostra. A fonte de iluminação utilizada foi uma lâmpada halógena de 50 W, com feixe não colimado para o plano visado, sendo posicionada a 35 cm da amostra e com um ângulo zenital de 30°. Utilizou-se como padrão de referência, uma placa branca de sulfato de bário, considerada como padrão de 100% de reflectância. A partir dos dados obtidos foram criadas representações gráficas (curvas de reflectância espectral).

Para análise química convencional dos teores argila foi usado o método do densímetro (Camargo et. al, 1986). Já os teores de ferro total ( $\text{Fe}_2\text{O}_3$ ) foram obtidos através do ataque sulfúrico (Camargo et al., 1986).

Através do programa *Unscrambler 9.7* foram retirados os *outliers* dos dados da biblioteca espectral. Com a finalidade de diminuir a dimensionalidade de informações contidas no banco de dados e verificar a variabilidade dos dados espectrais foi realizada a análise de componentes principais (PCA) (Wold, 1982) através do programa estatístico *JMP 7.0.1* (SAS, 2007). Quando aplicamos um algoritmo de PCA num conjunto de curvas espectrais de solos, o conjunto original das mesmas é substituído por um novo conjunto de variáveis (ou *scores*) denominado de componentes principais (CPs). A principal característica deste novo conjunto é a ortogonalidade, porém o mesmo é facilmente reconstruído a partir da combinação linear das variáveis originais (curvas espectrais). Como vantagem, o novo conjunto de variáveis (CPs), geralmente concentra a maior parte da informação (variância) em poucas variáveis (*scores*), diminuindo assim a dimensionalidade dos dados, sem perda significativa da informação.

Foram selecionadas aleatoriamente, através do programa estatístico *JMP 7.0.1* (SAS, 2007), 70% dos dados (953 amostras para argila e 167 para o ferro) para calibração dos modelos, denominado aqui grupo de calibração (Gc). Os 30% restantes (471 amostras para argila e 81 para o ferro) constituíam o grupo de validação (Gv). Os parâmetros estatísticos dos teores de argila e ferro total para o Gc e Gv (Tabela 2), indicam sua representatividade.

Tabela 1 - Parâmetros estatísticos dos teores de argila e Fe<sub>2</sub>O<sub>3</sub>

Parâmetros	Argila (g kg <sup>-1</sup> )		Fe <sub>2</sub> O <sub>3</sub> (g kg <sup>-1</sup> )	
	Dados de calibração	Dados de validação	Dados de calibração	Dados de validação
média	448.0	436.1	15.9	15.5
desvio padrão	251.7	248.2	10.2	10.0
mediana	400.0	380.0	18.0	17.6
máximo	930.0	920.0	36.1	32.7

Os modelos regionais para predição de argila foram realizados em quatro passos:

**(1) Subdivisão da biblioteca espectral - gerando os subgrupos regionais:** Para a identificação e subdivisão das amostras de acordo com suas características espectrais, utilizou-se a técnica de agrupamento cluster k-médias (Wold, 1982), implementada no programa JMP 7.0.1 (SAS, 2007). Foram utilizados como variáveis para o agrupamento, os cinco primeiros componentes principais. Consideraram-se três subgrupos como o número ótimo para o agrupamento. Isso se deve a quantidade de amostras, uma vez que se dividindo em mais classes, os subgrupos regionais de calibração poderiam deter poucos indivíduos e consequentemente não serem representativos.

**(2) Calibração de modelos de predição da argila:** Os modelos de predição foram construídos através da técnica regressão de mínimos quadrados parciais (PLSR) (Martens e Naes, 1989). A técnica é amplamente utilizada, apresentando bom desempenho na estimativa de atributos baseada no comportamento espectral do solo (Viscarra Rossel et al., 2008; Vasques et al., 2008). O algoritmo PLSR seleciona fatores ortogonais que maximizam a covariância entre variáveis preditoras (X, espectro) e a variável resposta (y, dados de laboratório). Ao ajustar um modelo espera-se encontrar poucos fatores PLSR que expliquem a maior parte da variação em X e y.

De forma resumida, o algoritmo decompõe X e y em “scores” (T) e “loadings” (P e q):

$$X = TP' + E \quad (1)$$

$$y = Tq + f \quad (2)$$

A decomposição é realizada de forma que os primeiros fatores explicam a maior parte da variação em X e y. Logo, fornecendo ao modelo um novo espectro X, a propriedades do solo y pode ser estimada através de uma combinação linear dos scores e loadings de X. Esta técnica foi aplicada, separadamente, a cada um dos três subgrupos obtidos na etapa anterior.

**(3) Modelo de regionalização das amostras:** Nesta etapa criou-se um modelo estatístico, sendo este responsável por indicar em qual dos grupos (definidos na etapa 1) as novas amostras pertenceriam. Para tanto, foi utilizada a análise discriminante (Wold, 1982), a qual se baseando nos componentes principais do espectro indicava automaticamente se as mesmas pertenciam ao grupo 1, 2 ou 3. Para sua calibração utilizou-se o Gc.

**(4) Validação:** As amostras do Gv foram inicialmente submetidas ao modelo de regionalização (etapa 3). Em seguida, cada um dos três subgrupos de validação teve seu teor de argila predito, conforme seu respectivo modelo de calibração.

Por fim, a avaliação dos modelos foi feita com base nos valores dos coeficientes de determinação ( $R^2$ ) e determinação ajustado ( $R^2_{ajus}$ ), erro médio (ME), raiz quadrada do erro médio (RMSE) e desvio padrão do erro (RPD).

### 3. Resultados e Discussão

Para a argila, foi possível separar os dados espectrais em três grupos distintos com 453, 364 e 136 amostras, respectivamente (Figura 1). Para cada um dos grupos foi gerado um modelo de quantificação deste atributo no programa ParLeS (Viscarra-Rossel, 2008).

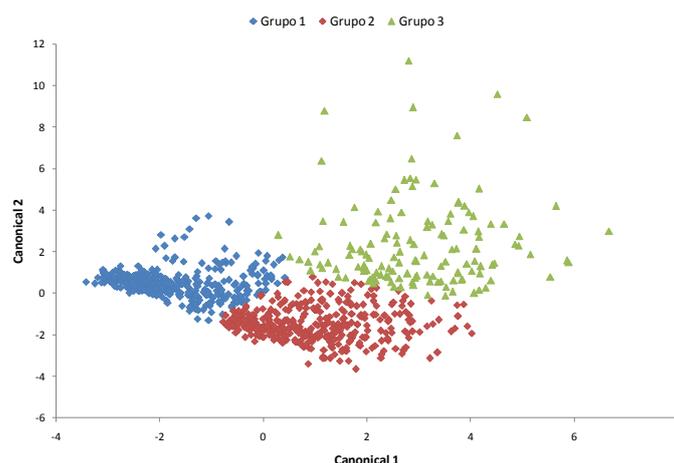


Figura 1 - Representação gráfica da análise discriminante realizada nos dados de calibração de argila

Os resultados demonstraram que o melhor modelo foi aquele obtido com a utilização de 364 amostras (Grupo 2), com um erro de no máximo 5,3 % para o teor de argila,  $RPD > 3$  e  $R^2$  significativo de 0,93 (Figura 2). Para construção destes modelos foram utilizados 18 fatores PLS.

Saeys et al. (2005) estabeleceram que valores de  $R^2$  entre 0,50 e 0,65 indicam a possibilidade de discriminação das altas e baixas concentrações do atributo no modelo, enquanto que, valores de  $R^2$  de aproximadamente 0,66 a 0,81, de 0,82 a 0,90 e acima de 0,90 indicam modelos quantitativos de predição de atributos químicos do solo aceitáveis, bons e excelentes, respectivamente. Dunn et al. (2002) e Chang et al. (2001) sugerem que valores de RPD menores que 1,5 sejam considerados insuficientes para a maioria das aplicações, enquanto que modelos com valores maiores que 2,0 sejam considerados excelentes. Já modelos com valores de RPD entre 1,5 e 2,0 são julgados úteis em relação à acurácia das predições. Portanto, os três modelos gerados são aceitáveis para predizer o atributo físico do solo, argila.

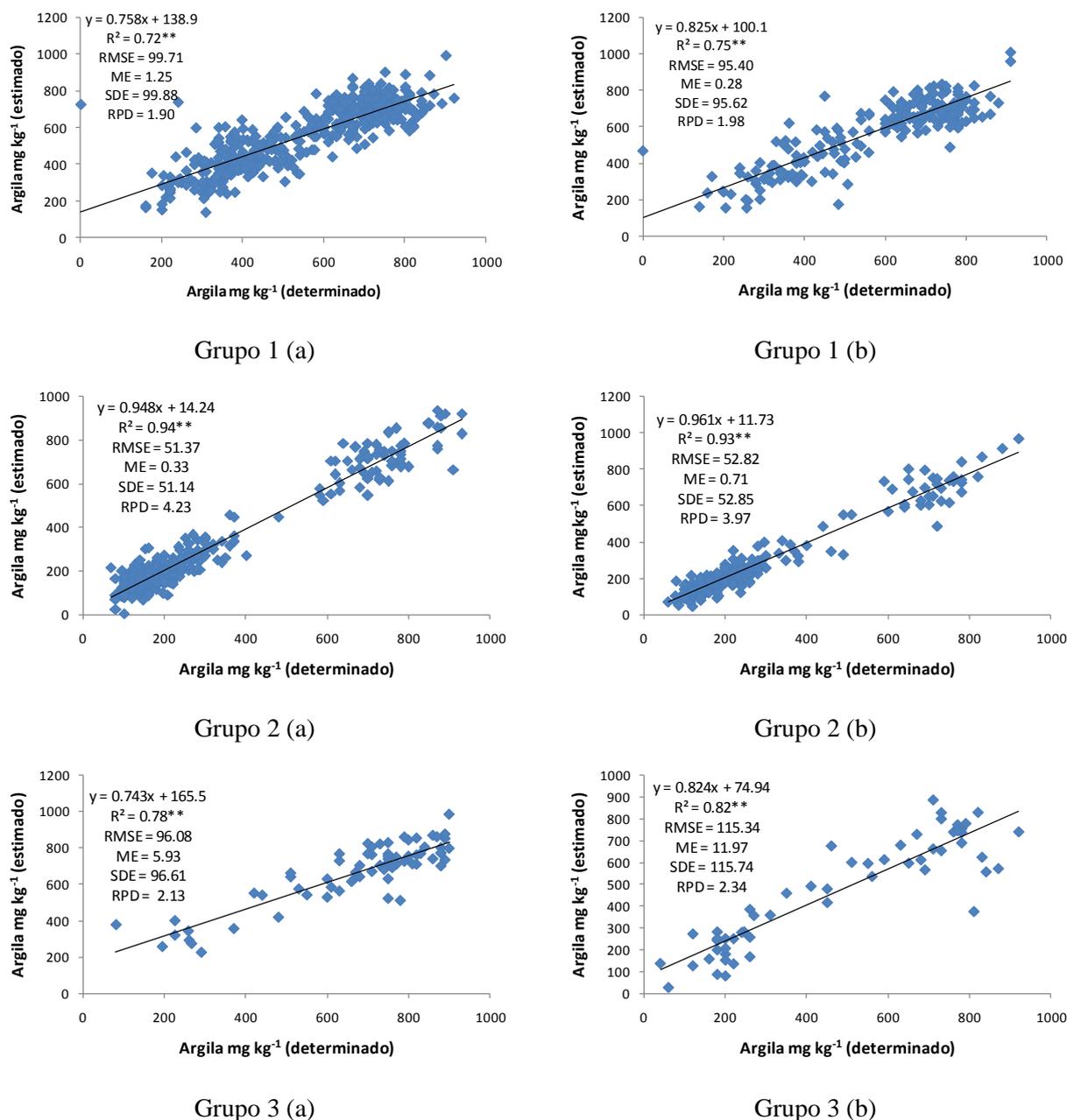


Figura 2 - (a) Modelos PLS da estimativa de argila gerados para cada grupo; (b) validação do modelo.

$R^2$ : coeficiente de determinação significativo a 1%; RMSE: raiz quadrada do erro médio; RPD: desvio padrão do erro; ME: erro médio; SDE = desvio padrão da distribuição dos erros;

No entanto, ao avaliar este atributo, considerando o conjunto total de dados e gerando um modelo único de predição, os resultados não foram diferentes (Figura 3). Isso demonstra que a subdivisão em grupos permitiu boas predições com um número reduzido de amostras. Desta forma, torna-se possível a construção de modelos de calibração regionais (solos com mesmo material de origem).

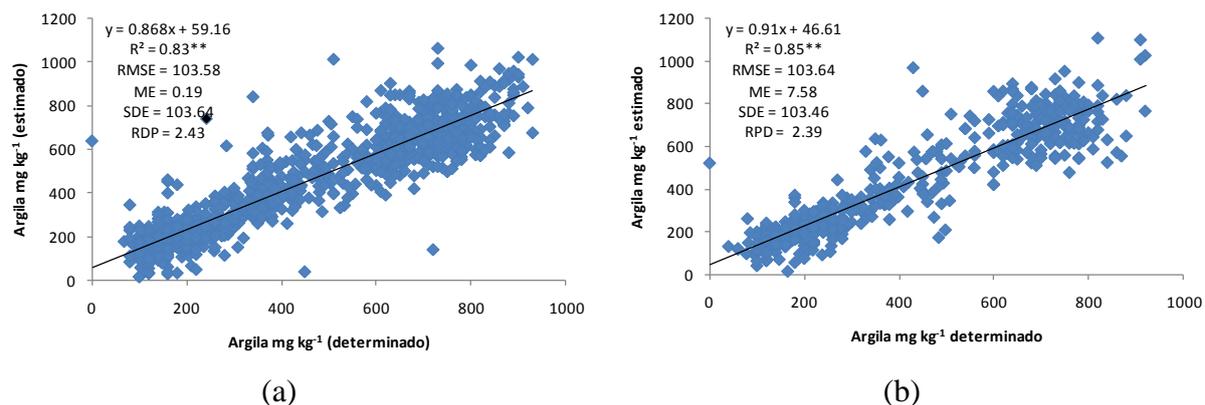


Figura 3 - (a) Modelo PLS da estimativa de argila avaliado para o conjunto total de dados; (b) validação do modelo.

$R^2$ : coeficiente de determinação significativo a 1%; RMSE: raiz quadrada do erro médio; RPD: desvio padrão do erro; ME: erro médio; SDE = desvio padrão da distribuição dos erros;

Os dados de ferro total não foram separados em grupos com a análise de componentes principais e discriminante. Portanto, também foi gerado um modelo de quantificação de ferro total (ParLeS 3.1), considerando todos os dados. Foi realizada a validação cruzada (valores obtidos das análises químicas convencionais versus valores preditos pelo modelo) a fim de se obter um número de componentes para gerar um modelo com maiores coeficientes de determinação ( $R^2$ ), menores valores da raiz quadrada do erro médio (RMSE) e maiores valores de desvio padrão do erro (RPD). Com a utilização de 16 fatores PLS, foi gerado o modelo de calibração de ferro total. A fim de testar a confiabilidade deste modelo de predição, foram utilizadas 81 amostras de solos, referentes aos 30% do total do banco. Os resultados demonstram que este modelo quantitativo de predição do ferro do solo é aceitável ( $R^2 = 0,65$ ) útil em relação à acurácia das predições ( $RPD > 1,5$ ).

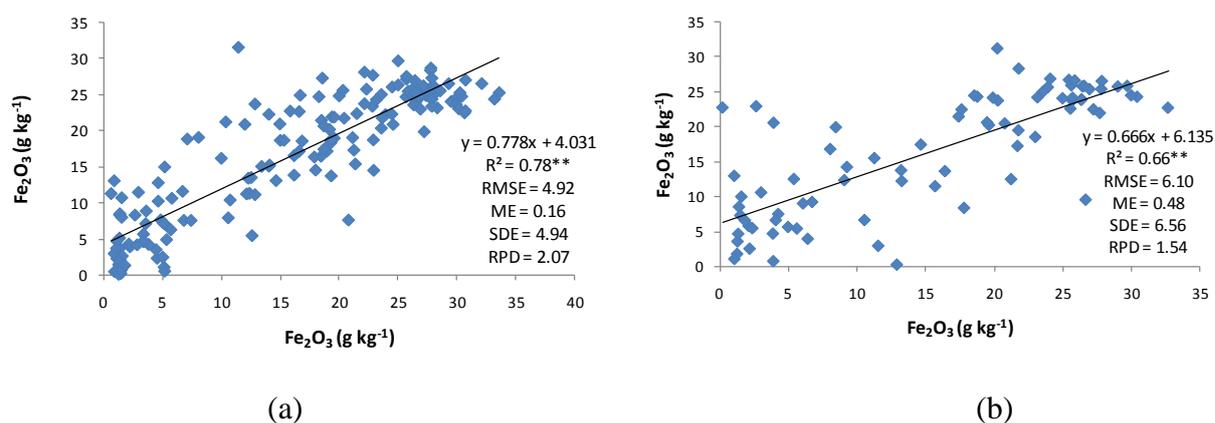


Figura 3 - (a) Modelo PLS da estimativa de  $Fe_2O_3$  avaliado no conjunto total de dados; (b) validação do modelo.

$R^2$ : coeficiente de determinação significativo a 1%; RMSE: raiz quadrada do erro médio; RPD: desvio padrão do erro; ME: erro médio; SDE = desvio padrão da distribuição dos erros;

#### 4. Conclusões

É possível prever os teores de argila e ferro ( $Fe_2O_3$ ) de solos, através de modelos estatísticos gerados a partir de bibliotecas espectrais. A subdivisão dos dados espectrais de acordo com a

região e material de origem do solo é uma boa alternativa para a geração de modelos, com elevada acurácia, de quantificação de atributos.

### **Referências Bibliográficas**

Camargo, A.O. de; A.C. Moniz; J.A. Jorge; J.M. Valadares. **Métodos de análise química, mineralógica e física de solos do IAC**. Campinas: Instituto Agrônomo, 1986. 94p. (IAC. Boletim técnico, 106).

Chang, C.W.; Laird, D.A.; Mausbach, M.J.; Hurburgh Junior, C.R. Near-infrared reflectance spectroscopy – principal components regression analyses of soil properties. **Soil Science Society of America Journal**, Madison, v.65, p.480-490, 2001.

Cohen, M.; Mylavarapu, R.S.; Bogrecki, I.; Lee, W.S.; Clark, M.W. Reflectance Spectroscopy For Routine Agronomic Soil Analyses. **Soil Science**, Baltimore, v.172, p.469-485, 2007.

Dunn, B. W.; Beecher, H. G.; Batten, G. D.; Ciavarella, S. The potential of near-infrared reflectance spectroscopy for soil analysis – a case study from the Riverine Plain of south-eastern Australia. **Australian Journal of Experimental Agriculture**, Melbourne, v.42, p.607-614, 2002.

EMBRAPA. **Manual de métodos de análise de solo**. 2 ed. Rio de Janeiro: Centro Nacional de Pesquisa de Solos, 1997. 212p.

Henderson, T.L.; Baumgardner, M.F.; Franzmeier, D.P.; Stott, D.E.; Coster, D.C. High dimensional reflectance analysis of soil organic matter. **Soil Science Society of America Journal**, Madison, v.56, p.865-872, 1992.

Martens, H., Næs, T. Multivariate Calibration. John Wiley and Sons, Chichester, p. 419, 1989.

Minasny, B.; McBratney, A. B. Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. **Chemometrics and Intelligent Laboratory Systems**, 2008.

Raij, B. van.; Andrade, J.C.; Cantarella, H.; Quaggio, J.A. **Análise química para avaliação de solos tropicais**. Campinas: IAC, 2001. 285p.

SAS Institute. 2007. JMP Version 7.0.1, SAS Institute, Cary, NC.

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

Viscarra-Rossel, R.A.; Jeon, Y.S.; Odeh, I.O.A.; McBratney, A.B. Using a legacy soil sample to develop a mid-IR spectral library. **Australian Journal of Soil Research**, Sidney, v. 46, p. 1-16, 2008.

Viscarra Rossel, R.A. ParLeS: Software for chemometric analysis of spectroscopic data. **Chemometrics Intelligent Laboratory Systems**, Amsterdam, v.90, p.72-83, 2008.