

Avaliação de modelos de calibração PLS e SVM na determinação do carbono orgânico do solo por espectroscopia NIR

Paulo R. Filgueiras^{1*} (PG), André M. Souza² (PQ), Ronei J. Poppi¹ (PQ), Maurício R. Coelho² (PG), Jerônimo G. Parés² (IC), Thais A. F. Cunha² (IC), Ricardo O. Dart² (PQ).

¹Instituto de Química, Universidade Estadual de Campinas, Caixa Postal 6154, Campinas, SP, CEP 13083-970;

²Embrapa Solos, Rua Jardim Botânico, 1.024 - Jardim Botânico Rio de Janeiro, RJ - CEP 22460-000. pauloiuna@hotmail.com

Palavras Chave: Quimiometria, Calibração Multivariada.

Introdução

A espectroscopia no infravermelho próximo (NIR) tem sido cada vez mais utilizada para análises quantitativas em amostras de solos. Dentre suas vantagens, destacam-se métodos analíticos rápidos, não invasivos, não destrutivos, limpos e que fornecem uma relação custo/benefício mais favorável quando comparado aos métodos analíticos tradicionais. Sua utilização em análises quantitativas requer, invariavelmente, a utilização de métodos multivariados de tratamento de dados.

Nos últimos anos, vários trabalhos sobre aprendizagem de máquinas foram desenvolvidos para modelar dados não lineares^{1,2}. Esses métodos possuem uma vantagem especial sobre os métodos lineares tradicionais (MLR, PCR e PLS), de modelar dados de forma generalizada. Cortes & Vapnik³ desenvolveram um método de aprendizagem de máquina para solução de problemas de classificação binária denominada Máquinas de Vetores de Suporte (SVM). Posteriormente, a técnica foi estendida para tratar problemas de classificação multiclasses e regressão. O SVM pode ser aplicado a problemas de calibração não linear, em função do mapeamento do espaço original dos dados em um espaço de alta dimensão, denominado de espaço de características, pela aplicação de uma função Kernel não linear³.

Este trabalho tem como objetivo avaliar os modelos de calibração multivariada PLS e SVM na determinação do carbono orgânico do solo (CO) por espectroscopia NIR.

Resultados e Discussão

Analisou-se o teor de CO de 87 amostras de solos do município de Manga-MG, segundo o procedimento preconizado por Embrapa (1997). As amostras foram obtidas de 11 perfis de solos amostrados para fins de mapeamento. Tomou-se espectros NIR em duplicata para cada uma das 87 amostras e utilizou-se o espectro médio para elaboração dos modelos. As amostras foram divididas em dois conjuntos: 60 para calibração e 27 para validação dos modelos. A variação na linha de

base dos espectros foi corrigida pela técnica de correção de espalhamento multiplicativo e em seguida aplicou-se os modelos PLS e SVM. A avaliação foi seguida pela análise da raiz quadrada dos erros de calibração RMSECV e previsão RMSEP.

A Figura 1 mostra a relação entre os valores modelados e previstos pelos modelos.

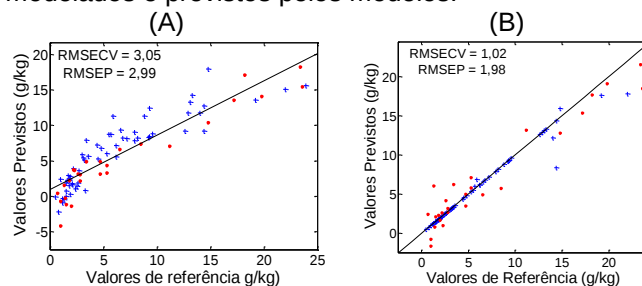


Figura 1. Gráfico dos valores de referência vs valores previstos pelos modelos: (A) PLS e (B) SVM. + amostras de calibração e • amostras de previsão.

O modelo PLS (Figura 1A) apresentou distribuição dos pontos com tendência não linear, indicando não se adequar aos dados. O mesmo não ocorre para o modelo SVM (Figura 1B), onde as amostras distribuíram-se de forma aleatória em torno da reta de relação linear. O modelo SVM foi o de melhor exatidão, com valores de 1,98g/kg contra 2,99 g/kg do modelo PLS.

Conclusões

A relação entre a resposta instrumental (espectro NIR) e o CO é não linear, sendo necessária a utilização de técnicas de calibração multivariadas específicas. Para este fim, o SVM mostrou ser uma excelente ferramenta quimiométrica de modelagem.

Agradecimentos

Os autores agradecem o CNPq.

¹ Xu, Q. S.; Liang, Y. Z.; Shen, H. L. *J. Chemometrics*. **2001**, 15: 135–148.

² Fong, S. S.; Kiss, V. S.; Brereton, R. G. *Talanta*. **2011**, 83, 1269–1278.

³ Cortes, C. & Vapnik, V. N. *Machine Learning*, **1995**, 20, 273–297.