

Aplicação do SPA e LS-SVM para determinação de parâmetros químicos de solos

Diego M. Souza (PG)¹, Beáta E. Madari (PQ)¹, Freddy F. Guimarães (PQ)² e Anderson da Silva Soares (PQ)², diego@cnpaf.embrapa.br

¹EMBRAPA, Centro Nacional de Pesquisa de Arroz e Feijão, GO 462, Km 12, Sto. Antônio de Goiás/GO, 75375-000

²Instituto de Química, Universidade Federal de Goiás, Campus Samambaia, Goiânia/GO, 74001-970

Palavras Chave: calibração multivariada, desvio da linearidade, SPA, RLM, LS-SVM.

Introdução

Em química analítica as metodologias são geralmente realizadas em regiões onde as calibrações têm comportamento linear. A razão disto deve-se principalmente a duas vantagens que tal região proporciona: (i) facilidade de obtenção da equação que correlaciona o teor do analito, $[x]$, e a absorbância (Abs), (ii) potencialização da sensibilidade do método, já que nesta região a $dAbs/d[x]$ é máxima. Em quimiometria, não é diferente, as técnicas mais difundidas de calibração multivariada pressupõe esta correlação como linear. São elas: Regressão Linear Múltipla (RLM); Regressão em Componentes Principais (PCR); Mínimos Quadrados Parciais (PLS). Contudo, ferramentas capazes de modelar fenômenos não lineares, como Redes Neurais Artificiais e Máquinas de Vetores de Suporte com Mínimos Quadrados (LS-SVM)¹, também têm sido aplicadas em calibrações multivariadas e apresentado resultados estatisticamente equivalentes ou superiores aos métodos lineares.

Quando há muitas variáveis independentes para construção das calibrações, pode existir multicolinearidade, ou seja, variáveis que tenham alta correlação, também denominado de informação redundante. A inclusão de variáveis multicolineares diminui a capacidade de generalização dos modelos. Técnicas de seleção de variáveis reduzem a dimensionalidade das variáveis independentes, produzindo modelos mais simples, robustos, e fáceis de interpretar. Para este fim, aplica-se principalmente: PLS por intervalos (i-PLS), Eliminação de Variáveis não Informativas por PLS (UVE-PLS), e Algoritmo Genético (GA). O Algoritmo das Projeções Sucessivas (SPA), proposto em 2001 por Ugulino Araújo *et al.*², é uma nova técnica de seleção de variáveis para modelos por RLM.

Apesar do SPA ter sido proposto para seleção de variáveis a serem utilizadas na RLM, esta apresenta a desvantagem de não ajustar possíveis casos de não-linearidade. Neste estudo, avaliou-se a associação do SPA com a técnica de ajuste multivariado LS-SVM. As calibrações consideraram um conjunto de 1173 solos brasileiros para predição dos teores de Matéria Orgânica (MO), Carbono total

(C_t), e Nitrogênio total (N_t), a partir dos espectros de infravermelho médio (MID-IR).

Resultados e Discussão

As amostras de solo selecionadas são representativas do território brasileiro, o que facilitou a identificação de desvios da linearidade das calibrações por RLM para MO e C_t, Fig. 1. Modelos restritos consideram amostras de características e concentrações próximas, situações nas quais os modelos lineares apresentam resultados satisfatórios. Nos três parâmetros estudados a substituição da RLM por LS-SVM melhorou a predição dos modelos, diminuindo a Raiz Quadrada do Erro Médio Quadrático da Previsão (RMSEP), ver Fig. 1.

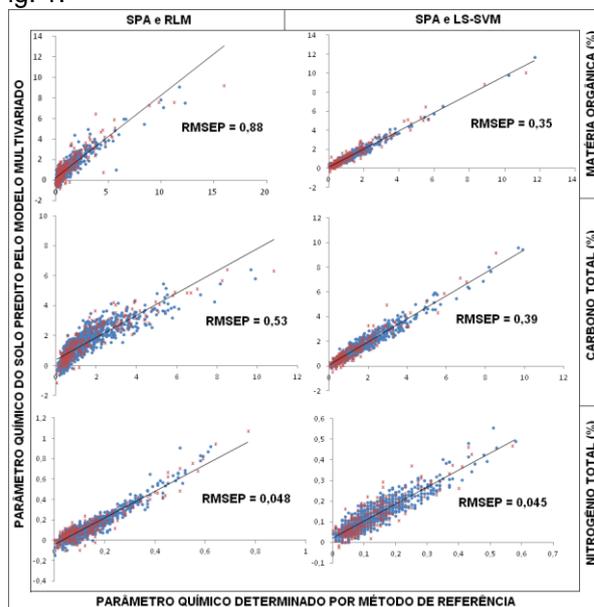


Figura 1. Desempenho dos modelos multivariados

Conclusões

Foi vantajoso associar o SPA ao LS-SVM, técnica capaz de modelar fenômenos não lineares, na determinação de MO e C_t. Os resultados são superiores aos obtidos por RLM, e são aplicáveis à rotina laboratorial de análise de fertilidade de solo.

¹ Cogdill, R. P. e Dardenne, J. *Near Infrared Spectrosc.* **2004**, *12*, 93-100.

² Araújo, M. C. U.; Saldanha, T. C. B.; Galvão, R. K. H.; Yoneyama, T.; Chame, H. C. e Visani, V. *Chem. and Intelligent Lab.* **2001**, *57*, 65-73.