

Um Novo Algoritmo de Seleção de Variáveis Aplicado à Determinação de Viscosidade de Óleos Vegetais por Espectrometria NIR

Sófacles F. C. Soares (IC)^{1*}, Alessandra F. C. Pereira (PG)¹, Francisco F. G. Neto (PG)¹, Márcio J. C. Pontes (PG)¹, Edvan C. Silva (PQ)¹, Mário C. U. Araújo (PQ)¹, Wallace D. Fragoso (PQ)¹, Sérgio R. B. Santos (PQ)² e Roberto K. H. Galvão (PQ)³
sofacles@gmail.com

¹Depto. de Química - Universidade Federal da Paraíba; ²Coord. De Química - Centro Federal de Educação Tecnológica de Alagoas; ³Divisão de Engenharia Eletrônica - Instituto Tecnológico da Aeronáutica, Brasil.

Palavras Chave: Seleção de variáveis, Calibração multivariada, Óleos vegetais, Viscosidade e Espectrometria NIR

Introdução

Regressão por mínimos quadrados parciais (PLS) e regressão linear múltipla (MLR) são técnicas usadas para calibração multivariada. Todavia, os modelos PLS podem não permitir uma interpretação física direta em virtude da regressão ser realizada no domínio dos dados transformados^[1]. Ao contrário, a MLR opera no domínio original, obtendo-se modelos mais simples e fáceis de interpretar.

Não obstante as vantagens dos modelos MLR, a regressão é sensível a problemas de colinearidade. Esse problema tem sido minimizado pelo algoritmo das projeções sucessivas (APS) em análises espectrométricas UV-Vis^[2], ICP-OES^[3] e NIR^[4].

Neste trabalho, propõe-se um novo algoritmo de seleção de variáveis baseado na busca dos maiores ângulos intersetores (ou variáveis) no espaço das amostras. O algoritmo proposto foi aplicado à determinação de viscosidade de óleos vegetais por espectrometria NIR e calibração MLR.

Materiais e Métodos

Aquisição e partição das amostras

70 amostras de óleos vegetais (soja, milho, canola e girassol) de diferentes lotes e fabricantes foram adquiridas em supermercados na cidade de João Pessoa. Esse conjunto de amostras foi dividido em 40, 15 e 15, respectivamente, para calibração, validação e previsão usando o algoritmo SPXY^[5].

Obtenção e pré-processamento dos espectros

Os espectros NIR foram registrados na região de 1628 a 2422 nm com resolução de 4cm⁻¹ e média de 16 varreduras, utilizando um espectrofotômetro FTIR Perkin Elmer, modelo Spectrum-GX. Em seguida, foram derivados após suavização usando um filtro Savitzky-Golay com polinômio de 2ª ordem e janela de 21 pontos. O resultado é mostrado na Fig. 1.

Seleção de variáveis e calibração multivariada

A seleção de variáveis pelo APS e o algoritmo proposto, bem como a calibração MLR, foram realizadas por softwares escritos em Matlab[®] 6.5. A calibração PLS foi realizada no Unscrambler[®] 9.6.

A escolha final das variáveis selecionadas pelos algoritmos e das latentes pelo PLS, foi realizada como base no menor valor do erro médio quadrático de previsão (RMSEP) definido como:

$$RMSEP = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$$

onde \hat{y}_i e y_i são os valores de viscosidade obtidos pelos modelos e método de referência. N é o número de amostras da série de teste (ou previsão).

Resultados e Discussão

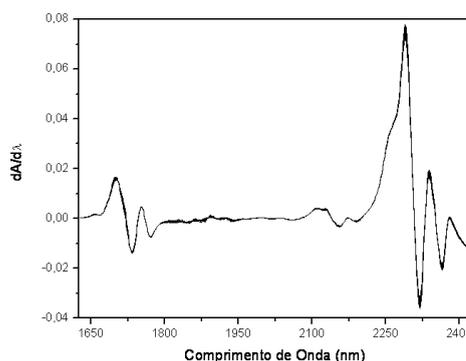


Fig. 1. Espectros derivativos NIR das amostras dos óleos vegetais.

A Tab. 1 mostra que os parâmetros de desempenho dos modelos são comparáveis, com exceção do número de variáveis selecionadas. De fato, o algoritmo proposto produziu um modelo MLR com maior parcimônia (menos variáveis).

Tab. 1-Valores de RMSEP de previsão, correlação (R²) e número de variáveis (entre parênteses) obtidos pelos modelos.

MODELO	PLS	MLR-APS	MLR-Algoritmo proposto
RMSEP	0,7	0,7 (22)	0,6 (8)
R ²	0,9820	0,9743	0,9797

Conclusões

O algoritmo proposto permitiu construir modelo MLR mais simples que o APS sem perda de capacidade de previsão do parâmetro de interesse.

Agradecimentos

- [1] BEEBE, K. R. et al Chemometrics-A Practical Guide, Wiley, New York, 1998.
- [2] ARAUJO, M.C.U.; et al. Chemom. Intell. Lab. Syst. 57 (2001) 65-73.
- [3] GALVÃO, R.K.H.; et al. Anal. Chim. Acta 443 (2001) 107-115.
- [4] BREITKREITZ, M.C. et al. Analyst 128 (2003) 1204-1207.
- [5] R.K.H. GALVÃO, et al. Talanta 67 (2005) 736-740.