

# DETERMINAÇÃO NÃO DESTRUTIVA DE PROTEÍNA POR DRIFTS/PLS-AG

Pedro A. Konzen<sup>1</sup>; Marco Flôres Ferrão<sup>2\*</sup>; João Carlos Furtado<sup>3</sup>;  
Marcelo A. Morgano<sup>4</sup>, Neura Bragagnolo<sup>5</sup> e Márcia M.C. Ferreira<sup>6</sup>

## RESUMO

Métodos de determinação não destrutivos e não geradores de resíduos nocivos ao ambiente vem sendo cada vez mais estudados, principalmente empregando informações provenientes de técnicas espectroscópicas por reflexão. Neste trabalho o teor de proteína de amostras de farinha de trigo e de café crú moído foram simultaneamente modeladas empregando os espectros de reflexão no infravermelho médio com transformada de Fourier (DRIFTS) juntamente com a técnica de regressão por mínimos quadrados parciais (PLS) e otimização com algoritmos genéticos (AG). Modelos robustos de calibração com coeficientes de correlação entre 0,987-0,992 foram obtidos e validados tanto com amostras de trigo como de café sendo os valores de erro percentual médio inferiores à 2,6%. Estas novas metodologias permitem o desenvolvimento de métodos de regressão empregando mais de uma matriz alimentícia de natureza semelhante.

**Palavras-chave:** proteína, grãos, DRIFTS, quimiometria, regressão multivariada, algoritmos genéticos.

## SUMMARY

NON-DESTRUCTIVE PROTEIN CONTENT DETERMINATION USING DRIFTS/PLS-GA. In this study are combined the potentialities of chemometrics techniques and the facilities in the acquisition of spectra presented by the technique Diffuse Reflectance Infrared Fourier Transform Spectroscopy (DRIFTS), bearing in mind the prediction of protein levels in wheat and green coffee samples. To build the models the Partial Least-squares regression method (PLS) was used with multiplicative scatter correction spectra in log (1/R) form and genetic algorithm routine for select spectral range. To validate the models, Standard Error Validation (SEV) was used. The results reflect the importance to choose the set variable selection, as well as the kind of pre-processing applied. These results let us conclude that good

---

<sup>1</sup> Depto. de Matemática da Universidade de Santa Cruz do Sul, m20254@mx.unisc.br

<sup>2</sup> Depto. de Química e Física da Universidade de Santa Cruz do Sul, ferrao@dquimfis.unisc.br

<sup>3</sup> Depto. de Informática da Universidade de Santa Cruz do Sul, jcarlosf@polaris.unisc.br

<sup>4</sup> Instituto de Tecnologia de Alimentos – Campinas, morgano@ital.org.br

<sup>5</sup> Faculdade de Engenharia de Alimentos da Universidade Estadual de Campinas, neura@fea.unicamp.br

<sup>6</sup> Instituto de Química da Universidade Estadual de Campinas, marcia@iqm.unicamp.br

models aiming the prediction of protein levels can be obtained, and that reflection techniques are adequate to allow a fast obtain of spectra of the similar types of milling agricultural products and not to create wastes which are harmful to the environment.

## 1. INTRODUÇÃO

Segundo Martens e Naes [5] a análise dos dados obtidos por técnicas espectroscópicas, podem ser realizadas por métodos multivariados, que permitem o estudo de várias espécies presentes, sem importar a existência de variações marcantes nos espectros, ou mesmo a presença de alta correlação entre os mesmos. Estes métodos de análise multivariada vem sendo empregados para o controle de qualidade de produtos alimentícios [1,4]. Em particular o emprego da espectroscopia no infravermelho próximo, tem apresentado resultados satisfatórios no controle de adulterações via análise discriminante [4,7], ou na substituição de métodos tradicionais para a quantificação de determinado parâmetro de qualidade via métodos de regressão multivariada [6,9].

Durante os anos 90 a aplicação da espectroscopia por reflexão difusa no infravermelho médio com transformada de Fourier (DRIFTS) na determinação quantitativa de fibra e proteína foi amplamente explorada. A grande maioria dos trabalhos publicados tratam da comparação entre o desempenho da técnica consagrada NIRR com a recente aplicação da técnica DRIFTS para esta finalidade como por exemplo da pesquisa publicada por Reeves III *et alii* [8].

Métodos mais robustos, que apresentem melhor desempenho, vem sendo cada vez mais estudados, principalmente quando empregam métodos de otimização visando a seleção dos comprimentos de onda (variáveis) mais adequados para a quantificação de determinado analito, sendo o algoritmo genético um dos mais recentemente empregados [3]. O Algoritmo Genético (AG) é uma técnica de busca aleatória direcionada, desenvolvida por Holland [2], capaz de obter a solução ótima global num espaço de busca complexo multi-dimensional. O AG é baseado na evolução natural das espécies, usando operadores inspirados no processo de evolução natural. Estes operadores, conhecidos como operadores genéticos, manipulam indivíduos de uma população, através de gerações, para melhorar (aperfeiçoar) a adaptação (fitness) gradativamente. Os indivíduos numa população, também denominados de cromossomos, são representados por cadeias ("strings") de números binários. A função de avaliação (fitness) estabelece a relação entre o AG e o problema de otimização.

## 2. MATERIAL E MÉTODOS

**2.1 Amostras.** Para a construção dos modelos de regressão foram empregadas 46

amostras de café cru, da variedade arábica, procedentes de diferentes regiões do Brasil, sendo primeiramente homogeneizadas em moinho de facas e peneiradas em peneira de 0,84 mm de tamanho de abertura de poro; e 46 amostras de farinha de trigo *Triticum aestivum L.*, cedidas pela empresa Filler S.A..

**2.2 Método de referência.** O método de referência utilizado para a determinação do teor de proteína total foi o de Kjeldhal da *Association of Official Analytical Chemists*.

**2.3 Espectros no Infravermelho.** Os espectros no infravermelho para as amostras de café foram coletados em um espectrofotômetro NICOLET 520 e das amostras de farinha de trigo num espectrofotômetro NICOLET Magna 550, sendo realizadas 3 réplicas para cada amostra, com resolução de  $4\text{ cm}^{-1}$  e 16 varreduras. Cada conjunto de espectros foi normalizado e tratados com a correção multiplicativa de sinal (MSC).

**2.4 Modelagem dos dados.** Na modelagem dos dados foi empregado o método dos mínimos quadrados parciais (PLS) disponível no ToolBox do ambiente MATLAB. A matriz de espectros de calibração foi construída com 23 amostras de café e 23 amostras de farinha de trigo, ambas em duplicata, totalizando 92 espectros compreendidos entre  $800\text{-}4000\text{ cm}^{-1}$ . A matriz de validação foi construída de forma análoga também totalizando 92 espectros. Para cada conjunto de espectros foi aplicada a correção do espalhamento de luz cujo resultado é ilustrado na figura 1.

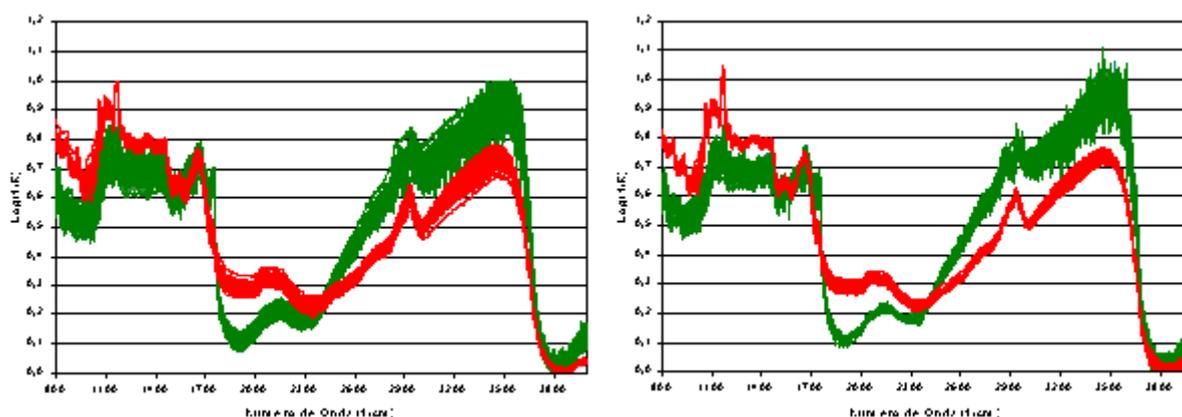


Figura 1 – Conjunto dos espectros do café (verde) e farinha (vermelho) sem correção (à esquerda) e com correção do espalhamento de luz (à direita).

A rotina de algoritmo genético foi desenvolvida em ambiente MATLAB pelo grupo, sendo utilizado como *fitness* o  $(SEV+SEC)/2$ , onde SEV é o erro padrão de validação e SEC o erro padrão de calibração. Os demais parâmetros do GA foram:

Número de cromossomos = 50; Total de variáveis selecionadas para iniciar = 10%,  
Número de iterações = 1000; % de mutação na população = 30%;  
% de mutação de genes = 5%; % de mutação da variáveis latentes = 30%;  
Número máximo de variáveis latentes = 20.

### 3. RESULTADOS E DISCUSÃO

Quatro repetições para a rotina do algoritmo genético foram realizadas, sendo os resultados da otimização apresentados na figura 2, onde F01, F02, F03 e F04 correspondem ao desempenho da *fitness* para cada uma das repetições. Um exemplo da curva de calibração obtida para estes modelos é apresentada na figura 3, onde foram utilizadas 20 variáveis latentes.

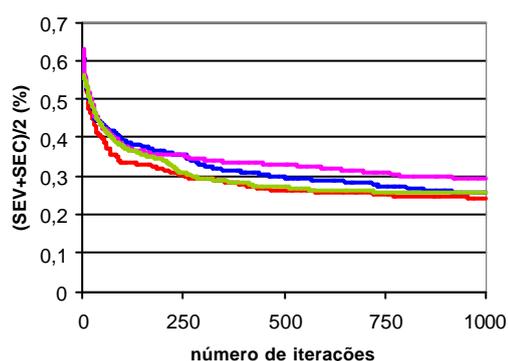


Figura 2 – Variação do fitness em função do número de iterações realizadas para o DRIFTS-PLS-AG

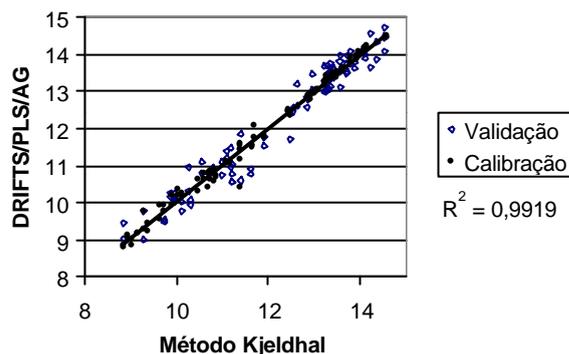


Figura 3 – Curva de calibração para estimar a teor de proteína empregando DRIFTS-PLS-AG.

Estes resultados demonstram o excelente desempenho da modelagem DRIFTS-PLS otimizada com rotinas de AG, mesmo quando são empregadas diferentes matrizes de espectros, viabilizando a realização de modelos de calibração que empreguem diferentes alimentos e que centrem sua performance na determinação de um parâmetro analítico. Além disso estes métodos são ideais para o monitoramento de insumos da indústria de alimentos por não serem destrutivos e não gerarem resíduos nocivos ao ambiente.

### 4. REFERÊNCIAS BIBLIOGRÁFICAS

[1] FEINBERG, M. & BUGNER, E. Chemometrics and food chemistry: Data validation. Anal. Chim. Acta, v. 223, p. 223-235, 1989.

- [2] HOLLAND, J.H. Adaptation in natural and artificial systems. Cambridge:MIT Press, p. 11-147, 1975.
- [3] LEARDI, R. Genetic algorithms in chemometrics and chemistry: a review. J. Chemometrics, v. 15, p. 559-569, 2001.
- [4] MARK,H. Chemometrics in near-infrared spectroscopy. Anal. Chim. Acta, v. 223, p. 75-93, 1989.
- [5] MARTENS, H. & NAES, T. 1989. Multivariate calibration. John Wiley & Sons, New York.
- [6] MOBLEY,P.R.; KOWALSKI,B.R.; WORKMAN JR.,J.J. & BRO,R. Review of chemometrics applied to spectroscopy: 1985-95, part 2. Appl. Spectrosc. Rev., v. 31, n.4, p. 347-368, 1996.
- [7] OSBORNE, B. G.; FEARN, T. & HINDLE, P. H. 1993. Practical NIR spectroscopy with applications in food and beverage analysis. Longman Scientific & Technical, Singapore.
- [8] REEVES III,J.B. Mid- versus near-infrared spectroscopic analysis of diversely treated feedstuffs. J. Near-infrared Spec., v. 7, p. 89-100, 1999.
- [9] WORKMAN JR.,J.J.; MOBLEY,P.R.; KOWALSKI,B.R. & BRO,R. Review of chemometrics applied to spectroscopy: 1985-95, part I. Appl. Spectrosc. Rev., v. 31, n. 1&2, p. 73-124, 1996.