

ANÁLISE DE UMIDADE E PROTEÍNA DE UMA AMOSTRA A PARTIR DE DADOS DE ESPECTROMETRIA NIR USANDO O ALGORITMO ASA-MLR

Pedro Germano Antonino Nunes¹, Edvan Cirino da Silva²
CCHSA-UFPB¹; CCEN-UFPB² pedro@cft.ufpb.br

Área: 02 – Ciência e tecnologia de alimentos

Introdução

Os métodos físicos e químicos de análise de produtos alimentícios apresentam sérias desvantagens como o uso de vidrarias específicas, reagentes químicos, equipamentos de alto custo, mão-de-obra altamente capacitada, elevado tempo para obtenção de resultados, cada método analisa apenas um constituinte, etc. A espectroscopia no infravermelho próximo (NIR – Near InfraRed) tem demonstrado ser um método capaz de suprir muitas dessas desvantagens. Este método vem sendo aplicado com sucesso no controle de qualidade de diversos produtos agroalimentares, tanto de origem animal quanto de origem vegetal, fornecendo informações analíticas sobre constituintes dos alimentos exigidos pelo Ministério da Agricultura (proteínas, gorduras, fibra, matéria seca, cinzas, umidade), e outros não obrigatórios, mas de interesse nutricional como valores energéticos, aminoácidos, digestibilidade dos nutrientes, etc. (Borges et al., 2001). A aceitação dessa técnica vem ganhando importância nas últimas duas décadas em virtude do desenvolvimento de equipamentos mais sensíveis e aplicação de métodos quimiométricos para tratamento dos espectros obtidos. A espectroscopia no infravermelho próximo compreende a radiação eletromagnética cujo comprimento de onda se situa entre 800 e 2500 nm. A interação dessa radiação é empregada para estabelecer uma relação qualitativa/quantitativa com os constituintes da amostra em estudo. As principais vantagens da espectrometria NIR como ferramenta de análise qualitativa e quantitativa são: a técnica é não destrutiva e não invasiva; a manipulação de amostra é mínima (a possibilidade de realizar medidas tanto no estado líquido como sólido, permite a minimização da manipulação prévia da amostra e, com isso, aumenta a rapidez das análises); baixo custo de análise (como não usa reagentes ou outros tipos de materiais para preparo das amostras, o custo de análise é praticamente nulo); a técnica permite a determinação de vários analitos da amostra sem a necessidade de um procedimento analítico para cada um deles separadamente; é possível determinar parâmetros não químicos de uma amostra; a resistência dos materiais utilizados e a ausência de partes móveis no sistema de detecção permitem que esta técnica seja usada no controle de processos em plantas industriais; em muitos campos de aplicação, a exatidão da técnica NIR é comparável a outras técnicas analíticas (Osborn, 1998). A Quimiometria pode ser definida como a parte da química que utiliza métodos matemáticos e estatísticos para planejar ou selecionar experimentos de forma otimizada e para fornecer o máximo de informação química na análise de dados de natureza multivariada (Massart et al., 1988). O sinal analítico (espectro) fornecido por um instrumento NIR não poderá ser utilizado de forma direta. É necessário um procedimento de calibração, que neste caso, é denominada calibração multivariada, pois utiliza mais de um comprimento de onda do espectro. Essa calibração pode ser realizada por métodos que utilizam todo o espectro (p. ex. regressão por mínimos quadrados parciais – PLS) ou por métodos tais como Regressão Linear Múltipla (MLR), que utilizam algoritmos para selecionar as variáveis (comprimentos de onda) mais informativas.

Objetivo

Este trabalho tem como objetivo determinar os teores de umidade e proteína em amostras de farinha de trigo, a partir de dados obtidos por espectrometria no infravermelho próximo (NIR) e utilização do método de seleção de variáveis denominado ASA e calibração multivariada MLR.

Metodologia

O ASA (Algoritmo de busca angular) foi desenvolvido com o intuito de minimizar problemas de correlação e multicolinearidade em calibração multivariada usando MLR. Para isso, esse algoritmo calcula os cossenos dos ângulos entre as variáveis da matriz de respostas instrumentais das amostras de calibração (\mathbf{X}_{cal}). Em seguida, o algoritmo seleciona aquelas variáveis que apresentam os menores cossenos (ou maiores ângulos) entre si. Após o cálculo dos cossenos, o ASA seleciona subconjuntos de variáveis usando o procedimento mínimo-máximo, iniciando a partir de cada variável. Para assegurar que as variáveis finais selecionadas pelo ASA tenham multicolinearidade desprezível, utiliza-se a Análise de Inflação de Variância (Variance Inflation Factors-VIF) como critério para minimizar a multicolinearidade entre as variáveis menos correlacionadas. Descartam-se, então, as variáveis que apresentam um valor de VIF a partir de um determinado limiar. Após este procedimento, os subconjuntos com as variáveis mantidas serão empregados para construir os modelos MLR-VIF. O resultado dos modelos é comparado em termos do erro médio quadrático obtido em um conjunto de validação independente (RMSEV). Para as N_{val} amostras, o RMSEV é definido como:

$$RMSEV = \sqrt{\frac{1}{N_{val}} \sum_{n=1}^{N_{val}} (y_n - \hat{y}_n)^2}$$

onde y_n e \hat{y}_n são os valores de referência e previstos do parâmetro determinado na n^{th} amostra de validação. O subconjunto que produz o menor valor de RMSEV é então escolhido (Nunes, 2008).

Os espectros NIR, no modo de reflectância difusa, foram obtidos para 100 amostras de farinha de trigo, na faixa de 1101 a 2502 nm com resolução de 2 nm. Os teores de umidade e proteína, usados como referências, foram determinados pelos métodos oficiais. Estes dados foram obtidos no site <ftp://ftp.clarkson.edu/pub/hopkepk/Chemdata/Kalivas/> (Kalivas, citado por Forina et al., 2007). Os dados foram pré-processados usando a segunda derivada de Savitzky-Galay com polinômio de 3ª ordem e janela de 11 pontos. O algoritmo SPXY-*Sample set Partitioning based on joint X-Y distances* (Galvão et al., 2005) foi empregado para dividir as amostras em conjuntos de calibração (50 amostras), validação (25 amostras) e previsão (25 amostras). As faixas de concentração para cada conjunto são mostradas na Tabela 1. O modelo de calibração, para cada analito, foi obtido usando o algoritmo ASA, para selecionar as variáveis, e o método MLR para realizar a calibração. Os modelos foram validados com o conjunto de validação e utilizados para determinar a umidade e proteína no conjunto de previsão. Outros algoritmos que realizam seleção de variáveis, tais como Algoritmo genético (AG) e Algoritmo das Projeções Sucessivas (APS), foram usados para comparação dos resultados. Além desses, comparamos os resultados como o algoritmo PLS.

Resultados e Discussão

A Figura 1(a) mostra os espectros NIR das 100 amostras de trigo. Podemos observar que os espectros apresentam um deslocamento da linha de base. Para utilizar estes dados, é necessário remover este deslocamento, neste caso, usamos o método da segunda derivada (Figura 1 (b)). As Tabelas 1 e 2 mostram os valores de RMSEP na determinação dos teores de umidade e proteínas, respectivamente, obtidos pelo algoritmo ASA e demais algoritmos usados. Em termos de erro relativo (RMSEP), todos os algoritmos apresentaram resultados bastante satisfatórios, com erros abaixo de 0,4%. Em relação ao número de variáveis selecionadas, verifica-se que o ASA selecionou apenas 4 variáveis para a umidade e 5 para a proteína, de um total de 681 variáveis. O algoritmo APS selecionou 4 variáveis para a proteína, no entanto, para a umidade o número foi bem maior (14 variáveis). Com relação ao comportamento do algoritmo AG, verifica-se que o mesmo seleciona sempre um número relativamente elevado de variáveis, porém, produzindo modelos com a mesma capacidade

preditiva. Isto significa que todas estas variáveis não portam informação relevante para o modelo de calibração. Na Figura 2 verifica-se a correlação entre valores de referência e previstos pelo modelo MLR-ASA, para o conjunto de previsão (25 amostras), na determinação dos parâmetros estudados. Observa-se uma correlação acima de 0,9, o que indica uma relação muito próxima entre os valores obtidos pelos métodos de referência (oficiais) e os valores obtidos pelo modelo aqui desenvolvido.

Considerações finais

Os resultados obtidos pelo modelo de calibração MLR-ASA, a partir dos dados da espectrometria no Infravermelho Próximo (NIR), são muitos semelhantes aos obtidos pelos métodos oficiais, tanto na determinação da umidade quanto na determinação da proteína da farinha de trigo. Os erros apresentados estão abaixo de 0,4%. O desempenho do algoritmo usado, o ASA, é bastante satisfatório, pois utiliza um número pequeno de variáveis, o que produz modelos mais simples e de mais fácil interpretação. Em consequência das vantagens citadas para a técnica de Infravermelho Próximo (NIR), aliada a métodos quimiométricos eficientes, esta vem sendo, cada vez mais, utilizada na análise de produtos agroindustriais, em substituição aos métodos tradicionais.

Referências

- BORGES, F. M. O.; FERREIRA, W. M.; SALIBA, E. O. S. Espectroscopia de Reflectância no Infravermelho Próximo NIRS. **Revista CFMV – Suplemento Técnico**, n.24, set/out/nov/dez/2001.
- FORINA, M.; LANTERI, S.; CASALE, M. Multivariate calibration: review. **Journal of Chromatography A**, v.1158, p.61, 2007.
- GALVÃO, R.K.H.; ARAÚJO, M.C.U.; JOSÉ, G.E.; PONTES, M.J.C.; SILVA, E.C.; SALDANHA, T.C.B.; A method for calibration and validation subset partitioning. **Talanta**, v. 67, p. 736, 2005.
- MASSART, D.L.; VANDEGISTE, B.G.M.; DEMING, S.N.; MICHOTTE, Y.; KAUFMAN, L. *Chemometrics: a textbook*, Elsevier, Amsterdam, 1988.
- NUNES, P. G. A. Uma nova técnica para seleção de variáveis em calibração multivariada aplicada às espectrometrias UV-VIS e NIR. Tese de Doutorado, Universidade Federal da Paraíba, João Pessoa, 2008.
- OSBORN, B. G. *Near-Infrared Spectroscopy in Food Analysis*. Ed. Robert A. Meyers, John Wiley & Sons Ltd., Australia, 1998.

Tabela 1. Faixas de concentração dos conjuntos de dados NIR de farinha de trigo.

Conjunto	Propriedade	
	Proteína [%]	Umidade [%]
calibração	7,75 – 14,28	12,45 – 17,36
validação	10,36 – 12,96	12,69 – 16,59
previsão	10,59 – 14,02	12,70 – 16,94

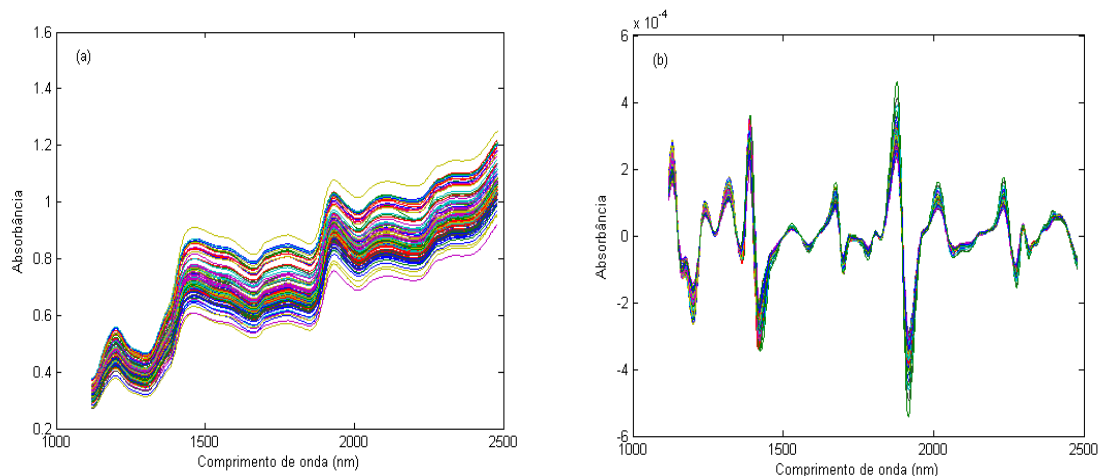


Figura 1. Espectros NIR de farinha de trigo: (a) espectros originais; (b) espectros derivativos.

Tabela 2. Valores de RMSEP [%] na determinação de umidade de farinha de trigo.

Modelo	Nº de variáveis selecionadas	RMSEP _{prev}
MLR-ASA-VIF	4	0,27
MLR-APS	14	0,28
MLR-AG	22	0,28
PLS	4*	0,27

* Variáveis latentes

Tabela 3. Valores de RMSEP [%] na determinação de proteína de farinha de trigo.

Modelo	Nº de variáveis selecionadas	RMSEP _{prev}
MLR-ASA-VIF	5	0,35
MLR-APS	4	0,32
MLR-AG	18	0,32
PLS	10*	0,28

* Variáveis latentes

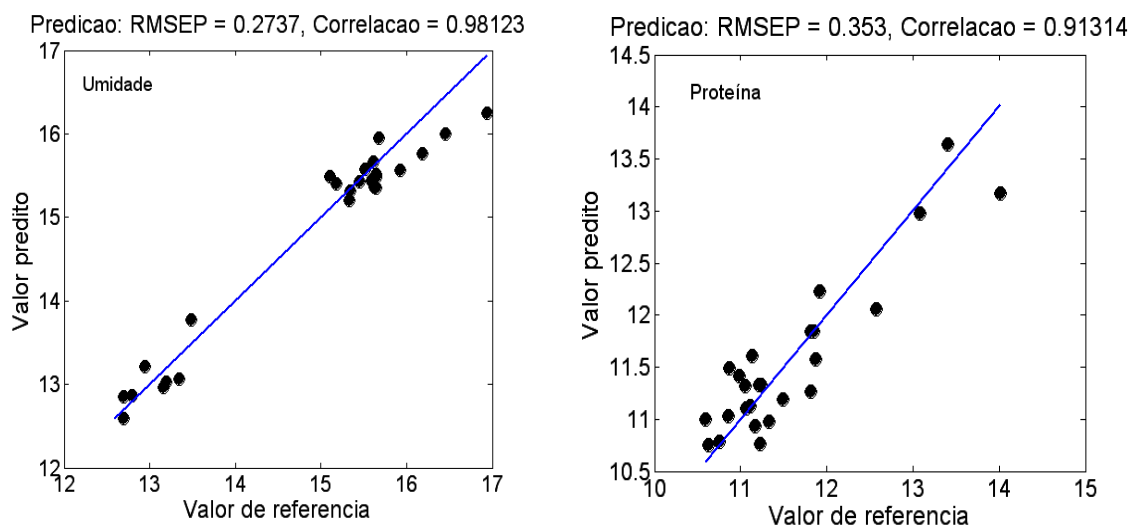


Figura 2. Valores de referência e previstos pelo modelo MLR-ASA para umidade e proteína do trigo.