

DETERMINAÇÃO SIMULTÂNEA DOS TEORES DE CINZA E PROTEÍNA EM FARINHA DE TRIGO EMPREGANDO NIR-PLS E DRIFT-PLS¹

Marco Flôres FERRÃO^{2,*}, Claudia Wollmann CARVALHO²,

Edson Irineu MÜLLER², Celso Ulysses DAVANZO³

RESUMO

As técnicas de espectroscopia por reflexão no infravermelho próximo (NIRRS) e por reflexão difusa no infravermelho médio com transformada de Fourier (DRIFTS) foram empregadas com o método de regressão multivariada por mínimos quadrados parciais (PLS) para a determinação simultânea dos teores de proteína e cinza em amostras de farinha de trigo da variedade *Triticum aestivum* L. Foram coletados espectros no infravermelho em duplicata de 100 amostras, empregando-se acessórios de reflexão difusa. Os teores de proteína (8,85-13,23%) e cinza (0,330-1,287%), empregados como referência, foram determinados pelo método Kjeldhal e método gravimétrico, respectivamente. Os dados espectrais foram utilizados no formato $\log(1/R)$, bem como suas derivadas de primeira e segunda ordem, sendo pré-processados usando-se os dados centrados na média (MC) ou escalados pela variância (VS) ou ambos. Cinquenta e cinco amostras foram usadas para calibração e 45 para validação dos modelos, adotando-se como critério de construção os valores mínimos do erro padrão de calibração (SEC) e do erro padrão de validação (SEV). Estes valores foram inferiores a 0,33% para proteína e a 0,07% para cinza. Os métodos desenvolvidos apresentam como vantagens a não agressão ao ambiente, bem como permitem uma determinação direta, simultânea, rápida e não destrutiva dos teores de proteína e cinza em amostras de farinha de trigo.

Palavras-chave: cinza; proteína; farinha de trigo; regressão multivariada; espectroscopia no infravermelho; quimiometria.

SUMMARY

SIMULTANEOUS DETERMINATION OF ASH CONTENT AND PROTEIN IN WHEAT FLOUR USING INFRARED REFLECTION TECHNIQUES AND PARTIAL LEAST-SQUARES REGRESSION (PLS). Partial Least Square (PLS) multivariate calibration associated to Near Infrared Reflection Spectroscopy (NIRRS) or Diffuse Reflectance Infrared Fourier Transform Spectroscopy (DRIFTS) were used to establish methods for simultaneous determination of protein and ash content in commercial wheat flour samples of *Triticum aestivum* L. Duplicate spectra of 100 samples with protein content between 8.85-13.23% (Kjeldahl method) and ash content between 0.330-1.287% (gravimetric method) were employed to build calibration methods. The spectra were used in $\log(1/R)$ format as well as in the first and second derivatives and they were also pre-processed by mean centering (MC) or variance scaling (VS) or both. The samples were divided in two sets, one with 55 samples and the other with 45 used as calibration and validation sets, respectively. Calibration methods were obtained by using minimum values of root mean square error of calibration (RMSEC) and root mean square error of validation (RMSEV). Methods of calibration with RMSEC and RMSEV values lower than 0.33 in the determination of protein content and lower than 0.07 in the determination of ash content. The proposed methods are environmentally non aggressive as they avoid the use of reagents and do not generate hazardous waste. Furthermore, they allow fast and non destructive simultaneous and direct determination of ash and protein content in wheat flour samples.

Keywords: infrared spectroscopy; PLS regression; wheat flour; protein content; ash content; NIR spectroscopy; chemometrics.

1 - INTRODUÇÃO

A espectroscopia por reflexão no infravermelho médio para a análise qualitativa e quantitativa de alimentos vem sendo amplamente utilizada desde o início dos anos 90. WILSON [35], em 1990, introduz as potencialidades da região do infravermelho médio e as técnicas de reflexão: reflexão total atenuada (ATR), detecção fotoacústica e reflexão difusa (DRIFTS), para análise de alimentos e sua utilização no controle industrial.

Em 1991, VAN DE VOORT & ISMAIL [30] revelam as vantagens do emprego da espectroscopia por reflexão total atenuada (ATR) na análise quantitativa de leite.

VAN DE VOORT [29], no ano seguinte, descreve as principais aplicações da espectroscopia ATR na análise de leite, carne, óleos, gorduras, manteiga, margarina, leite condensado e frutas, avaliando tanto aspectos quantitativos quanto qualitativos, como a certificação da qualidade. Nos últimos anos, REEVES III & ZAPF [24] empregam a espectroscopia por reflexão difusa no infravermelho com transformada de Fourier (DRIFTS) para discriminar vários ingredientes alimentares.

Em 1999, WILSON & TAPP [34] trazem uma revisão para o emprego da espectroscopia no infravermelho médio na análise de alimentos, com a reflexão difusa sendo principalmente utilizada na certificação dos alimentos e por reflexão total atenuada na determinação de açúcares.

A quantificação simultânea dos teores de cinza e de proteína em farinha de trigo, utilizando espectroscopia por reflexão no infravermelho médio, foi recentemente publicada por CARVALHO et al. [4].

Todavia, o emprego da espectroscopia no infravermelho próximo (NIRRS) é consagrada desde o início da década de sessenta quando HART, NORRIS & GOLUBIC [9], apresentam um método para determinar umidade em sementes baseado na análise espectrofotométrica

¹ Recebido para publicação em 25/05/2002. Aceito para publicação em 22/06/2004 (000900).

² Departamento de Química e Física, Universidade de Santa Cruz do Sul, Caixa Postal 188, CEP: 96815-900, Santa Cruz do Sul - RS, Brasil. E-mail: ferrao@unisc.br

³ Instituto de Química - Universidade Estadual de Campinas, Caixa Postal 6154, CEP: 13083-970, Campinas - SP, Brasil. E-mail: celso@iqm.unicamp.br

* A quem a correspondência deve ser enviada.

no infravermelho próximo. Diversas outras aplicações são descritas por WATSON [31], onde o infravermelho próximo é empregado na análise de produtos agrícolas.

A aplicação da espectroscopia no infravermelho próximo na análise de alimentos foi inicialmente restrita ao emprego da reflexão difusa (NIRR), sendo amplamente empregada na análise de grãos, farinhas e produtos industrializados [10, 22, 28]. Entre estas aplicações tem destaque o emprego da espectroscopia NIRR na determinação de proteína em substituição ao método Kjeldahl, que, segundo SCHUSTER [25], apesar de ser bastante preciso apresenta procedimento lento, além de gerar resíduos.

No início dos anos 80, OSBORNE [15] apresenta umas das primeiras retrospectivas da espectroscopia por reflexão no infravermelho próximo discorrendo sobre suas principais aplicações e, particularmente, enfocando o trigo, a farinha e os produtos industrializados. Ainda neste período são apresentadas vantagens na utilização do NIRR, principalmente no que se refere ao baixo tempo de análise, do emprego de técnicas de reflexão no infravermelho próximo no controle de qualidade da farinha de trigo [16, 17, 19].

Em 1984, DAVIES [5] reporta os principais avanços da técnica NIRR discutindo os novos métodos de calibração e regressão multivariada, de seleção de comprimentos de onda para a calibração e as novas formas de utilização da espectroscopia NIRR na indústria de alimentos. MCSHANE [12], em 1989 reforça a perspectiva de utilização da espectroscopia NIRR no controle de processos na indústria alimentícia.

Já, nos anos 90, os princípios das aplicações da técnica NIRR foram publicados por WILLIAMS & STEVENSON [33], onde também apresentam algumas das aplicações típicas desta técnica na análise de alimentos. Recentemente, ARCHIBALD et al. [1, 2] revisitam a espectroscopia NIRR, comparando-a com a espectroscopia Raman, e descrevem métodos de análise e de regressão multivariada para a determinação de fibra em cereais.

Este acentuado crescimento da utilização da espectroscopia no infravermelho médio nos últimos anos e o contínuo desenvolvimento de métodos no infravermelho próximo se devem, no nosso entendimento, à procura de métodos analíticos rápidos e limpos. Neste caso evitando-se o emprego de reagentes agressivos ou a geração de resíduos danosos ao ambiente. O uso da análise por reflexão difusa no infravermelho médio e próximo pode satisfazer os dois quesitos acima, pois requerem pouco ou nenhum tratamento das amostras. Entretanto, é imperioso reconhecer que a análise multivariada de dados foi uma ferramenta valiosa que contribuiu para o emprego mais extensivo daquelas técnicas. Neste contexto, a aparição de programas computacionais disponíveis comercialmente, que podem ser utilizados sem a necessidade de conhecimento mais aprofundado do tema, impulsionou o estudo de novos métodos analíticos usando a região do infravermelho.

Este artigo tem como objetivo ilustrar como novos métodos de análise podem ser desenvolvidos utilizando a reflexão difusa no infravermelho médio e próximo em combinação com a análise multivariada. Para isso, escolhemos como objeto de estudo a farinha de trigo por ser uma matéria-prima muito utilizada na indústria de alimentos. Dentre as muitas análises necessárias para o seu emprego adequado selecionamos duas que são o teor de proteína e o de cinzas, cujos métodos oficiais de análise são demorados e, no caso do teor de proteína o método faz gerar muito resíduo perigoso. Mostraremos, ainda, que os padrões poderão ser formulados pelo próprio usuário, pois usaremos amostras de farinha adquiridas no comércio por uma indústria alimentícia. Além disso, utilizaremos um programa computacional fornecido pela própria empresa que fabrica um dos equipamentos utilizados para obtenção dos espectros. Finalmente, segundo nosso conhecimento, é a primeira vez que há uma comparação entre os resultados do infravermelho médio e próximo avaliando os teores de cinzas e proteína em amostras de farinha de trigo.

1.1 - Reflexão difusa

A reflexão difusa é principalmente observada em amostras sólidas pulverizadas, sendo sua aplicação prática muito difundida na literatura, podendo ser utilizada em equipamentos que operam na região do infravermelho próximo [18, 21, 22, 26, 27, 32], associada a sigla NIRS, *Near Infrared Reflectance Spectroscopy*, ou simplesmente NIRRS. No final dos anos 70 e início dos 80 foi demonstrada a utilidade deste fenômeno quando acessórios de reflexão difusa foram acoplados com espectrômetros interferométricos com transformada de Fourier, sendo conhecida a técnica resultante pela sigla DRIFTS, *Diffuse Reflection Infrared Fourier Transform Spectroscopy* [7].

A reflexão difusa é observada quando uma luz incide em uma matriz descontínua penetra na amostra (tipo amostra em pó, papel) e reflete contendo informações espectrais. O caminho percorrido pela luz no interior da matriz pode ser considerado aleatório devido a múltiplas reflexões, algumas das quais após percorrer o interior de algumas partículas que constituem a amostra. Desta forma, a luz refletida pode ser atenuada por absorção e o espectro resultante é similar ao obtido através da técnica no infravermelho por transmissão utilizando KBr. Uma importante diferença entre a transmissão e a reflexão é devida ao diferente caminho óptico percorrido pela luz. Enquanto na transmissão o caminho óptico é constante para todo número de onda, na reflexão o caminho pode ser variável. É bem sabido que, em regiões do espectro, onde a amostra absorve fracamente, a luz penetra mais profundamente na matriz, enquanto o contrário acontece onde há forte absorção. Portanto, ao se comparar o espectro obtido por transmissão (pastilha de KBr) com o obtido por reflexão, as intensidades relativas das bandas serão diferentes. Por exemplo, as bandas fracas no espectro por transmissão aparecerão mais fortes na reflexão.

É importante salientar que num experimento de reflexão difusa será também observada a reflexão especular, sendo mais importante aquela que ocorre na interface ar/superfície da matriz, mas também ocorrendo nas faces das partículas, mas cujo efeito pode ser minimizado pela redução do tamanho da partícula. A reflexão especular é de maior intensidade na região onde a amostra apresenta forte absorção e, neste caso pode ocorrer severas distorções no espectro obtido [6]. Deve-se ressaltar a diferença marcante entre infravermelho médio e próximo. No infravermelho próximo as distorções nos espectros são quase imperceptíveis, enquanto que no infravermelho médio as distorções são mais freqüentes. Uma maneira de contornar esse efeito não desejado é diluindo a amostra numa matriz não absorvente, como KBr [13, 14].

1.2 - Regressão multivariada

1.2.1 - Quimiometria

A Quimiometria pode ser entendida como uma área multidisciplinar do conhecimento, onde a matemática e a estatística são utilizadas para análise de dados relativos aos processos químicos, de natureza multivariada, como por exemplo informações provenientes da espectroscopia, cromatografia, etc.

As técnicas quimiométricas visam o delineamento e análise de experimentos, tendo como objetivo utilizar de forma eficiente, correta e econômica os recursos disponíveis.

Um grande problema, muitas vezes enfrentado pelos químicos, é a necessidade de conhecer as propriedades e a composição de sistemas, que apresentam difícil solução pelos métodos convencionais de laboratório [20]. Segundo BROWN [3], estas propriedades podem ser obtidas de forma indireta, a partir de outras variáveis descritas como variáveis latentes. Desta forma, um dos objetivos dos métodos quimiométricos é o de encontrar as variáveis latentes, bem como as relações existentes entre os dados físico-químicos e o sistema em questão.

1.2.2 - Regressão por mínimos quadrados parciais (PLS)

Este método foi desenvolvido nos anos 70 por Herman Wold e um excelente guia prático foi escrito por GELADI & KOWALSKI [8]. Como usual em procedimentos analíticos primeiro desenvolve-se um modelo de calibração [23]. Uma matriz \mathbf{X} com \mathbf{m} colunas, correspondendo ao sinal espectral (p. e. absorbância) para cada número de onda, e \mathbf{n} linhas, representando cada amostra (padrão). Uma segunda matriz \mathbf{Y} , é definida com \mathbf{k} colunas, com as informações químicas (p. e. concentrações) e com \mathbf{n} linhas, correspondendo a cada amostra. Utilizando técnicas de análise fatorial a matriz \mathbf{X} é decomposta numa soma de matrizes menores (equação 1).

$$\mathbf{X} = \mathbf{M}_1 + \mathbf{M}_2 + \dots + \mathbf{M}_a + \mathbf{E} \quad (1)$$

Sendo \mathbf{M}_i os componentes principais ou variáveis latentes, com dimensão 1, 2... a , onde a corresponde ao número de fatores (componentes principais ou variáveis latentes) selecionado para truncar a igualdade. \mathbf{E} corresponde a matriz de resíduos, relacionada ao número de fatores.

A matriz \mathbf{Y} é decomposta de maneira análoga. Finalmente, determina-se a correlação entre as variáveis latentes usando os componentes das matrizes menores.

No modelo PLS as decomposições das matrizes \mathbf{X} e \mathbf{Y} são feitas simultaneamente por meio iterativo e se acredita que a informação contida em \mathbf{Y} é transferida para \mathbf{X} .

Após a calibração é possível prever as propriedades de novas amostras a partir de medidas espectrais. Os sinais espectrais são definidos numa nova matriz \mathbf{X}_p , que contém \mathbf{m} colunas mas \mathbf{t} linhas, correspondendo as \mathbf{t} novas amostras. Usando a correlação construída na fase de calibração as propriedades desconhecidas são coletadas numa nova matriz \mathbf{Y}_p .

1.2.3 - Critérios para avaliação de modelos de regressão

Um modelo é considerado eficiente se ele descreve da melhor forma possível uma situação real, considerando o maior número possível de variações e, quanto maior o número de fatores, menor o desvio da reta de calibração, porém, o aumento de fatores também aumenta o ruído e os erros de modelagem.

O número ótimo de fatores a ser usado num modelo será o número de fatores que corresponda ao ponto no qual a diminuição do erro (produzida pelo aumento da complexidade do modelo) é compensada pelo aumento do erro de superavaliação.

O número apropriado de fatores, ou variáveis latentes (VL), na calibração PLS pode ser determinada pelo PRESS (*predicted residual error sum of squares*), que é a soma dos desvios quadrados para os valores estimados em relação aos valores reais (equação 2).

$$\text{PRESS} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2)$$

Nesta equação, y_i é o valor de referência para a i ésima amostra, e \hat{y}_i é a previsão para o valor desta amostra. Um decréscimo no PRESS indica um modelo "robusto", ou seja, com habilidade de previsão. O modelo otimizado é formado usando o número mínimo de fatores que tem um PRESS com um erro padrão mínimo.

Outra medida que é empregada para avaliar a habilidade de previsão de um modelo é o SEC (*standard error of calibration*), que apresenta informações semelhantes aos do PRESS, sendo expressa conforme equação 3, onde n representa o número de amostras utilizadas.

$$SEC = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (3)$$

Na avaliação do conjunto de amostras de validação podemos usar a expressão 3, onde se substitui os valores para as amostras deste conjunto, denominando agora de SEV (*standard error of validation*). Este critério é muito útil quando se deseja avaliar se o modelo tende a ser superdimensionar o número de variáveis latentes a serem empregados. Neste trabalho este critério será usado para se obter o melhor modelo.

2 - MATERIAIS E MÉTODOS

2.1 - Amostras

Amostras de farinha de trigo (*Triticum aestivum* L.) procedentes de diferentes moinhos foram coletadas entre agosto de 1998 e abril de 1999, na empresa Filler S.A., Santa Cruz do Sul, RS. Foram selecionadas 104 amostras para elaborar uma rotina que possibilitasse estimar os parâmetros de qualidade em amostras de farinha de trigo que continham entre 8,85-13,23% de teor de proteína e entre 0,330-1,287% de teor de cinza.

2.2 - Métodos de referência

Os métodos empregados como referência foram realizados segundo os métodos oficiais descritos na *Tabela 1*, bem como o número de réplicas realizadas para cada parâmetro.

TABELA 1. Métodos empregados e o número de replicatas.

Parâmetro	Método	Réplicas
umidade	44-15A da AACC ^a (1981)	1
proteína	2055 da AOAC ^b (1984)	3
cinzas	08-03 da AACC ^a (1984)	3

a - AACC - American Association of Cereal Chemists
b - AOAC - Association of Official Analytical Chemists

Para os métodos onde foram realizadas replicatas, os valores de referência foram considerados como a média. Os teores de proteína e cinzas encontrados foram usados para construção das matrizes **Y**.

2.3 - Obtenção dos espectros

Os espectros no infravermelho foram obtidos em duplicata utilizando os parâmetros de análise descritos na *Tabela 2*. O espectro de referência no DRIFTS foi obtido utilizando o acessório de reflexão difusa e no lugar da amostra empregou-se uma peça maciça construída em alumínio comercial, enquanto que no NIRRS empregou-se politetrafluoretileno comercial. As superfícies de ambas as peças foram cuidadosamente lixadas de forma que as mesmas se tornassem foscas. A quantidade de massa de farinha, a ser colocada no acessório, foi determinada experimentalmente através de uma série de espectros.

TABELA 2. Instrumentação e regiões empregadas na aquisição dos espectros.

Espectrofotômetro	Nicolet Magna 550	Bomem DA-08
Região Espectral	450-4000 cm ⁻¹	5555-10000 cm ⁻¹
Resolução	4 cm ⁻¹	4 cm ⁻¹
Laser de Referência	15798 cm ⁻¹	15798 cm ⁻¹
Apodização	Happ-Genzel	Bartlet
Número de Pontos	1842	2307
Detetor	DTGS KBr	InSb
Divisor de Feixe	KBr	Quartzo
Acessório	Reflexão Difusa Easydif [®] Pike	Reflexão Difusa Spectra-Tech

2.4 - Modelagem dos dados

Todos os modelos de regressão apresentados neste trabalho foram desenvolvidos utilizando o aplicativo *Turbo Quant Analyst v.1.1a* da Nicolet, sendo baseados no método dos mínimos quadrados parciais (PLS). Com base nos resultados obtidos para a validação dos dados foram inicialmente retiradas 4 amostras, uma vez que seus espectros apresentaram anomalias que podem gerar problemas nas futuras modelagens a que serão submetidos os espectros. As 100 amostras restantes foram divididas em dois conjuntos: 55 para a construção dos modelos de regressão e 45 para a validação. Os dados espectrais usados para construção das matrizes **X** foram de 3 tipos, log (1/R) e suas respectivas primeira e segunda derivadas. Além disso, 3 tipos de pré-processamentos foram testados: i) centrado na média (MC), onde a matriz **X** teve para cada coluna calculada a média e para cada valor da coluna foi subtraído a média; ii) escalonado pela variância (VS), isto é, para cada coluna foi calculada a variância e cada um dos valores da coluna foi dividido pela variância e iii) auto escalonamento (MS/Vs), onde os procedimentos i) seguido de ii) foram efetuados. Estes 3 tipos de processamento também foram aplicados as matrizes **Y**. Finalmente foram retiradas as regiões espectrais onde se evidenciou pouca informação, bem como aquelas onde o ruído no espectro devido ao vapor de água e ao dióxido de carbono estavam presentes. As regiões usadas para modelagem foram 800-1800cm⁻¹ e 2500-3500cm⁻¹ para o infravermelho médio e 5600-8800cm⁻¹ para o infravermelho próximo.

3 - RESULTADOS E DISCUSSÃO

Os modelos obtidos para o teor de proteína, utilizando-se os dados na forma de log (1/R), primeira e segunda derivadas, empregando os dados DRIFTS e NIRRS, estão nas *Tabelas 3, 4 e 5*, respectivamente. Nelas são apresentados os valores do coeficiente de correlação (R²), do erro padrão de calibração (SEC), do erro padrão de validação (SEV) e o número de variáveis latentes (ou fatores) utilizado, correspondente ao menor PRESS.

Os três diferentes modelos apresentados em cada uma das tabelas diferem nos pré-processamentos empregados. Para a *Tabela 3* os modelos DRIFTS-A1 e NIRRS-A1 os dados foram apenas centrados na média

(MC); já para o modelos DRIFTS-B1 e NIRRS-B1 os dados foram auto escalonados, isto é, centrados na média e escalonados pela variância simultaneamente (MC/VS) e para os modelos DRIFTS-C1 e NIRRS-C1 os dados foram somente escalonados pela variância (VS). Vale as mesmas convenções para as outras tabelas.

TABELA 3. Resultados da modelagem empregando os espectros DRIFTS e NIRRS para o teor de proteína.

Técnica	DRIFTS			NIRRS		
	A1	B1	C1	A1	B1	C1
Modelo						
R ²	0,938	0,935	0,915	0,928	0,909	0,907
Variáveis Latentes	9	9	10	7	7	7
SEC	0,247	0,253	0,290	0,265	0,299	0,302
SEV	0,284	0,276	0,238	0,339	0,406	0,391
Centrados na média	sim	sim	não	sim	sim	não
Escalonados pela variância	não	sim	sim	não	sim	sim

TABELA 4. Resultados da modelagem empregando os espectros DRIFTS e NIRRS com primeira derivada para o teor de proteína.

Técnica	DRIFTS			NIRRS		
	D1	E1	F1	D1	E1	F1
Modelo						
R ²	0,868	0,925	0,914	0,997	0,998	0,998
Variáveis Latentes	4	3	4	7	6	7
SEC	0,360	0,271	0,291	0,049	0,035	0,039
SEV	0,283	0,324	0,454	0,446	0,426	0,435
Centrados na média	sim	sim	não	sim	sim	não
Escalonados pela variância	não	sim	sim	não	sim	sim

TABELA 5. Resultados da modelagem empregando os espectros DRIFTS e NIRRS com segunda derivada para o teor de proteína.

Técnica	DRIFTS			NIRRS		
	G1	H1	I1	G1	H1	I1
Modelo						
R ²	0,996	0,921	0,845	0,997	0,998	0,998
Variáveis Latentes	9	2	3	7	6	7
SEC	0,059	0,279	0,405	0,049	0,035	0,039
SEV	0,415	0,431	0,564	0,446	0,426	0,435
Centrados na média	sim	sim	não	sim	sim	não
Escalonados pela variância	não	sim	sim	não	sim	sim

A seleção dos modelos melhores ajustados deve considerar os valores de R², SEC e SEV. Primeiramente, deve-se escolher os modelos com os maiores coeficientes de correlação. Por isso, os modelos D1 (Tabela 4) e I1 (Tabela 5), ambos para a técnica DRIFTS, não serão considerados por apresentarem R² < 0,9. O próximo quesito para análise do ajuste do modelo é o valor de SEC. A expressão 2.2 indica se tratar de um erro padrão para o modelo de calibração. O SEC pode ser analisado em termos de erros percentuais se for considerado o intervalo de 8,85-13,23% para o teor de proteína utilizado neste trabalho. Como exemplo, adotar-se-

á para cálculo o maior valor de SEC entre os modelos em consideração, isto é, o valor 0,302 para o modelo C1 para NIRRS. Aplicando-se este SEC para o intervalo de teor de proteína, encontra-se, aproximadamente, o erro percentual no intervalo 3-2%. Embora o erro percentual do método de referência seja menor que estes valores, pode-se considerar todos os modelos bem ajustados, principalmente aqueles que usam NIRRS, pois os valores analisados representam o maior erro possível.

O próximo critério de análise do modelo é o valor do SEV, que é o erro padrão de validação obtidos com as amostras deixadas fora do modelo de calibração. Para que um modelo esteja adequadamente ajustado o conjunto de validação deve produzir um valor de SEV similar ao SEC. Valores de SEV muito maior que o SEC indicam modelos sobre-ajustados, isto é, a regressão encontrada considera dados que não estão realmente correlacionados, como por exemplo, ruídos, erros sistemáticos, etc. Se for considerada a razão SEC/SEV no intervalo 0,5-1, como critério de similaridade se notará que apenas os modelos A1, B1 (Tabela 3), E1, F1 (Tabela 4) e H1 (Tabela 5) para DRIFTS e A1, B1 e C1 (Tabela 3) para NIRRS apresentam a relação SEC/SEV no intervalo de similaridade e podem ser considerados bem ajustados. É interessante notar que alguns modelos NIRRS (D1, E1, F1, G1, H1 e I1) com valores muito baixo de SEC, portanto modelos de calibração muito bem ajustados, devem ser considerados modelos sobre-ajustados pelos valores de SEV e não podem ser escolhidos como bons modelos de análise. A razão para isto será discutida mais adiante em conexão com os modelos para teores de cinzas.

Os modelos obtidos para o teor de cinzas, utilizando-se os dados na forma de log (1/R), primeira e segunda derivadas, empregando os dados DRIFTS e NIRRS, estão nas Tabelas 6, 7 e 8, respectivamente. A denominação dos modelos seguem os mesmos critérios adotados para as Tabelas 3, 4 e 5. Também serão adotados os mesmos critérios usados na análise do teor de proteínas para verificação do ajuste dos modelos do teor de cinzas. Dessa forma, somente serão analisados os modelos com coeficientes de correlação R²>0,9, isto é, os modelos B1, E1 e F1 para DRIFTS e todos os modelos NIRRS. Valores de SEC maiores que 0,05 produzem erros percentuais no intervalo de teores de cinzas (0,330-1,287%) entre 15-4%, que poderiam ser adotados como os maiores aceitáveis. Por este critério seriam tomados como modelos bem ajustados, B1 para DRIFTS e para NIRRS apenas seriam excluídos B1 e C1. Introduzindo-se o quesito similaridade dos valores SEC e SEV para verificação de ajuste dos modelos, somente poderia ser selecionado o modelo B1 para DRIFTS.

Para resumir as discussões precedentes fazemos notar que com 100 amostras de farinha de trigo, separadas em 55 para construção de modelos de calibração e 45 para testes de validação dos modelos obtidos, pudemos construir bons ou excelentes modelos de análise. Todos os modelos foram selecionados pelos valores

mínimos de PRESS, SEC e SEV e pela similaridade destes dois últimos valores. Obtivemos valores de SEC menores que 0,3 para teor de proteínas e menor que 0,05 para teor de cinzas. Estes valores representam erros de análise aceitáveis quando comparados com os métodos de referência.

TABELA 6. Resultados da modelagem empregando os espectros DRIFTS e NIRRS para o teor de cinza.

Técnica	DRIFTS			NIRRS		
	A1	B1	C1	A1	B1	C1
Modelo						
R ²	0,854	0,954	0,825	0,993	0,918	0,913
Variáveis Latentes	7	10	7	12	7	7
SEC	0,082	0,046	0,091	0,017	0,062	0,063
SEV	0,067	0,044	0,052	0,060	0,085	0,078
Centrados na média	sim	sim	não	sim	sim	não
Escalados pela variância	não	sim	sim	não	sim	sim

TABELA 7. Resultados da modelagem empregando os espectros DRIFTS e NIRRS com primeira derivada para o teor de cinza.

Técnica	DRIFTS			NIRRS		
	D1	E1	F1	D1	E1	F1
Modelo						
R ²	0,827	0,913	0,934	0,995	0,997	0,997
Variáveis Latentes	4	3	4	6	5	6
SEC	0,089	0,063	0,056	0,014	0,011	0,010
SEV	0,054	0,055	0,061	0,079	0,075	0,076
Centrados na média	sim	sim	não	sim	sim	não
Escalados pela variância	não	sim	sim	não	sim	sim

TABELA 8. Resultados da modelagem empregando os espectros DRIFTS e NIRRS com segunda derivada para o teor de cinza.

Técnica	DRIFTS			NIRRS		
	G1	H1	I1	G1	H1	I1
Modelo						
R ²	0,797	0,883	0,886	0,995	0,997	0,997
Variáveis Latentes	4	2	3	6	5	6
SEC	0,097	0,073	0,073	0,015	0,011	0,012
SEV	0,085	0,077	0,083	0,090	0,084	0,085
Centrados na média	sim	sim	não	sim	sim	não
Escalados pela variância	não	sim	sim	não	sim	sim

Para teor de proteína foram gerados 18 modelos (9 DRIFTS e 9 NIRRS), sendo encontrados 8 modelos igualmente válidos, sendo 5 deles usando dados DRIFTS e 3 NIRRS. Dois modelos para DRIFTS foram obtidos com espectros na forma de $\log(1/R)$, sendo um deles com dados centrados na média (Tabela 3, A1) e um com dados auto-escalados (Tabela 3, B1). Dois outros modelos foram obtidos com a primeira derivada de $\log(1/R)$ com dados auto-escalados (Tabela 4, E1) e com dados escalados na variância (VS). O quinto modelo DRIFTS foi obtido com a segunda derivada de $\log(1/R)$ e com dados auto-escalados. Para a técnica NIRRS os 3 modelos foram obtidos com espectros na forma $\log(1/R)$,

porém com dados centrados na média (CM), auto-escalados e escalados pela variância (VS).

Para teor de cinzas também foram gerados 18 modelos, mas apenas um foi encontrado com similaridade dos valores de SEC e SEV foi encontrado para a técnica DRIFTS, com espectros na forma $\log(1/R)$ e com dados auto-escalados (Tabela 6, B1). É importante ressaltar, no entanto, que se os rigorosos critérios usados na seleção dos modelos forem abrandados os modelos NIRRS com dados $\log(1/R)$ (Tabela 5, B1 e C1). Estes modelos apresentam valores de SEC no intervalo 0,06-0,07 e satisfazem o critério de similaridade entre SEC e SEV. Aqueles modelos também podem ser considerados razoáveis, especialmente para valores mais elevados de teor de cinzas.

Dos resultados relatados acima ressaltamos que todos os modelos válidos foram obtidos com espectros centrados na média ou escalados na variância ou auto-escalados. Estas correções são necessárias, pois os equipamentos FT-IR comportam-se como espectrofotômetros de feixe único, podendo produzir espectros deslocados horizontalmente no eixo de intensidade. Além disso, a falta de reprodutibilidade no preenchimento do porta-amostra também pode produzir variações de mesma natureza na intensidade. Em poucas palavras, os pré-tratamentos indicam que não há necessidade de severo controle de reprodutibilidade dos espectros, o que é vantajoso pois não requer um operador especializado.

Julgamos interessante comentarmos os resultados obtidos com a técnica NIRRS, pois podem indicar a necessidade de outros caminhos para melhorar os modelos de calibração.

Para todos os modelos com a técnica NIRRS obtivemos $R^2 > 0,9$. Com exceção dos modelos B1 e C1, para o teor de cinzas, todos os outros modelos obtidos apresentaram valores de SEC baixos ou muito baixos. Os valores de R^2 e SEC indicaram que obtivemos excelentes curvas de calibração. No entanto, com exceção dos modelos A1, B1 e C1 para proteínas e B1 e C1 para teor de cinzas, todos os outros apresentaram valores de SEV muito diferentes do SEC, apontando para modelos sobre ajustados e, portanto, inaceitáveis para utilização na análise de cinzas e proteínas em farinha de trigo. Para o teor de proteína os modelos inaceitáveis foram obtidos quando se empregou sinais correspondentes à 1ª e 2ª derivada de $\log(1/R)$. Este tratamento do dado espectral introduz ruído maior que o do sinal original e a grande diferença entre SEC e SEV parece indicar que na etapa de calibração o ruído introduzido está sendo considerado, isto é, sinais não correlacionados com a propriedade em estudo afetam o modelo de calibração. Para o teor de cinzas mesmo os espectros no formato $\log(1/R)$ produziram modelos aceitáveis, mas produzindo erros maiores que a técnica DRIFTS. Vale lembrar que o espectro no infravermelho próximo não deve conter informações sobre matéria que pode produzir cinzas, como por exemplo carbonatos, sílica, silicatos, etc. Por isso, a técnica NIRRS tende a produzir modelos sobre

ajustados mesmo sem a obtenção da derivada do sinal. Portanto, novamente temos a indicação de que a curva de calibração é ajustada com sinais não correlacionados com a propriedade considerada.

Os fatos analisados acima podem facilmente ser notados para a técnica DRIFTS, especialmente para os modelos onde as derivadas dos sinais foram consideradas e o aumento do ruído poderia também justificar a construção de modelos inaceitáveis.

A discussão dos modelos inadequados obtidos com as técnicas DRIFTS e NIRRS apontam claramente para a necessidade de outros critérios para construção de modelos de calibração. Por exemplo, IWAMOTO et al. [11] obtiveram um bom modelo de calibração para cinzas em farinha de trigo, porém não usaram toda a região espectral como nos modelos aqui construídos.

Pelas razões discutidas, um cuidadoso estudo de métodos de seleção de variáveis está em andamento e em breve será publicado.

4 – CONCLUSÕES

As técnicas de espectroscopia por reflexão no infravermelho próximo (NIRRS) e por reflexão difusa no infravermelho médio com transformada de Fourier (DRIFTS) foram empregadas em conjunto com o método de regressão multivariado por mínimos quadrados parciais (PLS) para a construção de modelos para determinação simultânea dos teores de proteína e cinza em amostras de farinha de trigo da variedade *Triticum aestivum L.*, como encontradas no comércio. Os dados espectrais para construção dos modelos de calibração foram obtidos de 100 amostras de farinha de trigo com teores de proteína (intervalo 8,85-13,23%) e cinza (intervalo 0,330-1,287%), determinados pelo método Kjeldhal e método gravimétrico, respectivamente. Foi possível construir modelos utilizando-se dados espectrais no formato $\log(1/R)$, bem como suas derivadas de primeira e segunda ordem, sendo necessário que eles fossem pré-processados, usando-se os dados centrados na média (MC) ou escalados pela variância (VS) ou auto-escalados. A avaliação dos modelos obtidos foi efetuada separando-se as amostras e dois conjuntos, sendo um com 55 amostras usadas para calibração e outro com 45 para validação. Adotando-se como critério de construção dos modelos os valores mínimos do erro padrão de calibração (SEC) e do erro padrão de validação (SEV) foi possível selecionar excelentes métodos de análise com valores de SEC e SEV bastante similares. Estes valores foram inferiores a 0,3 para proteína e usando ambas as técnicas. Para teor de cinzas o valor de SEV encontrado para DRIFTS foi da ordem de 0,05, tratando-se de excelente resultado. No entanto, para NIRRS os valores de SEC e de SEV são da ordem de 0,06-0,07, porém também gerando modelos razoavelmente adequados. Para os métodos que não se ajustaram ao critério de similaridade entre SEC e SEV foi feita uma cuidadosa discussão dos fatores envolvidos na sua construção.

Os métodos desenvolvidos apresentam como vantagens a não agressão ao ambiente, por não empregarem reagentes ou gerarem resíduos, bem como permitem uma determinação direta, simultânea, rápida e não destrutiva dos teores de proteína e cinza em amostras de farinha de trigo.

5 – REFERÊNCIAS BIBLIOGRÁFICAS

- [1] ARCHIBALD, D.D.; KAYS, S.E.; HIMMELSBACH, D.S.; BARTON II, F.E. Raman and NIR Spectroscopic Methods for Determination of Total Dietary Fiber in Cereal Foods: A Comparative Study. **Applied Spectroscopy**, v. 52, p. 22-31, 1998.
- [2] ARCHIBALD, D.D.; KAYS, S.E.; HIMMELSBACH, D.S.; BARTON II, F.E. Raman and NIR Spectroscopic Methods for Determination of Total Dietary Fiber in Cereal Foods: Utilizing Model Differences. **Applied Spectroscopy**, v. 52, p. 32-41, 1998.
- [3] BROWN, S.D. Chemical Systems Under Indirect Observation: Latent properties and chemometrics. **Applied Spectroscopy**, v. 49, p. 14A-31A, 1995.
- [4] CARVALHO, C.W.; MÜLLER, E.I.; FERRÃO, M.F.; DAVANZO, C.U. Desenvolvimento de uma Rotina para Controle de Qualidade de Farinhas de Trigo Empregando Dados de Espectroscopia por Reflexão Difusa no Infravermelho com Transformada de Fourier (DRIFTS) e Métodos de Regressão Multivariada. **Redes**, v. 7, p. 141-167, 2002.
- [5] DAVIES, A.M.C. Progress in Human Food Analysis by Near-Infrared. **Analytical Proceedings**, v. 21, p. 488-491, 1984.
- [6] FERRÃO, M.F. Técnicas de Reflexão no Infravermelho Aplicadas na Análise de Alimentos. **Tecno-Lógica**, v. 5, p. 63-85, 2001.
- [7] FULLER, M.P.E.; GRIFFITHS, P.R. Diffuse Reflectance Measurements by Infrared Fourier Transform Spectrometry. **Analytical Chemistry**, v. 50, p. 1906-1910, 1978.
- [8] GELADI, P.E.; KOWALSKI, B.R. Partial Least-Squares Regression: A Tutorial. **Analytica Chimica Acta**, v. 185, p. 1-17, 1986.
- [9] HART, J.R.; NORRIS, K.H.; GOLUMBIC, C. Determination of the Moisture Content of Seeds by Near-Infrared Spectrophotometry of Their Methanol Extracts. **Cereal Chemistry**, v. 39, p. 94-99, 1962.
- [10] HOOTON, D.E. The Versatility of Near-Infrared Reflectance Devices. **Cereal Foods World**, v. 23, p. 176-179, 1978.
- [11] IWAMOTO, M.; KONGSERRE, N.; UOZUMI, J.; SUZUKI, T. Determination of Ash Content in Home-grown Wheat Flour in Japan by Near-infrared Diffuse Reflectance Analysis. **Journal of the Japanese Society for Food Science and Technology**, v. 33, p. 842-847, 1986.
- [12] MCSHANE, W.J., Applying NIR to Process Control. **Journal of the American Oil Chemists' Society**, v. 66, p. 641-643, 1989.
- [13] OLINGER, J.M.E.; GRIFFITHS, P.R. Effects of Sample Dilution and Particle Size/Morphology on Diffuse Reflection Spectra of Carbohydrate Systems in the Near- and Mid-Infrared. Part I: Single Analytes. **Applied Spectroscopy**, v. 47, p. 687-694, 1993.
- [14] OLINGER, J.M.E.; GRIFFITHS, P.R. Effects of Sample Dilution and Particle Size/Morphology on Diffuse Reflection Spectra of Carbohydrate Systems in the Near- and Mid-Infrared. Part II: Durum Wheat. **Applied Spectroscopy**, v. 47, p. 695-701, 1993.

- [15] OSBORNE, B.G. Principles and Practice of Near-Infrared (NIR) Reflectance Analysis. **Journal of Food Technology**, v. 16, p. 13-19, 1981.
- [16] OSBORNE, B.G.; FEARN, T. Collaborative Evaluation of Near-Infrared Reflectance Analysis for the Determination of Protein, Moisture and Hardness in Wheat. **Journal of the Science of Food and Agriculture**, v. 34, p. 1011-1017, 1983.
- [17] OSBORNE, B.G.; FEARN, T. Collaborative Evaluation of Universal Calibrations for the Measurement of Protein and Moisture in Flour by Near-Infrared Reflectance. **Journal of Food Technology**, v. 18, p. 453-460, 1983.
- [18] OSBORNE, B.G.E.; FEARN, T. **NIR Spectroscopy in Food Analysis**. New York, NY, USA: John Wiley and Sons, 1986, 200p.
- [19] OSBORNE, B.G.; DOUGLAS, S.; FEARN, T. The Application of Near-Infrared Reflectance Analysis to Rapid Flour Testing. **Journal of Food Technology**, v. 17, p. 355-363, 1982.
- [20] ÖZBOY, Ö.; KÖKSEL, H. An Application of Linear Regression Technique for Predicting Bulgar Yield and Quality of Wheat Cultivars. **Nahrung/Food**, v. 46, p. 21-24, 2002.
- [21] PASIKATAN, M.C.; HAQUE, E.; SPILLMAN, C.K.; STEELE, J.L.; MILLIKEN, G.A. Granulation Sensing of First-break Ground Wheat Using a Near-infrared Reflectance Spectrometer: studies with soft red winter wheats. **Journal of the Science of Food and Agriculture**, v. 83, p. 151-157, 2003.
- [22] PASQUINI, C. Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications. **Journal of the Brazilian Chemical Society**, v. 14, p. 198-219, 2003.
- [23] PIMENTEL, M.F.; BARROS NETO, B. Calibração: Uma revisão para químicos analíticos. **Química Nova**, v. 19, p. 268-277, 1996.
- [24] REEVES III, J.B.; ZAPF, C.M. Mid-Infrared Diffuse Reflectance Spectroscopy for Discriminant Analysis of Food Ingredients. **Journal of Agricultural Food Chemistry**, v. 46, p. 3614-3622, 1998.
- [25] SCHUSTER, J.H. Semi-Automated Proximate Analysis. **Cereal Foods World**, v. 23, p. 180-182, 1978.
- [26] STARK, E.E.; LUCHTER, K. Near-Infrared Analysis (NIRA): A Technology for Quantitative and Qualitative Analysis. **Applied Spectroscopy Reviews**, v. 22, p. 335-339, 1986.
- [27] STEUER, B.; SCHULZ, H. Near-infrared Analysis of Fennel (*Foeniculum vulgare* Miller) on Different Spectrometers – Basic Considerations for a Reliable Network. **Phytochemical Analysis**, v. 14, p. 285-289, 2003.
- [28] TKACHUK, R. **Analysis of Whole Grains by Near-Infrared Reflectance**. In: Williams P., Norris K. H., editors. Near-infrared technology in the agricultural and food industries. St. Paul, MN, USA: American Association of Cereal Chemists, 1987, p. 233-240.
- [29] VAN DE VOORT, F.R. Fourier Transform Infrared Spectroscopy Applied to Food Analysis. **Food Research International**, v. 25, p. 397-403, 1992.
- [30] VAN DE VOORT, F.R.; ISMAIL, A.A. Proximate Analysis of Foods by Mid-FTIR Spectroscopy. **Trends in Food Science and Technology**, v. 2, p. 13-17, 1991.
- [31] WATSON, C.A. Near-Infrared Reflectance Spectrophotometric Analysis of Agricultural Products. **Analytical Chemistry**, v. 49, p. 835A-840A, 1977.
- [32] WILLIAMS, P.C.E.; NORRIS, K.H. **Near Infrared Technology in the Agricultural and Food Industries**. St. Paul, MN, USA: American Association of Cereal Chemists, 1990, 330p.
- [33] WILLIAMS, P.C.; STEVENSON, S.G. Near-Infrared Reflectance Analysis: Food Industry Applications. **Trends in Food Science and Technology**, v. 1, p. 44-48, 1990.
- [34] WILSON, R.H.; TAPP, H.S. Mid-Infrared Spectroscopy for Food Analysis: Recent New Applications and Relevant Developments in Sample Presentation Methods. **Trends in Analytical Chemistry**, v. 18, p. 85-93, 1999.
- [35] WILSON, R.H. Fourier Transform Mid-Infrared Spectroscopy for Food Analysis. **Trends in Analytical Chemistry**, v. 9, p. 127-131, 1990.

6 – AGRADECIMENTOS

Os autores agradecem a colaboração da Empresa Filler S.A. pelo empenho em fornecer as amostras e financiar parte dos custos do projeto, à FAPESP pelo auxílio financeiro a pesquisa (Processo 95/6043-3) e à CAPES pela concessão de bolsa de doutorado.