



Ciência e Tecnologia de Alimentos

ISSN: 0101-2061

revista@sbcta.org.br

Sociedade Brasileira de Ciência e  
Tecnologia de Alimentos  
Brasil

MORGANO, Marcelo A.; Gomes FARIA, Cristiano; FERRÃO, Marco F.; BRAGAGNOLO, Neura; FERREIRA, Márcia M.C.

DETERMINAÇÃO DE PROTEÍNA EM CAFÉ CRU POR ESPECTROSCOPIA NIR E  
REGRESSÃO PLS

Ciência e Tecnologia de Alimentos, vol. 25, núm. 1, enero-marzo, 2005, pp. 25-31

Sociedade Brasileira de Ciência e Tecnologia de Alimentos  
Campinas, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=395940073005>

- Como citar este artigo
- Número completo
- Mais artigos
- Home da revista no Redalyc

redalyc.org

Sistema de Informação Científica

Rede de Revistas Científicas da América Latina, Caribe, Espanha e Portugal

Projeto acadêmico sem fins lucrativos desenvolvido no âmbito da iniciativa Acesso Aberto

# DETERMINAÇÃO DE PROTEÍNA EM CAFÉ CRU POR ESPECTROSCOPIA NIR E REGRESSÃO PLS<sup>1</sup>

Marcelo A. MORGANO<sup>2,5,\*</sup>, Cristiano Gomes FÁRIA<sup>2,5</sup>, Marco F. FERRÃO<sup>3</sup>,

Neura BRAGAGNOLO<sup>4</sup>, Márcia M.C. FERREIRA<sup>5</sup>

## RESUMO

A combinação da espectroscopia no infravermelho próximo (NIR) e calibração multivariada (método dos mínimos quadrados parciais – PLS) para a determinação do teor de proteína total em amostras de café cru, foi investigada. Os teores de proteína total foram inicialmente determinados usando-se como método de referência o de Kjeldhal, e, posteriormente foram construídos modelos de regressão a partir dos espectros na região do infravermelho próximo das amostras de café cru. Foram coletados 159 espectros das amostras de café cru utilizando um acessório de reflectância difusa, na faixa espectral de 4500 a 10000cm<sup>-1</sup>. Os espectros originais no NIR sofreram diferentes transformações e pré-tratamento matemático, como a transformação Kubelka-Munk; correção multiplicativa de sinal (MSC); alisamento (SPLINE); derivada primeira; média móvel e o pré-tratamento dos dados escalados pela variância. O método analítico proposto possibilitou a determinação direta, sem destruição da amostra, com obtenção de resultados rápidos e sem o consumo de reagentes químicos de forma a preservar o meio ambiente. O método proposto forneceu resultados com boa capacidade de previsão do teor de proteína total, sendo que os erros médios foram inferiores a 6,7%.

**Palavras-chave:** proteína; espectroscopia no infravermelho próximo; café; PLS.

## SUMMARY

DETERMINATION OF PROTEIN IN RAW COFFEE FOR NIR SPECTROSCOPY AND REGRESSION PLS. The combination of near infrared spectroscopy (NIR) and multivariate calibration using the partial least square – PLS method for the determination of the total protein level in raw coffee samples was investigated. The total protein levels were initially determined using the Kjeldhal method as the reference method. Regression models were built from the spectra in the NIR region of the raw coffee samples. Spectra of 159 samples were recorded, using an accessory of diffuse reflectance, in the range of 4500 and 10000cm<sup>-1</sup> with 4cm<sup>-1</sup> resolution. To the raw spectral data, different transformations and mathematical pretreatment such as Kubelka-Munk transformation; multiplicative sign correction (MSC); spline smoothing; first derivative and boxcar averaging were applied. The data was also preprocessed by scaling each column by its variance. The analytical method proposed is non invasive and the cost and time of analysis are very much reduced, making possible the fast and direct determination of the total protein content. The prediction error of protein levels given by the regression model were in average lower than 6.7%, indicating that this method is a good alternative for protein determination in raw coffee samples.

**Keywords:** protein; near infrared spectroscopy; coffee; PLS.

## 1 – INTRODUÇÃO

As análises convencionais para a determinação do teor de proteína total em café utilizam métodos laboriosos, que consomem muito tempo e geram muitos resíduos químicos, como o método de Kjeldhal [1, 20]. Por outro lado, os métodos que empregam a técnica de espectroscopia no infravermelho próximo (NIRS) associada a métodos multivariados de análise fornecem resultados rápidos e exatos com diversas aplicações em análises químicas de alimentos e produtos da agricultura [2, 3, 5, 11, 18, 21].

Nos trabalhos de WILLIAMS & NORRIS [21], um grande número de metodologias analíticas desenvolvidas com base em espectros no infravermelho, demons-

tra a habilidade desta técnica de lidar com problemas tão diversos como a determinação do teor de proteínas, umidade, carboidratos e gordura com rapidez e segurança, em diferentes tipos de alimentos.

Dentre as diversas técnicas de reflexão no infravermelho pode-se destacar a reflexão difusa, já bem fundamentada na literatura [7, 12]. Esta técnica vem sendo amplamente aplicada utilizando-se equipamentos que operam na região do infravermelho próximo em análises de alimentos, devido à facilidade de se obter espectros na região do infravermelho. O sucesso da tecnologia NIRS, também se deve aos avanços da computação. Os equipamentos modernos podem gerar mais de 1000 dados para uma amostra em menos de 30 segundos [17]. Diferentes *softwares* têm sido desenvolvidos para tratar os dados da NIRS (NSAS, ISI, IDAS, Pirouette, etc.), e todos eles utilizam métodos matemáticos e estatísticos para a obtenção do máximo de informações dos dados químicos gerados. Estes métodos são denominados de quimiométricos [9, 10, 15, 16] e um dos mais empregados, é o de regressão por quadrados mínimos parciais (PLS) [8].

Assim, o objetivo deste trabalho foi o de desenvolver uma metodologia para a determinação do teor de proteína total em amostras de café cru, utilizando-se dos espectros de reflectância no infravermelho próximo (NIR) e aplicação do método de regressão por quadrados mínimos parciais (PLS).

<sup>1</sup> Recebido para publicação em 17/07/2002. Aceito para publicação em 07/12/2004 (000884).

<sup>2</sup> Centro de Química de Alimentos e Nutrição Aplicada, ITAL, Caixa Postal 139, CEP 13073-001 Campinas, SP. E-mail: morgano@ital.org.br.

<sup>3</sup> Depto de Química e Física, UNISC, CEP: 96815-000 Santa Cruz do Sul, RS. E-mail: ferrao@polaris.unisc.br.

<sup>4</sup> Depto de Ciência de Alimentos – FEA, UNICAMP, Caixa Postal 6121, CEP: 13083-970 Campinas, SP. E-mail: neura@fea.unicamp.br.

<sup>5</sup> Instituto de Química – IQ, UNICAMP. Caixa Postal 6154, CEP: 13081-970 Campinas, SP. E-mail: marcia@iqm.unicamp.br.

\* A quem a correspondência deve ser enviada.

## 2 – MATERIAL E MÉTODOS

### 2.1 – Amostras

Para a construção dos modelos de regressão foram usadas 53 amostras de café cru, da variedade arábica, procedentes de diferentes regiões produtoras de café dos Estados de São Paulo, Minas Gerais e Bahia. Todas as amostras foram inicialmente homogeneizadas em moinho de facas até a obtenção de partículas de tamanho reduzido e passaram por peneira de 0,84mm de tamanho de abertura de poro.

### 2.2 – Método de referência

O método de referência utilizado para a determinação do teor de proteína total foi o de Kjeldhal [1]. Foram transferidos para um balão de Kjeldhal  $0,5000 \pm 0,0001$ g das amostras de café cru. Juntou-se 18g de sulfato de potássio, 1g de sulfato de cobre e 25mL de ácido sulfúrico P.A. Digeriu-se a quente até à destruição completa da matéria orgânica. Resfriou-se o balão e juntou-se 150mL de água. Deixou-se escorrer pelas paredes do balão 90mL de solução de hidróxido de sódio 50% (m/v), conectou-se o balão ao condensador e destilou-se, recolhendo cerca de 100mL do destilado em erlenmeyer e titulou-se com solução de ácido clorídrico  $0,1 \text{ Mol L}^{-1}$  usando solução de verde de bromocresol até mudança de coloração de verde para rosado. Foi determinado o teor de nitrogênio total e o teor de proteína total foi calculado multiplicando-se o teor de nitrogênio encontrado pelo fator 5,75 [4]. Todas as determinações do teor de proteína total foram realizadas em duplicata. Os valores para os teores de proteína, para as amostras de café utilizadas neste trabalho, estão compreendidos entre 11,094 e 14,555g/100g.

### 2.3 – Espectroscopia no infravermelho próximo

Os espectros de reflectância na região do infravermelho próximo foram coletados em um espectrômetro BOMEM DA-08, sendo realizadas 3 réplicas para cada amostra, com resolução de  $4 \text{ cm}^{-1}$ , 16 varreduras para cada espectro e a região espectral foi de  $4500$  a  $10000 \text{ cm}^{-1}$ . O instrumento foi equipado com um acessório de reflectância difusa (JASCO), sendo os sinais expressos em  $\log(1/R)$ .

### 2.4 – Análise dos dados

O método de regressão multivariada utilizado para o tratamento de dados deste estudo foi o método PLS (regressão por quadrados mínimos parciais) e o software utilizado foi o Matlab 5.1 [14], aplicando no tratamento dos dados e na construção dos modelos as rotinas presentes no toolbox QUIMIO-PAC [6]. Este é um método bem conhecido da comunidade científica [13], e, como qualquer outro método de regressão, tem como objetivo encontrar uma relação entre a matriz (X) contendo os espectros das amostras e o vetor que armazena as respectivas concentrações (y). O resultado é uma equação semelhante à equação 1:

$$y = Xb + e \quad (1)$$

onde  $b$  é o vetor de regressão e  $e$  é o vetor que representa os erros do modelo. O método PLS é especialmente indicado, quando  $X$  contém variáveis altamente correlacionadas como neste trabalho (dados de espectroscopia). Outra vantagem do método é que pode ser usado mesmo quando as amostras contêm interferentes (que devem estar presentes no conjunto de calibração).

No modelo PLS, a matriz  $X$  é decomposta em escores,  $t$ , e pesos,  $w$ , i.e.  $XW=T$ , onde  $W = (w_1, w_2, \dots, w_k)$  é escolhido de maneira que  $T = (t_1, t_2, \dots, t_k)$  apresente covariância máxima com  $y$  ( $k$  é o número de variáveis latentes). Assim, as informações espectrais e as concentrações são usadas ao mesmo tempo na fase de calibração. Um fator de suma importância na construção de um modelo PLS é a escolha do número de variáveis latentes  $k$  (LVs) a serem incluídas no modelo.

### 2.5 – Estatística

O conjunto total de espectros foi dividido em dois subconjuntos: um de calibração e outro de validação externa. Para se determinar o número de LVs utilizadas no modelo é feita uma validação cruzada (validação interna) no conjunto de calibração: uma amostra do conjunto de calibração é excluída, o modelo é construído e então estimado o seu teor de proteína. O processo é repetido até que todas as amostras sejam previstas para 1, 2, ... variáveis latentes. A habilidade do modelo de calibração para estimar (ou prever) o teor de proteína total baseado nos dados dos espectros NIR gerados foi avaliada usando os erros de previsão e os coeficientes de correlação entre os valores dos teores de proteína estimados pelo modelo utilizando espectros NIR e os valores do método de referência das amostras do conjunto de calibração. Os parâmetros de erro empregados foram:

1. Soma dos quadrados dos erros de previsão (PRESS):

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

onde  $n$  representa o número de amostras do conjunto de calibração,  $y_i$  é o valor de referência e  $\hat{y}_i$  é o valor previsto pelo modelo para a  $i$ -ésima amostra.

2. Erro quadrático médio (MSE) ou a raiz quadrada do mesmo (RMSE) denominado também de erro padrão de previsão (SEP):

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

onde  $k$  é igual ao número de variáveis latentes.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - k - 1}} \quad (4)$$

Como estes erros podem ser medidos tanto para o conjunto de calibração quanto para o de validação exter-

na é comum adicionar no final da sigla destes erros a letra C, indicando serem estes relativos à calibração C (MSEC ou RMSEC), ou P, quando forem relativos à previsão do conjunto de validação externa (MSEP ou RMSEP).

### 3. Variância relativa (REV):

$$REV = 1 - MSE/s^2 \quad (5)$$

A estimativa da variância total dos dados ( $s^2$ ) para  $n$  amostras, utilizada na expressão acima, é determinada segundo a expressão:

$$s^2 = \sum_{i=1}^n (y_i - \bar{y})^2 / (n-1) \quad (6)$$

### 4. Erro relativo percentual (ER (%)) entre o método de referência e o método desenvolvido (NIR-PLS):

$$ER (\%) = \frac{(y_i - \hat{y}_i)}{y_i} \times 100 \quad (7)$$

### 5. Coeficiente de correlação entre os valores estimados e os valores experimentais do método de referência:

$$r = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\left[ \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2 \sum_{i=1}^n (y_i - \bar{y})^2 \right]^{1/2}} \quad (8)$$

## 3 - RESULTADOS E DISCUSSÃO

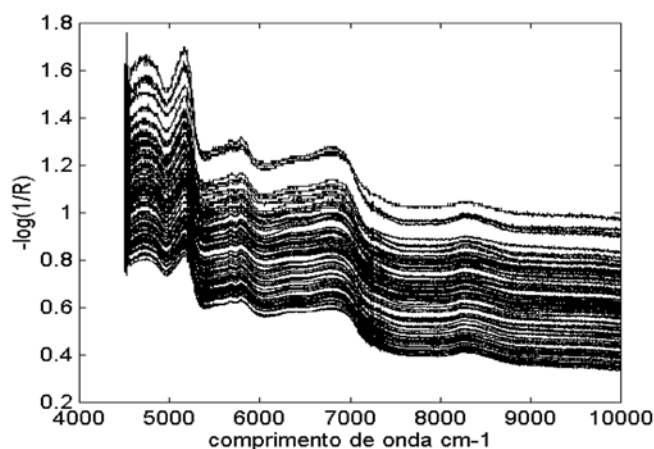
Para a construção dos modelos de regressão foram utilizados os valores médios do teor de proteína obtido na determinação empregando o método de Kjeldhal e os 159 espectros das amostras de café, sendo que 87 espectros, correspondentes a 29 amostras, foram utilizados na etapa de calibração e os 72 espectros restantes, correspondentes a 24 amostras, foram usados na validação externa dos modelos.

Na tentativa de minimizar os efeitos causados pela dificuldade de obtenção de um espectro ideal, ou seja, espectros sem variações aleatórias ou sistemáticas indesejáveis de diferentes origens, foram empregadas técnicas de tratamentos matemáticos nos espectros, antes da construção do modelo propriamente dito, para que estas variações não tivessem influência nos resultados finais. Assim, foram utilizados os seguintes tratamentos em cada espectro: a transformação Kubelka-Munk e a correção multiplicativa de sinal (MSC) [13] para minimizar os efeitos de espalhamentos de luz; o alisamento com splines (SPLINE) [19] e a média móvel para melhorar a razão sinal-ruído; finalmente a derivada primeira para a correção de deslocamento de linha de base. As absorvâncias referentes a cada comprimento de onda foram escaladas pela variância.

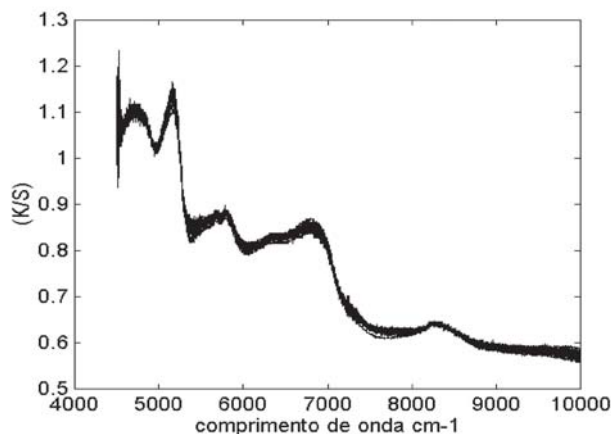
Na escolha da região espectral foram usados critérios objetivos para a avaliação do melhor conjunto de comprimentos de onda que permitiram ao modelo de calibração minimizar os erros de previsão. Este critério

foi avaliado pelos erros de previsão, como a raiz quadrada dos erros médios de previsão (RMSE). As regiões espectrais usadas correspondem, principalmente, aos estiramentos simétricos e assimétricos dos grupos N-H da proteína e estas regiões estão compreendidas entre 6623 cm<sup>-1</sup> e 5051 cm<sup>-1</sup>.

A Figura 1 mostra os espectros originais sem tratamento matemático. Os resultados após as transformações Kubelka-Munk, correção multiplicativa de sinal (MSC) e alisamento pelos splines (SPLINE) aplicadas nos espectros de reflectância podem ser visualizados na Figura 2. A seguir, foi feito um alisamento com média móvel para reduzir o número de variáveis e calculada a primeira derivada de cada espectro. A partir dos espectros assim transformados a matriz de dados resultante foi escalada pela variância (cada coluna foi dividida pelo respectivo desvio padrão).



**FIGURA 1.** Conjunto de espectros originais (sem transformações matemáticas) de reflectância na região do NIR.



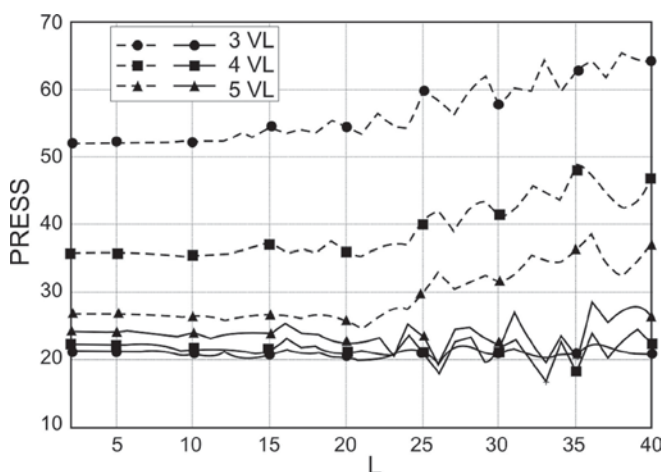
**FIGURA 2.** Conjunto de espectros NIR após transformações Kubelka-Munk, correção multiplicativa de sinal (MSC) e alisamento por SPLINE.

Para a determinação do tamanho da janela móvel e do número de variáveis latentes (VLs) foram testados modelos construídos com uma até dez variáveis latentes. Para cada um destes modelos variou-se o tamanho da janela móvel de 2 até 40. Estes resultados estão



apresentados na *Figura 3*. Observa-se que o PRESS de validação interna diminui com o aumento da complexidade do modelo, ou seja, com o aumento do número de variáveis latentes utilizado. Este comportamento foi observado também, em relação aos modelos construídos utilizando um número de variáveis latentes inferior a 3 e superior a 5. Esta tendência apresentada no PRESS de validação interna, não é observada com relação ao PRESS do conjunto de validação externa.

Com base na *Figura 3* escolheu-se o tamanho apropriado da janela móvel ( $L = 33$ , menor PRESS de validação externa) e o modelo que apresentou o menor PRESS de validação foi construído com 4 VLs. O uso de uma janela com estas dimensões permitiu reduzir o número de variáveis de 2250 (que são os comprimentos de onda lidos) para 69 variáveis. Destas 69 variáveis foram selecionadas 62 para compor o modelo final. Esta seleção foi feita a partir do valor de alavancagem (*leverage*) [13] das variáveis. Como a alavancagem está intimamente relacionada com o poder de modelagem que estas apresentam, a seleção foi feita descartando as variáveis que apresentavam um baixo valor de *leverage*, ou seja, as variáveis que contribuíam pouco na construção do modelo.



**FIGURA 3.** Representação gráfica do PRESS de validação interna (---) e de validação externa (—) em função do tamanho da janela ( $L$ ), para modelos com 3, 4 e 5 variáveis latentes (VLs).

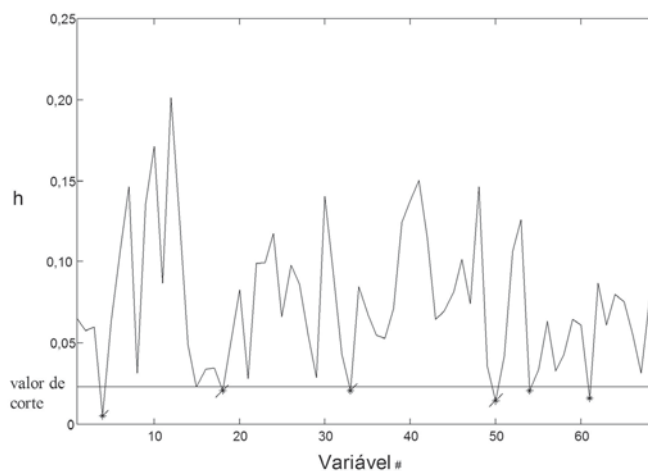
A *Figura 4* mostra o gráfico de *leverage* das variáveis. Foram descartadas as variáveis que apresentaram um valor de *leverage* inferior a 0,023 (valor de corte). A *Tabela 1* mostra os parâmetros de erro e ajuste para os modelos com e sem seleção de variáveis. Os menores valores de erro e um ligeiro aumento no valor do coeficiente de correlação linear ( $r$ ) apresentados no conjunto de calibração para o modelo II mostram que com a seleção de variáveis foi possível obter um modelo mais parcimonioso e mais ajustado. O valor superior do parâmetro REV evidencia um melhor conjunto de variáveis com melhor capacidade de previsão para este modelo. A habilidade de previsão do teor de proteína do modelo II, pode ser constatada nos baixos valores de

erro no conjunto de validação externa quando comparados com os valores do modelo I. O gráfico do PRESS do conjunto de validação externa (*Figura 5-A*) também indica ser 4 o número ótimo de VLs a serem consideradas nos modelos.

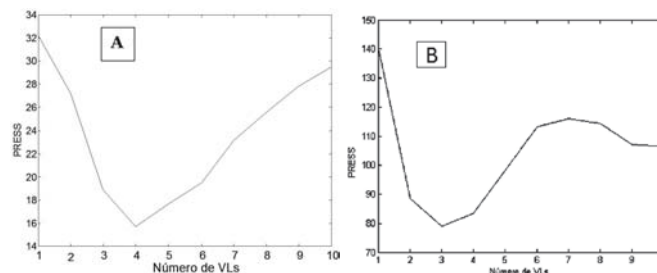
**TABELA 1.** Parâmetros de ajuste e erro dos modelos PLS utilizando 4 VLs.

Modelo	Conjunto de calibração					Conjunto de validação externa			
	PRESS	MSE	RMSE	REV	$r$	PRESS	MSE	RMSE	$r$
I (69 variáveis)	19,836	0,239	0,489	0,481	0,690	15,984	0,236	0,486	0,626
II (62 variáveis)	19,316	0,232	0,482	0,495	0,701	15,706	0,234	0,484	0,633

Press = soma dos erros quadráticos médios; MSE = erro quadrático médio; RMSE = raiz quadrada do erro quadrático médio; REV = erro relativo percentual;  $r$  = coeficiente de correlação.



**FIGURA 4.** Gráfico de *leverage* ( $h$ ) das variáveis para o modelo com 4 VLs mostrando as variáveis eliminadas.



**FIGURA 5.** **A** – Representação gráfica do PRESS do conjunto de validação externa em função do número de VLs consideradas nos modelos. **B** – Representação gráfica do PRESS do conjunto de validação interna em função do número de VLs consideradas nos modelos.

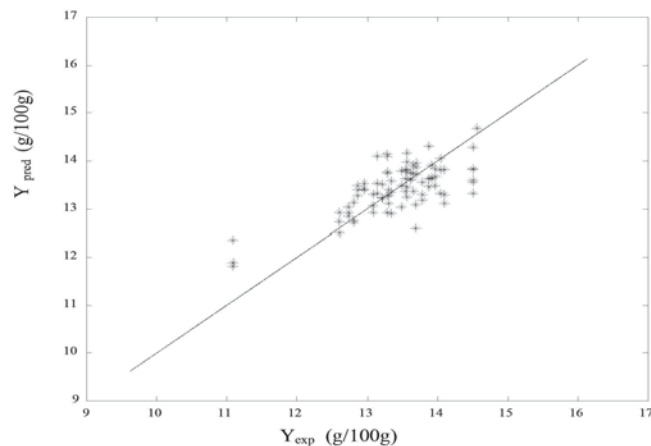
Embora a validação interna sugira a utilização de 3 VLs (*Figura 5-B*), optou-se por modelos utilizando 4 VLs, como indica o gráfico do PRESS da validação externa, uma vez que estes resultados avaliam de maneira mais apropriada a habilidade de predição do modelo frente a amostras com teor de proteína desconhecido.

Com a utilização de 4 VLs no modelo PLS foi possível captar praticamente toda a estrutura dos dados, tanto do bloco-X como do bloco-Y (vetor  $y$ ) (Tabela 2).

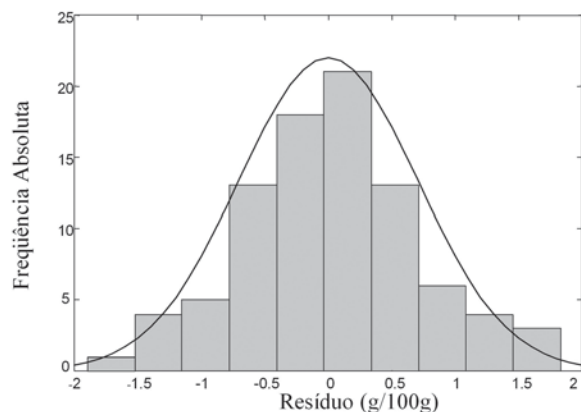
**TABELA 2.** Porcentagem de variância capturada por cada variável latente do modelo PLS.

VL#	Bloco-X		Bloco-Y	
	Variância (%)	Variância acumulada (%)	Variância (%)	Variância acumulada (%)
1	98,75	98,75	99,64	99,64
2	0,20	98,94	0,15	99,79
3	0,20	99,15	0,05	99,83
4	0,06	99,20	0,04	99,88

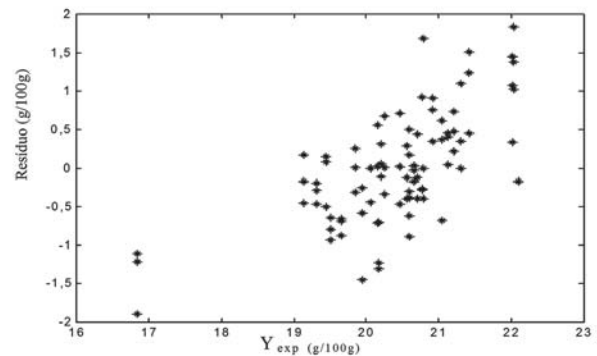
A Figura 6 mostra o gráfico dos valores experimentais versus valores preditos para as amostras do conjunto de calibração. A distribuição normal dos resíduos de calibração (Figura 7) é um indicativo que o modelo é robusto e apresenta um bom ajuste. Apesar dos baixos valores de erro, observa-se no modelo final uma tendência de superestimar o teor de proteína para as amostras de baixo teor protéico, e subestimar os valores de teor de proteína para as amostras de alto teor protéico, este aspecto indesejável do modelo fica mais evidente no gráfico dos resíduos versus o valor experimental (Figura 8).



**FIGURA 6.** Valores experimentais ( $Y_{exp}$ ) versus valores preditos ( $Y_{pred}$ ) do conjunto de calibração do modelo com 4 variáveis latentes.



**FIGURA 7.** Histograma dos resíduos de calibração.



**FIGURA 8.** Gráfico dos valores experimentais ( $Y_{exp}$ ) versus valores dos erros de calibração para o modelo II usando 4 variáveis latentes.

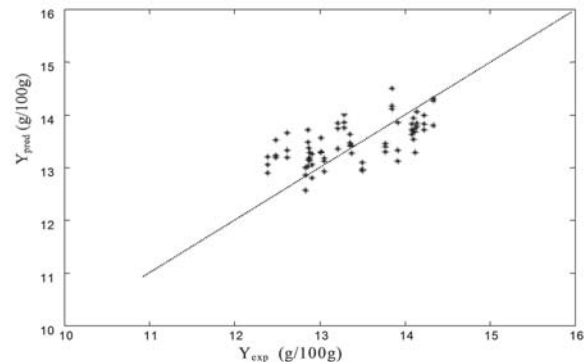
**Validação do modelo.** A validação do modelo feita por validação externa contou com um conjunto de 24 amostras tomadas em triplicatas. A Figura 9 mostra os valores experimentais versus valores preditos pelo modelo PLS (modelo II) para o conjunto de validação externa.

Na Figura 10, é mostrado o gráfico dos valores experimentais versus resíduos. Pode-se observar que a mesma tendência apresentada no conjunto de calibração foi encontrada também no conjunto de validação externa, em que as amostras que apresentam alto teor de proteína têm os seus valores subestimados pelo modelo. Já as amostras que apresentam baixo teor de proteína têm os seus valores superestimados. A Figura 11 apresenta os valores médios do teor de proteína das triplicatas versus os valores experimentais para as 24 amostras utilizadas no conjunto de validação externa e a Tabela 3 mostra os parâmetros de erro para este conjunto. Os baixos valores dos erros apresentados neste conjunto revelam uma boa habilidade de previsão do modelo.

**TABELA 3.** Parâmetros de erro para as amostras do conjunto de validação externa

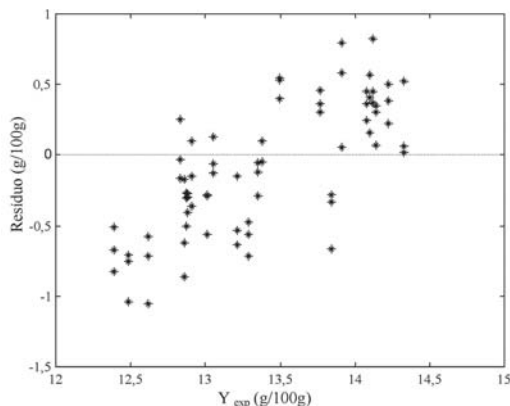
PRESS	MSE	RMSE	ER(%)	r
4,611	0,206	0,454	2,864	0,694

Press = soma dos erros quadráticos médios; MSE = erro quadrático médio; RMSE = raiz quadrada do erro quadrático médio; r = coeficiente de correlação.

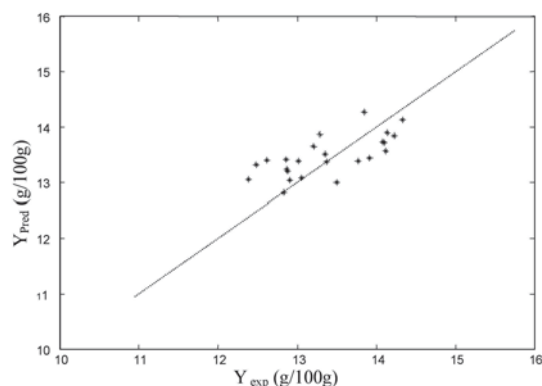


**FIGURA 9.** Curva de calibração dos valores experimentais ( $Y_{exp}$ ) versus valores preditos ( $Y_{pred}$ ) para o conjunto de validação externa do modelo com 4 variáveis latentes.

A Tabela 4 traz, para fins comparativos, os valores experimentais, os valores médios preditos, os resíduos e o erro percentual para as amostras do conjunto de validação externa. A maior discordância entre o valor experimental e o valor predito pelo modelo encontrado neste conjunto foi de 6,67%, que é um valor relativamente baixo para o erro percentual. A melhor habilidade de previsão do modelo também pode ser avaliada pelo baixo valor do erro relativo percentual apresentado pelo conjunto de validação externa.



**FIGURA 10.** Gráfico dos valores experimentais ( $Y_{exp}$ ) versus resíduos para o conjunto de validação externa.



**FIGURA 11.** Curva de calibração dos valores experimentais ( $Y_{exp}$ ) versus valores médios previstos ( $Y_{pred}$ ) das triplicatas.

Os erros relativos médios entre o método químico e o método de referência foram inferiores a 6,7%, o que indica uma boa capacidade de previsão dos teores de proteína total pelo método proposto para as amostras de café cru.

A precisão das medidas em triplicatas obtidas através do modelo PLS foram comparadas com a precisão das medidas do teor de proteína para os métodos de referência (tomadas em duplicatas). Para comparar a precisão entre estes dois conjuntos de dados foi aplicado o teste-F, utilizando a distribuição F com o nível de 5% de probabilidade. Excetuando uma medida, todas as outras medidas mostram que ao nível de 5% de probabilidade não há diferença significativa entre a precisão dos dados obtidos pelo modelo e a precisão dos dados obtidos a partir do método de referência.

**TABELA 4.** Comparação dos resultados de proteína total (g/100g) para as amostras do conjunto de validação externa obtidos pelos métodos de Kjeldhal e NIRS – PLS aplicando o modelo de regressão de melhor ajuste obtido.

Valor experimental (Método de Kjeldhal)	Valor médio predito (Método NIRS-PLS)	Resíduo	Erro (%)
13,845±0,064	14,272±0,412	-0,427	3,08
14,330±0,156	14,131±0,564	0,199	-1,39
14,080±0,111	13,727±0,202	0,353	-2,51
13,290±0,057	13,875±0,246	-0,584	4,40
14,100±0,141	13,723±0,411	0,377	-2,68
13,350±0,071	13,505±0,236	-0,155	1,16
12,614±0,069	13,396±0,490	-0,782	6,20
12,859±0,137	13,410±0,701	-0,551	4,29
13,769±0,054	13,393±0,155	0,376	-2,73
14,222±0,091	13,850±0,278	0,372	-2,61
14,138±0,088	13,898±0,296	0,240	-1,70
12,906±0,193	13,044±0,465	-0,138	1,07
12,878±0,367	13,207±0,142	-0,329	2,55
13,210±0,135	13,652±0,510	-0,442	3,34
13,378±0,231	13,377±0,171	0,001	-0,01
12,484±0,159	13,316±0,353	-0,832	6,67
14,121±0,342	13,572±0,490	0,549	-3,89
13,915±0,196	13,437±0,760	0,478	-3,43
13,496±0,159	13,005±0,161	0,491	-3,64
13,011±0,093	13,389±0,315	-0,378	2,90
13,056±0,056	13,077±0,267	-0,022	0,16
12,833±0,238	12,814±0,430	0,019	-0,15
12,873±0,151	13,231±0,249	-0,358	2,78
12,385±0,120	13,055±0,309	-0,670	5,41

#### 4 – CONCLUSÕES

A espectroscopia no infravermelho próximo associada ao método de calibração multivariada (PLS) é uma técnica alternativa simples e rápida, para a determinação do teor de proteína total em amostras de café cru. Os modelos de regressão construídos apresentaram melhor desempenho quando são empregados os seguintes tratamentos: transformação Kubelka-Munk; correção multiplicativa de sinal (MSC); alisamento (SPLINE); derivada primeira; média móvel e o pré-processamento de escalamento dos dados pela variância.

As principais vantagens do método proposto quando comparado com o método tradicional de Kjeldhal são: a redução do tempo de análise, a não destruição da amostra, pouca manipulação da amostra, resíduos químicos não são formados e o custo reduzido por amostra.

#### 5 – REFERÊNCIAS BIBLIOGRÁFICAS

- [1] ASSOCIATION OF OFFICIAL ANALYTICAL CHEMISTS. **Official methods of analysis of the A.O.A.C.** 13<sup>th</sup>

- Ed. Washington D.C. US Government Printing Office, 1980. (15.048).
- [2] BARTON, F.E. Analytical application to fibrous foods and commodities. In: **Near-infrared technology in the agricultural and food industries**; Willians, P., Norris, K., Eds.; American Association of Cereal Chemists, Inc.: St. Paul, MN, 1987.
- [3] BARTON, F.E.; WINDHAW, W.R. Determination of acid-detergent fiber and crude protein in forages by near-infrared reflectance spectroscopy: Collaborative study. **J. Assoc. Off. Anal. Chem.**, v. 71, p. 1162-1167, 1988.
- [4] BRASIL, Leis, Decretos, etc. Resolução n. 360 de 23 de dezembro de 2003. Disponível em: [http://www.anvisa.gov.br/legis/resol/2003/rdc/360\\_03rdc.htm](http://www.anvisa.gov.br/legis/resol/2003/rdc/360_03rdc.htm). Acesso em: 01 de março de 2004.
- [5] DELWICHE, S.R.; PIERCE, R.O.; CHUNG, O.K.; SEABOURN, W. Protein content of wheat by near-infrared spectroscopy of whole grain: Collaborative study. **J. Assoc. Off. Anal. Chem. Int.**, v. 81, p. 587-603, 1998.
- [6] FARIA, C.G.; FERREIRA, M.M.C. **QUIMIOPAC, LQTA**, Instituto de Química, Universidade Estadual de Campinas, lqta.iqm.unicamp.br., 2002.
- [7] FERRÃO, M.F. Técnicas de reflexão no infravermelho aplicadas na análise de alimentos. **Tecno-lógica**, v. 5, n. 1, p. 63-85, 2001.
- [8] FERREIRA, M.M.C.; ANTUNES, A.M.; MELGO, M.S.; VOLPE, P.L.O. Quimiometria I: Calibração multivariada, um tutorial. **Química Nova**, v. 22, n. 5, p. 724-731, 1999.
- [9] FERREIRA, M.M.C.; MORGANO, M.A.; QUEIROZ, S.C.N.; MANTOVANI, D.M.B. Relationships of the Mineral and Fatty Acid Contents in Processed Turkey Meat. **Food Chemistry**, v. 69, n. 3, p. 259-265, 2000.
- [10] GELADI, P. An overview of multivariate spectral analysis. In: **Near infrared spectroscopy: The future waves**, ed. A.M. Davies and P. Willians. p. 165. NIR Publications. Chichester, UK, 1996.
- [11] HONG, J.H.; IKEDA, K.; KREFT, I.; YASUMOTO, K. Near-infrared diffuse reflectance spectroscopic analysis of the amounts of moisture, protein, starch, amylose and tannin in buckwheat flours. **J. Nutr. Sci. Vitaminol.**, v. 42, p. 359-366, 1996.
- [12] HOOTON, D.E. The versatility of near-infrared reflectance devices. **Cereal Food World**, v. 23, p. 176-179, 1978.
- [13] MARTENS, H.; NAES, T. **Multivariate Calibration**. Johan Wiley & Sons: Chichester, USA, p. 418, 1989.
- [14] MATLAB – The Language of Technical Computing. **MATLAB User's guide**, Ver.5.1. The MathWorks Inc., 1997.
- [15] MORGANO, M.A.; QUEIROZ, S.C.N.; FERREIRA, M.M.C. Aplicação da Análise exploratória na diferenciação de vegetais. **Brazilian Journal of Food Technology**, v. 2, n. 1-2, p. 73-79, 1999.
- [16] MORGANO, M.A.; QUEIROZ, S.C.N.; FERREIRA, M.M.C. Determinação dos teores de minerais em sucos de frutas por espectrometria de emissão óptica em plasma indutivamente acoplado (ICP-OES). **Ciênc. Tecnol. Aliment.**, v. 19, n. 3, p. 344-348, 1999.
- [17] MROCZYK, W.B.; MICHALSKI, K.M. Application of modern computer methods for recognition of chemical compounds in NIRS. **Computers & Chemistry**, v. 22, n. 1, p. 119-122, 1998.
- [18] OSBORNE, B.G.; FEARN, T. Applications of near infrared spectroscopy in food analysis. In: **Near-infrared spectroscopy in food analysis**; Longman Scientific & Technical: Harlow, UK., 1986.
- [19] RAMSEY, J. Psychology Department, McGill University. Disponível em: <http://www.psych.mcgill.ca/faculty/ramsay/ramsay.html> Acesso em: 13 nov. 2001.
- [20] WILES, P.G.; GRAY, I.K.; KISSLING, R.C. Routine analysis of proteins by Kjeldahl and Dumas methods: Review and interlaboratory study using dairy products. **J. Assoc. Off. Anal. Chem. Int.**, v. 81, p. 620-632, 1998.
- [21] WILLIAMS, P.; NORRIS, K. Near Infrared Technology in the Agricultural and Food Industries. **American Association of Cereal Chemists**, Inc. Minnesota, USA., 1990.

## 6 – AGRADECIMENTOS

Ao Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café e à FUNAPE, pelo apoio financeiro para a realização deste trabalho.