

## MODELOS DE PREVISÃO DE NOTA DE SEIS ATRIBUTOS SENSORIAIS DE BEBIDA DE CAFÉ ARÁBICA BASEADOS EM ESPECTROS DE INFRAVERMELHO PRÓXIMO DO GRÃO TORRADO

Juliano Souza Ribeiro<sup>2</sup>; Terezinha de Jesus Garcia Salva<sup>3</sup> e Márcia Miguel Castro Ferreira<sup>4</sup>

<sup>1</sup> Trabalho financiado pela Capes, FINEP e FAPESP, CBP&D/café

<sup>2</sup> M.Sc., Instituto Agronômico de Campinas –SP e Unicamp, [jribeiro@iac.sp.gov.br](mailto:jribeiro@iac.sp.gov.br)

<sup>3</sup> Pesquisador, Dr., Instituto Agronômico de Campinas –SP, [tsalva@iac.sp.gov.br](mailto:tsalva@iac.sp.gov.br)

<sup>4</sup> Prof. Titular Instituto de Química - Unicamp – SP, [marcia@iqm.unicamp.br](mailto:marcia@iqm.unicamp.br)

**RESUMO:** A análise de espectros de infravermelho próximo de grãos de café vem sendo empregada como uma metodologia de avaliação de várias propriedades e determinação da composição química das amostras. Neste trabalho 51 amostras de cafés arábicas torrados foram analisadas sensorialmente e por reflectância difusa no infravermelho próximo. Os modelos gerados pelo método de regressão PLS (quadrados mínimos parciais), considerando a relação entre as variáveis selecionadas pelo algoritmo OPS (*Ordered Predictors Selection*) e as notas atribuídas pela avaliação sensorial da bebida, permitiram a previsão de nota de acidez com erro médio de 0,3, de amargor com erro médio de 0,37, de aroma com erro médio de 0,25, de corpo de 0,30, de sabor (bebida) de 0,37 e de qualidade global igual a 0,42, numa escala de 1 a 5.

**Palavras-chave:** Infravermelho próximo, reflectância difusa, café arábica e quimiometria.

## PREDICTION MODELS FOR EVALUATION OF SIX SENSORY ATTRIBUTES OF ARABICA COFFEE BEVERAGE BASED ON THE NEAR INFRARED SPECTRA OF THE ROASTED BEAN

**ABSTRACT:** The analyses of the near infrared beans have been used as a methodology for evaluation of several properties and chemical composition of samples. In this work 51 Arabica roasted coffee samples were tasted and analyzed by diffuse reflectance near infrared spectroscopy. The models from the PLS (partial least squares) regression method, using the relationship between the selected variables by the OPS (*Ordered Predictors Selection*) algorithm and the scores from the sensory evaluation of the beverage allowed previewing values with mean errors of 0,3 for acidity, 0,37 for bitterness, 0,25 for flavour, 0,30 for body, 0,37 for cleanliness and 0,42 for overall quality, in a scale ranging from 1 to 5.

**Key-words:** Diffuse reflectance near infrared, Arabica coffee and chemometrics.

## INTRODUÇÃO

A espectroscopia é o processo instrumental de medição baseado nas propriedades de absorção e emissão de energia eletromagnética incidente em uma amostra em alguma região do espectro eletromagnético.

A região espectral que corresponde ao infravermelho (IV) compreende a radiação em intervalo de comprimentos de onda aproximado de 700 a  $1 \times 10^6$  nm ou com número de onda entre 12800 a  $10 \text{ cm}^{-1}$ . Dentro desta faixa espectral o infravermelho pode ser dividido em três regiões distintas sendo elas: infravermelho próximo (NIR, *Near infrared* - 700-2500 nm), médio (MID, *Middle infrared* - 2500- $10^4$  nm) e distante (FAR, *Far infrared* -  $10^4$  -  $10^6$  nm).

As medidas feitas na região do IV possibilitam a análise de amostras sólidas, líquidas e gasosas. Na grande maioria das vezes, não requer grande preparo de amostra, podendo ser não destrutiva e relativamente rápida.

As análises de amostras por infravermelho próximo se baseiam na frequência de vibração das ligações químicas dos compostos que as constituem, e têm aplicação quantitativa e qualitativa. Essa técnica tem sido amplamente empregada como rotina em indústrias químicas (Li et al., 2006), alimentícias (Chen et al., 2006; Cozzolino et al., 2006; Karoui et al., 2006; Cen & He, 2007; Pizarro et al., 2007) e farmacêuticos (Reich, 2005; Roggo et al., 2007).

Para a exploração adequada das informações fornecidas pelas análises de NIR, são construídos modelos de regressão, que relacionam os dados espectrais com as informações obtidas por outros métodos de análises (métodos de referência). A regressão por quadrados mínimos parciais (PLS, *Partial Least Squares*) é uma ferramenta quimiométrica de análise multivariada bastante empregada na construção desses modelos.

Contudo, devido ao elevado número de variáveis que as análises de infravermelho próximo geram, é necessário selecioná-las para a construção do modelo de regressão. A seleção de variáveis é uma importante etapa preliminar

muito bem aceita na análise multivariada de dados, particularmente porque muitas dessas variáveis contêm informações redundantes ou ruídos experimentais (Teófilo et al., 2009). Em outras palavras, essa operação consiste na escolha de determinadas regiões do espectro (comprimentos de onda), que melhor se correlacionam com a resposta analítica em desejada.

A grande vantagem de uma seleção adequada de variáveis é que ela permite minimizar os erros de calibração, melhorando significativamente sua eficiência, resultando assim em um modelo mais robusto, simples de interpretar e com melhor exatidão nas previsões.

Um grande número de procedimentos para seleção de variáveis está disponível na literatura (Osborne et al., 199; Forina et al., 2004; Chong & Jun, 2005; Hoskuldsson, 2001; Teófilo et al., 2009), a maioria focada na seleção de comprimentos de onda de dados espectroscópicos.

O objetivo desse trabalho foi construir modelos multivariados de regressão para a previsão de notas de acidez, amargor, aroma, bebida (sabor), corpo e qualidade global de bebida de café, empregando o método PLS e análises de cafés torrados por infravermelho próximo.

## MATERIAL E MÉTODOS

Cinquenta e uma amostras de café arábica com e sem defeitos visíveis foram utilizadas neste trabalho. Para a análise de infravermelho, os grãos de café foram torrados ao ponto de torra média (Agtron # 55) em um torrador rotatório a gás (Pinhalense S/A Máquinas Agrícolas).

Os grãos torrados foram moídos e analisados no espectrômetro Nirsystem 6500 (Foss NIRSystems, Raamsdonksveer, Holanda), equipado com acessório de reflectância difusa e módulo de transporte de amostra. Para o controle do instrumento foi empregado o software Vision 2.22 (Foss NIRSystems, Raamsdonksveer, Holanda). As análises foram efetuadas com 256 varreduras na faixa de 1100 - 2500 nm, com 4 cm<sup>-1</sup> de resolução. Um padrão de poliestireno (4 picos) existente no próprio aparelho foi utilizado como referência. As amostras foram analisadas em triplicatas constituídas de porções diferentes de cerca de 4 gramas da café torrado e moído.

As análises sensoriais das bebidas das amostras de café foram feitas por 5 provadores pela “prova de xícara”, seguindo o preparo e avaliação determinados pela Instrução Normativa n° 8 de 11/06/2003 (<http://www.agricultura.gov.br>). Foram atribuídas notas de 1 a 5 para os atributos acidez, amargor, aroma, bebida (sabor), corpo e qualidade global. A escala foi crescente com o aumento da qualidade/intensidade do atributo considerado.

Os espectros adquiridos foram transformados em uma matriz  $\mathbf{X}$  ( $I \times J$ ) na qual cada replicata foi utilizada como uma amostra. O processamento dos dados foi realizado utilizando o software Matlab 6.5 (The MathWorks, Co., Natick, MA, USA) e o pacote computacional PLS\_Toolbox (Eigenvector Research, Inc. – PLS\_Toolbox version 3.02) (Wise et al., 2004).

O modelo matemático de regressão multivariada utilizada foi a regressão por quadrados mínimos parciais (PLS) (Ferreira et al., 1999). Para reduzir as variações aleatórias (ruído experimental) foi aplicado o alisamento, segundo Savitzky-Golay (1964), com janela tamanho 5. As variações sistemáticas foram reduzidas utilizando a função da 1ª derivada e os dados foram centrados na média. A seleção das variáveis para os modelos de regressão foi feita pelo algoritmo OPS (Teófilo et al., 2009).

O número de variáveis latentes empregados nos modelos foi determinado a partir da análise dos valores do erro de validação cruzada (RMSECV, *root mean square error of cross validation*) calculados para cada modelo (eq. 1), e pelo coeficiente de correlação da validação cruzada,  $r_{cv}$  (eq. 2). A validação externa foi avaliada pelos valores de RMSEP (*root mean square error of prediction*) (eq.3) obtidos para cada modelo.

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

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

$$RMSEP = \sqrt{\frac{\sum_{i=1}^I (y_i^p - \hat{y}_i^p)^2}{I_p}} \quad (3)$$

onde na (eq. 1),  $y_i$  é a resposta medida para cada amostra  $i^{\text{ésima}}$  do modelo de calibração,  $\hat{y}_i$  é a resposta estimada para a amostra retirada para a realização da validação cruzada,  $I$  é o número de amostras do conjunto de calibração e  $k$  é o número de variáveis latentes utilizadas no modelo. Na (eq. 2)  $\bar{\hat{y}}$  é a resposta media estimada e na (eq. 3)  $I_p$  é o número

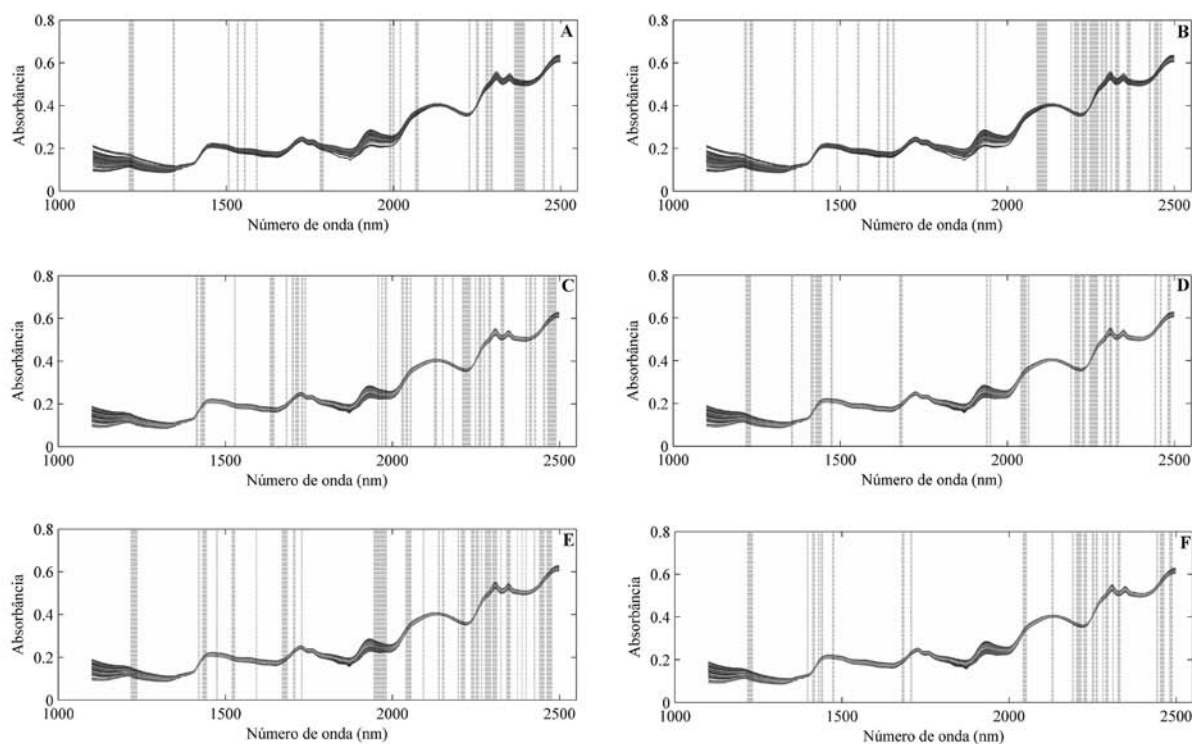
de amostras do conjunto de previsão,  $\hat{y}_i^p$  e  $y_i^p$  são valores preditos e medidos das respostas utilizadas nesse conjunto de previsão.

## RESULTADOS E DISCUSSÃO

Para a construção dos modelos de regressão foram utilizados os valores médios das notas de acidez, amargor, aroma, bebida corpo e qualidade global ( $y$ ) dadas pelos provadores e os espectros de infravermelho das amostras de café arábica torrado ( $X$ ).

Para formar o conjunto de calibração de cada modelo foram selecionados aleatoriamente 123 espectros, correspondentes a 41 amostras, de forma a cobrir toda a faixa de resposta das análises sensoriais. A validação cruzada (interna) realizada com cada conjunto de calibração foi realizada retirando-se aleatoriamente 5 amostras por vez (15 replicatas). Os 30 espectros restantes para cada modelo, correspondentes a 10 amostras, foram utilizados para formar o conjunto de previsão (calibração externa).

A partir da matriz de dados pré-tratada  $X_p$  (153x700), o método OPS foi utilizado para a seleção de variáveis de cada modelo. Desse modo, de um conjunto inicial de 700 variáveis foram selecionadas 76 para a construção do modelo de acidez (A), 116 para amargor (B), 118 para o aroma (C), 99 para a bebida (D), 143 para o corpo (E) e 85 para qualidade global (F). Estas variáveis estão indicadas em linhas verticais na Figura 1.



**Figura 1** - Espectros sobrepostos dos cafés e regiões do infravermelho próximo selecionadas para a construção dos modelos de calibração. A (acidez), B (amargor), C (Aroma), D (sabor), E (corpo), F (qualidade global)

O número de variáveis latentes utilizado nos modelos PLS foi determinado a partir dos valores de RMSECV exibidos nos conjuntos de validação cruzada. Esses valores de RMSECV foram atingidos em cada modelo quando se utilizou o número de variáveis latentes indicado na Tabela 2. Nessa tabela são indicados ainda os valores de RMSECV e os coeficientes de correlação da validação cruzada ( $r_{vc}$ ).

**Tabela 2** - Parâmetros estatísticos calculados para os modelos de regressão na validação cruzada.

Modelos	Nº VL	RMSECV	$r_{vc}$
Acidez	8	0,28 ± 0,01	0,84 ± 0,02
Amargor	8	0,35 ± 0,01	0,87 ± 0,01
Aroma	7	0,31 ± 0,01	0,93 ± 0,00
Bebida	8	0,38 ± 0,01	0,91 ± 0,01
Corpo	9	0,27 ± 0,01	0,88 ± 0,01
Qualidade global	8	0,39 ± 0,01	0,91 ± 0,00

Em média, 99 % e 86 % da variância dos blocos  $Y$  e  $X$  utilizados nas calibrações foram necessários para descrever cada modelo de regressão.

Os modelos para a predição das notas foram validados por validação externa, utilizando o conjunto de 10 amostras cada, relativos a 30 replicatas. A Tabela 3 apresenta os valores de referência dos provadores e os valores preditos pelos modelos de regressão para as amostras de previsão. Os valores dos erros de previsão (RMSEP) dos modelos foram 0,30 para acidez, 0,37 para o amargor, 0,25 para o aroma, 0,30 para corpo, 0,37 para bebida e 0,42 para qualidade global.

**Tabela 3** - Valores médios e desvios padrão das notas dos seis atributos sensoriais conferidos pelos provadores e previstos pelos modelos de regressão.

Amostras	Acidez		Amargor		Aroma	
	Medido	Previsto	Medido	Previsto	Medido	Previsto
1	3,00 ± 0,06	2,56 ± 0,08	4,13 ± 0,63	3,65 ± 0,22	2,75 ± 0,29	2,70 ± 0,09
2	3,70 ± 0,02	3,35 ± 0,05	4,38 ± 0,75	4,40 ± 0,06	4,19 ± 0,52	4,16 ± 0,06
3	3,69 ± 0,06	3,54 ± 0,11	4,13 ± 0,63	4,12 ± 0,11	3,63 ± 0,48	3,39 ± 0,05
4	3,38 ± 0,03	3,10 ± 0,02	4,13 ± 0,63	4,15 ± 0,08	3,75 ± 0,29	3,84 ± 0,12
5	2,56 ± 0,05	3,01 ± 0,05	2,69 ± 0,63	3,36 ± 0,08	2,25 ± 0,00	2,78 ± 0,01
6	2,19 ± 0,03	2,38 ± 0,05	3,25 ± 0,50	3,31 ± 0,03	2,63 ± 0,48	2,75 ± 0,07
7	2,91 ± 0,06	2,85 ± 0,02	3,50 ± 0,50	3,46 ± 0,20	4,50 ± 0,48	4,24 ± 0,10
8	3,00 ± 0,04	2,53 ± 0,05	3,50 ± 0,63	3,42 ± 0,15	4,00 ± 0,21	4,08 ± 0,02
9	3,70 ± 0,02	3,68 ± 0,04	4,00 ± 0,75	3,71 ± 0,10	4,50 ± 0,52	4,17 ± 0,10
10	3,00 ± 0,04	2,99 ± 0,08	4,00 ± 0,63	3,28 ± 0,18	4,50 ± 0,65	4,34 ± 0,13

	Bebida		Corpo		Qualidade global	
	Medido	Previsto	Medido	Previsto	Medido	Previsto
1	3,83 ± 0,29	3,51 ± 0,07	3,33 ± 0,58	3,15 ± 0,19	3,38 ± 0,48	2,93 ± 0,08
2	4,31 ± 0,38	3,82 ± 0,07	3,44 ± 0,38	3,29 ± 0,15	4,38 ± 0,25	3,96 ± 0,04
3	3,63 ± 0,48	3,52 ± 0,07	3,66 ± 0,58	3,33 ± 0,16	4,13 ± 0,25	3,99 ± 0,05
4	3,88 ± 0,14	3,47 ± 0,05	2,69 ± 0,31	2,71 ± 0,10	4,00 ± 0,00	3,54 ± 0,12
5	1,50 ± 0,20	1,83 ± 0,06	1,94 ± 0,24	2,52 ± 0,01	1,50 ± 0,20	2,31 ± 0,09
6	2,38 ± 0,43	2,54 ± 0,02	2,38 ± 0,43	2,50 ± 0,02	2,19 ± 0,13	2,45 ± 0,04
7	3,75 ± 0,29	3,88 ± 0,21	3,50 ± 0,48	3,50 ± 0,33	3,50 ± 0,41	3,94 ± 0,04
8	4,00 ± 0,31	3,77 ± 0,05	3,50 ± 0,25	3,35 ± 0,06	4,00 ± 0,25	3,80 ± 0,07
9	4,50 ± 0,24	4,25 ± 0,12	4,00 ± 0,50	3,61 ± 0,07	4,50 ± 0,25	4,17 ± 0,08
10	4,50 ± 0,41	4,35 ± 0,07	4,00 ± 0,41	3,82 ± 0,18	4,50 ± 0,50	4,35 ± 0,18

## CONCLUSÕES

O método de seleção de variáveis empregado indicou 76 regiões do infravermelho próximo relacionadas com a acidez da bebida de café, 116 relacionadas com o amargor, 118 relacionadas com o aroma, 99 com o sabor, 143 com o corpo e 85 com a qualidade global. Numa escala de 1 a 5, os modelos de regressão PLS gerados utilizando dados espectroscópicos de cafés arábicas torrados, previram as notas com um erro médio de 0,3 para a acidez, 0,37 para o amargor, 0,25 para o aroma, 0,30 para o corpo, 0,37 para bebida e 0,42 para qualidade global.

## REFERÊNCIAS BIBLIOGRÁFICAS

- CEN, H.; HE, Y. Theory and application of near infrared reflectance spectroscopy in determination of food quality. **Trends Food Sci. Technol.** 2007, 18, 72-83.
- CHEN, Q.; ZHAO, J.; ZHANG, H.; WANG, X. Feasibility study on qualitative and quantitative analysis in tea by near infrared spectroscopy with multivariate calibration. **Anal. Chim. Acta** 2006, 572, 77-84.
- CHONG, I. G.; JUN, C. H. Performance of some variable selection methods when multicollinearity is present. **Chemometr. Intell. Lab. Syst.** 2005, 78, 103-112.
- COZZOLINO, D.; SMYTH, H. E.; LATTEY, K. A.; CYNKAR, W.; JANIK, L.; DAMBERGS, R. G.; FRANCIS, I. L.; GISHEN, M. Combining mass spectrometry based electronic nose, visible-near infrared spectroscopy and chemometrics to assess the sensory properties of Australian Riesling wines. **Anal. Chim. Acta** 2006, 563, 319-324.
- FERREIRA, M. M. C.; ANTUNES, A. M.; MELGO, M. S., VOLPE, P. L. O. Chemometrics I – Multivariate calibration, a tutorial. **Quim. Nova** 1999, 22, 724-731.
- FORINA, M.; LANTERI, S.; OLIVEROS, M.; MILLAN, C. P. Selection of useful predictors in multivariate calibration. **Anal. Bioanal. Chem.** 2004, 380, 397-418.
- HOSKULDSSON, A. Variable and subset selection in PLS regression. **Chemometr. Intell. Lab. Syst.** 2001, 55, 23-38.

- KAROUI, R.; PILLONEL, L.; SCHALLER, E.; BOSSET, J. O.; DE BAERDEMAEKER, J. Prediction of sensory attributes of European Emmental cheese using near-infrared spectroscopy: A feasibility study. **Food Chem.** 2006, 101, 1121-1129.
- LI, W.; HUANG, D.; LIU, L. The application of near infrared spectroscopy in the quality control analysis of glass/phenolic resin prepreg. **J. Mater. Sci.** 2006, 41, 7183-7189.
- OSBORNE, S. D.; JORDAN, R. D.; KUNNEMEYER, R. Method of wavelength selection for partial least square. **Analyst** 1997, 122, 1531-1537.
- PIZARRO, C.; ESTEBAN-DÍEZ, I.; GONZÁLEZ-SÁIZ, J. M.; FORINA, M. Use of Near-Infrared Spectroscopy and Feature Selection Techniques for Predicting the Caffeine Content and Roasting Color in Roasted Coffees. **J. Agric. Food Chem.** 2007, 55, 7477-7488.
- REICH, G. Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications. **Advanced Drug Deliv. Rev.** 2005, 57, 1109-1143.
- ROGGO, Y.; CHALUS, P.; MAURER, L.; LEMA-MARTINEZ, C.; EDMOND, A.; JENT, N. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. **J. Pharm. Medical Anal.** 2007, 44, 683-700.
- SAVITZKY, A.; GOLAY, M. J. E. Smoothing + Differentiation of data by Simplified Least Squares Procedures. **Anal. Chem.** 1964, 36, 1627-&.
- SKOOG, D. A.; ROLLER, F. J.; NIEMAN, T. A. Princípios de Análise instrumental. 5ª ed. Bookman, São Paulo, 2002.
- TEÓFILO, R. F.; MARTINS, J. P. A.; FERREIRA, M. M. C. Sorting variables by using informative vectors as a strategy for featuring selection in multivariate regression. **J. Chemom.** 2009, 23, 32-41.
- WISE, B. M.; GALLAGHER, N. B.; BRO, R.; SHAVER, J. M.; WINDIG, W.; KOCH, R. S. PLS\_Toolbox 3.5, for Use with Matlab™, Eigenvector Research. 2004.