

MODELO QUIMIOMÉTRICO PARA A PREVISÃO DA QUALIDADE GLOBAL DE BEBIDA DE CAFÉ ARÁBICA BASEADO EM ESPECTROSCOPIA DE INFRAVERMELHO PRÓXIMO E REGRESSÃO POR QUADRADOS MÍNIMOS PARCIAIS¹

Juliano Souza Ribeiro²; Terezinha de Jesus Garcia Salva³

¹ Trabalho financiado pela FINEP e CBP&D/café

² M.Sc., Instituto Agronômico de Campinas –SP e Unicamp, jribeiro@iqm.unicamp.br

³ Pesquisador, Dr., Instituto Agronômico de Campinas –SP, tsalva@iac.sp.gov.br

RESUMO: A análise de infravermelho próximo de amostras de café cru e torrado tem se mostrado bastante promissora para a identificação de espécies, para a definição do ponto de torra e para a quantificação de alguns compostos químicos da semente. A espectroscopia de infravermelho próximo tem sido proposta também como um método objetivo de avaliação da qualidade da bebida. Neste trabalho apresenta-se um modelo quimiométrico para a previsão de qualidade global da bebida de café arábica construído a partir da correlação entre espectros de infravermelho próximo por reflectância difusa e análises sensoriais de cafés arábica cru. Para isso, cento e dezenove amostras de cafés arábica cru foram submetidas à avaliação sensorial para o atributo qualidade global e analisadas em infravermelho próximo por reflectância difusa. A regressão pelo método de quadrados mínimos parciais (PLS) foi utilizada para a construção do modelo. Através da seleção de variáveis realizada pelo método OPS foram identificadas 24 regiões espectrais relacionadas com a qualidade global. Para o modelo calculado com 5 variáveis latentes, o erro de validação cruzada (RMSECV) foi de $0,70 \pm 0,04$ e o coeficiente de correlação (r_{vc}) foi de $0,93 \pm 0,01$. Usando uma escala de 1 a 10, o erro médio dos valores previstos foi igual a 0,4, enquanto o erro médio da prova de xícara foi de 0,5.

Palavras-chave: Infravermelho próximo, café arábica, quimiometria, qualidade global.

CHEMOMETRIC MODEL FOR OVERALL QUALITY PREDICTION OF ARABICA COFFEE BEVERAGE BASED ON NEAR INFRARED SPECTROSCOPY AND PARCIAL LEAST SQUARES REGRESSION

ABSTRACT: The near infrared analysis of samples of raw coffee has been proposed for the species identification, degree of roasting determination and for the quantification of some chemical compounds of the seed. The near infrared spectroscopy has been also tried as an objective method for beverage quality evaluation. In this work a chemometric prediction model for the Arabica coffee beverage overall quality was constructed by the correlation between near infrared diffuse reflectance spectra and sensory score for the attribute. One hundred and nineteen samples of crude arabica coffees were submitted to the sensory evaluation regarding the overall quality attribute and analyzed by diffuse reflectance near infrared. The partial least squares (PLS) regression method was used for the construction of the model. Through, OPS variable selection method, 24 spectral regions related to overall quality were identified. For the model built with 5 latent variables, the root mean square error of cross validation (RMSECV) was $0,70 \pm 0,04$ and the correlation coefficient (r_{vc}) was $0,93 \pm 0,01$. When using an scale ranging from 1 to 10, the mean error of the previewed values was 0,4, while the cupping mean error was 0,5.

Key words: Near infrared, arabica coffee and chemometrics.

INTRODUÇÃO

O termo “qualidade” empregado a um determinado produto é uma das palavras-chave mais utilizadas no comércio mundial. A “qualidade” de um produto, no entanto, pode assumir diferentes significados para os consumidores, produtores e órgãos reguladores.

No caso do café, por exemplo, a qualidade pode se referir desde ao sistema de produção agrícola empregado e às características físicas e químicas dos grãos crus e torrados até às características e propriedades da bebida já na mesa do consumidor.

Tanto as caracterizações químicas quanto as sensoriais do café compreendem análises elaboradas, normalmente onerosas e demoradas. Uma metodologia substitutiva às usuais que é bastante promissora, principalmente por ser rápida e poder fornecer resultados sob vários aspectos das amostras é a espectroscopia de infravermelho próximo (NIRS).

Essencialmente, na região do NIR, bandas vibracionais de absorção de compostos orgânicos são determinadas por sobretoms (do inglês, *overtones*) e bandas de combinação envolvendo ligações C-H, O-H e N-H (Olinger & Griffiths, 1993). A sobreposição destas bandas de combinação e sobretoms é um complicador à atribuição precisa das regiões do infravermelho próximo (1100-2500 nm) a um determinado evento no grão de café (Workman, 1996).

Contudo, os métodos quimiométricos de tratamento de dados facilitam a interpretação do grande número de variáveis geradas na análise espectroscópica.

A literatura tem apresentado resultados de estudos empregando NIRS associado à quimiometria que visam à diferenciação de *Coffea arabica* e *Coffea canephora* em variedades puras e em *blends* (Kemsley et al., 1995; Pizarro et al., 2007a; Esteban-Díez et al., 2007), à quantificação de cafeína (Pizarro et al., 2007b), trigonelina, ácidos clorogênicos (Morgano et al., 2001; Bertrand et al 2005), açúcares totais (Morgano et al., 2007), minerais (Morgano et al., 2002), à definição do grau de torra, à determinação de umidade (Alessandrini et al., 2008) e à classificação de cafés expressos (Esteban-Díez et al., 2004).

Oficialmente, a qualidade sensorial da bebida de café deve ser determinada pela prova “de xícara”, como estabelecem as normas brasileiras em uso (Instrução Normativa nº 8 de 11/06/2003 <http://www.agricultura.gov.br>). Mesmo reconhecendo a “prova de xícara” como o método de referência para fins de avaliação da qualidade sensorial da bebida, no entanto, cientistas vêm procurando com diferentes finalidades desenvolver métodos analíticos simples e rápidos, que possam ser implementados em análises rotineiras da bebida do café.

Neste contexto, o objetivo desse trabalho foi desenvolver um modelo matemático de previsão da qualidade global baseado em espectros de infravermelho próximo e análises sensoriais de cafés arábica.

MATERIAL E MÉTODOS

Cento e dezenove amostras de café arábica cru, de diferentes procedências e livres de defeitos visíveis, foram utilizadas neste trabalho. Uma parte de cada amostra foi encaminhada para a “prova de xícara” por provadores enquanto outra parte foi moída em moinho Knifetec 1095 da marca Foss Tecator.

A escala utilizada pelos provadores para classificar o café segundo a sua qualidade global foi de 10 pontos. Nesta escala numérica, aos cafés de excelente qualidade global foram atribuídos 10 pontos.

Os espectros de infravermelho próximo foram adquiridos em espectrômetro modelo Nirsystem 6500 (Foss NIRSystems, Raamsdonksveer, Holanda) de reflectância difusa com módulo de transporte de amostra e controlado pelo software Vision 2.22 (Foss NIRSystems, Raamsdonksveer, Holanda).

Os espectros foram obtidos diretamente sobre as amostras moídas de café cru usando 256 varreduras na faixa de 1100 - 2500 nm, com 4 cm⁻¹ de resolução. Um padrão interno de poliestireno (4 picos) foi utilizado como branco. Três diferentes porções de cada amostra eram analisadas (triplicatas).

Os dados obtidos na análise espectroscópica foram transformados em matrizes 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).

A regressão por quadrados mínimos parciais (PLS) foi utilizada como método de regressão (Ferreira et al., 1999). Para reduzir as variações aleatórias (ruído experimental) foi realizado um alisamento, segundo Savitzky-Golay (1964), com janela tamanho 7. As variações sistemáticas foram reduzidas utilizando a 1ª derivada e os dados foram centrados na média. A seleção das variáveis para o modelo de regressão (PLS) foi realizada utilizando-se o algoritmo OPS (*Ordered Predictors Selection*) (Teófilo et al., 2009).

RESULTADOS E DISCUSSÃO

Na construção do modelo de regressão foram utilizados os valores médios das notas de qualidade global obtidas através das análises sensoriais descritivas pelos provadores e 357 espectros referentes à 119 amostras de café arábica.

Para formar o conjunto de calibração foram selecionados aleatoriamente 321 espectros, correspondentes a 107 amostras, de forma a cobrir toda a faixa de 0 a 10 estabelecida como escala das respostas da avaliação sensorial. A validação cruzada (interna) realizada com cada conjunto de calibração (321 replicatas) foi realizada retirando-se aleatoriamente 5 amostras por vez (15 replicatas). Os 36 espectros restantes para cada modelo, correspondentes a 12 amostras, foram utilizados para formar os conjuntos de previsão (calibração externa).

A partir da matriz de dados pré-tratada (363x700), das 700 variáveis originais, o método OPS selecionou 206 variáveis para a construção do modelo de qualidade global, indicadas como linhas verticais na Figura 1.

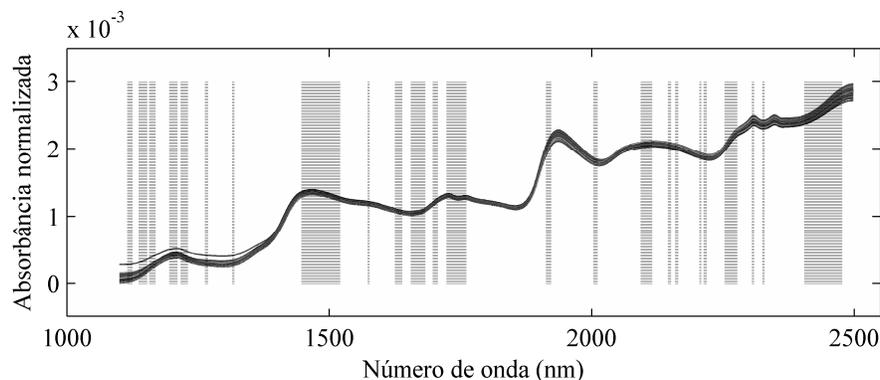


Figura 1- Espectros sobrepostos das amostras de café cru moído e regiões do infravermelho próximo selecionadas para a construção do modelo de calibração para a qualidade global.

Para o conjunto de dados considerado neste trabalho, o modelo de regressão PLS construído com 5 variáveis latentes apresentou algumas replicatas com valores elevados de *leverage* e outras com valores elevados de resíduo de *Student*, porém não foram identificadas amostras atípicas (Figura 2) para o limite de confiança do valor de *leverage* de 95 %.

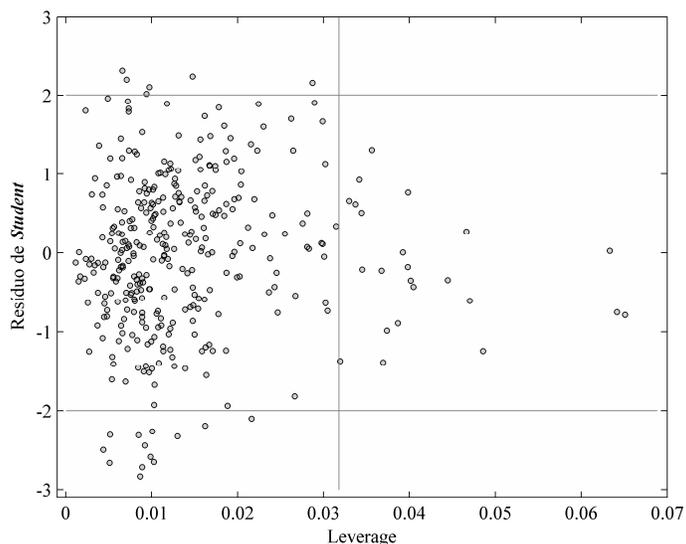


Figura 2 - Gráfico de *leverage* vs resíduo de *Student* para visualização de possíveis amostras atípicas.

Esse modelo apresentou erro de validação cruzada (RMSECV, *root mean square error of cross validation*) igual a $0,70 \pm 0,04$ e coeficiente de correlação da validação cruzada (r_{vc}) igual a $0,93 \pm 0,01$. Com as 5 variáveis latentes utilizadas foi possível descrever 87,82 % e 89,5 % da variância dos blocos **Y** e **X** utilizados nas calibrações.

O modelo foi validado por validação externa, utilizando o conjunto de 12 amostras, relativas às 36 replicatas. A Figura 3 representa os valores atribuídos pelos provedores *versus* os valores preditos pelo modelo para cada amostra de previsão distribuídos entre os valores do conjunto de validação cruzada. A Tabela 1 apresenta numericamente esses valores para facilitar a comparação dos resultados. O erro de previsão das amostras externas (RMSEP, *root mean square error of prediction*) foi de 0,4, enquanto o erro médio observado entre os provedores para esse atributo foi igual a 0,5, na mesma escala.

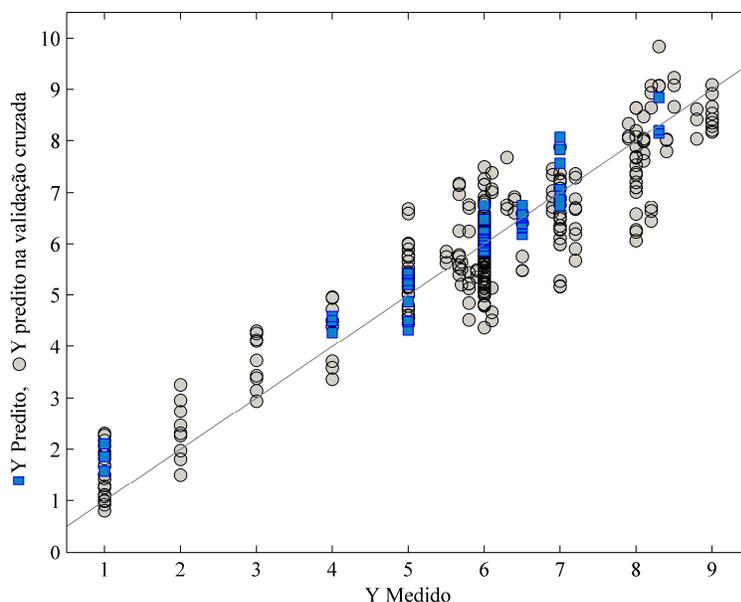


Figura 3 - Representação gráfica das notas de qualidade global dos provadores para as amostras de validação cruzada (○) e previsão (□) versus as notas previstas pelo modelo de regressão.

Tabela 1 - Valores médios de qualidade global da bebida atribuídos pelos provadores e previstos pelo modelo de regressão.

Amostras de previsão	Notas médias dos provadores	Notas previstas pelo modelo
1	7,0 ± 0,5	6,97 ± 0,17
2	7,0 ± 0,5	7,79 ± 0,26
3	6,0 ± 0,5	6,46 ± 0,27
4	5,0 ± 0,5	4,47 ± 0,25
5	5,0 ± 0,5	5,28 ± 0,11
6	8,3 ± 0,5	8,37 ± 0,42
7	6,0 ± 0,5	5,98 ± 0,08
8	6,0 ± 0,5	5,95 ± 0,13
9	6,5 ± 0,5	6,55 ± 0,24
10	6,5 ± 0,5	6,28 ± 0,13
11	1,0 ± 0,5	1,81 ± 0,21
12	4,0 ± 0,5	4,44 ± 0,16

CONCLUSÕES

Existe elevada relação linear entre as notas atribuídas à qualidade global por provadores e determinados comprimentos de onda na faixa de infravermelho próximo do café cru. O modelo de regressão PLS (quadrados mínimos parciais) gerado a partir das 206 variáveis selecionadas pelo método OPS e utilizando 5 variáveis latentes previu notas de qualidade global com erro médio de 0,4 numa escala de 1 a 10, enquanto o erro médio dos provadores foi de 0,5, na mesma escala.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALESSANDRINI, L.; ROMANI, S.; PINNAVAIA, G.; ROSA, M. D.; Near infrared spectroscopy: an analytical tool to predict coffee roasting degree. *Anal. Chim. Acta* 2008, 625, 95-102.
- BERTRAND, B.; ETIENNE, H. ; LASHERMES, P. ; GUYOT, B. ; DAVRIEUX F. Can near-infrared reflectance of green coffee be used to detect introgression in *Coffea arabica* cultivars ? *J. Sci. Food Agric.* 2005, 85, 955-962.
- ESTEBAN-DÍEZ, I.; GONZÁLEZ-SÁIZ, J. M.; PIZARRO, C. Prediction of sensory properties of espresso from roasted coffee samples by near-infrared spectroscopy. *Anal. Chim. Acta* 2004, 525, 171- 182.
- ESTEBAN-DÍEZ, I.; GONZÁLEZ-SÁIZ, J. M.; SÁENZ-GONZÁLEZ, C.; PIZARRO, C. Coffee varietal differentiation based on near infrared spectroscopy. *Talanta* 2007, 71, 221–229.
- 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.

- KEMSLEY, E. K.; RUAULT, S.; WILSON, R. H.; Discrimination between *Coffea arabica* and *Coffea canephora* variant robusta beans using infrared spectroscopy. **Anal. Nutr. Clinical Meth. Sec.** 1995, 54, 321-326.
- MORGANO, M. A.; FARIA, C. G.; FERRÃO, M. F.; FERREIRA, M. C. Determinação de açúcares total em café cru por espectroscopia no infravermelho próximo e regressão por mínimos quadrados parciais. **Quim. Nova** 2007, 30, 346-350.
- MORGANO, M. A.; PAULUCI, L. F.; MANTOVANI, B. M. V.; MORY, E. E. M. Determinação de minerais em café cru. **Ciênc. Tecnol. Aliment.** 2002, 22, 19-23.
- MORGANO, M.; CAMARGO, C.; PAGEL, A. P.; FERRAO, M. F.; FERREIRA, M. M. C. Determinação simultânea dos teores de cafeína, trigonelina e ácido clorogênico em amostras de café cru por análise multivariada (PLS) em dados de espectroscopia por reflexão difusa no infravermelho próximo. In: II Simpósio de Pesquisa dos Cafés do Brasil, Vitória, ES. **Anais do II Simpósio de Pesquisa dos Cafés do Brasil.** 2001,1502-1510.
- OLINGER, J. M.; GRIFFITHS, P. R. Effects of sample dilution and particle size/morphology on diffuse reflectance spectra of carbohydrate systems in the near- and mid-infrared. Part I: Single Analytes. **Appli. Spectros.** 1993, 47, 687-694.
- PIZARRO, C.; ESTEBAN-DÍEZ, I.; GONZÁLEZ-SÁIZ, J. M. Mixture resolution according to the percentage of robusta variety in order to detect adulteration in roasted coffee by near infrared spectroscopy. **Anal. Chim. Acta** 2007a, 585, 266-276.
- 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.** 2007b, 55, 7477-7488.
- SAVITZKY, A.; GOLAY, M. J. E. Smoothing + Differentiation of data by Simplified Least Squares Procedures. **Anal. Chem.** 1964, 36, 1627-&.
- 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.
- WORKMAN, J. J. Interpretive spectroscopy for near infrared. **Appli. Spectros. Reviews** 1996, 31, 251-320.