



Multivariate calibration and moisture control in yerba mate by near infrared spectroscopy

Larize Mazur^{1*}, Garbrieli Alves de Oliveira¹, Milene Oliveira Pereira Bicudo¹, Rosemary Hoffmann Ribani¹, Noemi Nagata² and Patrício Peralta-Zamora²

¹Programa de Pós-graduação em Engenharia de Alimentos, Universidade Federal do Paraná, Av. Cel. Francisco H. dos Santos, s/n, 81531-980, Curitiba, Paraná, Brazil. ²Programa de Pós-graduação em Química, Departamento de Química, Universidade Federal do Paraná, Curitiba, Paraná, Brazil. *Author for correspondence. E-mail: larizemazur@yahoo.com.br

ABSTRACT. This work describes the development of a multivariate model based on near infrared reflectance spectroscopy (NIR) and partial least squares regression for the prediction of the moisture content in yerba mate samples. The multivariate model based on derivatized and multiplicative sign correction (MSC) spectral signals (4000-8500 cm⁻¹) was elaborated with 3 latent variables, allowing the fast evaluation of the moisture content with average prediction errors of about 2.5%. The minimal manipulation of the samples permits a high analytical speed that facilitates the implementation of quality control operations.

Keywords: NIR, online analysis, PLSR.

Controle de umidade em erva-mate por calibração multivariada e espectroscopia de infravermelho próximo

RESUMO. Este trabalho descreve o desenvolvimento de um modelo multivariado baseado na espectroscopia de reflectância no infravermelho próximo (NIR) e na regressão por quadrados parciais para a previsão do teor de umidade em amostras de erva-mate. O modelo multivariado, usando derivação seguido de correção multiplicativa do sinal (MSC) nos espectros (4000-8500 cm⁻¹), foi elaborada com três variáveis latentes, o que permite a avaliação rápida do teor de umidade com erros de previsão média de cerca de 2,5%. A manipulação mínima das amostras permite alta velocidade de análise o que facilita a realização das operações de controle de qualidade.

Palavras-chave: NIR, análises on-line, PLSR.

Introduction

Quality control operations are extremely important in all industrial processes, in accordance with the necessity to ensure quality and safety of marketed products. This is of particular relevance in the production of foodstuffs, due to quality and safety being closely related to issues of public health.

In recent years, the availability of instrumental techniques that have the ability to provide numerous signal-responses, combined with the widespread use of computers and availability of specialized software, has transformed the operations of control, particularly in the development of routines in multivariate analysis (BOGOMOLOV, 2011). However, although the multivariate calibration domain is nearly universal, with about 1300 applications registered between 1990 and 2011 (Source: sciencedirect.com, Keyword: Multivariate Calibration), few (32 articles) studies have been published in the area of food (Source:

sciencedirect.com, Keywords: Multivariate Calibration and Foods) and fewer than (5 articles) on the control of food quality (Source: sciencedirect.com, Keywords: Multivariate Calibration and Foods Quality Control).

Recent studies have shown the convenience of multivariate methods of analysis, mainly when directed to the analysis of complex matrices, in the simplification of conventional wet-way routines. Within this context, attention should be given to spectroscopic methods of analysis (electronic and infrared spectroscopy), which in combination with tools of multivariate analysis, have provided many relevant applications in the area of food quality control (FERNÁNDEZ-CABANÁS et al., 2011; LI et al., 2007; RIBEIRO et al., 2011; SINIJA; MISHRA, 2009; SHAO et al., 2011).

Although, various interesting reviews and applications have been recently published, it is noted that the potential application of multivariate

spectroscopic methods has not been adequately explored for quality control operations of food products. Thus, this is the motivation for this work, which takes as an example, the determination of moisture in samples of yerba mate, using near infrared spectroscopy (NIR) in diffuse reflectance mode (FORINA et al., 2009; NI et al., 2011).

Yerba mate (*Ilex paraguariensis*) is a traditional and very typical South American beverage and has a significant input into the international market, due to its antioxidant, diuretic, stimulant, hypocholesterolemic and hepatoprotective properties (HECK; MEJIA, 2007). Although several studies have been conducted to evaluate the therapeutic properties and characterization of some of the physical and chemical properties of yerba mate (e.g., the antioxidant capacity (MEJÍA et al., 2010)), nothing was found in the literature about the multivariate spectroscopic measurement of moisture, the influence of which can be significantly reflected in the quality of the product (JENSEN et al., 2011).

Material and methods

Yerba mate samples

A total of 38 commercial yerba mate samples were provided by the Ervateira Baldo S.A. Trade, Industry and Exportation of São Mateus do Sul, Paraná State, Brazil. Samples were collected from six different producers in the region, dried, fragmented and placed in polypropylene bags (500 g) with a weight of 0.062 gm^{-2} until the analysis was performed.

Moisture analysis

The moisture content of the samples was measured using the AOAC method (AOAC, 2007) that is based upon the loss of weight after the sample has been dried in a conventional oven. Between 3 to 5 g of yerba mate samples were placed in porcelain dishes, weighed and dried at 100°C to constant weight. The samples were cooled in a desiccator and weighed and the moisture content was determined by the loss of weight. For each sample, three repetitions of the moisture analysis were performed.

Near infrared spectroscopy

The NIR analysis was performed using a TENSOR™ 37 Fourier-Transform IR spectrometer (Bruker Optics Inc., Billerica, USA) in diffuse reflectance mode with a detector operating at a resolution of 4 cm^{-1} for all the readings. The spectra were collected in the range $10000\text{-}4000 \text{ cm}^{-1}$ and were displayed in terms of reflectance as $\log(1/R)$,

where R represents the reflected energy. Before scanning each sample, the background spectrum was taken against a blank optical path (reference spectrum) and stored in the computer. For each sample, a total of three spectra were measured.

Chemometric analysis

Spectra were exported from the Origin 6.1® (OriginLab Inc., Northampton, USA) in ASCII format into Matlab 7.0 software (Mathworks Inc., Massachusetts, USA) for chemometric analysis. The NIR spectra were transformed by first derivative and multiplicative sign correction (MSC) pre-treatment to correct scatter effects. The transformed spectra were analyzed by Partial Least Square Regression (PLSR) that was cross-validated (leave-one-out) to generate calibration models. In this study, 90 samples were used to generate the calibration models and 24 samples were used for external validation. Models were evaluated in terms of loading vectors, standard error of cross validation (SECV), correlation coefficient (R^2).

Results and discussion

Infrared spectra

It is important to note that infrared spectroscopy has advanced in recent years, which among other things, has allowed systems miniaturization based on diffuse reflection phenomenon. This facilitates the attainment of instrumental answers directly in solid phases, practically without any pretreatment of samples. Additionally, the spectral signals are relatively simple, corresponding to overtones and combination vibrations of C-H, O-H and N-H. In general, these signals have been used successfully in the development of multivariate calibration models for the evaluation of properties such as; lipids (SHIROMA; SAONA, 2009), sugars (SOROL et al., 2010), proteins (PI et al., 2009), and moisture (BRÁS et al., 2005) in various kinds of foods.

The NIR spectra of the 114 yerba mate samples used in the study are presented in Figure 1. Firstly, it is possible to observe a good homogeneity in all spectra, represented by low relevant analytical information in the region between 8000 and 10000 cm^{-1} and a characteristic signal in the region between 4000 and 8000 cm^{-1} . The large signal observed between 7000 and 8000 cm^{-1} and the signal centered at 5300 cm^{-1} can be attributed to the O-H group of water, whilst the signal near 6000 cm^{-1} can be attributed to first overtone of the C-H group, a signal usually associated with the presence of lipids. Other signs recorded between 4000 and 5000 cm^{-1} may correspond to vibrations of combination of C-H groups, mainly associated with

the presence of amino acid and fatty acids (COZZOLINO et al., 2010).

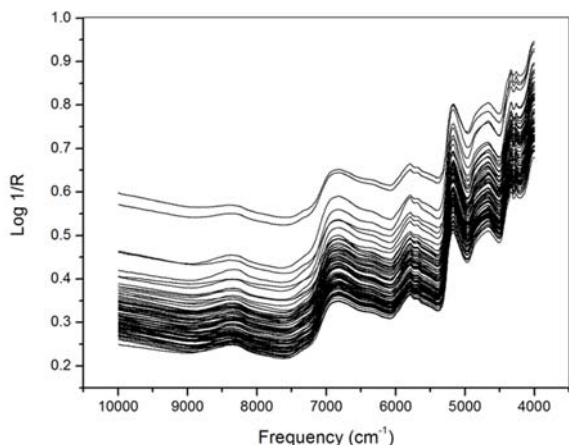


Figure 1. Near-infrared reflectance spectra of yerba mate samples.

Another important feature of the recorded spectral signals is represented by the near absence of instrumental noise, such as a relative equivalence in the relative magnitude of the most significant signals. These characteristics suggest, in a first analysis, that it is unnecessary to manipulate (pre-process) preliminary spectral data. However, it is quite common to use pre-processing routines to focus on the average (correction of baseline shifts), followed by smoothing derivation (disclosure of evidence of small magnitude), and of particular interest in signal processing obtained by diffuse reflectance, MSC, which minimizes the effects due to scattered light (SABIN et al., 2004).

Development of multivariate models

Multivariate calibration was preliminarily developed from 90 spectra, using different spectral regions and various types of pre-signal processing (mean centering, auto-scaled, smoothed-derived and submitted to MSC). In general, the best results were obtained in processing the spectral region between approximately 4000 and 8500 cm^{-1} , using a pre-processing MSC followed by derivation. Such processing has shown consistency, because mostly the derivation shows low intensity signs and allows removal of additive effects at baseline (CONZEN, 2006), whilst the MSC addresses the effect of light scattering caused by a lack of homogeneity in the samples (SABIN et al., 2004).

This model will be discussed in detail below as it provides the best preliminary results.

To define the optimal number of latent variables (LV), the internal cross-validation method was used,

particularly the routine called 'Leave one out'. In this procedure, a sample of the calibration set is excluded from the development phase of the model, being reserved as an element of prediction. This process is repeated n (numerous) times, so as to allow all calibration standards (n) to participate as elements of forecasting. Finally, the prediction error (Root Mean Square Error of Validation RMSECV) is obtained by comparing the predicted concentration for each standard with its true value, a value that is presented according to the number of LV used in the models preparation (Figure 2).

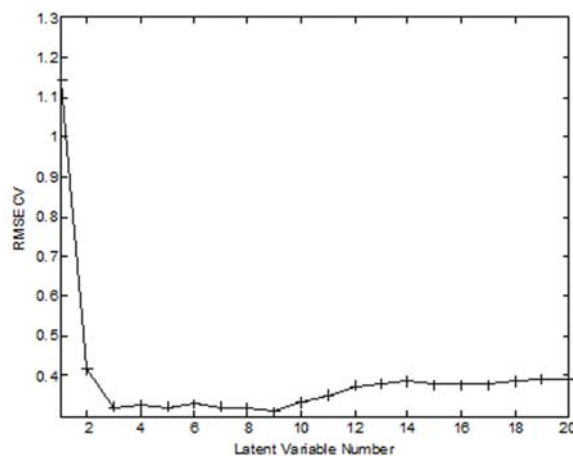


Figure 2. Evolution of the RMSECV value and the captured variance as a function of the number of latent variables.

From these results, it is possible to observe that three LV were necessary to describe the model which leads to smaller forecast errors in the internal validation phase, whilst it is observed that most of the variance concentration data (Y matrix) can be represented by only two LV.

It is important to observe that the selection of a small number of LV can delete an important variable (relevant analytical information) of the model, whilst a large number of LV can generate overestimated models and lead to the impairment of prediction in samples not involved in the set of calibration (FERREIRA et al., 1999). Thus, models were constructed with both 2 and 3 LV, observing the best prediction results were obtained with three LV.

In addition, it is very important to observe that 3 LV are responsible for representation of 99.87% of the concentration data variance, using the 99.90% spectral data variance. This means that all the spectral information, consisting of 875 original variables (frequencies) can be represented by only three LV, which result from the linear combination of original variables (MARTENS; NAES, 1989).

Another important aspect of optimizing the model is related to the identification of deficiencies in the calibration set. The 'leverage' criteria and Student residuals were used to identify anomalous samples (outliers). The first represents the influence of each of the samples in the regression model, with a threshold equal to $3 \text{ LV} / n$ (where LV represents the number of latent variables and n the number of samples), whilst the second indicates whether the sample is within a normal distribution with a confidence level of 95%, assuming threshold values of ± 2.5 (FERREIRA et al., 1999).

For the developed model with 3 LV (Figure 3), the limit value of 'leverage' (0.1) suggests no outlier data, whereas the limit value of the residuals indicates two outlier samples (32 and 83), probably due to errors in the moisture determination. In light of finding high residual values, both samples were removed from the calibration set. Normally, the exclusion of outlier samples provides a more homogeneous dataset, which allows the building of a more predictable, efficient and accurate model (FERREIRA et al., 1999; VANDEGINSTE et al., 1998).

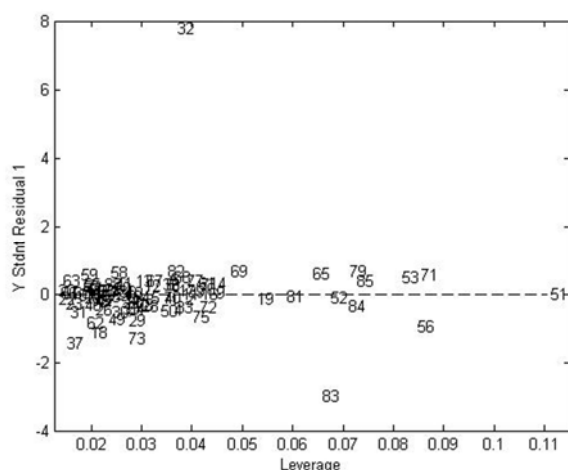


Figure 3. Leverage vs. Studentized residues chart (model developed with 3 LV).

Subsequently, the model was repeated with the same conditions (3 LV), using only 88 samples and under these conditions, no other outlier samples were observed. The value of R^2 cross validation found for this model was 0.96.

The predictability of the model was finally assessed by external validation, using the 24 samples reserved for this purpose. The results (Table 1) indicate that, with few exceptions, the multivariate method can reproduce the results of the reference method, with errors of less than 5%. In general, we

have observed that the forecast errors are of the same order of magnitude as the relative standard deviations from the standard method, which demonstrates the reliability of the multivariate analysis method considered.

Table 1. Moisture content in yerba mate samples of the external validation set.

Moisture content (%)		Error (%)
Method of reference (\pm standard deviation)	Multivariate method (\pm standard deviation)	
6.63	6.62	0.15
7.84	6.61 \pm 0.01	15.69
6.60	6.63	0.45
6.85	6.89	0.55
6.97	6.80 \pm 0.044	2.36
6.92	6.85	1.00
7.43	7.50	0.92
7.41	7.54 \pm 0.066	1.67
7.48	7.41	0.96
9.07	8.98	1.01
9.06	8.99 \pm 0.009	0.74
8.81	8.98	1.96
8.08	8.32	2.98
8.01	7.73 \pm 0.301	3.51
7.85	7.95	1.32
7.63	8.39	10.07
7.89	7.81 \pm 0.299	0.99
7.76	8.21	5.79
8.61	8.41	2.37
8.38	8.28 \pm 0.111	1.21
8.50	8.50	0.03
9.66	9.44	2.26
9.76	9.76 \pm 0.276	0.01
10.11	9.99	1.21
Average Error (%)		2.5%

To check the consistency between the spectral regions used to establish the regression and the parameter in question, Figure 4 presents the plot of regression coefficients as a function of the original variables (frequency). For comparison, Figure 4 also presents an original spectrum and the same submitted to the derivation.

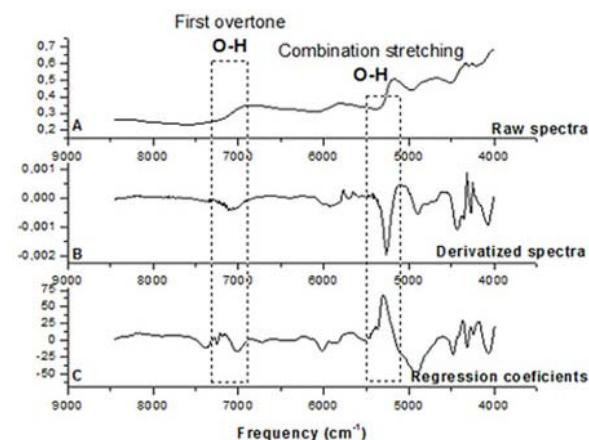


Figure 4. Near-infrared reflectance spectra of yerba mate (A), first-order derivative transformed spectra (B) and regression coefficients for the model developed with 3 LV.

It is possible to identify from figure 4, several regions of high correlation, especially for regions located at approximately 5250 and 7000 cm^{-1} . The first corresponds to the first overtone of the OH stretching, whilst the second corresponds to a combination band involving the OH group. This correspondence demonstrates the development of a model based on relevant spectral regions, not the generation of an artifact based on purely mathematical correlations.

Finally, it is important to note that the average error observed with the external validation (2.4%) is comparable with those reported in other studies of a similar nature (BRÁS et al., 2005; NI et al., 2011), which additionally demonstrates the convenient association between NIR and PLSR for the establishment of routine quality control, mainly due to the reliability of the results and the speed that comes from minimal sample manipulation.

Conclusion

Basically all the relevant analytical information present in the NIR spectra of yerba mate samples can be represented by a small number of latent variables, which facilitates the development of multivariable models of great foresight.

Models developed using spectral data (4000-8500 cm^{-1}) derivative, subject to correction by MSC, and based upon the use of three latent variables, allow a quick assessment of yerba-mate moisture content, giving average errors of the order of 2.5% during the external validation phase.

The proposed multivariate method proves convenient for implementation of routine quality control.

References

- AOAC-Association of Official Analytical Chemists. **Gravimetric method:** moisture. 18th ed. Washington, D.C.: AOAC, 2007.
- BOGOMOLOV, A. Multivariate process trajectories: capture, resolution and analysis. **Chemometrics and Intelligent Laboratory Systems**, v. 108, n. 1, p. 49-63, 2011.
- BRÁS, L. P.; BERNARDINO, S. A.; LOPES, J. A.; MENEZES, J. C. Multiblock PLS as an approach to compare and combine NIR and MIR spectra in calibrations of soybean flour. **Chemometrics and Intelligent Laboratory Systems**, v. 75, n. 1, p. 91-99, 2005.
- CONZEN, J. P. **Multivariate calibration.** A practical guide for the method development in the analytical chemistry. Ettlingen: Bruker Optick GmbH, 2006.
- COZZOLINO, D.; RESTAINO, E.; FASSIO, A. Discrimination of yerba mate (*Ilex paraguayensis* St. Hil.) samples according to their geographical origin by means of near infrared spectroscopy and multivariate analysis. **Food Quality and Safety**, v. 4, n. 2, p. 67-72, 2010.
- FERNÁNDEZ-CABANÁS, V. M.; POLVILLO, O.; RODRÍGUEZ-ACUÑA, R.; BOTELLA, B.; HORCADA, A. Rapid determination of the fatty acid profile in pork dry-cured sausages by NIR spectroscopy. **Food Chemistry**, v. 124, n. 1, p. 373-378, 2011.
- FERREIRA, M. M.; 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.
- FORINA, M.; CASALE, M.; OLIVERI, P. **Application of chemometrics to food chemistry, in comprehensive chemometrics.** Amsterdam: Elsevier, 2009.
- HECK, C. I.; MEJIA, E. G. Yerba mate tea (*Ilex paraguayensis*): a comprehensive review on chemistry, health implications, and technological considerations. **Journal of Food Science**, v. 72, n. 9, p. 138-151, 2007.
- JENSEN, S.; MELEIRO, L. A. C.; ZANOELO, E. F. Soft-sensor model design for control of a virtual conveyor-belt dryer of mate leaves (*Ilex paraguayensis*). **Biosystems Engineering**, v. 108, n. 1, p. 75-85, 2011.
- LI, X.; HE, Y.; WU, C.; SUN, D. Nondestructive measurement and fingerprint analysis of soluble solid content of tea soft drink based on Vis/NIR spectroscopy. **Journal of Food Engineering**, v. 82, n. 3, p. 316-323, 2007.
- MARTENS, H.; NAES, T. **Multivariate calibration.** Chichester: J. Wiley and Sons, 1989.
- MEJÍA, E. G.; SONG, Y. S.; HECK, C. I.; RAMÍREZ-MARES, M. Yerba mate tea (*Ilex paraguayensis*): Phenolics, antioxidant capacity and in vitro inhibition of colon cancer cell proliferation. **Journal of Functional Foods**, v. 2, n. 1, p. 23-34, 2010.
- NI, Y.; MEI, M.; KOKOT, S. Analysis of complex, processed substances with the use of NIR spectroscopy and chemometrics: Classification and prediction of properties. The potato crisps example. **Chemometrics and Intelligent Laboratory Systems**, v. 105, n. 2, p. 147-156, 2011.
- PI, F.; SHINZAWA, H.; OZAKI, Y.; HAN, D. Non-destructive determination of components in processed cheese slice wrapped with a polyethylene film using near-infrared spectroscopy and chemometrics. **International Dairy Journal**, v. 19, n. 10, p. 624-629, 2009.
- RIBEIRO, J. S.; FERREIRA, M. M. C.; SALVA, T. J. G. Chemometric models for the quantitative descriptive sensory analysis of Arabica coffee beverages using near infrared spectroscopy. **Talanta**, v. 83, n. 5, p. 1352-1358, 2011.
- SABIN, J. G.; FERRÃO, M. F.; FURTADO, J. C. Análise multivariada aplicada na identificação de fármacos antidepressivos. Parte II: Análise por componentes principais (PCA) e o método de classificação SIMCA. **Brazilian Journal of Pharmaceutical Sciences**, v. 40, n. 3, p. 387-396, 2004.
- SHAO, Y.; CEN, Y.; HE, Y.; LIU, F. Infrared spectroscopy and chemometrics for the starch and protein prediction in irradiated rice. **Food Chemistry**, v. 126, n. 4, p. 1856-1861, 2011.

SHIROMA, C.; SAONA, L. R. Application of NIR and MIR spectroscopy in quality control of potato chips. **Journal of Food Composition and Analysis**, v. 22, n. 6, p. 596-605, 2009.

SINIJA, V. R.; MISHRA, H. N. FT-NIR spectroscopy for caffeine estimation in instant green tea powder and granules. **LWT - Food Science and Technology**, v. 42, n. 5, p. 998-1002, 2009.

SOROL, N.; ARANCIBIA, E.; BORTOLATO, S. A.; OLIVIERI, A. C. Visible/near infrared-partial least-squares analysis of Brix in sugar cane juice: A test field for variable selection methods. **Chemometrics and Intelligent Laboratory Systems**, v. 102, n. 2, p. 100-109, 2010.

VANDEGINSTE, B. G. M.; MASSART, D. L.; BUYDENS, L. M. C.; DE JONG, S.; LEWI, P. J.; SMEYERS-VERBEKE, J. **Handbook of chemometrics and qualimetrics**. Part B. Amsterdam: Elsevier, 1998.

Received on July 2, 2012.

Accepted on January 30, 2013.

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.