

Classification of Brazilian mate tea using near infrared spectroscopy and chemometrics

S.S. Simões,^a T.C.B. Saldanha,^b F.A. Hansel,^c
C. Pasquini^d and K.M. dos Santos^d

^a*Departamento de Engenharia Química, Universidade Federal de Pernambuco, Pernambuco 50740-521 Brazil*

^b*Departamento de Química, Universidade Federal da Paraíba, Paraíba 58051-970 Brazil*

^c*Embrapa Florestas, Embrapa, Paraná 83411-000 Brazil*

^d*Instituto de Química, Universidade Estadual de Campinas, São Paulo 13084-971 Brazil*

Introduction

The tea made of mate is highly appreciated and consumed by people of South America including Brazil. The Brazilian mate product is usually classified in the mate companies by experts following the taste preferences of the consumers as “Bitter” (B) and “Mild” (M). Since mate products are increasing nowadays, the development of an instrumental technique for mate classification is required. At present this type of classification is performed by people with special skills, who are not able to classify the mate product processed in the mate industries quickly enough to achieve effective classification consistently. Therefore, a fast and robust technique for mate classification is a high requirement in the mate tea industry. Near Infrared (NIR) spectroscopy has been intensively used as source of information for qualitative purposes. An important advantage, is the relatively simple acquisition of reflectance spectra of natural samples such as ground mate leaves in the NIR spectral region. However, the efficient use of the information has become possible only by employing multivariate data analysis, due the high overlapping of the spectral features in the NIR wavelength region. This paper evaluates the use of NIR spectroscopy, allied to variable selection techniques and various multivariate algorithms, in order to classify the mate leaves according to the sensorial preference of the consumers. In particular, this work compares the performances of well-established classification algorithms (SIMCA, PLS-DA) with that of the Linear Discriminant Analysis (GA-LDA, SPA-LDA), also a well-established and robust classification algorithm, and variable selection made by Successive Projection Algorithm (SPA)¹ and Genetic Algorithm.

Table 1. Summary of results for classification of tea using different strategies of classification.

Classification algorithm	Variables used to construct the model*	Percentage of error in the test set samples
<i>SIMCA</i>	–	100%
PLS-DA	7	20%
GA-LDA	11	18
SPA-LDA	7	13%

*To PLS-DA models corresponds to latent variable and to GA-LDA and SPA-LDA to spectral variables.

Materials and methods

Two hundred samples of mate (*Ilex paraguariensis* St. Hilarie) leaves selected from two main producing regions of Paraná State, Brazil were dried in a commercial microwave oven for 5 minutes and pulverized to a particle size lower than 0.5 mm. The samples were previously classified as “bitter” (152) and “mild” (65) by an expert mate taster. The reflectance spectra were obtained in a Bomem 145 FT-NIR instrument in a range of 750 to 2500 nm by using the reflectance accessory (Powder SamplIR). This set of spectra was split into a calibration, validation and test representative subsets containing 99, 58 and 60 samples, respectively, by using the Kennard-Stone algorithm³. The spectral range from 750 to 1095 nm of NIR spectra was excluded due to poor signal-to-noise ratio observed in this region. Several data pretreatments were tested and the best results were obtained using the first derivative using the Savitzky-Golay polynomial algorithm (2nd order and 21 point window). Classification models were constructed based on SIMCA, PLS-DA, GA-LDA and SPA-LDA and the results were compared.

Results and discussion

Table 1 shows a summary of the results obtained by using all strategies for data treated using first derivative using Savitzky-Golay polynomial algorithm (2nd order and 21 point window). In general the performance of the Linear Discriminant Analysis Algorithms (GA-LDA and SPA-LDA) were better than the others classification algorithms. The use of SPA-LDA presented the best performance, with an error of 13% in the classification of test samples.

Reference

1. M. C. U. Araújo, T. C. B. Saldanha, R. K. H. Galvão, T. Yoneyama, H. C. Chame, V. Visani, *Chemometr. Intell. Lab. Syst.* **57**, 65 (2001).