Available online at www.sciencedirect.com
ScienceDirect
Talanta 71 (2007) 221–229
Talanta
www.elsevier.com/locate/talanta

# Coffee varietal differentiation based on near infrared spectroscopy

I. Esteban-Díez, J.M. González-Sáiz, C. Sáenz-González, C. Pizarro*

*Department of Chemistry, University of La Rioja, C/Madre de Dios 51, 26006 Logroño (La Rioja), Spain*

Received 21 November 2005; received in revised form 16 March 2006; accepted 23 March 2006
Available online 27 April 2006

**Abstract**

Near infrared spectroscopy (NIRS) was used to discriminate between *arabica* and *robusta* pure coffee varieties and blends of varied varietal composition. Direct orthogonal signal correction (DOSC) pre-processing method was applied on a set of 191 roasted coffee NIR spectra from both pure varieties and blends varying the final *robusta* content from 0 to 60% (w/w) in order to remove information unrelated to the actual varietal composition of samples. The corrected NIR spectra, as well as raw NIR spectra, were used to develop separate classification models using the potential functions method as class-modelling technique, exploring several options more or less restrictive according to the final number of considered categories. All constructed classification models were compared to evaluate their respective qualities and to show the suitability of applying DOSC method as pre-processing step for developing improved classification models for coffee varietal identification purposes.


*Keywords:* Food authentication; NIR spectroscopy; Direct orthogonal signal correction; Potential functions; Coffee varieties

## 1. Introduction

Food authentication is one of the most crucial issues in food quality control and safety. Food industry, regulatory authorities and consumer groups are all interested in authenticate raw materials and food products in order to satisfy the food quality and food safety requirements, in such a way that food quality assurance has become an essential tool to meet consumer demands and expectations [1,2].

Coffee identification or classification has gained increasing attention as a means to control and avoid coffee adulteration, mainly considering the great variability of the final sale price depending on coffee varietal or geographic origin. Most commercially available coffees are produced from *arabica* and *robusta* roasted beans or blends of these two species. Both varieties differ not only in relation to their botanical, chemical and organoleptic characteristics, but also in terms of commercial value, with *arabica* coffees achieving market prices 20–25% higher than *robusta*. Therefore, suitable methods are required, for quality and economical reasons, in order to differentiate these varieties and to avoid the possible mixing with other cheaper coffees, thus ensuring coffee authenticity.

Varietal classification for coffee authenticity has been tackled using different types of compositional data including metal content [3], volatile compounds [4], chlorogenic acids and caffeine content [5], fatty acids profile [6], sterolic profile [7], diterpenic alcohols [8], amino acids enantiomers [9], tocopherols and triglycerides [10]. In spite of the relative success showed by many of these approaches in coffee varietal identification, it is important to consider that many analytical reference methods used to assess the chemical components to be later used as discriminant parameters between coffee varieties in the classification model development may be quite expensive, elaborate and/or time-consuming.

For this reason, industry is looking for faster methods, and a rapid, clean and low cost technique, such as an automated classification on the basis of NIR spectra directly acquired on untreated samples could be a very useful tool. In fact, near infrared spectroscopy (NIRS) has emerged in the last years as a very promising alternative method for constructing on the basis of spectral features, and in combination with pattern recognition methods, reliable classification models for assessing the quality of a given product in many food applications [11–17], including several approaches applying NIR to the problem of coffee varietal authentication [18–20].

In this context and in order to search for an optimal classification model, some of us have recently proposed a strategy for developing improved classification models for differenti-

* Corresponding author. Tel.: +34 941299626; fax: +34 941299621.
*E-mail address:* consuelo.pizarro@dq.unirioja.es (C. Pizarro).

0039-9140/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2006.03.052

ate between *arabica* and *robusta* pure coffee varieties based on their NIR spectra and using potential functions Method as class-modelling technique [21]. In this former paper and to minimise certain physical light effects, thus enhancing the relevant chemical information contained in the spectra, two orthogonal signal correction methods were applied on raw NIR spectra to remove information not related to the caffeine content of samples, which was selected as response variable precisely due to its high discriminating power between coffee varieties. Although it was shown that original NIR spectra of roasted coffee samples might be used directly to develop a classification model with a moderately high discrimination ability between pure varieties, the classification models constructed after applying the orthogonal signal correction methods yielded excellent classification results also with a notable reduction in model complexity.

Therefore, given that this strategy appeared to have great promise for coffee varietal authentication purposes, we decided to deeper explore its actual applicability options, and to evaluate in the present work if a similar strategy (combination of direct orthogonal signal correction (DOSC)) method [22] and potential function class-modelling method [23–25]) can be successfully applied not only to discriminate between pure coffee varieties, but also between pure varieties and different blends of the two species on the basis of NIR spectra.

Moreover, it should be noticed that in this modified application no additional chemical variable need to be determined for those samples forming the calibration set in order to perform on its basis the orthogonal correction of spectra, since the response variable to be used in DOSC application is directly the percentage of *robusta* coffee contained in each sample, so it is easy to be realized the notable simplification and cheapening that this fact implies.

One decisive design parameter that has to be set prior to any classification model development is the number of categories to be considered and the particular requirements that a sample has to fullfit in order to be assigned to a certain class. In this study, two different approaches were explored differing in the number and particular properties of classification categories to be taken into account. In an initial and more flexible approach three separate classes were defined a priori: (1) pure *arabica*; (2) *arabica*–*robusta* blends; (3) pure *robusta*. Next, a more constrained approach taking into account five different categories was evaluated, in such a way that the only class previously defined as '*arabica*–*robusta* blends' was split into three separate classes in order to try to differentiate between blends with a low *robusta* content (from ≈5 to 20%), blends with a medium *robusta* content (from ≈25 to 40%) and blends with a high *robusta* content (from ≈45 to 60%). In both cases, for evaluating the effect of the orthogonal correction applied on NIR spectra on the quality of the final classification model constructed, the results obtained before and after transforming the spectra were analysed and compared.

It should be clarified that the aim of this study was not to provide a definitive and immutable NIR classification model for discriminating between pure coffee varieties and blends, but to propose a novel strategy capable of accomplishing this task and proving its reliability and effectiveness to assess the genuineness of coffee samples. Although the data set used in this work was designed to cover insofar as possible the great natural variability inherent to commercially available coffees by considering different roasting conditions and degrees, and varied geographical origins, it is quite clear that the dynamic nature of the coffee market and the particular needs and production lines of a given coffee company would demand a more exhaustive or specific collection of the calibration samples to develop the final class-model with the methodology here presented.

## 2. Experimental

### 2.1. Samples

The data set used in the present study comprised 83 roasted coffee samples from varied origins and varieties (36 *arabica* and 47 *robusta* coffees), which were processed under different roasting conditions. In addition, 108 blends of *arabica* and *robusta* coffee varieties were prepared in laboratory by combining the three coffee samples most representative of each variety which were previously selected. The application of PCA on mean centred NIR spectra provided an effective and easy-to-implement tool for selecting representative samples from *arabica* and *robusta* varieties. Fig. 1 shows a bidimensional representation of PC1 and PC2 scores accounting for 88.33% of the variance in the roasted coffee NIR spectral data, labelled according to their coffee variety: (1) *arabica* coffees; (2) *robusta* coffees. Two sample groups appeared slightly separated by the first bisectrix of the two component axes, suggesting the presence of two different clusters just associated with the two varieties considered. Thus, the centroid and two extreme samples within each class were selected in order to later generate on their basis suitable coffee blends with a *robusta* content in the final blends ranged from 0 to 60% (w/w). This specific range of blending (based on adding increasing *robusta* amounts to final blends) was precisely studied as an attempt to develop reliable classification models mainly focused on detecting *arabica* coffees

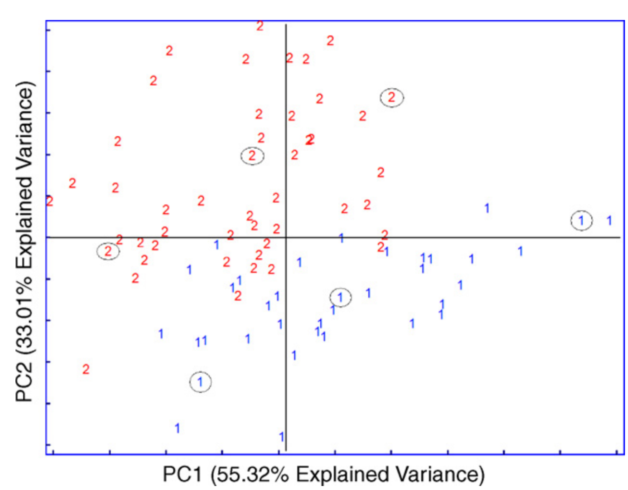


Fig. 1. Scores of the 83 roasted coffee samples from *arabica* (labelled as 1) and *robusta* (labelled as 2) pure varieties on the first two principal components explaining the variability in the NIR spectral data.

adulteration, since when a fraudulent practice is committed in the elaboration procedure of coffee blends from *arabica* and *robusta* pure coffee varieties the motivation behind is purely economical, i.e., to reduce costs due to the lower price of *robusta* coffee as compared to *arabica* coffee, increasing the actual *robusta* percentage in the final blends despite they will be sold as mostly based on *arabica* variety.

The resulting data set was split into two independent subsets: a calibration set with 100 samples and a test set with 91 samples. The main cautions taken in order to select a suitable composition of the external test set were to include samples of both pure varieties and different compositional blends, in such a way that the contained blends covered the whole range of *robusta* percentages studied.

### 2.2. Apparatus and software

NIR spectra were recorded on a near infrared spectrophotometer NIRSystems 5000 (Foss NIRSystems, Raamsdonksveer, The Netherlands) equipped with a reflectance detector and a sample transport module. The instrument was controlled by a compatible PC, and Vision 2.22 (Foss NIRSystems, Raamsdonksveer, The Netherlands) was used to acquire the data.

Data pre-processing treatments and potential functions class-modelling technique were applied by means of V-PARVUS 2004 (M. Forina et al., Dipartimento di Chimica e Tecnologie Farmaceutiche ed Alimentari, Università di Genova, Italy). The DOSC routine was implemented in MATLAB 6.5 (Mathworks, Natick, USA). Specifically, the DOSC method developed by Westerhuis et al. [21] was used for the DOSC calculations. Data for isopotential lines obtained from PARVUS were later mapped using Surfer 8 (Golden Software Inc.).

### 2.3. Recording of NIR spectra

Reflectance spectra were obtained directly from untreated samples. Due care was taken to ensure that the same amount of sample was always used to fill up the sample cell. Each spectrum was obtained from 32 scans performed at 2 nm intervals within the wavelength range 1100–2500 nm, with five replicates for each individual sample. The samples were decompacted between recordings. An average spectrum was subsequently computed from the collected replicates.

### 2.4. Validation of classification models

Usually, the potential functions class-modelling technique validates the predictive ability of the constructed classification models by cross-validation, since when working with potential functions is not recommended to waste objects to make up an external evaluation set. It is, however, well-known that orthogonal signal correction methods can produce a notable overfitting when applied on the spectra forming the calibration set. For this reason, although all potential functions classification models were constructed by cross-validation, we decided to also validate the actual predictive abilities of resulting models by testing their performance on an external test set to guard against overfitting.

### 2.5. Data processing

The whole data matrix was composed of 191 objects, 700 spectral variables (NIR absorbance values within the wavelength range 1100–2500 nm), and a response variable (percentage of *robusta* content in samples) to be used in the orthogonal correction. The initial set of 191 roasted coffee samples was divided into two subsets: the calibration set (100 objects) used to develop the classification models and the external test set (91 objects) used to evaluate the actual predictive ability of the constructed models.

The models used to classify roasted coffee samples were constructed by using the potential functions method in its modified form as a class-modelling technique. The optimal value of the smoothing parameter was selected by means of a cross-validation procedure, in such a way that the amplitude of each individual potential, defined by this smoothing parameter, was the same for all the objects in the category (fixed potential functions). Model boundaries were computed from the estimate of the equivalent determinant. The class-models were constructed at a level of significance corresponding to 95%. The same a priori class probability was applied to both categories (equal to 1). When classification models were developed on the basis of NIR data, given the large number of spectroscopic variables, a prior step of dimensionality reduction computing a small number of principal components was required. One crucial step in modelling based on NIR spectra is the selection of the optimal number of PCs to be used in the model development. The optimal complexity of each model was assessed by cross-validation (all the classification models were built by cross-validation using five deletion groups). When the DOSC method was applied to find classification models with a high predictive ability and avoid over-fitted solutions, the suitable orthogonal correction degree to be applied was determined as follows: the number of orthogonal-PCs to be removed from raw spectra was varied from 1 to 5, in such a way that the optimal number of orthogonal components to be subtracted was chosen according to the results obtained in the cross-validation procedure. All models were constructed on centred data.

The quality of the results provided by the different class-models constructed was compared according to several evaluation parameters:

- total classification (prediction) rate (TR)

$$\mathrm{TR} = \frac{\sum_c m_{\mathrm{cc}}}{N} \quad (1)$$

- category $c$ rate ($R_c$)

$$R_c = \frac{m_{\mathrm{cc}}}{N_c} \quad (2)$$

These equations were applied in both classification and prediction, where $m_{\mathrm{cc}}$ is the number of correct classifications (predictions) for a certain category $c$, $\sum_c m_{\mathrm{cc}}$ is the total number of

Table 1
Percentages of correctly classified samples

| PCs | Classification (%) | | | | External prediction (%) | | | |
|---|---|---|---|---|---|---|---|---|
| | RC1 | RC2 | RC3 | TR | RC1 | RC2 | RC3 | TR |
| Raw NIR spectra | | | | | | | | |
| 1 | 19.1 (17) | 43.1 (33) | 52.4 (10) | 40.0 (60) | 26.7 (11) | 44.0 (28) | 65.4 (9) | 47.3 (48) |
| 3 | 81.0 (4) | 93.1 (4) | 81.0 (4) | 88.0 (12) | 86.7 (2) | 98.0 (1) | 88.5 (3) | 88.5 (6) |
| 5 | 85.7 (3) | 94.8 (3) | 95.2 (1) | 93.0 (7) | 93.3 (1) | 94.0 (3) | 100 | 95.6 (4) |
| 7 | 85.7 (3) | 98.3 (1) | 100 | 96.0 (4) | 93.3 (1) | 94.0 (3) | 100 | 95.6 (4) |
| DOSC-corrected NIR spectra (three orthogonal-PCs removed) | | | | | | | | |
| 3 | 100 | 100 | 100 | 100 | 100 | 100 | 96.2 (1) | 98.9 (1) |

Total rate (TR) and category rates (RC1, RC2 and RC3) both in classification and external prediction, working on original and DOSC-corrected spectra. The number of samples incorrectly classified appears in brackets.

correct classifications (predictions), $N_c$ is the number of classifications (predictions) for a certain category $c$ and N is the number of total classifications (total predictions). It should be noticed that $N_c$ does not always equal the number of objects belonging to class $c$, in the same way as $N$ does not always represent the number of total objects, since, for instance, during cross-validation, an object can be classified several times.

Graphical tools, such as isopotential lines and Coomans plots, were also used to analyse the goodness of the models.

## 3. Results and discussion

### 3.1. Three-class approach

In the first part of this study three separate categories ((1) pure *arabica* coffees; (2) *arabica–robusta* blends; (3) pure *robusta* coffees) were considered, accounting for 191 samples that were divided at random into calibration (21 *arabica* coffees; 58 blends; 21 *robusta* coffees) and test (15 *arabica* coffees; 50 blends; 26 *robusta* coffees) sets.

Table 1 summarised the classification and prediction rates corresponding to the class-models developed on the basis of mean-centred raw NIR spectra of roasted coffee samples using the potential functions method with model complexities from 1 to 7 PCs for the three-class problem analysed. The table shows the influence on the number of PCs used to construct the class-model, in such a way that 7 PCs were used to compute the model to achieve the maximum correct classification rates in the cross-validation procedure, accounting for the 99.99% of the variance in the data. Table 2 shows the results corresponding to the same classification models developed from raw data expressed in terms of sensitivity and specificity for the three classes studied. Sensitivity is the proportion of samples belonging to a certain category correctly identified by the mathematical model corresponding to that class, i.e., it is a measure of the ability to correctly predict ‘true’ positives. Specificity is the proportion of samples not belonging to a certain class classified as foreign, i.e., it is a measure of the ability to discriminate against ‘false’ positives. In spite of the relatively good results obtained in both classification and prediction when considering the 7-PCs class-model developed from raw NIR spectra, the specificity of the model of category 1 for category 2 is 34.5%, whereas the specificity of the model of category 3 for category 2 is 10.4%, which indicates that many blends can be classified as ‘false’ pure varieties, leading to a high ‘false’ positives probability. Likewise, the observed specificity of the model of category 3 for category 1 is not neither too high (only 61.9%) which may suggest potential problems to discriminate even between pure varieties. These two parameters (sensitivity and specificity) are a very valuable diagnostic tool, since a class-model should not only accept samples belonging to the considered category but also it should reject foreign samples.

These numerical results can be also confirmed graphically. Fig. 2(a) shows the Coomans plot corresponding to the class-model computed with 7 PCs from raw NIR spectra. Bearing in mind the categories considered, the axes of the Coomans plot

Table 2
Sensitivity and specificity values for the potential functions models constructed for categories 1 (pure *arabica*), 2 (*arabica–robusta* blends) and 3 (pure *robusta*)

| PCs | Category 1 | | | Category 2 | | | Category 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Sensitivity | Specificity for class | | Sensitivity | Specificity for class | | Sensitivity | Specificity for class | |
| | | 2 | 3 | | 1 | 3 | | 1 | 2 |
| Raw NIR spectra | | | | | | | | | |
| 1 | 95.2 | 0.0 | 0.0 | 96.6 | 4.8 | 4.8 | 95.2 | 4.8 | 17.2 |
| 3 | 95.2 | 5.2 | 52.4 | 94.8 | 90.5 | 90.5 | 95.2 | 33.3 | 17.2 |
| 5 | 90.5 | 46.6 | 100 | 86.2 | 90.5 | 100 | 90.5 | 85.7 | 25.9 |
| 7 | 95.2 | 34.5 | 100 | 81.0 | 90.5 | 100 | 90.5 | 61.9 | 10.4 |
| DOSC-corrected NIR spectra (three orthogonal-PCs removed) | | | | | | | | | |
| 3 | 100 | 100 | 100 | 94.8 | 85.7 | 100 | 90.5 | 100 | 100 |

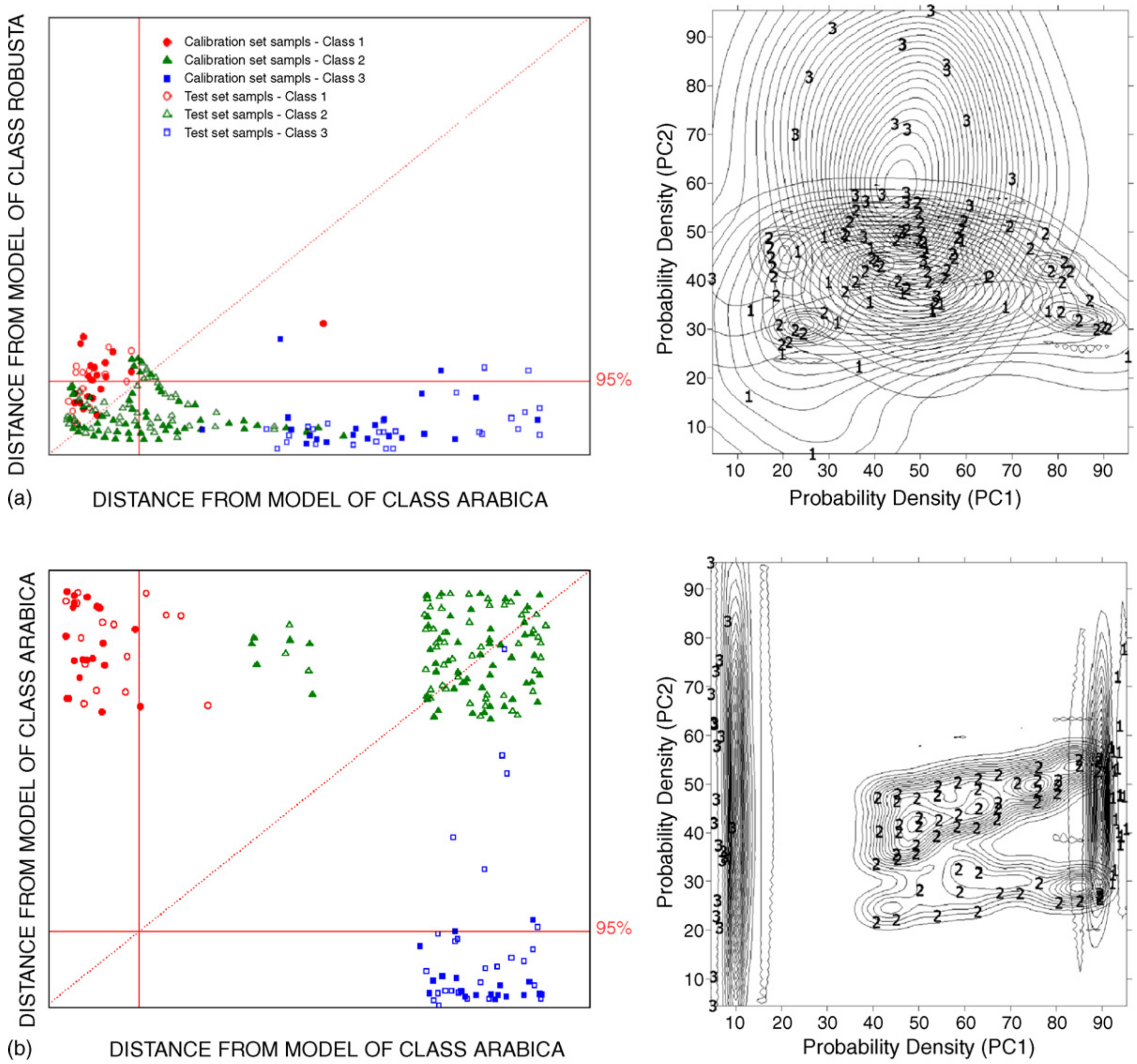


Fig. 2. Coomans and isopotential lines plots corresponding to a: (a) 7-PCs class-model developed from original NIR spectra, (b) 3-PCs class-model constructed from NIR spectra after removing three orthogonal PCs by DOSC; for the three-class problem analysed.

represent the class-models of *arabica* coffees (labelled as 1) in abscissas and *robusta* coffees (labelled as 3) in ordinates. Thus, the upper right quadrant will correspond to the samples rejected by the two models represented. In this case, the samples expected to be located in this quadrant would be the *arabica*–*robusta* blends (labelled as 2). Samples contained in the test set are plotted as 0. The large number of samples plotted in the class-space common to the two models representing pure

Table 3
Percentages of correctly classified samples

| PCs | Classification (%) | | | | | | External prediction (%) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | RC1 | RC2 | RC3 | RC4 | RC5 | TR | RC1 | RC2 | RC3 | RC4 | RC5 | TR |
| Raw NIR spectra | | | | | | | | | | | | |
| 1 | 0.0 (21) | 55.6 (8) | 16.7 (15) | 0.0 (22) | 0.0 (21) | 13.0 (87) | 13.3 (13) | 83.3 (3) | 16.7 (15) | 7.1 (13) | 34.6 (17) | 33.0 (61) |
| 3 | 66.7 (7) | 83.3 (3) | 72.2 (5) | 86.4 (3) | 76.2 (5) | 77.0 (23) | 86.7 (2) | 88.9 (2) | 72.2 (5) | 78.6 (3) | 88.5 (3) | 83.5 (15) |
| 5 | 85.7 (3) | 83.3 (3) | 83.3 (3) | 72.7 (6) | 95.2 (1) | 84.0 (16) | 86.7 (2) | 94.4 (1) | 66.7 (6) | 85.7 (2) | 100 | 87.9 (11) |
| 7 | 85.7 (3) | 88.9 (2) | 83.3 (3) | 68.2 (7) | 100 | 85.0 (15) | 93.3 (1) | 94.4 (1) | 66.7 (6) | 85.7 (2) | 100 | 89.0 (10) |
| DOSC-corrected NIR spectra (three orthogonal-PCs removed) | | | | | | | | | | | | |
| 1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

Total rate (TR) and category rates (RC1, RC2, RC3, RC4 and RC5) both in classification and external prediction, working on original and DOSC-corrected spectra. The number of samples incorrectly classified appears in brackets.

Table 4
Sensitivity and specificity values for the potential functions models constructed for categories 1 (pure *arabica*), 2 (low *robusta* content blends), 3 (medium *robusta* content blends), 4 (high *robusta* content blends) and 5 (pure *robusta*)

| PCs | Category 1 | | | | | Category 2 | | | | | Category 3 | | | | | Category 4 | | | | | Category 5 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Sensitivity | Specificity for class | | | | Sensitivity | Specificity for class | | | | Sensitivity | Specificity for class | | | | Sensitivity | Specificity for class | | | | Sensitivity | Specificity for class | | | |
| | | 2 | 3 | 4 | 5 | | 1 | 3 | 4 | 5 | | 1 | 2 | 4 | 5 | | 1 | 2 | 3 | 5 | | 1 | 2 | 3 | 4 |
| Raw NIR spectra | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 95.2 | 0.0 | 0.0 | 0.0 | 0.0 | 100 | 33.3 | 33.3 | 18.2 | 28.5 | 94.4 | 4.8 | 5.6 | 0.00 | 4.8 | 90.9 | 4.8 | 27.8 | 16.7 | 4.8 | 95.2 | 4.8 | 33.3 | 22.2 | 0.0 |
| 3 | 95.2 | 0.0 | 11.1 | 4.6 | 52.4 | 83.3 | 85.7 | 72.2 | 95.5 | 100 | 100 | 85.7 | 72.2 | 40.9 | 90.5 | 100 | 85.7 | 88.9 | 33.3 | 76.2 | 95.2 | 33.3 | 33.3 | 16.7 | 4.6 |
| 5 | 90.5 | 0.0 | 33.3 | 95.5 | 100 | 83.3 | 90.5 | 77.8 | 100 | 100 | 94.4 | 100 | 72.2 | 63.6 | 100 | 86.4 | 100 | 100 | 55.6 | 100 | 90.5 | 85.7 | 44.4 | 33.3 | 4.5 |
| 7 | 95.2 | 0.0 | 27.8 | 68.2 | 100 | 72.2 | 90.5 | 83.3 | 100 | 100 | 94.4 | 100 | 72.2 | 63.6 | 100 | 86.4 | 100 | 100 | 55.6 | 100 | 90.5 | 61.9 | 33.3 | 0.0 | 0.0 |
| DOSC-corrected NIR spectra (three orthogonal-PCs removed) | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 88.9 | 100 | 100 | 100 | 100 | 95.5 | 100 | 100 | 100 | 100 | 95.2 | 100 | 100 | 100 | 100 |

varieties demonstrated a low degree of specificity of the model based on raw spectra. It deserves special attention the fact that most samples belonging to the blends category appear in this area. The potential functions method also enabled us to obtain potentials for contour plots (isopotential lines plots) considering all samples and categories to be used as an additional visualizing tool. Fig. 2(a) displays the isopotential lines plot relative to the potential functions model of the three considered categories. As can be seen, class-models did not appear clearly separated, showing a high degree of overlapping between all categories.

In sight of the numerical and graphical results derived from the analysed raw spectra based classification model, it could be stated that, despite its acceptable discrimination ability which might be considered satisfactory for differentiating between *arabica* coffees, blends and *robusta* coffees, the lack of specificity showed between classes could reveal potential problems for classification of future samples. This finding stresses the relevance of the aims pursued in the present study, i.e., trying to improve the quality of the final classification model constructed in terms of both sensitivity and specificity to enable a more accurate practical application.

In order to try to solve the already discussed specificity problems associated with models based on raw data, DOSC was applied on NIR spectra taking into account the percentage of *robusta* variety contained in each sample as response variable and varying the number of orthogonal-PCs to be removed from 1 to 5 to determine the suitable orthogonal correction degree to be applied. The corrected resulting spectra were then used to develop the respective potential functions class-models. Considering all the models based on spectra corrected by the DOSC method, it was observed that when the number of orthogonal-PCs to be removed from the raw data increased, the quality of the respective potential functions models improved significantly and their corresponding complexity decreased once each orthogonal component had been removed. Nevertheless, this gradual improvement in class-model performance was limited, and thus, the model developed after removing three orthogonal-PCs by DOSC was considered the most suitable one, since further corrections did not provide any substantial advantage. The results corresponding to the classification model finally selected expressed as correct classification and prediction rates are shown in Table 1. It can be observed that the optimal class-model constructed after the removal of three orthogonal-PCs not only reduced the model complexity to only three components (explaining almost the 100% of the variance in the data), but also exhibited excellent results in both classification and prediction. It must be particularly underlined the great improvement achieved in terms of sensitivity, and more remarkably concerning specificity between categories (Table 2). Coomans and isopotential lines plots showed in Fig. 2(b) also confirmed the high quality of the classification model developed from DOSC-corrected NIR spectra for discrimination between pure coffee varieties and blends. Both plots showed a high degree of interclass specificity and a patently clear separation between class-models, considerably improved with regard to that provided by the model constructed from raw spectra.

### 3.2. Five-class approach

At this point, and without underestimating the very good results obtained in the first part of the study when considering only three separate categories to deal with the problem of varietal classification for coffee authenticity, we decided to study the same problem in more depth considering a higher number of a priori defined categories by limiting the qualitative requirements to be met for belonging to a certain class. The increase in the number of considered categories was aimed at discriminating, at least roughly between several degrees of *robusta* contents in a blend.

In this way, five separate categories were defined: (1) pure *arabica* coffees; (2) blends with a low *robusta* content; (3) blends with a medium *robusta* content; (4) blends with a high *robusta* content; (5) pure *robusta* coffees. The 191 total roasted coffee samples were split randomly into calibration (21 *arabica* coffees; 18 low *robusta* content blends; 18 medium *robusta* content blends; 22 high *robusta* content blends; 21 *robusta* coffees) and test (15 *arabica* coffees; 18 low *robusta* content blends; 18 medium *robusta* content blends; 14 high *robusta* content blends; 26 *robusta* coffees).

Table 3 shows the results in both classification and prediction (expressed as percentages of correctly classified samples) provided by the potential functions class-models developed from raw spectra, varying the number of PCs computed in the model development from 1 to 7 for the five-class problem studied. Table 4 summarised the results corresponding to the same class-models in terms of sensitivity and specificity percentages. Taking into account the numerical results collected in these tables, several conclusions can be drawn about the classification models developed from raw NIR spectra. Even though when a large number of PCs were used in the model construction (e.g. 7 PCs explaining for the 99.99% of the variance in the system), the fact of dividing the category of *arabica*–*robusta* blends (which initially has a 'global' nature) into three sub-categories according to the greater or lesser *robusta* content of samples gave rise to a notable worsening of the quality of the final classification

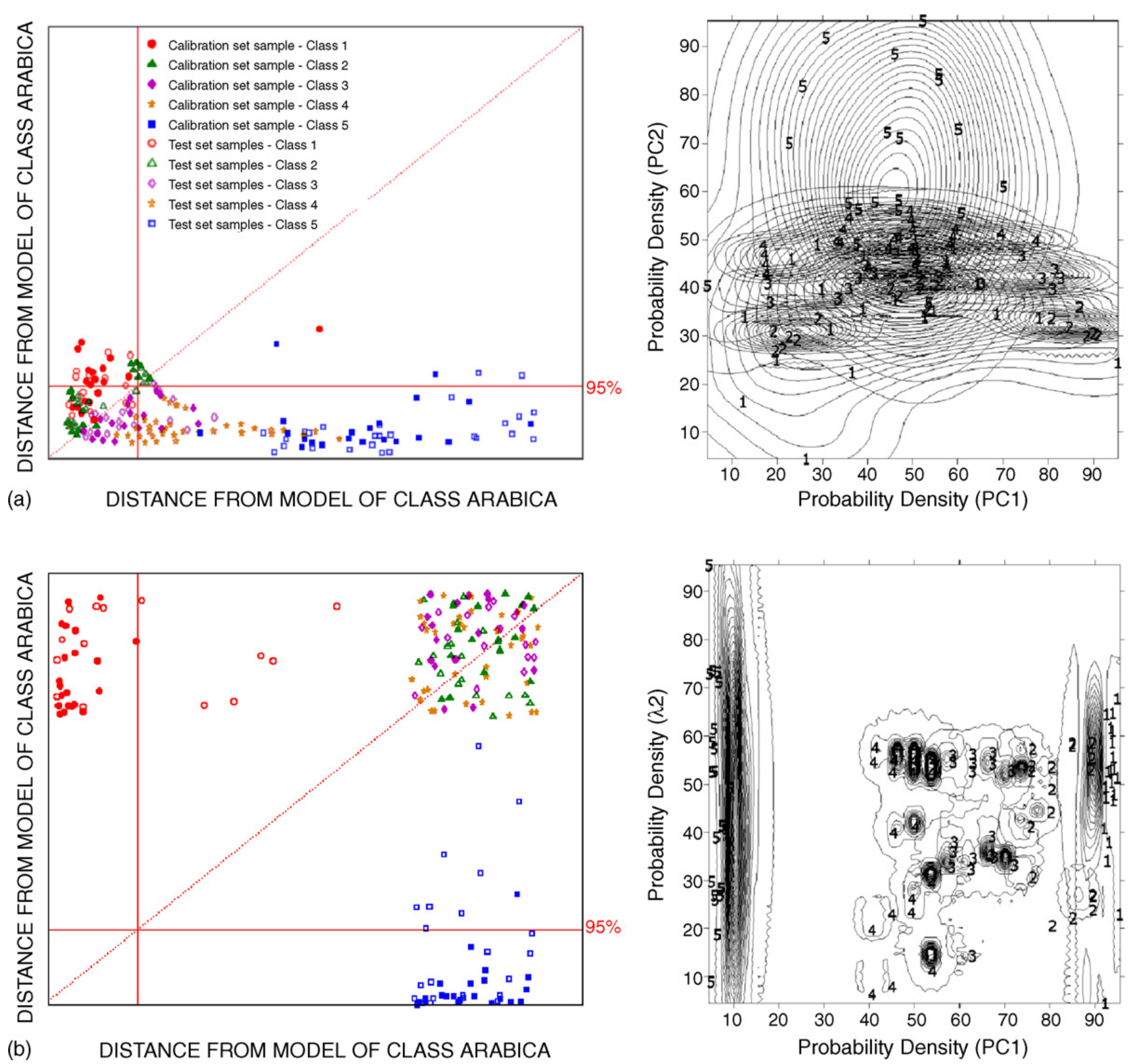


Fig. 3. Coomans and isopotential lines plots corresponding to a: (a) 7-PCs class-model developed from raw NIR spectra, (b) 1-PC class-model constructed from NIR spectra after the removal of three orthogonal PCs by DOSC; for the five-class problem studied.

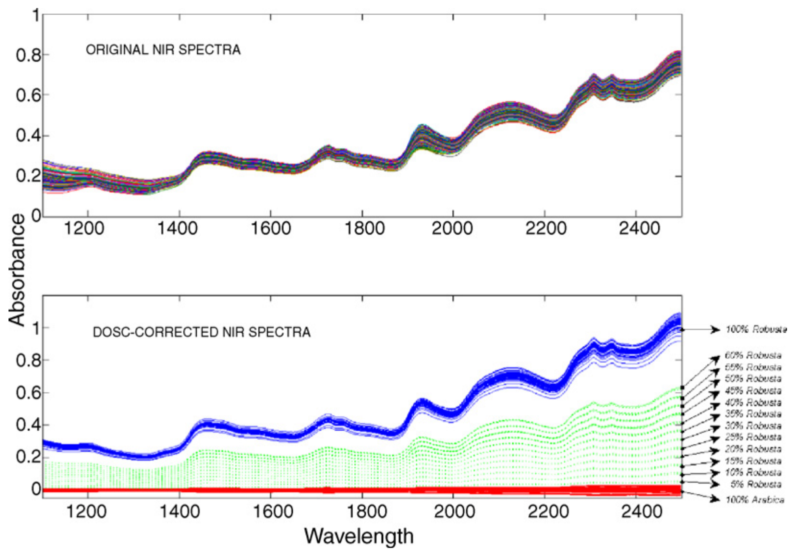


Fig. 4. Spectra of the 191 samples contained in the roasted coffee data set: (a) without pre-treatment and (b) after applying DOSC.

model. The category and total rates decreased appreciably in both calibration and prediction (particularly in the case of blends of medium and high *robusta* content), and the obtained specificity values between categories indicated an even higher risk of 'false' positives than in the case of only considering three classes. Fig. 3(a) show the Coomans plot and the isopotential lines plot corresponding to the 7-PCs potential functions model developed from raw data to discriminate between the five analysed categories. Again, in the Coomans plot, the class-model of *arabica* coffees (1) was represented in abscissas and the class-model of *robusta* coffees (5) was plotted in ordinates, so samples expected to be rejected by both displayed class-models (i.e., coffee blends belonging to categories 2, 3 and 4) should appear in the upper right quadrant. As in the previous case, external test samples are labelled as 0. From the samples distribution observed in the Coomans plot, it can be concluded that the model constructed on the basis of raw spectra was not able to neatly separate the five categories. The extremely strong overlapping between the five classes showed in the isopotential lines plot confirmed the serious lack of specificity of the model and its unsuitability to ensure an accurate classification.

Notably improved results were yielded by the potential functions model constructed after the application of DOSC preprocessing method on NIR spectra considering the *robusta* content in each sample as response variable to be used in the orthogonal correction (once selected the optimal number of orthogonal-PCs to be subtracted). As can be seen, the finally selected classification model developed after removing three orthogonal-PCs (we decided not to perform more exhaustive corrections because they did not provide any appreciable improvement in terms of model quality) and with the minimum complexity (only 1 PC explaining the 99.99% of the variance in the spectral data) not only provided 100% correct classifications in both calibration and prediction (Table 3), but also a great specificity between the five considered classes (Table 4), visually confirmed in Coomans and isopotential lines plot (Fig. 3(b)). Obviously, a slight longitudinal overlapping mainly between the three categories dealing with *arabica–robusta* blends, which can be observed in the isopotential lines plot, was not only expected but logical, due to the narrow margin that existed between the quantitative limits defining each one of these qualitative categories of blends.

### 3.3. Spectral profiles

The main objective of the present study was focused on minimising certain physical light effects that occur inherently in diffuse reflectance near-infrared spectroscopy, to develop improved classification models for coffee varietal authentication purposes.

The results obtained and discussed in this study, in both approaches analysed, have already demonstrated the practical usefulness and efficiency of applying an orthogonal signal correction method prior to the development of a high quality classification model. Nevertheless, a visual comparison between NIR spectral profiles before and after applying the orthogonal correction on roasted coffee NIR spectra (Fig. 4) can contribute to additionally show up the goodness of the strategy used here. A clear distinction is observed between the DOSC pre-treated spectra, but not for the original spectra, in such a way that corrected spectral profiles appeared perfectly grouped according to the particular percentage of *robusta* coffee contained in samples. Since DOSC method was applied directly on raw spectra (no pre-treatment was previously performed on data), it works as a kind of background subtraction: for the *arabica* coffee (0% *robusta* content), the spectrum is filtered away.

## 4. Conclusions

The results reported in this study have demonstrated that the combination of near-infrared spectroscopy with an orthogonal signal correction method and with a powerful class-modelling technique, such as potential functions method can be used as an optimal strategy not only for discriminating between *arabica* and *robusta* pure coffee varieties, but also for differentiating between pure varieties and blends of the two species, even when separate categories of blends were defined depending on the actual robusta content in samples. In both cases studied (varying the number of separate categories to be considered) a very substantial improvement in the global quality of the finally constructed class-model (particularly in terms of specificity between classes) was achieved in comparison with the respective model developed from original NIR spectra. The applied strategy did not imply any additional reference analysis, since the response variable to be considered in the orthogonal signal correction was precisely the percentage of *robusta* coffee in each sample (already known for calibration samples). Therefore, the applied classification strategy only relied on NIR measurements. This fact, together with the great results obtained, could make it very suitable for use in authenticity assessment of coffee. It should, however, be noticed that the work here presented is only a feasibility study, so further studies are required to properly evaluate its actual performance.

## Acknowledgements

The authors thank the Ministry of Science and Technology (Project No. 2FD1997-0491), the Autonomous Government of La Rioja – *Consejería de Educación, Cultura, Juventud y Deportes* (Project No. ACPI2000/08) and the University of La Rioja (Research grant FPI-2001) for their financial support, as well as Professor Michele Forina for providing us with the last version of Parvus package.

## References

[1] M. Lees (Ed.), Food Authenticity and Traceability, Woodhead Publishing Ltd., Cambridge, UK, 2003.
[2] C. Cordella, I. Moussa, A.C. Martel, N. Sbirrazzuoli, L. Lizzani-Cuvelier, J. Agric. Food Chem. 50 (2002) 1751.
[3] M.J. Martin, F. Pablos, A.G. Gonzalez, Food Chem. 66 (1999) 365.
[4] C.P. Bicchi, M.P. Ombretta, G. Pellegrino, A.C. Vanni, J. Agric. Food Chem. 45 (1997) 4680.
[5] M.J. Martin, F. Pablos, A.G. Gonzalez, Talanta 46 (1998) 1259.
[6] M.J. Martin, F. Pablos, A.G. Gonzalez, M.S. Valdenebro, M. Leon-Camacho, Talanta 54 (2001) 291.
[7] F. Carrera, M. Leon-Camacho, F. Pablos, A.G. Gonzalez, Anal. Chim. Acta 370 (1998) 131.
[8] N. Frega, F. Bocci, G. Lercker, Industrie Alimentari 34 (1995) 705.
[9] S. Casal, M.R. Alves, E. Mendes, M.B.P.P. Oliveira, M.A. Ferreira, J. Agric. Food Chem. 51 (2003) 6495.
[10] A.G. Gonzalez, F. Pablos, M.J. Martin, M. Leon-Camacho, M.S. Valdenebro, Food Chem. 73 (2001) 93.
[11] G. Downey, J. Near Infrared Spectrosc. 4 (1996) 47.
[12] D.G. Evans, C.N.G. Scotter, L.Z. Day, M.N. Hall, J. Near Infrared Spectrosc. 1 (1993) 33.
[13] B.G. Osborne, B. Mertens, M. Thompson, T. Fearn, J. Near Infrared Spectrosc. 1 (1993) 77.
[14] P.J. Gemperline, L.D. Webber, F.O. Cox, Anal. Chem. 61 (1989) 138.
[15] U.G. Indahl, N.S. Sahni, B. Kirkhus, T. Næs, Chemom. Intell. Lab. Syst. 49 (1999) 19.
[16] K.I. Hildrum, T. Isaksson, T. Næs, A. Tandberg, Near Infra-Red Spectroscopy. Bridging the Gap Between Data Analysis and NIR Applications, Ellis Horwood, Chichester, 1992.
[17] M.J. Sáiz-Abajo, J.M. González-Sáiz, C. Pizarro, J. Agric. Food Chem. 52 (2004) 7711.
[18] G. Downey, J. Boussion, D. Beauchêne, J. Near Infrared Spectrosc. 2 (1994) 85.
[19] G. Downey, J. Boussion, J. Sci. Food Agric. 71 (1996) 41.
[20] G. Downey, R. Briandet, R.H. Wilson, E.K. Kemsley, J. Agric. Food Chem. 45 (1997) 4357.
[21] I. Esteban-Díez, J.M. González-Sáiz, C. Pizarro, Anal. Chim. Acta 514 (2004) 57.
[22] J.A. Westerhuis, S. de Jong, A.K. Smilde, Chemom. Intell. Lab. Syst. 56 (2001) 13.
[23] D. Coomans, I. Broeckaert, Potential Pattern Recognition in Chemical and Medical Decision-Making, Research Studies Press, Letchworth, 1986.
[24] M. Forina, C. Armanino, R. Leardi, G. Drava, J. Chemom. 5 (1991) 435.
[25] D.L. Massart, L. Kaufman, The Interpretation of Analytical Chemical Data by the Use of Cluster Analysis, Wiley, New York, 1983.