Authentication of three coffee cultivars from Costa Rica by NIR: preliminary study

F. Davrieux[a], B. Bertrand, D. Bastianelli and B. Guyot

[a] *CIRAD Centre de Coopération Internationale en Recherche Agronomique pour le Développement, TA30/A, 34398 Montpellier cedex 5, France. Fabrice.davrieux@cirad*

Introduction

The continuing decline in coffee prices and the deterioration of coffee plantations has led Central American producers to reconsider their production, particularly their choice of arabica cultivars, with a view to winning new markets (*terroirs*, origins, etc.). In order to guarantee certification for these coffees and a quality bonus, reliable and rapid control methods need to be developed.

Near infrared reflectance spectroscopy was investigated as a means of discriminating between three cultivars (CR95, Caturra and Sarchimor) distributed in Costa Rica. Caturra is a traditional cultivar well known for its good cup quality. The other two cultivars are suspected of being poorer quality.

Experimental procedure

The study was carried out in 1999, on Arabica (*Coffea arabica* L.) trees of three cultivars (Caturra, CR95 and Sarchimor) planted at the Coffee Research Centre (CICAFE), Heredia, Costa Rica, in the same plot of land divided into three sets of 250 plants.

Harvesting was carried out in four steps (four picking dates 2/11/99, 1/12/99, 16/12/99 and 05/01/00). For each step, three degrees of fruit ripeness (yellow, red and blackish red) were collected. The 36 samples (three cultivars × four dates × three degrees of ripeness) underwent the same post-harvest treatment (wet processing, sun drying) to obtain green coffee. Six additional samples (four Caturra, one CR95 and one Sarchimor) were also available and kept for testing models.

Methods

Near infrared spectroscopy

NIR: Spectrometer NIRSystem 6500. Foss-Perstorp. Software: ISI NIRS 2 version 4.11 (InfraSoft International).

Three grams of homogenized ground green coffee were analysed, in random order, in diffuse reflectance from 400 nm to 2500 nm (2 nm steps). Each analysis was duplicated and the average spectrum saved.

Wet chemical methods

After extraction and purification, caffeine, trigonelline and sucrose contents were determined by HPLC (UV or electrochemical detector), chlorogenic acid contents by UV spectroscopy and fat and moisture contents by gravimetry.

Results

Wet chemistry values

Analysis of variance

An analysis of variance (Table 1) of this experimental design showed, in particular, that the three factors had no effect on the moisture content. This is important as the experimental design was set up all other things being equal, including post-harvest processing. There was a significant cultivar effect on caffeine, chlorogenic acid, fat and sucrose contents.

Table 1. Synthesis of the analysis of variance.

			Source of variation	
Dependent variable	Model	Cultivar	Harvest	Degree of ripeness
Dry matter	**None**	**None**	**None**	**None**
Caffeine	Yes	Yes	Yes	**None**
Trigonelline	Yes	**None**	Yes	**None**
Chlorogenic acids	Yes	Yes	Yes	Yes
Fat	Yes	Yes	Yes	**None**
Sucrose	Yes	Yes	Yes	Yes

Yes: significant for critical value (α) 0.05

Discriminat analysis

Linear discriminant analysis based on wet chemistry values (Figure 1) enabled classification of the samples according to cultivar groups with a success rate of 91.7%. This rate was 63% for harvest dates and 66.7% for degrees of ripeness.

It was thus only possible to partially discriminate between cultivars according to their biochemical composition.

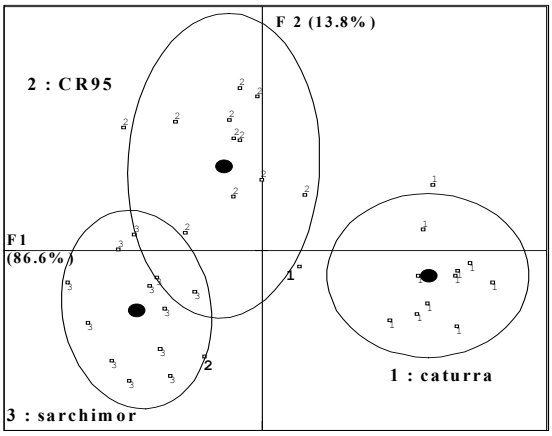

Figure1. Scatter plot of the sample scores for the two discriminant functions based on constituent contents (95% confidence ellipses).

Spectral data

These analyses were based on second derivatives of the spectra for wavelengths between 900 and 2500 nm (maths: 2,5,5,1). *SNVD* correction was applied to each spectrum.

Principal components analysis

The first to third PCs explained 54.75%, 30.88%, and 9.28% of total inertia respectively.

It was not possible to separate the cultivar groups (Figure 2), harvest dates and ripeness according to the sample scores for the different PCs.

Significant correlations between PCs and constituent contents (Table 2) were highlighted (for example, PC1 and fat content r = -0.81, PC2 and moisture content r = -0.89).

Table 2. Correlation Matrix between PCs and constituent contents.

	PC1	PC2	PC3	PC4	PC5	PC6
Caffeine	**0.382**	–0.234	0.324	–0.279	–0.289	–0.235
Dry matter	–0.316	**–0.886**	–0.308	–0.064	–0.047	–0.052
Trigonelline	**0.392**	0.088	0.257	**–0.604**	0.227	0.180
Chlorogenic acids	**0.505**	0.208	**0.490**	**–0.382**	–0.248	0.031
Fat	**–0.813**	**0.537**	0.042	–0.136	–0.140	–0.008
Sucrose	0.258	**–0.334**	–0.146	**0.575**	0.037	0.226

In bold, significant values. Critical value $\alpha = 0.05$. Pearson correlation

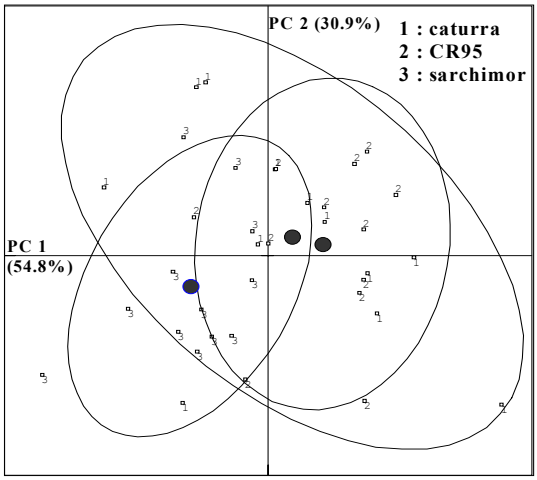

Figure 2. Scatter plot of sample scores for the two principal components (95% confidence ellipses).

Discriminant analysis using partial least squares regression

The PLS2 model was calculated using one dummy-variable for each group. Each of the variables was set to "2" for samples of the actual group and "1" otherwise. The optimum number of PLS terms was estimated using full cross-validation.

Seven PLS terms, with a *SECV* of 0.176, enabled the cultivars to be classified (Figure 3) with a success rate of 97.2%.

Six additional independent samples, from the same plantation, were successfully (100%) assigned to the right cultivar group.

Using this approach, sample classification according to harvest dates (55% of hits) and degree of ripeness (50% of hits) was not possible.

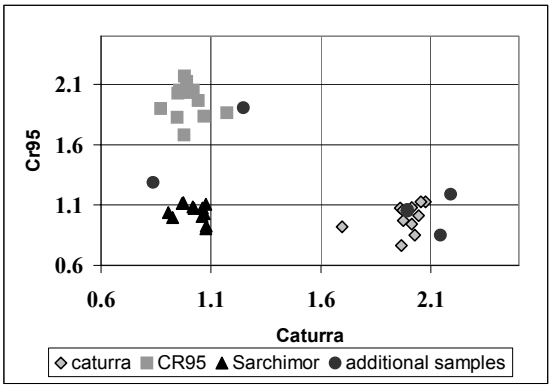

Figure 3. Scatter plot of the PLS2 predicted values for the 36 samples and the six additionnal samples. Score for Caturra and CR95 groups.

Discriminant analysis based on principal components

From the first eight PCs, extracted by principal component analysis (PCA) on the spectral data, a step by step linear discriminant analysis enabled cultivar classification with a 97.2% success rate. The leave-one-out cross-validation gave 88.9% of well-classified samples.

In the step by step LDA, at each step the variable that maximised the Mahalanobis distance between the closest two groups was introduced into the model. In this analysis seven PCs were introduced. The additional six samples were successfully (100%) assigned to the right cultivar group, according to their PCA scores obtained by projection on the PCA axes. The two discriminant functions explained 83.1% and 16.9% of total inertia (Figure 4).

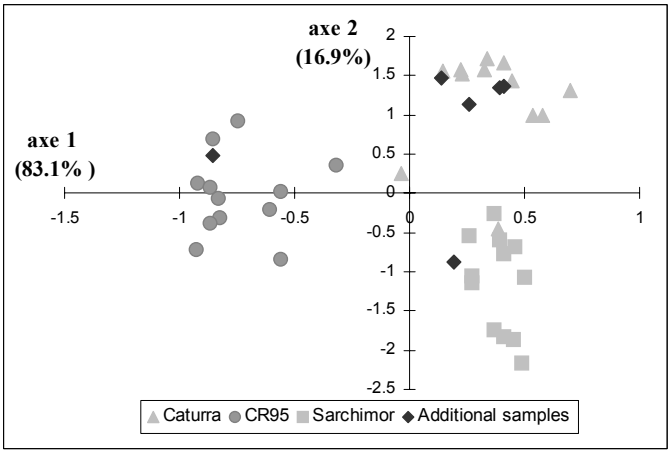

Figure 4. Scatter plot of sample sores for the two discriminant functions based on PCs.

The correlation between the PCs and the discriminant axes provided a way of identifying the weight of each variable in axis construction.

Axis 1 separating CR95 from Sarchimor-Caturra was correlated to PC1 (r = –0.43), which was itself correlated to fat content (Table 2). Axis 2 separating CR95-Sarchimor from Caturra was correlated to PC4 (r = -0.57), itself correlated to sucrose, trigonelline and chlorogenic acid contents.

Using this approach, sample classification according to harvest dates (52.8% of hits) and degree of ripeness (no variable retained) was not possible.

Conclusion

- Discrimination between the three cultivars based on biochemical contents was less efficient than that based on spectral data.
- The PLS2 and LDA approaches were similar: in both cases we tried to maximize between–group inertia and minimize within-group inertia. Moreover, the results were identical.
- Discrimination between the three cultivars was possible directly from spectral data for samples from the same plantation, which underwent the same post-harvest treatment, independently of the harvest date and degree of ripeness.
- These results are promising for establishing a way of checking green coffees. However, they need to be confirmed by increasing the number of samples, diversifying the geographical origins/post-harvest treatments and assessing the ability of the models to predict cultivar mixes.