

# *Coffea arabica* var. *laurina* Authentication Using Near Infrared Spectroscopy

F. DAVRIEUX<sup>1</sup>, B. GUYOT<sup>1</sup>, E. TARDAN<sup>1</sup>, F. DESCROIX<sup>2</sup>

<sup>1</sup>CIRAD PERSYST, Montpellier, France

<sup>2</sup>CIRAD PERSYST, St Pierre, La Réunion

## SUMMARY

On the island of Réunion, a project designed to develop a high added-value coffee brand was launched in 2002. The project focused on *Coffea Arabica* var. *Laurina* called 'Bourbon Pointu', and set out to determine growing practices and post-harvest treatments enabling the production of an inimitable coffee. The project involved the development of 106 experimental cultivation plots. The first phase of the project, from 2002 to 2007, was used to establish technical/commercial references for the production and processing of high added-value "grands crus" coffees. Soils were characterized and farms potentially suitable for the production of quality coffee were identified. Over that period more than 1126 samples were collected and analysed in diffuse reflectance from 1100 nm to 2500 nm using a NIRSystem FOSS 5000 spectrometer (Port Matilda, USA). In that way, a Near Infrared (NIR) spectral database specific to Bourbon Pointu (BP) coffees was compiled. A principal components analysis was performed on the spectral database in order to calculate Mahalanobis distances from the average spectrum for each spectrum and then fix boundaries for the population. Based on these results, samples differing from pure BP, such as "Bourbon rond", were removed from the database. The database was constructed by removing samples corresponding to abandoned trials such as certain post-harvest treatments or fertilizer tests. This led to a specific database for BP comprising 670 spectra of green coffees. From that collection, 250 random samples were analysed at the CIRAD laboratories, using reference methods, for their caffeine, chlorogenic acids, trigonelline, fat, sucrose and moisture contents. Those values were added to CIRAD's Arabica NIR database to develop specific calibration for BP coffees using Partial Least Squares Regression. This approach revealed that it was possible to authenticate BP coffee from Réunion using NIRS fingerprinting. The calibrations developed in this study made it possible to quantify the content of 6 major quality-related chemical compounds in BP green coffee.

## INTRODUCTION

The island of Réunion is characterized by a wide diversity of agricultural situations. The differences are due to biophysical characteristics, such as soil, temperature, rainfall and social aspects, including types of farming and production systems. The Coffee project falls under a long-term programme to create coffee sector production as part of the Réunion region agricultural diversification programme.

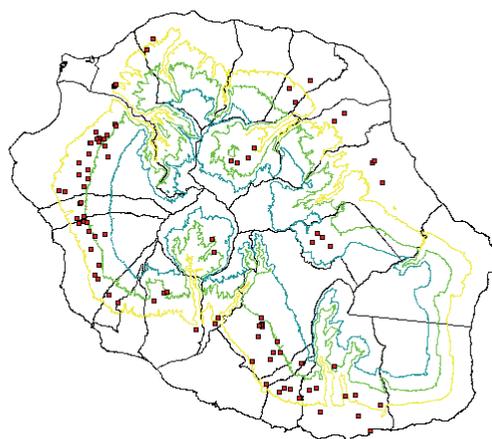
This project, scheduled over a 10-year period, has focused on *Coffea Arabica* var. *Laurina*, called 'Bourbon Pointu', and set out to determine growing practices and post-harvest treatments enabling the production of an inimitable coffee. The first phase of the project (2002-2007) was devoted to 1) the selection of coffee trees (within Réunionnais gardens) exhibiting specific agronomic, biochemical and organoleptic profiles, 2) identification of the different *terroirs*, 3) the elaboration of technical and economic references for coffee

production, 4) the study of technical-economic farming conditions suitable for the production of high quality coffee. This study was launched in 2005, the year when the 106 experimental cultivation plots produced their first yields. The aim of the study was to characterize Bourbon coffees for their biochemical properties using NIRS. The specific objectives were 1) compile a specific spectral database for Bourbon Pointu coffee (BP), 2) develop predictive equations for caffeine, moisture, trigonelline, chlorogenic acids, fat and sucrose contents. In this scheme, more than 1126 samples were collected and analysed for their NIR fingerprint, using a NIRSystem FOSS 5000 spectrometer (Port Matilda, USA). A NIR-predicted model developed for Arabica (Davrieux et al., 2003) coffees was applied to the database, samples with low or high predicted moisture contents and samples with predicted caffeine contents over 0.9% were discarded. The database was then structured using PCA analysis and Mahalanobis distances (GH) (Williams and Norris, 1990); samples with GH values over 3 were discarded. This led to a specific database for BP comprising 670 green coffee spectra. From that collection, 228 random samples were analysed for their caffeine, chlorogenic acids, trigonelline, fat, sucrose and moisture contents, at the CIRAD laboratories using a reference method. These values were added to CIRAD's Arabica NIR database to develop specific calibration for BP coffees using Partial Least Squares Regression. This approach showed that it was possible to authenticate BP coffee from Réunion using NIRS fingerprinting. The calibrations developed in this study made it possible to quantify the content of 6 major quality-related chemical compounds of BP green coffee.

## MATERIAL AND METHOD

### Plant material

The study was carried out over 3 production years, 2005, 2006 and 2007. A total of 1126 samples was collected over that period: 401 in 2005, 325 in 2006 and 400 in 2007. Sampling covered the 106 experimental orchards (Figure 1) and was representative of the different factors studied: *terroirs*, fertilizers, agricultural practices and post-harvest treatments. All the samples were wet processed.



**Figure 1. Réunion BP coffee experimental orchard (106) design.**

After sun drying, parchment coffees were stored in a climate chamber (60% RH and 28 °C). In that way, the expected moisture content should have been around 11%-11.5%. Parchment was removed prior to analysis (Africa Hullers McKINNON) and the coffee beans were graded (vibro grader, Spectrum Industries); only grade 14 beans were analysed. The beans were cooled with liquid nitrogen and ground (< 0.5 mm) using a Rescht ZM200 grinder.

## Near infrared spectroscopy

About 3 grams of homogenized powder were analysed in NIR using a FOSS 5000 spectrometer equipped with a transport module and small ring cups. Spectra were recorded as  $\log(1/R)$  in diffuse reflectance from 1100 nm to 2500 nm, in 2 nm steps.

The spectra were mathematically transformed using WINISI 1.5 software (Infrasoft International, Port Matilda, USA): a second derivative of the standard normal variate and a detrend corrected spectrum (SNVD) calculated on five data points and smoothed (Savitzky and Golay smoothing) on five data points.

The spectral population was structured (Shenk and Westerhaus, 1991) using a Principal Components Analysis (PCA) and Mahalanobis distances (H) calculated on extracted PCs. The matrix expression of H calculation was:  $H = X(X'X)^{-1}X'$ , where H was the matrix of H distances, X the matrix of centred spectra data and  $(X'X)^{-1}$  the reverse matrix of variance covariance. The Mahalanobis distance is the distance of each sample from the average sample and takes into account the total variability of the population. Generalized H distances (each individual distance divided by the average distance) were expressed as standard deviations, making it possible to define population limits and associate a probability with the H distance. A sample with an H distance over 3 had a probability of less than 1% of belonging to the population.

The parameters studied here were calibrated using Partial Least Squares Regression (mPLS). The calibration statistics used to evaluate model performances included the standard error of calibration (SEC), the coefficient of determination ( $R^2$ ), and the standard error of cross-validation (SECV). The cross-validation method resulted in calculation of the SECV. For SECV, 25% of the samples were used to validate a calibration model developed with the other 75%. SECV was repeated four times and the average calculated. The Student test (t) was used to identify t-outlier samples.

The statistics, multivariate analyses and regression model were performed using WINISI 1.5 software, STATGRAPHIC Centurion XV (StatPoint, Inc., Usa) and XLSTAT version 2008 6.02 (Addinsoft, Paris, France)

## Laboratory analyses

After extraction (water reflux with magnesium oxide), caffeine and trigonelline contents were determined by HPLC and UV detection at 280 nm. Polysaccharides were extracted with water (reflux) then separated and quantified using HPLC and pulsed amperometric detection. Total chlorogenic acids were extracted with methanol-water (70% w/w) then purified through a polyamide column, eluted using alkaline methanol and quantified using a spectrophotometer at 324 nm. Fat content was determined by gravimetry using a Soxtech (FOSS) extractor and petroleum ether. Moisture content was quantified by gravimetry at 103 °C for 16 hours using a Chopin oven.

## RESULTS AND DISCUSSION

The database was compiled with all the information relative to the coffee grown under the project: agronomic data, environmental data, genetic resources, post-harvest processing, qualitative data, chemical data and spectral data. The database was used to define a BP profile and understand the quality determinants. Based on this, a specific database for BP coffees was set up and used to authenticate new samples as BP. This approach required the compilation of

a database that was perfectly known (all kept samples certified BP), robust (all the possible cases foreseen, such as different cultural zones) and precisely defined (sampling, sample preparation and NIR analyses follow specific protocols).

### **Spectral database structuring**

Based on the previously defined criteria, the following methodology was applied to set up a database representative of certified BP coffees:

- 1) Selection of samples based on administrative data: samples from non-retained orchards, abandoned trials, specific and/or single trials, samples unclearly tagged and off-type samples (such as “Bourbon rond” coffee) were discarded.
- 2) Elimination of samples with an atypical  $\log(1/R)$  spectrum
- 3) A PCA analysis was carried out on the remaining samples and H distances calculated as a function of extracted PCs.
- 4) The remaining samples were predicted for their caffeine and moisture content using the NIR calibration developed at CIRAD for Arabica and BP coffees. A two-sided Grubb test with a 5% confidence level (Grubbs F., 1969), was performed on the predicted caffeine and moisture contents in order to identify and eliminate extreme values.
- 5) A final PCA was performed on the cleaned database in order to fix explained variability and calculated final H values for the optimum number of PCs adopted.

A total number of 1126 samples was analysed in NIR over 3 years, in accordance with the described procedure: 419 samples from abandoned trials or without clear tags, 2 samples with an atypical spectrum, 10 samples with H values over 3, 6 Arabica coffee samples from different origins, 8 samples with extreme caffeine values (Grubb’s test) and 11 samples with extreme moisture values were discarded.

After 2 iterations (2 successive PCA), the calculation of H values based on the 32 CPs extracted enabled the elimination of 9 outliers corresponding to specific *terroirs* (3), unidentified plant material (1), unsuitable soil (1) over-fermented and overripe coffees (4).

Grubb’s test, performed on predicted caffeine and moisture contents, identified 10 extreme caffeine values and 11 extreme moisture values respectively.

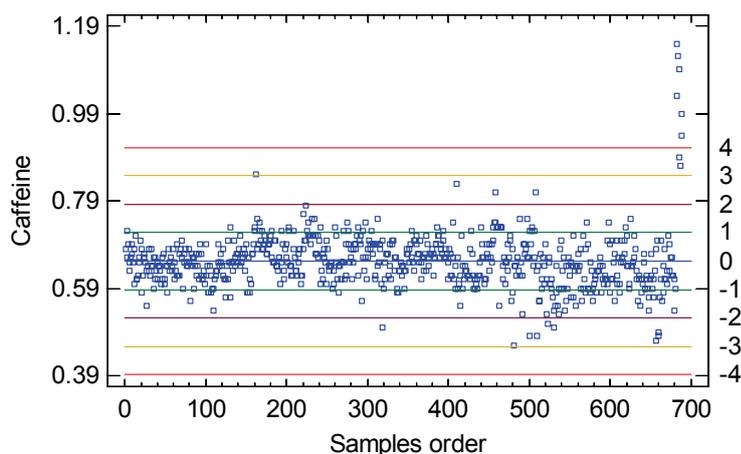
Among the 10 extreme caffeine values (Figure 2), 8 were over 0.85%, one equal to 0.85% and one under 0.46%. In accordance with the limit of 3 times the standard deviation (0.065%) around the mean value (0.65%), the 8 samples with a caffeine content over 0.85% were discarded. Those samples were tagged as “off-type” and corresponding to “Bourbon rond” and Arabica coffees grown in Réunion.

The 11 samples with extreme moisture values, 8 over 13.5% and 3 under 8%, were also removed from the database.

The final database, structured in this way, contained 670 samples: 131 in 2005, 323 in 2006 and 216 in 2007. The descriptive statistics for predicted values using the CIRAD equation are shown in Table 1; the average moisture value was constant over the year (11.4%), and the maximum value was 13.5%.

The caffeine content was between 0.46% and 0.85% with an average value of 0.65% constant over the year. The fat content ranged from 12% to 17%. Chlorogenic acids ranged from 5.7%

to 8.4% and the average value observed in 2007 (6.6%) was smaller than in 2005 and 2006 (7.1%).



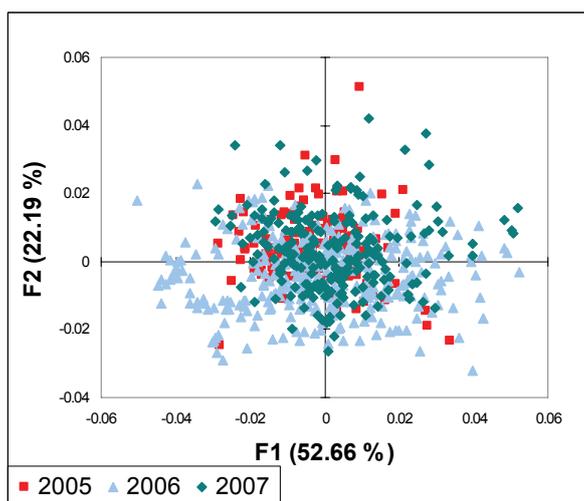
**Figure 2. Grubb’s test for extreme caffeine content values.**

The trigonelline content was constant over the year and ranged from 0.68% to 1.14%, with an average of 0.92%. The sucrose content increased from 2005 to 2007, from 7.6% /7.8% to 8.1% on average.

**Table 1. Descriptive statistics of the 670 BP coffee samples. Predicted values\*.**

	Caffeine	Trigonelline	Fat	Sucrose	Chlorogenic acids	Moisture
Minimum:	0.46	0.68	12.08	6.04	5.69	8.22
Maximum:	0.85	1.14	17.33	10.30	8.39	13.55
Mean:	0.65	0.92	15.14	7.88	6.94	11.35
Std. Deviation:	0.05	0.06	0.73	0.61	0.48	0.99
Total Values:	670	670	670	670	670	670

\*Expressed as a % of dry matter



**Figure 3. Scatter plot of the 670 sample scores for the first 2 PCs.**

A final PCA was performed on the 670 spectra; the first 3 PCs explained 87.6% of total inertia, 52.6%, 22.2% and 12.8% respectively. The H distances were calculated using 21 PCs, the maximum H value was equal to 2.7 with no outlier. The scatter plot of the sample scores

for the first 3 PCs (Figure 3) showed a uniform distribution of the samples without any year effect.

### Calibration

Of the 670 BP coffees, 250 random samples were analysed in the CIRAD laboratory for the 6 constituents and for their NIR spectra using CIRAD's FOSS 6500 spectrometer. These new data were merged with the CIRAD NIR Arabica database to create an Arabica database adapted to the specificities of BP coffees (Table 2).

Partial Least Squares Regression (PLSR) was used to establish quantitative relations between NIR spectral bands and caffeine, trigonelline, chlorogenic acid, fat, sucrose and moisture contents. The performances of the equations, in terms of  $R^2$ , SEC and SECV (Table 3) for caffeine, dry matter and fat, were highly satisfactory and the equations could be used for routine analyses. The predictive models developed for total chlorogenic acid and trigonelline contents were less efficient and those equations could be used for fine screening of samples based on these constituent predictions.

**Table 2. Descriptive statistics for wet chemistry Arabica – BP coffee data base.**

	Sucrose	DM	Caffeine	Trigonelline	Fat	Chlorogenic acids
Minimum:	5.26	85.10	0.45	0.67	10.46	5.10
Maximum:	10.66	93.99	1.79	1.39	17.70	11.04
Mean:	7.60	89.42	1.04	0.98	14.85	7.50
Std. Deviation:	0.94	1.49	0.32	0.11	1.52	0.90
N	551	831	606	496	356	466

The adjusted model for sucrose content was not satisfactory, probably due to a lack of fit in the wet chemistry method and the time delay between NIR and chemical analyses. This equation could be used for a rapid estimation of sucrose content and thereby the detection of extreme samples.

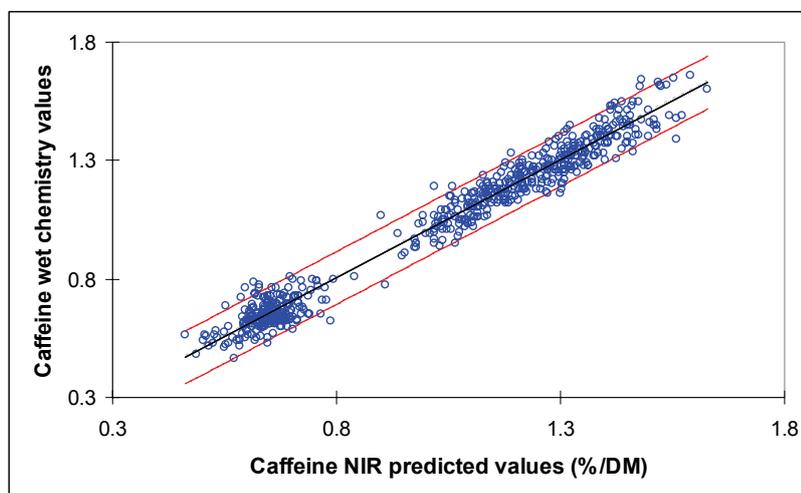
The scatter plot (Figure 4) for caffeine wet chemistry and NIR-predicted values highlighted the quality of the fit and showed the potential of the model for separating BP coffees with low caffeine contents from conventional Arabica coffees.

**Table 3. Statistics parameters for Arabica – BP coffee NIR equations.**

Constituent	N	Mean	SD	SEC	$R^2$	SECV
DM	751	89.49	1.47	0.13	0.99	0.14
Caffeine	577	1.03	0.32	0.05	0.97	0.06
Trigonelline	480	0.97	0.10	0.05	0.75	0.06
Fat	341	14.87	1.50	0.40	0.93	0.45
Sucrose	523	7.52	0.84	0.48	0.67	0.57
Chlorogenic Acids	449	7.48	0.84	0.42	0.75	0.47

*N*: Number of samples used for computation (*t* test); *SD*: standard deviation of reference values; *SEC*: standard error of calibration;  $R^2$ : coefficient of multiple determination; *SECV*: standard error of cross validation.

In the same way, PLSR models were developed for the 670 BP coffee spectra. Due to the reduced range in variation with the cleaned database specific to BP, the  $R^2$  obtained were low (except for moisture and fat,  $R^2 > 0.80$ ); for caffeine,  $R^2$  was equal to 0.15 with a SD equal to 0.06%. Although the SECV and SEC error terms were of the same order of magnitude (data not shown) as those obtained on the whole database (SECV 0.05% and 0.06% for caffeine), these models were not usable for routine analysis.



**Figure 4. Scatter plot for measured caffeine vs. NIR- predicted values.**

This was illustrated with the predicted caffeine content for Arabica and off-type samples removed from the BP coffee database, using the two models (Table 4).

**Table 4. Caffeine content for Arabica and off-type coffees with the 2 models.**

Samples	Predicted caffeine values (% DM)		Laboratory value	H distance vs BP database
	Global equation	BP equation		
Arabica1	1.18	0.584	1.27	10.5
Arabica2	1.07	0.563	1.19	15.6
Arabica3	1.15	0.646	1.25	3.9
Arabica4	1.11	0.648	1.19	3.3
Arabica5	1.15	0.659	1.24	2.1
Arabica6	1.15	0.661	1.17	2.2
Off-type1	1.15	0.694	ND	1.9
Off-type2	1.03	0.653	ND	2.5
Off-type3	1.12	0.701	ND	1.6
Off-type4	1.09	0.679	ND	0.9
Off-type5	0.89	0.623	ND	2.1
Off-type6	0.87	0.63	ND	5.0
Off-type7	0.94	0.683	1.09	2.2
Off-type8	0.99	0.693	ND	1.7

*ND: non determined.*

The predicted caffeine contents, using the specific BP coffee model, were underestimated for Arabica and off-type coffees, while they were correctly estimated using the global model based on numerous Arabica and BP coffee references.

At the same time, the H values calculated after projection on the final BP database scores were all higher than the outlier limit ( $H = 3$ ) for Arabica 1 and 2 (Costa Rica), Arabica 3 (India), Arabica 4 (Cuba) and under 3 for Arabica 5 (Ethiopia, Moka) and Arabica 6 (Brazil). Among the off-type samples, only one had an H value over 3; off-type sample 6 was a Bourbon rond coffee from Réunion.

These results showed the need to combine a qualitative approach (H value) with a quantitative approach (PLS models) to characterize a specific coffee. In our study, the combination of the 2 strategies led to perfect identification and selection of pure BP coffee from Réunion.

## CONCLUSION

The performance of the predictive models based on a broad database of Arabica coffees adapted to Bourbon pointu coffees from Réunion enabled fine characterization of these specific coffees, especially in terms of caffeine content ( $SECV = 0.06\%$ ,  $R^2 = 0.97$ ). The NIR fingerprint of BP coffee can therefore be used to identify fair commercial green BP coffees (mature, fair fermented, dried) without defects by comparison with a specific BP coffee database (H value) and through its predicted biochemical profile. This study demonstrated the need to have a well-known and highly variable database to develop efficient NIR equations; calibration for BP coffees only was not possible.

Further studies will try to strengthen authentication by combining NIR fingerprints, biochemical profiles and sensory profiles. We can imagine achieving a classification by production zone using NIR and sensory analyses.

## REFERENCES

- Barnes R.J., Dhanoa M.S. Standard normal variate transformation and detrending of near-infrared diffuse reflectance spectra. *Soc. Appl. Spectrosc.* 1989, 43 (5), 772-777.
- Davrieux F., Manez J.C., Durand N., Guyot B. (2004). Determination of the content of six major biochemical compounds of green coffee using near infrared spectroscopy. *Near Infrared Spectroscopy: Proceedings of 11 th International Conference*. Eds A.M.C. Davies, A. Garrido, NIR Publications Norfolk, UK, 441-444.
- Grubbs F. (1969). Procedures for Detecting Outlying Observations in Samples, *Technometrics*, Vol. 11, No. 1, pp. 1-21.
- Shenk J.S. and Westerhaus M.O. (1991). Populations Structuring of Near Infrared Spectra and Modified Partial Least Squares Regression. *Crop Sci.* 1991; 31: 1548-1555.
- Sinnaeve G., Dardenne P., Agneessens P., Biston R. The use of near infrared spectroscopy for the analysis of fresh grass silage. *J. Near Infrared. Spectrosc.* 1994, 2, 79-84.
- Williams P., Norris K., (1990). *Near Infrared Technology in the Agricultural and Food industries*. Eds. Williams P., Norris K., American Association of Cereal Chemists, Inc, St Paul, Minnesota, USA, 330 p.