# Use of near infrared spectroscopy in the development, qualification and certification of coffee Bourbon pointu

Fabrice Davrieux[1]*, Frédéric Descroix[1], Denis Bastianelli[2], Mathilde Hoarau[1] Jean-Michel Roger[3], Noel Durand[1] and Véronique Bellon-Maurel[3]

[1]CIRAD, UMR Qualisud, 34398 Montpellier cedex 05, France
[2]CIRAD, UMR SELMET, 34398 Montpellier cedex 05, France
[3]CEMAGREF- Montpellier Supagro, UMR ITAP, 34196 Montpellier cedex 5
*Corresponding author: davrieux@cirad.fr

## Introduction

On Reunion island a development project designed to develop a high added-value coffee brand was launched in 2002. The project focused on the cultivar *Coffea Arabica* var. *Laurina*, also called 'Bourbon pointu' (BP), with the aim of determining growing practices and post-harvest treatments which would enable the production of an inimitable coffee. The project involved the development of 113 experimental cultivation plots planted all around the island in various ecological conditions.

Producing coffee on Reunion Island is not new; the first coffee plants were introduced in the 18th century from Yemen. Coffee cultivation has contributed to the development of the island. Bourbon pointu is the result of a natural mutation of a Yemen Arabica coffee plant. Bourbon pointu has relatively low caffeine levels and is well known for its special citrus fruity taste.

The first phase of the project (2002-2007) was devoted to: 1) the selection of coffee trees exhibiting specific agronomic profiles, 2) the identification of the different *terroirs*, 3) the elaboration of technical and economic references for coffee production, and 4) the biochemical and sensorial characterisation of BP coffee. The second phase of the project (2007-2012) aimed to maintain and enhance BP quality, to increase production through new plantations and to obtain a protected designation of origin (PDO).

The present study was launched in 2005, when the 113 experimental cultivation plots produced their first fruits. The aim of the study was to characterise the biochemical properties of BP coffees using near infrared (NIR) spectroscopy. The specific objectives were: 1) compile a specific spectral database for BP coffee, 2) develop predictive equations for caffeine, moisture, trigonelline, chlorogenic acids, fat and sucrose contents, and 3) to support the PDO certification.

## Materials and Methods

This study was carried out over 6 years of production, 2005-2010. More than 2000 samples were collected over that period. Sampling covered the 113 experimental orchards (Figure 1) and was representative of the different factors studied: *terroirs*, fertilisers, agricultural practices and post-harvest treatments.

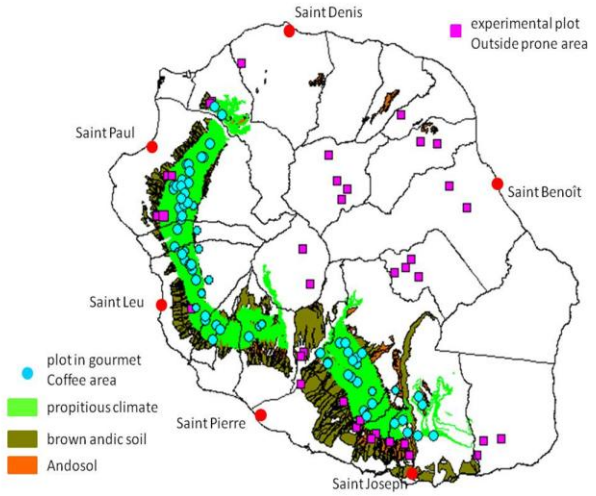


**Figure 1.** Reunion Island. BP coffee experimental orchards (113) design.

The database was compiled with all the information relevant to the coffee grown under the project: agronomic data, environmental data, genetic resources, post-harvest processing, qualitative data, chemical

Reference paper as:
F. Davrieux, F. Descroix, D. Bastianelli, M. Hoarau, J. Roger, N. Durand and V. Bellon-Maurel (2012).Use of near infrared spectroscopy in the development, qualification and certification of coffee Bourbon pointu, in: Proceedings of the 15th International Conference on Near Infrared Spectroscopy, Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 386-391.

and spectral data. The database was used to define a BP typical profile and understand the quality determinants. Based on previously defined criteria, the following methodology was applied to set up a spectral database representative of certified BP coffees: 1) selection of samples based on administrative data, i.e. elimination of samples from non-retained orchards, abandoned trials, specific and/or single trials, samples unclearly tagged and off-type samples (e.g. "Bourbon rond" coffee), 2) elimination of samples with an atypical spectrum, 3) PCA was carried out on the remaining database to remove samples with Mahalanobis H distances higher than 3, 4) specific calibrations (using PLS regression and external data set for validation) were developed for 6 constituents (caffeine, dry matter, chlorogenic acids, trigonelline, fat and sucrose) using a general Arabica calibration[1] and selected BP coffees, 5) a two-sided Grubb[2] test, with a 5% confidence level, was performed on the predicted caffeine and moisture contents in order to identify and eliminate samples with extreme values, and 6) a final PCA was performed on the cleaned database.

The accuracy and robustness of the caffeine calibration developed for BP were estimated by predicting a set of 48 BP and Arabica blends w/w. The blends covered the range between 0 and 100% w/w of Arabica, and completed the gap in caffeine content observed between BP (0.85% maximum) and Arabica coffees (1.01% minimum). In addition a set of 38 Arabica coffees from different origins was compared to the spectral database and their proximity to the BP coffees was estimated using the H distance value.

## Samples

All the coffee samples were wet processed. After sun drying, parchment coffees were stored in a climate chamber (60% RH and 28°C). Parchment was removed prior to analysis (Africa Hullers, Wm. McKinnon & Co. Limited, Aberdeen, United Kingdom and the coffee beans were graded (vibro grader, Spectrum Industries, Karnataka India); only grade 14 beans were analysed. The green beans were cooled with liquid nitrogen and ground (< 0.5 mm) using a Retsch ZM200 grinder. More than 2000 samples were analysed using NIR over 6 years. The first step consisted of removing samples from abandoned trials or without clear tags, and samples with atypical spectra. The second step, focused on samples which had H values higher than 3. This procedure was done after each year of sampling The final BP database consisted of 1254 samples: 219 in 2005, 451 in 2006, 102 in 2007, 165 in 2008, 204 in 2009 and 113 in 2010.

At the same time, a random selection of samples was carried out each year for laboratory analysis. The total number of BP samples analysed with reference methods was: 345 for DM, 237 for caffeine, 142 for trigonelline, 190 for fat and 200 for CGA.

The 48 blends between BP and Arabica coffees were mixed by weight (± 0.01 g) from powders, a minimum of 20 g was mixed per blend. Blends were homogenised using a home mixer coffee grinder (Seb, Ecully, France). Blends consisted of 4 different coffees taken from a set of 6 BP coffees and 8 different Arabica coffees (originating from Ethiopia, Brazil, Colombia, Nicaragua, Laos, Sumatra, China and Salvador).

An additional 38 pure Arabica coffees with different origins and varieties, 6 BP coffees grown in New Caledonia and one BP coffee from the Reunion island 2007 harvest, were kept as references and stored at -80°C.

## Near infrared spectroscopy

About 3 g of homogenised powder were analysed using a FOSS 5000 spectrometer (FOSS, Port Matilda, USA) equipped with a transport module and small ring cups. Spectra were recorded as log (1/R) in diffuse reflectance from 1100 to 2500 nm, in 2 nm steps. The spectra were mathematically transformed using WINISI 1.5 software (Infrasoft International, Port Matilda, USA). The second derivative of the standard normal variate and detrend corrected spectrum (SNVD) was calculated on five data points and then smoothed (Savitzky-Golay smoothing) over five data points.

The parameters studied were correlated with NIR spectra using partial least squares (PLS) regression. The calibration statistics used to evaluate model performances included the standard error of calibration (SEC), the coefficient of determination ($R^2$), the standard error of cross-validation (SECV) and the Ratio Performance to Deviation (RPD). The Student's t-test was used to identify outlier samples.

## Wet chemistry

After extraction (water reflux with magnesium oxide), caffeine and trigonelline contents were determined by HPLC and UV detection at 280 nm (LC20A binary pump and SPD-10AV detector, Shimadzu, Kyoto, Japan). Sucrose was extracted with water (reflux) then separated and quantified using HPLC and pulsed amperometric detection (LC20A binary pump and Decade II detector, Shimadzu, Kyoto, Japan). Total chlorogenic acids (CGA) were extracted with methanol-water (70% w/w) then purified through a polyamide

Reference paper as:
F. Davrieux, F. Descroix, D. Bastianelli, M. Hoarau, J. Roger, N. Durand and V. Bellon-Maurel (2012).Use of near infrared spectroscopy in the development, qualification and certification of coffee Bourbon pointu, in: Proceedings of the 15th International Conference on Near Infrared Spectroscopy, Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 386-391.

column, eluted using alkaline methanol and quantified using a spectrophotometer at 324 nm (.DU640 B, Beckman Coulter, Villepinte, France) Fat content was determined by gravimetry using a Soxtec (FOSS, Port Matilda, USA) extractor and petroleum ether. Moisture content (expressed as a percentage of dry matter, DM) was quantified by gravimetry after drying at 103°C for 16 hours using a Chopin oven (Gefran 800, Chopin, Boulogne, France).

## Results and Discussion

These spectra and wet chemistry values were added to the CIRAD Arabica database and specific Arabica+BP calibrations were developed. The descriptive statistics for the calibration set and equations performances parameters are reported in table 1. The performances of the equations, in terms of $R^2$, SEC and SECV for caffeine, dry matter and fat, were highly satisfactory and the equations could be used for routine analyses. The predictive models developed for CGA and trigonelline contents were less efficient and those equations could be used for screening of samples based on these constituent predictions. The adjusted model for sucrose content was not satisfactory, probably due to a poor repeatability of the wet chemistry method and the evolution of samples during storage between NIR and chemical analyses. This model can however be used to identify extreme values.

These calibrations were applied to the whole BP database, and then Grubb's tests were performed on predicted caffeine and moisture contents. Subsequently, 10 samples with extreme caffeine values and 11 samples with extreme moisture values were identified.

**Table 1.** Calibration set descriptive statistics and PLSr equations performance parameters.

| | | DM | Caffeine | Trigonelline | Fat | CGA | Sucrose |
|---|---|---|---|---|---|---|---|
| Calibration set (Arabica and BP coffees) | Number of samples | 1779 | 1228 | 876 | 854 | 831 | 848 |
| | Minimum | 85.10 | 0.45 | 0.44 | 10.37 | 5.10 | 5.26 |
| | Maximum | 93.99 | 1.79 | 1.45 | 17.70 | 11.04 | 11.48 |
| | Mean | 89.37 | 0.94 | 0.94 | 15.11 | 7.44 | 7.73 |
| | SD | 1.30 | 0.32 | 0.13 | 1.28 | 0.88 | 0.96 |
| | SEL* | 0,10 | 0,06 | 0,03 | 0,35 | 0,30 | 0,40 |
| Equations parameters | Number of samples** | 1568 | 1181 | 830 | 673 | 793 | 815 |
| | Mean | 89.38 | 0.93 | 0.93 | 15.15 | 7.45 | 7.67 |
| | SD | 1.30 | 0.31 | 0.12 | 1.34 | 0.81 | 0.88 |
| | SEC | 0.17 | 0.06 | 0.06 | 0.39 | 0.41 | 0.54 |
| | $R^2$ | 0.98 | 0.97 | 0.77 | 0.92 | 0.75 | 0.62 |
| | SECV | 0.18 | 0.06 | 0.07 | 0.43 | 0.46 | 0.59 |
| | RPD=SD/SECV | 7.06 | 5.03 | 1.87 | 3.14 | 1.78 | 1.49 |

All values were as % of dry matter
* SEL: standard error of laboratory
** Number of samples retained after a two passes *t* test.

Among the 10 extreme caffeine values, 8 were over 0.85%, one equal to 0.85% and one under 0.46%. The 8 samples with caffeine content over 0.85% were discarded. Those samples were tagged as "off-type" and corresponded to "Bourbon rond" and Arabica coffees grown in on Reunion Island. Of the 11 samples with extreme moisture values, 8 were over 13.5% and 3 under 8%, these were also removed from the database.

A final PCA was performed on the remaining 1254 spectra; the first 3 PCs explained 84.94% of total variance (PC1: 54.82%, PC2: 21.55% and PC3: 8.57%). The H distances were calculated using 32 PCs, the maximum H value was equal to 3.0. The scatter plot of the sample scores for the first 2 PCs (Figure 2) showed a uniform repartition of the samples without any specific year effect; 2005 and 2006 samples had the widest distribution. These results reflected the fact that the first two years of the project had several trials to identify the optimal cultural practices and post-harvest treatments. Samples harvested after 2008 corresponded to well regulated plantations and more standardised post-harvest conditions.

Reference paper as:
F. Davrieux, F. Descroix, D. Bastianelli, M. Hoarau, J. Roger, N. Durand and V. Bellon-Maurel (2012).Use of near infrared spectroscopy in the development, qualification and certification of coffee Bourbon pointu, in: Proceedings of the 15th International Conference on Near Infrared Spectroscopy, Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 386-391.

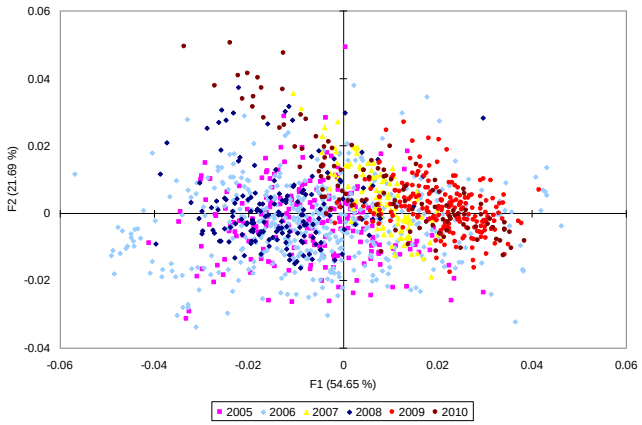


**Figure 2.** PC1 vs. PC2 scores plot of the PCA applied to the final 1254 BP coffee samples.

The average DM value was stable over the year (88.92%), and the minimum value was 86.53%. The caffeine content ranged between 0.41% and 0.83% with an average value of 0.63% constant over the year. The fat content ranged from 11.18% to 17%. Chlorogenic acids ranged from 5.9% to 9.6%. The trigonelline content was constant over the year and ranged from 0.53% to 1.08%, with an average of 0.83%. The sucrose ranged from 6.55% and 9.64% with an average of 7.97%, sucrose content increased over years, except in 2008, with an average maximum value of 8.41% observed in 2010.

The calibration strategy adopted by mixing Arabica and BP coffees was a good choice given the narrow variation of biochemical criteria within BP coffees; this was illustrated by caffeine content calibration. The scatter plot (Figure 3) for caffeine wet chemistry and NIR-predicted values highlights the quality of the fit and shows the potential of the model to discriminate low caffeine BP coffees from conventional Arabica coffees. This figure highlights the gap in caffeine values between Arabica and BP coffees; despite this, the residuals were centred on zero over the whole range (Figure 4).

The caffeine content predictions of the 48 blends between BP and Arabica were efficient even for samples with caffeine content between 0.85% and 1.00%, the Standard Error of Prediction (SEP) was equal to 0.039% with a $R^2$ equal to 0.94 (Figure 3).

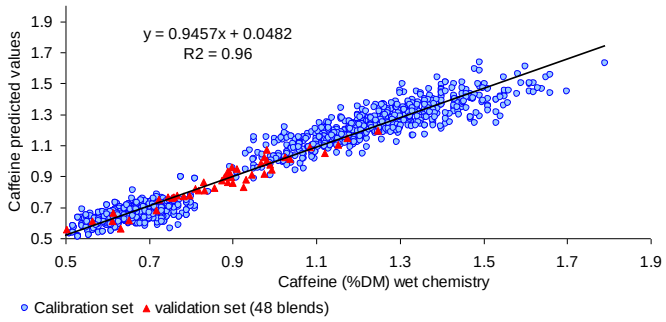


**Figure 3.** Scatter plot between wet chemistry and NIR predicted values for calibration and validation set.

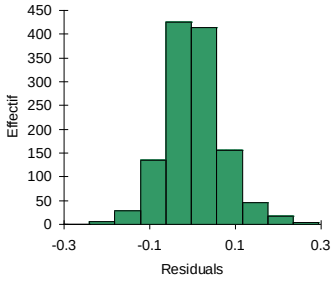


**Figure 4.** Histogram of residuals for calibration set.

The comparison of the 48 blends and the 38 Arabica coffees to the BP coffees database was done by calculation of their Mahalanobis distances (H) to the average spectra after projection on the PCs from the PCA done on BP spectral data. H values increased with Arabica proportion, with all blends above 60% Arabica having H>3. Below 20% of Arabica in the blend, the distance H was lower than 3 (Figure 5) and only 2 of the 30% Arabica-BP blends presented H values higher than 3 with a maximum of 3.5. Half of the blends ranged between 50% and 60% Arabica presented H values higher than 3, with a maximum of 4.4. The 6 BP coffees from New Caledonia were close to the BP coffees from Reunion Island. The reference BP (Reunion, 2007) sample had an H value equal to 1.2.

Reference paper as:
F. Davrieux, F. Descroix, D. Bastianelli, M. Hoarau, J. Roger, N. Durand and V. Bellon-Maurel (2012).Use of near infrared spectroscopy in the development, qualification and certification of coffee Bourbon pointu, in: Proceedings of the 15th International Conference on Near Infrared Spectroscopy, Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 386-391.

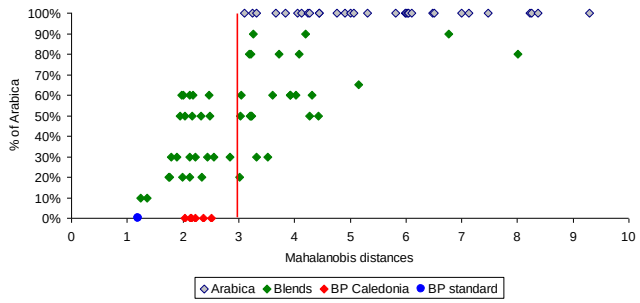


**Figure 5.** Mahalanobis distances to the centroid of BP spectral database as a function of Arabica content.

## Conclusion

The performance of the predictive models based on a broad database of Arabica coffees, and Bourbon Pointu coffees from the Reunion Island, enabled fine characterisation 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 commercial green BP coffees (mature, fair fermented, dried).

This study demonstrated the use of spectral characterisation at early stages in the development of a new product. After 6 years, we have obtained a robust database and efficient calibrations for moisture and caffeine content for commercial green coffees. Moisture content is a part of the commercial quality chart and should be lower than 12%. Caffeine content is fully describe in the specifications for a protected designation of origin (PDO), caffeine should be lower than 0.85%, our model will help to certify BP coffees.

Further studies will try to strengthen authentication by combining NIR fingerprints, biochemical profiles and sensory profiles. Future work will focus on the use of NIR spectroscopy to improve BP coffee quality by identifying indicators of cherry ripeness.

## Acknowledgements

We thank all the staff of Coopérative Bourbon Pointu (Saint Pierre, Reunion), and all the farmers involved in the project. This study was funded by the Région Reunion, the French government and Europe.

## References

1. F. Davrieux, J.C. Manez, N. Durand and B. Guyot, "Determination of the content of six major biochemical compounds of green coffee using near infrared spectroscopy", in *Proceedings of the 11th International Conference on Near Infrared Spectroscopy*, Ed by A.M.C. Davies and A. Garrido-Varo, NIR Publications, Norfolk, UK, pp.441-444, (2004).
2. F. Grubbs, Technometrics, **11**, 1, 1-21 (1969).

Reference paper as:
F. Davrieux, F. Descroix, D. Bastianelli, M. Hoarau, J. Roger, N. Durand and V. Bellon-Maurel (2012).Use of near infrared spectroscopy in the development, qualification and certification of coffee Bourbon pointu, in: Proceedings of the 15th International Conference on Near Infrared Spectroscopy, Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 386-391.