Evaluation of green coffee beans quality using near infrared spectroscopy: A quantitative approach

João Rodrigo Santos, Mafalda C. Sarraguça, António O.S.S. Rangel, João A. Lopes

Abstract

Characterisation of coffee quality based on bean quality assessment is associated with the relative amount of defective beans among non-defective beans. It is therefore important to develop a methodology capable of identifying the presence of defective beans that enables a fast assessment of coffee grade and that can become an analytical tool to standardise coffee quality. In this work, a methodology for quality assessment of green coffee based on near infrared spectroscopy (NIRS) is proposed. NIRS is a green chemistry, low cost, fast response technique without the need of sample processing. The applicability of NIRS was evaluated for Arabica and Robusta varieties from different geographical locations. Partial least squares regression was used to relate the NIR spectrum to the mass fraction of defective and non-defective beans. Relative errors around 5% show that NIRS can be a valuable analytical tool to be used by coffee roasters, enabling a simple and quantitative evaluation of green coffee quality in a fast way.

1. Introduction

Near infrared spectroscopy (NIRS) has been successfully used in the coffee industry, focusing on water content determination (Reh, Gerber, Prodolliet, & Vuataz, 2006), discrimination between Arabica and Robusta varieties (Downey, Boussion, & Beauchêne, 1994), roasting control (Alessandrinì, Romani, Pinnavaìa, & Rosa, 2008), and organoleptic characterisation (Pizarro, Esteban-Diez, & Gonzalez-Saiz, 2004) of roasted coffee.

The accurate assessment of green coffee quality is of major importance to coffee roasters and distributors. For each coffee bean shipment, beyond the initial identification of the coffee variety, further characterisation is required to verify if the purchased coffee quality matches the required coffee quality. Green coffee quality depends on the relative amount of defective beans and/or foreign bodies among non-defective beans. In this context, defects are described as foreign materials of non-coffee (e.g., stones/sticks) and non-bean origin (e.g., husks/hulls), abnormal beans regarding shape and visual appearance, such as black beans or any defects that impair brewed coffee taste and flavour (Leroy et al., 2006). Coffee bean defects can be divided in primary and secondary defects. Black beans and sour beans are usually identified as primary coffee bean defects. Regarding secondary defects, one can find a plethora of examples, the most common being partially black, partially sour, broken, insect-damaged, faded, green, mouldy and silver skinned. Bee et al. (2005) and Agresti, Franca, Oliveira, and Augusti (2008) stressed several bean defects, their origins and implications upon coffee cup quality.

Based on the aforementioned parameters, classification of coffee is not uniform and follows distinct guidelines. The International Coffee Organization (ICO) recommends that a maximum of 86 defects/300 g for Arabica variety and 150 defects/300 g for Robusta variety should not be exceeded for exportation purposes (ICO, 2011). The Speciality Coffee Association of America (SCAA) method for coffee grading qualifies five different coffee grades, according to the number of defects (per 350 g of green coffee sample), moisture content and 83 cup assessment (SCAA, 2011). The Brazilian classification for green coffee beans encompasses a more extensive number of coffee grades, using shorter intervals regarding the number of defects found in 300 g of green coffee samples (ABIC, 2011).

Compared to non-defective beans, defective beans roast differently and will also acquire a different colour after roasting (Franca, Mendonça, & Oliveira, 2005a). More importantly, roasting of defective beans impart to coffee, organoleptic characteristics that decrease cup quality (Franca, Oliveira, Mendonça, & Silva, 2005b; Salva, Ribeiro, & Ferreira, 2011). For the abovementioned reasons, it is important to develop strategies for fast assessment and standardisation of coffee quality.

The strategies referred to in the literature for defective green beans discrimination focus on the identification of chemical non-volatile and volatile markers, as well as evaluation of physical characteristics of the beans. Regarding non-volatile markers, chemical differences between non-defective and defective green...
beans have been shown regarding pH, water sucrone and amine compounds levels before roasting (Franca, Vasconcelos, Gloria, & Mendonca, 2007). Regarding the assessment of green and roasted coffee beans using volatile compounds as potential defective coffee bean markers, Toci et al. (Farah & Toci, 2008) proposed several chemical markers for immature, black and sour beans in both green and roasted coffee using SPME-GC–MS. Mendonça, Franca, and Oliveira (2009) evaluated the parameters volume, density and colour before and after roasting in order to identify defective from non-defective coffee beans in Arabica and Robusta varieties. For Arabica variety, discrimination could be attained through sieving either in green or roasted beans. For Robusta variety sieving allowed the discrimination between defective from non-defective green beans while colour measurement allowed discrimination within roasted beans. Still, none of the abovementioned strategies provide a quantitative evaluation for quantifying the presence of defective beans in raw coffee. To the best of our knowledge, a methodology focused on the characterisation and quantification of defective beans has not yet been reported. Craig, Franca, and Oliveira (2012) evaluated the potential of Fourier transform infrared (FTIR) spectroscopy to discriminate defective and non-defective beans. The objective was to discriminate black, immature, sour and non-defective beans from an Arabica variety using PCA and cluster analysis. However, the cited work presents a qualitative approach applied to one coffee variety. In this work, a methodology for quality assessment of green coffee beans based on NIRS is proposed. The applicability of NIRS was evaluated for Arabica and Robusta varieties from different geographical regions in a qualitative and quantitative approach.

2. Materials and methods

2.1. Coffee sampling

Three hundred grams of four green coffee beans of different provenances, were sampled from a 60-kg coffee bag, and visually inspected with respect to coffee bean defects. Two Robusta (Indonesia = RI and Vietnam = RV) and two Arabica (Colombia = AC and Nicaragua = AN) provenances were used. For each variety, besides the non-defective beans, six different types of defective beans were further identified (see Table 1). For Robusta variety, it was possible to visually identify the following defects: insect-damaged, broken, mouldy, black, partially black and sour. For Arabica variety it was possible to visually identify: insect-damaged, broken, silver skin, faded, immature and sour. The beans were carefully selected by visual inspection assure the presence of beans of similar size. In this way, the volume of each bean could be considered similar and the mass fraction of beans present in the sample is only due to the number of beans present.

2.2. Spectra acquisition

Near infrared spectra were recorded on a Fourier-transform near infrared spectrometer (FTLA 2000, ABB, Québec, Canada) equipped with an indium–gallium–arsenide (InGaAs) detector and a spinning accessory coupled to a powder sampling accessory (ACC101, ABB), enabling diffuse reflectance measurements over the sampling window with 6 mm diameter illumination area. Each spectrum was acquired in diffuse reflectance mode with a 2 cm⁻¹ resolution over a wavenumber interval between 10000 cm⁻¹ and 4000 cm⁻¹, and recorded as the average of 64 scans. The equipment was controlled via Bomen-Grams software (version 7, ABB).

Green coffee beans were poured into borosilicate flasks and then placed in the spinning accessory for a continuous rotation of the sample during spectral acquisition. This procedure guarantees a more precise measurement accounting for the intrinsic non-homogeneity of green coffee beans samples. At the beginning of the measurements and every hour, a background spectrum was taken by placing a 100% reflectance reference material PTFE (Teflon) over the sampling window.

2.3. Multivariate modelling

Principal component analysis (PCA) (Jolliffe, 2002) was used to verify the possibility of NIRS to distinguish the defective and non-defective beans in the coffee provenances tested. Coffee bean samples were produced according to experimental designs (o-optimal). The mass fraction of each coffee bean quality was varied in the experimental design, in order to maximise the information in the selected set of experimental runs (Eriksson, Johansson, Kettaneh-Wold, Wikstrom, & Wold, 2000). For the quantification of coffee bean qualities a full factorial matrix was created considering the seven qualities and six levels, yielding a total 279936 possible experiments (matrix D). Since the number of experiments to use was fixed at 39, randomly selected subsets of 39 experiments out of the 279936 were tested, considering the G = det(Di/Di) criterion (where Di is the experimental design matrix restricted to the 39 selected experiments). This procedure was repeated 10,000 times. The subset yielding the maximum value of G was selected as the experimental design. The same strategy was proceeded to produce the other experimental designs (changing the number of experiments and the number of factors).

The coffee bean qualities mass fraction values were modelled from the NIR spectra with partial least squares (PLS) regression. The optimal number of latent variables for each model was estimated with the leave-one-out cross-validation method (Tormod, Tomas, Fearn, & Tony, 2002). Other type of cross-validation methods were also tested, namely, Venetian blinds with 10 data splits, continuous blocks with 10 data splits and random subsets with 10 data splits and five interactions. It was found that for all the cross-validation methods tested the optimum value of latent variables (LV) was the same.

The PLS regression model is commonly used in chemometrics with the objective of establishing regression models (for physical or chemical properties) from spectral data (Geladi & Kowalski, 1986; Miller, 2000). To assess the PLS model error, the root mean square error of cross-validation (RMSECV) was estimated according to Eq. (1).

\[
\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{N}}
\]
In Eq. (1), \( \hat{y}_{CV} \) is the model estimate for sample \( i \) based on a model calibrated without sample \( i \) deleted and \( N \) is the number of samples.

The model error for unknown samples was estimated in terms of the root mean square error of prediction (RMSEP).

\[
RMSEP = \sqrt{\frac{1}{N_p} \sum_{i=1}^{N_p} (\hat{y}_i - y_i)^2}
\]  

In Eq. (2), \( \hat{y}_i \) is the estimate for \( y_i \) for the \( i \)th test sample and \( N_p \) is the number of samples in the test set.

The relative error is calculated by dividing the RMSECV or RMSEP by the average value of the parameter.

To reduce this spectral non-homogeneity, the following strategy was applied. Six instrumental replicas were made for each sample and the best replica for each sample was chosen, applying a genetic algorithm to the cross-validation procedure. The genetic algorithms simulate the evolution where random variations in the genetic make-up of a population combined with selection of the fittest individuals led to improvements in the breed (Leardi, 2001; Tomod et al., 2002). In this case the method adopted was based on a population of 30 individuals. Each individual is a chromosome with 39 genes (for the quantification of beans quality) or 20 genes (for the quantification of non-defective genes) with six positions (replicates) in each gene. The chromosomes have 0s when the sample is not included and 1s when the sample is included. Within each gene only one position can have the value 1 (replicate included). The mutation rate probability for each gene was set to 5%. Crossover was not used. The mutation rate decreased 0.01% after each generation. Each chromosome was evaluated using 100 different models, each considering a different calibration/test set. Only the test set results were considered and the value assigned to each chromosome evaluation was the mean RMSEP. This way prevents in some degree the level of overfitting. It has the disadvantage of making the convergence to the solution slower, since for the same chromosome different median RMSEP can be obtained (due to different train/test sets). One hundred times were chosen because of computational time limitations. A reasonable convergence was obtained after approximately 300 iterations. This procedure was only adopted to define the best calibration set. For prediction this procedure cannot be done. However, the Hotelling and SPE statistic can be checked to verify if every new sample spectrum fits the model or not. If not, a measurement repetition can be made.

In multivariate regression, the estimation of uncertainty is not straightforward. To overcome this difficulty, a statistical technique called bootstrapping was adopted. This procedure generates a set of samples by sampling with replacement from the original data set (Wehrens & Vander Linden, 1997). By repeating this procedure, a large number of “new” samples sets are generated. Here, the size of the original data set was kept when generating the new samples. Modelling these sets, yields an ensemble of estimates that can be used to obtain statistical parameters such as the standard deviation and relative standard deviation (RSD) of model errors (Wehrens, Putter, & Buydens, 2000).

All calculations were carried out using Matlab Version 7.9 (MathWorks, Natick, MA) and the PLS Toolbox version 5.5.1 (Eigenvector Research, Inc. Seattle, WA).

2.4. NIR spectral ranges and pre-processing methods

Four different spectral ranges were tested throughout the work: (i) 9980.0–4184.0 cm\(^{-1}\) (full spectra); (ii) 9980.0–7229.6, 7085.0–5590.3 and 5108.1–4184.0 cm\(^{-1}\) (removal of water bands); (iii) 8338.6–4184.0 cm\(^{-1}\) (removal of physical information); (iv) 8338.6–7229.6 cm\(^{-1}\), 7085.0–5590.3 and 5108.1–4184.0 cm\(^{-1}\) (a combination of the second and third spectral range). Several spectra pre-processing methods were also tested; namely, standard normal variates (SNV), multiple scatter correction (MSC) and Savitzky–Gooley filter (15 points) fitted with a second-order polynomial in combination with first or second derivatives.

3. Results and discussion

3.1. Qualification of coffee bean qualities

For each coffee provenience, a manual count of the beans’ qualities revealed that the Arabica variety has a percentage of non-defective beans around 95%, insect damage (ID) being the defect present in the largest percentage, 1.2% and 2.0% in AC and AN proveniences, respectively (Table 1). The percentage of non-defective beans in the Robusta variety ranged between 91% for RV and 70% for RI. The main difference between these two proveniences is the large amount (16.2%) of mouldy (M) defect found in RI compared with the 1.9% found in RV (Table 1).

NIRS spectra for the seven bean qualities for each coffee provenience are very similar, although some differences can be seen in the region between 6000 and 4000 cm\(^{-1}\), corresponding to the spectral region with more chemical information (see Figs. 1 and 2). This region of the NIRS spectrum correspond to the combination and first overtone regions that comprise the CH + CC, CH + CH, NH + OH, NH and OH combination bands and C=O second overtone bands and SH and CH first overtone bands.

To verify the possibility of beans quality distinction within each coffee provenience, the NIRS spectra of each one of the seven bean qualities were analysed by PCA.

The first observation that can be done is that the instrumental replicates of each sample appear scattered in the score plots (Fig. 3). This scatter is an indication of the irreproducibility inherent to the analytical method sampling, though this fact was somewhat expected due to several factors. For example, the defects may be present only in some parts of the beans (e.g., partially black beans). This circumstance creates irreproducibility in the spectral acquisition. Another factor is that using the spinning accessory for spectral acquisition does not assure that all the surface area of the beans will face the acquisition window. These two factors contribute to the irreproducibility that can be seen in the PCA score plots among the instrumental replicates.

Still, for the Robusta variety, a separation between the black and the mouldy beans and the others can be seen. However, only these two defects are separated in Robusta variety. In the Arabica variety, the separation is not very clear. For the Arabica AC, it seems that in the first principal component (PC 1) there is a separation between the sour, faded, insect-damaged and broken defects and the others. In the Arabica AN, this separation cannot be seen although there is a clear separation between the sour and faded from the others on PC 1.

Analysing the PCA loadings for the first component (Fig. 4) it can be seen that the spectral range that contributes to the separation in PC 1 is between 4600 and 4000 cm\(^{-1}\). The spectral region between 5000 and 4000 cm\(^{-1}\) is mainly due to C–H and C–C combination bands. However, since coffee is a complex mixture of organic molecules it is difficult to identify the molecules responsible for the defects separation.

The PCA modelling is not conclusive since some of the defects can be easily differentiated but others could not be separated.

3.2. Quantification of bean qualities

For the initial approach focussed on the quantification of beans qualities for each coffee provenience, a D-optimal experimental design was made for each variety (see Supplementary material
Fig. 1. NIRS spectra for each bean quality of the Robusta variety: B–black; BR–broken; ND–non-defective; ID–insect damage; M–moldy; PB–partially black; S–sour.

Fig. 2. NIRS spectra for each bean quality of the Arabica variety: BR–broken; F–faded; ND–non-defective; I–immature; ID–insect damage; SS–silver skin; S–sour.
The number of calibration samples is always an issue when dealing with NIRS calibrations, and several studies have been realised considering different matrices (Blanco, Coello, Iturriaga, Maspoch, & Pou, 2001; Sarraguça & Lopes, 2009). In the particular case of quantification of coffee beans quality no indication is given in the literature. The number of calibration samples should be as large as to produce enough concentration variability aiming at model robustness. Nevertheless, a minimum number of samples per significant component is always recommended. For each coffee provenance thirty-nine samples were prepared by mixing the respective amount of each quality of coffee bean indicated by the experimental design. For this particular study, a higher mass fraction range of the different defective beans was selected than that normally present in real samples, in order to better evaluate the ability of the NIRS to identify each bean defect. Thereafter, each sample was analysed six times using NIRS.

The different spectral ranges and spectra pre-processing methods already mentioned were tested and analysed through PLS regression with cross-validation (see Supplementary material Table S2 and S3). The number of latent variables (LV) was found by minimising the RMSECV. The best pre-processing method was SNV followed by mean centre for all cases.

As already discussed in the previous section, the spectra from instrumental replicates of the same sample show a slight variation. To reduce this spectral non-homogeneity, the following strategy was applied: six instrumental replicas were made for each sample and the best replica for each sample was chosen, applying a genetic algorithm to the cross-validation method (see Section 2.3 for more details on the genetic algorithm).

Comparing the PLS regression cross-validation results obtained through the selection of the best replicate (Supplementary material Table S2 and S3) with the results obtained by averaging all replicates (not shown) revealing that the selection of the best replicates always yielded better results. For the Robusta variety, the decrease in the relative error is in some cases in the order of 30% (e.g., mouldy beans quantification for the Robusta RI). For the Arabica the decrease in the relative error reaches 25% (e.g., silver skin quantification for Arabica AN).

For the four coffee provenances, the lowest error was found for the quantification of the non-defective quality, with errors around 30%. This result is an indication that although the defects could not be quantified individually it could be possible to quantify the non-defective quality against the rest of the remaining qualities.

### 3.3. Quantification of defective and non-defective beans

Because of the results described in the last section a new approach was attempted. The beans were separated into two groups: non-defective and defective beans. The non-defective group comprised the non-defective and broken beans, and the defective group comprised the rest of the defects. The broken beans were added to the non-defective beans since the only difference between them is that some small part of the beans is missing. This is a physical and not a chemical difference.

For each coffee provenance, two new experimental designs were made, one for calibration (20 samples) and one for testing (8 samples) (Supplementary material Table S4). The mass fraction range was set as 10% around the nominal mass fraction (see Table 1...
and Supplementary material Table S4). This range encompassed the usual ratio of non-defective vs defective beans of the coffee provenances used throughout this work, and is in conformity with the general range described by international standards of coffee grading.

It was ensured that the test samples comprised all the variability and had a smaller mass fraction range than the calibration samples. Prior to the preparation of each sample, a batch of non-defective beans and a batch of defective beans were prepared, both intending to encompass the usual quantities of each quality for the corresponding coffee provenance. Thereafter, for the preparation of each sample, a defined number of non-defective and defective beans were sampled and mixed before data acquisition. For the preparation of the following sample, the previous one was firstly separated, placing the respective non-defective and defective beans in the respective batch. The samples were analysed using NIRS, with each sample being measured six times. The spectral range and spectra pre-processing tested were the same as in the previous sections. As in the previous section the best pre-processing method was SNV followed by mean centre.

As before, to reduce the effects of the method sampling irreproducibility, six replicates were made for each sample and the method described in the last section to determine the best replicates for modelling was also applied in this case.

The calibration results (Table 2) obtained were considerably better (with respect to the estimation of individual defects), with relative errors around 5% for the Robusta and 4% for the Arabica variety. The coefficients of determination also improved with values between 0.60 and 0.62 for the Robusta variety and 0.74 and 0.77 for the Arabica variety. The calibration precision, given by the mean relative standard deviation (RSD) is very similar for all coffee provenances with this parameter ranging between 2.2% and 2.3%.

The calibration models were then tested using the samples in the test sets (Table 2). The test relative errors range between 3% and 4% for the Robusta variety and around 5% for the Arabica variety. The coefficients of determination were considerably high for all provenances, with the exception of the Arabica AC with a value lower than 0.5. The relative standard deviation for the validation models were in all cases below 2% (1.6% for Robusta variety)

Table 2
PLS cross-validation and test results for the prediction of non-defective beans mass fraction of each coffee provenance.

<table>
<thead>
<tr>
<th>Provenance</th>
<th>Spectral range</th>
<th>LV</th>
<th>RMSECV</th>
<th>Relative Error</th>
<th>$R^2_{CV}$</th>
<th>RSD (%) (mean)</th>
<th>RMSEP</th>
<th>Relative Error</th>
<th>$R^2_{test}$</th>
<th>RSD (%) (mean)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robusta RI</td>
<td>(ii)</td>
<td>4</td>
<td>0.045</td>
<td>5.00</td>
<td>0.602</td>
<td>2.24</td>
<td>0.038</td>
<td>4.27</td>
<td>0.709</td>
<td>1.58</td>
</tr>
<tr>
<td>Robusta RV</td>
<td>(ii)</td>
<td>6</td>
<td>0.044</td>
<td>4.87</td>
<td>0.625</td>
<td>2.31</td>
<td>0.027</td>
<td>3.08</td>
<td>0.865</td>
<td>1.60</td>
</tr>
<tr>
<td>Arabica AC</td>
<td>(iv)</td>
<td>5</td>
<td>0.035</td>
<td>3.89</td>
<td>0.771</td>
<td>2.15</td>
<td>0.049</td>
<td>5.55</td>
<td>0.472</td>
<td>1.32</td>
</tr>
<tr>
<td>Arabica AN</td>
<td>(iv)</td>
<td>4</td>
<td>0.037</td>
<td>4.16</td>
<td>0.737</td>
<td>2.31</td>
<td>0.045</td>
<td>5.07</td>
<td>0.666</td>
<td>1.20</td>
</tr>
</tbody>
</table>

* (i) 9980.9–4184.0 cm$^{-1}$; (ii) 9980.9–7229.6, 7085.0–5590.3 and 5108.1–4184.0 cm$^{-1}$; (iii) 8338.6–4184.0 cm$^{-1}$; (iv) 8338.6–7229.6, 7085.0–5590.3 and 5108.1–4184.0 cm$^{-1}$. 

**Fig. 4.** PCA first component loadings of the coffee beans quality for each coffee provenance. Spectral range: 9980.9–7229.6, 7085.0–5590.3 and 5108.1–4184.0 cm$^{-1}$ (removal of water bands). Spectral pre-processing: Savitzky–Golay 15 points filter fitted with a second-order polynomial and a first derivative followed by mean-centre.
Fig. 5. PLS X scores ($t$) in function of PLS y scores ($u$) for the four latent variables for the Robusta RI model for the quantification of non-defective beans.

Fig. 6. PLS regression coefficient for the quantification of non-defective beans.
and between 1.2% and 1.3% for Arabica AN and Arabica AC, respectively).

In Fig. 5 is shown a plot of the PLS spectra scores (t) against the mass fraction scores (u) for Robusta RI. It can be concluded that for all latent variables there is a linear relation between the scores, which indicates that the PLS model is modelling the non-defective beans mass fraction. For the rest of the provenances the results were similar.

By analysing the regression coefficient of the four PLS models (Fig. 6) it can be seen that the region between 5000 and 4000 cm⁻¹ is the most important. This spectral region is mainly due to C–H and C–C combination bands and is the one that contains most information regarding the chemical composition of the defective and non-defective beans, although, as already stated, coffee is a combination of numerous organic molecules and therefore it is difficult to indicate which are the molecules responsible for the information contained in the spectral region referred to previously.

The methodology, based on the assumption of considering broken beans as non-defective beans, and using a genetic algorithm to select replicates, presented relative errors <5% for both coffee varieties. This strategy is to the best of our knowledge, the only published analytical methodology able to quantify defective beans within a coffee batch.

4. Conclusion

In this work the use of NIRS with chemometric modelling was exploited for coffee bean quality assessment. It was shown that the approach followed can be a valuable tool to be used by coffee roasters, able to qualify green coffee quality with low relative errors concerning the quantification of non-defective against defective beans following general international guidelines of coffee classification. Additionally, the strategy herein presented is, low cost, fast, and requires no sample processing, features that make this strategy suitable to be implemented on a routine basis. The possibility of using hand-held NIR spectrometers to implement this strategy in field analysis can additionally be foreseen. These features diverge with the traditional method for coffee qualification, which is based on eye inspection, is time consuming, and most of all, dependent on operator judgment.

Acknowledgements

The authors are grateful to Tenco Cafés, Lda, Portugal, for the helpful discussion and collaboration. J.R. Santos and M.C. Sarraçuca acknowledge the grants SFRH/BDP/63492/2009 and SFRH/BDP/74788/2010, respectively funded by QREN, POPH, FSE and MCTES. This work has also been supported by Fundação para a Ciência e a Tecnologia through projects nos. PEst-C/EBQB/LA0006/2011 and PEst-OE/EBQB/LA0016/2011.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.foodchem.2012.06.059.

References