Classification of Arabica Java Coffee Beans Based on Their Origin using NIR Spectroscopy

F Kurniawan¹, I W Budiastra², Sutrisno², S Widyotomo³

¹Indonesian Center for Agricultural Post Harvest Research and Development, Ministry of Agriculture
Jl. Tentara Pelajar No. 12 Bogor 16111, West Java Indonesia
²Department of Mechanical and Biosystem Engineering, Faculty of Agricultural Engineering and Technology,
Bogor Agricultural University, Kampus IPB Darmaga Bogor 16680, West Java Indonesia
³Indonesian Coffee and Cocoa Research Institute, Jember, East Java Indonesia

*Email: fajar.koer@gmail.com

Abstract. The NIR spectroscopy for characterization and classification of intact Java arabica coffee beans based on their origin is proposed. Commonly, NIR spectroscopy is used to determine composition of agricultural products including coffee. Three kinds of Java arabica coffee beans namely Arabica Java Preanger, Arabica Bondowoso and Arabica Malang were used in this research. Three hundred samples, each consisted of 100 g coffee beans were prepared. The coffee beans of 100 g were placed in petri dish and the light reflectances of intact coffee beans were measured by FT-NIR spectrometer at the wavelength of 1000-2500 nm. After reflectance measurement, the samples were subjected to composition analysis using proximate and Liquid Chromatography Mass Spectrometry. The reflectance and absorbance spectra were processed by five spectra data pretreatments (smoothing, first derivative, second derivative, Standard Normal Variate (SNV), Multiple Scatter Correction (MSC)) and then Principle Component Analysis (PCA) were carried out. Discriminant analysis (DA) of Principle Components were developed to classify coffee beans based on their origin. The results show that the spectra data treatments of SNV and MSC of reflectance spectra and PC analysis using PC1 and PC2 gave the best results for discriminating three kinds of coffee beans. The DA of three principle components of reflectance data processed by SNV and MSC could classify arabica coffee beans accurately (100%). This results shows that NIR spectroscopy can be used as a nondestructive method to classify arabica coffee beans based on their origin.

1. Introduction

Coffee is number two of the most important commodity in international trade based on the volume of trade. Coffee trade involves a network of international trade from developing countries to developed countries which are the main consumers. The price of coffee beans depends on the quality that has a direct correlation to the taste of the final product. Beside quality, the originality and traceability of coffee
origin become important issues in global trade. Therefore, the determination of quality and originality of coffee is necessary. Commonly, quality of coffee is determined by chemical method. This method is time consuming and expensive so do not comply with coffee industry need. In recent years, the use of NIR spectroscopy (NIRS) to determine the characteristics and the quality of agricultural products including coffee has been studied.

NIRS was able to estimate catechin content in gambir flour from dried gambir leaves using the PLS model [1][2], and also was able to estimate the oil and ALB content in jathropa fruits [3]. Also, NIRS has been successfully used for predicting the concentration of caffeine, theobromine and theophylline in ground green coffee bean [4], predicting caffeine concentration in ground roasted coffee [5] and predicting caffeine in intact coffee bean [6].

Beside quality, NIR has been studied for discriminating coffee varieties. The NIRS method can discriminate between Arabica and Robusta varieties in both pure and mixed conditions, which combined with multivariate classification [7], NIRS is also able to discriminate the fermentation levels of the cocoa powder and bean [8]. However, NIR assessment for determining the originality of intact coffee beans has not been carried out yet.

The purpose of this study was to assess NIR Spectroscopy to characterize and classify of arabica Java coffee beans based on their origin using Principal Component Analysis (PCA) and Discriminant Analysis (DA). This research will be useful for coffee industries since it will provide a new rapid and nondestructive method for determination the originality and traceability of coffee beans.

2. Material and Method

2.1 Material and Apparatus

In this study, three selected arabica Java coffee were used those are Arabica Java Preanger, Arabica Bondowoso and Arabica Malang. The coffee beans used were in green bean form and obtained from Indonesian Coffee and Cocoa Research Institute, Jember. Coffee beans samples from each origin were weighed and packed into 100 g. The number of samples prepared for this study was 300 samples (100 samples per origin).

The instrument used for reflectance measurement were a FT-NIR Spectrometer (NIRFlex N-500, manufactured by BÜCHI Labortechnik AG, Switzerland), a petri dish, an oven, a digital scale (0-200 gram, accuracy 0.001 gram), a grinder, an oven, a spectrophotometer and a LC-MS for determining the caffeine, chlorogenic acid and trigonelline content of coffee bean.

2.2 Coffee Bean Spectra Acquisition

The acquisition of reflectance spectra was conducted to the intact coffee beans. Coffee beans of 100 g were placed in a petri dish by being stacked evenly and tightly. The reflectance of samples were measured by then scanning the samples 3 times at 3 different points by setting the petri dish to rotate 360 degrees during the sample scanning. The wavelength interval used was 1000-2500 nm with 2.0 nm intervals.

2.3 Chemical Analysis

The chemical analysis was assessed to get chemical content of coffee beans (green bean). The proximate analysis was estimated according to [9]. Caffeine, chlorogenic acid and trigonelline content was measured by a liquid chromatography–mass spectrometry (LC-MS) method. As much as 2 ppm of caffeine, trigonelline and CGA standard, respectively was injected into the LC-MS. Chromatographic separation was achieved on a bridged ethylene hybrid C18 column, maintained at a constant temperature of 40 °C. The LC two-phase mobile system consisted of 0.1% aqueous formic acid (eluent A) and acetonitrile (eluent B). The gradient was programmed with a flow rate of 0.3 mL/min, as follows. Subsequently, the peak of each concentration was obtained at any given time. After that, a 0.500 g portion of each coffee powder, weighed directly into a nylon centrifuge tube was extracted with 7.00 mL of 95% aqueous methanol by shaking for 30 min on a platform shaker at 200 rpm. The supernatant
was decanted carefully into a second centrifuge tube. The residue was resuspended in 7.00mL of the solvent. After combination of the supernatants, the volume was adjusted to 14.0 mL. A 0.25 mL portion of the extract was diluted five-fold by adding deionized water [10]. The diluted sample solution was treated with 15 mL each of Carrez reagent I and II to remove polymeric components [11]. After centrifuging the mixture for 5 min at 12 000 rpm, the supernatant was filtered directly into a chromatographic vial through a 0.22 mm syringe filter for analysis into LC-MS in the same condition.

2.4 Spectra Processing
The reflectance and absorbance spectra were processed with several spectra data pretreatments i.e Smoothing Savitzky-Golay [12], derivatives [13], SNV and MSC. The first and second derivatives applied to the standard spectrum are intended to sharpen their profile and eliminate the interference which is caused by baseline shifts and noise disturbances. The uneven surface effects of the sample, multiplication effects including scattering, particle size and multi-colinearity changes can cause a large variations in the reflectance spectrum [14]. Thus, it is important to apply wavelength or mathematical pretreatment options, such as standard normal variate transformation (SNV), multiplicative scatter correction (MSC), to minimize or avoid such effects [15]. Spectra data treatment are also required to obtain a better discrimination [16]. Spectrum data processing was carried out using Unscrambler software X 10.3 (trial version, CAMO).

Principal Component Analysis (PCA) was used to cluster three kinds of coffee beans based on principal components of processed spectra data. Display of scores plot and loading spectra of various peaks at the specific wavelengths of spectrum analysis results by PCA will be plotted and become a characteristic or finger print of coffee beans and become a basic in grouping the coffee bean cultivars [17].

Discriminant analysis of the result of qualitative analysis by PCA were constructed and applied for classification of three kinds of coffee beans. Discriminant analysis create discriminant functions based on a linear combination of predictor variables to produce the best discrimination between groups. The discriminant function is generated from the sample where the group identity of the subject has been known. The input data used in the discriminant analysis were the score of PC1, PC2 and PC3.

3. Result and Discussion

3.1 The Original Spectra of Arabica Java Coffee Beans
Figure 1 showed the absorbance spectra of Arabica Java coffee beans with several peak and valley in specific wavelength. The valley existed related to electromagnetic wave and response of the molecular bonds of O-H, C-H, C-O and C-C. The response subjected to vibrational energy, both of stretch vibration and bent vibration [16].

Characteristic of the absorbance spectra of samples from three Java arabica coffee have an almost similar shape. In general it can be said that the chemical characteristics of the three coffees are similar. It is supposed by results of chemical analysis (Table 1), which indicate similarity of major chemical component among the coffee origin.

| Table 1. Chemical characteristic of coffee beans (green bean) |
|---------------------------------------------------------------|
| **Parameters** | **Arabica Java Preanger** | **Arabica Bondowoso** | **Arabica Malang** |
| Water content (%) | 12.22 | 9.87 | 12.92 |
| Ash content (%) | 4.71 | 3.99 | 3.52 |
| Protein (%) | 13.16 | 14.59 | 14.17 |
| Lipid (%) | 13.47 | 13.15 | 13.71 |
| Carbohydrate (%) | 57.52 | 58.40 | 55.67 |
| Caffeine (%) | 1.32 | 1.42 | 1.31 |
| Trigonelline (%) | 0.96 | 0.99 | 0.98 |
| CGA (%) | 6.04 | 9.10 | 8.32 |
Although the spectra are likely similar, but if reviewed in detail, some interesting differences are shown. The chlorogenic acid content in three coffee beans has a significant difference. Arabica Bondowoso coffee has the highest chlorogenic acid content followed by Arabica Malang and the lowest is Java Preanger Arabica, which are 9.10%, 8.32% and 6.04%, respectively.

There are several absorbance peaks and valley that exist around 1100-1150 nm, 1200-1220 nm, 1290-1300 nm, 1450-1480 nm, 1650-1680 nm, 1850-1900 nm, 1920-1940 nm and 2030-2150 nm. Based on [18], the wavelength of 1934 nm correspond to water and protein structure; 2128 nm correspond to carbohydrate structure; 1477, 1726, and 2128 nm corresponds to lipid structure; 1128, 1298, 1672, 1726, and 1934 nm corresponds to caffeine structure; 1128 nm correspond to trigoneline structure; 1477, 1726, 1934, and 2128 nm corresponds to chlorogenic acid.

3.2 Clustering Analysis Using PCA of Spectra Data Pretreatment
The result of clustering analysis using PCA from 5 types of spectra data pretreatment for reflectance and absorbance spectra are shown in Fig. 2 and Fig. 3 respectively. SNV and MSC methods are able to cluster three kinds of coffees based on reflectance and absorbance spectra, although there is still overlapped data among three coffee beans. Reflectance spectra are clearly better in differentiating the characteristics of three coffee beans. Referring to [19], the particle size and scattering are the first principal components (PC1) of the calibration model using reflectance spectra. So it can be concluded that the differentiation of coffee beans characteristics using reflectance spectra in this study was strongly influenced by the particle size and scattering (physical factor) of the coffee beans, beside the chemical content. Meanwhile, the absorbance spectra more represented the condition of the chemical content of the coffee beans. Cluster analysis and the scatter plot could discriminate three kinds of coffee beans, however, it only provides discriminant information qualitatively.

Figure 4 shows the loading spectra of first and second principal component. The first and second principal components (PC1 & PC2) represent most variation in original spectra. Furthermore, the principal component emphasized the spectra signal that was centered in the wavelength of 1880-1936
nm and the spectra signal allegedly showing the O-H group of water molecules. Meanwhile the spectra signals centered at 2140-2200 nm highlight the signal characteristics of phenolic groups. This phenolic group is predicted as chlorogenic acid. This is in accordance with [20] that the spectra signal centered at 4670 cm$^{-1}$ is characterized as a phenolic group present in yerba mate.

**Figure 2.** Cluster analysis using PCA of pretreated reflectance spectra using SGs, D1, D2, SNV and MSC

**Figure 3.** Cluster analysis using PCA of pretreated absorbance spectra using SGs, D1, D2, SNV and MSC
Chlorogenic acid (CGA) is one of the functional quality characteristics in coffee as a phenolic compound of important concern which has antioxidant properties. CGA has been shown to play an important role in human health, and contributes to the acidity and overall quality of final cup coffee such as bitterness [21][22].

Figure 4. Loading spectra of PC1 and PC2 from SNV pretreated reflectance spectra

3.3 Discriminant Analysis
The result of classification using discriminant analysis showed that the overall prediction (calibration) accuracy value is in the range of 95-100%. The DA of PCA of pretreated reflectance spectra has more higher accuracy than that of the absorbance spectra. Even in the validation, the reflectance spectra showed 100% accuracy for 4 spectra data pretreatment i.e. 1st derivative, 2nd derivative, SNV and MSC. Meanwhile, only one absorbance spectrum could achieve 100% accuracy i.e. 1st derivative pretreatment. (Table 2). The results of this discriminant analysis are in line with the PCA results, that the reflectance spectrum was able to classify the coffee beans accurately.

Table 2. Recapitulation of Classification Result of Coffee Beans on The Input Data of PC1, PC2 and PC3

| Data Pretreatment | Cumulative Reliabilities (PCA) (%) | DA (Calibration) n=201(%) | DA (Validation) n=99(%) |
|-------------------|-----------------------------------|--------------------------|------------------------|
| **Reflectance**   |                                   |                          |                        |
| Smoothing-SG      | 100.0                             | 98.0                     | 84.8                   |
| 1st Derivative    | 92.0                              | 100.0                    | 100.0                  |
| 2nd Derivative    | 55.0                              | 100.0                    | 100.0                  |
| SNV               | 99.0                              | 100.0                    | 100.0                  |
| MSC               | 99.0                              | 100.0                    | 100.0                  |
| **Absorbance**    |                                   |                          |                        |
| Smoothing-SG      | 99.0                              | 95.3                     | 88.9                   |
| 1st Derivative    | 96.0                              | 100.0                    | 100.0                  |
| 2nd Derivative    | 69.0                              | 96.7                     | 97.9                   |
| SNV               | 99.0                              | 99.0                     | 98.9                   |
| MSC               | 99.0                              | 99.0                     | 97.9                   |

4. Conclusions
The results showed that based on spectra data pretreatment and PCA analysis, the reflectance spectra gave a better characterization result than the absorbance spectra. The best spectra data pretreatment that able to classify the three types of arabica Java coffee beans based on PC clearly and separately were SNV and MSC. The DA of three principal components of reflectance data processed by SNV and MSC data pretreatment could classify Arabica Java coffee beans accurately (100%). This results shows that NIR spectroscopy can be used as nondestructive method to classify arabica coffee beans based on their origin.
5. References

[1] Andasuryani, YA Purwanto, I W Budiastra, and K Syamsu, “Non destructive and rapid analysis of catechin content in gambir (*uncaria gambir* roxb.) using NIR spectroscopy”, International Journal of Scientific & Engineering Research, vol. 4, pp. 383-389, 2013.

[2] Andasuryani, Y A Purwanto, I W Budiastra, and K Syamsu, “Determination of catechin content in gambir powder from dried gambir leaves quickly using FT NIR PLS model”, International Journal on Advanced Science Engineering Information Technology, vol. 4(5), pp. 2088-5334, 2014.

[3] Lengkey Lady CE Ch, I W Budiastra, K B Seminar, and B S Purwoko, “Prediction model of water content, fat and free fatty acids in three provenan Jatropha Curcasl using NIR spectroscopy with PLS method”, Littri Journals, vol. 19(4), pp. 203-211, 2013.

[4] Huck CW, Guggenbichler W, and Bonn G K, “Analysis of caffein, theobromin, theophylline in coffee by near infrared spectroscopy (NIRS) compared to high performance liquid chromatography (HPLC) coupled to mass spectrometry”, Analytica Chimica Acta, vol. 538, pp. 195-203, 2005.

[5] Zhang Xuan, Wei Li, Bin Yin, Weizhong Chen, Declan P. Kelly, Xiaoxin Wang, Kaiyi Zheng, and Yiping Du, “Improvement of near infrared spectroscopic (NIRS) analysis of caffeine in roasted Arabica coffee by variable selection method of stability competitive adaptive reweighted sampling (SCARS)”, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, vol. 114, pp. 350–356, 2013.

[6] Rosita R, I W Budiastra, and Sutrisno, “Prediction of caffeine content of arabica coffee bean by Near Infrared Spectroscopy”, JTEP Jurnal Keteknikan Pertanian, vol. 4, pp. 179-186, 2016.

[7] Esteban-D’iez I., J.M. Gonzalez-S’aiz, and C. Pizarro, “Prediction of sensory properties of espresso from roasted coffee samples by near-infrared spectroscopy”, Analytica Chimica Acta, vol. 525, pp. 171–182, 2004.

[8] Zulfahrizal, Sutrisno, I W Budiastra, K. B. Seminar, and AA Munawar, “Acquisition of near infrared reflection spectra on cocoa beans”, Littri Journals, vol. 4, pp. 1-10, 2013.

[9] Anonymous, *Official method of analysis of the association of official analytical of chemist*, Washington DC: AOAC, 2006.

[10] Bewketu Mehari, Mesfin Redi-Abshiro, Bhagwan Singh Chandravanshi, Sandra Combrinck, Minaleshewa Atlabachew, and Rob McGrindle, “Profiling of phenolic compounds using UPLC-MS for determining the geographical origin of green coffee beans from Ethiopia”, Journal of Food Composition and Analysis, vol. 45, pp. 16–25, 2016.

[11] Moon JK, Yoo HS, Shibamoto T, “Role of roasting conditions in the level of chlorogenic acid content in coffee beans: correlation with coffee acidity”, J. Agric. Food Chem., vol. 57, pp. 5365–5369, 2009.

[12] Savitzky A, and Golay MJE, “Smoothing and differentiation of data by simplified least squares procedures”, Analytical Chemistry, vol. 36(8), pp. 1627–1631, 1964.

[13] Candolfi A, Maesschalck R D, Jouan-Rimbaud D, Hailey PA, and Massart DL, “The influence of data pre-processing in the pattern recognition of excipients near-infrared spectra”, Journal of Pharmaceutical and Biomedical Analysis, vol. 21(1), pp. 115–132, 1999.

[14] Martens H, and Naes T, *Multivariate calibration by data compression*, In: *Near-Infrared Technology in the Agricultural and Food Industries*, Williams B, Norris B, Ed. St. Paul, Minnesota (USA): American Association of Cereal Chemists. 1987.

[15] Li Xiaoli and Yong He, “Discriminating varieties of tea plant based on Vis/NIR spectral characteristics and using artificial neural networks”, Biosystems Engineering, vol. 99, pp. 313 – 321, 2008.

[16] Cen H and He Y, “Theory and application of near infrared reflectance spectroscopy in determination of food quality”, Trends in Food Science & Technology, vol. 18(2), pp. 72–83, 2007.
[17] Li Xiaoli, Yong He, Changqing Wu, and Da-Wen Sun, “Nondestructive measurement and fingerprint analysis of soluble solid content of tea soft drink based on Vis/NIR spectroscopy”, Journal of Food Engineering, vol. 82, pp. 316–323, 2007.

[18] Ribeiro JS, Ferreira MMC, and Salva, TJG, “Chemometric models for the quantitative descriptive sensory analysis of Arabica coffee beverages using NIR spectroscopy”, Talanta, vol. 83, pp. 1352-1358, 2011.

[19] Murray I, PC William. 1990. Chemical Principles of Near-Infrared Technology, In: Near-Infrared Technology in the Agricultural and Food Industries, 2nd ed., Williams P., Norris K., Ed. St. Paul, Minnesota (USA): American Association of Cereal Chemists, 1990.

[20] C N T Frizon, Gabrieli A, Oliveira, Camila A. Perussello, Patrício G. Peralta-Zamora, Ana M O Camlofski, Überson B Rossa, and Rosemary Hoffmann-Ribani, “Determination of total phenolic compounds in yerba mate (Ilex paraguariensis) combining near infrared spectroscopy (NIR) and multivariate analysis”, Food Science and Technology, vol. 60, pp. 795-801, 2015.

[21] Shan Jiajia, Tetsuhito Suzuki, Diding Suhandy, Yuichi Ogawa, and Naoshi Kondo, “Chlorogenic acid (CGA) determination in roasted coffee beans by Near Infrared (NIR) spectroscopy”, Engineering in Agriculture, Environment and Food, vol. 7, pp. 139-142, 2014.

[22] Tfouni SAV, Serrate CS, Carreiro LB, Camargo MCR, Teles CRA, Cipolli KMVAB, et al., “Effect of roasting on chlorogenic acid, caffeine and polycyclic aromatic hydrocarbons levels in two coffee cultivars: coffea Arabica cv.”, Int J Food Sci Technol, vol. 47, pp. 406-415, 2012.

Acknowledgment

The authors would like to thank the Indonesian Agency for Agricultural Research and Development, the Indonesian Ministry of Agriculture, Directorate of Research Ministry of Research, Technology and Higher Education (STRANAS, FY 2016-2017) and the Indonesian Coffee and Cocoa Research Institute in Jember for the funding and assistance in conducting the research.