

Prediction of Caffeine Content in Java Preanger Coffee Beans by NIR Spectroscopy Using PLS and MLR Method

I W Budiastra¹, Sutrisno¹, S Widyotomo² and P C Ayu¹

¹ Department of Mechanical and Biosystem Engineering, Bogor Agricultural University, Indonesia

² Indonesian Coffee and Cocoa Research Institute, Jember, Indonesia

E-mail: wbudiastra@gmail.com

Abstract. Caffeine is one of important components in coffee that contributes to the coffee beverages flavor. Caffeine concentration in coffee bean is usually determined by chemical method which is time consuming and destructive method. A nondestructive method using NIR spectroscopy was successfully applied to determine the caffeine concentration of Arabica gayo coffee bean. In this study, NIR Spectroscopy was assessed to determine the caffeine concentration of java preanger coffee bean. A hundred samples, each consist of 96 g coffee beans were prepared for reflectance and chemical measurement. Reflectance of the sample was measured by FT-NIR spectrometer in the wavelength of 1000-2500 nm (10000-4000 cm⁻¹) followed by determination of caffeine content using LCMS method. Calibration of NIR spectra and the caffeine content was carried out using PLS and MLR methods. Several spectra data processing was conducted to increase the accuracy of prediction. The result of the study showed that caffeine content could be determined by PLS model using 7 factors and spectra data processing of combination of the first derivative and MSC of spectra absorbance ($r=0.946$; $CV=1.54\%$; $RPD=2.28$). A lower accuracy was obtained by MLR model consisted of three caffeine and other four absorption wavelengths ($r=0.683$; $CV=3.31\%$; $RPD=1.18$).

1. Introduction

Coffee is one of important commodities for Indonesia. It contributes to Indonesia economy and community welfare. Indonesia is rich in arabica coffee, namely arabica java preanger, arabica gayo, mandheling, toraja, and kintamani. These coffee have a unique taste and flavor which influenced by its chemical composition such as carbohydrate, protein, caffeine, trigonelline and chlorogenic acid.

Caffeine is one of the important chemical components of coffee. Caffeine has stimulant properties on our central nervous system and becomes most widely consumed psychoactive substance in the world [1]. Caffeine concentration in coffee bean is usually determined by chemical method which is time consuming and destructive method. Some efforts to find a fast and nondestructive method to determine caffeine content of coffee have been done; one of them is using NIR spectroscopy. However, most of the research was carried out in ground coffee beans [2; 3]. A few researches on NIR spectroscopy to determine caffeine content of coffee bean. Rosita et.al. 2016 successfully applied NIR spectroscopy and PLS to determine the caffeine concentration of Arabica gayo coffee bean. In this research, NIR spectroscopy was assessed to determine caffeine content in java preanger green coffee bean using PLS and MLR method. The inclusion of MLR method in this research aimed to develop a



fastest detection and sorting system for coffee beans in industry in the near future, since MLR method can reduce wavelength numbers and thus reduce processing time compared to PLS method.

2. Materials and methods

2.1. Material and Apparatus

A batch of Arabica java preanger green coffee bean (water content 12–14%) was obtained from Indonesian Coffee and Cocoa Research Institute (ICCRI) Jember, Indonesia. Apparatus used in this study were NIR spectrometer type NIRFlex N-500 (BUCHI Labortechnik AG, Switzerland), LC-MS 2020 (Shimadzu, Kyoto, Japan), computer equipped with unscrambler software v 10.3 (CAMO, Norway) and Microsoft Excel, digital scales, petri dish, and grinder.

2.2. NIRS Measurement

A sample of coffee bean was weighed about 96 g for NIR measurement. The sample was put in a petri dish and arranged into 4 layers [4]. Measurement of NIR spectra was conducted using NIR spectrometer in the wavelength of $10000 - 4000 \text{ cm}^{-1}$ ($1000 - 2500 \text{ nm}$) with 4/cm interval and scan speed of 3 scans/s (Temperature around $22 - 25 \text{ }^{\circ}\text{C}$). The number of samples for NIR testing was 100 samples.

2.3. Chemical Analysis

After NIR spectra measurement, the sample was subjected to chemical analysis. Each sample was grounded using grinder; about 0.2 g of ground coffee was used for caffeine analysis. Caffeine standard, formic acid, acetonitrile, and distilled water were used in chemical analysis. Determination of caffeine by chemical method was conducted using LCMS procedure. To express the amount of caffeine on a dry weight basis (db.), the water content was first determined using AOAC method [5]. Afterwards, samples of 0.2 g of ground coffee were extracted before LC-MS analysis. The LC equipment comprised ESI-APCI combination source (DUIS) and C-18 water was used as column in LCMS analysis with temperature of 40°C . Subsequently, mobile phase were divided into mobile phase A (40% of water/formic acid) and mobile phase B (60% of acetonitrile). The flow rate was 0.2 ml/min , and volume of sample injected was $2 \mu\text{l}$. Determination of caffeine content was achieved by comparing peak area measurement and the calibration graph of caffeine standard with correlation coefficient of 0.999.

2.4. NIR Data Processing and Analysis

The reflectance spectra of samples, consisted of 100 reflectance observation, was transformed into absorbance spectra ($\log(1/R)$) for a linear correlation of NIR values and the caffeine content. In this study, the calibration methods were compared between PLS and MLR method using unscramble software. To obtain the best prediction, several NIR spectra data processing method like first and second derivative, multiple scatter correction (MSC), standard normal variate (SNV) and combination of both first and second derivative with MSC and SNV were performed. Samples were divided into calibration and validation samples that were 67 and 33 samples respectively. Hence, cross-validation was used to test the performance of calibration model. The parameters used to evaluate the capability of model included coefficient correlation (r), standard error (SE), coefficient of variation (CV) and ratio of standard deviation to SE (RPD).

3. Results and discussion

3.1. NIR Spectra and Caffeine Concentration of Java Preanger Coffee Beans

Figure 1 shows the original absorbance spectra of the lowest and highest caffeine concentration in Arabica java preanger green coffee bean. The spectra shape was influenced by several aspects, in particular major chemical concentration (water, carbohydrate, etc.) and the particle size. The peak of

spectra indicated in 1450 nm and 1940 nm that due to water absorption. The peak in 1210 nm and 1780 nm indicate carbohydrate and fat absorption. The caffeine absorption 1128, 1298 and 1672 nm could not be observed in original NIR absorption spectra. Therefore, to analyze the caffeine concentration, the data processing such as derivative method may be required. The difference in absorption between the highest and the lowest caffeine content is very small indicated that determination of caffeine will be difficult.

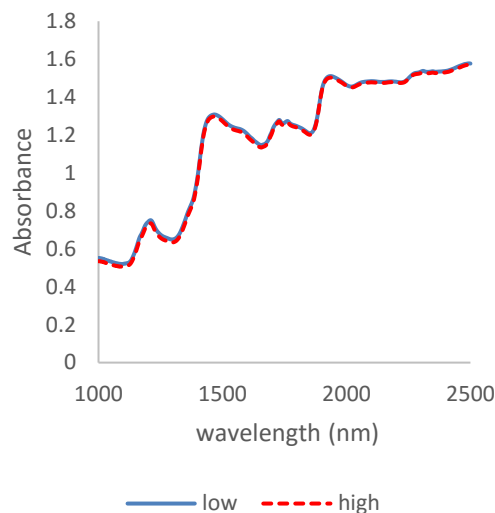


Figure 1. NIR absorbance spectra of the lowest and highest caffeine concentration in java preanger coffee bean

The caffeine concentration in java Preanger coffee bean was low around 1.18 – 1.41 % (table 1). This concentration is higher than concentration of Arabica gayo coffee bean which around 0.194–0.328% [6]. This is due to the difference of origin place, variety, and complex reaction that happened along the processing stage. Also, the water loss in materials also influences the caffeine concentration [3].

Table 1. Caffeine concentration in java preanger coffee beans

Process	Range (%)	Mean (%)	Standard deviation (%)
Calibration	1.18 – 1.41	1.31	0.056
Validation	1.21 – 1.40	1.32	0.047

3.2. Results of calibration and validation

The best calibration and validation result of caffeine using PLS and MLR method is shown in table 2 and 3, respectively. In PLS method, the best calibration model was achieved using combination of first derivative and MSC data processing and 7 PLS factors indicated by the high correlation coefficient (0.946), high RPD (2.28) and low CV (1.54 %). The first derivative was used to enhance the spectral resolution since changes in the gradient are examined so that this data processing suitable for complex spectra [7]. Also, MSC was performed to remove scatter and baseline shift in the obtained spectra [8], afterward the value of error validation decrease which influences RPD and consistency. According to Mouazen et al. [9], RPD value must be higher than 1.5 so that the calibration model can be used.

However, in a study on predicting caffeine concentration in Arabica gayo green coffee bean using NIRS [6], the second derivative method was selected as the best calibration model using PLS method indicated by high correlation coefficient (0.97), low SEC and SEP (0.005 and 0.007%) and low coefficient of variation (1.76%). There is a difference between error value in this research and [6] due to the difference of caffeine concentration, particle size of sample, and other properties in sample. The caffeine concentration in java preanger was around 1.18 – 1.41% while in gayo coffee was around 0.194 – 0.328%. Another properties concentration in sample such as carbohydrate, protein, lipid, and water are also different. This is due to the difference of origin place, variety, and complex reaction that happened along the processing stage [3]. Higher concentration will show a higher absorbance value in the spectra. Moreover, the size of gayo and java preanger coffee is also different. The size of java preanger coffee is bigger than gayo coffee; a bigger size will show bigger absorbance value in the spectra.

Table 2. Result of calibration and validation for caffeine using PLS method

Data Processing	Factors	r	SEC (%)	SEP (%)	CV (%)	RPD	Consistency (%)
Original	7	0,582	0,045	0,042	3,18	1,11	106,76
1st derivative	6	0,943	0,018	0,023	1,72	2,05	81,02
2nd derivative	5	0,953	0,017	0,023	1,7	2,06	74,62
1st derivative ; MSC	7	0,946	0,018	0,0206	1,54	2,28	87,57
2nd derivative ; MSC	4	0,942	0,019	0,023	1,71	2,06	82,33
1st derivative ; SNV	7	0,947	0,018	0,0207	1,55	2,27	87,09
2nd derivative ; SNV	5	0,958	0,016	0,022	1,65	2,13	72,69

On the other hand, MLR method could not give the accurate and robustness prediction (table 3). Increasing variables input of absorption in the number of selected wavelengths could increase accuracy, but the accuracy is still lower than PLS method. The wavelengths selected for predicting caffeine content using MLR method were a combination of caffeine wavelengths (1128, 1298, 1672 nm) [10] and several wavelengths which accommodate the scatter and intercorrelation effects between chemical compounds [11].

The MLR calibration method also produces the models with lower accuracies in predicting the caffeine content of Java Preanger coffee beans, even with variables input of processed absorption spectra using dg1, dg2 and normalization (table 4).

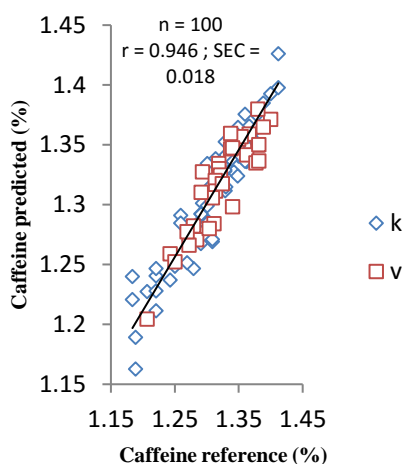
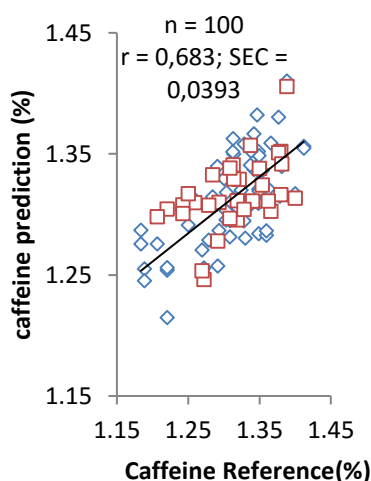
Table 3. Result of calibration and validation for caffeine using MLR method

Combination of selected wavelength	r	SEC (%)	SEP (%)	CV (%)	RPD	Consistency (%)
1128	0,119	0,0535	0,0501	3,81	1,03	106,79
1128, 1066	0,385	0,0497	0,0517	3,93	0,99	96,09
1128, 1066, 1940	0,388	0,0496	0,0524	3,98	0,98	94,70
1128, 1066, 1940, 1298	0,508	0,0462	0,0516	3,92	0,99	89,60
1128, 1066, 1940, 1298, 1672	0,423	0,0488	0,0533	4,05	0,97	91,55
1128, 1066, 1940, 1298, 1672, 1959	0,508	0,0464	0,0495	3,77	1,04	93,62
1128, 1066, 1940, 1298, 1672, 1959, 1032	0,683	0,0393	0,0435	3,31	1,18	90,36
1128, 1066, 1940, 1298, 1672, 1959, 1032, 1125	0,683	0,0393	0,0433	3,29	1,19	90,73
1128, 1066, 1940, 1298, 1672, 1959, 1032, 1125, 1120	0,685	0,0392	0,0437	3,32	1,18	89,76
1128, 1066, 1940, 1298, 1672, 1959, 1032, 1125, 1120, 1063, 1060, 1053	0,695	0,0382	0,0425	3,23	1,21	89,98
1128, 1066, 1940, 1298, 1672, 1959, 1032, 1125, 1120, 1063, 1060, 1053, 1040, 1029	0,709	0,0379	0,0418	3,18	1,23	90,77
1128, 1066, 1940, 1298, 1672, 1959, 1032, 1125, 1120, 1063, 1060, 1053, 1040, 1029, 1114, 1096, 1088, 1074	0,761	0,0380	0,0349	2,66	1,47	108,71
1128, 1066, 1940, 1298, 1672, 1959, 1032, 1125, 1120, 1063, 1060, 1053, 1040, 1029, 1114, 1096, 1088, 1074, 1018, 1015, 1010, 1007, 1006, 1020, 1019	0,844	0,0289	0,0338	2,57	1,52	85,45
1006, 1007, 1010, 1015, 1018, 1019, 1020, 1026, 1029, 1032, 1039, 1052, 1060, 1063, 1066, 1075, 1083, 1088, 1089, 1096, 1106, 1111, 1114, 1120, 1125	0,908	0,0233	0,0255	1,93	1,84	91,32

Table 4. Result of calibration and validation for caffeine using NIR data processing and MLR method

Data processing	r	SEC (%)	SEP (%)	CV (%)	RPD	Consistency (%)
dg1	0,600048	0,043065	0,044899	3,41	1,14	95,916
dg2	0,354821	0,050331	0,051564	3,92	1,00	97,610
n01	0,680787	0,039432	0,044899	3,41	1,14	87,825

Fig. 2 and Fig. 3 showed the scatter plot of reference and prediction values of caffeine using PLS and MLR method (using 7 selected wavelengths), respectively. figure 2 showed a high correlation and a lower SEC between caffeine predicted using NIRS and caffeine reference using LCMS, indicated by r-value > 0.9 and SEC $< 0.023\%$. The results indicate that the obtained calibration model using PLS method could predict caffeine concentration in java preanger green coffee bean. On the other hand, the model developed by MLR method gives a lower accuracy in predicting the caffeine concentration of java preanger coffee beans.

**Figure 2.** Plots of caffeine referenced vs. predicted using PLS method**Figure 3.** Plots of caffeine referenced vs. predicted using MLR method

4. Conclusion

The best calibration model of PLS for good prediction of the caffeine concentration of java preanger coffee bean was obtained by combination of first derivative and MSC data processing and using 7 PLS factors ($r = 0.946$, $CV = 1.54\%$, $RPD = 2.28$ and consistency 87.57%). The accuracy of the PLS model is higher and better than the MLR model. This study showed that NIR spectroscopy and the PLS model can be used to determine the caffeine concentration in Arabica java preanger green coffee bean.

Acknowledgement

We would like to thanks the Directorate of Research, Ministry of Research, Technology and Higher Education for supporting this research (STRANAS Research Grant of FY 2017).

References

- [1] Campo GD, Borrego I, Caracena R and Zuriarrain J 2010 *Talanta* **81** 367-371
- [2] Huck CW, Guggenbichler W and Bonn GK 2005 *Analytica Chimica Acta* **538** 195-203
- [3] Zhang X, Li W, Yin B, Chen W, Kelly DP, Wang X, Zheng K and Du Y 2013 *Molecular and Biomolecular Spectroscopy* **114** 350-356
- [4] Rosita R 2016 *Non-destructive Determination of Arabika Gayo Coffee Chemical Composition with Near Infrared Spectroscopy* (in Indonesian) (Bogor: Graduate School of Bogor Agricultural University)
- [5] [AOAC] Association of Official Analytical Chemyst (USA) 2005 *Official Method of Analysis of the Association of Official Analytical of Chemyst* (Arlington: The association of official analytical chemyst, Inc)
- [6] Rosita R, Budiastira I W and Sutrisno 2016 *JTEP Jurnal Keteknikaan Pertanian* **4** 179-186.
- [7] Stuart B 2004 *Infrared Spectroscopy; Fundamentals and Applications* (John Wiley & Sons, Ltd)
- [8] Shan J, Suzuki T, Ogawa Y, Kondo N 2015 *Engineering in Agriculture, Environment and Food* **8** 195-199.
- [9] Mouazen A M, W Saeys, J Xing, J De Baerdemaeker and H Ramon 2005 *J Near Infrared Spectroscopy* **13** 87-97.
- [10] Ribeiro J S, Ferreira M M C and Salva T J G 2011 *Talanta* **83** 1352-58.
- [11] Budiastira I W, Ikeda Y and Nishizu T 1998 *Journal of JSAM* **60** 117 – 127.