Prediction of water content in Lintong green bean coffee using FT-NIRS and PLS method

Sahfitri, A Rindang* and P C Ayu

Department of Agricultural Engineering, Faculty of Agriculture, Universitas Sumatera Utara, Medan, Sumatera Utara, Indonesia.

E-mail: *adian.rindang@usu.ac.id

Abstract. Water content is one of the coffee compositions that affects the coffee flavour, texture and appearance. Measurement of chemical composition is usually conducted using the chemical method which is known to be expensive and time consuming. Nowadays, Fourier Transform Near Infrared Spectroscopy (FT-NIRS) method is developed in measuring the chemical content of coffee in a short time. The purpose of this study was to predict the water content of Lintong green bean using FT-NIRS and PLS method. The spectra measurement was conducted in the wavelength of Near Infrared of 1000-2500 nm, then followed by the water content measurement using chemical methods. Several data pretreatment was used, namely first derivative, second derivative, multiple scatter correction (MSC), standard normal variate (SNV), first derivative + MSC and second derivative + MSC. Result showed that the best calibration model was obtained using MSC with 7 factors of PLS, indicated by $r = 0.860; \text{SEC} = 0.12; \text{SEP} = 0.14; \text{CV} = 1.14; \text{and RPD} = 2.09$.

1. Introduction

Coffee from various origin is known to have different percentage of chemical content, which cause different flavour and also price. In coffee storage and distribution, water is one of the components that is needed to be controlled to ensure the quality of coffee. In coffee trading, the water content of green bean coffee must be below 12.5% [1].

Determination of chemical compounds in coffee is usually conducted destructively which is known to have several deficiencies such as time consuming and expensive, so it is not appropriate to be used in coffee industry. That is the reason a new method which can answer the destructive method problems is needed.

FT- NIRS is a non-destructive method that can analyse the content of coffee in a short time, free chemical components and cheap. The chemical composition of organic materials is measured using NIR wavelengths of 1000-2500 nm which contains more complex information structures due to the combination pattern of chemical bonds [2]. Nowadays, FT-NIRS has been used to determine the chemical content such as caffeine in Java Preanger Arabica coffee beans non-destructively [3] and also to determine the water content, lipid content, sucrose, caffeine, trigonelline and chlorogenic acid content in coffee beans [4].
2. Material and method

2.1 Materials and equipments
The material used in this study was Arabica Lintong green bean coffee obtained from Humbang Hasundutan Regency. The equipments used in this study were digital scales, petri dishes, grinders, 50 mesh sieves, micro pipettes, measuring cups, ovens, desiccators, FT-NIR spectrometers (NIR Flex N-500, produced by BUCHI Labortechnik AG, Switzerland), computers which is equipped with software unscramble v 10.3 (CAMO, Norway) and Ms. Excel software.

2.2 Methods

2.2.1 Acquisition of coffee beans. The acquisition of reflectance spectra was carried out on each sample of green bean coffee. A total of 96 grams coffee beans [1] were placed in petri dish evenly and tightly with 4 layers of piles [5]. Reflectance measurements were carried out by scanning each sample 3 times at 3 different points by adjusting the petri dish to rotate about 360 degrees during the sample transfer. The NIR wavelength range used was 1000-2500 nm with an interval of 2 nm.

2.2.2 Chemical analysis. Chemical analysis was carried out to obtain the percentage of water content of coffee beans by the thermogravimetric method.

2.2.3 Research data analysis. Spectra data obtained from measurements were reflectance data, then transformed into absorbance (log (I / R)) to obtain a linear correlation between NIR absorption values and chemical data. Several data pretreatment methods such as first and second derivative (dg1 and d2), multiple scatter correction (MSC), standard normal variate (SNV) and a combination of two pretreatments were used to reduce the noise caused by external influences. Moreover, PLS method was used to build the prediction model of water content of Lintong green bean coffee. The optimum PLS factors selection was conducted based on the value of Predicted Residual Error Sum Square (PRESS) in the validation set (V-set-PRESS), where the smallest V-PRESS value (close to zero) and the consistency value between 80-110% then the PLS factor was chosen as optimum PLS factor. After the calibration model was obtained, validation was then performed.

3. Result and discussion

Figure 1 showed the absorbance spectra of Lintong green coffee bean which consisted of several peaks and valleys of absorption of chemical contents to facilitate interpreting the data. The form of NIR graph is influenced by the particle size, the used wavelength and the chemical content of coffee beans. Water content is shown in the wavelengths of 1450 nm and 1940 nm [6].

![Figure 1. Average absorbance of NIRS on green beans Arabica Lintong coffee beans](image-url)
Table 1. Distribution of water content of Arabica green beans

| Process   | N  | Mean% | SD%  | Min% | Max% |
|-----------|----|-------|------|------|------|
| Calibration | 60 | 10.74 | 0.26 | 10.26 | 11.26 |
| Validation | 30 | 10.72 | 0.25 | 10.26 | 11.26 |
| Total      | 90 | 10.73 | 0.26 | 10.26 | 11.26 |

Note: N: number, Mean: Average, SD: Standard deviation, Min: minimum, Max: maximum

Table 1 showed that the average of water content of the Arabica Lintong green bean sample was 10.73% which was lower than 12.5%, which indicated that the coffee samples was suitable with the requirement percentage of coffee quality from Association of Indonesia Coffee Exporters and Industries (AICE) [1].

3.1 Result of NIR data pretreatment

Figure 2. Results of several spectra data processing (a. First derivative of absorbance; b second descendant of absorbance; c. MSC of absorbance; d SNV of absorbance; e. Combination of first derivative and MSC; f. Combination of second derivative and MSC
Figure 2 showed the spectra after being processed using several NIR data pretreatment. Spectra processed by the first derivative method produced peaks and valleys of absorption of chemical contents more than the original spectra (Figure 2a). Pretreatment with the second derivative method produced more peak and valley spectra and was clearer than the first derivative (Figure 2b), because the first and second derivative methods were able to separate the overlapping spectra so that it would bring up the desired chemical content.

Data processing with MSC (Figure 2c) and SNV (Figure 2d) produced almost the same regression line. Initial data processing using MSC and SNV were able to eliminate the effects of multiplicative interference on data distribution, changing beam distance, and particle size, and can improve the effect of multiplicative and additive scatter [1]. Furthermore, the results of a combination of the first or second derivative with MSC (Figures 2e, 2f) produced spectra that were almost similar as shown in the spectra from the first and second derivatives.

3.2 Calibration model with PLS method

Spectra data retrieval in data processing was conducted randomly using the PLS method. The number of PLS factors used will affect the model that will be generated. If taking PLS factors that were too high will cause a low predictive value, so the data will be overfitting due to excess factor x [5]. On the other hand, if the selected PLS factors were too low, it will produce an underfitting model. Then, the optimum PLS factor selection is a consideration to build a good PLS model.

Evaluation of the calibration model was needed to be conducted to obtain the best calibration model. The parameters used for calibration and validation were r, SEC, SEP, CV, RPD and consistency. According to [7] a good prediction model is r value approaching 1, with a consistency value range of 80-110%, and SEC and SEP values must be small close to 0. Lammertyn et al (2000) stated that the difference in value is small between SEC and SEP values indicates a good calibration model, conversely if it has a large difference in value, it indicates that the calibration set does not represent a validation set. The RPD value will also show the accuracy of a good prediction model that is> 2. If the RPD value produced <1.5 calibration results cannot be used [8].

Table 2. Calibration and validation results for the content of water content in coffee beans

| Point | Pre-treatment | Factor | R   | SEC | SEP | CV  | RPD | Consistency |
|-------|---------------|--------|-----|-----|-----|-----|-----|-------------|
| Original | 7 | 0.815 | 0.13 | 0.15 | 1.30 | 1.84 | 88.93% |
| MSC    | 7 | 0.860 | 0.12 | 0.14 | 1.14 | 2.09 | 86.82% |
| SNV    | 7 | 0.860 | 0.12 | 0.14 | 0.14 | 2.09 | 86.63% |
| dg₁    | 3;4 | 0.845 | 0.12 | 0.14 | 1.20 | 2.00 | 85.94% |
| dg₂    | 1 | 0.848 | 0.12 | 0.16 | 1.19 | 2.00 | 79.24% |
| dg₁ + msc | 3;3 | 0.86 | 0.11 | 0.16 | 1.11 | 2.15 | 73.96% |
| dg₂ + msc | 1;1 | 0.78 | 0.15 | 0.18 | 1.40 | 1.70 | 81.96% |
Results of calibration and validation of water content of ArabicaLintong green bean coffee can be seen in Table 2. The best calibration model was using MSC pre-treatment with 7 PLS factors produced a value of $r = 0.860$; SEC $= 0.12\%$; SEP $= 0.14\%$; CV $= 1.14$; and RPD $= 2.09$. These results indicated that the PLS model that was built has a good accuracy.

Figure 3 showed the plot of FT-NIR result and the result of chemical methods for water content. It can be seen that the reference data (chemical data) was not much different from the results of FT-NIRS predictions, although there were some inaccurate data. However, the model obtained from this study could already be used to predict the water content in ArabicaLintong green coffee bean.

4. Conclusion
The water content of ArabicaLintong green bean coffee can be predicted using FT-NIRS and PLS method with 7 PLS factors, indicated by value of $r$ of 0.860; SEC of 0.12; SEP of 0.14; CV of 1.14 and RPD of 2.09.

References
[1] Ayu P C 2017 Pengembangan model penentuan kandungan kimia utama pembentuk flavour biji kopi java preanger menggunakan FT-NIRS [Development of determination model of main flavour attributes in java preanger green coffee beans using FT-NIRS] Thesis Bogor: IPB University
[2] Cen H and He Y 2007 Theory and Application of Near Infrared Reflectance Spectroscopy in Determination of Food Quality *J Trends in Food Sci & Technol* 18 pp 72-83
[3] Budiastra I W, Sutrisno, Widiotomo S and Ayu P C 2018 Prediction of caffeine content in Java Preanger coffee beans by NIR spectroscopy using PLS and MLR method *IOP Conf Series: Earth and Environmental Science* 147 (2018) 012004 doi :10.1088/1755-1315/147/1/012004
[4] Caporaso N, Martin B W, Stephen G and Ian D Fisk 2018 Rapid prediction of single green coffee bean moisture and lipid content by hyperspectral imaging *Journal of Food Engineering* 227 pp 18-29
[5] Rosita R, Budiastra I W and Sutrisno 2016 Prediction of caffeine content of Arabica coffee bean by Near Infrared Spectroscopy *JTEP Jurnal Keteknikan Pertanian* 4 (2) pp 179-86
[6] Heruwati I 2011 Deteksi gejala chilling injury buah belimbing (Averrhoa Carambola L) yang disimpan pada suhu rendah dengan NIR spectroscopy [Detection of chilling injury symptoms of star fruits (Averrhoa Carambola L) stored at low temperature by NIR spectroscopy] Undergraduate *Thesis* (Bogor: IPB University)
[7] Nircal 5 5 manual 2013 *Nircal manual* (Switzerland: Buchi Labortechnik AG, CH Flawil)
[8] Mouazen A M, W Saeys, J Xing, J De Baerdeamaker and H Ramon 2005 Near infrared spectroscopy for agricultural materials: an instrument comparison J Near Infrared Spectroscopy 13 pp 87-97

Acknowledgements
The research was funded by Universitas Sumatera Utara through the Non PNBP Fund Fiscal in 2019 based on the Implementation Contract of Universitas Sumatera Utara TALENTA Research Fiscal in 2019 between the Rector and the Chairman of Universitas Sumatera Utara Research Institute Number: 4167/UN5.1.R/PPM/2019 dated April 1st 2019.