Improving the accuracy of near-infrared (NIR) spectroscopy method to predict the oil content of oil palm fresh fruits

I Novianty¹, K B Seminar², Irzaman³, I W Budiastra²

¹Agricultural Engineering Sciences Program, Department of Mechanical and Biosystem Engineering, IPB University (Bogor Agricultural University), Bogor, Indonesia
²Department of Mechanical and Biosystem Engineering, IPB University (Bogor Agricultural University), Bogor, Indonesia
³Department of Physics, IPB University (Bogor Agricultural University), Bogor, Indonesia
⁴Surfactant and Bioenergy Research Center, IPB University (Bogor Agricultural University), Bogor, Indonesia

Email: inna.novianty@gmail.com

Abstract. Oil content is an important factor to determine the price of the oil palm fruit. In this study, the NIR Spectroscopy method was conducted to determine the oil content of oil palm fresh fruits. Several studies that related to the NIR spectroscopy method for predicting the oil content of oil palm fruits showed that the accuracy was still optimal yet. This study is the initial stage to develop the optical portable instrument that can predict the oil content of oil palm fresh fruits. Five hundred samples which are divided into ten groups based on maturity levels were prepared for reflectance and oil content chemical measurement. The reflectance of the sample was measured by the FT-NIR spectrometer in the wavelength of 1000-1500 nm. The obtained spectrum of oil palm fresh fruits was transformed to absorbance (Log 1/R) and several spectra data processing was conducted to increase the accuracy of prediction. The calibration and validation of processed NIR spectra and the oil content were conducted using Partial Least Square (PLS) and MLR methods. Generally, PLS and MLR methods can be used to improve the accuracy of the NIR spectroscopy method. Then, the result showed that the MLR method is less accurate than the PLS method to predict the oil content of oil palm fresh fruits. In this study, the best result could be determined by the PLS model using 5 factors and spectra data processing of the first derivative of spectra absorbance (R=0.879; CV=19.8%; RPD= 2.46). A lower accuracy was obtained by MLR model (R=0.677; CV=28.33%; RPD =1.72).

1. Introduction
The oil palm is an important energy crop plentiful in Southeast Asia. Indonesia is one of the world leaders in palm oil production and export for oil palm production in Southeast Asia [1]. Palm oil plays an important role as a bioenergy source, it can be used as a fuel for generating internal combustion engines in vehicles and for generating heat or electricity [2]. Palm oil became the world's second most important vegetable oil after soybean oil in 1980 [3] and mainly used in many food categories, such as oil, cooking margarine, white butter, ice cream, milk, cream, health supplements such as vitamin A and vitamin E, etc [2].
The oil content is one of the important chemical components of the oil palm fruit. The term oil content (OC) means the fraction of the weight of extracted oil per weight of dried mesocarp [1]. Oil content is usually determined by the chemical method (soxhlet extraction) which is time consuming and destructive method. Nowadays, many researchers are trying to estimate OC quickly using different techniques. Near-infrared spectroscopy (NIR) became one of the techniques to determine the OC in fresh palm fruits. NIR spectral analysis has been widely used for assessment of agricultural products’ quality as non-destructive technique, rapid and cost-efficient techniques. Kasemeumran et al. 2012 were successfully investigated the feasibility of non-destructive near-infrared (NIR) spectroscopy and partial least squares regression (PLSR) models for the determination of palm oil in oil palm fruits.

The objective of the present study is applying NMR spectroscopy for the quantitative determination of the oil content in oil palm fresh fruits. The oil content chemical measurement using soxhlet extraction was conducted after NIR spectroscopy measurement. Then, several spectra data pre-processing was conducted to increase the accuracy of prediction such as the first derivative, normalized, a combination between the first derivative and normalized. This study is the initial stage to develop the optical portable instrument that can predict the oil content of oil palm fresh fruits. In this research, NIR spectroscopy was assessed to determine oil content in oil palm fresh fruits using partial least squares (PLS) and multilinear regression (MLR) method. MLR method was used because this method can reduce wavelength numbers and reduce processing time compared to (PLS) method [4].

2. Materials and Methods

2.1. Materials
Five hundred samples which are divided into ten groups based on maturity levels (3, 4, 5, 6 months old, 4 months 1 week old, 4 months 2 weeks old, 4 months 3 weeks old, 5 months 1 week old, 5 months 2 weeks old, 5 months 3 weeks old). Fresh oil palm bunches (FPB) of the Tenera species were obtained from the Cikabayan Experimental Garden Departement of Agronomy and Horticulture IPB University, Indonesia. The apparatus used in this study were NIR spectrometer type NIRFlex N-500 (BUCHI Labortechnic AG. Switzerland), the spectra of each sample were collected for pretreatment by Unscrambler v 10.4 classifier chemometrics software package (CAMO, Trondheim, Norway), Microsoft Excel, filter paper, electrical heater, oven, digital scales.

2.2. NIR Spectra Measurements
A sample of oil palm fresh fruit was taken from fresh oil palm bunches (FPB). The number of samples for NIR testing was five hundred samples (oil palm fresh fruits) which are divided into ten groups based on the maturity levels. Measurement process of NIR spectra was conducted using the NIR spectrometer in the wavelength of 1000-1500 nm with a 4/cm interval, a scan speed of 3 scans/s and the ambient temperature around 22 – 25 °C.

2.3. Soxhlet Extraction
In this study, the oil content destructive method was conducted by chemical method (soxhlet extraction) In the oil content determination, samples were weighed as W (7-10 g), and then extracted at 80 °C for 6 h by Soxhlet extractor with hexane as a chemical solvent. The residues were dried at 105 °C in a vacuum oven to the constant weight as W0 (g). The result was recorded for mesocarp oil content calculation OC% = W0/W ×100%. This gravimetric procedure was defined in accordance with standards established by the National Standardization Body of Indonesia SNI 01-2891-1992.
2.4. **NIR Data Processing and Analysis**

Multiple linear regression (MLR) and partial least square (PLS) analysis with respect to the logarithms of the reflectance reciprocal or absorbance spectra (log (1/R)) are widely used for calibrating reflectance NIR spectral analysis model[5]. The model performance was evaluated by comparing the oil content prediction results and oil content measured values. The main performance statistics used for model validation were the coefficient of correlation (R), the standard error of calibration (SEC), the standard error of prediction (SEP), coefficient of variation (CV), residual predictive deviation (RPD) and consistency. Samples were divided into calibrations and cross-validations data sets that were 333 and 167 samples. The performance of the model on calibration was highly related to the SEC results while the validation process and the SEP results confirmed the validation of the model. The model was considered appropriately accurate when the R-value was high while SEC and SEP values were low. Then, the model was having small differences in the value of SEC and SEP [6]. In order to increase model performance, the spectral reflectance data were pre-processed such as the first derivative, normalized, a combination between the first derivative and normalized. In this study, the PLS method was compared to the MLR method using unscrambler software.

3. **Results and Discussion**

3.1. **NIR Spectra and Oil Content Concentration of Oil Palm Fresh Fruits**

Figure 1 shows the original absorbance NIR spectra in the 1000-1500 nm of different maturity levels in the oil palm fruit samples. 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 around 1200 nm due to oil content and the intense water band absorption around 1450 nm. The peak absorption that occurs at the wavelength of 1210 nm, 1161 nm, 1188 nm, 1212 nm, 1387 nm indicates the oil absorption [7].

In previous researches, NIR spectroscopy utilizes the spectral range from 780 to 2500 nm and provides complex structural information related to the vibration behavior of combinations of bonds [8]. The oil content which is dominated by triglycerides (TAG) have molecular bonds such as CH, CH₂, dan R-CO₂ [9]. Based on the original absorbance of NIR spectra, the oil content was dominated by triglycerides which are composed of CH and CH₂ bonds indicated at the wavelength of 1190-1219 nm [10]. The bands in the 1150–1200 nm region are assigned to the second overtone of the C–H stretching modes, those in the 1400–1450 nm are due to the first overtone of the O–H stretching modes and the combination of the C–H stretching and C–H deformation modes [11].
The oil content of oil palm fresh fruits (Tenera species) in this research was around 1.6-35.2% as shown in Table 1. Zaqlul Iqbal 2015 assessed the oil content of the oil palm for 4 months until 7 months old maturity level which around 1.25% - 53.91%. This difference of the oil content is due to the difference of origin place, with a higher value of the oil content in the wettest climate and the other reaction during the ripening process. In the other research, the distribution of the palm oil content in all samples was around 4.21–50.00% and 50.01–88.13% representing the groups of unripe and ripe oil palm fruits[15]. The oil palm fresh fruits of their research were collected from oil palm estates of PT SMART Tbk (Sinar Mas Agribusiness Resources and Technology) while our research was conducted in the Cikabayan Experimental Garden Departement of Agronomy and Horticulture IPB University, Indonesia. It can be seen from Fig 1, the higher the palm oil content the higher the absorbance value of spectra. Moreover, the oil palm fruits have different stages of ripeness, the more mature oil palm fruits are, the bigger absorbance value of spectra.

| Process   | Range (%) | Mean(%) | Standard deviation (%) |
|-----------|-----------|---------|------------------------|
| Calibration | 1.6-35.2 | 19.88   | 6.17                   |
| Validation  | 1.5-30.4 | 16.93   | 8.24                   |

3.2. Result of calibration and validation
In many NIRS analysis, many calibration models of NIR spectra from derivative were often built and gave good prediction results[8]. Generally, it is very important to build a reliable calibration model for quantitative or qualitative analysis in an agricultural material analysis combined with chemometrics to extend the NIRS applications [8].

The best calibration and validation result of oil content using PLS and MLR method is shown in Tables 2 and 3, respectively. To analyze the oil content, the data processing such as derivative method
(dg1), normalization method (n01) or combination of normalization method (n01) and derivative method (dg1) may be required. The first derivative was used to enhance the spectral resolution since changes in the gradient are examined then this data processing suitable for complex spectra [13]. While the normalization data processing was designed to reduce baseline variations caused by differences in grain size [14]. The result of data processing methods is shown in Table 2.

In the PLS method, the best calibration model was achieved using first derivative data processing and five PLS factors indicated by the high correlation coefficient (0.879), high RPD (2.46) and CV value (19.8 %). This model was considered to be sufficiently accurate with a high R-value (0.87), even though the CV value was still large (19.8 %). This value can be reduced by divided oil content data range into two parts (1.5-17 % and 18% - 35.2%). However, the oil content data range used in this study was 1.5%-35.2%. Then, the model was having small differences in the value of SEC and SEP (0.4112 %). The value of R (0.70 to 0.99) was considered appropriately accurate which indicates a good correlation between actual data with spectra data [12].

| Data processing          | Calibration (n=333) | Validation (n=167) | Consistency (%) |
|--------------------------|---------------------|--------------------|-----------------|
|                          | R                   | SEC (%)            | SEP (%)         | RPD  | CV  | (%)
| original                 | 0.662               | 4.6285             | 4.9275          | 1.67 | 29.11 | 93.93 |
| dg1                      | 0.638               | 4.7519             | 5.5343          | 1.49 | 32.69 | 85.86 |
| n01                      | 0.677               | 4.5435             | 4.7954          | 1.72 | 28.33 | 94.74 |
| combination of n01 and dg1| 0.656               | 4.6581             | 5.4081          | 1.52 | 31.95 | 86.13 |

On the other hand, the MLR method was employed to improve model accuracy and efficiency by selecting suitable wavelengths. The wavelengths selected for predicting the oil content using the MLR method were a combination of the oil wavelengths absorption (1210, 1161, 1188, 1212, and 1387 nm) [7]. This method could not give an accurate prediction (Table 3). The MLR calibration method also produces the models with lower accuracies in predicting the oil content of oil palm fresh fruits, even with variables input of processed absorption spectra using the derivative method (dg1), normalization (n01) and combination of dg1 and normalization.

| Data processing          | Factors | Calibration (n=333) | Validation (n=167) | Consistency (%) |
|--------------------------|---------|---------------------|--------------------|-----------------|
|                          |         | R                   | SEC (%)            | SEP (%)         | RPD  | CV  | (%)
| original                 | 7       | 0.673               | 4.5649             | 5.2562          | 1.57 | 31.05 | 91.04 |
| dg1                      | 5       | 0.879               | 2.9412             | 3.3524          | 2.46 | 19.8  | 87.74 |
| n01                      | 7       | 0.802               | 3.6893             | 3.9769          | 2.07 | 23.49 | 90.28 |
| combination of n01 and dg1| 7       | 0.861               | 2.5728             | 3.5426          | 2.33 | 20.93 | 72.62 |
The scatter plot of reference and prediction value of the oil content using PLS and MLR method (using 10 selected wavelengths) as shown in Fig. 2 and Fig. 3 respectively. Figure 2 showed a high correlation and a lower SEC between oil content predicted using NIRS and oil content reference using soxhlet extraction, indicated by an R-value of 0.879. The obtained SEP (3.35%) was close to the SEC value (2.94%). Furthermore, the R-value of the original spectra was 0.673, while the first derivative (dg1) data processing has an R-value of 0.879. This indicated that the PLS method is reliable for predicting other samples within the full range of oil content and the PLS method successfully improves the accuracy of the NIRS method. On the other hand, the model developed by the MLR method gives a lower accuracy in predicting the oil content of oil palm fresh fruits, but R-value was not improved significantly.

Figure 2. Plots of oil content referenced vs. predicted using MLR method

Figure 3. Plots of oil content referenced vs. predicted using the PLS method
4. Conclusion

The study reports that the application of NIR spectroscopy for determining the oil content of oil palm fresh fruits samples was successfully achieved with good prediction. The optimum calibration model of the PLS method of oil palm fresh fruits was successfully obtained using five PLS factors and spectra data processing of the first derivative of spectra absorbance (R= 0.879; CV=19.8%; RPD= 2.46 and consistency 87.74%). The accuracy of the PLS model is higher and significantly improving the accuracy of the oil content prediction than the MLR model. Thus, this study showed that NIR spectroscopy and the PLS model can be used to determine the oil content of oil palm fresh fruits.

References

[1] Nokkaewa R and Punsuvon V 2016 Microwave drying combined with magnetic stirring for quick analysis of oil content of fresh palm bunch Science Asia 42 2017-212
[2] Kasemsumran S, Thanapase W, Punsuvonb V and Ozakic Y 2012 A feasibility study on non-destructive determination of oil content in palm fruits by visible–near infrared spectroscopy J. Near Infrared Spectrosc 20 687–694
[3] Sambanthamurthi R, Sundram K and Tan Y A 2000 Chemistry and biochemistry of palm oil progress in Lipid Research 39 507-558
[4] Budiadstra 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 IP Conf. Series: Earth and Environmental Science 147 012004
[5] Liu Y, Chen X, and Ouyang A 2008 Nondestructive determination of pear internal quality indices by visible and near-infrared spectrometry LWT – Food Science and Technology 41 1720–1725
[6] Makky M and Soni P 2014 In situ quality assessment of intact oil palm fresh fruit bunches using rapid portable non-contact and non-destructive approach journal of Food Engineering 120 248–259
[7] Instrumentation Research Laboratory Beltsville Agricultural Research Center U.S. Departement of Agriculture Beltsville MD
[8] Cen H and He Y 2007 Theory and application of near infrared reflectance spectroscopy in determination of food quality Trends in Food Science & Technology 18 72-83
[9] Ketaren S 1986 Minyak dan Lemak Pangan (Jakarta: UI-Press)
[10] Iqbal Z 2015 Prediction of Oil Palm Fresh Fruit Bunch (FFB) Chemical Content Using NIR Spectroscopy (in Indonesian) (Bogor: Graduate School of Bogor Agricultural University)
[11] Osborne BG, Fearn T and Hindle P H 1993 Practical NIR Spectroscopy With Application in Food and Beverage Analysis (Singapore: Longman Singapore Publishers (Pte) Ltd)
[12] Saranwong S, Sornsrisvichai J, and Kawano S 2004 Prediction of ripe stage eating quality of mango fruit from its harvest quality measured nondestructively by near infrared spectroscopy journal of Near Infrared Spectroscopy 131 137-145
[13] Stuart B 2004 Infrared Spectroscopy: Fundamentals and Applications (John Wiley & Sons, Ltd)
[14] Udelhoven T, Emmerling C and Jarmer T 2003 Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial least-square regression: A feasibility study. Plant and Soil 251 319–329
[15] Sudarno, Silalahi D D, Risman T, Widyastuti B L, Davrieux F, and Yuan YY 2017 Caliman Rapid determination of oil content in dried-ground oil palm mesocarp and kernel using near infrared spectroscopy Journal of Near Infrared Spectroscopy 25 338–347
[16] Zhang X, Li W, Yin B, Chen W, Kelly D P, Wang X, Zheng K, and Du Y 2013 Molecular and Biomolecular Spectroscopy 114 350-356