Near-Infrared Technology in Agriculture: Non-Destructive Determination of Inner Quality Parameters in Intact Cocoa Beans

Zulfahrizal Zulfahrizal* and Agus Arip Munawar

Department of Agricultural Engineering, Syiah Kuala University, Banda Aceh 23111, Indonesia

(*Corresponding author’s e-mail: zulfahrizal@unsyiah.ac.id)

Received: 1 June 2020, Revised: 1 May 2021, Accepted: 19 May 2021

Abstract

This present study aimed to apply the near-infrared technology based on reflectance spectroscopy or NIRS in determining 2 main quality attributes on intact cocoa beans namely fat content (FC) and moisture content (MC). Absorbance spectral data, in a wavelength range from 1000 to 2500 nm were acquired and recorded for a total of 110 bulk cocoa bean samples. Meanwhile, actual reference FC and MC were obtained using standard laboratory approaches and Soxhlet and Gravimetry methods. Samples were split onto calibration and validation datasets. The prediction models, used to determine both quality attributes were developed from the calibration dataset using 2 regression methods: Principal component regression (PCR) and partial least square regression (PLSR). To obtain more accurate and robust prediction performance, 4 different spectra correction methods namely baseline shift correction (BSC), mean normalization (MN), standard normal variate (SNV), and orthogonal signal correction (OSC) were employed. The results showed that PLSR was better than PCR for both quality parameters prediction. Moreover, spectra corrections enhanced the prediction accuracy and robustness from which OSC was found to be the best correction method for FC and MC determination. The prediction performance using validation dataset generated a correlation coefficient (r), ratio prediction to deviation (RPD), and ratio error to range (RER) indexes for FC were 0.93, 3.16 and 7.12, while for MC prediction, the r coefficient, RPD and RER indexes were 0.96, 3.43 and 9.25, respectively. Based on obtained results, it may conclude that NIRS combined with proper spectra correction and regression approaches can be used to determine inner quality attributes of intact cocoa beans rapidly and simultaneously.

Keywords: Cocoa, NIRS, Prediction, Quality, Technology

Introduction

Cocoa beans are mainly used as raw materials in chocolate products industries which is very popular among people worldwide. In the highly competitive market, chocolate is generally preferred by people as a solid-state of cocoa materials with additional sugar and other important ingredients [1]. Moreover, chocolate can be served as cold and warm beverages in daily morning or afternoon. If it is consumed in a proper amount, chocolate was to be believed to have a health impact [2,3]. Before being processed onto chocolate products, cocoa beans were taken from the pod of the cocoa tree, roasted, fermented, and grounded [4]. Recently, chocolate plays a strategic and promising position in the food industry segment, since this product can be consumed directly and also in form of other foodstuffs. Agricultural products and the food industry need to be ensured they are being supplied with high-quality raw materials to be processed. For cocoa beans, FC and MC are 2 of the main inner quality attributes which are considered and related to chocolate product qualities [5-7].

To determine the inner quality parameters of cocoa beans and other agricultural products, several methods were widely employed. However, most of these methods are based on liquid and solvent extraction followed by other laboratory procedures [8]. These procedures are often time-consuming, required a complicated sample preparation, laborious, destructive and may cause environmental pollution, since some methods involve additional chemical liquids and materials [8-10]. Therefore, it is unsuitable to be applied in commercial industries. The industry needs to be facilitated with a proper and ideal rapid method that can be able to monitor real-time cocoa processing steps. Thus, facilitating urgent and important decisions to be taken as early as possible [11,12].
The need for fast, robust, and non-destructive methods for the analysis of cocoa raw material, has been one of the essential objectives of the cocoa industry and manufacturer over the last decades. In the last few decades, literature reported that near-infrared technology based on reflectance spectroscopy or known as NIRS has become one of the most promising and significant alternative methods that can be used as a rapid and non-destructive approaches for quality attributes determination of several foods and agricultural products [10,13,14]. The main advantages of NIRS are fast, cost-effective, simple preparation, non-destructive and environmental friendly since no chemical materials are used. More importantly, NIRS has the potential ability to determine several inner quality attributes of studied samples simultaneously [8,15].

Numerous studies and publications have been carried out reported related to NIRS applications in agriculture for rapid quality assessment of foods [16-18], soil nutrients and contaminations [19-21], intact fruits [13,22-24]. Particularly, for cocoa quality assessment, NIRS has been also employed for the inner quality determination and classification based on fermentation stages [25-28]. Based on those reported studies, we may argue that the NIRS technology is feasible to determine the inner quality parameters of organic materials. Reported previous research found that NIRS can be applied and used as a rapid and non-destructive method in determining several quality attributes of raw intact organic materials. Therefore, the main aim of this present study is to study and apply the NIRS as a fast, robust, and simultaneous method for cocoa bean quality assessment especially in determining FC and MC of intact cocoa bean samples. We also investigated the impact of 4 different spectral data correction algorithms on the prediction accuracy and robustness.

The proposed novelty offered in this present study is the instrument that we used is a self-developed portable sensing device of near-infrared spectroscopy (PSD NIRS i16 Iptek). The advantage of this instrument is that the size and price are smaller and cheaper compared to common other NIRS instruments with dimensions 21×17×2 cm³ [29]. Yet, the instrument has similar a wavelength range in the near-infrared region (around 780 - 2500 nm) with a maximum of 8× optical gain. Moreover, we use an original cocoa cultivar from Indonesia and the models are developed for unfermented and fermented cocoa beans.

**Materials and methods**

**Cocoa bean samples**

The cocoa samples used in this present study are cocoa beans, cultivar *Lindak* which are harvested from June to August in the same cocoa plantations in East Java, Indonesia. We used a total of 110 bulk cocoa bean samples amounted to 60 g per bulk. Cocoa bean samples were dried using a mechanical dryer to obtain cocoa beans which are suitable for short-period storage. Samples contain unfermented and fermented beans with various fermentation stages (1, 3, 5 and 7 fermentation days). Furthermore, samples were split into 2 datasets namely the calibration dataset (72 samples) and validation (38 samples).

**Spectral data acquisition**

The near-infrared spectral data of all samples were taken in form of diffuse reflectance or also known as absorbance spectrum using the portable near-infrared instrument PSD NIRS i16 Iptek. Spectral data were obtained in a wavelength range from 1000 to 2500 nm with a resolution of 0.02 nm and co-added 32 scans per spectra data acquisition.

**Actual reference fat and moisture content measurement**

After spectral data collection was completed, all cocoa bean samples were taken directly to measure their inner quality attributes in form of FC and MC and to measure their inner quality parameters in form of FC and MC. At first, FC was measured using the Soxhlet method [30, 31]. Ten g of sample was mixed in the tube with a maximum of 150 mL n-hexane and extracted in Soxhlet apparatus at a temperature of 95 ºC for 6 h. FC was then determined by evaporating those solvents using a rotary evaporator until it is only fat liquid is left in the tube and expressed in percentage (%) FC. On the other hand, MC of cocoa bean samples was measured using a Gravimetry method and measured based on ISO 6673 in duplicate and then averaged [32]. A forced-air electrical oven (Thermicon type UT6120, Heraeus Instruments GmbH, Hanau, Germany) was used to dry approximately 15 g whole intact beans in open glass petri dishes (diameter: 14 cm, height: 2.3 cm) at 120 ºC for 18 h. After drying is completed, the petri dishes were closed with glass lids immediately to avoid exposure and then stored in desiccators for 1 h in to equilibrate samples into ambient temperature. The MC is expressed in percentage (%) dry bulb.
Building prediction models
In the next step, after spectra data acquisition and actual FC and MC measurement, we established prediction models from the calibration dataset that are used to determine both mentioned quality parameters. At first, we attempted to develop prediction models using original untreated and uncorrected spectral data by regressing absorbance spectra data (X variables) and actual quality parameters (Y variables). We employed 2 regression approaches namely PCR and PLSR to develop the models. The prediction results were compared and chose the best one between them. The chosen regression approach was then used to develop other prediction models using corrected spectra data and investigated the impact of those corrected spectra on the prediction performances.

Spectral data correction and enhancement
Spectral data may contain noise due to light scattering and may interfere with prediction accuracy. Thus, we need to correct and enhance those spectra data to improve prediction accuracy and robustness. In this study, we employed 4 different spectra correction methods, namely: BSC, MN, SNV and OSC.

Prediction performance evaluation
To obtain robust prediction performance, we conducted also a cross-validation using K-fold cross-validation where K = 10 which means, there are 7 randomized samples were excluded during calibration, and those 7 samples were used to test the models. This step was repeated 10 times until all folds is completed with different samples each time the cross-validation is performed. Prediction model’s performance was evaluated based on calibration and cross-validation results according to the correlation coefficient (r), the root mean square error (RMSE), and the residual predictive deviation (RPD) index obtained by dividing the standard deviation of reference data with the RMSE value. Last, but not least, we took into account the number of latent variables (LVs) required to develop prediction models. Ideal and robust models should have a higher r coefficient and RPD index, lower RMSE and fewer LVs [8,34]. We then systematically compared the prediction performances for FC and MC prediction based on those statistical indicators. The best spectra correction method was chosen and used to predict other external cocoa bean samples. For more certain applications, external validation was conducted using a validation dataset consisting of a total of 38 unknown cocoa bean samples.

Results and discussion
Spectra features of cocoa bean
The typical absorbance or diffuse reflectance spectrum for intact cocoa beans in the NIR region is presented in Figure 1. It has chemical information buried and highly correlated with the presence of related quality attributes as derived from the bands resulted from the interaction between electromagnetic radiation and organic material like cocoa bean samples. These bands correspond to specific molecular bonds of O-H, C-H, C-H-O, C-O and N-H. This information can be revealed by a specific method which is called chemometrics. It is the use of mathematical and statistical methods like regression, normalization, and validation to correlate the spectral data with actual reference-quality parameters measured using standard laboratory procedures [8,35].

Figure 1 Near-infrared reflectance spectra features of intact cocoa bean sample in wavelength range from 1000 to 2500 nm.
As we knew, the inner quality parameters of intact cocoa beans like FC and MC were constructed by molecular bonds of C=H-O and O-H, respectively. Thus, we may argue that near-infrared reflectance spectroscopy can predict these quality parameters satisfactorily. The peak and valley of the absorbance spectra data in a certain wavelength region correspond to the strength of vibrated molecules of chemical bonds due to the interaction between electromagnetic energy and biological object. This is also in agreement with preceding research confirming that the absorbance and peak are related to overtone and bending vibrated due to light strike or in other words, by electromagnetic energy [6,36].

**Fat and moisture content prediction**

In this reported study, we attempted to develop prediction models used to determine both inner quality parameters of intact cocoa beans simultaneously. Two different regression approaches namely PCR and PLSR were applied to establish those models by regressing original untreated spectra data as independent variable (X) and FC and MC data as the dependent variables (Y). Prediction results for both quality parameters are shown in Table 1. In general, FC and MC of intact cocoa bean samples can be predicted quite satisfactorily with a maximum correlation coefficient was 0.85 for MC and 0.82 for FC prediction.

<table>
<thead>
<tr>
<th>Quality parameters</th>
<th>Method</th>
<th>Statistical indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>r</td>
</tr>
<tr>
<td>Fat content</td>
<td>PCR</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>PLSR</td>
<td>0.82</td>
</tr>
<tr>
<td>Moisture content</td>
<td>PCR</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>PLSR</td>
<td>0.85</td>
</tr>
</tbody>
</table>

PCR: Principal component regression, PLSR: Partial least square regression, r: Correlation coefficient, RMSE: Root mean square error, RPD: Residual predictive deviation.

The RPD index for FC and MC prediction were 1.81 and 1.97, respectively. Based on the literature, the RPD index between 1.5 and 2.0 was categorized as coarse sufficient prediction models and needs to be enhanced [8,14,37]. Moreover, we found that the PLSR regression approach generated and achieved better prediction performance than PCR as shown in Figure 2.

![Figure 2 PCR and PLSR calibration used to predict FC of the cocoa bean.](image-url)

Judging from the prediction performance, we proposed the PLSR approach to develop further prediction models using corrected spectral data and also investigate the impact of 4 different spectra correction methods on the prediction accuracy of FC and MC determinations. The PLSR seeks to find the
best correlation between reference and infrared spectra data during the transformation onto LVs in the regression process.

For FC prediction, PLSR achieved a correlation coefficient of 0.82 and RPD index 1.81. Both PLSR and PCR require 5 LVs during prediction models development (calibration) to achieve that prediction accuracy and robustness. Meanwhile for MC (Figure 3), PLSR is also better than PCR with a correlation coefficient of 0.85 and RPD index of 1.97.

![Figure 3](image-url) PCR and PLSR calibration to determine MC of the cocoa bean.

We chose the PLSR method because we found that the PLSR regression approach provides better prediction results than the PCR method. This is also in agreement with other findings reported that partial least squares achieve more accurate and robust prediction compared to PCR [14,38,39]. Cocoa beans are biological objects that may interfere with the inner-quality parameters such as FC and MC during ripening, storage, and distribution phases. External factors such as temperature and relative humidity will also affect inner quality of cocoa beans or other agricultural products. Thus, it may interfere prediction model’s accuracy and robustness. Those effects need to be treated to achieve more robust and accurate prediction results. Therefore, it is strongly recommended to pre-process or enhance spectra data prior to prediction models calibration. We attempted to apply and investigate the impact of spectra correction prior to calibration to the prediction accuracy and robustness of inner quality attributes prediction on intact cocoa bean samples.

**The impact of spectra correction on the prediction performance**

To study the impact of spectra correction and enhancement method on the prediction performance, we applied and systematically compared 4 different spectra correction methods, namely: BSC, MN, SNV and OSC coupled with the PLSR approach to enhance and increase prediction performance in terms of its accuracy and robustness. Prediction models for both quality parameters (FC and MC) were established using 72 spectral data in the calibration dataset corrected by those 4 spectra correction methods. The best spectra correction algorithm was chosen based on their prediction performances.

The 1st spectra correction method to employ was the BSC method. It seeks to enhance the spectra by removing baseline and multiplicative effects due to physical error during spectra acquisition. Prediction results for FC and MC using BSC spectra correction were presented in Table 2. The achieved prediction accuracy in a form correlation coefficient (r) and RPD index was significantly improved after BSC correction compared using raw uncorrected spectra data. The correlation coefficient for FC prediction was increased from 0.82 to 0.90 and the RPD index was also increased from 1.81 to 2.77, while RMSE was decreased from 1.18 to 0.79. A similar finding was also observed for MC prediction, where prediction models established using corrected spectra had improved its correlation to 0.92 and the RPD index to 2.92 whilst the RMSE index was also decreased to 0.44.

Moreover, the SNV correction method generated slightly better than the BSC method. As presented in Table 2, correction coefficient and RPD achieved from SNV were for both quality attributes were
improved. BSC obtained its ideal spectrum from its mean spectra data for all samples while SNC obtained its ideal spectrum from scaling algorithm.

### Table 2 Prediction performance for FC and MC using different spectra correction methods.

<table>
<thead>
<tr>
<th>Quality parameters</th>
<th>Spectra correction</th>
<th>Statistical indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r</td>
<td>RMSE</td>
</tr>
<tr>
<td>Fat content</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSC</td>
<td>0.90</td>
<td>0.79</td>
</tr>
<tr>
<td>MN</td>
<td>0.87</td>
<td>0.82</td>
</tr>
<tr>
<td>SNV</td>
<td>0.89</td>
<td>0.80</td>
</tr>
<tr>
<td>OSC</td>
<td>0.92</td>
<td>0.71</td>
</tr>
<tr>
<td>Moisture content</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSC</td>
<td>0.92</td>
<td>0.44</td>
</tr>
<tr>
<td>MN</td>
<td>0.89</td>
<td>0.46</td>
</tr>
<tr>
<td>SNV</td>
<td>0.92</td>
<td>0.44</td>
</tr>
<tr>
<td>OSC</td>
<td>0.95</td>
<td>0.38</td>
</tr>
</tbody>
</table>

BSC: Baseline shift correction, MN: Mean normalization, SNV: Standard normal variate, OSC: Orthogonal signal correction, r: Correlation coefficient, RMSE: Root mean square error, RPD: Ratio prediction to deviation.

On the other hand, MN spectra correction was taken into account since we learned from the body of the literature that this spectra correction method was fit to be applied when dealing with bulk samples. In this study we also MN correction method to improve PLSR prediction accuracy for FC and MC. As presented in those Table 2 above, MN correction seems to be less accurate compared to the other 2 (BSC and SNV). Nonetheless, if we compared to raw un-corrected spectra, MN obviously can improve prediction accuracy and robustness for both quality parameters of intact cocoa beans. The best prediction performance for FC and MC was achieved when spectra data were corrected and enhanced using the OSC method as presented in Table 2. It achieved the maximum correlation coefficient and RPD index for FC prediction were 0.91 and 3.14 respectively, while for MC prediction the maximum r and RPD coefficients were 0.95 and 3.41 respectively. Therefore, based on these results, we may argue that spectra correction was significantly improved prediction accuracy and robustness for both inner quality parameters of intact cocoa bean samples. Spectra correction and enhancement are used to remove any irrelevant information such as noises and background information that occurred during spectra data acquisitions.

To evaluate the established prediction models, external validation was conducted to assure our findings. We employed the OSC spectra correction in tandem with PLSR to determine FC and MC of 38 cocoa bean samples separated in the validation dataset. Validation performance for FC and MC prediction is shown in Table 3. The scatter plot between predicted and measured FC and MC using is presented in Figure 4. Near-infrared technology based on spectroscopy combined with OSC spectra correction seems to be able to predict FC and MC with a correlation coefficient of 0.89 for FC, and 0.90 for MC. The RPD index for these both quality parameters were 2.72 and 2.85, while range to error ratio indexes were 6.84 and 7.38 for FC and MC respectively

### Table 3 Validation performance for FC and MC prediction using OSC spectra data.

<table>
<thead>
<tr>
<th>Quality parameters</th>
<th>Statistical indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r</td>
</tr>
<tr>
<td>Fat content</td>
<td>0.89</td>
</tr>
<tr>
<td>Moisture content</td>
<td>0.90</td>
</tr>
</tbody>
</table>

r: Correlation coefficient, RMSEP: Root mean square error for prediction, RPD: Ratio prediction to deviation, RER: Range to error ratio.
In brief, NIRS technology can be used as an alternative fast and robust method in predicting inner quality parameters of intact cocoa bean samples. In terms of spectra correction approaches, OSC is found to be the best correction method and more superior than BSC, SNV, and MN in predicting FC and MC of intact cocoa beans. Thus, spectra correction needs to be performed before prediction models development to obtain and achieve more accurate and robust prediction performances. The limitation related to this presented study is the cocoa cultivar used is only 1 cultivar. Therefore, the established prediction models are probably accurate for the same cocoa bean cultivar. Further study can be attempted to develop global prediction models using various cocoa cultivars with different fermentation stages as well. Hence, it may be applicable to be applied directly for rapid quality assessment of intact cocoa beans.

Conclusions

Based on obtained prediction results, near-infrared technology appears optimistic to be applied as a rapid, non-destructive, and simultaneous method to determine inner quality attributes on intact cocoa bean samples. The results showed that NIR technology combined with the PLSR approach can predict FC and MC of intact cocoa beans. The maximum correlation coefficient (r) and RPD indexes for FC were 0.92 and 3.14, while for MC, the r coefficient and RPD index were 0.95 and 3.41 respectively. Further, spectra correction can significantly improve prediction accuracy and robustness. Hence, we recommend performing spectra correction prior to prediction models development. The OSC was found to be the best correction method and can be able to predict FC and MC of independent cocoa bean samples with a correlation coefficient of 0.89 for FC, and 0.90 for MC.

Acknowledgements

We would like to sincerely acknowledge to LPPM Universitas Syiah Kuala for funding support through PHI research scheme 2021.

References


[27] M Zareef, MM Hassan, M Arslan, W Ahmad, S Ali, Q Ouyang, H Li, X Wu and Q Chen. Rapid prediction of caffeine in tea based on surface-enhanced Raman spectroscopy coupled multivariate


