Development of a low-cost modular VIS/NIR spectroscopy for predicting soluble solid content of banana

M F R Pahlawan*, R K Wati and R E Masithoh
Department of Agricultural and Biosystems Engineering, Faculty of Agricultural Technology, Universitas Gadjah Mada, Jl. Flora No. 1 Bulaksumur, Yogyakarta 55281, Indonesia

*Email: mfrezap@mail.ugm.ac.id

Abstract. Soluble solids content (SSC) is one of the most important parameters of banana associated with taste and consumer acceptance. NIR spectroscopy has been applied for nondestructive determination of SSC, but limited studies were conducted for a low-cost and modular VIS/NIR spectroscopy. This study was conducted to develop a calibration model to predict SSC in bananas using a modular type of VIS/NIR spectroscopy in the range of 350-1000 nm by varying distances of fiber optic probe to samples. Two varieties of bananas, namely Musa acuminata × balbisiana and Musa acuminata 'Lady Finger' were used. Partial least square regression (PLSR) was used to build a calibration model and to predict SSC of bananas. Normalization, baseline correction, standard normal variate (SNV), and multiple scattered (MSC) correlation were used for spectra preprocessing. The research showed that using 2 cm probe-sample distance and SNV method resulted in the best model with the coefficient correlation of calibration ($R^2_C$) and prediction ($R^2_P$) of 0.95 and 0.87, respectively. This study proved that probe-sample distances affected the efficiency of the model for VIS/NIR spectroscopy. This work concluded that the low-cost modular VIS/NIR spectroscopy is a promising tool for SSC measurement.

1. Introduction
Banana is the world's largest fruit produced and the second most favorite fruits in the world [1]. In 2019, global exports of bananas (exclude plantain) are expected to reach 20.2 million tones [2]. Not only can be consumed as a table-fruit, but banana is also popular as processed foods like chips, snacks, cakes, or beverages [3]. Due to high demand and popularity, it is important to produce high-quality bananas. One of the most important parameters of banana that is associated with taste and consumer acceptance is soluble solid content (SSC).

Commonly, to determine SSC in fruit is done by refractometer [4]. While assessing SSC using a refractometer considered as a destructive method, a non-destructive method for assessing SSC using visible and near-infrared (Vis-NIR) spectroscopy has been previously reported [5]–[7]. Vis-NIR spectroscopy is a type of near Infrared (NIR) spectroscopy that has a region range between 400-1000 nm.
Compared to other NIR spectroscopy, Vis-NIR spectroscopy has a cheaper price, but it has stronger penetration energy and lowers heating [6] since Vis-NIR has a lower wavelength compared to others. Vis-NIR spectroscopy can record the amount of energy that absorbed, reflected, or transmitted by the molecule of sugar in fruit at each wavelength [8] based on 3rd or 4th overtones of vibration [6]. Since sugars are the major content of soluble solids in fruit, so it could be used to determine SSC in fruit in the term of °Brix [4]. Therefore, this low-cost spectroscopy offers rapid analysis, limited preparation, and nondestructive measurement of SSC in fruits.

Partial least squares (PLS) regression is commonly used for NIR spectral chemometric analysis [9]. To improve the performance of the model, PLS regression was combined with several preprocessing techniques [10] which were used to remove undesired physical phenomena in the spectra [11]. The preprocessing technique that commonly used for spectral data is normalization, baseline correction, Savitsky-Golay 1st and 2nd derivatives, standard normal variate (SNV), and multiple scattering correction (MSC) [12].

Many studies were conducted using Vis-NIR spectroscopy to predict SSC in fruits, such as apple [6], melon [13], pineapple [5], [8], [14] and other fruits [15]–[17]. However, only a few studies reported the use of modular and low-cost Vis/NIR spectrometer like one developed in this study. The performance efficiency of this modular Vis/NIR spectrometer is still under study, especially when determining the best probe-sample distance which provides the best SSC prediction. Hence, this study was aimed to examine the potential of the modular and low-cost Vis-NIR spectroscopy in predicting SSC in bananas by varying probe-sample distances.

2. Materials and methods

2.1. Sample preparation

‘Kapas’ bananas (Musa acuminata × balbisiana) and ‘Mas’ bananas (Musa acuminata ‘Lady Finger’) were collected from bananas local market in Kaliurang, Sleman, Daerah Istimewa Yogyakarta, Indonesia. To obtain variations in SSC, the samples were stored at room temperature for two days. For ‘Kapas’ bananas, on the day the bananas were purchased (day 1), 30 green bananas were assessed, while on days 2 and 3, 35 bananas were assessed each day. For ‘Mas’ bananas, on day 1, 40 green bananas were assessed, while on day 2 and day 3, 30 bananas were assessed per day. There were 100 ‘Kapas’ and 100 ‘Mas’ bananas used which made in a total of 200 samples.

2.2. Spectra acquisition

The samples were analyzed using Vis-NIR Miniature Spectrometer (Flame-T-VIS-NIR Ocean Optics, 350-1000 nm) with tungsten halogen light (360-2400 nm, HL-2000-HP-FHSA Ocean Optics) and reflection probe (QR400-7-VIS-NIR Ocean Optics). A black chamber was built for image acquisition to minimize outside light, and the schematic of the spectral measurement set-up was shown in Figure 1. The distances between the probe and the samples were adjusted using three movable layers made of a triplex with a hole in its center that will not affect the spectra measurement. Three probe-sample distances evaluated in this study were 0, 1, and 2 cm.
Reflectance spectra were collected using OceanView 1.67 Software with an integration time of 150 ms, scan to average of 5, and boxcar width of 1. The white and black reference spectrum was measured before each sample measurement. Samples were analyzed at room temperature (around 28 °C).

2.3. Soluble solids content (SSC) measurement

Before SSC measurement, bananas were cut and made into juice using a blender (HR2106, Phillips, Indonesia). A certain amount of liquid juice was placed onto a digital refractometer (HI96801, Hanna Instrument, Koper, Slovenia) to obtain SSC in °Brix. SSC was measured in triplicate and was then averaged.

2.4. Preprocessing spectra and analysis

All spectra collected were then transformed into Ms. Excel. Spectra preprocessing and partial least square (PLS) regression were done using Unscrambler®X version (10.5.1, CAMO, Oslo, Norway). Preprocessing techniques that were applied were normalization, multiple scattering correction (MSC), standard normal variate (SNV), and baseline correction. Before the PLS regression, 200 samples were randomly divided into 134 samples for the calibration set and 66 samples for the prediction set. Full-cross validation was applied for the calibration set. Vis/NIR spectrometer recorded reflectance spectra at the wavelength range of 345-1033 nm. However, the spectra used for analysis were only in the range of 450-950 nm due to the presence of severe noise at wavelengths below 450 nm and above 950 nm. The total number of variables in the wavelength range of 450-950 nm was 2640 variables.

3. Results and discussion

3.1. SSC value of banana and spectra analysis

Table 1 showed the statistics summary of the SSC of the samples. SSC values for 200 banana samples ranged from 5.5 to 28.4 with a mean of 19.9 and a standard deviation of 5.479. This ranged of SSC was nearly similar to those reported by [18] at a range of 1.8-22.8 °Brix, but wider than [19] at a range of 6.5-24.1. In this paper, bananas were grouped into three according to SSC contents, i.e. 5.5-13.47, 13.47-20.83, and 20.83-28.2 °Brix.

| Quality Parameters | Minimum | Maximum | Mean | Standard Deviation |
|--------------------|---------|---------|------|--------------------|
| SSC (°Brix)        | 5.5     | 28.4    | 19.88| 5.48               |

The average reflectance spectra of banana at 0, 1, and 2 cm of probe-sample distances were displayed in Figure 2 which showed similar trends for all distances. Regardless the distances of probe to sample, there were increase in reflectance values at 500-550 nm and there was a reflectance peak around 550 nm
correlated to carotenoid pigment which also appeared in spectra of mango [20]. It was then followed by
decrease at 675 nm at which the highest absorbance of chlorophyll took place. Relatively high and stable
reflectances were noticeable above 720 nm similar to those reported by [8], [20], [21]. High reflectance at
760 nm was correlated with the third overtone of O-H stretching [6] or the second overtone of water [14].
The absorption values at 950-1000 nm were the second overtone of water spectra [6], [7], [20].

The absorption energy by chlorophyll might be correlated with the ripening process. During ripening,
chlorophyll in bananas decreased [22] and SSC increased [18]. Moreover, [23] reported that chlorophyll
content was inversely correlated to SSC content. At 675 nm, the highest absorbance values which
indicated the highest chlorophyll content were owned by bananas with the lowest SSC of 5.5-13.47 °Brix.
On the other hand, at 675 nm the low absorbance values belonged to bananas with the high SSC
of 13.47-28.2 °Brix.
Figure 2. Average spectra of banana captured by Vis/NIR spectrometer with probe-samples distances (a) 0 cm, (b) 1 cm, and (c) 2 cm.

Figure 2 showed that the distance between the probe and samples affects the reflectance spectra in which the farther distance between the probe and sample, the more noise appeared in the spectra around 450-500 nm and 800-950 nm. It also showed that by using 1 and 2 cm probe-sample distance resulted in lower reflectance intensity compared to 0 cm. Greater distances reduce the amount of light captured by the detector due to light scattering and absorption of the sample component [24]. [25] also reported the effect of sample and detector distance to the performance of the chemometrics model.

3.2. PLS regression model

Table 2 showed coefficient correlation of calibration and prediction ($R^2_C$ and $R^2_P$) as well as root mean square error of calibration and prediction (RMSEC and RMSEP) of PLS modeling for SSC value of banana. The best $R^2_C$ was obtained from SNV preprocessing spectra with a 2 cm distance between the probe and samples. The $R^2_C$ of 0.96 means that the model was applicable for screening and approximate calibration [26]. The $R^2_C$ obtained from this research was better than $R^2_C$ of 0.83 obtained by [7] to predict SSC of intact satsuma mandarin, $R^2_C$ of 0.94 obtained by [14] to predict SSC of pineapple, and better than those obtained by [18], [19] to predict SSC in bananas with wavelength range of 1000-2400 nm and 650-2500 nm. High calibration model was achieved using the Vis/NIR instrument since water molecules which are the largest component in fruits are weakly absorbed. Therefore, the instrument is suitable for predicting SSC without influenced by the water spectrum [27].

| Probe-sample distance | Transformation process   | Calibration (134 samples) | Prediction (66 samples) |
|-----------------------|--------------------------|---------------------------|-------------------------|
|                       |                          | $R^2_C$  | RMSEC  | $R^2_P$ | RMSEP  |
| 0 cm                  | Original                 | 0.86     | 2.04   | 0.83    | 2.23    |
|                       | Area Normalization       | 0.86     | 2.05   | 0.87    | 1.98    |
|                       | Unit Vector Normalization| 0.86     | 2.02   | 0.85    | 2.12    |
|                       | Mean Normalization       | 0.86     | 2.05   | 0.87    | 1.98    |
|                       | Max Normalization        | 0.80     | 2.44   | 0.79    | 2.54    |
|                       | Range Normalization      | 0.84     | 2.18   | 0.78    | 2.56    |
|                       | Baseline                 | 0.88     | 1.85   | 0.89    | 1.83    |
|                       | SNV                      | 0.82     | 2.29   | 0.76    | 2.69    |
|                       | MSC                      | 0.84     | 2.18   | 0.82    | 2.33    |
| 1 cm                  | Original                 | 0.89     | 1.84   | 0.85    | 2.16    |
Figure 3 showed the average coefficient correlation of the calibration and prediction model. The average $R^2_C$ with 0, 1, and 2 cm of the probe-samples gap were 0.847, 0.921, and 0.947, respectively, while the average $R^2_P$ with 0, 1, and 2 cm gap were 0.829, 0.858, and 0.890, respectively. Therefore, it can be concluded that adding the distance between the probe and samples improve the PLS model performance for SSC calibration and prediction model. However, the distance above 2 cm was not studied, therefore it was not discussed in this research.

![Figure 3](image)

**Figure 3.** The average coefficient correlation of calibration, $R^2_C$ (red) and prediction, $R^2_P$ (green).

Figure 4 shows the regression coefficient of PLSR using SNV spectra to determine SSC in bananas. The spectra were heavily noisy, but several peaks could be noticed at wavelengths at 550 nm which correlated with anthocyanin, at 670 nm correlated with chlorophyll, and 720 nm correlated with water. SSC values were inversely correlated with chlorophyll content in the peak around 670 nm [23], [28]. However, it was unclear which wavelengths contributed the most to the determination of SSC model. Nevertheless, the results showed that this low-cost modular Vis/NIR spectroscopy

| Method                  | Area Normalization | Unit Vector Normalization | Mean Normalization | Max Normalization | Range Normalization | Baseline | SNV       | MSC       |
|-------------------------|--------------------|--------------------------|-------------------|------------------|--------------------|----------|------------|-----------|
| Original                | 0.95               | 1.28                     | 0.89              | 1.78             |                    |          |            |           |
| 2 cm                    | 0.95               | 1.27                     | 0.90              | 1.73             |                    |          |            |           |
| Area Normalization      | 0.94               | 1.38                     | 0.85              | 2.16             |                    |          |            |           |
| Unit Vector Normalization| 0.93              | 1.46                     | 0.83              | 2.29             |                    |          |            |           |
| Mean Normalization      | 0.94               | 1.38                     | 0.85              | 2.16             |                    |          |            |           |
| Max Normalization       | 0.92               | 1.59                     | 0.89              | 1.86             |                    |          |            |           |
| Range Normalization     | 0.89               | 1.76                     | 0.87              | 1.95             |                    |          |            |           |
| Baseline                | 0.93               | 1.46                     | 0.88              | 1.94             |                    |          |            |           |
| SNV                     | 0.93               | 1.49                     | 0.85              | 2.09             |                    |          |            |           |
| MSC                     | 0.94               | 1.36                     | 0.87              | 2.01             |                    |          |            |           |
| SNV                     | 0.94               | 1.38                     | 0.85              | 2.16             |                    |          |            |           |
| MSC                     | 0.93               | 1.46                     | 0.88              | 2.09             |                    |          |            |           |
| MSC                     | 0.94               | 1.36                     | 0.87              | 2.01             |                    |          |            |           |
Figure 4. The regression coefficient of PLSR using SNV spectra with 2 cm of the probe-samples distance.

4. Conclusions

The soluble solid content (SSC) of bananas were measured with the Vis/NIR spectrometer at wavelength of 450-950 nm. Several preprocessing spectra were used and the 0, 1, and 2 cm distance of detector to sample were evaluated. The best calibration model for SSC resulted in $R^2$ of 0.96 and RMSEC of 1.15 obtained from SNV preprocessing spectra with 2 cm distance. With proper selection of chemometric and sample-detector distance, this instrument is a promising tool to measure SSC in fruit.

References

[1] Statista, “• Fruit: world production by type 2016 | Statistic,” Statista. 2018.
[2] FAO, “EST: Bananas,” Food and Agriculture Organization of the United Nations. 2018.
[3] D. Mohapatra, S. Mishra, C. B. Singh, and D. S. Jayas, “Post-harvest Processing of Banana: Opportunities and Challenges,” Food Bioprocess Technol., vol. 4, no. 3, pp. 327–339, 2011, doi: 10.1007/s11947-010-0377-6.
[4] D. Garner, C. H. Crisosto, P. Wiley, and G. M. Crisosto, “Measurement of Soluble Solids Content,” Qual. Eval. Methodol., 2005.
[5] K. S. Chia, H. Abdul Rahim, and R. Abdul Rahim, “Prediction of soluble solids content of pineapple via non-invasive low cost visible and shortwave near infrared spectroscopy and artificial neural network,” Biosyst. Eng., vol. 113, no. 2, pp. 158–165, 2012, doi: 10.1016/j.biosystemseng.2012.07.003.
[6] Z. Guo, W. Huang, Y. Peng, Q. Chen, Q. Ouyang, and J. Zhao, “Color compensation and comparison of shortwave near infrared and long wave near infrared spectroscopy for determination of soluble solids content of ‘Fuji’ apple,” Postharvest Biol. Technol., vol. 115, pp. 81–90, 2016, doi: 10.1016/j.postharvbio.2015.12.027.
[7] R. E. Masithoh, R. Haff, and S. Kawano, “Determination of soluble solids content and titratable acidity of intact fruit and juice of satsuma Mandarin using a hand-held near infrared instrument in transmittance mode,” J. Near Infrared Spectrosc., vol. 24, no. 1, pp. 83–88, 2016, doi: 10.1255/jnirs.1196.
[8] K. S. Chia, H. Abdul Rahim, and R. Abdul Rahim, “Evaluation of common pre-processing approaches for visible (VIS) and shortwave near infrared (SWNIR) spectroscopy in soluble solids content (SSC) assessment,” Biosyst. Eng., vol. 115, no. 1, pp. 82–88, 2013, doi: 10.1016/j.biosystemseng.2013.02.008.
[9] S. Sharma, M. Goodarzi, H. Ramon, and W. Saeyes, “Performance evaluation of preprocessing techniques utilizing expert information in multivariate calibration,” Talanta, vol. 121, pp. 105–112,
[10] A. C. Dotto, R. S. D. Dalmolin, A. ten Caten, and S. Grunwald, “A systematic study on the application of scatter-corrective and spectral-derivative preprocessing for multivariate prediction of soil organic carbon by Vis-NIR spectra,” Geoderma, vol. 314, no. May 2017, pp. 262–274, 2018, doi: 10.1016/j.geoderma.2017.11.006.

[11] Á. Rinnan, F. van den Berg, and S. B. Engelsen, “Review of the most common pre-processing techniques for near-infrared spectra,” TrAC - Trends in Analytical Chemistry. 2009, doi: 10.1016/j.trac.2009.07.007.

[12] S. Verboven, M. Hubert, and P. Goos, “Robust preprocessing and model selection for spectral data,” J. Chemom., vol. 26, no. 6, pp. 282–289, 2012, doi: 10.1002/cem.2446.

[13] R. Hu, L. Zhang, Z. Yu, Z. Zhai, and R. Zhang, “Optimization of soluble solids content prediction models in ‘Hami’ melons by means of Vis-NIR spectroscopy and chemometric tools,” Infrared Phys. Technol., vol. 102, no. May, p. 102999, 2019, doi: 10.1016/j.infrared.2019.102999.

[14] D. Suhandy, “Nondestructive measurement of soluble solids content in pineapple fruit using short wavelength near infrared (SW-NIR) spectroscopy,” Int. J. Appl. Eng. Res., vol. 4, no. 1, pp. 107–114, 2009.

[15] A. Wang, D. Hu, and L. Xie, “Comparison of detection modes in terms of the necessity of visible region (VIS) and influence of the peel on soluble solids content (SSC) determination of navel orange using VIS-SW-NIR spectroscopy,” J. Food Eng., vol. 126, pp. 126–132, 2014, doi: 10.1016/j.jfoodeng.2013.11.011.

[16] Y. Huang, R. Lu, and K. Chen, “Assessment of tomato soluble solids content and pH by spatially-resolved and conventional Vis/NIR spectroscopy,” J. Food Eng., vol. 236, no. May, pp. 19–28, 2018, doi: 10.1016/j.jfoodeng.2018.05.008.

[17] S. Travers, M. G. Bertelsen, K. K. Petersen, and S. V. Kucheryavskiy, “Predicting pear (cv. Clara Frijs) dry matter and soluble solids content with near infrared spectroscopy,” LWT - Food Sci. Technol., vol. 59, no. 2P1, pp. 1107–1113, 2014, doi: 10.1016/j.lwt.2014.04.048.

[18] C. Y. Liew and C. Y. Lau, “Determination of quality parameters in Cavendish banana during ripening by NIR spectroscopy,” Int. Food Res. J., vol. 19, no. 2, pp. 751–758, 2012.

[19] M. M. Ali, R. B. Janius, N. M. Nawi, and N. Hashim, “Prediction of total soluble solids and pH in banana using near infrared spectroscopy,” J. Eng. Sci. Technol., vol. 13, no. 1, pp. 254–264, 2018.

[20] V. Cortés, C. Ortiz, N. Aleixos, J. Blasco, S. Cubero, and P. Talens, “A new internal quality index for mango and its prediction by external visible and near-infrared reflection spectroscopy,” Postharvest Biol. Technol., vol. 118, pp. 148–158, 2016, doi: 10.1016/j.postharvbio.2016.04.011.

[21] J. Fernández-Novalles, T. Garde-Cerdán, J. Tardáguila, G. Gutiérrez-Gamboa, E. P. Pérez-Álvarez, and M. P. Diago, “Assessment of amino acids and total soluble solids in intact grape berries using contactless Vis and NIR spectroscopy during ripening,” Talanta, vol. 199, no. February, pp. 244–253, 2019, doi: 10.1016/j.talanta.2019.02.037.

[22] X. tang Yang, Z. qi Zhang, D. Joyce, X. mei Huang, L. ying Xu, and X. qun Pang, “Characterization of chlorophyll degradation in banana and plantain during ripening at high temperature,” Food Chem., vol. 114, no. 2, pp. 383–390, 2009, doi: 10.1016/j.foodchem.2008.06.006.

[23] J. Sugiyama, “Visualization of sugar content in the flesh of a melon by near-infrared imaging,” J. Agric. Food Chem., vol. 47, no. 7, pp. 2715–2718, 1999, doi: 10.1021/jf981079i.

[24] M. Rupawala, H. Dehghani, S. J. E. Lucas, P. Tino, and D. Cruse, “Shining a light on awareness: A review of functional near-infrared spectroscopy for prolonged disorders of consciousness,” Front. Neurol., vol. 9, no. MAY, pp. 1–17, 2018, doi: 10.3389/fneur.2018.00350.

[25] A. Rady, J. Fischer, S. Reeves, B. Logan, and N. J. Watson, “The Effect of Light Intensity , Sensor
Height, and Spectral Pre-processing Methods when using NIR Spectroscopy to Identify Different Allergen-Containing Powdered Foods,” *Sensors*, vol. 20, no. 230, 2020, doi: 10.3390/s20010230.

[26] P. C. Williams, “Implementation of near-infrared technology,” in *Near-Infrared Technology in the Agricultural and Food Industries*, 2001.

[27] A. Guelpa, F. Marini, A. du Plessis, R. Slabbert, and M. Manley, “Verification of authenticity and fraud detection in South African honey using NIR spectroscopy,” *Food Control*, vol. 73, pp. 1388–1396, 2017, doi: 10.1016/j.foodcont.2016.11.002.

[28] P. Carlini, R. Massantini, F. Mencarelli, P. Carlini, R. Massantini, and F. Mencarelli, “Vis-NIR Measurement of Soluble Solids in Cherry and Apricot by PLS Regression and Wavelength Selection Vis-NIR Measurement of Soluble Solids in Cherry and Apricot by PLS Regression and Wavelength Selection,” *J. Agric. Food Chem.*, vol. 48, no. 11, pp. 5236–5242, 2000, doi: 10.1021/jf000408f.