Non-destructive determination of SSC and pH of banana using a modular Vis/NIR spectroscopy: comparison of Partial Least Square (PLS) and Principle Component Regression (PCR)

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

*E-mail: evi@ugm.ac.id

Abstract. Determination of soluble solid content (SSC) and pH of banana was investigated using a modular Vis/NIR spectroscopy in reflectance mode. Vis/NIR spectroscopy has been applied for non-destructive SSC or pH measurement, but limited studies were conducted for a modular VIS/NIR spectroscopy. This study was conducted to develop a calibration model to predict SSC and pH in bananas using a modular type of VIS/NIR spectroscopy at wavelength of 350-1000 nm. Two chemometrics analysis namely partial least square (PLS) and principle component regression (PCR) were used to develop calibration models and to predict SSC and pH of bananas. Normalization, baseline correction, standard normal variate (SNV), and multiplicative scatter correction (MSC) pre-processing were used for spectra transformation. Research showed that PLS regression produced better models compared to PCR in determining SSC and pH contents. PLS regression resulted in $R^2_C$ of 0.95, RMSEC of 1.27, $R^2_P$ of 0.85, RMSEP of 1.98, and bias of -0.09 for SSC and $R^2_C$ of 0.96, RMSEC of 0.05, $R^2_P$ of 0.82, RMSEP of 0.11, and bias of 0.11 for pH. PCR resulted in $R^2_C$ of 0.78, RMSEC of 2.63, $R^2_P$ of 0.76, RMSEP of 2.71, and bias of -0.12 for SSC and $R^2_C$ of 0.71, RMSEC of 0.14, $R^2_P$ of 0.62, RMSEP of 0.16, and bias of -0.02 for pH. This modular Vis/NIR instrument combined with proper pre-processing method and chemometric analysis is promising to be used for determination of SSC and pH of fruits.

1. Introduction
Banana is widely consumed in tropical and non-tropical countries and rich with carbohydrate, vitamin, mineral, flavonoids and phenolic compounds [1]. Banana is preferred due to its taste which is the combination of sour and sweet. Sourness and sweetness of banana are determined from its acidity (titratable acidity and pH) [2]. Physiological changes during fruit ripening involve conversion of starch to sugars, formation of flavor, and increase of pH [3]. Banana contains sugars, acids, vitamin C, amino acids, and pectin which are soluble in water forming soluble solids content (SSC) [4]. SSC of banana increases during maturation or ripening and is used as maturity index. Therefore, pH and SSC are important parameters used to determine quality of banana.

Usually, pH or acidity is determined using pH meter, acid meter, or titration method, which involve subjective measurement and intense-work analysis. For SSC, measurement is usually done using refractometer by which although the method is easy but requires sample preparation. Those common pH and SSC methods are impractical for routine analysis and vast samples. With the development of...
spectroscopy technology, fruity qualities including SSC and pH are determined using infrared spectroscopy. Determination of SSC and pH using Vis/NIR/MIR or hyperspectral imaging (HIS) spectroscopy were reported for satsuma mandarin [5], orange and grapefruit [6], pineapple [7], passion fruit [8], tomato [9], or limes [10].

Those methods provide rapid and accurate results but require high investment and operational costs, especially for Raman HSI. Spectrometer working in the visible and short-wave near infra-red (Vis/NIR) at 350 - 1000 nm is also available. Since water is the largest component in fruits, water spectrum will dominate the spectra which affects in developing model for other quality parameters in fruit. Weak absorption of water molecules in Vis/NIR wavelength gives advantage for detection of low concentration of components [11]. Currently, price of spectrometers has declined especially for modular type working in the range of 350-1000 nm, thus it can be potential for determining quality parameters of fruits implemented for small farmers. Therefore, this research aimed at evaluating the potential of a low-cost Vis/NIR spectroscopy for pH and SSC determination. Two MVA methods i.e. PCR and PLS regression were used to obtain the most accurate calibration model.

2. Materials and Methods

2.1. Sample preparation
Two types of banana, i.e. ‘Kapas’ (Musa acuminata Balbisiana) and ‘Mas’ (Musa acuminata Lady Finger) were collected from local market in Kalirang, Indonesia. Each type of banana was 100 making the total of 200 samples consisted of green and yellow color of bananas (Figure 1). After purchasing, bananas were cleaned and stored at room temperature (28-29°C) without any specific treatments. After purchasing, 30 and 40 of Kapas and Mas bananas were scanned for spectra acquisition and followed by SSC and pH measurement. Other remaining samples were analyzed on the day-2 and day-3.

Figure 1. Green and yellow bananas used in this research

2.2. Sample spectra acquisition
Spectra of samples were acquired 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-VIS-NIR Ocean Optics). The spectral measurement set-up was shown in Figure 2. The distances between the probe and sample were set at 2 cm. Spectra in reflectance mode 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. Each sample was scanned three times and averaged. Spectra, pH, and SSC measurement were done at room temperature.

2.3. SSC and pH analysis
Bananas were made into juice using a blender prior to SSC measurement which was measured using a digital refractometer (HI96801, Hanna Instrument, Koper, Slovenia) to obtain SSC in °Brix. SSC was measured in triplicate and was then averaged. After SSC measurement, pH was determined using a digital pH Meter (KL- 009(I), China) which was measured in triplicate and averaged.
2.4. Spectra analysis

Spectra of bananas in reflectance mode were converted into Excel file (Microsoft® Excel®) which were then imported to the Unscrambler software v10.5.1 (CAMO Software AS, Norway) for pre-processing and analysis. Several pre-processing methods were applied, such as Savitzky-Golay smoothing, unit normalization, baseline correction, multiplicative scatter correction (MSC), and standard normal variate correction (SNV). Original and pre-processing spectra were used to develop calibration model using PCR and PLS regression to predict SSC and pH of bananas. Of 200 spectra of bananas, 120 and 80 spectra were used as calibration and prediction data sets.

![Figure 2. The spectral measurement set-up.](image)

PCR combines multiple linear regression (MLR) and principal component analysis (PCA) for developing quantitative models. By using decomposed principal components from PCA method to calculate regression equation, more robust model for predicting concentration is achieved [12]. PLS regression also combines MLR and PCA which constructs new latent variables known as principal components in developing linear regression. PLS is similar to PCR method, but during the decomposition process it uses the concentration information. By doing this, spectra having high concentrations are weighted more heavily than those with low concentrations [12].

Both PCR and PLS regression models for predicting SSC and pH were built using original and pre-processed spectra. The best model was assessed based on the highest coefficient of determination of calibration and validation ($R_c^2$ and $R_v^2$) and the lowest root mean square error of calibration and cross validation (RMSEC and RMSECV). Models which have small differences between calibration and validation are considered to be robust. The calibration models were then applied to prediction data sets which were evaluated based on coefficient of determination of prediction ($R_p^2$), root mean square error of prediction (RMSEP), and bias.

3. Results and Discussion

3.1. SSC, pH, and spectra exploration

SSC and pH profiles of banana is shown in Table 1 in which minimum, maximum, mean, and standard deviation values are described. SSC values of all banana samples ranged from 5.5 to 28.4, with a mean of 19.9, and a standard deviation of 5.48 °Brix. While pH ranges from 4.6-5.6, a mean of 4.92, and a standard deviation of 0.26. SSC used in this samples were higher than those reported by several studies [13, 14], while pH values were lower than those reported by Jaiswal et al [15].

The reflectance spectra of several banana based on SSC and pH content are displayed in Figure 3 which show similar trends. Of 350-1000 nm wavelength supported by this fibre optic spectroscopy, only wavelength at 450-950 nm were used for analysis since the spectra below 450 nm and above 950 nm were very noisy and cluttered. As shown in Figure 3(a), the higher SSC values the higher reflectance values, as opposed to Figure 3(b) in which the lower pH the higher reflectance values. In general, there are increase in reflectance values at 500-550 nm which are then followed by decrease at 675 nm.
However, the decrease is distinct for spectra of banana with low SSC and pH. Above 700 nm, relatively flat reflectance is noticeable. Those spectra trends at 500-850 nm is similar to those reported by [16]. A reflectance at 550 nm correlated to pigments and chlorophyll [17], while at 760 nm correlated with the third overtone of O-H stretching [18] and 800-950 nm related to water [19].

### Table 1. Statistics of minimum, maximum, mean, and standard deviation of SSC and pH of bananas.

| Parameters | Range   | Mean  | Standard Deviation |
|------------|---------|-------|--------------------|
| SSC (°Brix)| 5.5 – 28.4 | 19.88 | 5.48               |
| pH         | 4.6 – 5.6  | 4.92  | 0.26               |

**Figure 3.** Original spectra of banana at 500-850 nm according to (a) SSC, and (b) pH level

3.2. **PCA analysis**

Since spectroscopy data contain large variables, PCA can be used to reduce dimensions which closely related to samples [20]. By using PC-1 and PC-2, bananas can be classified based on SSC classes, as shown in Figure 4(a). Bananas having high SSC are located at negative axis of PC-1, while bananas having low SSC are located at positive axis. Figure 4(b) shows score of PC-1 and PC-2 which are able to differentiate bananas based on pH although some samples grouped together. Bananas with high pH (5.3-5.6) and low pH (4.6-4.99) are located at negative and positive axis of PC-2, respectively.

Banana during ripening experiences decrease in chlorophyll [21] and increase in SSC [22]; thus the relationship between those two variables are inversely as also found in melons [23]. Figure 4 is PC-1 to PC-4 loadings which show several peaks at wavelength of 450-950 nm. A peak at 670 nm can be seen from Figure 4 in which PC-1 separated low SSC from high SSC. PC-1 loading value shown in Figure 5 showed high reflectance at 670 nm which means low absorption of chlorophyll belongs to bananas with low SSC. Therefore, the low absorption band of chlorophyll at 670 nm can be indirectly correlated to SSC content.

Organic acids, such as citric, malic, and tartaric acids, are source of nutrition and affect sourness and sweetness of taste in fruits. Since organic acids have different capabilities in separating hydrogen ion, the pH, which is free hydrogen ions calculated by -log [H⁺], is often used to express the acidity in fruits. Pigments significantly influence the absorbance variation in visible region, but the relationship between pH and pigments can be used to develop calibration model of the pH content [25]. Values of pH both of peel and pulp banana were higher for unripe compared to ripe bananas due to high production of malic acid during ripening [22]. Therefore, during ripening the trend of pH is conversely with SSC. A
reflectance at 550 nm (Figure 5) refers to absorbance of carotenoid in banana. In this study, bananas with low pH are green (unripen) while those with high pH are yellow (ripen). This result corresponds to study reported by [26], but [27] found the opposite results. Although carotenoid was not measured in this study, yellow (ripen) banana have higher carotenoid compared to green (unripen) banana [28]. Figure 4(b) shows PCA scores of PC-1 and PC-2 of banana based on pH content. It can be observed that by using PC-2, low pH bananas are located on the positive axis while bananas with high pH are located on negative axis. Thus, the band at 550 nm of carotenoid can be used to differentiate banana based on the pH values; in this study are pH of 4.6-5.29 and 5.3-5.6.

![Figure 4](image1.png)

**Figure 4.** PCA of bananas based on (a) SSC and (b) pH

![Figure 5](image2.png)

**Figure 5.** PCA loadings which show several prominent reflectance bands

3.3. **PCR and PLS regression**

The calibration and validation results of PCR and PLS regression are shown in Table 2. By applying pre-processing methods to original spectra, does not significantly increase the PLS performances for SSC; applying SNV and MSC resulted in lower performances. However, by applying pre-processing methods increase the PLS performance for pH, except for MCS method. The best PLS models for SSC
and pH having R² of 0.94 and 0.96 are obtained by applying baseline and unit vector normalization, respectively. The PLS model for SSC banana is higher than the ones for SSC in pomegranate [20] and mango [29]. For predicting pH banana, the PLS models is higher than studies reported for tomatoes [9] and grapes [30].

Unlike PLS regression, studies conducted for qualitative and quantitative analysis using PCR are limited. In this study, PCR methods resulted in lower performances for SSC and pH determination compared to PLS regressions. The best PCR models are obtained by applying UV normalization and MSC per-processing method to original spectra with R² of 0.78 and 0.71, respectively. Although the R² from PCR model were not high (less than 0.8), they are usable for screening and approximate applications [31]. [32] and [33] reported high PCR result for corn syrup and margarine but both research employed mid-infrared which have higher wavelength providing more important variables for determining SSC and pH.

**Table 2.** PLS and PCR results of calibration and prediction of SSC and pH of banana

| Pre-processing method | PLS | PCR |
|-----------------------|-----|-----|
|                      | Calibration | Validation | Prediction | Calibration | Validation | Prediction | Bias |
|                      | R² | RMSEC | RMSECV | R² | RMSEP | Bias | R² | RMSEC | RMSECV | R² | RMSEP | Bias |
| SSC                  |    |       |       |    |       |      |    |       |       |    |       |      |
| Original             | 0.94 | 1.32 | 0.85 | 2.18 | 0.85 | 2.05 | -0.07 | 0.77 | 2.68 | 0.75 | 2.84 | 0.72 | 2.76 | -0.25 |
| UV Norm.             | 0.94 | 1.32 | 0.85 | 2.16 | 0.84 | 2.06 | -0.08 | 0.78 | 2.63 | 0.76 | 2.79 | 0.76 | 2.71 | -0.12 |
| Baseline             | 0.95 | 1.27 | 0.86 | 2.15 | 0.85 | 1.98 | -0.09 | 0.78 | 2.65 | 0.75 | 2.85 | 0.72 | 2.74 | 0.09 |
| SNV                  | 0.92 | 1.63 | 0.82 | 2.43 | 0.78 | 2.44 | 0.03 | 0.74 | 2.87 | 0.72 | 2.98 | 0.68 | 2.95 | 0.09 |
| MSC                  | 0.92 | 1.54 | 0.82 | 2.39 | 0.79 | 2.33 | -0.05 | 0.76 | 2.78 | 0.74 | 2.91 | 0.69 | 2.9  | 0.09 |
| pH                   |    |       |       |    |       |      |    |       |       |    |       |      |
| Original             | 0.93 | 0.07 | 0.82 | 0.11 | 0.81 | 0.11 | -0.01 | 0.68 | 0.15 | 0.62 | 0.16 | 0.57 | 0.17 | -0.04 |
| UV Norm.             | 0.96 | 0.05 | 0.84 | 0.11 | 0.82 | 0.11 | -0.01 | 0.58 | 0.17 | 0.54 | 0.18 | 0.50 | 0.18 | -0.03 |
| Baseline             | 0.94 | 0.06 | 0.81 | 0.12 | 0.81 | 0.11 | 0.00  | 0.67 | 0.15 | 0.62 | 0.16 | 0.59 | 0.17 | -0.03 |
| SNV                  | 0.96 | 0.06 | 0.83 | 0.11 | 0.83 | 0.11 | -0.01 | 0.58 | 0.17 | 0.55 | 0.18 | 0.52 | 0.18 | -0.02 |
| MSC                  | 0.92 | 0.07 | 0.82 | 0.11 | 0.84 | 0.10 | -0.01 | 0.71 | 0.14 | 0.67 | 0.15 | 0.62 | 0.16 | -0.02 |

Note: UV Norm: unit vector normalization; SNV: standard normal variate correction; MSC: multiplicative scatter correction; R²: coefficient of determination; RMSEC: root mean square error of calibration; RMSECV: root mean square error of cross validation; RMSEP: root mean square error of prediction.

New datasets, which were not included in developing calibration model, were used to predict SSC and pH of bananas. PLS regression resulted in R² of 0.79-0.85 and 0.81-0.84 for SSC and pH, respectively by using several pre-processing methods. PCR resulted in R² of 0.68-0.76 and 0.50-0.62 for SSC and pH, respectively. In addition, relatively low bias obtained from PCR and PLS regression for both SSC and pH. Those results show that PCR is more consistent in predicting SSC and pH compared to PLS regression based on the difference between calibration and prediction results.

**4. Conclusions**

SSC and pH of bananas were measured using a Vis/NIR spectrometer in reflectance mode. PLS regression models resulted in better performance based on R² and RMSE for calibration, validation, and prediction, compared to PCR models. The Vis/NIR spectrometer using PLS regression is a potential method for measurements of SSC and pH of bananas.
Acknowledgement
Great appreciation delivered to Universitas Gadjah Mada which provides financial support under Recognisi Tugas Akhir Program No. 2488/UN1.P.III/DIT-LIT/PT/2020.

References
[1] Kookal S K and Thimmaiah A 2018 Am. J. Plant Sci. 9
[2] Etienne A et al 2013 Sci. Hortic. (Amsterdam). 162
[3] Maftoonazad N and Ramaswamy H 2019 Polym. Test. 75
[4] Tourky M N et al 2014 Am. J. Plant Physiol. 9
[5] Masithoh R et al 2016 J. Near Infrared Spectrosc. 24
[6] Neama K et al 2017 J. Food Eng. 193
[7] Chia K S et al 2012 Biosyst. Eng. 113
[8] Maniwara P et al 2019 Sci. Hortic. (Amsterdam). 257
[9] Huang Y et al 2018 J. Food Eng. 236
[10] Teerachaichayut S and Ho H T 2017 Postharvest Biol. Technol. 133
[11] Manley et al 2007 J. Near Infrared Spectrosc. 188
[12] Chen J and Wang X Z 2001 J. Chem. Inf. Comput. Sci. 10
[13] Liew C Y and Lau C Y 2012 Int. Food Res. J. 19
[14] Subedi P P and Walsh K B 2011 Postharvest Biol. Technol. 62
[15] Jaiswal P et al 2012 Sci. Hortic. (Amsterdam). 135
[16] Munera S et al 2017 J. Food Eng. 214
[17] Rambo M K D et al 2013 Anal. Chim. Acta 775
[18] Guo Z et al 2016 Postharvest Biol. Technol. 115
[19] Rajkumar P et al 2012 J. Food Eng. 108
[20] Khodabakhshian R et al 2017 Int. J. Food Prop. 20
[21] Yang X T et al 2009 Food Chem. 114
[22] Newilah G N et al 2009 Fresh Prod. 3
[23] Sugiyama J and Tsuta M 2010 Visualization of sugar distribution of melons by hyperspectral technique (Hyperspectral Imaging for Food Quality Analysis and Control) ed D W Sun (San Diego: Academic Press) pp 349–368
[24] Qin J and Lu R 2008 Postharvest Biol. Technol. 49
[25] Cen H et al 2006 J. Agric. Food Chem. 54
[26] Alkarkhi A F M et al 2011 Food Chem. 129
[27] Adi D D et al 2019 Sci. African 6
[28] Vu H T et al 2019 Sci. Hortic. (Amsterdam). 253
[29] Cortés V et al 2016 Postharvest Biol. Technol. 118
[30] Cozzolino D et al 2004 J. Near Infrared Spectrosc. 12
[31] Williams P C 2001 Implementation of near-infrared technology (Near-infrared technology in the agricultural and food industries) eds P C Williams and K Noris (Minnesota: The American Association of Cereal Chemist, Inc.)
[32] Dhaulaniya A S et al 2020 LWT 131
[33] Rácz A et al 2018 Anal. Methods 10