Prediction of surface color of ‘crystal’ guava using UV-Vis-NIR spectroscopy and multivariate analysis

Kusumiyati*, W Sutari, Farida, S Mubarok, J S Hamdani

Department of Agronomy, Agricultural Faculty, Universitas Padjadjaran, Jatinangor, Indonesia

*Email: kusumiyati@unpad.ac.id

Abstract. Ultraviolet, visible, and near infrared (UV-Vis-NIR) spectroscopy technique has been increasingly applied in sorting of agricultural products, especially fruits. This study aimed to evaluate the use of UV-Vis-NIR spectroscopy and multivariate analysis to predict the surface color (L*, a*, and b*) of ‘crystal’ guava. Wavelengths covering the UV-Vis (300-699 nm) and UV-Vis-NIR (300-1065 nm) area were investigated to obtain the best prediction accuracy. A total of 120 samples of guava fruits were harvested at the same maturity level, the samples were divided into 3 groups, which consisted of 40 samples per group. The first group was stored for 0 days. The second and third groups were stored for 4 and 8 days, respectively. Spectra data acquisition was performed at wavelengths of 300-1065 nm and interval of 3 nm. Multiplicative scatter correction (MSC) and standard normal variate (SNV) spectra correction methods were applied to improve prediction accuracy. Partial least squares regression (PLSR) was used as calibration method. Validation was done by the k fold cross validation. The best prediction of L* was obtained at UV-Vis-NIR wavelength with SNV correction method. The UV-Vis wavelength and SNV acquired the best prediction of a*. Original spectra and UV-Vis-NIR wavelengths resulted the best prediction of b*.

1. Introduction
Determination of surface colour is usually done by visual assessment. This method does not have good accuracy and sometimes inaccurate. Researchers evaluate the surface colour using chromameter, but not many of them use the ultraviolet, visible, and near infrared (UV-Vis-NIR) spectroscopy to measure the surface colour. Spectroscopy is a technique that utilize interaction between light at certain wavelength and organic materials. This technique could predict the surface colour of fruits fast and accurate. Ultraviolet radiation contains wavelength ranged from 10-400 nm, while visible from 400-700 nm, whereas near infrared radiation from 700-2500 nm. Each wavelength radiation gives different response to the organic materials. The spectra absorption of samples and measured values of desired quality attributes is used to obtain calibration model through regression analysis.

UV-Vis and UV-Vis-NIR wavelengths have been performed to predict the quality attributes of agricultural products. UV-Vis spectroscopy was used to detect olive oil colour, chlorophyll and lutein affect the color of olive oil [1], [2]. UV-Vis-NIR spectroscopy technique was used to predict sugar content and water content [3], [4]. Wavelengths of 312-1050 nm were used to predict L*, a*, and b* of sapodilla surface color and resulted high prediction accuracy [5].

The accuracy of calibration model is evaluated by coefficient of determination (R²), ratio of performance to deviation (RPD), root mean squares error of calibration (RMSEC), root mean squares error of cross-validation (RMSECV), and principal component (PC). Partial least squares regression (PLSR) is a method to develop the calibration model. PLSR is able to analyse predictor variable (X) and response variable (Y) simultaneously. This method is widely used to build prediction model of fruit quality attributes, such as guava [6], mango [7] and peaches [8]. This technique was able to differentiate fruit species from other species [9]. Color is a fruit quality attribute that can be used as
indicator of maturity level. ‘Crystal’ guava is climacteric fruit that changes the color during ripening phase due to high production of ethylene. The color of the fruit changes from green to the yellowish.

Spectroscopy based prediction of fruit color can be done at various wavelengths, such as UV-Vis and UV-Vis-NIR. Spectra data obtained from the spectrometer still contain noise. This can reduce the value of prediction accuracy. However, noise can be reduced using the spectra correction method, such as multiplicative scatter correction (MSC), and standard normal variate (SNV). The purpose of this study was to examine the use of UV-Vis-NIR spectroscopy and multivariate analysis to predict the surface color of ‘crystal’ guava fruit.

2. Materials and Method

2.1. Samples

120 samples of fresh ‘crystal’ guava fruits were harvested from Sumedang, West Java, Indonesia. All samples were collected at same maturity level. The samples were split into 3 groups, each group consisted of 40 samples. 0 day storage duration was run for first group, 4 days storage duration for second group, and 8 days storage duration for third group.

2.2 Spectra Measurement

Handheld UV-Vis-NIR instrument (NirVana AG410) was performed to obtain the absorbance spectra. The measurement of spectra was done from 300-1065 nm. Six spectra measurements were made to each sample. The spectra were split for UV-Vis (300-699 nm) and UV-Vis-NIR area (300-1065 nm).

2.3. Spectra Correction

Once the spectra were acquired, spectra correction is necessary to improve the prediction accuracy. SNV and MSC spectra corrections were operated individually to the original spectra.

2.4. Surface Color Measurement

Chromameter was used to obtain to reference data of surface color. The measured surface color included the values of L*, a*, and b*. L* indicates the level of lightness, while a* denotes the color gradation between green-red, whereas b* defines the color gradation between blue-yellow.

2.5. Data Analysis

Multivariate analysis is useful to process the data with more than one variable. In this research, the data contain a single response variable (measured surface color) and large numbers of predictor variables (spectra data). The purpose of using multivariate analysis is to find correlation between predictor variable and response variable. PLSR was utilized to perform the regression analysis.

3. Results and Discussion

3.1. Spectra Data Analysis

Absorption bands on each commodity showed different response due to the difference physical and chemical properties (molecular structure) found in the organic materials being irradiated. Peaks and valleys indicate the wavelength that plays important role in detecting the chemical content of guava fruit. Fig. 1 presents the UV-Vis spectra from 300-699 nm.

Chlorophyll a and chlorophyll b were detected at 640 nm and 480 nm, respectively. Chlorophyll content is responsible for the green colour of guava fruit. The green of surface color turns yellowish during maturity. Discoloration is caused by the degradation of chlorophyll content based on the level of fruit maturity. The decreased chlorophyll is replaced by the increase of rubberoid content, causing the color changes from green to yellowish [10].
The use of a wider wavelength range is generally useful to detect more nutrient contents. Wavelength is inversely proportional to frequency. The higher the wavelength, the smaller the frequency of energy released. Fig. 2 shows the UV-Vis-NIR absorbance spectra of ‘crystal’ guava. Chlorophyll a was detected at 640 nm, while chlorophyll b was at 480 nm, whereas water content absorption was at 930 nm. The sample is partially reflected, absorbed, and transmitted light. However, the sample of intact fruits is not translucent, hence the spectrum obtained is reflectance light, then transformed into absorbance data.

The original spectra were corrected using MSC and SNV methods. Basically, the purpose of spectra correction is to remove the irrelevant information of original spectra. MSC method processes each spectrum based on the reference spectrum, which is the mean of all acquired spectra. SNV
method gives similar result to the MSC. In addition, SNV does not standarize spectrum based on reference spectrum, but the spectrum itself.

3.2. Surface Color Prediction

Surface color assessment is displayed in $L^*$, $a^*$, and $b^*$ color identification. $L^*$ is defined as value of lightness ranged from 0 to 100. Higher $L^*$ value describe the lighter fruit color. $a^*$ indicates gradation from green to red. Negative $a^*$ means green, while positive $a^*$ is assumed as the color of red. The value of $b^*$ explains gradation from blue to yellow. Negative $b^*$ represents blue, whereas positive $b^*$ expresses yellow color.

Table 1. Prediction of $L^*$ using UV-Vis wavelength and UV-Vis-NIR wavelength

| Component | Wavelength | Spectra | PC | $R^2_{\text{cal}}$ | RMSEC | $R^2_{\text{cv}}$ | RMSECV | RPD |
|------------|------------|---------|----|-------------------|--------|-------------------|--------|-----|
| $L^*$      | UV-Vis     | Original| 1  | 0.8               | 1.49   | 0.79              | 1.54   | 2.18 |
|            |            | MSC     | 1  | 0.83              | 1.37   | 0.82              | 1.41   | 2.39 |
|            |            | SNV     | 1  | 0.83              | 1.38   | 0.81              | 1.43   | 2.36 |
| UV-Vis-NIR | Original   | 1       | 0.79| 1.49              | 0.79   | 1.54              | 2.18   |     |
|            | MSC        | 1       | 0.85| 1.24              | 0.85   | 1.28              | 2.59   |     |
|            | SNV        | 1       | 0.85| 1.28              | 0.85   | 1.31              | 2.6    |     |

Prediction of $L^*$ using UV-Vis and UV-Vis-NIR wavelengths gave different prediction accuracy (Table 1). Based on Table 1, spectra correction of SNV and UV-Vis-NIR wavelength produced the best prediction for $L^*$. $R^2_{\text{cal}}$, RMSEC, $R^2_{\text{cv}}$, RMSECV, RPD, and PC from regression analysis showed the values of 0.85, 1.28, 0.85, 1.31, 2.6, and 1, respectively. Calibration model that generates R>0.71 ($R^2>$0.5) is feasible for further prediction [11].

The purpose of storage for 0, 4, and 8 days was to obtain variation of samples. Scatter plot of $L^*$ prediction is shown on Fig. 3. $R^2$ for calibration and validation yielded similar value, therefore the distribution data between calibration and validation on scatter plot of $L^*$ prediction showed similar results as well. Calibration model is the result of regression analysis that involves variable $X$ (spectra) and variable $Y$ (measured data). Validation was performed using cross-validation. The samples were devided into 10 segments. 9 segments were determined as train set, while 1 segment was made as test set.

![Figure 3. Scatter plot of $L^*$ prediction using UV-Vis-NIR wavelength and SNV correction method](image)
The best prediction of \( a^* \) was shown by UV-Vis wavelength and SNV spectra correction with \( R^2_{\text{cal}}, \text{RMSEC}, R^2_{\text{cv}}, \text{RMSECV}, \text{RPD} \) and PC of 0.90, 0.76, 0.91, 0.81, 3.07 and 5, respectively. Although UV-Vis and UV-Vis-NIR wavelength showed similar results, but UV-Vis + SNV generated highest RPD and \( R^2 \) value. In addition, the lowest of RMSEC and RMSECV were displayed by UV-Vis and SNV. Another experiment advised that it was important to check the RPD value even though there was significant correlation between NIR prediction and actual data values [12], [13]. RPD value <1.5 indicates the calibration model is unusable, ranged from 1.5-1.9 denotes the model is able to distinguish the variation between sample, whereas 2.0-2.9 explains that the model is good, and RPD above 3 means the model is excellent [14].

### Table 2. Prediction of \( a^* \) using UV-Vis wavelength and UV-Vis-NIR wavelength

| Component     | Wavelength     | Spectra | PC | \( R^2_{\text{cal}} \) | RMSEC | \( R^2_{\text{cv}} \) | RMSECV | RPD |
|---------------|----------------|---------|----|----------------------|-------|----------------------|--------|-----|
| UV-Vis        | Original       | 2       | 0.9 | 0.79                 | 0.89  | 0.85                 | 3.05   |     |
|               | MSC            | 5       | 0.9 | 0.91                 | 0.88  | 0.86                 | 2.98   |     |
|               | SNV            | 5       | 0.9 | 0.76                 | 0.91  | 0.81                 | 2.87   | 3.07|
| UV-Vis-NIR    | Original       | 3       | 0.9 | 0.93                 | 0.87  | 1.03                 | 2.87   |     |
|               | MSC            | 4       | 0.9 | 0.97                 | 0.89  | 0.85                 | 2.98   |     |
|               | SNV            | 4       | 0.9 | 0.78                 | 0.89  | 0.86                 | 3.03   |     |

Scatter plot of \( a^* \) prediction using UV-Vis +SNV is displayed on Fig. 4. The principal component for this calibration model was determined as 5. PC is usually used to decrease the dimensionality of large data set being analyzed. The optimum number of PC was assigned on the lowest value of the estimated sum of the predicted residual error sum of squares [15]. The selection of an optimal principal component number provides an advantage to make a good accuracy of calibration model.

![Figure 4. Scatter plot of \( a^* \) prediction using UV-Vis wavelength and SNV correction method](image)

Original spectra and UV-Vis-NIR wavelength presented the best prediction accuracy of \( b^* \) with \( R^2_{\text{cal}}, \text{RMSEC}, R^2_{\text{cv}}, \text{RMSECV}, \text{RPD} \) and PC of 0.73, 0.98, 0.67, 1.11, 1.73 and 7, respectively (Table 3). MSC and SNV correction method were not able to improve prediction accuracy in UV-Vis-NIR wavelength. On the other side, MSC and SNV were useful to increase prediction accurac in UV-
Vis wavelength. In this study, it was found that the wavelength range should be considered in the selection of the spectra correction method.

Table 3. Prediction of $b^*$ using UV-Vis wavelength and UV-Vis-NIR wavelength

| Component | Wavelength | Spectra | PC | $R^2_{cal}$ | RMSEC | $R^2_{cv}$ | RMSECV | RPD |
|-----------|------------|---------|----|-------------|--------|------------|--------|-----|
| UV-Vis    | Original   | 3       | 0.58 | 1.12        | 0.51   | 1.19       | 1.46   |
| UV-Vis    | MSC        | 3       | 0.61 | 1.1         | 0.56   | 1.16       | 1.53   |
| UV-Vis    | SNV        | 3       | 0.6  | 1.09        | 0.57   | 1.15       | 1.52   |
| b*        | UV-Vis-NIR | Original| 7   | 0.73        | 0.98   | 0.67       | 1.11   | 1.73|
| b*        | UV-Vis-NIR | MSC     | 7   | 0.72        | 1      | 0.6        | 1.21   | 1.57|
| b*        | UV-Vis-NIR | SNV     | 7   | 0.73        | 1      | 0.65       | 1.15   | 1.69|

Figure 5 shows the scatter plot of $b^*$ prediction. Reference data of $b^*$ was ranged from 23.5 to 34.5, which means that the surface color of samples were closer to yellow. ‘Crystal’ guava color changed during ripening phase. The color of fruit is affected by pigments. Chlorophyll content decreases while the carotenoid content increases during ripening phase [16]. Another research reported that the fruit color change during ripening is caused by the loss of chlorophyll content [17].

Figure 5. Scatter plot of $b^*$ prediction using UV-Vis-NIR wavelength and original spectra

4. Conclusion

Surface color of ‘crystal’ guava fruit was able to be predicted using the UV-Vis-NIR spectroscopy and multivariate analysis. UV-Vis-NIR wavelength showed the most feasible model for predicting $L^*$ and $b^*$. UV-Vis wavelength gave the best model accuracy for $a^*$ prediction. Spectra correction method of SNV increased the $L^*$ and $a^*$ prediction accuracy.

References
[1] D. Giuffrida, F. Salvo, A. Salvo, L. L. Pera, and G. Dugo, “Pigments Composition in Monovarietal Virgin Olive Oils from Various Sicilian Olive Varieties,” Food Chem., vol. 101, no. 2, pp. 833–837, 2007.
[2] F. C. G. B. S. Alves, A. Coqueiro, P. H. Março, and P. Valderrama, “Evaluation of Olive Oils from the Mediterranean Region by UV–Vis Spectroscopy and Independent Component Analysis,” Food Chem., vol. 273, no. 1, pp. 124–129, 2019.
[3] M. Zude-Sasse, “Non-Destructive Prediction of Banana Fruit Quality Using VIS/NIR Spectroscopy,” *Fruits*, vol. 58, no. 3, pp. 135–142, 2003.

[4] Kusumiyati, Y. Hadiwijaya, and I. E. Putri, “Determination of Water Content of Intact Sapodilla Using Near Infrared Spectroscopy,” in *IOP Conference Series: Earth and Environmental Science, Volume 207.*, 2018, pp. 1–8.

[5] Kusumiyati et al., “Evaluation of Sapodilla Fruit Quality Using Near-Infrared Spectroscopy,” *J. Food, Agric. Environ.*, vol. 16, no. 1, pp. 49–53, 2018.

[6] T. R. Viegas, A. L. M. L. Mata, M. M. L. Duarte, and K. M. G. Lima, “Determination of Quality Attributes in Wax Jambu Fruit Using NIRS and PLS,” *Food Chem.*, vol. 190, pp. 1–4, 2016.

[7] Y. A. Purwanto et al., “Non Destructive Prediction of pH in Mango Fruits cv. Gedong Gincu Using NIR Spectroscopy,” *Int. J. Eng. Technol.*, vol. 13, no. 3, pp. 70–73, 2013.

[8] X. P. Fu, Y. B. Ying, Y. Zhou, L. J. Xie, and H. R. Xu, “Application of NIRS Spectroscopy for Firmness Evaluation of Peaches,” *J. Zhejiang Univ. Sci. B*, vol. 9, no. 7, pp. 552–557, 2008.

[9] Kusumiyati, Y. Hadiwijaya, and I. E. Putri, “Non-destructive Classification of Fruits Based on UV-Vis-NIR Spectroscopy and Principal Component Analysis,” *J. Biodjati*, vol. 4, no. 1, pp. 89–95, 2019.

[10] N. Jain, K. Dhand, S. Maholtra, and R. Singh, “Biochemistry of Fruit Ripening of Guava (Psidium guajava L.): Compositional and Enzymatic Changes,” *Plant Foods Hum. Nutr.*, vol. 58, no. 4, pp. 309–315, 2003.

[11] Y. Ozaki, W. F. McClure, and A. A. Christy, *Near-Infrared Spectroscopy in Food Science and Technology*. New Jersey: John Wiley & Sons, Inc, 2006.

[12] M. W. Davey, W. Saeys, E. Hof, H. Ramon, R. L. Swennen, and J. Keulemans, “Application of Visible and Near-Infrared Reflectance Spectroscopy (Vis/NIRS) to Determine Carotenoid Contents in Banana (Musa spp.) Fruit Pulp,” *J. Agric. Food Chem.*, vol. 57, no. 5, pp. 1742–1751, 2009.

[13] L. S. Magwaza et al., “Prediction of ‘Nules Clementine’ Mandarin Susceptibility to Rind Breakdown Disorder Using VIS/NIR Spectroscopy,” *Postharvest Biol. Technol.*, vol. 74, pp. 1–10, 2012.

[14] W. Saeys, A. M. Mouazen, and H. Ramon, “Potential for Onsite and Online Analysis of Pig Manure Using Visible and Near Infrared Reflectance Spectroscopy,” *Biosyst. Eng.*, vol. 91, no. 4, pp. 393–402, 2005.

[15] D. M. Haaland and E. V. Thomas, “Partial Least Squares Methods for Spectral Analyses. 1. Relation to Other Quantitative Calibration Methods and The Extraction of Qualitative Information,” *Anal. Chem.*, vol. 60, no. 11, pp. 1193–1202, 1988.

[16] A. Anbarasan and C. Tamilmani, “Effect of Calcium Pectate on the Biochemical and Pigment Changes During The Ripening of Bitter Gourd Fruit (Momordica charantia L. Var-Co-1),” *Asian J. Plant Sci. Res.*, vol. 3, no. 4, pp. 51–58, 2013.

[17] A. Aberoumand, “A Review Article on Edible Pigments Properties and Sources as Natural Biocolorants in Foodstuff and Food Industry,” *World J. Dairy Food Sci.*, vol. 6, no. 1, pp. 71–78, 2011.