Water content prediction of ‘crystal’ guava using visible-near infrared spectroscopy and chemometrics approach

Kusumiyati¹, Y Hadiwijaya², I E Putri² and S Mubarok¹

¹ Department of Agronomy, Agriculture Faculty, Universitas Padjadjaran, Sumedang 45363, Indonesia
² Graduate Student in Agronomy, Agriculture Faculty, Universitas Padjadjaran, Sumedang 45363, Indonesia

E-mail: kusumiyati@unpad.ac.id

Abstract. The purpose of the research was to apply the visible-near infrared (vis-nir) spectroscopy and chemometrics approach to predict the water content of crystal guava rapidly and non-destructively. The absorbance spectra were collected from intact ‘crystal’ guava fruits in wavelength from 381 to 1065 nm. Spectra pre-treatment was used to enhance the prediction accuracy. Several spectra pre-treatment methods were arranged to the original absorbance spectra such as multiplicative scatter correction (MSC), standard normal variate (SNV), second derivative absorbance (d2a), MSC+d2a, and SNV+d2a. Calibration model was developed by partial least squares regression (PLSR). Validation was done by K-fold cross-validation. The results presented that vis-nir spectroscopy combined with chemometrics approach gave accurate water content prediction of ‘crystal’ guava. The best calibration model was provided by spectra pre-treatment of SNV+d2a with coefficient of determination (R²) and ratio performance deviation (RPD) of 0.74 and 1.60, respectively. It concluded that vis-nir spectroscopy and chemometrics approach can be performed to predict the water content of ‘crystal’ guava rapidly and non-destructively by replacing the standard laboratory analysis.

1. Introduction
Generally, most researchers assess the quality of fruit by doing destructive analysis. This technique results damaged fruit, hence the fruit can not be marketed afterwards. Destructive analysis also requires some chemical materials and consumes times and costs [1]. Visible-near infrared (vis-nir) spectroscopy is new non-destructive technique to predict fruit quality. It is a spectroscopic technique based on atomic vibrations of the molecules. The vis-nir spectrum is generated by passing light radiation onto the sample and determining the absorbed radiation. The advantages of using vis-nir spectroscopy are damageless prediction, and also many sample forms can be investigated such as fluids, solvents, pastes, powders, fibers, and gases can be identified by appropriate sample selection techniques [2]. In addition, vis-nir spectroscopy technique is environmentally friendly.

The spectroscopy technique was first used in agriculture to measure the moisture content of grain [3]. Nowadays, this technique has been done to determine the total soluble solids, firmness, acidity, surface color, and dry matter of fruits [4], [5], [6]. In another study vis-nir spectroscopy technique was used to predict water content of intact sapodilla [7]. Other research reported that this technique was performed to differentiate fruit species from other species [8]. Once the spectra measurement was completed, the spectra need pretreatment to remove the noises. Standard normal variate (SNV),
multiplicative scatter correction (MSC), and second derivative absorbance (d2a) are the commonly used spectra pretreatment methods. Storage is one of the important things in postharvest handling of fruits, because the quality of the fruit may change during storage. The factors that affecting fruit quality changes are storage duration and temperature. The research objective was to predict the water content of ‘crystal’ guava fruit using vis-nir spectroscopy and chemometrics approach, which stored for 0, 4 and 8 days at room temperature.

2. Materials and Method

2.1. Raw Materials
Guava fruits for this study were obtained from Tomo, Sumedang, Indonesia. Each guava sample was numbered before the analysis, then stored for 0, 4, and 8 days at room temperature (27°C). The aim of the storage was to acquire data variation of water content. A total of 150 fruit samples were used in this research.

2.2. Spectra Data Acquisition
Guava fruit samples were irradiated using spectrometer (NirVana AG410) with wavelength of 381-1065 nm and 3 nm interval. Each sample was illuminated 3 points on the frontside and 3 points on the backside vertically starting from the top to the bottom of the fruit. Once the spectra data were obtained, multivariate calibration of chemometric was performed to extract the information from vis-nir spectra data [9]. Afterwards, the spectra data along with reference data were used to build the calibration model using partial least squares regression (PLSR).

2.3. Reference Data Measurement
The water content of guava is measured by calculating fresh wet weight compared with dried fruit weight. The guava fruits were sliced then dried in the oven with temperature of 80°C. Dried sample was calculated every day until the weight showed constant value. The water content value is expressed by percentage (%) of water content.

2.4. Chemometrics Analysis
Chemometrics is a technique of extracting data using mathematics and statistics to analyze chemical data. In spectroscopic analysis, chemometrics is composed of three aspects, pretreatment of spectra data, model development, and transfer model. One of the most important chemometric analysis is to build calibration model through a pattern recognition technique to recognize the major similarities and data patterns. In addition, this model is performed to predict unknown samples [10]. Currently, chemometrics is utilized to analyze multivariate data that involving multiple variables and each variable can be mutually correlated.

2.5. Partial Least Squares Regression
Partial least squares regression (PLSR) is the most used multivariate calibration techniques in chemometrics due to the quality of calibration model and the ease of application. PLSR is able to analyze large numbers of data, having high level of collinearity, a lot numbers of X variables and some Y response variables [11]. This is dimensional data reduction method to find the most relevant factors in predicting and interpreting data. PLSR removes predictor data that is not relevant to the data diversity. Furthermore, this method can also be used for classification purposes. PLSR analysis was done using absorbance spectra and destructive analysis data of each sample to build the calibration model.

2.6. Model Accuracy
Root mean squares error (RMSE) is a method to assess the prediction technique that is used to evaluate the accuracy level of prediction model. RMSE measures the difference of data predicted by model and the actual reference data. Root mean squares error of calibration (RMSEC) is RMSE for calibration
model, while root mean squares error of cross validation (RMSECV) is the RMSE for cross validation model. RMSE reveals the mean of inaccuracy that can be expected for prediction of unknown samples. Besides, coefficient of determination (R squared or R²) is also important to determine the model accuracy. R² explains the goodness of model.

Ratio performance deviation (RPD) is obtained by dividing the standard deviation of reference data in cross validation set with root mean squares error of cross validation [12]. RPD ranged from 1.5 to 2 indicates the model is able to distinguish low from high values of the reference data; value ranged from 2 to 2.5 indicates the decent quantitative model prediction; RPD value above 2.5 defines the excellent model prediction [13], [14]. RPD is acquired using the following equation:

\[
\text{RPD} = \frac{\text{SD}}{\text{RMSECV}}
\]

3. Results and Discussion

3.1. Vis-NIR Spectra of ‘Crystal’ Guava

The absorbance spectra denote the intensity of light absorption by the samples. Amount of light absorption is different on each fruit species due to different physical and chemical properties. Radiation between 381-1065 nm describes the vis-nir region. Fig 1. displays absorbance spectra of the samples. Noises of spectra appeared on almost the entire wavelengths. Spectra pretreatment is necessary to correct the spectra and improve the prediction accuracy. MSC, SNV, and d2a spectra preprocessing were practiced to the original spectra. Moreover, combination of MSC+d2a and SNV+d2a were expected to give better results.

3.2. Destructive Analysis

Table 1 presents destructive data of water content. Based on this study, water content of ‘crystal’ guava ranged from 80 to 88%, while the average generated the value of 85%. Destructive measurement data were used as response variable, while spectra data were taken as predictor variable for the regression analysis.
Table 1. Reference data of water content

| Numbers of Samples | Range (%) | Mean (%) | Standard Deviation |
|--------------------|-----------|----------|--------------------|
| 150                | 80-88     | 85       | 0.01               |

3.3. Calibration Model and Validation

![Figure 2](image)

Fig. 2 exposes the calibration and validation plot using PLSR combined with cross-validation. The calibration models that have been built need to be reassessed to verify the calibration accuracy, therefore validation analysis was required. Validation was performed using K-fold cross-validation method. The samples were splitted into 20 segments, each segment contained about 7-8 samples. A total of 19 segments were used as training sets, while 1 segment was made as test set. Cross-validation was applied to prevent the model overfitting, which could lead to the failure of unknown samples prediction. Table 2 provides the calibration and validation of water content prediction. The obtained results showed the validation did not differ much from the calibration model. The best spectra correction method was shown by SNV combined with d2a that yielded $R^2$ of 0.74 and 0.61 for calibration and validation. Besides, the RPD produced 1.60 value or the highest result than other spectra correction method. A good model shall provide high $R$ value in calibration and validation set, as well as having low RMSEC and RMSEP values. In addition, RMSEC and RMSEP are expected to have almost the same value. RPD values alone can not be used as benchmark for calibration potential, it requires to consider other statistical analysis results such as $R^2$ value of validation set, RMSECV, and the number of latent variables used in the calibration set [15].
**Table 2.** Calibration and validation of water content prediction based on Vis-NIR spectroscopy and spectra pretreatment

| Spectra     | Latent Variable | Calibration | Validation | RPD |
|-------------|-----------------|-------------|------------|-----|
|             |                 | R²          | RMSEC      | R²  | RMSECV |
| Original    | 8               | 0.59        | 0.01       | 0.52| 0.01   | 1.43 |
| MSC         | 8               | 0.63        | 0.01       | 0.51| 0.01   | 1.43 |
| SNV         | 10              | 0.66        | 0.01       | 0.55| 0.01   | 1.50 |
| d2a         | 8               | 0.70        | 0.01       | 0.55| 0.01   | 1.48 |
| MSC+d2a     | 9               | 0.72        | 0.01       | 0.55| 0.01   | 1.49 |
| SNV+d2a     | 9               | 0.74        | 0.01       | 0.61| 0.01   | 1.60 |

4. Conclusion

The results showed that non-destructive technique using vis-nir spectroscopy and chemometrics approach produced high accuracy values to predict the water content of ‘crystal’ guava. The best spectra correction method was displayed by combination of SNV+d2a.

Acknowledgement

This research was financially supported by DRPMI of Universitas Padjadjaran.

References

[1] Gómez A H, He Y and Pereira G 2006 *J. Food Engineering* 77 313–319
[2] Stuart B 2004 *Infrared spectroscopy: Fundamental and Applications* (Philadelphia: Saunders College Pub.)
[3] Norris K H 1964 *J. Agric. Eng.* 45 370-2
[4] Schmilovitch Z, Mizrach A, Hoffman A, Egozi H and Fuchs Y 2000 *Postharvest Biol. Technol.* 19 245–52
[5] Nagle M, Mahayothee B, Rungpichayapichet P, Janjai S and Müller J 2010 *Sci. Hortic.* 125 771–4
[6] Kusumiyati, Mubarok S, Hamdani J S, Farida, Sutari W, Hadiwijaya Y, Putri I E and Mutiarawati T 2018 *J. Food Agric. Environ.* 16 49–53
[7] Kusumiyati, Hadiwijaya Y and Putri I E 2018 IOP Conf. Ser. Earth Environ. Sci. 207
[8] Kusumiyati, Hadiwijaya Y and Putri I E 2019 *J. Biodjati* 4 89-95
[9] Stchur P, Cleveland D, Zhou J and Michel R. 2002 *Appl. Spectrosc. Rev.* 37 383–428
[10] Africia S M 2007 Kombinasi Metode Spektrofotometri dan Kalibrasi Multivariat untuk Penentuan Simultan Natrium Benzoat dan Kalsium Sorbat (Bogor: IPB University) [Master Thesis]
[11] Kuno A and Matsuo M 2000 *Anal. Sci.* 16 597-602
[12] Williams P 2007 Application to agricultural and marine products *Near-Infrared Spectroscopy in Food Science and Technology* eds Ozaki Y, McClure W F and Christy A A (New Jersey: John Wiley and Sons, Inc. Pub.) pp 165–218
[13] Nicolaï B M, Beullens K, Bobelyn E, Peirs A, Saey S, Theron K I and Lammertyna J 2007 *Postharvest Biol. Technol.* 46 99–118
[14] Viscarra R R V, McGlyn R N, and McBretney A B 2006 *Geoderma* 146 403-11
[15] Ignat T, Schmilovitch Z, Fefoldi J, Steiner B and Alkalai-Tuvia S 2012 *Postharvest Biol. Technol.* 74 91–9