Analytical Methods

Determination of quality attributes in wax jambu fruit using NIRS and PLS

Thayna R. Viegas [a], Ana L.M.L. Mata [a], Márcia M.L. Duarte [a], Kássio M.G. Lima [b,*]

[a] Federal University of Rio Grande do Norte, Department of Chemical Engineering, 59072-970 Natal, RN, Brazil
[b] Federal University of Rio Grande do Norte, Institute of Chemistry, PPGQ, Biological Chemistry and Chemometrics, 59072-970 Natal, RN, Brazil

CrossMark

ARTICLE INFO

ABSTRACT

The aim of this work was to develop an analytical method to predict total anthocyanins content (TAC) and total phenolic compounds (TPC) in intact wax jambu fruit [*Syzygium malaccense* (L.) Merryl et Perry] using near-infrared spectroscopy (NIRS) and partial least squares (PLS). The estimation accuracy was based on parameters such as root mean square error of prediction (RMSEP), correlation coefficients [calibration (r_c) and prediction (r_p) set] and ratio of performance to deviation (RPD). TAC, $r_p = 0.98$, RMSEP = 9.0 mg L^{-1} and RPD = 5.19 were attained using second derivative pre-treatment. TPC, $r_p = 0.94$, RMSEP = 22.18 (mg gallic acid equivalents (GAE)/100 g) and RPD = 3.27 (excellent accuracy) were also obtained using second derivative pre-treatment. These findings suggest that the NIRS and PLS algorithms can be used to determine TCA and TPC in intact wax jambu fruit.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Syzygium malaccense (L.) Merr. and L.M. Perry, commonly known as wax jambu or Malay rose apple, or simply Malay apple, is an evergreen tree with origins in Asia (Malaysia, Indonesia, Vietnam and Thailand). It has been introduced and grown throughout the tropics and sub-tropical parts of the word, such as Panamá, Costa Rica, Venezuela, Puerto Rico and Brazil. The tree can grow to 12–18 m in height with four fleshy calyx lobes and 1–4 seeds (1–2 cm in diameter). The fruit is a pear-shaped bacca, juicy and indehiscent. The epicarp is thin, smooth and red dish; the mesocarp and the endocarp are whitish and succulent (Costa, Oliveira, Môro, & Martins, 2006).

Previous research on wax jambu has reported quality attributes of the intact or fresh-cut fruit such as total soluble solids (TSS), total anthocyanin content (TAC), browning index, firmness, titratable acidity (TA), sugar acid ratio (TSS/TA), total phenolic content (TPC), and ascorbic acid content (Khandaker, Boyce, & Osman, 2012; Shü, Chu, Hwang, & Shieh, 2001; Supapvanich, Pimsaga, & Srisujan, 2011). However, despite the reliability of quality attributes for wax jambu fruit, the analytical methods are inherently destructive, expensive, require specialized reagents and generate chemical waste.

In recent years, research has aimed to develop non-destructive techniques for measuring directly quality attributes of intact samples without any sample preparation. In this way, near-infrared spectroscopy (NIRS) with diffuse reflectance is a prominent technique for non-destructive assessment of fruit quality. NIRS has been used to determine quality attributes of a number of fruit including açaí (Inácio, de Lima, Lopes, Pessoa, & de Almeida Teixeira, 2013); plums (Costa & Lima, 2013); apples (Nicolaï, Lötze, Peirs, Scheerlinck, & Theron, 2006); melons (Greensill, Wolfs, Spiegelman, & Walsh, 2001); grapes (Janik, Cozzolino, Dambergs, Cynkar, & Gishen, 2007); oranges (Cayuela, 2008); mangos (Saranwong & Sornsrivichai, 2003); strawberries (Sánchez et al., 2012); kiwifruit (Clark et al., 2004); and pears (Nicolaï et al., 2008).

Although NIRS is becoming an increasingly attractive analytical technique for measuring quality parameters of fruits, several complicating factors remain. The major difficulties in analyzing quality attributes in fruits by NIRS are: (i) weakness of the NIRS signals from fruit components compared with those from other components; (ii) complexity of overlapping bands due to the chemical composition of a typical fruit; and (iii) wavelength-dependent scattering effects (tissue heterogeneities, instrumental noise, ambient effects and other sources of variability). To overcome these difficulties, sophisticated multivariate statistical techniques are used to extract useful information from an NIR spectrum. Essentially, these include regression techniques such as partial least squares (PLS) (Dupuy, Galtier, Ollivier, Vanloot, & Artaud,

* Corresponding author.
E-mail address: kassiolima@gmail.com (K.M.G. Lima).

2010), principal component regression (PCR) (Xie & Kalivas, 1997), artificial neural networks (ANN) (Makino, Ichimura, Oshita, Kawagoe, & Yamanaka, 2010) and least squares–vector support machine (LS–SVM) (Shao, Zhao, Bao, & He, 2012), coupled with spectral preprocessing such as averaging (Nicolaï et al., 2006), smoothing (Næs, Isaksson, Fearn, & Davies, 2004), standardization (Greensill et al., 2001), and transformation (Griffiths, 1995).

The quality attributes of wax jambu [*S. malaccense* (L.) Merryl et Perry] fruit have never been calibrated using NIR spectroscopy or any other rapid technique. Herein, we have attempted to evaluate the feasibility of rapidly measuring intact wax fruit with a fiber optic probe FT-NIR spectrometer. The specific objectives of our research were to: (1) determine relationships between FT-NIR measurements with total anthocyanins (TA) and total phenolic compounds (TPC) of wax fruit based on PLS method; (2) calculate prediction performance of calibration and prediction models using PLS method and establish the best calibration models.

2. Experimental

2.1. Sampling

All the samples were harvested by hand (October, 2013 and March, 2014) from a metropolitan area of Northeast Brazil (Natal). Fruits were harvested from different trees, then pooled and randomized. The wax jambu used for each experiment were selected on the basis of uniform size and color, and being free from any diseases and physical damage. The sorted wax jambu were stored under ambient room conditions (26–30 °C, RH 60–80%) before NIR diffuse reflectance spectral measurements were performed. All measurements, including spectral collection and parameter determination (TAC and TPC values), were carried out on the same or following day. In the laboratory, we measured fruit morphological characteristics including weight and diameter (average values). The weight (24.87 g) was determined using an electric analytical balance, while the diameter (3.28 cm) was determined using a tape measure. After weighing, we cut the fruit in half with a sharp knife and each half was cut at the exposed edge into four equal pieces. The cavity tissue and the bottom of the pieces were removed. The number of samples (*N*), minimum (Min), maximum (Max), median and standard deviation (S.D.) for each parameter are presented in Table 1.

2.2. Reference methods

2.2.1. TAC

We used the pH differential method applicable for the determination of monomeric anthocyanins, such cyanidin-3-glucoside, in fruit as a reference method (Lee, Durst, & Wrolstad, 2005). Total anthocyanin extraction was carried out using fresh-cut wax jambu. Two portions of macerated pulp (0.050 g each) were weighed out. One was mixed with 0.025 mol L^{-1}, pH 1.0, potassium chloride buffer, and the other with 0.4 mol L^{-1}, pH 4.5, sodium acetate buffer. After two at room temperature ($\sim$25 °C), the samples were passed through a Whatman No. 1 filter paper, and the absorbance of both solutions recorded using a spectrophotometer (SP-22 UV–Vis, Curitiba, Brazil) at 520 and 700 nm, respectively. All analyses were performed in triplicate.

2.2.2. TPC

TPC was measured using a colorimetric Folin–Ciocalteu method (Li, Hydamaka, Lowry, & Beta, 2009). Wax jambu extract (1 mL) or gallic acid (standard) were mixed with 0.5 mL of Folin–Ciocalteu reagent (prediluted 10-fold with distilled water) and allowed to stand at room temperature for 5 min before 1.0 mL of sodium bicarbonate (7.5%) was added to the mixtures. After standing for 60 min at room temperature, the absorbance was measured at 765 nm. Results were expressed as mg gallic acid equivalents (GAE)/100 g sample. All analyses were performed in triplicate.

2.3. NIRS

Spectral measurements were performed using an Antaris MX FT-NIR spectrophotometer (Thermo Fisher Scientific Inc, USA) equipped with a NIR fiberoptic probe, interferometer, cooled InGaAs detector, and a wide-band quartz halogen light source (50 W). The NIR spectra were obtained across the ranges 10,000–4166 cm^{-1} or 1000–2400 nm, and were recorded with a spectral resolution of 32 cm^{-1} and an average of 32 scans. Each spectrum (32 scans) took 26 s. For each wax jambu fruit, we measured a diffuse reflectance spectrum on three evenly spaced equatorial positions but only an averaged spectrum (three spectra per fruit) was used for analysis, giving a total of 50 NIR spectra. Absorbance spectra of wax jambu samples were obtained against a spectrum for spectralon as the background. Spectral measurements were obtained in a room acclimatized to 22 °C and 60% relative air humidity.

2.4. Chemometrics procedure and software

We performed data loading, pre-processing Savitzky-Golay smoothing with different windows (3, 5 and 7) (Nicolaï et al., 2007), multiplicative scattering correction (MSC) (Isaksson & Naes, 1988), and derivatization – first and second derivatives (Cen & He, 2007), chemometric regression model construction (PLS) and validation in a MATLAB®version7.10 environment (Math-Works, Natick, USA) with the PLS-toolbox (Eigenvector Research, Inc., Wenatchee, WA, USA, version 7.8). The number of calibration and validation samples selected were 50 and 35, respectively, applying the classic Kennard-Stone (KS) selection algorithm to the NIR spectra (Kennard & Stone, 1969).

We tested the predictability of the model by determining the RMSECV (root mean square error of cross validation), RMSEP (root mean square error of prediction) and correlation coefficients of each model for calibration data set (r_c) and prediction data set (r_p).

The RMSECV, RMSEP and r were calculated as follows:

$$\text{RMSECV} = \sqrt{\frac{1}{I_c - 1} \sum_{i=1}^{I_c} (\hat{y}_i - y_i)^2} \tag{1}$$

where $\hat{y}_i$ is the predicted value of the *i*th observation, y_i the measured value of *i*th observation and I_c is the number of observations in the calibration set.

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{I_p} (y_i - \tilde{y}_i)^2}{I_p}} \tag{2}$$

where $\tilde{y}_i$ is the predicted value for predict set sample *i*, y_i the measured value for predict sample *i* and I_p is the number of observations in the prediction set.

$$r = \sqrt{1 - \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}} \tag{3}$$

Table 1
Summary of the samples and reference (TAC/TPC) concentrations showing, number of samples (N), minimum (Min), maximum (Max), mean and standard deviation (S.D.).

Parameter	N	Min	Max	Mean	S.D.
TAC	50	9.07	218.7	66.0	47.0
TPC	50	31.24	374.6	135.6	72.6

where $\hat{y}_i$, y_i are the predicted and measured values of sample i in the calibration and prediction sets, $\bar{y}$ the mean of the reference measurement results for all samples in the calibration and prediction sets, and n is the number of observations in the calibration and prediction sets.

The ratio of S.D. (standard deviation) to the RMSECV or the RMSEP, called the ratio of prediction to deviation (RPD), is the factor by which the prediction accuracy increases compared with the mean composition for all samples. Ideally, this ratio should be greater than two for a good calibration.

3. Results and discussion

3.1. Spectrum description

Fig. 1A shows the raw wax jambu average spectra in the NIR spectral region from 1000 to 2400 nm. The overtone and combination bands observed in the NIR spectra are due to the C–H, N–H and O–H bonds. The spectra show five broad absorption peaks around the 1190, 1450, 1790, 1940 and 2380 nm regions (see arrows in Fig. 1). At 1190 nm, this could be attributed to the first overtone bands of C–H groups present in sugars, and the prominent absorption bands around 1450 and 1940 nm were identified as water absorption. The peak at 1790 nm overlaps with the first C–H overtone region, which is also sugar related. The small peak at 2380 nm falls within the regions associated with the C–H and C–H combination.

Initially, the PLS method was performed on the entire range of the raw spectra to examine the NIR model for non-destructively predicting TAC and TPC. As shown in Fig. 1A, raw spectral data needed to undergo spectral preprocessing. There were undesirable systematic variations in the data, such as baseline drift and random noise. In this study, three spectral preprocessing methods were applied comparatively, as follows: Savitzky-Golay smoothing with different windows (3, 5 and 7), multiplicative scatter correction (MSC), and derivatization – first and second derivatives. The resulting second-derivative spectra are shown in Fig. 1B.

3.2. PLS models for predicting of TAC

The calibration models based on PLS were developed in the 1000–2400 nm region. The results obtained for the calibration models in the NIR region for TAC for wax jambu fruit are shown in Table 2. The PLS models for TAC determination shown in Table 2 had RMSECV and RMSEP values between 7.56–38.78 (mg L^{-1}) and 9.04–40.15 (mg L^{-1}) for calibration and prediction sets, respectively. The correlation coefficients for the prediction set ranged from 0.45 to 0.98 for all the models. However, in analyzing the results obtained from all the pre-treatment methods, the application of the second derivative produced excellent results where the RMSECV and RMSEP values decreased strongly in comparison with
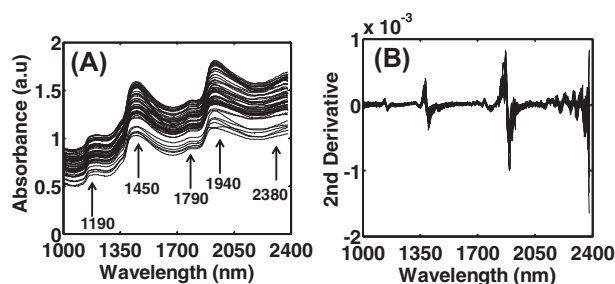
Table 2
Summary of statistics for calibrations and predictions for TAC (mg L^{-1}) in wax jambu fruit using FT–NIR spectra and PLS.

Spectral pre-treatments	Calibration			Prediction		
	r_c	RMSEC V	r_p	RMSEP	RPD	VL
Raw	0.50	38.78	0.45	40.1	1.17	10
Smoothing	0.52	37.95	0.51	38.6	1.21	10
MSC	0.75	23.91	0.72	24.3	1.93	8
1st derivative	0.88	15.32	0.85	16.0	2.92	8
2nd derivative	0.99	7.56	0.98	9.0	5.19	8

the direct regression model on raw spectra. RMSECV and RMSEP values for second derivative pre-treatment were 7.56 and 9.04, respectively. The correlation coefficient for the prediction set was 0.98 and was obtained using eight latent variables.

Furthermore, when the second derivative was performed, we obtained a RPD of 5.19, which ensured the accuracy and robustness of the model for the prediction of TAC in wax jambu. Fig. 2 shows the goodness of fit using second derivative and PLS, presented by plotting measured and predicted values for TAC in wax jambu. Lastly, the PLS model (second derivative) was not significantly different using prediction samples for TAC when compared with the reference values according to a paired t-test ($t_{calculated} = 0.58$, $t_{critical} = 1.76$, 95% confidence level).

3.3. PLS models for predicting of TPC

The results obtained in calibration and prediction models after application of spectral pre-treatment are summarized in Table 3. The RMSECV values for TPC were 51.93% and 19.01% (mg gallic acid equivalents (GAE)/100 g), respectively, while the RMSEP values ranged between 79.15% and 22.18% (mg gallic acid equivalents (GAE)/100 g). The correlation coefficients for the prediction set ranged from 0.43 to 0.98 for all the models. From the table, it can be seen that the second derivative produced a significant improvement in terms of both cross-validation and external prediction (smaller RMSECV/RMSEP and larger r_c/r_p). Evidently, this PLS model achieved the best prediction precision with RMSEP = 22.18 and RPD = 3.27.

These results are corroborated by the graph of predicted versus reference values obtained using the PLS model (second derivative) with nine latent variables, and we also found a correlation coefficient of 0.98 for the prediction set, as shown in Fig. 3. The presence of relevant bias was tested with prediction results for prediction samples using the PLS model (second derivative) with the t-test. The results showed the bias included in the model was not significant, since the t-value obtained (0.49 for TPC) was less than the critical value of 1.76 with 95% confidence.
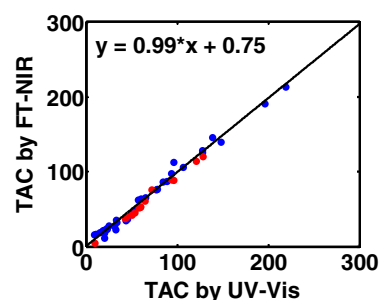

Fig. 1. Set of original NIR average spectra (1000–2400 nm) containing samples of wax jambu: (A) raw spectra; (B) second derivative pretreatment.

Fig. 2. Correlation plot of calibration and prediction set for the determination of TAC (mg L^{-1}) in wax jambu using PLS model after second derivative pretreatment. (●) calibration set; (●) prediction set. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Summary of statistics for calibrations and predictions for TPC (mg gallic acid equivalents (GAE)/100 g) in wax jambu fruit using FT-NIR spectra and PLS.

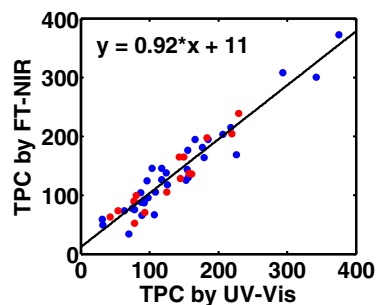| Spectral pre-treatments	Calibration			Prediction		
	r_c	RMSEC V	r_p	RMSEP	RPD	VL
Raw	0.45	51.93	0.43	79.15	0.96	10
Smoothing	0.48	50.69	0.47	77.12	0.98	10
MSC	0.71	35.87	0.68	39.58	1.92	9
1st derivative	0.90	25.98	0.86	27.56	2.76	9
2nd derivative	0.95	19.01	0.94	22.18	3.27	9

Fig. 3. Correlation plot of calibration and prediction set for the determination of TPC (expressed as mg kg^{-1} of gallic acid) in wax jambu using PLS model after second derivative pretreatment. (●) calibration set; (●) prediction set. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. Conclusions

Results reported here demonstrate the capability of NIRS combined with multivariate analysis (PLS) as a tool for the rapid and non-destructive prediction of two properties (TAC and TPC) of wax jambu fruit for the first time. The strength of the calibrations for each parameter varied with the different mathematical treatments, but the second derivative data produced the strongest results for predictions. The strong RPD values (5.19 for TAC and 3.19 for TPC) suggest that NIR spectroscopy could be an effective method for the measurement of important parameters in wax jambu.

References

Cayuela, J. A. (2008). Vis/NIR soluble solids prediction in intact oranges (*Citrus sinensis* L.) cv. Valencia Late by reflectance. *Postharvest Biology and Technology, 47*(1), 75–80.
Cen, H., & He, Y. (2007). Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends in Food Science & Technology, 18*(2), 72–83.
Clark, C. J., McGlone, V. A., De Silva, H. N., Manning, M. A., Burdon, J., & Mowat, A. D. (2004). Prediction of storage disorders of kiwifruit (*Actinidia chinensis*) based on visible-NIR spectral characteristics at harvest. *Postharvest Biology and Technology, 32*(2), 147–158.
Costa, R. C., & Lima, K. M. G. de L. (2013). Prediction of parameters (soluble solid and pH) in intact plum using NIR spectroscopy and wavelength selection. *Journal of Brazilian Chemical Society, 24*(8), 1351–1356.
Costa, R. S., Oliveira, I. V. de M., Môro, F. V., & Martins, A. B. G. (2006). Morphological aspects and influence of the seed size in the germination of wax jambu. *Revista Brasileira de Fruticultura Jaboticabal, 28*(16), 117–120.
Dupuy, N., Galtier, O., Ollivier, D., Vanloot, P., & Artaud, J. (2010). Comparison between NIR, MIR, concatenated NIR and MIR analysis and hierarchical PLS model. Application to virgin olive oil analysis. *Analytica Chimica Acta, 666*(1–2), 23–31.
Greensill, C. V., Wolfs, P. J., Spiegelman, C. H., & Walsh, K. B. (2001). Calibration transfer between PDA-Based NIR Spectrometers in the NIR Assessment of Melon Soluble Solids Content. *Journal of Applied Spectroscopy, 55*(5), 647–653.
Griffiths, P. R. (1995). Letter: practical consequences of math pre-treatment of near infrared reflectance data: Log(1/R) vs. F(R). *Journal of Near Infrared Spectroscopy, 3*, 60–62.
Inácio, M. R. C., de Lima, K. M. G., Lopes, V. G., Pessoa, J. D. C., & de Almeida Teixeira, G. H. (2013). Total anthocyanin content determination in intact açaí (*Euterpe oleracea* Mart.) and palmitero-juçara (*Euterpe edulis* Mart.) fruit using near infrared spectroscopy (NIR) and multivariate calibration. *Food Chemistry, 136*(3–4), 1160–1164.
Isaksson, T., & Naes, T. (1988). The effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy. *Applied Spectroscopy, 42*(7), 1273–1284.
Janik, L. J., Cozzolino, D., Dambergs, R., Cynkar, W., & Gishen, M. (2007). The prediction of total anthocyanin concentration in red-grape homogenates using visible-near-infrared spectroscopy and artificial neural networks. *Analytica Chimica Acta, 594*(1), 107–118.
Kennard, R., & Stone, L. (1969). Computer aided design of experiments. *Technometric, 11*(1), 137–148.
Khandaker, M. M., Boyce, A. N., & Osman, N. (2012). The influence of hydrogen peroxide on the growth, development and quality of wax apple (*Syzygium samarangense,* [Blume] Merrill & L.M. Perry var. jambu madu) fruits. *Plant Physiology and Biochemistry, 53*, 101–110.
Lee, J., Durst, R. W., & Wrolstad, R. E. (2005). Determination of total monomeric anthocyanin pigment content of fruit juices, beverages, natural colorants, and wines by the pH differential method: Collaborative study. *Journal of AOAC International, 88*(5), 1269–1278.
Li, W., Hydamaka, A. W., Lowry, L., & Beta, T. (2009). Comparison of antioxidant capacity and phenolic compounds of berries, chokecherry and seabuckthorn. *Central European Journal of Biology, 4*(4), 499–506.
Makino, Y., Ichimura, M., Oshita, S., Kawagoe, Y., & Yamanaka, H. (2010). Estimation of oxygen uptake rate of tomato (*Lycopersicon esculentum* Mill.) fruits by artificial neural networks modelled using near-infrared spectral absorbance and fruit mass. *Food Chemistry, 121*(2), 533–539.
Næs, T., Isaksson, T., Fearn, T., & Davies, T. (2004). *A User-friendly Guide to Multivariate Calibration and Classification.* Charlton, Chichester, UK: NIR publications.
Nicolaï, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., et al. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biology and Technology, 46*(2), 99–118.
Nicolaï, B. M., Lötze, E., Peirs, A., Scheerlinck, N., & Theron, K. I. (2006). Non-destructive measurement of bitter pit in apple fruit using NIR hyperspectral imaging. *Postharvest Biology and Technology, 40*(1), 1–6.
Nicolaï, B. M., Verlinden, B. E., Desmet, M., Saevels, S., Saeys, W., Theron, K., et al. (2008). Time-resolved and continuous wave NIR reflectance spectroscopy to predict soluble solids content and firmness of pear. *Postharvest Biology and Technology, 47*(1), 68–74.
Sánchez, M.-T., De la Haba, M. J., Benítez-López, M., Fernández-Novales, J., Garrido-Varo, A., & Pérez-Marín, D. (2012). Non-destructive characterization and quality control of intact strawberries based on NIR spectral data. *Journal of Food Engineering, 110*(1), 102–108.
Saranwong, S., & Sornsrivichai, J. (2003). Performance of a portable near infrared instrument for Brix value determination of intact mango fruit. *Journal of Near Infrared Spectroscopy, 11*(3), 175–181.
Shao, Y., Zhao, C., Bao, Y., & He, Y. (2012). Quantification of nitrogen status in rice by least squares support vector machines and reflectance spectroscopy. *Food and Bioprocess Technology, 5*(1), 100–107.
Shü, Z., Chu, C., Hwang, L., & Shieh, C. (2001). Light, temperature, and sucrose affect color, diameter, and soluble solids of disks of wax apple fruit skin. *HortScience, 36*(2), 279–281.
Supapvanich, S., Pimsaga, J., & Srisujan, P. (2011). Physicochemical changes in fresh-cut wax apple (*Syzygium samarangense* [Blume] Merrill & L.M. Perry) during storage. *Food Chemistry, 127*(3), 912–917.
Xie, Y.-L., & Kalivas, J. H. (1997). Local prediction models by principal component regression. *Analytica Chimica Acta, 348*(1–3), 29–38.