pH Prediction of Perlis Sunshine Mango Using NIR Spectrometer

P S Woo¹ and M F Ibrahim*¹,²
¹School of Bioprocess Engineering, Universiti Malaysia Perlis, Perlis, Malaysia
²Centre of Excellence for Advance Sensor Technology, Universiti Malaysia Perlis, Perlis, Malaysia

*Corresponding author: firdausibrahim@unimap.edu.my

Abstract. To evaluate the various internal quality attributes of fruits, NIR spectroscopy techniques are undoubtedly quick and non-destructive tools. Acidity (pH) is one of the main quality attributes in mango fruits. Generally, it is being predicted in destructive way. The aim of this research was to develop calibration model and prediction of pH in Perlis Sunshine mangoes using NIR spectrometer. The transmission spectra of Sunshine mangoes were acquired in the wavelength range from 300 to 1000 nm. The effects of different types of pre-processing methods and spectra treatments, such as baseline correction, multiplicative scatter correction (MSC), Savitzky-Golay (SG) smoothing, second order derivative (SG) and normalisation were analyzed. The prediction models were developed by partial least squares (PLS) regression. The coefficient of determination (R²) of pH was 0.928 and the standard error of cross-validation (SECV) was 0.153. The results indicated that by using the NIR measurement system, in the suitable spectral range, it is possible to predict the pH of mango fruits by non-destructively.

1. Introduction
Mango (Mangifera indica L.) is abundantly available in subtropical and tropical countries. It carries a good price in the world market due to its delicious flavor, high nutritional value, low calories and excellent aroma [1]. Perlis Sunshine mango also known as ‘Sala’ mango is a type of mango which grows in Perlis, Malaysia. Normally, the length of Sunshine mango is around 12 – 15cm while the average weight is between 300 – 450 g. Sunshine mango is green in colour when it is unripe. Its colour turns to yellowish green when it is ripe. Fascinatingly, the taste for both ripe and unripe Sunshine mango is sour.

Nowadays, the agro-food industry is growing rapidly around the world. It is crucial to select fruit with the suitable degree of maturation in order to ensure the production of high-quality fruits. By measuring size, colour and firmness, maturity of mango could be predicted [2]. Various researchers have considered maturity from different point of views [3]. In mango fruits, acidity (pH) is one of the internal quality attributes used in determining the maturity. However, destructive methods are generally carried out to measure this internal quality. Hence, it is necessary to develop effective non-destructive methods so that the internal quality of fruits such as acidity (pH) can be measured.

Near-infrared (NIR) spectrometry is a rapid and non-destructive tool which has been widely used to evaluate various internal quality attributes of fruits [4]. The radiation of NIR spectrometry covers a spectra wavelength ranging from 780 to 2500 nm [5]. Compared to other conventional methods, NIR spectrometry offers decisive benefits. It involves simple sample preparation, fast analysis, chemical-free, pollution-free as well as can be carried out on-line [6, 7].
Partial least square (PLS) regression is the computation of the optimal least-squares fit to part of a correlation or covariance matrix. It is normally carried out to relate spectral data to quality attributes. PLS is quite identical to Principal Component Regression (PCR). Despite, it defines the latent variables (principal components) based on the covariance between the independent and dependent variables, rather than on the variance in the independent variables alone [8]. Hence, the aim of this research was to develop calibration model and prediction of pH in Perlis Sunshine mangoes using NIR spectrometer. The prediction models were developed by partial least squares (PLS) regression.

2. Experimental methods

2.1. Design of storage box

A simple storage box was designed to store Perlis Sunshine mango fruits under ambient condition. Temperature and humidity sensor (DHT22) were used to measure temperature and air humidity inside the storage box. The main purpose of designing the storage box was to control temperature and air humidity inside the storage box. This was because the temperature and humidity might affect the spectral data collection of NIR spectrometry [9]. Arduino is a microcontroller that was used to control the temperature and humidity inside the storage box. The bulb was fixed on top of the storage box. Its function was to act as the heat source. Two 12V DC fans were fixed at both sides of the storage box. The fans were used to control the temperature inside the storage box by taking cold air in and hot air out. The storage box was designed with the dimensions of 550 mm x 400 mm x 320 mm. Figure 1 shows the simple circuit diagram of mangoes storage box. Figure 2 shows the 3D drawing of mangoes storage box and Figure 3 shows the top view drawing of mangoes storage box.

![Figure 1. Circuit diagram of mangoes storage box.](image1)

![Figure 2. 3D drawing of mangoes storage box.](image2)

![Figure 3. Top view drawing of mangoes storage box.](image3)
2.2. Sample preparation
After the designing the storage box, a total of 120 unripe Sunshine mangoes were used. At first, 40 unripe mango samples were stored inside the storage box with temperature of 27 ± 3 °C and relative humidity of 65 ± 5 %. A total of three blocks replication for the sample preparation were carried out. 40 unripe mango samples were used for each replication.

2.3. NIR spectra acquisition
The acquisition of spectra was performed using a NIR spectrometer (Ocean Optics USB4000-XR1-ES), equipped with a 400 μm reflection probe fiber bundle (6 illumination fibers around 1 read fiber). The fiber optic variable attenuator was used to control the amount of light transmitted between two fibers. Tungsten halogen light beam (360 – 2000 nm) with power rating of 5-V/12-W was applied to the selected sampling area and then a stray-light protection cover was closed. It was used as lighting for this instrument. After that, a flexible rubber pad was fixed to the end of the probe as to prevent any light noise and light radiation leakage from outside.

The transmission spectra of mango samples were acquired in the range of 300 – 1000 nm with an optical resolution of 2 nm. Each day, a set of NIR reflectance spectra was acquired for every 5 mango samples. Three scans were taken at three spots in a single side of the mango in order to remove the spatial variability [10]. The readings were averaged to obtain one spectrum signal for each sample. The white reference measurement which has reflectivity above 98% for 250 – 1500 nm was obtained using Polytetrafluoroethylene (PTFE) Diffuse Reflectance Standard. This PTFE standard was specifically designed for reflection studies. Before carrying out the measurement process, the reference was taken only once.

2.4. Determination of actual pH value
After acquiring the spectra, the actual pH value of mango samples was determined using standard destructive measurement for reference values. Three portions of 15 mm width and 15 mm deep mango pulp were taken from three different spots which have been illuminated by the NIR radiation. The mango pulp was homogenized for 3 minutes using a small blender. After that, a piece of muslin cloth was used for filtration as to avoid any suspended solid on the juice. Then, a digital pH meter (Hanna HI model 2213, Padova, Italy), calibrated with pH 4.0 and 7.0 buffers was used to measure the actual pH value of mango juice and the average values were noted. After acquired the NIR spectra and actual pH value, the mango samples used were discarded. All mango samples were not used and examined repeatedly.

2.5. Spectra data analysis
Spectral data and actual pH value obtained were imported to Unscrambler (Version 10.3, Camo, Oslo, Norway) software to carry out multivariate analysis. Partial least squares regression (PLS) was used to develop the models for predicting the pH of mango samples. The wavelength ranges between 300 – 1000 nm were analyzed to select the best wavelength ranges that would obtain the best correlation between the spectral data and pH of mango samples.

Before carried out the modelling, it was necessary to pre-process spectral data in order to obtain stable and accurate calibration models [11]. Irrelevant information such as uncertainties, noise and unrecognized features were needed to be discarded. Several pre-processing techniques such as such as baseline correction, multiplicative scatter correction (MSC), Savitzky-Golay (SG) smoothing, second order derivative (SG) and normalisation were used to improve the predictability of PLS model. With the help of software, a few outlier samples were identified and eliminated for further improvement in the model.
3. Results and discussion

3.1. Spectral characteristics of Perlis Sunshine mango

Figure 4 shows the original NIR diffuse reflectance spectra ranged 300 – 1000 nm for 120 Perlis Sunshine mangoes. Among these mango samples, there were many cross-over points and significant number of overlapping with each other. Through visual inspection, the original spectra showed a homogenous trend without any outliers. The general outline of the absorption spectra for Sunshine mangoes was very much alike to that of other kinds of fruit materials such as lemon [12] and kiwi [13]. The selection of wavelength range was important in developing a strong calibration model. The spectral data which produced noise and little predictive ability was discarded. With original spectra, it was very hard to determine the relationship between NIR spectra characteristics and pH of the mango samples. Baseline drift and noise were existed in the NIR absorbance spectra in the wavelength range of 300 – 500 nm and 950 – 1000 nm. These features were quite common to be found in NIR spectra which collected by diffuse reflectance method. Thus, several pre-processing techniques were used in the analysis of NIR absorbance spectra.

![XY Graph](image)

**Figure 4.** Original NIR diffuse reflectance spectra ranged 300 - 1000 nm for 120 Perlis Sunshine mangoes.

Figure 5 shows the original NIR diffuse reflectance spectra ranged 300 – 1000 nm for 3 randomly selected Perlis Sunshine mangoes with low, medium and high pH value. Based on Figure 5, the NIR reflectance spectra taken from the mango sample with high pH value showed the most significant fluctuation compared to the mangoes with medium and low pH value. This showed that the relationship between NIR spectra characteristics and pH of Perlis Sunshine mangoes might have existed between each other. However, noises could be observed at the beginning (300 – 500 nm) and at the end of the spectral data (950 – 1000 nm). Thus, the wavelength data within these two spectral ranges were eliminated in the selection of wavelength group with the best performance so that the accuracy of the results can be enhanced.
Figure 5. Original NIR diffuse reflectance spectra ranged 300 - 1000 nm for 3 randomly selected Perlis Sunshine mangoes with low, medium and high pH values.

3.2. Selection of wavelength group with the best performance
The spectral range of (500 – 950 nm) was calibrated using partial least squares (PLS) regression. The models were developed in order to predict the pH of mango samples non-destructively. By using PLS regression, coefficient of determination ($R^2$) of the calibration and validation set were 0.880 and 0.663 respectively. $R^2$ value obtained for the calibration set was acceptable, however, for the validation set was quite low. Table 1 shows the statistical parameters and results of PLS model for pH prediction using 500 – 950 nm wavelength range.

| Wavelength range, nm | Calibration | Validation |
|----------------------|-------------|------------|
|                      | SEC | $R^2$ | SECV | $R^2$ |
| PLS model            | 0.096 | 0.880 | 0.163 | 0.663 |

Due to the low $R^2$ value obtained for the validation set, the models were then tested using a smaller wavelength range. The number of wavelengths can be minimized by selecting the wavelength group with the best performance in order to increase the model’s accuracy and reduce errors of the results [14]. The wavelength range of 500 – 950 nm was divided into two sets. The first set wavelength range was started from 500 to 699 nm while the second set was started from 700 to 950 nm. These wavelength range was separated based on the visible spectrum (500 – 699 nm) and NIR spectrum (700 – 950 nm). Table 2 shows the statistical parameters and results of PLS model for pH prediction using different NIR wavelength range. Figure 6 and Figure 7 show the scatter plot of pH using PLS model in the wavelength range of 500 – 699 nm and 700 – 950 nm.

| Wavelength range, nm | Calibration | Validation |
|----------------------|-------------|------------|
|                      | SEC | $R^2$ | SECV | $R^2$ |
| PLS model            |      |      |      |      |
| 500 - 950            | 0.096 | 0.880 | 0.163 | 0.663 |
| Wavelength Range | $R^2$ Calibration | $R^2$ Validation | SEC | SECV |
|------------------|------------------|------------------|-----|------|
| 500 - 699 nm     | 0.089            | 0.897            | 0.184 | 0.571 |
| 700 - 950 nm     | 0.075            | 0.928            | 0.153 | 0.705 |

**Figure 6.** Scatter plot of pH using PLS model in the wavelength range of 500 - 699 nm.

**Figure 7.** Scatter plot of pH using PLS model in the wavelength range of 700 - 950 nm.

For wavelength range of 700 – 950 nm, the $R^2$ value for the calibration and validation set using PLS model were 0.928 and 0.705 respectively. Based on Table 2, the result indicated that the PLS model tested using the wavelength range of 700 – 950 nm was more suitable to predict the pH of mango samples compared to the wavelength range of 500 – 699 nm. This was because the wavelength range of 700 – 950 nm produced better statistical results. The results showed that the values of SEC and SECV decreased with an increasing of $R^2$ values for both calibration and validation set.

Based on Figure 6 and 7, the scatter plot of pH using PLS model in the wavelength range of 700 – 950 nm showed stronger correlation between the predicted and measured pH compared to the wavelength range of 500 – 699 nm. This was due to fewer outliers were detected in the scatter plot of pH using the wavelength range of 700 – 950 nm. The truncation of wavelength tested from 700 – 950 nm was found to improve the results of PLS model. Thus, the results proved that the performance of PLS model can be enhanced by selecting the wavelength group with the best performance. Same improvement was also found in the pH prediction of mangoes from seven different mango cultivars [14].

The removal of visible wavelength region of 500 – 699 nm proved that there was no significant correlation existed between the skin colour and pH of mango samples. Based on Figure 4, the peak in the NIR diffuse reflectance spectra indicated that there were strong reflectance characteristics existed in
the wavelength range of 700 – 950 nm. The peak was used to create relationship between the NIR spectra and pH of mango samples. Thus, the NIR wavelength range of 700 – 950 nm was used to develop the next PLS model by using different pre-processing methods. Similar wavelength range (700 – 1100 nm) was also being used by [15] to determine the acidity (pH) of mango ‘Nam Dokmai’.

3.3. Prediction of pH using different pre-processing methods

There were several pre-processing methods that can be processed together with PLS model such as baseline correction, multiplicative scatter correction (MSC), Savitzky-Golay (SG) smoothing, second order derivative (SG) and normalisation [14]. These pre-processing methods were usually used to remove the irrelevant information from spectra which derived from unknown sources such as irregular surface morphology as well as varied distance between sample and detector [16]. Table 3 shows the results of PLS model with different pre-processing methods for pH prediction using the wavelength range of 700 – 950 nm.

Table 3. Results of PLS model with different pre-processing methods for pH prediction using the wavelength range of 700 - 950 nm.

| Pre-processing method                        | No. of PLS factor | Calibration | Validation |
|---------------------------------------------|-------------------|-------------|------------|
|                                            |                   | SEC        | R²         | SECV       | R²         |
| PLS model (700 - 950 nm)                    |                   |            |            |            |            |
| No data treatment                           | 5                 | 0.114      | 0.833      | 0.169      | 0.640      |
|                                             | 6                 | 0.096      | 0.880      | 0.160      | 0.677      |
|                                             | 7                 | 0.075      | 0.928      | 0.153      | 0.705      |
| Baseline correction                         | 5                 | 0.116      | 0.827      | 0.175      | 0.617      |
|                                             | 6                 | 0.101      | 0.867      | 0.167      | 0.652      |
|                                             | 7                 | 0.074      | 0.929      | 0.161      | 0.677      |
| Multiplicative scatter correction (MSC)     | 5                 | 0.081      | 0.916      | 0.177      | 0.597      |
|                                             | 6                 | 0.065      | 0.945      | 0.175      | 0.606      |
|                                             | 7                 | 0.053      | 0.964      | 0.175      | 0.608      |
| Smoothing (SG)                              | 5                 | 0.135      | 0.763      | 0.182      | 0.577      |
|                                             | 6                 | 0.104      | 0.861      | 0.161      | 0.668      |
|                                             | 7                 | 0.090      | 0.894      | 0.158      | 0.680      |
| 2nd order derivative (SG)                   | 5                 | 0.078      | 0.921      | 0.187      | 0.560      |
|                                             | 6                 | 0.067      | 0.939      | 0.190      | 0.546      |
|                                             | 7                 | 0.055      | 0.961      | 0.198      | 0.507      |
| Normalisation                               | 5                 | 0.102      | 0.866      | 0.172      | 0.614      |
|                                             | 6                 | 0.081      | 0.916      | 0.172      | 0.616      |
|                                             | 7                 | 0.068      | 0.941      | 0.171      | 0.620      |

According to [17], the suitable number of regression factor to be used for the PLS model was selected based on the minimum standard error of cross-validation (SECV). Based on Table 3, the PLS model with no data treatment showed the minimum SECV which was 0.153. Thus, the performance of PLS model processed with other types of pre-processing methods were evaluated based on the regression factor of 7.
For the calibration set, PLS model with regression factor of 7 treated with MSC pre-processing method showed the best correlation between the predicted and measured pH. The $R^2$ value obtained was 0.964 and SEC was 0.053. The results showed that SEC decreased with an increasing of $R^2$. For baseline correction, MSC, second order derivative (SG) and normalisation, the $R^2$ values obtained for the calibration set were slightly higher than the PLS model with no data treatment. However, the $R^2$ values obtained for the validation set by using these four types of pre-processing methods were all lower than the PLS model with no data treatment. Low $R^2$ values indicated that a weak correlation was existed between the predicted and measured pH of mango samples.

For the validation set, the PLS model with no data treatment showed the lowest SECV and highest $R^2$ value compared to the other types of pre-processing methods. The SECV obtained was 0.153 and $R^2$ value was 0.705. Besides, the value of SECV and SEC obtained for the PLS model with no data treatment was closed to each other. This result indicated that the loss in the accuracy was small. A low standard error indicated that the PLS model with no data treatment showed reasonably accurate results in predicting the pH of mango samples. [18] discovered that PLS model was suitable to be used for pH prediction of mangoes due to good calibration results were found in their research.

The application of different types of pre-processing methods such as baseline correction, MSC, smoothing (SG), second order derivative (SG) and normalisation on the PLS model were found to be less or no effect in improving the predictability of the calibration and validation set of mango samples. Similar results were also reported by [14] and [12] for the pH prediction of mango and lemon samples using NIR spectroscopy.

Figure 7 shows the scatter plot of pH using PLS model with no data treatment in the wavelength range of 700 – 950 nm. There was a strong positive correlation existed between the predicted and measured pH of mango samples. From the scatter plot, most of the data points were closed to the target line and the slope of the line fitted to the data points was closed to 45°. This indicated that the predicted and measured pH were highly correlated with each other.

4. Conclusion
PLS regression was used to determine the correlation between the measured and predicted pH of mango samples. The results showed that the PLS model with no data treatment in the wavelength range of 700 – 950 nm produced the best regression coefficient values compared to the other types of pre-processing methods. The application of several types of pre-processing methods such as baseline correction, MSC, smoothing (SG), second order derivative (SG) and normalisation on PLS model were found to be less or no effect in improving the predictability of the calibration and validation set of mango samples. The results indicated that PLS model with no data treatment in the wavelength range of 700 – 950 nm yielded good calibration and validation results. The $R^2$ value obtained for calibration set was 0.928 and validation set was 0.705. There was a strong correlation existed between the predicted pH and measured pH of mango samples. Therefore, it is concluded that by using the NIR measurement system, in the suitable spectral range, it is possible to predict the pH of mango fruits by non-destructively.

Acknowledgement
The authors acknowledge financial support from Universiti Malaysia Perlis (UniMAP) through 9001-00582 Internal Social Project Via Research (INSPIRE) 2019.

References
[1] Schmilovitch Z, Mizrach A, Hoffman A, Egozi H and Fuchs Y 2000 Determination of mango physiological indices by near-infrared spectrometry Postharvest Biology and Technology 19 245-52
[2] Jha S, Kingsly A and Chopra S 2006 Physical and mechanical properties of mango during growth and storage for determination of maturity Journal of Food engineering 72 73-6
[3] Peacock B 1986 Influence of Harvest Maturity of Mangoes on Storage Potencial and Ripe Fruit Quality
[4] Costa RC and de Lima KM 2013 Prediction of parameters (soluble solid and pH) in intact plum using NIR spectroscopy and wavelength selection *Journal of the Brazilian Chemical Society* **24** 1351-6

[5] Sheppard N, Willis H and Rigg J 1985 Names, symbols, definitions and units of quantities in optical spectroscopy (Recommendations 1984) *Pure and Applied Chemistry* **57** 105-20

[6] Saleh B 2012 Biochemical and genetic variation of some Syrian wheat varieties using NIR, RAPD and AFLPs techniques *Journal of Plant Biology Research* **1** 1-11

[7] Pissard A, Baeten V, Romnée J-M, Dupont P, Mouteau A and Lateur M 2012 Classical and NIR measurements of the quality and nutritional parameters of apples: a methodological study of intra-fruit variability *BASE*

[8] Wold S, Sjöström M, Eriksson L 2001 PLS-regression; a basic tool of chemometrics. Chemometrics and intelligent laboratory systems **58** 109-30

[9] Burns DA and Ciurczak EW 2001 *Handbook of Near-Infrared Analysis* (CRC press)

[10] Blahovec J and Kutílek M 2003 *Physical Methods in Agriculture: Approach to Precision and Quality* (Springer Science & Business Media)

[11] Cen H and He Y 2007 Theory and application of near infrared reflectance spectroscopy in determination of food quality *Trends in Food Science & Technology* **18** 72-83

[12] Reddy NS, Nivetha D and Yadav B 2016 Non-destructive quality assessment of citrus fruits using ft-near-infrared spectroscopy *International Journal of Science, Environment and Technology* **5** 1850-60

[13] Lee JS, Kim S-C, Seong KC, Kim C-H, Um YC and Lee S-K 2012 Quality prediction of kiwifruit based on near infrared spectroscopy *Korean Journal of Horticultural Science and Technology* **30** 709-17

[14] Jha SN, Jaiswal P, Narsaiah K, Gupta M, Bhardwaj R and Singh AK 2012 Non-destructive prediction of sweetness of intact mango using near infrared spectroscopy *Scientia Horticulturae* **138** 171-5

[15] Rungpichayapichet P, Mahayothee B, Nagle M, Khuwijitjaru P and Müller J 2016 Robust NIRS models for non-destructive prediction of postharvest fruit ripeness and quality in mango *Postharvest Biology and Technology* **111** 31-40

[16] Lu R 2001 Predicting firmness and sugar content of sweet cherries using near–infrared diffuse reflectance spectroscopy *Transactions of the ASAE* **44** 1265

[17] Moghimi A, Aghkhani MH, Sazgarnia A and Sarmad M 2010 Vis/NIR spectroscopy and chemometrics for the prediction of soluble solids content and acidity (pH) of kiwifruit. *Biosystems Engineering* **106** 295-302

[18] Purwanto YA, Zainal PW, Ahmad U, Sutrisno M, Makino Y, Oshita S, Kawagoe Y, Kuroki S 2013 Non destructive prediction of pH in mango fruit cv. Gedong Gincu using NIR spectroscopy *International Journal of Engineering and Technology IJET-IJENS* **13** 70-3