EASR

Engineering and Applied Science Research

https://www.tci-thaijo.org/index.php/easr/index

Published by the Faculty of Engineering, Khon Kaen University, Thailand

# Wavelengths selection based on genetic algorithm (GA) and successive projections algorithms (SPA) combine with PLS regression for determination the soluble solids content in Nam-DokMai mangoes based on near infrared spectroscopy

Kanvisit Maraphum[1, 2], Artjima Ounkaew[3, 4], Pornnapa Kasemsiri[3, 4], Salim Hiziroglu[5] and Jetsada Posom*[1, 2]

[1]Department of Agricultural Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen 40002, Thailand
[2]Applied Engineering for Important Crops of the North East Research Group, Department of Agricultural Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen 40002, Thailand
[3]Department of Chemical Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen 40002, Thailand
[4]Sustainable Infrastructure Research and Development Center, Faculty of Engineering, Khon Kaen University, Khon Kaen 40002, Thailand
[5]Department of Natural Resource Ecology and Management, Oklahoma State University, 303-G Agricultural Hall, Stillwater, OK 74078, USA

Received 13 March 2021
Revised 7 May 2021
Accepted 19 May 2021

**Abstract**

The objective of this work was to search for an optimal wavelength selection for near infrared (NIR) spectroscopy for quality measurement of Nam-Dokmai mangoes. In this study, NIR spectroscopy has been applied to grading management systems for commercial mangoes export. Near infrared spectra were collected using a near infrared instrument incorporating a wavelength region of 860-1760 nm. Genetic algorithm (GA) and successive projections algorithms (SPA) was employed for selecting the spectra wavelengths. The selected wavelengths were also used to generate the prediction models via partial least square (PLS) regression. The optimal pretreatment was obtained from the second derivative. The model of full wavelengths rendered effective the best performance with $r^2$ of 0.66-0.74, RMSEP of 0.72-0.80 °Brix and RPD equal to 1.8-2.0. The SPA-PLS resulted in values of $r^2$, RMSEP and RPD were 0.43-0.70, 0.77-1.01°Brix and 1.4-1.9, respectively. Meanwhile, the result of GA-PLS performed efficiency with $r^2$, RMSEP and RPD were 0.52-0.72, 0.74-0.96°Brix and 1.5-1.9, respectively. The outcome, the GA-PLS model (50 variables) is suitable for use in the measuring soluble solids content (SSC) in mangoes. This model could be used as screening purpose. It also was not different significantly when compared to the best model. Hence, authors suggested that the prediction model by GA-PLS with 50 variables can be effectively used for evaluating SSC in mangoes.

**Keywords:** Mango, Wavelengths selection, Brix, PLS

## 1. Introduction

Internal quality of mango plays an important role on market price and consumer satisfaction. Higher quality induces a customer's willingness to pay premium prices for a product. Nam Dokmai mango or golden mango is popular fruit larger fruit than all other mango varieties in Thailand. Each fruit weighs from 250 to 550 grams, whereas other varieties may be only 150 to 250 grams [1, 2]. Moreover, the internal quality, sweetness in the ripe stage of Nam Dokmai mango is around 17-20 °Brix reading. While the other varieties have corresponding value of 15-16 °Brix [1]. In the year 2006, Thailand has a plantation area of 0.28 Mha and 2.2 Mton productions. However, they are mainly produced for domestic consumption, although about 10% of market are being exported in fresh condition [3]. The properties of mango to be exported include skin colour, size ripe stage and most importantly, sweetness [2]. However, the ripe mangoes and ripe impulsive, look similar each other in colour and size. Mango is a climacteric fruit, their quality properties change rapidly which can affect the purchasing decisions of consumer. Sombatpraiwan et al. [4] studied the factors effected to ripe stages of Nam Dokmai mango after harvesting and classifying. Several factors were studied including specific gravity (SG), firmness, peel colour, total soluble solids (TSS), titratable acidity (TA) and the ratio of betwee TSS and TA (TSS/TA), respectively. The results shown that using fruit firmness and TSS/TA could be used to classify the ripening stages of mangoes, including unripe, ripe, and over ripe with accuracy of 91.7%. Hence, quality grading during packing is quite difficult.

Near infrared (NIR) spectroscopy has been used for evaluating the chemical characteristics of food and agricultural products [5]. Due to advantages of NIR spectroscopy including the rapid measurement, repeatability, fact that multiple attributes i.e. soluble solids content (SSC), pH and acid levels can be measured simultaneously, and non-destructively [6]. At present, the NIR spectroscopy have been used for different purposes such as breeding programme [7, 8] and products guarantee [9-12], respectively. Taira et al. [13] used the spectrometer wavelength of 600-1000 nm to evaluate the SSC of Iwarin mango. They obtained $r^2$ and RMSECV were 0.76 and

*Corresponding author. Tel.: +666 1773 9618
Email address: Jetspo@kku.ac.th
doi: 10.14456/easr.2022.14

0.70 °Brix. Jha et al. [14] also applied the NIR spectroscopic method with visual wavelength (400-700 nm) to determine the sweetness of mangoes (cv. Dashehari). The resulting showed that correlation (r) and SEP were 0.90 and 1.97 °Brix. Both reports above show that the model developed in those using the vis-NIR range of wavelengths revealed as good potential for evaluating the internal quality of mango non-destructively.

However, according to the above research, there is no selection of wavelengths for modelling. If the relevant wavelength is selected for the quality measurement, it might be able to increase the accuracy of the model and it is also possible to reduce the cost of sensors (NIR filter) for building instrument. According to previous report, Pitak et al. [15] applied the wavelengths selection of successive projections algorithms (SPA) and genetic algorithms (GA) to develop the elemental composition model of biomass pellet. Selected wavelength method reduces the independent variable number which consist of the most significant spectra [16]. Therefore, wavelength selection is still needed before build calibration.

The purpose of this experiment was to use the NIR spectroscopy method to measure and monitor the internal quality change in Nam Dokmai mango from the unripe stage to ripe stage. This information is directly related to mango maturity change, which has the potential to inform precision agriculture and storage management. It is expected data from this looks should help mango produces to grade overall fruit quality prior export.

## 2. Materials and methods

### *2.1 Sample preparation*

Different Twenty-two 'Nam DokMai' mangoes were randomly purchased from a local store for experiment. Samples were taken from 22 fully ripe fruits for the tests within 1 day of collection. Before they were held in the room temperature for 1 hour before the experiment were taken.

### *2.2 NIR Spectral acquisition*

Before spectral data scan, the NIR spectrometer (AvaSpec-NIR256/512-1.7-EVOAvantes BV, Netherlands) across wavelength range of 860 – 1760 nm was calibrated as reference using a Teflon plate. An integration time should be specified to achieve optimum system sensitivity. Based on our chamber environment, the integration time was set to 12.3 ms, yielding approximately 90% full-scale Analog-to-Digital Converter (ADC) of the reference material reflectance. Figure 1 shows a scanning measurement position, each fruit was two sides; front side was sanned 4- positions and backside was scanned 4 positions. The total spectra was 8-positions per fruit. Each position was scanned four times. Then, the spectrum of each position was computed by averaging the four successive scans. Therefore, each mango collected 8 spectra.

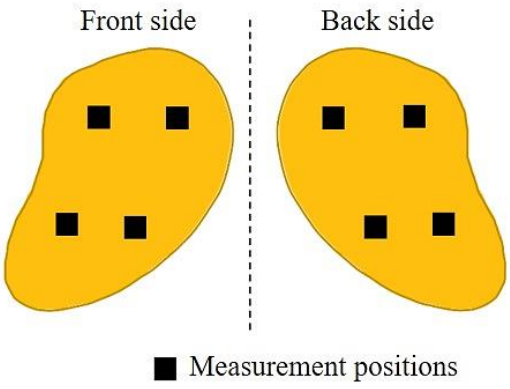


**Figure 1** Measurement positions of NIR spectra and cutting area for measurement SSC of mangoes Nam Dokmai.

The NIR spectral data of each sample were recorded in absorbance (log1/R) unit and its were saved to CSV format file. Then spectral data were extracted for mathematical processing. Principle component analysis (PCA) was used to screening outliers' spectra. Moreover, the spectral data were checked both of repeatability (Rep) and reproducibility (Repro). The Repro of spectral data was calculated as the standard deviation of absorbance values, when the sample was scanned with re-loading and re-scanning for 10 times in the same position [17]. Meanwhile, Rep of spectra was the standard deviation of 10 spectra when the sample was scanned with re-scanning. The samples number 6, 12 and 18 were selected randomly for scanning to get Rep and Repro, respectively.

### *2.3 Analysis of physico-chemical properties of mango fruit*

After scanning the NIR spectra, each scanned area was cut with a width of ~20 mm and a depth of ~10 mm and immediately squeezed juice. Then, the SSC values were determined by dropping juice into the test hole of a portable refractometer (Nar-3Ta, ATAGO, Tokyo, Japan). The reference value was done duplicate. Table 1 displays data of the calibration set and validation sets.

**Table 1** Statistical of the calibration set and validation set of Nam Dokmai mango

| Set | No | Max | Min | Mean | SD |
|---|---|---|---|---|---|
| Calibration set | 129 | 22.6 | 15.6 | 18.18 | 1.15 |
| Validation set | 44 | 21.9 | 16.3 | 18.56 | 1.42 |

For determinaning precision of laboratory, the repeatability (Rep) of reference method was presented. The Rep was the standard deviation of different value between duplicate. Rep was used to calculate the maximum coefficient of determination ($R^2_{max}$). Rep and $R^2$max were calculated as follows [18].

$$\mathrm{Rep} = \sqrt{\frac{\sum_1^N \left( |e| - \frac{\sum_1^N |e|}{N} \right)^2}{N-1}} \tag{1}$$

$$R^2_{max} = \frac{SD_y^2 - Rep^2}{SD_y^2} \tag{2}$$

When N is the sample amount, e is the absolute difference between the duplicate value.

Also, the outliers of SSC were detected by using equation 3. The information of SSC values which obtained used for multivariate analysis are provided in Table 1.

$$\frac{Y_i - \overline{Y}}{SD} \geq \pm 3 \tag{3}$$

Where $Y_i$ is SSC of sub-sample, $\overline{Y}$ is mean and SD is the standard deviation of reference data. If the sample beyond the range of -3 to +3, it would be removed.

*2.4 Multivariate analysis*

All samples were sorted in ascending order according to SSC value, and then samples were separated into calibration and validation sets, where 75% of total sample was calibration set and 25% was validation set, respectively. The maximum value and minimum value was assigned as calibration set. Due to spectra were contained a lot of noise and baseline offset effect, so then before making the prediction models, NIR spectra were pre-treated via several techniques which consist of SNV, first derivative (D1) and second derivative (D2), both derivatives were pre-processed by used Norris gab and segment were 7 and 7, respectively. Either full wavelength or selected wavelength were utilized for model creation. To discover the best variables used for evaluating the SSC of mangoes, the successive projections algorithms (SPA) and genetic algorithms (GA) were also used in this experiment. SPA is the most widely used for selected wavelengths before multivariate analysis. SPA and GA is a flexible technique for variable selection in multivariate calibration [19]. It is a forward selection method which starts with one wavelength and incorporates a new one at each iteration, yielding wavelengths whose information content is minimally redundant. Genetic Algorithms (GA) is a mathematical model inspired by Charles Darwin's idea of natural selection, and GA also is a relatively new optimization technique which can be used as an alternative method for variable selection.

After wavelengths were selected, these were performed to generate models via the partial least square (PLS) regression via MATLAB program (MathWorks, Natick, MA, USA). Furthermore, the calibration set was used to generate by PLS regression with leave-one-out full cross-validation for determining the PLS factor. The factor of PLS regression was selected from the minimized standard error of cross-validation (RMSECV) [19]. The performance of model for predicting are used as indexes of coefficients of determination of calibration set ($R^2$), root mean square error of calibration (RMSEC), coefficients of determination of validation set ($r^2$), root mean square error of prediction (RMSEP) and ratio of performance to standard deviation (RPD), respectively. These indicators can be calculated as equations per below.

$$R^2, r^2 = 1 - \frac{\sum_{i=1}^N \left(Y_i - \hat{Y}\right)^2}{\sum_{i=1}^N (Y_i - \overline{Y})^2} \tag{4}$$

$$\mathrm{RMSEC, RMSEP} = \sqrt{\frac{\sum_{i=1}^N \left(Y_i - \hat{Y}\right)^2}{N}} \tag{5}$$

$$\mathrm{RPD} = \frac{SD_y}{\mathrm{RMSEP}} \tag{6}$$

Where $Y_i$ is reference data, $\hat{Y}$ is predicted value, N is sample number and $SD_y$ is standard deviation of the reference value, which obtained from the validation set.

The RPD has been used for determining performance of NIR model. [20]. The model is reliable if PRD ≥ 2, a fair model if 1.4≤ RPD < 2, and a non-reliable model if <1.4 [21], respectively. The RPD in of Chang et al. [21]'s scale is suitable for directly scanning in the inhomogeneous materials. Many reports in the field of near-infrared spectroscopy and hyper-spectroscopy such as Wang et al. [22], Ncama et al. [23] and Maraphum et al. applied portable NIR spectrometer to determine the quality of agronomy such as sugarcane stalk and cassava tubers [7, 24]. Moreover, Maraphum et al. [25] also applied the portable hyperspectral camera to visualize the brix distribution in sugarcane stalks. These reports from above indicate that RPD in this scale accepted use in several the measurement field.

## 3. Results and discussion

### *3.1 Characteristics of NIR spectral data*

Figure 2(a) and 2(b) shows that the raw NIR spectra and second derivative in the range of 860- 1760 nm. The obvious peak displayed at the wavelength of 910, 1200, 1450 nm which related to sugar [10] water [26]. These results were consistent with those of a previous report in which the peak of 970 nm was assigned to the second overtone of the O-H stretching mode [26].

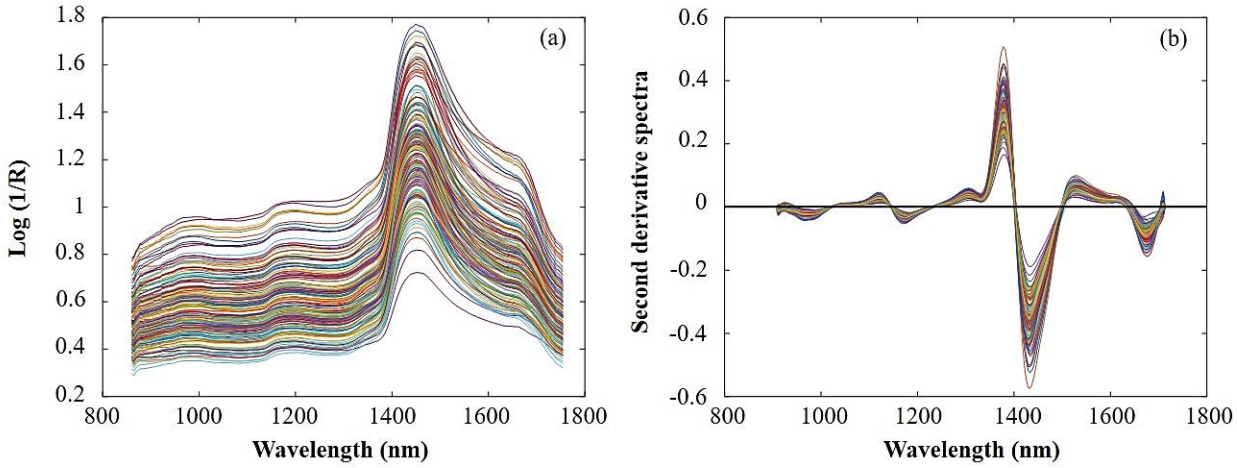


**Figure 2** NIR Spectra of mango in the wavelength range of 860-1760 nm. (a) Raw spectra (b) second derivative spectra.

### *3.2 Calibration models on optimal wavelengths*

Table 2 demonstrates repeatability, reproducibility, and $R^2_{max}$ of soluble solid content (SSC), which rendered an $R^2_{max}$ of 0.81. Table 3 illustrates the repeatability and reproducibility of absorption at 970 nm, 1200 nm, and 1450 nm of Nam Dokmai mango. The mean of absorption in repeatability was nearly the same as the reproducibility, but the standard deviation or the reproducibility amounted to approximately four times than the repeatability.

Table 4 displayed the predictions of the SSC models of mangoes, the results were calculated from a different algorithm, which consisting of full-length PLS, SPA-PLS and GA-PLS, respectively. For the results for the full lengths combined by PLS regression method which gave $r^2$ were 0.66-0.74, RMSEP were 0.72-0.80 °Brix and RPD were 1.8-2.0, respectively. The results of PLS model which developed from wavelength selection by SPA, those wavelengths were selected which separate into three categories including of 100, 50 and 25 wavelengths, respectively. Each category was performed models via PLS method, and they provided $r^2$ were 0.38-0.70, RMSEP were 0.77-1.11°Brix and RPD were 1.3-1.9. The best model of GA-PLS provided the highest $r^2$ and RPD; and lowest RMSEP, which were 0.72, 2 and 0.74°Brix, respectively. This model were developed using pre-processed spectra of the second derivative. However, full-length PLS model was also provided with a slight performance, which gave $r^2$, RPD and RMSEP and RPD values of 0.74, 2.0 and 0.72 °Brix, respectively.

**Table 2** Repeatability, reproducibility and $R^2_{max}$ of SSC in mango samples.

| Parameter | Repeatability | | Reproducibility | | $R^2_{max}$ |
|---|---|---|---|---|---|
| | Mean of the different of duplicate | Standard deviation of the different of duplicate | Mean of the different of duplicate | Standard deviation of the different of duplicate | |
| SSC | 0.32 | 0.65 | 0.2 | 0.73 | 0.81 |

SSC: Soluble solids content of mango

**Table 3** Repeatability and reproducibility of absorption at 970 nm, 1200 nm, and 1450 nm of Nam-Dokmai mangoes of sample numbers 6, 12 and 18.

| Fruit number | Repeatability | | | Reproducibility | |
|---|---|---|---|---|---|
| | Absorption value | Mean of absorption | Repeatability | Mean of absorption | Reproducibility |
| 6 | At 970 nm | 0.660 | 0.00137 | 0.668 | 0.04382 |
| | At 1200 nm | 0.708 | 0.00156 | 0.714 | 0.00406 |
| | At 1450 nm | 1.374 | 0.00119 | 1.318 | 0.01041 |
| | Average | 0.914 | 0.00138 | 0.900 | 0.01943 |
| 12 | At 970 nm | 0.523 | 0.00152 | 0.604 | 0.00851 |
| | At 1200 nm | 0.548 | 0.00161 | 0.649 | 0.00969 |
| | At 1450 nm | 1.165 | 0.00308 | 1.219 | 0.04335 |
| | Average | 0.745 | 0.00207 | 0.824 | 0.02051 |
| 18 | At 970 nm | 0.486 | 0.00069 | 0.451 | 0.00102 |
| | At 1200 nm | 0.532 | 0.00102 | 0.586 | 0.00126 |
| | At 1450 nm | 0.973 | 0.00218 | 0.940 | 0.01018 |
| | Average | 0.664 | 0.00130 | 0.659 | 0.00415 |

**Table 4** PLS Regression result of Nam-Dokmai mango

| Algorithms | Wavelengths | Pre-treatment | LVs | $R^2$ | RMSEC | $r^2$ | RMSEP | RPD |
|---|---|---|---|---|---|---|---|---|
| PLS | Full | Raw | 9 | 0.54 | 0.79 | 0.66 | 0.80 | 1.8 |
| | Full | SNV | 9 | 0.54 | 0.78 | 0.69 | 0.78 | 1.8 |
| | Full | D1 | 10 | 0.59 | 0.74 | 0.70 | 0.77 | 1.8 |
| | Full | D2 | 9 | 0.60 | 0.73 | 0.74 | 0.72 | 2.0 |
| SPA-PLS | 100 | Raw | 9 | 0.57 | 0.76 | 0.70 | 0.77 | 1.9 |
| | 100 | SNV | 8 | 0.46 | 0.85 | 0.56 | 0.88 | 1.6 |
| | 50 | Raw | 8 | 0.59 | 0.74 | 0.58 | 0.90 | 1.6 |
| | 50 | SNV | 7 | 0.43 | 0.87 | 0.61 | 0.83 | 1.7 |
| | 50 | D1 | 8 | 0.38 | 0.91 | 0.54 | 0.97 | 1.5 |
| | 50 | D2 | 6 | 0.48 | 0.83 | 0.48 | 1.02 | 1.4 |
| | 25 | Raw | 7 | 0.53 | 0.79 | 0.43 | 1.01 | 1.4 |
| | 25 | SNV | 7 | 0.54 | 0.79 | 0.53 | 0.94 | 1.5 |
| | 25 | D1 | 6 | 0.34 | 0.94 | 0.52 | 0.98 | 1.5 |
| | 25 | D2 | 7 | 0.41 | 0.89 | 0.38 | 1.11 | 1.3 |
| GA-PLS | 100 | Raw | 9 | 0.54 | 0.78 | 0.67 | 0.79 | 1.8 |
| | 100 | SNV | 9 | 0.49 | 0.83 | 0.65 | 0.81 | 1.8 |
| | 100 | D1 | 10 | 0.56 | 0.77 | 0.66 | 0.82 | 1.7 |
| | 100 | D2 | 9 | 0.55 | 0.78 | 0.72 | 0.74 | 2.0 |
| | 50 | Raw | 9 | 0.53 | 0.80 | 0.62 | 0.83 | 1.7 |
| | 50 | SNV | 9 | 0.45 | 0.86 | 0.61 | 0.87 | 1.6 |
| | 50 | D1 | 10 | 0.54 | 0.79 | 0.61 | 0.86 | 1.7 |
| | 50 | D2 | 9 | 0.54 | 0.78 | 0.70 | 0.76 | 1.9 |
| | 25 | Raw | 9 | 0.41 | 0.89 | 0.58 | 0.85 | 1.7 |
| | 25 | SNV | 9 | 0.45 | 0.86 | 0.57 | 0.90 | 1.6 |
| | 25 | D1 | 6 | 0.38 | 0.91 | 0.52 | 0.96 | 1.5 |
| | 25 | D2 | 8 | 0.43 | 0.88 | 0.58 | 0.92 | 1.6 |

LVs: Latent variable of PLS; $R^2$: Coefficient of determination of calibration set; RMSEC: root mean square error of calibration; $r^2$: Coefficient of determination of validation set.

After the calibration models were developed, it was used to evaluate their performance. Almost model have range of RPD were 1.4-2.0 which indicated model are fair [21]. The best model is the full-length which developed from second derivative resulting in $r^2$, RMSEP and RPD values of 0.74, 0.72 °Brix and 2.0, respectively. However, a few models in Table 4 can provide RPD closely the best model, i.e. PLS-GA with 100 wavelengths which pre-treated by second derivative and 50 wavelengths pre-treated by the second derivative. They gave RMSEP are 0.74-0.76 Brix and both models gave RPD was 1.9, respectively. These models were tested for a significant difference at 95% confidence level according to the ANOVA. The results indicate that these three models were not significantly different from each others. Figure 3 illustrates the spectral data pre-treated by D2 and wavelength selection via GA method also showing the band occurring for model development for SSC mango evaluation.

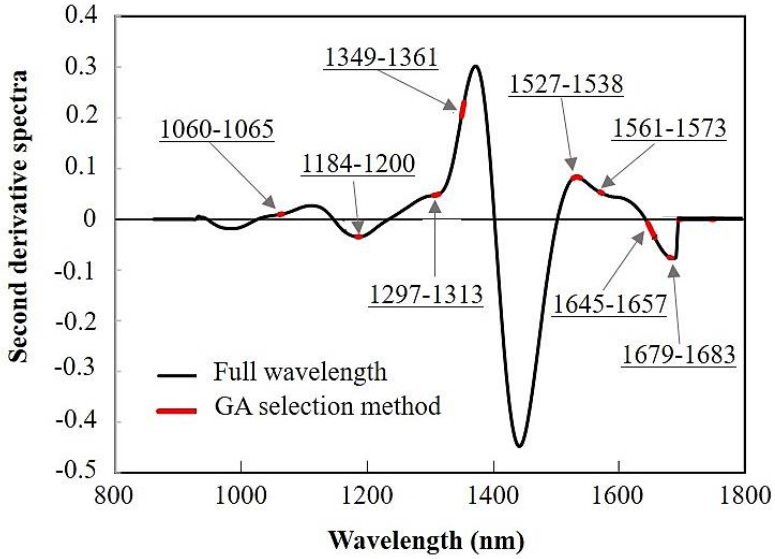


**Figure 3** Wavelength pre-treated by second derivative technique and interval wavelength selected by GA algorithms.

The reflectance bands of second derivative spectra selected for model development by GA is shown in Table 5. It appears that the model developed with D2 spectra of GA selected wavelength with 50 variables was more suitable than the model developed with D2 spectra of full wavelength having 256 variables. Figure 4 shows that the scatter plot between actual values which obtained from the refractometer and predicted values obtained from the GA-PLS model.

**Table 5** The absorbance bands with second derivative spectra selected by GA algorithms.

| Wavelength (nm) selection by GA | Wavelength referred from Ref. | Bond Vibration | Structure |
|---|---|---|---|
| 1060-1065 | 1060 [27] | N-H str. second overtone | $RNH_2$ |
| 1184-1200 | 1194 [27] | C-H methyl C-H, ($CH_3$) | $CH_3$, aromatic C-H |
| | 1195 [27] | C-H methyl ($CH_3$) | C-H, C-$H_3$ |
| 1297-1313 | - | | |
| 1349-1361 | 1360 [27] | 2× C-H str. + C-H def | $CH_3$ |
| 1527-1538 | 1530 [28] | C-H methyne C-H as (R-C-C≡C-H) | C-H, methyne (1-hexyne) |
| 1561-1573 | 1570 [28] | N-H amide. NH | N-H, CONHR |
| 1645-1657 | 1654 [28] | C-H methyl C-H, nitro ($CH_3NO_2$) | C-H methyl C-H, |
| 1679-1683 | 1680 [28] | C-H aromatic (ArCH) | C-H, aromatic C-H |
| | 1682 [28] | C-H methyl C-H, carbonyl associated as one C removed (C=$OCH_2CH_3$) | C-H methyl |
| 1698 | - | - | - |

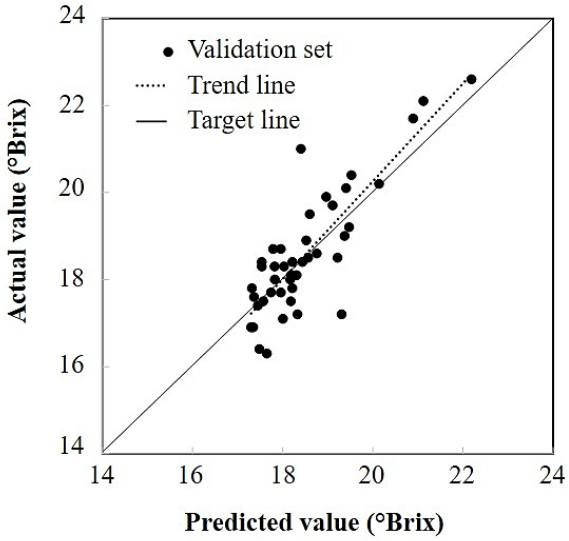


**Figure 4** Scatter plot of NIR predicted value versus actual Brix value obtained from PLS calibration model of PLS-GA 50 wavelengths in prediction set.

*3.3 Evaluation of model robustness*

After calibration model the NIR model for measuring SSC of mangoes was optimized successfully. The results were similar to those found in previous study carried out by Taira et al. [13] In that study it was developed the calibration model from portable NIR spectrometer with wavelength 600-1000 nm for predicting the SSC and skin colour of Iwrin mangoes [13]. The $r^2$cv and RPD values were found 0.76 and 2.1, respectively. Posom et al. [29] who applied the NIR spectroscopy in diffuse interactance mode for measuring the internal quality of Marian Plum (it is the same family with mangoes). The $r^2$cv, brix and RPD were 0.66, 0.86 °Brix and 1.73, respectively. Williams [30] described the $r^2$ between 0.50-0.64 indicated that model could be used for rough screening purpose, while the $r^2$ between 0.66-0.81 meaning the model could be utilized for screening. According both reports have $r^2$ were 0.66 and 0.70, which was indicated that the models could be used as screening purpose. The RPD was 1.73, indicated that the performance of prediction model was fair [20].

Based on the finding in this work it was determined that higher model accuracy of the model compared to Nagle et al. [31]. Nagle et al. [31] used NIR to predict SSC of mango ($r^2$ = 0.49, RMSEP = 0.60 and RPD = 1.58). Sharma et al. [32] applied online Vis-NIR spectroscopy to measure TSS of intact Nm Dokmai mango, which purpose for assessing internal quality before exportation. They founded that the optimal model for evaluating TSS of mango, provided correlation (r), RMSEP and RPD of 0.74, 0.77 °Brix and 1.44, respectively. Moreover, comparing the performance from Rungpichayapichet et al. [33] who studied the SSC in mango variety of 'Nam Dokmai' and 'Si Thong' using a VIS/NIR spectrometer (HandySpec Campo 1000, tec5AG, Oberursel, Germany) on the range of 700-1100 nm (SWNIR), They got an RMSEP was 1.2 °Brix, which higher than our's result.

Hence, the result of this study was similar to the reports above, it is could be acceptably used in the mangoes grading industry. The wavelengths which were selected also could be used for making an increasing accuracy of the calibration model. Moreover, these wavelengths could be used as criteria to build low-cost spectrometer.

## 4. Conclusions

This study has demonstrated that NIR instrument with a wavelength range between 860 and 1760 nm could be applied to measure the SSC value on mango directly. The spectral data were intact scanned at the mango skin. The calibration model generates from the wavelengths which selected by three conditions such as full wavelength, GA and SPA algorithm which helped to reduce the number of variables. The model for prediction of SSC which was developed from the second derivative and selecting wavelengths by GA with 50 variables was suitable due to fewer variables than the full wavelength having 50 variables for GA and 256 variables for full

wavelength. The NIR could have to be used as a non-destructive measurement approach for the direct screening of the internal quality of mango. It could also help the farmers in grading or sorting during packing process non-destructively.

## 5. Acknowledgements

The authors would like to thank the Research and Graduate Studies Khon Kaen University, Thailand, the National Science and Technology Development Agency (NSTDA), RD&E Funding Contract (FDA-CO-2562-10130-TH), Thailand, the Research EN KKU and Applied Engineering for Important Crops of the North East research group, Khon Kaen University, for providing research fund of this experiment. Lastly, we would like to thanks to Salim Hiziroglu to edit and revise manuscript.

## 6. References

[1] Produce Report Partners. Mango Nam Dok Mai: A global favorite that must be tried, at least once in a lifetime [Internet]. 2020 [cited 2020 May 24]. Available from: https://www.producereport.com/article/mango-nam-dok-mai-global-favorite-must-be-tried-least-once-lifetime.
[2] Tropical Green. Golden Namdokmai Mango [Internet]. 2021 [cited 2021 Jan 25]. Available from: https://www.tgfresh.com/product/thai-fresh-mango/.
[3] Chomchalow N, Na Songkhla P. Thai mango export: a slow-but-sustainable development. AU J T. 2008;12(1):1-8.
[4] Sombatpraiwan S, Tipyavimol T, Treeamnuk K. Factors related to ripening-stages of Nam Dok-mai mango after harvesting. J Thai Soc Agr Eng. 2012;18(1):52-8.
[5] Williams PC. Implementation of near-infrared technology. In: Williams PC, Norris KH, editors. Near-infrared technology in the agricultural and food industries. Saint Paul: AACC Inc; 2001. p. 145-71.
[6] Lammertyn J, Nicolai BM, De Smedt V, De Baerdemaeker J. Non-destructive measurement of acidity, soluble solids and firmness of Jonagold apples using NIR-spectroscopy. Trans ASAE. 1998;41(4):1089-94.
[7] Maraphum K, Chuan-Udom S, Saengprachatanarug K, Wongpichet S, Posom J, Phuphaphud A, et al. Effect of waxy material and measurement position of a sugarcane stalk on the rapid determination of Pol value using a portable near infrared instrument. J Near Infrared Spectrosc. 2018;26(5):287-96.
[8] Phuphaphud A, Saengprachatanarug K, Posom J, Maraphum K, Taira E. Prediction of the fibre content of sugarcane stalk by direct scanning using visible-shortwave near infrared spectroscopy. Vib Spectrosc. 2019;101:71-80.
[9] Kawano S, Watanabe H, Iwamoto M. Determination of sugar content in intact peaches by near infrared spectroscopy with fiber optics in interactance mode. J JPN Soc Hortic Sci. 1992;61(2):445-51.
[10] Kawano S, Fujiwara T, Iwamoto M. Nondestructive determination of sugar content in satsuma mandarin using near infrared (NIR) transmittance. J JPN Soc Hortic Sci. 1993;62(2):465-70.
[11] Miyamoto K, Kitano Y. Non-destructive determination of sugar content in satsuma mandarin fruit by near infrared transmittance spectroscopy. J Near Infrared Spectrosc. 1995;3(4):227-37.
[12] Temma T, Hanamatsu K, Shinoki F. Development of a portable near infrared sugar-measuring instrument. J Near Infrared Spectrosc. 2002;10(1):77-83.
[13] Taira E, Nakamura S, Hiyane R, Honda H, Ueno M. Development of a nondestructive measurement system for mango fruit using near infrared spectroscopy. Eng Appl Sci Res. 2017;44(3):189-92.
[14] Jha SN, Chopra S, Kingsly ARP. Determination of sweetness of intact mango using visual spectral analysis. Biosyst Eng. 2005;91(2):157-61.
[15] Pitak L, Sirisomboon P, Saengprachatanarug K, Wongpichet S, Posom J. Rapid elemental composition measurement of commercial pellets using line-scan hyperspectral imaging analysis. Energy. 2021;220:119698.
[16] Liu D, Sun DW, Zeng XA. Recent advances in wavelength selection techniques for hyperspectral image processing in the food industry. Food Bioprocess Technol. 2014;7(2):307-23.
[17] Posom J, Sirisomboon P. Evaluation of lower heating value and elemental composition of bamboo using near infrared spectroscopy. Energy. 2017;121:147-58.
[18] Dardenne P. Some considerations about NIR spectroscopy: closing speech at NIR-2009. NIR News. 2010;21(1):8-14.
[19] Araujo MCU, Bezerra STC, Galvao RKH, Yoneyama T, Chame HC, Visani V. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. Chemometr Intell Lab Syst. 2001;57(2):65-73.
[20] Williams P, Norris K. Near-infrared technology in the agricultural and food industries. 2nd ed. Saint Paul: Amer Assn of Cereal Chemists; 1987.
[21] Chang CW, Laird DA, Mausbach MJ, Hurburgh CR. Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. Soil Sci Soc Am J. 2001;65(2):480-90.
[22] Wang J, Wang J, Chen Z, Han D. Development of multi-cultivar models for predicting the soluble solid content and firmness of European pear (*Pyrus communis* L.) using portable vis-NIR spectroscopy. Postharvest Biol Technol. 2017;129:143-51.
[23] Ncama K, Opara UL, Tesfay S, Fawole OA, Magwaza LS. Application of Vis/NIR spectroscopy for predicting sweetness and fla- vour parameters of 'Valencia' orange (*Citrus sinensis*) and 'Star Ruby' grapefruit (*Citrus x paradisi Macfad*). J Food Eng. 2017;193:86-94.
[24] Maraphum K, Saengprachatanarug K, Wongpichet S, Phuphaphud A, Posom J. In-field measurement of starch content of cassava tubers using handheld vis-near infrared spectroscopy implemented for breeding programmes. Comput Electron Agr. 2020;175(1):105607.
[25] Maraphum K, Saengprachatanarug K, Aparatana K, Izumikawa Y, Taira E. Spatial mapping of brix and moisture content using hyperspectral imaging system in sugarcane stalk. J Near Infrared Spectrosc. 2020;28(4):167-74.
[26] Osborne BG, Fearn T, Hindle PH. Practical NIR spectroscopy with applications in food and beverage analysis. UK: Addison-Wesley Longman Ltd; 1993.
[27] Osborne BG, Fearn T. Near infrared spectroscopy in food analysis. London: Longman Scientific and Technical; 1986.
[28] Workman J, Weyer L. Practical guide to interpretive near-infrared spec- troscopy. Boca Raton: CRC Press; 2007.

[29] Posom J, Klaprachan J, Rattanasopa K, Sirisomboon P, Saengprachatanarug K, Wongpichet S. Predicting marian plum fruit quality without environmental condition impact by handheld visible-near-infrared spectroscopy. ACS Omega. 2020;5(43): 27909-21.
[30] Williams P. Near-infrared technology? Getting the best out of light edition: a short course in the practical implementation of near-infrared spectroscopy for the user. 2nd ed. Nanaimo: PDK Projects Inc; 2007.
[31] Nagle M, Mahayothee B, Rungpichayapichet P, Janjai S, Muller J. Effect of irrigation on near-infrared (NIR) based prediction of mango maturity. Sci Hortic. 2010;125(4):771-4.
[32] Sharma S, Sirisomboon P, Pornchaloempong P. Application of a Vis-NIR spectroscopic technique to measure the total soluble solids content of intact mangoes in motion on a belt conveyor. Hort J. 2020;89(5):545-52.
[33] Rungpichayapichet P, Mahayothee B, Nagle M, Khuwijitjaru P, Muller J. Robust NIRS models for non-destructive prediction of postharvest fruit ripeness and quality in mango. Postharvest Biol Technol. 2016;111:31-40.