Engineering Journal issn: 0125-8281

Article

Two Different Portables Visible-Near Infrared and Shortwave Infrared Region for On-Tree Measurement of Soluble Solid Content of Marian Plum Fruit

Jetsada Posom[1,2,3,a,*], Navavit Soonnamtiang[1,b], Patcharapong Kotethum[1,c], Pakhpoom Konjun[1,d], Panmanas Sirisomboon[4,e], Khwantri Saengprachatanarug[1,2,3,f], and Seree Wongpichet[1,3,g]

1 Department of Agricultural Engineering, Khon Kaen University, Khon Kaen 40002, Thailand
2 Applied Engineering for Important Crops of the North East Research Group, Department of Agricultural Engineering, Khon Kaen University, Khon Kaen 40002, Thailand
3 Bio-Sensing and Field Robotic Laboratory, Khon Kaen University, Khon Kaen 40002, Thailand
4 Department of Agricultural Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Chalongkrung Road, Ladkrabang, Bangkok 10520, Thailand
E-mail: [a]jetspo@kku.ac.th (Corresponding author), [b]navavit@kkumail.com, [c]patcharapong.0441@gmail.com, [d]pakhpoom@kkumail.com, [e]panmanas@gmail.com, [f]khwantri@kku.ac.th, [g]serwon123@gmail.com

Abstract. The goal of this study was to predict the soluble solid content (SSC) of on-tree Marian plum fruit using two different wavelength range and algorithm. One of these was the commercial dispersion NIR spectrometer (MicroNIR 1700), providing shortwave infrared (SWIR), while the other was a making diode array spectrometer giving visible-near infrared (Vis-NIR). To search optimal model, the analytical ability of the two wavelength ranges spectrometers coupled with two algorithms: i.e. partial least squares regression (PLSR) and support vector machine regression (SVR), were investigated. Different spectral pre-processing methods were tested. The model providing the lowest root mean square errors of prediction (RMSEP) was selected. Overall, the proposed outcome was that the performance of SWIR was more accurate than Vis-NIR spectrometer, and that both SWIR and Vis-NIR coupled with PLSR algorithm had a higher accuracy than SVR algorithm. The best model for on-tree evaluation SSC was the SWIR constructed using the PLSR algorithm with the spectral pre-processing of the 2nd derivative, providing a coefficient of determination of calibration set (R^2) of 0.81, a coefficient of determination of validation set (r^2) of 0.76, RMSEP of 0.69 °Brix, and a relative standard error of prediction (RSEP) of 4.43%. The outcome showed that a portable SWIR spectrometer developed with PLSR could be used for monitoring the SSC of individual Marian plum fruit on-tree for quality assurance.

Keywords: Handheld NIR spectrometer, soluble solids content, partial least squares regression, support vector machine regression, on-tree measurement.

ENGINEERING JOURNAL Volume 24 Issue 5
Received 16 February 2020
Accepted 3 August 2020
Published 30 September 2020
Online at https://engj.org/
DOI:10.4186/ej.2020.24.5.227

1. Introduction

The Marian plum (*Bouea macrophylla*) is a fruiting tree planted in Southeast Asia, with most plantations being found in Malaysia, Thailand, and Indonesia [1, 2]. Currently, the price of Marian plum is set following its appearance. For example, if fruit is a big size, ellipse shape, absence of defects, skin and flesh color, it is sold with a high price. However, quality of fruit is not only the physical properties but also chemical properties. Soluble solid content (SSC) is the one parameter indicated the quality of Marian plum fruit. It should be checked before selling and SSC should be used to set the price of fruit. Therefore, if a Marian plum has a lower SSC, it should be sold with a lower price. On the contrary, fruits giving a high SSC value should be sold with a high price. At present, the SSC of fruit in each plantation is randomly checked using the refrectometry method. This leads to destruction, wastage, and lost time. Even though the examined fruits are from the same tree, area and age, the fruits' quality are varied due to different on-tree conditions, which is uncontrolled [3–5]. Therefore, to harvest with a quality of fruit, the SSC of all fruits should be checked using an easy-to-use device to be 100-percent verified. Hence, a rapid and non-destructive technique such as the NIR spectrometer is a key. Moreover, the evaluation of the quality of individual fruits in the field also benefits field management and fruit storage [6].

NIR spectroscopy has been applied to estimate the quality of various fruits and crops, Asian pears [5], apple [6, 7], passion fruit pulp [8], onion bulbs [9, 12], mango [10], and sugarcane stalk [11]. Unfortunately, no studies have been conducted analyzing Marian plum fruit in the field. Therefore, obtaining the necessary knowledge of how to develop the NIR predictive model is important. The factors of interest used to develop the model include wavelength ranges, algorithm, spectral and pre-processing techniques [9]. Because these factors are very important, they urgently need to be examined. In the field, the focus is placed on the analytical performance of a portable NIR spectrometer. Because the Marian plum is a climacteric fruit, this focus is necessary, due to the fact that its effective utilization can lead to differences in the properties of the fruits during the harvesting process and during transportation to the customer [2].

For model development, algorithm and spectra preprocessing is significant. The best model can be achieved if algorithm and spectra pretreatment methods are suitable. Both linear and non-linear algorithms were preferred for model development [13–16]. PLSR and SVR represent the linear and non-linear algorithms, respectively. PLSR creates an inner relationship between the absorbance information and the reference value, using the maximizing covariance of the independent variable (X, absorption value) and the dependent variable (reference value) [17, 18]. For example, PLSR was proposed for use in the application of NIR spectroscopic method on sugarcane [11, 19], avocado fruit [20], and onion [12], giving a high performance. SVR represents non-linear correlation, and at present, only a few research studies have applied SVR to NIR spectroscopic techniques. It is able to provide excellent performance when using small sample sizes [21]. It is very interesting to study occasions in which SVR has been compared to PLSR, and as such, a comparison of the performance of PLSR and SVR using SWIR was studied by Malegori et al. [13], who found that SVR algorithm had given better results when evaluating titratable acidity in acerola fruit. PLSR is not always a perfect technique [22] if the relationship between the spectra and interests are not linear. As a result, the non-linear model was proposed. Pretreatment of spectral data has become an integral part of chemometrics modeling, it is used to remove physical phenomena in the spectra in order to improve the multivariate regression [23]. Spectral pretreatment is divided into two categories: derivatives methods (e.g. sd1, sd2, smoothing (Savitzky-Golay)) and scatter correction methods (e.g. SNV, MSC, baseline correction). For example, baseline offset was applied in order to solve the baseline shift resulting from differences in particle sizes. The 1st and 2nd derivatives were able to separate the overlapping peaks, which benefitted the modelling [24].

SNV and MSC were able to reduce the multiplicative scattering effects of sample surfaces resulting from the differences in particle size. SNV reduces the additive in spectra and the effect of multiplicative scatter, it is used to treat individual spectrum by dividing with its standard deviation [25]. While MSC helps to adjust the slope and the intercept of individual spectrum to the averaged slope and intercept [25]. The spectral pre-treatments of SNV and MSC give very similar results [26]. However, their geometry in spectral space is not the same [26]. Optimal technique of spectral data collection is significant, but, after data collection, pre-processing of spectral data is the most important step before model creation [23]. Therefore, the application of different combinations of spectral pretreatment method is necessary in order to improvement of calibrations and the selection of the best model on the basis of validation results [27].

Therefore, the goal of this work was to compare the analytical performance of the two different wavelength ranges (Vis-NIR and SWIR) of spectra, which had been collected from different portable spectrometers, and two algorithms (PLSR and SVR) on their accuracy to examine the SSC of Marian plum fruit on-tree. Several spectra pretreatment method including single and combination pre-treatments was applied for modeling. Utilizing this knowledge could be the best way to make good predictions about the SSC of 'tree-ripening' fruits in the field. If this work succeeds, it will be able to benefit the commercial fruit industry by allowing them to generate higher profits, as faulty decisions, which tend to be made at harvest time, will be reduced.

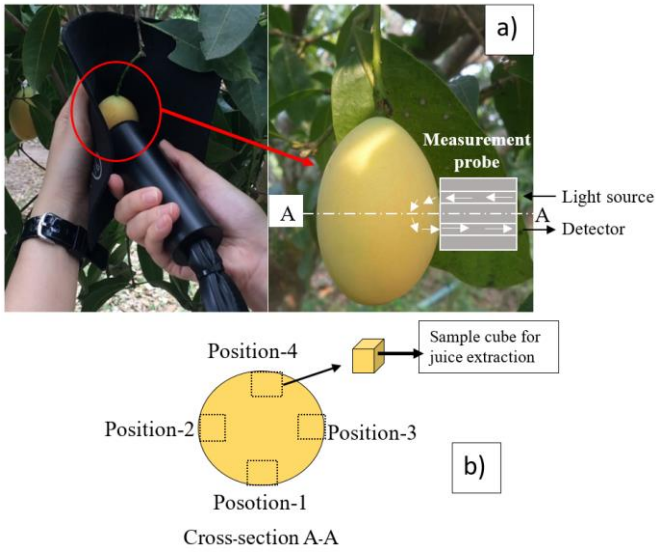

Fig. 1. a) Marian plum fruit ripening on-tree and spectral acquisition, b) The measurement position and its cross-section.

2. Materials and Methods

2.1. Sample Collection

Marian plums of the 'Tool Kraw' variety were collected from five plantations located in Nakhon Nayok Province, Thailand. In order to obtain a wide variance of SSC values, Marian plums at five different stages of ripeness, including <60 days (13 fruits), 60 days (7 fruits), 70 days (13 fruits), 75 days (54 fruits), and >75 days (13 fruits) days after blooming, were randomly collected from different plantations i.e. orchard -A, -B, -C, -D, -E, -F, -G and -blind. The total sample equaled 100 fruits. Fig. 1a illustrates the Marian plum fruit ripening on-tree.

2.2. Spectral Acquisition

NIR spectral acquisition for the Vis-NIR and SWIR were scanned using an interactance mode. The scanning was performed on-tree in the field under real environmental conditions. Fig. 1a illustrated on-tree NIR spectral acquisition. The following two spectrometers were used:

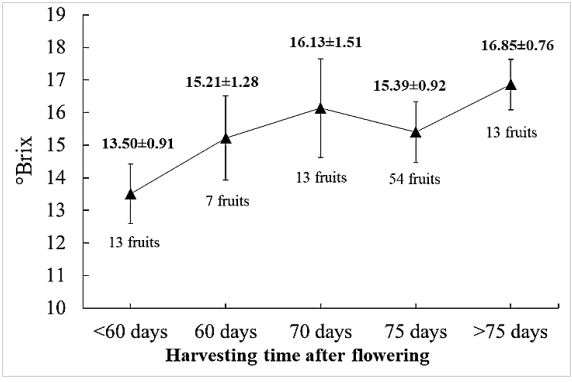

Fig. 2. Variation of SSC value at each ripening stage.

• A portable NIR instrument (P-TF1, HNK Engineering Co., Ltd., Hokkaido, Japan) in an interactance mode with a short wavelength range (Vis-NIR) of between 570 to 1031 nm.
• A MicroNIR Pro1700 in an interactance mode with a long wavelength range (SWIR) of between 908-1676 nm.

Each fruit was scanned on-tree using the two spectrometers with an integration time of 30 ms at the center of fruit. During spectral acquisition, temperature samples was measured using infrared digital thermometer (GM 320, benetech, China). The measurement positions were at Position-1, Position -2, Position -3, and Position -4, respectively (see Fig. 1b). In order to avoid specular reflection, a light source window and a detector were adjoined to the surface of the fruit. Position-1 was randomly chosen and scanned. Next, Position-2, Position-3, and Position-4 were pointed at 90°, 180°, and 270° degree rotations, respectively from Position-1. Each position was scanned one time. The absorbance was calculated from log 1/R, in which R is the relative reflectance. It was calculated using the formula [28]: $R = \frac{(R_{Sam}-R_{Dark})}{(R_{White}-R_{Dark})}$, in which R_{Sam} represents the reflection of the Marian plum sample, R_{White} represents the reflection of the white reference (Teflon), and R_{Dark} represents the reflection when the light source was turned off.

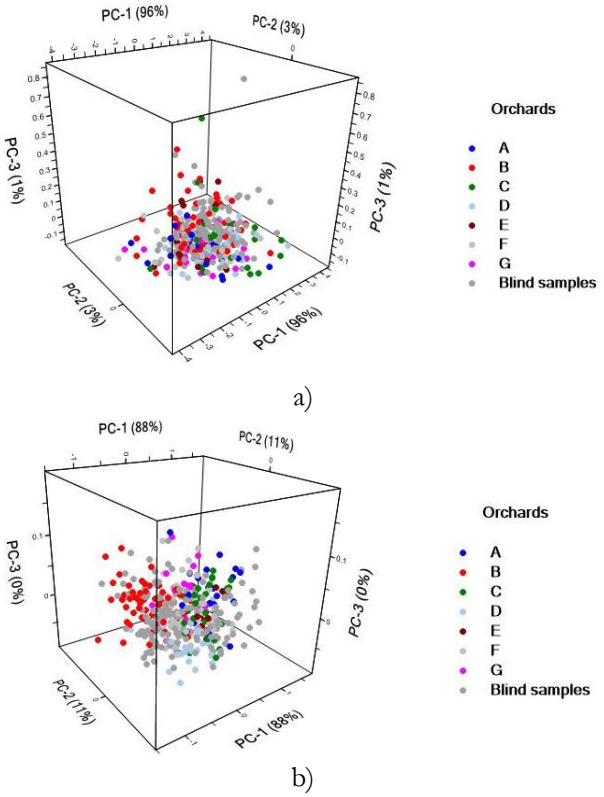

Fig. 3. PCA scores plot (PC1, PC2, and PC3) for all spectral data acquired on the a) shortwave and b) longwave.

2.3. Reference Analysis

After spectral acquisition, the fruit at each position was cut 1cm^3 around the scanned point. Then the 1 cm^3 sample was immediately squeezed in order to obtain its juice (See Fig. 1b for details of the juice extraction and scanning procedures). The SSC values were collected by dropping juice into the test slot of a portable refractometer (Nar–3Ta, ATAGO, Tokyo, Japan).

2.4. NIR Construction

The SSC predictive models were developed using Vis-NIR and SWIR spectra coupled with two different multivariate algorithms (i.e., PLSR and SVR). The two algorithms and spectral pre-treatments were performed using multivariate software analysis (Unscrambler X 10.3, Camo, Norway). Before model development, either raw spectra or pre-processed spectra was used for the multivariate analysis. The spectral pre-processing methods were performed utilizing any of the following methods: a) smoothing (Savitzky-Golay, smoothing points=7, polynomial order=2); b) the 1st derivative (Savitzky-Golay, smoothing points=7, polynomial order=2); c) the 2nd derivative (Savitzky-Golay, smoothing points=7, polynomial order=2); d) the baseline offset; e) the standard normal variate (SNV); f) the multiplicative scatter correction (MSC); g) the baseline offset + SNV; h) MSC+SNV; and i) SNV+MSC.

The SSC of each position and its corresponding spectrum was assigned as a data set. The total sample equaled one hundred fruits scanned with four positions per fruit. Therefore, the total data set was four hundred (100 fruits × 4 positions = 400 data set). The data set was sorted in ascending order of SSC value. One in every four samples was selected to be a validation set while the remaining samples were assigned to be a calibration set. The calibration set covered the maximum SSC range. The model was validated using test set method: 75% was for the calibration set and 25% was for the validation.

Table 1. Results of calibration and the validation model for NIR (700-1000 nm) developed using PLSR and SVR algorithms in the evaluation of total soluble solids of Marian plum fruits.

Spectral	Algorithms	**Calibration Set**				**Validation Set**			
Pre-processing		Factor	R^2	RMSE	r^2	RMSEP	bias	RPD	RSEP,%
Raw spectra	PLSR	9	0.73	0.72	0.66	0.81	-0.097	1.7	5.22
	SVR	-	0.30	1.24	0.31	1.28	0.0098	1.1	8.32
[a]Smoothing (Savitzky-Golay)	**PLSR**	**10**	**0.75**	**0.70**	**0.70**	**0.76**	**-0.12**	**1.8**	**4.91**
	SVR	-	0.30	1.24	0.31	1.28	0.0094	1.1	8.32
[a]1st derivative	PLSR	9	0.77	0.67	0.65	0.82	-0.080	1.7	5.28
	SVR	**-**	**0.75**	**0.72**	**0.69**	**0.85**	**0.080**	**1.6**	**5.54**
[a]2nd derivative	PLSR	11	0.67	0.80	0.50	0.98	-0.055	1.4	6.30
	SVR	-	0.52	1.00	0.42	1.13	0.027	1.2	7.34
Baseline offset	PLSR	10	0.75	0.70	0.67	0.80	-0.13	1.8	5.15
	SVR	-	0.47	1.04	0.48	1.13	0.018	1.2	7.31
SNV	PLSR	7	0.71	0.75	0.63	0.84	-0.086	1.7	5.41
	SVR	-	0.68	0.80	0.61	0.92	0.11	1.5	6.02
MSC	PLSR	8	0.74	0.71	0.65	0.82	-0.15	1.7	5.27
	SVR	-	0.68	0.80	0.60	0.93	0.11	1.5	6.08
Baseline offset + SNV	PLSR	7	0.71	0.75	0.63	0.84	-0.086	1.7	5.41
	SVR	-	0.68	0.80	0.61	0.92	0.11	1.5	6.02
MSC+SNV	PLSR	6	0.69	0.78	0.59	0.88	-0.068	1.6	5.70
	SVR	-	0.68	0.80	0.61	0.92	0.11	1.5	6.02
SNV+MSC	PLSR	8	0.74	0.71	0.65	0.82	-0.15	1.7	5.27
	SVR	-	0.68	0.80	0.60	0.93	0.11	1.5	6.08

[a]Smoothing point=3, Polynomial order=2, R^2: coefficient of determination of calibration set, r^2: coefficient of determination of validation set, RMSE: root mean square error of estimation, RMSEP: root mean square error of prediction, RSEP: relative standard error of prediction, PLSR: partial least square regression, SVM: support vector machine

2.5. Model Performance

After the modeling, the accuracy of each condition was compared using the statistical term of the coefficient of determination (R^2: calibration set, r^2: validation set), the root mean square error of prediction (RMSEP), the bias, the residual predictive deviation (RPD), and the relative standard error of prediction (RSEP) [11]. The effective model was selected based on the highest r^2 or the lowest RMSEP. These parameters can be calculated as follows [25]

$$R^2 = 1 - \frac{\sum_1^n (y_i - y_{pre})^2}{\sum_1^n (y_i - \bar{y})^2} \quad (1)$$

$$RMSEP = \sqrt{\frac{\sum_i^n [(y_i - y_{pre}) - bias]^2}{n}} \quad (2)$$

$$\text{bias} = \frac{\sum_i^n (y_i - y_{pre})}{n} \quad (3)$$

$$\text{RPD} = \frac{SD_y}{\text{RMSEP}} \quad (4)$$

$$\text{RSEP} = \sqrt{\sum \left[\frac{(Y_i - Y_{pre})^2}{\sum Y_i^2}\right]} \times 100 \quad (5)$$

in which y_i, y_{pre}, and $\bar{y}$ are the measured SSC value, the predicted SSC value, and the average measured SSC value, respectively. Bias is the mean difference between the reference SSC values and the predicted value. The percentage of absolute error of prediction and its reference value is shown in RSEP. For example, if the RSEP was equal to 5 %, it meant that every prediction could approximately give a 5% error [11].

To search for the effective model, the wavelength range (Vis-NIR and SWIR) and the regression algorithm (PLSR and SVR) were crossed, including Vis-NIR-PLRS; SWIR-PLSR; Vis-NIR-SVR; and SWIR-SVR.

Table 2. Results of calibration and validation model for SWIR (908-1676 nm) developed using PLSR and SVR algorithms in the evaluation of the SSC of Marian plum fruits.

		Calibration Set			Validation Set				
Spectral Pre-processing	Algorithms	Factor	R^2	RMSE	r^2	RMSEP	Bias	RPD	RSEP,%
Raw spectra	PLSR	11	0.81	0.61	0.72	0.74	0.00046	1.9	4.75
	SVR	-	0.29	1.19	0.32	1.16	0.072	1.2	7.57
[a]Smoothing (Savitzky-Golay)	PLSR	11	0.81	0.61	0.72	0.74	0.0023	1.9	4.75
	SVR	-	0.29	1.19	0.32	1.16	0.072	1.2	7.57
[a]1st derivative	PLSR	8	0.80	0.63	0.74	0.70	-0.0026	2.0	4.52
	SVR	-	0.73	0.74	0.70	0.78	-0.0077	1.8	5.07
[a]2nd derivative	PLSR	9	0.81	0.60	0.76	**0.69**	-0.011	2.0	4.43
	SVR	-	0.79	0.65	0.73	0.73	0.029	1.9	4.76
Baseline offset	PLSR	10	0.81	0.61	0.72	0.73	0.0044	1.9	4.74
	SVR	-	0.51	1.00	0.47	1.05	0.052	1.3	6.82
SNV	PLSR	12	0.82	0.60	0.75	**0.69**	-0.019	2.0	4.48
	SVR	-	0.47	1.05	0.47	1.06	0.063	1.3	6.92
MSC	PLSR	11	0.81	0.61	0.74	0.70	-0.014	2.0	4.53
	SVR	-	0.47	1.05	0.48	1.06	0.080	1.3	6.92
Baseline offset+ SNV	PLSR	12	0.82	0.60	0.75	**0.69**	-0.019	2.0	4.48
	SVR	-	0.47	1.05	0.47	1.06	0.063	1.3	6.92
MSC+SNV	PLSR	10	0.76	0.63	0.72	0.73	-0.011	2.0	4.74
	SVR	-	0.47	1.05	0.47	1.06	0.062	1.3	6.92
SNV+MSC	PLSR	11	0.81	0.61	0.74	0.70	-0.014	2.0	4.53
	SVR	-	0.47	1.05	0.47	1.06	0.063	1.3	6.92

3. Result and Discussions

3.1. Reference Data

Figure 2 delineates the statistical data of the Marian plum fruits at the different harvesting times used for developing the SSC predictive model. The variability of the fruit samples had in fact depended upon the differences in their stages of maturity. As the results show, there were different SSC values at each of the ripening stages. The lowest SSC was found at the age at <60 days after flowering, and the highest SSC was at the age of >75 days. The SSC values increased as the harvesting time increased. However, because the harvests had been carried out in different orchards, a lower SSC value was discovered at 75 days than was found at 70 days. It is possible that the differences in the SSC of the Marian plum fruits could have arisen if any of the farmers had given different attention or care to the trees or if there had been differences in the external factors, such as location and soil fertility. Therefore, if a plant had grown in good field conditions and had received good attention and care, it is possible that fruits harvested after a lower number of harvesting days could have a higher SSC than those with a higher number of flowering days. This confirms the fact that the classification of the SSC of the Marian plum cannot be determined visually or predicted based on the age of the fruit.

3.2. Spectral Data

Figures 3a and 3b demonstrate the scatter plots of PC1, PC2, and PC3 of raw spectra obtained from Vis-NIR and SWIR spectra. The result shows that there was no effect of different orchards. The in-field temperature sample during experiment ranged between 29 to 39 °C, and average temperature sample was 32.9 °C. As reported by Acharya et al. [29], found that the temperature sample between 20 to 40 °C did not affect the accuracy (R value and SEP) of tomato dry matter content model predicted by Vis-NIR spectrometer. Figures 4a, 4b, 4c, and 4d show the spectra of the Marian plum samples for the Vis-NIR

spectra, the 2nd derivative Vis-NIR spectra, the SWIR spectra, and the 2nd derivative SWIR spectra, respectively. The Vis-NIR spectra contained higher noise in the wavelength range of 571-700 nm due to effects from the color of the fruit's surface. Meanwhile, 1000-1031 nm contained noise checked by visualizing (see Fig. 4b). These wavelength ranges were then not recommended. Sanseechan et al. [11] found that the wavelength range of 700–1000 nm was able to give good predictions for the assessment of cane density, whereas a wavelength ranges of 670–1031 nm gave poor predictions. The model was improved by removing the confounding effects of the visible wavelength range. For the development of the model, a wavelength of 700–1000 nm was used. This range was also suggested by Luo et al. [7] who found that the wavelength range of 861-1074 nm had provided the most robust model for predicting the SCC of apples. Meanwhile, SWIR was determined to be a good spectrum. By carrying out a visual check, it was found that there was no noise in the raw and the 2nd derivative spectra. SWIR had many obvious peaks which had occurred in the 2nd derivative spectra, such as at 1370 nm, 1409 nm, 1450 nm, and 1651 nm, which are related to the C-H ($2\nu CH_3$ and δCH_3) combinations of hydrocarbons [30]; the C-H ($2\nu CH_2$ and δCH_2) combinations of hydrocarbons, O-H (2v), the O-H of water and starch [30] and C-H methyl, and CH_3NO_2 [30], respectively. In addition, these obvious peaks demonstrated that there was much change when the composition of the samples differed. This characteristic can give a higher variance of absorbance, which is useful for modeling.

3.3. Analytical Performance

Table 1 demonstrates the PLSR and SVR results developed from Vis-NIR across either the raw or the pre-treated spectra. Overall, the PLSR algorithm was proven to be more accurate than the SVR algorithm. The best model for PLSR algorithm, which provided the lowest RMSEP, was optimized with spectral pre-treatment of smoothing. The R^2, RMSE, r^2, RMSEP, bias, RPD, and RSEP were 0.75, 0.70 °Brix, 0.70, 0.76 °Brix, -0.12 °Brix, 1.8, and 4.91%, respectively. For the SVR algorithm, generation with spectral pre-processing of the 1st derivative was also able to provide a fair prediction. The R^2, RMSE, r^2, RMSEP, bias, RPD, and RSEP were 0.75, 0.72 °Brix, 0.69, 0.85 °Brix, 0.080 °Brix, 1.6, and 5.54%, respectively.

For SWIR results, the PLSR and the SVR algorithms utilizing different pre-treatment techniques are illustrated in Table 2. The overall outcome found was that the PLSR proved to be more accurate than SVR algorithm. The best result in PLSR algorithm was developed with spectral pre-processing of the 2nd derivative, giving a PLS factor of 9, providing an R^2 value of 0.81, an RMSE of 0.60 °Brix, an r^2 of 0.76, an RMSEP of 0.69 °Brix, a bias of -0.011 °Brix, an RPD of 2.0, and an RSEP of 4.43%. Furthermore, the best result in SVR algorithm was pre-treated with the 2nd derivative, provided an R^2 of 0.79, an RMSE of 0.65 °Brix, an r^2 of 0.73, an RMSEP of 0.73 °Brix, a bias of 0.029 °Brix, an RPD of 1.9, and an RSEP of 4.76%, respectively. As the result, it was shown that the SVR algorithm for Vis-NIR and SWIR could be optimized with derivative technique.

The ability of the validation model indicated by the RPD value was as follows: an RPD $\leq$1.5 means that it will give poor predictions, an 1.5<RPD$\leq$2.0 means that it can be used for rough screening, 2.0<RPD$\leq$2.5 can be used for quantitative predictions, 2.5<RPD$\leq$3.0 are fit to be used as good predictors, and an RPD>3.0 are appropriate to create an excellent model [20, 31, 32].

Therefore, the SWIR-PLSR model gave a good prediction and was able to be used for screening as a quantitative and qualitative assessment, while the SWIR-SVR, Vis-NIR-PLSR, and the Vis-NIR-SVR had a fair models and could be used for rough screening. However, all of the models showed a low RSEP (approximately 4.43 - 5.54%), which was able to be accepted for quality assessment. For example, if the RSEP equals 4.43%, it means that every predicted value will give an error of approximately 4.43% based on measured SSC value. The scatter plots between the predicted values and measured values of the effective model in NIR were developed using Vis-NIR-PLSR, SWIR-PLSR, Vis-NIR-SVR, and SWIR-SVR and are illustrated in Figs. 5a, 5b, 5c, and 5d, respectively. The optimal algorithm which could be used to predict SSC in Marian plum fruit was PLSR.

Figures 6a and 6b show the regression coefficient plot and the pre-treated spectra of the best model optimized from the PLSR algorithm of Vis-NIR and SWIR. If the regression coefficient value had shown high value at any wavelength, it would mean that the wavelength had had a high impact on the prediction.

Soluble solid content (SSC) in fruit are mainly composed of soluble sugars, including fructose, glucose and sucrose [33]. This leads to SSC is a key parameter in the quality of fruit. Sugars molecules have a generic formula of $C_nH_{2n}O_n$ [34].

The obvious peaks in regression coefficient plot and the Vis-NIR spectra, which had been pre-treated by smoothing, were at 876, 912, and 940 nm and are related to the vibrational band of the C-H aromatic of hydrocarbon [30], the C-H methyl of CH_3 [30], and the C-H stretching third overtone of CH_2 [35], respectively. The waveband at 910 nm is related to the sugar band [36]. The result can be seen that important peak related to vibrational band of hydrocarbon, which are in sugar structure. Maraphum et al. [37] found that the vibrational bands of 910, 953, 836, 928, and 970 nm had impacted the prediction of Pol in cane stalks. The important peaks used to predict the density of sugarcane stalks were found at 800 to 818 nm, 912, 960, and 980 nm as reported by Sanseechan et al. [11].

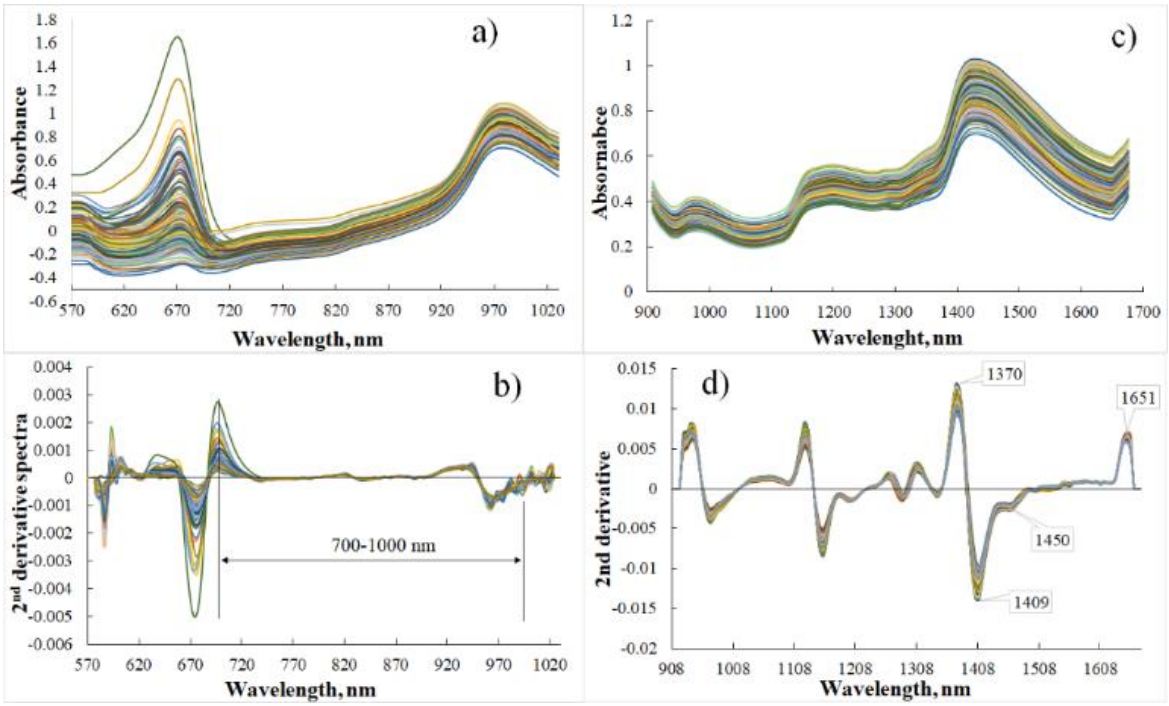

Fig. 4. The spectra of the Marian plum samples: a) shortwave NIR spectra, b) 2nd derivative shortwave NIR spectra, c) longwave NIR spectra, and d) 2nd derivative longwave NIR spectra.

For the SWIR model, which had been pre-processed from the 2nd derivative, obvious peaks in the regression coefficient plot were at 982, 1155, 1409, and 1465 nm, and they corresponded to the O-H stretching of the second overtone of starch, the C-H stretching of the second overtone of CH_3, and the O-H stretching of the first overtone of ROH [35], respectively. These peaks related to sugar structure.

For Vis-NIR and SWIR model, the accuracy is not much difference between raw spectra and pre-treated spectra. Therefore, the raw spectral model could be assigned as the best for the application of detecting SSC on trees because of processed fast.

Comparing to the previous study, our study showed a high degree of accuracy, which has also been reported by Huang et al. [38], who predicted the soluble solids content (SSC) of 'Sun Bright' tomatoes by visible and shortwave near-infrared Vis-NIR (400-1100 nm) and SWIR (900-1300 nm) spectroscopy. In their study, they gave r^2, RMSEP, and RPD values of 0.53, 0.45 °Brix, and 1.4; and 0.66, 0.37 °Brix, and 1.7, respectively.

Comparing the performance of the SSC prediction model with the results of other studies using portable devices, the prediction precision obtained in the present study was better than that reported by Neto et al. [39], who studied 'Palmer' mango (*Mangifera indica* L.) using a commercial portable device in the spectral range of 310–1100 nm, the best SSC calibration model was developed using spectra pre-processed with SNV, sd1 and window of 699–999 nm. It was observed a RMSECV of 1.39%. On the other hand, lower RMSEP values were reported by Nordey et al. [40] applied longwave (800-2200 nm) NIR portable for prediction of °Brix value in mango fruits (*Mangifera indica* cv. 'Cogshall'), the best model provided RMSEP of 0.6 ◦Brix.

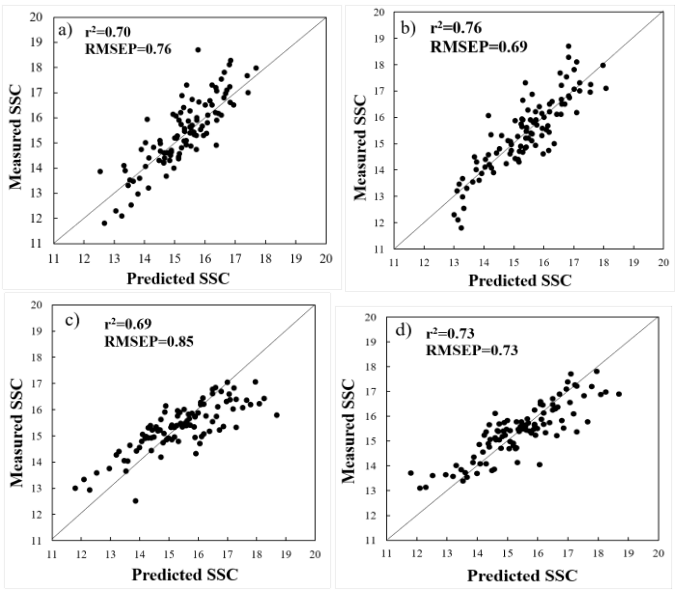

Fig. 5. A comparison of the SSC: a) Vis-NIR developed using PLSR with spectral pre-processing of smoothing, b) SWIR developed using PLSR with spectral pre-processing of the 2nd derivative, c) Vis-NIR developed using SVR with spectral pre-processing of the 1st derivative, and d) SWIR developed using SVR with spectral pre-processing of the 2nd derivative) of Marian plum fruit predicted by effective NIR spectroscopy and measured by a reference laboratory.

DOI:10.4186/ej.2020.24.5.227

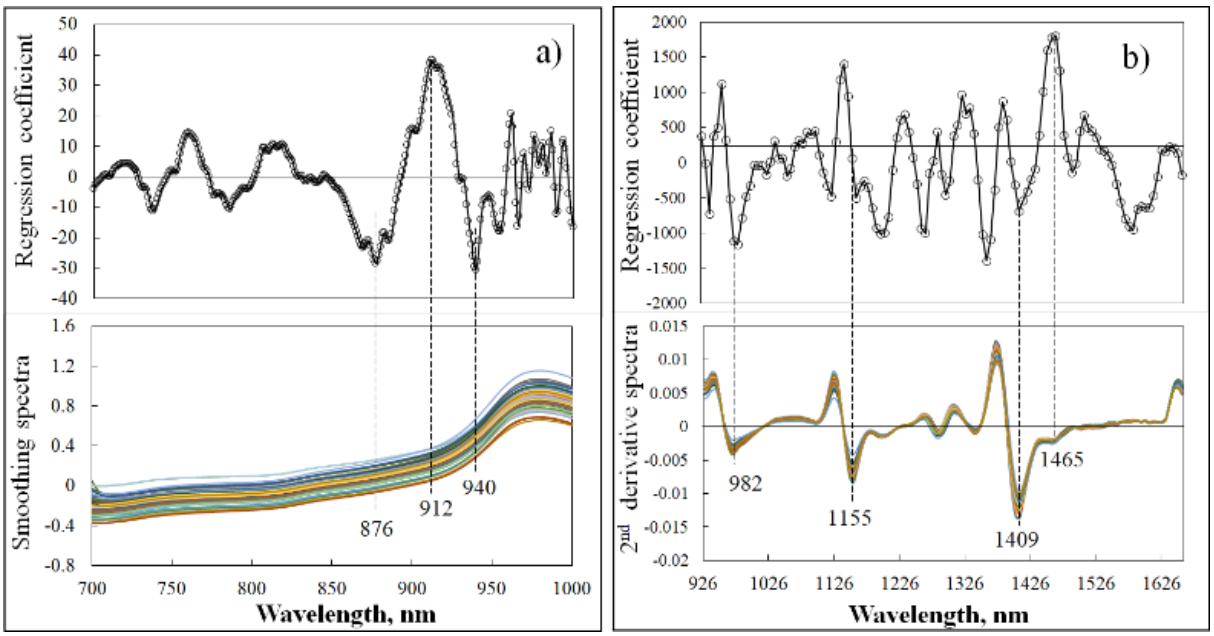

Fig. 6. a) Regression coefficient plot and the smoothing spectra used in the Vis-NIR model developed from the PLSR, b) the regression coefficient plot and the 2nd derivative spectra used in the SWIR model developed from PLSR.

Marques et al. [41] use commercial handheld NIR spectrometer (900-1700 nm) for predicting SSC of 'Tommy Atkins' mango, r^2 and RMSEP values were, respectively: 0.92 and 0.55 °Brix for SSC. From three studies, it was noted that longwave NIR portable had accuracy than shortwave NIR, and the PLS model was a promising tool to predict the SSC of mangoes.

4. Conclusions

It was necessary to use the in-field NIR spectrometer to assess the SSC of Marian plum fruits on tree so that the optimal model development (i.e., wavelength range, algorithm utilization, and spectral pre-processing) could be studied. This study illustrated that wavelength range (Vis-NIR and SWIR) across the algorithms had influenced the model's ability. Considering the algorithm, the PLSR was optimal over the SVR algorithm. For wavelength, SWIR had a higher performance than Vis-NIR. It was found that the pre-treatment of the derivative technique was suitable for intact fruit using the scanning mode of interactance. However, author opinions that raw spectral model was recommended for in-field application because not much different accuracy and rapid analysis. The SVR algorithm was not suggested for the Vis-NIR because this region was a flat spectrum after the derivative displayed a small obvious peak. This is the reason why SVR is not recommended for Vis-NIR. The most effective model was SWIR coupled with the PLSR algorithm, which can be used to make quantitative predictions and to carry out quality assessment. This is a sustainable way to determine the optimal conditions that should be used for fruit assessment in the field.

Acknowledgements

This work was financial supported by the young researcher development project of Khon Kaen University, Thailand, and Sustainable Infrastructure Research and Development Center, Faculty of Engineering, Khon Kaen University, Thailand. The authors would like to thank the Applied Engineering for Important Crops of the North East Research Group at the Department of Agricultural Engineering (AENE) of Khon Kaen University, Thailand, for their equipment support. In addition, the authors would like to express their gratitude to the NIRS Research Center for Agricultural Products and Food (www.nirsresarch.com) at the Department of Agricultural Engineering in the Faculty of Engineering at King Mongkut's Institute of Technology located at Ladkrabang in Bangkok, Thailand for supporting the Micro-NIR Pro spectrometer and the multivariate software analysis (Unscrambler X 10.3, Camo, Norway).

References

[1] S. Subhadrabandhu. (2001). *Under-Utilized Tropical Fruits of Thailand* [Online]. Available: http://www.fao.org/docrep/004/ab777e/ab777e04.htm#bm4.3. (assessed on 5 Aug. 2018).

[2] V. L. Mavuso and C. Yapwattanaphun, "Effect of environmental conditions on flower induction of Marian plum (Bouea burmanica Griff)," *Agriculture and Natural Resources*, vol. 51, pp. 243–246, 2017.

[3] V. Tavallali, S. Esmaili, and S. Karimi, "Nitrogen and potassium requirements of tomato plants for the optimization of fruit quality and antioxidative capacity during storage," *J Food Meas Charact*, vol. 12, pp. 755–762, 2018.

[4] X. An, W. Li, Y. Liang, L. Mu, and T. Bao, "Fruit quality components of balsam pear (Momordica charantia L.) and soil respiration in response to soil moisture under two soil conditions," *J Food Meas Charact*, vol. 12, pp. 710–720, 2018.

[5] J.-H. Choi, P.-A. Chen, B. H. N. Lee, S.-H. Yim, M.-S. Kim, Y.-S. Bae, D.-C. Lim, and H.-J. Seo, "Portable, non-destructive tester integrating VIS/NIR reflectance spectroscopy for the detection of sugar content in Asian pears," *Sci Hortic*, vol. 220, pp. 147–153, 2017.

[6] S. Fan, B. Zhang, J. Li, W. Huang, and C. Wang "Effect of spectrum measurement position variation on the robustness of NIR spectroscopy models for soluble solids content of apple," *Biosyst Eng*, vol. 143, pp. 9–19, 2016.

[7] X. Luo, Z. Ye, H. Xu, D. Zhang, S. Bai, and Y. Ying, "Robustness improvement of NIR-based determination of soluble solids in apple fruit by local calibration," *Postharvest Biol Technol*, vol. 139, pp. 82–90, 2018.

[8] G. Oliveira-Folador, M. O. Bicudo, E. F. Andrade, C. M. G. C. Renard, S. Bureau, and F. Castilhos, "Quality traits prediction of the passion fruit pulp using NIR and MIR spectroscopy," *LWT-Food Sci Technol*, vol. 95, pp. 172–178, 2018.

[9] M. N. Islam, G. Nielsen, S. Stærke, A. Kjær, B. Jørgensen, and M. Edelenbos, "Novel non-destructive quality assessment techniques of onion bulbs: a comparative study," *J Food Sci Technol*, vol. 55, no. 8, pp. 3314–3324, 2018.

[10] T. Nordey, J. Joas, F. Davrieux, M. Chillet, and M. Léchaudel, "Robust NIRS models for non-destructive prediction of mango internal quality," *Sci Hortic*, vol. 216, pp. 51–57, 2017.

[11] P. Sanseechan, L. Panduangnate, K. Saengprachatanarug, S. Wongpichet, E. Taira, and J. Posom, "A portable near infrared spectrometer as a non-destructive tool for rapid screening of solid density stalk in a sugarcane breeding program," *Sens Biosensing Res*, vol. 20, pp. 34–40, 2018.

[12] S. Kuroki, M. Nishino, S. Nakano, Y. Deguchi, and H. Itoh, "Positioning in spectral measurement dominates estimation performance of internal rot in onion bulbs," *Postharvest Biol Technol,* vol. 128, pp. 18–23, 2017.

[13] C. Malegori, E. J. N. Marques, S. T. Freitas, M. F. Pimentel, C. Pasquini, and E. Casiraghi, "Comparing the analytical performances of Micro-NIR and FT-NIR spectrometers in the evaluation of acerola fruit quality, using PLS and SVM regression algorithms," *Talanta*, vol 165, pp. 112–116, 2017.

[14] C. Chen, H. Li, X. Lv, J. Tang, C. Chen, and X. Zheng. "Application of near infrared spectroscopy combined with SVR algorithm in rapid detection of cAMP content in red jujube," *Optik*, vol. 194, pp. 163063, 2019.

[15] J. Chen. S. Zhu, and G. Zhao, "Rapid determination of total protein and wet gluten in commercial wheat flour using siSVR-NIR," *Food Chem*, vol. 221, pp. 1939–1946, 2017.

[16] S. Xu, B. Lu, M. Baldea, TF. Edgar, and M. Nixon, "An improved variable selection method for support vector regression in NIR spectral modeling," *J Process Contr*, vol. 67, pp. 83–93, 2018.

[17] N. Shetty and R. Gislum, "Quantification of fructan concentration in grasses using NIR spectroscopy and PLSR," *Field Crops Res*, vol. 120, no. 1, pp. 31–37, 2011.

[18] P. S. Sampaio, A. Soares, A. Castanho, A. S. Almeida, J. Oliveira, and C. Brites, "Optimization of rice amylose determination by NIR-spectroscopy using PLS chemometrics algorithms," *Food Chem*, vol. 242, pp. 196–204, 2018.

[19] K. Phetpan, V. Udompetaikul, and P. Sirisomboon, "An online visible and near-infrared spectroscopic technique for the real-time evaluation of the soluble solids content of sugarcane billets on an elevator conveyor," *Comput Electron Agr*, vol. 54, pp. 460–466, 2018.

[20] O. O. Olarewaju, I. Bertling, and L. S. Magwaza, "Non-destructive evaluation of avocado fruit maturity using near infrared spectroscopy and PLS regression models," *Sci Hortic*, vol. 199, pp. 229–236, 2016.

[21] D. Zhu, B. Ji, C. Meng, B. Shi, Z. Tu, and Z. Qing, "The performance of ν-support vector regression on determination of soluble solids content of apple by acousto-optic tunable filter near-infrared spectroscopy," *Anal Chim Acta*, vol. 598, pp. 227–234, 2007.

[22] M. Barbanera, E. Lascaro, V. Stanzione, A. Esposito, R. Altieri, and M. Bufacchi, "Characterization of pellets from mixing olive pomace and olive tree pruning," *Renew Energ*, vol. 88, pp. 185–191, 2016.

[23] Å. Rinnan, F. Berg, and SB. Engelsen. "Review of the most common pre-processing techniques for near-infrared spectra," *Trac-trend anal chem*, vol. 28, no. 10, 2009.

[24] M. B. Romìa and M. A. Bernàrdez, "Multivariate calibration for quantiative analysis," in *Infrared Spectroscopy for Food Quality Analysis and Control*. Academic Press, 2009, pp. 74–75.

[25] M. N. Islam, G. Nielsen, S. Stærke, A. Kjær, B. Jørgensen, and M. Edelenbos, "Noninvasive determination of firmness and dry matter content of stored onion bulbs using shortwave infrared imaging with whole spectra and selected wavelengths," *Appl Spectrosc*, vol. 72, no. 10, pp. 1467–1478, 2018.

[26] T. Fearn, C. Riccioli, A. Garrido-Varo, and J. E. Guerrero-Ginel. "On the geometry of SNV and MSC," *Chemometr intell lab*, vol. 96, pp. 22–26, 2009.

[27] V. M. Fernández-Cabanás, A. Garrido-Varo, D. Pérez-Marín, and P. Dardenne. "Evaluation of pretreatment strategies for near-infrared spectroscopy calibration development of unground and ground compound feeding stuffs," *Appl. Spectrosc*, vol. 60, no. 1, pp. 17–23, 2006.

[28] A. Phuphaphud, K. Saengprachatanarug, J. Posom, K. Maraphum, and E. Taira. "Prediction of the fibre content of sugarcane stalk by direct scanning using visible-shortwave near infrared spectroscopy," *Vib Spectrosc*, vol. 101, pp. 71–80, 2019.

[29] U. K. Acharya, K. B. Walsh, and P. P. Subedi, "Effect of temperature on SWNIRS based models of fruit DM and colour," in *NIR2013 Proceedings*, 2–7 June, La Grande-Motte, France, 2014, pp. 674–676.

[30] J. Workman and J. R. L. Weyer, *Practical Guide to Interpretive Near-Infrared Spectroscopy*. Boca Raton, FL: Taylor and Francis, 2007, pp. 240–262.

[31] K. Ncama, S. Z. Tesfay, O. A. Fawole, U. L. Opara, and L. S. Magwaza, "Non-destructive prediction of 'Marsh' grapefruit susceptibility to postharvest rind pitting disorder using reflectance Vis/NIR spectroscopy," *Sci Hortic*, vol. 231, pp. 265–271, 2018.

[32] P. Williams, "Near-infrared technology–getting the best out of light," in *A Short Course in the Practical Implementation of Near-Infrared Spectroscopy for the User*. 5th ed. PDK Grain, Nanaimo, Canada, 2007.

[33] K. Wei, C. Ma, K. Sun, Q. Liu, N. Zhao, Y. Sun, K. Tu, and L. Pan, "Relationship between optical properties and soluble sugar contents of apple flesh during storage," *Postharvest Biol Tec*, vol. 159, pp. 111021, 2020.

[34] P. Shapley, "Structure of sugars," University of Illinois, 2012. Available: http://butane.chem.uiuc.edu/pshapley /GenChem2/B4/1.html

[35] B. G. Osborne and T. Fearn, *Near Infrared Spectroscopy in Food Analysis*. London: Longman Science and Technical, 1986.

[36] S. Kawano, T. Fujiwara, and M. Iwamoto, "Nondestructive determination of sugar content in satsuma mandarian using near infrared (NIR) transmittance," *J Jpn Soc Hort Sci*, vol. 62, pp. 465–470, 1993.

[37] K. Maraphum, S. Chuan-Udom, K. Saengprachatanarug, S. Wongpichet, J. Posom, A. Phuphaphud, and E. Taira, "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 Spec*, vol. 26, no. 5, pp. 287–296, 2018.

[38] Y. Huang, R. Lu, and K. Chen, "Assessment of tomato soluble solids content and pH by spatially-resolved and conventional Vis/NIR spectroscopy," *J Food Eng*, vol. 236, pp. 19–28, 2018.

[39] J. P. S. Neto, M. W. D. Assis, I. P. Casagrande, L. C. C. Júnior, and G. H. A. Teixeira, "Determination of 'Palmer' mango maturity indices using portable near infrared (VIS-NIR) spectrometer," *Sci. Hortic*, vol. 130, pp. 75–80, 2017.

[40] E. J. N. Marques, S. T. Freitas, M. F. Pimentel, and C. Pasquini, "Rapid and non-destructive determination of quality parameters in the 'Tommy Atkins' mango using a novel handheld near infrared spectrometer," *Food Chem*, vol. 197, pp. 1207–1214, 2016.

[41] T. Nordey, J. Joas, F. Davrieux, M. Chillet, and M. Léchaudel, "Robust NIRS models for non-destructive prediction of mango internal quality," *Sci. Hortic,* vol. 216, pp. 51–57, 2017.

Jetsada Posom, photograph and biography not available at the time of publication.

Navavit Soonnamtiang, photograph and biography not available at the time of publication.

Patcharapong Kotethum, photograph and biography not available at the time of publication.

Pakhpoom Konjun, photograph and biography not available at the time of publication.

Panmanas Sirisomboon, photograph and biography not available at the time of publication.

Khwantri Saengprachatanarug, photograph and biography not available at the time of publication.

Seree Wongpichet, photograph and biography not available at the time of publication.