

## Empirical approach to improve the prediction of soluble solids content in mango using near-infrared spectroscopy

Phuangsombut, K., Phuangsombut, A. and \*Terdwongworakul, A.

*Department of Agricultural Engineering, Faculty of Engineering at Kamphaengsaen, Kasetsart University, Kamphaengsaen, Nakhonpathom, Thailand*

### Article history

Received: 3 July 2019  
Received in revised form:  
17 January 2020  
Accepted:  
2 March 2020

### Keywords

mango,  
empirical approach,  
peel effect,  
soluble solids content,  
near infrared spectroscopy

### Abstract

The present work studied the effect of peel on the performance of near-infrared (NIR) spectroscopy for estimating the soluble solids content (SSC) of mangoes and on SSC estimation improvement using the empirical approach developed from our previous work. The results showed that a partial least squares (PLS) regression model based on intact fruit or unpeeled NIR spectra had lower accuracy (correlation coefficient,  $r = 0.84$ ; root mean squared error of prediction, RMSEP = 1.50 °Brix) than using flesh or peeled fruit NIR spectra ( $r = 0.88$ ; RMSEP = 1.27 °Brix). Improved prediction was obtained ( $r = 0.87$  and RMSEP = 1.36 °Brix) by applying the empirical approach (which partially removed the difference between the flesh and the peel, and the flesh spectra) to the original peel and flesh spectrum prior to developing the PLS model. The empirical approach proved beneficial in improving the prediction of the SSC in mango fruit which was affected by the peel.

© All Rights Reserved

### Introduction

The quality of mango (*Mangifera indica* L.) fruit for consumption is determined by the sweetness of the flesh. One of the sweetness indicators of the mango is the soluble solids content (SSC) (Lakshminarayana, 1980). Properly harvested mango fruit at the mature stage has attained a favourable sweetness when ripe (Medlicott *et al.*, 1988). On the other hand, low sweetness is obtained from immature mango fruit. Therefore, assessment of sweetness in terms of the SSC for mango fruit, especially in a non-destructive way, is very important from the consumer viewpoint.

Near-infrared (NIR) spectroscopy has been studied for non-destructive evaluation of soluble solids in many intact fruits such as cantaloupe (Dull *et al.*, 1989), apple (Ventura *et al.*, 1998; Park *et al.*, 2003), kiwi fruit (Martinsen and Schaare, 1998), and tomatoes (Slaughter *et al.*, 1996). The Brix value for mango was reported to be predicted with good accuracy with a standard error of prediction (SEP) of 0.40 °Brix (Saranwong *et al.*, 2003a) and 0.67 °Brix for ripe fruit (Subedi *et al.*, 2007). Schmilovitch *et al.* (2000) assessed the use of NIR spectrometry to measure the physiological properties and quality indices of intact mango fruit including softening of the flesh, the SSC content, and acidity. They reported a high likelihood for NIR spectrometry to assess the TSS producing an SEP and coefficient of determination ( $r^2$ ) of 1.22 °Brix and 0.93, respectively.

The peel has been known to have an effect on light penetration into the flesh and thus influences the prediction performance of the internal chemical content. For example, Lu *et al.* (2000) compared predictions of the sugar content in unpeeled and peeled 'Empire' apples using NIR diffuse reflectance between 800 and 1700 nm. They reported better prediction of the sugar content for peeled apples (SEP = 1.95 °Brix) than for unpeeled apples (SEP = 2.54 °Brix). Another report on the effect of the peel in the prediction of the SSC in oranges using visible / NIR spectroscopy by Jamshidi *et al.* (2014) found that the model developed from intact NIR reflectance spectra in the range from 700 to 1000 nm yielded a slightly higher root mean squared error of prediction (RMSEP = 0.309 °Brix) than the peeled model (RMSEP = 0.289 °Brix). Krivoshiev *et al.* (2000) reported on a technique for the elimination of peel interference in visible / NIR spectroscopy for the non-destructive determination of the internal quality of fruits and vegetables. They showed that the flesh optical density of potato can be predicted non-destructively using the whole potato transmittance and the diffuse reflectance measurements at the irradiated site, and at an opposite site where the light is transmitted.

There has been a limited number of researches that have attempted to reduce the effect of the peel in order to improve the prediction performance of the chemical content. In our previous work, we introduced an empirical technique to enhance the model

\*Corresponding author.  
Email: fengant@ku.ac.th

performance in predicting the dry matter content in durian fruit by using the NIR reflectance of both partially peeled fruit and the flesh of the fruit (peeled fruit) (Phuangsoambut *et al.*, 2018). The developed empirical approach reduced the difference between the peeled fruit and unpeeled fruit spectra using information on the flesh dry matter content. In the present work, we applied this empirical technique to prove its potential in improving the prediction performance for the soluble solids in mango using the NIR reflectance of peeled and unpeeled mango fruit.

## Materials and methods

A total of 100 mango fruits (cv. Nam Dokmai) were purchased from three different local markets during a five-month period to obtain variation in the SSC. The sample had means ( $\pm$  standard deviation) of weight, length, width, and thickness of  $355.08 \pm 16.92$  g,  $151.31 \pm 4.92$  mm,  $70.81 \pm 2.10$  mm, and  $51.61 \pm 1.85$  mm, respectively. The condition of the separate fruits in the sample was a mixture of early ripe with a yellow-green surface colour and ripe with a yellow surface colour. All fruits were transported to the laboratory and kept at  $25^{\circ}\text{C}$  in an air-conditioned laboratory overnight to attain temperature equilibrium prior to carrying out measurements on the following day.

### *Spectral acquisition*

Scanning of each sample was carried out using a portable spectrophotometer (FQA-NIR GUN, FANTEC Research Institute, Kosai, Japan) in the range of 600 - 1100 nm with a resolution of 2 nm in interactance mode. The integration time of the spectrophotometer was set at 30 ms for optimal absorbance. After scanning every ten fruits, a Teflon disc was scanned and used as a reference.

Each fruit was scanned to collect reflectance spectra at the middle location of the cheek on opposite sides. The reason that the cheek was selected for scanning was because in the future production line application, each mango would be placed on a moving conveyor and the cheek would be presented upward for scanning in a stable orientation. At each location, the scan was conducted three times and the averaged spectrum was computed for each location. Therefore, a total of 200 spectra were acquired from all intact or unpeeled fruit which contained both peel and flesh absorption. After scanning completion of the intact or unpeeled sample, a small portion of the peel (about 0.5 mm thick at the scanned location) was then sliced off using a slicer machine (Globe, Dayton, OH, USA). The flesh or the peeled fruit at the peeled location was scanned to obtain the NIR flesh spectrum. The acquired flesh

spectra were used to develop a prediction model of the SSC of the flesh which was used as the reference model for maximum potential accuracy.

### *Measurement of soluble solids content*

Immediately following the NIR measurements, the flesh at the peeled location of each fruit was removed and its juice was extracted. The juice was filtered and measured for SSC in  $^{\circ}\text{Brix}$  using a digital refractometer (PR-32, Palette Series, Atago Co. Ltd., Tokyo, Japan). The average of two readings was used as a representative value for each location.

### *Development of prediction models*

All samples were divided into calibration and prediction sets, having a range of similarly uniform SSC values by ranking the samples based on the SSC and alternately selecting into the calibration and prediction sets with an equal number of samples in each set. The minimum, maximum, and average values of the SSC in the calibration set were 10.00, 21.30 and  $14.89^{\circ}\text{Brix}$ ; and in the prediction set were 10.10, 20.60 and  $14.83^{\circ}\text{Brix}$ , respectively.

First, the flesh-SSC prediction model was built based on the NIR spectral data of the flesh from the calibration set. All the SSC prediction models were created using partial least squares regression (PLSR) (The Unscrambler v. 9.8, Camo, Oslo, Norway). The spectra, which were used in the model development, were pre-treated with an individual pre-treatment or a combination of the pre-treatments including the standard normal variate (SNV), the multiplicative scatter correction, and the second derivative to reduce the effect of scattering. The number of PLS factors used in the model development was obtained using leave-ten-out cross validation. The developed model was validated on the prediction set and the performance was judged based on the correlation coefficient ( $r$ ) and root mean squared error of prediction (RMSEP). This flesh model was regarded as a reference model.

The second model was also developed using the NIR spectral data from the unpeeled samples to compare the accuracy of SSC prediction.

### *Empirical approach to improve spectral data*

In the present work, the empirical approach from our previous work (Phuangsoambut *et al.*, 2018) was applied to adjust the NIR spectrum of the unpeeled mango to match the NIR spectrum of the mango flesh as closely as possible, since the flesh spectrum was regarded as the reference that could be used to predict the SSC with maximum accuracy.

Practically, this approach used the SSC value

to adjust the unpeeled fruit or the peel and flesh (PF) spectra to be similar to the flesh spectrum (F). This was achieved by initially determining the difference between the unpeeled ( $A_{PF}$ ) and flesh ( $A_F$ ) spectrum by simple subtraction. Then the relationship between the obtained difference in spectrum ( $\Delta A$ ) and the SSC was developed at each wavelength using regression analysis. It was hypothesised that the  $\Delta A$  was dependent on wavelength. Therefore, the SSC was needed to estimate the  $\Delta A$ . The PLS model based on the PF spectrum was used to predict the SSC which was then applied to the regression equation to estimate the  $\Delta A$ . Finally, the  $\Delta A$  was added to the original  $A_{PF}$  to adjust it to be equivalent to  $A_F$ . The steps were as follows:

i. The difference spectrum ( $\Delta A$ ) of the sample for adjustment of the original PF spectrum was calculated by subtraction of the PF spectrum ( $A_{PF}$ ) from the flesh spectrum ( $A_F$ ), as shown in Eq. 1.

$$\Delta A = A_F - A_{PF} \quad (\text{Eq. 1})$$

ii. The  $\Delta A$  value was deemed to represent the peel absorbance. However, the  $\Delta A$  value appeared to vary with wavelength. Therefore, at each wavelength, there needed to be a corresponding equation to predict the  $\Delta A$  value. The investigation by Jha *et al.* (2013) showed that the peel attributes were affected by the ripening period and stage of harvesting. The SSC of the flesh was also related to the harvesting stage and ripening period. Thus, there was a relationship between the flesh SSC and the peel properties. Therefore, the SSC of each sample could be used to predict the  $\Delta A$  value. The SSC value of each sample could be predicted by a PLSR model which was developed from the SSC and the original PF spectra of all samples. So, at each wavelength, an equation for prediction of the  $\Delta A$  value from the SSC was created using regression analysis as shown in Eq. 2.

$$\Delta A_j = b_{0j} + b_j (\text{SSC}) \quad (\text{Eq. 2})$$

where,  $b_{0j}$  and  $b_j$  = regression coefficients. Thus, the number of regression equations relating the SSC with  $\Delta A$  was the same as the number of wavelengths.

iii. The original PF spectrum was adjusted one wavelength at a time. So, at each wavelength, J, initially the PLSR model was used to predict the SSC of the sample based on the original PF spectrum of the sample. The predicted SSC was then substituted into Eq. 2 to compute the  $\Delta A_j$ . Finally, the original

PF spectrum ( $A_{PFj}$ ) was adjusted to the adjusted PF spectrum ( $A_{APFj}$ ) by substituting the obtained  $\Delta A_j$  and  $A_{PFj}$  into Eq. 3.

$$A_{APFj} = A_{PFj} + \Delta A_j \quad (\text{Eq. 3})$$

iv. For adjustment of the original PFS spectrum at the remaining wavelengths, the above-mentioned steps were repeated using the same SSC of the sample.

PLSR was applied subsequently to the adjusted PF spectra to create the SSC prediction model, and the performance was compared with that obtained from the PLSR models based on the original PF spectrum and the flesh spectrum.

### Statistical analysis

Significant relationship between the actual SSC and NIR predicted SSC was tested for each regression model by the *F*-test of the overall significance which assessed the regression coefficients mutually at 0.05 significance level. In addition, the 95% prediction limits were also calculated around the NIR predicted SSC for each actual SSC. The NIR predicted SSC corresponding with the actual SSC lied within the prediction limits with 95% probability.

## Results and discussion

### Spectral characteristics

The difference between the spectral absorbance of the peel and flesh (intact or unpeeled), and of the flesh (peeled) of mango fruit is illustrated in Figure 1. Generally, both spectra were characterised by broad absorption peaks at around 650, 740, and 970 nm. In the visible region, the peak at 650 nm was close to the chlorophyll band (Zude, 2003) due to the colour and pigments of the peel. The strong absorption peak at 970 nm and the broad peak at around 740 nm were in the vicinity of water absorption wavelengths (Workman Jr. and Weyer, 2012), which were associated with the water as the main component in the fruit.

In Figure 1, the intact mango (peel and flesh) absorbed more light in both the visible and NIR regions than mango flesh. This finding was in contrast to previous reports that the intact fruit had lower absorption than flesh in the NIR region (Lu *et al.*, 2000; Jamshidi *et al.*, 2014) because of the presence of peel. The contrary result might have been due to the fact that the mango peel caused lower scattering and less reflection than the mango flesh.

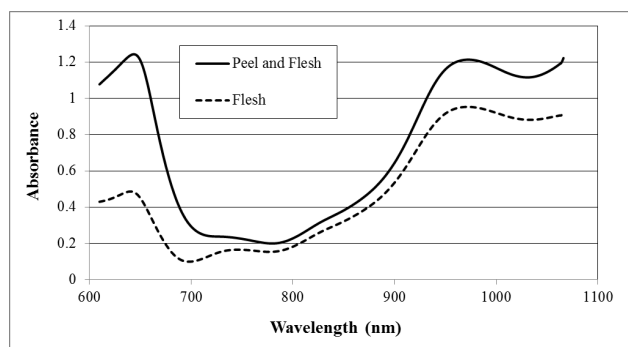


Figure 1. Examples of absorbance spectra for peel and flesh (unpeeled fruit) and flesh (peeled fruit) of mangoes.

#### Prediction of soluble solids content

All the optimal SSC prediction models were achieved using the spectra pre-treated with the SNV before being applied for model development. Table 1 shows the prediction performance for the SSC for the PLSR models developed from the SNV flesh (F model), the SNV peel and flesh (PF model), and the SNV adjusted peel and flesh spectra (APF model).

Table 1 summarises the statistical results for prediction of the SSC using the PLSR model developed from the flesh, from the peel and flesh, and from the adjusted peel and flesh spectra pre-treated with SNV. As expected, the F model gave the best accuracy for the SSC prediction ( $r = 0.88$  and  $RMSEP = 1.27$  °Brix) with good fitted data between the actual and the NIR predicted SSC as shown in Figure 2(a). When the PF model was used to predict the SSC, the prediction errors increased ( $r = 0.84$  and  $RMSEP = 1.50$  °Brix) as shown in Table 1 and Figure 2(b). The results agreed with previous reports on the determination of the sugar content of apples (Lu *et al.*, 2000) and oranges (Jamshidi *et al.*, 2014), that better prediction was obtained from the peeled fruit spectra or flesh spectra. The peel affected the

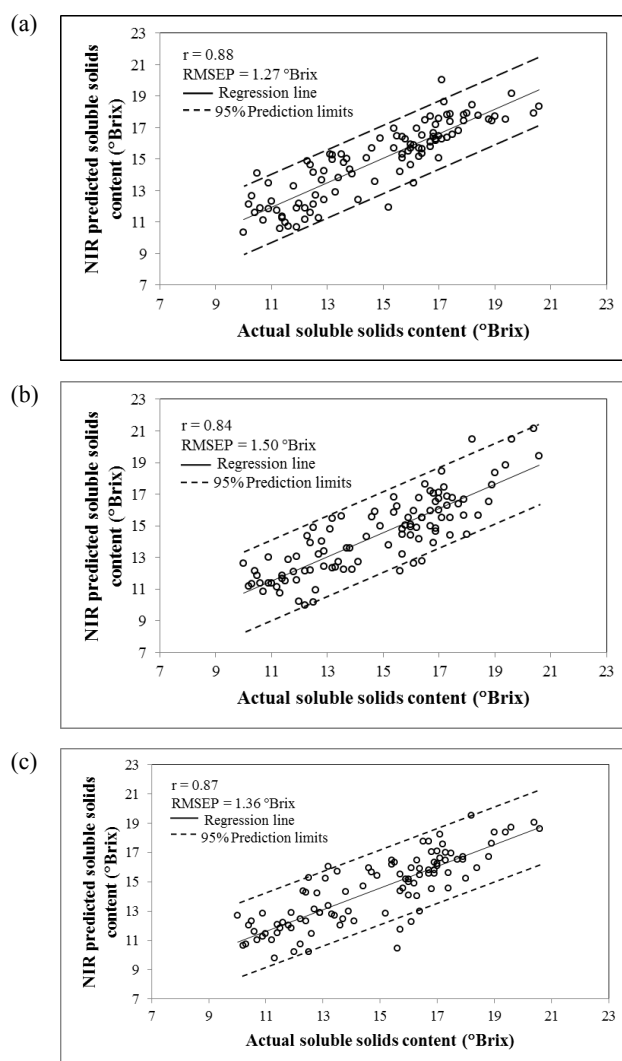


Figure 2. Prediction of soluble solids content using near infrared spectra based on (a) flesh, (b) peel and flesh, and (c) adjusted peel and flesh spectra.

Table 1. Summary of prediction performance of optimal calibration models developed from flesh spectra, peel and flesh spectra, and adjusted peel and flesh spectra for evaluation of soluble solids content. All spectra were pre-treated with the standard normal variate prior to partial least squares analysis.

NIR spectrum used to develop model	Number of factors	Calibration		Prediction		RPD <sup>e</sup>	Sig. <sup>f</sup>
		$r^b$	RMSEC <sup>c</sup> (°Brix)	$r$	RMSEP <sup>d</sup> (°Brix)		
SNV <sup>a</sup> flesh spectra or peeled fruit spectra	10	0.94	0.94	0.88	1.27	2.08	$5.1 \times 10^{-33}$
SNV peel and flesh spectra or unpeeled fruit spectra	10	0.89	1.25	0.84	1.50	1.82	$3.8 \times 10^{-28}$
SNV adjusted unpeeled-fruit spectra	10	0.91	1.14	0.87	1.36	1.94	$9.3 \times 10^{-29}$

<sup>a</sup>Standard normal variate (SNV); <sup>b</sup>Correlation coefficient; <sup>c</sup>Root mean squared error of calibration (RMSEC); <sup>d</sup>Root mean squared error of prediction (RMSEP); <sup>e</sup>Standard deviation / SEP; and <sup>f</sup>Regression significance of the ratio of the mean regression sum of squares to the mean error sum of squares.

prediction results by restricting light penetration into the flesh in which the SSC was predicted. The ratio of the standard deviation to the standard error of prediction (RPD; Williams, 2007) as a measure of the accuracy of the SSC prediction by the PF model (RPD = 1.77) was comparable to the result of previous research (RPD = 1.78; Saranwong *et al.*, 2003b) although different varieties of mango were used.

Then, the empirical approach as described in the previous section was applied to adjust the PF spectrum to match the F spectrum. The results of adjustment using the empirical approach are shown in Figure 3. It could be observed that originally the PF spectrum was quite different from the F spectrum. When applying the empirical approach to the PF spectrum, it was clear that the PF spectrum was changed to the adjusted PF or APF spectrum which was closer to the F spectrum as exemplified by a sample for comparison in Figure 3. This showed that the empirical approach was acceptably successful in removing the difference between the F and PF spectra and adjusting the PF spectrum to be relatively similar to the flesh spectrum.

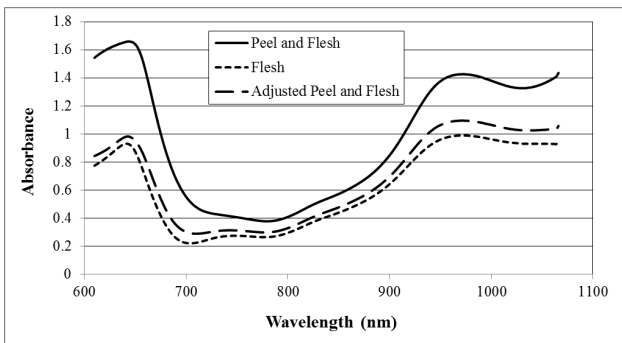


Figure 3. Example of comparison between flesh, peel and flesh, and adjusted peel and flesh spectra using the empirical approach.

The APF spectra were next used to develop another SSC prediction model using PLS and the prediction performance was evaluated using comparison with the PF model. As shown in Table 1, the APF-spectral-based model provided improvement in the prediction accuracy ( $r = 0.87$  and  $RMSEP = 1.36$  °Brix) over the model based on the PF spectrum ( $r = 0.84$  and  $RMSEP = 1.50$  °Brix).

The differences between the RMSEP values of the PF model and the F model, as well as the RMSEP differences from the APF and F models were comparatively tested using the methodology of Snedecor and Cochran (1989). The methodology is based on the calculation of the lower and upper limits of 95% confidence and tests whether the interval includes one. The results showed that RMSEP values

from the F model and the PF model were significantly different. However, there was no significant difference between the RMSEP values of the F and the APF models. This significant test of difference between RMSEP values justified the advantage of the empirical approach in improving both the PF spectra and the prediction performance.

The APF model had an RPD (standard deviation / RMSEP) of 1.96 which was a 10.3% increase from the RPD value of 1.77 for the PF model. In our previous work on durian (Phuangsoambut *et al.*, 2018), the empirical approach increased (17.9%) the RPD from 1.55 in the PF model to 1.83 in the APF model. The difference in the percentage increase in the RPD suggested that this empirical approach offered more advantage for fruit with a thicker peel. The peel thickness of durian in the previous research was 2 - 3 mm as compared to 0.8 - 1.0 mm for mangoes.

The regression coefficient was explored on the APF spectral-based model as shown in Figure 4. The maximum at around 1038 nm was in the proximity of the absorption wavelength (1042 nm) of a combination of 2 C-H stretching and 2 C-H deformation vibrations of oil (Osborne and Fearn, 1986). The second highest coefficient was observed at around 1000 nm which corresponded with one of the wavelengths used in the calibration equation for the prediction of the SSC in Philippine mango fruits of the Carabao variety (Saranwong *et al.*, 2001).

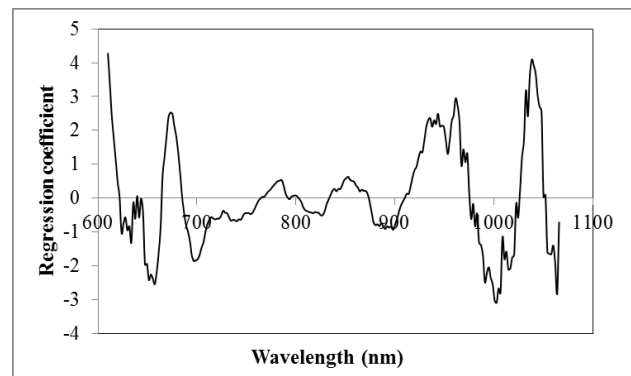


Figure 4. Regression coefficient of prediction model for soluble solids content based on adjusted peel and flesh absorbance.

## Conclusion

The prediction of the SSC in mango fruit using the model based on the peel and flesh or on the intact fruit NIR spectra had a lower accuracy ( $RMSEP = 1.50$  °Brix) than that using the peeled fruit or the flesh NIR spectra ( $RMSEP = 1.27$  °Brix). Application of the empirical approach, which aimed to remove the difference between the peel and flesh

spectra, and the flesh spectra, proved helpful in improving both the original peel and flesh spectra, and the prediction performance (RMSEP = 1.36 °Brix). However, the improvement in prediction performance of the SSC was lower than that for the prediction of the dry matter content in the thicker peel of durian fruit. Further investigation is needed to compare the benefit of this empirical approach between the same kinds of fruit with different varieties having different peel of thickness.

#### Acknowledgement

The authors would like to acknowledge the Kasetsart University Research and Development Institute (KURDI), Bangkok, Thailand for financially supporting the present work, and the Centre of Advanced Studies in Industrial Technology, Faculty of Engineering, Kasetsart University, Bangkok, Thailand for the research facility.

#### References

- Dull, G. G., Birth, G. S., Smittle, D. A. and Leffler, R. G. 1989. Near infrared analysis of soluble solids in intact cantaloupe. *Journal of Food Science* 54(2): 393-395.
- Jamshidi, B., Minaei, S., Mohajerani, E. and Ghassemian, H. 2014. Prediction of soluble solids in oranges using visible / near-infrared spectroscopy: effect of peel. *International Journal of Food Properties* 17(7): 1460-1468.
- Jha, S. N., Jaiswal, P., Narsaiah, K., Kaur, P. P., Singh, A. K. and Kumar, R. 2013. Textural properties of mango cultivars during ripening. *Journal of Food Science and Technology* 50(6): 1047-1057.
- Krivoshiev, G. P., Chalucova, R. P. and Moukarev, M. I. 2000. A possibility for elimination of the interference from the peel in nondestructive determination of the internal quality of fruit and vegetables by VIS / NIR spectroscopy. *LWT - Food Science and Technology* 33(5): 344-353.
- Lakshminarayana, S. 1980. Mango. In Nagy, S. and Shaw, P. E. (eds). *Tropical and Subtropical Fruits: Compositions, Properties and Uses*. United States: AVI Publishing.
- Lu, R., Guyer, D. E. and Beaudry, R. M. 2000. Determination of firmness and sugar content of apples using near-infrared diffuse reflectance. *Journal of Texture Studies* 31(6): 615-630.
- Martinsen, P. and Schaare, P. 1998. Measuring soluble solids distribution in kiwifruit using near-infrared imaging spectroscopy. *Postharvest Biology and Technology* 14(3): 271-281.
- Medlicott, A. P., Reynolds, S. B., New, S. W. and Thompson, A. K. 1988. Harvest maturity effects on mango fruit ripening. *Tropical Agriculture* 65(2): 153-157.
- Osborne, B. G. and Fearn, T. 1986. *Near infrared spectroscopy in food analysis*. United Kingdom: Longman Scientific and Technical.
- Park, B., Abbott, J. A., Lee, K. J., Choi, C. H. and Choi, K. H. 2003. Near-infrared diffuse reflectance for quantitative and qualitative measurement of soluble solids and firmness of delicious and gala apples. *Transactions of the ASAE* 46(6): 1721-1732.
- Phuangsoambut, K., Phuangsoambut, A., Talabnark, A. and Terdwongworakul, A. 2018. Empirical reduction of rind effect on rind and flesh absorbance for evaluation of durian maturity using near infrared spectroscopy. *Postharvest Biology and Technology* 142: 55-59.
- Saranwong, I., Sornsrivichai, J. and Kawano, S. 2001. Improvement of PLS calibration for Brix value and dry matter of mango using information from MLR calibration. *Journal of Near Infrared Spectroscopy* 9(4): 287-295.
- Saranwong, I., Sornsrivichai, J. and Kawano, S. 2003a. Performance of a portable near infrared instrument for Brix value determination of intact mango fruit. *Journal of Near Infrared Spectroscopy* 11(3): 175-181.
- Saranwong, I., Sornsrivichai, J. and Kawano, S. 2003b. On-tree evaluation of harvesting quality of mango fruit using a hand-held NIR instrument. *Journal of Near Infrared Spectroscopy* 11(4): 283-293.
- Schmilovitch, Z., Mizarach, A., Hoffman, A., Egozi, H. and Fuchs, Y. 2000. Determination of mango physiological indices by near-infrared spectrometry. *Postharvest Biology and Technology* 19(3): 245-252.
- Slaughter, D. C., Barrett, D. and Boersig, M. 1996. Nondestructive determination of soluble solids in tomatoes using near infrared spectroscopy. *Journal of Food Science* 61(4): 695-697.
- Snedecor, G. W. and Cochran, W. G. 1989. *Statistical methods*. 8<sup>th</sup> ed. United States: Iowa State University Press.
- Subedi, P. P., Walsh, K. B. and Owens, G. 2007. Prediction of mango eating quality at harvest using short-wave near infrared spectrometry. *Postharvest Biology and Technology* 43(3): 326-334.
- Ventura, M., De Jager, A., De Putter, H. and Roelofs, F. P. M. M. 1998. Non-destructive determination of soluble solids in apple fruit by near infrared

spectroscopy (NIRS). *Postharvest Biology and Technology* 14(1): 21-27.

Williams, P. 2007. *Near-infrared technology - getting the best out of light*. Edition 5.0. A short course in the practical implementation of near-infrared spectroscopy for the user. Canada: PDK Grain.

Workman Jr., J. and Weyer, L. 2012. *Practical guide and spectral atlas for interpretive near-infrared spectroscopy*. 2<sup>nd</sup> ed. United States: CRC Press.

Zude, M. 2003. Comparison of indices and multivariate models to non-destructively predict the fruit chlorophyll by means of visible spectrometry in apple fruit. *Analytica Chimica Acta* 481(1): 119-126.