

Article

Estimation of Sugar Content in Sugarcane (*Saccharum spp.*) Variety Lumpang 92-11 (LK 92-11) and Khon Kaen 3 (KK 3) by Near Infrared Spectroscopy

Khanita Kamwilaisak^{1,a,*}, Pasakorn Jutakridsada¹, Wimonporn Iamamornphanth¹, Khwantri Saengprachatanarug², Pornnapa Kasemsiri¹, Supasit Konyai², Jetsada Posom², and Prinya Chindaprasirt^{3,*}

¹ Department of Chemical Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen, 40002, Thailand

² Department of Agricultural Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen, 40002, Thailand

³ Department of Civil Engineering, Faculty of Engineering, Sustainable infrastructure Research and Development Center, Khon Kaen University, Khon Kaen, 40002, Thailand

E-mail: ^{a,*}khanita@kku.ac.th (Corresponding author)

Abstract. In this study, a non-destructive measuring method, near-Infrared (NIR) technique was used to evaluate the quality of sugarcane. Two sugarcane (*Saccharum spp.*) varieties viz., Lumpang 92-11 and Khon Kaen 3 were chosen for the test. The samples were collected for 3 years. The sugar contents were measured in terms of °Brix, %Pol, %Fiber, and Commercial Cane Sugar (CCS) values using the NIR technique and conventional laboratory testing for comparison. The Partial Least Squares Regression (PLSR) model was performed using 400 samples for each variety. The NIR models showed the coefficient of determination (R^2) of 0.97, 0.90, 0.86 and 0.82 for °Brix, %Pol, %Fiber and CCS, respectively with the corresponding root mean square error of prediction (RMSEP) of 0.246, 0.512, 0.353 and 0.542. The results indicated that the modelling using °Brix gave the best estimation with the highest R^2 and lowest RMSEP, indicating high accuracy and reliability. The modelling with %Pol and %Fiber gave the moderate estimations and that with CCS value gave the lowest accuracy. However, all the four modelling predictions were within the acceptable range and could thus be used in crops trading instead of the traditional method. It was more reliable, quicker, more comfortable and more environmentally friendly than the traditional method as it did not involve the use of the chemical.

Keywords: Near infrared spectroscopy, sugarcane, non-destructive, partial least squares regression.

ENGINEERING JOURNAL Volume 25 Issue 3

Received 16 May 2020

Accepted 30 January 2021

Published 31 March 2021

Online at <https://engj.org/>

DOI:10.4186/ej.2021.25.3.69

1. Introduction

Sugarcane (*Saccharum officinarum*) is considered an industrial plant because its crop yields food and bioenergy products. It is a fast-growing plant whose chief product is sugar. About 6% of the world's total sugarcane production comes from Thailand (103 Mt/year in 2016) with the corresponding values of approximately 18 million dollars per year. Also, the most popular sugarcane varieties viz., Lumpang 92-11 (LK 92-11) and Khon Kaen 3 (KK3) originated from Thailand [1]. These two sugarcane varieties are suitable for various soil conditions and have higher production levels compared to other varieties. It has been reported that the LK 92-11 and KK 3 productions were 35 tons/acre (approximately 5.13 CCS/acre), while the other varieties produce only 33 tons/acre (approximately 3.28 CCS/ acre).

The commercial cane sugar (CCS) value is one of the critical parameters used to determine the price of sugarcane for trading [2] based on the sugar content parameters viz., °Brix, %Pol, and %Fiber. Traditionally, sugar content is measured in laboratories on sugarcane juice samples [3]. Chemical reagents were added to the sugarcane juice to clarify the sample before measuring °Brix and %Pol. °Brix indicates the sugar content as a percentage of soluble solid in a liquid measured by the refraction index of light as it passes through the liquid sample [3]. %Pol is the percentage of sucrose, determined by the optical activity property obtained from polarized light as it passes through the sample. %Fiber is obtained from the dry weight of sugarcane after the juice is extracted [3]. Furthermore, the laboratory methods of °Brix, %Pol and %Fiber are time-consuming and require chemical reagents [4].

However, the use of near-infrared spectroscopy (NIR) technique has been growing within the chemometrics circle as it is a non-destructive, non-chemical with the capability for fast analysis and high accuracy [5-7]. NIR provides direct analysis of samples with light scattering at a wavelength of 700-2500 nm [8]. However for NIR, the measurement of the liquid sample has limitation since the liquid sample contains water molecule with broad vibration bands at 1000-1300 nm [5]; OH group bonded absorbing at 1160 nm, OH group in H-bonded absorbing at 1200 nm [9]. The water molecule with two OHs and H-bonded absorbing can interfere with other OH groups in the sample and affects the measurement accuracy [10, 11]. The application of NIR technique for °Brix and %Pol measurement in the sugar industry has been extended from laboratory measurement to portable type measurement [12, 13].

Furthermore, compared to the traditional method, NIR measurement saves time and chemical usage.

Moreover, the accuracy of NIR measurement is improving from the extensive studies of various factors, including many samples, moisture content, and measuring temperature [14-17]. However, sugarcane varieties affect the measurement accuracy because they contain different fiber, juice and sugar contents [18]. To provide excellent

accuracy and high precision with the NIR measurement, the correlations between the NIR range and °Brix and %Pol in sugarcane are essential. The chemometric algorithms of partial least-squares regression (PLSR) method are created to predict °Brix and %Pol in sugarcane [19]. The NIR measurements of °Brix, %Pol, %Fiber and CCS have been reported by Sanseechan et al. (2018). However, the influence of sugarcane varieties on the NIR measurement has not yet been reported.

Thus, this work aims to develop the prediction models using the NIR technique to evaluate the sugar contents viz., °Brix, %Pol, %Fiber, and CCS. Furthermore, the prediction model of LK 92-11 and KK 3 are compared. The calibration and validation were optimized from the 400 samples of each variety. It is expected that this work could improve the NIR measurement resulting in the better alternative method for the estimation of sugar cane quality without the use of chemicals.

2. Materials And Methods

2.1. Preparation of Sugarcane Sample

The procedure in experiment was demonstrated in Fig. 1. Two hundred samples of sugarcane LK 92-11 and KK 3 each were collected from the Official Cane and Sugarcane Board of Thailand and sugarcane growers for 3 years from 2016-2019. All sugarcane was at the age of 11-13 months. Topped, root, and leaf of cane stalks were eliminated by hand cutting. The cleaned stalk samples were chopped in the length of 50 cm. The chopped cane stalks were crushed by a shatter machine and passed through a sieve to obtain sizes between 2-4 mm. The shattered sugarcane was ready to sue in the next experiment

2.2. NIR Measurement

The shattered sample was analyzed using an FT-NIR spectrometer (MPA, Bruker Optics, Ettinger, Germany). Six hundred grams of shattered sugarcane was placed in the rotating drum having 244 mm in diameter (see Fig. 1). The analysis was conducted in diffuse reflectance mode at 25 ± 3 °C. The spectral was obtained between 3600 and 12500 cm^{-1} (800-2777 nm), and the samples were scanned 32 times at a resolution of 32 cm^{-1} . Before sample scanning, as a reference material which is the gold plate used was scanned for background. It was assumed that the gold material provides 100% reflection. The NIR spectra was recorded in term of absorption which was calculated by $\log(1/R)$, where R was reflectance value of the sample. Each sample was scanned in duplicate and the averaged spectra was used for data analysis. These data are the independent variables. The scanned sample spectra were analyzed and recorded in duplicate. OPUS 7.8 was used to collect the spectral data, and the data was saved as an OPUS file. After the scanning, the shattered sugarcane was used to determine °Brix, %Pol, %Fiber, and CCS.

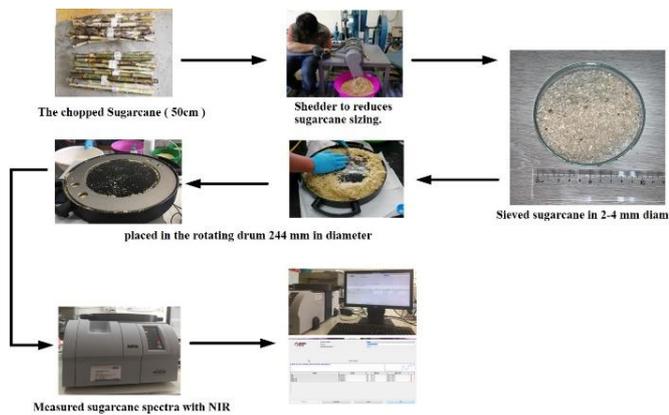


Fig. 1. NIR data collecting process.

2.3. Reference Methods

The sugarcane juice was measured °Brix and %Pol following the methods described by the International Commission for Uniform Methods of Sugar Analysis (ICUMSA). Two hundred grams of shattered cane was pressed to extract a juice. The °Brix was measured by using a refractometer (RX 5000 CX, ATAGO, Japan) at the °Brix range of 0-50% at 20 ± 3 °C. For %Pol measurement, 50 mL of sugarcane juice was mixed with 8 mL of 200 g/L aluminium sulfate solution ($\text{Al}_2(\text{SO}_4)_3 \cdot 18\text{H}_2\text{O}$), and 5 mL of 2 M sodium hydroxide (NaOH) solution in a volumetric flask. The solution was then filled with distilled water to 200 mL. The obtained solution was agitated and filtered through filter paper (45 mm in diameter, 11 μm in pore size, and 180 μm in thickness). The filtered solution was used to determine the degree of optical rotation (OR) using a polarimeter and polarization (POLAX 2L, ATAGO, Japan). %Pol was calculated from °Brix and OR based on Engelke's method [3] following Eq. (1):

$$\%Pol = \frac{-6.517 + (25.3 \times OR) - 0.0118(OR)^2 + (2.937 \times \text{°Brix}) - 0.207(\text{°Brix})^2}{100} \quad (1)$$

where OR is the optical rotation, and °Brix is the concentration of soluble solids in sugarcane juice (%). %Fiber (FC%) is determined following Method 4 of BSES (2001). The shattered sugarcane was placed into a cotton cambric bag and then placed in bathwater at 60 °C for 30 min for sugar extraction. The bag was then washed for three cycles in an automatic washing machine and dried at 100-105°C for 36 h until a constant weight was obtained. The fiber content for each sample was calculated following the mathematical method proposed by Watson et al. (1999):

$$\%Fiber = \frac{(100^2 - 100 \times \text{moisture}\% - 97 \times \text{°Brix})}{(100 + \text{°Brix})} \quad (2)$$

The CCS value was then calculated from %Pol, %Brix and %Fiber according to the formula given below:

$$(CCS) = \frac{0.9433(\%Pol)(100 - FC\%)}{100} - \frac{1}{2} \left(\frac{0.9660(\text{°Brix})(100 - \%Fiber)}{100 - 0.9433(\%Pol)} - \frac{(100 - \%Fiber)}{100} \right) \quad (3)$$

The precision of measured values (°Brix, %Pol, %Fiber and CCS) were indicated by using the repeatability (Rep) value as the reference data. The Rep value was calculated from the standard deviation of the different results of the reference methods in duplication. The repeatability was used to calculate the maximum coefficient of determination (R_{max}^2) using Eq. (4).

$$R_{\text{max}}^2 = \frac{(SD_y^2 - \text{Rep}^2)}{SD_y^2} \quad (4)$$

where SD_y is the standard error of reference value.

The rep indicates the precision of the laboratory reference methods. The R_{max}^2 is the means to the maximum possible coefficients of examinations when no error from the spectral collection is present [12]. For example, if R_{max}^2 is higher than 0.90 is defined to goodness fitting of reference method [1], mean that the reference method is precise.

2.4 Spectral Pre-processing and Model Development

In order to increase the precision of modelling, it is imperative that the obtained data needed to be treated before modelling [20]. The NIR spectra were thus pretreated using constant offset elimination, straight-line subtraction, vector normalization, min-max normalization, MSC (multiplicative scatter correction), first derivatives, second derivatives, first derivatives + straight line subtraction, first derivatives + vector normalization and first derivatives + MSC. The spectral pre-processing was used based on the previous study by Ramírez-Morales, Rivero [13].

The data analysis was conducted to establish a correlation between the response matrix composed of the reference values (°Brix, %Pol, %Fiber, and CCS) and the spectral matrix (reflectance measurements), with the number of lines equalled to the number of sugarcane samples, and the number of columns equalled to the number of wavelengths. The OPUS software (Bruker, Germany) was used for partial least squares regression (PLSR) calibration and validation. The validation samples (test set) were obtained from the same sample set at 50% of the total samples.

PLSR is a multivariate statistical covariance-based method used to optimize both the variance and correlation with the response variables. Within this algorithm, an orthogonal basis of latent variables is constructed one by one, so they are positioned along the direction of maximal covariance between the response and the spectral matrices. Thus, PLSR models find new variables, which are the estimates of latent variables or their rotations. The new variables are named X-scores and are predictors of the responses. Deviations between the

measured and predicted responses are also generated and named Y-residuals [21]. The final model is given by Eq. (5) as follows:

$$Y=B \cdot X+F \quad (5)$$

where Y is the matrix of response variables ($^{\circ}$ Brix, %Pol, %Fiber and CCS), X is the matrix of predictor variables or scores ($^{\circ}$ Brix, %Pol, %Fiber and CCS), B is the matrix of regression coefficients (dimensionless), and F is the matrix of Y-residuals ($^{\circ}$ Brix, %Pol, %Fiber and CCS).

The prediction models were evaluated by OPUS software (Bruker, Germany). Root mean square error of calibration (RMSEC), statistical parameters mean bias error (BIAS), the ratio of the standard error of prediction to standard deviation of reference values (RPD), and coefficient of determination (R^2) were the performance of NIR model calculated from Eq. (6), (7), (8) and (9), respectively:

$$\text{RMSEP and RMSEC} = \sqrt{\frac{\sum_{i=1}^i (y_i - y_{\text{pre}})^2}{i}} \quad (6)$$

$$\text{BIAS} = \frac{\sum_{i=1}^i (y_i - y_{\text{pre}})}{i} \quad (7)$$

$$\text{RPD} = \frac{\text{SD}}{\text{RMSEP}} \quad (8)$$

$$R^2 = \frac{\sum_{i=1}^i (y_i - y_{\text{pre}})^2}{\sum_{i=1}^i (y_i - \hat{y}_i)^2} \quad (9)$$

where y_i is the measured value from the traditional method ($^{\circ}$ Brix, %Pol, %Fiber and CCS), y_{pre} is the predicted value from NIR model, \hat{y}_i are the average values measured ($^{\circ}$ Brix, %Pol, %Fiber, and CCS), and i is the number of samples (dimensionless). SD is the means to standard deviation of Y_i .

3. Result and Discussion

3.1. Measured Value

Table 1 shows the rep and R^2_{max} of the measured value for two varieties including $^{\circ}$ Brix, %Pol, %Fiber, and CCS. It was found that the R^2_{max} of all parameters showed no significant difference in both sugarcane varieties. Also, the highest R^2_{max} of $^{\circ}$ Brix and %Pol were obtained at 0.997 for two sugarcane varieties, whereas the R^2_{max} of CCS and %Fiber content was at 0.993 and 0.962, respectively. It is important to note that the R^2_{max} of all sugar content parameters was above 0.95, which is an indicative factor of reliable predictive models [22].

Table 1. Repeatability of reference method and maximum R^2 for $^{\circ}$ Brix, %Pol, %Fiber and CCS.

Parameter	variety	SD	Repeatability	R^2_{max}
$^{\circ}$ Brix	KK	1.60	0.08	0.997
	LK	0.83	0.06	0.994
	KK and LK	1.42	0.07	0.997
%Pol	KK	1.75	0.08	0.997
	LK	1.13	0.08	0.995
	KK and LK	1.50	0.08	0.997
%Fiber	KK	1.15	0.26	0.950
	LK	1.13	0.22	0.962
	KK and LK	1.15	0.24	0.962
CCS	KK	1.92	0.16	0.993
	LK	1.16	0.10	0.992
	KK and LK	1.61	0.13	0.993

3.2. Spectral Information and Principal Components Analysis (PCA) Analysis

Figure 2 presents the raw NIR spectra of 80 sugarcane samples in each variety (KK and LK) in the range of 12500 to 3600 cm^{-1} . The NIR spectral graph shows the relation of the x-axis as the wavenumber in cm^{-1} and the y-axis is the absorbance value, both of which were obtained from the $\log(1/R)$. The NIR spectral graph of sugarcane variety KK is illustrated as a red line in Fig. 2a. The leading bands were between 6000 to 7500 cm^{-1} and 4000 to 5500 cm^{-1} , which is an O-H stretching and HOH bending combination of polysaccharides at 6940 cm^{-1} (1449 nm). These could be a sucrose molecule [23]. The second band shows a significant constituent of sucrose, glucose and fructose in sugarcane. The band at 5181 cm^{-1} (1930 nm) demonstrates the stretching of the second overtone of O-H, which corresponds to polysaccharides [23]. There were also prominent bands at 4252 cm^{-1} and 4019 cm^{-1} (2352 and 2488 nm). The band at 4252 cm^{-1} is assigned to C-H stretching in the polysaccharide molecule, and the band at 4019 cm^{-1} is a combination of C-H and C-C stretching [24]. The NIR spectra of the sugarcane variety LK is presented in Fig. 2b. It has similar results to the sugarcane variety KK, and no significant difference was noted. This is because sugarcane contains the same major sugar molecules, resulting in the absorbance value of both varieties to be within the same range of 1111 to 18181 cm^{-1} ; however, both sugarcane varieties could have different amounts of sugar content in each sample. Similar results have been reported by Neto et al. 2017 and Nawi et al. 2014 [14, 25] wherein both reported absorption bands of sugarcane in the wavelength region of 11111 to 18181 cm^{-1} , including famous bands near 4000 and 5296 cm^{-1} , which account for the overtone of CH stretching and OH stretching from polysaccharides [24]. Figure 2c demonstrated comparison raw spectral of two sugarcane varieties.

In order to gain a better insight into the influence of sugarcane varieties on NIR measurement. The correlate of NIR spectra from sugarcane varieties was subjected to PCA analysis which is performed using second derivative spectra. Figure 3 shows the performed PCA analysis from the second derivative spectra. The spectral pretreatment of derivative is used to solve the overlapping of NIR spectra. The cluster data was indicated the spectra of sugarcane KK and LK as displayed in black dot and red dot, respectively. The total variability was explained by the first two principal components (PCs), with PC1 and PC2 accounting for 50 and 50 % of the variation in data, respectively. PCA analysis was showed a distribution of sugarcane spectra data. It was found that the scatter plot of two sugarcane varieties performed as a cluster was in a dot line circle. The cluster of data was indicated that similar properties of two sugarcane varieties. As the result of NIR spectra, it was found that the model may be developed using combination spectra of two varieties. The non-different NIR spectra from sugarcane varieties were exhibited advantage on the prediction model; this is because of the combination of NIR spectra from two sugarcane varieties were extended sugarcane parameter ranges ($^{\circ}$ Brix, %Pol, %Fiber and CCS). The wide parameter range would be a contribution to a wide range of predictive model from NIR measurement.

3.3. The Influence of Sugarcane Varieties on the Prediction Model

Table 2 shows the statistical data of the sugarcane KK and LK variety. If the reference value range of calibration is more extensive than the prediction set, it is considered plausible because the sample distribution in the calibration set is vital for the accuracy of the prediction model. Also, the combination of two sugarcane varieties increases the range of reference data, which can help improve the prediction model performance since the range of the sample distribution is more extensive compared to using just one sugarcane variety. The $^{\circ}$ Brix reference values of sugarcane variety KK are distributed to mean values of 21.34%, standard deviation (SD) of 2.01, and the range of $^{\circ}$ Brix value at ca. 12.77%, while the sugarcane variety LK shows the $^{\circ}$ Brix reference values of 21.52%, standard deviation (SD) of 1.55, and the range of $^{\circ}$ Brix range at 3.45, meaning the mean values of $^{\circ}$ Brix reference from the sugarcane KK variety is higher than the sugarcane LK variety. With regards to SD, the distribution of $^{\circ}$ Brix value from KK is higher than the distribution of $^{\circ}$ Brix from LK. In order to, it can be seen the combination of two sugarcane variety that a standard deviation (SD) is slightly increased to about 0.2% by using separate calibration for different sugarcane variety. It is possible to improve upon prediction models. When comparing the two sugarcane varieties, the total distribution of the $^{\circ}$ Brix reference decreased to 24.42%, and the range of the $^{\circ}$ Brix value remained at 12.77. However, the calibration of the two sugarcane varieties illustrates that the range of the $^{\circ}$ Brix slightly increased from the LK variety, this can be

confirmed the dependability of $^{\circ}$ Brix prediction model in the combination of two varieties.

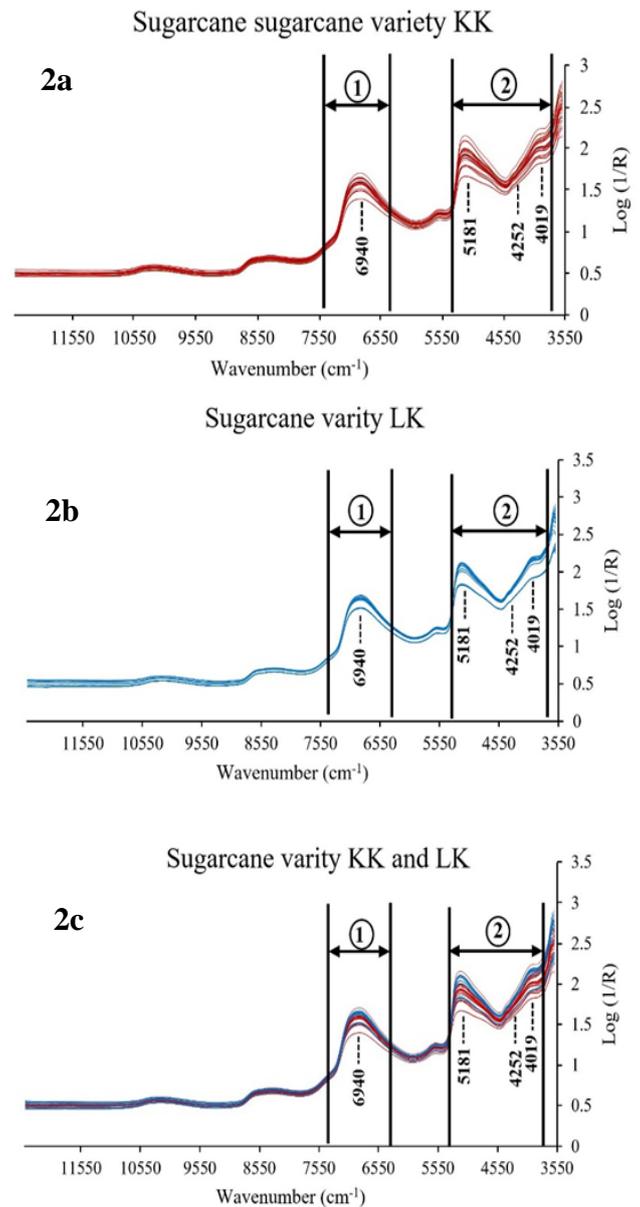


Fig. 2. NIR raw spectra of 80 samples of sugarcane variety KK as a red line (2a), LK(2b) as a blue line and comparison spectral of two sugarcane varieties (2c).

The %Pol reference values indicate the quality of sucrose in sugarcane. The mean of %Pol reference values of two sugarcane varieties are distributed around mean values of 17 to 18%, the range of %Pol is ca. 6.10 and 13.28, and the standard deviation is at 1.36%. For calibration, the mean of %Pol, the range of %Pol, and SD from two sugarcane varieties are 13.28%, 6.10%, and 1.54, respectively. %Fiber reference values show the percentage of fiber composition in sugarcane. %Fiber reference values of two sugarcane varieties obtained mean values of 13.61%, the range of %Fiber was 10.61 to 6.70, and standard deviation of 1.11%. The calibration of %Fiber showed a significant increase from 6.70 to 8.70 of LK.

CCS is the source of payment in the sugarcane industry. It is not a direct measure of sucrose content but estimates the total sugar (%) with °Brix, %Pol, and %Fiber in sugarcane. Besides, CCS of KK and LK varieties distribute reference value around mean values of 13.22, range of 11.88, and standard deviations of 1.64. For the calibration of two sugarcane varieties, the CCS value shows a narrow range; this may be because the CCS value is an estimated value from the source data such as °Brix, %Pol, and %Fiber.

The total distribution of °Brix, %Pol, %Fiber and CCS reference illustrates that the range of the values is slightly higher for the LK variety compared to the KK variety. The prediction model will be obtained more reliable if the range of data in reference value is more extensive than that of the prediction model. Thus, the prediction model of the LK variety has higher reliability than that of the KK variety. However, the combination of two varieties in the prediction model rise the range of reference value, leading to improve the dependability of the prediction model.

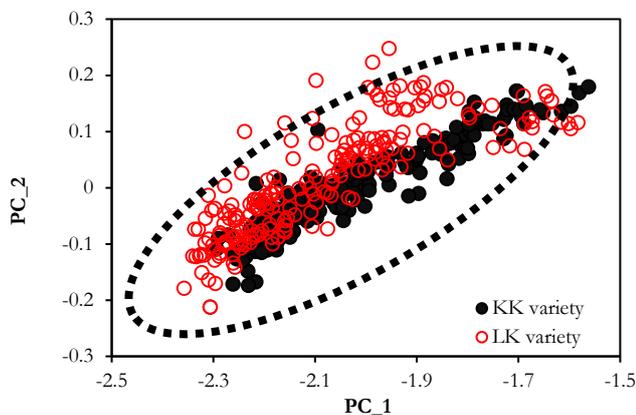


Fig. 3. PC1 versus PC2 score plot of 80 NIR analyses of sugarcane from two different varieties (Sugarcane variety KK (as a black dot) and LK (as a blue dot)).

3.4. Prediction Model of °Brix, %Pol, %Fiber and CCS

Table 3 illustrates the optimized results of the model for determination of °Brix, %Pol, %Fiber, and CCS. The table was included in the wavenumber range, spectral pretreatment methods, number of PLS factor or latent variable, coefficient of determination (R^2) of both calibration and validation, root mean square error of calibration (RMSEC) and prediction (RMSEP), and bias of effective PLSR models. The wavenumber ranges are effective for the determination of values measurement since the wavenumber ranges are meant to absorb the stretching or vibration of chemical bonding, which is measured by the NIR detector [26]. Overall, the wavenumber ranges of °Brix, %Pol and CCS were similar at 4150 to 4875 and 5415 to 6481 cm^{-1} , while the percentage of fiber in sugarcane (%Fiber) showed a wavenumber range of 4150 to 8038 cm^{-1} . The difference in the wavenumbers indicates the various measured values; this may be because the measurements of °Brix and %Pol were

determined by the sugar content in sugarcane, which is related to OH in the sucrose structure. The vibration band of OH in sucrose banding is associated with O-H stretching and OH the first overtone at a wavenumber of 4800 and 6274 cm^{-1} [27]. Nevertheless, the %Fiber presents a difference in the wavenumber range. This may be because %Fiber is calculated from dried shattered sugarcane, which is related to the vibration band of C-H of the second overtone, O-H and C-H stretching at wave range of 4213, 4367 and 6050 cm^{-1} , respectively [28]. The results of determination of cellulose by the NIR has been reported by He et al. (2013) [29]. They found that cellulose contains an important vibration band with spectral regions between 5685 and 4300 cm^{-1} , and is associated with C-H stretching and O-H stretching with a wavenumber of 4300 and 6025 cm^{-1} , respectively [30]. Overall, a pre-treatment method of the PLSR models for °Brix, %Pol, %Fiber, and CCS have several latent variables around 6-7. The pre-treatment method of °Brix, %Pol and CCS indicates a second derivative, while the %Fiber used the spectral pre-treatment of first derivative and vector normalization method. R^2 , RMSEP, RPD and bias of all sugar contents were in range of 0.90 - 0.97, 0.24-0.35, 3.41-6.40 and -7.67×10^{-6} - 0.01, respectively. R^2 means that the percentage of the measured value could be explained by variance of NIR spectra; however, the percentage residuum of R^2 was the unexplained variance. As a result, the best of PLSR is the prediction of °Brix. The optimized PLSR model for °Brix shows that R^2 of 0.97. It indicates that 97.3% of °Brix variance could be explained by adsorption variance, and 2.70% is the unexplained variance, and the root means a square error of prediction (RMSEP) of °Brix is 0.24. The lowest RMSEP selected the best model due to it shows the different values between reference and prediction model. Then, the optimal sub wavenumber ranges and spectrum pretreatment method were considered to improve the calibration. If the model suffers from not being complex enough (underfitting), calibration error approximates prediction error. The ratio of the standard error of prediction to a standard deviation of reference values (RPD) is used by Williams et al. 2004 [31], which is used to assess the goodness of the NIR calibration. RPD of °Brix is 6.40 while the RPD of %Pol, %Fiber and CCS are 3.46, 3.41 and 3.75, respectively. RPD of all sugar contents were above 3, which were indicated high accuracy prediction model. [32]. Bias constants value between prediction and value from the reference method is an indicator of the overall accuracy of the calibration model for the prediction model. Bias value was obtained at -0.011 by the PLSR model of °Brix value.

The best PLSR model for %Pol was optimized using the second derivative with a latent variable number of 7. The model showed that R^2 , RMSEC, RPD, and Bias was 0.93, 0.33, 3.46, and -0.001, respectively. Furthermore, Zornorza et al., (2008) [33] suggested that $R^2 > 0.90$ and $\text{RPD} > 3$ (or between 2.5 and 3) corresponds to good and excellent prediction accuracy. Thus, it can be supported

that this prediction model is high accuracy and reliable using for sugar contents measurement.

The optimization of the PLSR model for %Fiber used the first derivative and vector normalization with a seven latent variable number. R^2 , RMSEC, RPD and Bias was 0.91, 0.29, 3.41, and -7.67×10^{-6} , respectively. All values were similar to Hoang et al. 's' work (2017)[34], who studied the prediction of %Fiber and sugar in sugarcane.

Furthermore, the best PLSR model of CCS is obtained by using the second derivative with a latent number of 7. The model showed that R^2 , RMSEC, RPD and bias was 0.92, 0.35, 3.75, and -0.001, respectively. Also, PLSR mode of CCS is the lowest values, which can account for the fact that the CCS was calculated from °Brix, %Pol and %Fiber by the ratio of RMSEP and mean value of prediction set. However, the CCS value was at 4.27%. It has been reported by Mevik & Cederkvist (2004) that CCS prediction model was suitable for price evaluating value with PLSR mode of CCS higher than 3 % [35].

The comparison between the scatter plot of measured values and the predicted value of the calibration set and prediction set, regression coefficient plot, and X-loading plot is shown in Fig. 4a, b, c, and d. This also shows the target line with its slope equal to 1.00. For the regression coefficient and X-loading plot, if a high value at various wavenumbers exists, it suggests the vibration of a molecular bond at that wavenumber (in FS1-FS4 and TS1-TS4 [36-37]). Also, this has been the primary influence on the model prediction. The dependent variable positively or negatively changes with the composition of the sugarcane, and this relates to the change in vibration bonding in the sugarcane composition. These results are essential for developing spectrometers for this specific application [38-40].

RMSEP is the most efficient measure of uncertainty in the NIR predictions as it measures the average uncertainty that can be expected when predicting new samples (unknown samples). This measurement is valid if the new samples are like those used for the calibration; otherwise, the prediction error might be much higher. In the present study, the RMSEP of °Brix, %Pol, %Fiber, and CCS of two sugarcane varieties were 0.25, 0.35, 0.28, and 0.33, respectively. When compared to other works, the RMSEP in this study was in range (RMSEP<0.3), meaning the PLSR model can be applied for the prediction of essential parameters such as °Brix, %Pol, %Fiber, and CCS in future development models [20].

CCS is a payment parameter for sugarcane trading, which is calculated from °Brix, %Pol and %Fiber content of sugarcane. The precisely accurate CCS prediction

model is needed for commercial trading because of its users to decided price for customers.

Table 2. Statistical data of °Brix, %Pol, %Fiber and CCS in model development.

Parameter	variety	Data set		N	Max	Min	Mean	Range	SD
°Brix	KK	Total	200	25.22	12.45	21.34	12.77	2.01	
		Calibration set	100	25.22	12.45	21.97	12.77	1.97	
		Validation set	100	25.22	12.56	21.98	12.77	1.98	
	LK	Total	200	23.0	19.47	21.81	3.83	0.83	
		Calibration set	105	24.42	19.47	21.60	4.95	0.88	
		Validation set	100	24.42	19.47	21.61	4.95	0.89	
	KK + LK	Total	400	25.22	12.45	21.56	12.77	1.77	
		Calibration set	200	25.22	12.45	21.79	12.77	1.55	
		Validation set	200	24.42	9.53	14.92	14.88	3.27	
%Pol	KK	Total	200	22.18	8.90	17.76	13.28	1.91	
		Calibration set	100	22.18	8.9	18.26	13.28	1.54	
		Validation set	100	22.18	8.9	18.26	13.28	1.54	
	LK	Total	200	20.32	14.22	18.63	6.10	1.13	
		Calibration set	100	20.32	14.22	18.63	6.10	1.54	
		Validation set	100	20.32	14.22	18.63	6.10	1.54	
	KK + LK	Total	400	20.32	8.90	18.18	13.28	1.73	
		Calibration set	200	22.18	8.90	18.26	13.28	1.54	
		Validation set	200	22.18	8.90	18.26	13.28	1.54	
%Fiber	KK	Total	200	16.90	6.29	13.50	10.61	1.10	
		Calibration set	100	16.24	9.53	13.40	6.70	0.94	
		Validation set	100	16.24	9.53	13.40	6.70	0.94	
	LK	Total	200	16.25	8.18	13.71	8.07	1.13	
		Calibration set	100	16.24	9.53	13.40	6.70	0.95	
		Validation set	100	16.24	9.53	13.40	6.70	0.94	
	KK + LK	Total	400	16.64	6.29	13.61	10.61	1.11	
		Calibration set	200	16.24	9.53	13.40	6.70	0.94	
		Validation set	200	16.24	9.53	13.40	6.70	0.94	
CCS	KK	Total	200	17.58	5.69	12.80	11.88	1.79	
		Calibration set	100	16.58	8.6	13.48	7.98	1.23	
		Validation set	100	16.58	8.6	13.48	7.98	1.23	
	LK	Total	200	15.31	9.32	13.67	5.99	1.16	
		Calibration set	100	16.58	8.60	13.48	7.98	1.23	
		Validation set	100	16.58	8.60	13.48	7.98	1.23	
	KK + LK	Total	400	17.58	5.69	13.22	11.88	1.64	
		Calibration set	200	16.58	8.60	13.45	7.98	1.23	
		Validation set	200	16.58	8.60	13.48	7.98	1.23	

In this work, the model was compared the determination of CCS from NIR prediction and calculation of CCS from NIR prediction values (°Brix, %Pol, %Fiber). Table 4 shows RMSEP of the determination of CCS from NIR prediction (RMSEP_{pre}) and the calculation of CCS from NIR prediction values (RMSEP_{cal}). In the present study, RMSEP_{pre} were obtained at 0.35, 0.34, and 0.33 for sugarcane KK, LK and two varieties (KK and LK), respectively. In the case of the RMSEP_{cal}, there were 0.67, 0.72 and 0.61 for sugarcane KK, LK and two varieties, respectively. RMSEP_{cal} was higher RMSEP values than that of RMSEP_{pre}. This means model from the determination of CCS from NIR prediction was more precise than that of the model from the calculation of CCS from NIR prediction. This could be that RMSEP_{cal} was obtained by calculation of CCS formula while RMSEP_{pre} was used PLSR model for optimizing prediction value. This could be supported that the determination of CCS from NIR prediction can be used for CCS prediction.

3.5. Comparison of Results with Other Works

Figure 4 shows scattering plots of calibration and validation of sugar contents in the sugarcane prediction model. It can be seen similarly scattering between calibration and validation data from NIR and the reference

method (Laboratory data). Most of NIR technique work for sugar content prediction models have focused on °Brix, %Pol and %Fiber value. No such work has reported in CCS value prediction. In this work, %Fiber prediction model of sugarcane variety Lumpang 92-11 (LK 92-11) and Khon Kaen 3 (KK 3) shows the lowest R^2 (0.90), comparing to the other parameters (0.97, 0.91, and 0.92 for °Brix, %Pol and CCS, respectively). Comparison to Sanseechan P et al.'s work, (2018) [12], R^2 value was the same value at 0.90. In additional, R^2 of °Brix, %Pol and %Fiber from validation method were 0.99, 0.98 and 0.87, respectively. From the set of prediction and validation data, it can be confirmed that this prediction model was precisely accuracy and reliable for sugar content measurement and price evaluating.

Table 3. Results of partial least squares regression models (PLSR) for determination of °Brix, %Pol, %Fiber and CCS of two sugarcane varieties.

Parameters	Wavenumber range (cm ⁻¹)	Pretreatment	variety	Calibration			Validation			
				L V	R _c ²	RMS EC	R _p ²	RMS EP	RP D	Bias
°Brix	6481-5415-4875-4150	Second derivative	KK	6	0.97	0.29	0.97	0.30	5.8	0.047
			LK	6	0.96	0.25	0.97	0.23	3.52	0.031
			KK + LK	6	0.97	0.24	0.98	0.25	6.40	-0.011
%Pol	6418-5415-4875-4150	Second derivative	KK	7	0.93	0.43	0.93	0.45	3.17	-6.00 × 10 ⁻⁴
			LK	7	0.90	0.36	0.90	0.40	3.07	-0.004
			KK + LK	7	0.93	0.33	0.92	0.35	3.46	-0.001
%Fiber	8038-4150	First derivative + Vector normalization	KK	7	0.91	0.29	0.90	0.28	3.54	4.25 × 10 ⁻⁵
			LK	7	0.89	0.37	0.89	0.37	2.43	2.45 × 10 ⁻⁵
			KK + LK	7	0.91	0.29	0.91	0.28	3.41	-7.67 × 10 ⁻⁶
CCS	6803-5291-4751-3826	Second derivative	KK	7	0.93	0.34	0.93	0.35	3.77	-0.002
			LK	7	0.92	0.33	0.93	0.34	3.58	-0.002
			KK and LK	7	0.92	0.35	0.93	0.33	3.75	-0.001

Table 4. Comparison of the root means a square error of prediction (RMSEP) form the determination of CCS by NIR prediction and calculation of CCS from NIR prediction values (°Brix, %Pol, %Fiber).

CCS	RMSEP _{pre}	RMSEP _{cal}
KK	0.35	0.67
LK	0.34	0.72
KK + LK	0.33	0.61

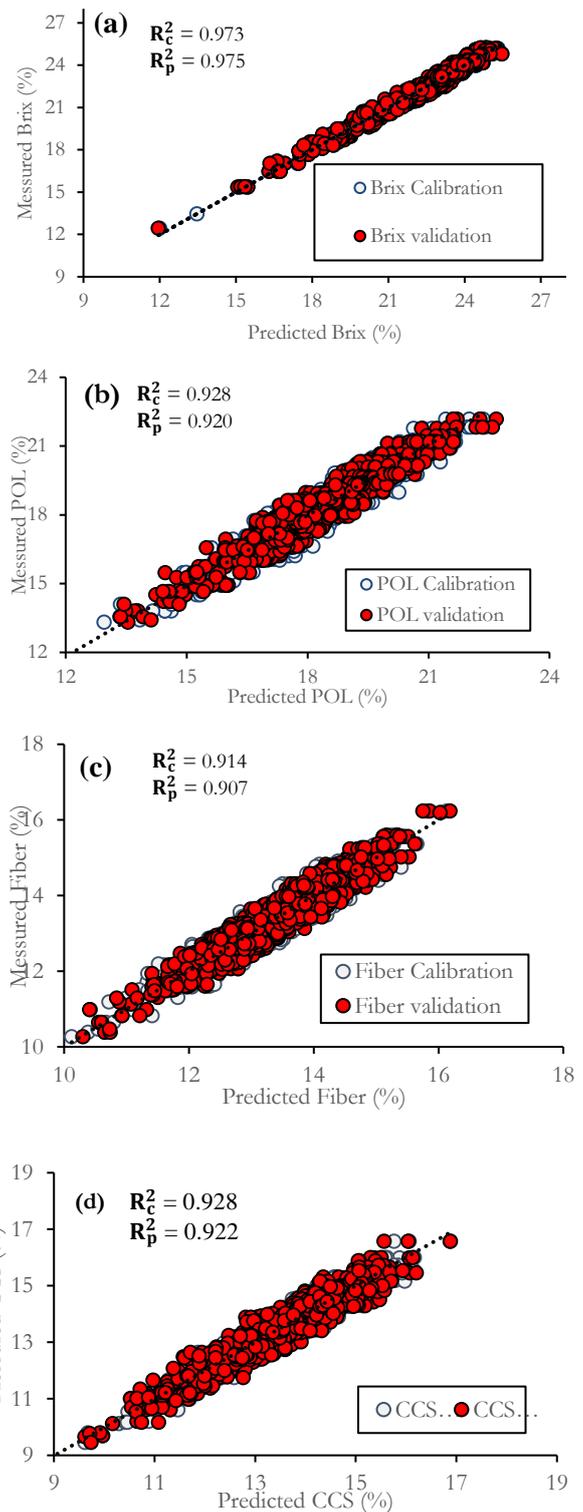


Fig. 4. Scatter plot of the measured value of °Brix (a), %Pol (b), %Fiber (c) and CCS (d) of two sugarcane varieties versus the predicted value of calibration and validation set.

4. Conclusion

In this study, the NIR technique was successfully applied to use for measuring sugar content in sugarcane. Sugarcane (*Saccharum spp.*) variety Lumpang 92-11 (LK 92-11) and Khon Kaen 3 (KK 3) were compared in term of °Brix, %Pol, %Fiber and CCS in both of Laboratory and NIR techniques primary and secondary methods. It

was no significant difference in both laboratory and NIR methods of two varieties. By using the set of data for the prediction model, the best PLSR model for CCS is obtained by using the second derivative with a latent number of 7. The model showed that R^2 , RMSEP, RPD and bias were 0.92, 0.35, 3.75, and -0.001, respectively. Furthermore, the RMSEP of °Brix, %Pol, %Fiber, and CCS of two sugarcane varieties were 0.25, 0.35, 0.28, and 0.33, respectively. When compared to other works, the RMSEP in this study was in range (RMSEP<0.3), meaning the PLSR model can be applied for the prediction of important parameters such as °Brix, %Pol, %Fiber, and CCS in future development models. Comparison of CCS determination form NIR prediction and CCS calculation form prediction values, RMSEPcal from the calculation of CCS from NIR prediction was higher than that of RMSEPpre from the determination of CCS from NIR prediction. This could be that RMSEPcal was obtained by calculation of CCS formula while RMSEPpre was used PLSR model for optimizing prediction value. This could be supported that the determination of CCS from NIR prediction can be used for CCS prediction.

Acknowledgement

This work was supported by Research and Graduate Studies Khon Kaen University, Thailand. We gratefully acknowledge Sustainable Infrastructure Research and Development Center (SIRDC), Farm Engineering and Automatic Control Technology Research Group (FEAT Group), Applied Engineering for Important Crops of the North East Research group (AENE Group), and the Graduate School at Khon Kaen University for the financial support.

References

- [1] *Thailand Sugarcane Production Report (2017/18)*, 2018.
- [2] P. B. Dixon and D. T. Johnson, "Pricing of Queensland sugar cane: Appraisal of the present formula and a suggestion for reform," *Review of Marketing and Agricultural Economics*, vol. 56, no. 1, pp. 27-35, 1988.
- [3] J. Engelke, "Sugarcane: Measuring commercial quality, Western Australia Department of Agriculture, 2002.
- [4] S. Escribano, et al., "Non-destructive prediction of soluble solids and dry matter content using NIR spectroscopy and its relationship with sensory quality in sweet cherries," *Postharvest Biology and Technology*, vol. 128, p. 112-120, 2017.
- [5] J. Véstia, et al., "Predicting calcium in grape must and base wine by FT-NIR spectroscopy," *Food Chemistry*, vol. 276, pp. 71-76, 2019.
- [6] P. S. Sampaio, et al., "Optimization of rice amylose determination by NIR-spectroscopy using PLS chemometrics algorithms," *Food Chemistry*, vol. 242, pp. 196-204, 2018.
- [7] G. Bázár, et al., "NIR detection of honey adulteration reveals differences in water spectral pattern," *Food Chemistry*, vol. 194, pp. 873-880, 2016.
- [8] S. Ghidini, et al., "Rapid authentication of European sea bass (*Dicentrarchus labrax* L.) according to production method, farming system, and geographical origin by near infrared spectroscopy coupled with chemometrics," *Food Chemistry*, vol. 280, pp. 321-327, 2019.
- [9] H. Cen and Y. He, "Theory and application of near infrared reflectance spectroscopy in determination of food quality," *Trends in Food Science & Technology*, vol. 18, no. 2, pp. 72-83, 2007.
- [10] R. Giangiaco, "Study of water-sugar interactions at increasing sugar concentration by NIR spectroscopy," *Food Chemistry*, vol. 96, no. 3, pp. 371-379, 2006.
- [11] C. Collell, et al., "NIR technology for on-line determination of superficial aw and moisture content during the drying process of fermented sausages," *Food Chemistry*, vol. 135, no. 3, pp. 1750-1755, 2012.
- [12] P. Sansechan, et al., "A portable near infrared spectrometer as a non-destructive tool for rapid screening of solid density stalk in a sugarcane breeding program," *Sensing and Bio-Sensing Research*, vol. 20, pp. 34-40, 2018.
- [13] I. Ramírez-Morales, et al., "Optimization of NIR calibration models for multiple processes in the sugar industry," *Chemometrics and Intelligent Laboratory Systems*, vol. 159, pp. 45-57, 2016.
- [14] A. J. S. Neto, et al., "Prediction of mineral contents in sugarcane cultivated under saline conditions based on stalk scanning by Vis/NIR spectral reflectance," *Biosystems Engineering*, vol. 156, pp. 17-26, 2017.
- [15] L. E. Agelet, and C. R. Hurburgh, "A tutorial on near infrared spectroscopy and its calibration," *Critical Reviews in Analytical Chemistry*, vol. 40, no. 4, pp. 246-260, 2010.
- [16] F. Linow, *Near-Infrared Technology in the Agriculture and Food Industries*. St. Paul, Minnesota, USA: American Association of Cereal Chemists, Inc., 1987.
- [17] T. Giordanengo, et al., "Correction of moisture effects on near infrared calibration for the analysis of phenol content in eucalyptus wood extracts," *Annals of Forest Science*, vol. 65, no. 8, p. 1, 2008.
- [18] M. Kim, and D. F. Day, "Composition of sugar cane, energy cane, and sweet sorghum suitable for ethanol production at Louisiana sugar mills," *J Ind Microbiol Biotechnol*, vol. 38, no. 7, pp. 803-7, 2011.
- [19] C. A. dos Santos, et al., "A review on the applications of portable near-infrared spectrometers in the agro-food industry," *Appl Spectrosc*, vol. 67, no. 11, pp. 1215-33, 2013.
- [20] M. Naderi-Boldaji, et al., "Dielectric power spectroscopy as a potential technique for the non-destructive measurement of sugar concentration in sugarcane," *Biosystems Engineering*, vol. 140, pp. 1-10, 2015.

- [21] S. Wold, et al., "Some recent developments in PLS modeling," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 2, pp. 131-150, 2001.
- [22] H. Mark, and J. Workman Jr., *Chemometrics in Spectroscopy*. Elsevier, 2010.
- [23] D. A. Burns and E. W. Ciurczak, *Handbook of Near-Infrared Analysis*. CRC Press, 2007.
- [24] J. -W. Hao, et al., "Rapid quantification of polysaccharide and the main monosaccharides in *Dendrobium huoshanense* by near-infrared attenuated total reflectance spectroscopy," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 151, pp. 331-338, 2018.
- [25] N. M. Nawi, G. Chen, and T. Jensen, "In-field measurement and sampling technologies for monitoring quality in the sugarcane industry: A review," *Precision Agriculture*, vol. 15, no. 6, pp. 684-703, 2014.
- [26] J. A. Wahr, et al., "Near-Infrared spectroscopy: Theory and applications," *Journal of Cardiothoracic and Vascular Anesthesia*, vol. 10, no. 3, pp. 406-418, 1996.
- [27] S. R. Djafari Petroudy, E. Rasooly Garmaroody, and H. Rudi, "Oriented Cellulose Nanopaper (OCNP) based on bagasse cellulose nanofibrils," *Carbohydrate Polymers*, vol. 157, pp. 1883-1891, 2017.
- [28] A. T. Badaró, et al., "Identification of fiber added to semolina by near infrared (NIR) spectral techniques," *Food Chemistry*, vol. 289, pp. 195-203, 2019.
- [29] W. He and H. Hu, "Prediction of hot-water-soluble extractive, pentosan and cellulose content of various wood species using FT-NIR spectroscopy," *Bioresource Technology*, vol. 140, pp. 299-305, 2013.
- [30] J. Xue, et al., "On-line measurement of proximates and lignocellulose components of corn stover using NIRS," *Applied Energy*, vol. 137, pp. 18-25, 2015.
- [31] P. Williams, et al., *Near-Infrared Technology: Getting the Best Out of Light*. Value Added Wheat CRC, Limited, 2004.
- [32] B. M. Nicolai, et al., "Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review" *Postharvest Biology and Technology*, vol. 46, no. 2, pp. 99-118, 2007.
- [33] R. Zornoza, et al., "Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils," *Soil Biology & Biochemistry*, vol. 40, no. 7, pp. 1923-1930, 2008.
- [34] N. V. Hoang, et al., "High-throughput profiling of the fiber and sugar composition of sugarcane biomass," *BioEnergy Research*, vol. 10, no. 2, pp. 400-416, 2017.
- [35] B.-H. Mevik and H. R. Cederkvist, "Mean squared error of prediction (MSEP) estimates for principal component regression (PCR) and partial least squares regression (PLSR)," *Journal of Chemometrics*, vol. 18, no. 9, pp. 422-429, 2004.
- [36] J. Workman, Jr. and L. Weyer, *Practical Guide to Interpretive Near-Infrared Spectroscopy*. CRC Press, 2008.
- [37] B. G. Osborne, "Near-infrared spectroscopy in food analysis," in *Encyclopedia of Analytical Chemistry*. 2006.
- [38] G. A. de Oliveira, et al., "Comparison of NIRS approach for prediction of internal quality traits in three fruit species," *Food Chemistry*, vol. 143, pp. 223-230, 2014.
- [39] S. Sahamishirazi, et al., "Development of a near-infrared spectroscopy method (NIRS) for fast analysis of total, indolic, aliphatic and individual glucosinolates in new bred open pollinating genotypes of broccoli (*Brassica oleracea* convar. botrytis var. italica)," *Food Chemistry*, vol. 232, pp. 272-277, 2017.
- [40] M. Ye, et al., "Rapid detection of volatile compounds in apple wines using FT-NIR spectroscopy," *Food Chemistry*, vol. 190, pp. 701-708, 2016.





Khanita Kamwilaisak was born on 24th of May 1971 at Bangkok. She graduated B.Eng in chemical Engineering at Khon Kaen University, Thailand in 1993, M.Eng in Chemical Engineering at KMUTT, Thailand in 1999 and PhD. in Chemical and Process Engineering at The University of Sheffield, United Kingdom.

She was a lecturer in chemical engineering at Mahanakorn University of Technology, Thailand from June 1995 to May 2005. Now She is working at chemical engineering at Khon Kaen University, Thailand since April 2009. She is interested in photocatalytic reaction, sustainable product, circular economy and waste management.

Pasakorn Jutakradsada, photograph and biography not available at the time of publication.

Wimonporn Iamamornphanth, photograph and biography not available at the time of publication.

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

Pornnapa Kasemsiri, photograph and biography not available at the time of publication.

Supasit Konyai, photograph and biography not available at the time of publication.

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



Prof. Prinya Chindaprasirt received B.Eng. from University of Tasmania Australia in 1974, M.Eng. and Ph.D. in Civil Engineering from The University of New South Wales Australia in 1977 and 1980, respectively. Professor Prinya Chindaprasirt is the Director of the Sustainable Infrastructure Research and Development Center (SIRDC), a research unit in Khon Kaen University. He is now also the head of the Advanced Functional Materials research cluster of Khon Kaen University. He was appointed a full professor in 2007. In 2009 he was the first person in civil engineering who was appointed the highest rank professor in Thailand. He has set up the Thai Geopolymer Network in 2005 to promote the research and collaboration among Thai researchers

in this field.

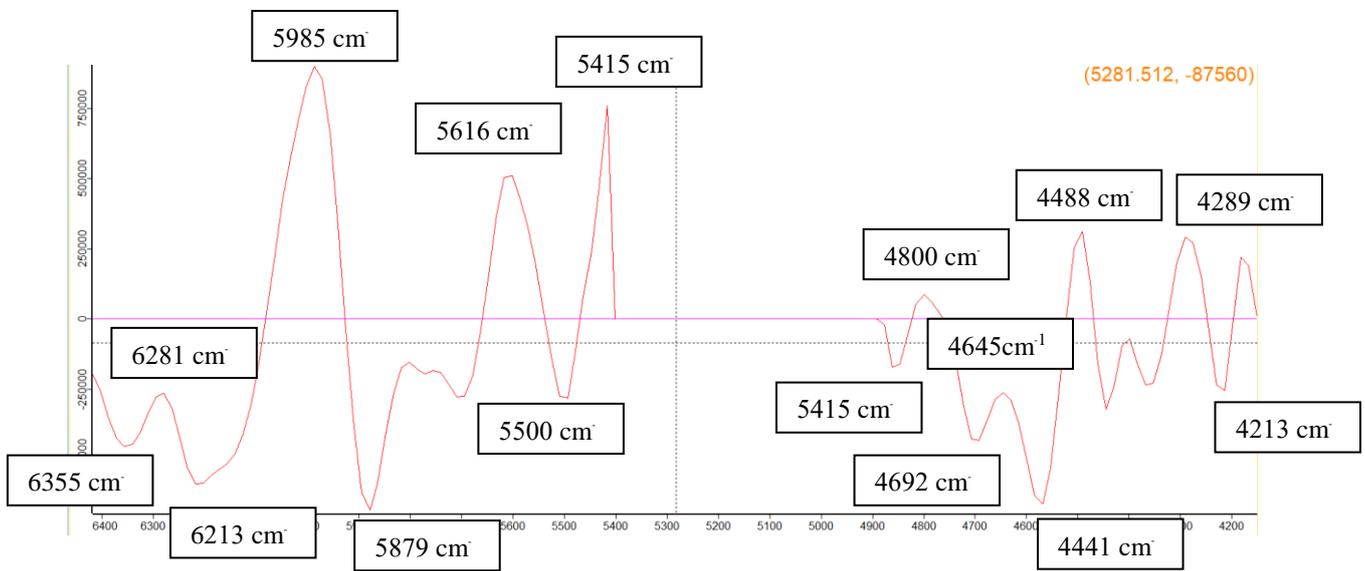
Appendix: Supplementary Data

The vibration of a molecular bond at that wavenumber of °Brix, POL, Fiber contents and CCS value.

TS1. The summarized of vibration of a molecular bond in wavelength of Brix prediction

Wavenumber (cm ⁻¹)	Wavelength (nm)	Wavelength from reference (nm)	Bond vibration	Structure	Reference
4213	2373	2380	O-H def second overtone	ROH	[36]
4289	2326	2323	C-H stretching +C-H def	CH ₂	[36]
4441	2251	2252	O-H stretching O-H def	Strach	[36]
4645	2152	2150	2*amide I+ amide II	CONHR	[36]
4692	2131	2132	N-H stretching, C=O stretching	Amino acid	[36]
4800	2083	2080	O-H stretching O-H def	R-OH Sucrose	[36]
5415	1846	1820	O-H stretching+2xC-O stretching	Cellulose	[36]
5523	1810	1820	O-H/C-H combination	Cellulose	[36]
5616	1780	1780	C-H stretching first overtone	Cellulose	[36]
5710	1751	1765	C-H stretching first overtone	CH ₂	[36]
5817	1719	1725	C-H stretching first overtone	CH	[36]
5879	1700	1701	C-H methyl C-H, OH associated	ROHCH ₃	[36]
5985	1670	1671	C-H aromatic	CH	[36]
6000	1667	1664	C-H aromatic	ROHCH ₃	[36]
6213	1609	1620	C-H stretching first overtone	=CH ₂	[37]
6355	1573	1580	O-H stretching first overtone	Glucose	[37]

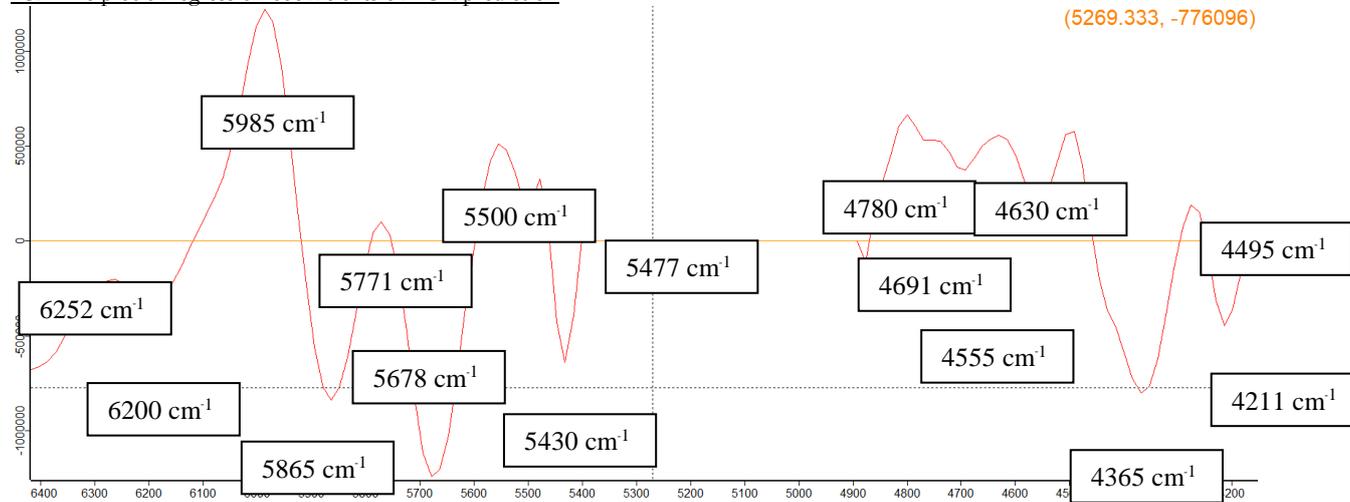
FS1. The plot of regression coefficients of °Brix prediction



TS2. The summarized of vibration of a molecular bond in wavelength of POL prediction

Wavenumber (cm ⁻¹)	Wavelength (nm)	Wavelength from reference (nm)	Bond vibration	Structure	Reference
4211	2374	2380	O-H second overtone	ROH	[36]
4365	2290	2294	N-H stretching + C=O stretching	Amino acid	[36]
4495	2224	2200	C-H stretching + C=O stretching	-CHO	[36]
4555	2195	2190	CH ₂ asymmetric + C=O stretching	HC=CH	[36]
4630	2159	2160	2xAmide I + Amide III	CONHR	[36]
4691	2131	2132	N-H stretching, C=O stretching	Amino acid	[36]
4780	2092	2080	O-H stretching O-H	R-OH Sucrose	[36]
5430	1841	1860	C-Cl chlorinated organics (C-Cl group)		[36]
5477	1825	1820	O-H stretching + 2xC-O stretching	Cellulose	[36]
5500	1818	1820	O-H stretching + 2xC-O stretching	Cellulose	[36]
5678	1761	1762	C-H methylene (CH ₂) symmetric		[36]
5710	1751	1765	C-H stretching first overtone	CH ₂	[37]
5771	1732	1725	C-H stretching first overtone	CH ₂	[36]
5865	1705	1705	C-H stretching first overtone	CH ₃	[36]
5985	1670	1671	C-H aromatic	CH	[36]
6200	1612	1620	C-H stretching first overtone	=CH ₂	[36]
6252	1599	1580	O-H stretching first overtone (Intermol H-bond)	Glucose, starch	[36]

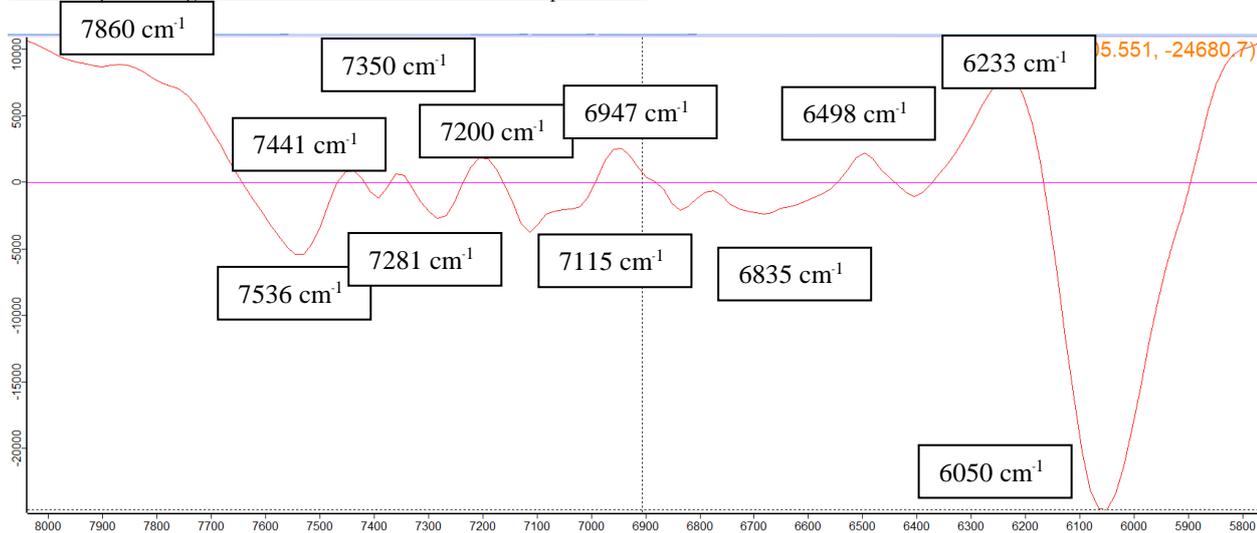
FS2. The plot of regression coefficients of POL prediction



TS3. The summarized of vibration of a molecular bond in wave length of Fiber content prediction

Wavenumber (cm)	Wavelength (nm)	Wavelength from reference (nm)	Bond vibration	Structure	Reference
6050	1652	1660	C-H stretching first overtone	<i>Cis</i> -RCH=CHR ¹	[36]
6233	1604	1598	C=O/N-H combination from polyamide II	C=O/N-H	[36]
6498	1538	1540	O-H stretching first overtone	Starch	[36]
6835	1463	1460	N-H stretching first overtone	CONHR	[36]
6947	1439	1440	O-H stretching first overtone	Sucrose	[36]
7115	1405	1410	O-H stretching first overtone	ROH	[36]
7200	1388	1395	2x C-H stretching + C-H def	CH ₂	[36]
7281	1373	1370	C-H methyl C-H, associated with linear aliphatic	Hydrocarbon	[36]
7350	1360	1360	C-H methyl C-H, associated with aromatic (ArCH ₃)	Hydrocarbon	[36]

FS3. The plot of regression coefficients of Fiber contents prediction



TS4. The summarized of vibration of a molecular bond in wave length of CCS value prediction

Wavenumber (cm)	Wavelength (nm)	Wavelength from reference (nm)	Bond vibration	Structure	Reference
4213	2373	2352	C-H def Second overtone	Cellulose	[36]
4367	2289	2270	O-H stretching + C-H def	Cellulose	[37]
4494	2225	2200	C-H stretching + C=O stretching	HC=CH	[36]
4552	2196	2190	CH ₂ asymmetric + C=O stretching	HC=CH	[36]
4658	2146	2140	=C-H stretching + C=C stretching	HC=CH	[36]
4770	2096	2100	2x O-H def + 2x C-O stretching	Strach	[36]
4800	2083	2080	O-H stretching O-H def	R-OH Sucrose	[36]
5786	1728	1725	C-H stretching first overtone	CH ₂	[36]
5878	1701	1705	C-H stretching first overtone	CH ₃	[36]
5987	1670	1660	C-H stretching first overtone	<i>Ci</i> -RCH=CHR ¹	[36]
6173	1619	1620	C-H stretching first	=CH ₂	[36]
6274	1593	1580	O-H stretching first over tone (Intramolecular H-bond)	Glucose	[36]
6515	1534	1533	C-H stretching first over tone	CH	[36]
6666	1500	1490	O-H stretching first over tone (Intramolecular H-bond)	Cellulose	[36]
6725	1486	1480	O-H stretching first over tone (Intramolecular H-bond)	Glucose	[36]

FS4. The plot of regression coefficients of CCS value prediction

