EASR

Engineering and Applied Science Research

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

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

Evaluation of informative spectral wavelengths for estimating soluble solids content in sugarcane billets

Vasu Udompetaikul[1)], Kittisak Phetpan*[2)] and Panmanas Sirisomboon[1)]

[1)]Department of Agricultural Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand
[2)]Department of Engineering, King Mongkut's Institute of Technology Ladkrabang, Prince of Chumphon Campus, Chumphon 86160, Thailand

Received 28 December 2021
Revised 13 March 2022
Accepted 26 April 2022

Abstract

This study proposed individual spectral wavelengths significant to estimate soluble solids content (SSC) in sugarcane billets moving on the conveyor. At the same time, an all-in-one quality and yield monitor using those wavelengths for a sugarcane harvester was also proposed. Seven wavelengths, 475, 560, 668, 717, 755, 840, and 890 nm, were arranged into three groups for modeling. Group 1, consisting of 475, 560, 668, 717, and 840 nm, was based on the spectral responses of a commercial multispectral camera, while group 2 (717 and 840 nm) was based on the invisible (RedEdge and near-infrared or NIR) responses of the camera. For group 3, two sugar-related wavelengths at 755 and 890 nm were selected as the candidates for modeling. Partial least squares regression (PLSR) was employed to model those three groups with corresponding soluble solids content (SSC). The results showed that the developed models based on two sugar-related wavelengths at 755 and 890 nm provided the best performance, explaining 80.2 % of the variance in the SSC and displaying a root mean square error of calibration (RMSEC) of 0.32 ºBrix. The predictive performance had the root mean square error of prediction (RMSEP) of 0.33 ºBrix. This finding confirmed the effectiveness of the sugar wavelengths and conveyed the possibility to develop the sugarcane quality and yield monitor.

Keywords: Sugarcane, Yield, Quality, Monitoring system, Near-infrared, Soluble solids

1. Introduction

Modern agricultural production focusing on improving yield and quality aspects, particularly in sugarcane production, is on the rise without enlarging the areas used for growing crops. A key factor is a management through accessing the information on production elements and the variation of such aspects across a given field. Based on the accessed information, it leads to proper agricultural practices. Indeed, improving yield and quality with excellent practice helps farmers increase efficiency and reduce fertilizer costs and natural resources-water, land, and energy. This management form is well known as "Precision agriculture (PA) technology," which has already been applied in the sugarcane industry. Accessing the site-specific yield outcome of sugarcane across the field and constructing variability maps is not an arduous process but still presents challenges for getting the site-specific quality [1]. In the future, figuring out yield and quality variations across the field will lead to proper farm management and a fair payment system to the growers [2]. Besides, those variations can be utilized for dealing with cooperative farming on a large scale. Hence, an on-the-go quality monitoring system for sensing the variation across the sugarcane field is needed to deal with the PA technology to improve yield and quality in sugarcane production. However, there are still no commercially available sensors.

Among several alternative techniques such as electronic refractometer [3], microwave [4-6], and near-infrared (NIR) spectroscopy [3, 7] conducted for the sugarcane quality measurement in the field, the NIR seems to have more potential than others. Using this technique based on visible and shortwave spectral region (400-1000 nm) to detect the non-contact sample-based spectra can evaluate sugar content (soluble solids content, SSC) in cane billets [3]. Nawi et al. [3] presented an impressive result with a good coefficient of determination (R^2) and root means square error of prediction (RMSEP) of 0.91 and 0.721 °Brix, respectively. This result prompted Nawi et al. [8] to present that the NIR sensor could be installed on the conveyor and used for real-time measurement. Their result led to Phetpan et al.'s [9] research, whose work was to develop a laboratory-scale prototype for the sugarcane's real-time sweetness measurement on the conveyor. Based on their prototype, the typical speed of the conveyor was set up for the experiment. This work exhibited the R^2 of the calibration model of 0.85 and the error term (RMSEP) of 0.30 ºBrix. Interestingly, two effective wavelengths at 755 and 890 nm exhibited high influence to be used for the sugar content prediction of the sugarcane. Indeed, this could be a message to the sugarcane industry in promoting the hope to access the sugarcane quality across the field.

Based on the possibility, the development of an all-in-one system for monitoring the quality and yield of the sugarcane at the same time during harvesting seems promising. One of the exciting concepts proposed and researched by Price et al. [10] is the optic fiber-based system designed for a sugarcane chopper harvester to obtain the yield during harvesting. This concept employed a duty-cycle

*Corresponding author. Tel.: +669 9324 2127
Email address: kittisak.ph@kmitl.ac.th
doi: 10.14456/easr.2022.56

manner accompanied by three optic sensors mounted at the conveyor floor to determine the depth of sugarcane billets profile on the slats. Based on the depth information, the weight could be estimated using a calibration model employed the linear relationship between the optical responses and the actual yields. The field test result showed the R^2 and average prediction error of 0.98 and 7.5%, respectively. With this concept, two exciting and possible ways which can lead to the development of the all-in-one system are as follows. The first is to use an optic probe providing the sugar-related wavelengths to sense the sugarcane's sweetness and simultaneously figure out the profile of the sugarcane billets being moved on the conveyor. The second is to employ a camera providing such wavelengths to evaluate the sweetness and estimate the profile of the billets on the conveyor. Indeed, such two proposed ways need to be researched for exploring the significantly informative wavelengths before further steps.

Therefore, this study explored the informative spectral wavelengths significant to the sugar content prediction in sugarcane billets moving on the conveyor. Seven wavelengths, either available in the commercial sensors or related to the vibration bands of sugar molecules, including 475, 560, 668, 717, 755, 840, and 890 nm, were evaluated their significance via mathematical model analysis. Close to those, some wavelengths exhibited a strong influence on the sugar content prediction of agriculture crops, e.g., at 676 nm for melon [11], at 677, 765, 665, 727, 758, 750, 721, and 879 nm for pears [12], at 674, 716, 748, 850 and 894 nm for citrus fruit [13], at 890 nm for apples, peaches and pears [14], and at 889 nm for sugarcane [15].

2. Materials and methods

2.1 Sample and dataset preparations

Fifty clumps of "Khonkaen 3" sugarcane variety at different maturity stages of 11 and 12 months after planting were collected randomly from fields in Suphan Buri, Thailand, in February 2017. They were brought to King Mongkut's Institute of Technology Ladkrabang, Bangkok, around 150 km away from the fields. This caused the samples took time around 24 h at ambient temperature (30 ± 5 °C) from harvesting until scanning. For the preparation of the samples, the leaves and sheath were stripped off without scraping off wax covering the sample surface. Each clump, twenty sugarcane stalks were chopped into billets with an approximate length of 20 centimeters (the same size as that chopped by the cane harvester), and it was defined as a sample group in this paper. So, there were 50 sample groups in total.

Note that the dataset used for the analysis throughout this work was the same as our previous published work [9]. The reflectance spectral data were collected using the online system [9] (Figure 1a) having a vis/NIR spectrometer (AvaSpec-2048-USB2, Avantes BV, Netherlands) as a spectral collector. The spectrometer works across a wavelength range of 350-1100 nm with a spectral resolution of 2.4 nm. Based on the environment of the spectral measurement chamber, the integration time of the spectrometer was set to 14 ms, giving approximately 90% full-scale Analog-to-Digital Converter (ADC) of the reference material reflectance. With a conveying speed of 2 m/s (typical speed of the elevator of the sugarcane harvester), the spectrometer could collect 19 spectral scans covering the area between 2 slats (a distance of 52 cm). The spectral collection of each sample group was done at an ambient temperature around 30°C for two replications with two repetitions, leading to 4 spectral sets (19 spectra each). After each replication, the scanned cane billets were collected for °Brix (Soluble solids content, SSC) measurement. Therefore, there were 200 sets of spectra (4 sets × 50 sample groups) and 100 corresponding SSC values (2 values from 2 replications × 50 sample groups) in total. There were typically 3 types of the spectral pattern (Figure 1b): sugarcane spectra, slat spectra, and elevator floor spectra, except some spectral sets that provided only floor spectra. Hence, one hundred and fifty sets of spectra remained after employing the spectral filtration process proposed in the previous work. They were then preprocessed to deal with noises and the scattering effect in the spectra. Based on the preprocessing steps, the spectral datasets (150 sets) across the wavelengths of 450-900 nm based on the moving average (MA) smoothing and the standard normal variate (SNV) were obtained. Each spectral set was averaged to acquire a spectral representative. Overview of spectral data acquisition, preprocessing steps, and modeling process is shown in Figure 2. Finally, one hundred and fifty samples matching the averaged spectra and their corresponding SSC were obtained for model analysis.

As the dataset provided (150 objects consisted of 150 spectra and their corresponding SSC), independent variables (450-900 nm based spectral wavelengths) were chosen and arranged into three groups for modeling in this study. Seven individual wavelengths were selected, including 475, 560, 668, 717, 755, 840, and 890 nm (Figure 3). Five wavelengths consisting of 475, 560, 668, 717, and 840 nm were selected based on the consideration of a commercial multispectral camera which corresponds to blue, green, red, red-edge, and near-infrared (NIR) bands, respectively. These five wavelengths cover the visible and invisible ranges. The variables for group 1 were composed of all those five wavelengths, whereas group 2 was composed of only 717 nm (defined as RedEdge) and 840 nm (defined as NIR_{840}). The wavelengths selected for the latter group were to represent the invisible spectral bands in the camera. For group 3, two wavelengths at 755 and 890 nm were selected as they had a strong influence on the prediction of sugar content in sugarcane billets, according to the reports of Phetpan et al. [9] and Udompetaikul et al. [16]. These two wavelengths were denoted as "NIR_{755}" and "NIR_{890}".

2.2 Mathematical modeling and the model validation

A random split of the dataset into two sets was done to create and optimize the models: two-thirds of all the objects into a training set (100 objects) and the remaining objects (50 objects) into a validation set. Three groups (Group 1, 2, and 3) of the independent variables from the training set were used to model with their corresponding SSC values. The models were then optimized to obtain good prediction performance using the validation set. Partial least squares regression (PLSR) approaches were employed to model this work using Statistics and Machine Learning Toolbox in MATLAB R2020b. The "plsregress" function (Partial least-squares regression) was applied for the PLSR analysis.

The PLSR modeling approach is based on the linear regression method that relates independent variables (X-variables) and a response (y-variable). The model analysis simultaneously decomposes X-variables and a y-variable as the X-weight by maximizing covariance between these two variables and then estimates new latent variables (LVs) as linear combinations of the X with the weight [17]. These LVs are subsequently used for the least square regression with a variable y [18]. The PLSR is a robust linear regression model insensitive to collinear variables and accepted for multi variables [19]. In this work, the PLSR was performed by setting a modeling parameter "cv" as a default ("resubstitution").

The statistical terms consisted of the coefficient of determination (R^2 for the calibration and r^2 for the validation) and root mean square error of calibration (RMSEC) and prediction (RMSEP) were employed to consider the performance of the developed models.

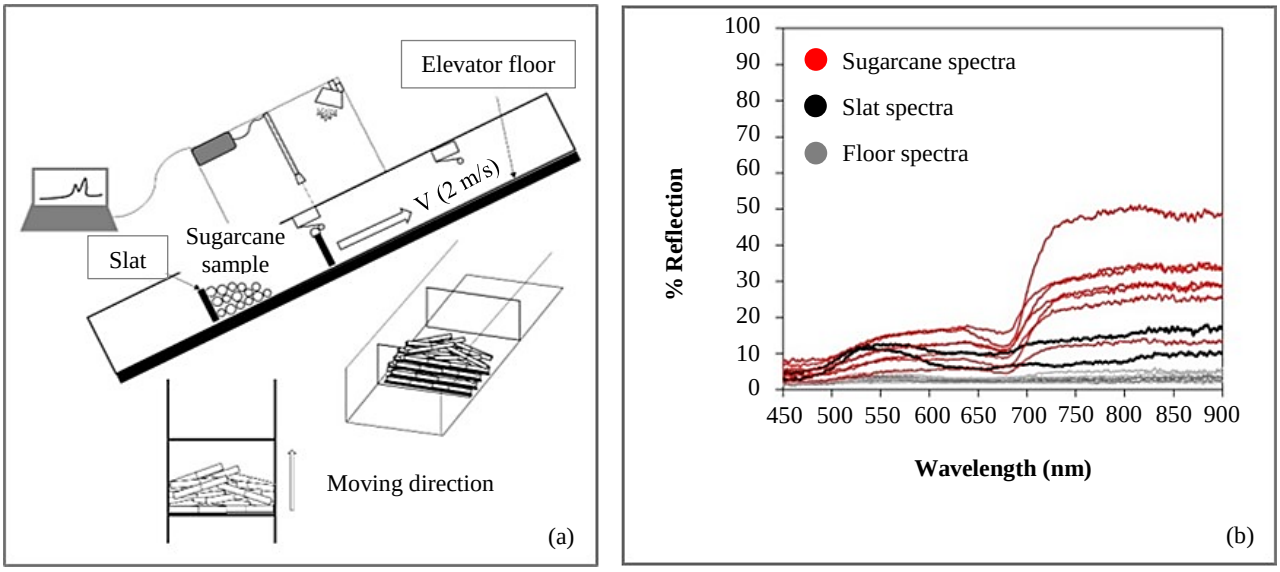


Figure 1 Scheme of an online measurement system and presentation of spectral patterns

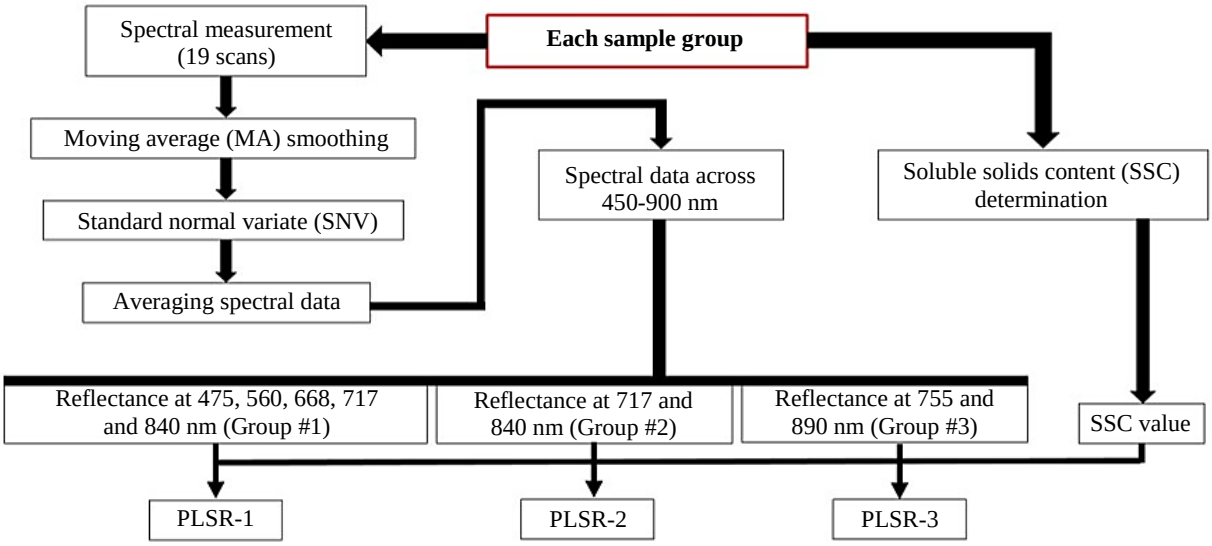


Figure 2 Overview of spectral data acquisition, preprocessing steps, and modeling process

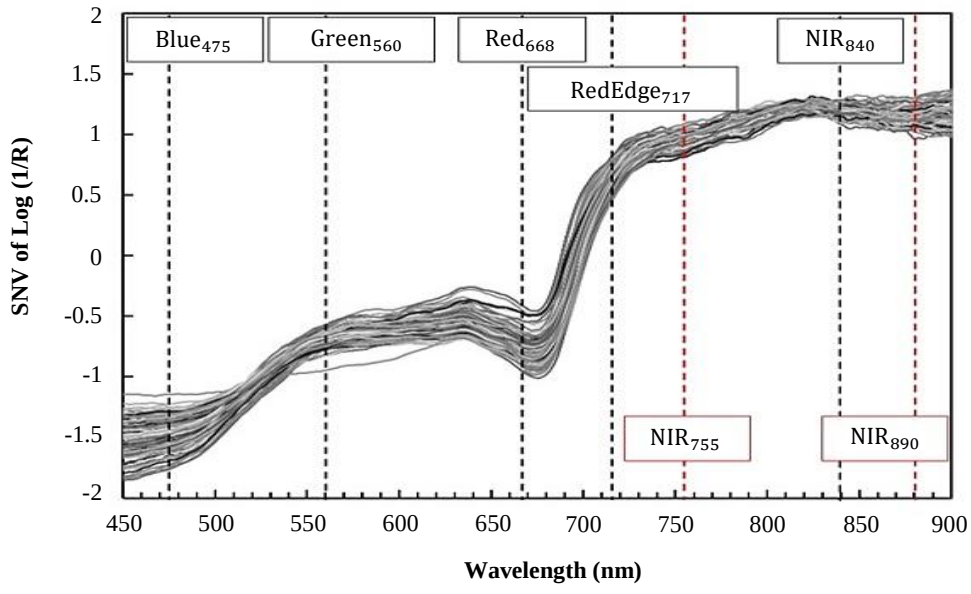
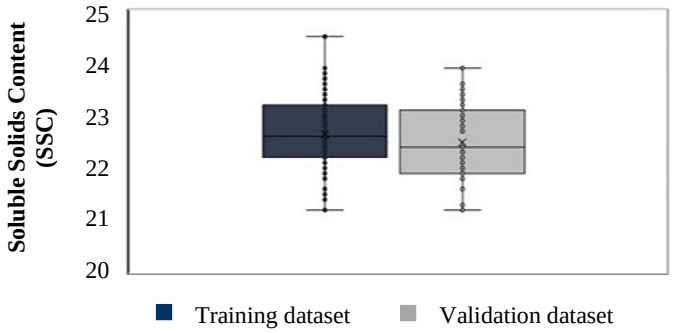


Figure 3 Presentation of the seven wavelengths selected for modeling

3. Results

3.1 Mathematical models and prediction performance

Figure 4 shows the distribution of the SSC values accompanied by their summaries of statistical characteristics for modeling and validating the models. Although the distribution comes out within the narrow range due to one cane cultivar sampling in this study, this is adequate for the study to complete the intention of this work. As selected and arranged the independent variables into three groups, each group in the training dataset was employed to model against the corresponding SSC values. The developed models were evaluated using the validation dataset. Figure 5 shows the modeling and validating results in which models 1, 2, and 3 were denoted for the modeling based on groups 1, 2, and 3, respectively. Compared to each others, it was clear that model 3 was the best one having the highest R squared (R^2) and lowest RMSEC values at 0.80 and 0.32 ºBrix, respectively. Also, it displayed the best SSC prediction performance having the highest r squared (r^2) and lowest RMSEP values at 0.79 and 0.33 ºBrix, respectively.



Dataset	N	Maximum	Minimum	Mean	SD
Training set	100	24.5	21.2	22.6	0.76
Validation set	50	23.9	21.4	22.6	0.66

Figure 4 Box plot and summaries of statistical characteristics of the soluble solids content

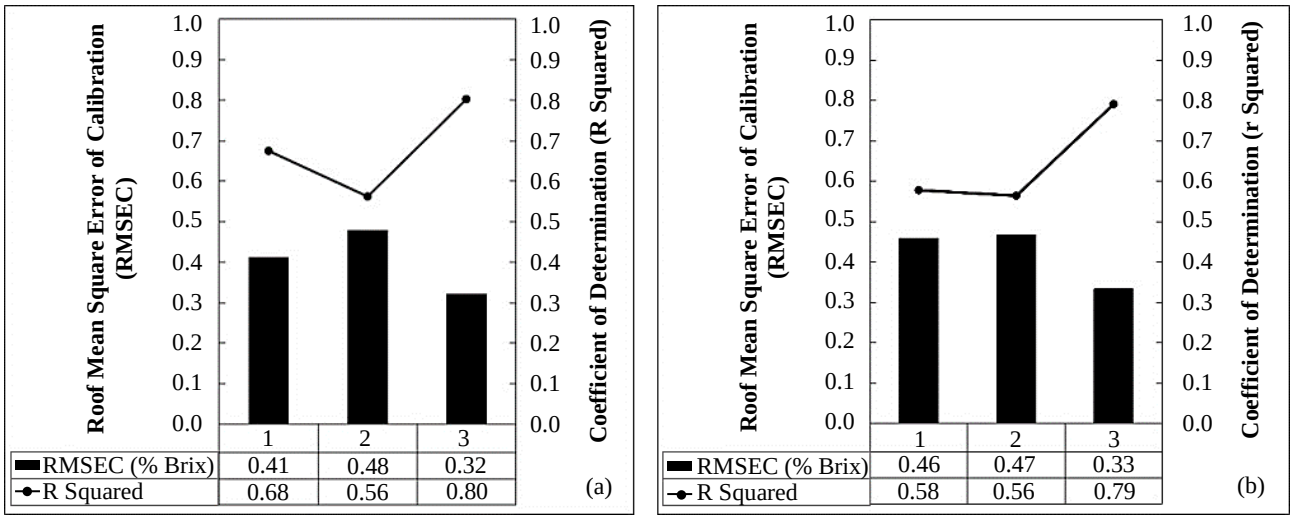


Figure 5 Results of the developed models in predicting the soluble solids content based on different wavelengths; (a) for modeling and (b) for validating the models

3.2 Consideration of the developed models

Based on the PLSR approach, the summaries are shown in Figure 6. Figure 6 (a) shows the plots of the percent variance of each latent variable (LV) explained in X and y variables. The PLSR-1, modeled based on all five wavelengths appearing in the multispectral camera, employed 5 LVs to explain 100 percent variance in X. Based on 5 LVs, it could account for only 67.5 percent variance y (SSC values). For the PLSR-2, the model was constructed using the bands of RedEdge and NIR_{840} as the independent variables. It used 2 LVs to extract 100 percent variance in the variable X, but only 56.2 percent variance in the SSC values was accounted for. In modeling based on two sugar-related wavelengths (PLSR-3), one hundred percent variance in X extracted through 2 LVs could be used to account for 80.2 percent variance in the SSC values. The optimizations of these 3 PLSR models based on the validation dataset are displayed through the scatter plots shown in Figure 6 (b-d). Table 1 shows the summaries of the regression and validation results of the PLSR models developed from different datasets. Also, the paired samples t-test results with 95% confidence were considered for judging the difference between the observed SSC values and the predicted ones. As the results, there were no significant differences between the observed and predicted values in all those PLSR models.

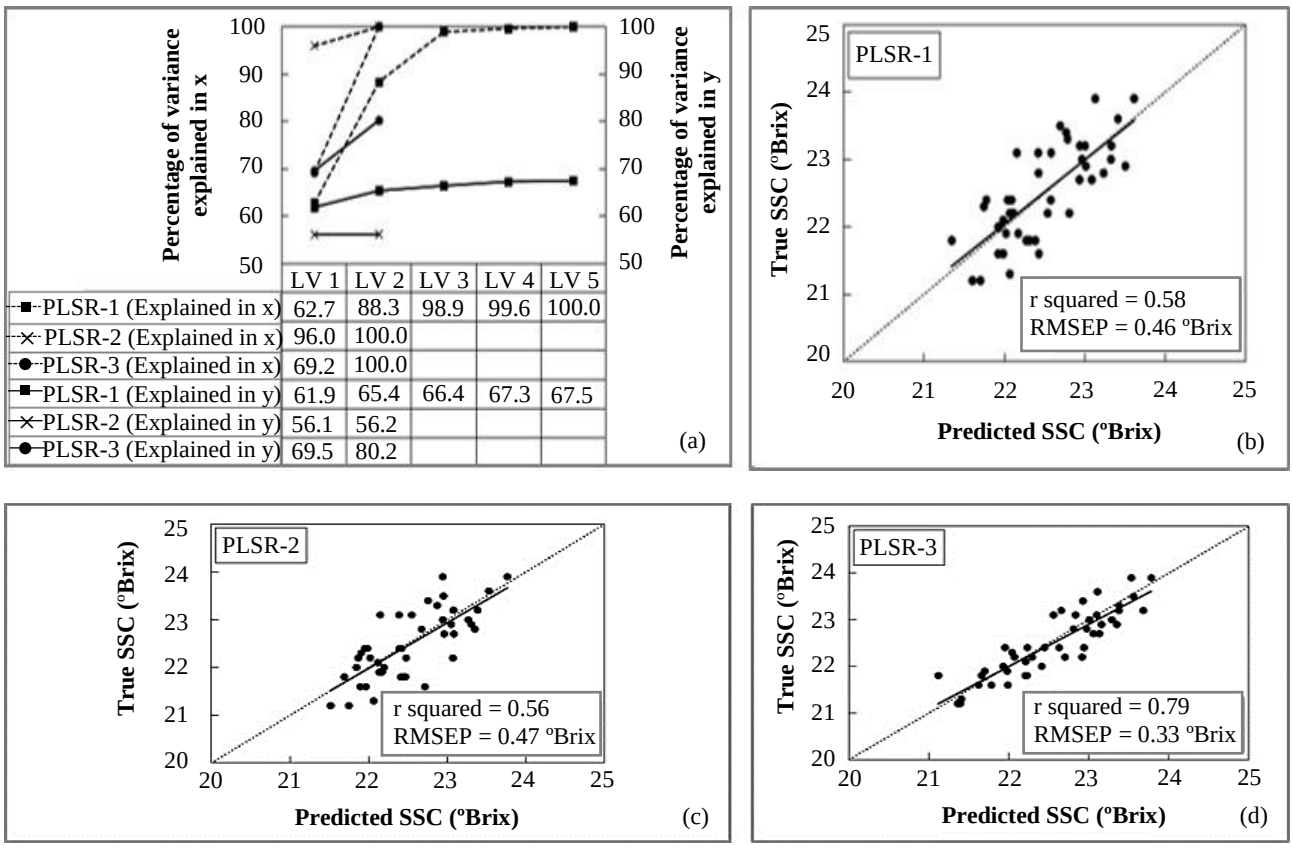


Figure 6 Analysis of the PLSR models; (a) plots of the percent variance explained in x and y variables, (b) validation scatter plots of PLSR-1, (c) validation scatter plots of PLSR-2, (d) validation scatter plots of PLSR-3

Table 1 Regression and validation results of the PLSR approach developed with different datasets

Models	Latent variables (LVs)	R^2	RMSEC (°Brix)	r^2	RMSEP (°Brix)	RPD	P(T<=t) two-tail
PLSR-1	5	0.68	0.41	0.58	0.46	1.54	0.79
PLSR-2	2	0.56	0.48	0.56	0.47	1.51	0.54
PLSR-3	2	0.80	0.32	0.79	0.33	2.14	0.27

Note: R^2 is the coefficient of determination for the calibration.
r^2 is the coefficient of determination for the validation.
RMSEC is a root mean square error of calibration.
RMSEP is a root mean square error of prediction.
RPD is a residual predictive deviation.
PLSR-1 means modeling based on the reflectance at 475, 560, 668, 717, and 840 nm.
PLSR-2 means modeling based on the reflectance at 717 and 840 nm.
PLSR-3 means modeling based on the reflectance at 755 and 890 nm.

4. Discussion

As reported above, model 3 provided the best prediction performance with the r^2 and RMSEP of 0.79 and 0.33 °Brix, respectively. It could be compared directly to Phetpan et al. [9], whose work reported the performance with the r^2 and RMSEP values of 0.79 and 0.30 °Brix, respectively. It confirmed that modeling based on only two sugar-related variables expressed no difference in the predictive results compared to those on the total wavelength variable (450-900 nm) as presented in Phetpan et al. [9].

A key point of the PLSR analysis is to create new latent variables (LVs) in which the first one accounts for the greatest amount of variance. As seen in Figure 6 (a), the amount in LV 1 stood at 62.7 % and 61.9 %, 96 % and 56.1 %, and 69.2 % and 69.5 % in explaining the variance of X and y (SSC values) for the models PLSR-1, PLSR-2, and PLSR-3, respectively. From this point, the PLSR-3 seems much more reasonable, especially when the amount of variance in the next LV (LV 2) is considered, in which it rose significantly to the maximum values at 100 % and 80.2 %, in explaining the variance of X and SSC values, respectively. It could identify and emphasize that the two wavelengths selected at NIR_{755} and NIR_{890} affected the sugar content prediction in sugarcane billets. Because the wavelength at 755 nm might be shifted from the fourth overtone of C-H stretching of sugar at 762 nm [20] or the third overtone of O-H stretching of sucrose in water at 740 nm [21] and the wavelength at 890 nm might be from the third overtone of C-H stretching of sucrose in water at 910 nm [21]. Furthermore, it means that the model (PLSR-3) could predict the sugar content in the cane billets without the interference of color pigments. In this case, some research also identified that the predictor variables related to the chlorophyll absorption in the peel of fruits could compromise model accuracy [16, 22, 23].

The spectral responses at 755 and 890 nm were significant and reasonable to utilize for further research steps from the findings of this work. As introduced regarding the development of the all-in-one system, we propose two promising ideas for future work-investigating the performance of the sugarcane billet's sweetness prediction by using either 1) the optic sensor that responds to those two wavelengths or 2) the digital camera modified for capturing only the NIR_{755} and NIR_{890} based imageries. If the prediction performance still exhibits satisfactory results, we believe that this will play a crucial role in the success of the all-in-one monitor

development for the sugarcane quality and yield. Nevertheless, the dataset employed for modeling in this work was collected from only one variety, area, and period. Therefore, improving the robustness of the model for practical uses with a dataset consisting of a great number of cultivars and agro-climatic conditions (different growing zones) is necessary for further work.

5. Conclusions

This work intended to explore the informative spectral wavelengths significant to predicting the sugarcane billets' sugar content to propose the guidelines for developing a quality and yield monitor for a sugarcane harvester. Based on the exploration of seven wavelengths, two effective wavelengths at 755 and 890 nm were more significant than others. Modeling with these two individual wavelengths provided R^2 and RMSEC of 0.80 and 0.32 °Brix, respectively. The model's predictive performance based on the external validation (test set used) gave r^2, RMSEP, and RPD of 0.79, 0.33 °Brix, and 2.14, respectively. For the modeling based on the spectral responses of the commercial multispectral camera through the dataset of group 1 (475, 560, 668, 717, and 840 nm) and group 2 (717 and 840 nm), the models presented R^2 and RMSEC between 0.56-0.68 and 0.41-0.48 °Brix, respectively. Based on the employed LVs and the predictive performance, the modeling based on 717 and 840 nm representing the invisible bands was better. The model employed 2 LVs and exhibited r^2, RMSEP, and RPD of 0.56, 0.47 °Brix, and 1.51, respectively.

Based on the findings, the reflectance at 755 and 890 nm were the best and could be the informative spectral bands significant to predicting the sugarcane billets' sugar content. They were worthy for developing the sugarcane quality and yield monitor based on either the optic sensor or the modified digital camera system.

6. Acknowledgements

The authors would like to acknowledge the financial support from the Royal Golden Jubilee PhD Scholarship (PhD/0102/2558) of Thailand Research Fund (TRF). We also thank you the Near Infrared Spectroscopy Research Center for Agricultural Products and Food (www.nirsresearch.com) and the Precision Agriculture Laboratory at King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand, for laboratory space and instruments, as well as Mr. Ampol Jongsomboonpokha for supporting the sugarcane samples.

7. References

[1] Bramley RGV. Lessons from nearly 20 years of precision agricultural research, development, and adoption as a guide to its appropriate application. Crop Pasture Sci. 2009;60(3):197-217.
[2] Nawi NM, Chen G, Jensen T. Visible and shortwave near infrared spectroscopy for predicting sugar content of sugarcane based on a cross-sectional scanning method. J Near Infrared Spectrosc. 2013;21(4):289-97.
[3] McCarthy S, Billingsley J. A sensor for the sugar cane harvester topper. Sens Rev. 2002;22(3):242-6.
[4] Klute U. Microwave measuring technology for the sugar industry. Int Sugar J. 2007;109(1308):749-55.
[5] Nelson SO. Potential agricultural applications for RF and microwave energy. Trans ASABE. 1987;30(3):818-31.
[6] Shah S, Joshi M. Modeling microwave drying kinetics of sugarcane bagasse. Int J Electr Eng. 2010;2(1):159-63.
[7] Nawi NM, Chen G, Jensen T, Mehdizadeh SA. Prediction and classification of sugar content of sugarcane based on skin scanning using visible and shortwave near infrared. Biosyst Eng. 2013;115(2):154-61.
[8] Nawi NM, Chen G, Jensen T. In-field measurement and sampling technologies for monitoring quality in the sugarcane industry: a review. Precision Agric. 2014;15:684-703.
[9] Phetpan K, Udompetaikul V, Sirisomboon P. 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 Agric. 2018;154:460-6.
[10] Price RR, Johnson RM, Viator RP, Larsen J, Peters A. Fiber optic yield monitor for a sugarcane harvester. Trans ASABE. 2011;54(1):31-9.
[11] Sugiyama J. Visualization of sugar content in the flesh of a melon by near-infrared imaging. J Agric Food Chem. 1999;47(7): 2715-8.
[12] Xu H, Qi B, Sun T, Fu X, Ying Y. Variable selection in visible and near-infrared spectra: application to on-line determination of sugar content in pears. J Food Eng. 2012;109(1):142-7.
[13] Wang A, Xie L. Technology using near infrared spectroscopic and multivariate analysis to determine the soluble solids content of citrus fruit. J Food Eng. 2014;143:17-24.
[14] Liu R, Qi S, Lu J, Han D. Measurement of soluble solids content of three fruit species using universal near infrared spectroscopy models. J Near Infrared Spectrosc. 2015;23(5):301-9.
[15] Phuphaphud A, Saengprachatanarug K, Posom J, Maraphum K, Taira E. Prediction of the fibre content of sugarcane stalk by direct scanning using visible-shortwave near infrared spectroscopy. Vib Spectrosc. 2019;101:71-80.
[16] Udompetaikul V, Phetpan K, Sirisomboon P. Development of the partial least-squares model to determine the soluble solids content of sugarcane billets on an elevator conveyor. Measurement. 2021;167:107898.
[17] Varmuza K, Filzmoser P. Introduction to multivariate statistical analysis in chemometrics. Boca Raton: CRC Press; 2009.
[18] Abdi H. Partial Least Squares (PLS) regression. In: Lewis-Beck M, Bryman A, Futing T, editors. Encyclopedia of social sciences research methods. Thousand Oaks: SAGE; 2003.
[19] Miller JN, Miller JC. Statistics and chemometrics for analytical chemistry. 6th ed. Hoboken: Prentice Hall; 2010.
[20] Osborne BG, Fearn T, Hindle PT. Practical NIR spectroscopy with applications in food and beverage analysis. 2nd ed. UK: Longman Scientific and Technical; 1993.
[21] Golic M, Walsh K, Lawson P. Short-wavelength near-infrared spectra of sucrose, glucose, and fructose with respect to sugar concentration and temperature. Appl Spectrosc. 2003;57(2):139-45.
[22] Guo Z, Huang W, Peng Y, Chen Q, Ouyang Q, Zhao J. Color compensation and comparison of shortwave near infrared and long wave near infrared spectroscopy for determination of soluble solids content of Fuji apple. Postharvest Biol Technol. 2016;115:81-90.
[23] Travers S, Bertelsen MG, Petersen KK, Kucheryavskiy SV. Predicting pear (cv. Clara Frijs) dry matter and soluble solids content with near infrared spectroscopy. LWT-Food Sci Technol. 2014;59(2):1107-13.