Spectroscopic measurement approaches in evaluation of dry rubber content of cup lump rubber using machine learning techniques

Amorndej Puttipipatkajorn[1], Amornrit Puttipipatkajorn[2*]

(1. *Department of Food Engineering, Faculty of Engineering at Kamphaeng Saen, Kasetsart University,*
Nakhon Pathom, 73140, Thailand;
2. *Department of Computer Engineering, Faculty of Engineering at Kamphaeng Saen, Kasetsart University,*
Nakhon Pathom, 73140, Thailand)

Abstract: Dry rubber content (DRC) is an important factor to be considered in evaluating the quality of cup lump rubber. The DRC analysis requires prolonged laboratory validation. To develop fast and effective DRC determination methods, this study proposed methods to evaluate the DRC of cup lump rubber using different spectroscopic measurement approaches. This involved a complete fundamental analysis leading to an efficient measurement method based on either point-based measurement using NIR reflectance spectrometer or area-based measurement using hyperspectral imaging. A dataset was prepared that 120 samples were randomly divided into a calibration set of 90 samples and a validation set of 30 samples. To obtain an average spectrum to represent a cup lump rubber sample, the spectral data were collected by locating and scanning for point-based and area-based measurement, respectively. The spectral data were calibrated using partial least squares regression (PLSR) and the least-squares support vector machine (LS-SVM) methods against the reference values. The experiments showed that the area-based measurement approach with both algorithms performed outstandingly in predicting the DRC of cup lump rubber and was clearly better than the point-based measurement approach. The best predictions of PLSR represented by the coefficient of determination (R^2), the root mean square error of prediction (RMSEP) and the residual predictive deviation (RPD) were 0.99, 0.72% and 15.17, while the best prediction of LS-SVM were 0.99, 0.64% and 16.83, respectively. In summary, the area-based measurement based on the LS-SVM prediction model provided a highly accurate estimate of the DRC of cup lump rubber.

Keywords: cup lump rubber, dry rubber content, spectroscopic measurement, machine learning, partial least squares regression, least-squares support vector machine

DOI: 10.25165/j.ijabe.20211403.6298

1 Introduction

Natural rubber is an important economic crop in southeast Asia. Thailand is one of the leading producers of natural rubber with about a 35% share of the total production. According to the Office of Agricultural Economics[1], Thailand exported 2.2 million t of natural rubber in 2019 which provided revenue from both export and employment. Natural rubber is used extensively by many manufacturing companies in the rubber industry, either alone or in combination with other materials. Currently, farmers produce rubber not only in the form of latex but also in various types of rubber products, for example Standard Thai Rubber (STR), ribbed smoked sheet (RSS), air dried sheet (ADS), crepe, skim rubber and concentrated latex. STR 20 grade is a high grade of STR rubber and makes up approximately 60% of all STR rubber in the export market. Due to its physical properties, this high-graded rubber (STR 20) is suitable for manufacturing high quality rubber products for the automobile industry, for surgical and

pharmaceutical use and molded and extruded rubber products in general.

STR 20 rubber originates from cup lump rubber under a quality control process. In the major areas of Thailand, cup lump rubber is formulated by coagulating fresh latex with acid in a collection cup. The coagulated cup lumps are collected daily and processed into dry forms for trading. The process to produce cup lump rubber is uncomplicated with respect to other forms of natural rubber such as ribbed smoked sheet or air dry sheet rubber; consequently, most rubber farmers prefer to sell their natural rubber as cup lump rubber rather than in other forms. The price of cup lump rubber depends mainly on its dry rubber content (DRC). In the current market, traders estimate roughly the DRC of cup lump rubber by seeing and touching samples without reference to any standards. The resultant difficulty in applying a consistent approach motivated the current research to determine a model that could predict the DRC of cup lump rubber rapidly and precisely.

Near-infrared spectroscopy (NIRS) is a type of vibrational spectroscopy based on the wavelength range of 750 to 2500 nm (wavenumbers: 13300 to 4000 cm^{-1}). As a fast and non-destructive method, NIRS has been widely used for rapid analysis of the moisture, protein and fat content of a wide variety of agricultural and food produce[2,3]. Also, it has been proven ability to quantify trace amounts of moisture in a raw rubber sheet. The results indicated that a strong combination of absorption bands for water was at approximately 1940 nm and the first, second and third

overtones were at 1450 nm, 970 nm and 760 nm, respectively[4]. NIRS has also been used successfully to evaluate the moisture content of natural rubber in various forms such as cup lump rubber[5] and concentrated latex[6].

Hyperspectral imaging, like other spectroscopic techniques, is usually carried out in the Vis-NIR (400-1000 nm) or NIR (1000-1700 nm) ranges for food and agricultural produce[7]. Hyperspectral images, known as a hypercube[8], are made up of hundreds of contiguous wavebands in the spatial dimension of the target being studied. The hypercube allows for visualization of the biochemical constituents of a sample, separated into given areas of the image, depending on their spectral properties. In recent years, there has been a rapid expansion in the use of hyperspectral imaging in various analytical processes over a broad range of food and agricultural produce such as poultry[9], pork[10] and fruits[11]. The current study proposed a point-based approach using NIR reflectance spectrometry and area-based measurement using a hyperspectral imaging system to demonstrate the capabilities of hyperspectral imaging compared to the normal spectroscopic measurement method. Besides, two algorithms (partial least squares regression (PLSR) and least-squares support vector machine (LS-SVM)) were compared to assess the performance of the proposed models for assessing the DRC of cup lump rubber.

2 Materials and methods

2.1 Cup lump rubber preparation

The fresh latex was sampled from a rubber plantation in eastern Thailand in October 2017. In total, 120 cup lump rubber samples were prepared from fresh latex. Each sample was coagulated using 0.3% diluted formic acid which took approximately 30 min. At this stage, the liquid serum inside the cup lump rubber was released and separated off. After the cup lump rubber had completely coagulated, all cup lump rubber pieces were weighed to determine the initial mass. The dry matter of each cup lump rubber was calculated by multiplying its initial mass by the DRC of fresh latex measured according to the laboratory method defined by AOAC standards[12]. All samples were stored and left to dry at ambient temperature for 1 d, 3 d, 5 d, 7 d, 9 d and 12 d. In each period, 20 samples (15 samples for calibration and 5 samples for validation) were weighed again for their final mass. Then, the DRC (%) was calculated using Equation (1). Table 1 shows the variation in the DRC of cup lump rubber in each period, with the DRC increasing with increasing storage time.

$$DRC = \frac{Dry\ matter}{Final\ mass} \times 100\% \qquad (1)$$

Table 1 Variation in the DRC of cup lump rubber in each period

Time /d	Number of samples		DRC range/%		Mean ±SD	
	Calibration	Validation	Calibration	Validation	Calibration	Validation
1	15	5	58.78-61.52	59.64-65.84	59.94±0.70	61.11±2.66
3	15	5	64.26-72.11	63.64-81.20	67.12±2.15	69.12±6.97
5	15	5	72.61-85.69	74.95-86.22	78.62±4.30	80.97±5.18
7	15	5	78.46-93.26	77.53-85.67	85.35±4.28	81.32±3.14
9	15	5	77.81-89.53	86.10-90.29	85.20±2.97	87.87±1.54
12	15	5	79.57-92.25	90.37-90.88	89.27±2.95	90.65±0.20

2.2 Spectral measurement methods

The cup lump spectrum was collected using two measurement approaches. The first approach, called point-based measurement, used NIR reflectance spectrometry in the range 901.35-1700.64 nm with 3.9 nm resolution produced by Texas Instruments, Inc., USA.

The NIR reflectance spectrometer was placed on the surface of the cup lump rubber sample to acquire a spectrum at four different positions as shown in Figure 1a. Each measurement point was based on an average of 6 individual readings. Afterward, the total spectrum was collected. Readings were averaged and transferred to the Matlab software package for spectral preprocessing and multivariate analysis.

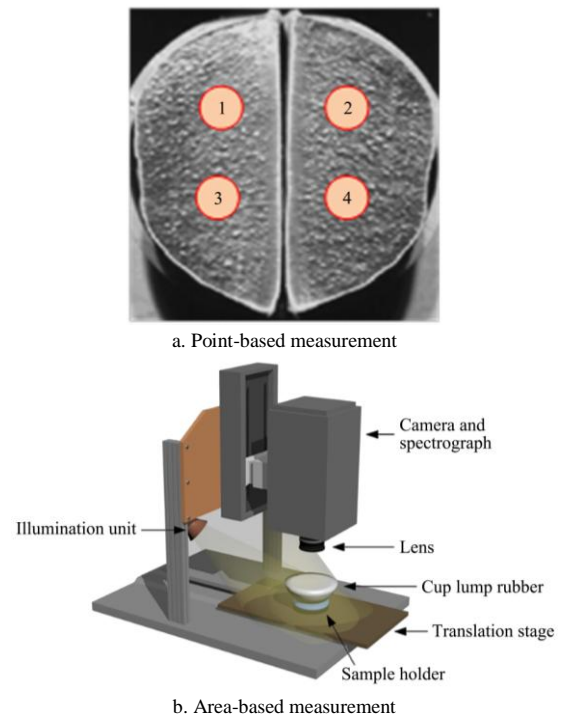a. Point-based measurement

b. Area-based measurement

Figure 1   Cup lump rubber measurement methods

The second approach called area-based measurement used the hyperspectral imaging system manufactured by SPECIM, Spectral Imaging Ltd., Finland. The system was composed of a temperature-stabilized InGaAs camera equipped with an imaging spectrograph, a fore objective lens (OLES15), an illumination unit (tungsten halogen lamps, 20 W), a translation stage and a computer equipped with the data acquisition and control software. The hyperspectral imaging system was configured to operate spatially at 10 mm/s. The illumination unit was inclined at 35 degrees to the vertical to allow the proper direction of the reflected light to the camera lens. The optimal spectrum came from scanning the surface of the cup lump rubber can be tuned by adjusting the camera position and focal length of the lens as shown in Figure 1b. It provided 256 bands through a wavelength range from 864.53 to 1695.08 nm with 3.3 nm spectral resolution and 320 pixels for spatial resolution. White and dark reference spectral images were acquired in each image acquisition. The raw reflectance image was based on Equation (2).

$$R = \frac{I_0 - I_d}{I_w - I_d} \qquad (2)$$

where, R is the raw reflectance image; I_0 is the sample image; I_d is the dark image acquired by completely closing the electro-mechanical shutter of the camera; I_w is the white reference image, which collects 99% diffuse reflectance from the white standard. The image spectra were stored in the computer using the SPECIM software interface and analyzed using the Matlab software.

2.3 Spectral preprocessing

In reflectance mode, the reflected light travels different distances from the sample to the detector. A longer light path

produces a lower relative reflectance value. This causes spectral translation and affects the spectral model. Pretreatment of spectra is a required part of spectral analysis and can improve the accuracy of analysis[13]. For this reason, spectral pretreatment was applied to the cup lump rubber spectra to reduce problems associated with noise, light-scattering and external effects prior to implementing the regression analysis. To reduce the light scattering, the signal normal variate (SNV) transformation[14] was primarily necessary for spectral pretreatment before applying the other spectral transformations. The first and second derivative transformations

were used to remove the additive and multiplicative effects in the spectra[15]. Thus, a combination of spectral pretreatments was selected (SNV, SNV followed by the first derivative and the SNV followed by the second derivative) to compare their influences on model performance.

Figure 2a shows the pre-processing procedures for the point-based measurement. The aim was to obtain the average spectrum of each sample from 4-point spectral measurement using different kinds of pretreatment methods. Area-based measurement as shown in Figure 2b used a more sophisticated procedure by
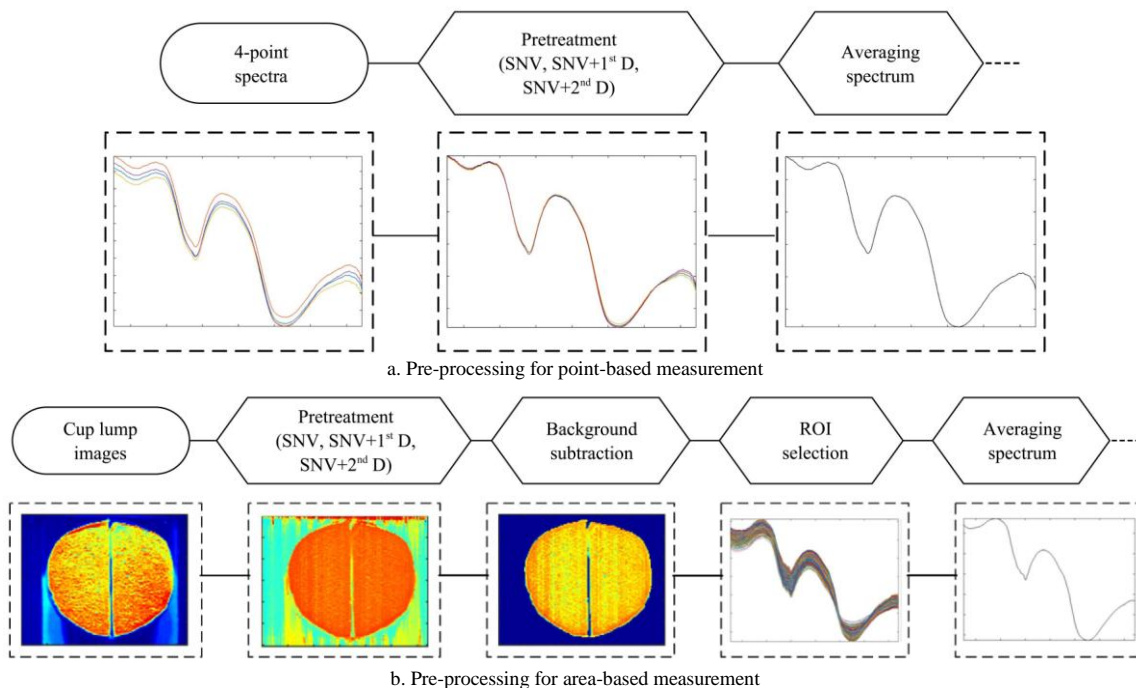

a. Pre-processing for point-based measurement

b. Pre-processing for area-based measurement

Figure 2   Pre-processing procedures to obtain the average spectrum for point-based and area-based measurement

a. Selected wavelengths for threshold value

b. SNV

c. SNV followed by 1st derivative

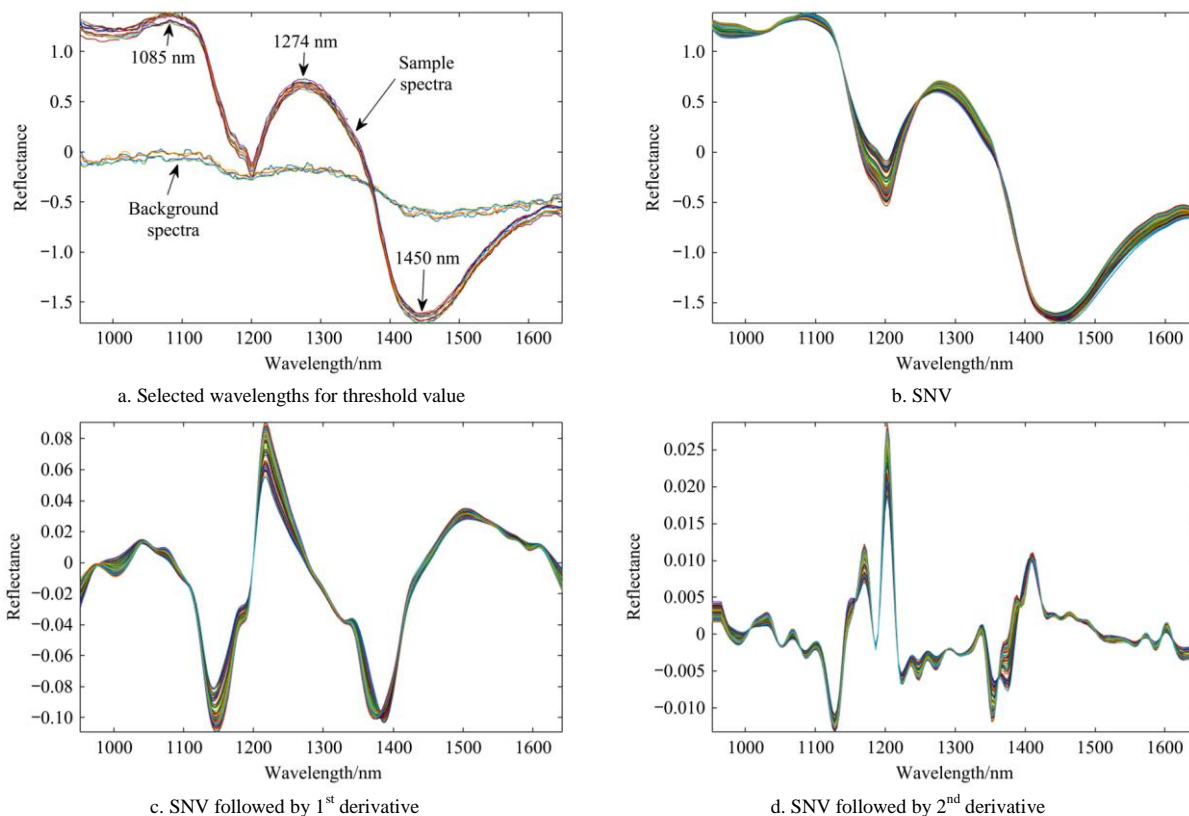
d. SNV followed by 2nd derivative

Figure 3   ROI selection using threshold at certain bands and an average spectrum of samples in calibration set when different spectral preprocessing methods were applied

considering the spatial dimension of the spectrum. All spectra of images were first preprocessed, and an average spectrum of the cup lump image was obtained based on a background subtraction using the thresholding technique, because of its simplicity and lower computational load. Inspection of the spectra indicated that at approximately 1085 nm, 1274 nm and 1450 nm, the reflectance values of the sample and the background were clearly different, as shown in Figure 3a. At these wavelengths, the threshold values were defined in terms of the upper and lower limits of the reflectance value that covered only the cup lump rubber spectra. Therefore, if a given pixel spectrum had a reflectance value in the range, that pixel was picked up as a region of interest (ROI). The averaged ROI was used to represent the spectrum of a cup lump sample. Finally, all average spectra obtained from both point-based and area-based measurements were extracted in the range 942 to 1650 nm for further analysis, because outside this range strong noise was found in the spectrum. Figure 3b-d shows a group of the average spectra preprocessed using different kinds of spectral pretreatments in the calibration set prepared for model development.

2.4   Calibration model

2.4.1   Partial least squares regression

The PLSR model has become the most frequently applied in NIR spectroscopy[16]. The PLSR algorithm can effectively remove the co-linearity problem of spectral data[17]. A PLSR model expresses the linear relationship between the independent variables or spectral variables (X) and the dependent variables or reference values (Y). The optimal number of latent variables (LVs) is a key parameter in construction of calibration model and directly affects the predictive performance of the model. The maximum LVs is accepted at approximately 10[18]. However, a low number of LVs is desirable to avoid the inclusion of signal noise in the model[19], whereas, on the other hand, using more LVs in the model causes over-fitting problems. The optimal number of LVs should be selected to avoid over-fitting and to maximize the covariance between the X and Y space. In this study, the PLSR algorithm was implemented on Matlab R2017a (The Math Works, Natick, Ma, USA).

2.4.2   Least-squares support vector machine

The support vector machine (SVM) was primarily designed for data classification tasks[20]. The modified version of SVM called LS-SVM[21] is an easier, robust approach for the classification and regression analysis of linear and nonlinear multivariate problems. Recently, it has been widely used in the area of chemometrics[22-25] and hyperspectral imaging applications[26-28]. In the current study, LS-SVM was used to create a calibration model for predicting the dry rubber content of cup lump rubber. The model parameters of LS-SVM regression were found by solving the optimization task shown in Equation (3).

$$\min_{w,b,e} J(w,e) = \frac{1}{2}w^T w + \gamma\frac{1}{2}\sum_{k=1}^{N} e_i^2 \qquad (3)$$

subject to $y_i = w^T \varnothing(x_i) + b + e_i$, $i = 1,...,n$.

where, e_i is the error variable; γ is the regularization parameter which penalizes the error; $\varnothing(\cdot)$ is kernel space mapping function that mapping the sample in origin space to a vector in the high dimensional feature space; w is the weighting vector in the dimension of feature space; b is the bias. According to the optimization equation, Lagrange function is obtained as follows:

$$L(w,b,e,a) = J(w,b,e) - \sum_{i=1}^{n} a_i[w^T \varnothing(x_i) + b + e_i - y_i] \qquad (4)$$

where, a_i are the Lagrange multipliers. When the Lagrange equation and the Karush-Kuhn-Tucker conditions for optimality are combined, a set of linear equations is formulated and solved. Eventually, the LS-SVM regression model can be expressed as Equation (5):

$$y(x) = \sum_{i=1}^{n} a_i K(x, x_i) + b \qquad (5)$$

where, $K(x, x_i)$ is the kernel function that in the current study was a Gaussian radial basis function (RBF) as shown in Equation (6):

$$K(x, x_i) = \exp(\frac{\| x - x_i \|^2}{2\sigma^2}) \qquad (6)$$

This algorithm was tested in Matlab R2017a software using the LS-SVMlab toolbox[29].

2.5   Model evaluation

A comparison of the reliability and efficient prediction from the above-mentioned models was based on the following statistics. The coefficient of determination (R^2) provides information on the goodness-of-fit of a model. It is interpreted as the proportion of the variance in the dependent variable that is predictable from the independent variable. The root means square error (RMSE) represents the mean absolute error of the time-series calculated by the model between the reference y_i and predicted $\hat{y}_i$ values. The residual predictive deviation (RPD) is a goodness-of-fit measure that can be used to analyze how well a set of data points fits with the actual model. The RPD is calculated by dividing the standard deviation (SD) by RMSE. Interpretation of the RPD depends on the material analyzed. According to certain literature, if the RPD is less than 2, the model performance is insufficient for prediction, but there is room for improvement[30]. An RPD between 2 and 5 provides adequate accuracy for good estimations, an RPD between 5 and 10 indicates high accuracy and is suitable for analysis analogous to that of the reference method and if the RPD is higher than 10, the model shows excellent performance[31,32]. In general, a good model should yield high values of R^2 and RPD while producing low values of RMSE. The parameters used are presented in Equations (7)-(9).

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \qquad (7)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \qquad (8)$$

$$\text{RPD} = \frac{\text{SD}}{\text{RMSE}} \qquad (9)$$

3   Results and discussion

The entire dataset of 120 samples used in this experiment was randomly divided into a calibration set of 90 samples for developing the models and a validation set of 30 samples for model testing. Suitable LVs or latent variables must be identified in the development of the PLSR model. In general, the optimal LV is discovered using a cross-validation technique[33]. In the current study, leave-one-out cross-validation (LOOCV) was implemented to locate optimal LVs. This algorithm will partition data into k randomly chosen subsets, where k is equal to the total number of observations in the calibration set so that all data will be used once as a test set. Models were created in each round of a given test set with the increase in the number of LVs in each being one. The LVs that resulted in the minimum mean square error (MSE) were chosen as the optimal LVs of the model. In this study, the PLSR

models originating from the different kinds of spectral pretreatment (SNV, SNV + the first derivative, and SNV + the second derivative) had optimal LVs of 10, 10 and 8 for the point-based measurement and had optimal LVs of 12, 10 and 15 for the area-based measurement, respectively.

Unlike the PLSR model, the LS-SVM model used Gaussian RBF as the kernel function and needed two parameters, namely the squared bandwidth of the Gaussian curve (σ^2) and the regularization parameter (γ) that determines the trade-off between training error minimization and smoothness[29]. These parameters were tuned to search for the optimal value using coupled simulated annealing for initial estimation and the simplex method for fine tuning. Table 2 shows the tuning parameters for the point-based and area-based measurements based on different kinds of spectral pretreatment.

Table 2   Tuning parameters of LS-SVM model variants

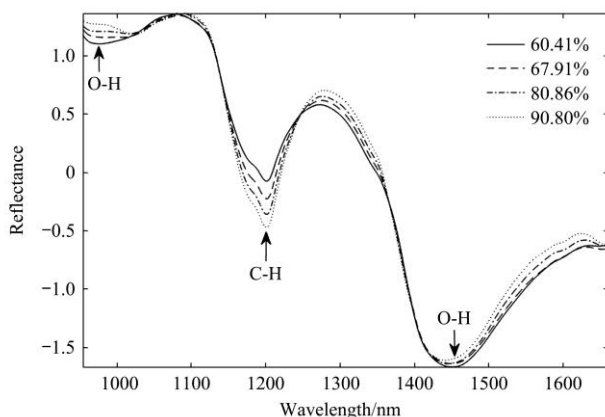| Pretreatment	Point-based measurement		Area-based measurement	
	γ	σ^2	γ	σ^2
SNV	786.67	1600.10	951.68	824.48
SNV + 1st derivative	8862.40	3938.65	56.06	459.59
SNV + 2nd derivative	1931.00	2474.16	634.76	2153.06

The performance of all proposed models was compared based on validation set testing and the results are listed in Table 3 indicates both the PLSR and LS-SVM models provided a good prediction of dry rubber content based on the respective values of RMSEP, R^2 and RPD for both measurement approaches, because there were high variations of reflectance values at 960-1000 nm, at approximately 1200 nm and 1450-1600 nm when the rubber content differed. At approximately 970 nm and 1450 nm, they could be associated with the second and first overtones of the O-H vibration of water[34,35]. The rubber content as the main ingredient of cup lump rubber is a polymer partially constructed with C-H

bonding that was attributed to the second overtone vibration at approximately 1200 nm[6], as shown in Figure 4a. This allowed the calibration models to capture those differences and to subsequently accurately estimate the DRC of the cup lump rubber. In addition, the predictive ability of the area-based measurement method was significantly better than that of the point-based measurement, because the spectrum shape derived from area-based measurement was more complete than from the point-based approach that generated deviation in the spectrum. These differences are evident by comparing point-based measurement with various spectral pretreatments in Figures 4b, 4d and 4f with those of area-based measurement in Figures 4a, 4c and 4e. For example, the spectrum shapes in Figure 4e at approximately 1000-1100 nm, 1200-1400 nm and 1500-1600 nm were smoother and clearer than the spectrum shapes in Figure 4f. For point-based measurement, the models with SNV pretreatment had the best accuracy with both the PLSR and LS-SVM algorithms perhaps because there was less deviation in the spectra than for others. In contrast, for the area-based measurement, the spectra pre-processed using SNV followed by the 2nd derivative produced the best results, because their spectra were more explicit so that the additional derivative technique could effectively reduce additive and multiplicative effects in the spectrum. Therefore, for point-based measurement, the PLSR model with SNV provided the best prediction with values of R^2 = 0.96, RMSEP = 2.20% and RPD = 4.84, while the best prediction values using LS-SVM were 0.97, 1.84% and 5.61, respectively. For area-based measurement, the PLSR model with SNV+2nd derivative represented the best prediction with values for R^2, RMSEP and RPD of 0.99, 0.72% and 15.17, respectively, and the best prediction values using LS-SVM were 0.99, 0.64% and 16.83, respectively. Additionally, the results of area-based measurement had very high RPD values (*RPD* > 10), indicating that the models were acceptable for use in all analytical tasks.

Table 3   Performance comparison of PLSR and LS-SVM models based on different measurement methods

Method	Spectra pretreatments	PLSR			LS-SVM		
		R^2	RMSEP/%	RPD	R^2	RMSEP/%	RPD
Point-based	SNV	0.96	2.20	4.84	0.97	1.84	5.61
	SNV+1st derivative	0.96	2.28	4.66	0.97	1.90	5.54
	SNV+2nd derivative	0.95	2.40	4.26	0.96	2.10	5.03
Area-based	SNV	0.99	0.90	11.86	0.99	0.90	12.10
	SNV+1st derivative	0.99	0.97	11.16	0.99	0.74	14.69
	SNV+2nd derivative	0.99	0.72	15.17	0.99	0.64	16.83

Note: RMSEP is the root mean square error of prediction.

a. Area-based measurement (SNV)

b. Point-based measurement (SNV)

c. Area-based measurement (SNV+1st derivative)

d. Point-based measurement (SNV+1st derivative)

e. Area-based measurement (SNV+2nd derivative)
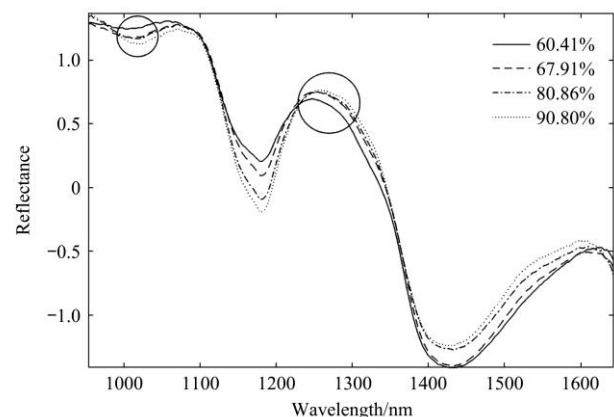
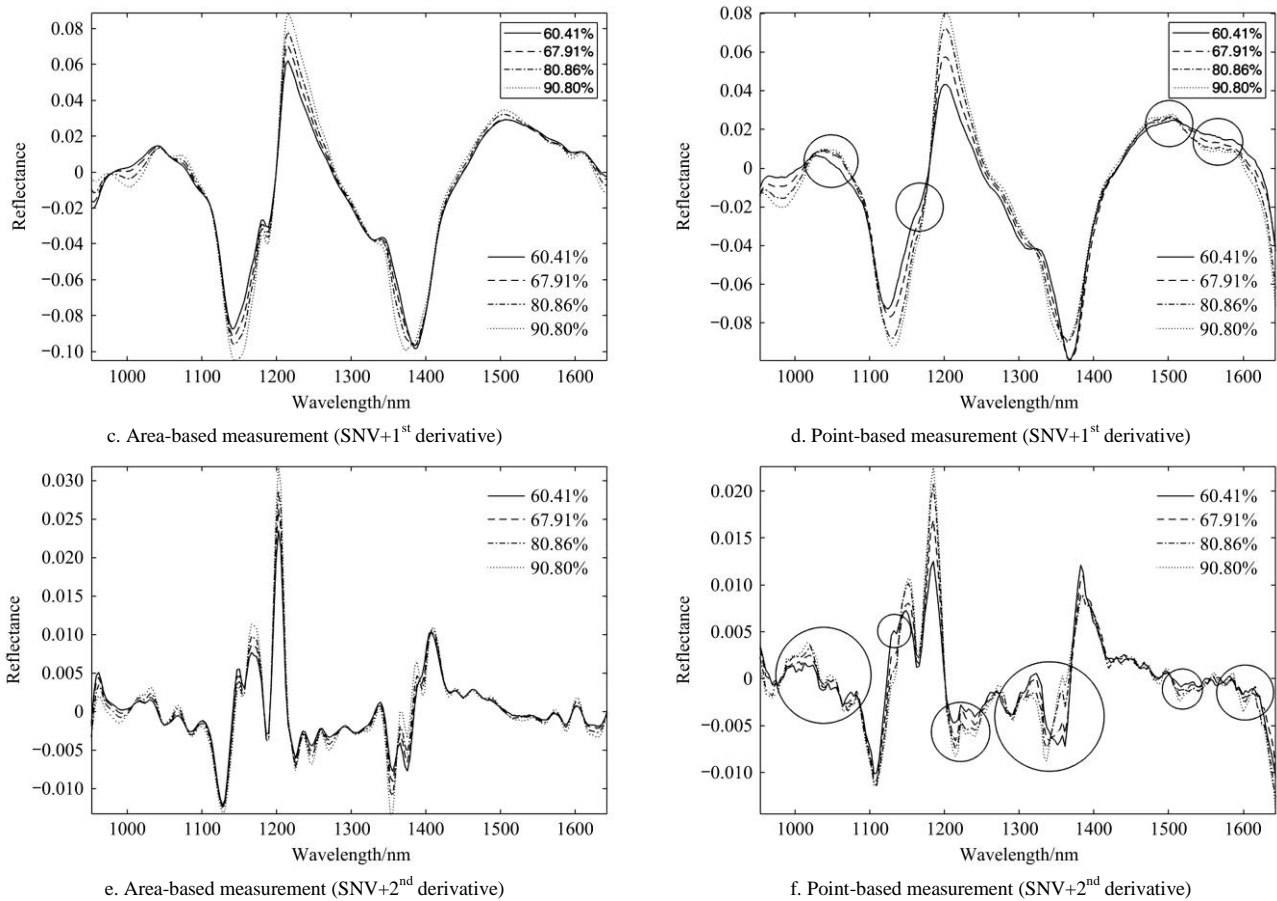f. Point-based measurement (SNV+2nd derivative)

Figure 4   Average spectra of cup lump rubber for different degrees of DRC

The results indicated that the LS-SVM model was slightly better than the PLSR model. The nonlinearity of the LS-SVM model meant that it had higher sensitivity than a PLSR model so that the LS-SVM models could track the spectra more closely than the PLSR models and eventually predicted the DRC better than the PLSR models. These findings confirmed that a nonlinear regression method such as LS-SVM was a good alternative algorithm when dealing with spectroscopy. The scatter plots in Figure 5 depict the best outcomes of the PLSR and LS-SVM models for both point-based and area-based measurements. These plots show that models derived from area-based measurement could predict the percentage of DRC closely.
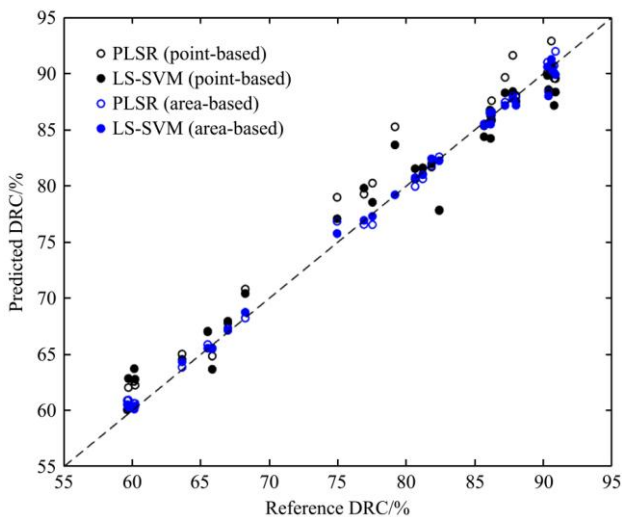

Figure 5   Scatter plots of DRC for PLSR and LS-SVM models based on different measurement approaches

4 Conclusions

The experiments showed that the area-based measurement method performed outstandingly in predicting the DRC of cup lump rubber and was more efficient than the point-based measurement. The best prediction of PLSR had values of R^2=0.99, RMSEP = 0.72% and RPD = 15.17, while the best prediction of LS-SVM had values of 0.99, 0.64% and 16.83, respectively. Although the point-based measurement method could predict the DRC of cup lump rubber with good accuracy, it still had problems of reliability and repeatability due to the lack of a spatial dimension. On the other hand, the area-based measurement approach was able to accommodate the spatial dimension which resulted in obtaining extensive spectral data from a sample so that this approach could provide more accurate results. Nevertheless, data analysis using area-based measurement needs more computational resources and consumes extra processing time due to its large spectral dataset and includes many processing steps such as image processing, spectral processing and multivariate regression. In addition, the hyperspectral imaging system was not portable and was quite costly. However, its implementation would be possible in a large-scale rubber industry with a short payback period. Finally, this study confirmed the potential of the area-based measurement method using hyperspectral imaging and machine learning techniques for predicting the DRC with very high accuracy.

[References]

[1] Office of Agricultural Economics (OAE). Rubber Research Institute Department of Agriculture. Available: http://www.oae.go.th/view/1/ Information/EN-US. Accessed on [2020-06-06].

[2] Norris K H. Design and development of a new moisture meter. Agric. Eng., 1964; 45(7): 370–372.

[3] Maneewan W, Limpiti S, Theanjumpol P. The application of near infrared reflectance spectroscopy to determine the moisture and protein content in soybean seeds. The 14th International Conference of NIR Spectroscopy, Bangkok, Thailand, 2009.

[4] Rittiron R, Seehalak W. Moisture content in raw rubber sheet analyzed by transflectance near infrared spectroscopy. J. Innov. Opt. Health Sci., 2014; 7(4): 1350068. doi: 10.1142/S1793545813500685.

[5] Suchata S, Theanjumpolb P, Karrilaa S. Rapid moisture determination for cup lump natural rubber by near infrared spectroscopy. Industrial Crops and Products, 2015; 76: 772–780.

[6] Sirisomboon P, Kaewkuptong A, Williams P. Feasibility study on the evaluation of the dry rubber content of field and concentrated latex of Para rubber by diffuse reflectance near infrared spectroscopy. Journal of Near Infrared Spectroscopy, 2013; 21: 81–88.

[7] Davies A M, Grant A. Review: near-infra-red analysis of food. Int. J. Food Sci. Technol., 1987; 22: 191–207.

[8] Lu R F, Chen Y R. Hyperspectral imaging for safety inspection of food and agricultural products. In Proceedings of SPIE Conference on Pathogen Detection and Remediation for Safe Eating, Boston, US, 1998.

[9] Park B. Quality inspection of poultry carcasses. In D.W. Sun (Ed.). Computer vision technology for food quality evaluation. Amsterdam: Academic Press, 2008; pp. 157–187.

[10] Barbin D, ElMasry G, Sun D W, Allen P. Near infrared hyperspectral imaging for grading and classification of pork. Meat Science, 2012; 90(1): 259–268.

[11] Ariana D, Guyer D E, Shrestha B. Integrating multispectral reflectance and fluorescence imaging for defect detection on apples. Computers and Electronics in Agriculture, 2016; 50(2): 148–161.

[12] AOAC. Official Methods of Analysis of AOAC International. 2000. Available: http://www.aoac.org/aoac_prod_imis/AOAC/Publications/ Official_Methods_of_Analysis/AOAC_Member/Pubs/OMA/AOAC_Offici al_Methods_of_Analysis.aspx. Accessed on [2020-04-01].

[13] Nawi N M, Chen G, Jensen T, Mehdizadeh S A. Prediction and classification of sugar content of sugarcane based on skin scanning using visible and shortwave near infrared. Biosystems Engineering, 2013; 115: 154–161.

[14] Barnes R J, Dhanoa M S, Lister S J. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Appl. Spectrosc., 1989; 43: 772–777.

[15] Dos Santos C A T, Lopo M, Páscoa R N M J, Lopes J A. A review on the applications of portable near-infrared spectrometers in the agro-food industry. Appl. Spectrosc., 2013; 67: 1215–1233.

[16] Martens H, Næs T. Multivariate calibration (2nd edition). Chichester: John Wiley & Sons. 1991; 438p.

[17] Shao Y Y, Gao C, Xuan G T, Gao X M, Chen Y Q, Hu Z C. Determination of damaged wheat kernels with hyperspectral imaging analysis. Int J Agric & Biol Eng, 2020; 13(5): 194–198.

[18] Moghimi A, Aghkhani M H, Sazgarnia A, Sarmad M. Vis/NIR spectroscopy and chemometrics for the prediction of soluble solids content and acidity (pH) of kiwifruit. Biosystems Engineering, 2010; 106: 295–302.

[19] Zhao X B, Zhao J W, Huang X Y, Li Y X. Use of FTNIR spectrometry in non-invasive measurements of soluble solid contents (SSC) of "Fuji" apple based on different PLS models. Chemometrics and Intelligent Laboratory Systems, 2007; 87: 43–51.

[20] Vapnik V, Lerner A. Pattern recognition using generalized portrait method. Automation and Remote Control, 1963; 24: 774–780.

[21] Suykens J A K, Van Gestel T, De Brabanter J, De Moor B, Vandewalle J, Van Gestel T. Least squares support vector machine (Vol. 4). Singapore: World Scientific, 2002.

[22] Sun T, Lin H, Xu H, Ying Y. Effect of fruit moving speed on predicting soluble solids content of 'Cuiguan' pears (Pomaceae pyrifolia Nakai cv. Cuiguan) using PLS and LS-SVM regression. Postharvest Biol. Technol., 2009; 51(1): 86–90.

[23] Yua H Y, Niua X Y, Lina H J, Yinga Y B, Lib B B, Panc X X. A feasibility study on on-line determination of rice wine composition by Vis–NIR spectroscopy and least-squares support vector machines. Food Chem., 2009; 113(1): 291–296.

[24] Morellosa A, Pantazia X E, Moshou D, Alexandridisc T, Whettonb R, Tziotziosa G, et al. Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. Biosystems Eng., 2016; 152: 104–116.

[25] Han Z, Cai S, Zhang X, Qian Q, Huang Y, Dai F, et al. Development of predictive models for total phenolics and free p-coumaric acid contents in barley grain by near-infrared spectroscopy. Food Chem., 2017; 227: 342–348.

[26] Chen J H, Sun D W. Data fusion and hyperspectral imaging in tandem with least squares-support vector machine for prediction of sensory quality index scores of fish fillet. Food Sci. Technol., 2015; 63(2): 892–898.

[27] Kamruzzaman M, Makino Y, Oshita S. Hyperspectral imaging for real-time monitoring of water holding capacity in red meat. Food Sci. Technol., 2016; 66: 685–691.

[28] Yang Q, Sun D W, Cheng W. Development of simplified models for nondestructive hyperspectral imaging monitoring of TVB-N contents in cured meat during drying process. Food Eng., 2017; 192: 53–60.

[29] De Brabanter K, Karsmakers P, Ojeda F, Alzate C, De Brabanter J, Pelckmans K. LS-SVM toolbox user's guide: Version 1.8. ESAT-SISTA Technical Report 10–146, 2011; 110p.

[30] Abu-Khalaf N, Hmidat M. Visible/Near Infrared (VIS/NIR) spectroscopy as an optical sensor for evaluating olive oil quality. Computer and Electronics in Agriculture, 2020; 173: 105445. doi: 10.1016/ j.compag.2020.105445.

[31] Williams P. Tutorial: The RPD statistic: A tutorial note. NIR News, 2014; 25(1): 22.

[32] Esbensen K H, Geladi P, Larsen A. The RPD myth. NIR News, 2014; 25(5): 24.

[33] Redaellia R, Alfieria M, Cabassi G. Development of a NIRS calibration for total antioxidant capacity in maize germplasm. Talanta, 2016; 154: 164–168.

[34] Ye M, Gao Z, Li Z, Yuan Y, Yue T. Rapid detection of volatile compounds in apple wines using FT-NIR spectroscopy. Food Chemistry, 2016; 190: 701–708.

[35] Hans G, Leblon B, Cooper P, Rocque A L, Nader J. Determination of moisture content and basic specific gravity of Populus tremuloides (Michx.) and Populus balsamifera (L.) logs using a portable near-infrared spectrometer. J. Wood Mater Sci Eng., 2015; 10: 3–16.