Asia-Pacific Journal of Science and Technology: Volume: 29. Issue: 02. Article ID.: APST-29-02-12. Research Article

APST

Asia-Pacific Journal of Science and Technology

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

Published by the Research and Graduate Studies Division,
Khon Kaen University, Thailand

Fusion of multimode near-infrared detections using multiblock data analysis for prediction of mixed sugar content

Chanat Thanavanich[1], Nutthatida Phuangsaijai[1], Sujitra Funsueb[1], Parichat Theanjumpol[2,3] and Sila Kittiwachana[1,*]

[1] Department of Chemistry, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand
[2] Postharvest Technology Research Center, Faculty of Agriculture, Chiang Mai University, Chiang Mai 50200, Thailand
[3] Postharvest Technology Innovation Center, Ministry of Higher Education, Science, Research and Innovation, Bangkok 10400, Thailand
[*]Corresponding author: silacmu@gmail.com; sila.k@cmu.ac.th

Received 15 August 2022
Revised 16 November 2022
Accepted 30 November 2022

Abstract

This study combined NIR spectral data from different detection modes to enhance the quantification performance of the instrument owing to the benefits of various detection modes in the near-infrared (NIR) spectroscopic technique. Mixed sugar samples, used for the demonstration, composed of glucose, fructose, maltose, and sucrose, were composed based on a mixture design. NIR spectra were recorded using portable Vis-NIR (360.9–1078.3 nm) and benchtop NIR (400–2500 nm) spectrometers using both transmittance and transflectance detection modes. A multiblock principal component analysis (MB-PCA) was applied to exploratorily analyze the multiblock data. Multiblock regression models, including concatenated partial least squares (C-PLS), serial-PLS (S-PLS), and multiblock-PLS (MB-PLS), were employed to quantify the sugar sample concentrations. MB-PCA could easily distinguish between the sugars, accounting for 98.51 % and 100.00 % for the portable Vis-NIR and benchtop NIR spectra on the first two principal components (PCs). The spectral fusion using the multiblock data analysis could improve the predictive performance. The best results were based on the use of MB-PLS, with values of R^2 of 1.00, Q^2 of 0.96, and 1.00; root mean square error of calibration (RMSEC) of 0.02 and 0.05; root mean square error of cross validation (RMSECV) of 0.22, and 0.06; ratio of prediction to deviation (RPD) of 5.03 and 18.43; and relative standard deviation (RSD) values of 5.63 and 1.54 using the portable Vis-NIR and benchtop NIR spectral data, respectively.

Keywords: Chemometrics, Multiblock data analysis, Near-infrared spectrometer, Non-destructive detection, Sugar content

1. Introduction

Near-infrared (NIR) spectroscopy is a spectroscopic technique that measures electromagnetic absorption in the region of 800–2500 nm [1]. The signals detected in the NIR region are overtone and combination bands derived from common infrared (IR) absorptions. The main advantage of NIR measurement is that the analysis is rapid, non-destructive, non-invasive, and requires little or no sample preparation [1, 2]. NIR spectroscopy has been widely used in various fields, including pharmaceutical, fuel, food and beverage production, and agriculture [3-6]. When considering how light interacts with samples, NIR detection can be categorized into two modes: transmittance and reflectance. The reflectance detection mode is suitable for solids, whereas the transmittance detection mode is often used for liquid samples. A measurement called transflectance is a subdivision of the transmittance detection mode. The transflectance measurement allows the NIR radiation to pass through the sample before being reflected by the reflection plates, resulting in a doubling of the optical path length [1, 7].

Nowadays, it is possible to merge or combine spectroscopic data from different sources to enhance predictive performance in both qualitative and quantitative analyses. Multiblock data analysis is a chemometric method that can be used to combine measured data from various sources into a single analysis, resulting in clearer comprehension and enhanced predictive performance of multivariate models[8, 9]. For example, Biancolillo et al. [10] utilized a multiblock data analysis called mid-level data-fusion (score concatenation) of partial least squares-discriminant analysis (PLS-DA) to authenticate a high-quality beer (*Reale*) from an Italian craft brewery (*Birra del Borgo*) based on five different instrumental data: the thermogravimetric profile, mid-infrared, near-infrared, ultraviolet, and visible spectra. Dai et al. [11] combined NIR and Raman spectral data with low-, mid-, and high-level data fusions using various techniques, including PLS-DA and support vector machine (SVM), for the identification of four rice varieties from different origins, resulting in improved classification accuracy. Laxalde et al. [12] reported that the quantitative determination of saturate, aromatic, resin, and asphaltene (SARA) quantities in heavy oils could be improved by combining NIR and MIR spectroscopic data.

Different information can be obtained from the NIR spectra recorded by different detection modes, and there is a significant possibility that these data could be utilized simultaneously for prediction in multivariate prediction models. In this study, NIR spectra were recorded for both the transmittance and transflectance detection modes. Multiblock-PCA (MB-PCA) was performed to exploratorily analyze the data. In addition, the multiblock data were combined using various multiblock methods, including partial least squares (PLS), consensus-PLS (C-PLS), serial-PLS (S-PLS), and multiblock-PLS (MB-PLS). The developed methods were applied to quantify the concentrations of sugars in the mixture solutions, and two different NIR instruments were used for comparison.

2. Materials and methods

2.1 Preparation of sugar solution samples

The mixture samples consisted of four different sugars: glucose [$C_6H_{12}O_6$, 180.16 g/mol] (Ajax Finechem, Australia), fructose [$C_6H_{12}O_6$, 180.16 g/mol] (Ajax Finechem, Australia), maltose [$C_{12}H_{22}O_{11}\cdot H_2O$, 360.32 g/mol] (Ajax Finechem, Australia), and sucrose [$C_{12}H_{22}O_{11}$, 342.30 g/mol] (RANKEM, India). All chemicals were of analytical grade. To prepare stock solutions, the sugars were dissolved in deionized water at concentrations of 5.52, 5.55, 2.03, and 2.56 M for glucose, fructose, maltose, and sucrose, respectively. To create mixture solutions from the sugar samples, the mixture design method [13] was applied to generate 15 samples of sugar mixture solutions. The concentrations of the mixture samples used in the study are shown in Table 1.

Table 1 Code values and concentrations of mixture samples in mixture design used in this research

Sample No.	Code value of each sample				Conc. (M)
	Glucose	Fructose	Maltose	Sucrose	
1	1.00	0.00	0.00	0.00	5.52
2	0.00	1.00	0.00	0.00	5.55
3	0.00	0.00	1.00	0.00	2.03
4	0.00	0.00	0.00	1.00	2.56
5	0.50	0.50	0.00	0.00	5.54
6	0.50	0.00	0.50	0.00	3.78
7	0.50	0.00	0.00	0.50	4.04
8	0.00	0.50	0.50	0.00	3.79
9	0.00	0.50	0.00	0.50	4.06
10	0.00	0.00	0.50	0.50	2.30
11	0.33	0.33	0.33	0.00	4.32
12	0.33	0.33	0.00	0.33	4.50
13	0.33	0.00	0.33	0.33	3.34
14	0.00	0.33	0.33	0.33	3.35
15	0.25	0.25	0.25	0.25	3.92

2.2 NIR spectral detection

In this study, two different NIR spectrometers were used: a portable Vis-NIR spectrometer and a benchtop NIR spectrometer. The portable spectrometer captured NIR light using a Hamamatsu C10083CAH sensor (Hamamatsu Photonics, Japan). This optical sensor detected the light region between 360.9–1078.3 nm, providing 1,658 data points. An LS-1 tungsten halogen lamp (Ocean Optics, USA) was used as the light source and connected to the system using TP300-series fiber optics (Ocean Optics, USA). The second spectrometer used as a reference instrument was a Foss NIR System 6500 benchtop spectrometer (Multi-Mode™ Analyzer, Foss, USA). Spectral data were collected in the wavelength region of 400–2500 nm resulting in 1,050 data points. The samples were maintained at 25 °C for at least 6 hr before the analysis in a temperature-controlled room at 25 °C.

During the NIR spectroscopic analysis, each sugar mixture was poured into a 1-cm quartz cuvette cell. The samples were analyzed using the transmittance and transflectance detection methods. In the transflectance mode, a mirror and aluminum foil was used as the reflection plates for the portable and benchtop spectrometers, respectively. Figure 1(A) shows an instrumentation layout of the portable spectrometer, where the 1-cm quartz cuvette cell was placed in the center of a portable NIR box. The transmittance and transflectance detection modes are presented in Figure 1(B) and 1(C), respectively, where the white arrows indicate the direction of the NIR radiation passing through the cuvette cell. Additionally, the NIR background spectrum was measured using deionized water to prevent solvent interference.

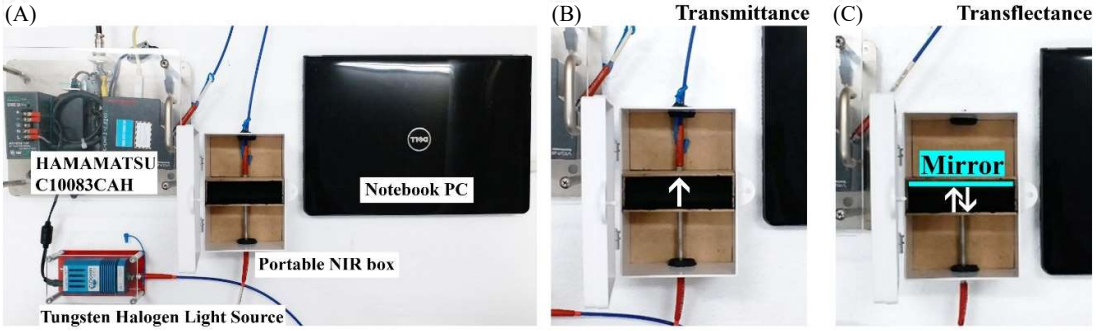

Figure 1 Instrumentation setup of the portable spectrometer (A). The white arrows indicated the directions of the NIR radiation passing through the cuvette cell for the transmittance (B) and transflectance (C) detection modes.

2.3 Chemometric analyses

2.3.1 Multiblock exploratory data analysis

Initially, the recorded NIR spectra were analyzed using PCA and MB-PCA. PCA is a well-known multivariate data exploratory method based on the nonlinear iterative partial least squares (NIPALS) algorithm that facilitates the exploration and visualization of data. PCA calculations were based on the literature [13] using the following equation:

$$X = T \cdot P + E, \tag{1}$$

where X denotes the data matrix. T is the data score, P is the data loading, and E is a residual matrix, which is the resulting PCA decomposition of the spectral data X.

MB-PCA is an extension of PCA proposed by Westerhuis et al. [14], allowing the investigation of the correlation of samples within the blocks using block scores (T_b) and block loadings (P_b). Additionally, super block (T), super score (t_T), and super weight (w_T) can be employed to examine global variation patterns [14-16]. In this study, PCA analysis of the NIR spectra was performed using both transmittance and transflectance NIR detection modes, defined as X_1 and X_2, respectively. The calculation of MB-PCA can be described as follows.

$$X_b = T_b \cdot P_b + E_b, \tag{2}$$

$$T = [T_1 \dots T_b], \tag{3}$$

$$t_T = T \cdot w_T. \tag{4}$$

2.3.2 Multiblock regression analysis

Regression methods aim to investigate the relationship between predictive data (X) and response data (y), wherein related information can be utilized to subjectively estimate the response data of unknown samples. PLS is among the most common multivariate regression techniques that combine the features of PCA and multiple regression methods. In the PLS model, the data matrix (X) is represented by scores (T) and data loading (P). Moreover, the response data (y) are represented by scores (T) and response loadings (q). However, it should be noted that T and P for PLS differ from T and P for PCA and that scores and loadings are generated uniquely for each compound in the dataset. The PLS calculation is as follows [13].

$$X = T \cdot P + E, \tag{5}$$

$$y = T \cdot q + f, \tag{6}$$

where q is a response loading and f is a residual of a response.

C-PLS is the simplest low-level data-fusion method [12]. The calculation of C-PLS is the same as that of traditional PLS. Data matrix X is a single augmented matrix of a concatenating data block $[X_1 \dots X_b]$. Thus, in this study, both transmittance (X_1), transflectance (X_2), and NIR spectra were concatenated into a single data block prior to applying the C-PLS regression technique for the prediction of total sugar concentrations.

S-PLS, a multiblock extension of PLS, is a mid-level data-fusion method [12]. It is a regression model based on the assumption that the algorithm performs a sequence of PLS calculations depending on the response residuals from the previous PLS model. Additionally, each PLS model for each data block can be utilized with a different number of latent variables. The calculation of S-PLS can be written as [12, 17, 18]

$$X_1 = T_1 \cdot P_1 + E_1, \tag{7}$$

$$X_2 = T_2 \cdot P_2 + E_2, \tag{8}$$

$$y = T_1 \cdot q_1 + T_2 \cdot q_2 + f. \tag{9}$$

MB-PLS is a mid-level data fusion method. Similar to MB-PCA, MB-PLS contains two levels: block level and super level. At the block level, block scores (T_b), block loadings (P_b), and block weights (w_b) were calculated and contained information on each data point. In contrast, at super level, the super score (t_T) and super weight (w_T) were calculated from the super block (T), which contains global information from all data blocks. The response (y) can be predicted using the MB-PLS correlation between the super score t_T and the latent variable of the response (q). The MB-PLS algorithm can be expressed as [14, 19]

$$X_1 = T_1 \cdot P_1 + E_1, \tag{10}$$

$$X_2 = T_2 \cdot P_2 + E_2, \tag{11}$$

$$y = \beta \cdot t_T + f, \tag{12}$$

where β is a regression coefficient

For the regression models of the sugar concentration analysis, the coefficient of determination for the calibration (R^2) was calculated to investigate the goodness-of-fit of the calibration models. In this research, root mean square error of calibration (RMSEC) method was used to determine the optimal number of latent variables for the regression models [13]. It should be noted that some other methods, such as cross validation, can also be used for the optimization task. However, in this case, the optimization process was based on auto-prediction to avoid the greater number of significant latent variables where the number of the studied samples was limited. The coefficient of determination for the prediction (Q^2) and root mean square error of cross validation (RMSECV) in the leave-one-out cross validation (LOOCV) were calculated to evaluate the predictive performance of the models. The ratio of prediction to deviation (RPD) and relative standard deviation (RSD) values were used to evaluate the predictive quality of different models by standardizing the predictive accuracy [20]. All model evaluations and statistical analyses were performed using in-house MATLAB scripts (MATLAB R2014a, The MathWorks Inc., Natick).

3. Results and discussion

3.1 Comparison of transmittance and transflectance modes

Figure 2(A) and 2(B) show the transmittance and transflectance spectra, respectively, recorded using a portable Vis-NIR spectrometer. The absorption bands in the region of 800–1100 nm are in the shortwave NIR region related to the 3rd overtones and combinations of the common IR absorptions. The absorption peaks from both detection methods were slightly different because of the different integration time of the light sensor; 100 ms and 20 ms for the transmittance and transflectance modes, respectively. The longer integration time generally results in the higher absorption light intensity. The absorption peaks at 400 nm could be related to the light-yellow color

of the solution samples. The NIR region around 970 nm represents the absorption of water in the samples and shows an opposite direction of the curve. This was that, in this study, the NIR background spectrum was measured using deionized water. The presenting NIR spectra were obtained by subtracting the background from the NIR spectra of the sugar mixtures. Since there was less water in the solution samples due to the sugar concentrations, the water absorption peaks in the solution samples were lower [21]. As a result, the water absorption peaks at 970 nm in Figure 2 have negative absorbance values for the portable detection.

The transmittance and transflectance NIR spectra of the benchtop spectrometer are shown in Figure 2(C) and 2(d), respectively. The transmittance spectra in Figure 2(C) show major absorption bands at 1450 nm: 2nd overtones of -CH, -$CONH_2$, -OH, and water; 1700–1800 nm: 1st overtones of -CH_3, -CH_2, and -CH; 1900–2000 nm:1st overtones of -$CONH_2$, -COOH, -COOR, -OH, and water; 2100 nm: combination bands of -$CONH_2$ and -OH; and 2250–2450 nm: combination bands of -CH_3, -CH_2, and -CH. In contrast, in Figure 2(D), the transflectance spectra show major absorption bands at 400–500, 750–800, and 900–1000 nm: 3rd overtones of -CH_2, -CH, -OH, and water; and 1150–1250 nm: 2nd overtones of -CH_3, -CH_2, and -CH. In addition, the transflectance spectra contain a relatively high absorbance, resulting in nearly flat spectra and low informative bands at 1400–2500 nm, which could be due to the double path length of the transflectance detection mode.

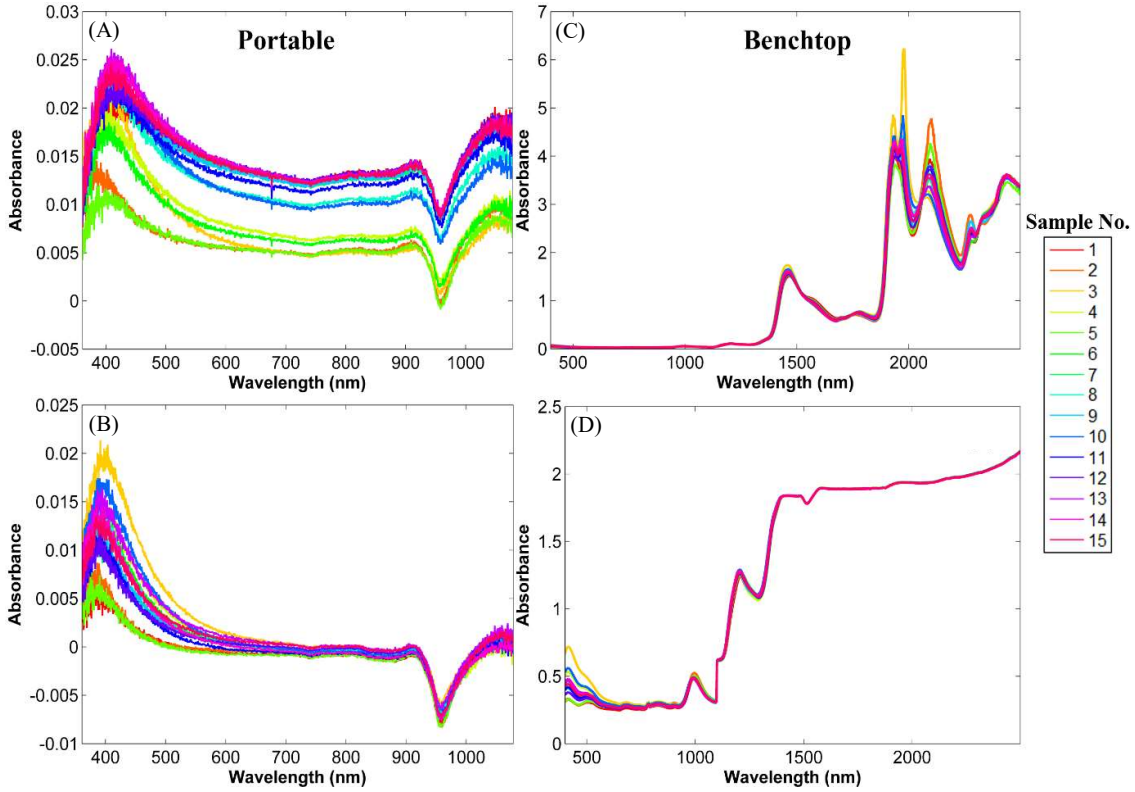

Figure 2 NIR spectra from portable (left) and benchtop (right) spectrometers: (A) and (C) = transmittance mode, (B) and (D) = transflectance mode.

3.2 Chemometric analysis

3.2.1 Exploratory data analysis

3.2.1.1 NIR detection using portable spectrometer

The variables obtained using the MB-PCA method can be categorized into two levels: the block level and the global (super) level. At block level, block scores (T_b) and block loadings (P_b) were calculated for each data block. Meanwhile, at super level, super score (t_T) and super weight (w_T) were obtained from super block (T), which contains the information of all data blocks. Additionally, block scores (T_b) from each data block can be used to visualize block score plots, while the super score (t_T) can be used to visualize a global score plot. Consequently,

the advantage of MB-PCA is that it allows not only the correlation of samples within blocks to be explored but also the correlation of samples with global variation patterns. Figure 3(A) and 3(B) show the PCA score plots of the NIR spectra from the transmittance and transflectance modes, respectively, using a portable spectrometer. The PCA score plots successfully differentiated mixtures containing different sugar concentrations. Figure 3(C) shows the MB-PCA global score plot, in which the result is the same as that of the transmittance mode. The samples were labeled from yellow to red according to their sugar concentrations from low to high (Figure 3(D)). The samples with lower concentrations are displayed at the top, whereas those with higher concentrations are displayed at the bottom of the MB-PCA plane.

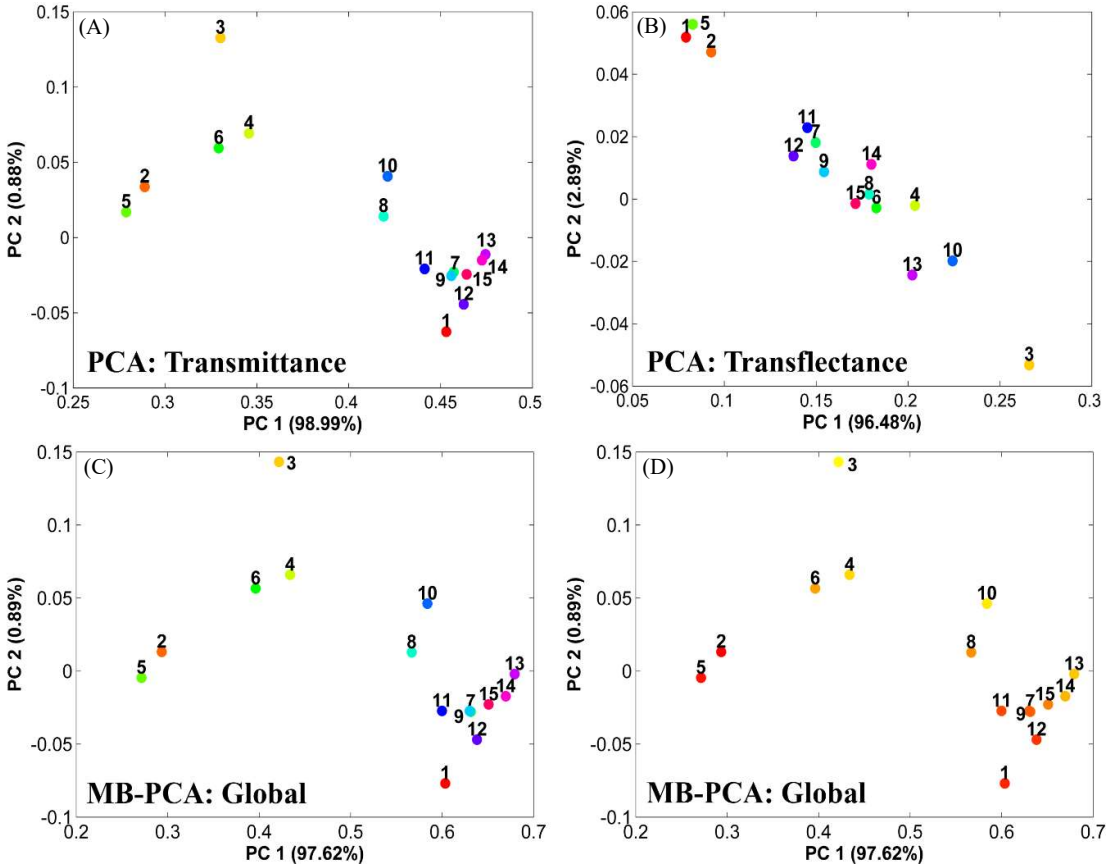

Figure 3 PCA and MB-PCA score plots from portable Vis-NIR spectrometer: (A) PCA of transmittance, (B) PCA of transflectance, and (C-D) MB-PCA global score plots.

3.2.1.2 NIR detection using benchtop spectrometer

Figure 4(A) and 4(B) illustrate the PCA score plot of the NIR spectra using the transmittance and transflectance modes from the benchtop spectrometer. Samples 1 to 4 contained pure sugars, as listed in Table 1. Therefore, the samples have different chemical characteristics, and the scores are mostly located outside the edges of the main clusters. The MB-PCA block score plot in Figure 4(C) is not markedly different from the PCA score plots of the transmittance mode in Figure 4(A).

Furthermore, the samples containing mixtures of the two types of sugars are positioned approximately between the two pure sugar samples. For instance, sample no. 5, which contains both glucose and fructose, is located approximately between samples no. 1 and 2, which contain glucose and fructose, respectively. The position of sample no. 6, which contains both glucose and maltose, is approximately between that of samples no. 1 and 3, which contain glucose and maltose, respectively. Similarly, sample no.15, which consists of all four types of sugar samples, is located approximately in the center of the MB-PCA space. Figure 4(D) shows that the samples with lower sugar concentrations are located at the bottom, and those with higher sugar concentrations are located at the top of the MB-PCA plane.

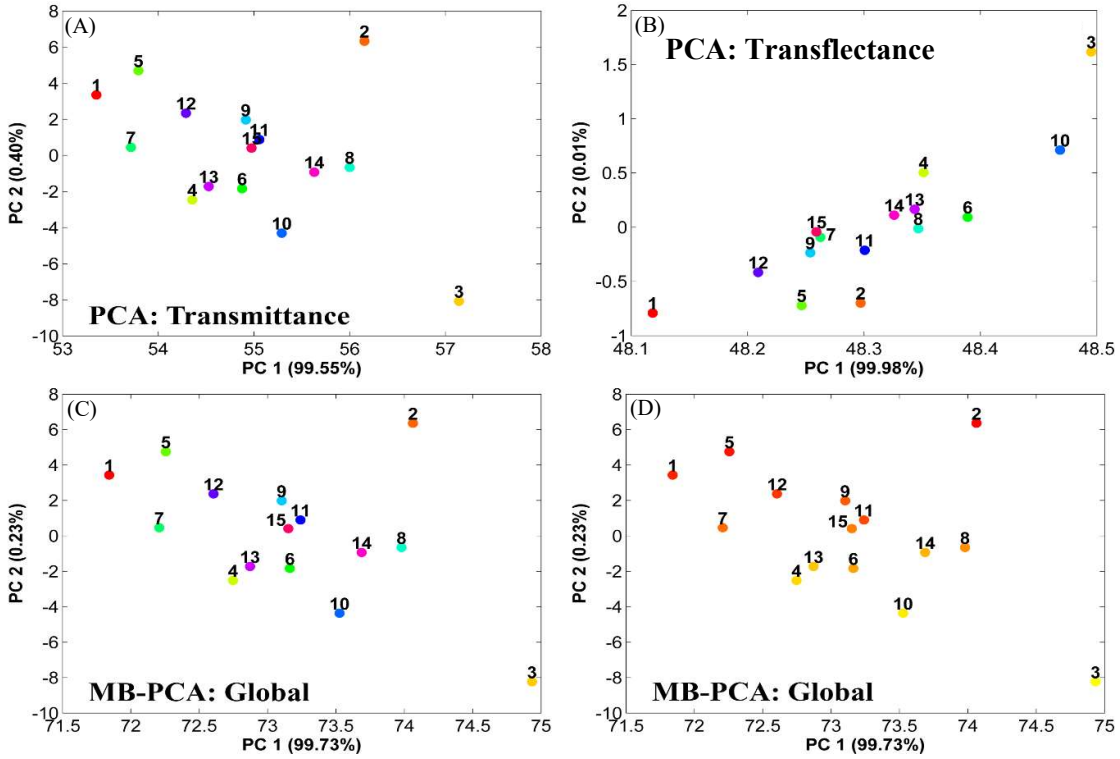

Figure 4 PCA and MB-PCA score plots from benchtop NIR spectrometer: (A) PCA of transmittance, (B) PCA of transflectance, and (C-D) MB-PCA global score plots.

3.2.2 Regression analysis

3.2.2.1 Predictions using traditional PLS

The NIR spectra of the mixed samples were used to establish prediction models to estimate sugar concentrations. Ideally, various data preprocessing methods, such as normalization, standardization, standard normal variates (SNV), and multiplicative scatter correction (MSC), could be applied to the NIR spectra. In this research, these methods have been tested where the optimal results were obtained from the use of SNV followed by mean centering. Therefore, the discussion was based on using SNV with mean centering to demonstrate the comparison of the calibration models. The statistical model results of the generated regression models are summarized in Table 2. The results show that, from the portable spectra, the traditional PLS model using transflectance spectra outperforms the PLS model using transmittance spectra in terms of predictive performance, with a high R^2 and Q^2 of 1.00 and 0.96, low RMSEC and RMSECV of 0.01 and 0.22, and higher RPD and lower RSD of 5.03 and 5.63, respectively. This could be due to the double path length of the transflectance detection mode, which produces more NIR-relevant information. Whereas, using the benchtop spectra, the PLS model using transmittance spectra outperforms the PLS model using transflectance spectra with an R^2 and Q^2 of 1.00 and 0.99, low RMSEC and RMSECV of 0.04 and 0.09, and higher RPD and lower RSD of 12.28 and 2.30, respectively. The results can be explained by the fact that the transmittance spectra are highlighted in the NIR region (over 1400 nm) using SNV and mean centering. In contrast, the transflectance spectra are highlighted in the visible-NIR range (below 1400 nm). Therefore, the transmittance spectra contain more of the expected chemical information. It was noted here that the dominant wavelengths in the visible region of the portable spectrometer could be attributable to the fact that the Hamamatsu C10083CAH sensor was designed for light detection in the visible region. The absorbance after 700 nm was optional but practical for the detection in the user manual.

3.2.2.2 Predictions using multiblock data analysis

Several multiblock regression models, including C-PLS, S-PLS, and MB-PLS, were employed to determine whether the predictive performance could be enhanced. The C-PLS, modeled using portable spectra, produced better results than the traditional PLS model using transmittance spectra. Using the benchtop spectra, the statistical results show that the C-PLS model performed better than the traditional PLS based on the use of the transflectance spectra but performed worse based on the transmittance spectra.

Using the S-PLS models, the statistical model results show that changing the block order has no significant impact on the predictive performance for both portable and benchtop spectra. The predictive abilities of the S-PLS model from the portable spectra were slightly lower than those of the C-PLS model. In contrast, the S-PLS model of the benchtop spectra outperformed both C-PLS and traditional PLS models. The statistical results indicated that the MB-PLS models, developed from both portable and benchtop NIR spectra, provided the most accurate predictive ability, offering the lowest RMSECV of 0.06.

The percentage of explained variance in *y* was first calculated to select the appropriate number of LVs for each MB-PLS model. The explained variation in the y data was determined using the cumulative product of the super score (t_T) and the response weight (q). For instance, Figure 5(A) shows that using MB-PLS of the portable spectra with six LVs, the model predicts and explains most of the variance in the response y. Additionally, compared with other multiblock regression methods, this could be one of the benefits of the MB-PLS method because it improves predictability while simultaneously preventing model overfitting by applying the right number of LVs.

Table 2 Regression model statistic values

Regression models		No. of LVs	Model statistics					
			R^2	Q^2	RMSEC	RMSECV	RPD	RSD
Portable spectra								
PLS	Transmittance (TM)	5	1.00	0.95	0.02	0.25	4.42	6.40
	Transflectance (TF)	5	1.00	0.96	0.01	0.22	5.03	5.63
C-PLS		5	1.00	0.95	0.02	0.24	4.61	6.14
S-PLS	TM + TF	5 + 1	1.00	0.94	0.02	0.27	4.09	6.91
	TF + TM	1 + 5	1.00	0.94	0.02	0.27	4.09	6.91
MB-PLS		6	1.00	0.96	0.02	0.22	5.03	5.63
Benchtop spectra								
PLS	Transmittance (TM)	4	1.00	0.99	0.04	0.09	12.28	2.30
	Transflectance (TF)	5	1.00	0.99	0.03	0.11	10.05	2.82
C-PLS		4	1.00	0.99	0.05	0.10	11.06	2.56
S-PLS	TM + TF	5 + 1	1.00	1.00	0.03	0.07	15.79	1.79
	TF + TM	1 + 5	1.00	1.00	0.04	0.07	15.79	1.79
MB-PLS		5	1.00	1.00	0.05	0.06	18.43	1.54

3.2.2.3 Interpretation of MB-PLS models

Another advantage of the MB-PLS model is that the percentages of the block importance values can be calculated. Block importance is the MB-PLS super weight (w_T) utilized for each data block and can be used to explain the weight of each data block applied for the MB-PLS calculation. Figure 5(D) explains the percentage of the block importance of the MB-PLS model from the benchtop NIR spectra. The results indicate that in the first LV, the MB-PLS weighs the relevance of the transmittance spectra (93 %) greater than that of the transflectance spectra (7 %), yielding a percentage variance of y of 92 %. The explained variances in y were thereafter increased from two to five LVs to 100 % by weighing the transmittance spectra over the transflectance spectra. In this case, it was noticeable that the prediction using the transmittance spectra could be more important and resulted in a lower RMSECV value.

The result, shown in Figure 5(B), reveals that, in the first LV of the portable spectra, the %block importance values of both detection modes on the first LV were relatively the same, 51% and 49%, respectively, for the transmittance and transflectance detections, yielding a percentage variance of y of 94 %. The comparatively same amount of the %block importance in the most significant LV (LV no. 1) posed the comparative level of the variations in both detection methods. As a result, in this experiment, the transflectance spectra somehow resulted in a slightly lower RMSECV value.

Additionally, Figure 6 and 7 show the MB-PLS super-loading of both the transmittance and transflectance spectra. MB-PLS super-loading can be used to explain the behavior of each variable (wavelength) in the MB-PLS model. Figure 6 shows the super-loading of each LVs from the portable Vis-NIR spectra. The results show that at an LV of 1, both transmittance and transflectance spectra are highlighted in the NIR region rather than in the visible region because the super-loading values are greater than zero. At an LV of 2, the transmittance spectra are highlighted in the visible region and the Vis-NIR region. In contrast, in the transflectance spectra, the spectra are highlighted in the NIR region. At an LV of 3, both the transmittance and transflectance spectra are highlighted in the visible and NIR regions. However, the spectra primarily contain noise at an LV larger than 4. Figure 7 shows the super-loading of each LV from the benchtop NIR spectra. The results indicate that for every LV utilized, the transmittance spectra are emphasized in the region above 1400 nm, while the transflectance spectra are emphasized in the region below 1400 nm. Therefore, the percentage of block importance and the range of super loadings used are correlated for both the portable and benchtop MB-PLS models, with a wider range of super loadings related to the higher percentage of block importance.

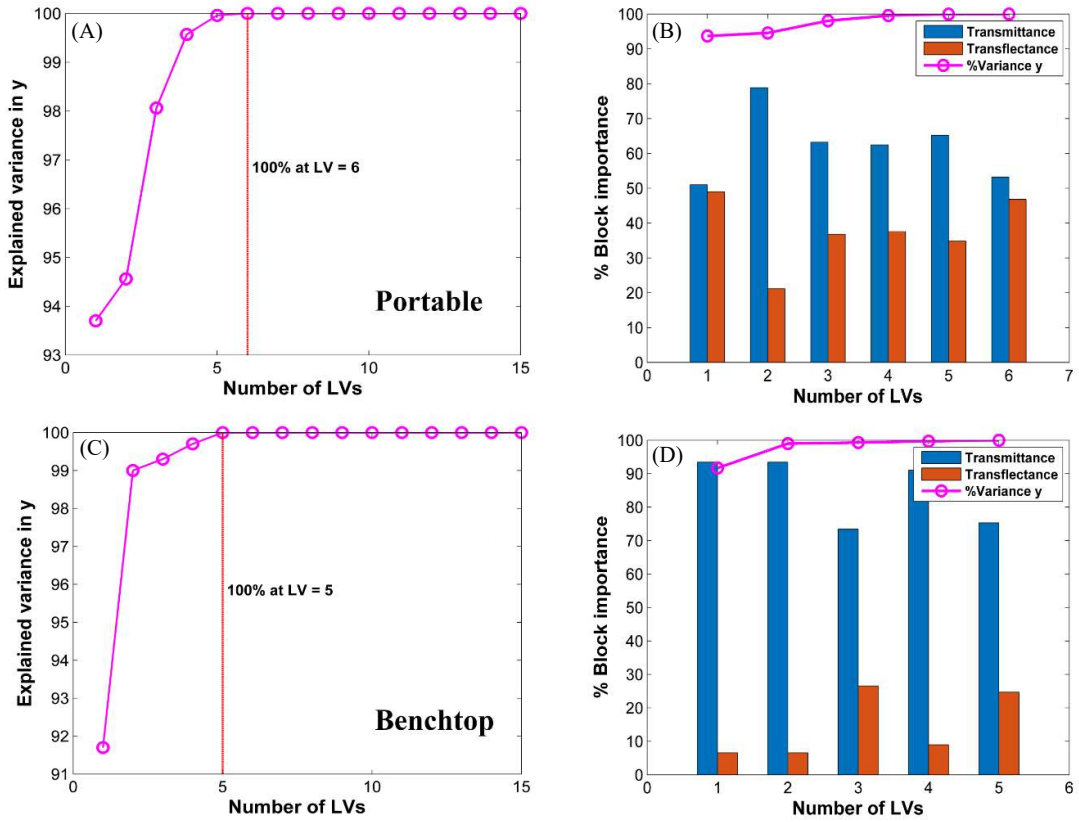

Figure 5 MB-PLS explained variance in *y* (left) and percentage of block importance (right) from (A-B) portable and (C-D) benchtop NIR spectra.

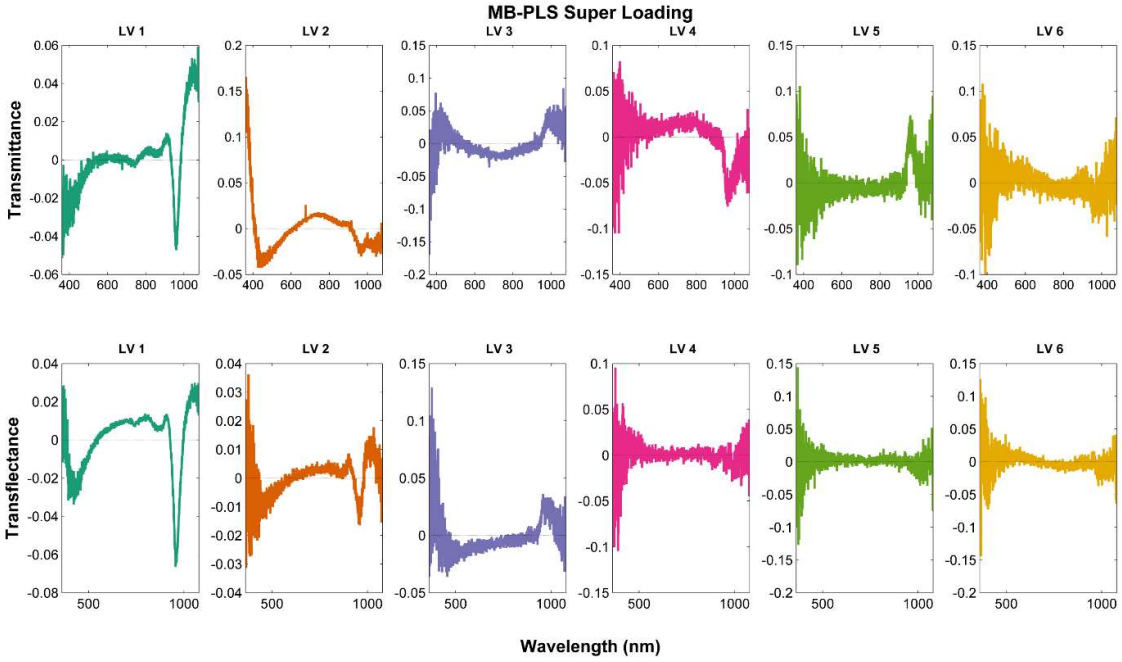

Figure 6 MB-PLS super-loading plots of portable Vis-NIR spectra.

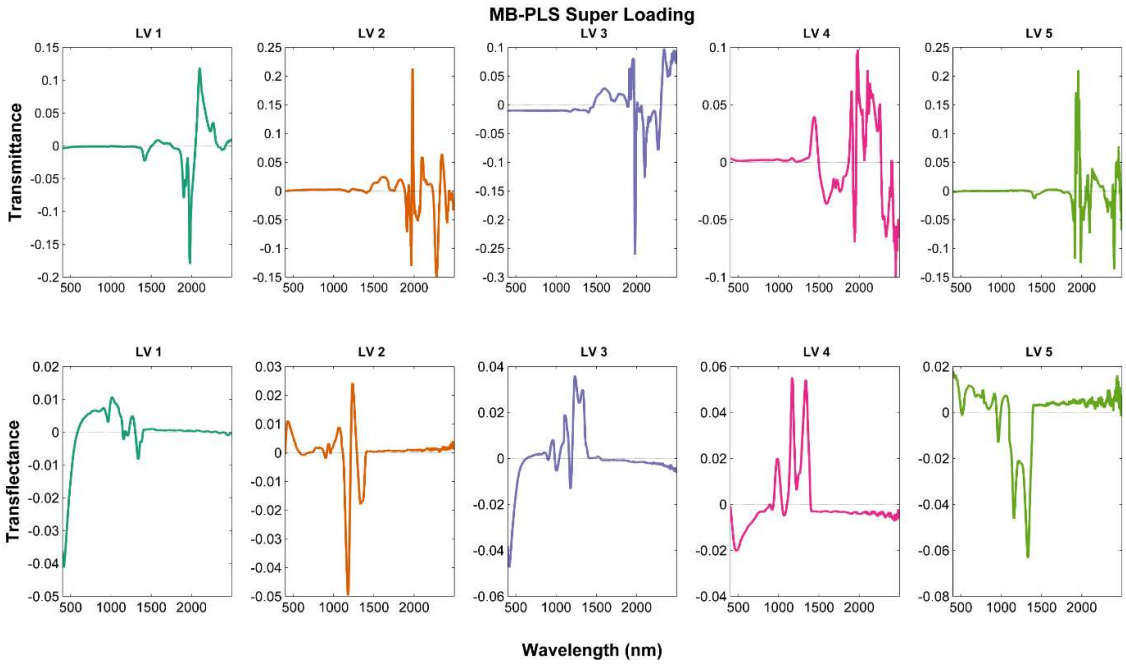

Figure 7 MB-PLS super-loading plots of the benchtop NIR spectra.

4. Conclusion

In this study, NIR spectra were obtained for both the transmittance and transflectance detection modes. The transflectance spectra from the spectrometers exhibit higher absorption peaks than the transmittance spectra owing to the double path length of the transflectance detection mode. The PCA and MB-PCA score plots were able to differentiate mixture samples containing varying sugar contents. In regression methods, when compared with other regression models, the MB-PLS algorithm could enhance the ability to predict total sugar concentrations. Furthermore, information on block importance and explained variance in responses can be obtained. The analysis data supported the selection of the optimal detection mode in NIR analysis.

5. Acknowledgement

This project is funded by National Research Council of Thailand (NRCT) and Chiang Mai University: N42A650304. Postharvest Technology Innovation Centre, Office of the Higher Education Commission, Bangkok, Thailand is also gratefully acknowledged.

6. References

[1] Ozaki Y, Morisawa Y. Principles and characteristics of NIR spectroscopy. In: Ozaki Y, Huck C, Tsuchikawa S, Engelsen SB, editors. Near-Infrared Spectroscopy. Singapore: Springer; 2021. p. 11-35.

[2] Agelet LE, Hurburgh CR. A tutorial on near infrared spectroscopy and its calibration. Crit Rev Anal Chem. 2010;40(4):246-260.

[3] Alamprese C, Casale M, Sinelli N, Lanteri S, Casiraghi E. Detection of minced beef adulteration with turkey meat by UV-vis, NIR and MIR spectroscopy. LWT - Food Sci Technol. 2013;53(1):225-232.

[4] Borràs E, Ferré J, Boqué R, Mestres M, Aceña L, Busto O. Data fusion methodologies for food and beverage authentication and quality assessment - A review. Anal Chim Acta. 2015;891:1-14.

[5] Osborne BG. Near-infrared spectroscopy in food analysis. Encyclopedia of Analytical Chemistry. Chichester: John Wiley & Sons; 2006.

[6] Pasquini C, Hespanhol MC, Cruz KAML, Pereira AF. Monitoring the quality of ethanol-based hand sanitizers by low-cost near-infrared spectroscopy. Microchem J. 2020;159:105421.

[7] Pasquini C. Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. J Braz Chem Soc. 2003;14(2):198-219.

[8] Tao L, Via B, Wu Y, Xiao W, Liu X. NIR and MIR spectral data fusion for rapid detection of Lonicera japonica and Artemisia annua by liquid extraction process. Vib Spectrosc. 2019;102:31-38.

[9] Smolinska A, Engel J, Szymanska E, Buydens L, Blanchet L. Chapter 3 - General framing of low-, mid-, and high-level data fusion with examples in the life sciences. In: Cocchi M, editor. Data Handling in Science and Technology, vol. 31. Amsterdam: Elsevier; 2019, p. 51-79.
[10] Biancolillo A, Bucci R, Magrì AL, Magrì AD, Marini F. Data-fusion for multiplatform characterization of an italian craft beer aimed at its authentication. Anal Chim Acta. 2014;820:23-31.
[11] Dai Y, Dai Z, Guo G, Wang B. Nondestructive identification of rice varieties by the data fusion of Raman and Near-Infrared (NIR) spectroscopies. Anal Lett. 2023;56(5):730-743.
[12] Laxalde J, Caillol N, Wahl F, Ruckebusch C, Duponchel L. Combining near and mid infrared spectroscopy for heavy oil characterisation. Fuel 2014;133:310-316.
[13] Brereton RG. Chemometrics: Data analysis for the laboratory and chemical plant. Chichester: John Wiley & Sons; 2003.
[14] Westerhuis JA, Kourti T, MacGregor JF. Analysis of multiblock and hierarchical PCA and PLS models. J Chemom. 1998;12:301-321.
[15] Hassani S, Hanafi M, Qannari EM, Kohler A. Deflation strategies for multi-block principal component analysis revisited. Chemometr Intell Lab Syst. 2013;120:154-168.
[16] Qin SJ, Valle S, Piovoso MJ. On unifying multiblock analysis with application to decentralized process monitoring. J Chemom. 2001;15(9):715-742.
[17] Felício CC, Brás LP, Lopes JA, Cabrita L, Menezes JC. Comparison of PLS algorithms in gasoline and gas oil parameter monitoring with MIR and NIR. Chemometr Intell Lab Syst. 2005;78(1-2):74-80.
[18] Gaydou V, Kister J, Dupuy N. Evaluation of multiblock NIR/MIR PLS predictive models to detect adulteration of diesel/biodiesel blends by vegetal oil. Chemometr Intell Lab Syst. 2011;106(2):190-197.
[19] Campos MP, Sousa R, Pereira AC, Reis MS. Advanced predictive methods for wine age prediction: Part II – A comparison study of multiblock regression approaches. Talanta 2017;171:132-142.
[20] Wongsaipun S, Krongchai C, Jakmunee J, Kittiwachana S. Rice grain freshness measurement using rapid visco analyzer and chemometrics. Food Anal Methods. 2018;11:613-623.
[21] Giangiacomo R. Study of water–sugar interactions at increasing sugar concentration by NIR spectroscopy. Food Chem. 2006;96(3):371-379.