

# Integration of Independent Component Analysis with Near Infrared Spectroscopy for Rapid Quantification of Sugar Content in Wax Jambu (*Syzygium samarangense* Merrill & Perry)

YUNG-KUN CHUANG<sup>1,3</sup>, SUMING CHEN<sup>1,2\*</sup>, Y. MARTIN LO<sup>3</sup>, CHAO-YIN TSAI<sup>2</sup>,  
I-CHANG YANG<sup>3</sup>, YU-LIANG CHEN<sup>1</sup>, PEI-JU PAN<sup>1</sup> AND CHIEN-CHIH CHEN<sup>1</sup>

<sup>1</sup>. Department of Bio-Industrial Mechatronics Engineering, National Taiwan University, Taipei, Taiwan, R.O.C.

<sup>2</sup>. Bioenergy Research Center, National Taiwan University, Taipei, Taiwan, R.O.C.

<sup>3</sup>. Department of Nutrition and Food Science, University of Maryland, College Park, MD, U.S.A.

(Received: December 13, 2011; Accepted: July 25, 2012)

## ABSTRACT

Independent component analysis (ICA) was integrated with near infrared (NIR) spectroscopy for rapid quantification of sugar content in wax jambu (*Syzygium samarangense* Merrill & Perry). The JADE algorithm (Joint Approximate Diagonalization of Eigenmatrices) and linear regression with spectral pretreatments were incorporated to analyze the NIR spectra of wax jambu against sucrose solutions. Unlike other multivariate approaches, ICA enabled comprehensive quantification of sugar content in wax jambu. In the present study, ICA was applied as the sole tool to build the NIR calibration model of internal quality of intact wax jambu without any other multivariate analysis methods. The best spectral calibration model of wax jambu (600 to 700 nm and 900 to 1,098 nm) yielded  $r_c = 0.988$ ,  $SEC = 0.243$  °Brix,  $r_v = 0.971$ ,  $SEV = 0.381$  °Brix, and  $RPD = 4.15$  using the normalized first derivative spectra and 9 independent components (ICs). All ICA results were better than those of partial least squares regression (PLSR). Thus, ICA can quickly identify and effectively quantify the sugar contents in wax jambu with calibration models achieving high predictability.

Key words: near infrared (NIR) spectroscopy, calibration model, wax jambu, sugar content, independent component analysis (ICA)

## INTRODUCTION

Near infrared (NIR) spectroscopy, a nondestructive inspection method based on specific absorptions within a given range of wavelength corresponding to the constituents in the sample<sup>(1)</sup>, has been widely applied for the evaluation of internal quality of agricultural products<sup>(2,3)</sup>. Since NIR spectra of a mixture is the linear summation of individual spectra of the constituents in the mixture, such a mixture spectra thus can be regarded as 'blind sources'<sup>(4)</sup> as the proportion of constituents in the samples remains unknown. Many attempts have been made in recent years to extract critical features from the spectra using multivariate analysis<sup>(5,6)</sup>, including multiple linear regression (MLR)<sup>(7)</sup>, principal component regression (PCR)<sup>(8)</sup>, and partial least squares regression (PLSR)<sup>(9)</sup>. However, these methods were not designed for resolving the 'blind source' problem and may not correlate well with

the properties of constituents in the mixture, consequently hindering the applicability of the spectra for chemometric analysis of the constituents<sup>(10-12)</sup>.

A multiuse statistical approach originally used to implement 'blind source separation' in signal processing<sup>(13,14)</sup>, independent component analysis (ICA) is capable of disassembling the mixture signals of Gaussian distribution into non-Gaussian independent constituents with only a little loss of information and does not require any information to be added to the source<sup>(15)</sup>. In practice, multiple ICA algorithms have been developed, including JADE algorithm (Joint Approximate Diagonalization of Eigenmatrices)<sup>(16,17)</sup> and FastICA algorithm<sup>(18,19)</sup>, making ICA a high-speed and reliable tool<sup>(20)</sup> for analytical chemistry<sup>(21,10)</sup>, biomedical signal processing, telecommunications, econometrics, audio processing, and image processing<sup>(4)</sup>.

Application of ICA for spectrum analysis has been demonstrated by Chen and Wang<sup>(11)</sup> to separate the pure spectra of various constituents from the NIR spectra of the

\* Author for correspondence. Tel: +886-2-3366-5350;  
Fax: +886-2-2362-7620; E-mail: schen@ntu.edu.tw

mixture and to build relationship between the estimated independent components and the constituents. Such a capability also enabled complete explanation of the constituents' properties for NIR qualitative analyses<sup>(22)</sup>. In addition, ICA was used to obtain statistically independent and chemically interpretable latent variables (LVs) in multivariate regression<sup>(23)</sup>. It was also noted that the number of independent components extracted from the spectra of mixtures is related to the performance of ICA<sup>(24)</sup>. Moreover, ICA was employed to identify the infrared spectrum of mixtures containing two pure materials<sup>(25)</sup> as well as the constituents in commercial gasoline<sup>(26,27)</sup>. Equally noteworthy is that the calibration model built through MLR, after using ICA to extract independent components of aqueous solutions, gave good predictability<sup>(12)</sup>, whereas NIR estimation of glucose concentration<sup>(10)</sup> was enhanced by using ICA.

While application of ICA for spectral analysis appeared promising, available literatures still focused mainly on (1) identification of constituents in the mixture, especially for chemical compounds<sup>(10-12,25-28)</sup>; (2) a preprocessing method for improving predictability of calibration model<sup>(29)</sup>; and (3) combination of ICA and other multivariate analysis methods, such as PCA-ICA<sup>(26)</sup>, ICA-MLR<sup>(12,30)</sup>, ICA-PLS<sup>(30)</sup>, ICA-LS-SVM<sup>(31)</sup> and ICA-NNR<sup>(32)</sup> to deal with linear or nonlinear problems. However, no literature exists for ICA with NIR spectroscopy to be applied as the sole tool to quantify internal quality of intact fruit without any other multivariate analysis methods. Wax jambu (*Syzygium samarangense* Merrill & Perry), an endemic fruit in Taiwan and parts of southeast Asia (Figure 1) has very unique surface and texture that are easily bruised or damaged, hence requiring wax jambu to be handled delicately from harvest to shipping and distribution. To date, several researches aimed to develop a non-invasive and rapid detection method for the analysis of internal quality of wax jambu<sup>(33-35)</sup>. For further applications of ICA as the sole tool with NIR spectroscopy in the inspection of fruits, wax jambu is suitable to serve as sample for discussion. In the present study, ICA was integrated for NIR spectral analysis to quantify the sugar content in intact wax jambu. The results of wax jambu were also compared with those of sucrose solutions. Spectral pretreatments and linear regression were then

used to build spectral calibration models of sugar content. The analysis results of ICA were also compared with those of PLSR to assess the abilities in predicting sugar content in wax jambu.

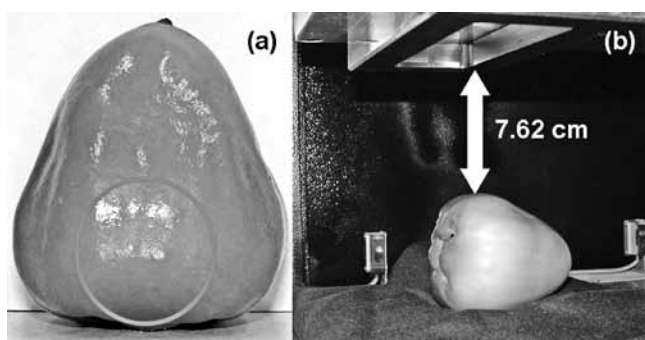
## MATERIALS AND METHODS

### I. Sample Preparation

Sucrose ( $C_{12}H_{22}O_{11}$ , FW: 342.30) powder was solubilized in de-ionized water to prepare 78 sucrose solutions with sugar content ranging from 0.4 to 19.0 °Brix. The average sugar content was 9.83 °Brix, and the standard deviation was 5.48 °Brix. A total of 114 wax jambu (*S. samarangense* Merrill & Perry) samples purchased from Fangliao, Pingtung County in Taiwan were employed for the study. Before measuring the sugar content, wax jambu was first pressed to extract 15 mL juice which was centrifuged for 15 min at 2,500 rpm and 22°C to clarify the sample in centrifuge KUBOTA 2700 (KUBOTA Corporation Co., Ltd., Osaka City, Osaka, Japan).

### II. NIR Spectra and Sugar Content Measurement

A NIRS 6500 spectrophotometer and sample transport (FOSS NIRSystems, Laurel, MD, USA) with quartz cuvette were used to measure the transmittance spectra of sucrose solutions. The wavelength ranged from 400 to 2,498 nm with 2 nm intervals. The quartz cuvette (light path: 1 mm; external dimensions: length = 3.0 cm, width = 0.2 cm, and height = 3.5 cm) was filled with sucrose solution for transmittance measurements. An on-line NIRS 6500 spectrophotometer (FOSS NIRSystems, Laurel, MD, USA) was used to measure the reflectance spectra of the wax jambu samples. The wavelength range was from 400 to 2,498 nm and the interval was 2 nm. As shown in Figure 1(a), the wax jambu was examined to find an area with no defects, which was then selected as the location for reflectance measurements. The wax jambu was placed horizontally in line with the spectrophotometer in a dark compartment, and the distance between the light source and the top of sample was adjusted to the suggested value of 7.62 cm, as shown in Figure 1(b). The spectrophotometer was controlled by a personal computer to perform NIR acquisition and spectrum editing. All spectral data were recorded as the logarithm of reciprocal of reflectance ( $\log 1/R$ ), and NIR spectrum of each sample was the average of 32 scans. A digital refractometer (PR-101, ATAGO Co., Ltd., Itabashi-ku, Tokyo, Japan) was used to measure the sugar content as the reference values. The index "°Brix" used for PR-101 refractometer is a parameter that denotes the total amount of soluble solids in the sample. For fruits such as wax jambu, most of soluble solids in the juice are sugars, mainly sucrose, fructose and glucose. Therefore, the value of °Brix measured from wax jambu can be regarded as the total sugar content.



**Figure 1.** A wax jambu (*Syzygium samarangense* Merrill & Perry) sample (a) side view and the NIR measurement location, and (b) sample placement with suggested distance 7.62 cm between the light source and the top of sample in the on-line NIRS 6500 spectrophotometer.

### III. Data Analysis

### (I) Independent Component Analysis (ICA)

Independent component analysis (ICA) is a method used to transform the observed multivariate data to statistically independent components (ICs) and to present them as a linear combination of observation variables. The number of receptors defined by ICA algorithm must be more than or equal to the number of sources, and the signals emitted by the sources are in non-Gaussian distribution<sup>(20)</sup>. The ICs are latent variables; therefore, they cannot be directly observed, indicating that the mixing matrix is also unknown. The purpose of the ICA algorithm is to determine the mixing matrix (**M**) or the separating matrix (**W**). In order to predict the unknown source, it is assumed that  $\mathbf{W} = \mathbf{M}^{-1}$ ,

$$\hat{\mathbf{s}} = \mathbf{W}\mathbf{x} = \mathbf{M}^{-1}\mathbf{M}\mathbf{s} \quad (1)$$

where  $\hat{\mathbf{s}}$  is the estimation of the sources (**s**) and **x** represents the observed spectra of the objects.

In the present study JADE (joint approximate diagonalization of eigenmatrices) algorithm<sup>(16,17)</sup> was employed to conduct ICA analysis. In general, JADE offers rapid performance for dealing with spectra data due to it works off-the-shelf, an improvement over other multivariate approaches like PCR and PLSR. Assuming that the spectra obtained through measurement of the unknown mixtures were the linear combination of various components' spectra, it can be expressed as:

$$\mathbf{A} = \mathbf{M}\mathbf{I} \quad (2)$$

The spectra of samples were all linearly composed of *m* ICs. Matrix  $\mathbf{A}_{l \times n}$  stands for *l* samples containing *n* values;  $\mathbf{I}_{m \times n}$  stands for the matrix of ICs, including *m* independent components.  $\mathbf{M}_{l \times m}$  stands for the mixing matrix, which is related to the component concentration in the mixture. The linear relationship between the mixing matrix (**M**) and the component concentration (**C**) can be expressed as:

$$\mathbf{C} = \mathbf{M}\mathbf{B} \quad (3)$$

Among them, **B** referred to the matrix of regression coefficient. In doing so, the concentration of each component in the mixture could be determined by the combination of ICA and linear regression.

### (II) Partial Least Squares Regression (PLSR)

Partial least squares regression (PLSR), a typical method in chemometrics<sup>(9)</sup>, has been widely applied to chemical and engineering fields. When PLSR is applied to spectral analysis, the spectra can be regarded as the composition of several principal components (PCs), and be expressed as a 'factor' in the PLSR algorithm. The factors' sequence is determined by their influences; the more important factor is ranked earlier in the order, such as factor 1 and factor 2. Since information from spectral bands was used in PLSR analysis, the analysis

results can be improved by selecting appropriate number of factors and specific wavelength ranges. To avoid overfitting of the PLSR model's results with too many factors, the factors were selected based on the following principles in this study: (1) A maximum factor limit was set at 1/10 of calibration set data + 2 to 3 factors; (2) new factors were not added if they caused a rise in the prediction error; and (3) new factors were not added if they resulted in a standard error of validation (SEV) smaller than the standard error of calibration (SEC).

### (III) Spectral Pretreatments

The purpose of spectral pretreatments was to eliminate the spectral variation, which was not caused by chemical information contained in the samples<sup>(36)</sup>. For the raw NIR spectra of sucrose solutions and wax jambu, three different spectral pretreatments were employed in this study: (1) normalization; (2) first derivative with normalization; and (3) second derivative with normalization. Normalization scaled the spectrum absorbance of all samples to fall within an interval of -1 to 1. For further applications of ICA in fast on-line inspection of fruits, the procedure of selecting best pretreatment parameters, including points of smoothing and gap of derivative, were not employed to save computational time. The gap of derivative was set at a minimal value of 2, so as to maintain the most wavelength values as inputs for the model.

### (IV) Model Establishment

This study used the mathematic software MATLAB (The MathWorks, Inc., Natick, MA, USA) to write ICA programs based on toolbox of JADE algorithm (available at [http://perso.telecom-paristech.fr/~cardoso/RR\\_ECG/jadeR.m](http://perso.telecom-paristech.fr/~cardoso/RR_ECG/jadeR.m)) for establishing ICA spectral calibration models. The results of ICA were compared with the spectral calibration models of PLSR built by WinISI II (Infrasoft International, LLC., Port Matilda, PA, USA) chemometric software package. The analysis procedure of both ICA and PLSR for wax jambu and sucrose solution samples included: (1) selecting calibration set and validation set, (2) spectral pretreatments, and (3) determining best calibration model. Since the sucrose solutions were mixtures of sucrose powder and water, their composition were rather simple. Therefore, the data of full wavelength range (400 to 2,498 nm) were used for comparing the tolerance abilities of ICA and PLSR since spectral bands with more noises (e.g. 2,200 to 2,498 nm) often affect the analysis results. Identification of specific wavelength ranges was needed for wax jambu because their composition was more complicated than that in sucrose solutions, which required additional correlation analysis between wavelengths and sugar content. All of the sucrose solutions and wax jambu samples were respectively used for analysis to assess the tolerance abilities of ICA and PLSR. A ratio of calibration to validation samples of 2:1 was adopted according to the sugar content in the sample. All samples were ranked ascendantly according to their sugar content. Number 1 and 2 were

assigned for calibration and 3 for validation, with subsequent numbers following the same alternating sequence. The same sets of calibration and validation were used for both ICA and PLSR analyses.

After the respective spectral calibration models of sucrose solution and wax jambu were built, these models were then used to predict the sugar contents of the calibration and the validation set. The evaluation of predictability was based on the following statistical parameters, including coefficient of correlation of calibration set ( $r_c$ ), standard error of calibration (SEC), coefficient of correlation of validation set ( $r_v$ ), standard error of validation (SEV), bias, and ratio of [standard error of] performance to [standard] deviation (RPD), as defined by:

$$SEC = \left[ \frac{1}{n_c} \sum_{i=1}^{n_c} (Y_r - Y_c)_i^2 \right]^{1/2} \quad (4)$$

$$SEV = \left\{ \frac{1}{n_v} \sum_{i=1}^{n_v} [(Y_r - Y_v) - \text{Bias}]_i^2 \right\}^{1/2} \quad (5)$$

$$\text{Bias} = \frac{1}{n_v} \sum_{i=1}^{n_v} (Y_r - Y_v)_i \quad (6)$$

$$RPD = \frac{SD}{SEV} \quad (7)$$

where  $Y_c$  and  $Y_v$  represent the estimated sugar contents of the calibration set and the validation set, respectively.  $Y_r$  is the reference sugar content,  $n_c$  and  $n_v$  are the number of samples in the calibration set and validation set, and  $SD$  is the standard deviation of sugar content within the validation set. RPD is one of the indices used to evaluate the performance of a model. An RPD value greater than 3 is considered adequate for analytical purposes in most of NIR spectroscopy applications for agricultural products<sup>(37)</sup>.

## RESULTS AND DISCUSSION

### I. Sucrose Solution

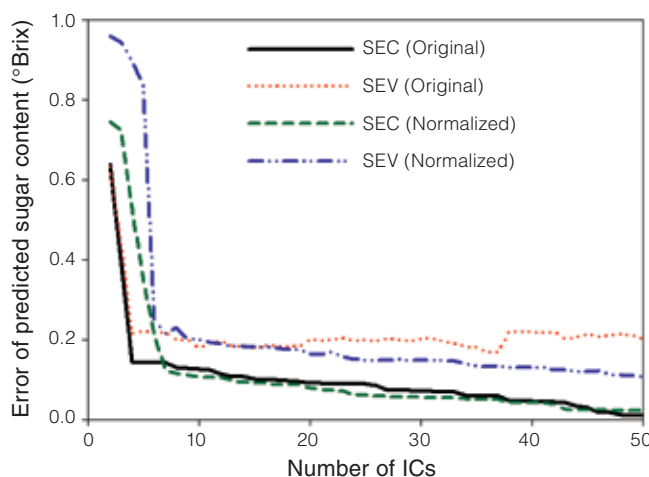
The 78 sucrose solution samples were divided into 52 calibration samples and 26 validation samples with a ratio of 2 : 1. The distribution of their sugar content (°Brix) is shown in Table 1. For all the samples within the calibration and validation sets, the difference between maximum values of two sets was 0.2 °Brix; the differences for other items including minimum, average, standard deviation, and coefficient of variation (CV), were all smaller than 0.5 °Brix. The above sets of samples were conforming to the consistent requirement of sugar content distributions.

#### (I) Selection of the Most Appropriate Number of ICs

According to the definition of ICA, the observed

**Table 1.** Summary of sucrose solutions and sample sugar contents. Total samples ( $n = 78$ ), calibration set ( $n = 52$ ) and validation set ( $n = 26$ ) were arranged to have consistent distributions of sugar content

Group	n	Sucrose Solutions				
		Sugar Content (°Brix)				
		Max.	Min.	Mean	SD	CV
Total Samples	78	19.00	0.40	9.83	5.48	0.56
Calibration Set	52	19.00	0.40	9.72	5.52	0.57
Validation Set	26	18.80	0.90	10.06	5.52	0.55



**Figure 2.** Relationship between the numbers of ICs and errors of the predicted sugar content for sucrose solutions. The most appropriate number of ICs for normalized spectra was determined by the tendency of SEC (green-short dash line) and SEV (blue-dash dot dot line) values.

receptor signals can be decomposed at most into a number of ICs (independent components) equal to the number of samples<sup>(20)</sup>. This study used the data of full range of wavelength (400 to 2,498 nm) as the inputs of ICA, conducted ICA for the original spectra of 52 calibration samples of sucrose solution by selecting 1 to 52 ICs, and observed the prediction error by using the calibration model. Both situations with and without normalization were examined. When only one IC applied, the prediction error was high, so the results were only shown by applying 2 to 50 ICs. As shown in Figure 2, when the number of ICs increased to 4, SEC of the case without normalization sharply decreased to 0.14 °Brix, and SEV fell to 0.21 °Brix, indicating that different numbers of ICs can influence the predictability of the spectral calibration model. However, application of more ICs did not necessarily help improve the ability of the calibration model, hence only the initial 4 ICs were applied in the calibration model.

The results of ICA with normalized spectra can be observed in Figure 2. The prediction error greatly reduced as the number of ICs increased to 7; the SEC and SEV with 7 ICs were 0.12 and 0.22 °Brix, respectively. Normalization apparently gave less variations of SEV compared with that of original spectra.

(II) Spectra Decomposition and Correlation Analysis of Sugar Content

Based on ICA analysis it is critical to examine whether these 7 ICs were statistically independent. To illustrate the operation, IC 1 and 4 were selected and their correlation was shown in Figure 3, with the coefficient of determination ( $r^2$ ) being only  $4.0 \times 10^{-8}$ . This indicated that IC 1 and 4 were independent of each other. Diagrams of every two ICs among the 7 ICs also showed a similar distribution to that in Figure 3, with all of the  $r^2$  smaller than 0.243, conforming to the mutually independent characteristics of ICs<sup>(20)</sup>.

Equation (5) shows that the constituent information 'sugar content' should mainly correspond to a specific IC, and there should be a high correlation between the values of the IC in the mixing matrix and the sugar content. So a diagram was made with the reference sugar content and the values of each column (each IC) in the mixing matrix. As shown in Figure 4, the correlation coefficient ( $r$ ) between IC 1 and the reference sugar content could reach 0.977, which meant that with 7 ICs extracted, the IC 1 among all 7 ICs could reveal the most information resulted from the sugar content in the spectra. The results were in agreement with Westad<sup>(24)</sup>. Therefore, selection of the numbers of ICs is important since it influences how the information is used after spectra decomposition.

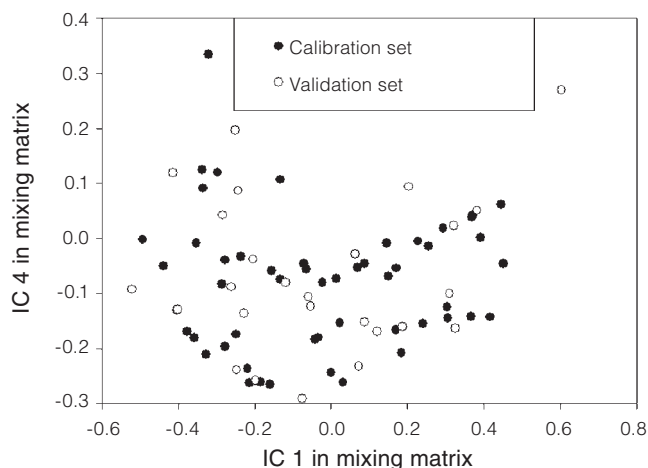
The regression coefficient matrix by the NIR spectra and the reference sugar content of calibration sets was shown in Table 2, and the values from the top to the bottom referred to IC 1 to 7. All values were compared in terms of absolute values. It was found that the value of the first row (IC 1) was the largest, closely followed by the value of IC 4. The results agreed with the order of correlation between each IC and the reference sugar content, and indicated that the importance of each IC was independent of the IC sequence. Each major constituent had its corresponding IC decomposed by ICA, in which IC contribution was clearly defined, so that

all constituents of the mixtures could be distinguished by ICA<sup>(11,25-27,12)</sup>.

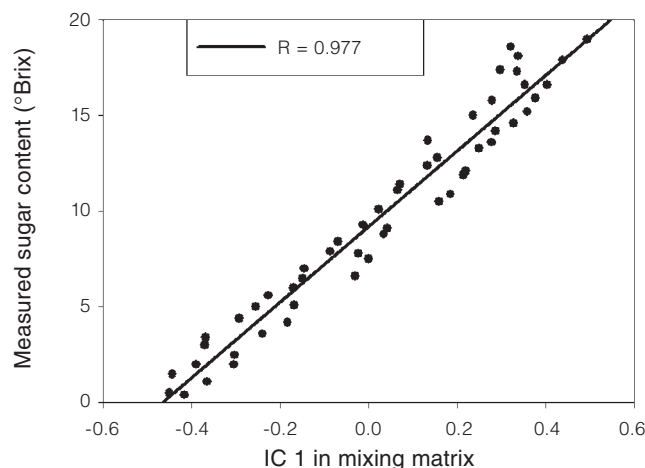
The ICs, decomposed from the spectra by ICA, reflected the spectral characteristics of the unknown mixture and constituted the pure materials' spectra of this mixture under an ideal state<sup>(11,25-27)</sup>. Since the sucrose solutions were mixtures of sucrose and water, and the spectra was comprised of both constituents, the ICs decomposed by ICA should reflect the characteristics of these two pure substances. For the original spectra of the normalized calibration set, among the 7 ICs applied for ICA, the order of the 7 ICs, according to the correlation with reference sugar content, was IC 1, 4, 2, 5, 3, 7, and 6. The NIR original spectra of the calibration set and IC 1 were shown in Figure 5(a) and (b). The peak positions of IC 1 (964, 1,090, 1,436, 2,100, and 2,276 nm) matched the specific wavelength ranges of sugar content (C-H band)<sup>(7,38,25)</sup>. So IC 1 can be considered to respond mainly to the sugar content, conforming to the above results. The other ICs had poor correlation with reference sugar content, and the absolute values in the regression coefficient matrix were much smaller than that of IC 1, so they exerted an assisting function.

**Table 2.** Regression coefficient matrix of sucrose solutions with 7 ICs were extracted from the NIR spectra of calibration sets. Correlation between the absolute value of each IC in regression coefficient matrix and sugar content was examined.

IC #	Regression Coefficient
1	-2.1811
2	-0.2843
3	-0.1843
4	1.2976
5	0.1876
6	-0.1334
7	-0.1416



**Figure 3.** Distribution of calibration and validation samples of sucrose solutions in IC 1-IC 4 space. IC 1 and IC 4 were randomly selected from the 7 ICs.



**Figure 4.** Correlation between the values of IC 1 in the mixing matrix and the reference sugar contents of sucrose solutions.

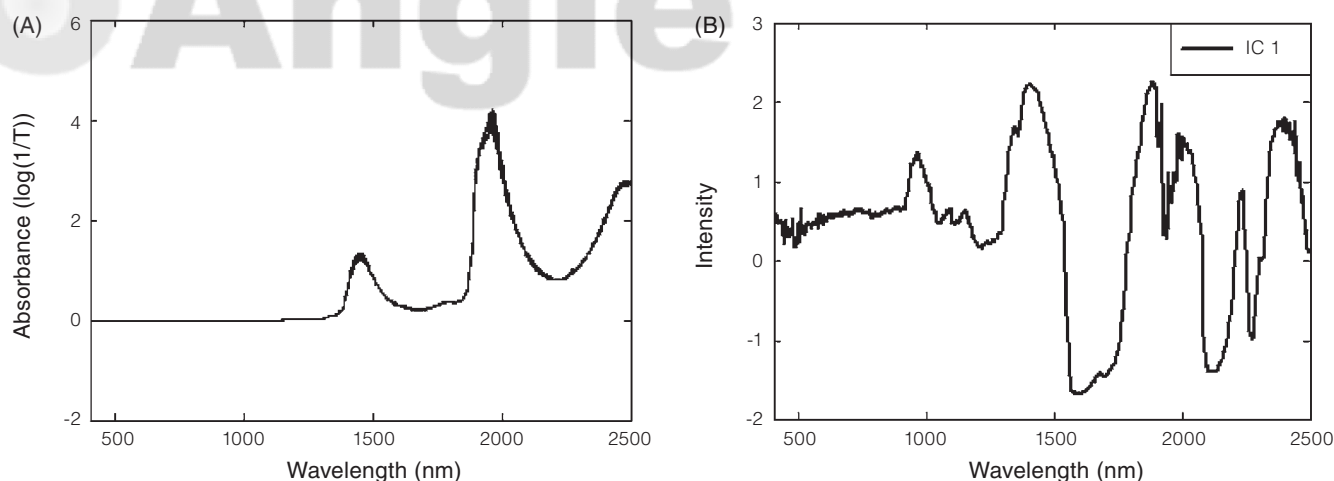


Figure 5. Original NIR spectra of sucrose solutions and IC 1 decomposed from calibration sets.

Table 3. Regression results by ICA and PLSR analyses for sucrose solutions

Method	Spectrum	ICs / Factors	Calibration Set (52)			Validation Set (26)				
			Mean: 9.715, SD: 5.515			Mean: 10.058, SD: 5.515				
			$r_c$	SEC (°Brix)	RSEC (%)	$r_v$	SEV (°Brix)	RSEV (%)	bias (°Brix)	RPD
ICA	Original	4	0.9997	0.144	6.97	0.9995	0.215	3.57	0.045	25.69
	Original + Normalization	7	0.9998	0.124	4.01	0.9993	0.216	3.68	0.014	25.54
	1 <sup>st</sup> Derivative + Normalization	4	0.9994	0.193	13.71	0.9984	0.331	10.34	0.028	16.66
	2 <sup>nd</sup> Derivative + Normalization	5	0.9983	0.321	19.66	0.9973	0.409	16.20	-0.014	13.48
PLSR	Original	2	0.9995	0.181	11.41	0.9985	0.300	8.78	0.069	18.38
	Original + Normalization	4	0.9990	0.218	11.59	0.9975	0.399	8.68	0.022	13.82
	1 <sup>st</sup> Derivative + Normalization	3	0.9995	0.192	11.50	0.9950	0.546	12.92	0.031	10.10
	2 <sup>nd</sup> Derivative + Normalization	2	0.9990	0.243	20.96	0.9869	0.899	34.99	0.013	6.14

(III) Sugar Content Quantification Based on ICA and PLSR

Quantitative analyses of sugar content in sucrose solutions were conducted by ICA and PLSR using the full range of wavelength from 400 to 2,498 nm. The results of spectral calibration models built by ICA are shown in Table 3. It was found that the best spectral calibration model was the original spectra normalized, with 7 ICs applied. The results were  $r_c = 0.9998$ ,  $SEC = 0.124$  °Brix,  $r_v = 0.9993$ ,  $SEV = 0.216$  °Brix,  $bias = 0.014$  °Brix, and  $RPD = 25.54$ . A comparison was made in light of the result of the original spectra with and without normalization, and it was found that the calibration model yielded similar outcomes in the validation sets, whereas the SEC value was improved when normalization was applied. Although derivatives can improve baseline shift of the original spectra and amplify the signal characteristics, noise interference may also be enhanced at the same time, making it unsuitable for spectral bands with much noises. The spectrum in the range of 2,200 to 2,498 nm contained more noises; therefore, the predictability of the spectral calibration models would decrease as derivatives were attempted.

The results of spectral calibration models built by PLSR indicated that the best spectral calibration model was acquired when the original spectra and 2 factors were employed, and the results were as follows:  $r_c = 0.9995$ ,  $SEC = 0.181$  °Brix,  $r_v = 0.9985$ ,  $SEV = 0.300$  °Brix,  $bias = 0.069$  °Brix, and  $RPD = 18.38$  (Table 3). Moreover, with the  $SEC = 0.192$  °Brix and the  $SEV = 0.546$  °Brix for the first derivative with normalization, and the  $SEC = 0.243$  °Brix and the  $SEV = 0.899$  °Brix for the second derivative with normalization, it is apparent that the SEV values of both first and second derivatives were many times higher than SEC. The results showed that the PLSR spectral calibration models had poor predictability when applied to validation sets.

Comparing the quantitative analysis results of ICA and PLSR, all ICA spectral calibration models had better ability than PLSR in predicting calibration and validation sets. This means that ICA extracts the characteristic information from the spectra more effectively, not only improving the expository ability of calibration models for the calibration sets, but also increasing the tolerance for the validation sets. Results also showed that ICA was preferable to PLSR due to much

lower bias (Table 3). This finding became more obvious with normalization, indicating that ICA had a better tolerance to the influences caused by factors other than chemical characteristics of the constituents in the samples, which helped to build more robust spectral calibration models. In summary for the sucrose solutions, ICA achieved better quantitative analysis of sugar content than PLSR did, while selecting a suitable number of ICs and spectral pretreatments could help improve the predictability of spectral calibration models. The results of sucrose solutions also helped establish proper procedures with useful information applicable when conducting ICA analysis of wax jambu.

## II. Wax Jambu

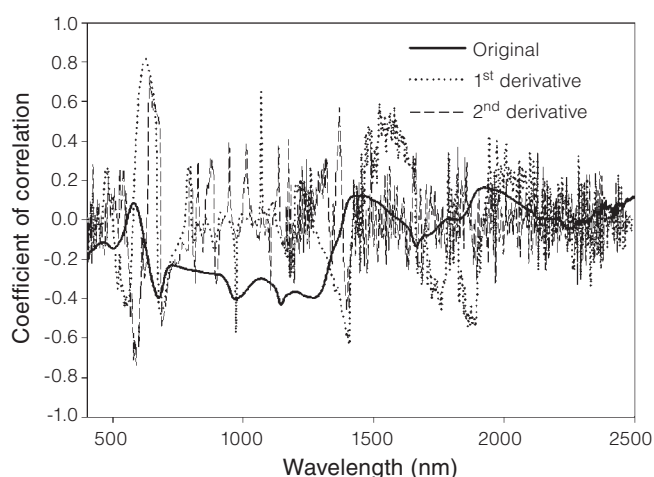
Wax jambu samples totaling 114 were used; their sugar contents ranged from 6.4 to 14.5 °Brix. The average sugar content was 9.92 °Brix with the standard deviation of 1.61 °Brix. All the samples were divided in a 2 : 1 ratio into 76 and 38 calibration and validation samples (Table 4).

### (I) Correlation Analysis of NIR Spectra and Sugar Content

Figure 6 showed the distribution of the correlation

**Table 4.** Summary of wax jambu (*Syzygium samarangense* Merrill & Perry) and sample sugar contents. Total samples (n = 114), calibration set (n = 76) and validation set (n = 38) were arranged to have consistent distributions of sugar content

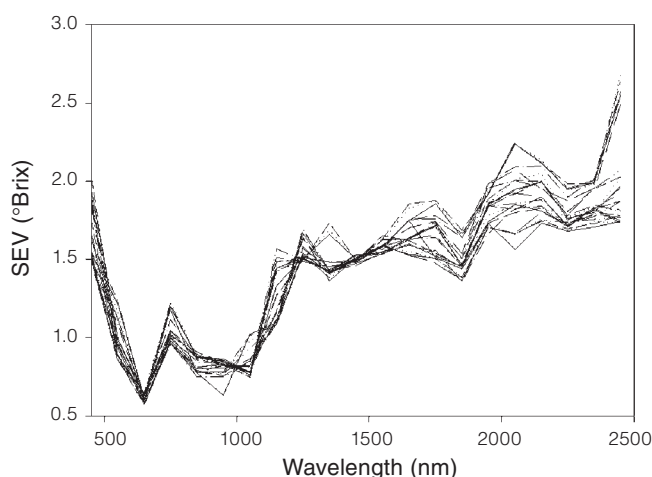
Group	n	Wax Jambu				
		Sugar Content (°Brix)				
		Max.	Min.	Mean	SD	CV
Total Samples	114	14.50	6.40	9.92	1.61	0.16
Calibration Set	76	14.50	6.40	9.89	1.61	0.16
Validation Set	38	14.00	7.10	9.99	1.62	0.16



**Figure 6.** Correlation coefficient distributions of the spectra and the sugar content of wax jambu through three different spectral pretreatments (original spectra, first derivative spectra and second derivative spectra).

coefficients for the original, the first derivative and the second derivative spectra of the wax jambu samples and their sugar contents. The main absorption wavelengths of the original spectra were 676, 968, and 1,144 nm, of which 676 nm was located within the visible region of red light, whereas 968 and 1,144 nm in the NIR region, belonging to the 2<sup>nd</sup> overtone of the C-H bond. The main absorption wavelengths of the first derivative spectra were 626, 974, 1,070, and 1,406 nm, of which 626 nm was located in the visible region of orange light, with the correlation up to 0.808, while the remaining wavelengths in the NIR region. The main absorption wavelengths of the second derivative spectra were located in the visible region between orange light and red light, namely 594, 642, and 692 nm. Figure 6 showed that the wavelength range of 600 to 1,098 nm was the major absorption band, and the first derivative spectra were most significantly correlated to the sugar content<sup>(35)</sup>. As for the spectral band 650 to 700 nm, which belonged to the absorption band of red light, it was consistent with the color of wax jambu skin, indicating that color information was also reflected in the spectrum.

The NIR spectra of wax jambu samples were analyzed by taking every 100 nm as a band region, and full spectrum range from 400 to 2,498 nm was divided into 21 band regions, in which they were separately analyzed. Analysis of the 76 wax jambu calibration samples could have been decomposed into 76 ICs; however, applying too many ICs could easily lead to overfitting of the model. Hence, in this study ICA was conducted with the limit of 30 ICs. The SEV showed no obvious trend when applying 1 to 6 ICs, and was greatly influenced by water (O-H bond). There is a high proportion of water in the wax jambu samples, so it was necessary to avoid using the spectral bands of 1,450 and 1,900 nm that represent primarily water absorption. When applying 7 to 30 ICs (Figure 7), the SEV values in the ranges of 600 to 700 nm and 800 to 1,098 nm were less than 1 °Brix, so were the results of the first and the second derivative spectra. All three spectra



**Figure 7.** Relationship between spectral bands and errors of the predicted sugar content for wax jambu when applying 7 to 30 ICs. Full spectrum range from 400 to 2498 nm was divided into 21 band regions by taking every 100 nm as a band region.

fitted the spectral bands of higher correlation in Figure 6, so the specific wavelength regions for spectrum analyses of wax jambu were selected from the wavelength range of 600 to 700 nm and 800 to 1,098 nm<sup>(35)</sup>.

(II) Sugar Content Quantification Based on ICA and PLSR

1. Analysis without Spectral Pretreatment

The ICA results of the spectral calibration model for wax jambu are shown in Table 5. The best spectral calibration model was found with the normalized first derivative spectra and 10 ICs, resulting in  $r_c = 0.956$ ,  $SEC = 0.471$  °Brix,  $r_v = 0.954$ ,  $SEV = 0.489$  °Brix,  $bias = -0.013$  °Brix,  $RPD = 3.32$ . Among the 10 ICs applied for ICA, the order of the initial 4 ICs, according to the correlation with reference sugar content, is IC 3, 7, 8, and 6, with respective correlation coefficient ( $r$ ) of -0.805, 0.647, -0.612, and 0.279. IC 3, 7, and 8 can be considered to respond mainly to the information of sugar content (including fructose, glucose and sucrose)<sup>(39,40)</sup> as the composition of wax jambu is rather complicated than that of sucrose solution alone. Since the specific wavelengths used were within the wavelength range of 600 to 700 nm and 800 to 1,098 nm, the spectra covered the 3<sup>rd</sup> overtone of C-H bond, conforming to the results of Figure 6 and 7. Additionally, the spectral calibration models built after normalization used the characteristic information of 10 ICs, which is in line with the SEV trend observed in Figure 7. Moreover, the small values of bias indicated that ICA had good tolerance to the influence caused by factors other than the internal chemical composition of the samples.

The PLSR results of the spectral calibration model are shown in Table 5, with the best spectral calibration model found in the normalized original spectra with 5 factors, yielding  $r_c = 0.884$ ,  $SEC = 0.753$  °Brix,  $r_v = 0.867$ ,  $SEV = 0.816$  °Brix,  $bias = 0.238$  °Brix, and  $RPD = 1.99$ . The specific wavelength regions used were within the wavelength range of 600 to 700 nm and 800 to 1,098 nm, consistent with the aforementioned results.

After comparing the results of ICA and PLSR quantitative analysis, it was found that the ICA calibration model performed better than PLSR, since not only did it enhance the predictability of the model but it also reduced the bias. The specific wavelengths used in ICA and PLSR showed a high degree of coincidence. When applied to wax jambu samples, the correlation analysis between NIR spectra and sugar content provided a basis to select the appropriate specific wavelength regions.

2. Analysis with Spectral Pretreatment

To evaluate the best predictability of ICA models for wax jambu, ICA analysis was further performed with pretreatment and outlier procedures. After selecting the best pretreatment parameters (points of smoothing and gap of derivative were both 3) and eliminating 1/10 outliers (11 samples) from the total of 114 samples, the best spectral calibration model was found, as shown in Table 6, with the normalized first derivative spectra and 9 ICs, resulting in  $r_c = 0.988$ ,  $SEC = 0.243$  °Brix,  $r_v = 0.971$ ,  $SEV = 0.381$  °Brix,  $bias = 0.001$  °Brix,  $RPD = 4.15$ . The PLSR analysis results under the same conditions were  $r_c = 0.983$ ,  $SEC = 0.287$  °Brix,  $r_v = 0.963$ ,

Table 5. Regression results by ICA and PLSR analyses for wax jambu (without spectral pretreatment)

Method	Spectrum	Wavelength Range (nm) <sup>a</sup>	ICs / Factors	Calibration Set (76)			Validation Set (38)				
				Mean: 9.891, SD: 1.610			Mean: 9.990, SD: 1.624				
				$r_c$	SEC (°Brix)	RSEC (%)	$r_v$	SEV (°Brix)	RSEV (%)	bias (°Brix)	RPD
ICA	Original	600-700, 2 1000-1098, 2	7	0.930	0.591	6.37	0.919	0.642	7.11	-0.024	2.53
	Original + Normalization	600-700, 2 800-1000, 2	10	0.948	0.515	5.51	0.940	0.568	5.73	0.054	2.86
	1 <sup>st</sup> Derivative + Normalization	600-700, 2 900-1098, 2	10	0.956	0.471	4.88	0.954	0.489	4.98	-0.013	3.32
	2 <sup>nd</sup> Derivative + Normalization	600-700, 2 800-1000, 2	10	0.943	0.535	5.85	0.944	0.553	5.77	-0.017	2.94
PLSR	Original	800-1098, 2	2	0.428	1.455	15.56	0.237	1.599	15.80	0.351	1.02
	Original + Normalization	600-700, 2 800-1098, 2	5	0.884	0.753	8.11	0.867	0.816	8.90	0.238	1.99
	1 <sup>st</sup> Derivative + Normalization	600-700, 2 800-1098, 2	3	0.876	0.777	8.44	0.880	0.782	8.89	-0.020	2.08
	2 <sup>nd</sup> Derivative + Normalization	600-700, 2 800-1000, 2	2	0.797	0.973	11.58	0.803	0.975	10.95	0.182	1.67

<sup>a</sup> Interval is 2 nm.



**Table 6.** Regression results by ICA and PLSR analyses for wax jambu (with spectral pretreatment)

Method	Spectrum	Wavelength Range (nm) <sup>a</sup>	Smoothing Points / Gap	ICs / Factors	Calibration Set (68)			Validation Set (35)				
					Mean: 9.953, SD: 1.556			Mean: 9.953, SD: 1.580				
					r <sub>c</sub>	SEC (°Brix)	RSEC (%)	r <sub>v</sub>	SEV (°Brix)	RSEV (%)	bias (°Brix)	RPD
ICA	1 <sup>st</sup> Derivative + Normalization	600-700, 2 900-1098, 2	3, 3	9	0.988	0.243	2.46	0.971	0.381	3.91	0.001	4.15
PLSR	1 <sup>st</sup> Derivative + Normalization	600-700, 2 800-1098, 2	4, 4	8	0.983	0.287	2.81	0.963	0.426	4.41	-0.039	3.71

<sup>a</sup> Interval is 2 nm.

SEV = 0.426 °Brix, bias = -0.039 °Brix, RPD = 3.71. The ICA spectral calibration model had better results than PLSR results with pretreatment and outlier procedures in predicting calibration and validation sets.

Compared to the previous literatures<sup>(33-35)</sup>, the spectral calibration models built by ICA had higher predictability for wax jambu since the SEC values reported by You (2002), Chung *et al.* (2004) and Lin (2002) were 0.413 °Brix, 0.388 °Brix and 0.252 °Brix, respectively. Among them, the SEP values reported by Chung *et al.* (2004), 0.262 °Brix, 0.207 °Brix and 0.322 °Brix, were all lower than its SEC value (0.388 °Brix); these MLR analysis results seemed unreasonable because that the prediction sets were unknown to the calibration model, thus the SEP values should be higher than SEC value. Even though, our ICA results listed in Table 6 were better than those reported by Chung *et al.* (2004) and Lin (2002) in terms of r<sub>c</sub>, SEC, r<sub>p</sub> and RPD.

The results of ICA sugar content quantification based on NIR spectroscopy showed that ICA can effectively extract the characteristic information in the spectra, and build the spectral calibration models with desirable abilities to evaluate the concentration of the constituents. It thus can be expected that integration of ICA with NIR spectroscopy could become a powerful tool for quantitative analysis of specific targets.

## CONCLUSIONS

Independent component analysis (ICA) was applied as the sole tool to integrate with NIR spectroscopy for rapid quantification of sugar content in sucrose solutions and wax jambu. ICA gave a comprehensive approach to characterize the NIR spectra with respect to the sugar content in wax jambu and sucrose solutions that other multivariate analysis methods cannot deal with. The spectral calibration models built by ICA had high predictability for both wax jambu and sucrose solutions. Compared to PLSR, ICA can identify the sugar features in the spectra of wax jambu and then evaluate their concentrations more effectively. Therefore, it offers a reliable tool for quantitative analysis of sugar content in wax jambu by NIR spectroscopy. ICA in conjunction with NIR spectroscopy also has a potential to be applied to identify multiple constituents and evaluate their concentrations of agricultural products.

## REFERENCES

- McClure, W. F. 2003. 204 years of near infrared technology: 1800-2003. *J. Near Infrared Spec.* 11: 487-518.
- Davey, M. W., Saeys, W., Hof, E., Ramon, H., Swennen, R. L. and Keulemans, J. 2009. Application of visible and near-infrared reflectance spectroscopy (Vis/NIRS) to determine carotenoid contents in banana (*Musa spp.*) fruit pulp. *J. Agric. Food Chem.* 57: 1742-1751.
- Lebot, V., Malapa, R. and Bourrieau, M. 2011. Rapid estimation of taro (*Colocasia esculenta*) quality by near-infrared reflectance spectroscopy. *J. Agric. Food Chem.* 59: 9327-9334.
- Hyvärinen, A., Karhunen, J. and Oja, E. 2001. *Independent Component Analysis*. John Wiley and Sons. New York, U.S.A.
- Blanco, M. and Villarroya, I. 2002. NIR spectroscopy: a rapid-response analytical tool. *Trac-Trends Anal. Chem.* 21: 240-250.
- Burns, D. A. and Ciurczak, E. W. 2008. *Handbook of near-infrared analysis*. 3<sup>rd</sup> ed. CRC Press. Florida, U.S.A.
- Chang, W. H., Chen, S. and Tsai, C. C. 1998. Development of a universal algorithm for use of NIR in estimation of soluble solids in fruits juices. *Trans. ASAE* 41: 1739-1745.
- Wold, S. 1987. Principal component analysis. *Chemometrics Intell. Lab. Syst.* 2: 37-52.
- Wold, S., Sjöström, M. and Eriksson, L. 2001. PLS-regression: a basic tool of chemometrics. *Chemometrics Intell. Lab. Syst.* 58: 109-130.
- Al-Mbaideen, A. and Benaissa, M. 2011. Determination of glucose concentration from NIR spectra using independent component regression. *Chemometrics Intell. Lab. Syst.* 105: 131-135.
- Chen, J. and Wang, X. Z. 2001. A new approach to near-infrared spectral data analysis using independent component analysis. *J. Chem. Inf. Comput. Sci.* 41: 992-1001.
- Kaneko, H., Arakawa, M. and Funatsu, K. 2008. Development of a new regression analysis method using independent component analysis. *J. Chem. Inf. Model.* 48: 534-541.
- Herault, J. and Jutten, C. 1986. Space or time adaptive signal processing by neural network models. In "AIP Conference Proceedings 151 on Neural Networks for

- Computing". pp. 206-211. Denker, J. S. ed. American Institute of Physics Inc. New York, U.S.A.
14. Vittoz, E. A. and Arreguit, X. 1989. CMOS integration of Heralut-Jutten cells for separation of sources. In "Analog VLSI Implementation of Neural Systems" pp. 57-83. Mead, C. and Ismail, M. ed. Kluwer Academic Publishers. Massachusetts, U.S.A.
  15. Comon, P. 1994. Independent component analysis, a new concept? *Signal Process.* 36: 287-314.
  16. Cardoso, J. F. and Souloumiac, A. 1993. Blind beamforming for non-Gaussian signals. *IEE Proceedings F* 140: 362-370.
  17. Cardoso, J. F. 1999. High-order contrasts for independent component analysis. *Neural Comput.* 11: 157-192.
  18. Hyvärinen, A. and Oja, E. 1997. A fast fixed-point algorithm for independent component analysis. *Neural Comput.* 9: 1483-1492.
  19. Hyvärinen, A. 1999. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Netw.* 10: 626-634.
  20. Hyvärinen, A. and Oja, E. 2000. Independent component analysis: algorithms and applications. *Neural Netw.* 13: 411-430.
  21. Lathauwer, L. D., Moor, B. D. and Vandewalle, J. 2000. An introduction to independent component analysis. *J. Chemometr.* 14: 123-149.
  22. Westad, F. and Kermit, M. 2003. Cross validation and uncertainty estimates in independent component analysis. *Anal. Chim. Acta* 490: 341-354.
  23. Gustafsson, M. G. 2005. Independent component analysis yields chemically interpretable latent variables in multivariate regression. *J. Chem. Inf. Model.* 45: 1244-1255.
  24. Westad, F. 2005. Independent component analysis and regression applied on sensory data. *J. Chemometr.* 19: 171-179.
  25. Hahn, S. and Yoon, G. 2006. Identification of pure component spectra by independent component analysis in glucose prediction based on mid-infrared spectroscopy. *Appl. Opt.* 45: 8374-8380.
  26. Pasadakis, N. and Kardamakis, A. A. 2006. Identifying constituents in commercial gasoline using Fourier transform-infrared spectroscopy and independent component analysis. *Anal. Chim. Acta* 578: 250-255.
  27. Kardamakis, A. A., Mouchtaris, A. and Pasadakis, N. 2007. Linear predictive spectral coding and independent component analysis in identifying gasoline constituents using infrared spectroscopy. *Chemometrics Intell. Lab. Syst.* 89: 51-58.
  28. Liu, F., He, Y. and Wang, L. 2008. Determination of effective wavelengths for discrimination of fruit vinegars using near infrared spectroscopy and multivariate analysis. *Anal. Chim. Acta* 615: 10-17.
  29. Zou, X. and Zhao, J. 2006. Improving of prediction model for apple sugar content using independent component analysis. *Chin. J. Anal. Chem.* 34: 1291-1294. In Chinese.
  30. Liu, F., Jiang, Y. and He, Y. 2009. Variable selection in visible/near infrared spectra for linear and nonlinear calibrations: A case study to determine soluble solids content of beer. *Anal. Chim. Acta* 635: 45-52.
  31. Wu, D., Feng, S. and He, Y. 2008. Short-wave near-infrared spectroscopy of milk powder for brand identification and component analysis. *J. Dairy Sci.* 91: 939-949.
  32. Fang, L. M. and Lin, M. 2008. Prediction of active substance contents in pharmaceutical tablet using ICA and NIR. *Acta Chim. Sin.* 66: 1791-1795. In Chinese.
  33. You, C. C. 2002. Import photo-sensor grader for wax apple. *Journal Taiwan Agri. Mach.* 17: 1-4. In Chinese.
  34. Lin, Y. C. 2002. Evaluation of internal quality of wax apple and papaya using near-infrared technology. National Taiwan University Master Thesis. In Chinese.
  35. Chung, C. T., Tseng, K. J., Wang, H. C. and Cheng, C. W. 2004. Study of applying near infrared spectroscopy for determinating the sugar content of wax apple fruits. *J. Agri. Mach.* 13: 13-25.
  36. de Noord, O. E. 1994. The influence of data preprocessing on the robustness and parsimony of multivariate calibration models. *Chemometrics Intell. Lab. Syst.* 23: 65-70.
  37. Williams, P. C. and Sobering, D. C. 1993. Comparison of commercial near-infrared transmittance and reflectance instruments for analysis of whole grains and seeds. *J. Near Infrared Spectrosc.* 1: 2532.
  38. Park, B., Abbott, J. A., Lee, K. J., Choi, C. H. and Choi, K. H. 2003. Near-infrared diffuse reflectance for quantitative and qualitative measurement of soluble solids and firmness of delicious and gala apples. *Trans. ASAE* 46: 1721-1731.
  39. Moneruzzaman, K. M., Al-Saif, A. M., Alebidi, A. I., Hossain, A. B. M. S., Normaniza, O. and Boyce, A. N. 2011. An evaluation of the nutritional quality evaluation of three cultivars of *Syzygium samarangense* under Malaysian conditions. *Afr. J. Agric. Res.* 6: 545-552.
  40. Tehrani, M., Chandran, S., Hossain, A. B. M. S. and Nasrulhaq-Boyce, A. 2011. Postharvest physico-chemical and mechanical changes in jambu air (*Syzygium aqueum* Alston) fruits. *Aust. J. Crop Sci.* 5: 32-38.