

國立臺灣大學生物資源暨農學院生物產業機電工程學系



博士論文

Department of Bio-Industrial Mechatronics Engineering

College of Bioresources and Agriculture

National Taiwan University

Doctoral Dissertation

應用獨立成分分析法於生物材料之近紅外光分析

Near Infrared Analysis of Biomaterials

Using Independent Component Analysis

莊永坤

Yung-Kun Chuang

指導教授：陳世銘 博士

Advisor: Suming Chen, Ph.D.

中華民國 102 年 7 月

July, 2013

誌 謝

博士的訓練過程是充滿回憶的，很高興能夠順利完成博士論文的撰寫。一路走來承蒙指導教授陳世銘博士的提攜與教誨，老師無論於身教或言教皆樹立良好的典範，使我自懵懂青澀的大學專題生，逐步邁向具有獨立思考與解決問題能力的博士，在此衷心地感謝。

承蒙博士論文指導委員臺灣大學生物機電系盧福明榮譽教授、中興大學生物機電系盛中德教授、中興大學生物機電系謝廣文副教授、宜蘭大學生物機電系邱奕志院長對於本論文的細心審閱與斧正，謹於此致上最誠摯的感謝。美國馬里蘭大學營養與食品科學系 Dr. Y. Martin Lo 與美國農業部 Beltsville 農業研究中心 Dr. Stephen R. Delwiche 於本論文的指導與建議亦功不可沒，我對二位的感激之情難以言喻。很感謝國家科學委員會提供為期一年的千里馬計畫（100-2917-I-002-004）機會，使我有幸前往美國馬里蘭大學與美國農業部，接受兩位知名學者完整而紮實的訓練與指導，並拓展我的國際觀，著實感激與珍惜。


在實驗室從事研究工作的過程中，感謝加增學長與宜璋學長帶領我入門，而一路上亦師亦友的蔡兆胤博士，更是我學習的榜樣，經由共同處理實驗室內的多項工作，使我受惠成長，特此感謝。也感謝世傑學長、翁雯學姐、毓良學長、錦銘學長及詠惠，無論是維持實驗室內各項工作與計畫的順利運作，或對於我的協助與鼓勵，皆功不可沒。很感謝宇帆、慶茵學姐、易平、育菘及俊吉學長提供我論文研究中實驗與數據的幫助，以及美國農業部 Beltsville 農業研究中心工程師 Diane E. Chan 小姐協助潤飾論文的文字。至於在我求學過程中曾經一同共事的研究助理、學長姐、同學及學弟妹們，你們都是我最好的同伴，也唯有長期以來良好的團隊合作與氣氛，才能創造實驗室豐碩的研究成果，謝謝。

最後將此博士論文獻給我最親愛的父親、母親、兄長、女友及親友們，您們總是無怨無悔地提供我生活與精神上的支持與鼓勵，使我能一路堅持到取得博士學位，謹以此論文作為最好的禮物呈獻給您們，謝謝。

中文摘要



本論文使用獨立成分分析法為唯一之核心演算法，應用於三種生物材料之近紅外光定量分析，包含蓮霧 (*Syzygium samarangense Merrill & Perry*)、藥用植物龍膽 (*Gentiana scabra Bunge*) 及白米之研究，亦對不同型態樣本 (蔗糖水溶液、蓮霧完整果、龍膽乾燥粉末及白米粒) 進行分析探討。第一部分研究結合獨立成分分析法與近紅外光光譜於蓮霧糖度之快速定量分析，結合 JADE 演算法、線性迴歸及光譜前處理方法，分別對蓮霧與蔗糖溶液之近紅外光光譜進行分析。相較於其他多變量分析方法，獨立成分分析法可提供更完整之蓮霧糖度資訊，其最佳光譜檢量模式使用一次微分光譜搭配正規化處理，光譜範圍為 600~700 nm 與 900~1098 nm， $R_c = 0.988$ ， $SEC = 0.243^\circ\text{Brix}$ ， $SEV = 0.381^\circ\text{Brix}$ ，顯示獨立成分分析法可快速準確地擷取蓮霧光譜中之糖度資訊，並建立具高預測能力之光譜檢量模式，更有效地定量蓮霧糖度。第二部分研究應用獨立成分分析法於龍膽指標成分龍膽苦苷與當藥苦苷之近紅外光分析，對 94 個組織培養瓶苗與 68 個植株樣本 (包含 68 個地上部與 68 個地下部) 進行探討。選擇與兩種指標成分高度相關之獨立成分後，組織培養瓶苗、植株地上部及植株地下部清楚分佈於獨立成分空間之三個位置，可觀察龍膽苦苷與當藥苦苷含量之變化趨勢。龍膽苦苷之最佳光譜檢量模式使用二次微分光譜，光譜範圍為 600~700 nm、1600~1700 nm 及 2000~2300 nm，其 $R_c = 0.847$ ， $SEC = 0.865\%$ ， $SEV = 0.909\%$ ；當藥苦苷之最佳光譜檢量模式使用一次微分光譜，光譜範圍為 600~800 nm 與 2200~2300 nm，其 $R_c = 0.948$ ， $SEC = 0.168\%$ ， $SEV = 0.216\%$ ，皆具有良好之預測能力。本研究成功建立龍膽苦苷與當藥苦苷之定性與定量關係，可針對不同生長時期之龍膽進行兩種指標成分含量之檢測，作為快速且準確之龍膽品質評估工具。第三部分研究應用獨立成分分析法於稻米新鮮度之快速定性與定量分析，新鮮度為決定稻米品質之重要指標，稻米貯藏時間會影響其外觀、食味及營養價值。本研究對六個收穫時期 (95




年第一期、96 年第一期、97 年第一期、98 年第一期、99 年第一期及 99 年第二期) 之白米進行探討，結果顯示不同新鮮度之白米清楚分佈於三維之獨立成分空間中，對酸鹼度 pH 值所建立之光譜檢量模式亦具有高預測能力，其 $R_c = 0.939$ ， $SEC = 0.202$ ， $SEP = 0.233$ ，表示結合獨立成分分析法與近紅外光光譜可有效評估稻米之新鮮度，且 pH 值與脂肪酸度較年份期別為更合適之評量指標。結合獨立成分分析法與近紅外光光譜可快速且正確地評估生物材料之內部成分，獨立成分分析法提供近紅外光光譜於生物材料內部成分定量分析一項快速可靠之工具，應用於評估生物材料內部品質具有重大貢獻。

關鍵詞：近紅外光光譜、獨立成分分析法、蓮霧、糖度、龍膽、龍膽苦苷、當藥苦苷、稻米新鮮度

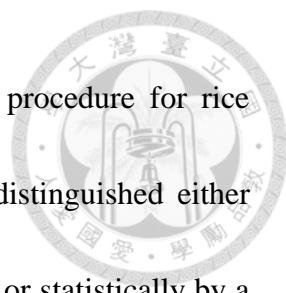
ABSTRACT



In this study, independent component analysis (ICA) was first adopted as the sole tool in conducting NIR quantitative analyses of biomaterials, including wax jambu fruit (*Syzygium samarangense Merrill & Perry*), medicinal plant *Gentiana scabra* Bunge, and milled white rice, to evaluate the applicability of this method. The influence due to various types of samples (sucrose solution, intact fruit, dry powder of *Gentiana scabra* Bunge, and rice kernel) was also studied. In the first part, ICA was integrated with near infrared (NIR) spectroscopy for rapid quantification of sugar content in wax jambu. The JADE algorithm (Joint Approximate Diagonalization of Eigenmatrices) and linear regression with spectral pretreatments were incorporated to analyze the NIR spectra of wax jambu as well as sucrose solutions. Unlike other multivariate approaches, ICA enabled comprehensive quantification of sugar content in wax jambu. In the present study, ICA was used as the sole tool to build the NIR calibration model of internal quality of intact wax jambu without any other assisted multivariate analysis methods. The best spectral calibration model of wax jambu (600 to 700 nm and 900 to 1098 nm) yielded $R_c = 0.988$, $SEC = 0.243$ °Brix, and $SEV = 0.381$ °Brix using the normalized first derivative spectra. Thus, ICA can quickly identify and effectively quantify the sugar contents in wax jambu with calibration models achieving high predictability. In the second part, ICA was applied to NIR spectroscopy on the analysis of gentiopicroside



and swertiamarin, the two bioactive components of *Gentiana scabra* Bunge. Independent components (ICs) that are highly correlated to the two bioactive components were selected for the analysis of tissue cultures, shoots and roots, which were found to distribute in three different positions within the domain (2D and 3D) constructed by the ICs. This setup could be used for quantitative determination of respective contents of gentiopicroside and swertiamarin within the plants. For gentiopicroside, the spectral calibration model based on the 2nd derivative spectra produced the best effect in the wavelength ranges of 600 to 700 nm, 1600 to 1700 nm, and 2000 to 2300 nm ($R_c = 0.847$, $SEC = 0.865 \%$, and $SEV = 0.909 \%$). For swertiamarin, spectral calibration model based on the 1st derivative spectra gave the best effect in the wavelength ranges of 600 to 800 nm and 2200 to 2300 nm ($R_c = 0.948$, $SEC = 0.168 \%$, and $SEV = 0.216 \%$). Both models showed a satisfactory predictability. This study successfully established qualitative and quantitative correlations for gentiopicroside and swertiamarin with NIR spectra, enabling rapid and accurate inspection on the bioactive components of *Gentiana scabra* Bunge at different growth stages. Furthermore, determination of freshness is an important issue for rice quality. The storage time of rice has an enormous effect on its appearance, flavor, and quality of the nutrients. A total of 180 white rice samples were collected from 6 crop seasons (2nd crop of 2010, 1st crop of 2010, 1st crop of 2009, 1st crop of 2008, 1st crop of 2007 and 1st



crop of 2006) for the purpose of developing an ICA NIR based procedure for rice freshness as quantified by pH. Freshness of white rice could be distinguished either visually by a 3-dimensional diagram composed from ICs 2, 3 and 4, or statistically by a calibration model ($R_c = 0.939$, $SEC = 0.202$, and $SEP = 0.233$). The results showed that ICA with NIR has the potential to be a useful tool for evaluating rice freshness. Compared to harvest time, pH value and fat acidity were more appropriate to serve as indicators of rice freshness. By combining ICA with NIR spectroscopy, fast and accurate evaluation of constituents in biomaterials could be achieved. ICA offers a rapid and reliable tool for quantitative analyses of constituents in biomaterials by NIR spectroscopy. The obtained results contribute substantially to identify multiple constituents of biomaterials and evaluate their concentrations.

Keywords: Near infrared spectroscopy, Independent component analysis, Wax jambu, Sugar content, *Gentiana scabra* Bunge, Gentiopicroside, Swertiamarin, Rice freshness

CONTENTS



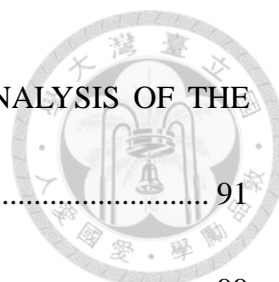
誌謝.....	i
中文摘要.....	ii
ABSTRACT.....	iv
CONTENTS.....	vii
LIST OF FIGURES.....	xii
LIST OF TABLES.....	xv
CHAPTER 1. GENERAL INTRODUCTION.....	1
1.1 INTRODUCTION.....	1
1.1.1 WAX JAMBU.....	4
1.1.2 GENTIANA SCABRA BUNGE.....	5
1.1.3 RICE.....	7
1.2 GENERAL OBJECTIVE.....	8
1.3 DISSERTATION ORGANIZATION.....	9
CHAPTER 2. INTEGRATION OF INDEPENDENT COMPONENT ANALYSIS WITH NEAR INFRARED SPECTROSCOPY FOR RAPID QUANTIFICATION OF SUGAR CONTENT IN WAX JAMBU.....	10
2.1 INTRODUCTION.....	10
2.2 MATERIALS AND METHODS.....	14



2.2.1 SAMPLE PREPARATION.....	14
2.2.2 NIR SPECTRA AND SUGAR CONTENT MEASUREMENT.....	14
2.2.3 DATA ANALYSIS.....	16
2.3 RESULTS AND DISCUSSION	22
2.3.1 SUCROSE SOLUTION	22
2.3.2 WAX JAMBU.....	35
2.4 CONCLUSIONS	46
CHAPTER 3. QUANTIFICATION OF BIOACTIVE GENTIOPIROSIDE IN A MEDICINAL PLANT <i>GENTIANA SCABRA</i> BUNGE BY NEAR INFRARED SPECTROSCOPY.....	48
3.1 INTRODUCTION	48
3.2 MATERIALS AND METHODS.....	50
3.2.1 <i>G. SCABRA</i> BUNGE SAMPLE PREPARATION	50
3.2.2 NIR SPECTRA AND HPLC MEASUREMENT.....	50
3.2.3 DATA ANALYSIS	51
3.3 RESULTS AND DISCUSSION	56
3.3.1 GENTIOPIROSIDE CONCENTRATION AND DISTRIBUTION IN <i>G. SCABRA</i> BUNGE.....	56
3.3.2 CORRELATION BETWEEN NIR SPECTRA AND	



GENTIOPIROSIDE CONTENT	57
3.3.3 GENTIOPIROSIDE QUANTIFICATION USING SPECIFIC WAVELENGTH RANGES	61
3.3.4 GENTIOPIROSIDE QUANTIFICATION USING CCD CAMERA WAVELENGTH SPECTRA.....	67
3.4 CONCLUSIONS	75
CHAPTER 4. INTEGRATION OF INDEPENDENT COMPONENT ANALYSIS WITH NEAR INFRARED SPECTROSCOPY FOR ANALYSIS OF BIOACTIVE COMPONENTS IN A MEDICINAL PLANT <i>GENTIANA SCABRA</i> BUNGE	76
4.1 INTRODUCTION	76
4.2 MATERIALS AND METHODS	78
4.2.1 <i>GENTIANA SCABRA</i> BUNGE SAMPLE PREPARATION	78
4.2.2 NIR SPECTRA AND HPLC MEASUREMENT.....	79
4.2.3 DATA ANALYSIS	80
4.3 RESULTS AND DISCUSSION	85
4.3.1 DISTRIBUTIONS OF THE TARGET CONSTITUENTS IN <i>GENTIANA SCABRA</i> BUNGE.....	85
4.3.2 CORRELATION BETWEEN NIR SPECTRA AND TARGET CONSTITUENTS' CONTENTS	86



4.3.3 NIR SPECTRA DECOMPOSITION AND ICA ANALYSIS OF THE TARGET CONSTITUENTS	91
4.4 CONCLUSIONS	99
CHAPTER 5. INTEGRATION OF INDEPENDENT COMPONENT ANALYSIS WITH NEAR INFRARED SPECTROSCOPY FOR EVALUATION OF RICE FRESHNESS	101
5.1 INTRODUCTION	101
5.2 MATERIALS AND METHODS	103
5.2.1 SAMPLE PREPARATION.....	103
5.2.2 NIR SPECTRA AND PH VALUE MEASUREMENT	104
5.2.3 DATA ANALYSIS	106
5.3 RESULTS AND DISCUSSION	108
5.3.1 RELATIONSHIP BETWEEN FAT ACIDITY AND PH VALUE.....	108
5.3.2 DISTRIBUTIONS OF THE PH VALUE IN RICE.....	110
5.3.3 NIR SPECTRA DECOMPOSITION AND ICA ANALYSIS OF THE PH VALUE	112
5.4 CONCLUSIONS	117
CHAPTER 6. GENERAL CONCLUSIONS	119
6.1 GENERAL DISCUSSION	119

6.2 RECOMMENDATIONS FOR FUTURE RESEARCH.....	122
REFERENCES	123



LIST OF FIGURES



- Fig. 2.1** 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. 16
- Fig. 2.2** 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. 24
- Fig. 2.3** 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. 26
- Fig. 2.4** Correlation between the values of IC 1 in the mixing matrix and the reference sugar contents of sucrose solutions. 27
- Fig. 2.5** (A) Original NIR spectra of sucrose solutions, (B) IC 1 decomposed from calibration sets, and (C) the reflectance spectrum of sucrose powder post-Detrend. 30
- Fig. 2.6** Correlation coefficient distributions of the spectra and the sugar content of wax jambu through three different spectral pretreatments (original spectra, 1st derivative spectra, and 2nd derivative spectra). 38



Fig. 2.7 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..... 40

Fig. 3.1 The spectra of *G. scabra* Bunge powder post-MSC (A) tissue culture and (B) grown plants. 58

Fig. 3.2 Correlation coefficient distributions between absorbance values of the spectra and gentiopicroside contents of the *G. scabra* Bunge powder (A) tissue culture and (B) grown plants. 61

Fig. 4.1 (A) The spectra of *Gentiana scabra* Bunge powder post-MSC; (B) correlation coefficient distributions between the spectra and gentiopicroside; and (C) correlation coefficient distributions between the spectra and swertiamarin..... 90

Fig. 4.2 The three ICs decomposed from the original spectra of *Gentiana scabra* Bunge powder post-MSC that has higher correlation with gentiopicroside and swertiamarin. 92

Fig. 4.3 Scores of tissue culture, shoot, and root in IC 4-IC 5 space established with calibration samples. (A) = calibration set, (B) = validation set. Scores of gentiopicroside and swertiamarin in IC 4-IC 5 space established with calibration samples. (C) = calibration set, (D) = validation set..... 94



Fig. 4.4 Scores of tissue culture, shoot, and root in IC 4-IC 5-IC 6 space established with calibration samples. (A) = calibration set, (B) = validation set. Scores of gentiopicroside and swertiamarin in IC 4-IC 5-IC 6 space established with calibration samples. (C) = calibration set, (D) = validation set..... 96

Fig. 4.5 Relationship between the estimated contents and the reference contents of (A) gentiopicroside; and (B) swertiamarin in *Gentiana scabra* Bunge. 99

Fig. 5.1 Relationship between fat acidity and pH value established by the 18 selected rice samples (Hu, 2011; Chen *et al.*, 2011). 109

Fig. 5.2 Distributions of pH values for six crop seasons of white rice samples. 111

Fig. 5.3 Relationship between the numbers of independent components and the model standard errors for pH value. 113

Fig. 5.4 Scores of white rice with 6 crop seasons in the vector space of independent components 2, 3, and 4 established with calibration samples. a = calibration set, b = validation set. 115

Fig. 5.5 Scatter plot of the reference pH values and the predicted pH values by independent component analysis of the NIR spectra..... 117

LIST OF TABLES

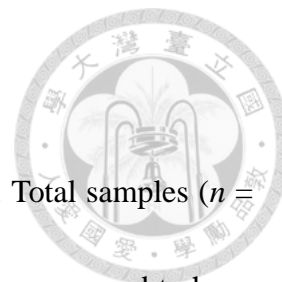


Table 2.1 Summary of sucrose solutions and sample sugar contents. Total samples ($n = 78$), calibration set ($n = 52$) and validation set ($n = 26$) were arranged to have consistent distributions of sugar content.	22
Table 2.2 Regression coefficient matrix of sucrose solutions with 7 ICs were extracted from the NIR spectra of calibration sets. Correlation between the absolute value of each IC in regression coefficient matrix and sugar content was examined.....	28
Table 2.3 Regression results by ICA and PLSR analyses for sucrose solutions.....	32
Table 2.4 Summary of wax jambu (<i>Syzygium samarangense Merrill & Perry</i>) 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.	36
Table 2.5 Regression results by ICA and PLSR analyses for wax jambu (without spectral pretreatment).	42
Table 2.6 Regression results by ICA and PLSR analyses for wax jambu (with spectral pretreatment).....	45
Table 3.1 The gentiopicroside content in tissue culture and grown plants of <i>G. scabra</i> Bunge.....	57

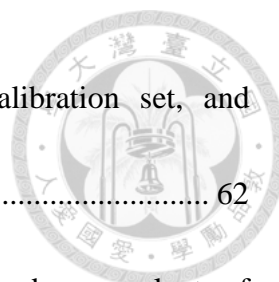


Table 3.2 The gentiopicroside content of effective samples, calibration set, and validation set in tissue culture and grown plants.....	62
Table 3.3 Prediction of the gentiopicroside content in tissue culture and grown plants of <i>G. scabra</i> Bunge by MPLSR models in the wavelength of 400 to 2498 nm..	65
Table 3.4 Prediction of the gentiopicroside content in tissue culture and grown plants of <i>G. scabra</i> Bunge by MPLSR models in the wavelength of 400 to 1098 nm..	69
Table 3.5 Prediction of the gentiopicroside content in tissue culture and grown plants of <i>G. scabra</i> Bunge by SMLR models in the wavelength of 400 to 1098 nm. ...	73
Table 4.1 Contents and distributions of the target constituents in <i>Gentiana scabra</i> Bunge.....	86
Table 4.2 The target constituents' contents of effective samples, calibration set, and validation set in <i>Gentiana scabra</i> Bunge.	87
Table 4.3 Prediction of the target constituents' contents in <i>Gentiana scabra</i> Bunge by ICA models.....	98
Table 5.1 Regression results by ICA analyses for white rice.....	115

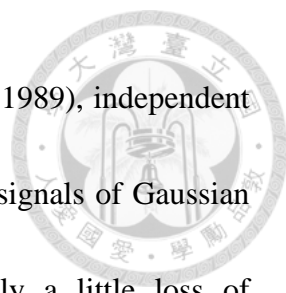
CHAPTER 1. GENERAL INTRODUCTION



1.1 INTRODUCTION


Near infrared (NIR) spectroscopy, a nondestructive sensing method based on specific absorptions within a given range of wavelength corresponding to the constituents in the sample (McClure, 2003), has been widely applied for the evaluation of internal quality of agricultural products (Davey *et al.*, 2009; Lebot *et al.*, 2011). 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’ (Hyvärinen *et al.*, 2001) 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 (Blanco and Villarroya, 2002; Burns and Ciurczak, 2008), including multiple linear regression (MLR) (Chang *et al.*, 1998), principal component regression (PCR) (Wold, 1987), and partial least squares regression (PLSR) (Wold *et al.*, 2001). 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 (Al-Mbaideen and Benaissa, 2011; Chen and Wang, 2001; Kaneko *et al.*, 2008).

A multiuse statistical approach originally used to implement ‘blind source separation’



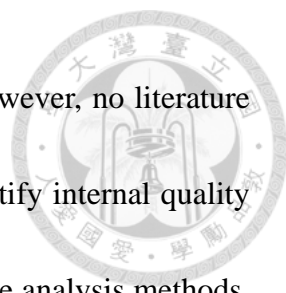
in signal processing (Herault and Jutten, 1986; Vittoz and Arreguit, 1989), independent component analysis (ICA) is capable of disassembling the mixture signals of Gaussian distribution into non-Gaussian independent constituents with only a little loss of information and does not require any information to be added to the source (Comon, 1994). In practice, multiple ICA algorithms have been developed, including JADE algorithm (joint approximate diagonalization of eigenmatrices) (Cardoso and Souloumiac, 1993; Cardoso, 1999) and FastICA algorithm (Hyvärinen and Oja, 1997; Hyvärinen, 1999), making ICA a high-speed and reliable tool (Hyvärinen and Oja, 2000) for analytical chemistry (Lathauwer *et al.*, 2000; Al-Mbaideen and Benaissa, 2011), biomedical signal processing, telecommunications, econometrics, audio processing, and image processing (Hyvärinen *et al.*, 2001).

Application of ICA for spectrum analysis has been demonstrated by Chen and Wang (2001) to separate the pure spectra of various constituents from the NIR spectra of the mixture and to build qualitative relationship between the estimated independent components and the constituents. Such a capability also enabled complete explanation of the constituents' properties for NIR qualitative analyses (Westad and Kermit, 2003). In addition, ICA was used to obtain statistically independent and chemically interpretable latent variables (LVs) in multivariate regression (Gustafsson, 2005). It was



also noted that the number of independent components extracted from the spectra of mixtures is related to the performance of ICA (Westad, 2005). Moreover, ICA was employed to identify the infrared spectrum of mixtures containing two pure materials (Hahn and Yoon, 2006) as well as the constituents in commercial gasoline (Pasadakis and Kardamakis, 2006; Kardamakis *et al.*, 2007). Equally noteworthy is that the calibration model built through MLR, after using ICA to extract independent components of aqueous solutions, gave good predictability (Kaneko *et al.*, 2008), whereas NIR estimation of sucrose concentration (Chuang *et al.*, 2010) and glucose concentration (Al-Mbaideen and Benaissa, 2011) were enhanced by using ICA.

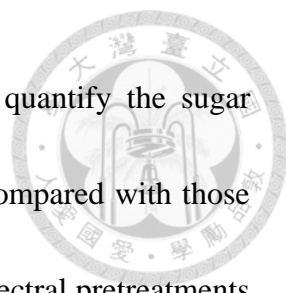
While application of ICA for spectral analysis appears promising, available literature still focuses mainly on chemical samples or non-natural products, such as (1) identification of constituents in the mixture, especially for chemical compounds (Chen and Wang, 2001; Hahn and Yoon, 2006; Pasadakis and Kardamakis, 2006; Kardamakis *et al.*, 2007; Liu *et al.*, 2008; Kaneko *et al.*, 2008; Al-Mbaideen and Benaissa, 2011); (2) a preprocessing method for improving predictability of calibration model (Zou and Zhao, 2006); and (3) combination of ICA and other multivariate analysis methods, such as PCA-ICA (Pasadakis and Kardamakis, 2006), ICA-MLR (Kaneko *et al.*, 2008; Liu *et al.*, 2009), ICA-PLS (Liu *et al.*, 2009), ICA-LS-SVM (Wu *et al.*, 2008) and ICA-NNR



(Fang and Lin, 2008) to deal with linear or nonlinear problems. However, no literature exists by using ICA with NIR spectroscopy as the sole tool to quantify internal quality or constituents of biomaterials without any other assisted multivariate analysis methods. The applicability of ICA for quantitative inspection of biomaterials thus should be evaluated and studied. In this dissertation, ICA was first applied as the sole tool in conducting NIR quantitative analyses of biomaterials, including wax jambu fruit (see CHAPTER 2), medicinal plant *Gentiana scabra* Bunge (see CHAPTER 3 and 4), and milled white rice (see CHAPTER 5), to evaluate the applicability of this method. Influence due to various types of sample (sucrose solution, intact fruit, dry powder of *Gentiana scabra* Bunge, and cargo rice) was also studied.

1.1.1 WAX JAMBU

Wax jambu (*Syzygium samarangense* Merrill & Perry), an endemic fruit in Taiwan and parts of southeast Asia 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 (You, 2002; Lin, 2002; Chung *et al.*, 2004). For further applications of ICA with NIR spectroscopy in the inspection of fruits, wax jambu is suitable to serve as an example for discussion. In the



present study, ICA was integrated with 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 — mixtures of sucrose and de-ionized water. 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.

1.1.2 GENTIANA SCABRA BUNGE


Medicinal plants have always been considered an important and reliable source of pharmacy, since they are rich in many bioactive components. The international trade market for medicinal plant products continues to expand and covers food, beverages, drugs, cosmetics, and skin care products. *Gentiana scabra* Bunge, a perennial herbaceous plant, is mainly grown in temperate regions such as Taiwan, China, Japan, South Korea, and Russia. Dried root and rootstock of *Gentiana scabra* Bunge are commonly used as pharmaceutical raw materials, since they are rich in many secoiridoid glycosides such as gentiopicroside, swertiamarin and sweroside (Kakuda *et al.*, 2001). In particular, gentiopicroside has been shown to protect liver, inhibit liver dysfunction, and promote gastric acid secretion in addition to its antimicrobial and anti-inflammatory effects, making it a popular ingredient in Chinese herbal medicine

and health products (Kim *et al.*, 2009).



In early days, *Gentiana scabra* Bunge was mainly collected from the wild. As the demand for *Gentiana scabra* Bunge increases, the wild resources diminish gradually, thus restoration of *Gentiana scabra* Bunge became an important issue (Zhang *et al.*, 2010). Studies in recent years used tissue culture technology to cultivate of *Gentiana scabra* Bunge (Cai *et al.*, 2009), by domesticating the tissue culture of *Gentiana scabra* Bunge, then transplanting it to the greenhouse for cultivation. In order to monitor the change of *Gentiana scabra* Bunge during the growth process, it is necessary to measure the bioactive components of *Gentiana scabra* Bunge. However, the commonly used methods such as micellar electrokinetic capillary chromatography (MECC) (Glatz *et al.*, 2000), high performance liquid chromatography (HPLC) (Szücs *et al.*, 2002; Kikuchi *et al.*, 2005; Carnat *et al.*, 2005; Kušar *et al.*, 2010; Hayta *et al.*, 2011a; Hayta *et al.*, 2011b), liquid chromatography-mass spectrometry (LC-MS) (Aberham *et al.*, 2007; Aberham *et al.*, 2011), and ultra-performance liquid chromatography (UPLC) (Nastasijević *et al.*, 2012) are all time-consuming and energy-intensive, hence cannot be applicable for daily quality inspection of *Gentiana scabra* Bunge during cultivation.

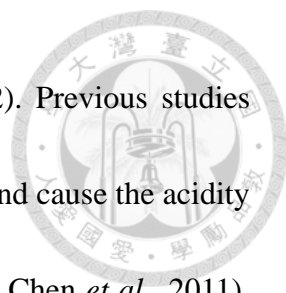
NIR spectroscopy has been widely used in dispensation, such as herbal component



analysis of Chinese herbal plants *Angelicae gigantis Radix* (Woo *et al.*, 2005), Rhubarb (Zhang and Tang, 2005), licorice (*Glycyrrhizia uralensis* Fisch.) (Wang *et al.*, 2007), Panax Species (Chen *et al.*, 2011), and *Lonicera japonica* (Wu *et al.*, 2012), as well as the content detection of active pharmaceutical ingredients (APIs) in tablets (Paris *et al.*, 2006; Jamrógiewicz, 2012; Porfire *et al.*, 2012). However, it has not been employed to qualitatively monitor the growth of *Gentiana scabra* Bunge. In recent years, ICA has been used in medicinal tests (Fang and Lin, 2008; Wang *et al.*, 2009; Shao *et al.*, 2009). Considering there hasn't been any study applying NIR spectroscopy in inspection on internal components of *Gentiana scabra* Bunge currently, it is the intent of this study to apply ICA, which could analyze various components simultaneously, in NIR spectroscopy analysis on gentiopicroside and swertiamarin to discuss qualitative and quantitative relationships of the two bioactive components. Efforts were also made to build spectral calibration models with high predictability in order to evaluate the potentiality of NIR for quality inspection on *Gentiana scabra* Bunge.

1.1.3 RICE

Rice is one of the most important and popular food crops in the world, and freshness of rice depends on the storage conditions such as storage time, storage temperature, and relative humidity. Among them, the storage time of rice has an enormous effect on its



appearance, flavor, and quality of the nutrients (Zhou *et al.*, 2002). Previous studies demonstrated that most lipids in rice hydrolyze into free fatty acids and cause the acidity of rice to increase with prolonged storage (Takano, 1989; Hu, 2011; Chen *et al.*, 2011). Therefore, the determination of rice freshness is one of the main goals in site examination. There is a strong need to develop a non-invasive, rapid detection method for the analysis of freshness. Therefore, the objective of the current study was to inspect rice freshness in terms of qualitative and quantitative approaches using NIR spectroscopy. Rice freshness was expressed by both pH value and fat acidity. The pH values were determined by bromothymol blue - methyl red (BTB-MR) method (Hsu and Song, 1988) and fat acidity by AACC International method 02-02.02 (AACC International, 2000). By means of a calibration curve, a relationship between pH and fat acidity was established. ICA was subsequently integrated with NIR spectral analysis to quantify the pH in rice. Linear regression was then used to build spectral calibration models of pH value.

1.2 GENERAL OBJECTIVE

The objective of the dissertation was to apply ICA as the sole tool in conducting NIR quantitative analyses of biomaterials, including wax jambu fruit, medicinal plant *Gentiana scabra* Bunge, and milled white rice, to evaluate its applicability. Influence

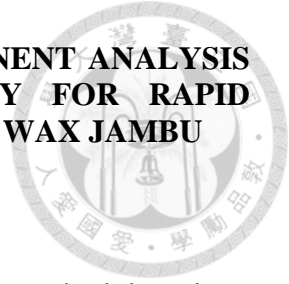
due to various types of sample (sucrose solution, intact fruit, dry powder of *Gentiana scabra* Bunge, and cargo rice) was also studied in the dissertation.



1.3 DISSERTATION ORGANIZATION


The dissertation is written in the alternative format. The “GENERAL INTRODUCTION” section is followed by chapters containing manuscripts of four research papers: (1) Integration of independent component analysis with near infrared spectroscopy for rapid quantification of sugar content in wax jambu (*Syzygium samarangense* Merrill & Perry), (2) Quantification of bioactive gentiopicroside in a medicinal plant *Gentiana scabra* Bunge using near infrared spectroscopy, (3) Integration of independent component analysis with near infrared spectroscopy for analysis of bioactive components in a medicinal plant *Gentiana scabra* Bunge, and (4) Integration of independent component analysis with near infrared spectroscopy for evaluation of rice freshness. These are followed by “GENERAL CONCLUSIONS” section.

CHAPTER 2. INTEGRATION OF INDEPENDENT COMPONENT ANALYSIS WITH NEAR INFRARED SPECTROSCOPY FOR RAPID QUANTIFICATION OF SUGAR CONTENT IN WAX JAMBU



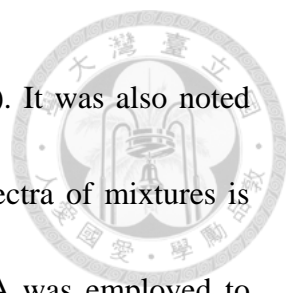
2.1 INTRODUCTION

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 (McClure, 2003), has been widely applied for the evaluation of internal quality of agricultural products (Davey *et al.*, 2009; Lebot *et al.*, 2011). 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' (Hyvärinen *et al.*, 2001) 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 (Blanco and Villarroya, 2002; Burns and Ciurczak, 2008), including multiple linear regression (MLR) (Chang *et al.*, 1998), principal component regression (PCR) (Wold, 1987), and partial least squares regression (PLSR) (Wold *et al.*, 2001). 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 (Al-Mbaideen and Benaissa, 2011; Chen and Wang, 2001; Kaneko *et al.*, 2008).




A multiuse statistical approach originally used to implement ‘blind source separation’ in signal processing (Herault and Jutten, 1986; Vittoz and Arreguit, 1989), 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 (Comon, 1994). In practice, multiple ICA algorithms have been developed, including JADE algorithm (joint approximate diagonalization of eigenmatrices) (Cardoso and Souloumiac, 1993; Cardoso, 1999) and FastICA algorithm (Hyvärinen and Oja, 1997; Hyvärinen, 1999), making ICA a high-speed and reliable tool (Hyvärinen and Oja, 2000) for analytical chemistry (Lathauwer *et al.*, 2000; Al-Mbaideen and Benaissa, 2011), biomedical signal processing, telecommunications, econometrics, audio processing, and image processing (Hyvärinen *et al.*, 2001).

Application of ICA for spectrum analysis has been demonstrated by Chen and Wang (2001) to separate the pure spectra of various constituents from the NIR spectra of the 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 (Westad and Kermit, 2003). In addition, ICA was used to obtain statistically independent and chemically interpretable



latent variables (LVs) in multivariate regression (Gustafsson, 2005). It was also noted that the number of independent components extracted from the spectra of mixtures is related to the performance of ICA (Westad, 2005). Moreover, ICA was employed to identify the infrared spectrum of mixtures containing two pure materials (Hahn and Yoon, 2006) as well as the constituents in commercial gasoline (Pasadakis and Kardamakis, 2006; Kardamakis *et al.*, 2007). Equally noteworthy is that the calibration model built through MLR, after using ICA to extract independent components of aqueous solutions, gave good predictability (Kaneko *et al.*, 2008), whereas NIR estimation of sucrose concentration (Chuang *et al.*, 2010) and glucose concentration (Al-Mbaideen and Benaissa, 2011) were 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 (Chen and Wang, 2001; Hahn and Yoon, 2006; Pasadakis and Kardamakis, 2006; Kardamakis *et al.*, 2007; Liu *et al.*, 2008; Kaneko *et al.*, 2008; Al-Mbaideen and Benaissa, 2011); (2) a preprocessing method for improving predictability of calibration model (Zou and Zhao, 2006); and (3) combination of ICA and other multivariate analysis methods, such as PCA-ICA (Pasadakis and Kardamakis, 2006), ICA-MLR (Kaneko *et al.*, 2008; Liu *et al.*, 2009), ICA-PLS (Liu *et al.*, 2009),



ICA-LS-SVM (Wu *et al.*, 2008) and ICA-NNR (Fang and Lin, 2008) 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 (Fig. 2.1) 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 (You, 2002; Lin, 2002; Chung *et al.*, 2004). 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.



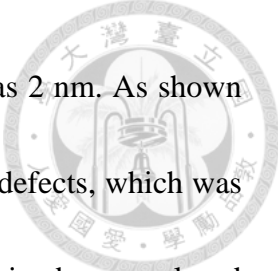
2.2 MATERIALS AND METHODS

2.2.1 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 2500 rpm and 22°C to clarify the sample in centrifuge KUBOTA 2700 (KUBOTA Corporation Co., Ltd., Osaka City, Osaka, Japan).

2.2.2 NIR SPECTRA AND SUGAR CONTENT MEASUREMENT

A NIRS 6500 spectrophotometer and sample transport (FOSS NIRSystems, Laurel, MD, U.S.A.) with quartz cuvette were used to measure the transmittance spectra of sucrose solutions. The wavelength ranged from 400 to 2498 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, U.S.A.) was used to measure the reflectance spectra of the wax jambu samples.



The wavelength range was from 400 to 2498 nm and the interval was 2 nm. As shown in Fig. 2.1(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 Fig. 2.1(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.

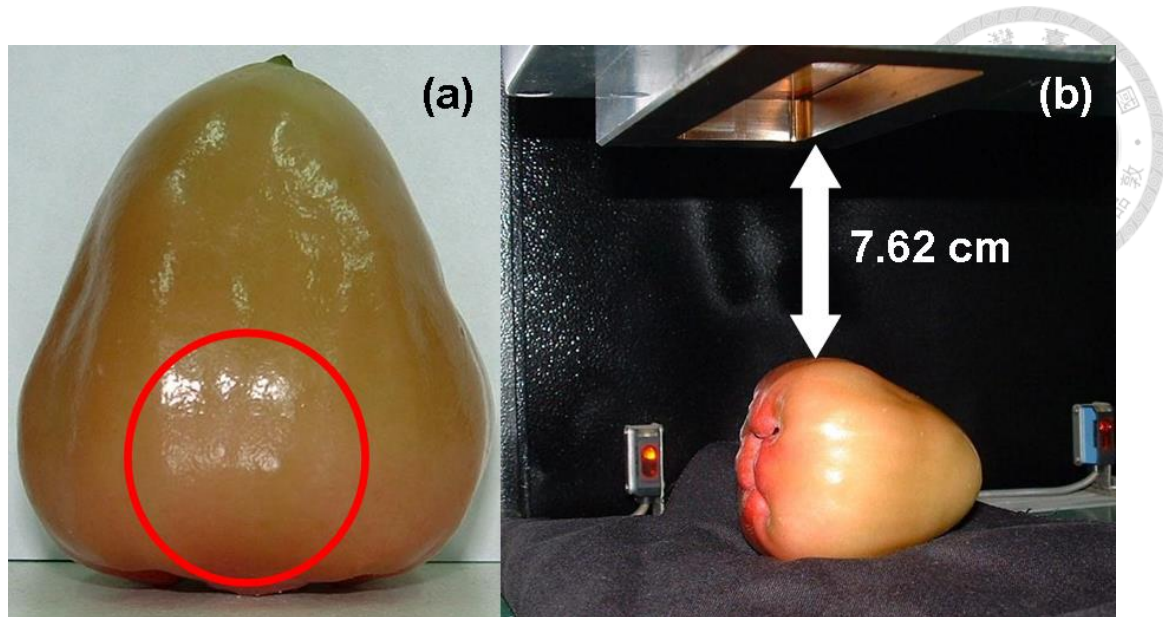


Fig. 2.1 A wax jambu (*Syzygium samarangense* Merrill & Perry) sample (A) side view and the NIR measurement location, and (B) sample placement with suggested distance 7.62 cm between the light source and the top of sample in the on-line NIRS 6500 spectrophotometer.

2.2.3 DATA ANALYSIS

2.2.3.1 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 (Hyvärinen and Oja, 2000). 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 (\mathbf{M}) or the separating matrix (\mathbf{W}). In order to predict the unknown source, it is assumed that $\mathbf{W} = \mathbf{M}^{-1}$,

$$\hat{\mathbf{s}} = \mathbf{W}\mathbf{x} = \mathbf{M}^{-1}\mathbf{M}\mathbf{s} \quad (2.1)$$

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

In the present study JADE (joint approximate diagonalization of eigenmatrices) algorithm (Cardoso and Souloumiac, 1993; Cardoso, 1999) 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:

$$\mathbf{A} = \mathbf{M}\mathbf{I} \quad (2.2)$$

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 (\mathbf{M}) and



the component concentration (**C**) can be expressed as:

$$\mathbf{C} = \mathbf{MB} \quad (2.3)$$

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.

2.2.3.2 PARTIAL LEAST SQUARES REGRESSION (PLSR)

Partial least squares regression (PLSR), a typical method in chemometrics (Wold *et al.*, 2001), 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).

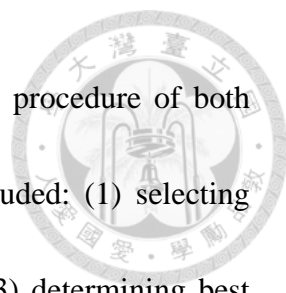


2.2.3.3 SPECTRAL PRETREATMENTS

The purpose of spectral pretreatments was to eliminate the spectral variation, which was not caused by chemical information contained in the samples (de Noord, 1994). For the raw NIR spectra of sucrose solutions and wax jambu, three different spectral pretreatments were employed in this study: (1) normalization; (2) 1st derivative with normalization; and (3) 2nd 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.

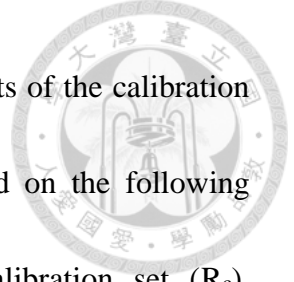
2.2.3.4 MODEL ESTABLISHMENT

This study used the mathematic software MATLAB (The MathWorks, Inc., Natick, MA, U.S.A.) to write ICA programs based on JADE algorithm 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, U.S.A.) 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 2498 nm) were used for comparing the tolerance abilities of ICA and PLSR since spectral bands with more noises (e.g. 2200 to 2498 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:

$$SEC = \left[\frac{1}{n_c} \sum_{i=1}^{n_c} (Y_r - Y_c)_i^2 \right]^{1/2} \quad (2.4)$$

$$SEV = \left\{ \frac{1}{n_v} \sum_{i=1}^{n_v} [(Y_r - Y_v) - \text{Bias}]_i^2 \right\}^{1/2} \quad (2.5)$$

$$\text{Bias} = \frac{1}{n_v} \sum_{i=1}^{n_v} (Y_r - Y_v)_i \quad (2.6)$$

$$RPD = SD/SEV \quad (2.7)$$

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. The greater the value of RPD is considered adequate for analytical purposes in most of NIR spectroscopy applications for agricultural products (Williams and Sobering, 1993).



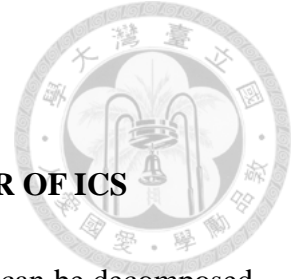
2.3 RESULTS AND DISCUSSION

2.3.1 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 ($^{\circ}$ Brix) is shown in Table 2.1. For all the samples within the calibration and validation sets, the difference between maximum values of two sets was 0.2 $^{\circ}$ Brix; the differences for other items including minimum, average, standard deviation, and coefficient of variation (CV), were all smaller than 0.5 $^{\circ}$ Brix. The above sets of samples were conforming to the consistent requirement of sugar content distributions.

Table 2.1 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.

Sucrose Solutions						
Group	n	Sugar Content ($^{\circ}$ 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



2.3.1.1 SELECTION OF THE MOST APPROPRIATE NUMBER OF ICS

According to the definition of ICA, the observed receptor signals can be decomposed at most into a number of ICs (independent components) equal to the number of samples (Hyvärinen and Oja, 2000). This study used the data of full range of wavelength (400 to 2498 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 Fig. 2.2, 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 because the sucrose solutions were mixtures of sucrose and water, hence only the initial 4 ICs were applied in the calibration model.

The results of ICA with normalized spectra can be observed in Fig. 2.2. 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.

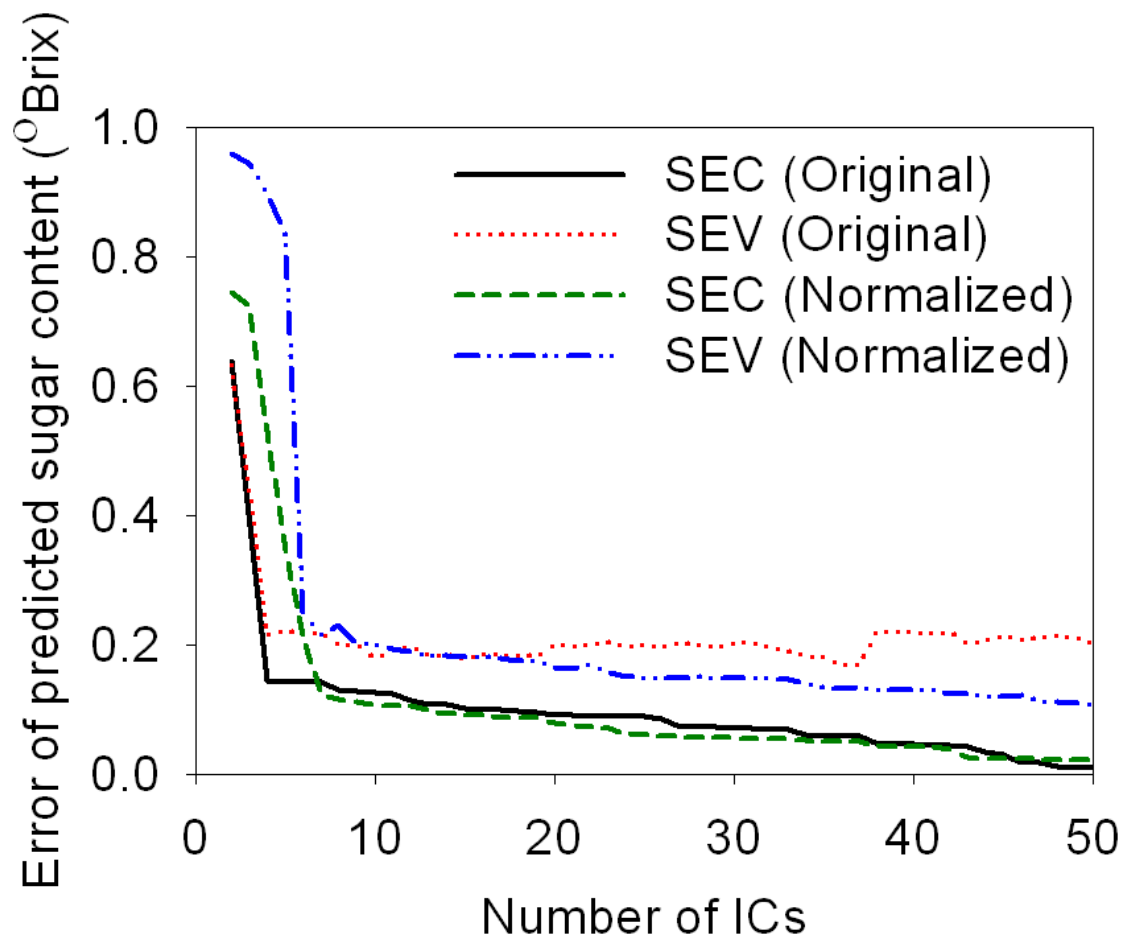


Fig. 2.2 Relationship between the numbers of ICs and errors of the predicted sugar content for sucrose solutions. The most appropriate number of ICs for normalized spectra was determined by the tendency of SEC (green-short dash line) and SEV (blue-dash dot dot line) values.

2.3.1.2 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 Fig. 2.3, with the coefficient of determination (r^2) being only 4.0×10^{-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 Fig. 2.3, with all of the r^2 smaller than 0.243, conforming to the mutually independent characteristics of ICs (Hyvärinen and Oja, 2000).

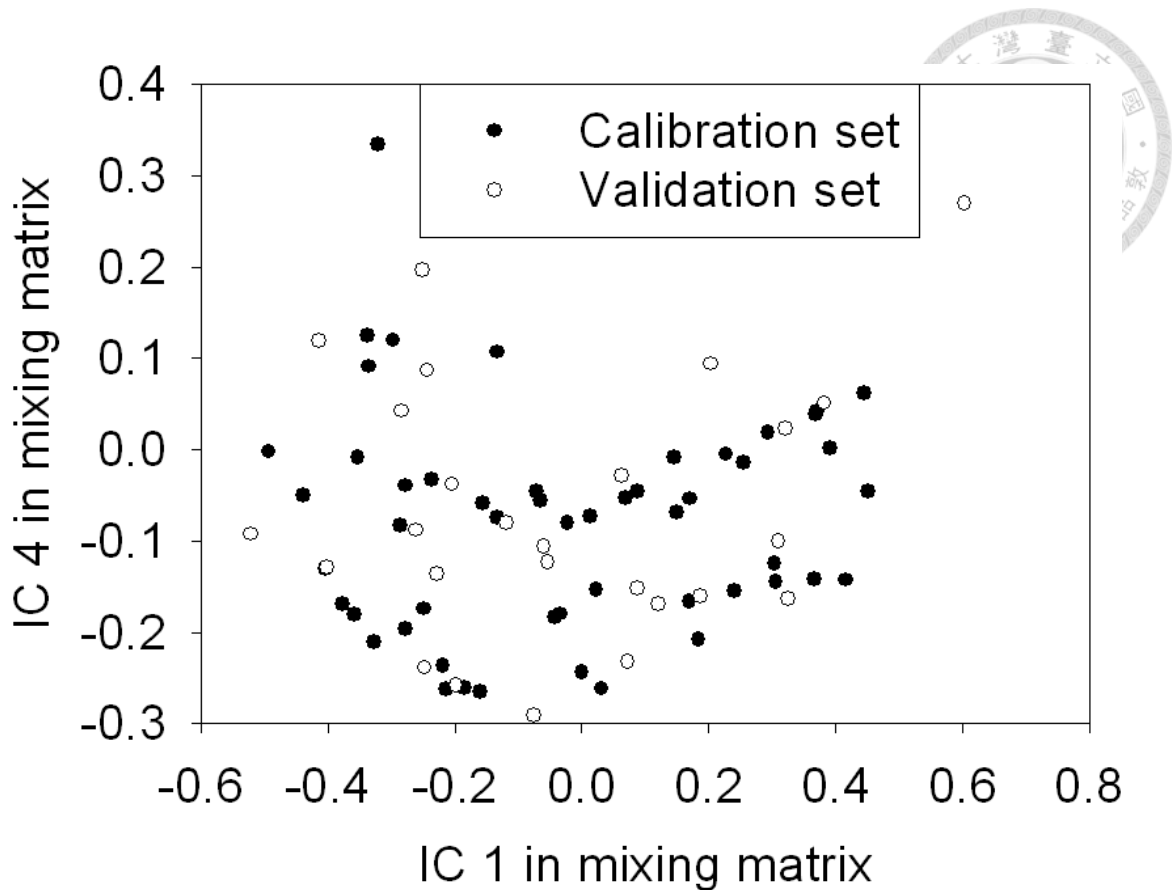


Fig. 2.3 Distribution of calibration and validation samples of sucrose solutions in IC

1-IC 4 space. IC 1 and IC 4 were randomly selected from the 7 ICs.

Eq. 2.5 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 Fig. 2.4, 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 (2005). Therefore, selection of the numbers of ICs is important since it influences how the information is used after spectra decomposition.

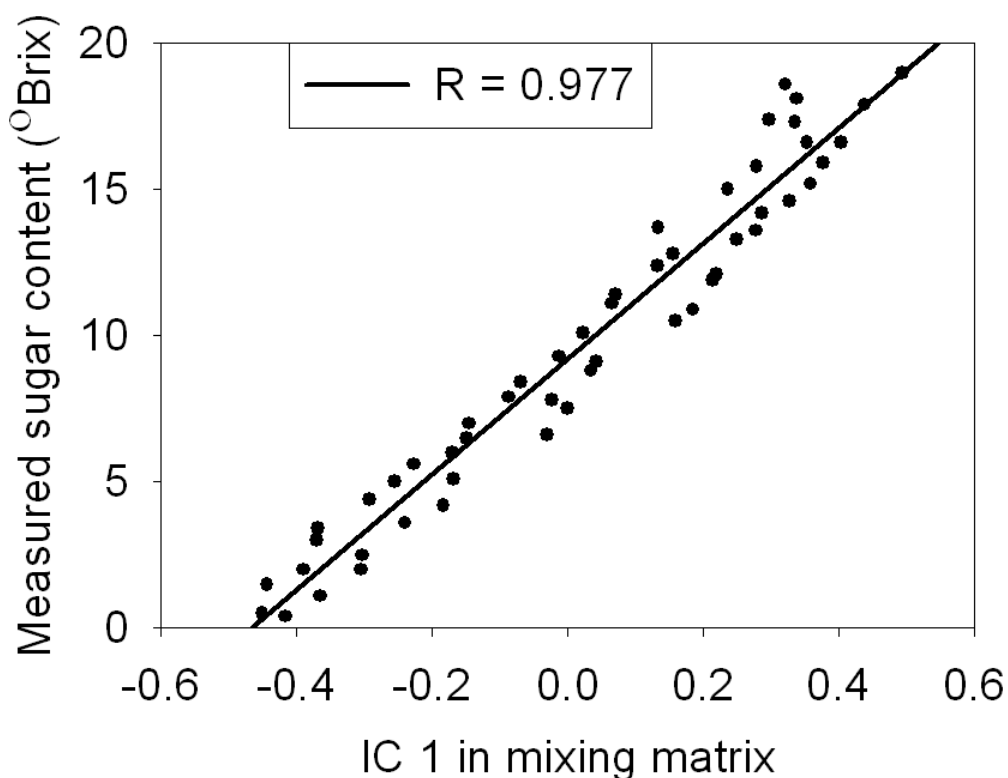
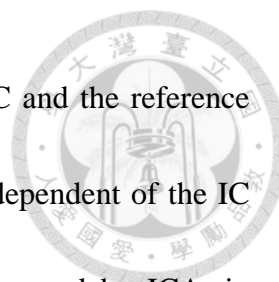


Fig. 2.4 Correlation between the values of IC 1 in the mixing matrix and the reference sugar contents of sucrose solutions.

The regression coefficient matrix by the NIR spectra and the reference sugar content of calibration sets was shown in Table 2.2, 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 (Chen and Wang, 2001; Hahn and Yoon, 2006; Pasadakis and Kardamakis, 2006; Kardamakis *et al.*, 2007; Kaneko *et al.*, 2008).

Table 2.2 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



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 (Chen and Wang, 2001; Hahn and Yoon, 2006; Pasadakis and Kardamakis, 2006; Kardamakis *et al.*, 2007). 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 Fig. 2.5(A) and (B), and the reflectance spectrum of sucrose powder post-Detrend was shown in Fig. 2.5(C). The peak positions of IC 1 (964, 1090, 1436, 2100, and 2276 nm) matched the specific wavelength ranges of sugar content (C-H band) (Chang *et al.*, 1998; Park, 2003; Hahn and Yoon, 2006), which was also consistent with the absorption bands seen in Fig. 2.5(C). 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.

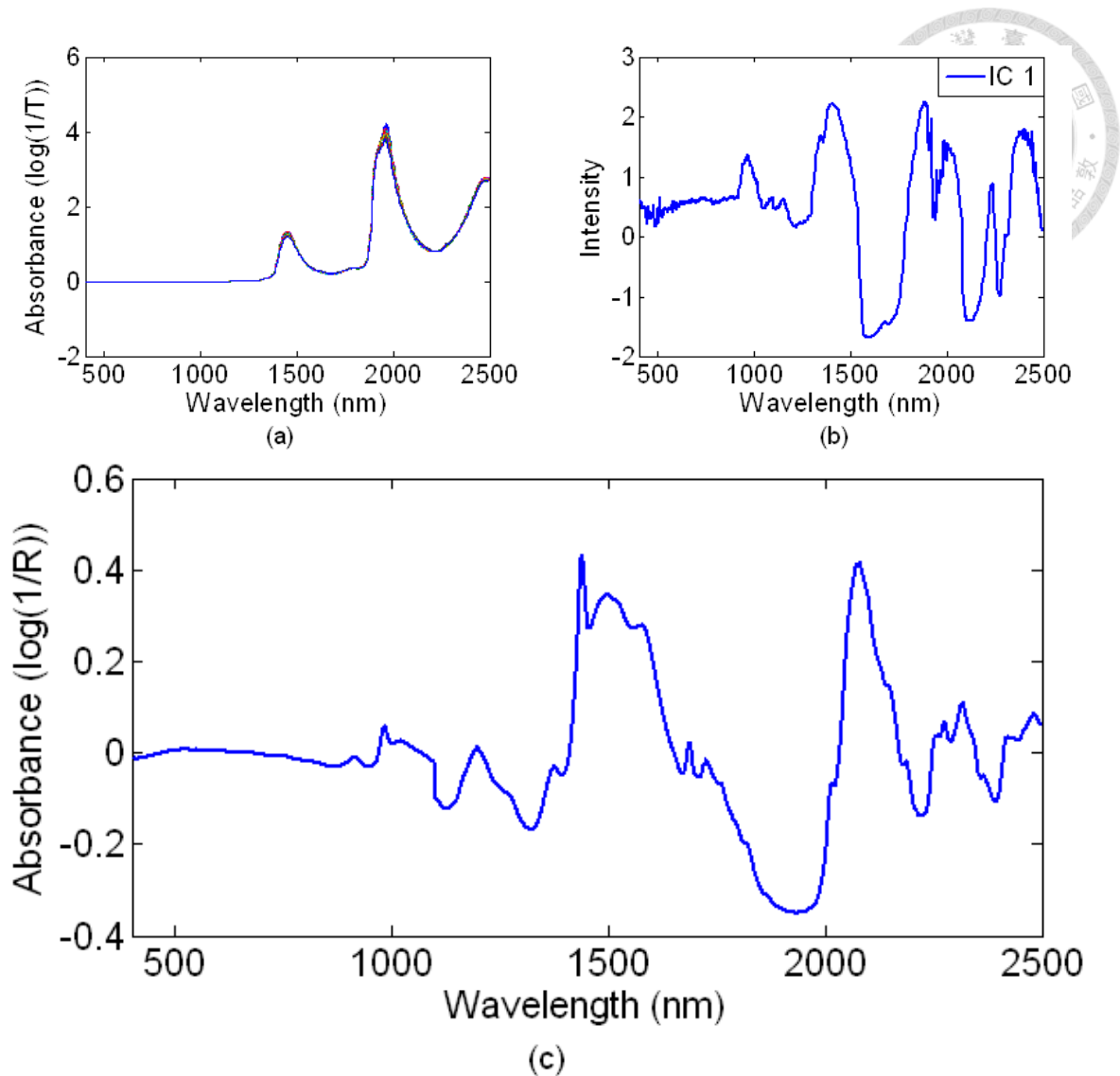
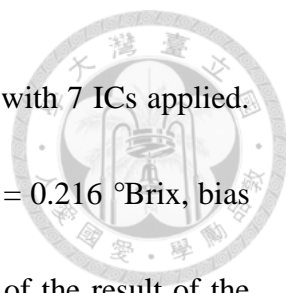


Fig. 2.5 (A) Original NIR spectra of sucrose solutions, (B) IC 1 decomposed from calibration sets, and (C) the reflectance spectrum of sucrose powder post-Detrend.

2.3.1.3 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 2498 nm. The results of spectral calibration models built by ICA are shown in Table 2.3. 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 2200 to 2498 nm contained more noises; therefore, the predictability of the spectral calibration models would decrease as derivatives were attempted.

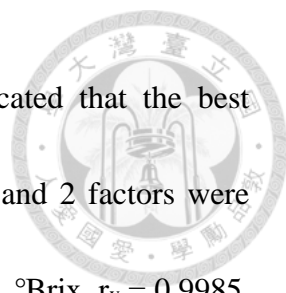


Table 2.3 Regression results by ICA and PLSR analyses for sucrose solutions.

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 st Derivative + Normalization	4	0.9994	0.193	13.71	0.9984	0.331	10.34	0.028	16.66
	2 nd 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 st Derivative + Normalization	3	0.9995	0.192	11.50	0.9950	0.546	12.92	0.031	10.10

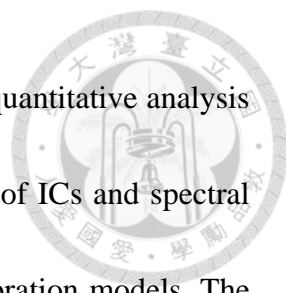
2 nd Derivative + Normalization	2	0.9990	0.243	20.96	0.9869	0.899	34.99	0.013	6.14
--	---	--------	-------	-------	--------	-------	-------	-------	------





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 2.3). Moreover, with the $SEC = 0.192$ °Brix and the $SEV = 0.546$ °Brix for the 1st derivative with normalization, and the $SEC = 0.243$ °Brix and the $SEV = 0.899$ °Brix for the 2nd derivative with normalization, it is apparent that the SEV values of both 1st and 2nd 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 2.3). 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.

2.3.2 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 2.4).

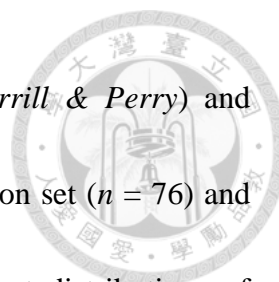
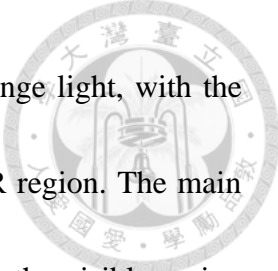


Table 2.4 Summary of wax jambu (*Syzygium samarangense* Merrill & Perry) and sample sugar contents. Total samples ($n = 114$), calibration set ($n = 76$) and validation set ($n = 38$) were arranged to have consistent distributions of sugar content.

Wax Jambu						
Group	n	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

2.3.2.1 CORRELATION ANALYSIS OF NIR SPECTRA AND SUGAR CONTENT

Fig. 2.6 showed the distribution of the correlation coefficients for the original, the 1st derivative and the 2nd derivative spectra of the wax jambu samples and their sugar contents. The main absorption wavelengths of the original spectra were 676, 968, and 1144 nm, of which 676 nm was located within the visible region of red light, whereas 968 and 1144 nm in the NIR region, belonging to the 2nd overtone of the C-H bond. The main absorption wavelengths of the 1st derivative spectra were 626, 974, 1070, and



1406 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 2nd derivative spectra were located in the visible region between orange light and red light, namely 594, 642, and 692 nm. Fig. 2.6 showed that the wavelength range of 600 to 1098 nm was the major absorption band, and the 1st derivative spectra were most significantly correlated to the sugar content (Chung *et al.*, 2004). 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.

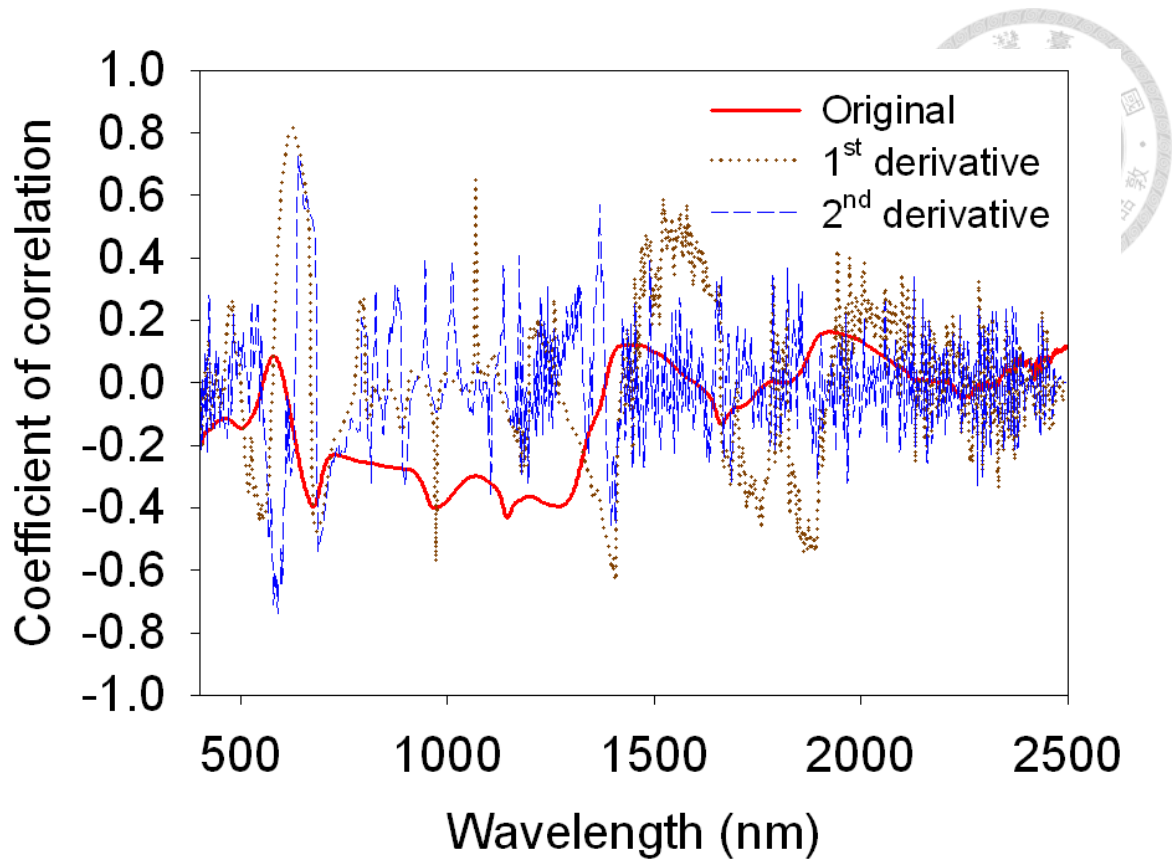



Fig. 2.6 Correlation coefficient distributions of the spectra and the sugar content of wax jambu through three different spectral pretreatments (original spectra, 1st derivative spectra, and 2nd derivative spectra).

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 2498 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 1450 and 1900 nm that represent primarily water absorption. When applying 7 to 30 ICs (Fig. 2.7), the SEV values in the ranges of 600 to 700 nm and 800 to 1098 nm were less than 1 °Brix, so were the results of the 1st and the 2nd derivative spectra. All three spectra fitted the spectral bands of higher correlation in Fig. 2.6, 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 1098 nm (Chung *et al.*, 2004).

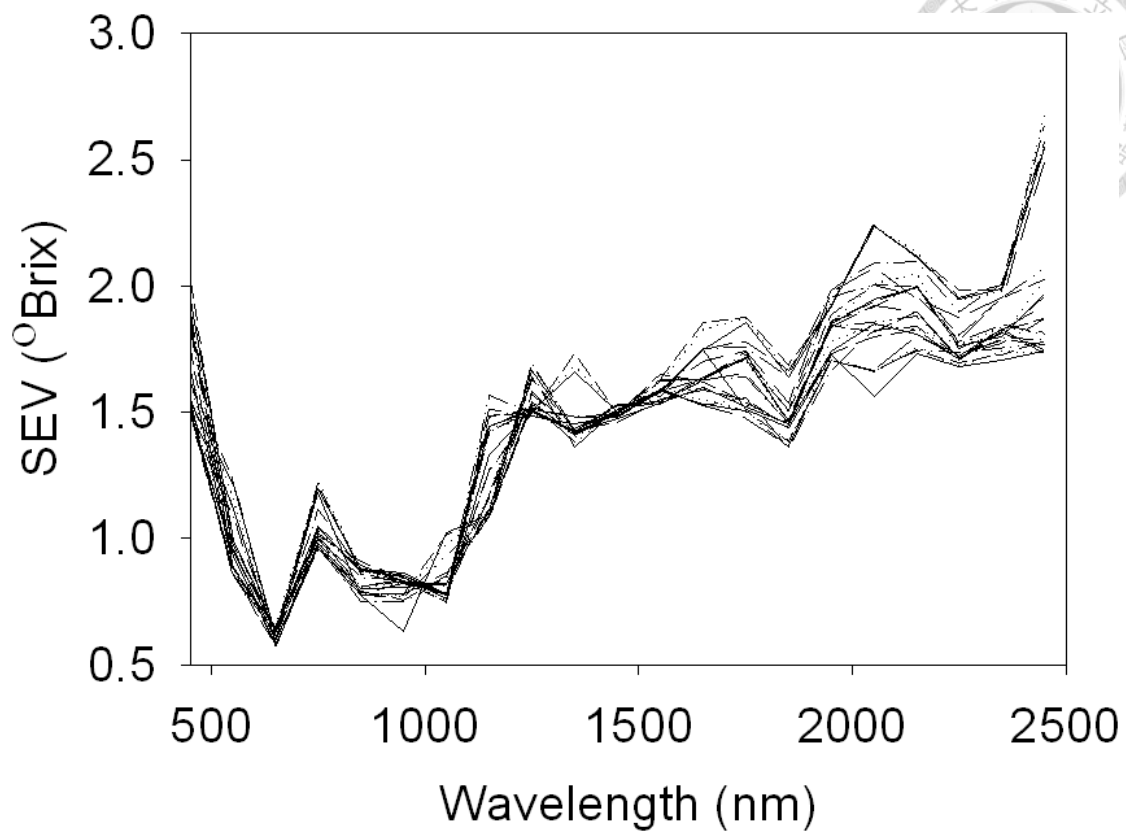
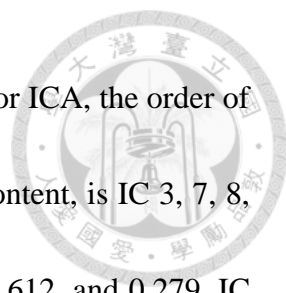


Fig. 2.7 Relationship between spectral bands and errors of the predicted sugar content for wax jambu when applying 7 to 30 ICs. Full spectrum range from 400 to 2498 nm was divided into 21 band regions by taking every 100 nm as a band region.

2.3.2.2 SUGAR CONTENT QUANTIFICATION BASED ON ICA AND PLSR

2.3.2.2.1 ANALYSIS WITHOUT SPECTRAL PRETREATMENT

The ICA results of the spectral calibration model for wax jambu are shown in Table 2.5. The best spectral calibration model was found with the normalized 1st derivative spectra and 10 ICs, resulting in $R_c = 0.956$, $SEC = 0.471$ °Brix, $r_v = 0.954$, $SEV = 0.489$




$^{\circ}\text{Brix}$, bias = -0.013 $^{\circ}\text{Brix}$, 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) (Moneruzzaman *et al.*, 2011; Tehrani *et al.*, 2011) 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 1098 nm, the spectra covered the 3rd overtone of C-H bond, conforming to the results of Fig. 2.6 and 2.7. 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 Fig. 2.7. 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 2.5, with the best spectral calibration model found in the normalized original spectra with 5 factors, yielding $R_c = 0.884$, $\text{SEC} = 0.753$ $^{\circ}\text{Brix}$, $r_v = 0.867$, $\text{SEV} = 0.816$ $^{\circ}\text{Brix}$, and bias = 0.238 $^{\circ}\text{Brix}$. The specific wavelength regions used were within the wavelength range of 600 to 700 nm and 800 to 1098 nm, consistent with the aforementioned results.



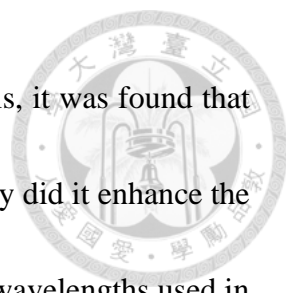
Table 2.5 Regression results by ICA and PLSR analyses for wax jambu (without spectral pretreatment).

Method	Spectrum	Wavelength Range (nm) ^a	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	7	0.930	0.591	6.37	0.919	0.642	7.11	-0.024	2.53
		1000-1098, 2									
	Original + Normalization	600-700, 2	10	0.948	0.515	5.51	0.940	0.568	5.73	0.054	2.86
		800-1000, 2									
	1 st Derivative + Normalization	600-700, 2	10	0.956	0.471	4.88	0.954	0.489	4.98	-0.013	3.32
		900-1098, 2									



	2 nd Derivative +	600-700, 2									
			10	0.943	0.535	5.85	0.944	0.553	5.77	-0.017	2.94
	Normalization	800-1000, 2									
	Original	800-1098, 2	2	0.428	1.455	15.56	0.237	1.599	15.80	0.351	1.02
	Original +	600-700, 2									
			5	0.884	0.753	8.11	0.867	0.816	8.90	0.238	1.99
	Normalization	800-1098, 2									
PLSR	1 st Derivative +	600-700, 2									
			3	0.876	0.777	8.44	0.880	0.782	8.89	-0.020	2.08
	Normalization	800-1098, 2									
	2 nd Derivative +	600-700, 2									
			2	0.797	0.973	11.58	0.803	0.975	10.95	0.182	1.67
	Normalization	800-1000, 2									

^a Interval is 2 nm.



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.

2.3.2.2.2 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 2.6, with the normalized 1st 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, $r_v = 0.963$, $SEV = 0.426$ °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.

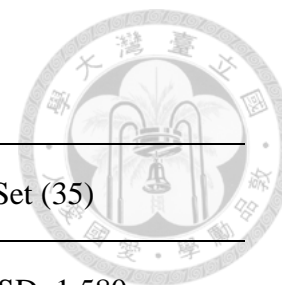
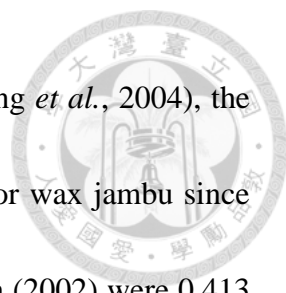


Table 2.6 Regression results by ICA and PLSR analyses for wax jambu (with spectral pretreatment).

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

^a Interval is 2 nm.

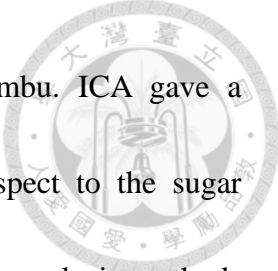


Compared to the previous literatures (You, 2002; Lin, 2002; Chung *et al.*, 2004), the spectral calibration models built by ICA had higher predictability for wax jambu since the SEC values reported by You (2002), Chung *et al.* (2004) and Lin (2002) were 0.413 °Brix, 0.388 °Brix and 0.252 °Brix, respectively. 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 2.6 were better than those reported by Chung *et al.* (2004) and Lin (2002) in terms of R_c , SEC, r_p 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.

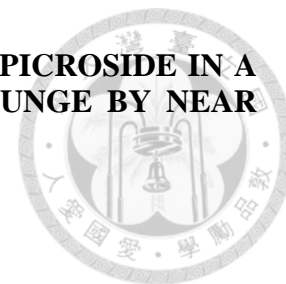
2.4 CONCLUSIONS

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.

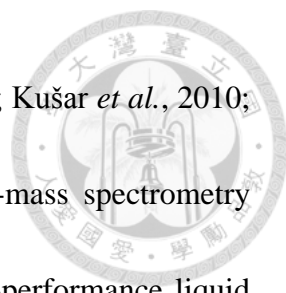
CHAPTER 3. QUANTIFICATION OF BIOACTIVE GENTIOPICROSIDE IN A MEDICINAL PLANT *GENTIANA SCABRA* BUNGE BY NEAR INFRARED SPECTROSCOPY



3.1 INTRODUCTION

Dried root and rootstock of *Gentiana scabra* Bunge are commonly used as pharmaceutical raw materials, since they are rich in many secoiridoid glycosides, such as gentiopicroside, swertiamarin and sweroside (Kakuda *et al.*, 2001). In particular, gentiopicroside has been shown to protect liver, inhibit liver dysfunction, and promote gastric acid secretion in addition to its antimicrobial and anti-inflammatory effects, making it a popular ingredient in Chinese herbal medicine and health products (Kim *et al.*, 2009).

In early days, *G. scabra* Bunge was mainly collected in the wild. As the demand for *G. scabra* Bunge increases and the wild resources diminish, the restoration of *G. scabra* Bunge thus became an important issue (Zhang *et al.*, 2010). Studies in recent years used tissue culture technology to artificial cultivate *G. scabra* Bunge (Cai *et al.*, 2009), by domesticating the tissue culture of *G. scabra* Bunge, then transplanting it to the greenhouse for cultivation. In order to monitor the change of *G. scabra* Bunge during the growth process, it is necessary to measure the bioactive components of *G. scabra* Bunge. However, the commonly used methods, such as micellar electrokinetic capillary chromatography (MECC) (Glatz *et al.*, 2000), high performance liquid chromatography



(HPLC) (Szücs *et al.*, 2002; Kikuchi *et al.*, 2005; Carnat *et al.*, 2005; Kušar *et al.*, 2010; Hayta *et al.*, 2011a; Hayta *et al.*, 2011b), liquid chromatography-mass spectrometry (LC-MS) (Aberham *et al.*, 2007; Aberham *et al.*, 2011), and ultra-performance liquid chromatography (UPLC) (Nastasijević *et al.*, 2012), are all time-consuming and energy-intensive, hence not applicable for daily quality inspection of *G. scabra* Bunge during cultivation.

Near infrared (NIR) spectroscopy is a nondestructive inspection method that can rapidly measure the target object. Its spectrum contains characteristic spectral information of the internal constituents in the sample, so it has been widely used in dispensation, such as herbal component analysis of Chinese herbal plants *Angelicae gigantis Radix* (Woo *et al.*, 2005), *Rhubarb* (Zhang and Tang, 2005), *licorice* (*Glycyrrhizia uralensis* Fisch.) (Wang *et al.*, 2007), *Panax* species (Chen *et al.*, 2011), and *Lonicera japonica* (Wu *et al.*, 2012), as well as the content detection of active pharmaceutical ingredients (APIs) in tablets (Paris *et al.*, 2006; Jamrógiewicz, 2012; Porfire *et al.*, 2012). However, it has not been employed to qualitatively monitor the growth of *G. scabra* Bunge.

The present study was aimed to explore the NIR feature of gentiopicroside, the

bioactive component of *G. scabra* Bunge, in order to build the spectral calibration models. Moreover, the applicability of silicon CCD sensing band when using multi-spectral imaging technology to inspect the quality of *G. scabra* Bunge was evaluated.



3.2 MATERIALS AND METHODS

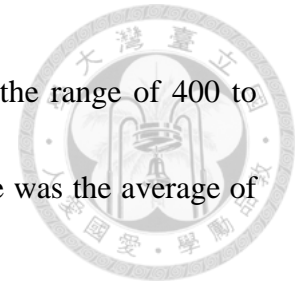
3.2.1 G SCABRA BUNGE SAMPLE PREPARATION

The samples of *G. scabra* Bunge were provided by Taiwan Sugar Research Institute (TSRI; Tainan, Taiwan). A total of 94 tissue culture samples and 68 grown plant samples of different cultivation time was acquired (Yang *et al.*, 2008; Cheng, 2009). The shoot and root of the grown plant samples were measured separately in order to compare their differences. The *G. scabra* Bunge samples were first dried for 48 hours in a dryer (50 °C), then milled with a high speed grinder (RT-02A, Sun-Great Technology Co., Ltd., New Taipei City, Taiwan). The dried powder was filtered with a 100 mesh sieve and stored in amber sample vials to avoid light exposure (Yang *et al.*, 2008; Cheng, 2009).

3.2.2 NIR SPECTRA AND HPLC MEASUREMENT

Dry powder of *G. scabra* Bunge was gently poured into a small ring cup (i.d. 5 cm) and subjected to NIR measurement (NIRS 6500, FOSS NIRSystems, Inc., Laurel, MD,

U.S.A.). The reflectance spectra of the samples were collected in the range of 400 to 2498 nm with 2 nm intervals, and the NIR spectrum of each sample was the average of 32 scans (Yang *et al.*, 2008; Cheng, 2009).



To attain the reference value of the bioactive component, gentiopicroside was measured by HPLC (DX 500 ion chromatograph, Dionex Corporation, Sunnyvale, CA, U.S.A.) equipped with a DIONEX C18 column (250 mm × 4.6 mm i.d.). The peak of gentiopicroside appeared at 250 nm when methanol : water (20:80) was used as the mobile phase at a flow rate of 1 mL/min. A high-precision scale was used to measure the gentiopicroside standard powder, and diluted into 1000, 500, and 250 ppm with 70% methanol as the standard solutions for the three-point calibration of HPLC. A quantitative linear relationship was established between the standard concentration and the peak area (Yang *et al.*, 2008; Cheng, 2009).

3.2.3 DATA ANALYSIS

In order to apply the specific wavelengths identified to multi-spectral imaging inspection of *G. scabra* Bunge, the spectra of the full wavelength range (400 to 2498 nm) and the silicon CCD sensing band (400 to 1098 nm) were analyzed. Modified partial least squares regression (MPLSR) and stepwise multiple linear regression (SMLR)

methods were employed to build the calibration models of gentiopicroside.

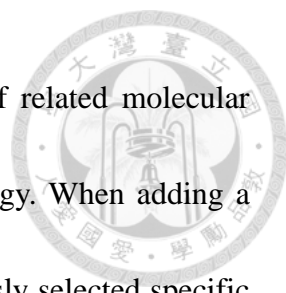


3.2.3.1 MPLSR

An extension of partial least squares regression (PLSR), MPLSR abides by the principle of normalization of the spectra and constituent values prior to PLSR, which is a standard tool in chemometrics and has been widely used in the pharmaceutical, chemical, and agricultural fields (Wold *et al.*, 2001). 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, i.e., the more important factor is ranked earlier in the order. Since PLSR analysis uses information from spectral bands, the analysis results can be improved by selecting appropriate number of factors and specific wavelength ranges.

3.2.3.2 SMLR

SMLR selects the specific wavelengths according to the F-test ($F \geq 3$) of null hypothesis testing (Chang *et al.*, 1998). To build the calibration model with numerous wavelengths, the SMLR algorithm chooses the most important specific wavelength from the major molecular bonding region of the objects, and the second most important



specific wavelength is usually chosen between the combination of related molecular bonding, or the overtone of complementary bonding, and by analogy. When adding a new wavelength for training, the algorithm will base on the previously selected specific wavelengths to continue finding the wavelength, which can allow the highest multiple coefficient of determination (r^2) and the minimum prediction error, and determine whether such wavelength can replace the current specific wavelength or not. In case of poor competency of the newly-added wavelength for training, the algorithm will stop training.

3.2.3.3 SPECTRAL PRETREATMENTS

The purpose of spectral pretreatments was to eliminate the spectral variation not caused by chemical information contained in the samples (de Noord, 1994; Fearn, 2001). Since inevitable light scattering could be added into the spectra when using NIR to measure powder samples, especially when the particle size is not uniform, multiplicative scatter correction (MSC) was used to allow additive and multiplicative transformation of the spectra (Eq. 3.1). It was conducted using the average spectrum of all samples as the reference value, and calculating the parameters a and b with the least square. After MSC treatment, the spectra of *G. scabra* Bunge powder not only reduced the physical impact of non-uniform particles (Helland *et al.*, 1995; Maleki *et al.*, 2007),



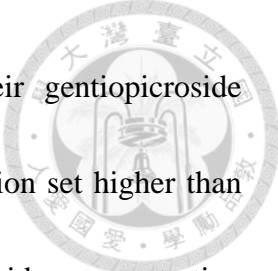
but also confirmed the linearity of the spectral information (Isaksson and Næs, 1988), which would contribute to subsequent linear regression analysis (Thennadil *et al.*, 2006).

$$x_{ik}(new) = \frac{[x_{ik}(old) - a_i]}{b_i} \quad (3.1)$$

The spectra of *G. scabra* Bunge powder post-MSC was subjected to three independent treatments, namely (1) smoothing; (2) smoothing with 1st derivative; and (3) smoothing with 2nd derivative, in order to choose the best pretreatment parameters, including the smoothing points and the gap ranging from 2 to 50, with the gap being greater than or equal to the smoothing points.

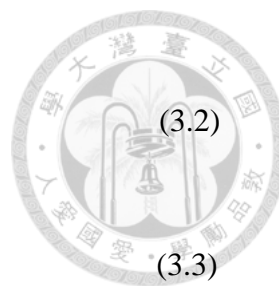
3.2.3.4 MODEL ESTABLISHMENT

The spectral calibration models of MPLSR and SMLR were built by WinISI II chemometric software (Infrasoft International, LLC., Port the Matilda, PA, U.S.A.). The MPLSR analysis procedure included: (1) spectral pretreatments; (2) selecting the specific wavelength regions; (3) selecting calibration set and validation set; and (4) determining best calibration model. In steps 1 and 2, 3-fold cross validation (CV) was used to enable objective selection of the parameters. A 2:1 ratio of calibration to validation samples was adopted according to the gentiopicroside concentration in the



sample. All samples were ranked ascendantly according to their gentiopicroside concentration, with the gentiopicroside concentration in the calibration set higher than the validation set, yet both sets contained similar gentiopicroside concentration distribution of all samples. When selecting the best calibration model, in order to avoid over-fitting caused by use of excessive factors, the following principles were adhered to: (1) the maximum number of factors is one tenth of the number of calibration sets + 2 to 3; (2) stop if the adding of a new factor makes the SEV rise; and (3) when the SEV is lower than the SEC, stop adding new factor. The SMLR analysis procedure was: (1) selecting calibration set and validation set; (2) spectral pretreatments; and (3) determining best calibration model and the specific wavelengths. The same calibration and validation sets were used for both MPLSR and SMLR analyses.

After the respective spectral calibration models of MPLSR and SMLR were built, these models were then used to predict the gentiopicroside concentration of the calibration and the validation set. The predictability of the models was evaluated based on the following statistical parameters, including coefficient of correlation of calibration set (R_c), standard error of calibration (SEC), standard error of validation (SEV), bias and the ratio of the standard error of performance to the standard deviation of the reference values (RPD), as defined below:



$$SEC = \left[\frac{1}{n_c} \sum_{i=1}^{n_c} (Y_r - Y_c)_i^2 \right]^{1/2} \quad (3.2)$$

$$SEV = \left\{ \frac{1}{n_v} \sum_{i=1}^{n_v} [(Y_r - Y_v) - \text{Bias}]_i^2 \right\}^{1/2} \quad (3.3)$$

$$\text{Bias} = \frac{1}{n_v} \sum_{i=1}^{n_v} (Y_r - Y_v)_i \quad (3.4)$$

$$RPD = SD/SEV \quad (3.5)$$

where Y_c and Y_v represent the estimated gentiopicroside concentration of the calibration set and the validation set, respectively. Y_r is the reference gentiopicroside concentration; n_c and n_v are the number of samples in the calibration set and validation set, respectively; SD is the standard deviation of gentiopicroside concentration within the validation set.

3.3 RESULTS AND DISCUSSION

3.3.1 GENTIOPICROSIDE CONCENTRATION AND DISTRIBUTION IN *G*

SCABRA BUNGE

The gentiopicroside contents in different parts (94 tissue culture, 68 shoot and 68 root samples) of *G. scabra* Bunge samples are shown in Table 3.1. It can be seen that the gentiopicroside content in *G. scabra* Bunge whole grown plant (including shoot and root) increased after *G. scabra* Bunge tissue culture was transplanted into the greenhouse for cultivation. Within the grown plant itself, the gentiopicroside content was significantly

higher in root than in shoot, indicating that during greenhouse cultivation, the gentiopicroside was mainly stored in the root.

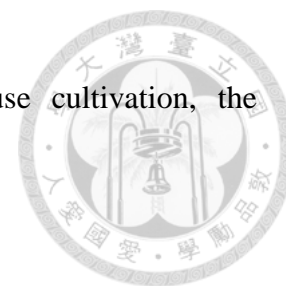


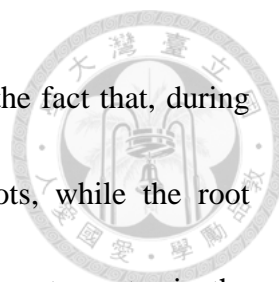
Table 3.1 The gentiopicroside content in tissue culture and grown plants of *G. scabra*

Bunge.

Sample	n	Gentiopicroside Content (%)				
		Min.	Max.	Mean	SD	CV
Tissue Culture	94	2.69	8.18	5.35	1.29	0.24
Grown Plant						
Shoot	68	1.34	5.90	3.26	0.91	0.28
Root	68	2.24	8.77	4.68	1.62	0.35

3.3.2 CORRELATION BETWEEN NIR SPECTRA AND GENTIOPIROCOSIDE CONTENT

The NIR spectra of the 94 *G. scabra* Bunge tissue culture samples and the 136 grown plant samples (68 shoot and 68 root) were acquired using the MSC treatment. As shown in Fig. 3.1(A) and 3.1(B), there were absorption peaks in both visible region of blue light (452 nm) and red light (666 nm) because the chlorophyll in *G. scabra* Bunge would absorb the majority of blue and red light during photosynthesis. The spectra of



tissue culture and shoot were similar, which could be attributed to the fact that, during the domestication period, the tissue is mainly composed of shoots, while the root development of *G. scabra* Bunge is not obvious at that time. The root spectra in the visible region showed a significant difference, with high absorption from green to yellow light (492 to 586 nm) and low absorption (flat waveform) from orange to red light (606 to 700 nm). This could be due to lack of chlorophyll in the roots of *G. scabra* Bunge plant, which reduces absorption of blue and red light, and reflects green light.

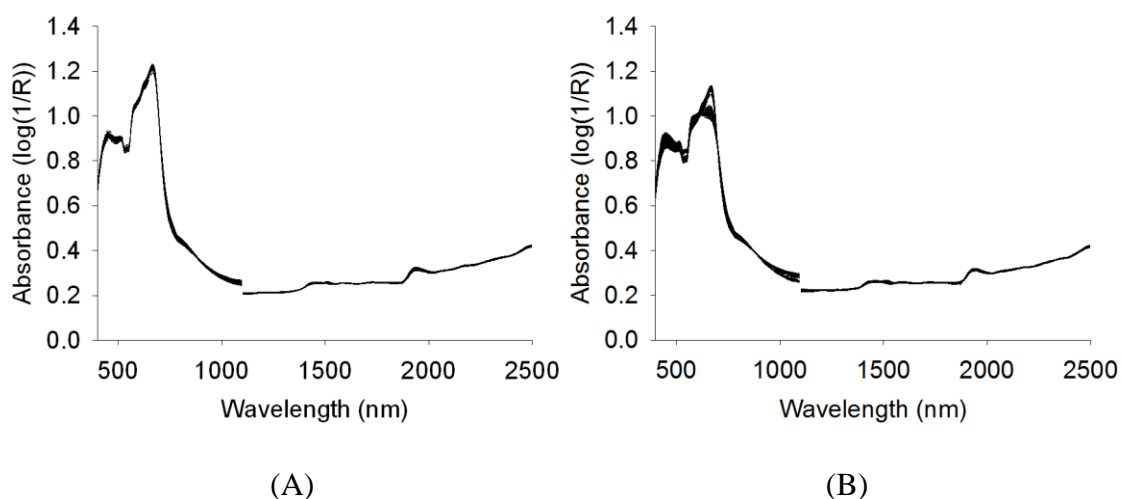
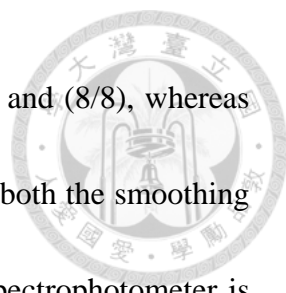


Fig. 3.1 The spectra of *G. scabra* Bunge powder post-MSC (A) tissue culture and (B) grown plants.

After MSC treatment, the spectra of *G. scabra* Bunge tissue culture and grown plant were analyzed using the following pretreatments: (1) smoothing; (2) smoothing with 1st derivative; and (3) smoothing with 2nd derivative. The best pretreatment parameters of



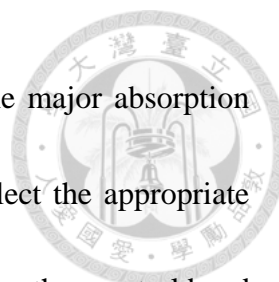
the tissue culture spectra (smoothing points / gap) were (1/0), (6/6) and (8/8), whereas the best ones of the grown plant spectra were (1/0), (2/2) and (3/3); both the smoothing points and the gap were less than 10, indicating that NIRS 6500 spectrophotometer is stable, and the spectra of *G. scabra* Bunge powder exhibits minimal noise.

The correlation between the spectra and gentiopicroside of *G. scabra* Bunge powder was analyzed before selecting the specific wavelength regions. The gentiopicroside correlation coefficient distribution of *G. scabra* Bunge tissue culture samples and grown plant samples were compared using the original spectra, 1st derivative spectra, and 2nd derivative spectra, and the threshold value ($|r| > 0.55$) was set to determine the degree of correlation. It is unnecessary to avoid the O-H bond absorption band around 1450 nm and 1900 nm because the influence of water absorption on the spectra of *G. scabra* Bunge powder has been eliminated. Fig. 3.2(A) shows that the bands of high correlation between the spectra and gentiopicroside of tissue culture were mainly distributed in the NIR region, with only a few in the visible region. The absorption bands of the original spectra were located in the 1st overtone of the C-H bond and C-C bond, whereas the absorption bands of the 1st derivative spectra were located in the orange light and the combination of the 1st overtone of C-H bond. Moreover, the absorption bands of the 2nd derivative spectra were found to locate in the 2nd overtone of C=O bond stretch.



The correlation coefficient distributions between absorbance values of the spectra and gentiopicroside contents of the *G. scabra* Bunge grown plants were also compared using the original spectra, the 1st derivative spectra, and the 2nd derivative spectra (Fig. 3.2(B)).

It can be seen that there were highly correlated bands in both visible region and the NIR region. The absorption bands in the original spectra were located between the yellow and orange light, as well as the combination of two C-H bonds. The absorption bands in the 1st derivative spectra were located between the orange and red light, the 4th overtone of C-H bond, the 3rd overtone of C-H bond, the 1st overtone of C-H bond, and the combination of two C-H bonds, whereas the absorption bands of the 2nd derivative spectra were located in the blue and red light, the 3rd overtone of N-H bond, and the combination of two C-H bonds. Because the spectra of shoot and root showed obvious differences in the visible region, the correlation of blue and red light to gentiopicroside was improved, indicating that the amount of chlorophyll contained in different parts of the grown plant also affect the performance of the specific wavelength regions. The specific wavelength regions of both tissue culture and grown plant in the NIR region were located in the combination of two C-H bonds and the overtones of C-H bond, indicating the C-H bonds are the main absorption of NIR. According to the absorption bands of the C-H bonds in the spectrum, Fig. 3.2 showed that the wavelength range of



900 to 1300 nm, 1500 to 1800 nm, and 2200 to 2300 nm were the major absorption bands, and these wavelengths can be used to provide a basis to select the appropriate specific wavelength regions when conducting MPLSR analysis. As for the spectral band 400 to 650 nm, which belonged to the absorption band of blue to red light, color information was also reflected in the spectra.

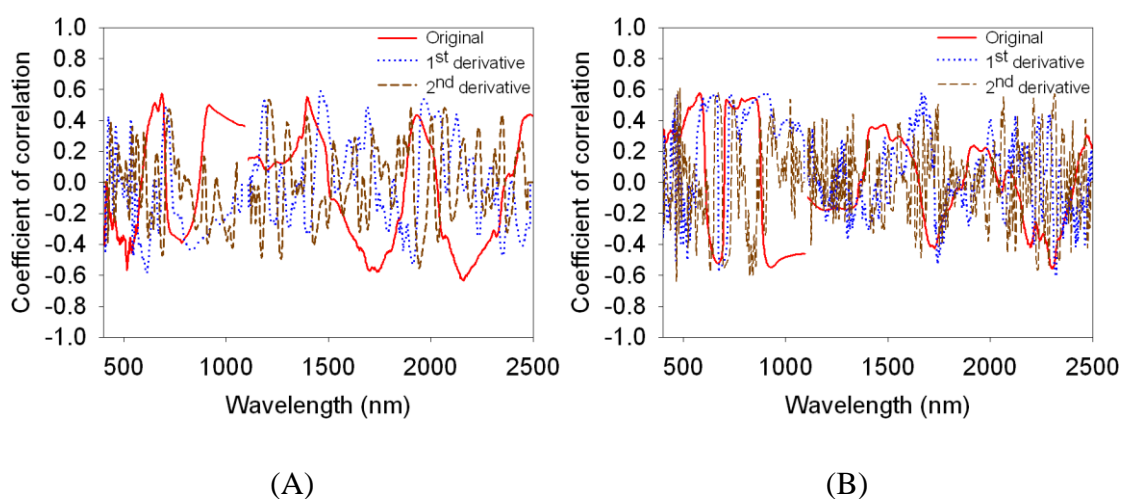


Fig. 3.2 Correlation coefficient distributions between absorbance values of the spectra and gentiopicroside contents of the *G. scabra* Bunge powder (A) tissue culture and (B) grown plants.

3.3.3 GENTIOPICROSIDE QUANTIFICATION USING SPECIFIC WAVELENGTH RANGES

Out of the 89 and 126 valid *G. scabra* Bunge tissue culture and grown plant samples, respectively, were retained for statistical calibration and validation of the

gentiopicroside content (Table 3.2). It was found that there was no significant difference in the mean, standard deviation, and coefficient of variation (CV) of the effective samples, the calibration set, and the validation set, indicating that the gentiopicroside content distribution of the two sample groups were consistent.

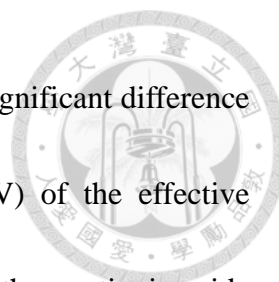


Table 3.2 The gentiopicroside content of effective samples, calibration set, and validation set in tissue culture and grown plants.

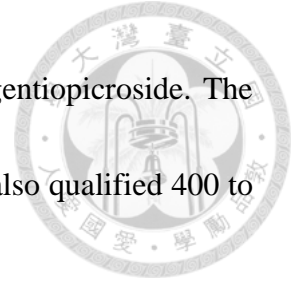
Sample	n	Gentiopicroside Content (%)				
		Min.	Max.	Mean	SD	CV
Tissue Culture						
Effective Samples	89	2.69	7.83	5.26	1.19	0.23
Calibration Set	60	2.69	7.83	5.26	1.22	0.23
Validation Set	29	3.12	7.35	5.26	1.14	0.22
Grown Plant						
Effective Samples	126	1.34	8.77	4.01	1.51	0.38
Calibration Set	84	1.34	8.77	4.01	1.52	0.38
Validation Set	42	1.59	8.19	4.01	1.50	0.37

The MPLSR analysis results of FWR spectra (400 to 2498 nm) were shown in Table

3.3. The best calibration model of *G. scabra* Bunge tissue culture was found with the 1st derivative spectra and 6 factors, with both smoothing points and gap set at 6, using the wavelength range of 900 to 1000 nm, 1200 to 1300 nm, and 1600 to 1700 nm, and resulting in $R_c = 0.868$, $SEC = 0.606\%$, $SEV = 0.862\%$, $bias = -0.215\%$, and $RPD =$

1.32. Due to the spectral difference between the calibration set and the validation set, the prediction result of the validation set was a little worse than the calibration set when using the calibration model. The best calibration model of *G. scabra* Bunge grown plant was identified when using the 2nd derivative spectra and 5 factors, with both smoothing points and gap set at 3, using the wavelength range of 400 to 500 nm, 1100 to 1200 nm, 1600 to 1800 nm, and 2200 to 2300 nm. The results were $R_c = 0.944$, $SEC = 0.502\%$, $SEV = 0.685\%$, $bias = -0.162\%$, and $RPD = 2.19$. The calibration models built based on 1st derivative spectra and 2nd derivative spectra were both better than those based on the original spectra, indicating that the heterogeneous particles of *G. scabra* Bunge powder really affect the spectral absorption. The calibration models of grown plant were all better than those of the tissue culture, even with fewer spectral pretreatments, because more grown plant samples can build more stable calibration models. The specific wavelength regions of tissue culture and grown plant were mainly distributed in 900 to 1300 nm and 1600 to 1800 nm, and the calibration model of grown plant also incorporated the spectral information within 400 to 500 nm and 2200 to 2300 nm,

indicating that the NIR region contained more information about gentiopicroside. The absorption differences between shoot and root in the visible region also qualified 400 to 500 nm employable as a specific wavelength region.



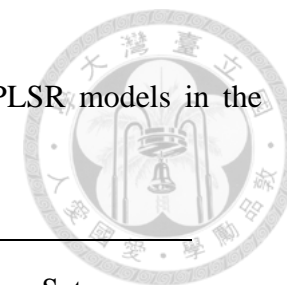


Table 3.3 Prediction of the gentiopicroside content in tissue culture and grown plants of *G. scabra* Bunge by MPLSR models in the wavelength of 400 to 2498 nm.

Sample	Spectrum	Wavelength Range (nm) ^a	Smoothing Points / Gap	Factors	Calibration Set		Validation Set		
					R _c	SEC (%)	SEV (%)	bias (%)	RPD
Tissue Culture	Original	900-1000, 2 1600-1700, 2	1 / 0	5	0.752	0.804	0.943	-0.137	1.21
	1 st Derivative	900-1000, 2 1200-1300, 2	6 / 6	6	0.868	0.606	0.862	-0.215	1.32
	2 nd Derivative	500-600, 2 1050-1098, 2	8 / 8	4	0.852	0.638	0.830	-0.123	1.37
	1100-1300, 2								



		1550-1750, 2							
Grown Plant	Original	400-500, 2							
		1600-1700, 2	1 / 0	7	0.881	0.717	0.775	-0.054	1.94
			2200-2300, 2						
	1 st Derivative	400-500, 2							
		1100-1200, 2	2 / 2	5	0.919	0.597	0.726	-0.141	2.07
			1600-1700, 2						
			2200-2300, 2						
	2 nd Derivative	400-500, 2							
		1100-1200, 2	3 / 3	5	0.944	0.502	0.685	-0.162	2.19
			1600-1800, 2						
		2200-2300, 2							

3.3.4 GENTIOPIICOSIDE QUANTIFICATION USING CCD CAMERA WAVELENGTH SPECTRA



MPLSR analysis of silicon CCD sensing bands (400 to 1098 nm) was shown in Table

3.4. The best calibration model of *G. scabra* Bunge tissue culture was acquired when the 2nd derivative spectra and 3 factors were employed, where both smoothing points and gap were at 2, with wavelength range of 400 to 500 nm and 800 to 1000 nm, and the results were $R_c = 0.865$, $SEC = 0.611\%$, $SEV = 0.772\%$, of bias = 0.025%, $RPD = 1.47$.

The best calibration model of *G. scabra* Bunge grown plant was found with the 1st derivative spectra and 5 factors, smoothing points and gap at 2, with wavelength range of 400 to 600 nm and 900 to 1098 nm, resulting in $R_c = 0.904$, $SEC = 0.649\%$, $SEV = 0.724\%$, bias = -0.089%, $RPD = 2.08$. Regardless of the samples being tissue culture or grown plant, the calibration models built based on 1st derivative spectra and 2nd derivative spectra were better than those based on the original spectra, indicating that spectral pretreatments indeed enhanced the predictability of the calibration models. The spectral calibration models of grown plant were all better than those of the tissue culture with fewer spectral pretreatments, which was consistent with the results shown in Table 3.3. The specific wavelength regions of tissue culture and grown plant were mainly distributed in 400 to 600 nm (blue and red light) and 800 to 1098 nm (the 2nd and 3rd overtone of C-H bond). The absorption capacity of these bands was a little lower than

the combination and the 1st overtone of C-H bond, producing a fewer spectral absorption performances of gentiopicroside, so the predictability declined slightly when using silicon CCD sensing band to build the calibration models.



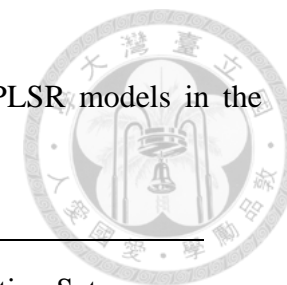



Table 3.4 Prediction of the gentiopicroside content in tissue culture and grown plants of *G. scabra* Bunge by MPLSR models in the wavelength of 400 to 1098 nm.

Sample	Spectrum	Wavelength Range (nm) ^a	Smoothing Points / Gap	Factors	Calibration Set		Validation Set		
					R _c	SEC (%)	SEV (%)	bias (%)	RPD
Tissue Culture	Original	550-650, 2	9 / 0	4	0.704	0.866	1.084	-0.047	1.05
		900-1050, 2							
	1 st Derivative	600-700, 2	6 / 6	6	0.764	0.786	0.906	-0.061	1.26
		900-1000, 2							
	2 nd Derivative	400-500, 2	2 / 2	3	0.865	0.611	0.772	0.025	1.47
		800-1000, 2							
Grown Plant	Original	400-600, 2	1 / 0	5	0.840	0.823	1.089	-0.015	1.38
		950-1050, 2							

1 st Derivative	400-600, 2	2 / 2	5	0.904	0.649	0.724	-0.089	2.08
	900-1098, 2							
2 nd Derivative	400-650, 2	3 / 3	3	0.888	0.697	0.750	-0.100	2.00
	950-1098, 2							



^a Interval is 2 nm



In addition, the SMLR analysis results of silicon CCD sensing band (400 to 1098 nm) was shown in Table 3.5. The best calibration model of *G. scabra* Bunge tissue culture was found when the 2nd derivative spectra were used. Both smoothing points and gap were at 3, with the specific wavelength of 846 nm and 932 nm, of which $R_c = 0.750$, $SEC = 0.806\%$, $SEV = 0.990\%$, $bias = 0.270\%$, $RPD = 1.15$ were achieved. The best calibration model of grown plant was attained when the 2nd derivative spectra were employed, where both smoothing points and gap were set at 3, in the combination of the 4 wavelengths of 670 nm, 786 nm, 474 nm and 826 nm, of which $R_c = 0.860$, $SEC = 0.775\%$, $SEV = 0.848\%$, $bias = -0.134\%$, $RPD = 1.77$ were achieved. The calibration models built based on the 1st and 2nd derivative spectra were all better than those based on the original spectra, indicating that the spectral pretreatments reduced the noise influence and made the combination of selected wavelengths more consistent when the number of wavelengths increased. The specific wavelengths selected in Table 3.5 were similar to those in Table 3.3 and Table 3.4, only with a small number of specific wavelengths beyond those selected through the MPLSR analysis. Since the silicon CCD sensing band contains fewer information of gentiopicroside, and SMLR built the spectral calibration model based on the combination of a small number of wavelengths, which gives less spectral information than MPLSR, so the analysis results seemed a little worse than those in Table 3.3 and Table 3.4. Compared to the tissue culture which

can only apply 2 wavelengths at most for inspection, grown plant can apply 4 wavelengths to build the calibration model, consequently improving its predictability.



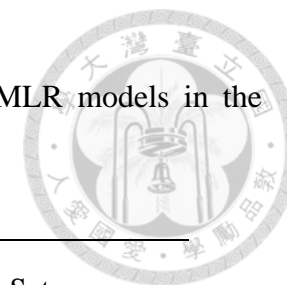


Table 3.5 Prediction of the gentiopicroside content in tissue culture and grown plants of *G. scabra* Bunge by SMLR models in the wavelength of 400 to 1098 nm.

Sample	Spectrum	Specific Wavelength (nm)	Smoothing Points / Gap	Calibration Set		Validation Set		
				R _c	SEC (%)	SEV (%)	bias (%)	RPD
Tissue Culture	Original	684	4 / 0	0.613	0.963	1.028	-0.064	1.11
		910, 512		0.643	0.934	0.999	0.000	1.14
	1 st Derivative	612	2 / 2	0.654	0.922	1.060	-0.116	1.07
	2 nd Derivative	848	3 / 3	0.632	0.946	1.016	0.376	1.12
		846, 932		0.750	0.806	0.990	0.270	1.15
Grown Plant	Original	580	2 / 0	0.588	1.227	1.249	-0.076	1.20
		690, 480		0.689	1.099	1.329	-0.112	1.13
	436, 690, 420	0.759	0.988	1.284	-0.154	1.17		

	966, 420, 408, 436		0.802	0.906	1.186	-0.178	1.27
1 st Derivative	730		0.590	1.225	1.265	-0.074	1.19
	462, 676		0.725	1.044	0.889	-0.041	1.69
		2 / 2					
	684, 780, 462		0.806	0.897	0.936	0.072	1.61
	650, 780, 462, 512		0.850	0.799	0.823	0.008	1.83
2 nd Derivative	468		0.626	1.182	1.122	0.001	1.34
	460, 634		0.736	1.027	1.011	-0.250	1.49
		3 / 3					
	666, 788, 474		0.834	0.838	0.897	-0.144	1.67
	670, 786, 474, 826		0.860	0.775	0.848	-0.134	1.77





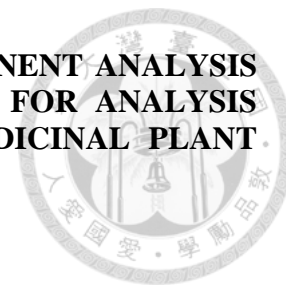
3.4 CONCLUSIONS

This study applied NIR for quantitative analysis of gentiopicroside in the medicinal plant *G. scabra* Bunge. It was found that the spectral pretreatments of MSC in combination of 2nd derivative reduced the spectral noise caused by the heterogeneous particle size of *G. scabra* Bunge powder. The specific wavelength regions or specific wavelengths selected based on their characteristic response to gentiopicroside could effectively improve the predictability of the calibration models. This study successfully built the spectral calibration models for *G. scabra* Bunge tissue culture and grown plant, which enable quantitative inspection of the bioactive component gentiopicroside in *G. scabra* Bunge during different growth stages. The specific wavelengths selected in Silicon CCD sensing band can be used as the foundation to establish a nondestructive and rapid method to assess the quality of *G. scabra* Bunge using multi-spectral imaging.

ACKNOWLEDGMENTS

This work was supported by Industrial Development Bureau, Ministry of Economic Affairs (09611101087-9601). I would like to thank Mr. Cheng-Wei Huang, Mr. Yu-Song Chen, and Mr. Chun-Chi Chen for their assistance.


CHAPTER 4. INTEGRATION OF INDEPENDENT COMPONENT ANALYSIS WITH NEAR INFRARED SPECTROSCOPY FOR ANALYSIS OF BIOACTIVE COMPONENTS IN A MEDICINAL PLANT. *GENTIANA SCABRA* BUNGE



4.1 INTRODUCTION

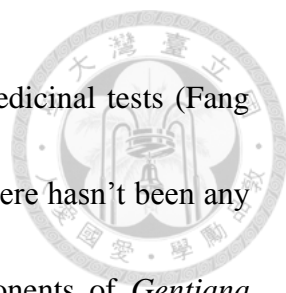
Medicinal plants have always been considered an important and reliable source of pharmacy, since they are rich in many bioactive components. The international trade market for medicinal plant products continues to expand and covers food, beverages, drugs, cosmetics, and skin care products. *Gentiana scabra* Bunge, a perennial herbaceous plant, is mainly grown in temperate regions such as Taiwan, China, Japan, South Korea, Russia, and North America. *Gentiana scabra* Bunge has been proven to achieve good effect in pharmacology, its dried root and rootstock are commonly used as pharmaceutical raw materials, since they are rich in many secondary metabolites such as gentiopicroside, swertiamarin and sweroside (Kakuda *et al.*, 2001). In particular, gentiopicroside has been shown to protect liver, inhibit liver dysfunction, and promote gastric acid secretion in addition to its antimicrobial and anti-inflammatory effects, and swertiamarin is anti-inflammatory, antiepileptic, analgesic, and sedative, making it a popular ingredient in Chinese herbal medicine and health products (Kim *et al.*, 2009).

In early days, *Gentiana scabra* Bunge was mainly collected in the wild. As the demand for *Gentiana scabra* Bunge increases, change of natural environment and climate, the wild resources diminish rapidly, thus restoration of *Gentiana scabra* Bunge



became an important issue in order to protect and sustainably utilize rare plants (Zhang *et al.*, 2010). Studies in recent years used tissue culture technology to make artificial cultivation of *Gentiana scabra* Bunge (Cai *et al.*, 2009), by domesticating the tissue culture of *Gentiana scabra* Bunge, then transplanting it to the greenhouse for cultivation. In order to monitor the change of *Gentiana scabra* Bunge during the growth process, it is necessary to measure the bioactive components of *Gentiana scabra* Bunge. However, the commonly used methods such as micellar electrokinetic capillary chromatography (MECC) (Glatz *et al.*, 2000), high performance liquid chromatography (HPLC) (Kikuchi *et al.*, 2005) and liquid chromatography-mass spectrometry (LC-MS) (Aberham *et al.*, 2011) are all time-consuming and energy-intensive, hence cannot be applicable for daily quality inspection of *Gentiana scabra* Bunge during cultivation.

Near infrared (NIR) spectroscopy is a nondestructive inspection method that has been widely used in dispensation (Zhang *et al.*, 2005; Wang *et al.*, 2007; Chen *et al.*, 2011). Generally, spectrum of a mixture is a linear combination of spectra of various components and can be considered as the 'blind sources' when the components are unknown. A fast and reliable algorithm - independent component analysis (ICA) can deal with the issue of blind source separation (BSS) (Hyvärinen and Oja, 2000) and identify components of a mixture effectively (Pasadakis and Kardamakis, 2006;



Kardamakis *et al.*, 2007). In recent years, ICA has been used in medicinal tests (Fang and Lin, 2008; Wang *et al.*, 2009; Shao *et al.*, 2009). Considering there hasn't been any study applying NIR spectroscopy in inspection on internal components of *Gentiana scabra* Bunge currently, it is the intent of this study to apply ICA, which could analyze various components simultaneously, in NIR spectroscopy analysis on gentiopicroside and swertiamarin to discuss qualitative and quantitative relationships of the two bioactive components. Efforts were also made to build spectral calibration models with high predictability in order to evaluate the potentiality of NIR for quality inspection on *Gentiana scabra* Bunge.

4.2 MATERIALS AND METHODS

4.2.1 GENTIANA SCABRA BUNGE SAMPLE PREPARATION

The samples of *Gentiana scabra* Bunge were provided by Taiwan Sugar Research Institute (TSRI; Tainan City, Taiwan). The total of 94 tissue culture samples and 68 grown plant samples of different cultivation time was acquired (Yang *et al.*, 2008; Cheng, 2009). The shoot and root of the grown plant samples were measured separately in order to compare their differences. The *Gentiana scabra* Bunge samples were first dried for 48 hours in a dryer (50°C), then milled with a high speed grinder (RT-02A, Sun-Great Technology Co., Ltd., New Taipei City, Taiwan). The dried powder was

filtered with 100 mesh and stored in amber sample vials to avoid light exposure that may cause degradation of the functional ingredients in *Gentiana scabra* Bunge (Yang *et al.*, 2008; Cheng, 2009).



4.2.2 NIR SPECTRA AND HPLC MEASUREMENT

Dry powder of *Gentiana scabra* Bunge was gently poured into a small ring cup (i.d. 5 cm) and subjected to NIR measurement (NIRS 6500, FOSS NIRSystems, Inc., Laurel, MD, U.S.A.). The reflectance spectra of the samples ranged from 400 to 2498 nm with 2 nm intervals, and the NIR spectrum of each sample was the average of 32 scans (Yang *et al.*, 2008; Cheng, 2009).

To attain the reference value of the two bioactive components, the authors measured gentiopicroside and swertiamarin using HPLC (DX 500 ion chromatograph, Dionex Corporation, Sunnyvale, CA, U.S.A.) equipped with a DIONEX C18 column (250 mm × 4.6 mm i.d.). The peak of gentiopicroside and swertiamarin showed up at 250 nm when applying methanol:water (20:80) in the mobile phase at a flow rate of 1 mL/min. A high-precision scale was used to measure the gentiopicroside and the swertiamarin standard powder, and diluted into 1000, 500, and 250 ppm with 70 % methanol as the standard solutions for the three-point calibration of HPLC. A quantitative linear

relationship was established between the standard concentration and the peak area (Yang *et al.*, 2008; Cheng, 2009).

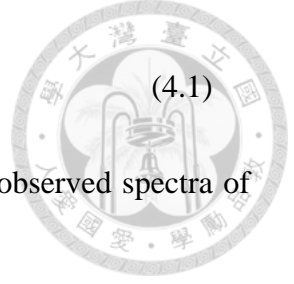


4.2.3 DATA ANALYSIS

After the reflectance spectra of *Gentiana scabra* Bunge powder and the contents of two bioactive components of *Gentiana scabra* Bunge were measured, ICA was applied to explore the relationship between them and spectral calibration models of the two bioactive components were then built, respectively.

4.2.3.1 INDEPENDENT COMPONENT ANALYSIS (ICA)

Independent component analysis (ICA) is a method used to transform the observed multivariate data to statistically independent components (ICs) and 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 (Hyvärinen and Oja, 2000). 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 (\mathbf{M}) or the separating matrix (\mathbf{W}). In order to predict the unknown source, it is assumed that $\mathbf{W} = \mathbf{M}^{-1}$,



$$\hat{\mathbf{s}} = \mathbf{W}\mathbf{x} = \mathbf{M}^{-1}\mathbf{M}\mathbf{s} \quad (4.1)$$

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

In the present study JADE (joint approximate diagonalization of eigenmatrices) algorithm (Cardoso and Souloumiac, 1993; Cardoso, 1999) 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:

$$\mathbf{A} = \mathbf{M}\mathbf{I} \quad (4.2)$$

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 (\mathbf{M}) and the component concentration (\mathbf{C}) can be expressed as:

$$\mathbf{C} = \mathbf{M}\mathbf{B} \quad (4.3)$$

Among them, \mathbf{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.

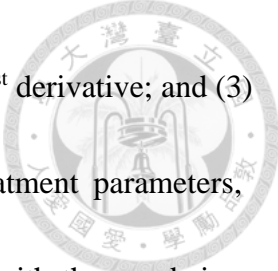


4.2.3.2 SPECTRAL PRETREATMENTS

The purpose of spectral pretreatments was to eliminate the spectral variation not caused by chemical information contained in the samples (de Noord, 1994; Fearn, 2001). Since inevitable light scattering could be added into the spectra when using NIR to measure powder samples, especially when the particle size is not uniform, multiplicative scatter correction (MSC) was used to allow additive and multiplicative transformation of the spectra (Eq. 4.4). It was conducted using the average spectrum of all samples as the reference value, and calculating the parameters a and b with the least square. After MSC treatment, the spectra of *Gentiana scabra* Bunge powder not only reduced the physical impact of non-uniform particles (Helland *et al.*, 1995; Maleki *et al.*, 2007), but also confirmed the linearity of the spectral information (Isaksson and Næs, 1988), which would contribute to subsequent linear regression analysis (Thennadil *et al.*, 2006).

$$x_{ik}(new) = \frac{[x_{ik}(old) - a_i]}{b_i} \quad (4.4)$$

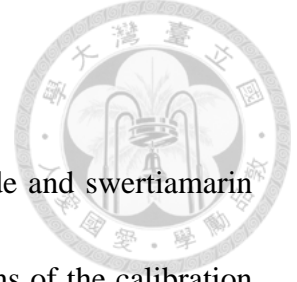
The spectra of *Gentiana scabra* Bunge powder post-MSC was subjected to three



independent treatments, namely (1) smoothing; (2) smoothing with 1st derivative; and (3) smoothing with 2nd derivative, in order to choose the best pretreatment parameters, including the smoothing points and the gap ranging from 2 to 50, with the gap being greater than or equal to the smoothing points.

4.2.3.3 MODEL ESTABLISHMENT

This research used MATLAB version 7.5.0 (The MathWorks, Inc., Natick, MA, U.S.A.) to edit program of ICA spectra analysis. The ICA analysis procedure included: (1) selecting calibration set and validation set; (2) spectral pretreatments; (3) selecting the specific wavelength regions; and (4) determining best calibration model. A 2:1 ratio of calibration to validation samples was adopted according to the concentrations of bioactive components in the sample. All samples were ranked ascendantly according to their concentrations of gentiopicroside and swertiamarin, with the concentrations in the calibration set higher than the validation set, yet both sets contained similar concentration distributions of all samples. When selecting the best calibration model, in order to avoid over-fitting caused by use of excessive ICs, the following principles were adhered to: (1) the maximum number of ICs is one tenth of the number of calibration sets + 2 to 3; (2) stop if the adding of a new IC makes the SEV rise; and (3) when the SEV is lower than the SEC, stop adding new IC.



After the respective spectral calibration models of gentiopicroside and swertiamarin were built, these models were then used to predict the concentrations of the calibration and the validation set. The predictability of the models was evaluated based on the following statistical parameters, including coefficient of correlation of calibration set (R_c), standard error of calibration (SEC), standard error of the validation (SEV), bias and the ratio of the standard error of performance to the standard deviation of the reference values (RPD), as defined below:

$$SEC = \left[\frac{1}{n_c} \sum_{i=1}^{n_c} (Y_r - Y_c)_i^2 \right]^{1/2} \quad (4.5)$$

$$SEV = \left\{ \frac{1}{n_v} \sum_{i=1}^{n_v} [(Y_r - Y_v) - \text{Bias}]_i^2 \right\}^{1/2} \quad (4.6)$$

$$\text{Bias} = \frac{1}{n_v} \sum_{i=1}^{n_v} (Y_r - Y_v)_i \quad (4.7)$$

$$RPD = SD/SEV \quad (4.8)$$

where Y_c and Y_v represent the estimated concentration of the calibration set and the validation set, respectively. Y_r is the reference concentration; n_c and n_v are the number of samples in the calibration set and validation set; and SD is the standard deviation of concentration within the validation set.



4.3 RESULTS AND DISCUSSION

4.3.1 DISTRIBUTIONS OF THE TARGET CONSTITUENTS IN *GENTIANA*

SCABRA BUNGE

Table 4.1 shows the contents of bioactive components of 230 *Gentiana scabra* Bunge samples including 94 tissue cultures, 68 shoots, and 68 roots. The gentiopicroside content was found significantly higher than swertiamarin in all parts of the samples studied, indicating the dominance of gentiopicroside as the main bioactive component in *Gentiana scabra* Bunge. It is interesting to note that gentiopicroside was more abundant in the whole grown plant (including shoot and root) than in the tissue culture, suggesting that accumulation of gentiopicroside in the grown plant was increased after the tissue culture was deflasked and transplanted into the greenhouse for cultivation. In addition, the gentiopicroside content in the root was higher than in the shoot, indicating that gentiopicroside was mainly stored in the root when the grown plant of *Gentiana scabra* Bunge was cultivated in the greenhouse. On the other hand, swertiamarin in the whole grown plant was far lower than in the tissue culture, suggesting that swertiamarin in the grown plant was reduced significantly after the tissue culture was deflasked and transplanted into the greenhouse for cultivation. Since the level of swertiamarin in shoot and root were both low, it is reasonable to postulate that swertiamarin might distribute evenly in stem node, shoot, and root when the grown plant of *Gentiana scabra* Bunge

was cultivated in the greenhouse.



Table 4.1 Contents and distributions of the target constituents in *Gentiana scabra*

Bunge.

Sample	#	Gentiopicroside Content (%)			Swertiamarin Content (%)		
		Mean (Min. - Max.)	SD	CV	Mean (Min. - Max.)	SD	CV
Tissue Culture	94	5.35 (2.69 - 8.18)	1.29	0.24	1.18 (0.60 - 2.15)	0.28	0.24
Grown Plant							
Shoot	68	3.26 (1.34 - 5.90)	0.91	0.28	0.27 (0.10 - 0.59)	0.11	0.42
Root	68	4.68 (2.24 - 8.77)	1.62	0.35	0.24 (0.01 - 0.34)	0.07	0.28

4.3.2 CORRELATION BETWEEN NIR SPECTRA AND TARGET CONSTITUENTS' CONTENTS

After eliminating 1/10 outliers (23 samples) from 230 *Gentiana scabra* Bunge samples, the remaining 207 effective samples were divided respectively into 138 and 69 calibration and validation samples in the ratio of 2:1. Statistical assessments on the gentiopicroside and swertiamarin contents in each data set are shown in Table 4.2. The differences of average, standard deviation, and coefficient of variation (CV) of the effective samples in the calibration and validation set were all less than 0.05 %.



Table 4.2 The target constituents' contents of effective samples, calibration set, and validation set in *Gentiana scabra* Bunge.

Sample	#	Gentiopicroside Content (%)			Swertiamarin Content (%)		
		Mean (Min. - Max.)	SD	CV	Mean (Min. - Max.)	SD	CV
Effective Samples	207	4.72 (1.59 - 8.77)	1.52	0.32	0.69 (0.12 - 2.15)	0.49	0.72
Calibration Set	138	4.73 (1.59 - 8.77)	1.53	0.32	0.69 (0.12 - 2.15)	0.49	0.72
Validation Set	69	4.72 (1.92 - 8.19)	1.51	0.32	0.68 (0.12 - 1.72)	0.49	0.72


The NIR spectra of the 207 *Gentiana scabra* Bunge samples were acquired by using the MSC treatment. As shown in Fig. 4.1(A), absorption peaks were found in both the visible region of blue light (452 nm) and red light (666 nm), since the chlorophyll in *Gentiana scabra* Bunge absorbs the majority of blue and red light when involved in photosynthesis. The spectra of tissue culture and the shoot were similar, which could be attributed to the fact that during the domestication period the tissue is mainly composed of shoots, since the root development of *Gentiana scabra* Bunge is not obvious at that time. Contrarily, the root spectra in the visible region showed a significant difference, with high absorption occurring from green to yellow light (492 to 586 nm) and low absorption (flat waveform) from orange to red light (606 to 700 nm). This could be due

to lack of chlorophyll in the roots of *Gentiana scabra* Bunge plant, hence reducing the absorption of blue and red light, while reflecting green light.



After MSC treatment, the spectra of *Gentiana scabra* Bunge were analyzed using the following pretreatments: (1) smoothing; (2) smoothing with 1st derivative; and (3) smoothing with 2nd derivative. The best pretreatment parameters (smoothing points / gap) of the gentiopicroside analysis were (3/0), (2/2), and (6/6), whereas the best of the swertiamarin analysis were (1/0), (2/2), and (6/6); both the smoothing points and the gap were less than 10, indicating that NIRS 6500 spectrophotometer was stable, and the spectra of *Gentiana scabra* Bunge powder exhibited minimal noise.

The correlation between the spectra of *Gentiana scabra* Bunge powder and the bioactive components were assessed at first when selecting specific wavelength regions of spectra. As for original spectra, the 1st derivative spectra, and the 2nd derivative spectra, the correlation coefficients of gentiopicroside of effective samples were distributed as shown in Fig. 4.1(B), and the threshold value ($|r| > 0.50$) was set to determine the degree of correlation. Because the influence of water absorption on the spectrum of *Gentiana scabra* Bunge powder had been eliminated, it's unnecessary to avoid the O-H bond absorption band around 1450 and 1900 nm. In both the visible and

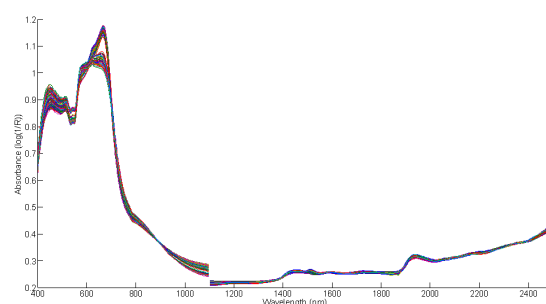


the NIR region, there were highly correlated bands, with the original spectra located between the orange and red light region as well as the O-H bond region. The 1st derivative spectra were located throughout the regions of red light, the 4th overtone of C-H bond, the combination of 1st overtone of C-H bond, and the combination between C-H bonds. On the other hand, the 2nd derivative spectra were located in the regions of red light, the 4th overtone of C-H bond, the 1st overtone of C-H bond, and the combination between N-H bond and O-H bond.

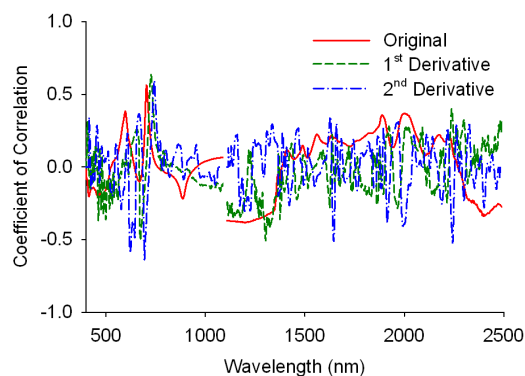
The correlation coefficients between the spectra of *Gentiana scabra* Bunge powder and swertiamarin are shown in Fig. 4.1(C) with the threshold value ($|r| > 0.75$) set to determine the degree of correlation. The original spectra were located in different regions, including red light, the 1st overtone of C-H bond, the combination between N-H bond and O-H bond, and the combination between C-H bond and C-C bond. The 1st derivative spectra were located in the regions of the 4th overtone of C-H bond, the 2nd overtone of N-H bond, the 2nd overtone of C-H bond, the combination of 1st overtone of C-H bond, the 1st overtone of C-H bond, and the combination between C-H bond and C-C bond; whereas the 2nd derivative spectra were located in the red light and the 4th overtone of C-H bond regions. As indicated by Fig. 4.1(B) and 4.1(C), the 4th overtone of C-H bond was the main absorption band for both gentiopicroside and swertiamarin. It



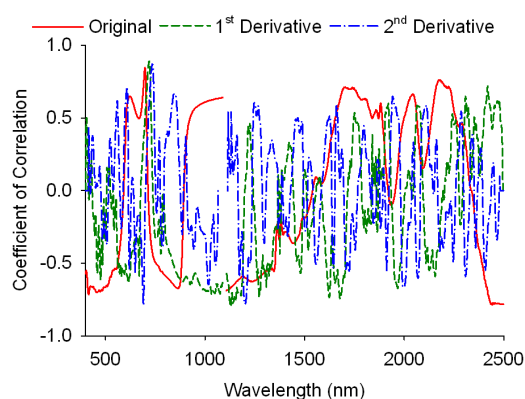
is noteworthy that the dominance of red light in the visible region of the original spectra could be attributed to the differences in the color of tissue culture, shoot, and root.



(a)



(b)



(c)

Fig. 4.1 (A) The spectra of *Gentiana scabra* Bunge powder post-MSC; (B) correlation coefficient distributions between the spectra and gentiopicroside; and (C) correlation coefficient distributions between the spectra and swertiamarin.

4.3.3 NIR SPECTRA DECOMPOSITION AND ICA ANALYSIS OF THE TARGET CONSTITUENTS



According to the definition of ICA, the observed signal of receiver can be decomposed into ICs of which the number is the same as that of training samples at most (Hyvärinen and Oja, 2000). In order to avoid over-fitting of calibration model caused by use of excessive ICs, appropriate ICs were selected under the condition that calibration models were built only by using 1 to 17 ICs when ICA analysis was conducted for original spectra (400 to 2498 nm) of the calibration set. The SEV of the calibration models continued to drop and then rise when 7 ICs were applied, indicating that incorporation of more IC will not necessarily be helpful to the analysis as it is sufficient to decompose the spectra into 7 ICs.

After the original spectra (400 to 2498 nm) of the calibration set was decomposed into 7 ICs, correlations between each IC and the two bioactive components were checked. ICs 4 and 5 presented the higher correlation coefficients, followed by IC 6, suggesting that the spectral information about gentiopicroside and swertiamarin was typically stored in these three ICs. There were peaks for IC 4 in the wavelength of 704 nm, IC 5 in the wavelengths of 692 and 740 nm, and IC 6 in the wavelengths of 494, 1838, 1944, 2058, and 2132 nm (Fig. 4.2), which was consistent with the absorption

bands seen in Fig. 4.1(B) and 4.1(C). This suggests that the spectral characteristics of gentiopicroside and swertiamarin were mainly reflected in ICs 4, 5, and 6 (Chen and Wang, 2001; Hahn and Yoon, 2006; Pasadakis and Kardamakis, 2006; Kardamakis *et al.*, 2007). These wavelengths will be taken as the reference for selection on specific wavelength region of spectra when building calibration models.

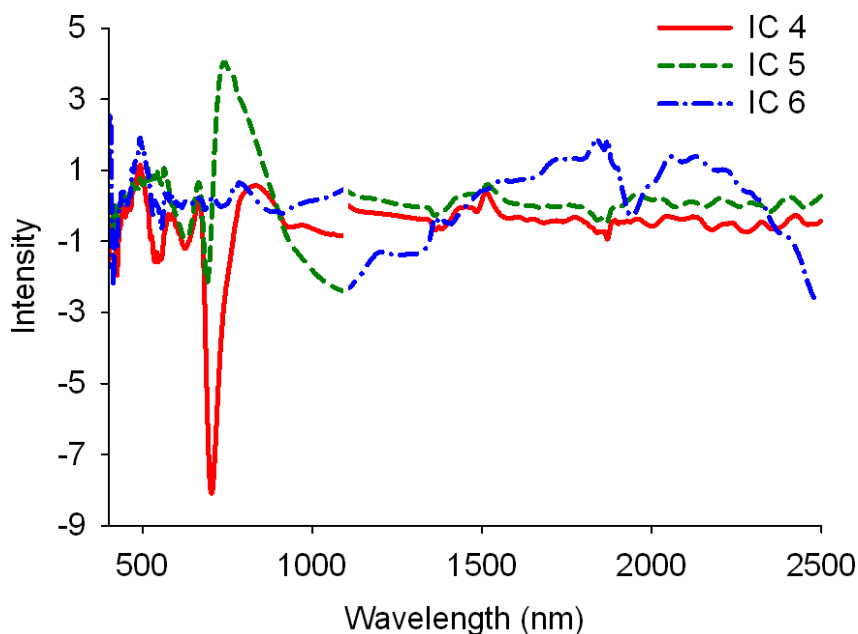



Fig. 4.2 The three ICs decomposed from the original spectra of *Gentiana scabra* Bunge powder post-MSC that has higher correlation with gentiopicroside and swertiamarin.

As shown in Eq. 4.2, the mixing matrix contained concentration information of the two bioactive components in each sample. Since the spectral information of gentiopicroside and swertiamarin was mainly reflected in ICs 4 and 5, the values of



these two ICs in the mixing matrix were used to configure 2-D distributions. As can be seen in Fig. 4.3(A) and 4.3(B), tissue culture, shoot, and root were distributed in three distinct locations of the IC 4-IC 5 space. The values of tissue culture and shoot were close to each other and the root presented a higher value in IC 5, showing differences among different parts of *Gentiana scabra* Bunge presented in the spectra, which are consistent with the result in Fig. 4.1(A). If the average contents of gentiopicroside and swertiamarin were taken as the threshold values, the samples could be classified into four groups, namely A: gentiopicroside and swertiamarin at high contents; B: gentiopicroside at high content and swertiamarin at low content; C: gentiopicroside at low content and swertiamarin at high content; and D: gentiopicroside and swertiamarin at low contents. The distributions of calibration and validation sets in the IC 4-IC 5 space are shown in Fig. 4.3(C) and 4.3(D), of which the gentiopicroside contents of most tissue cultures were higher than the mean value, suggesting that the production of gentiopicroside of *Gentiana scabra* Bunge was sufficient during the domestication period. As the grown plants of *Gentiana scabra* Bunge were collected at different growth stages, their gentiopicroside content in root varied. The gentiopicroside content in shoot was low, indicating that gentiopicroside was mainly stored in the root for *Gentiana scabra* Bunge plant during greenhouse cultivation. On the other hand, the swertiamarin content in tissue culture was higher than the mean value, but lower than



the mean value in shoot and root, indicating that swertiamarin in *Gentiana scabra* Bunge plant was reduced during greenhouse cultivation; therefore it is preferable to extract swertiamarin from tissue culture.

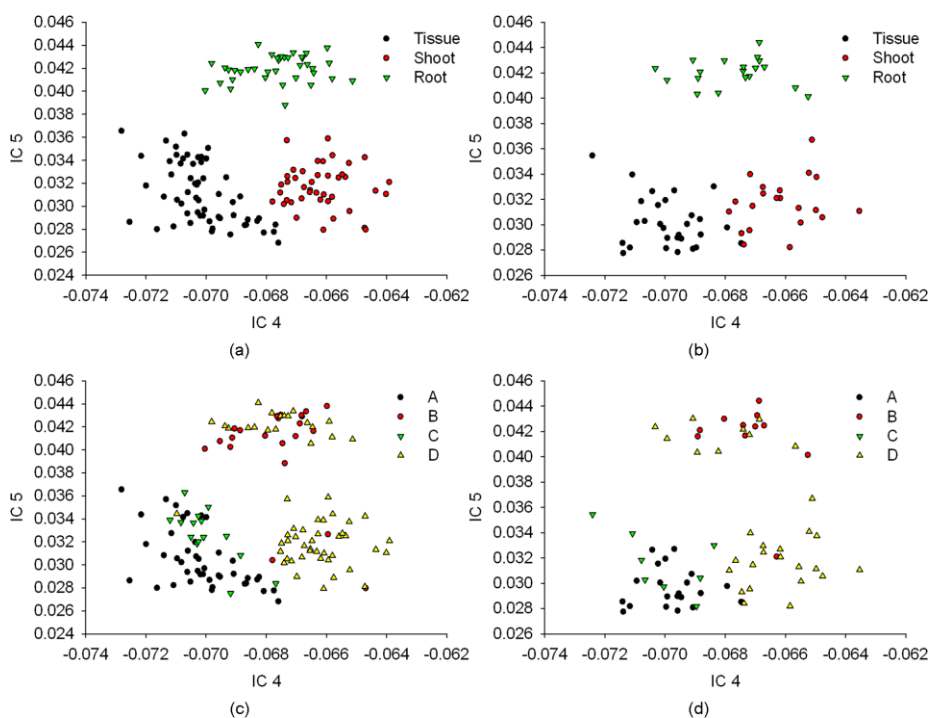



Fig. 4.3 Scores of tissue culture, shoot, and root in IC 4-IC 5 space established with calibration samples. (A) = calibration set, (B) = validation set. Scores of gentiopicroside and swertiamarin in IC 4-IC 5 space established with calibration samples. (C) = calibration set, (D) = validation set.

According to the discussion foregoing, IC 6 also contains spectral information about gentiopicroside and swertiamarin; so the values of ICs 4, 5, and 6 in the mixing matrix were used for 3-D plotting. As shown in Fig. 4.4(A) and 4.4(B), tissue culture, shoot,



and root were clearly distributed in three locations of the IC 4-IC 5-IC 6 space, indicating that even if the correlation between IC 6 and the two bioactive components was lower than that of ICs 4 and 5, the information could still be helpful to the analysis.

If the average contents of gentiopicroside and swertiamarin were used for sample grouping, the distributions of calibration and validation sets in the IC 4-IC 5-IC 6 space could be constructed, as shown in Fig. 4.4(C) and 4.4(D). The lower the value of IC 4 is, the higher the value of IC 6, hence the higher the gentiopicroside content. Similarly, the lower the values of ICs 4 and 5 are, the higher the value of IC 6, thus the higher the swertiamarin content. Fig. 4.3 and 4.4 indicate that the differences among various parts of *Gentiana scabra* Bunge could be clearly identified by the change in the trend of two bioactive components from the space of ICs, making the information useful in qualitative and quantitative analysis of NIR spectroscopy.

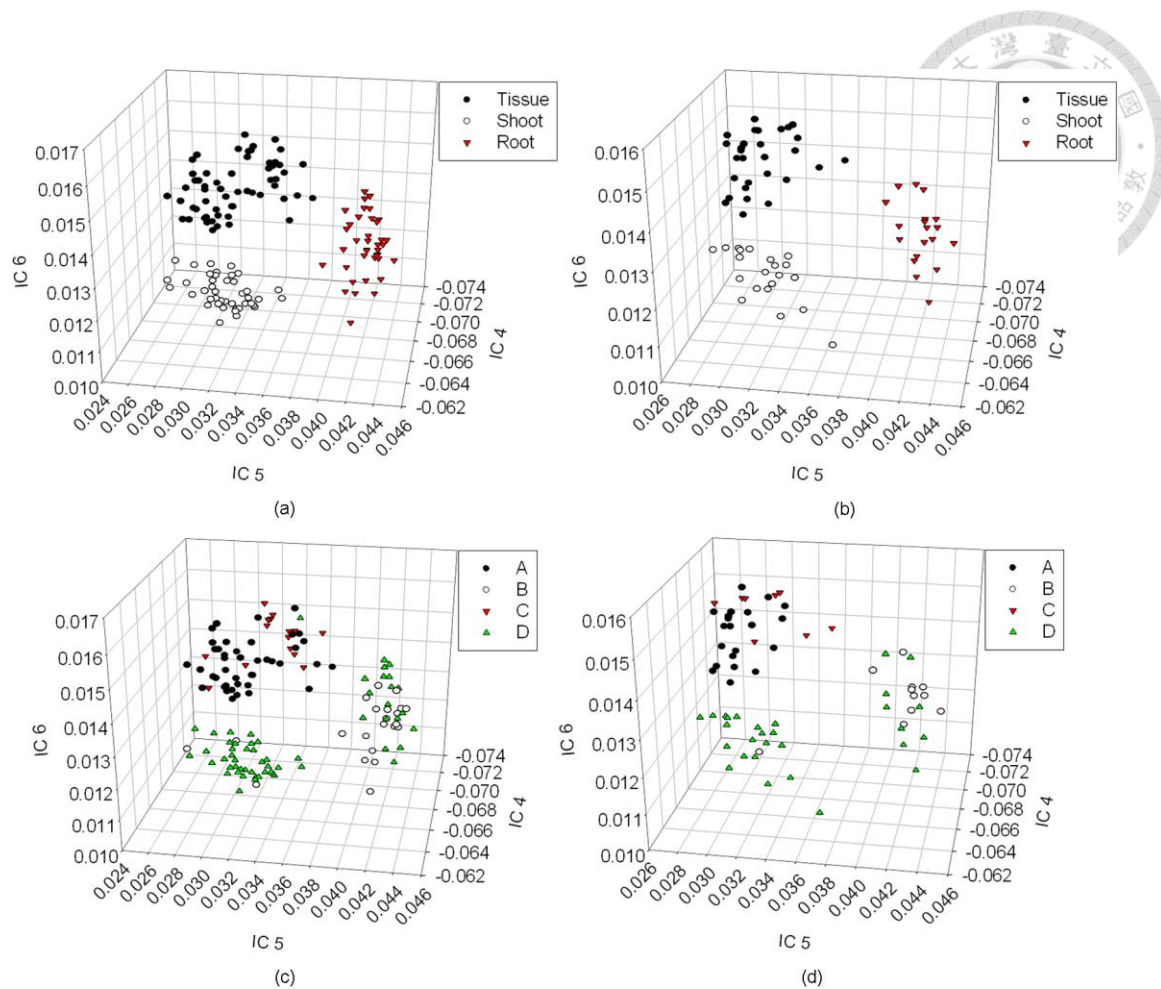
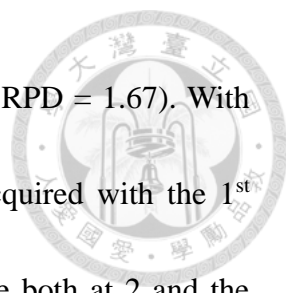


Fig. 4.4 Scores of tissue culture, shoot, and root in IC 4-IC 5-IC 6 space established with calibration samples. (A) = calibration set, (B) = validation set. Scores of gentiopicroside and swertiamarin in IC 4-IC 5-IC 6 space established with calibration samples. (C) = calibration set, (D) = validation set.

The ICA analysis results of the two bioactive components are shown in Table 4.3.

The best spectral calibration model of gentiopicroside was attained when applying the 2nd derivative spectra, of which the smoothing points and the gap were both 6 and the wavelength ranged 600 to 700 nm, 1600 to 1700 nm, and 2000 to 2300 nm ($R_c = 0.847$,



SEC = 0.865 %, $r_v = 0.756$, SEV = 0.909 %, bias = -0.395 %, and RPD = 1.67). With regard to swertiamarin, the best spectral calibration model was acquired with the 1st derivative spectra, of which the smoothing points and the gap were both at 2 and the wavelength ranged 600 to 800 nm and 2200 to 2300 nm ($R_c = 0.948$, SEC = 0.168 %, $r_v = 0.898$, SEV = 0.216 %, bias = 0.003 %, and RPD = 2.28). Satisfied outcomes were acquired for both gentiopicroside and swertiamarin. The relationship between the predicted and reference concentrations of both bioactive components are shown in Fig. 4.5. Since the content of gentiopicroside predicted by the calibration model was mainly affected by bias, the predictability can be improved by eliminating the bias calculated from a set of representative samples. As for the prediction accuracy of swertiamarin content in the calibration model, it is clear that the error mainly came from minor outlier samples because swertiamarin content in *Gentiana scabra* Bunge is relatively low, which is also why the quantity and equitability of *Gentiana scabra* Bunge powder are both important.

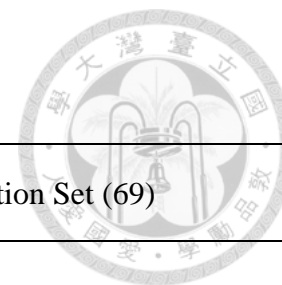


Table 4.3 Prediction of the target constituents' contents in *Gentiana scabra* Bunge by ICA models.

Constituent	Spectrum	Smoothing Points / Gap	Wavelength Ranges (nm), Interval	ICs	Calibration Set (138)		Validation Set (69)				
					R _c	SEC (%)	r _v	SEV (%)	Bias (%)	RPD	
Gentiopicroside	2 nd Derivative		600 - 700, 2								
		6 / 6	1600 - 1700, 2	16	0.847	0.865	0.756	0.909	-0.395	1.67	
			2000 - 2300, 2								
Swertiamarin	1 st Derivative		600 - 800, 2								
		2 / 2	2200 - 2300, 2	17	0.948	0.168	0.898	0.216	0.003	2.28	

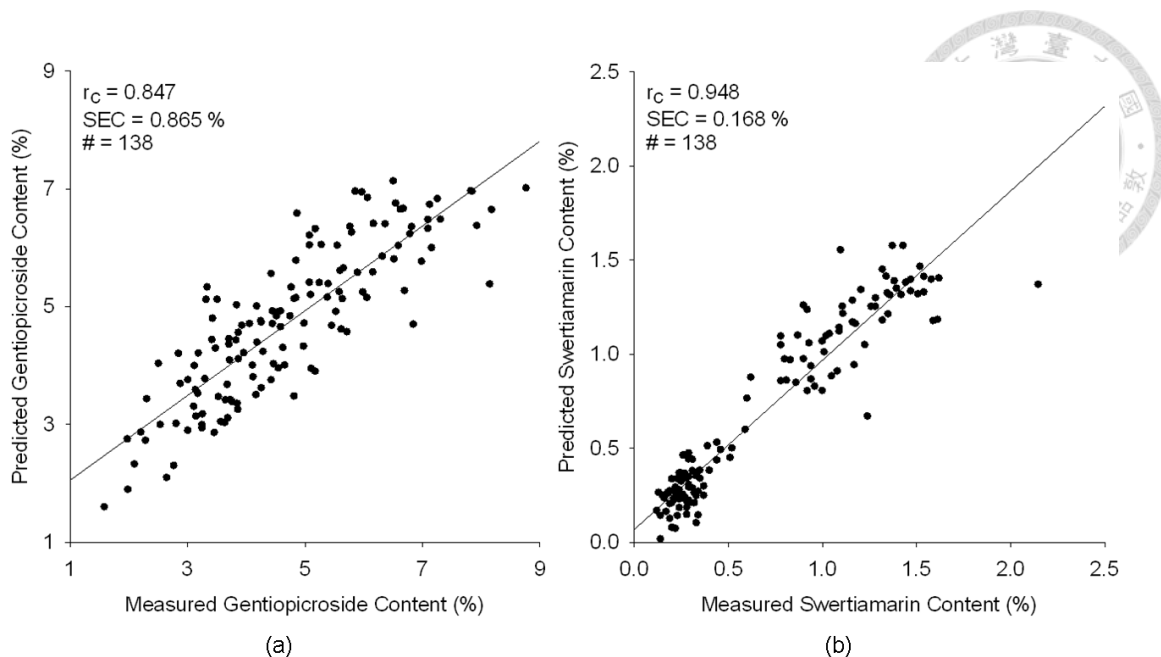


Fig. 4.5 Relationship between the estimated contents and the reference contents of (A) gentiopicroside; and (B) swertiamarin in *Gentiana scabra* Bunge.

4.4 CONCLUSIONS

This study applied ICA in NIR spectroscopy analysis on gentiopicroside and swertiamarin - bioactive components of *Gentiana scabra* Bunge and discussed relevant tissue culture and grown plant (including shoot and root). By selecting ICs that were highly correlated to the bioactive components, the space of ICs could clearly show the distribution of gentiopicroside and swertiamarin in different parts of *Gentiana scabra* Bunge. Additionally, the predictability of the spectral calibration models on the two bioactive components was adequate for establishing qualitative and quantitative correlations. Therefore, by combining ICA with NIR spectroscopy, fast and accurate inspection of gentiopicroside and swertiamarin in *Gentiana scabra* Bunge at different

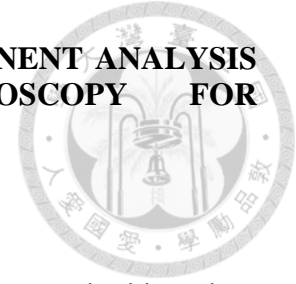
growth stages could be achieved. This technology could contribute substantially to the quality management of *Gentiana scabra* Bunge during and post cultivation.



ACKNOWLEDGMENT

I would like to thank Mr. Cheng-Wei Huang, Mr. Yu-Song Chen, and Mr. Chun-Chi Chen for their assistance.


CHAPTER 5. INTEGRATION OF INDEPENDENT COMPONENT ANALYSIS WITH NEAR INFRARED SPECTROSCOPY FOR EVALUATION OF RICE FRESHNESS



5.1 INTRODUCTION

Near infrared (NIR) spectroscopy, a rapid nondestructive inspection method based on specific absorptions within a given range of wavelength corresponding to the constituents in the sample, has been widely applied for evaluation of internal quality of agricultural products (Delwiche, 1998; Delwiche and Graybosch, 2002; Bao *et al.*, 2007; Chen and Huang, 2010; Salgó and Gergely, 2012). Because an NIR spectrum of a mixture on first approximation is the linear addition of individual spectra of the constituents in the mixture, such a spectrum thus can be regarded as an assembly of ‘blind sources’ as the proportion of constituents in the samples remains unknown (Hyvärinen *et al.*, 2001). A multiuse statistical approach originally used to implement ‘blind source separation’ in signal processing (Herault and Jutten, 1986; Vittoz and Arreguit, 1989), known as independent component analysis (ICA), is capable of disassembling the mixture’s signals from a Gaussian distribution into non-Gaussian independent constituents with only a small loss of information and does not require any additional information from the source (Comon, 1994).

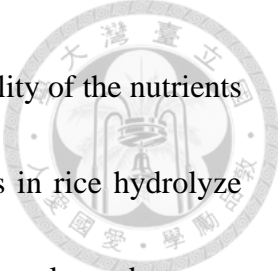
Application of ICA for spectrum analysis has been demonstrated by Chen and Wang (2001) in separating the pure spectra of various constituents from the NIR spectra of the



mixtures, whereupon relationships were established between the estimated independent components and the constituents. Such a capability also enabled complete explanation of the constituents' properties for NIR qualitative analyses (Westad and Kermit, 2003).

In addition, ICA was used to obtain statistically independent and chemically interpretable latent variables (LVs) in multivariate regression (Gustafsson, 2005). It was also noted that the number of independent components extracted from the spectra of mixtures is related to the performance of ICA (Westad, 2005). Moreover, ICA was employed to identify the infrared spectra of mixtures containing two pure materials (Hahn and Yoon, 2006) as well as the constituents in commercial gasoline (Pasadakis and Kardamakis, 2006; Kardamakis *et al.*, 2007). Equally noteworthy is the observation that the calibration model built through multiple linear regression (MLR), after using ICA to extract independent components of aqueous solutions, gave good predictability (Kaneko *et al.*, 2008). In other work, the accuracy of the NIR estimation of sucrose concentration (Chuang *et al.*, 2010) and glucose concentration (Al-Mbaideen and Benaissa, 2011) were enhanced by using ICA.

While application of ICA for spectral analysis appears promising, available literature still focuses mainly on chemical samples or non-natural products. To date, ICA has not been applied to NIR quantitative analysis of the internal quality of rice. The storage




time of rice has an enormous effect on its appearance, flavor, and quality of the nutrients (Zhou *et al.*, 2002). A previous study demonstrated that most lipids in rice hydrolyze into free fatty acids and cause the acidity of rice to increase with prolonged storage (Takano, 1989; Hu, 2011; Chen *et al.*, 2011). Therefore, the determination of rice freshness is one of the main goals in site examination. There is a strong need to develop a non-invasive, rapid detection method for the analysis of freshness. Therefore, the objective of the current study was to examine rice freshness in terms of qualitative and quantitative approaches using NIR spectroscopy. Rice freshness was expressed by both pH value and fat acidity (Hu, 2011; Chen *et al.*, 2011). The pH values were determined by bromothymol blue - methyl red (BTB-MR) method (Hsu and Song, 1988) and fat acidity by AACC International method 02-02.02 (AACC International, 2000). By means of a calibration curve, a relationship between pH and fat acidity was established (Hu, 2011; Chen *et al.*, 2011). ICA was subsequently integrated with NIR spectral analysis to quantify the pH in rice. Linear regression was then used to build spectral calibration models of pH value.

5.2 MATERIALS AND METHODS

5.2.1 SAMPLE PREPARATION

A total of 180 (= 6 cargo lots × 30 draws per lot) Tainan 11 (TN-11) paddy rice

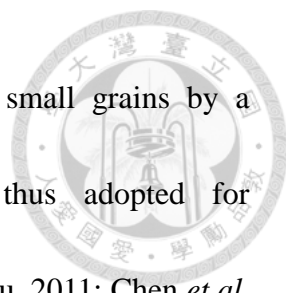


samples stored at 10-15°C were provided by the Erlin Farmers' Association, Changhua County (a central-west coastal county in Taiwan) and Agricultural Research and Extension Station, Taichung in Taiwan, including 6 crop seasons (1 lot per season): 2nd crop of 2010, 1st crop of 2010, 1st crop of 2009, 1st crop of 2008, 1st crop of 2007 and 1st crop of 2006. All samples were collected at one time and then dehulled and milled soon thereafter (Hu, 2011; Chen *et al.*, 2011).

5.2.2 NIR SPECTRA AND PH VALUE MEASUREMENT

A NIRS 6500 spectrophotometer and sample transport module (FOSS NIRSystems, Laurel, MD, U.S.A.) with coarse granular sample cell was used to measure rice reflectance spectra (Hu, 2011; Chen *et al.*, 2011). The wavelength ranged from 400 to 2498 nm in 2 nm intervals. The NIR spectrum of each sample was the average of 32 scans.

Two chemical methods, AACC method 02-02.02 and the BTB-MR method, were used to evaluate rice freshness (Hu, 2011; Chen *et al.*, 2011). Natural fats are mixtures of esters of fatty acids with glycerol and are essentially neutral. However, partial hydrolysis of the glycerides may be caused by unfavorable storage conditions. The resultant free fatty acids increase the acidity, which is an indication of deterioration in



quality. AACC method 02-02.02 determines total fatty acids in small grains by a titrimetric procedure (AACC International, 2000) and was thus adopted for quantification of total fat acidity (FA) content of rice in this study (Hu, 2011; Chen *et al.*, 2011).

The BTB-MR method, a rapid method compared to AACC method 02-02.02, determines the pH value of rice by distinguishing the color of a solution of rice in which standardized color solutions and pH have been established (Hsu and Song, 1988; Hu, 2011; Chen *et al.*, 2011). The BTB-MR method has also been adopted as a standard method for the examination of rice freshness by Agriculture and Food Agency, Council of Agriculture, Executive Yuan in Taiwan. A 200 mL standard solution was first made by mixing 0.1 g methyl red (MR), 0.3 g bromothymol blue (BTB), 150 mL ethyl alcohol, and distilled water. The standard solution was subsequently diluted by volume ratio 1:50 with distilled water. After mixing 10 mL of the diluted solution and 5 g of white rice, a shaker was employed to evenly disperse the rice in solution. The relationship between colors of rice solutions and pH values was established by standard color checks of known pH. Upon establishing the relationship between the AACC and BTB-MR methods, the latter, because of its ease of use and good precision, served as the standard method for assessing rice freshness in this study (Hu, 2011; Chen *et al.*, 2011).



5.2.3 DATA ANALYSIS

5.2.3.1 INDEPENDENT COMPONENT ANALYSIS (ICA)

Independent component analysis (ICA) is a method used to transform the observed multivariate data to statistically independent components (ICs) and present them as a linear combination of observation variables. The number of receptors defined by the ICA algorithm must be more than or equal to the number of sources, and the signals emitted by the sources are in non-Gaussian distributions (Comon, 1994; Hyvärinen and Oja, 2000). ICA supposes that all components (sources) are statistically independent. The ICs are latent variables; therefore, they cannot be directly observed. This indicates that the mixing matrix, the intensity of the sources among the observed signals, is also unknown. The purpose of the ICA algorithm is to determine the mixing matrix (\mathbf{M}) or its inverse, the separating matrix (\mathbf{W}). The unknown source, s , is approximated as

$$\hat{\mathbf{s}} = \mathbf{W}\mathbf{x} = \mathbf{M}^{-1}\mathbf{M}\mathbf{s} \quad (5.1)$$

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

In the present study, a JADE (joint approximate diagonalization of eigenmatrices) algorithm (Cardoso and Souloumiac, 1993; Cardoso, 1999) was employed to conduct



ICA. In general, the JADE approach offers rapid performance when dealing with spectral data compared to other multivariate approaches like principal component regression (PCR) and partial least squares regression (PLSR). Assuming that the spectra obtained through measurement of the unknown mixtures were the linear combination of the spectra of various components, these can be expressed as

$$\mathbf{A} = \mathbf{M}\mathbf{I} \quad (5.2)$$

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 (\mathbf{M}) and the component concentration (\mathbf{C}) can be expressed as:

$$\mathbf{C} = \mathbf{M}\mathbf{B} \quad (5.3)$$

where \mathbf{B} refers to the regression coefficient matrix. In doing so, the concentration of each component in the mixture can be determined by the combination of ICA and linear regression.

5.2.3.2 MODEL ESTABLISHMENT

The ICA algorithm was coded in MATLAB (The MathWorks, Inc., Natick, MA,

U.S.A.), which produced the spectral calibration equations. The 180 rice samples were divided into 120 calibration samples (2nd crop of 2010, 1st crop of 2009, 1st crop of 2008, and 1st crop of 2006) and 60 prediction samples (1st crop of 2010 and 1st crop of 2007).

After the respective spectral calibration equations for rice were built, these equations were then used to predict the pH values in the calibration and the prediction sets. The evaluation of predictability was based on the following statistical parameters: multivariate coefficient of determination of the calibration (R^2), standard error of calibration (SEC), standard error of prediction (SEP), bias, and the ratio of the standard error of performance to the standard deviation of the reference values (RPD).

5.3 RESULTS AND DISCUSSION

5.3.1 RELATIONSHIP BETWEEN FAT ACIDITY AND PH VALUE

In the present study the BTB-MR method served as a standard method for assessing rice freshness due to its rapidness and convenience compared to AACC method 02-02.02 (Hu, 2011; Chen *et al.*, 2011). The pH measurement procedure was performed on each of the 180 samples, with values ranging from 5.20 to 6.96. The average pH from a total of 180 samples was 5.90, and the standard deviation was 0.53. To establish the relationship between fat acidity and pH value, a total of 18 (= 6 crop seasons \times 3 samples per season) rice samples were randomly selected to receive AACC method

02-02.02 (Hu, 2011; Chen *et al.*, 2011). A calibration curve of fat acidity versus pH value from the selected samples is shown in Fig. 5.1, with the coefficient of determination (r^2) being 0.924 according to the linear regression equation (Hu, 2011; Chen *et al.*, 2011).

$$\text{FA (KOH mg/100g)} = -1.6213 \times \text{pH} + 44.141 \quad (5.4)$$

The fat acidity of the other samples thus can be accurately calculated by the calibration curve, however, for the remainder of this study just the BTB-MR pH value is used.

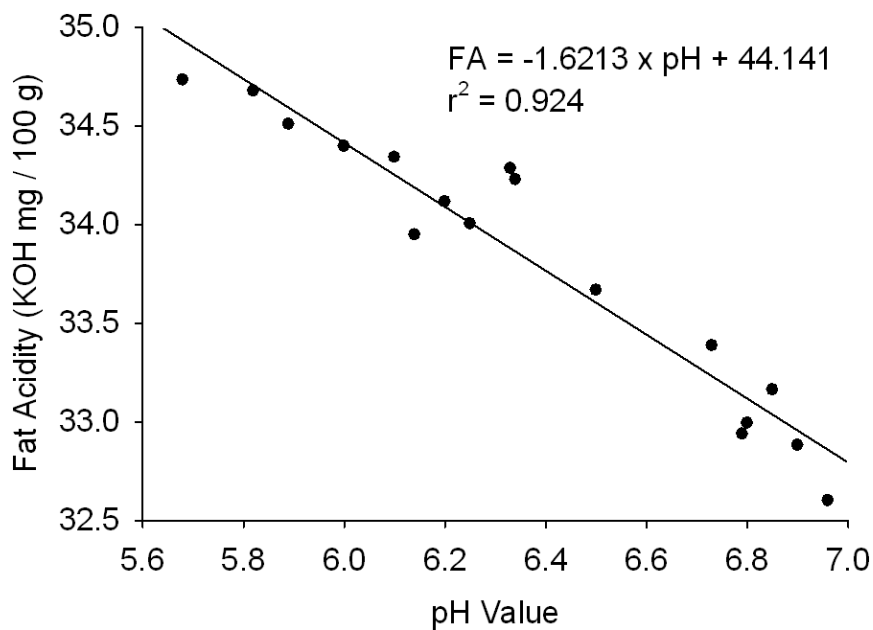


Fig. 5.1 Relationship between fat acidity and pH value established by the 18 selected rice samples (Hu, 2011; Chen *et al.*, 2011).

5.3.2 DISTRIBUTIONS OF THE PH VALUE IN RICE



The distribution of pH values of six crop season samples is shown in Fig. 5.2. It can be clearly found that pH value of rice decreases with increasing storage time. For the six crop season samples, four groups can be identified according to their pH values, with group 1 (6.5-7.0 pH, $n = 31$) containing the 2nd crop of 2010, group 2 (6.0-6.5 pH, $n = 30$) containing the 1st crop of 2010, group 3 (5.5-6.0 pH, $n = 63$) containing the 1st crop of 2009 and 1st crop of 2008, and finally group 4 (5.0-5.5 pH, $n = 56$) containing the 1st crop of 2007 and 1st crop of 2006. The average pH levels of the calibration and prediction sets are 5.93 and 5.86, and their standard deviations are 0.58 and 0.41, respectively.

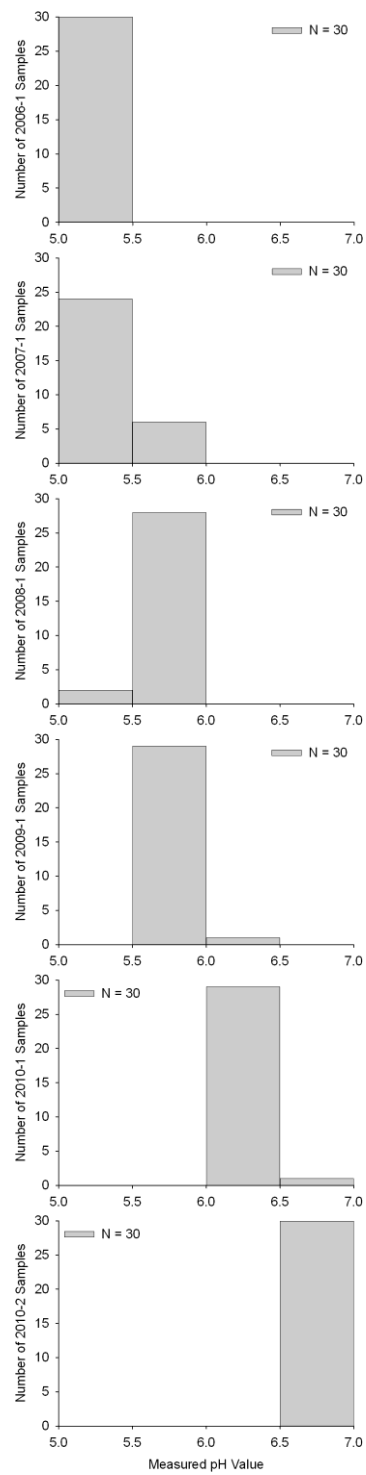


Fig. 5.2 Distributions of pH values for six crop seasons of white rice samples.

5.3.3 NIR SPECTRA DECOMPOSITION AND ICA ANALYSIS OF THE PH VALUE



The full visible through near infrared wavelength range (400 to 2498 nm) was used as the inputs of ICA. According to the definition of ICA, the observed receptor signals can be decomposed at most into a number of ICs (independent components) equal to the number of samples (Hyvärinen and Oja, 2000). Therefore, the calibration samples could have been decomposed into 120 ICs; however, applying too many ICs could easily lead to over fitting of the model as well as be time-consuming. Hence, ICA was conducted for the original spectra by selecting 1 to 15 ICs, and observing the prediction error of the calibration samples. As shown in Fig. 5.3, when the number of ICs increased to 5, SEC decreased to 0.202, and SEP fell to 0.233, indicating that different numbers of ICs can influence the predictability of the calibration model. However, application of more ICs did not necessarily improve the ability of the calibration model; hence, only the initial 5 ICs were applied.

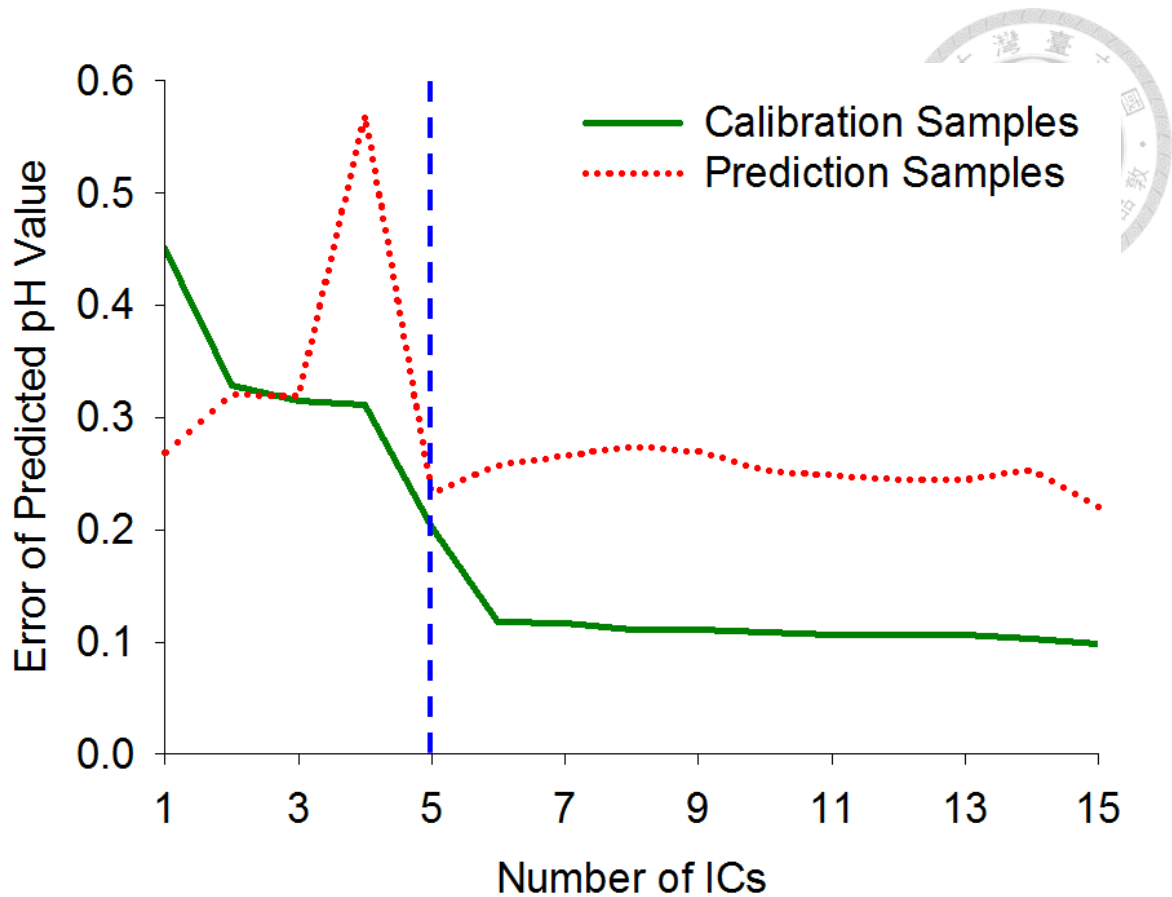
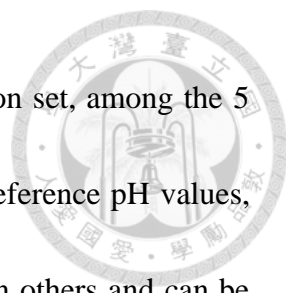


Fig. 5.3 Relationship between the numbers of independent components and the model standard errors for pH value.

Selection of the numbers of ICs influences how the information is used after spectral decomposition. The ICs reflected the spectral characteristics of the unknown mixture and constituted the pure materials' spectra of this mixture under an ideal state (Chen and Wang, 2001; Hahn and Yoon, 2006; Pasadakis and Kardamakis, 2006; Kardamakis *et al.*, 2007). Since white rice is a mixture of carbohydrates, lipids, storage proteins, enzymes, nucleic acids and other macro molecules, and the spectra inherently represent these constituents, the ICs decomposed by ICA should reflect, among other components, the



characteristics of the lipids. For the original spectra of the calibration set, among the 5 ICs used, the order of the 5 ICs, according to the correlation with reference pH values, was IC 4, 3, 2, 5 and 1. ICs 4, 3 and 2, have higher correlation than others and can be considered to respond mainly to the information of lipids, as the composition of lipids in white rice is complicated. Distributions of calibration and prediction sets of white rice in IC-2 IC-3 IC-4 space are shown in Fig. 5.4. Calibration samples are clearly separated into 3 groups, as are the prediction samples. The results are consistent with the groups seen in Table 5.1. Among calibration set, samples of 2010 gathered very close, indicating that the quality of new rice was consistent and uniform. Samples of 2009 and 2008 were interspersed, which means that the degrees of deterioration of them were very similar. Some samples of 2008 spread widely due to their quality decaying gradually. The variation trend for the samples in IC-2 IC-3 IC-4 space can be readily observed. The longer storage time of rice is, the lower the values of ICs 2 and 3 are, and the higher the value of IC 4 is. Hence, this is consistent with the direct relationship between acidity (low pH) and level of deterioration. It is interesting to note that samples of 2007 spread widely, suggesting that they underwent continuous quality decaying. Samples of 2006 clustered together but out of the variation tendency by reason of too long storage time. This indicates that white rice with different pH levels can be distinguished by ICA.

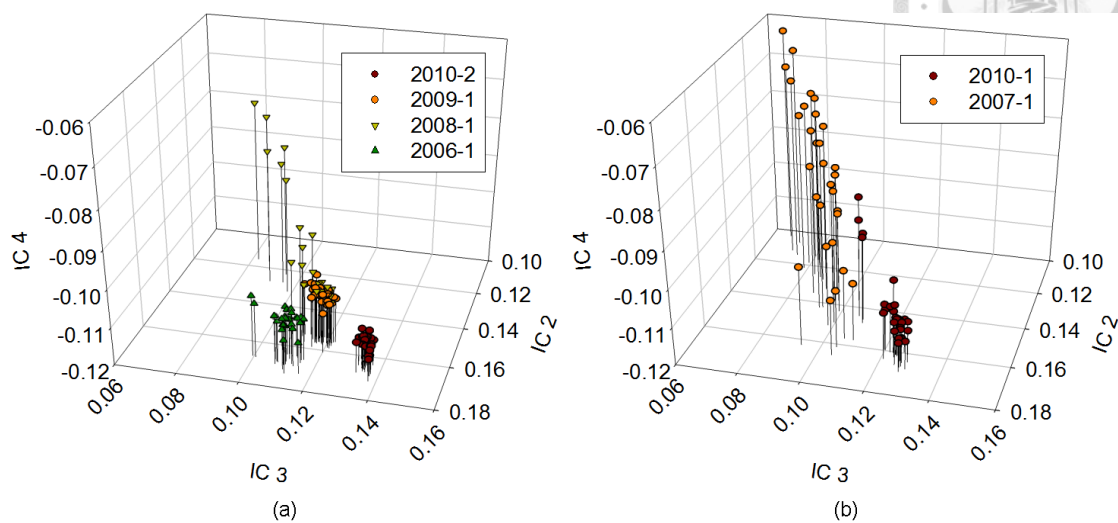
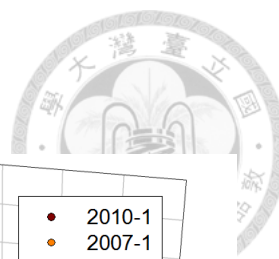


Fig. 5.4 Scores of white rice with 6 crop seasons in the vector space of independent components 2, 3, and 4 established with calibration samples. a = calibration set, b = validation set.

Table 5.1 Regression results by ICA analyses for white rice.

Calibration Set (120)			Prediction Set (60)		
Mean: 5.93, Std Deviation: 0.58			Mean: 5.86, Std Deviation: 0.41		
ICs	R ²	SEC	SEP	bias	RPD
5	0.882	0.20	0.23	0.068	1.75

Quantitative analyses of pH value in white rice were conducted by ICA using the full wavelength range of 400 to 2498 nm. The calibration model built by ICA is shown in Table 5.1 with 5 ICs applied. In Fig. 5.5, a scatter plot is made with the reference pH

value and the predicted pH value of each sample in the calibration set and prediction set.

The results are $R^2 = 0.882$, and in units of pH, $SEC = 0.20$, $SEP = 0.23$, and bias =

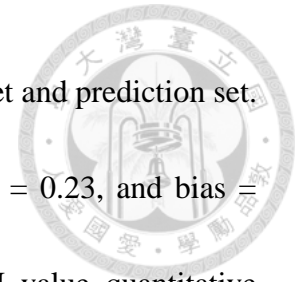
0.068. This produced an $RPD = 1.75$. Hence, ICA achieved pH value quantitative

analysis results at a level suitable for screening. Moreover, the small value for bias

indicated that ICA had a high level of tolerance from the influence caused by factors

other than the internal chemical composition of the samples. Satisfactory outcomes

were achieved in modeling pH in stored rice.



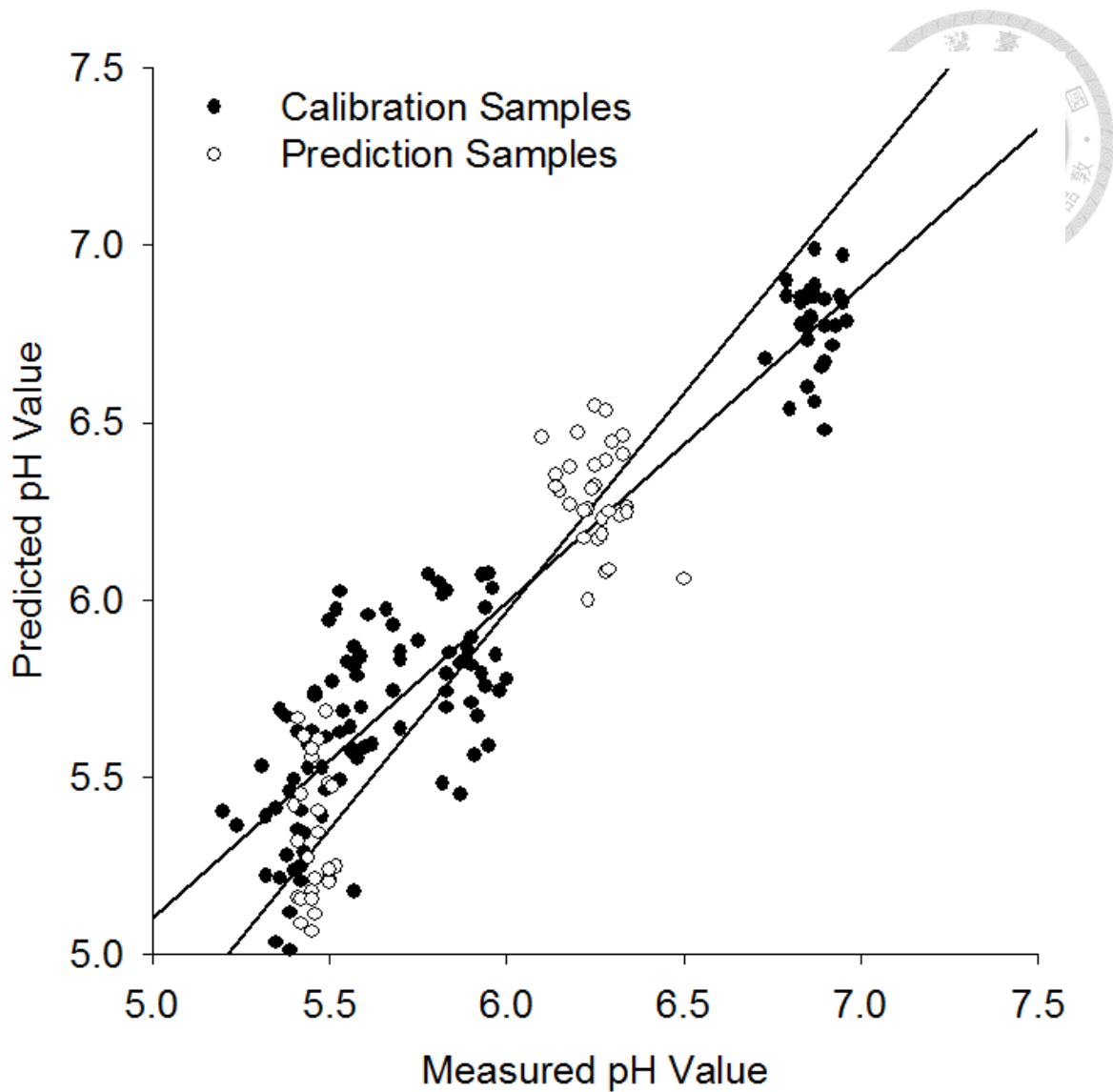


Fig. 5.5 Scatter plot of the reference pH values and the predicted pH values by independent component analysis of the NIR spectra.

5.4 CONCLUSIONS

ICA was integrated with NIR spectral analysis to quantify the internal quality of rice. A quantitative model was developed using ICA factors to predict the pH value of ground white rice in solution as a proxy for rice freshness. The results show that ICA quantitative analysis methods with near infrared spectroscopy can successfully

distinguish rice freshness and can serve as a nondestructive rapid analytical screening tool.



CHAPTER 6. GENERAL CONCLUSIONS



6.1 GENERAL DISCUSSION

In the dissertation, independent component analysis (ICA) was first adopted as the sole tool for NIR quantitative analyses of biomaterials, including wax jambu fruit (Chuang *et al.*, 2010; Chuang *et al.*, 2012c), medicinal plant *Gentiana scabra* Bunge (Chen *et al.*, 2010; Chuang *et al.*, 2012b; Chuang *et al.*, 2013), and milled white rice (Chuang *et al.*, 2012a), to evaluate the applicability of this method. Influence due to various format types of samples (sucrose solution, intact fruit, dry powder of *Gentiana scabra* Bunge, and rice kernel) was also studied.

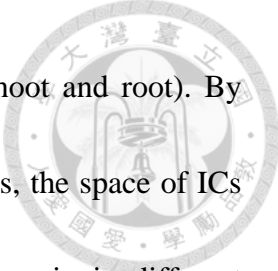
In the first part, ICA was applied as the sole tool to integrate with NIR spectroscopy for rapid quantification of sugar content in sucrose solutions and wax jambu. ICA gave a comprehensive approach to characterize the NIR spectra with respect to the sugar content in wax jambu and sucrose solutions that 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.



Regarding medicinal plants, NIR was applied for quantitative analysis of gentiopicroside which was one of the bioactive components in the medicinal plant *Gentiana scabra* Bunge. It was found that the spectral pretreatments of MSC in combination of 2nd derivative reduced the spectral noise caused by the nonuniform particle sizes of *Gentiana scabra* Bunge powder. The specific wavelength regions or specific wavelengths selected based on their characteristic response to gentiopicroside could effectively improve the predictability of the calibration models. This study successfully built the spectral calibration models for *Gentiana scabra* Bunge tissue culture and grown plant, which enable quantitative inspection of the bioactive component gentiopicroside in *Gentiana scabra* Bunge during different growth stages. The specific wavelengths selected in Silicon CCD sensing band can be used as the foundation to establish a nondestructive and rapid method to assess the quality of *Gentiana scabra* Bunge using multi-spectral imaging.

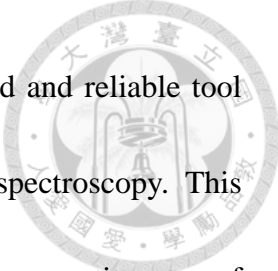
For further evaluation, this study applied ICA in NIR spectroscopy analysis on gentiopicroside and swertiamarin - bioactive components of *Gentiana scabra* Bunge



and discussed relevant tissue culture and grown plant (including shoot and root). By selecting ICs that were highly correlated to the bioactive components, the space of ICs could clearly show the distribution of gentiopicroside and swertiamarin in different parts of *Gentiana scabra* Bunge. Additionally, the predictability of the spectral calibration models on the two bioactive components was adequate for establishing qualitative and quantitative correlations. Therefore, by combining ICA with NIR spectroscopy, fast and accurate inspection of gentiopicroside and swertiamarin in *Gentiana scabra* Bunge at different growth stages could be achieved. This technology could contribute substantially to the quality management of *Gentiana scabra* Bunge or other medicinal plants (e.g. *Herba Saussureae Involucratae*) during and post cultivation.

On the other hand, ICA was integrated with NIR spectral analysis to quantify the internal quality of rice. A quantitative model was developed using ICA factors to predict the pH value of ground white rice in solution as a proxy for rice freshness. The results show that ICA quantitative analysis methods with near infrared spectroscopy can successfully distinguish rice freshness and can serve as a nondestructive rapid analytical screening tool.

In conclusion, by combining ICA with NIR spectroscopy, fast and accurate evaluation



of constituents in biomaterials could be achieved. ICA offers a rapid and reliable tool for quantitative analysis of constituents in biomaterials by NIR spectroscopy. This technology could contribute substantially to identify multiple constituents of biomaterials and evaluate their concentrations.

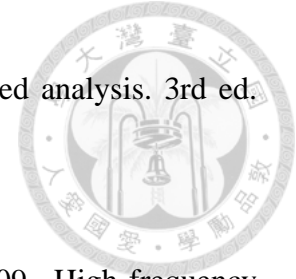
6.2 RECOMMENDATIONS FOR FUTURE RESEARCH

There are several ways for the application of ICA for future research. First, ICA can be applied to deal with more research topics according to their needs or requirements. The analysis results of ICA by conducting JADE algorithm can also be assessed by comparing to other algorithms like FastICA and kernel ICA. Second, the combination of ICA and other multivariate analysis methods such as ICA-ANN, ICA-LS-SVM, and ICA-SVM may be available to deal with nonlinear problems instead of using ICA. Third, ICA can be integrated with spectral imaging or fluorescence imaging technology for inspection of biomaterials and food products in food safety and quality assurance issues. Finally, the relationship between the values of CV and the ICA calibration models could be explored for further research.

REFERENCES



- AACC International. 2000. Method 02-02.02. Fat Acidity – Rapid Method, for Small Grains. Approved Methods of Analysis, 11th Ed. St. Paul, MN: AACC International. Available at: dx.doi.org/10.1094/AACCIntMethod-02-02.02. Accessed 2000.
- Aberham, A., S. Schwaiger, H. Stuppner, and M. Ganzera. 2007. Quantitative analysis of iridoids, secoiridoids, xanthenes and xanthone glycosides in *Gentiana lutea* L. roots by RP-HPLC and LC-MS. *J. Pharm. Biomed. Anal.* 45(3): 437-442.
- Aberham, A., V. Pieri, Jr. E. M. Croom, E. Ellmerer, and H. Stuppner. 2011. Analysis of iridoids, secoiridoids and xanthenes in *Centaurium erythraea*, *Frasera caroliniensis* and *Gentiana lutea* using LC-MS and RP-HPLC. *J. Pharm. Biomed. Anal.* 54(3): 517-525.
- Al-Mbaideen, A., and M. Benaissa. 2011. Determination of glucose concentration from NIR spectra using independent component regression. *Chemometrics Intell. Lab. Syst.* 105(1): 131-135.
- Bao, J., Y. Shen, and L. Jin. 2007. Determination of thermal and retrogradation properties of rice starch using near-infrared spectroscopy. *J. Cereal Sci.* 46(1): 75-81.
- Blanco, M., and I. Villarroya. 2002. NIR spectroscopy: a rapid-response analytical tool. *Trac-Trends Anal. Chem.* 21(4): 240-250.



Burns, D. A., and E. W. Ciurczak. 2008. Handbook of near-infrared analysis. 3rd ed.

Florida, U.S.A.: CRC Press.

Cai, Y., Y. Liu, Z. Liu, F. Zhang, F. Xiang, and G. Xia. 2009. High-frequency

embryogenesis and regeneration of plants with high content of gentiopicroside

from the Chinese medicinal plant *Gentiana straminea* Maxim. *In Vitro Cell. Dev.*

Biol.-Plant 45(6): 730-739.

Cardoso, J. F., and A. Souloumiac. 1993. Blind beamforming for non-Gaussian signals.

IEE Proceedings F 140(6): 362-370.

Cardoso, J. F. 1999. High-order contrasts for independent component analysis. *Neural*

Comput. 11(1): 157-192.

Carnat, A., D. Fraisse, A. P. Carnat, C. Felgines, D. Chaud, and J. L. Lamaison. 2005.

Influence of drying mode on iridoid bitter constituent levels in gentian root. *J. Sci.*

Food Agric. 85(4): 598-602.

Chang, W. H., S. Chen, and C. C. Tsai. 1998. Development of a universal algorithm for

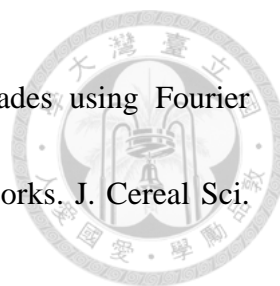
use of NIR in estimation of soluble solids in fruit juices. *Trans. ASAE* 41(6):

1739-1745.

Chen, J., and X. Z. Wang. 2001. A new approach to near-infrared spectral data analysis

using independent component analysis. *J. Chem. Inf. Comput. Sci.* 41(4):

992-1001.



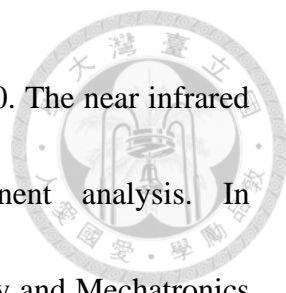
Chen, K. J., and M. Huang. 2010. Prediction of milled rice grades using Fourier transform near-infrared spectroscopy and artificial neural networks. *J. Cereal Sci.* 52(2): 221-226.

Chen, S., Y. K. Chuang, C. Y. Tsai, I. C. Yang, and Y. L. Chen. 2010. Near infrared analysis of biomaterials using independent component analysis. In "Proceedings of the 6th International Workshop on Nondestructive Quality Evaluation of Agricultural, Livestock and Fishery Products", 32-44. Taipei, Taiwan: Taiwan Agricultural Mechanization Research and Development Center.

Chen, S., Y. P. Hu, I. C. Yang, C. Y. Tsai, Y. K. Chuang, J. S. Huang, and C. W. Wang. 2011. Near infrared analysis of rice freshness. In "Proceedings of the 7th International Workshop on Nondestructive Quality Evaluation of Agricultural, Livestock and Fishery Products", 22-37. Taipei, Taiwan: Taiwan Agricultural Mechanization Research and Development Center.

Chen, X., D. Wu, Y. He, and S. Liu. 2011. Nondestructive differentiation of panax species using visible and shortwave near-infrared spectroscopy. *Food Bioprocess Technol.* 4(5): 753-761.

Cheng, Y. F. 2009. Detection of marker compounds of *Gentiana scabra* using hyper-spectral imaging. Master thesis. Taipei, Taiwan: National Taiwan University, Department of Bio-Industrial Mechatronics Engineering. In Chinese.



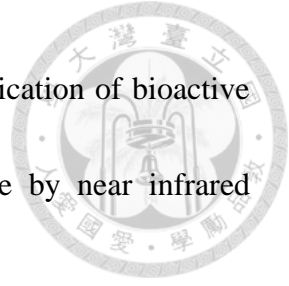
Chuang, Y. K., S. Chen, C. Y. Tsai, I. C. Yang, and Y. L. Chen. 2010. The near infrared estimation of sugar content using independent component analysis. In “Proceedings of the 5th International Symposium on Machinery and Mechatronics for Agricultural and Biosystems Engineering”, 47. Fukuoka, Japan: Kyushu University.

Chuang, Y. K., S. Chen, S. R. Delwiche, Y. M. Lo, C. Y. Tsai, I. C. Yang, and Y. P. Hu. 2012. Integration of independent component analysis with near infrared spectroscopy for evaluation of rice freshness. In “Proc. SPIE 8369, Sensing for Agriculture and Food Quality and Safety IV”, Vol. 8369, p 83690X-1~7. Baltimore, MD, U.S.A.: The International Society for Optical Engineering.

Chuang, Y. K., S. Chen, S. R. Delwiche, Y. M. Lo, I. C. Yang, and C. Y. Tsai. 2012. Assessment of bioactive components in the medicinal plant *Gentiana scabra* Bunge by near infrared spectroscopy. IFT Annual Meeting. Poster No. 164-01. Las Vegas, NV, U.S.A.

Chuang, Y. K., S. Chen, Y. M. Lo, C. Y. Tsai, I. C. Yang, Y. L. Chen, P. J. Pan, and C. C. Chen. 2012. Integration of independent component analysis with near infrared spectroscopy for rapid quantification of sugar content in wax jambu (*Syzygium samarangense* Merrill & Perry). *J. Food Drug Anal.* 20(4): 855-864.

Chuang, Y. K., S. Chen, Y. M. Lo, I. C. Yang, Y. F. Cheng, C. Y. Wang, C. Y. Tsai, R. M.



- Hsieh, K. H. Wang, C. C. Lai, and W. C. Chen. 2013. Quantification of bioactive gentiopicroside in a medicinal plant *Gentiana scabra* Bunge by near infrared spectroscopy. *J. Food Drug Anal.* In Press.
- Chung, C. T., K. J. Tseng, H. C. Wang, and C. W. Cheng. 2004. Study of applying near infrared spectroscopy for determinating the sugar content of wax apple fruits. *J. Agri. Mach.* 13(4): 13-25. In Chinese.
- Comon, P. 1994. Independent component analysis, a new concept? *Signal Process.* 36(3): 287-314.
- Davey, M. W., W. Saeys, E. Hof, H. Ramon, R. L. Swennen, and J. Keulemans. 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(5): 1742-1751.
- de Noord, O. E. 1994. The influence of data preprocessing on the robustness and parsimony of multivariate calibration models. *Chemometrics Intell. Lab. Syst.* 23(1): 65-70.
- Delwiche, S. R. 1998. Protein content of single kernels of wheat by near-infrared reflectance spectroscopy. *J. Cereal Sci.* 27(3): 241-254.
- Delwiche, S. R., and R. A. Graybosch. 2002. Identification of waxy wheat by near-infrared reflectance spectroscopy. *J. Cereal Sci.* 35: 29-38.

Fang, L. M., and M. Lin. 2008. Prediction of active substance contents in pharmaceutical tablet using ICA and NIR. *Acta Chim. Sin.* 66(15): 1791-1795. In Chinese.



Fearn, T. 2001. Standardisation and calibration transfer for near infrared instruments: a review. *J. Near Infrared Spectrosc.* 9(4): 229-244.

Glatz, Z., J. Pospíšilová, and P. Musil. 2000. Determination of gentiopicroside in extracts of *Centaureum erythraea* and *Gentiana lutea* by micellar electrokinetic capillary chromatography. *J. Liq. Chromatogr. Relat. Technol.* 23(12): 1831-1839.

Gustafsson, M. G. 2005. Independent component analysis yields chemically interpretable latent variables in multivariate regression. *J. Chem. Inf. Model.* 45(5): 1244-1255.

Hahn, S., and G. Yoon. 2006. Identification of pure component spectra by independent component analysis in glucose prediction based on mid-infrared spectroscopy. *Appl. Opt.* 45(32): 8374-8380.

Hayta, S., I. H. Akgun, M. Ganzera, E. Bedir, and A. Gurel. 2011. Shoot proliferation and HPLC-determination of iridoid glycosides in clones of *Gentiana cruciata* L. *Plant Cell Tissue Organ Cult.* 107(1): 175-180.

Hayta, S., A. Gurel, I. H. Akgun, F. Altan, M. Ganzera, B. Tanyolac, and E. Bedir. 2011. Induction of *Gentiana cruciata* hairy roots and their secondary metabolites.



Biologia 66(4): 618-625.

Helland, I. S., T. Næs, and T. Isaksson. 1995. Related versions of the multiplicative scatter correction method for preprocessing spectroscopic data. *Chemometrics Intell. Lab. Syst.* 29(2): 233-241.

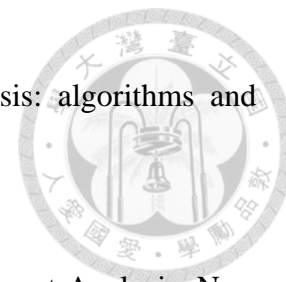
Herauld, J., and C. Jutten. 1986. Space or time adaptive signal processing by neural network models. In "AIP Conference Proceedings 151 on Neural Networks for Computing", 206-211. J. S. Denker, eds. New York, U.S.A.: American Institute of Physics Inc.

Hsu, A. N., and S. Song. 1988. Studies on the method of identification of new and old milled rices. In "Proceeding of Symposium on Rice Grain Quality", 341-349. S. Song and M. C. Hong, eds. Taichung, Taiwan: Taichung District Agricultural Research and Extension Station. In Chinese.

Hu, Y. P. 2011. Study of rice freshness using near-infrared spectroscopy. Master thesis. Taipei, Taiwan: National Taiwan University, Department of Bio-Industrial Mechatronics Engineering. In Chinese.

Hyvärinen, A., and E. Oja. 1997. A fast fixed-point algorithm for independent component analysis. *Neural Comput.* 9(7): 1483-1492.

Hyvärinen, A. 1999. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Netw.* 10(3): 626-634.



Hyvärinen, A., and E. Oja. 2000. Independent component analysis: algorithms and applications. *Neural Netw.* 13(4-5): 411-430.

Hyvärinen, A., J. Karhunen, and E. Oja. 2001. *Independent Component Analysis*. New York, U.S.A.: John Wiley and Sons.

Isaksson, T., and T. Næs. 1988. The effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy. *Appl. Spectrosc.* 42(7): 1273-1284.

Jamrógiewicz, M. 2012. Application of the near-infrared spectroscopy in the pharmaceutical technology. *J. Pharm. Biomed. Anal.* 66: 1-10.

Kakuda, R., T. Iijima, Y. Yaoita, K. Machida, and M. Kikuchi. 2001. Secoiridoid glycosides from *Gentiana scabra*. *J. Nat. Prod.* 64(12): 1574-1575.

Kaneko, H., M. Arakawa, and K. Funatsu. 2008. Development of a new regression analysis method using independent component analysis. *J. Chem. Inf. Model.* 48(3): 534-541.

Kardamakis, A. A., A. Mouchtaris, and N. Pasadakis. 2007. Linear predictive spectral coding and independent component analysis in identifying gasoline constituents using infrared spectroscopy. *Chemometrics Intell. Lab. Syst.* 89(1): 51-58.

Kikuchi, M., R. Kakuda, M. Kikuchi, and Y. Yaoita. 2005. Secoiridoid glycosides from *Gentiana scabra*. *J. Nat. Prod.* 68(5): 751-753.



Kim, J. A., N. S. Son, J. K. Son, Y. Jahng, H. W. Chang, T. S. Jang, M. K. Na, and S. H.

Lee. 2009. Two new secoiridoid glycosides from the rhizomes of *Gentiana scabra*

Bunge. *Arch. Pharm. Res.* 32(6): 863-867.

Kušar, A., H. Šircelj, and D. Baričević. 2010. Determination of seco-iridoid and

4-pyrone compounds in hydro-alcoholic extracts of *Gentiana lutea* L. subsp.

symphyandra Murb. leaves and roots by using high performance liquid

chromatography. *Isr. J. Plant Sci.* 58(3-4): 291-296.

Lathauwer, L. D., B. D. Moor, and J. Vandewalle. 2000. An introduction to independent

component analysis. *J. Chemometr.* 14(3): 123-149.

Lebot, V., R. Malapa, and M. Bourrieau. 2011. Rapid estimation of taro (*Colocasia*

esculenta) quality by near-infrared reflectance spectroscopy. *J. Agric. Food Chem.*

59(17): 9327-9334.

Lin, Y. C. 2002. Evaluation of internal quality of wax apple and papaya using

near-infrared technology. Master Thesis. Taipei, Taiwan: National Taiwan

University, Department of Bio-Industrial Mechatronics Engineering. In Chinese.

Liu, F., Y. He, and L. Wang. 2008. Determination of effective wavelengths for

discrimination of fruit vinegars using near infrared spectroscopy and multivariate

analysis. *Anal. Chim. Acta* 615(1): 10-17.

Liu, F., Y. Jiang, and Y. He. 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(1): 45-52.

Maleki, M. R., A. M. Mouazen, H. Ramon, and J. De Baerdemaeker. 2007.

Multiplicative scatter correction during on-line measurement with near infrared spectroscopy. *Biosyst. Eng.* 96(3): 427-433.

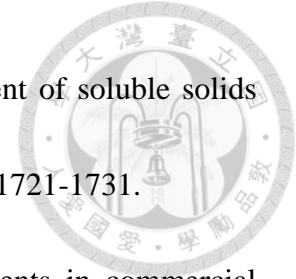
McClure, W. F. 2003. 204 years of near infrared technology: 1800-2003. *J. Near Infrared Spec.* 11(6): 487-518.

Moneruzzaman, K. M., A. M. Al-Saif, A. I. Alebidi, A. B. M. S. Hossain, O. Normaniza, and A. N. Boyce. 2011. An evaluation of the nutritional quality evaluation of three cultivars of *Syzygium samarangense* under Malaysian conditions. *Afr. J. Agric. Res.* 6(3): 545-552.

Nastasijević, B., T. Lazarević-Pašti, S. Dimitrijević-Branković, I. Pašti, A. Vujačić, G. Joksić, and V. Vasić. 2012. Inhibition of myeloperoxidase and antioxidative activity of *Gentiana lutea* extracts. *J. Pharm. Biomed. Anal.* 66: 191-196.

Paris, I., A. Janoly-Dumenil, A. Paci, L. Mercier, P. Bourget, F. Brion, P. Chaminade, and A. Rieutord. 2006. Near infrared spectroscopy and process analytical technology to master the process of busulfan paediatric capsules in a university hospital. *J. Pharm. Biomed. Anal.* 41(4): 1171-1178.

Park, B., J. A. Abbott, K. J. Lee, C. H. Choi, and K. H. Choi. 2003. Near-infrared



- diffuse reflectance for quantitative and qualitative measurement of soluble solids and firmness of delicious and gala apples. *Trans. ASAE* 46(6): 1721-1731.
- Pasadakis, N., and A. A. Kardamakis. 2006. Identifying constituents in commercial gasoline using Fourier transform-infrared spectroscopy and independent component analysis. *Anal. Chim. Acta* 578(2): 250-255.
- Porfire, A., L. Rus, A. L. Vonica, and I. Tomuta. 2012. High-throughput NIR-chemometric methods for determination of drug content and pharmaceutical properties of indapamide powder blends for tableting. *J. Pharm. Biomed. Anal.* 70: 301-309.
- Salgó, A., and S. Gergely. 2012. Analysis of wheat grain development using NIR spectroscopy. *J. Cereal Sci.* 56(1): 31-38.
- Shao, X., Z. Liu, and W. Cai. 2009. Extraction of chemical information from complex analytical signals by a non-negative independent component analysis. *Analyst* 134(10): 2095-2099.
- Szücs, Z., B. Dános, and Sz. Nyiredy. 2002. Comparative analysis of the underground parts of *Gentiana* species by HPLC with diode-array and mass spectrometric detection. *Chromatographia* 56(1): S-19-S-23.
- Takano, K. 1989. Studies on the mechanism of lipid-hydrolysing in rice bran. *J. Jpn. Soc. Food Sci. Technol.-Nippon Shokuhin Kagaku Kogaku Kaishi* 36(6): 519-524.

Tehrani, M., S. Chandran, A. B. M. S. Hossain, and A. Nasrulhaq-Boyce. 2011.

Postharvest physico-chemical and mechanical changes in jambu air (*Syzygium aqueum* Alston) fruits. *Aust. J. Crop Sci.* 5(1): 32-38.

Thennadil, S. N., H. Martens, and A. Kohler. 2006. Physics-based multiplicative scatter correction approaches for improving the performance of calibration models. *Appl. Spectrosc.* 60(3): 315-321.


Vittoz, E. A., and X. Arreguit. 1989. CMOS integration of Herault-Jutten cells for separation of sources. In "Analog VLSI Implementation of Neural Systems", 57-83. C. Mead and M. Ismail, eds. Massachusetts, U.S.A.: Kluwer Academic Publishers.

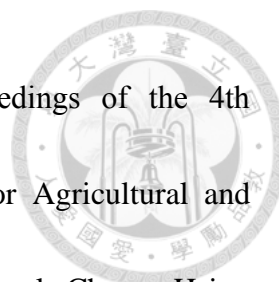
Wang, G., C. Dong, Y. Shang, Y. Sun, D. Fu, and J. Zhao. 2009. Characterization of radix rehmanniae processing procedure using FT-IR spectroscopy through nonnegative independent component analysis. *Anal Bioanal Chem* 394(3): 827-833.

Wang, L., F. S. C. Lee, and X. Wang. 2007. Near-infrared spectroscopy for classification of licorice (*Glycyrrhizia uralensis* Fisch) and prediction of the glycyrrhizic acid (GA) content. *LWT-Food Sci. Technol.* 40(1): 83-88.

Westad, F., and M. Kermit. 2003. Cross validation and uncertainty estimates in independent component analysis. *Anal. Chim. Acta* 490(1-2): 341-354.

Westad, F. 2005. Independent component analysis and regression applied on sensory

- 
- data. *J. Chemometr.* 19(3): 171-179.
- Williams, P. C., and D. C. Sobering. 1993. Comparison of commercial near-infrared transmittance and reflectance instruments for analysis of whole grains and seeds. *J. Near Infrared Spectrosc.* 1(1): 25-32.
- Wold, S. 1987. Principal component analysis. *Chemometrics Intell. Lab. Syst.* 2(1-3): 37-52.
- Wold, S., M. Sjöström, and L. Eriksson. 2001. PLS-regression: a basic tool of chemometrics. *Chemometrics Intell. Lab. Syst.* 58(2): 109-130.
- Woo, Y. A., H. J. Kim, K. R. Ze, and H. Chung. 2005. Near-infrared (NIR) spectroscopy for the non-destructive and fast determination of geographical origin of *Angelicae gigantis Radix*. *J. Pharm. Biomed. Anal.* 36(5): 955-959.
- Wu, D., S. Feng, and Y. He. 2008. Short-wave near-infrared spectroscopy of milk powder for brand identification and component analysis. *J. Dairy Sci.* 91(3): 939-949.
- Wu, Z., B. Xu, M. Du, C. Sui, X. Shi, and Y. Qiao. 2012. Validation of a NIR quantification method for the determination of chlorogenic acid in *Lonicera japonica* solution in ethanol precipitation process. *J. Pharm. Biomed. Anal.* 62: 1-6.
- Yang, I. C., S. Chen, C. Y. Wang, Y. S. Chen, Y. F. Cheng, R. M. Hsieh, C. C. Lai, and W. C. Chen. 2008. Pharmaceutical ingredients detection of *Gentiana scabra* in



powder form using near infrared spectroscopy. In “Proceedings of the 4th International Symposium on Machinery and Mechatronics for Agricultural and Biosystems Engineering”, FE-P9. Taichung, Taiwan: National Chung Hsing University.

You, C. C. 2002. Import photo-sensor grader for wax-apple. *Journal Taiwan Agri. Mach.* 17(4): 1-4. In Chinese.

Zhang, H. L., S. H. Xue, F. Pu, R. K. Tiwari, and X. Y. Wang. 2010. Establishment of hairy root lines and analysis of gentiopicroside in the medicinal plant *Gentiana macrophylla*. *Russ. J. Plant Physiol.* 57(1): 110-117.

Zhang, Z., and Y. Tang. 2005. Identification of rhubarb samples by using NIR spectrometry and Takagi-Sugeno fuzzy system. *Spectr. Lett.* 38(4-5): 447-457.

Zhou, Z., K. Robards, S. Helliwell, and C. Blanchard. 2002. Ageing of stored rice: changes in chemical and physical attributes. *J. Cereal Sci.* 35(1): 65-78.

Zou, X., and J. Zhao. 2006. Improving of prediction model for apple sugar content using independent component analysis. *Chin. J. Anal. Chem.* 34(9): 1291-1294. In Chinese.