JOURNAL OF FOOD AND DRUG ANALYSIS 22 (2014) 336–344

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.jfda-online.com

Original Article

Integration of independent component analysis with near-infrared spectroscopy for analysis of bioactive components in the medicinal plant *Gentiana scabra* Bunge

Yung-Kun Chuang [a,b], I-Chang Yang [b], Yangming Martin Lo [b], Chao-Yin Tsai [c], Suming Chen [a,c,*]

[a] Department of Bio-Industrial Mechatronics Engineering, National Taiwan University, Taipei, Taiwan
[b] Department of Nutrition and Food Science, University of Maryland, College Park, Maryland, USA
[c] Bioenergy Research Center, National Taiwan University, Taipei, Taiwan

ARTICLE INFO

Article history:
Received 5 June 2013
Received in revised form
24 July 2013
Accepted 25 September 2013
Available online 17 February 2014

Keywords:
Gentiana scabra Bunge
Gentiopicroside
Independent component analysis
Near-infrared spectroscopy
Swertiamarin

ABSTRACT

Independent component (IC) analysis was applied to near-infrared spectroscopy for analysis of gentiopicroside and swertiamarin; the two bioactive components of *Gentiana scabra* Bunge. ICs that are highly correlated with 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 [two-dimensional (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 second derivative spectra produced the best effect in the wavelength ranges of 600–700 nm, 1600–1700 nm, and 2000–2300 nm (correlation coefficient of calibration = 0.847, standard error of calibration = 0.865%, and standard error of validation = 0.909%). For swertiamarin, a spectral calibration model based on the first derivative spectra produced the best effect in the wavelength ranges of 600–800 nm and 2200–2300 nm (correlation coefficient of calibration = 0.948, standard error of calibration = 0.168%, and standard error of validation = 0.216%). Both models showed a satisfactory predictability. This study successfully established qualitative and quantitative correlations for gentiopicroside and swertiamarin with near-infrared spectra, enabling rapid and accurate inspection on the bioactive components of *G. scabra* Bunge at different growth stages.

* *Corresponding author*. Department of Bio-Industrial Mechatronics Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei 10617, Taiwan.
E-mail address: schen@ntu.edu.tw (S. Chen).

http://dx.doi.org/10.1016/j.jfda.2014.01.021

1. Introduction

Medicinal plants have always been considered an important and reliable source for pharmacy, because 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. *G. scabra* Bunge has been proven to achieve good pharmacological effects, and its dried root and rootstock are commonly used as pharmaceutical raw materials, because they are rich in many secondary metabolites such as gentiopicroside, swertiamarin, and sweroside [1]. 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 [2].

In the early days, *G. scabra* Bunge was mainly collected in the wild. Increasing demand for *G. scabra* Bunge, and changes in the natural environment and climate, have rapidly diminished the wild sources, thus restoration of *G. scabra* Bunge has become an important issue in order to protect and sustainably utilize rare plants [3]. Studies in recent years have used tissue culture technology to make artificial cultivation of *G. scabra* Bunge [4], by domesticating the tissue culture of *G. scabra* Bunge, then transplanting it to the greenhouse for cultivation. In order to monitor the change in *G. scabra* Bunge during the growth process, it is necessary to measure its bioactive components. However, the commonly used methods such as micellar electrokinetic capillary chromatography [5] and liquid chromatography–mass spectrometry [6] are all time-consuming and energy intensive, hence they cannot be used for daily quality inspection of *G. scabra* Bunge during cultivation.

Near-infrared (NIR) spectroscopy is a nondestructive inspection method that has been widely used in dispensation [7–9]. Generally, the 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 [10] and identify components of a mixture effectively [11,12]. In recent years, ICA has been used in medicinal tests [13–15] and agriculture [16]. However, previous research [16] has only focused on the quantitative analysis of one constituent in intact fruit in the ICA with NIR spectral analysis. There has not been any study applying ICA with NIR spectroscopy to inspection of ICs of *G. scabra* Bunge, thus, it was the intention of the present study to apply ICA, which can analyze multiple components simultaneously, to NIR spectroscopy of gentiopicroside and swertiamarin, to discuss the qualitative and quantitative relationships of the two bioactive components. The qualitative discrimination of various parts (tissue culture, shoots, and roots) and the target constituents (gentiopicroside and swertiamarin) within the plants was attempted. Efforts were also made to build spectral calibration models with high predictability in order to evaluate the quality of *G. scabra* Bunge. The predictability of ICA of gentiopicroside and swertiamarin constituents was also investigated in this study.

2. Materials and methods

2.1. *G. scabra Bunge sample preparation*

Samples of *G. scabra* Bunge were provided by Taiwan Sugar Research Institute (Tainan, Taiwan). A total of 94 tissue culture samples and 68 grown plant samples of different cultivation time were acquired. The shoots and roots 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), and 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 of *G. scabra* Bunge.

2.2. *NIR spectra and high-performance liquid chromatography measurement*

Dry powder of *G. scabra* Bunge was gently poured into a small ring cup (5 cm internal diameter) and subjected to NIR measurement (NIRS 6500; FOSS NIRSystems, Laurel, MD, USA). The reflectance spectra of the samples ranged from 400 nm to 2498 nm with 2-nm intervals, and the NIR spectrum of each sample was the average of 32 scans.

To attain the reference value of the two bioactive components, we measured gentiopicroside and swertiamarin using high-performance liquid chromatography (HPLC; DX 500 ion chromatograph; Dionex Corporation, Sunnyvale, CA, USA) equipped with a DIONEX C18 column (250 mm × 4.6 mm internal diameter). 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/minute. A high-precision scale was used to measure the gentiopicroside and swertiamarin standard powder, which was diluted into 1000 ppm, 500 ppm, 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 [17].

2.3. *Data analysis*

After the reflectance spectra of *G. scabra* Bunge powder and the contents of the two bioactive components of *G. 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.

2.3.1. *ICA*

ICA is a method used to transform the observed multivariate data to statistically independent components 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 [10]. 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 (M) or its inverse, the separating matrix (W). The unknown source, s, is approximated as:

$$\hat{s} = Wx = M^{-1}Ms \quad (1)$$

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

In the present study, a JADE (joint approximate diagonalization of eigenmatrices) algorithm [18,19] was used to conduct ICA. The more detailed data analysis procedure was introduced in the previous study of our research group [16].

2.3.2. *Spectral pretreatment*

Inevitable light scattering can be added to the spectra when using NIR to measure powder samples, especially when the particle size is not uniform, therefore, multiplicative scatter correction (MSC) was used to allow additive and multiplicative transformation of the spectra. After MSC treatment, the spectra of *G. scabra* Bunge powder not only reduced the physical impact of nonuniform particles [20], but also confirmed the linearity of the spectral information [21], which contributed to subsequent linear regression analysis [22].

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

2.3.3. *Model establishment*

This research used MATLAB version 7.5.0 (MathWorks, Natick, MA, USA) to edit the program of ICA spectrum analysis. The ICA procedure included: (1) selecting a calibration set and validation set; (2) spectral pretreatment; (3) selecting the specific wavelength regions; and (4) determining the best calibration model. The 207 effective samples were divided into 138 calibration samples and 69 validation samples at the ratio of 2:1. A total of 69 subsets from the 207 samples were made at first by taking every three samples to form each subset. Samples 1 and 3 from the first subset were assigned to the calibration set and Sample 2 to the validation set, and subsequent subsets followed the same alternating sequence to make both calibration and validation sets.

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 validation sets. The predictability of the models was evaluated based on the following statistical parameters, including correlation coefficient of calibration (R_c), standard error of calibration (SEC), standard error of the validation (SEV), and bias. When selecting the best calibration model, in order to avoid overfitting caused by use of excessive ICs, the following principles were adhered to: (1) the maximum number of ICs was 10% of the number of calibration sets + 2–3; (2) stop if the addition of a new IC increased the SEV; and (3) when the SEV was lower than the SEC, stop adding new ICs.

3. Results and discussion

3.1. *Distributions of the target constituents in Gentiana scabra Bunge*

Table 1 shows the bioactive components of 230 *G. scabra* Bunge samples including 94 tissue cultures, 68 shoots, and 68 roots. The gentiopicroside content was significantly higher than that of swertiamarin in all parts of the samples studied, indicating the dominance of gentiopicroside as the main bioactive component in *G. scabra* Bunge. It is interesting to note that gentiopicroside was more abundant in the whole grown plant (including shoots and roots) 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 roots was higher than in the shoots, indicating that gentiopicroside was mainly stored in the roots when the grown plant of *G. scabra* Bunge was cultivated in the greenhouse. By contrast, swertiamarin in the whole grown plant was 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. The level of swertiamarin in the shoots and roots was low, therefore, it is reasonable to postulate that swertiamarin might have been distributed evenly in the stem nodes, shoots, and roots when the grown plant of *G. scabra* Bunge was cultivated in the greenhouse.

3.2. *Correlation between NIR spectra and target constituent content*

After eliminating 10% of the outliers (23 samples) from 230 *G. scabra* Bunge samples, the remaining 207 effective samples

Table 1 – Contents and distributions of the target constituents in *Gentiana scabra* Bunge

Sample	No.	Gentiopicroside content (%)			Swertiamarin content (%)		
		Mean (range)	SD	CV	Mean (range)	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

CV = coefficient of variation; SD = standard deviation.

Table 2 – Target constituent content of calibration set and validation set in *Gentiana scabra* Bunge.

Calibration set						Validation set		
Sample no.	Gentiopicroside content (%)	Swertiamarin content (%)	Sample no.	Gentiopicroside content (%)	Swertiamarin content (%)	Sample no.	Gentiopicroside content (%)	Swertiamarin content (%)
1	1.59	0.12	70	4.53	1.17	1	1.92	0.12
2	2.09	0.25	71	4.55	0.21	2	2.10	0.16
3	2.18	0.17	72	4.60	0.14	3	2.35	0.28
4	2.32	0.13	73	4.61	0.92	4	2.51	0.29
5	2.37	0.19	74	4.63	0.83	5	2.52	0.14
6	2.39	0.16	75	4.70	0.78	6	2.55	0.18
7	2.59	0.22	76	4.74	0.18	7	2.64	0.29
8	2.61	0.23	77	4.82	0.24	8	2.76	0.22
9	2.77	0.62	78	4.83	0.29	9	2.99	0.24
10	2.79	0.34	79	4.91	0.86	10	3.12	0.25
11	2.85	0.26	80	4.93	1.16	11	3.15	0.24
12	2.88	0.31	81	4.95	1.02	12	3.22	0.27
13	2.89	0.23	82	5.10	1.36	13	3.23	0.29
14	2.90	0.32	83	5.16	1.00	14	3.23	1.24
15	2.93	0.29	84	5.16	0.29	15	3.33	0.18
16	2.95	0.28	85	5.16	1.18	16	3.43	0.32
17	3.06	0.14	86	5.18	1.01	17	3.53	0.25
18	3.09	0.22	87	5.19	0.51	18	3.80	0.22
19	3.13	1.10	88	5.23	0.22	19	3.81	0.38
20	3.15	0.29	89	5.26	1.24	20	3.83	0.24
21	3.15	0.27	90	5.26	0.29	21	3.92	0.36
22	3.22	0.87	91	5.31	1.32	22	3.92	0.31
23	3.22	0.24	92	5.36	1.11	23	3.94	0.91
24	3.26	0.24	93	5.38	0.37	24	3.95	0.20
25	3.28	0.80	94	5.42	1.35	25	3.96	0.42
26	3.33	0.92	95	5.49	1.51	26	3.97	0.94
27	3.33	0.20	96	5.54	0.29	27	4.17	0.29
28	3.34	0.35	97	5.58	0.29	28	4.23	0.29
29	3.36	0.26	98	5.61	1.08	29	4.28	0.45
30	3.37	0.37	99	5.64	1.38	30	4.40	0.23
31	3.38	0.24	100	5.66	2.15	31	4.48	0.81
32	3.42	0.28	101	5.68	1.28	32	4.48	0.35
33	3.51	0.32	102	5.70	0.19	33	4.52	1.09
34	3.55	0.78	103	5.72	1.09	34	4.56	0.27
35	3.56	1.00	104	5.74	0.93	35	4.61	0.89
36	3.56	0.25	105	5.80	0.20	36	4.61	0.19
37	3.57	0.25	106	5.85	1.42	37	4.67	1.32
38	3.59	0.35	107	5.85	1.47	38	4.70	0.67
39	3.61	0.23	108	5.91	1.04	39	4.71	0.78
40	3.65	0.78	109	5.98	0.59	40	4.71	0.52
41	3.65	0.27	110	6.07	0.18	41	4.74	1.01
42	3.67	0.90	111	6.09	1.32	42	4.76	1.10
43	3.72	0.34	112	6.14	0.20	43	4.81	0.26
44	3.73	0.15	113	6.15	1.35	44	4.87	1.13
45	3.76	0.96	114	6.24	0.19	45	4.99	0.98
46	3.79	0.29	115	6.24	0.27	46	5.14	0.84
47	3.79	0.22	116	6.25	1.54	47	5.28	1.37
48	3.80	1.11	117	6.35	1.21	48	5.29	1.02
49	3.81	0.40	118	6.40	1.40	49	5.48	0.92
50	3.82	0.24	119	6.45	1.26	50	5.50	1.28
51	3.84	0.26	120	6.50	1.44	51	5.69	1.21
52	3.94	1.09	121	6.59	1.43	52	5.75	1.08
53	3.95	0.60	122	6.61	1.05	53	5.86	1.30
54	3.99	1.37	123	6.63	1.34	54	6.41	1.38
55	4.01	0.44	124	6.68	1.47	55	6.48	1.42
56	4.04	0.27	125	6.87	1.61	56	6.52	0.29
57	4.19	0.44	126	7.18	1.54	57	6.55	0.20
58	4.23	0.33	127	7.21	1.28	58	6.57	1.55
59	4.26	1.16	128	7.24	0.28	59	6.59	1.52
60	4.27	0.27	129	7.30	1.59	60	6.65	1.35
61	4.31	0.21	130	7.34	1.62	61	6.71	1.49

(*continued on next page*)

Table 2 – (*continued*)

Calibration set						Validation set		
Sample no.	Gentiopicroside content (%)	Swertiamarin content (%)	Sample no.	Gentiopicroside content (%)	Swertiamarin content (%)	Sample no.	Gentiopicroside content (%)	Swertiamarin content (%)
62	4.34	0.81	131	7.39	1.23	62	6.74	0.21
63	4.34	0.46	132	7.52	0.31	63	6.88	0.31
64	4.34	0.33	133	7.91	1.58	64	6.95	0.34
65	4.35	0.94	134	7.93	0.34	65	6.97	1.17
66	4.36	0.94	135	8.01	0.33	66	6.99	0.33
67	4.37	0.39	136	8.23	1.52	67	7.24	1.46
68	4.51	0.90	137	8.27	0.29	68	8.04	1.72
69	4.51	0.52	138	8.77	0.31	69	8.19	1.59
			Min.	1.59	0.12	Min.	1.92	0.12
			Max.	8.77	2.15	Max.	8.19	1.72
			Mean	4.73	0.69	Mean	4.72	0.68
			SD	1.53	0.49	SD	1.51	0.49
			CV	0.32	0.72	CV	0.32	0.72

CV = coefficient of variation; Max. = maximum; Min. = minimum; SD = standard deviation.

were divided into 138 calibration samples and 69 validation samples in the ratio of 2:1. As shown in Table 2, the target constituent contents of the calibration set and validation set, and their statistical assessments in each data set are listed. The levels of gentiopicroside and swertiamarin in all the *G. scabra* Bunge powder samples were measured by HPLC as the reference values for NIR analysis. All samples of the two data sets were ranked in ascendance according to their gentiopicroside concentrations. There were obvious differences in the gentiopicroside and swertiamarin levels between the calibration and validation sets. The distributions of gentiopicroside and swertiamarin contents of the two sample groups were consistent.

The NIR spectra of the 207 *G. scabra* Bunge samples were acquired using MSC treatment. As shown in Fig. 1A, absorption peaks were found in both the visible region of blue light (452 nm) and red light (666 nm), because the chlorophyll in *G. scabra* Bunge absorbs the majority of blue and red light when involved in photosynthesis. The spectra of tissue culture and the shoots were similar, which could be attributed to the fact that during the domestication period, the tissue is mainly composed of shoots, because the root development of *G. scabra* Bunge is not obvious at that time. By contrast, the root spectra in the visible region showed a significant difference, with high absorption occurring from the green to yellow light (from 492 nm to 586 nm) and low absorption (flat waveform) from the orange to red light (from 606 nm to 700 nm). This could have been due to a lack of chlorophyll in the roots of *G. scabra* Bunge, hence reducing the absorption of the blue and red light, while reflecting the green light.

After MSC treatment, the spectra of *G. scabra* Bunge were analyzed using the following pretreatments: (1) smoothing; (2) smoothing with first derivative; and (3) smoothing with second 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 <10, indicating that the NIRS 6500 spectrophotometer was stable, and the spectra of *G. scabra* Bunge powder exhibited minimal noise.

The correlation between the spectra of *G. scabra* Bunge powder and the bioactive components were assessed at first when selecting specific wavelength regions of spectra. As for the original spectra, the first derivative spectra and the second derivative spectra, the correlation coefficients of gentiopicroside of effective samples were distributed as shown in Fig. 1B, 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 *G. scabra* Bunge powder had been eliminated, it was not necessary to avoid the O-H bond absorption band around 1450 nm 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 first derivative spectra were located throughout the regions of red light, the fourth overtone of the C-H bond, combination of the first overtone of the C-H bond, and combination between the C-H bonds. By contrast, the second derivative spectra were located in the regions of the red light, the fourth overtone of the C-H bond, first overtone of the C-H bond, and combination between the N-H bond and O-H bond.

The correlation coefficients between the spectra of *G. scabra* Bunge powder and swertiamarin are shown in Fig. 1C, with a threshold value ($|r| > 0.75$) set to determine the degree of correlation. The original spectra were located in different regions, including the red light, the first overtone of the C-H bond, combination between the N-H bond and O-H bond, and combination between the C-H bond and C-C bond. The first derivative spectra were located in the regions of the fourth overtone of the C-H bond, second overtone of the N-H bond, second overtone of the C-H bond, combination of the first overtone of the C-H bond, first overtone of the C-H bond, and combination between the C-H bond and C-C bond, whereas the second derivative spectra were located in the red light and the fourth overtone of the C-H bond regions. As indicated in Fig. 1B and C, the fourth overtone of the 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 the tissue culture, shoots, and roots.

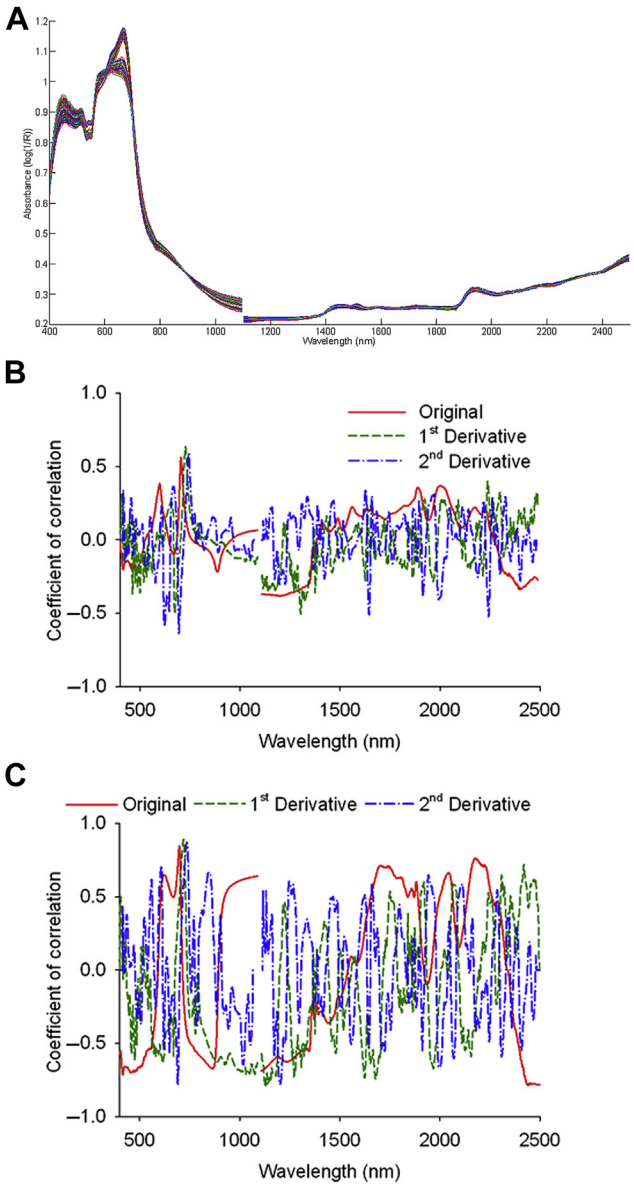

Figure 1 – (A) The spectra of *Gentiana scabra* Bunge powder post-multiplicative scatter correction. (B) Correlation coefficient distributions between the spectra and gentiopicroside. (C) Correlation coefficient distributions between the spectra and swertiamarin.

3.3. *NIR spectra decomposition and ICA of the target constituents*

According to the definition of ICA, the observed signal of the receiver can be decomposed into ICs of which the number is the same as that of training samples at most [10]. In order to avoid overfitting of the calibration model caused by use of excessive ICs, appropriate ICs were selected under the condition that calibration models were built only by using 1–17 ICs when ICA was conducted for the original spectra (400–2498 nm) of the calibration set. The SEV of the calibration models continued to drop and then increase when seven ICs were applied, indicating that incorporation of more ICs is not necessarily helpful to the analysis because it is sufficient to decompose the spectra into seven ICs.

After the original spectra (400–2498 nm) of the calibration set was decomposed into seven 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 nm and 740 nm, and IC 6 in the wavelengths of 494 nm, 1838 nm, 1944 nm, 2058 nm, and 2132 nm (Fig. 2), which was consistent with the absorption bands seen in Fig. 1B and C. This suggests that the spectral characteristics of gentiopicroside and swertiamarin were mainly reflected in ICs 4–6 [11,12]. These wavelengths will be taken as the reference for selection on a specific wavelength region of spectra when building calibration models.

As shown in Equation (1), the mixing matrix contained concentration information of the two bioactive components in each sample. The spectral information of gentiopicroside and swertiamarin was mainly reflected in ICs 4 and 5, thus, the values of these two ICs in the mixing matrix were used to configure two-dimensional (2D) distributions. As can be seen in Fig. 3A and B, tissue culture, shoots, and roots were distributed in three distinct locations of the IC 4–5 space. The values of tissue culture and shoots were close to each other and the roots presented a higher value in IC 5, showing differences among different parts of *G. scabra* Bunge in the spectra, which was consistent with the result in Fig. 1A. If the average contents of gentiopicroside and swertiamarin were taken as the threshold values, the samples could be classified into four groups: (1) gentiopicroside and swertiamarin at high concentrations; (2) gentiopicroside at high concentration and swertiamarin at low concentration; (3) gentiopicroside at low concentration and swertiamarin at high concentration; and (4) gentiopicroside and swertiamarin at low concentrations. The distributions of the calibration and validation sets in the IC

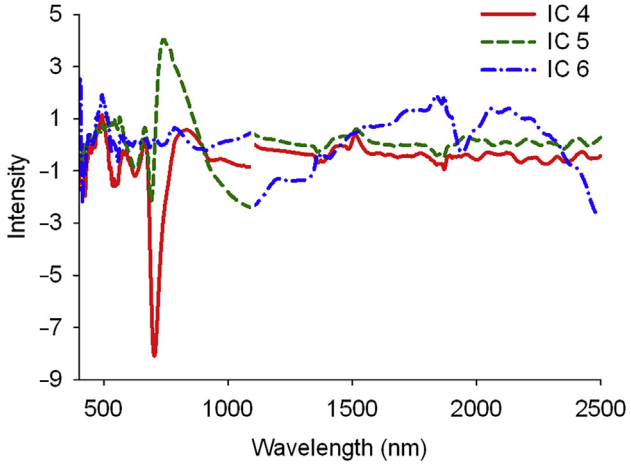

Figure 2 – The three independent components decomposed from the original spectra of *Gentiana scabra* Bunge powder post-multiplicative scatter correction that had higher correlation with gentiopicroside and swertiamarin.

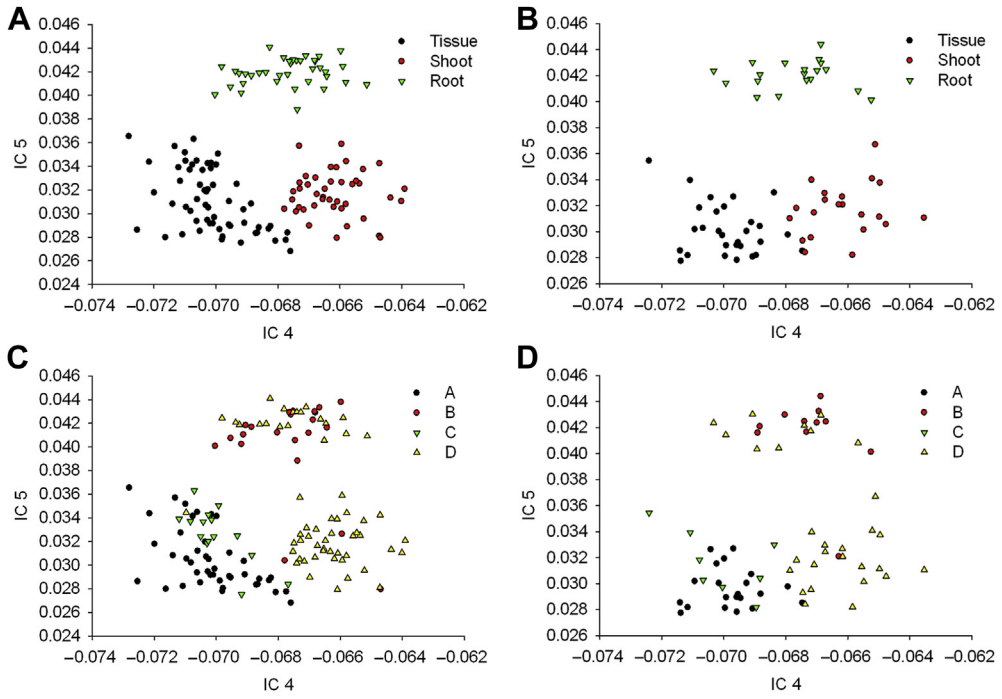

Figure 3 – Scores of tissue culture, shoots, and roots in IC 4–5 space established with calibration samples: (A) calibration set; (B) validation set. Scores of gentiopicroside and swertiamarin in IC 4–5 space established with calibration samples: (C) calibration set; (D) validation set. IC = independent component.

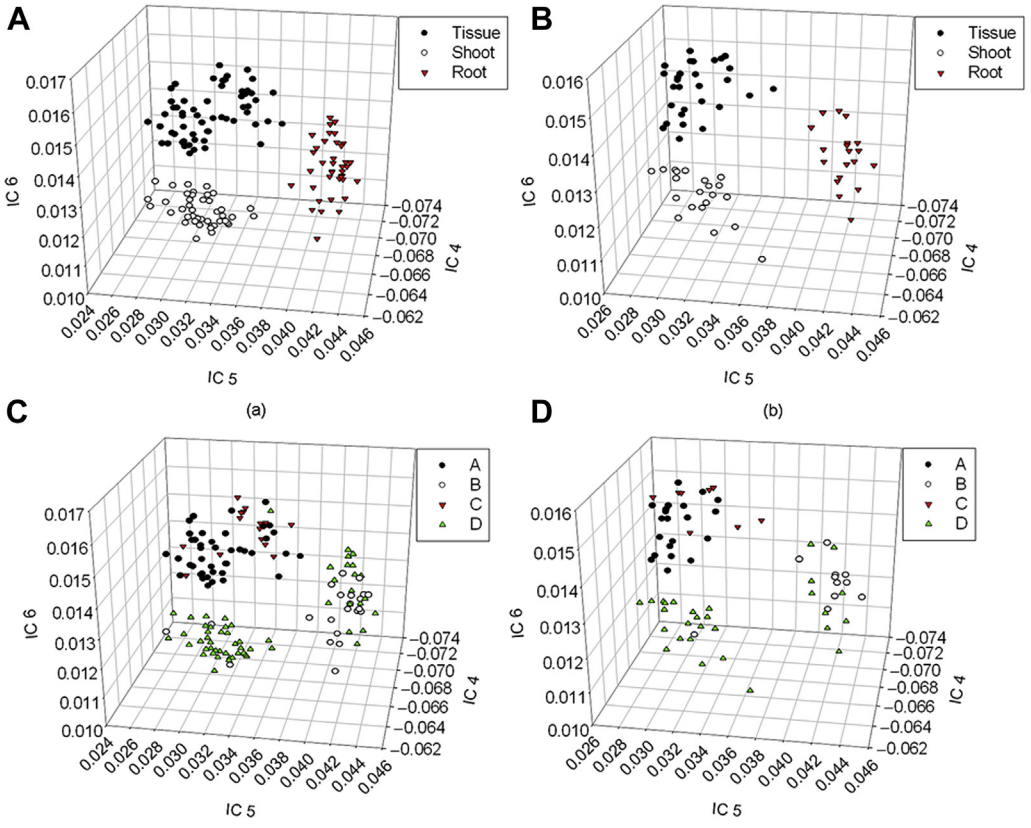

Figure 4 – Scores of tissue culture, shoots, and roots in IC 4–6 space established with calibration samples: (A) calibration set; (B) validation set. Scores of gentiopicroside and swertiamarin in IC 4–6 space established with calibration samples: (C) calibration set; (D) validation set.

Table 3 – Prediction of target constituent content in *Gentiana scabra* Bunge by ICA models.

Constituent	Spectrum	Smoothing points/gap	Wavelength range (nm), interval	Calibration set (138)		Validation set (69)	
				R_c	SEC (%)	SEV (%)	Bias (%)
Gentiopicroside	2nd derivative	6/6	600–700, 2 1600–1700, 2 2000–2300, 2	0.847	0.865	0.909	−0.395
Swertiamarin	1st derivative	2/2	600–800, 2 2200–2300, 2	0.948	0.168	0.216	0.003

ICA = independent component analysis; R_c = correlation coefficient of calibration; SEC = standard error of calibration; SEV = standard error of validation.

4–5 space are shown in Fig. 3C and D; of which, the gentiopicroside concentration in most tissue cultures was higher than the mean value, suggesting that production of gentiopicroside of *G. scabra* Bunge was sufficient during the domestication period. As the grown plants of *G. scabra* Bunge were collected at different growth stages, their gentiopicroside content in the roots varied. The gentiopicroside content in the shoots was low, indicating that gentiopicroside was mainly stored in the roots of *G. scabra* Bunge during greenhouse cultivation. By contrast, the swertiamarin content in tissue culture was higher than the mean value, but lower than the mean value in shoots and roots, indicating that swertiamarin in *G. scabra* Bunge was reduced during greenhouse cultivation; therefore, it is preferable to extract swertiamarin from tissue culture.

According to the foregoing discussion, IC 6 also contained spectral information about gentiopicroside and swertiamarin; therefore, the values of ICs 4–6 in the mixing matrix were used for 3D plotting. As shown in Fig. 4A and B, tissue culture, shoots, and roots were clearly distributed in three locations of the IC 4–6 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 the calibration and validation sets in the IC 4–6 space could be constructed, as shown in Fig. 4C and 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. Figs. 3 and 4 indicate that the differences among various parts (tissue culture, shoots, and roots) of *G. scabra* Bunge could be clearly identified by the change in the trend of the two bioactive components (gentiopicroside and swertiamarin) from the 2D and 3D spaces of ICs, making the information useful in qualitative and quantitative analyses of NIR spectroscopy.

The ICA results of the two bioactive components are shown in Table 3. The best spectral calibration model of gentiopicroside was attained when applying the second derivative spectra, of which the smoothing points and the gap were both 6 and the wavelength ranged 600–700 nm, 1600–1700 nm, and 2000–2300 nm (R_c = 0.847, SEC = 0.865%, SEV = 0.909%, and bias = −0.395%). With regard to swertiamarin, the best spectral calibration model was acquired with the first derivative spectra, of which the smoothing points and the gap were both 2 and the wavelength ranged 600–800 nm and 2200–2300 nm (R_c = 0.948, SEC = 0.168%, SEV = 0.216%, and bias = 0.003%). Satisfied outcomes were acquired for both gentiopicroside and swertiamarin. The

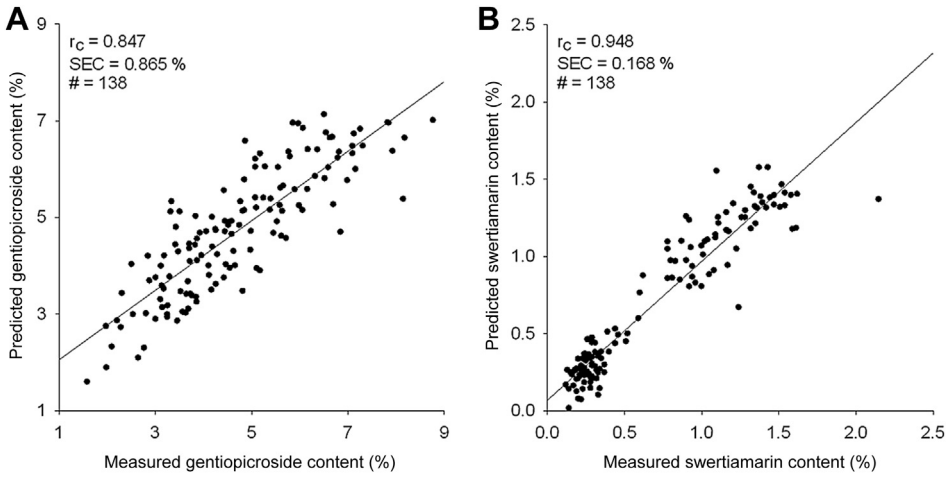

Figure 5 – Relationship between the estimated contents and the reference contents of (A) gentiopicroside and (B) swertiamarin in *Gentiana scabra* Bunge.

relationship between the predicted and reference concentrations of both bioactive components is shown in Fig. 5. The content of gentiopicroside predicted by the calibration model was mainly affected by bias, therefore, 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 *G. scabra* Bunge is relatively low, which is also why the quantity and equitability of *G. scabra* Bunge powder are both important.

In conclusion, this study applied ICA in NIR spectroscopy analysis of gentiopicroside and swertiamarin – bioactive components of *G. scabra* Bunge – and discussed relevant tissue culture and grown plants (including shoots and roots). Both qualitative and quantitative analyses of the bioactive components in *G. scabra* Bunge were performed. Qualitative discrimination of various parts (tissue culture, shoots, and roots) and the target constituents (gentiopicroside and swertiamarin) in the plants was conducted and presented. By selecting ICs that were highly correlated to the bioactive components, the 2D and 3D spaces of ICs clearly illustrated the distribution of gentiopicroside and swertiamarin in different parts of *G. scabra* Bunge. The predictability of ICA for dealing with various constituents at the same time was evaluated by analysis of gentiopicroside and swertiamarin, and satisfactory results were obtained. Therefore, by combining ICA with NIR spectroscopy, fast and accurate inspection of gentiopicroside and swertiamarin in *G. scabra* Bunge at different growth stages could be achieved. This technology could contribute substantially to the quality management of *G. scabra* Bunge during and after cultivation.

Conflicts of interest

All authors declare no conflicts of interest.

Acknowledgments

We thank Mr. Yu-Fan Cheng, Ms. Ching-Yin Wang, Mr. Cheng-Wei Huang, Mr. Yu-Song Chen, and Mr. Chun-Chi Chen for their assistance.

REFERENCES

[1] Kakuda R, Iijima T, Yaoita Y, et al. Secoiridoid glycosides from *Gentiana scabra*. J Nat Prod 2001;64:1574–5.
[2] Kim JA, Son NS, Son JK, et al. Two new secoiridoid glycosides from the rhizomes of *Gentiana scabra* Bunge. Arch Pharm Res 2009;32:863–7.
[3] Zhang HL, Xue SH, Pu F, et al. Establishment of hairy root lines and analysis of gentiopicroside in the medicinal plant *Gentiana macrophylla*. Russ J Plant Physiol 2010;57:110–7.
[4] Cai Y, Liu Y, Liu Z, et al. 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 2009;45:730–9.
[5] Glatz Z, Pospísilová J, Musil P. Determination of gentiopicroside in extracts of *Centaurium erythreae* and *Gentiana lutea* by micellar electrokinetic capillary chromatography. J Liq Chromatogr Relat Technol 2000;23:1831–9.
[6] Aberham A, Pieri V, Croom Jr EM, et al. Analysis of iridoids, secoiridoids and xanthones in *Centaurium erythraea*, *Frasera caroliniensis* and *Gentiana lutea* using LC-MS and RP-HPLC. J Pharm Biomed Anal 2011;54:517–25.
[7] Zhang Z, Tang Y. Identification of rhubarb samples by using NIR spectrometry and Takagi-Sugeno fuzzy system. Spectr Lett 2005;38:447–57.
[8] Wang L, Lee FSC, Wang X. Near-infrared spectroscopy for classification of licorice (*Glycyrrhizia uralensis* Fisch) and prediction of the glycyrrhizic acid (GA) content. LWT-Food Sci Technol 2007;40:83–8.
[9] Chen X, Wu D, He Y, et al. Nondestructive differentiation of panax species using visible and shortwave near-infrared spectroscopy. Food Bioprocess Technol 2011;4:753–61.
[10] Hyvärinen A, Oja E. Independent component analysis: algorithms and applications. Neural Netw 2000;13:411–30.
[11] Pasadakis N, Kardamakis AA. Identifying constituents in commercial gasoline using Fourier transform-infrared spectroscopy and independent component analysis. Anal Chim Acta 2006;578:250–5.
[12] Kardamakis AA, Mouchtaris A, Pasadakis N. Linear predictive spectral coding and independent component analysis in identifying gasoline constituents using infrared spectroscopy. Chemometrics Intell Lab Syst 2007;89:51–8.
[13] Fang LM, Lin M. Prediction of active substance contents in pharmaceutical tablet using ICA and NIR. Acta Chim Sin 2008;66:1791–5 [In Chinese, English abstract].
[14] Wang G, Dong C, Shang Y, et al. Characterization of radix rehmanniae processing procedure using FT-IR spectroscopy through nonnegative independent component analysis. Anal Bioanal Chem 2009;394:827–33.
[15] Shao X, Liu Z, Cai W. Extraction of chemical information from complex analytical signals by a non-negative independent component analysis. Analyst 2009;134:2095–9.
[16] Chuang YK, Chen S, Lo YM, et al. 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 2012;20:855–64.
[17] Chen S, Chuang YK, Tsai CY, et al. Near infrared analysis of biomaterials using independent component analysis. In: Lu FM, Chou JJ, Lee JC, editors. Proceedings of the 6th International Workshop on Nondestructive Quality Evaluation of Agricultural, Livestock and Fishery Products. Taipei: Taiwan Agricultural Mechanization Research and Development Center; 2010. pp. 32–44.
[18] Cardoso JF, Souloumiac A. Blind beamforming for non-Gaussian signals. IEE Proc F 1993;140:362–70.
[19] Cardoso JF. High-order contrasts for independent component analysis. Neural Comput 1999;11:157–92.
[20] Helland IS, Naes T, Isaksson T. Related versions of the multiplicative scatter correction method for preprocessing spectroscopic data. Chemometrics Intell Lab Syst 1995;29:233–41.
[21] Isaksson T, Naes T. The effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy. Appl Spectrosc 1988;42:1273–84.
[22] Thennadil SN, Martens H, Kohler A. Physics-based multiplicative scatter correction approaches for improving the performance of calibration models. Appl Spectrosc 2006;60:315–21.