Classification of cultivation years of ginseng using near infrared reflectance spectroscopy

Young-Ah Woo,*[a] Soo-Hyun Chang,[a] Hyo-Jin Kim[a] and Junghwan Cho[b]

[a]*College of Pharmacy, Dongduk Women's University, Seoul 136-714, Korea.*

[b]*College of Pharmacy, Sookmyung Women's University, Seoul 140-742, Korea.*

Introduction

Ginseng, the root of Panax ginseng C.A. Meyer, is one of the most widely used medicinal plants, particularly in East Asian countries. It has a wide range of pharmacological and physiological properties. It is known that the efficacy and properties change due to ginseng saponin based on the cultivation years of ginseng. Results have been obtained from different compositions or increases of ginseng saponin.[1,2] So far, the evaluation of cultivation years of ginseng has been dependent on visual inspection, based on the appearance of ginseng. The results of this method are different according to the visual inspector. Therefore, a rapid and accurate classification method is necessary. It has not been proven that the classification can be performed using chemical properties such as ginseng saponin or physical properties such as thickness of cork. Therefore, nondestructive and accurate analytical instrumentation for acquiring these properties simultaneously is required for the classification of the cultivation years of ginseng.

Near-infrared (NIR) spectroscopy has been applied in qualitative and quantitative analysis. NIR spectra can be utilised for classification of complex substances using pattern recognition techniques.[3] Principal component analysis (PCA)[4] and discriminant partial least squares (PLS)[5] were widely used for classification. PLS arecommonly used multivariate statistical methods for spectral analysis. Similar to PCA, the discriminant PLS creates new factors that account for most variances in the data set.[6] Discriminant PLS differs from PCA in that the former utilises the known group association of samples during the factor extracting process.[7] Furthermore, soft independent modelling of class analogy (SIMCA) is also utilised in this study. SIMCA has been already used combined with NIR data for qualitative analysis. If NIR reflectance spectroscopy using these pattern recognition techniques is successfully performed, it provides rapid and nondestructive method for classification of ginseng.

Experimental

Sample preparation

Ginseng radix, roots of *Panax ginseng* C.A. Meyer, was studied. All ginseng radix samples were acquired from The Experimental Station of the Natural Agriculture Products Inspection Office (NAPIO), Seoul, Korea. In this study, 188, 219 and 264 samples were collected for 4-, 5- and 6-year-old ginseng, respectively. The entire samples were dried when acquired. All samples were

Near Infrared Spectroscopy: Proceedings of the 9th International Conference

powdered using a cyclone mill (Udymill, USA) fitted with 1mm screen. The particle size of powder was controlled below 20 meshes.

NIR reflectance spectra

Visible-NIR reflectance spectra were collected over the 400 to 2500 nm spectral region with an NIRSystems model 6500 spectrometer (Foss NIRSystems, MD, USA) equipped with a quartz halogen lamp, silicon detector (for the visible region) and PbS detector (for the NIR region). The spectra were collected with 2 nm data intervals. The spectra were acquired with a circular sample cup with a quartz window (38 mm in diameter and 10 mm in thickness). Each sample spectrum was obtained by averaging 32 scans. All the spectra were recorded as log R^{-1} with respect to a ceramic reference standard.

Results and discussion

Near infrared (NIR) reflectance spectra features

The reflectance spectra were acquired over 400–2500 nm range including the visible and NIR regions. Visible ranges were included to identify the colour changes in samples. This can be used as one of the factors for classification, since the colour of ginseng samples are slightly different. Figure 1 shows the average NIR reflectance spectra of ginseng according to cultivation years. There are no significant differences in the spectra besides the intensity. In NIR reflectance spectra, the difference of intensity is due to particle size effect. Therefore, a derivative algorithm was used to enhance the spectral features. Figure 2 shows the first derivative spectra. The spectra

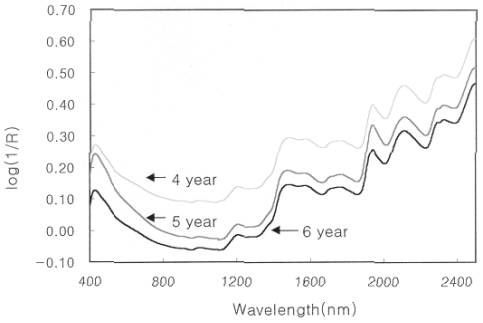

Figure 1. Average of NIR reflectance of 4-year-old, 5-year-old and 6-year-old ginseng samples.

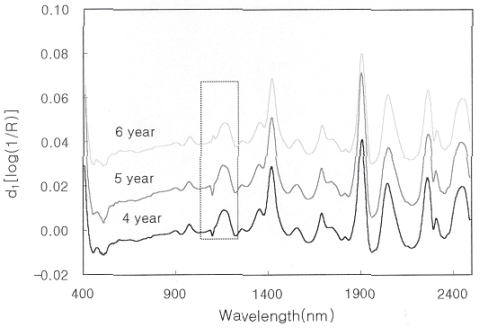

Figure 2. The first derivative spectra of Figure 1.

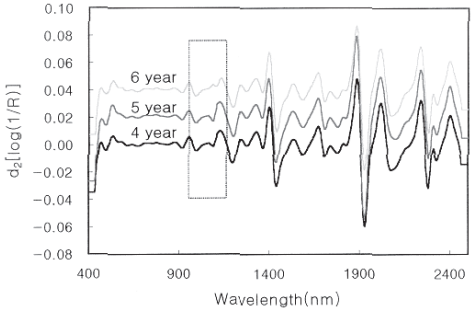

Figure 3. The second derivative spectra of Figure 2.

were linearly offset to enhance qualitative comparison. Around 1050 nm, the spectral differences, based on the cultivation year, are shown. As with the first derivative spectra, it is observed that the spectral differences occur in the same region of the second derivative spectra, which were linearly offset to enhance qualitative comparison in Figure 3. The qualitative properties, including visible spectral differences in NIR reflectance, can be utilised using a factor analysis such as principal component analysis, discriminant partial least squares (PLS) and soft independent modelling of class analogy (SIMCA).

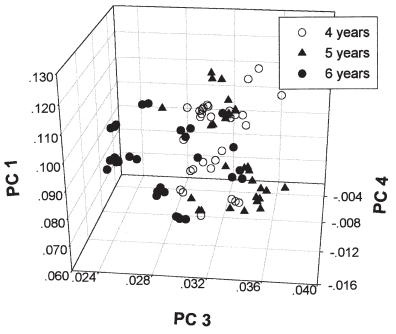

Figure 4. Three-dimensional score plot using PC1, PC3 and PC4.

Principal component analysis

Using significant factors for reconstructing a model of NIR reflectance spectra, the dimensions of spectra are reduced when principal component analysis is performed. Figure 4 shows the three-dimensional (3-D) score plot using 1st, 3rd and 4th PCs, using the full 400–2500 nm range of visible-NIR spectra. Thirty data were used for each 4- year, 5-year and 6-year samples in PCA analysis. The classification was not clear. However, it can be observed there are trends of the cultivation years of ginseng samples.

Discriminant partial least squares

Discriminant PLS is based on the projection of multivariate data sets into far fewer dimensions that are most relevant to the known associated group of samples. This analysis was performed as supervised learning, which have the priori knowledge of membership of class. Discriminant PLS utilises not only the spectral data itself but also the external knowledge of the geographical origin of samples. This method regards the whole data as two different groups based on external knowledge and places one group as 1 and the other group as 2. 2 corresponds with the sample in the desired group and 1 is for the sample not in the group. The general concept is that a predicted value of 2.0 is regarded as a perfect match, 1.0 as not in the group and 1.5 as uncertain, which means that it can go either way. This is performed with a confidence level of 95% in a t-test. In discriminant PLS, a training set for modelling and a test set for estimating the classification rule's performance is required. Data are divided into a training set and a test set with random selection as listed in Table 1.

In order to remove baseline drift, first and second derivatives of spectra were utilised. Then the full range of 400–2500 nm was used, including the visible and NIR regions. In order to find the optimal PLS2 model, first and second derivative algorithms were utilised and the used factor for each model is

Table 1. Data set preparation for discriminant PLS.

Sample	Number of spectra in training set	Number of spectra in test set
4 year	130	58
5 year	150	69
6 year	185	79

Table 2. Modeling parameters for discriminant PLS.

Model	Derivatisation	Spectral region (nm)	Number of factors
1	None	400–2500	18
2	First derivatives	400–2500	20
3	Second derivatives	400–2500	19

Table 3. Data set preparation and modeling parameters for SIMCA.

Sample	Number of spectra in training set	Number of spectra in test set	Spectral range for modelling (nm)	Principal component for modelling
4 year	50	138	400–2500	12
5 year	50	169	400–2500	10
6 year	50	214	400–2500	14

Table 4. Modelling by discriminant PLS.

Model		4	5	6	Total
Model 1	4	119	11	0	130
	5	9	141	0	150
	6	4	8	173	185
Model 2	4	129	1	0	130
	5	1	148	1	150
	6	0	0	185	185
Model 3	4	127	3	0	130
	5	0	147	3	150
	6	0	0	185	185

listed in Table 2. Models 1, 2 and 3 are the PLS model with training set using raw spectra, first derivative and second derivative spectra, respectively. Table 4 shows the calibration results for each model. According to the results, Model 2, using first derivative spectra, gave the best calibration results compared with using raw and second derivative spectra. There are 1 and 12 misses for 4-year and 5-year-old ginseng data, respectively. Six-year-old ginseng data are clearly classified. The prediction results of the test set, which was composed of data not used for the training set, are shown in Table 5. As witht he calibration, Model 2, with the use of first derivative spectra, provided a clear classification.

Soft independent modelling of class analogy

The SIMCA method uses independent principal components for each class. The full range of 400–2500 nm are utilised. A training set and a test set are randomly divided, as listed in Table 3.

Table 5. Prediction results using discriminant PLS and SIMCA.

Model		4	5	6	Total
Model 1	4	51	7	0	58
	5	3	66	0	69
	6	3	3	73	79
Model 2	4	58	0	0	58
	5	0	69	0	69
	6	0	0	79	79
Model 3	4	57	1	0	58
	5	2	67	0	69
	6	0	0	79	79

Twelve, ten and 14 principal components are selected for each class of 4-, 5- and 6-year old-ginseng, respectively. Table 5 shows the prediction results for the test set not used for the training set. Four and 6 year-old ginseng data are correctly classified whereas 15 of the 5-year-old ginseng data are classified as 4-year old data. Using the SIMCA method, the first or second derivatisation algorithm gives much worse results compared with raw spectra. It is noticeable that when classification rule is evaluated using SIMCA, better results are obtained by not using derivative spectra.

Conclusions

This study shows the possibility using near infrared reflectance spectroscopy for the classification of cultivation years of ginseng. With the use of near infrared reflectance spectra, the rapid, accurate and nondestructive analysis was performed for classification. For supervised learning, pattern recognition techniques using discriminant PLS and soft independent modelling of class analogy were applied. As a result, prediction results with both discriminant PLS and SIMCA show that near-infrared spectroscopy has potential for the classification of ginseng.

References

1. M.W. Kim, S.R. Ko, K.J. Choi and S.C. Kim, *Korean J. Ginseng Sci.* **11,** 10 (1987).
2. H. Yamaguchi, M. Takada, R. Kasai, H. Matsuura, O.Tanaka and T. Fuwa, *YAKUGAKU ZASSHI* **108,** 872 (1988).
3. M.K. Bewig, A.D. Clarke, C. Roberts and N. Unklesby, *JAOCS* **71,** 195 (1994).
4. M. Shimoyama, H. Maeda, H. Sato, T. Ninomiya and Y. Ozaki, *Appl. Spectrosc*. **51,** 1154 (1997).
5. R. Vong, P. Geladi, S.Wold and K. Esbensen, *J. Chemometrics* **2,** 281 (1988).
6. H. Martens and T. Næs, *Multivariate Calibration*. John Wiley, Chichester, UK (1989).
7. M. Sjostrom and S. Wold, *SIMCA*; *Pattern Recognition Practice*. North-Holland Publishing Company; Amsterdam, The Netherlands (1980).