Abstract
Improving the differentiation of two groups by utilising scores from different spectral ranges: agricultural application

S. Lee* and H. Chung
Department of Chemistry, Hanyang University, Seoul 133-791 Korea.
E-mail: lsu19@unitel.co.kr

Introduction

Although NIR spectroscopy has been widely applied for the discrimination of geographical origin of agricultural products, continuously improving the method will ensure more accurate evaluation of agricultural products. More selective spectroscopic methods may be used to enhance product evaluation, but only by sacrificing the speed and non-destructive nature of the NIR approach. Alternatively, the use of advanced chemometric methods that effectively describe the spectral features and improve discrimination accuracy is desirable. This is practicable, since the NIR spectral data can be utilised without serious consideration of other analytical methods. The aim of this study was to evaluate possible improvement in the differentiation of two groups by utilising scores independently generated from separate spectral ranges.

Materials and methods

Three different agricultural products (sesame, cnidium officinale, and carrot) were employed in this study. All samples were collected over more than six years to account for wider compositional variations from seasonal changes and cultivation areas. Imported products were mostly from diverse local origins in China. The samples were sliced, dried, and ground into powder. The powdered samples were finally prepared by passing the ground powder through a 20-mesh sieve. The NIR spectra were collected immediately following sample preparation. NIR diffuse reflectance spectra were collected over the spectral region of 1100–2500 nm with a Foss NIRSystems Model 6500 spectrometer equipped with a quartz halogen lamp and PbS detector. The spectra were collected at 2-nm intervals and acquired with a circular sample cup with a quartz window (38 mm in diameter and 10 mm thick) by gently packing the powdered samples. Each sample spectrum was obtained by averaging 32 scans. All spectra were recorded as log 1/R with respect

to a ceramic reference standard at room temperature. All algorithms were developed using Matlab Version 7.0 (The MathWorks Inc., MA, USA).

Results and discussion

It is generally known that agricultural products, including herbal medicines, are very complex mixtures of organic and inorganic compounds. Therefore, their NIR spectra may be severely overlapping, and it is often difficult to find the spectral features of a specific component, or group, that can be employed as the discrimination index. Because of this, normally the best spectral range for differentiation is found by evaluating the resulting performances from many different spectral ranges. When two different ranges are used for differentiation, one series of scores can be obtained by combining two ranges, or two separate series of scores can be acquired from each spectral range independently. The first case is a typical procedure in NIR analysis. We have investigated the possible improvement of differentiation by using two separate series of scores. The scores were used as the input for linear discriminant analysis (LDA). Table 1 shows the results for the differentiation of two geographical origins of sesame.

When comparing the results using the full, and the optimised single range for PCA-LDA, the identification error was improved when the optimised single range was used. The differentiation was further improved when the two ranges were simultaneously used for PCA-LDA. As shown, the best accuracy was achieved when optimally chosen scores from two separate spectral ranges were used for LDA. We have processed the spectral datasets of cnidium officinale and carrot, and could improve the differentiation for these two products, as observed in Table 1. When the spectral features of two groups are closer to each other, the proposed strategy would be superior, because minute spectral difference can be recognised more effectively.

Table 1. The results for differentiation of two geographical origins of sesame.

Method	Spectral range optimised	Scores used	Identification error
Full range PCA	1100–2500 nm	1, 4	13.4%
Single range PCA	1700–1800 nm	2, 5	11.6%
Dual range PCA	1700–1800 nm & 2300–2400 nm	1, 3	10.3%
PCA from two ranges	1700–1800 nm & 2400–2450 nm	2, 3	6.7%