Revista Brasileira de Farmacognosia
BRAZILIAN JOURNAL OF PHARMACOGNOSY
www.sbfgnosia.org.br/revista

Original article

Rapid identification of three varieties of *Chrysanthemum* with near infrared spectroscopy

Cun-wu Chen[a], *Hui Yan*[b], *Bang-xing Han*[a,*]

[a]*College of Biological and Pharmaceutical Engineering, Research Center of Research and Development of Traditional Chinese Medicine, West Anhui University, Lu'an, China*
[b]*School of Biological and Environmental Engineering, Jiangsu University of Science and Technology, Zhenjiang, China*

ARTICLE INFO

Article history:
Received 28 November 2013
Accepted 18 February 2014

Keywords:
Chrysanthemum morifolium
Near infrared spectroscopy
Rapid detection

ABSTRACT

A total of 139 batches of *Chrysanthemum* samples were randomly divided into calibration set (92 batches) and prediction set (47 batches). The near infrared diffuses reflectance spectra of *Chrysanthemum* varieties were preprocessed by a first order derivative (D1) and autoscaling, and a modelwas built using partial least squares analysis. In this study, three *Chrysanthemum* varieties were identified, the accuracy rates in calibration sets of Dabaiju, Huju, and Xiaobaiju are 97.60, 96.65, and 94.70%, respectively; And 95.16, 86.11, and 93.46% accuracy rate in prediction sets was obtained. The research results demonstrate that the qualitative analysis can be conducted by machine learning combined with Near-Infrared Spectroscopy, which provides a new method for rapid and non-invasive identification of *Chrysanthemum* varieties.

Introduction

The use of various *Chrysanthemum* species in medicine derives from the use of the plant's capitulum for expelling wind and heat, calming a hyperactive liver, and to improve vision acuity; also it is consumed as an edible plant. *Chrysanthemum* species as herbs have a long history of cultivation throughout China, culturing high quality ones and different varieties, such as Chuju, Boju, Hangju, and Gongju (The Pharmacopoeia Comittee of People's Republic of China, 2010). Hangju has three varieties: Dabaiju, Huju and Xiaobaiju. which are cultured in Macheng City, Tongxiang City and Yancheng City respectively. These varieties differ in quality, chemical composition, functions, and application (Jing et al., 2007). Therefore, *Chrysanthemum* varieties in demand precise identification the market for their reasonable and correct application as geoherbs.

Currently, the identification of the *Chrysanthemum* medicinal constituents depends on the observation of properties, chemical composition (Dong et al., 2007) and molecular biology (Yang et al., 2006). However, these methods have inevitable limitations, such as being difficult in promotion, difficult analysis, long time for processing, and expensiveness (Huang et al., 2009). Therefore, it is essential to develop an efficient, rapid, and comprehensive method to detect specific *Chrysanthemum* varieties at low cost.

* *Corresponding author.*
E-mail: hanbx1978@sina.com (B.X. Han).
0102-695X/$ - see front matter © 2014 Sociedade Brasileira de Farmacognosia. Published by Elsevier Editora Ltda. All rights reserved.
DOI: 10.1590/0102-695X20142413387

Near infrared spectroscopy (NIR) uses the electromagnetic spectrum ranging from the visible range and mid-infrared spectral region and its spectral range is 4,000-12,500 cm^{-1}, which is primarily the frequency multiplication and combination frequency absorption of hydrogen-containing radicals like C-H, N-H, and O-H. By scanning samples using NIR, the information about the samples can be obtained, including the chemical composition, physical and chemical properties, and even biological properties (Rodriguez-Saona et al., 2004). Along with the computational identification techniques, stoichiometry, and pattern recognition technology; NIR can rapidly, efficiently and properly analyze samples with easy processing without reagent use or contamination and multi-component detection, so it can be extensively applied in many fields, including Traditional Chinese Medicine (TCM), with good results (Hua et al., 2003; Woo et al., 2005; Chen et al., 2008; Han et al., 2009; Yan et al., 2011; Leng et al., 2013). In recent years, due to the development of computer technology and chemometric analysis softwares; especially in-depth research and wide application of stoichiometry, NIR has become one of the most eye catching spectroscopic technologies.

In this research three varieties of *Chrysanthemum* were studied, analyzed the NIR data of *Chrysanthemum* samples, classified accuracy rate as an evaluation parameter, and established a disaggregated model of discriminant partial least squares to obtain a classification algorithm.

Materials and methods

Apparatus

A WQF-400N FT-NIR analyzer (Beijing Rayleigh Analytical Instrument Corporation) was used to collect near-infrared spectrum, and a lead sulfide (PbS) probe was selected to diffuse reflect loading attachments.

Sample collection and preparation

Dabaiju samples were collected at Futianhe town from Macheng city of Hubei Province; Huju samples were collected at Shimen town, Tongxiang city of Zhejiang Province; Xiaobaiju samples were collected from Yanmar town, Yancheng city of Jiangsu province. From October 31 to December 31, 2011, 139 batches of samples were regularly harvested. They were identified as *Chrysanthemum morifolium* Ramat (Asteraceae), authenticated by Prof. Dequn Wang of Anhui College of TCM.

From Dabaiju variety 51 samples were used, 34 were randomly selected as calibration set and 17 for the prediction set. From Huju variety, 51 samples were used, 34 were randomly selected as calibration set and 17 for the prediction set. From Xiaobaiju variety, 37 samples were used, 24 were randomly selected as calibration set and 13 for the prediction set. All samples were air dried, powdered and sieved using a 40 mesh.

Collect NIR data

Data was recorded at room temperature of 20°C with relative humidity of 45%, the scanned area was 10,000-3,500 cm^{-1}, corrections were performed a total of 32 scans, with a resolution factor of 4 cm^{-1}, and light source was 10 W/6 V halogen tungsten lamp with air in the background. The samples were tested three times and the spectral data taken was the average value of three-time sampling.

Spectral data preprocessing

The methods taken for preprocessing are a standard normal variable transformation, a multiplicative scatter correction to get a first derivative (D1), and second derivative (D2). By comparing the four preprocessing methods on grain sizes, processing environment, and the machine's noise, the best preprocessing method would be obtained.

Partial least squares discriminant analysis modeling

Partial least squares discriminant analysis (PLS-DA) is a regression method based on characteristic variables. As a stable discriminant statistical analysis method, its use fits the situation with many variable numbers and multicollinearity, few samples of observation, and interference noise.

Spectra preprocessing and PLS-DA were carried out using a PLS toolbox 5.0 (American Eigenvector).

Results

Spectra preprocessing

The original spectra collected (Fig. 1), contains the information related to the sample composition and the noise signals produced by different factors. The noise signals would interfere with spectrum information, which is, sometimes, even serious, and thus influences the calibration model and the prediction of unknown samples compositions and properties. Hence, spectra preprocessing mainly aims at filtering spectral noise, data screening, optimization of spectral range and eliminating influence of other factors over obtained data, so as to lay the foundation of further establishment of calibration model and the precise prediction of unknown samples.

The original spectra were preprocessed using D1, D2, standard normal variate (SNV), and multiplicative scatter correction (MSC), and calibration set was established using PLS-DA model, while prediction set was used for testing the preciseness of the model. The results demonstrate that D1 + autoscale is the best parameter combination, achieving 100% predictive accuracy in calibration (leave-one-out cross-validation) and prediction set. The spectra preprocessed by D1 + autoscale are seen in Fig. 2. Comparing Figs. 1 and 2, the preprocessed spectra have many additional peaks in all bands, highlighting the spectral information.