Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 206 (2019) 23–30

Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

journal homepage: www.elsevier.com/locate/saa

# Rapid identification and quantification of *Panax notoginseng* with its adulterants by near infrared spectroscopy combined with chemometrics

Peng Liu [a], Jing Wang [a], Qian Li [a], Jun Gao [b], Xiaoyao Tan [a,*], Xihui Bian [a,*]

[a] State Key Laboratory of Separation Membranes and Membrane Processes, School of Environmental and Chemical Engineering, Tianjin Polytechnic University, Tianjin 300387, PR China
[b] College of Chemical and Environmental Engineering, Shandong University of Science and Technology, Qingdao, 266590, PR China

ARTICLE INFO

*Article history:*
Received 14 March 2018
Received in revised form 24 July 2018
Accepted 30 July 2018
Available online 01 August 2018

*Keywords:*
*Panax notoginseng*
Adulteration
Near infrared spectroscopy
Chemical pattern recognition
Multivariate calibration
Preprocessing

ABSTRACT

Traditional methods for identification of *Panax notoginseng* (PN) such as high performance liquid chromatography (HPLC) and gas chromatography (GC) are time-consuming, laborious and difficult to realize rapid and online analysis. In this research, the feasibility of identification and quantification of PN with rhizoma curcumae (RC), *Curcuma longa* (CL) and rhizoma alpiniae offcinarum (RAO) are investigated by using near infrared (NIR) spectroscopy combined with chemometrics. Five chemical pattern recognition methods including hierarchical cluster analysis (HCA), partial least squares-discriminant analysis (PLS-DA), artificial neural networks (ANN), support vector machine (SVM) and extreme learning machine (ELM) are used to build identification model of the dataset with 109 samples of PN and its three adulterants. Then seven datasets of binary, ternary and quaternary adulterations of PN are designed, respectively. Five multivariate calibration methods, i.e., principal component regression (PCR), support vector regression (SVR), partial least squares regression (PLSR), ANN and ELM are used to build quantitative model and compared for each dataset, separately. Finally, in order to further improve the prediction accuracy, SG smoothing, 1st derivative, 2nd derivative, continuous wavelet transform (CWT), standard normal variate (SNV), multiple scatter correction (MSC) and their combinations are investigated. Results show that PLS-DA and SVM can achieve 100% classification accuracy for identification of 109 PN with its three adulterants. PLSR is an optimal calibration method by comprehensive consideration of prediction accuracy, over-fitting and efficiency for the quantitative analysis of seven adulterated datasets. Furthermore, the predictive ability of the PLSR model for PN contents can be improved obvious by pretreating the spectra by the optimal preprocessing method, with correlation coefficients of which all higher than 0.99.



## 1. Introduction

*Panax notoginseng* (PN), the dried root and rhizome of *Panax notoginseng* (Burk.) F. H. Chen, is a precious traditional Chinese medicine (TCM) with important pharmacological values [1]. PN has been widely used to treat cardiovascular diseases, diabetes, fatty liver, swelling, pain and spot removing due to its blood stasis-removing, anticoagulation, anti-inflammatory, anti-diabetic, anti-fibrotic, as well as antioxidative effects [2]. It is the main material of many prescriptions and Chinese patent drugs, such as Compound Danshen dropping pills, Yunnnan Baiyao capsules, Notoginseng tablets, and so on.

Owing to the rising demands and shortage of resources, the price of PN keeps rising. Many counterfeit products such as rhizoma curcumae (RC), *Curcuma longa* (CL) and rhizoma alpiniae offcinarum (RAO) which have similar appearances and much lower prices are mixed with PN and sold in the market. Due to their completely different efficacy, these adulterants and frauds cannot cure or even have hazardous effects on human health. Thus, it is necessary to identify PN with its adulterants or frauds. However, because of the similar appearances, it is almost impossible for consumers to identify the purity of PN by naked eyes. Modern separating and analyzing methods such as high performance liquid chromatography (HPLC) [3,4] and gas chromatography (GC) [4,5] coupled with mass spectrometry have been used to analyze PN, but these methods are time-consuming, laborious and difficult to realize rapid and online analysis.

Near infrared (NIR) spectroscopy has been successfully proved as an efficient analytical method for rapid and nondestructive quality control and monitoring of agricultural products [6–8] and herbs [2,9–12]. NIR spectroscopy records the spectral bands 12,000 to 4000 $cm^{-1}$ which correspond to the overtone or combination vibrations of hydrogen bands such as C—H, O—H and N—H. However, single-wavelength analysis is not appropriate for NIR spectra because of the weakness, broadness and overlap of the absorption peaks. Consequently, varieties of chemometric pattern recognition [13–15] and multivariate calibration methods [16–19] have been applied to such complex samples for classification and simultaneous determination of multiple components. Finally, because the NIR spectral differences of adulterated samples are

* Corresponding authors.
*E-mail addresses:* tanxiaoyao@tjpu.edu.cn (X. Tan), bianxihui@163.com (X. Bian).

https://doi.org/10.1016/j.saa.2018.07.094
1386-1425/

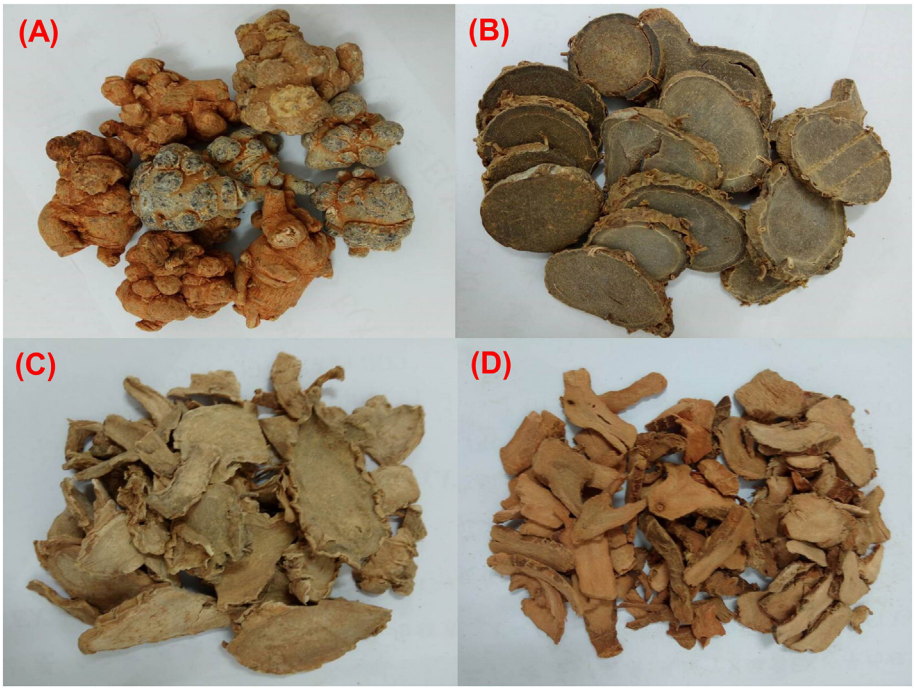


**Fig. 1.** The appearances of PN (A), RC (B), CL (C) and RAO (D).

usually not obvious, it is necessary to reduce the interference in the spectra such as noise, backgrounds, baselines and scattering effects. Preprocessing approaches [20–23] like smoothing, derivatives, continuous wavelet transform (CWT), standard normal variate (SNV) and multiplicative scatter correction (MSC) are often required.

The aim of present study is try to develop a rapid and accurate method for quality control of PN, including identification of PN with RC, CL and RAO and quantification of PN multiple adulterants. Five chemical pattern recognition methods are evaluated for the identification of PN with its three adulterants. Then the feasibility of quantification of PN multiple adulterated samples are investigated by using five multivariate calibration methods. Finally, based on the optimal multivariate calibration method, a series of spectral preprocessing methods, including smoothing, derivatives, SNV, MSC, CWT, and their combinations are also studied to enhance the prediction accuracy of adulterated PN samples.

## 2. Experimental

### 2.1. Preparing of PN and Its Three Adulterants

PN (25 samples), RC (28 samples), CL (28 samples) and RAO (28 samples) were collected from different authentic pharmacies in Tianjin. Fig. 1 shows the appearances of the four TCMs. It can be seen that, RC, CL and RAO have similar appearances and the colors as PN. PN usually used in the form of powder, which is more easily adulterated by the four adulterants because of their closely color. Since a certain amount of water exist in TCMs, all the samples were dried at 60 °C to constant weight in an oven. Then all the samples are ground into particles using a Dade portable grinder (DFT 50 g, Wenling Linda machinery Co. Ltd., Zhejiang, China) and filtered through a 1200-mesh sieve one by one. All the powdered samples were saved in 60 mm * 100 mm sealed plastic bags. Before NIR analysis, 109 TCMs were taken out 4 g

**Table 1**
the information of seven adulterated PN datasets.

| Datasets | Adulterant-ary | Number of samples | | | Components | Concentration range (%) |
|---|---|---|---|---|---|---|
| | | Total | Calibration set | Prediction set | | |
| 2 | Binary | 75 | 50 | 25 | PN | 0–100 |
| | | | | | RC | 0–100 |
| 3 | Binary | 75 | 50 | 25 | PN | 0–100 |
| | | | | | CL | 0–100 |
| 4 | Binary | 75 | 50 | 25 | PN | 0–100 |
| | | | | | RAO | 0–100 |
| 5 | Ternary | 66 | 44 | 22 | PN | 0.93–97.96 |
| | | | | | RC | 1.08–95.37 |
| | | | | | CL | 0.62–94.68 |
| 6 | Ternary | 66 | 44 | 22 | PN | 1.036–97.62 |
| | | | | | RC | 1.38–94.75 |
| | | | | | RAO | 0.61–94.90 |
| 7 | Ternary | 66 | 44 | 22 | PN | 1.14–97.43 |
| | | | | | CL | 0.99–94.82 |
| | | | | | RAO | 1.09–95.03 |
| 8 | Quaternary | 75 | 50 | 25 | PN | 1.08–94.62 |
| | | | | | RC | 1.03–94.30 |
| | | | | | CL | 0.38–94.32 |
| | | | | | RAO | 0.99–94.70 |