Geographical origin identification and quantitative calibration analysis of *Polygonum perfoliatum* L. by near infrared spectroscopy

Y. P. Zhang[1], Z. X. Jin[1*], Y. C. Yuan[2], X. S. Wang[3]

[1] Research Center on Life Sciences and Environmental Sciences, Harbin University of Commerce, Harbin 150076, China
[2] Chemical and Pharmaceutical Department, Henan University of Science and Technology, Luoyang 471023, China
[3] Forecasting and Analysis Center, Liming Chemical Research Institute, Luoyang 471001, China

Rapid near infrared (NIR) spectroscopic analysis models were developed to distinguish the geographical origin and determine the content of *Polygonum perfoliatum* L. The spectra of 94 samples of differing origin were recorded using two NIR instruments (Antaris II and Nicolet 6700). The spectral pretreatment methods used were the first-derivative (FD) and the standard normal variate correction (SNV). Distance match (DM) pattern was applied to classify the geographical origins of the samples, which had better predictive ability (>81%). For the quantitative calibration, the polyphenol and flavone models were established with a partial least squares (PLS) algorithm. The established models were evaluated with root mean square error of prediction (RMSEP) and correlation coefficients. The correlation coefficients of both calibration models were above 0.910, indicating satisfactory predictive abilities. This study demonstrated that NIR diffuse reflection spectroscopy can be used for rapid discrimination of geographical origin and content determination of *P. perfoliatum* L., and   is a beneficial quality control method for this raw material.

Key words: Near infrared spectroscopy, Polygonum perfoliatum L., distance match, partial least squares, polyphenols, flavones.

INTRODUCTION

Polygonum perfoliatum L. (PPL) is a traditional herb used to treat upper respiratory tract infections, bronchitis, whooping cough, acute tonsillitis, enteritis, dysentery, nephritis edema, as well as for the topical treatment of herpes zoster, eczema, acute folliculitis, snake bite [1]. Owing to its curative effects, it was recorded in the 2010 edition of the Chinese pharmacopoeia (ChP) as a new variety [2]. The plant is widely distributed in China. Accordingly, the quality of the herbal raw materials fluctuates, owing to varying ecological environment. Thus, a control method to distinguish and quantify samples is necessary. Currently, traditional Chinese medicines (TCMs) are identified by microscopic and physicochemical evaluations, which can depend on personal experience to make a judgment. Therefore, most quality control methods for TCMs involve the use of HPLC for the identification of one or two main compounds. However, it is difficult to comprehensively evaluate TCMs using this method [3, 5].

The main compounds in PPL are polyphenols and flavones [6-9]. According to the ChP, PPL quality is evaluated by microscopic identification, determination of caffeic acid by TLC, and detection of sophorin by HPLC [2]. However, these methods require tedious sample preparation, take much time, and require large solvent amounts, making them unsuitable for fast analysis during industrial production. Furthermore, sophorin and caffeic acid are poor indicators of product quality. Therefore, it is necessary to develop a low-cost, rapid, multi-component, and controlled method to address this issue.

Fourier transform near-infrared spectroscopy (FT-NIR) is a high-speed, non-destructive technique that has been widely applied for low-cost, reliable qualitative and quantitative analyses in the food, chemical, agrochemical, and petrochemical industries [10-14]. This method has been used for raw material testing [15, 16], product quality control, and process monitoring [17-19], and is being extended to TCM research [20-25].

In this study, two NIR models were established. One is a qualitative model, which can distinguish PPL obtained from different geographical origins, whereas the other is a quantitative calibration model for the rapid determination of polyphenols and flavones in PPL.

* To whom all correspondence should be sent:
E-mail: Jin_ai_pu@163.com

MATERIALS AND METHODS

Sample preparation

PPL samples (n = 94) were collected from 12 provinces and 22 locations in China. All samples were identified by Dr. ZhexiongJin (Commercial University of Harbin), and are listed in Table 1.

Table 1. Summary of the tested samples

Ser. no.	Origin	Sam. no.	Ser. no.	Origin	Sam. no.
1	Meizhou Guangdong	6	12	Guizhou	2
2	Jieyang Guangdong	6	13	Sanming, Fujian	2
3	Qingyuan Guangdong	6	14	Nanyang, Henan	3
4	Fuzhou Fujian	6	15	Changsha Hunan	2
5	Wuyishan Fujian	6	16	Nanjing Jiangsu	2
6	Hechi Guangxi	6	17	Fuzhou Jiangxi	1
7	Yizhou Guangxi	8	18	Yingtan Jiangxi	2
8	Yongjia Zhejiang	6	19	Yunnan	2
9	Wenzhou Zhejiang	6	20	Taizhou, Zhejiang	2
10	Luoyang Henan	10	21	Chongqing	2

The samples were milled into powder with a grinder. The final powder samples were obtained by passing the ground powder through a 50-mesh sieve. To ensure that moisture was not an interfering factor, all samples were dried in an oven at 40°C for 12 h. In order to reduce operating error, each sample was scanned three times.

NIR spectra collection

The NIR diffuse reflection spectra were recorded from 10000 to 4000 cm^{-1} at 4 cm^{-1} intervals using the Integrating Sphere module of the Antaris II FT-NIR analyzer (Thermo Scientific, Madison, WI, USA) with an InGaAs detector. Each sample was scanned 64 times, and the resolution was 16 cm^{-1}. The environmental temperature was held at 20-22°C and the relative humidity was 25-30%. Result workflow-based software (Thermo Scientific) was used to obtain the NIR spectra, and the spectra were recorded as the logarithm of the reciprocal reflectance, log (1/R). In order to reduce operating error, each sample was scanned three times.

To explore the application of the established quantitative calibration models on NIR instruments from different facilities, a second NIR analyzer (Nicolet 6700, Thermo Scientific) was used for NIR spectra collection, under the working conditions mentioned above. The NIR spectra collected by the two types of instruments are shown in Fig. 1.
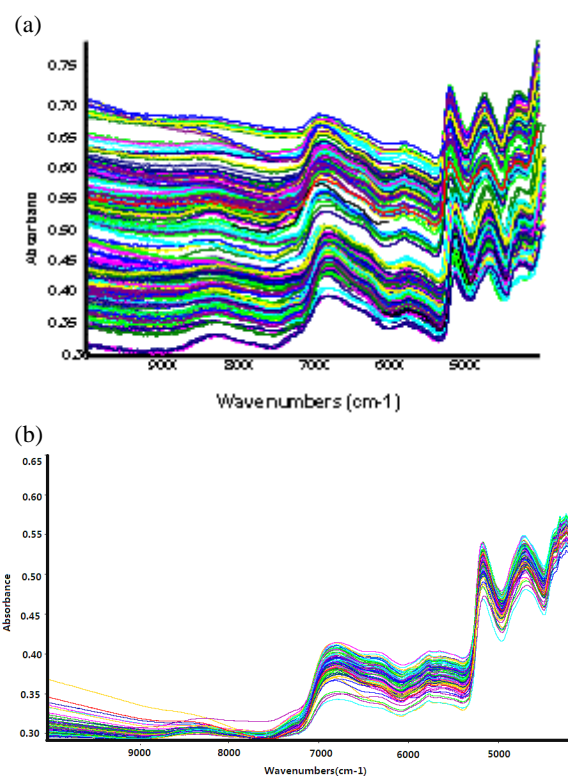

Fig. 1. Spectra of samples collected from Antaris II (A) and Nicolet 6700 (B) FT-NIR.

Reference analysis method

A UV spectrophotometer was used to detect total polyphenols and flavones. Samples (2 g) were reflux-extracted in a water bath with 50 mL of mixed solvent (70:30 [v/v] methanol and water) for 0.5 h. After filtration, samples were placed in 250 mL brown volumetric flasks and diluted with water to the mark. Folin-Ciocalteu colorimetry was used to determine total polyphenols at 760 nm by using gallic acid as a reference substance. AlCl$_3$ colorimetry was used to determine total flavones at 413 nm by using rutin as a reference substance.

Chemometric methods

For determining the geographical origin, distance match (DM) patterns were applied. For the quantitation of components, the partial least squares (PLS) method was adopted. Spectral pretreatment and chemometric analyses were implemented using the TQ Analyst (V8.0, Thermo Scientific) and Matlab (V7.0, Math Works, Natick, MA, USA).

RESULTS AND DISCUSSION

Content analysis

Polyphenols and flavones are the main active constituents of PPL, and are widely occurring

Table 2. Methodology parameters and calibration curves of the reference UV method.

Compound	Linearity range (mg/ml)	Calibration curve	R	Repeatability (RSD %, n=6)	Recovery (%,n=6)
Polyphenol	0.001-0.006	C=0.009A-0.001	0.999	1.532	97.65
Flavone	0.007-0.025	C=0.0302A-0.001	0.998	1.083	98.05

secondary metabolites. Using the UV method described in Section 2.3, all 94 samples were analyzed three times, using the main parameters listed in Table 2. The average content of samples from different provinces is listed in Table 3. We found that samples from different provinces have obvious differences. For example, samples from Jiangxi and Henan provinces had higher polyphenol and flavone content.

Table 3. Average contents of samples from different provinces.

Ser. no.	Province	Content of polyphenol (mg/g ± s)	Content of flavone (mg/g ± s)
1	Guangxi	34.91±11.25	15.68±7.01
2	Guangdong	37.95±14.76	16.36±6.22
3	Fujian	35.48±18.98	16.28±5.11
4	Chongqing	31.42±7.98	13.89±5.14
5	Guizhou	21.01±5.68	10.98±2.12
6	Yunnan	23.9±6.21	13.94±1.38
7	Hunan	30.19±11.75	17.69±4.42
8	Jiangxi	40.14±14.12	23.45±7.06
9	Zhejiang	32.82±12.44	17.68±2.65
10	Henan	44.12±11.94	23.56±5.23
11	Hubei	26.51±6.94	17.30±5.48
12	Jiangsu	27.69±0.99	15.92±2.95

NIR spectral characteristics

FT-NIR can reveal the transitions of corresponding chemical constituents, including O-H, N-H, and C-H. Fig. 1 shows the NIR spectra collected with two NIR instruments. Importantly, the Nicolet 6700 and Antaris II NIR analyzers have different interferometers. The former is equipped with a no wear electromagnetic interferometer, whereas the latter - with a Michelson interferometer. Compared with the Nicolet 6700 NIR analyzer, the spectral absorption intensity obtained using the Antaris II NIR analyzer was higher, and the between-sample intensity variation ranges were wider. However, the shapes of the fundamental spectra were analogous in appearance. For the further research we used the Antaris II NIR analyzer.

Origin identification model

PPL samples from twelve geographical origins with sample numbers above six were used to construct models. In this study, DM was carried out by the traditional authenticity recognition method on the basis of NIR spectral data.

Spectral pre-processing

Spectral pre-processing is used to reduce the effects of systematic noise, base line variation, light scattering, and path length differences. There are several different types of pre-processing available, including first-derivative (FD), second-derivative (SD), multiplicative scatter correction (MSC), standard normal variate correction (SNV), Savitzky-Golay (SG), and Norris derivative (ND). Through multiple optimization calculation, we selected FD, SNV, and 7500-4119 cm^{-1} to set up the models.

Results of the DM method

When establishing the origin identification analysis model, the threshold value is particularly important. The threshold values of the samples are shown in Table 4.

Table 4. Threshold values and identification accuracy rate

Ser. no.	Origin	Threshold	Accuracy rate (%)
1	Meizhou, Guangdong	0.3	90.9
2	Jieyang Guangdong	0.69	100
3	Qingyuan Guangdong	0.23	90.9
4	Fuzhou, Fujian	0.46	81.8
5	Wuyishan, Fujian	0.92	100
6	Hechi, Guangxi	0.1	100
7	Yizhou, Guangxi	0.46	90.9
8	Yongjia, Zhejiang	0.23	100
9	Wenzhou, Zhejiang	0.23	100
10	Luoyang, Henan	0.92	100
11	Suizhou, Hubei	0.92	0.909

These twelve models were added to the workflow in the form of measurements, selecting composite measurements in the classification. We used the rest of the samples for validation, obtaining a distance value of the sample spectrum after the software analysis. The identification accuracy rates are listed in Table 4. Six models

reached 100%, Fuzhou was 81.8%, and the other models were above 90%. From these values we found that DM can be used as an origin identification method, and that NIR can identify PPL samples from different geographical origins.

PLS QUANTITATIVE CALIBRATION MODEL

Division of calibration and validation sets

The 94 samples were randomly divided into calibration and validation sets, which were designed to examine the extrapolation performance. The polyphenol and flavone content in the calibration and validation sets must cover a wide range, which is helpful to develop robust models. The statistical values for the calibration and validation sets are listed in Table 5.

Table 5. Statistical values of the active ingredient contents in calibration and validation sets.

Active ingredients	Average value $(X\pm s,\%)$	Calibration sets (%)	Validation sets (%)
Polyphenol	32.11±14.69	5.6-84.92	15.3-54.67
Flavone	18.24±6.76	4.7-32.15	7.12-23.94

Selection of the most suitable conditions for building the calibration model

There are many methods of qualitative analysis, including simple Beer's law, classical least squares (CLS), stepwise multiple linear regression (SMLR), partial least squares (PLS), and principal component regression (PCR). For the natural products used in TCM, which have a complex composition, PLS is typically used.
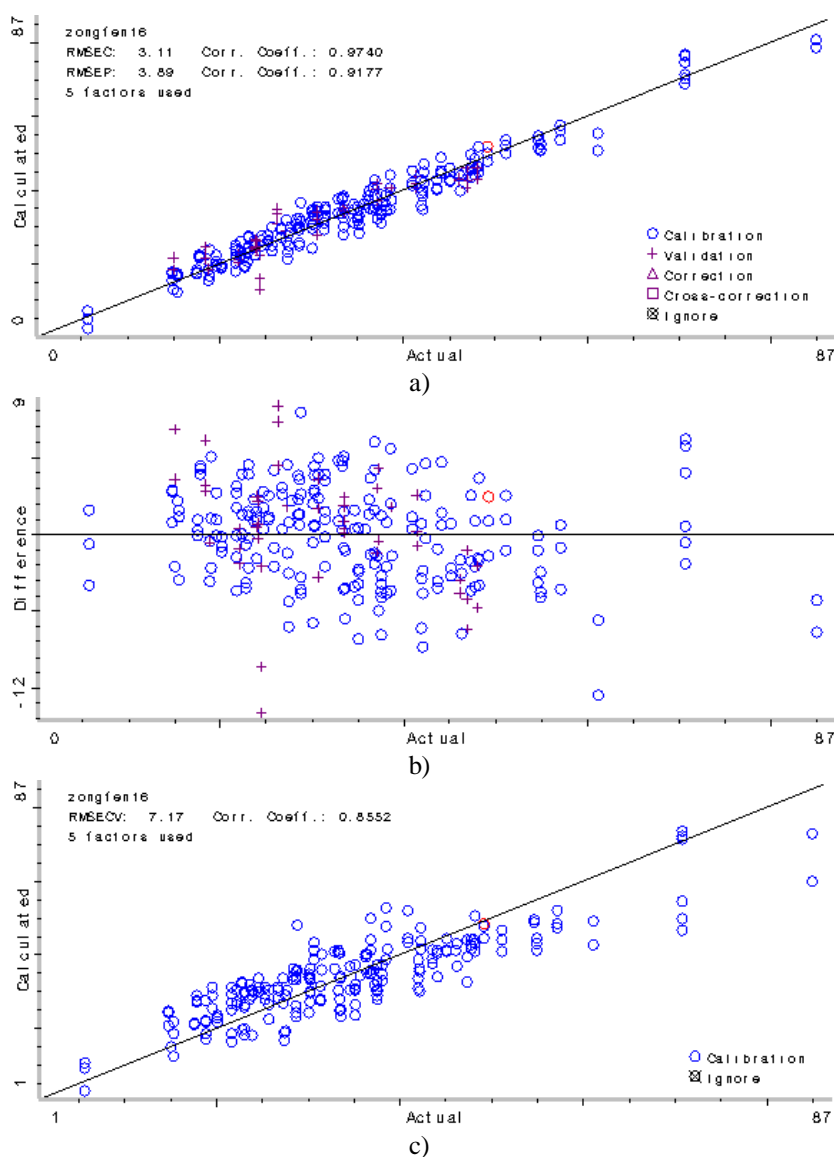

Fig. 2. Calculated model, difference and cross-validated chart of polyphenol.

Table 6. Performance parameters of the established models.

Component	Pretreatment method	PC no.s	Calibration		Cross-validation		External validation	
			RMSEC	R	RMSECV	R	RMSEP	R
Flavone	SNV+1-d	4	2.19	0.9463	3.96	0.8127	2.16	0.9139
Polyphenol	SNV+1-d	5	3.11	0.9740	7.17	0.8552	3.89	0.9177

A good NIR calibration model should have a low root mean square error of calibration (RMSEC) and root mean square error of prediction (RMSEP) value, high correlation coefficient (R), and small differences between root mean square error of cross-validation (RMSECV) and RMSEP. In the spectral pre-processing, FD, SD, MSC, SNV, SG, and ND were compared to optimize the models. These parameters are listed in Table 6. Fig. 2 is a chart of the calculated model (A), difference (B), and cross-validation (C) of polyphenols. The RMSEC and RMSEP correlation coefficients are above 0.91, indicating that these two models possess acceptable fitting results and predictive ability, and can be used for the rapid analysis of polyphenol and flavone content in PPL.

Accuracy and precision of the established NIR method

In order to validate the accuracy of the established NIR method, the sample content in the prediction set was determined with the NIR and UV method, and evaluated with the paired two sample *t*-test. In the 95% confidence interval, the polyphenols and flavones had statistical *t* values of -0.3268 and -0.7167, respectively, with absolute values less than the two-tail critical value, t = 2.179, 2.262. The two-tail p-values were greater than 0.05, indicating that there was no significant difference between the NIR and UV method.

CONCLUSIONS

In the present study, a set of qualitative and quantitative NIR spectroscopic models were developed for PPL quality control. Based on these models, we showed that no sample pretreatment was necessary. Furthermore, data on the content of polyphenols, flavonoids, and the sample origin could be obtained in 1 min. Compared with traditional methods, this method affords rapid and low-cost detection; it can distinguish geographical origin and detect multiple components in one analysis. The overall results show the feasibility of NIR spectroscopy for quality control of TCM pharmaceuticals.

REFERENCES

1. J. S. Guo. Recognition and application of Chinese herbal medicine. People's Medical Publishing House, Beijing, 2009.
2. Ministry of Public Health of the People's Republic of China, Pharmacopoeia of the People's Republic of China, China Pharmaceutical Technology Press, Beijing, 2010.
3. X. Y. Song, Y. D. Li, Y. P. Shi, L. Jin, J. Chen, *Chin. J. Nat. Med.*, **11**, 596 (2013).
4. S. P. Li, J. Zhao, B. Yang, *J. Pharm. Biomed, Anal.*, **55**, 802 (2011).
5. Y. Jiang, David B, P. Tu, Y. Barbin, *Anal. Chim. Acta.*, **657**, 9 (2010).
6. J. Lei, N. Yao, K. Wang, *Biochem. Syst. Ecol.*, **48**, 186 (2013).
7. K. W. Wang, J.R. Zhu, L. Q. Shen, *Nat. Prod. Res.*, **27**, 568 (2013).
8. G. H. Zhu, D. Y. Wang, G. J. Men, *Indian. J. Heterocycl. Ch.*, **10**, 41 (2000).
9. C. Zhao, H. G. Chen, X. J. Gong, *Chin. Herb. Med.*, **41**, 365 (2010).
10. X. M. Chen, X. Z. Yu, Y. G. Wang, Y. D. Yang, J. Y. Zhang, *J. Oleo Sci.*, **64**, 255 (2015).
11. G. Valentina, B. Roberto, G. Riccardo, *J. Food Eng.*, **142**, 80 (2014).
12. M. Romera-Fernandez, L. A. Berrueta, S. Garmon-Lobato, B. Gallo, F. Vicente, J. M.Moreda, *Talanta*, **88**, 303 (2012).
13. C. W. Huck, *J. Pharmaceut. Biomed. Anal.*, **87**, (2014).
14. Y. Yoshihisa, S. Toyofumi, M. Mika, O. Michiteru, H. Shuichi, F. Toshiro, T. Kazuo, *Chem. Pharm. Bull.*, **60**, 624 (2012).
15. B. Jamshidia, S.Minaeia, *Comput. Electron. Agr.*, **85**, 64 (2012).
16. J. R. Lucio-Gutiérrez, J. Coello, S. Maspoch, *Food Res. Int.*, **44**, 557 (2011).
17. M. Snežana, P, Ksenija, K. Janez, H. Matej, *Eur. J. Pharm. Biopharm.*, **88**, 847 (2014).
18. H. Xiong, X. Gong, H. Qu, *J. Pharm. Biomed. Anal.*, **70**, 178 (2012).
19. H. Huang, H. Qu, *Anal. Chim. Acta.*, **707**, 47 (2011).
20. C. Zhang, J. Su, *Acta. Pharm. Sinica B*, **4**, 182 (2014).
21. W. Li, Y. Wang, H.Qua, *Vib. Spectrosc.*, **62**, 159 (2012).
22. Q. Shao, A. Zhang, W.Ye, H. Guo, R. Hu, *Spectrochim. Acta. A*, **120**, 499 (2014).
23. Y. Lai, Y. Ni, S. Kokot, *Vib. Spectrosc.*, **56**, 154 (2011).
24. W. Li, L. Xing, Y. Cai, H., *Vib. Spectrosc.,* **55**, 58 (2011).
25. A. Rohman, A. Nugroho, E. Lukitaningsih, ?. Sudjadi, *Appl. Spectrosc. Rev.*, **49**, 603 (2014).

ИДЕНТИФИКАЦИЯ НА ГЕОГРАФСКИ ПРОИЗХОД И КОЛИЧЕСТВЕН АНАЛИЗ НА *Polygonum perfoliatum* L. ЧРЕЗ БЛИЗКА ИНФРАЧЕРВЕНА СПЕКТРОСКОПИЯ

И. П. Жанг[1], З. Кс. Джин[1*], И. Ц. Юан[2], Кс. С. Уанг[3]

[1] *Изследоеателски център за науки за живота и околната среда, Търговски университет в Харбин, Китай*

[2] *Департамент по фармация, Научно-технологичен университет Хенан, Луоянг, Китай*

[3] *Център за прогнози и анализи, Химически изследователски институт Лиминг, Луоянг, Китай*

(Резюме)

Разработен е модел за анализ, основан на бързата близка инфрачервена спектроскопия (NIR) за разпознаване на географския произход и определяне на съдържанието на *Polygonum perfoliatumL.* Определени са спектрите на 94 проби от различен произход с помощта на два NIR-уреда (Antaris II and Nicolet 6700). Приложени са специални методи за предварителна подготовка – на първата производна (FD) и стандартната корекция на нормалната дисперсия (SNV). Приложен е методът на дистанционното съпоставяне (DM) за класифицирането на географския произход на пробите с с по-добра предсказувмост (>81%). Количествената калибровка е направена с помощта на полифеноли и флавони по алгоритъма на най-малките квадрати (PLS). Корелационните коефициенти на двата калибрационни моделаса над 0.910, указвайки за добра предсказуемост. Тази статия демонстрира факта, че NIR-дифузионната отражателна спектроскопия може да се използва за бързо разграничаване на географския произход и количеественото определяне на *P. perfoliatum* L. и е удобен метод за качествен контрол на тази суровина.