HOSTED BY

ELSEVIER

Review Paper

Species authentication and geographical origin discrimination of herbal medicines by near infrared spectroscopy: A review ☆

CrossMark

Pei Wang [a,b], Zhiguo Yu [a,*]

[a] School of Pharmacy, Shenyang Pharmaceutical University, Shenyang 110016, China
[b] Center for Excellence in Post-Harvest Technologies, North Carolina Agricultural and Technical State University, North Carolina Research Campus, 500 Laureate Way, Kannapolis, NC 28081, USA

ARTICLE INFO

ABSTRACT

Near infrared (NIR) spectroscopy as a rapid and nondestructive analytical technique, integrated with chemometrics, is a powerful process analytical tool for the pharmaceutical industry and is becoming an attractive complementary technique for herbal medicine analysis. This review mainly focuses on the recent applications of NIR spectroscopy in species authentication of herbal medicines and their geographical origin discrimination.

Contents

1. Introduction

As one of the most traditional forms of health care, herbal medicine has been worldwidely used for over hundreds of years. The World Health Organization (WHO) estimates that about 65%–80% of the world's population, particularly in the developing countries, has limited access to modern medical care, and herbal medicine is still their primary source of health care [1]. Certain botanicals have been widely used in some societies, such as turmeric (*Curcuma longa* L.) and *Curcuma xanthorizza* Lam [2]. Active components like morphine, digitoxin, cocaine and taxol contained in herbal medicines are used in standard allopathic medicine, and related quality standards regarding the purity, safety and efficacy are carried out by the United States Food and Drug Administration (USFDA) [2,3]. In fact, it is estimated that over a quarter of modern medicines are directly or indirectly derived from higher plants [4].

Herbal medicine can be represented either as a single-herb or a multi-herb formula, and it is reported that about 92% of herbal medicine formulas are a combination of less than thirteen herbal medicines [5]. Traditionally, the identification of herbal medicine is carried out according to the differences in morphology, and/or thin layer chromatography (TLC) identification or content determination of one or two marker constituents [6,7]. The characteristics of systematism, multi-target and synergistic actions of traditional Chinese medicines (TCMs) originate from their multiple constituents, which can vary significantly in contents. The

chemical compositions in herbs may vary depending on the species, location of growth, age, harvesting season, drying processes and some other factors [8]. Consequently, to ensure the reliability and repeatability of pharmacological and clinical researches and to guarantee the consistency of the final product quality, the determination of all bioactive constituents of a herbal material is necessary [9,10]. However, elucidating all of the herb bioactive compounds is time-consuming, arduous and unsuitable for clarifying the synergies between herbal medicines. Thus, it is of utmost importance to formulate quality control protocols based on entire metabolome, which can be regarded as a 'pattern-oriented' method, especially for species authentication and geographical origin discrimination [6].

For herbal medicine species authentication, the WHO, the USFDA and the European Medicines Agency (EMEA) have updated their regulations and state that the identification of herbal medicines is one of the first assays that should be conducted to ensure their quality and discriminate from related species or adulterated samples [11–14]. However, species authentication is still not sufficient for quality control of herbal medicines. It is reported that, for herbal medicines even from the same species, the quality and efficacy are somewhat different according to their growing conditions such as cultivation soil and climates based on the geographical origins [15–17]. Therefore, rapid and accurate analytical approaches are essentially required the estimation of correct value and the prevention of illegal distribution.

Due to its capability of fingerprinting analysis, the modern vibrational spectroscopies [mid-infrared (mid-IR) and near-infrared (NIR) and Raman] fulfill the common requirements such as speed of analysis and ease of use, especially in combination with chemometric techniques, and are highly efficient in distinguishing types or species as well as geographical origins of herbal medicines [3,18]. Among these, NIR spectroscopy is widely applied owing to its high analytical speed, low cost and reliability for qualitative and quantitative analysis of various types of samples such as soil [19], food [20] and beverages [21]. Thus, NIR spectroscopy serves as an excellent candidate for herbal medicine analysis [15].

During the last few years several review articles dealing with NIR spectroscopy and its applications to analysis of natural products have been published [3,22,23]. With the aim of providing an up-to-date overview of the applications of NIR spectroscopy on medicinal plant analysis, the present review summarizes the recent applications of NIR spectroscopy to herbal medicine species identification and geographical origin discrimination during the past 15 years.

2. Near infrared technique

The American Society of Testing and Materials (ASTM) defines the NIR region from 780 to 2526 nm ($12821-3959\ cm^{-1}$), located between the red band of the visible light and the mid-IR region [24]. The most prominent absorption bands are a consequence of the absorbance of light due to molecular vibrations (overtones and combinations of the fundamental mid-IR bands) of hydrogen bonds like –C–H, –S–H, –N–H, and –O–H functional groups [25]. The NIR region was discovered by Herschel more than 200 years ago, and it has become a popular technique since 1960. The current triumph of NIR spectroscopy is attributed to Norrisr et al. who recognized the immense capability of NIR spectroscopy as a potential process analytical technology tool in industrial practice for measurements of certain types of food, agricultural components and product quality control [26].

NIR spectroscopy has gained wide acceptance in various fields since it has several advantages over other analytical techniques with respect to fast acquisition, low cost, and nondestructive character towards the analyzed sample, while the most noticeable feature of NIR spectroscopy is its ability to acquire spectra for solid, semi-solid, and liquid samples without or with only minimal sample preparation [27–30]. On the one hand, the interest in NIR has increased owing to the improvements of instrument and the advancement of intrinsically safe measurement probes and fiber optics which make the delocalization of the measurements a reality. On the other hand, the fast growing applications of NIR spectroscopy also have been stimulated by the advance in computer technology and the progress in new mathematical methods which make large-scale data processing possible [28]. However, like every scientific technique, NIR spectroscopy has its own disadvantages. For instance, in comparison to mid-IR spectra whose absorbance bands can be directly interpreted due to the specific absorption of organic functional groups, NIR spectra are more complex owing to the nature of NIR bands (overlapping overtones and combination bands for hydrogen bonds). Besides, the physical state of the sample and the testing environment also influence the spectra, which make the data interpretation more complicated [31]. In summary, it is particularly hard to discern 'relevant' information about the characteristics of target analytes from the raw spectra.

Therefore, for qualitative or quantitative NIR analysis, mathematical and statistical methods are required to extract 'relevant' information (i.e. spectral variables related to properties of the analyte) and reduce 'irrelevant' information (i.e. interfering parameters), which belongs to the research field of chemometrics [32]. Chromemtrics regroups several related topics including design and optimization of experimental procedures, information extraction strategies (modeling, classification and hypothesis validation) and techniques for obtaining knowledge about chemical systems [33]. Owing to the development of chromemtrics, NIR spectroscopy has found applications in a broad range of domains during the past decades, such as in the petrochemical [34,35], environmental [36,37], pharmaceutical [31,32,38], clinical [39,40], agricultural [41,42], food [40,43], biomedical [44], and herbal medicinal [22,23,45] sectors.

3. Selected applications of herbal medicine species authentication

Although some herbal medicines are of different species, the morphological characteristics are similar to each other, especially, among closely related species. Therefore, the rapid and sensitive recognition of herbal species plays a decisive role in herbal medicine quality control. The traditional test mainly depends on naked-eye inspection or TLC. These test methods are either subjective in nature or require operative skills and experience which are not efficient enough for screening huge volumes of herbal medicines [46]. Over the past decade, a large number of publications have been available in the literature, which are dedicated to the classification of herbal medicines based on their species using NIR spectroscopy.

Paris, which belongs to the Liliaceae family, contains about 24 species and is mainly distributed in Europe and Eastern Asia. However, only the rhizomes of *Paris polyphylla* var. *chinensis* and *P. polyphylla* var. *yunnanensis* are officially listed in Chinese Pharmacopoeia. It is hard to discriminate dry rhizomes of the same genus by traditional morphological identification methods, especially for the original powder form. Zhao et al. [47] used NIR spectroscopy in combination with partial least squares discriminated analysis (PLS–DA) to give a preliminary overview of the similarities and differences among the species, and the results indicate that wild *Paris* species exert a significant effect on the NIR

spectrum. These results show that *P. cronquisistii* var. *xichouensis*, *P. caobangensis*, *P. cronquistii*, *P. polyphylla* var. *alba,* and *P. polyphylla* var. *pseudothib* are clearly separated from the others.

Cortex Phellodendri (CP), Chinese name 'Huangbai', is a commonly used Chinese herb. There are two species of CP: one is *Cortex Phellodendri Chinensis* (PCS) and the other is *Cortex Phellodendri Amurensis* (PAR). With the aim to differentiate the two species of CP, NIR spectroscopy coupled with principal component analysis (PCA) was performed by Chan et al. [48]. After second derivative pretreatment, the spectral variations between PCS and PAR were explored through the NIR ranging from 4082 to 4545 cm^{-1}, and this spectral region was adopted in classification via PCA. Finally, all the samples were successfully separated into two different categories corresponding to PCS and PAR, respectively.

Kudo et al. [49] investigated the application of NIR spectroscopy for rapid identification of *Digitalis purpurea* from other four close species (*Digitalis lanata, D. mertonensis, D. ambigua, and D. orientalis*). Five methods including the maximum distance in wavelength space, correlation in wavelength space, correlation coefficients, two-wavelength plot and identification using nearest-neighbor were carried out and compared. It was found that the maximum distance in wavelength space was the most efficient method for the identification of *D. purpurea*, followed by the use of two-wavelength plot, which is also useful in pattern recognition and gives a good visual idea of the differences between species. In contrast, the use of correlation values did not seem to be very useful for the discrimination of the samples [49].

Radix *pocrariae*, known as 'Gegen' (GG), is an important edible herb used in oriental medicine. It has been widely used for the treatment of diarrhea, acute dysentery, deafness and cardiovascular diseases [50]. Two different species of GG, roots of Radix *pocrariae lobata* (Wild.) ohwi (Yege, YG) and Radix *thomsonii* benth (Fenge, FG) were officially recorded in Chinese Pharmacopoeia since the 2000 edition. The photochemistry comparison demonstrated that the amounts of major bioactive isoflavones of YG and FG are greatly different. It has been split into two entries since 2005 edition of Chinese Pharmacopoeia. Rapid NIR spectroscopy in conjunction with linear discriminant analysis (DA) and soft independent modeling class analogy (SIMCA) have been applied to the species authentication of YG and FG [51]. Clustering models using full spectrum and two selected regions (5556–6250 cm^{-1} and 4082–4878 cm^{-1}) were established and compared. It was found that models based on the intensities from the selected spectral regions were superior to those from full spectrum using either linear DA or SIMCA method.

Wang et al. [52] studied the use of two-dimensional NIR correlation spectroscopy for the discrimination of *Dendrobium densiflorum* Lindl. ex Wall. (Mihuashihu), *D. aurantiacum* Rchb.f. var. *denneanum* (Kerr.) Z. H. Tsi (Dieqiaoshihu) and *Dendrobium chrysotoxum* Lindl. (Guchuishihu), which evidently belong to three different species. PCA was carried out first to ascertain the possibility of discrimination using NIR reflectance spectroscopy. Then, temperature-induced generalized two-dimensional NIR correlation spectroscopy (2D NIR) was generalized. Compared with the one-dimensional NIR spectroscopy, the 2D NIR correlation spectroscopy is more powerful, with the ability to enhance spectral resolution, simplify the overlapped bands, and provide information about temperature-induced spectral intensity variations. For different species of *Dendrobium*, remarkable differences located in the range from 4750 to 5600 cm^{-1} were observed in the synchronous and asynchronous 2D correlation spectra, and this region was directly used to discriminate the three species of *Dendrobium*.

Authentication of *Ephedra* plants of different species using NIR spectroscopy was investigated by Fan et al. [53]. NIR diffuse reflectance spectra were collected from 37 pulverized samples of
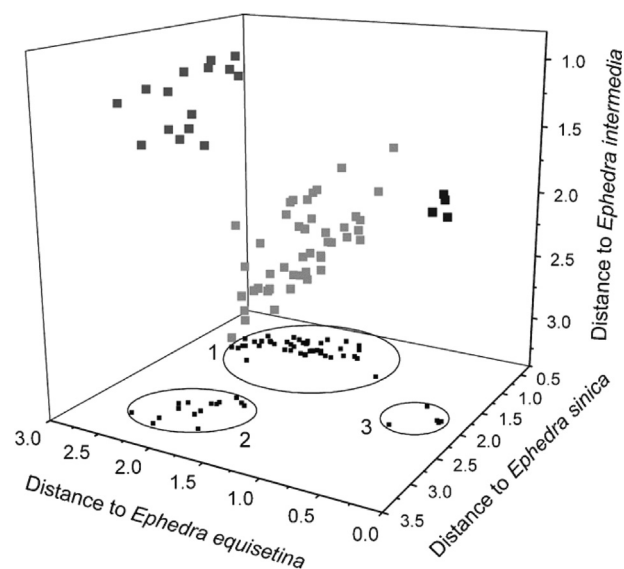

Fig. 1. Projection maps of the DA of the calibration samples of Ephedra plants of three species. *Ephedra sinica, Ephedra intermedia* and *Ephedra equisetina* are respectively labeled by (1), (2) and (3) in the two-dimensional map, and represented with light gray, gray and dark gray spots in the three-dimensional map. Reprinted from [53] with permission from Elsevier.

Ephedra plants. Three different multivariate analysis techniques, namely DA, self-organizing map and back-propagation artificial neural network (BP-ANN), were carried out for the spectral data analysis after spectra processing and data pre-processing. The performance indexes of the DA model were 84%–92%, and the prediction accuracies of both the self-organizing map and the BP-ANN models were also acceptable. Projection maps of the DA of the calibration samples of *Ephedra* plants of three species are outlined in Fig. 1.

One of the most famous herbal medicines analyzed using NIR spectroscopy is ginseng [46,54–58]. Ginseng is a widely used medicinal product that mainly grows in East Asia and North America. Asian ginseng (Radix et Rhizoma Ginseng, the root and rhizome of *Panax ginseng* (PG) C.A. Meyer, Araliaceae), cultivated mainly in China and Korea, has been widely used as a TCM for thousands of years. *Panax quinquefolium* (PQ). L (Araliaceae), known as American ginseng, has been widely used for the stress and blood sugar reduction and immunity adjustment [59]. In America, 'ginseng' can be used to refer to either Asian ginseng, or American ginseng, and even Siberian ginseng [*Eleutherococcus senticosus* (Rupr. and Maxim.) Maxim; botanical syn. *Acanthopanax senticosus* (Rupr. & Maxim.) Harms] which is not the same herb as American ginseng or Panax ginseng [60]. Obviously, rapid and accurate differentiation of ginseng is essential for the correct use of ginseng. The application of visible and short-wave NIR spectroscopy to differentiate the species of Panax was investigated by Chen et al. [56]. PCA was carried out prior to least-square support vector machine (LS-SVM) modeling; PCA could effectively reduce the vast majority of the spectral data. All the tested samples can be discriminated with 100% correct classification rate by the proposed PCA-LS-SVM method [56]. Different types of molecular spectroscopy (including NIR diffuse reflection, Raman and mid-IR spectroscopy) with OPUS/Ident software (Thermo Scientific, Waltham, MA) were utilized for cluster analysis of ginseng according to species and processing methods, and it was found that compared with IR spectra, Raman and NIR spectra are less affected by other factors, and obtained more accurate results when combined with chemometric analysis [54]. Woo et al. [61] reported that the availability of NIR fingerprinting using SIMCA would be adequate for the classification of Asian and American ginseng, and

compared with DA and PLS–DA, SIMCA has a better capability to detect debased samples [55]. SIMCA combined with NIR spectroscopy was also used for the differentiation of ginseng from *Austragali* Radix and *Smilacis* Rhizoma [46]. Apart from the species authentication of ginseng, different parts of ginseng, such as the epidermis, phloem and xylem were also successfully distinguished with score plots of PCA of diffuse reflectance NIR spectra [58].

NIR fingerprinting in combination with SIMCA, DA and PLS–DA was reported by Lucio-Gutiérrez et al. [62] for rapid identification of *E. senticosus* from other eight herbs, which were related and not related to the *Arialianceae* family, and good results were obtained in the detection of counterfeits and adulterations when using SIMCA and PLS–DA.

Chrysanthemum species as medicine herbs have a long history of cultivation throughout China. Three Chrysanthemum species of Hangju (Dabaiju, Huju, and Xiaobaiju) were identified by machine learning techniques combined with NIR spectroscopy [63]. For Dabaiju, Huju, and Xiaobaiju in calibration sets, the accuracy rates were 98%, 97% and 95%, respectively. While for those in prediction sets, the accuracy rates were 95%, 86% and 93%, respectively.

Rhubarb, one of the most ancient and best known traditional herbal medicines, has more than 40 species widespread in China. However, only three species among the rhubarbs are reported to have medicinal values, and are officially designated as authentic rhubarb, i.e. *Rheum palmatum*, *Rheum tanguticum*, and *Rheum officinale*. The other species of rhubarbs are designated as unauthentic rhubarbs. For the discrimination of authentic and unauthentic rhubarb samples, NIR spectroscopy technique and temperature-constrained cascade correlation models (TCCCNs) were developed by Wang et al. [64]. All of the powdered rhubarb samples were correctly classified by the TCCCN model.

Pelargonium sidoides, a species of the Geraniaceae family, is indigenous to South Africa and abundant in the Eastern Cape Province. Several commercial herbal products which are formulated with *P. sidoides* are marketed in Germany, with Umckaloabo® as probably the most popular and successfully one.

Maree and Viljoen [65] developed a method based on NIR spectroscopy to discriminate *P. sidoides* from *Pelargonium reniforme* (a closely related species). The NIR-spectroscopic data were analyzed using chemomoetrics approaches including PCA and orthogonal projections to latent structures discriminant analysis (OPLS–DA), and were found that OPLS–DA model from a special spectral region (ranging from 4400 to 7400 cm^{-1}) combined with multiplicative signal correlation (MSC) and centre scaled spectral filters was an efficient tool for the differentiation of *P. sidoides* and *P. reniforme* [65]. OPLS–DA score plot for the classification of *P. sidoides* and *P. reniforme* is shown in Fig. 2.

The species authentication of *Acorus calamus* L. (AC) and *Acorus tatarinowii* Schott (AT) using NIR spectroscopy was conducted by Ying et al. [66]. PCA and discriminant partial least squares (DPLS) were utilized. The DPLS models were constructed using a nonmetric dummy variable. AT samples were assigned a numeric value of 1, and AC samples were assigned 2. The classification of the AC and AT samples was on the basis of the 0.5 cut off value. For AT samples, if the predicted value was between 0.5 and 1.5, it meant that the AT sample was classified correctly; otherwise the sample was classified wrongly. And it was an AC sample if the value was between 1.5 and 2.5. Compared to the classification result of PCA, DPLS showed a better, more visual and effective prediction, and all samples were correctly classified.

4. Selected applications of herbal medicine geographical origin discrimination

Literature analysis shows a great number of papers dedicated to herbal medicine geographical origin discrimination using NIR spectroscopy. The potential of NIR spectroscopy method for the discrimination of *Rhizoma Corydalis* according to its geographical origins was evaluated [67]. A training set of such *Rhizoma Corydalis* spectral objects was modeled using LS-SVM, radial BP-ANN, PLS–DA and K-nearest-neighbor (KNN) methods. Comparisons of the
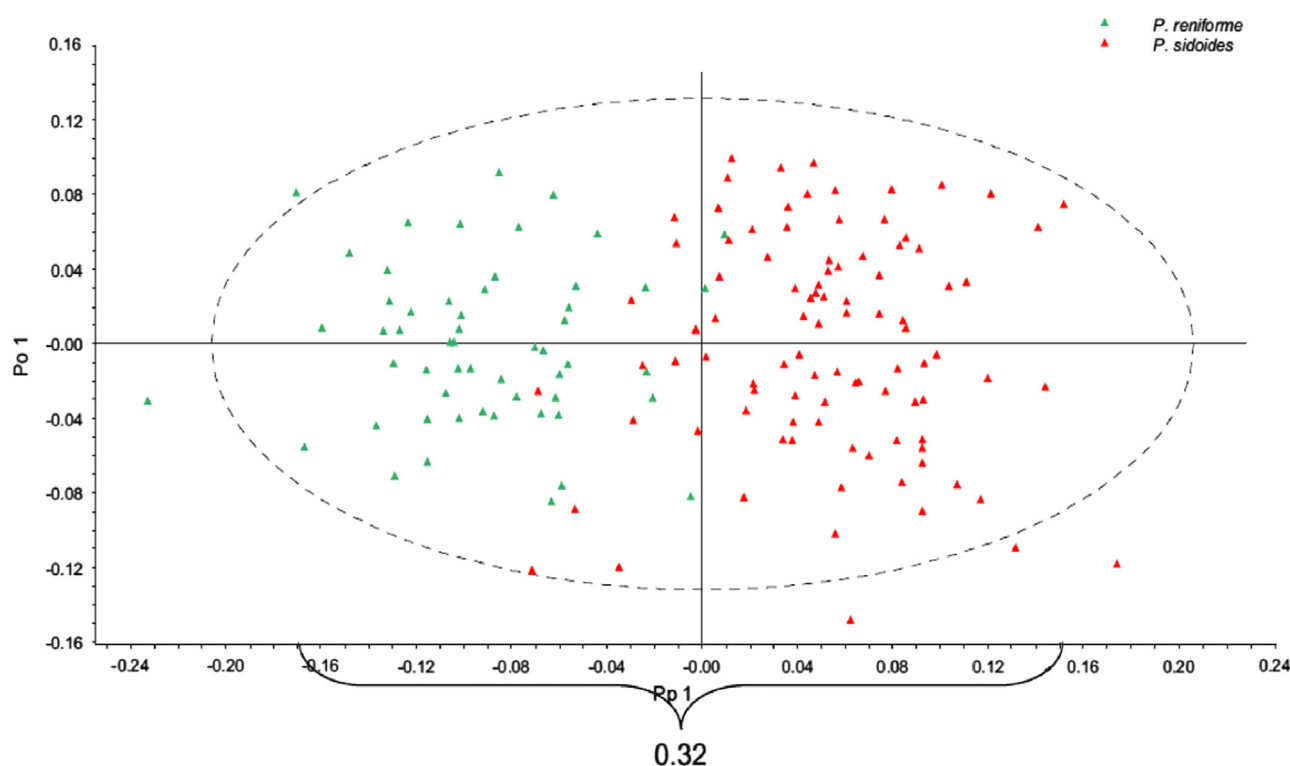

Fig. 2. OPLS-DA score plot for the classification of *Pelargonium sidoides* and *Pelargonium reniforme*. Reprinted from [65] with permission from Elsevier.
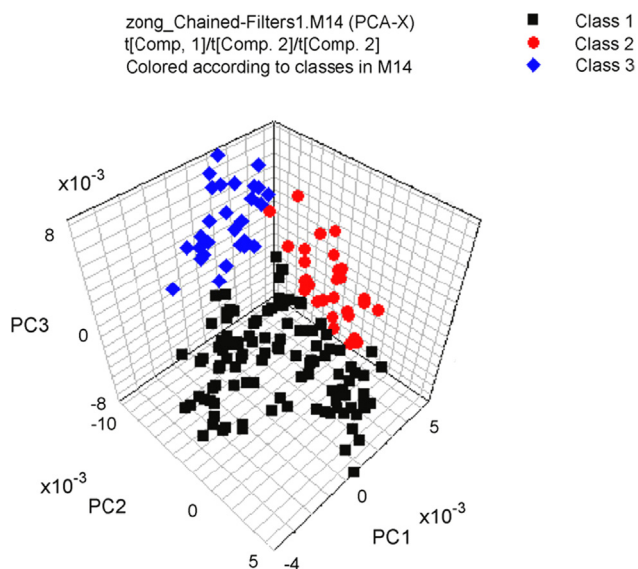
Fig. 3. Three-dimensional score plot using PC1, PC2, and PC3 for discrimination *Ganoderma lucidum* from three provinces, class 1, Shandong Province; class 2, Anhui Province; class 3, Zhejiang Province. Reprinted from [15] with permission from Elsevier.

four different approaches were carried out, and LS-SVM performed best with a correct discrimination rate of over 95%.

The feasibility of using NIR spectroscopy to discriminate *Ganoderma lucidum* according to cultivation area was reported [15]. PCA, discriminant partial least-squares (DPLS) and DA were applied to classify the geographical origins [15]. Excellent classification results can be obtained after optimization of spectral pretreatments. For the samples from three different provinces (Shandong, Anhui and Zhejiang, China), DPLS provided 100% correct discrimination (Fig. 3). Moreover, for samples from six different geographic regions (Jiaxiang, Huangshan, Taishan, Longquan, Jinzhai and Jingdangpu, China), after the standard normal variate correction (SNV) and the first derivative spectral pretreatment, the accuracy rate of the DA model for classifications of the calibration and the validation data set was 96%. Chen et al. [68] also perfomed a method based on the combination of NIR spectroscoy and two chemometrics (the partial least-squares (PLS) and radial basis function (RBF) network) for the quantitative analysis of total polysaccharides and triterpenoids in *G. lucidum* and *Ganoderma atrum* from different origins. Good predictability of the two quantitative models was obtained.

The potential of two-dimensional (2D) NIR correlation spectroscopy to discriminate the geographic regions of Fructus Lycii has also been evaluated [69]. Compared with one-dimensional NIR spectroscopy, 2D NIR correlation spectroscopy could enhance the spectrum with overlapped bands, simplify spectral resolution and provide useful information about temperature-induced spectral intensity variations which can hardly be obtained from one-dimensional NIR spectroscopy. The 2D synchronous and asynchronous spectra showed significant differences within the range from 4950 to 5700 cm^{-1} among samples from different geographic regions.

The combination of NIR spectra with SIMCA was evaluated as a method to predict the geographical origins (China and Korea) of *Angelicae gigantis* Radix, one of the most ancient and widely used herbal medicines in East Asia [70]. In order to suppress baseline variations observed in raw reflectance spectra and enhance spectral features, a second derivative using Savitsky–Golay algorithm with 5 points of smoothing was carried out. Major differences between Chinese and Korean samples could be identified based on the unique 1625 nm band of decursin. The resulting SIMCA model

performed excellently, achieving 100% accuracy for the classification of Korean and Chinese samples.

The potential of NIR spectroscopy was also investigated for the discrimination of Carthami Flos (saffron) geographical origins [71]. It was reported that the diagnosis of the three family tests (Iran, Greece, and Spain) showed a critical probability level of 1×10^{-4}. The interclass distances between different countries demonstrated that Iranian samples were very different from Greek and Spanish samples ($D_{\text{Iran–Greece}} = 180$; $D_{\text{Iran–Spain}} = 319$), whereas Greek and Spanish samples were much similar with lower interclass distances ($D_{\text{Greece–Spain}} = 22$). The proposed NIR approach showed excellent performance for saffron geographical origin discrimination, yielding 100%, 95% and 88% recognition accuracy for Iranian, Greek and Spanish samples, respectively.

A fiber optic diffuse reflectance NIR spectroscopy was applied for the classification of Licorice (*Glycyrrhizia uralensis* Fisch) according to their growing environments, geographic origins, and plant parts [72]. For the raw NIR spectra, different spectral pretreatment methods including MSC and Norris derivative filter were carried out to enhance the differences of NIR spectra among different licorice samples. Licorice samples could be moderately clustered in principle components spaces, and SIMCA provided satisfactory classification results. Additionally, a partial least squares quantitative analysis of glycyrrhizic acid in licorice was carried out, and acceptable results were obtained.

Radix *Salvia miltiorrhiza* Bge. var. alba, named Danshen in China, is one of the most widely used and important TCMs. Duan et al. [73] developed a rapid and nondestructive method based on Fourier transform-NIR spectroscopy for the discrimination of geographical origin. Four geographical origins (i.e. Taian, Laiwu, Rongcheng, and Guangrao) of raw *S. miltiorrhiza* var. *alba* samples were correctly discriminated using DA.

Lee et al. [74] investigated the potential of NIR spectroscopy for its ability to nondestructively discriminate the geographic origins of Scrophulariae Radix. It has been widely used in eastern Asia for the treatment of fever, swelling, neuritis, constipation, pharyngitis and laryngitis [75]. The application of PCA to NIR spectra leads to a clear separation of Andong sample from the others. And for the two major neuroprotective constituents (8-*O*-(*E-p*-methoxycinnamoyl)-harpagide, and *E-p*-methoxycinnamic acid) of Scrophularia spp., a quantitative PLS regression method was successfully established.

The ability of NIR spectroscopy was investigated to discriminate the geographical origins of *Scutellariae* radix, a widely used TCM [76]. Using the Integrating-Sphere (Thermo Fisher, Pittsburgh, PA, USA), the NIR spectra were collected in the diffused reflectance mode. Two different classification methods, DA and DPLS, were investigated and compared. Since for the DPLS method, the linear relationship and mutual influence of the spectra matrix and the origin information were considered, the DPLS was more effective than DA with an accuracy rate of 100% for the discrimination of the geographical origins of *Scutellariae* radix.

NIR spectroscopy was investigated as a method for the discrimination of *peucedanum* origins [77]. PCA was carried out for the extraction of relevant information; ANN with PCs as input variables (PC-ANN) and PLS–DA were used to build the classification models. The results showed that PCA could hardly serve the purpose of identifying the geographical origin of *peucedanum*. In comparison with PCA, both PC-ANN model based on 7 principal components (PCs) and PLS–DA model based on 3 latent variables achieved identification rate of 100%.

Nondestructive discrimination of *Fructus forsythiae* from different geographical origins was reported using NIR spectroscopy combined with clustering analysis and DA [78]. For clustering analysis, the best results were achieved in the NIR spectra ranging from 4092 to 8008 cm^{-1} after pretreatments with second

derivative and Norris smoothing. One hundred and thirty-three samples were divided into three categories corresponding to the *Fructus forsythiae* samples from three different provinces of China, but some Shanxi samples were mis-classified into samples from Henan, and some Shaanxi samples were misjudged into samples from Shanxi. For the DA model, the full NIR spectral range instead of specific spectral regions was used. After first derivative and Norris smoothing, PCA was performed, and the top 7 PCs were used to establish the DA model. The accuracy rate of the internal cross-validation identification was 97%.

The feasibility of NIR spectroscopy integrated with chemometrics to predict the geographic origins of *Codonopsis pilosula* was reported by Li et al. [79]. Two chemometric methods, random forests and KNN, were carried out for the classification model development and geographical origin prediction. The predictive capability of the classification models developed based on the raw and the SNV+ first derivative converted NIR spectra were compared. High accuracy rate of 94% could be obtained by both random forests and KNN for the independent test set.

Li. et al. [80] exploited a qualitative method based on NIR spectroscopy applied for the geographical origin identification of *Lonicerae Japonicae Flos* [80]. One hundred *Lonicerae Japonicae Flos* samples were collected from different origins, and NIR spectral acquisition was carried out on two NIR instruments form different manufacturers, one from Thermo Fisher Scientific Inc. and the other one from Buchi Inc. NIR model based on SIMCA was established for the differentiation of *Lonicerae Japonicae Flos* from different producing areas. Using the DA model above, all the samples from Henan Province could be predicted with no misjudgment, while for the samples from other origins, 6 in 68 were incorrectly judged. Partial least squares regression (PLSR) models were also developed, and the model transformation between two NIR instruments was also investigated and successfully applied for the quantification of six organic acids in *Lonicerae Japonicae Flos.* Li et al. [81] also developed a Wavelet-based classification and influence matrix analysis method for the rapid discrimination of *Salviae miltiorrhizae* radix according to the geographical origins

with NIR, with no misjudgment in both cross validation and prediction set.

Paeoniae Radix, which has a wide spectrum of pharmacological properties and physiological activities, is extensively used in China [82]. *Paeoniae* Radix from different cultivated regions has its own Chinese name. For *Paeoniae* Radix cultivated in Zhejiang, Sichuan and Anhui, China, the corresponding Chinese name is 'hangshao', 'chuanshao', and 'boshao', respectively. NIR spectroscopy combined with PCA was employed for the differentiation of *Paeoniae* Radix from the three cultivation areas mentioned above [83]. A quantitative approach based on NIR spectroscopy was also established for the determination of paeoniflorin, albiflorin, and benzoylalbiflorin in *Paeoniae* Radix [83].

In Table 1, a summary of applications related to geographical origin discrimination of herbal medicines is given.

5. Conclusions

The use of complementary and alternative medicine, especially herbal medicine, is becoming popular in the general population worldwide. Parallel to the growing global interest in alternative medical therapies, similar trends have also been conducted in research activities dealing with the evaluation of efficacy and safety of herbal medicines worldwide [84]. Traditionally, discrimination of herbal medicines is carried out based on its morphology, one or two specific compounds' chromatography identification, and/or quantification. However, according to the theory of herbal medicine, the quality of herbal medicine should be regarded as a whole. Conventional analytical methods can hardly provide a complete profile of the herbs, so they are usually useless for species authentication and geographical origin discrimination of herbal medicines. Thereby, over the past decades, the analysis of herbal medicines has begun to emphasize more on their basic theories, and their integrative and holistic properties [85]. Vibrational spectroscopy, including NIR, mid-IR and Raman, offers authentication analysis of herbal medicine as a whole matrix.

Table 1
NIR spectroscopy used for geographical origin discrimination of herbal medicines[a].

Herbal medicine	Wavelength range (cm^{-1})	Pretreatment method	Method	Correct discrimination (%)	Ref.
Rhizoma Corydalis	4000–10000	wavelet transform	LS-SVM, radial BP-ANN, PLS–DA, KNN	85–100	[67]
Ganoderma lucidum	4011–5114, 6996–7629	SNV+1st derivative	PCA	[b]	[15]
		SNV+2nd derivative	DPLS	100	
		SNV+1st derivative	DA	97	
Angelicae gigantis Radix	5882–6668	2nd derivative	SIMCA	100	[70]
Carthami Flos	4000–10000	2nd derivative	DA	88–100	[71]
Glycyrrhiza uralensis Fisch	4500–8500	MSC+1st derivative	PCA	[c]	[72]
			SIMCA	[b]	
Radix *Salvia miltiorrhiza*	4000–10000	MSC+1st derivative+Savitzky-Golay smoothing	DA	100	[73]
Scrophulariae Radix	4000–10000	1st derivative	PCA	[b]	[74]
Scutellariae Radix	4000–10000	SNV+2nd derivative+Savitzky-Golay smoothing	DA	92–94	[76]
			DPLS	100	
Peucedanum	3500–8500	1st derivative+autoscale	PC-ANN	100	[77]
			PLS–DA	100	
Fructus forsythiae	4100–11000	1st derivative+Norris smoothing	DA	97	[78]
Codonopsis pilosula	7503–6904, 5106–4017	SNV+1st derivative	Random forests, KNN	94	[79]
Lonicerae Japonicae Flos	4100–10000	SNV+2nd derivative	SIMCA	[b]	[80]
Paeoniae Radix	4000–10000	MSC+1st derivative	PCA	[b]	[82]

SNV: standard normal variate correction, MSC: multiplicative signal correlation, LS-SVM: least-square support vector machine, BP-ANN: back-propagation artificial neural network, PLS–DA: partial least squares discriminated analysis, KNN: K-nearest-neighbor, PCA: principal component analysis, DPLS: discriminant partial least squares, DA: discriminant analysis, SIMCA: soft independent modeling class analogy, PC-ANN: principal component-artificial neural network.

[a] The potential of two-dimensional (2D) NIR correlation spectroscopy to discriminate the geographic regions of Fructus Lycii [69] is not included in Table 1.
[b] Acceptable discrimination.
[c] Moderate discrimination.

Especially, modern NIR spectroscopy, in combination with chemometric methods, offers reliable species authentication and accurate geographical origin discrimination of herbal medicine. It is expected NIR spectroscopy in combination of chemometric methods to be further employed in the authentication and quality control of herbal medicines.

Acknowledgments

References

[1] A. Der Marderosian, Guide to popular natural products. Facts and Comparisons, Lippincott Williams & Wilkins, St. Louis, MO, USA, 1999.

[2] A.A. Bunaciu, H.Y. Aboul-Enein, S. Fleschin, Recent applications of fourier transform infrared spectrophotometry in herbal medicine analysis, Appl. Spectrosc. Rev. 46 (2011) 251–260.

[3] A. Rohman, A. Nugroho, E. Lukitaningsih, et al., Application of vibrational spectroscopy in combination with chemometrics techniques for authentication of herbal medicine, Appl. Spectrosc. Rev. 49 (2014) 603–613.

[4] O.F. Kunle, H.O. Egharevba, P.O. Ahmadu, Standardization of herbal medicines: a review, Int. J. Biodivers. Conserv. 4 (2012) 101–112.

[5] Y.D. Yi, I.M. Chang, An overview of traditional Chinese herbal formulae and a proposal of a new code system for expressing the formula titles, Evidence Based Complement. Altern. Med. 1 (2004) 125–132.

[6] Y. Jiang, B. David, P. Tu, et al., Recent analytical approaches in quality control of traditional Chinese medicines: a review, Anal. Chim. Acta 657 (2010) 9–18.

[7] P. Wang, L. Li, H. Yang, et al., Chromatographic fingerprinting and quantitative analysis for the quality evaluation of Xinkeshu tablet, J. Pharm. Anal. 2 (2012) 422–430.

[8] H.A. Gad, S.H. El-Ahmady, M.I. Abou-Shoer, et al., Application of chemometrics in authentication of herbal medicines: a review, Phytochem. Anal. 24 (2013) 1–24.

[9] Y.Z. Liang, P. Xie, K. Chan, Quality control of herbal medicines, J. Chromatogr. B 812 (2004) 53–70.

[10] P. Wang, B. Wang, J. Xu, et al., Detection and chemical profiling of Ling-Gui-Zhu-Gan decoction by ultra performance liquid chromatography-hybrid linear ion trap-Orbitrap mass spectrometry, J. Chromatogr. Sci. 53 (2) (2014) 263–273.

[11] World Health Organization, WHO Guidelines on Good Manufacturing Practices (GMP) for Herbal Medicines. 2007.

[12] FDA, Guidance for Industry, Botanical Drug Products, Rockville, MD, 2004.

[13] Herbal Medicinal Products Committee (HMPC). Guideline on Good Agricultural Polysaccharand Collection Practice (GACP) for Starting Materials of Herbal Origin. EMEA/HMPC/246816/2005, London, 2006.

[14] Herbal Medicinal Products Committee (HMPC), Guideline on Specifications: Test Procedures and Acceptance Criteria for Herbal Substances, Herbal Preparations and Herbal Medicinal Products/Traditional Herbal Medicinal Products. CPMP/QWP/2820/00 Rev 1, London, 2006.

[15] Y. Chen, M.Y. Xie, Y. Yan, et al., Discrimination of *Ganoderma lucidum* according to geographical origin with near infrared diffuse reflectance spectroscopy and pattern recognition techniques, Anal. Chim. Acta 618 (2008) 121–130.

[16] M. Papagianni, S.E. Nokes, K. Filer, Submerged and solid-state phytase fermentation by *Aspergillus niger*: effects of agitation and medium viscosity on phytase production, fungal morphology and inoculum performance, Food Technol. Biotechnol. 39 (2001) 319–326.

[17] R. Gobalakrishnan, M. Kulandaivelu, R. Bhuvaneswari, et al., Screening of wild plant species for antibacterial activity and phytochemical analysis of Tragia involucrata L, J. Pharm. Anal. 3 (2013) 460–465.

[18] M. Laasonen, T. Harmia-Pulkkinen, C.L. Simard, et al., Fast identification of *Echinacea purpurea* dried roots using Near-infrared spectroscopy, Anal. Chem. 74 (2002) 2493–2499.

[19] C.W. Chang, D.A. Laird, M.J. Mausbach, et al., Near-infrared reflectance spectroscopy–principal components regression analyses of soil properties, Soil Sci. Soc. Am. J. 65 (2001) 480–490.

[20] H. Cen, Y. He, Theory and application of near infrared reflectance spectroscopy in determination of food quality, Trends Food Sci. Technol. 18 (2007) 72–83.

[21] H. Huang, H. Yu, H. Xu, et al., Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: a review, J. Food Eng. 87 (2008) 303–313.

[22] C. Zhang, J. Su, Application of near infrared spectroscopy to the analysis and fast quality assessment of traditional Chinese medicinal products, Acta Pharm. Sin. 4 (2014) 182–192.

[23] D. Cozzolino, Near infrared spectroscopy in natural products analysis, Planta Med. 75 (2009) 746–756.

[24] Handbook of Near-infrared Analysis, Revised and Expanded, in: D.A. Burns, E.W. Ciurczak (Eds.), second ed.,Marcel Dekker, Inc., New York, Basel, Hingkong, 2001.

[25] H.W. Siesler, Y. Ozaki, S. Kawata, et al., Near-infrared Spectroscopy Principles Instruments Applications, John Wiley & Sons, New York, 2008, pp. 125–128.

[26] K. Norris, R. Barnes, J. Moore, et al., Predicting forage quality by infrared replectance spectroscopy, J. Anim. Sci. 43 (1976) 889–897.

[27] C. De Bleye, P.F. Chavez, J. Mantanus, et al., Critical review of near-infrared spectroscopic methods validations in pharmaceutical applications, J. Pharm. Biomed. Anal. 69 (2012) 125–132.

[28] Y. Roggo, P. Chalus, L. Maurer, et al., A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies, J. Pharm. Biomed. Anal. 44 (2007) 683–700.

[29] Z. Xiaobo, Z. Jiewen, M.J.W. Povey, et al., Variables selection methods in near-infrared spectroscopy, Anal. Chim. Acta 667 (2010) 14–32.

[30] Q. Dong, H. Zang, A. Liu, et al., Determination of molecular weight of hyaluronic acid by near-infrared spectroscopy, J. Pharm. Biomed. Anal. 53 (2010) 274–278.

[31] J. Luypaert, D.L. Massart, Y. Vander Heyden, Near-infrared spectroscopy applications in pharmaceutical analysis, Talanta 72 (2007) 865–883.

[32] G. Reich, Near-infrared spectroscopy and imaging: basic principles and pharmaceutical applications, Adv. Drug Delivery Rev. 57 (2005) 1109–1143.

[33] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, et al., Handbook of Chemometrics and Qualimetrics, Part A, Elsevier, Amsterdam, 1997.

[34] S. Naik, V.V. Goud, P.K. Rout, et al., Production of first and second generation biofuels: a comprehensive review, Renewable Sustainable Energy Rev. 14 (2010) 578–597.

[35] A. Murugesan, C. Umarani, T. Chinnusamy, et al., Production and analysis of bio-diesel from non-edible oils—a review, Renewable Sustainable Energy Rev. 13 (2009) 825–834.

[36] V. Bellon-Maurel, A. McBratney, Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils—critical review and research perspectives, Soil Biol. Biochem. 43 (2011) 1398–1410.

[37] M. Jimare Benito, C. Bosch Ojeda, F. Sanchez Rojas, Process analytical chemistry: applications of near infrared spectrometry in environmental and food analysis: an overview, Appl. Spectrosc. Rev. 43 (2008) 452–484.

[38] T. De Beer, A. Burggraeve, M. Fonteyne, et al., Near infrared and Raman spectroscopy for the in-process monitoring of pharmaceutical production processes, Int. J. Pharm. 417 (2011) 32–47.

[39] M. Ferrari, V. Quaresima, A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application, NeuroImage 63 (2012) 921–935.

[40] T.J. Huppert, S.G. Diamond, M.A. Franceschini, et al., HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain, Appl. Opt. 48 (2009) D280–D298.

[41] A.O. Awiti, M.G. Walsh, K.D. Shepherd, et al., Soil condition classification using infrared spectroscopy: a proposition for assessment of soil condition along a tropical forest-cropland chronosequence, Geoderma 143 (2008) 73–84.

[42] A. Sakudo, Y. Suganuma, T. Kobayashi, et al., Near-infrared spectroscopy: promising diagnostic tool for viral infections, Biochem. Bioph. Res. Co 341 (2006) 279–284.

[43] N. Prieto, R. Roehe, P. Lavín, et al., Application of near infrared reflectance spectroscopy to predict meat and meat products quality: a review, Meater. Sci. 83 (2009) 175–186.

[44] S. Landau, T. Glasser, L. Dvash, Monitoring nutrition in small ruminants with the aid of near infrared reflectance spectroscopy (NIRS) technology: a review, Small Ruminant Res. 61 (2006) 1–11.

[45] P. Wang, H. Zhang, H. Yang, et al., Rapid determination of major bioactive isoflavonoid compounds during the extraction process of Kudzu (Pueraria lobata) by near-infrared transmission spectroscopy, Spectrochim. Acta A 137 (2015) 1403–1408.

[46] Y.A. Woo, H.J. Kim, J. Cho, Identification of herbal medicines using pattern recognition techniques with near-infrared reflectance spectra, Microchem. J. 63 (1999) 61–70.

[47] Y. Zhao, J. Zhang, T. Yuan, et al., Discrimination of wild Paris based on near infrared spectroscopy and high performance liquidchromatography combined with multivariate analysis, PLoS One 9 (2014) e89100.

[48] C.O. Chan, C.C. Chu, D.K.W. Mok, et al., Analysis of berberine and total alkaloid content in Cortex Phellodendri by near infrared spectroscopy (NIRS) compared with high-performance liquid chromatography coupled with ultra-visible spectrometric detection, Anal. Chim. Acta 592 (2007) 121–131.

[49] M. Kudo, R.A. Watt, A.C. Moffat, Rapid identification of digitalis purpurea using near-infrared reflectance spectroscopy, J. Pharm. Pharmacol. 52 (2000) 1271–1277.

[50] Z. Zhang, T.N. Lam, Z. Zuo, Radix puerariae: an overview of its chemistry, pharmacology, pharmacokinetics, and clinical use, J. Clin. Pharmacol. 53 (2013) 787–811.

[51] C.C. Lau, C.O. Chan, F. Chau, et al., Rapid analysis of Radix puerariae by near-infrared spectroscopy, J. Chromatogr. A 1216 (2009) 2130–2135.

[52] C. Wang, B. Xiang, W. Zhang, Application of two-dimensional near-infrared (2D-NIR) correlation spectroscopy to the discrimination of three species of Dendrobium, J. Chemometr. 23 (2009) 463–470.

[53] Q. Fan, Y. Wang, P. Sun, et al., Discrimination of Ephedra plants with diffuse reflectance FT-NIRS and multivariate analysis, Talanta 80 (2010) 1245–1250.

[54] J. Mao, J. Xu, Discrimination of herbal medicines by molecular spectroscopy and chemical pattern recognition, Spectrochim. Acta A 65 (2006) 497–500.

[55] J.R. Lucio-Gutierrez, J. Coello, S. Maspoch, Expeditious identification and semi-quantification of Panax ginseng using near infrared spectral fingerprints and multivariate analysis, Anal. Methods 5 (2013) 857–865.

[56] X. Chen, D. Wu, Y. He, et al., Nondestructive differentiation of Panax species using visible and shortwave near-Infrared spectroscopy, Food Bioprocess Technol. 4 (2011) 753–761.

[57] Y.W. Huang, J.S. Jacqueline, L. Lei, Research on fast discrimination between Panax ginseng and Panax quinquefolium based on near infrared spectroscopy, Spectrosc. Spectra Anal. 30 (2010) 2954–2957.

[58] Y. Wu, Y. Zheng, Q. Li, et al., Study on difference between epidermis, phloem and xylem of Radix ginseng with near-infrared and infrared spectroscopy coupled with principal component analysis, Vib. Spectrosc. 55 (2011) 201–206.

[59] T.S. Li, Asian and American ginseng—a review, HortTechnology 5 (1995) 27–34.

[60] M.M Iwu, Handbook of African Medicinal Plants, CRC press, 2014.

[61] Y.A. Woo, C.H. Cho, H.J. Kim, et al., Classification of cultivation area of ginseng by near infrared spectroscopy and ICP-AES, Microchem. J. 73 (2002) 299–306.

[62] J.R. Lucio-Gutiérrez, J. Coello, S. Maspoch, Application of near infrared spectral fingerprinting and pattern recognition techniques for fast identification of *Eleutherococcus senticosus*, Food Res. Int. 44 (2011) 557–565.

[63] C.W. Chen, H. Yan, B.X. Han, Rapid identification of three varieties of Chrysanthemum with near infrared spectroscopy, Rev. Bras. Farmacogn. 24 (2014) 33–37.

[64] F. Wang, Z. Zhang, X. Cui, et al., Identification of rhubarbs by using NIR spectrometry and temperature-constrained cascade correlation networks, Talanta 70 (2006) 1170–1176.

[65] J.E. Maree, A.M. Viljoen, Fourier transform near- and mid-infrared spectroscopy can distinguish between the commercially important *Pelargonium sidoides* and its close taxonomic ally *P. reniforme*, Vib. Spectrosc. 55 (2011) 146–152.

[66] X. Ying, Y. Pei, M. Liu, et al., Discrimination and quantification analysis of Acorus calamus L. and Acorus tatarinowii Schott with near-infrared reflection spectroscopy, Anal. Methods 6 (2014) 4212–4218.

[67] Y. Lai, Y. Ni, S. Kokot, Discrimination of Rhizoma Corydalis from two sources by near-infrared spectroscopy supported by the wavelet transform and least-squares support vector machine methods, Vib. Spectrosc. 56 (2011) 154–160.

[68] Y. Chen, M. Xie, H. Zhang, et al., Quantification of total polysaccharides and triterpenoids in *Ganoderma lucidum* and *Ganoderma atrum* by near infrared spectroscopy and chemometrics, Food Chem. 135 (2012) 268–275.

[69] J. Lu, B. Xiang, H. Liu, et al., Application of two-dimensional near-infrared correlation spectroscopy to the discrimination of Chinese herbal medicine of different geographic regions, Spectrochim. Acta A 69 (2008) 580–586.

[70] Y.A. Woo, H.J. Kim, K.R. Ze, et al., Near-infrared (NIR) spectroscopy for the non-destructive and fast determination of geographical origin of *Angelicae gigantis* Radix, J. Pharm. Biomed. Anal. 36 (2005) 955–959.

[71] A. Zalacain, S.A. Ordoudi, E.M. Díaz-Plaza, et al., Near-infrared spectroscopy in saffron quality control: determination of chemical composition and geographical origin, J. Agr. Food Chem. 53 (2005) 9337–9341.

[72] L. Wang, F.S.C. Lee, X. Wang, Near-infrared spectroscopy for classification of licorice (*Glycyrrhizia uralensis* Fisch) and prediction of the glycyrrhizic acid (GA) content, LWT—Food Sci. Technol. 40 (2007) 83–88.

[73] X. Duan, D. Zhang, L. Nie, et al., Rapid discrimination of geographical origin and evaluation of antioxidant activity of *Salvia miltiorrhiza* var. alba by Fourier transform near infrared spectroscopy, Spectrochim. Acta A 122 (2014) 751–757.

[74] D.Y. Lee, S.H. Kim, Y.C. Kim, et al., Discrimination of Scrophulariae Radix according to geographical origin and determination of active constituents by near infrared spectroscopy (NIRS), Microchem. J. 99 (2011) 213–217.

[75] S.U. Park, Y.A. Chae, P. Facchini, Genetic transformation of the figwort, Scrophularia buergeriana Miq., an Oriental medicinal plant, Plant Cell Rep. 21 (2003) 1194–1198.

[76] W. Li, L. Xing, Y. Cai, et al., Classification and quantification analysis of Radix scutellariae from different origins with near infrared diffuse reflection spectroscopy, Vib. Spectrosc. 55 (2011) 58–64.

[77] J.Y. Zhu, B. Chen, H. Yan, et al., Rapid identification of peucedanum geographical growing areas through near infrared spectroscopy, in: 2011 Fourth International Conference on Biomedical Engineering and Informatics (BMEI), IEEE, 2011, pp. 1772–1776.

[78] Y. Bai, X. Wang, J. Lei, et al., Discrimination of *Fructus forsythiae* according to geographical origin with near-infared spectroscopy, in: 2012 International Conference on Biomedical Engineering and Informatics (BMEI), IEEE, 2012, pp. 175–178.

[79] B. Li, Y. Wei, H. Duan, et al., Discrimination of the geographical origin of *Codonopsis pilosula* using near infrared diffuse reflection spectroscopy coupled with random forests and k-nearest neighbor methods, Vib. Spectrosc. 62 (2012) 17–22.

[80] W. Li, Z. Cheng, Y. Wang, et al., Quality control of Lonicerae Japonicae Flos using near infrared spectroscopy and chemometrics, J. Pharm. Biomed. Anal. 72 (2013) 33–39.

[81] W. Li, H. Qu, Wavelet-based classification and influence matrix analysis method for the fast discrimination of Chinese herbal medicines according to the geographical origins with near infrared spectroscopy, J. Innovative Opt. Health Sci. 7 (2014).

[82] H.K. Wu, S.J. Sheu, Capillary electrophoretic determination of the constituents of Paeoniae Radix, J. Chromatogr. A 753 (1996) 139–146.

[83] X. Luo, X. Yu, X. Wu, et al., Rapid determination of Paeoniae Radix using near infrared spectroscopy, Microchem. J. 90 (2008) 8–12.

[84] R.C. Kessler, R.B. Davis, D.F. Foster, et al., Long-term trends in the use of complementary and alternative medical therapies in the United States, Ann. Intern. Med. 135 (2001) 262–268.

[85] P. Li, L.W. Qi, E.H. Liu, et al., Analysis of Chinese herbal medicines with holistic approaches and integrated evaluation models, TrAC—Trend Anal. Chem. 27 (2008) 66–77.