Journal of Innovative Optical Health Sciences
Vol. 13, No. 2 (2020) 2050006 (13 pages)

DOI: 10.1142/S1793545820500066

Rapid and simultaneous determination of moisture and berberine content in Coptidis Rhizoma and Phellodendri Chinensis Cortex by near-infrared spectroscopy and chemometrics

Siyu Zhang*, Ming Chen*, Yong Chen*, Wei Xiao†, Yerui Li‡, Jun Wang‡, Xuesong Liu*,§,‖ and Yongjiang Wu*,¶,‖

*Institute of Modern Chinese Medicine
College of Pharmaceutical Science, Zhejiang University
Hangzhou 310058, P. R. China

†Jiangsu Kanion Pharmaceutical Co., Ltd.
Lianyungang 222001, P. R. China

‡Suzhou ZeDaXingBang Pharmaceutical Co., Ltd.
Suzhou 215000, P. R. China
§liuxuesong@zju.edu.cn
¶yjwu@zju.edu.cn

Received 22 October 2019
Accepted 1 December 2019
Published 20 December 2019

Coptidis Rhizoma (Chinese: Huanglian) and Phellodendri Chinensis Cortex (Chinese: Huangbo) are widely used Traditional Chinese Medicine, and often used in combination because of their similar pharmacological effects in clinical practice. However, the quality control methods of the two drugs are different and complicated, which is time consuming and laborious in practical application. In this paper, rapid and simultaneous determination of moisture and berberine in Coptidis Rhizoma (CR) and Phellodendri Chinensis Cortex (PC) was realized by near-infrared spectroscopy (NIRs) combined with global models. Competitive adaptive reweighted sampling (CARS) and successive projection algorithm (SPA) method were applied for variable selection. Principal component analysis (PCA) and partial least squares regression method (PLSR) were applied for qualitative and quantitative analysis, respectively. The characteristic variables of berberine showed similarity and consistency in distribution, providing basis for the global models. For moisture content, the global model had relative standard error of prediction set (RSEP) value of 3.04% and 2.53% for CR and PC, respectively. For berberine content, the global model had RSEP value of 5.41% and 3.97% for CR and PC, respectively. These results indicated the global models based on CARS-PLS method achieved satisfactory prediction for moisture and berberine content, improving the determination efficiency. Furthermore, the greater range and larger

‖Corresponding authors.

number of samples enhanced the reliance of the global model. The NIRs combined with global models could be a powerful tool for quality control of CR and PC.

Keywords: Near-infrared spectroscopy; global model; Coptidis Rhizoma; Phellodendri Chinensis Cortex; berberine.

Abbreviations

CARS	:	Competitive adaptive reweighted sampling
CR	:	Coptidis Rhizoma
LVs	:	Latent variables
NIRs	:	Near-infrared spectroscopy
PC	:	Phellodendri Chinensis Cortex
PCA	:	Principal component analysis
PLSR	:	Partial least squares regression
Rc	:	Correlation coefficient of the calibration set
Rp	:	Correlation coefficient of the prediction set
RSEC	:	Relative standard error of calibration set
RSEP	:	Relative standard error of prediction set
RMSE	:	Root mean square error
RMSEC	:	Root mean square error of calibration set
RMSECV	:	Root mean square error of cross-validation
RMSEP	:	Root mean square error of prediction set
SPA	:	Successive projections algorithm
SPXY	:	Sample set portioning based on joint *x*-*y* distance
TCM	:	Traditional Chinese medicine

1. Introduction

Coptidis Rhizoma (CR) (Chinese: Huanglian) and Phellodendri Chinensis Cortex (PC) (Chinese: Huangbo) are both critical traditional Chinese medicine (TCM), widely used for heat-clearing and detoxification. The main pharmacological ingredients in these two TCM are similar and one of the most important is berberine. Berberine has various pharmacological effects such as antihyperlipidemic activity, inhibiting helicobacter pylori and anti-inflammatory activity.[1–3] Therefore, berberine are chosen to be quality indicators for both CR and PC according to Chinese Pharmacopeia (2015 edition, Vol. 1). Furthermore, CR and PC are known as matched pair of mutual assistance in TCM.[4] It is preferred to combine these two TCM clinically such as Huanglian-Jie-Du-Tang.[5,6] However, the ingredients contained in CR and PC are complicated, and their quality has a wide range of factors. Most medicinal plants are cultivated artificially, as a consequence, different origins and harvesting time lead to the significant quality difference. Furthermore, the approach of transportation and storage also influences quality greatly. Therefore, the quality control of CR and PC has become a major problem, gaining more and more attention. The traditional methods including microscopic identification and chromatographic analysis are time consuming and costly during the pretreatment and determination. The samples need to be pretreated such as weighted accurately, extracted and purified by harmful organic reagent many times. Therefore, the rapid and economical method for quality control of CR and PC are urgently needed.

Near-infrared spectroscopy (NIRs), with the characteristics of fast speed, high efficiency, low cost and nondestructiveness, has been widely used in many fields.[7–10] Compared with traditional analysis, the cumbersome pretreatments are eliminated. The samples are determined with their original state by NIR method. Furthermore, the prediction process only takes few seconds, saving a lot of time and money. The absorption of NIR mainly caused by C–H, N–H, O–H and S–H[11] has correspondence to chemical structure. The spectrum of the same chemical molecule could have similarity and versatility based on its structure characteristics. At present, the quality control of TCM has achieved efficient results by NIRs, in both qualitative identification and multi-index quantitative analysis.[12,13] However, the absorption peaks always overlap heavily and must be analyzed by chemometrics.[14]

When NIRs combined with chemometrics was used for quality control of TCM, specific models are usually constructed for different herbs and chemical components.[15–17] However, there are similar or even common components in different TCM. The non-global models between different TCM make it necessary to construct and switch the model according to specific TCM, which is inefficient and laborious. Global models that could achieve simultaneous determination of two or more herbs are expected. Furthermore, the predictive accuracy and robustness of the models could be improved by using larger and more varied sample sets.[18]

At present, many studies have proved the feasibility and superiority of global models using NIR

spectroscopy. Torres has developed robust and accurate universal model of two citrus species: oranges and mandarins, which is applicable to all fruits belonging to the genus Citrus.[19] Universal models for molecules in meat such as proteins, fat and moisture content have already been studied and recommended for official first action of Association of Official Analytical Chemists (AOAC), obviating the need to develop specific models for each species.[20] In the pharmaceutical fields, some researches of global models focused on different manufacturers and the determination of moisture content.[21,22] However, the global model for medicinal ingredients has not yet explored. Considering the common berberine ingredient in CR and PC and their extensive clinical applications, global models for the two herbs were expected to achieve rapid determination.

The object of this work was to perform the rapid and simultaneous quantitative analysis of moisture and berberine content in two herbs by developing global models using NIRs. To our best knowledge, this study was the first report on global models of CR and PC by NIRs coupled with chemometrics. The performance of the models for moisture and berberine content was optimized based on variable selection method, robust and efficient determination of the same quality control indicators in two herbs were achieved.

2. Materials and Methods

2.1. *Samples and reagents*

Different batches of CR were collected from Hubei and Sichuan provinces of China, with total 94 samples. Different batches of PC were collected from Sichuan, Anhui and Henan provinces, with total 99 samples.

Bruker Matrix-F Fourier transform NIR spectrometer (Bruker Optics Inc., Germany) was used for spectra collection. An Agilent 1200 HPLC system (Agilent Technologies, USA) with vacuum degasser, quaternary pump, autosampler, thermostatic column compartment and UV detector was employed to determine the berberine content in CR and PC. Standard of berberine (MUST-17110105; purity $> 98.0\%$) was purchased from Chengdu Must Bio-Technology Co., Ltd (Chengdu, P. R. China).

2.2. *Determination of reference values*

The content detection of moisture and berberine in CR and PC refers to the Chinese Pharmacopoeia (2015 edition, Vol. 1).[4] The moisture content of CR and PC was detected by oven drying method. The samples were accurately weighed and dried under 100–105° C for 5 h. After 5 h, the samples were placed in a desiccator, and cooled for 30 min. After they were accurately weighed, the samples were dried at the above temperature for 1 h again. The samples were accurately weighed again after 1 h and it was ensured that the difference between the two measurements is less than 5 mg. The moisture content was calculated based on the weight lost.

The berberine content was detected by HPLC method. The analysis was performed on an X Bridge Shield RP18 column (250 mm × 4.6 mm, 5 μm). The content detected by legal method was used as reference values.

Samples of CR were extracted by ultrasonic extraction with 50 mL of methanol-hydrochloric acid (100:1) mixed solution and detected with elution of acetonitrile-0.05 $\text{mol} \cdot \text{L}^{-1}$ potassium dihydrogen phosphate solution (50:50) at flow rate of $1\,\text{mL} \cdot \text{min}^{-1}$. The injection volume was 10 μL and the detective wavelength was 345 nm. The PC samples were extracted by ultrasound with 25 mL mobile phase. The berberine content of PC samples were performed with elution of acetonitrile −0.1% phosphoric acid solution (36:64) at flow rate of $1\,\text{mL} \cdot \text{min}^{-1}$ by HPLC method. The injection volume was 5 μL and the detective wavelength was 284 nm.

2.3. *Collection of spectral data*

Powder of CR and PC was crushed by grinder and sieving through 80 meshes. The spectra of the samples were collected at room temperature. The collection conditions were as follows: the scanning range was 12000–4000 cm^{-1}, the scanning times were 32. The resolution was set as $8\,\text{cm}^{-1}$. Each sample was scanned three times, and the average spectrum used for model construction was obtained by OPUS 7.0 software.

2.4. *Chemometrics and data analysis*

2.4.1. *Division of calibration and prediction set*

The calibration set and prediction set were obtained by sample set portioning based on joint x–y distance (SPXY) method. The SPXY method is

based on the distance between the samples, considering both the NIRs data and the reference value, which can improve the predictive ability of the model.[23]

2.4.2. *Principal component analysis*

The central purpose of principal component analysis (PCA) is to reduce the dimensionality of the data and convert the original variables, so that some new variables are linear combinations of the original variables.[24] At the same time, these variables should express the data characteristics of the original variables as much as possible without losing information. PCA decomposes the spectral matrix $X(n \times m)$ into the sum of the outer products of m vectors, which can be written in the following matrix[25]:

$$X = TP^T, \tag{1}$$

where T is called the score matrix and P is called the load matrix. For the PCA model, the score matrix T can be used for qualitative analysis to achieve classification of different samples. In this study, PCA was used to analyze the NIR spectroscopy of CR and PC for classification between these two herbs.

2.4.3. *Variable selection and model construction*

Successive projections algorithm (SPA) method could search for the variable group containing the minimum redundant information from the spectra, so that the collinearity between the variables is minimized. At the same time, the number of variables used in model construction can be greatly reduced, the speed and efficiency of the model could be improved.[26]

Competitive adaptive reweighted sampling (CARS) method is a commonly used method for variable selection. "Survival of the fittest", principle of Darwin's evolution theory is the main idea. CARS method was applied to select characteristic variables, eliminate irrelevant ones and reduce the influence of collinear variables.[27] The samples were selected based on Monte Carlo sampling method and root mean square error of cross-validation (RMSECV) values were calculated for each subset of variables. The variable subset with the smallest RMSECV value is the optimal and could be used for model construction.

Partial least squares regression (PLSR) is a linear method for model construction. It has been widely used in pharmaceutical, food and other fields.[28,29] PLSR has many advantages, including eliminating collinearity, overcoming band overlapping and mutual interference.[30] According to preliminary research, PLSR models are efficient and could be used for quality control of TCM.[31,32] Therefore, PLSR were selected for the model construction in this paper. The latent variables (LVs) are the only parameters of the PLS model, and the model optimization is based on leave-one-out cross-validation. The number of LVs with the lowest RMSECV value is considered to be the most optimal.

2.4.4. *Evaluation criteria*

Correlation coefficient of calibration set (Rc) and the correlation coefficient of prediction set (Rp) were calculated, the closer the model coefficient is to 1, the better the model performance is. The relative standard error of calibration (RSEC) set and the root mean square error of calibration (RMSEC) are used as the indicators to optimize modeling parameters. Furthermore, the root mean square error of prediction (RMSEP), the relative standard errors of prediction (RSEP) are used for validation. The smaller the RSEP, the higher the accuracy of the model. When the RSEP value was less than 15%, the accuracy of the model is generally considered to meet the requirements in practical applications.[33]

2.5. *Software*

For NIR spectra data acquisition, OPUS software (version 7.0 Bruker Optics Inc., Germany) was used. PCA and the qualitative discrimination models including CARS, SPA and PLSR were conducted using MATLAB software (R2018a, MathWorks, Natick, United States).

3. Results and Discussion

3.1. *Raw spectral analysis*

The raw NIR spectra of CR and PC are shown in Fig. 1. As can be seen from the figure, the peak shapes of the two herbs are similar. The spectra were collected by diffuse reflectance, so Savitzky–Golay filter algorithm, standard normal variate transformation and multiple scatter correction methods were applied for pretreatments considering the influence of spectra collection. After comparison,

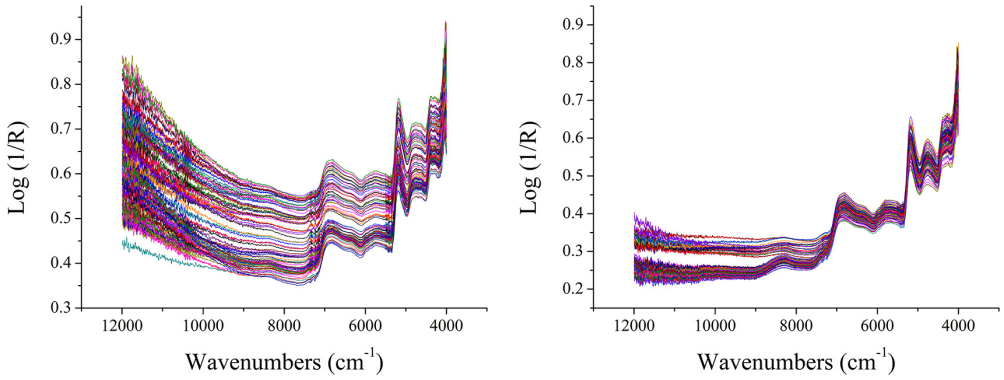


Fig. 1. NIR spectra of different samples (a) CR and (b) PC.

Table 1. Reference data by legal method.

Quality control indicators	Samples		Content range (%)	Mean	Standard deviation
Moisture	CR	94	8.627–11.69	9.924	0.80
	PC	99	8.427–11.29	10.02	0.68
Berberine	CR	94	5.314–7.664	6.508	0.62
	PC	99	6.699–13.45	9.961	1.4

the modeling results were similar and raw spectra showed better modeling performance with lower RSEC value and higher R value. This may be because some important spectral information was removed while spectral preprocessing was being performed. Therefore, in order to retain spectral information comprehensively, raw spectra were used and brought into model construction instead of pretreatment spectra. Because of the complicated information contained in and the severe band overlapping, further analysis in combination with chemometrics is needed.

3.2. *Results of the reference analysis method*

All samples were analyzed by reference method according to Chinese Pharmacopoeia (2015 edition, Vol. 1). The moisture content was detected by oven drying method, and the berberine content was detected by HPLC method. The analytical results of the CR and PC samples were summarized in Table 1. For different batches of herbs, the content of moisture and berberine varied greatly, indicating the necessity of quality control of TCM. Among them, the berberine content was determined by different HPLC methods. The representative chromatograms were shown in Figs. 2 and 3.

3.3. *Samples division*

The samples were divided into calibration set and prediction set according to a ratio of 2:1 using the sample set portioning based on joint x–y distance (SPXY) method. The calibration sets of CR and PC datasets were used for model construction in specific models and their combination were used for global models. The prediction sets were independent samples and only used for model validation.

3.4. *Principal component analysis*

Principal component analysis (PCA) was performed on CR and PC datasets. When PCA was applied to the spectra data, the whole dataset was used. The central purpose of PCA is to reduce the dimensionality of data by transforming the original variables and making a few new variables, which are linear combinations of the original variables, called the principal components. At the same time, these components should express the data characteristics of the original variables as much as possible and make sure no information is lost. The variance interpreted by PC1 and PC2 were 97.34% and 1.70%, respectively, which means that 97.34% and 1.70% information of the original data could be expressed by the first and second principle components. The accumulated variance contribution rate was up to 99.04% for the first two PCs, which means more than 99% chemical information of NIR spectroscopy was represented. The score plot for all the samples according to the first two PCs is shown

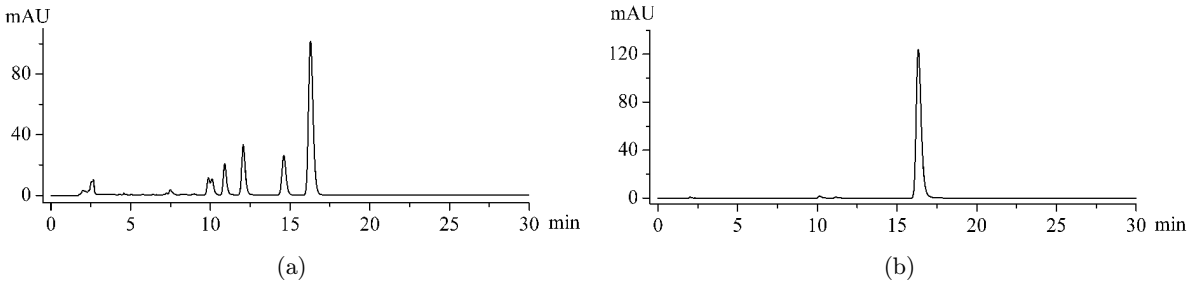


Fig. 2. HPLC chromatograms of berberine in CR. (a) Extract of CR sample and (b) Standard solution.

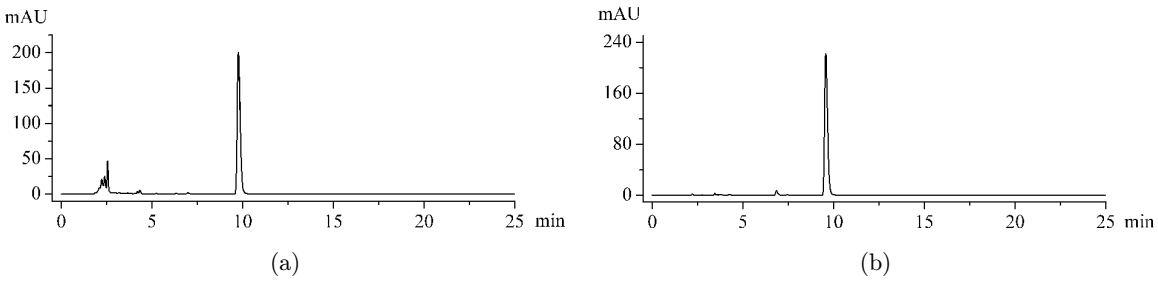


Fig. 3. HPLC chromatograms of berberine in PC. (a) Extract of PC sample and (b) Standard solution.

in Fig. 4. The boundaries of the two categories were very clear in space. With the zero on the abscissa as the boundary, CR and PC were well separated. As a result, NIRs combined with PCA can be used as a good tool for origin classification of CR and PC.

3.5. *Variable selection*

3.5.1. *Competitive adaptive reweighted sampling method*

When using CARS method for wavenumbers selection, the process was shown in Fig. 5. The trend of the RMSECV value reflected that as the number of sampling runs increased and the redundant information was eliminated, the RMSECV value became smaller. Continuing sampling after the minimum value was reached, the RMSECV value increased, indicating that the indicator information was eliminated along with the redundant information, so the subset of variables with the smallest RMSECV value was the most optimal. Based on CARS method, 125 and 117 variables were selected for moisture from CR and PC dataset, respectively. The characteristic variables of berberine were 95 and 125, respectively.

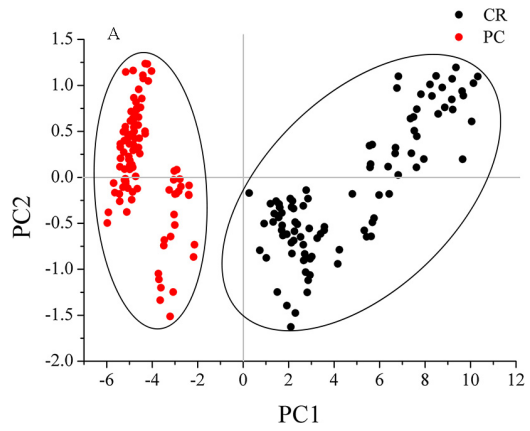


Fig. 4. Score cluster plot with top two principal components (PCs)

3.5.2. *Successive projections algorithm method*

As a classic forward period wavenumber selection method, the principle of filtering variables by the SPA method is to optimize the variable group with the smallest root mean square error (RMSE). Figure 6 shows the variation of the RMSE value and the number of key variables in the process of optimizing by SPA method. The results show that with the increase of the number of variables, the trend of RMSE value decreased and gradually flattened. 98 variables were screened for berberine from CR and PC datasets respectively and 89 variables were for moisture.

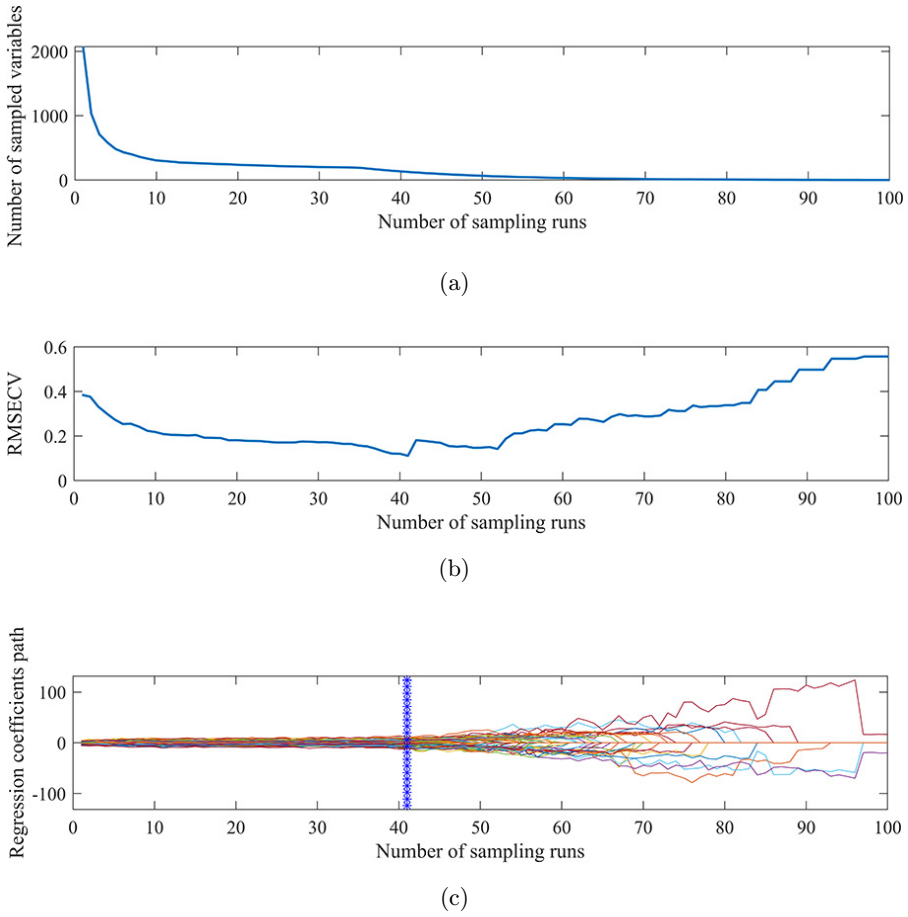


Fig. 5. CARS variable selection on spectra data of CR dataset.

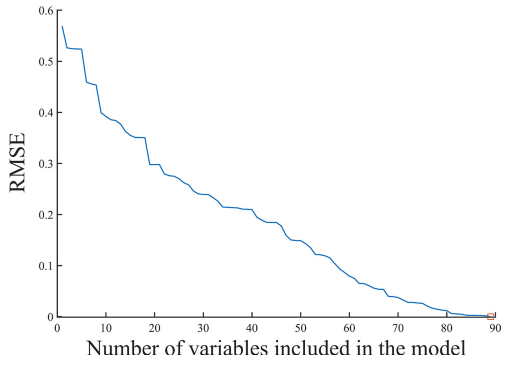


Fig. 6. SPA variable selection on spectra data of CR dataset.

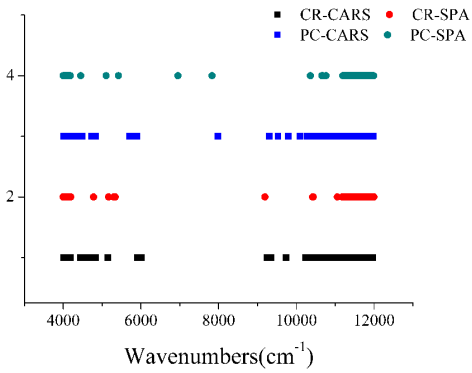


Fig. 7. Variables selected based on CARS and SPA method for berberine.

In order to explore the consistency of characteristic variables in two herbs, the characteristic variables of berberine selected from CR and PC datasets by CARS and SPA method were shown in Fig. 7. The distribution of characteristic variables selected by two methods were consistent and seemed to be similar between two herbs, which also provided conditions for global model of the common component. However, minor differences still existed.

The absorbance of 6999 cm^{-1} was selected by SPA method in PC and was known as deformation of O–H in water, which may lead to the poor prediction of SPA-PLS model.

Berberine hydrochloride is a quaternary ammonium isoquinoline alkaloids, containing benzene ring and methoxy group.[34] The most important absorption peak in NIRs comes from C–H and O–H.

Table 2. Modeling results for single species model of CR.

Quality control indicators	LVs		Calibration set				Prediction set		
			Rc	RMSEC	RMSECV	RSEC%	*Rp*	RMSEP	RSEP%
Moisture	RAW	6	0.9301	0.2500	0.4510	2.51	0.8327	0.3366	3.40
	SPA	6	0.9269	0.2554	0.3978	2.56	0.6808	0.4453	4.50
	CARS	6	0.9455	0.2217	0.3617	2.22	0.8653	0.3017	3.08
Berberine	RAW	6	0.8494	0.3068	0.4687	4.53	0.5314	0.3967	6.16
	SPA	5	0.8042	0.3349	0.5111	5.14	0.6176	0.3495	5.34
	CARS	9	0.9765	0.1420	0.3211	1.85	0.9174	0.1954	3.09

Notes: LVs: latent variables; *Rc*: correlation coefficient of the calibration set; RMSEC: root mean square error of calibration set; RMSECV: root mean square error of cross-validation; RSEC: relative standard error of calibration set; *Rp*: correlation coefficient of the prediction set; RMSEP: root mean square error of prediction set; RSEP: relative standard error of prediction set.

The characteristic variables selected by the CARS method also corresponds to the peak in the original spectrum, of which 6250–5555 cm^{-1} is the first-order frequency multiplication of the hydrocarbon C–H vibration, and the first combination frequency appears between 5000 and 4160 cm^{-1} which is strong and obvious. The absorption peak around 4050 cm^{-1} is the combined frequency of C–H stretching vibration and C–H bending vibration on the benzene ring. The series of absorption peaks at 4660 cm^{-1} are the combined frequency of C–C stretching vibration and C–H stretching vibration, and its combination appears at 9259 cm^{-1}.

3.6. *Construction of single species models*

3.6.1. *Single species model of CR*

For the moisture and berberine content of CR samples, single species models were first constructed, and the prediction results were shown in Table 2. It can be seen from the results that the model based on variable selection has improvement in prediction performance than the raw spectra model. Among them, the CARS-PLS model is the most suitable for the model construction of the CR datasets. For the optimal PLSR models of moisture and berberine content, the *Rc* values are all above 0.94 and the RSEP value are under 4%, indicating the accuracy of NIRs. The correlation between the predicted content by NIRs and the reference value analyzed by reference method of CR samples was shown in Fig. 8.

3.6.2. *Single species model of PC*

The prediction results of PC datasets were shown in Table 3. Similarly, the content of quality control indicators in PC was also well predicted by NIRs combined with chemometrics. According to the model performance, variables selected by SPA

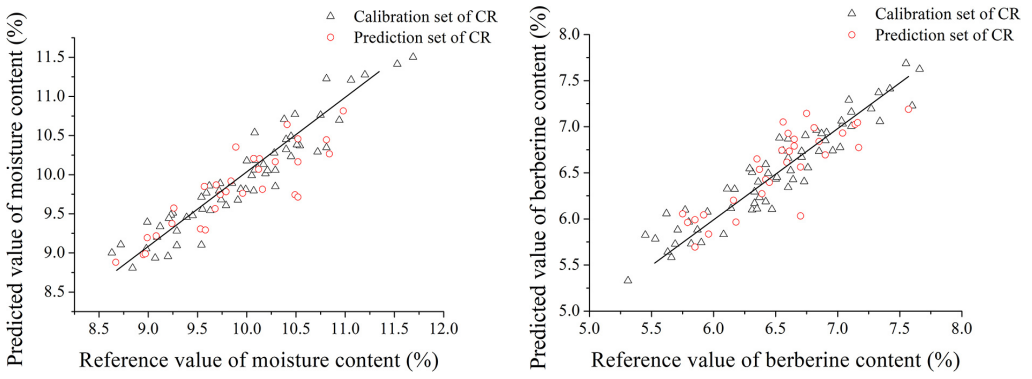


Fig. 8. Correlation diagram between NIRS predicted values and reference values of CR model. (a) Moisture content and (b) Berberine content.

Table 3. Modeling results for single species model of PC.

Quality control indicators	LVs		Calibration set				Prediction set		
			Rc	RMSEC	RMSECV	RSEC%	*Rp*	RMSEP	RSEP%
Moisture	RAW	4	0.8044	0.3416	0.4222	3.41	0.7149	0.2749	3.10
	SPA	2	0.3705	0.5342	0.5748	5.33	0.3948	0.4101	4.08
	CARS	6	0.9534	0.1735	0.2721	1.73	0.8575	0.2297	2.28
Berberine	RAW	5	0.8882	0.5732	0.8250	5.71	0.8454	0.7591	7.55
	SPA	6	0.8514	0.6543	1.1007	6.52	0.7551	0.9316	9.26
	CARS	7	0.9632	0.3354	0.5480	3.34	0.9459	0.4609	4.58

Notes: LVs: latent variables; *Rc*: correlation coefficient of the calibration set; RMSEC: root mean square error of calibration set; RMSECV: root mean square error of cross-validation; RSEC: relative standard error of calibration set; *Rp*: correlation coefficient of the prediction set; RMSEP: root mean square error of prediction set; RSEP: relative standard error of prediction set.

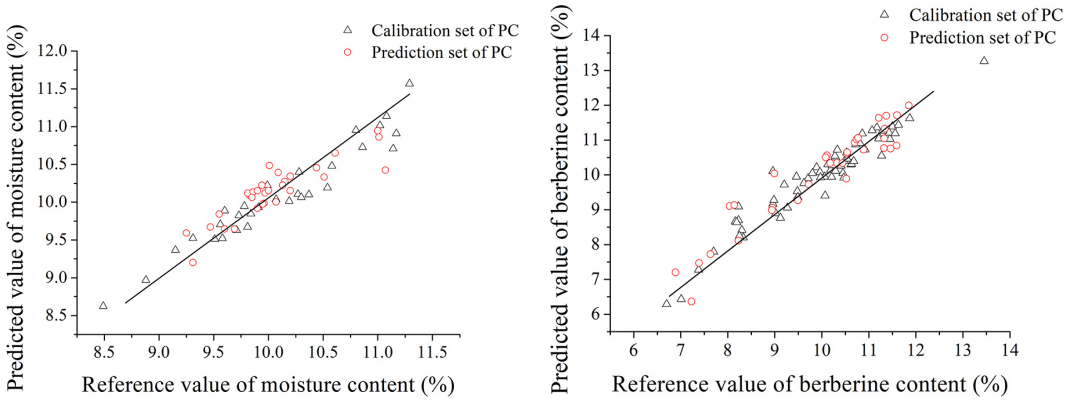


Fig. 9. Correlation diagram between NIRS predicted values and reference values of PC model. (a) Moisture content and (b) Berberine content.

method are not suitable for PC dataset, which resulted in poor prediction. SPA method aimed to minimize the collinearity of the variables, which may result in insufficient effective information extracted from the variables. The data was over-compressed during the process, and the main chemical information was removed, thus affecting the model performance. The correlation between the predicted content by NIRs and the reference value analyzed by reference method of PC samples was shown in Fig. 9.

3.6.3. *Construction of global models*

The global models for moisture and berberine content used the combination of two calibration sets from CR and PC datasets. Compared with single species models, the samples in calibration set were combined and the prediction sets were consistent. The modeling performance were shown in Table 4. Compared with the raw spectra model, the CARS-PLS model was simplified and the prediction accuracy was improved, with higher R values and lower RMSEC values, which was considered as the most suitable model for prediction of moisture and berberine content. Therefore, the global models based on CARS-PLS method were applied in prediction of moisture and berberine content in CR and PC, respectively. The prediction results were shown in Table 5.

In order to evaluate the global model accurately, the parameters of the prediction set were given more attention, such as RSEP values. The lower RSEP values correspond to better prediction performance. According to the parameters, the prediction results of the global model were accurate and robust, meeting the requirements of TCM in process analysis. The correlation between the

Table 4. Modeling results of global models.

Quality control indicators	LVs		Calibration set			
			Rc	RMSEC	RMSECV	RSEC%
Moisture	RAW	7	0.8413	0.3314	0.4852	3.31
	SPA	6	0.7174	0.4395	0.5474	4.39
	CARS	9	0.9528	0.1860	0.3119	1.86
Berberine	RAW	7	0.9570	0.5723	0.8233	6.73
	SPA	6	0.9246	0.7578	0.9768	8.91
	CARS	10	0.9880	0.3051	0.4948	3.59

Notes: LVs: latent variables; *Rc*: correlation coefficient of the calibration set; RMSEC: root mean square error of calibration set; RMSECV: root mean square error of cross-validation; RSEC: relative standard error of calibration set.

Table 5. Prediction results of CARS-PLS global models.

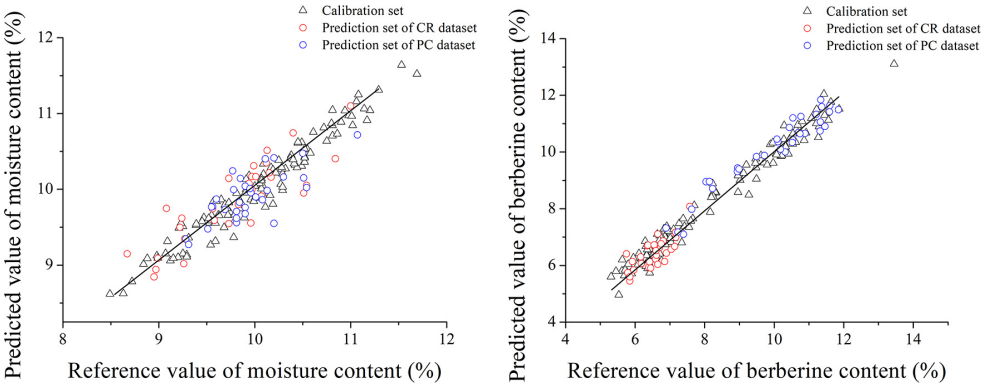
Quality control indicators	Prediction set of CR			Prediction set of PC		
	Rp	RMSEP	RSEP%	*Rp*	RMSEP	RSEP%
Moisture	0.8541	0.2982	3.04	0.7567	0.2524	2.53
Berberine	0.9596	0.4192	5.41	0.9823	0.3782	3.97

Notes: *Rp*: correlation coefficient of the prediction set; RMSEP: root mean square error of prediction set; RSEP: relative standard error of prediction set.

predicted content by NIRs and the reference value analyzed by HPLC method was shown in Fig. 10.

The content results showed that the moisture content of two herbs was very similar, and the prediction accuracy of the global model is satisfactory. For the same prediction set, the RSEP values of specific and global models for CR dataset are 3.08% and 3.04%, respectively, slightly decreased. The RSEP values of moisture prediction set by specific and global models in PC dataset are 2.28% and 2.54%, respectively. For CR dataset, global model performed better than specific model in prediction accuracy. The parameter of global model for PC showed a little bit worse, but the value of 2.54% is also very satisfactory in rapid analysis and meet the requirement of process analysis. Furthermore, the number of samples used for global models were increased significantly compared with single species ones, which means the distribution of samples are more even and the representation of the calibration set could be enhanced.

For berberine content, the content range of berberine in CR and PC were 5.314–7.664% and 6.699–13.45%, respectively, which means only a small part was overlapped. The global models of berberine achieved wider cover range from 5.314% to 13.45%, expanding the scope of application significantly. Compared with single specific models, the prediction error of berberine content in PC was decreased obviously from 4.58% to 3.97%. The better accuracy may attribute to larger number of samples and wider modeling range compared with the single species models. The RSEP value of CR was increased from 3.09% to 5.41%. The berberine content in CR was low compared with other indicators, the prediction accuracy is relatively worse. However, the global model showed good prediction accuracy and stability in general. The common components and similar substances contained in CR and PC provide a good basis for the establishment of the



(a) (b)

Fig. 10. Correlation diagram between NIRS predicted values and reference values of global model. (a) Moisture content and (b) Berberine content.

global model, and the experimental results also prove the feasibility of it.

Literature research showed that CR and PC have similar biological efficacy in clearing heat and detoxifying, and the common ingredients are also alkaloids. The spectra combination of two herbs provide a wealth of structural and material information because of the different chemical environments of CR and PC. According to the prediction results, the global models are more suitable for rapid detection of moisture content in CR and berberine content in PC dataset, the prediction accuracy were improved compared with single specific models. In general, the global models were constructed successfully. Models constructed showed excellent globality and compatibility, performing efficient and accurate prediction. For both herbs, the concentration range was expanded, exceeding the unusual samples that may be encountered in further prediction, for which the disadvantage of the narrow and less representative calibration was avoided. Besides, moderate samples outside the concentration range of single species could contribute to wider application and efficient update. What's more, with large number of samples came from different resources, the difficulty of collection was alleviated and the quantitative function between spectral variables and content could be more reliable. The global models constructed could be applied to the rapid quality control of CR and PC, especially for the combined use in the production of TCM.

4. Conclusion

In conclusion, the quantitative analysis of CR and PC by global models were explored. The classification of two herbs were realized clearly by PCA method, and the distribution of characteristic variables between two herbs was consistent. The established specific models and global models were accurate and meet the requirement in quantitative analysis. Furthermore, the global model showed better robustness and accuracy in prediction of moisture content for CR and berberine content for PC, compared with single species models. The number of samples and the cover range were expanded, which enhanced the robustness and reliability of the global models. The prediction efficiency was also significantly improved in practical application. Therefore, the global models constructed could be applied to the rapid quality control of CR and PC.

Conflict of Interest

The authors declare no competing financial interest.

Acknowledgment

This work was supported by National Major Scientific and Technological Special Project for "Significant New Drugs Development" (2018ZX09201010).

References

1. K. He, S. M. Kou, Z. Y. Zou, Y. R. Hu, M. Feng, B. Han, X. G. Li, X. L. Ye, "Hypolipidemic effects of alkaloids from *Rhizoma Coptidis* in diet-induced hyperlipidemic hamsters," *Planta Med.* **82**, 690–697 (2016).
2. C. L. Li, J. H. Xie, X. Y. Chen, Z. Z. Mo, W. Wu, Y. Liang, Z. Q. Su, Q. Li, Y. C. Li, Z. R. Su, X. B. Yang, "Comparison of helicobacter pylori urease inhibition by *Rhizoma Coptidis*, *Cortex Phellodendri* and Berberine: Mechanisms of interaction with the sulfhydryl group," *Planta Med.* **82**, 305–311 (2016).
3. R. Mohammadinejad, Z. Ahmadi, S. Tavakol, M. Ashrafizadeh, "Berberine as a potential autophagy modulator," *J. Cell. Physiol.* **234**, 14914–14926 (2019).
4. Chinese Pharmacopoeia Commission, *Pharmacopoeia of the People's Republic of China, Part 1*, Vol. 303, P. 305, Chemical Industry Press, Beijing (2015).
5. F. C. Meng, Z. F. Wu, Z. Q. Yin, L. G. Lin, R. B. Wang, Q. W. Zhang, "*Coptidis Rhizoma* and its main bioactive components: Recent advances in chemical investigation, quality evaluation and pharmacological activity," *Chin. Med.* **13**, 13 (2018).
6. J. Lu, J. S. Wang, L. Y. Kong, "Anti-inflammatory effects of Huang-Lian-Jie-Du decoction, its two fractions and four typical compounds," *J. Ethnopharmacol.* **134**, 911–918 (2011).
7. L. Cecillon, B. G. Barthes, C. Gomez, D. Ertlen, V. Genot, M. Hedde, A. Stevens, J. J. Brun, "Assessment and monitoring of soil quality using near-infrared reflectance spectroscopy (NIRS)," *Eur. J. Soil Sci.* **60**, 770–784 (2009).
8. Y. J. Wu, Y. Jin, Y. R. Li, D. Sun, X. S. Liu, Y. Chen, "NIR spectroscopy as a process analytical technology (PAT) tool for on-line and real-time monitoring of an extraction process," *Vib. Spectrosc.* **58**, 109–118 (2012).

9. L. S. Magwaza, U. L. Opara, H. Nieuwoudt, P. J. R. Cronje, W. Saeys, B. Nicolai, "NIR spectroscopy applications for internal and external quality analysis of citrus fruit-a review," *Food Bioprocess Technol.* **5**, 425–444 (2012).
10. L. L. Gao, T. T. Li, L. Li, Z. Li, L. Nie, H. Y. Zhou, Z. Y. Sun, S. Y. Ye, R. C. Liu, H. C. Zang, "Rapid determination of lumbrokinase potency in the earthworm extract intermediate by near-infrared spectroscopy," *Chemom. Intell. Lab. Syst.* **185**, 59–64 (2019).
11. Y. Roggo, P. Chalus, L. Maurer, C. Lema-Martinez, A. Edmond, N. Jent, "A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies," *J. Pharm. Biomed. Anal.* **44**, 683–700 (2007).
12. X. D. Lai, J. N. Li, X. Gong, X. J. Lin, G. Q. Tang, R. Li, C. C. Jia, D. Wang, S. G. Ji, "Rapid simultaneous determination of andrographolides in andrographis paniculata by near-infrared spectroscopy," *Anal. Lett.* **51**, 2745–2760 (2018).
13. W. Liu, Z. Z. Wang, J. P. Qing, H. J. Li, W. Xiao, "Classification and quantification analysis of peach kernel from different origins with near-infrared diffuse reflection spectroscopy," *Pharmacogn. Mag.* **10**, 441–448 (2014).
14. X. Zhang, W. Li, B. Yin, W. Z. Chen, D. P. Kelly, X. X. Wang, K. Y. Zheng, Y. P. Du, "Improvement of near infrared spectroscopic (NIRS) analysis of caffeine in roasted Arabica coffee by variable selection method of stability competitive adaptive reweighted sampling (SCARS)," *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **114**, 350–356 (2013).
15. J. T. Xue, Y. L. Shi, L. M. Ye, Q. W. Yang, C. Y. Li, X. Y. Chen, Y. Jing, "Near-infrared spectroscopy for rapid and simultaneous determination of five main active components in rhubarb of different geographical origins and processing," *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **205**, 419–427 (2018).
16. Y. Yang, Y. J. Wu, W. L. Li, X. S. Liu, J. Y. Zheng, W. T. Zhang, Y. Chen, "Determination of geographical origin and icariin content of Herba Epimedii using near infrared spectroscopy and chemometrics," *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **191**, 233–240 (2018).
17. J. Y. Wang, T. T. Li, H. L. Yang, T. Hu, L. Nie, F. Wang, M. Alcala, H. C. Zang, "Geographical origin discrimination and polysaccharides quantitative analysis of Radix codonopsis with micro near-infrared spectrometer engine," *J. Innovative Opt. Health Sci.* **11**, 1 (2018).
18. T. Casian, S. Iurian, A. Gavan, C. Revnic, S. Porav, A. Porfire, L. Vlase, I. Tomuta, "Near Infra-Red spectroscopy for content uniformity of powder blends - Focus on calibration set development, orthogonality transfer and robustness testing," *Talanta*, **188**, 404–416 (2018).
19. I. Torres, D. Perez-Marin, M. J. De la Haba, M. T. Sanchez, "Developing universal models for the prediction of physical quality in citrus fruits analysed on-tree using portable NIRS sensors," *Biosyst. Eng.* **153**, 140–148 (2017).
20. S. Anderson, "Determination of fat, moisture, and protein in meat and meat products by using the FOSS FoodScanTm near-infrared spectrophotometer with FOSS artificial neural network calibration model and associated database: Collaborative study," *J. AOAC Int.* **90**, 1073–1083 (2007).
21. X. L. Wang, Q. A. Fu, J. F. Sheng, X. Yang, J. Z. Jia, W. Du, "Construction of a universal quantitative model for ibuprofen sustained-release capsules from different manufacturers using near-infrared diffuse reflection spectroscopy," *Vib. Spectrosc.* **53**, 214–217 (2010).
22. X. B. Zhang, Y. C. Feng, C. Q. Hu, "Feasibility and extension of universal quantitative models for moisture content determination in beta-lactam powder injections by near-infrared spectroscopy," *Anal. Chim. Acta.* **630**, 131–140 (2008).
23. Z. Y. Sun, J. Y. Wang, L. Nie, L. Li, D. W. Cao, J. J. Fan, H. Y. Wang, R. C. Liu, Y. R. Zhang, H. C. Zang, "Calibration transfer of near infrared spectrometers for the assessment of plasma ethanol precipitation process," *Chemom. Intell. Lab. Syst.* **181**, 64–71 (2018).
24. Y. He, X. L. Li, X. F. Deng, "Discrimination of varieties of tea using near infrared spectroscopy by principal component analysis and BP model," *J. Food Eng.* **79**, 1238–1242 (2007).
25. S. Vaidyanathan, S. A. Arnold, L. Matheson, P. Mohan, B. McNeil, L. M. Harvey, "Assessment of near-infrared spectral information for rapid monitoring of bioprocess quality," *Biotechnol. Bioeng.* **74**, 376–388 (2001).
26. J. F. G. Martin, "Optical path length and wavelength selection using Vis/NIR spectroscopy for olive oil's free acidity determination," *Int. J. Food Sci. Technol.* **50**, 1461–1467 (2015).
27. Q. F. Sun, Z. Y. Sun, F. Wang, L. Li, R. H. Liu, L. Nie, J. Y. Wang, M. Y. Wang, H. C. Zang, "pH value monitoring during human albumin purification with near infrared spectroscopy and chemometrics," *J. Innovative Opt. Health Sci.* **11**, 1850009 (2018).
28. M. R. Akkaya, "Prediction of fatty acid composition of sunflower seeds by near-infrared reflectance spectroscopy," *J. Food Sci. Technol.* **55**, 2318–2325 (2018).
29. F. Y. H. Kutsanedzie, L. Hao, S. Yan, Q. Ouyang, Q. Chen, "Near infrared chemo-responsive dye

intermediaries spectra-based in-situ quantification of volatile organic compounds," *Sensors Actuators B Chem.* **254**, 597–602 (2018).
30. X. Z. Yu, Y. Gao, Z. Zhao, J. M. Gao, "Rapid determination of amino acids in Chinese wolfberry (Lycium bararum L.) fruit by using Fourier transform infrared spectroscopy and partial least square regression," *Food Anal. Methods* **10**, 2436–2443 (2017).
31. M. Ferreiro-Gonzalez, E. Espada-Bellido, L. Guillen-Cueto, M. Palma, C. G. Barroso, G. F. Barbero, "Rapid quantification of honey adulteration by visible-near infrared spectroscopy combined with chemometrics". *Talanta* **188**, 288–292 (2018).
32. M. Campos, G. Antolin, L. Debán, R. Pardo, "Assessing the influence of temperature on NIRS prediction models for the determination of sodium content in dry-cured ham slices," *Food Chem.* **257**, 237–242 (2018).
33. Y. Jin, Z. Wu, X. Liu, Y. Wu, "Near infrared spectroscopy in combination with chemometrics as a process analytical technology (PAT) tool for on-line quantitative monitoring of alcohol precipitation," *J. Pharm. Biomed. Anal.* **77**, 32–39 (2013).
34. J. Tang, Y. B. Feng, S. Tsao, N. Wang, R. Curtain, Y. W. Wang, "Berberine and Coptidis Rhizoma as novel antineoplastic agents: A review of traditional use and biomedical investigations," *J. Ethnopharmacol.* **126**, 5–17 (2009).