

### **Research** Article

# Rapid Quality Assessment of *Polygoni Multiflori* Radix Based on Near-Infrared Spectroscopy

## Bin Jia,<sup>1,2</sup> Ziying Mai,<sup>2</sup> Chaoqun Xiang,<sup>2</sup> Qiwen Chen,<sup>2</sup> Min Cheng,<sup>3</sup> Longkai Zhang,<sup>3</sup> and Xue Xiao <sup>1,2</sup>

<sup>1</sup>NMPA Key Laboratory for Rapid Testing Technology of Drugs, Guangdong Institute for Drug Control, Guangzhou 510663, China

<sup>2</sup>Guangdong Metabolic Diseases Research Center of Integrated Chinese and Western Medicine (Institute of Traditional Chinese Medicine), Guangdong Pharmaceutical University, Guangzhou 510006, China

<sup>3</sup>Hong Zheng Dao (China) Traditional Chinese Medicine Research Company Ltd, Guangzhou 510665, China

Correspondence should be addressed to Xue Xiao; xuexiao@gdpu.edu.cn

Received 8 February 2024; Revised 9 March 2024; Accepted 13 March 2024; Published 20 March 2024

Academic Editor: Michele Fedel

Copyright © 2024 Bin Jia et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The precise and prompt determination of quality control indicators such as moisture, stilbene glycosides, and anthraquinone glycosides is crucial in assessing the quality of Polygoni Multiflori Radix. Near-infrared spectroscopy is a nondestructive analytical technique that offers a more desirable approach than traditional methods for assessing content levels. In this study, various spectral preprocessing techniques were used to preprocess the raw spectral data. The spectral data were correlated with the determination of three-component contents using the partial least squares regression (PLSR) method. Then different algorithms, such as competitive adaptive weighted sampling (CARS), Monte Carlo uninformative variable elimination (MCUVE), and random frog hopping (RF), were used for model simplification and feature selection. The data suggest that the first-order deconvolution derivative (1<sup>st</sup> Dev.) processing of the spectral data is superior to other methods in all three model evaluation metrics. The PLSR model for moisture, stilbene glycosides, and anthraquinone glycosides produced the calibration coefficient of determination  $(R^2_C)$  of 0.82, 0.52, and 0.58, the root mean square error of cross validation (RMSE<sub>CV</sub>) of 0.91%, 0.77%, and 0.69%, the prediction coefficient of determination  $(R^2_P)$  of 0.72, 0.28, and 0.54, the root mean square error of prediction (RMSE<sub>P</sub>) of 0.65%, 0.81%, and 0.75%, and relative percentage differences (RPDs) of 1.7, 1.0, and 0.8. After optimizing the model using CARS,  $R^2_C$  increased by 0.15%, 0.41%, and 0.34%, RMSE<sub>CV</sub> decreased by 0.53%, 0.32%, and 0.24%,  $R^2_P$  increased by 0.21%, 0.63%, and 0.35%, RMSE<sub>P</sub> decreased by 0.36%, 0.41%, and 0.31%, and RPD increased by 1.1, 0.9, and 0.6, significantly improving the predictive capacity of the model. This research provides a feasible method for rapid compliance testing of Polygoni Multiflori Radix. To further improve the model's performance and applicability, it is necessary to continuously expand the sample set with different varieties and locations for wide variation.

#### 1. Introduction

*Polygoni Multiflori* Radix, the dried tuberous root of the plant *Polygonum multiflorum* Thunb. of the family Polygonaceae, has the effects of removing toxin for eliminating carbuncles, expulsing pathogen for preventing malaria, and moistening dryness for relaxing the intestines and hypolipidemic. Modern pharmacological studies have shown that it also has effects such as antiatherosclerosis [1], antioxidation [2], antiosteoporosis [3], antitumor [1],

hypoglycemic [4], cognitive enhancement [5], and antimicrobial activities [2]. The efficacy of traditional Chinese medicine (TCM) is closely related to its quality, and moisture content is one of the key indicators to evaluate the quality of TCM. The excessive moisture content in TCM can affect the concentration of active ingredients and even cause problems such as insect damage and mold. The concentration of active ingredients similarly affects the quality of the herbs. The active ingredients of *Polygoni Multiflori* Radix include stilbene glycosides and anthraquinone glycosides, which possess various beneficial effects such as antiaging [6], hypolipidemic [4], antiosteoporosis [7], and antiinflammatory [8] properties. The 2020 edition of the Pharmacopoeia of the People's Republic of China (Chinese Pharmacopoeia) specifically stipulates that the content of stilbene glucosides in Polygoni Multiflori Radix. It should not be less than 1.0%, and the glycoside content of anthraquinones should not be less than 0.05% [9]. Among the methods used to evaluate the content of stilbene glucosides and anthraquinones glycosides in Polygoni Multiflori Radix, high-performance liquid chromatography (HPLC) is considered complex, time-consuming, and labor-intensive, which makes it difficult to keep up with the modernization of TCM in batch production and routine production of tablets. Similarly, traditional moisture determination methods can also destroy samples and are also challenging to apply for online monitoring in the production of TCM in the industry [10]. Near-infrared spectroscopy (NIRS) is the most widely used process analytical technique in the pharmaceutical industry, and it has been increasingly emphasized in the rapid detection of product quality and online control of processes in TCM [11]. Compared to traditional analytical techniques, the NIRS analysis technology has numerous advantages. It enables measurement of multiple performance parameters in just a few minutes through a single measurement of the sample without the need for sample preparation. The analysis process does not consume any additional materials and is cost-effective. Currently, NIRS technology is undergoing significant advancements and widespread adoption in various industries including petrochemical [12], agricultural [13], food [14], and medical sectors [15].

Numerous studies have demonstrated the successful utilization of NIRS technology for rapid analysis of the chemical constituent of TCM. Jintao et al. used NIRS to determine the content of five major active components in rhubarb. In addition, they used this method to elucidate the influences of the source and processing methods, to identify counterfeit or adulterated rhubarb [16]. NIRS technology can also be used for the quantification of five alkaloids in different parts of Coptidis rhizome. This provides valuable technical support for the comprehensive development of Coptidis rhizome plant resources [17]. It can also be used in conjunction with chemometric methods to rapidly determine the puerarin content during the percolation and concentration process of Puerariae Lobatae Radix, with RMSE<sub>P</sub> of 0.0396 mg/ml and 0.0365 mg/ml and  $R^2$  of 97.79% and 98.47%, respectively [18]. The NIRS technique can be used for the rapid determination of two chemical indicators, 5-HMF and 420 nm absorbance values, with the  $RMSE_P$  less than 0.03 and the  $R^2$  greater than 0.99 in the TCM process. It was concluded that NIRs are expected to replace current subjective colour judgments and tedious HPLC or UV/Vis methods and are applicable to rapid analysis and quality control of TCM in the process of industrial production [19]. Using NIRS, Zhu et al. established a quantitative analysis model to determine the content of seven alkaloids in crude and processed products of Corydalis Rhizoma. In addition, a qualitative identification model was also developed to

distinguish between different processed products [20]. However, there is limited research on the use of nearinfrared spectroscopy in the study of *Polygoni Multiflori* Radix. Therefore, the main objective of this study is to attempt to establish a method for the rapid quantitative measurement of the components in *Polygoni Multiflori* Radix using NIRS, in order to provide technical support for the rapid detection and control of its quality.

#### 2. Materials and Methods

2.1. Experimental Materials. 149 batches of Polygoni Multiflori Radix were supplied by HongZhengDao (China) Traditional Chinese Medicine Research Company Ltd. We grond each batch of samples separately, sieved them through a 60 mesh screen, and subsequently sealed and stored them individually.

2.2. Near-Infrared Spectral Acquisition. About 3 grams of each sample was taken into a sample cup (injection vials with a mirror base, 520 × 2200/120 mm, Bruker, Germany) and measured. The laboratory temperature was maintained at  $23 \pm 2$ °C, with a relative humidity ranging from 45% to 60%. The MPA II type FT-NIR multipurpose analyzer (Bruker, Germany) was utilized for NIR spectroscopic analysis. The instrument adopts the integrating sphere diffuse reflectance sampling method. The spectral scanning range was from 11550 to  $3950 \text{ cm}^{-1}$ , with a sampling interval of  $10^7 \text{ cm}^{-1}$ . Each sample was scanned 32 times to generate a single spectrum. Each sample was measured six times to obtain an average spectrum for experimentation. This process resulted in 149 average spectra, each consisting of 2666 wavelength points. The parameters of the background compensation process should be the same as those of the sample measuring.

2.3. Determination of the reference Values. The moisture content of Polygoni Multiflori Radix was determined by drying according to the Chinese Pharmacopoeia, Edition 2020 (General Rule 0832). Take 5g of the sample and spread flat in a glass vial with lid and the initial mass was recorded. Open the lid and dry at 100°C-105°C for 5 h, close the lid and cool to room temperature and record the cooling mass. Afterwards, the drying time was set to 1 h and the cooling mass was recorded each time. The moisture content in the samples was calculated by stopping the drying until the difference in cooling mass after two consecutive dryings did not exceed 5mg. According to the 2020 edition of the Chinese Pharmacopoeia, Part I, the content of stilbene glycosides, total anthraquinone, and free anthraquinone in the samples were determined by the HPLC system (Alliance e2695, with2998 PDA Detector, Waters, USA). Dissolve 0.2 g of the sample in 25 ml of dilute ethanol (AR  $\ge$  99.7%, Tianjin ZhiYuan Heagent Co., Ltd., China) and reflux for 30 min. 10 µl of the solution was extracted, and the stilbene glycoside content of the sample was measured at a wavelength of 320 nm using acetonitrile (HPLC  $\geq$  99.9%, Merck, Germany) and pure water (Hangzhou Wahaha Group Co., Ltd., China) (25:75) as the mobile phase. 0.5 g of the sample was

dissolved in 25 ml of methanol (AR  $\geq$  99.7%, Tianjin Zhiyuan Reagent Co., Ltd, China) and refluxed for 60 min.  $10 \,\mu$ l of the solution was extracted, and the free anthraquinone content of the sample was measured at a wavelength of 254 nm using methanol (HPLC  $\geq$  99.9%, Merck, Germany) and 0.1% phosphoric acid (HPLC≥98%, Tianjin Kemiou Chemical Reagent Co., Ltd., China) (80:20) as the mobile phase. 0.5 g of the sample was dissolved in 25 ml of methanol and refluxed for 60 min. After evaporation of the solvent, redissolve in 20 ml of hydrochloric acid (36.0%≤AR≤38.0%, Guangzhou Chemical Reagent Factory, China) at 8% by mass. Use ultrasound to ensure complete dissolution. Then add an equal volume of chloroform (AR≥99.0%, Guangzhou Chemical Reagent Factory, China). The mixture was heated in a water bath for 60 min and extracted with chloroform. The bottom layer was evaporated and dissolved in 10 ml of methanol.  $10 \,\mu$ l of the solution was extracted, and the total anthraquinone content in the sample was measured at a wavelength of 254 nm using methanol and 0.1% phosphoric acid (80:20) as the mobile phase. The anthraquinone glycoside content was obtained by subtracting the free anthraquinone content from the total anthraquinone content.

2.4. Investigation of the NIRS Model. Bruker NIR spectral analysis software OPUS 8.5 (Bruker, Germany) and MATLAB R2020b (MathWorks, USA) were used to analyze the processing of spectra. Original data were processed using the methods included in the software to eliminate outliers. Subsequently, the authentic data selected were divided into a calibration set and an external prediction set in a 3:1 ratio using the sample set partitioning algorithm based on the Kennard-Stone algorithm. During the process of spectral acquisition, apart from the information related to the components in the measured sample, there are also noises originating from the measurement environment and instrument, due to the influence of the detection instrument. Background interference, noise, baselines, and stray light interference in NIRS signals can affect the quality of models [21]. Therefore, preprocessing of the spectra is necessary. The purpose of this preprocessing is to improve the predictive power and stability of the model. In this study, preprocessing techniques were performed on the spectra, including Savitzky-Golay smoothing (SG), standard normal variant (SNV), first-order deconvolutional derivative (1<sup>st</sup> Dev, number of points: 11), and second-order deconvolutional derivative (2<sup>nd</sup> Dev, number of Points: 11) to preprocess the NIR spectra [22]. Excessive redundancy and irrelevant information tend to mask relevant variables, increase complexity, and reduce model accuracy, leading to overfitting or underfitting problems commonly found in spectral models. Variable selection is an important method to address these challenges. It involves analysing nearinfrared spectra, determining the optimal wavelength range, and compressing the spectral data, all of which help to simplify the model, reduce computational overhead, and improve model quality. To further enhance computational efficiency, we employed three feature selection algorithms: competitive adaptive reweighted sampling (CARS), Monte

Carlo uninformative variable elimination (MCUVE), and random frog leaping (RF). The CARS algorithm screens and optimizes spectroscopic wavelengths with superior predictive capabilities in NIRS through a competitive mechanism and an adaptive reweighting strategy [23]. MCUVE evaluates each variable's contribution to the model's performance by simulating random data and automatically eliminates uninformative variables. This algorithm has several advantages, including high robustness, global search capability, and suitability for complex data [24]. RF is a simple and widely applicable heuristic algorithm characterised by its ease of implementation. It is known for its ability to avoid getting trapped in local optima, making it suitable for large scale problems. It can also be parallelized for efficient processing [25].

2.5. External Verification. Employing an external validation set facilitates an unbiased assessment of the performance of the model. Manipulations such as parameter tuning and feature selection are common during model training, tailor-made to the training data, yet bearing the risk of inciting overfitting. Using an external validation set, we can examine the model's ability to generalize on previously unseen data, enabling a more accurate evaluation of its authentic predictive capacity.

2.6. Evaluation Indicators. The quantitative model established in this study was developed using partial least squares regression (PLSR). We performed  $R^2$  evaluation for multiple components in different batches of *Polygoni Multiflori* Radix extract. The coefficient of determination  $R^2$  is a metric that assesses the goodness of fit of the model, with higher  $R^2$  values indicating closer proximity between the predicted and actual values. The evaluation encompasses three  $R^2$  coefficients: calibration determination coefficient ( $R^2_{CV}$ ), cross-validation determination coefficient ( $R^2_{P}$ ). The calculation formulas for these coefficients are given in the following equation:

$$R^{2} = 1 - \frac{\sum_{i=1}^{n} (y_{i} - \hat{y}_{i})^{2}}{\sum_{i=1}^{n} (y_{i} - \bar{y})^{2}}.$$
 (1)

The root mean square error (RMSE), an important metric that measures the quality of a quantitative model, calculates the error between the actual and predicted values, and a lower RMSE value indicates a smaller deviation and a better predictive performance of the model. Model evaluation typically includes three RMSE metrics: root mean square error of calibration (RMSE<sub>C</sub>), root mean square error of cross-validation (RMSE<sub>CV</sub>), and root mean square error of prediction (RMSE<sub>P</sub>). The calculation formulas for these metrics are as follows:

RMSE = 
$$\sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}}$$
. (2)

The relative percent difference (RPD) is a metric used to evaluate the performance of a model and is commonly used to assess the consistency between actual and predicted values. A higher RPD value indicates better consistency between actual and predicted values, while a lower RPD value indicates poorer consistency. If the RPD is greater than 2.0, it indicates that the model has excellent predictive ability. If the RPD is in the range of 1.4–2.0, it indicates that the model's predictive performance is reasonably good. However, if the RPD is less than 1.4, it indicates poor predictive performance and the results are unreliable. The formula for calculating this metric is shown in the following equation:

$$RPD = \frac{SD}{RMSE_P},$$
 (3)

where *n* is the number of samples,  $y_i$  is the reference measurement results for sample *i*,  $\hat{y}_i$  is the predicted measurement results for sample *i*,  $\bar{y}$  is the mean of the reference measurement results for all samples in the dataset, and SD is the standard deviation of the sample content of the prediction set.

#### 3. Results and Discussion

3.1. NIR Spectroscopy Analysis. The acquired spectral graph contains the essential information necessary for modeling. Figure 1(a) illustrates the original spectra of various *Polygoni* Multiflori Radix samples, which exhibit multiple absorption bands. However, on visual examination, the spectral contours appear to exhibit similarities without distinctive differences. Moreover, the spectral profiles seem to lack informative details, demanding the utilization of multivariate algorithms to aid in the establishment of qualitative and quantitative models for predicting relevant parameters. NIRS belongs to the realm of molecular vibrational spectroscopy, predominantly capturing the overtones and combinations of stretching vibrations induced by hydrogencontaining groups (C-H, O-H, and N-H). The primary components of Polygoni Multiflori Radix include stilbene glycosides, anthraquinones, flavonoids, and phenolic acids. These bioactive compounds are characterised by the presence of abundant hydroxyl groups, carboxyl moieties, and carbon-hydrogen bonds.

In Figure 1(a), the absorption peaks at 7140-7040 cm<sup>-1</sup> and 5210-5050 cm<sup>-1</sup> correspond to the overtone and combination absorption ranges, respectively, of the functional groups of OH. The overtone absorption range of the functional group C=O is found between 5230 and 5130 cm<sup>-1</sup>, while the overtone and combination absorption ranges of the functional groups C-H are observed between  $8700 \text{ and } 8200 \text{ cm}^{-1}, 7350-7200 \text{ cm}^{-1}, 7090-6900 \text{ cm}^{-1},$  $6020-5550\,{\rm cm^{-1}},~{\rm and}~4440-4200\,{\rm cm^{-1}}.$  However, the spectra shown in the graph exhibit a significant amount of overlap, making it difficult to distinguish the spectral information of the target components. In addition, during the acquisition of NIRS data, the spectra are prone to influences such as scattering, background interference, and noise. Therefore, prior to modeling, it is necessary to apply chemometric methods to preprocess the spectral data, such as SNV and 1st Dev shown in Figure 1, to mitigate these effects.

3.2. Analysis of reference Values. According to the 2020 edition of the Chinese Pharmacopoeia, Part I, quantitative analyzes of moisture, stilbene glycoside, and bound an-thraquinone were carried out in different batches of *Polygoni Multiflori* Radix, and the results are shown in Table 1. From the data in the table, it can be seen that there are unqualified samples of each index component in the sample set consisting of 149 batches of *Polygoni Multiflori* Radix. This indicates that it is appropriate to select this sample set to establish the quantitative model.

#### 3.3. Quantitative Model Construction

3.3.1. Outlier Detection and Sample Division. Abnormal samples may arise from issues related to sample collection, experimental procedures, or instrument sensor malfunction, among other factors. These abnormal samples can result in biases during the model training phase, affecting its accuracy and reliability. To improve the stability of the model and enhance its generalizability, this study employed the methods included in the software for identifying and removing outlier samples. As a result, up to 13 abnormal samples were excluded from the three quantitative models. The remaining samples were divided into a calibration set and an external validation set using the K-S algorithm in a ratio of 3:1(see Table 2). During the partitioning process, the concentration information of the target components was also considered to ensure that the concentration range of the training set covered that of the external validation set. In addition, this study incorporated an internal validation set into the establishment of the model, which was derived from random cross-validation of the calibration set samples.

3.3.2. Comparison of Spectral Pretreatment. The optimal model typically has the following characteristics: a close-tounity  $R^2_{C}$ , indicating a strong association between the predicted values and the reference values, ensuring increased reliability in the predictions. An  $R^2_C$  value exceeding 0.9 denotes a robust correlation between the predicted values and the target variable. A smaller  $RMSE_{C}$  indicates superior performance of the quantitative model, and a smaller RMSE<sub>CV</sub> indicates improved predictive capabilities of the model [26]. Table 3 shows that the preprocessed spectral models have varying degrees of improvement. The quantitative model, after undergoing first derivative deconvolution processing, performs better than other treatments. However, the results in the calibration set are better than those in the prediction set. The presence of excessive redundant information and numerous wavelength variables in the full spectrum near-infrared data may be the reason for this phenomenon. To simplify the model and enhance its accuracy, effective variable selection methods must be employed to eliminate redundant information in the spectra.

3.3.3. Selection of Wavelengths. The predictive capabilities of a model can decrease due to the abundance of spectral variables, making it necessary to screen spectral data for



FIGURE 1: The preprocessed NIRS spectra. (a) Original. (b) SG. (c) SNV. (d) 1<sup>st</sup> Dev. (e) 2<sup>nd</sup> Dev.

TABLE 1: Quantitative analysis of Polygoni Multiflori Radix content.

Component	Component content (%)	Mean content (%)	Standard (%)
Moisture	4.5-13.8	10.1	≤10.0
Stilbene glycosides	0.02-6.42	2.26	≥1.00
Anthraquinone glycosides	0.00-0.58	0.15	≥0.05

TABLE 2: Qualitative results of the classification of the sample set of the model.

In and diants	Calibration	set	Prediction s	Prediction set		
Ingredients	Number of samples	Range (%)	Number of samples	Range (%)		
Moisture	100	4.5-13.0	38	7.5-13.0		
Stilbene glycosides	102	0.02-4.75	34	0.08 - 4.00		
Anthraquinone glycosides	102	0.00-0.40	34	0.00-0.36		

irrelevant variables. This study used commonly used variable selection methods, such as competitive adaptive weighted sampling (CARS), Monte Carlo uninformative variable elimination (MCUVE), and random frog leaping (RF).

Table 4 presents the results, which show significant improvements across  $R^2_{C}$ ,  $R^2_{P}$ , RMSE<sub>CV</sub>, and RMSE<sub>P</sub> for all three algorithms compared to the full wavelength analysis. The quantitative model involved in CARS shows strong overall performance. However, the efficacy of MCUVEbased feature selection appears to be comparatively limited in this study, possibly due to increased result instability arising from the large sample size. In the context of largescale data, the computation of mutual information may be susceptible to noise and data volume. This can introduce challenges in the evaluation of results and reduce the effectiveness of the MCUVE algorithm. As a result, there is the possibility of accidentally misidentifying predictive features, which can affect their predictive capabilities.

NIRS provides significant information about the nonsimple harmonic vibrations of molecules. This includes bands typically associated with the NIR region, which originates from the telescopic vibrations of hydrogencontaining groups (CH, NH, and OH). Figure 2 shows the distribution of the variables selected for the three components in the red area, as determined by the optimization algorithm. The variables selected for moisture content modeling are concentrated in specific wavelength ranges. These ranges are  $10000-8850 \text{ cm}^{-1}$ ,  $7140-7040 \text{ cm}^{-1}$ , and  $5210-5050 \text{ cm}^{-1}$ . They correspond to the quadratic octave absorption, the primary octave absorption, and the combined frequency absorption of the O-H groups, respectively [27].

Ingredients		LVs		Calibratio	Prediction			
	Method		$R^2_C$	$\text{RMSE}_C$ (%)	$RMSE_{CV}$ (%)	$R^2_P$	$RMSE_P$ (%)	RPD
	2 <sup>nd</sup> Dev	6	0.66	0.42	1.25	0.60	0.99	1.1
Moisture	1 <sup>st</sup> Dev	7	0.82	0.24	0.91	0.72	0.65	1.7
	SNV	5	0.72	1.05	1.15	0.57	0.86	1.3
	SG	9	0.75	0.88	1.07	0.71	0.67	1.7
	None	8	0.74	0.93	1.08	0.71	0.68	1.7
	2 <sup>nd</sup> Dev	2	0.49	0.68	0.78	0.21	0.83	0.9
	1 <sup>st</sup> Dev	4	0.52	0.32	0.77	0.28	0.81	1.0
Stilbene glycosides	SNV	9	0.40	0.67	0.87	0.28	0.88	0.9
07	SG	11	0.50	0.62	0.79	0.47	0.66	1.2
	None	9	0.44	0.67	0.83	0.38	0.72	1.1
Anthraquinone glycosides	2 <sup>nd</sup> Dev	8	0.53	0.42	0.79	0.44	0.58	1.0
	1 <sup>st</sup> Dev	6	0.58	0.54	0.69	0.54	0.75	0.8
	SNV	7	0.51	0.96	1.15	0.47	0.79	0.8
	SG	9	0.47	0.93	0.97	0.33	0.88	0.7
	None	9	0.48	0.94	1.09	0.31	1.00	0.6

TABLE 3: Comparison of pretreatment with quantitative models.

The bold values in the table are the model parameter values under the optimal preprocessing approach, and are bolded to highlight them.

TABLE 4: Comparison of quantitative model wavelength selection.

Ingredients	Method V	<b>XA7 I</b>	LVs	Calibration			Prediction		
		wavenumber		$R^2_C$	$\text{RMSE}_C$ (%)	$RMSE_{CV}$ (%)	$R^2_P$	$RMSE_P$ (%)	RPD
Moisture	None	2666	7	0.82	0.24	0.91	0.72	0.65	1.7
	MCUVE	84	4	0.87	0.81	0.89	0.78	0.63	1.8
	CARS	68	4	0.97	0.29	0.38	0.93	0.29	3.8
	RF	113	4	0.96	0.30	0.41	0.93	0.30	3.7
Stilbene glycosides	None	2666	4	0.52	0.32	0.77	0.28	0.81	1.0
	MCUVE	126	4	0.76	0.63	0.69	0.61	0.71	1.1
	CARS	108	4	0.93	0.33	0.45	0.91	0.41	1.9
	RF	68	4	0.89	0.32	0.49	0.87	0.63	1.2
Anthraquinone glycosides	None	2666	6	0.58	0.54	0.69	0.54	0.75	0.8
	MCUVE	156	4	0.74	0.53	0.47	0.74	0.45	1.3
	CARS	59	5	0.92	0.31	0.45	0.89	0.44	1.4
	RF	98	5	0.86	0.40	0.56	0.77	0.45	1.3

The bold values in the table are the model parameter values under the best wavelength selection, and the bold is for emphasis.



FIGURE 2: The preferred wavelength selection distribution by CARS. (a) Moisture; (b) stilbene glycosides; (c) anthraquinone glycosides.

In addition, the variables selected by the content prediction model algorithms are distributed in the wavelength ranges  $11000-8500 \text{ cm}^{-1}$  and  $5000-4000 \text{ cm}^{-1}$  associated with vibrational absorption of C-H groups. The selected variables near the ranges  $4500-4100 \text{ cm}^{-1}$ ,  $6000-5500 \text{ cm}^{-1}$ , and  $8000 \text{ cm}^{-1}$  are mainly generated by

combination bands of C-H bending vibration, C-H stretching vibration, and C-H bending vibration combined with O-H stretching vibration within the wavelength range  $4500-4100 \text{ cm}^{-1}$  [28]. The wavelength points near  $6020-5500 \text{ cm}^{-1}$  and  $8000 \text{ cm}^{-1}$  are caused by the primary and secondary combination frequencies of the C-



FIGURE 3: Correlation between the predicted value and the reference value. (a) Moisture. (b) Stilbene glycosides. (c) Anthraquinone glycosides.

H stretching vibration in  $-CH_2$  and  $-CH_3$  [29]. It can be seen that the wavelength points selected by this algorithm are quite scattered. The composition of the samples is complex, with valuable information within each frequency band, so accurate band assignment is a significant challenge.

3.3.4. Algorithmic Advantages. The CARS algorithm, which is based on the Monte Carlo sampling method, treats each variable as an individual entity. It performs a phased selection process on these variables and adjusts the variable retention rate using an exponential decay function [30]. The algorithm also introduces model population analysis (MPA) and employs statistical techniques to effectively examine the parameters for model subset selection. The spectral wavelengths with the worst predictive performance are identified by comparing prediction errors based on performance metrics. By attenuating their weights, the algorithm mitigates their impact on the model. Furthermore, the reweighted spectral wavelengths are incorporated into the subsequent rounds of competition and selection, thereby boosting the model's performance. Consequently, it improves the performance of quantitative models and has extensive applications in the field of NIRS [31, 32].

The MCUVE algorithm assesses each variable's contribution to the model's performance by simulating data randomness and repeatability, achieving variable elimination without using information, and generates random datasets that disrupt the correlation between variables and responses, ensuring that the datasets generated do not have an actual information association between variables and responses. A model is constructed using randomly generated datasets, and performance metrics such as the RMSE or prediction residual sum of squares (PRESS) are used to evaluate the model's performance. To improve the model's overall performance, the MCUVE algorithm eliminates variables with small contributions or no significant impact. This is done through repeated iterations based on a certain threshold or variable ranking. However, its applicability in the industry is limited due to the significant computational requirements and challenges associated with selecting model parameters [33].

The RF algorithm is an intelligent optimization algorithm based on principles of biological evolution and simulates the foraging behaviour of amphibians, using their random and nondeterministic leaps to search for optimal solutions [34]. The algorithm begins by generating a random population, with each individual representing a potential solution. The individuals are then evaluated based on an objective function to obtain their fitness value. This process mimics the jumping process of amphibians in search of food. To find superior solutions, each individual undergoes random leaps within the solution space following certain strategies and rules. The fitness value of each frog is calculated on the basis of its new position, and the state of each individual in the population is updated accordingly. The algorithm terminates when either the maximum number of iterations is reached or a fitness threshold is achieved. The resulting solution obtained through the iterative process represents the algorithm's outcome. However, in complex high-dimensional optimization problems, the convergence rate of the RF algorithm may be slow, requiring more iterations to achieve an optimal solution [35, 36].

3.3.5. Quality of the Test Model with the External Validation Set. To provide a more comprehensive illustration of the preferred model results presented in the table, a scatterplot has been generated in Figure 3. As shown in the graph, all data points exhibit a dense clustering pattern near the diagonal line. The results presented in Table 4 and Figure 3 show notable improvements in the  $R^2_P$  values of all three quantitative models, demonstrating commendable robustness and predictive precision. Furthermore, the RMSE<sub>P</sub> values of the samples do not show any significant differences, indicating a well-optimized training of the models without any noticeable overfitting problems.

#### 4. Conclusion

In this study, an NIRS method was proposed for the quantitative analysis of *Polygoni Multiflori* Radix. After confirming the capability of the PLSR method, different

spectroscopic preprocessing and wavelength selection methods were extensively investigated and compared to obtain the optimal NIRS model. The experimental results indicate that appropriate spectral preprocessing algorithms and wavelength selection techniques significantly improve the predictive accuracy of the constituent content models. The optimised  $R_{P}^{2}$  values show significant improvements, while the RMSE<sub>P</sub> values decrease significantly, indicating improved model stability. This demonstrates the ability of the model to accurately predict moisture, stilbene glycosides, and anthraquinone glycoside content in Polygoni Multiflori Radix. The results also suggest that the proposed method meets the conventional control requirements of herb analysis and can be used for practical content determination. However, it should be noted that the quality of Polygoni Multiflori Radix is influenced by factors such as species, origin, cultivation management, and storage conditions. Therefore, further efforts are needed to increase the sample size to improve the performance and generalizability of the model.

#### **Data Availability**

The raw/processed data required to reproduce these findings cannot be shared at this time as the data also form part of an ongoing study.

#### **Conflicts of Interest**

The authors declare that they have no conflicts of interest.

#### Acknowledgments

The work was supported by the Open Research Fund of the Key Laboratory of NMPA for Rapid Drug Testing Technology, Guangdong Institute for Drug Control (No. KF2022006), and the Pearl River S & T Nova Program of Guangzhou (No. 201610010113).

#### References

- S. Jung, H. Son, C. E. Hwang et al., "The root of Polygonum multiflorum Thunb. Alleviates non-alcoholic steatosis and insulin resistance in high fat diet-fed mice," *Nutrients*, vol. 12, no. 8, p. 2353, 2020.
- [2] J. Y. Shin, Y. H. Choi, J. Kim et al., "Polygonum multiflorum extract support hair growth by elongating anagen phase and abrogating the effect of androgen in cultured human dermal papilla cells," *BMC Complement Med Ther*, vol. 20, no. 1, p. 144, 2020.
- [3] D. U. Kim, J. Y. Chung, S. C. Jin et al., "Effects of processed Polygonum multiflorum with KIOM patent on bone remodeling-related protein expression in human osteoblast-like SaOS-2 cells," *Evidence-based Complementary* and Alternative Medicine, vol. 2020, Article ID 4168535, 6 pages, 2020.
- [4] X. Y. Li, Z. Wang, J. G. Jiang, and C. Y. Shen, "Role of polyphenols from Polygonum multiflorum Caulis in obesityrelated disorders," *Journal of Ethnopharmacology*, vol. 294, Article ID 115378, 2022.

- [5] D. Gao, J. P. Hao, B. Y. Li et al., "Tetrahydroxy stilbene glycoside ameliorates neuroinflammation for Alzheimer's disease via cGAS-STING," *European Journal of Pharmacology*, vol. 953, Article ID 175809, 2023.
- [6] D. Kim, H. J. Kim, and H. S. Jun, "Polygonum multiflorum Thunb. Extract stimulates melanogenesis by induction of COX2 expression through the activation of p38 MAPK in B16F10 mouse melanoma cells," *Evidence-based Complementary and Alternative Medicine*, vol. 2020, Article ID 7642019, 10 pages, 2020.
- [7] J. B. Yang, C. S. Yang, J. Li et al., "Dianthrone derivatives from Polygonum multiflorum Thunb: anti-diabetic activity, structure-activity relationships (SARs), and mode of action," *Bioorganic Chemistry*, vol. 135, Article ID 106491, 2023.
- [8] C. Wang, S. Dai, L. Gong et al., "A review of pharmacology, toxicity and pharmacokinetics of 2,3,5,4'-tetrahydroxystilbene-2-O-β-D-Glucoside," *Frontiers in Pharmacology*, vol. 12, Article ID 791214, 2021.
- [9] C. P. Commission, *Pharmacopoeia of the People's Republic of China*, China Medical Science Press, Beijing, China, 2020.
- [10] M. Tian, X. Ma, M. Liang, and H. Zang, "Application of rapid identification and determination of moisture content of Coptidis rhizoma from different species based on data fusion," *Journal of AOAC International*, vol. 106, no. 5, pp. 1389–1401, 2023.
- [11] L. Ma, D. Liu, C. Du et al., "Novel NIR modeling design and assignment in process quality control of Honeysuckle flower by QbD," Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, vol. 242, Article ID 118740, 2020.
- [12] L. Strani, R. Vitale, D. Tanzilli et al., "A multiblock approach to fuse process and near-infrared sensors for on-line prediction of polymer properties," *Sensors*, vol. 22, no. 4, p. 1436, 2022.
- [13] Y. Pu, D. Pérez-Marín, N. O'Shea, and A. Garrido-Varo, "Recent advances in portable and handheld NIR spectrometers and applications in milk, cheese and dairy powders," *Foods*, vol. 10, no. 10, p. 2377, 2021.
- [14] K. B. Beć, J. Grabska, and C. W. Huck, "Miniaturized NIR spectroscopy in food analysis and quality control: promises, challenges, and perspectives," *Foods*, vol. 11, no. 10, p. 1465, 2022.
- [15] X. Wu, F. Yang, S. Cai, K. Pu, and G. Hong, "Nanotransducerenabled deep-brain neuromodulation with NIR-II light," *American Chemical Society Nano*, vol. 17, no. 9, pp. 7941– 7952, 2023.
- [16] X. Jintao, S. Yongli, Y. Liming et al., "Near-infrared spectroscopy for rapid and simultaneous determination of five main active components in rhubarb of different geographical origins and processing," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 205, pp. 419–427, 2018.
- [17] X. Jintao, L. Yufei, Y. Liming et al., "Rapid and simultaneous analysis of five alkaloids in four parts of Coptidis Rhizoma by near-infrared spectroscopy," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 188, pp. 611– 618, 2018.
- [18] W. Guangrui, X. Jintao, Y. Quanwei et al., "Rapid determination of puerarin by near-infrared spectroscopy during percolation and concentration process of Puerariae Lobatae Radix," *Pharmacognosy Magazine*, vol. 12, no. 47, p. 188, 2016.
- [19] L. Ye, K. Li, W. Wang, Y. Liu, S. Jiang, and G. Huang, "Nearinfrared spectroscopy as a process analytical technology tool for monitoring the parching process of traditional Chinese

medicine based on two kinds of chemical indicators," *Pharmacognosy Magazine*, vol. 13, no. 50, p. 332, 2017.

- [20] W. Zhu, H. Hong, Z. Hong et al., "Rapid quality identification of decoction pieces of crude and processed Corydalis rhizoma by near-infrared spectroscopy coupled with chemometrics," *Journal of Analytical Methods in Chemistry*, vol. 2021, Article ID 1936057, 11 pages, 2021.
- [21] H. Zhang, X. Hu, L. Liu, J. Wei, and X. Bian, "Near infrared spectroscopy combined with chemometrics for quantitative analysis of corn oil in edible blend oil," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 270, Article ID 120841, 2022.
- [22] C. Q. An, X. Yan, C. Lu, and X. H. Zhu, "Effect of spectral pretreatment on qualitative identification of adulterated bovine colostrum by near-infrared spectroscopy," *Infrared Physics and Technology*, vol. 118, Article ID 103869, 2021.
- [23] F. Ping, J. Yang, X. Zhou et al., "Quality assessment and ripeness prediction of table grapes using visible-near-infrared spectroscopy," *Foods*, vol. 12, no. 12, p. 2364, 2023.
- [24] T. Wang, Y. Zhang, Y. Liu, Z. Zhang, and T. Yan, "Intelligent evaluation of stone cell content of korla fragrant pears by vis/ NIR reflection spectroscopy," *Foods*, vol. 11, no. 16, p. 2391, 2022.
- [25] Z. Bai, M. Xie, B. Hu et al., "Estimation of soil organic carbon using vis-NIR spectral data and spectral feature bands selection in southern Xinjiang, China," *Sensors*, vol. 22, no. 16, p. 6124, 2022.
- [26] M. A. Berhow, M. Singh, M. J. Bowman, N. P. J. Price, S. F. Vaughn, and S. X. Liu, "Quantitative NIR determination of isoflavone and saponin content of ground soybeans," *Food Chemistry*, vol. 317, Article ID 126373, 2020.
- [27] A. Malvandi, H. Feng, and M. Kamruzzaman, "Application of NIR spectroscopy and multivariate analysis for Nondestructive evaluation of apple moisture content during ultrasonic drying," Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, vol. 269, Article ID 120733, 2022.
- [28] M. X. Li, Y. B. Shi, J. B. Zhang et al., "Rapid evaluation of Ziziphi Spinosae Semen and its adulterants based on the combination of FT-NIR and multivariate algorithms," *Food Chemistry X*, vol. 20, Article ID 101022, 2023.
- [29] Z. Lan, Y. Zhang, Y. Sun et al., "A mid-level data fusion approach for evaluating the internal and external changes determined by FT-NIR, electronic nose and colorimeter in Curcumae Rhizoma processing," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 188, Article ID 113387, 2020.
- [30] H. Li, Y. Liang, Q. Xu, and D. Cao, "Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration," *Analytica Chimica Acta*, vol. 648, no. 1, pp. 77–84, 2009.
- [31] S. A. Haruna, H. Li, W. Wei et al., "Simultaneous quantification of total flavonoids and phenolic content in raw peanut seeds via NIR spectroscopy coupled with integrated algorithms," Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, vol. 285, Article ID 121854, 2023.
- [32] L. Wang and R. Wang, "Determination of soil pH from Vis-NIR spectroscopy by extreme learning machine and variable selection: a case study in lime concretion black soil," Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, vol. 283, Article ID 121707, 2022.
- [33] Y. Hu, H. Zhang, W. Liang, P. Xu, K. Lou, and J. Pu, "Rapid and simultaneous measurement of praeruptorin A, praeruptorin B, praeruptorin E, and moisture contents in peucedani Radix using near-infrared spectroscopy and chemometrics," *Journal* of AOAC International, vol. 103, no. 2, pp. 504–512, 2020.

- [34] J. Sun, W. Yang, M. Feng, Q. Liu, and M. S. Kubar, "An efficient variable selection method based on random frog for the multivariate calibration of NIR spectra," *Royal Society of Chemistry Advances*, vol. 10, no. 28, pp. 16245–16253, 2020.
- [35] Y. Chen, W. Sun, S. Jiu et al., "Soluble solids content binary classification of miyagawa satsuma in chongming island based on near infrared spectroscopy," *Frontiers of Plant Science*, vol. 13, Article ID 841452, 2022.
- [36] S. A. Haruna, H. Li, W. Wei et al., "Intelligent evaluation of free amino acid and crude protein content in raw peanut seed kernels using NIR spectroscopy paired with multivariable calibration," *Analytical Methods*, vol. 14, no. 31, pp. 2989– 2999, 2022.