Zhao J-L *et al.* (2022)
Notulae Botanicae Horti Agrobotanici Cluj-Napoca
Volume 50, Issue 4, Article number 12945
DOI:10.15835/nbha50412945
Research Article

Estimation of nutrient contents in wolfberry (*Lycium barbarum* L.) based on hyperspectral analysis

Jin-Long ZHAO[1,2], Xue-Yi ZHANG[1,2*], Qi ZHANG[1,2], Xue-Jun NAN[1,2], Yun-Xia WANG[1,2], Yang LI[1,2*]

[1]*Key Laboratory for Meteorological Disaster Monitoring and Early Warning and Risk Management of Characteristic Agriculture in Arid Regions, CMA, Yinchuan 750002, China; zjl891229@163.com; 49793811@qq.com (*corresponding author); 13770860693@163.com; 1530679768@qq.com; WangYX112@163.com; lynuist68@163.com (*corresponding author)*
[2]*Ningxia Key Lab of Meteorological Disaster Prevention and Reduction, Ningxia Institute of Meteorological Sciences, Yinchuan 750002, China*

Abstract

Rapid and accurate determination of the nutrient contents in wolfberry (*Lycium barbarum* L.) is of great significance for identifying the quality and origin of the fruit. Compared to traditional chemical analysis methods, hyperspectral remote sensing has the advantages of high speed and low cost. In this study, the dried fruits of wolfberry (cultivar 'Ningqi No. 7') taken from the Huinong, Yinchuan, Zhongning, and Tongxin regions of Ningxia, China, in 2020 were selected as samples. Two methods, the variable importance measure (VIM) under random forest and the successive projection algorithm (SPA), were applied to select the hyperspectral characteristic variables. The test set coefficient of determination (R^2p), root mean square error of prediction (RMSEP), and relative percent deviation (RPD) for different pre-treatments and characteristic variable selection methods were compared. Finally, the optimal estimation models of partial least squares regression (PLSR) for the contents of eight nutrients in wolfberry were established. The results were as follows: (1) The variation trends of the original hyperspectral reflectance curves of wolfberries from different yield areas and harvest dates were similar, and the spectral characteristic absorption bands were significantly enhanced after first-order differential transformation. (2) After extracting the characteristic bands using SPA, the RPD values were all above 2.0, and the estimation performance was significantly better than that of the full bands. (3) The optimal estimation model for total sugar (TS), crude protein (CP), and Ca was found to be SG-MSC-SPA-PLSR; the optimal estimation model for *Lycium barbarum* polysaccharide (LBP), Mn, and Zn was found to be SG-2ndD-SPA-PLSR; and the optimal estimation model for Cu and Fe was found to be SG-1stD-SPA-PLSR. The results have important reference value for the quality evaluation and origin traceability of wolfberry.

Keywords: hyperspectral; nutrient content; PLSR; remote sensing; wolfberry

Introduction

Wolfberry (*Lycium barbarum* L.) is a deciduous shrub of *Lycium* L. in *Solanaceae*. This berry has been cultivated for more than 600 years in northwest China (Chen *et al.*, 2013). The dry and mature fruits of

Received: 11 Oct 2022. Received in revised form: 23 Oct 2022. Accepted: 31 Oct 2022. Published online: 02 Dec 2022.
From Volume 49, Issue 1, 2021, Notulae Botanicae Horti Agrobotanici Cluj-Napoca journal uses article numbers in place of the traditional method of continuous pagination through the volume. The journal will continue to appear quarterly, as before, with four annual numbers.

wolfberry are used in traditional Chinese medicine and can nourish the liver and kidney, reduce thirst due to internal heat (Chinese Pharmacopoeia Commission of the People's Republic of China, 2020), reduce blood sugar (Zou *et al*., 2010), and exert anti-cancer effects (Mao *et al*., 2011). In recent years, wolfberry has become popular in Europe and North America as a medicine and functional food, and many commercial products named after wolfberry have entered the health food market (Potterat, 2010). However, there are great differences in the medicinal quality of *Lycium barbarum* obtained from different yield areas. Therefore, the accurate and rapid acquisition of nutrient contents in *Lycium barbarum* is of important reference value for identifying the origin and quality of the berry.

However, the traditional chemical analysis method is time-consuming and complicated. Hyperspectral remote sensing technology has the advantages of being multi-band (350-2,500 nm) and offering high resolution results (3-10 nm). The diffuse reflection spectra obtained via hyperspectral remote sensing technology contain abundant information on the structure and composition of reflectors (Pu and Gong, 2003), thereby providing an effective method to obtain the contents of crop nutrients. As early as the 1960s and 1970s, researchers from the United States Department of Agriculture (USDA) performed detailed spectroscopic measurements of plant leaves, successfully revealing the characteristics of the 42 sub-absorption bands of chlorophyll, protein, water, and other notable leaf contents in the 400-2,400 nm band range (Curran, 1989). For more than half a century, with the rapid development of stoichiometry and the great improvements in the detection accuracy of hyperspectral instruments, studies on the use of hyperspectral technology in the agriculture and food fields have become increasingly based on multi-spectrum, multi-mode, and multi-platform technology. Especially in recent years, a combination of hyperspectral analysis and partial least squares regression (PLSR) has been widely used in crop quality estimation. Ye *et al*. (2020) estimated the nitrogen content of Fuji apple trees in Aomori, Japan, at different scales of leaf and canopy. Grieco *et al*. (2022) established a model to accurately estimate the nitrogen, phosphorus, and potassium concentrations in barley HEB-25 leaves from Halle, Germany. Ouyang *et al*. (2021) studied and identified the contents of caffeine, tea polyphenols, free amino acids, and chlorophyll in matcha from eastern China. Caporaso *et al*. (2022) established a model to effectively estimate the volatile compounds of Arabica and Robusta in commercial roasted coffee beans.

In terms of the hyperspectral inversion of nutrient contents in wolfberry, several scholars (Fatchurrahman *et al*., 2021; Li *et al*., 2017; Wang *et al*., 2016) have studied and established estimation models for total sugar (TS), *Lycium barbarum* polysaccharide (LBP), vitamin C, total antioxidants, phenols, anthocyanins, soluble solid content, and total acidity. However, studies on the hyperspectral estimation of crude protein (CP) and mineral elements mainly focused on herbage (Duranovich *et al*., 2020), rice (Liu *et al*., 2014), wheat (Hu *et al*., 2021), avocado (Kämper *et al*., 2020), orange (Osco *et al*., 2020), nuts (Bai *et al*., 2018), etc., with few reports on wolfberry.

The aim of the present study was to evaluate the application potential of hyperspectral technology for the detection of nutrient contents in *Lycium barbarum* and solve the time-consuming, labor-intensive, and high-cost issues of traditional chemical analysis methods. We used the wolfberry cultivar 'Ningqi No. 7' as experimental materials. The hyperspectral and chemical analysis data of 'Ningqi No. 7' collected in a laboratory were used to study the spectral response characteristics of TS, LBP, CP, Mn, Cu, Fe, Zn, and Ca with a combination of different pre-treatments and characteristic variable selection methods. Based on the above study, the overall aim was to establish optimal partial least squares regression models for the eight nutrients and provide a fast, accurate, and low-cost method for the quality identification of *Lycium barbarum*.

Materials and Methods

Study site

Ningxia Hui Autonomous Region (104°17'-107°40' E, 35°14'-39°23' N) is located in the inland of north central China. This region's unique light, heat, and soil moisture conditions are conducive to the photosynthesis, sugar accumulation, and fruit coloration of wolfberry, making this region one of the most important wolfberry cultivation and high-yield areas in China. The region's reputation can be described as follows: "the Yellow River is rich in Ningxia, and the *Lycium barbarum* is the best in the world". The yields and sales of wolfberry in Ningxia account for more than 60% of sales in China. Zhongning County of Ningxia, which is known as "the hometown of Chinese wolfberry", has *Lycium barbarum* with excellent medicinal quality. The efficacy of *Lycium barbarum* was recorded as early as the Ming Dynasty medical text "Compendium of Materia Medica", and wolfberry is the only cultivar listed in the 2020 edition of Chinese Pharmacopoeia (Chinese Pharmacopoeia Commission of the People's Republic of China, 2020). In 2020, *Lycium barbarum* was selected as part of the mutual recognition protection list in the "China-EU agreement on geographical indications", indicating that the quality of *Lycium barbarum* was recognized by the international market, and the international influence of the fruit was improved. In recent years, with the rapid development of the wolfberry industry in Ningxia, a new pattern of "one core and two belts" has formed. This pattern takes the Zhongning County as its core area and radiates throughout Qingshui River Basin and both sides of the Yellow River. The spatial distribution of sampling points at the study site is shown in Figure 1.

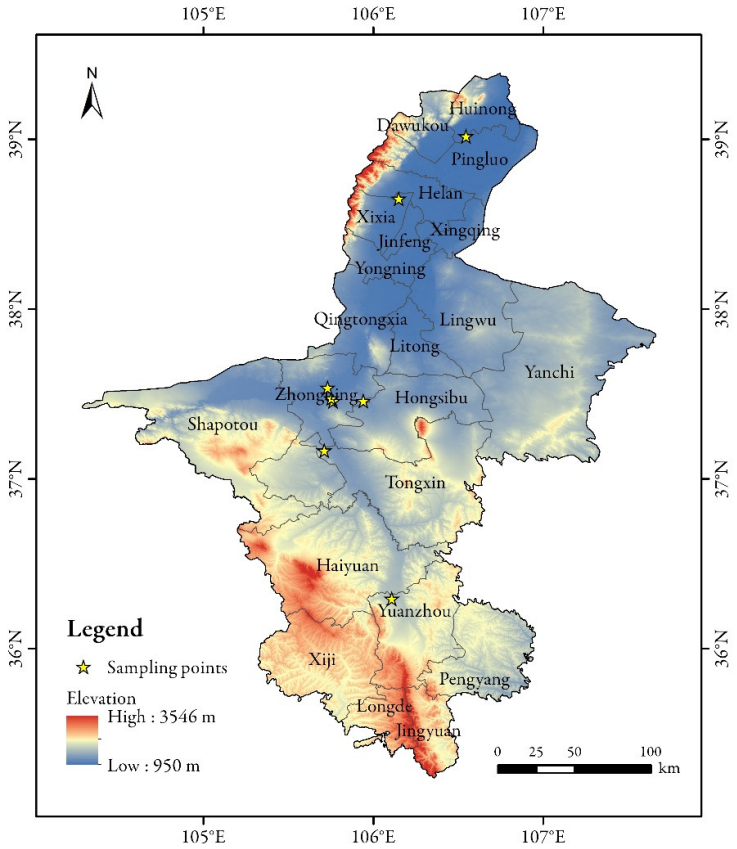

Figure 1. Study site and sampling point distribution

Sampling and drying

The cultivar used in this study was 'Ningqi No. 7', and the sampling points were located in the main yield areas of wolfberry in Ningxia, including Huinong, Yinchuan, Zhongning, Tongxin and Yuanzhou. The sampling period encompassed the summer fruit maturity period of *Lycium barbarum* in 2020 (June 18th, June 19th, June 28th, June 30th, July 6th, July 7th, July 8th, July 10th, July 16th, July 17th, July 23rd, July 24th, and August 3rd). The locations of sampling points in different batches were the same. We selected plants of similar ages and randomly took samples from the upper, middle, and lower parts of the canopy. Then, we put the samples into envelopes, stored them in a refrigerator, and quickly transported them to the laboratory.

To reduce the interference of external factors on the detection of nutrient contents of *Lycium barbarum*, a traditional Na_2CO_3 solution was not used to soak the wax layer on the surface of fresh fruit during the drying process. The whole process involved natural drying. After the fresh fruits were naturally dried to a constant weight, the stems and bad fruits were removed and then placed in a 60 °C oven to continue drying. When the moisture content of the wet basis of *Lycium barbarum* was less than 13%, drying was considered completed (El-Sebaii and Shalaby, 2013). Then, we used a high-speed crusher to crush the samples into powder and bagged and stored the powder in a dry container to prevent moisture return. Ultimately, we obtained 25 bags of samples, each about 250 g.

Hyperspectral measurement and chemical analysis

The dried fruit samples of *Lycium barbarum* were divided into two parts using a 90 × 90 mm petri dish. One part was about 10 mm thick and had a smooth and uniform surface for spectral measurement, while the other part was used for nutrient analysis. The pulverized wolfberries reduced the effects of texture and shadow on the reflectance of the dried fruit surface, making the samples more suitable for studying the spectral response characteristics of different nutrients based on the mechanism. Spectral measurements were carried out indoors using an SR-3500 portable spectrometer (Spectral Evolution Inc., Haverhill, MA, USA). The wavelength range of the SR-3500 was 350-2,500 nm with a resolution of 3.5 nm in the 350-1,000 nm range, 10 nm in the 1,000-1,900 nm range, and 7 nm in the 2,100-2,500 nm range. The instrument was preheated for 30 min before measurement. During the measurements, the probe was oriented vertically downward at a height of about 10 cm from the sample surface. The vertical height of the light source was placed at about 30 cm from the surface, and the zenith angle was about 15°. A white reflectance panel (5 × 5) was used to calculate the reflectance for each spectral measurement, and the process was carried out in a darkroom with a 100 W halogen tungsten lamp. Each sample was repeated 5 times, and the arithmetic mean was taken as the actual reflectance.

The chemical analysis indices of *Lycium barbarum* included total sugar (TS), *Lycium barbarum* polysaccharide (LBP), crude protein (CP), Mn, Cu, Fe, Zn, and Ca. TS was assayed by direct titration method reference to GB/T 18672-2014 Appendix B of the Chinese national standards (CNS) (Li *et al*., 2017). LBP was assayed by phenol-sulfuric acid method reference to GB/T 18672-2014 Appendix A of CNS (Yu *et al*., 2017). CP was assayed by Kjeldahl method reference to GB 5009.5-2016 of CNS (Tian *et al*., 2022). Mn was assayed by flame atomic absorption spectrophotometry (FAAS) at 279.5 nm reference to GB 5009.242-2017 of CNS (Wang *et al*., 2021). Cu was assayed by FAAS at 324.8 nm reference to GB 5009.13-2017 of CNS (Wang et al., 2021). Fe was assayed by FAAS at 248.3 nm reference to GB 5009.90-2016 of CNS (Wang *et al*., 2021). Zn was assayed by FAAS at 213.9 nm reference to GB 5009.14-2017 of CNS (Wang *et al*., 2021). Ca was assayed by FAAS at 422.7 nm reference to GB 5009.92-2016 of CNS (Wang *et al*., 2021). The mean, maximum (Max), minimum (Min), standard deviation (SD), and coefficient of variation (CV) in different sets are shown in Table 1.

Table 1. Basic nutrient contents of *Lycium barbarum* samples

Samples	Indices	TS[a]	LBP[a]	CP[a]	Mn[b]	Cu[b]	Fe[b]	Zn[b]	Ca[c]
Total (n=25)	Mean	52.83	4.27	14.52	7.96	7.25	52.16	12.69	0.84
	Max	62.40	5.78	17.20	14.50	11.00	87.40	20.40	1.10
	Min	41.80	3.10	12.00	4.10	2.00	16.50	7.60	0.57
	SD	4.96	0.68	1.54	2.48	2.22	15.85	2.92	0.13
	CV	0.09	0.16	0.11	0.31	0.31	0.30	0.23	0.16
Training set (n=17)	Mean	51.95	4.38	14.61	7.58	7.51	49.98	12.91	0.86
	Max	62.40	5.78	17.20	12.80	10.20	79.50	20.40	1.10
	Min	41.80	3.10	12.00	4.90	3.40	16.50	7.60	0.57
	SD	5.03	0.65	1.48	2.09	1.77	15.94	2.94	0.14
	CV	0.10	0.15	0.10	0.28	0.24	0.32	0.23	0.16
Test set (n=8)	Mean	54.70	4.04	14.31	8.79	6.71	56.80	12.21	0.80
	Max	61.20	5.28	16.70	14.50	11.00	87.40	17.40	1.05
	Min	46.00	3.14	12.70	4.10	2.00	32.40	9.20	0.67
	SD	4.23	0.69	1.65	3.00	2.89	14.61	2.80	0.11
	CV	0.08	0.17	0.12	0.34	0.43	0.26	0.23	0.14

Notes: [a], [b] and [c] represent the units of '$g \cdot 100g^{-1}$', '$mg \cdot kg^{-1}$' and '$g \cdot kg^{-1}$', respectively. CV is dimensionless.

Hyperspectral data processing method

The built-in DARWin SP software of the SR-3500 was used to extract the original spectral reflectance (OR). The R 4.1.2 "prospectr" package was used for data pre-processing. This package integrates many functions that are practical in near-infrared and infrared spectroscopy applications and can be used for spectral data pre-processing and the selection of representative samples or spectra. First, the original spectral data were smoothed using the Savitzky–Golay (SG) filter. The data were then subjected to first-order differential ($1^{st}D$), second-order differential ($2^{nd}D$), multivariate scattering correction (MSC), and standard normal variable transformation (SNV).

Estimation model of nutrient contents

Modelling method

The twenty-five samples were divided into a training set with 17 samples and a test set with 8 samples. The pre-treated spectral reflectance data of the full bands and characteristic bands were used to establish the models. The partial least squares regression models of total sugar, *Lycium barbarum* polysaccharide, crude protein, Mn, Cu, Fe, Zn, and Ca contents were established using the R 4.1.2 "PLS" package.

PLSR was proposed by Swedish statistician Herman Wold (Wold, 1966) in 1966 and integrates the advantages of three analysis methods: ordinary multiple linear regression, principal component analysis, and canonical correlation analysis. This method can accurately solve the problems of multicollinearity between independent variables, fewer samples than variables, and complex calculations (Yu *et al*., 2015).

Characteristic band

Hyperspectral data are multidimensional and contain a large amount of redundant data. By extracting the characteristic wavelengths, the interference of irrelevant information can be reduced, and the prediction ability of the model can be improved. In this study, we used the variable importance measure (VIM) under random forest and the successive projection algorithm (SPA) to select the characteristic bands.

The VIM indicates the contribution of each independent variable to the dependent variable. Here, we chose the mean decrease accuracy (MDA) as the evaluation indicator. MDA is a measure of the reduction in the accuracy of random forest predictions caused by changing the value of a variable into a random number. The higher the value, the greater the importance of the variable (Han *et al*., 2016). We ranked the results from

large to small according to importance and took the first 15 values as the characteristic variables. The built-in function "importance ()" in the R 4.1.2 "randomForest" package was used to select the variables.

The SPA uses variable projection operations in the data matrix to find the combination of spectral characteristic variables with the least amount of redundant information and the lowest collinearity. This method can extract a selection of variable data from a large number of original spectral data matrices to summarize most spectral information. This method also can avoid spectral information overlap to the greatest extent and improve modelling speed by reducing the amount of calculations and simplifying the model structure (Araújo *et al.*, 2001). The optimal combination of characteristic bands was determined according to the minimum root mean square error of prediction (RMSEP). We used the MATLAB 6.5 SPA toolkit (Araújo *et al.*, 2001; Galvão *et al.*, 2001) to select the characteristic variables.

Model accuracy test

The coefficient of determination (R^2), root mean square error (RMSE), and relative percent deviation (RPD) were used for model accuracy evaluation. R^2 represents the stability of the model. The closer R^2 is to 1, the better the stability of the model and the higher the degree of fitting. Here, RMSE is used to test the prediction ability of the model, where the smaller the RMSE, the stronger the prediction ability of the model. RPD is the ratio of the sample's standard deviation to the root mean square error, which is used to evaluate the prediction ability of the model. When RPD < 1.4, the model cannot predict the sample; when $1.4 \leq$ RPD < 2.0, the model prediction accuracy is general and the sample can be roughly predicted; when RPD $\geq$ 2.0, the model has excellent prediction ability (Bellon-Maurel *et al.*, 2010; Rossel *et al.*, 2006). The flow chart of the nutrient content estimation model is shown in Figure 2.

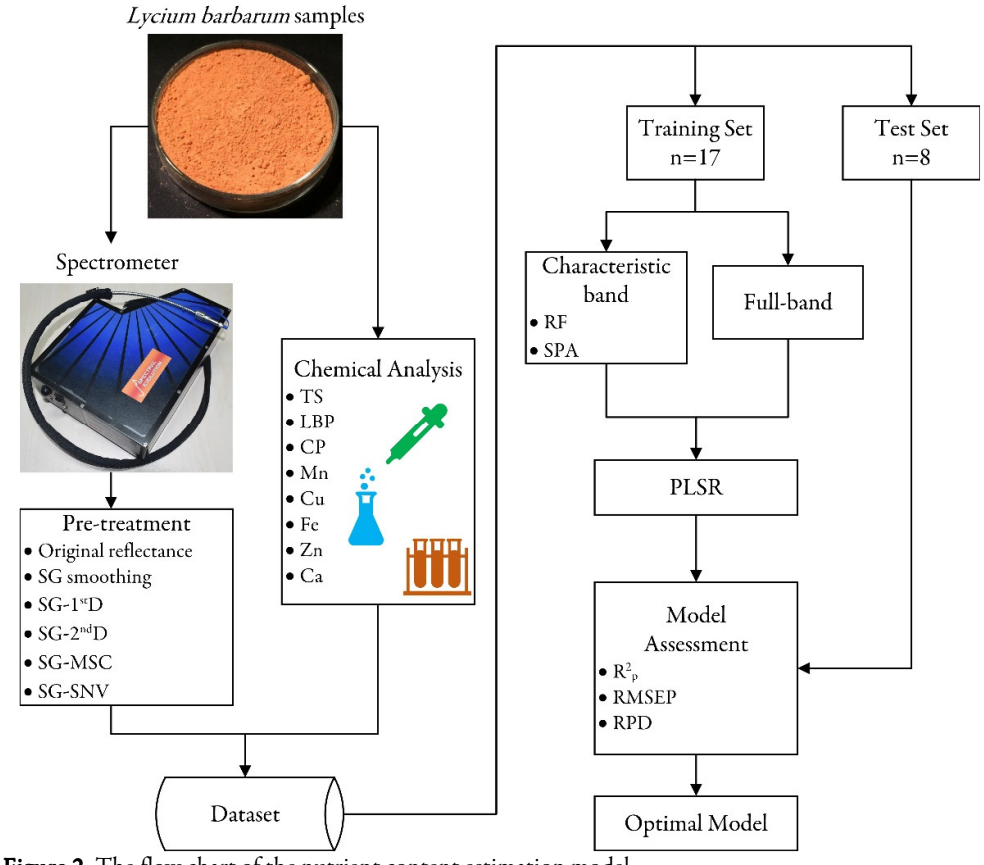

Figure 2. The flow chart of the nutrient content estimation model

Results

Hyperspectral characteristics and transformation

The original hyperspectral reflectance curves of *Lycium barbarum* were largely similar between different yield areas and harvesting periods (Figure 3a).

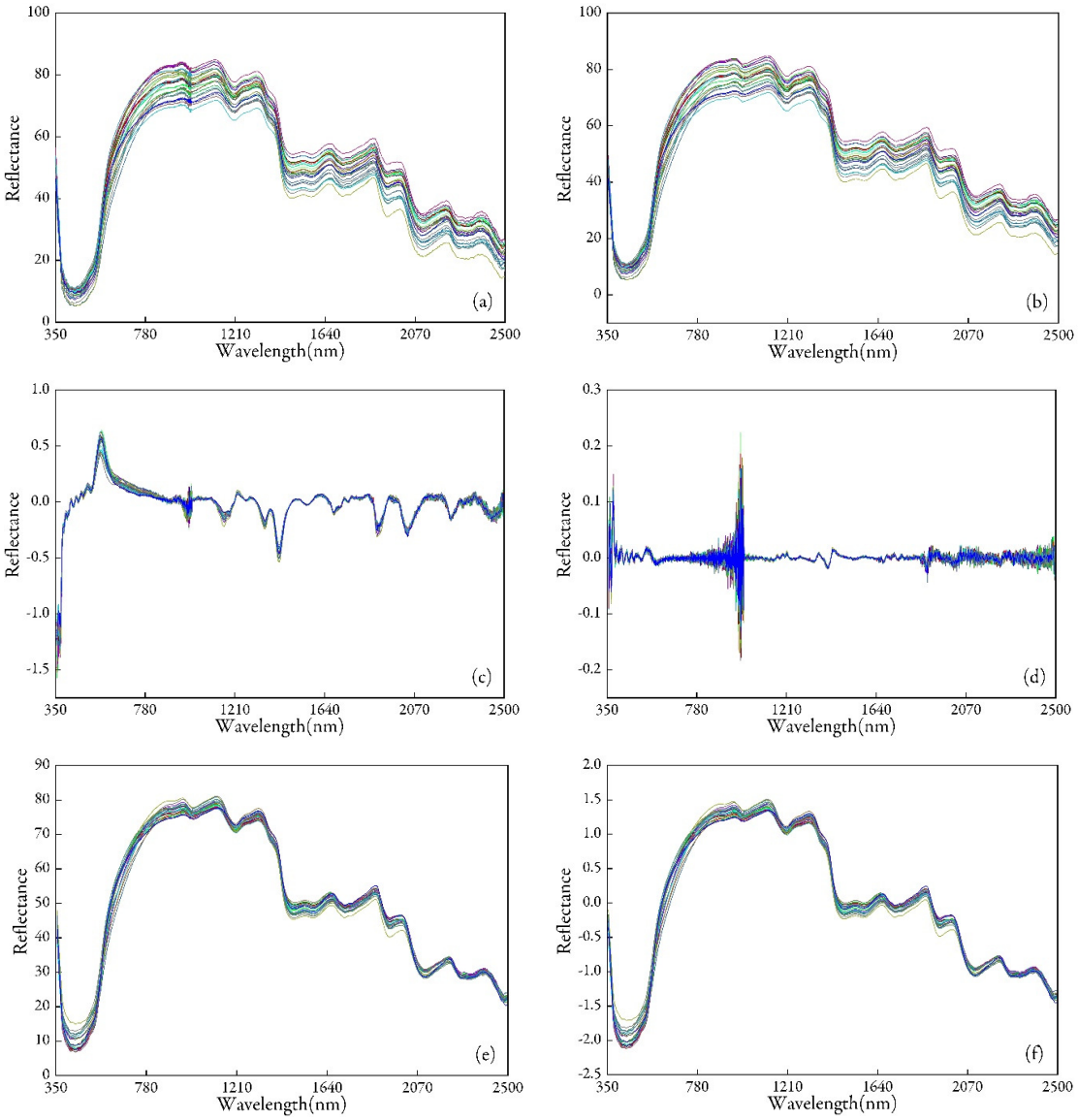

Figure 3. The original spectral curve and different pre-treatment spectral curves of the *Lycium barbarum* samples: (a) Original spectral reflectance, (b) SG smoothing, (c) SG-1stD, (d) SG-2ndD, (e) SG-MSC, and (f) SG-SNV

The curve slope was higher, and the reflectance rose rapidly, at 530-780 nm range of visible light. There was a peak region observed at 850-1,320 nm under near-infrared light, and the reflectance ranged from 65% to 85%. There were absorption peaks observed near 1,000 and 1,200 nm, as well as reflection peaks near 1,119 and 1,311 nm, which were mainly related to the second-order frequency doubling of the C–H bond stretching

vibration (Khodabux *et al*., 2007). The reflectance dropped rapidly at 1,320-1,450 nm and flattened at 1,450-1,875 nm (with 40%-60%), before showing fluctuating and decreasing trends. There were absorption peaks observed near 1,465, 1,723, 1,935, 2,105, and 2,310 nm. The peak region near 1,465 nm was related to the first-order frequency doubling of the O–H bond stretching vibration (Aernouts *et al*., 2011). The characteristics of the *Lycium barbarum* spectral curve found in this study were consistent with the results of Mu *et al*. (2021) for the dried fruits of wolfberries from Ningxia, Qinghai, Gansu, and Xinjiang of China. In addition, similar conclusions were reported by Zhang *et al*. (2020) who used hyperspectral analysis and predicted the contents of total phenols, total flavonoids, and total anthocyanins in the dried fruits of black goji berry (*Lycium ruthenicum* Murr.) in the Ningxia and Xinjiang yield areas.

Due to the influence of light source intensity, noise, and instrument performance, the original spectral data need to be pre-processed before use. SG smoothing can effectively reduce noise. The smoothing effect is determined by the number of smoothing points. The eleven points of the SG first-order polynomial were used for smoothing. After smoothing, the random noise in the regions of 838-1,023 and 1,735-2,500 nm was significantly improved (Figure 3b). Then, the data were processed via first-order differentiation (Figure 3c), second-order differentiation (Figure 3d), multivariate scattering correction (Figure 3e), and standard normal variable transformation (Figure 3f). This different combination of pre-processing methods was able to reduce the heterogeneity of the spectral curves of different *Lycium barbarum* samples, enhance the quality of spectral data, and improve the modelling accuracy. Spectral differential transformation can eliminate baseline and background interference, distinguish overlapping peaks, and improve spectral resolution and sensitivity; MSC can correct the baseline drift of the diffuse reflectance spectrum and effectively suppress the noise caused by sample heterogeneity; SNV can eliminate the scattering effect caused by the uneven particle size and distribution of the sample.

Based on the reflection and absorption characteristics of the *Lycium barbarum* spectrum, the positions of the absorption peaks did not change compared to the original spectral curve after MSC and SNV processing. After first-order differential transformation processing, the characteristic absorption bands were more prominent. There were especially strong absorption peaks near 1,160, 1,353, 1,422, 1,557, 1,685, 1,898, 2,038, and 2,246 nm, as well as strong reflection peaks near 566 nm. Overall, the heterogeneity of the spectral curve was significantly improved.

Full-band modelling

Taking the transformed full-band spectral data as the independent variable and the measured data of *Lycium barbarum* nutrient content as the dependent variable, we established partial least squares regression estimation models. The accuracy of full-band PLSR models in reflecting the contents of eight nutrients under different pre-treatment combinations is shown in Table 2. Overall, the model predicted (RPD $\geq$ 1.4) TS and CP, while the other six nutrients could not be predicted. The spectral data smoothed by the SG filter offered better prediction accuracy for TS and CP. Specifically, the model prediction ability for TS was average (R^2p = 0.731, RPD = 1.827), while that for CP was strong (R^2p = 0.820, RPD = 2.087). Compared to the original spectral reflectance model, the RPD values of TS and CP increased by 55.25% and 4.60%, respectively.

Characteristic band selection

To reduce the complexity and enhance the interpretability of the model, we used variable importance measure (VIM) and the successive projection algorithm (SPA) to select the characteristic bands. The band distribution maps are shown in Figure 4. The red dots represent RPD $\geq$ 2.0, the blue dots represent 1.4 $\leq$ RPD < 2.0, and the black dots represent RPD < 1.4. According to the distribution of red dots and blue dots, among the eight nutrients in *Lycium barbarum*, the sensitive bands of organic matter were mainly found in the 355-390 nm range of violet light, the 455-487 nm range of blue light, the 533-558 nm range of green light, the 622-667 nm range of red light, the 784-994 nm range of near-infrared short-wave light, and the 1,119-1,363 and

1,881-2,485 nm range of near-infrared long-wave light. The sensitive bands of mineral elements were mainly observed in the 361-455 nm range of violet light, the 455-492 nm range of blue light, the 492-577 nm range of green light, the 577-597 nm range of yellow light, the 597-615 nm range of orange light, the 622-774 nm range of red light, and the 1,314-2,495 nm range of near-infrared long-wave light.

Table 2. Accuracy of the full-band PLSR model with different pre-treatments for the contents of eight nutrients in *Lycium barbarum*

Pre-treatments	Indices	TS[a]	LBP[a]	CP[a]	Mn[b]	Cu[b]	Fe[b]	Zn[b]	Ca[c]
OR	R^2p	0.628	0.242	0.816	0.169	0.516	0.331	0.689	0.414
	RMSEP	2.750	0.680	0.720	2.900	2.060	13.200	2.150	0.096
	RPD	1.177	0.949	1.995	0.301	0.951	0.964	0.405	1.117
SG	R^2p	0.731	0.264	0.820	0.169	0.540	0.331	0.689	0.415
	RMSEP	2.560	0.650	0.700	2.900	2.170	13.200	2.150	0.096
	RPD	1.827	0.868	2.087	0.301	0.854	0.964	0.405	1.118
SG-1stD	R^2p	0.610	0.138	0.563	0.092	0.399	0.468	0.297	0.377
	RMSEP	2.730	0.690	1.100	3.860	2.490	13.000	2.460	0.089
	RPD	1.412	0.679	1.106	0.285	0.464	0.993	0.925	0.755
SG-2ndD	R^2p	0.068	0.119	0.214	0.114	0.240	0.027	0.162	0.038
	RMSEP	4.990	0.670	1.470	3.110	2.650	18.200	2.700	0.130
	RPD	0.526	0.478	0.541	0.382	0.304	0.435	0.713	0.558
SG-MSC	R^2p	0.621	0.130	0.690	0.212	0.048	0.354	0.347	0.587
	RMSEP	2.820	0.700	1.020	3.020	3.010	13.700	2.540	0.076
	RPD	1.282	0.699	0.945	0.166	0.413	0.846	0.212	0.837
SG-SNV	R^2p	0.623	0.131	0.687	0.170	0.058	0.354	0.356	0.594
	RMSEP	2.810	0.670	1.020	3.050	3.000	13.600	2.530	0.078
	RPD	1.292	0.713	0.963	0.153	0.456	0.842	0.225	0.720

Notes: [a], [b] and [c] represent the units of '$g{\cdot}100g^{-1}$', '$mg{\cdot}kg^{-1}$' and '$g{\cdot}kg^{-1}$', respectively. R^2p and RPD are dimensionless

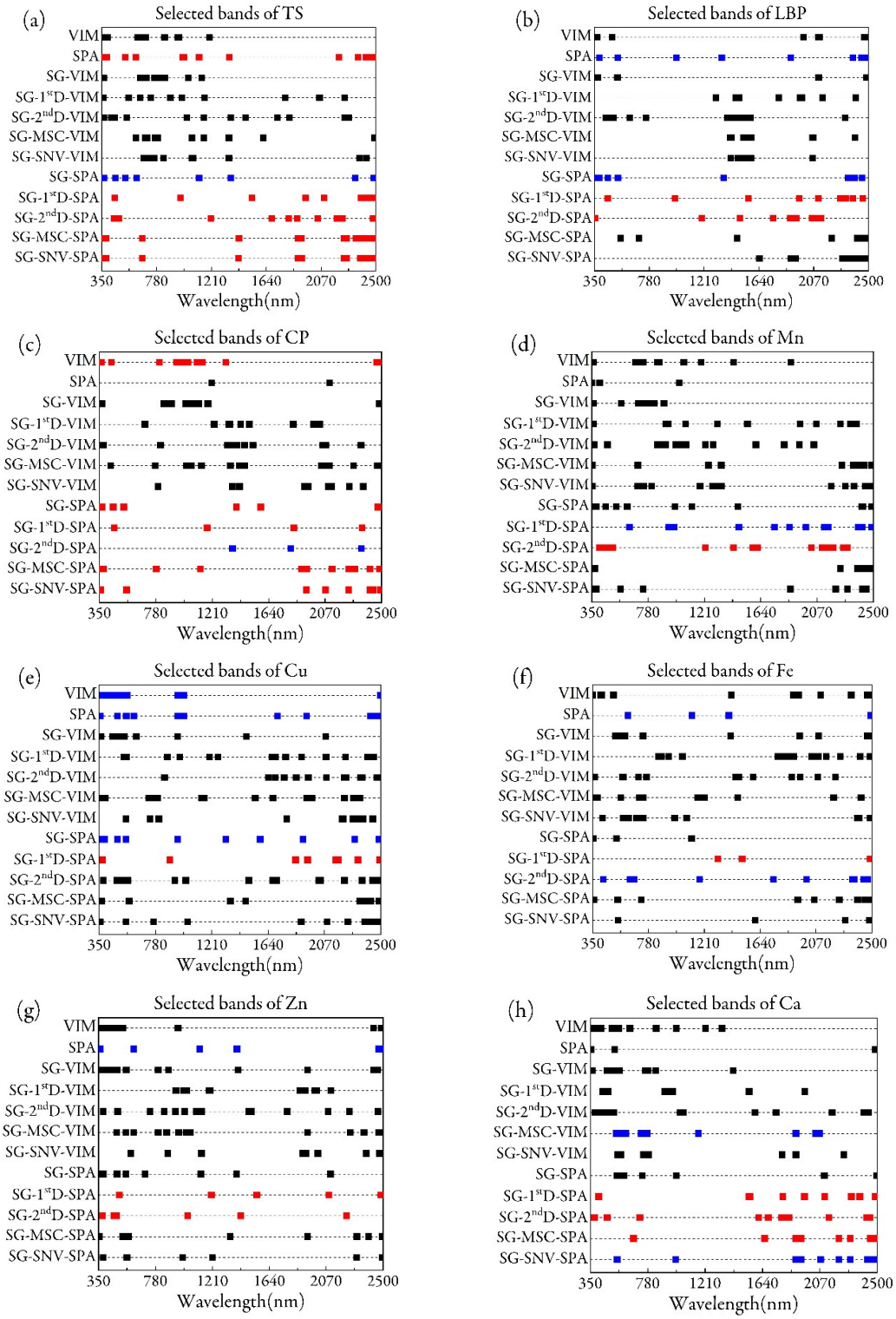

Figure 4. The characteristic bands of TS (a), LBP (b), CP (c), Mn (d), Cu (e), Fe (f), Zn (g), and Ca (h) in *Lycium barbarum* selected by different methods: VIM, SPA, SG-VIM, SG-1stD-VIM, SG-2ndD-VIM, SG-MSC-VIM, SG-SNV-VIM, SG-SPA, SG-1stD-SPA, SG-2ndD-SPA, SG-MSC-SPA, and SG-SNV-SPA

Optimal estimation models of nutrient contents

The RPD values for combinations of different pre-treatments and different characteristic variables for nutrient prediction models are shown in Figure 5. Compared to the full band estimation model (Table 2), after feature selection, the number of independent variables involved in modelling was significantly reduced, and the estimation performance was improved. The RPD values of TS and CP increased by 383.81% and 371.56%, respectively, while the RPD values of LBP, Mn, Cu, Fe, Zn, and Ca, which were unable to be predicted with the full band, reached more than 2.0 after selecting characteristic variables and showed strong estimation performance. In addition, compared to the VIM method, the overall accuracy was significantly improved after SPA band selection.

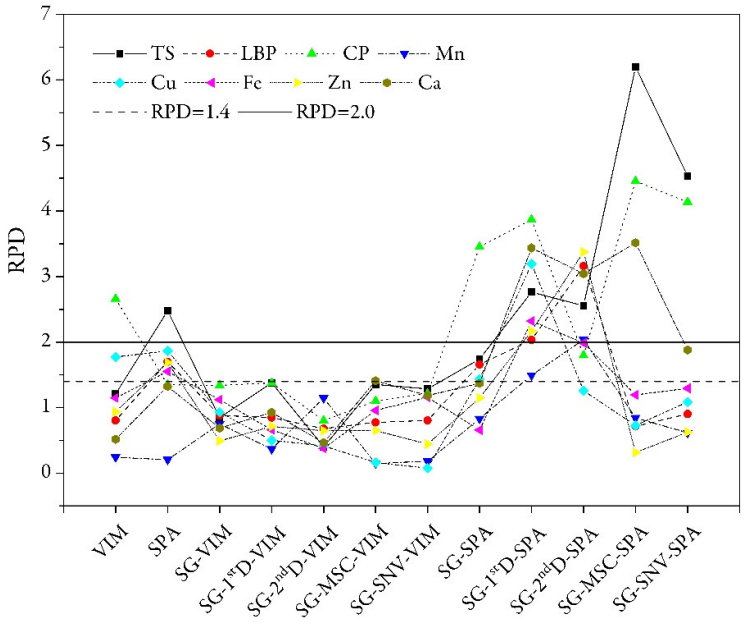

Figure 5. The RPD values of different pre-treatment combination models for nutrient contents

According to R^2p (Figure 6), RMSEP (Table 3) and RPD (Figure 5) values under the different pre-treatments and SPA characteristic band modelling, the spectral data indicate stronger response characteristics to LBP, Mn, Cu, Fe, and Zn after SG-1stD and SG-2ndD transformation. The R^2p was mainly concentrated at 0.763-0.938. Compared to the models without transformation, the RMSEP decreased by 14.88%-57.44%, and the RPD increased by 28.09%-890.51%. The spectral data after SG-MSC transformation were more sensitive to TS, CP, and Ca, with the R^2p mainly concentrated at 0.897-0.963. Compared to the models without transformation, the RMSEP decreased by 22.34%-63.95%, and the RPD increased by 3.10%-186.67%.

The model with the largest RPD value (Figure 5) was selected as the optimal estimation model for each nutrient content. The optimal estimation models for TS, LBP, CP, Mn, Cu, Fe, Zn, and Ca are shown in Figure 7, and the selected characteristic bands are shown in Table 4. The results showed that the optimal estimation model for TS was SG-MSC-SPA, with R^2p, RMSEP, and RPD values of 0.9751, 0.686, and 6.2, respectively. The most sensitive bands were 355-384 nm in violet light, 670 nm in red light, and 1,426 and 1,891-2,495 nm in near-infrared light. The optimal LBP estimation model was SG-2ndD-SPA, and its R^2p, RMSEP, and RPD values were 0.921, 0.206, and 3.16, respectively. The most sensitive bands were 357 nm in violet light and 1,193, 1,493, and 1,752-2,129 nm in infrared long-wave light. The optimal estimation model for CP was SG-MSC-SPA, and its R^2p, RMSEP, and RPD values were 0.954, 0.387, and 4.457, respectively. The most sensitive bands were 355-384 nm in violet light, 784 nm in near-infrared short-wave light, and 1,119 and 1,891-2,495 nm in near-infrared long-wave light.

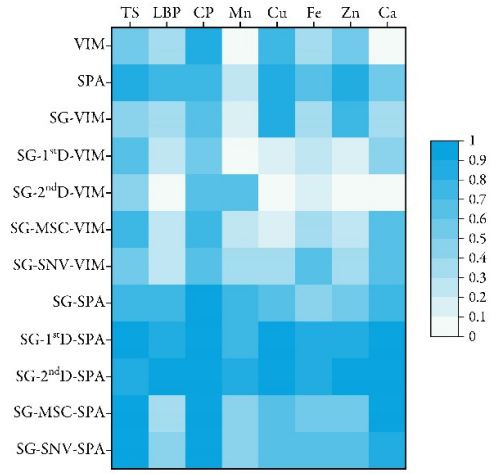

Figure 6. The R^2 of different pre-treatment combination models for nutrient contents

Table 3. The RMSEP under different pre-treatment combination models for nutrient contents

Pre-treatments	TS[a]	LBP[a]	CP[a]	Mn[b]	Cu[b]	Fe[b]	Zn[b]	Ca[c]
VIM	2.94	0.60	0.66	3.14	1.65	14.00	1.95	0.13
SPA	1.72	0.36	0.85	2.89	1.30	10.10	1.21	0.086
SG-VIM	3.21	0.59	0.96	2.96	1.68	14.10	2.00	0.093
SG-1stD-VIM	2.76	0.70	1.15	3.17	2.72	18.10	2.82	0.087
SG-2ndD-VIM	4.35	0.71	1.20	1.84	3.00	18.70	3.38	0.15
SG-MSC-VIM	2.15	0.73	0.96	3.02	2.89	14.10	2.36	0.074
SG-SNV-VIM	3.29	0.77	0.98	2.98	3.15	9.21	2.36	0.066
SG-SPA	2.14	0.40	0.47	2.45	1.80	11.70	1.97	0.067
SG-1stD-SPA	1.32	0.26	0.38	1.58	0.83	5.88	1.03	0.035
SG-2ndD-SPA	1.72	0.21	0.66	1.23	1.44	5.92	0.88	0.031
SG-MSC-SPA	0.69	0.58	0.39	2.18	1.98	10.20	2.33	0.037
SG-SNV-SPA	1.01	0.59	0.37	2.38	1.84	9.43	1.99	0.047

Notes: [a], [b] and [c] represent the units of 'g·100g^{-1}', 'mg·kg^{-1}' and 'g·kg^{-1}', respectively.

The optimal estimation model for Mn was SG-2ndD-SPA, and its R^2p, RMSEP, and RPD values were 0.869, 1.23, and 2.037, respectively. The most sensitive bands were relatively dispersed but mainly located in the 408 and 447 nm range of violet light; 475 nm range of blue light; 512 nm range of green light; and 1,219, 1,432-1,617, and 2,027-2,301 nm range of near-infrared long-wave light. The optimal estimation model for Cu was SG-1stD-SPA, and its R^2p, RMSEP, and RPD values were 0.938, 0.825, and 3.194, respectively. The most sensitive bands were found in the 375 nm range of violet light, 890 nm range of near-infrared short-wave, 1,850-2,491 nm range of near-infrared long-wave light. The optimal estimation model for Fe was SG-1stD-SPA, and its R^2p, RMSEP, and RPD values were 0.849, 5.88, and 2.322, respectively. The most sensitive bands were least numerous among the eight nutrients near the 1,314, 1,502, and 2,485 nm range of near-infrared long-wave light. The optimal estimation model for Zn was SG-2ndD-SPA, and its R^2p, RMSEP, and RPD values were 0.916, 0.881, and 3.376, respectively. The most sensitive bands were relatively dispersed but mainly located in the 380 nm range of violet light, 468 and 485 nm range of blue light, 1,024 nm range of near-infrared short-wave light, and 1,424 and 2,223 nm range of near-infrared long-wave light. The optimal estimation model for Ca was SG-MSC-SPA, and its R^2p, RMSEP, and RPD values were 0.948, 0.0366, and 3.515, respectively. The most sensitive bands were in the 674 nm range of red light and 1,656-2,495 nm range of near-infrared long-wave light.

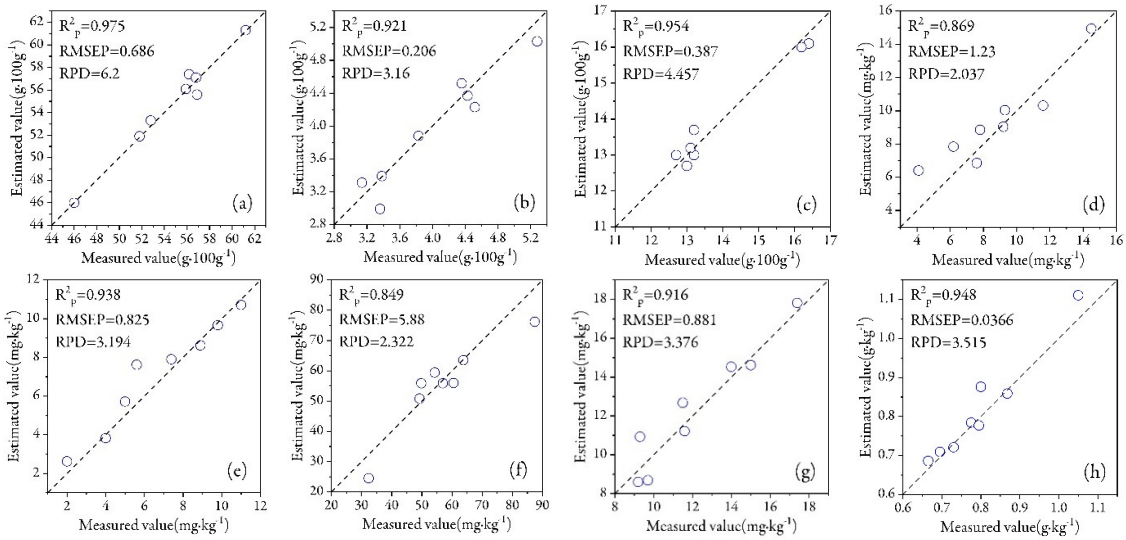

Figure 7. The optimal PLSR prediction model for each nutrient: (a) TS, (b) LBP, (c) CP, (d) Mn, (e) Cu, (f) Fe, (g) Zn, and (h) Ca

Table 4. Characteristic bands used in the optimal prediction model of each nutrient

Items	Optimal models	Optimal bands(nm)
TS	SG-MSC-SPA-PLSR	355, 361, 384, 670, 1,426, 1,891, 1,918, 2,250, 2,273, 2,341, 2,385, 2,412, 2,446, 2,483, 2,495
LBP	SG-2ndD-SPA-PLSR	357, 1,193, 1,493, 1,752, 1,884, 1,935, 2,055, 2,105, 2,129
CP	SG-MSC-SPA-PLSR	355, 361, 384, 784, 1,119, 1,891, 1,930, 2,123, 2,250, 2,300, 2,413, 2,485, 2,495
Mn	SG-2ndD-SPA-PLSR	408, 445, 475, 512, 1,219, 1,432, 1,581, 1,617, 2,027, 2,110, 2,160, 2,192, 2,270, 2,301
Cu	SG-1stD-SPA-PLSR	375, 890, 1,850, 1,943, 2,155, 2,177, 2,327, 2,491
Fe	SG-1stD-SPA-PLSR	1,314, 1,502, 2,485
Zn	SG-2ndD-SPA-PLSR	380, 468, 485, 1,024, 1,424, 2,223
Ca	SG-MSC-SPA-PLSR	674, 1,656, 1,892, 1,929, 2,213, 2,297, 2,446, 2,495

Discussion

Hyperspectral estimation of organic matter content

Although the basic trends for the spectral curves of *Lycium barbarum* from different yield areas were similar, genuine differences in nutrient contents still represented the main reason for changes of reflectance in the same band. Sugar plays an important role in achieving the medicinal quality of *Lycium barbarum*, with TS and LBP recognized as the most important medicinal contents. Several scholars have used near-infrared spectroscopy and PLSR to predict the contents of TS and LBP in *Lycium barbarum* in different regions of China. The R^2 values of the prediction set were 0.942 (Li *et al*., 2017) and 0.938 (Wang *et al*., 2016), respectively. Compared to previous studies, the prediction accuracy of TS was improved in our work by 3.5%, and the prediction accuracy of LBP was basically the same. Scholars have also evaluated the TS and LBP

contents in other crops via hyperspectral analysis and successfully predicted the TS contents of tobacco (R^2 = 0.67-0.86, RMSE = 4.28%-17.06%) (Soares *et al*., 2019), *Lolium perenne* L. (R^2 = 0.58, RMSE = 34.40 mg/g) (Shorten *et al*., 2019), and transverse sections of apple (R^2 = 0.81, RPD = 2.20) (Lan *et al*., 2021), as well as the LBP contents of chrysanthemum (R^2 = 0.93, RPD = 3.73) (He *et al*., 2018), licorice (R^2 = 0.9119, RMSEP = 0.4350) (Zhang *et al*., 2015), and lily (R^2 = 0.9455, RMSEP = 0.9098) (Huang *et al*., 2020). Compared to the above studies, our models showed better performance in TS and LBP. Crude protein is one of the most important indices for *Lycium barbarum* quality evaluation. Compared to previous hyperspectral studies on rice (R^2 = 0.917) (Liu *et al*., 2014) and forage grass (R^2 = 0.78) (Duranovich *et al*., 2020), the prediction accuracy of CP was improved by 4.0% and 22.31%, respectively, under our models. The spectral absorption characteristics in the near-infrared region might be related to the N–H and N=H bond vibrations in protein (Smith *et al*., 2020).

Under the action of visible near-infrared spectrum energy transfer, the frequency doubling absorption of the stretching vibrations of different combination modes of *Lycium barbarum* organic functional groups (such as C–H, C–O, O–H, and N–H) formed the fingerprint characteristics of spectral recognition. In previous studies, three main reasons were given for differences in the hyperspectral estimation performance of organic matter contents in crops: (1) each chemical substance has various forms—for example, the protein in chloroplast is not necessarily the same as that in cytoplasm (Curran, 1989); (2) most organic substances in crops, such as protein, starch, cellulose, and sugar, contain common molecular bonds (such as O–H and C–H), and the absorption peaks overlap; and (3) the spectral data quality can differ in various ways such as the strength of the light source, indoor and outdoor application scenarios, the materials of the dispersion components, and the spectral range and resolution of the detector.

Hyperspectral estimation of mineral element content

Mineral elements are indispensable components of enzymes, active proteins, vitamins, and hormones in the human body. These elements play an important biological role in the metabolism, growth, and development of the body (Shenkin, 2004). They also directly or indirectly affect the accumulation of secondary metabolites of Chinese medicinal materials (Konieczynski *et al*., 2016). The geographical indication of mineral elements is also an important indicator for the origin tracing of *Lycium barbarum*. Accurately acquiring the contents of mineral elements has important reference value for improving the quality of *Lycium barbarum*. Previous studies have shown that mineral nutrients do not absorb light energy in the visible near-infrared region (Manley, 2014). Therefore, visible near-infrared spectroscopy cannot be directly used to detect mineral content. However, the indirect prediction of inorganic content becomes possible when the inorganic content is combined with molecules containing covalent bonds (usually N–H, S–H, O–H, C–H, C–O, or C=C) (Manley, 2014; Ruano-Ramos *et al*., 1999).

Several scholars have predicted the mineral element contents in crops via hyperspectral analysis. For example, Kämper *et al*. (2020) predicted the contents of 14 mineral nutrients in Hass avocado fruit, finding that the Ca in flesh (R^2 = 0.53, RPD = 1.71) and skin (R^2 = 0.68, RPD = 1.57) offered the best performance. Bai *et al*. (2018) predicted the contents of Fe (R^2 = 0.75, RPD = 1.46), Mn (R^2 = 0.71, RPD = 1.84), and Zn (R^2 = 0.62, RPD = 1.37) in Canarium indicum nuts. Hu *et al*. (2021) predicted the contents of Ca (R^2 = 0.70) and Zn (R^2 = 0.77) in wheat kernels and flour. Osco *et al*. (2020) predicted the contents of Cu (R^2 = 0.861) and Zn (R^2 = 0.855) in Valencia orange leaves. Compared to previous studies, the prediction performance of Mn, Cu, Fe, Zn, and Ca improved in our work by 22.4%, 8.9%, 13.2%, 7.1%-47.7%, and 35.4%-78.9%, respectively.

Characteristic variable modelling

Due to the large amount of hyperspectral data, the high implied correlation between bands will lead to multicollinearity problems and distort the model. Therefore, determining how to select sensitive bands has long been the focus of hyperspectral inversion. In terms of modelling, we evaluated the spectral response characteristics of different nutrients in both full bands and characteristic bands. Because the group is selective in the absorption of the spectrum, not all spectra are highly correlated with the chemical indices being measured. Due to the excessive number of spectral variables and the presence of irrelevant spectral information, the robustness and accuracy of the full-band model were not ideal. By extracting characteristic bands, on the one hand, the information redundancy of independent variables was reduced, and the interpretability of the model was enhanced. On the other hand, these results have important reference value for the future development and popularization of production-oriented special equipment. In recent years, with the development of unmanned aerial vehicle (UAV) hyperspectral technology and deep learning methods, the development of chemometrics has ushered in a new era. With the coming era of artificial intelligence, we should give full play to the advantages of intelligent algorithms and high-performance instruments to facilitate the future development of hyperspectral technology and thereby improve the timeliness and accuracy of crop quality prediction.

Conclusions

In this study, we took Ningxia Hui Autonomous Region as the research area and the summer dried fruit of *Lycium barbarum* as the experimental materials. Hyperspectral remote sensing technology was used to obtain indoor spectral data for the samples. Two characteristic variable selection methods were used to extract the bands, and PLSR prediction models for the contents of eight nutrients were established. The conclusions of the study can be summarized as follows. (1) The original hyperspectral reflectance curves of *Lycium barbarum* from different yield areas and harvesting periods were similar in their overall variation trends. After first-order differential transformation, the heterogeneity of the spectral curve was improved. There were strong absorption peaks observed at the 1,160, 1,353, 1,422, 1,557, 1,685, 1,898, 2,038, and 2,246 nm range of near-infrared long-wave light. There was also a strong reflection peak near 566 nm for green light. (2) Among all nutrients, the inversion accuracy of the full band PLSR model was generally insufficient and could only predict the contents of TS and CP. However, the PLSR model established based on characteristic variables had RPD values above 2.0, indicating strong prediction performance. Compared to the VIM feature selection method, the overall accuracy of the PLSR prediction model established using SPA feature bands for each nutrient was improved. (3) Based on the PLSR model performance of different pre-treatment combinations, we found that the reflectance after SG-MSC-SPA conversion had the strongest prediction ability for TS, CP, and Ca, with RPD values of 6.2, 4.457, and 3.515, respectively. The reflectance after SG-2ndD-SPA conversion, however, had the strongest prediction ability for LBP, Mn, and Zn, with RPD values of 3.16, 2.037, and 3.376, respectively, and the reflectance after SG-1stD-SPA conversion had the strongest prediction ability for Cu and Fe, with RPD values of 3.194 and 2.322, respectively. The future work will focus on the real-time quality analysis method of *Lycium barbarum* based on hyperspectral remote sensing in the field. The combination of ground-based hyperspectral imaging, UAV and satellite remote sensing will further improve the quality monitoring efficiency and provide decision-making for the optimal harvest period of *Lycium barbarum*.

Authors' Contributions

Conceptualization, Jin-Long ZHAO and Xue-Yi ZHANG; Methodology, Jin-Long ZHAO; Software, Jin-Long ZHAO and Yun-Xia WANG; Validation, Jin-Long ZHAO, Qi ZHANG, Yun-Xia WANG and Yang LI; Formal analysis, Jin-Long ZHAO and Xue-Yi ZHANG; Investigation, Jin-Long ZHAO, Xue-Jun NAN and Qi ZHANG; Resources, Xue-Yi ZHANG; Data curation, Jin-Long ZHAO and Xue-Jun NAN; Writing - original draft preparation, Jin-Long ZHAO; Writing - review and editing, Jin-Long ZHAO, Xue-Yi ZHANG and Yang LI; Visualization, Jin-Long ZHAO and Yang LI; Supervision, Xue-Yi ZHANG and Yang LI; Project administration, Jin-Long ZHAO and Xue-Yi ZHANG; Funding acquisition, Xue-Yi ZHANG. All authors have read and approved the final manuscript.

Ethical approval (for researches involving animals or humans)

Not applicable.

Acknowledgements

This work was supported by The Special Fund of Chinese Central Government for Basic Scientific Research Operations in Commonweal Research Institutes, grant number IDM2018013; Ningxia Hui Autonomous Region Youth Top Talent Training Program, grant number RQ0033.

The authors would like to thank their team members and colleagues from the Ningxia Hui Autonomous Region Institute of Meteorological Sciences, Yinchuan, China, for drying wolfberries and spectral measurements.

Conflict of Interests

The authors declare that there are no conflicts of interest related to this article.

References

Aernouts B, Polshin E, Lammertyn J, Saeys W (2011). Visible and near-infrared spectroscopic analysis of raw milk for cow health monitoring: Reflectance or transmittance? Journal of Dairy Science 94(11):5315-5329. *https://doi.org/10.3168/jds.2011-4354*

Araújo MCU, Saldanha TCB, Galvão RKH, Yoneyama T, Chame HC, Visani V (2001). The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. Chemometrics and Intelligent Laboratory Systems 57(2):65-73. *https://doi.org/10.1016/S0169-7439(01)00119-8*

Bai SH, Tahmasbian I, Zhou J, Nevenimo T, Hannet G, Walton D, ... Wallace HM (2018). A non-destructive determination of peroxide values, total nitrogen and mineral nutrients in an edible tree nut using hyperspectral imaging. Computers and Electronics in Agriculture 151:492-500. *https://doi.org/10.1016/j.compag.2018.06.029*

Bellon-Maurel V, Fernandez-Ahumada E, Palagos B, Roger JM, McBratney A (2010). Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. Trends in Analytical Chemistry 29(9):1073-1081. *https://doi.org/10.1016/j.trac.2010.05.006*

Caporaso N, Whitworth MB, Fisk ID (2022). Prediction of coffee aroma from single roasted coffee beans by hyperspectral imaging. Food Chemistry 371:131159. *https://doi.org/10.1016/j.foodchem.2021.131159*

Chen JJ, Liu XL, Zhu LY, Wang Y (2013). Nuclear genome size estimation and karyotype analysis of *Lycium* species (Solanaceae). Scientia Horticulturae 151:46-50. *https://doi.org/10.1016/j.scienta.2012.12.004*

Chinese Pharmacopoeia Commission of the People's Republic of China (2020). Pharmacopoeia of the People's Republic of China (Part I). Beijing: China Medical Science Press, pp 260-261.

Curran PJ (1989). Remote sensing of foliar chemistry. Remote Sensing of Environment 30(3):271-278. *https://doi.org/10.1016/0034-4257(89)90069-2*

Duranovich FN, Yule IJ, Lopez-Villalobos N, Shadbolt NM, Draganova I, Morris ST (2020). Using proximal hyperspectral sensing to predict herbage nutritive value for dairy farming. Agronomy 10(11):1826. *https://doi.org/10.3390/agronomy10111826*

El-Sebaii AA, Shalaby SM (2013). Experimental investigation of an indirect-mode forced convection solar dryer for drying thymus and mint. Energy Conversion and Management 74:109-116. *https://doi.org/10.1016/j.enconman.2013.05.006*

Fatchurrahman D, Nosrati M, Amodio ML, Chaudhry MMA, de Chiara MLV, Mastrandrea L, Colelli G (2021). Comparison performance of visible-NIR and near-infrared hyperspectral imaging for prediction of nutritional quality of Goji berry (*Lycium barbarum* L.). Foods 10(7):1676. *https://doi.org/10.3390/foods10071676*

Galvão RKH, Pimentel MF, Araujo MCU, Yoneyama T, Visani V (2001). Aspects of the successive projection's algorithm for variable selection in multivariate calibration applied to plasma emission spectrometry. Analytica Chimica Acta 443(1):107-115. *https://doi.org/10.1016/S0003-2670(01)01182-5*

Grieco M, Schmidt M, Warnemünde S, Backhaus A, Klück HC, Garibay A, ... Pillen K (2022). Dynamics and genetic regulation of leaf nutrient concentration in barley based on hyperspectral imaging and machine learning. Plant Science 315:111123. *https://doi.org/10.1016/j.plantsci.2021.111123*

Han H, Guo XL, Yu H (2016). Variable selection using mean decrease accuracy and mean decrease Gini based on random forest. In: Proceedings of the 7th IEEE International Conference on Software Engineering and Service Science (ICESS). IEEE, Beijing, China pp 219-224. *https://doi.org/10.1109/ICSESS.2016.7883053*

He J, Chen LD, Chu BQ, Zhang C (2018). Determination of total polysaccharides and total flavonoids in *chrysanthemum morifolium* using near-infrared hyperspectral imaging and multivariate analysis. Molecules 23(9):2395. *https://doi.org/10.3390/molecules23092395*

Huang JH, Zhou RR, He D, Chen L, Yang YY, Xie HL, ... Huang LQ (2020). Rapid identification of *Lilium* species and polysaccharide contents based on near infrared spectroscopy and weighted partial least square method. International Journal of Biological Macromolecules 154:182-187. *https://doi.org/10.1016/j.ijbiomac.2020.03.109*

Hu NY, Li W, Du CH, Zhang Z, Gao YM, Sun ZC, ... Wang ZM (2021). Predicting micronutrients of wheat using hyperspectral imaging. Food Chemistry 343:128473. *https://doi.org/10.1016/j.foodchem.2020.128473*

Kämper W, Trueman SJ, Tahmasbian I, Bai SH (2020). Rapid determination of nutrient concentrations in Hass Avocado fruit by VIS/NIR hyperspectral imaging of flesh or skin. Remote Sensing 12(20):3409. *https://doi.org/10.3390/rs12203409*

Khodabux K, L'Omelette MSS, Jhaumeer-Laulloo S, Ramasami P, Rondeau P (2007). Chemical and near-infrared determination of moisture, fat and protein in tuna fishes. Food Chemistry 102(3):669-675. *https://doi.org/10.1016/j.foodchem.2006.05.057*

Konieczynski P, Arceusz A, Wesolowski M (2016). Essential elements and their relations to phenolic compounds in infusions of medicinal plants acquired from different European regions. Biological Trace Element Research 170(2):466-475. *https://doi.org/10.1007/s12011-015-0481-6*

Lan WJ, Jaillais B, Renard CMGC, Leca A, Chen SC, Le Bourvellec C, Bureau S (2021). A method using near infrared hyperspectral imaging to highlight the internal quality of apple fruit slices. Postharvest Biology and Technology 175:111497. *https://doi.org/10.1016/j.postharvbio.2021.111497*

Li Q, Yu XZ, Gao JM (2017). A novel method to determine total sugar of Goji berry using FT-NIR spectroscopy with effective wavelength selection. International Journal of Food Properties 20:S478-S488. *https://doi.org/10.1080/10942912.2017.1299759*

Liu MB, Li XL, Liu Y, Huang JF, Tang YL (2014). Detection of crude protein, crude starch, and amylose for rice by hyperspectral reflectance. Spectroscopy Letters 47(2):101-106. *https://doi.org/10.1080/00387010.2013.779283*

Manley M (2014). Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials. Chemical Society Reviews 43(24):8200-8214. *https://doi.org/10.1039/c4cs00062e*

Mao F, Xiao BX, Jiang Z, Zhao JW, Huang X, Guo JM (2011). Anticancer effect of *Lycium barbarum* polysaccharides on colon cancer cells involves G0/G1 phase arrest. Medical Oncology 28(1):121-126. *https://doi.org/10.1007/s12032-009-9415-5*

Mu QS, Kang ZL, Guo YJ, Chen L, Wang SY, Zhao YC (2021). Hyperspectral image classification of wolfberry with different geographical origins based on three-dimensional convolutional neural network. International Journal of Food Properties 24(1):1705-1721. *https://doi.org/10.1080/10942912.2021.1987457*

Osco LP, Marques Ramos AP, Faita Pinheiro MM, Saito Moriya ÉA, Imai NN, Estrabis N, ... Creste JE (2020). A machine learning framework to predict nutrient content in Valencia-Orange leaf hyperspectral measurements. Remote Sensing 12(6):906. *https://doi.org/10.3390/rs12060906*

Ouyang Q, Wang L, Park B, Kang R, Chen QS (2021). Simultaneous quantification of chemical constituents in matcha with visible-near infrared hyperspectral imaging technology. Food Chemistry 350:129141. *https://doi.org/10.1016/j.foodchem.2021.129141*

Potterat O (2010). Goji (*Lycium barbarum* and *L. chinense*): phytochemistry, pharmacology and safety in the perspective of traditional uses and recent popularity. Planta Medica 76(1):7-19. *https://doi.org/10.1055/s-0029-1186218*

Pu RL, Gong P (2003). Hyperspectral Remote Sensing and Its Applications. Beijing: Higher Education Press, pp 82-83.

Rossel RAV, McGlynn RN, McBratney AB (2006). Determing the composition of mineral-organic mixes using UV-vis-NIR diffuse reflectance spectroscopy. Geoderma 137(1-2):70-82. *https://doi.org/10.1016/j.geoderma.2006.07.004*

Ruano-Ramos A, García-Ciudad A, García-Criado B (1999). Near infrared spectroscopy prediction of mineral content in botanical fractions from semi-arid grasslands. Animal Feed Science and Technology 77(3-4):331-343. *https://doi.org/10.1016/S0377-8401(98)00245-4*

Shenkin A (2004). The role of minerals and trace elements in relation to long-term health and chronic disease. Nestlé Nutrition Workshop Series Clinical and Performance Programme 9:169-185. *https://doi.org/10.1159/000080662*

Shorten PR, Leath SR, Schmidt J, Ghamkhar K (2019). Predicting the quality of ryegrass using hyperspectral imaging. Plant Methods 15:63. *https://doi.org/10.1186/s13007-019-0448-2*

Smith C, Karunaratne S, Badenhorst P, Cogan N, Spangenberg G, Smith K (2020). Machine learning algorithms to predict forage nutritive value of in situ perennial ryegrass plants using hyperspectral canopy reflectance data. Remote Sensing 12(6):928. *https://doi.org/10.3390/rs12060928*

Soares FLF, Marcelo MCA, Porte LMF, Pontes OFS, Kaiser S (2019). Inline simultaneous quantitation of tobacco chemical composition by infrared hyperspectral image associated with chemometrics. Microchemical Journal 151:104225. *https://doi.org/10.1016/j.microc.2019.104225*

Tian YL, Xia T, Qiang X, Zhao YX, Li SP, Wang YM, ... Wang M (2022). Nutrition, bioactive components, and hepatoprotective activity of fruit vinegar produced from Ningxia wolfberry. Molecules 27(14):4422. *https://doi.org/10.3390/molecules27144422*

Wang YG, Gao Y, Yu XZ, Wang YY, Deng S, Gao JM (2016). Rapid determination of *Lycium barbarum* polysaccharide with effective wavelength selection using near-infrared diffuse reflectance spectroscopy. Food Analytical Methods 9:131-138. *https://doi.org/10.1007/s12161-015-0178-7*

Wang YM, Guo JP, Liu J, Yu F, Tian YM, Wu YJ, Liu LE (2021). Comparison and principal component analysis of main mineral elements in *Lycium ruthenicum* Murray from different habitats. Science and Technology of Food Industry 42(11):233-239 (in Chinese with English abstract). *https://doi.org/10.13386/j.issn1002-0306.2020070308*

Wold H (1966). Estimation of principal components and related models by iterative least squares. Multivariate Analysis 1:391-420.

Ye XJ, Abe S, Zhang SH (2020). Estimation and mapping of nitrogen content in apple trees at leaf and canopy levels using hyperspectral imaging. Precision Agriculture 21(1):198-225. *https://doi.org/10.1007/s11119-019-09661-x*

Yu HC, Wang RB, Yin Y, Liu YH (2017). Detection of polysaccharides and total sugar in Chinese wolfberry based on hyperspectral imaging in different wavebands. Food Science 38(8):191-197 (in Chinese with English abstract). *https://doi.org/10.7506/spkx1002-6630-201708030*

Yu L, Hong YS, Geng L, Zhou Y, Zhu Q, Cao JJ, Nie Y (2015). Hyperspectral estimation of soil organic matter content based on partial least squares regression. Transactions of the Chinese Society of Agricultural Engineering 31(14):103-109 (in Chinese with English abstract). *https://doi.org/10.11975/j.issn.1002-6819.2015.14.015*

Zhang CH, Yun YH, Fan W, Liang YZ, Yu Y, Tang WX (2015). Rapid analysis of polysaccharides contents in *Glycyrrhiza* by near infrared spectroscopy and chemometrics. International Journal of Biological Macromolecules 79:983-987. *https://doi.org/10.1016/j.ijbiomac.2015.06.025*

Zhang C, Wu WY, Zhou L, Cheng H, Ye XQ, He Y (2020). Developing deep learning based regression approaches for determination of chemical compositions in dry black goji berries (*Lycium ruthenicum* Murr.) using near-infrared hyperspectral imaging. Food Chemistry 319:126536. *https://doi.org/10.1016/j.foodchem.2020.126536*

Zou S, Zhang X, Yao WB, Niu YG, Gao XD (2010). Structure characterization and hypoglycemic activity of a polysaccharide isolated from the fruit of *Lycium barbarum* L. Carbohydrate Polymers 80(4):1161-1167. *https://doi.org/10.1016/j.carbpol.2010.01.038*