

Rapid estimation of soil heavy metal nickel content based on optimized screening of near-infrared spectral bands

Qian Lu¹ · Shijie Wang¹ · Xiaoyong Bai¹  · Fang Liu² · Shiqi Tian¹ · Mingming Wang¹ · Jinfeng Wang¹

Received: 6 March 2019/Revised: 16 May 2019/Accepted: 4 June 2019/Published online: 12 June 2019
© Science Press and Institute of Geochemistry, CAS and Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract In near-infrared spectroscopy, the traditional feature band extraction method has certain limitations. Therefore, a band extraction method named the three-step extraction method was proposed. This method combines characteristic absorption bands and correlation coefficients to select characteristic bands corresponding to various spectral forms and then uses stepwise regression to eliminate meaningless variables. Partial least squares regression (PLSR) and extreme learning machine (ELM) models were used to verify the effect of the band extraction method. Results show that the differential transformation of the spectrum can effectively improve the correlation between the spectrum and nickel (Ni) content. Most correlation

coefficients were above 0.7 and approximately 20% higher than those of other transformation methods. The model effect established by the feature variable selection method based on comprehensive spectral transformation is only slightly affected by the spectral transformation form. In five types of spectral transformation, the RPD values of the proposed method were all within the same level. The RPD values of the PLSR model were concentrated between 1.6 and 1.8, and those of the ELM model were between 2.5 and 2.9, indicating that this method is beneficial for extracting more complete spectral features. The combination of the three-step extraction method and ELM algorithm can effectively retain important bands associated with the Ni content of the soil. The model based on the spectral band selected by the three-step extraction method has better prediction ability than the other models. The ELM model of the first-order differential transformation has the best prediction accuracy ($R_p^2 = 0.923$, $RPD = 3.634$). The research results provide some technical support for monitoring heavy metal content spectrum in local soils.

✉ Xiaoyong Bai
baixiaoyong@vip.skleg.cn

Qian Lu
985479226@qq.com

Shijie Wang
wangshijie@vip.skleg.cn

Fang Liu
lfang6435@163.com

Shiqi Tian
1005794858@qq.com

Mingming Wang
1275166243@qq.com

Jinfeng Wang
490431493@qq.com

¹ State Key Laboratory of Environmental Geochemistry, Institute of Geochemistry, Chinese Academy of Sciences, 99 Lincheng West Road, Guiyang 550081, Guizhou Province, People's Republic of China

² College of Resource and Environment, Guizhou University, Guiyang 550000, China

Keywords Heavy metal · Band extraction · Partial least squares regression · Extreme learning machine · Near infrared spectroscopy

1 Introduction

Soil is the natural basis for human activities and survival and an important resource for agricultural production. Heavy metal nickel (Ni) is one of the essential micronutrients for tall plants and certain microorganisms. However, it is also a toxic element at high concentrations and causes serious harm to the environment and human health (Achary

et al. 2016; Pu et al. 2019). The traditional measurement of heavy metal in soils is accurate but expensive and time-consuming because it requires sample treatment. Rapid and cheap methods are necessary for predicting and monitoring changes in heavy metals in soils in specific areas and for overcoming the obstacles of traditional measurement. The hyperspectral inversion of the Ni content of the soil is a good method that is superior to laboratory analysis in terms of time and cost savings (Chakraborty et al. 2017).

Numerous researchers have explored and studied the spectral response of heavy metals and proven the feasibility of predicting heavy metals in soil by spectroscopic reflectance (Moros et al. 2009; Chakraborty et al. 2017). Several studies have suggested that the relationship between soil spectrum and heavy metal content may be supported by the nature of soil (Tang and Han 2017). The spectral response of heavy metals in soil and vegetation can be studied by correlation analysis for the identification of the most sensitive bands (Zhang et al. 2017). The soil spectra are weak and nonspecific because of the overlapping absorptions of some components. Hence, mathematical statistical methods are often needed to extract the spectral information of the soil (Wang et al. 2016). Moreover, variable extraction is important to the establishment of a model, where the prediction effect of the model will probably be affected. Several studies have proven that the selection of variables has many nonnegligible effects on the model (Jing et al. 2016). In the previous study, the band selection method was single and unexplained, and the spectral information contained was incomplete (Jiang et al. 2017). When using the multivariate stepwise regression method to establish a full-band model, the model only selects the relevant feature bands as modeling variables. The bands that are weakly related to heavy metals are ignored. The band is selected according to the mathematical selection rules of the variables, and the explanatory is weak (Zhang et al. 2015). When the regression model is built with the full band, the independent variable is excessively large and thereby generates data redundancy and affects model performance. The use of spectral reflectance and soil heavy metal correlation analysis results in a large number of significant correlation bands and thus increases the complexity of model processing and calculation and reduces the accuracy of the estimation of regression model parameters and the prediction of dependent variables (Ramachandran and Tsokos 2015). Therefore, an optimal spectral band selection method must be established. Traditional linear models have many shortcomings in considering the relationship between soil spectra and soil properties. Complex and stable models and methods are therefore needed to analyze this relationship. Many studies have used partial least squares regression (PLSR) models to estimate the heavy metal content of the soil. These methods can extract

variables with more characteristics (Wu et al. 2006; Pandit et al. 2010). In recent years, machine learning has attracted widespread attention and has been successfully applied to various issues such as regression and classification. Moreover, many studies have used extreme learning machines (ELM) or back propagation to perform hyperspectral predictions of heavy metal contents in soils (Shrivastava and Panigrahi 2014).

The present study took place at the Houzhai River Basin in Guizhou Province of China where research was done on the Ni content of the soil. Ni is among the eight key elements specified by the national standard for assessing heavy metal pollution in soils. First, the influence of variable selection method on the model is discussed. Then, the effects of ELM and PLSR algorithms on the model's effect are compared. The results have important scientific significance for the rapid and effective monitoring of soil content in karst areas.

2 Materials and methods

2.1 Research area

The Houzhai River Basin is located in Puding County, Guizhou Province, and is a typical plateau karst basin. The geographical position is 105°40'–105°49'E and 26°12'–26°18'N, with a total area of about 81 km². The terrain is high in the southeast and low in the northwest. It is open and flat, with an altitude of 1100–1400 m, belonging to the subtropical monsoon climate zone. The relative height difference is generally between 250 m and 300 m, with a maximum of 311 m. The Triassic carbonate rocks in the Houzhai area are widely distributed, while the underground river system in Houzhai is quite developed, with constant runoff all year round and is an important water supply source for the area. Figure 1 shows the location map of the study area.

2.2 Sample collection and measurement

The technical roadmap of this study is shown in Fig. 2. A total of 100 sampling points were set based on the topographic features, land cover types, and corresponding areas of each region. The samples included yellow earth, paddy soils, limestone soils. The sampling depth was 20 cm. The sampling site was divided into 4 layers such that each layer was 5 cm thick. In addition, 100 g of soil were collected per layer, and the 4 layers of soil samples were mixed to be the soil sample corresponding to the sample point. Sand and gravel were removed after the sample was dried. Each sample was divided into two sections and filtered using a 200 and a 100 mesh sieve, respectively. The obtained fine

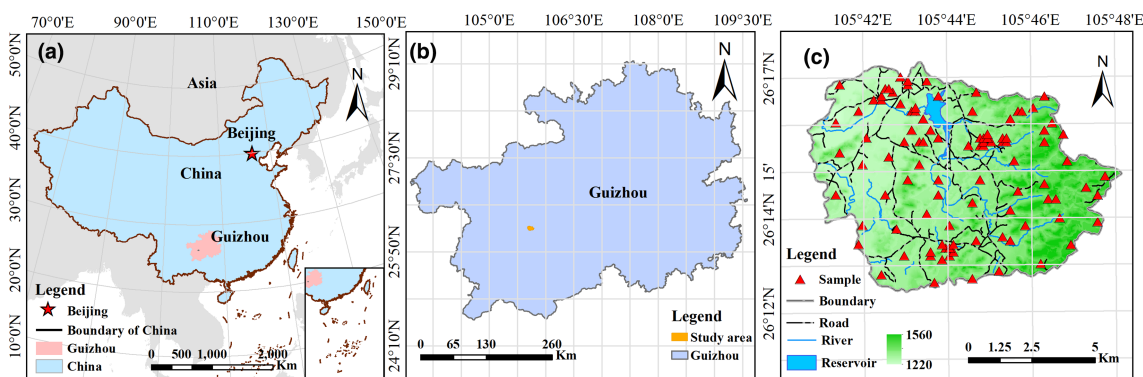


Fig. 1 Map of the study area

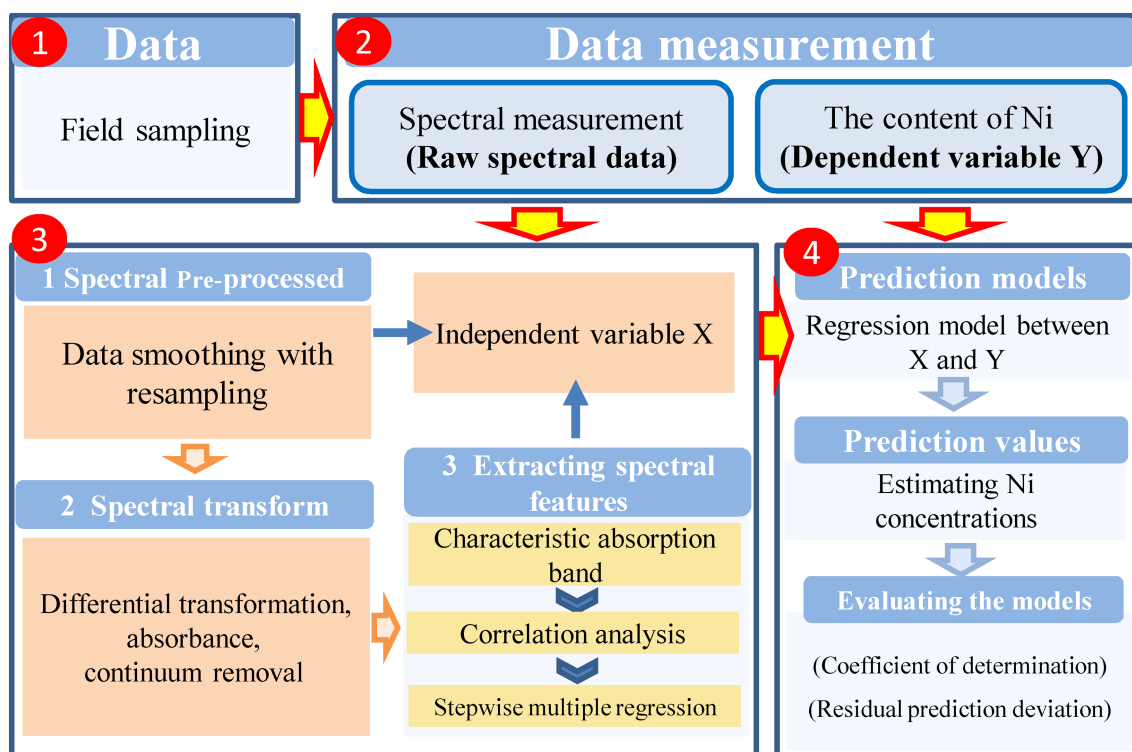


Fig. 2 Flowchart of the analysis

particles were used for heavy metal content measurement and spectral data measurement.

The soil samples were preprocessed by electric heating board acid (HNO₃–HCL–HClO₄–HF) digestion before the heavy metal concentrations were measured. The Ni content was measured by inductively coupled plasma mass spectrometry (ICP-MS, Thermo Electron, USA). The spectral data of the soil samples were measured using an ultraviolet–visible–near-infrared spectrophotometer covering a spectral range of 500–2500 nm. Each sample was tested 10 times in succession, and the 10 results were arithmetically averaged as the final spectral data of the soil sample. However, the error was inevitably affected by random

factors during the spectrometry. In order to reduce the error, the outliers were removed from the samples using The Unscrambler software. After removing 4 outliers, 96 samples remained.

2.3 Spectral pre-processing

Due to the interference of the spectrometer itself and external factors, there were many abrupt “burrs” in the shape of the obtained spectral curve. Savitzky–Golay convolution smoothing was adopted to remove the noise, and then resampled the smoothed spectra at an interval of 10 nm to reduce redundancy. The band reflectance

mentioned below is the reflectance after resampling, for example, the reflectance of the 500 nm band is the average of the 500–510 nm spectral data.

Spectral measurement is easily affected by many factors such as observation angle, illumination, and sample surface roughness. Low-frequency noise is also an important factor affecting the quality of spectral data. Pre-processing can eliminate the effects of baseline drift or smooth background interference and provides higher resolution and clearer spectral contour transformation than the original spectrum. On the basis of smoothing and resampling, the spectral data is further transformed into a differential transform, continuum removal, and reciprocal logarithmic transformation.

2.4 Establishment and validation of predication models

PLSR is a multiple regression analysis method that reveals the main control factor of the maximum change in content from the spectral data while reducing the spectral dimension (Wold et al. 2001). Thus, the established model exhibited better robustness. ELM is a new efficient learning algorithm (Guo et al. 2015). In our study, the excitation function of ELM was the “sig” function, and the number of best-hidden nodes of ELM was found by changing the number of nodes (the node was increased from 4 to 30 to increase 2). The results were unstable because of the random selection of the initial weight values of ELM. Therefore, each model structure was tested 600 times, and the number of hidden layer nodes was decided by considering the smallest average of iteration results. Finally, the best model was screened out. The models were built in Matlab2016.

The coefficient of determination (R_p^2), root mean square error (RMSEP) and residual prediction deviation (RPD) were used as standard methods for verifying the predictive model. The predictions were evaluated using the following criteria: RPD and R_p^2 values greater than 2.0 and 0.80 are good predictive; RPD values are between 1.5 and 2.0, and R_p^2 values are between 0.51 and 0.80, and approximate predictions are determined (Chang et al. 2001).

3 Results and discussion

3.1 Soil nickel content and spectral properties

The statistics of the samples are presented in Table 1. The Ni content in the study area was generally high. The Ni contents of 83 samples exceeded the soil background value of Guizhou Province (32.9 mg/kg). According to the

national soil quality standard (GB 15618-2018) risk screening value (100 mg/kg), the Ni contents of seven samples exceeded the standard, and the over-standard rate was 7.29%. In general, the risk of heavy metal pollution in the soil is low. The data from 96 samples were arranged from low to high Ni content, and one out of every three samples was used as a verification sample to test the accuracy of the model. Thus, 32 samples were verification samples. The remaining 64 models were used as modeling sample sets to ensure that the modeled samples were consistent with the scope of the validated samples and evenly distributed. The average and standard deviation of the modeled sample set and the test sample set were close. The distribution of the sample was balanced.

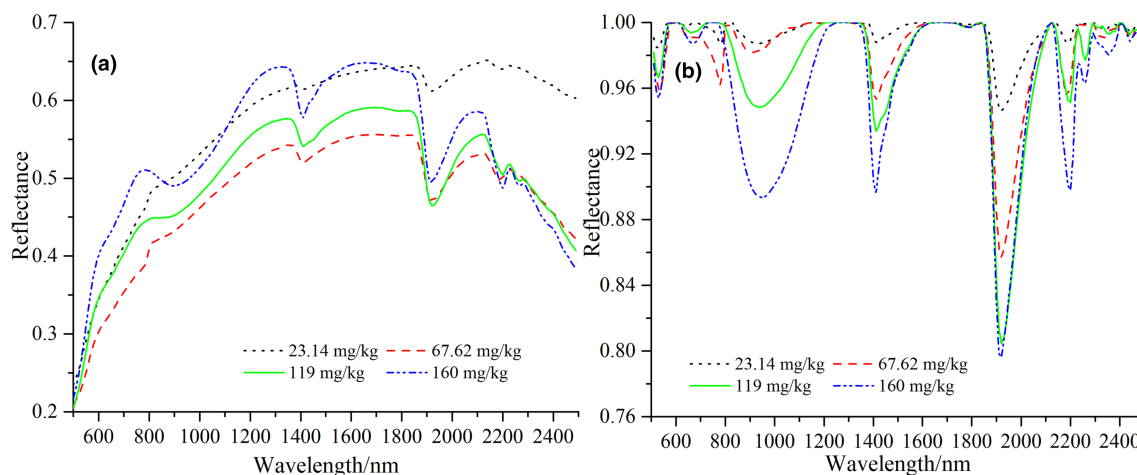
The original reflectance and continuum removal spectra of the representative samples of the four Ni concentration gradients were selected to plot trends (Fig. 3). The waveforms of the spectral curves were roughly the same; that is, the waveforms formed an upward parabolic shape (Fig. 3a). In the spectral range of 500–1300 nm, the curve rapidly rose. The curve changed gently after the wavelength exceeded 1300 nm. After the wavelength of 1800 nm, the Ni content of soil was higher than 67.62 mg/kg, and the spectral curve had a downward trend. The spectral curve of the 23.14 mg/kg sample was flat, and the change was small, but the absorption characteristics still exist. On this basis, the low-concentration sample had a weak response to the spectrum. The overall trends of the two samples at low concentrations was essentially the same (23.14 and 67.62 mg/kg). The samples with high concentrations (119 and 160 mg/kg) also had the same trend. The soil types of the two types of samples belong to lime soil and yellow soil. The overall change in the spectral curve was affected by the soil type. Studies have found that different soil components have different absorption levels. The adsorption of heavy metals by soil components provides an opportunity for estimating heavy metal content with the spectrum (Xu Binbin 1995).

As shown in Fig. 3b, the continuum removal transformation can effectively suppress the influence of the background light and enhance the absorption characteristics of the spectrum, such as absorption characteristics at wavelengths of 900, 1900, and 2200 nm. Three distinct absorption peaks appeared around 1400, 1900, and 2200 nm and are generally associated with kaolinite in clay minerals (Clark et al. 1990). Meanwhile, the spectral absorption characteristics of vermiculite were mainly at 1400 and 1900 nm. Small absorption peaks were present in the vicinity of the 500, 800, and 1000 nm bands. These absorption peaks were affected by the concentrations of heavy metals. The absorption characteristics became increasingly evident with increasing Ni content. Studies have shown that iron ions in the oxide cause some features

Table 1 Descriptive statistics of the Ni concentrations (mg/kg)

Set	Number of samples	Max	Min	Mean	SD	Skewness	Kurtosis	CV
Calibration	64	160	23.14	53.14	26.94	1.99	4.55	0.51
Validation	32	143	24.61	53.01	26.35	1.90	4.06	0.50
All	96	160	23.14	53.10	26.60	1.93	4.13	0.50

SD standard deviation, CV coefficient of variation

**Fig. 3** Spectral curves of four Ni concentration samples: **a** original reflectance spectrum and **b** continuum removal spectrum

in the spectral ranges of 500–600, 610–800, and 810–1240 nm (Shi et al. 2013). On this basis, the spectral characteristic absorption band is in the ranges of 500–580, 610–1240, 1320–1650, 1830–2100, and 2140–2500 nm.

3.2 Three-step extraction method of wavebands

The five characteristic absorption bands of the spectrum are determined through the continuum removal transformation of spectral reflectance. On the basis of the correlation analysis between the reflectivity of these bands and the Ni content of the soil, the correlation between the Ni content of the soil and the spectral reflectance is taken as the standard for the elimination of weak correlation bands and retention of significant correlation bands. Then, the multivariate stepwise regression model is used to eliminate the variables in the remaining bands that have no significant influence on the dependent variable (the Ni content of the soil). Finally, the characteristic bands of various transformed forms of the spectrum are obtained. The stepwise regression method can gradually select variables in accordance with the contribution values to screen the model variables with the best regression effect. The multivariate stepwise regression method selects and rejects the probability *P* value of the *F* statistic. In this method, the entry value is set to 0.05, and the removal value is set to 0.1. The criterion for correlation is a two-tailed test with *P* value less than 0.01.

As shown in Fig. 4, many bands in the near-infrared spectral range are strongly correlated with the Ni content of the soil. Spectral transformation can improve the correlation between the spectrum and the Ni content of the soil. The first-order differential (Fig. 4b) and second-order differential (Fig. 4c) of the spectral reflectance are in the interval between the two bands of 610–1240 nm and 1320–1650 nm, respectively. The correlation has a significant impact, and the correlation coefficient is significantly improved. The correlation coefficient between the original spectrum and the transformed spectrum is high in the intervals 183–2100 nm and 2140–2500 nm. Therefore, the bands in these two intervals play an important role in predicting the Ni content of the soil. A significant correlation exists between the spectral reflectance in the range of 500–580 nm and the Ni content of the soil. Heavy metal adsorption experiments show that soil organic matter and clay minerals have a certain retention capacity for Ni. The spectral characteristics of clay minerals are around the 1900, 2200, and 1400 nm bands. The 570 and 660 nm bands are important to the prediction of soil organic matter (Rossel et al. 2006). These bands, which have been shown to be correlated with the Ni content of the soil, are included in the selected five characteristic absorption bands. However, the possibility of using other bands in the absorption bands of these features remains unknown. Therefore, further screening is needed. By using the abovementioned methods, the characteristic bands corresponding to

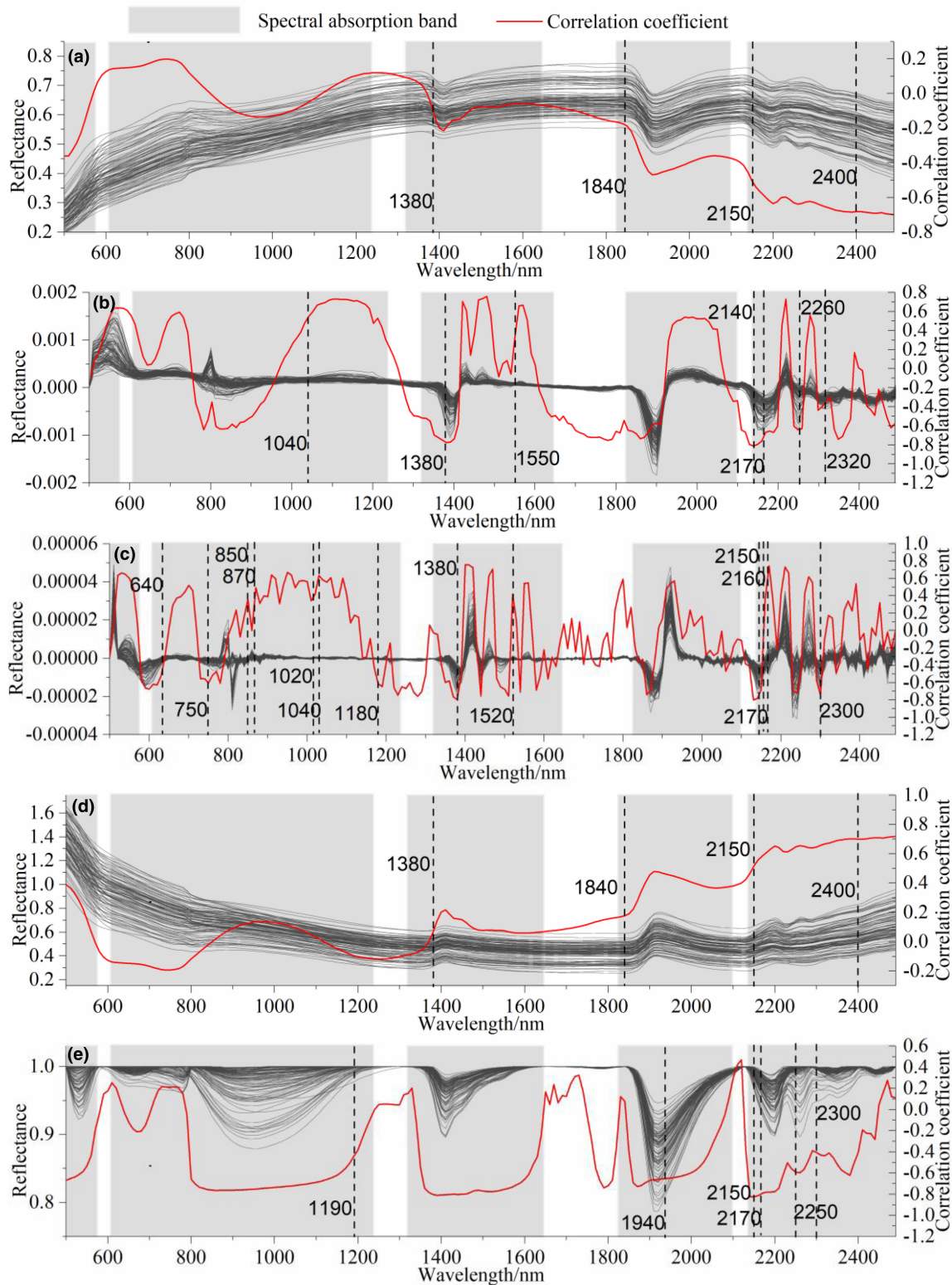


Fig. 4 The correlation coefficient between spectral reflectance of soil and Ni content and its spectral transformation. **a** Raw spectra, **b** first differential, **c** second differential, **d** absorbance transformation and **e** continuum removal

different spectral transformations are extracted separately. The bands are indicated by dashed lines in Fig. 4. Each characteristic band corresponds to a high correlation

coefficient value, and most of the bands correspond to the valleys of the spectral curve. On this basis, these characteristic bands are representative.

3.3 Influence of the band extraction methods on the performances of the Ni models

3.3.1 Models based on three-step extraction method

In accordance with the spectral bands selected by the different methods, the Ni content prediction model was established by PLSR and ELM methods, and the prediction accuracy of the model was compared. Table 2 shows that the Ni content was predicted with the highest precision by models established using the derivatives of the spectrum data. These preprocessing techniques can effectively eliminate noise and enhance the signal-to-noise ratio of the original spectrum without destroying its characteristics.

The model established by the significantly correlated band is better than the model established by the absorption band alone. Furthermore, the highly correlated band can effectively explain the relationship between the spectral reflectance and the Ni content of the soil. The prediction accuracy of the model can reach 0.7–0.8. The stability of the model established by the ELM method is greater than 2. Therefore, these models can roughly estimate the Ni content of the soil in the study area. However, as a whole, the stability of the model established by PLSR was poor. Meanwhile, the model established by CR spectral transformation had a low effect of only 0.569, and the error was large. On the basis of the defects of the two band selection methods, namely, the characteristic absorption band and the significant correlation band, a comprehensive three-step extraction method was proposed. As shown in Table 2, the band selected by the three-step extraction method contributes to the improvement of model accuracy and stability. PLSR and ELM methods have good prediction effects. The accuracy of the model R_p^2 established by PLSR is around 0.7, and the accuracy of the PLSR model established by CR spectral transformation reaches a good

level ($R_p^2 = 0.856$, $RMSEp = 9.856$, $RPD = 2.674$). In the ELM model, the three spectral transformations of FD, SD, and CR show a good performance. The FD-ELM model has the best performance ($R_p^2 = 0.923$, $RMSEp = 7.253$, $RPD = 3.634$) and accurately predicts the Ni content of the soil. The three transformation modes of FD, SD, and CR can enhance the characteristic information of the original spectrum. By combining the three-step variable extraction method, the spectral characteristic bands related to the Ni content of soil can be extracted completely, and an accurate prediction model can be established.

Figure 5 shows the comparison of the estimated values of the best PLSR and ELM models on the basis of the different variable selection methods. The modeling effect improves as the sample point line graph becomes close. Compared with the model SD-PLSR (Fig. 5a) and the SD-PLSR (Fig. 5c), the SD-PLSR (Fig. 5c) predicts Ni more accurately (Fig. 5f). The predicted value of the FD-ELM model is close to the measured value, and the fitting effect is the best. The highest R_p^2 and the lowest $RMSEp$ value was obtained ($R_p^2 = 0.923$, $RMSEp = 7.253$, $RPD = 3.634$). The results in Fig. 5 and Table 2 are mutually validated. The accuracy and stability of the model established by the ELM method are generally better than the model established by the PLSR method. The reason is that the ELM model not only achieves good accuracy but also maintains fast training speeds and responsiveness. By disregarding the iterative steps, the ELM can be trained quickly and a successful training model can be achieved.

3.3.2 Comprehensive spectral transformation selects the band

The characteristic bands selected by different spectral transformations were combined into a comprehensive feature variable set; that is, 22 bands, namely,

Table 2 The parameters of Ni content prediction models based on different feature variables selection methods

Model	Pre.p	Spectral absorption band			Significantly correlated band			Three-step extraction method		
		R_p^2	RMSEp	RPD	R_p^2	RMSEp	RPD	R_p^2	RMSEp	RPD
PLSR	SR	0.558	18.509	1.424	0.713	16.464	1.601	0.692	15.798	1.668
	FD	0.767	16.100	1.637	0.655	16.762	1.572	0.734	15.261	1.727
	SD	0.696	14.998	1.757	0.745	13.892	1.897	0.785	12.252	2.151
	AT	0.576	18.239	1.445	0.742	16.076	1.639	0.713	15.429	1.705
	CR	0.568	17.038	1.547	0.569	17.017	1.549	0.856	9.856	2.674
ELM	SR	0.764	12.748	2.067	0.807	12.026	2.192	0.850	10.831	2.433
	FD	0.815	11.404	2.311	0.865	11.177	2.358	0.923	7.253	3.634
	SD	0.796	11.740	2.245	0.844	11.216	2.350	0.912	7.757	3.397
	AT	0.774	12.683	2.078	0.818	11.481	2.296	0.842	10.930	2.411
	CR	0.769	12.953	2.035	0.809	11.517	2.288	0.911	7.975	3.305

Pre.p pre-processing, SR raw spectra, FD first derivative, SD second derivative, AT absorbance transformation, CR continuum removal

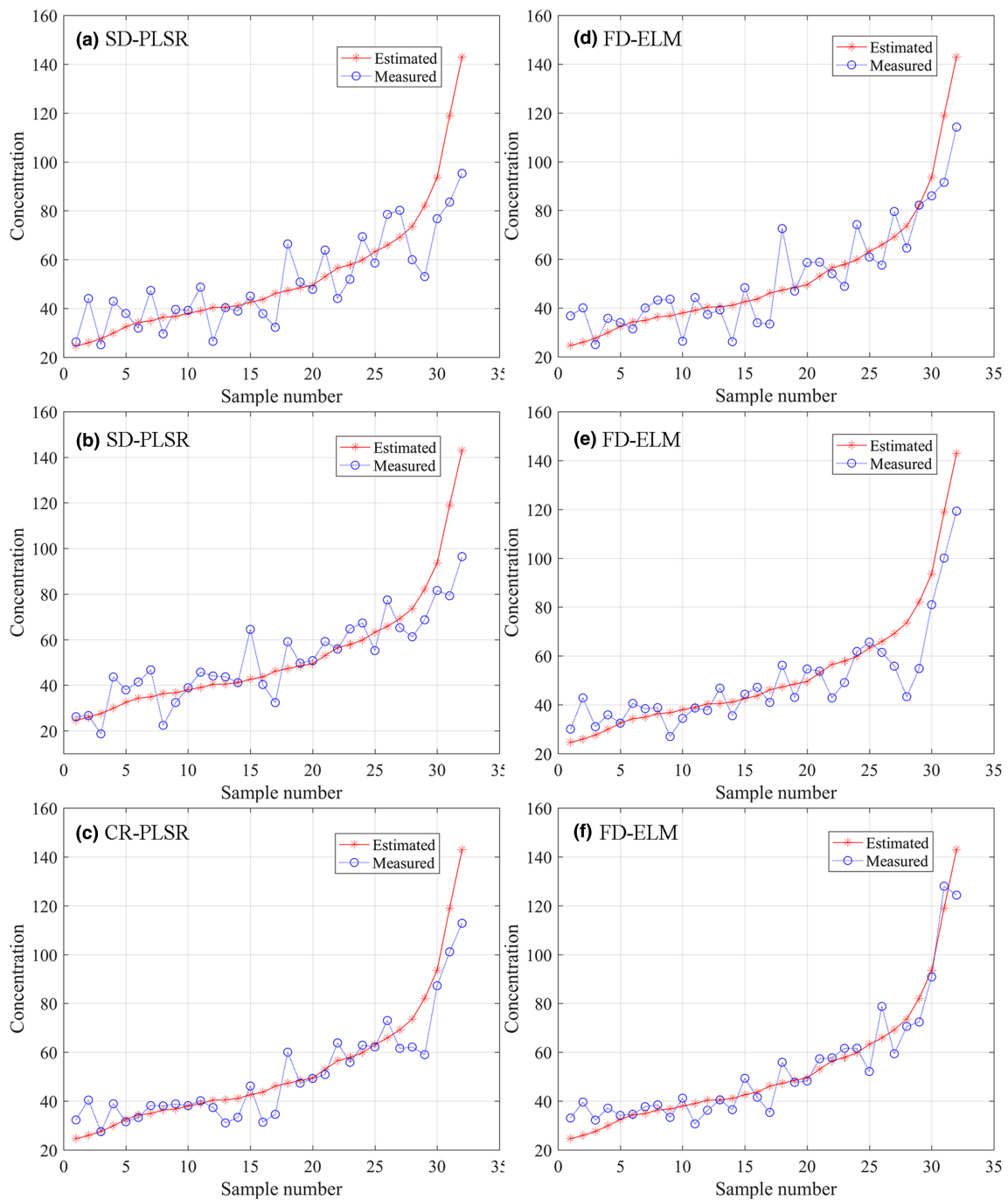


Fig. 5 Scatterplots of measured versus predicted nickel (Ni) concentrations of soil: **a, d** spectral absorption band; **b, e** significantly correlated band; **c, f** three-step extraction method

640, 750, 850, 870, 1020, 1040, 1180, 1190, 1380, 1520, 1550, 1840, 1940, 2140, 2150, 2160, 2170, 2250, 2260, 2300, 2320, and 2400 nm were selected. The models were established by PLSR and ELM methods with their corresponding reflectance as independent variables and soil Ni content as the dependent variable. The lowest RMSEP and maximum value of R_p^2 were the main criteria for calculating the models in this study.

As shown in Table 3, the results of the PLSR and ELM models have similar trends. The SD-PLSR had a high R_p^2 value, that is, 0.727 (Table 3). The FD-ELM model had good predictable results and the R_p^2 and RMSEP values were 0.877 and 9.114, respectively. In sum, the effect of the model established by the comprehensive spectral bands set under different spectral transformation conditions is

Table 3 The parameters of Ni content prediction models based on the method of selecting the feature variables of integrated spectral transformations

Model	EI	Extraction method of integrated spectral transformation				
		RS	FD	SD	AT	CR
PLSR	R_p^2	0.702	0.703	0.727	0.715	0.731
	RMSEp	15.906	15.923	14.285	15.596	15.214
	RPD	1.657	1.655	1.845	1.690	1.732
ELM	R_p^2	0.854	0.877	0.868	0.855	0.865
	RMSEp	10.897	9.114	9.659	10.311	9.595
	RPD	2.519	2.892	2.729	2.556	2.747

EI evaluation indicator

better than those of the other models, and the variation range is small. The RPD values of the PLSR model were concentrated between 1.6 and 1.8. The RPD values of the ELM model were between 2.5 and 2.9. Different spectral transformations have less influence on the accuracy of the model and are generally at the same level. Therefore, in terms of model promotion ability, the ELM model based on the comprehensive spectral transformation method has a good effect, that is, the prediction model can perform well without the mathematical transformation of the original spectrum. Of course, FD-ELM model has the best overall effect among the models.

The prediction model established by the comprehensive spectral band is weakly affected by the spectral transformation form. The prediction ability of the model is affected by the modeling method, and the spectral transformation is not dominant. These results are obtained possibly because the primary function of the spectral transformation is to amplify unapparent spectral features in the original spectrum to select characteristic spectral bands. The comprehensive band set is selected on the basis of various spectrally transformed spectral bands and contains most of the spectral features. Therefore, the model established using the comprehensive feature variable set based on comprehensive spectral transformation is almost unaffected by the spectral transformation mode. The differences in performance among the five models in Table 3 are small.

3.4 Evaluating the models

The traditional method of band selection is single, and its interpretation is weak (Choe et al. 2008). Therefore, some insignificant variables should be discharged, and an optimal subset of independent variables should be established. However, if the regression model is established using the least relevant or few bands, then other characteristic bands may be lost, and the established model might become unstable (Liu et al. 2011). As shown in Tables 2 and 3, the model based on the three-step extraction method generally has a better effect than the model based on the traditional

single-band selection method. For example, the accuracy of the PLSR model changes clearly. Hence, a large amount of spectral band information can be obtained by transforming the spectrum. The corresponding feature band information selected by a single method cannot effectively express the relationship between spectral reflectance and heavy metal content. The resulting model has weak predictive ability and low precision. As shown in Table 2, the PLSR model with the differential transformation of reflectance and absorbance as the independent variable has the best effect, and the fitting effect of the model is above 0.7. The prediction ability of the ELM model established by using reflectance and absorbance is poor. However, the prediction ability of the ELM model established by other spectral transformation forms is good. Hence, the differential transformation of the spectrum contributes to the improvement of the predictive ability of the model. As shown in Table 3, for the integrated spectral transformation method, although the influence of the spectral transformation is not obvious, the prediction effect of the model is approximately 0.7–0.8. Compared with the models established by the spectral absorption band and the significantly correlated band, the PLSR and ELM models have improved the model prediction ability. The selected bands after the comprehensive spectral transformation contain complete information. Thus, the excess bands are eliminated in the stepwise regression process, thereby leaving significant spectral information correlated with the heavy metal content. The different spectral transforms can highlight the different characteristic bands, and the number is limited. Therefore, the bands selected by combining multiple spectral transforms are more representative and can explain more spectral information than the other bands.

Regardless of the band selection method used to build the model, the prediction performance of the ELM model is always superior to that of the PLSR model. The advantages of the ELM can be explained by the ability of neural networks in handling nonlinear relationships between metal elements and spectra. Similar results that showed the best performance of ELM models have been reported (Chen

et al. 2015). The comparison of the results of the two models showed that the ELM model not only has the best accuracy but also has the fastest speed. As shown in Tables 2 and 3, the RPD values of the ELM model are all above 2.0. Thus, the model can predict the concentration of heavy metals in the soil well. In addition, as shown in Fig. 5, the predicted value of the ELM model is closer to the measured value, especially in the low concentration range. This phenomenon is probably due to the fact that the ELM algorithm with high generalization ability avoided many problems, such as learning rate and local minima (Zhu et al. 2005). ELM can be fast in the training model because no iterative step is required.

4 Conclusions

This article proposes a method for screening characteristic bands by combining with spectral features, a correlation coefficient method, and a stepwise regression method. On the basis of the reflectivity of the different forms of spectral transformation, the Ni content prediction model based on different variable selection methods was established by using PLSR and ELM methods, and the prediction ability of the model was verified. A comprehensive set of bands was formed by synthesizing the characteristic bands corresponding to the five spectral transformations. Finally, the influence of the comprehensive spectral transformation on the model was discussed. The following conclusions can be drawn:

More spectral band information can be obtained by spectral transformation processing. The spectral band of the differential transformation (namely, first and second differentials) is significantly correlated with Ni content. For example, at 1380 nm, the correlation coefficient is above 0.750. Therefore, the differential transformation can effectively acquire the spectral sensitivity band of the soil.

The prediction accuracy of the model based on the feature variable selection method based on comprehensive spectral transformation is less affected by the spectral transformation form. The RPD values of the PLSR model are concentrated between 1.6 and 1.8. Meanwhile, the RPD values of the ELM model are between 2.5 and 2.9, at the same level. The model change trend is stable under different spectral transformation modes. Hence, the variable selection method of the comprehensive spectral transformation is beneficial to extract complete features.

The model established based on the spectral band selected by the three-step extraction method has the best prediction ability. Meanwhile, the ELM model can improve the stability and accuracy of the prediction. The ELM model under the first-order differential

transformation has the best precision (R_p^2 is between 0.82 and 0.94), which is better than the prediction model based on the PLSR method. Therefore, the combination of the three-step extraction method and the ELM algorithm can effectively retain the important bands associated with the Ni content of the soil, and the model prediction effect is the best ($R_p^2 = 0.923$, RPD = 3.634).

Acknowledgements This research work was supported jointly by the National Key Research Program of China (Nos. 2016YFC0502102, 2016YFC0502300), “Western light” talent training plan (Class A), Chinese academy of science and technology services network program (No. KFJ-STX-ZDTP-036) and international cooperation agency international partnership program (Nos. 132852KYSB20170029, 2014-3), Guizhou high-level innovative talent training program “ten” level talents program (No. 2016-5648), United fund of karst science research center (No. U1612441), International cooperation research projects of the national natural science fund committee (Nos. 41571130074, 41571130042), Science and Technology Plan of Guizhou Province of China (No. 2017–2966).

References

- Achary MS, Satpathy KK, Panigrahi S, Mohanty AK, Padhi RK, Biswas S et al (2016) Concentration of heavy metals in the food chain components of the nearshore coastal waters of Kalpakkam, southeast coast of India. *Food Control* 72:232–243
- Chakraborty S, Weindorf DC, Deb S, Li B, Paul S, Choudhury A et al (2017) Rapid assessment of regional soil arsenic pollution risk via diffuse reflectance spectroscopy. *Geoderma* 289:72–81
- Chang C, Laird DA, Mausbach MJ, Chang CW et al (2001) Near infrared reflectance spectroscopy-principal components regression analyses of soil properties. *Soil Sci Soc Am J* 65(2):480–490
- Chen T, Chang Q, Clevers JGPW et al (2015) Rapid identification of soil cadmium pollution risk at regional scale based on visible and near-infrared spectroscopy. *Environ Pollut* 206:217–226
- Choe E, Meer FVD, Ruitenbeek FV, Werff HVD, Smeth BD, Kim KW (2008) Mapping of heavy metal pollution in stream sediments using combined geochemistry, field spectroscopy, and hyperspectral remote sensing: a case study of the rodalquilar mining area, se Spain. *Remote Sens Environ* 112(7):3222–3233
- Clark RN, King TVV, Klejwa M, Swayze GA, Vergo N (1990) High spectral resolution reflectance spectroscopy of minerals. *J Geophys Res Solid Earth* 95(B8):12653–12680
- Guo W, Shang L, Zhu X, Nelson SO (2015) Nondestructive detection of soluble solids content of apples from dielectric spectra with ann and chemometric methods. *Food Bioprocess Technol* 8(5):1126–1138
- Jiang Z, Yang Y, Sha J (2017) Application of GWR model in hyperspectral prediction of soil heavy metals. *Acta Geogr Sin* 72(3):533–544
- Jing T, Wei HZ, Yun NZ et al (2016) Spectral response and inversion models for prediction of total copper content in soil of Xifanping mining area. *Spectrosc Spectr Anal* 36:3637–3642
- Liu Y, Li W, Wu G, Xu X (2011) Feasibility of estimating heavy metal contaminations in floodplain soils using laboratory-based hyperspectral data: a case study along Le’an river, china. *Geospat Inf Sci* 14(1):10–16
- Moros J, Vallejuelo FOD, Gredilla A, Diego AD, Guardia MDL (2009) Use of reflectance infrared spectroscopy for monitoring the metal content of the estuarine sediments of the Nerbioi-

- Ibaizabal river (Metropolitan Bilbao, Bay of Biscay, Basque Country). *Environ Sci Technol* 43(24):9314–9320
- Pandit CM, Filippelli GM, Li L (2010) Estimation of heavy-metal contamination in soil using reflectance spectroscopy and partial least-squares regression. *Int J Remote Sens* 31(15):4111–4123
- Pu W, Sun J, Zhang F et al (2019) Effects of copper mining on heavy metal contamination in a rice agrosystem in the Xiaojiang River Basin, southwest China. *Acta Geochim*. <https://doi.org/10.1007/s11631-019-00321-5>
- Ramachandran KM, Tsokos CP (2015) Chapter 8: linear regression models. *Math Stat Appl R* 9(1–3):409–458
- Rossel RAV, Walvoort DJJ, Mcbratney AB, Janik LJ, Skjemstad JO (2006) Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* 131(1–2):59–75
- Shi T, Chen Y, Liu Y, Wu G (2013) Visible and near-infrared reflectance spectroscopy-an alternative for monitoring soil contamination by heavy metals. *J Hazard Mater* 265C(2):166–176
- Shrivastava NA, Panigrahi BK (2014) A hybrid wavelet-elm based short term price forecasting for electricity markets. *Int J Electr Power Energy Syst* 55:41–50
- Tang Y, Han G (2017) Characteristics of heavy metals in soils under different land use in a typical karst area, southwest China. *Acta Geochim* 03:165–168
- Wang J, Tiyp T, Ding J, Zhang D, Liu W (2016) Estimation of desert soil organic carbon content based on hyperspectral data preprocessing with fractional differential. *Trans Chin Soc Agric Eng* 32(21):161–169
- Wold S, Sjöström M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. *Chemom Intell Lab Syst* 58(2):109–130
- Wu Y, Chen J, Ji J, Gong P, Liao Q, Ma H (2006) Visible and near infrared reflectance spectroscopy for measuring soil heavy metal content as a quick method. *Acta Geochim* 25(s1):225–226
- Xu Binbin LDX (1995) A preliminary study on identification of clay minerals in soils with reference to reflectance spectra. *Pedosphere* 5(2):135–142
- Zhang L, Wu Y, Liu S, Zhou L, Cai D (2015) Research on monitoring of Pb content in the sewage irrigation area of Longkou city using hyperspectral data. *Environ Eng* 33(8):114–117 (in Chinese)
- Zhang Q, Zhang H, Liu W, Zhao S (2017) Inversion of heavy metals content with hyperspectral reflectance in soil of well-facilitated capital farmland construction areas. *Trans Chin Soc Agric Eng* 33(12):230–239
- Zhu QY, Qin AK, Suganthan PN, Huang GB, Patcog J (2005) Evolutionary extreme learning machine. *Pattern Recogn* 38(10):1759–1763