




# Hyperspectral estimation model of soil Pb content and its applicability in different soil types

Shiqi Tian<sup>1,2,3</sup> · Shijie Wang<sup>2,3</sup> · Xiaoyong Bai<sup>2,3</sup>  · Dequan Zhou<sup>1</sup> · Qian Lu<sup>2,3</sup> · Mingming Wang<sup>2,3</sup> · Jinfeng Wang<sup>2,3</sup>

Received: 24 June 2019 / Revised: 10 October 2019 / Accepted: 11 December 2019  
© Science Press and Institute of Geochemistry, CAS and Springer-Verlag GmbH Germany, part of Springer Nature 2019

**Abstract** In order to obtain Pb content in soil quickly and efficiently, a multivariate linear regression (MLR) and a principal component regression (PCR) Pb content estimation model were established on the basis of hyperspectral techniques, and their applicability in different soil types was evaluated. Results indicated that Pb exhibited strong spatial heterogeneity in the study area, and more than 82% of the samples exceeded the background value. In addition, the pollution range was large. Pb was sensitive in the near-infrared band, and the correlation of absorbance (AB) was most significant of all the transformed forms. Both models achieved optimal stability and reliability when AB was used as an independent variable. Compared with the PCR model, the stability, fitting accuracy, and predictive power of the MLR model were superior with a coefficient of determination, root mean square error, and mean relative error of 0.724%, 24.92%, and 28.22%, respectively. Both models could be applied to different soil types; however, MLR had better applicability compared with PCR. The PCR model that distinguished different soil types had better reliability than one that did not. Thus, the model

established via hyperspectral techniques can achieve large-area, rapid, and efficient soil Pb content monitoring, which can provide technical support for the treatment of heavy metal pollution in soil.

**Keywords** Hyperspectral data · Heavy metal · Pb · Estimation

## 1 Introduction

A large amount of domestic and industrial wastes have been discharged into soils. Chemical fertilizers are also used in agriculture. Therefore, the soils accumulate heavy metal Pb, which enters the human body through the food chain and other means and poses a serious threat to human health (Li et al. 2013a; Song et al. 2017). Therefore, the estimation of soil Pb content is crucial.

Traditionally, the determination of heavy metal content in soil requires extensive sampling and long-term laboratory measurement analysis. This method has high measurement accuracy but is costly in terms of time and economy and has low efficiency; monitoring the spatial distribution information of large-area and continuous elements is difficult and may damage the soil environment (Ren et al. 2009; Song et al. 2015; Wang et al. 2018). Hyperspectral technology can effectively estimate heavy metal content in soil in a nondestructive manner while saving substantial time and money and considerably improve estimation efficiency (Gholizadeh et al. 2015; Guo et al. 2015; Tao et al. 2016). Some scholars have achieved the indirect inversion of heavy metal content by using the intrinsic relationship between heavy metals and soil components, such as organic matter, iron manganese oxide, and clay minerals (Sun et al. 2018; Lu et al. 2019). Although

✉ Xiaoyong Bai  
baixiaoyong@vip.skleg.cn

<sup>1</sup> School of Geography and Environmental Sciences, Guizhou Normal University, Guiyang 550081, Guizhou Province, People's Republic of China

<sup>2</sup> 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

<sup>3</sup> Puding Karst Ecosystem Observation and Research Station, Chinese Academy of Sciences, Puding 562100, Guizhou Province, People's Republic of China

good results have been obtained, the degree of heavy metal adsorption by these soil components is often affected by differences in heavy metal elements and soil conditions. Therefore, the applicability of the model in studies of other elements or other soil conditions has to be examined further. Existing heavy metal prediction models are divided into linear and nonlinear models. Some scholars have predicted heavy metal Pb content based on nonlinear models, such as neural networks (Tian et al. 2019), random forests (Tan et al. 2020), and support vector machines (Tan et al. 2014). Although the results are satisfactory, the construction of these models often requires programming, which is difficult and prone to overfitting, thus resulting in inaccurate results. Other scholars have estimated Pb content using linear models, such as multiple linear regression (MLR), multiple linear stepwise regression, and partial least squares regression (Tan et al. 2014; Zhang et al. 2014; Yao et al. 2014; Maliki et al. 2014), which have also achieved the desired results. Moreover, the linear model is relatively easy to develop, simple to calculate, and easy to operate.

The current study on soil Pb content focuses on non-karst ecosystems; the soil distribution in karst areas is discontinuous, and spatial heterogeneity is strong (Zhang et al. 2010), making studying soil Pb content difficult. Therefore, research on soil Pb content in karst areas is relatively lacking. In addition, several problems arise in the application of hyperspectral technology in a karst region. On the one hand, the hyperspectral data commonly used today come from three types of remote sensing satellite imagery and field and laboratory spectrometers. The terrain of a karst area is complex and diverse, and rocks are exposed to nakedness. The low signal-to-noise ratio of the sensor and the complex error source may affect the quality of the field spectrum and remote sensing image, making effective soil spectral information difficult to obtain and current research strongly dependent on laboratory spectra. Moreover, spectral data based on laboratory analysis are prone to errors due to external factors and defects in the instrument itself. Therefore, preprocessing the soil spectrum and optimizing the hyperspectral signal are necessary to improve prediction accuracy. Whether the model established by indoor spectral data can be applied to field and remote sensing images remains unclear due to the regional uniqueness of soil and the uncertainty of the field environment. On the other hand, current research is either conducted on a single soil type or does not consider the soil type, and the applicability of the model under different soil types has yet to be studied.

In view of the above limitations, this study took the Houzhai River Watershed in Guizhou Province as a case study. On the basis of the measured Pb content and hyperspectral data, an MLR and a principal component

regression (PCR) estimation model was built. Moreover, the applicability of these two models in different soil types was evaluated. The aim was to explore the possibility of using hyperspectral techniques to estimate Pb content in karst areas and provide new ideas for the rapid monitoring of heavy metal elements in the soil of other ecosystems.

## 2 Materials and methods

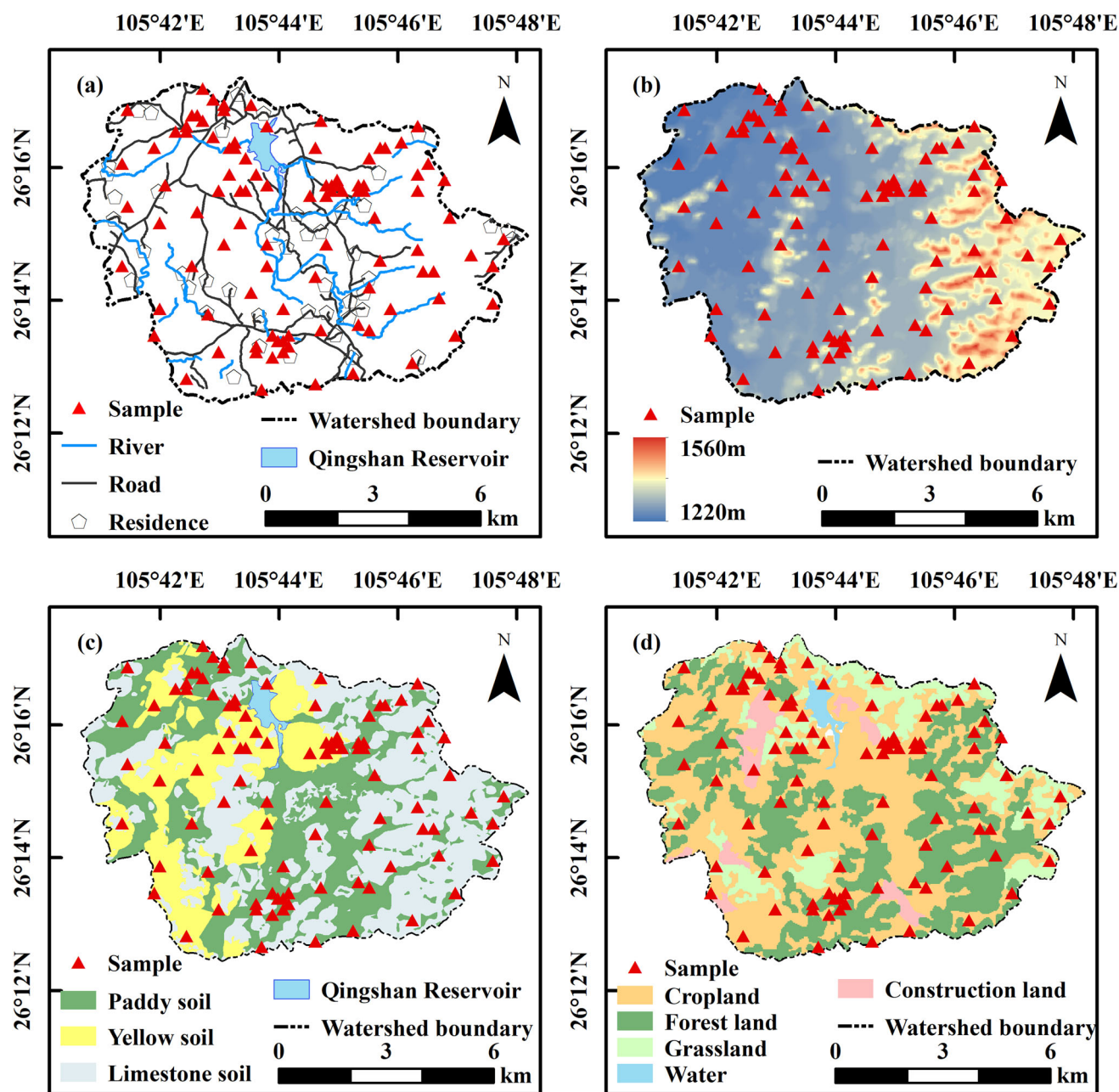
### 2.1 Study area and soil samples

Located in the west of the Qianzhong Plateau, the Houzhai River Watershed has a well-developed karst landform, which is the watershed between the Yangtze and the Pearl River systems (Li et al. 2013a, b). Its latitude and longitude ranges are 105° 41' E–105° 48' E and 26° 13' N–26° 17' N, respectively. The elevation is 1220–1560 m and the watershed area is approximately 75 km<sup>2</sup>. The terrain in the watershed is high in the southeast and low in the northwest. Land use includes forest land, cropland, grassland, construction land, and waters. The upper reaches are mainly forest land, whereas the middle and lower reaches are mainly cropland. The watershed has many soil types, including limestone, paddy, and yellow soils, and the spatial distribution pattern is complex (Zhang et al. 2018).

A total of 98 soil samples of different soil types weighing approximately 1 kg were collected in the watershed. The sampling depth was 0–20 cm, and the range was wide (Fig. 1). Each sample was divided into two parts after natural air-drying and removal of impurities. The samples were ground and filtered through a 200-mesh nylon sieve. One part was used for chemical analysis, and the other was for spectral analysis.

### 2.2 Determination of soil Pb content and spectral data

The soil samples were subjected to microwave digestion with hydrochloric–nitric–perchloric acid, and Pb content was determined via inductively coupled plasma mass spectrometry (PerkinElmer, Canada). Quality control was performed using the standard samples (AGV-2 = 12.191 mg/kg; AMH-1 = 9.822 mg/kg; GBPG = 13.054 mg/kg) to ensure the quality of the analysis. Spectral determination of soil samples was performed in the laboratory using a Cary 5000 UV–Vis–NIR spectrophotometer (Agilent Technologies, USA). The band ranged from 500 to 2500 nm, the sampling interval was 1 nm, and three spectra were collected for each soil sample. Unscrambler software was used to perform principal component analysis on 98 samples, thus eliminating two outliers; the average of the spectral reflectance of 96



**Fig. 1** Overview of the study area and distribution of samples. **a** distribution of samples, **b** elevation, **c** soil types and **d** land use types

samples was taken as the original reflectance spectral value (Xu et al. 2011; Li et al. 2017).

### 2.3 Data processing and analysis

The spectrometry process was susceptible to errors due to random factors. In contrast to the number of samples, the process had many variables, and the overlap of information between adjacent bands caused spectral data redundancy and some interference to the data analysis. The original spectral data were preprocessed via Savitzky–Golay

smoothing, resampling (RE), continuum removal (CR), reflectance first derivative (RFD), reflectance second derivative (RSD), absorbance (AB), first derivative of absorbance (AFD), and second derivative of absorbance (ASD), thereby effectively reducing the influence of noise, purifying the spectral information, and reducing errors (Fig. 2) (Shi 2014; Peng et al. 2014; Ma et al. 2016; Wang et al. 2018; Tu et al. 2018).

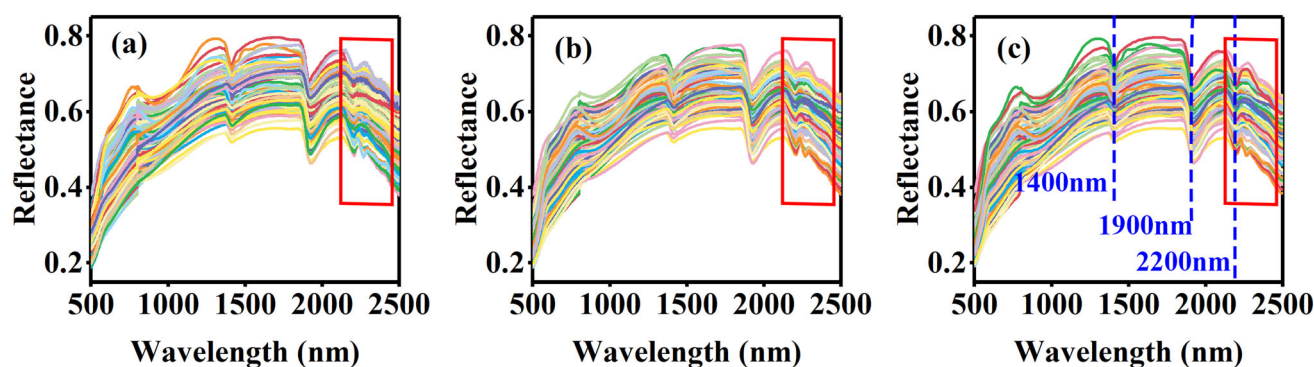


Fig. 2 Spectral data smoothing and resampling. **a** Original spectral curve, **b** spectral smoothing and **c** spectral resampling

## 2.4 Model building and accuracy evaluation

The soil samples were sorted in accordance with the Pb content in descending order. One out of every three samples was used as a validation sample (a total of 32, accounting for 33.3% of the total samples), and the remaining samples were used for calibration (a total of 64, accounting for 66.7% of the total samples). The correlation between Pb content and spectral data after RE, RFD, RSD, CR, AB, AFD, and ASD was analyzed, and the feature bands were screened. With the feature bands and soil Pb content taken as the independent and dependent variables, respectively, the MLR and PCR models were built. The applicability of both models in different soil types was evaluated. The model results were evaluated using the coefficient of determination  $R^2$  and root mean square error (RMSE). The larger the  $R^2$ , the better the stability of the model; the smaller the RMSE, the higher the accuracy of the model. The mean relative error (MRE) and 1:1 line were used as criteria for judging the estimation capability of the model. The smaller the MRE, the closer the sample point was to the 1:1 line, and the stronger the estimation capability of the model. The calculation formulas for  $R^2$ , RMSE, and MRE are as follows:

$$R^2 = \frac{\sum_{i=1}^n (Y_p - \bar{Y})^2}{\sum_{i=1}^n (Y_m - \bar{Y})^2}, \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_i (Y_m - Y_p)^2}, \quad (2)$$

$$MRE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_m - Y_p}{Y_m} \right| \times 100, \quad (3)$$

where  $Y_p$  is the predicted value,  $Y_m$  is the measured value,  $\bar{Y}$  is the average value of the measured values, and  $n$  is the number of samples.

## 3 Results

### 3.1 Soil Pb content analysis

The calibration and validation samples are similar in average and standard deviation and the range is relatively uniform and balanced (Table 1). The total sample set has a coefficient of variation of 0.60. The distribution of Pb in the soil is not uniform, showing significant spatial heterogeneity. Seventy-nine out of the 96 samples exceeded the national standard natural background value (China National Environmental Monitoring Center 1990), that is, 82% of the sample content exceeds the standard, and the total average value is 1.8 times the national background value. This result indicates that the soil in the Houzhai River Watershed is heavily polluted by heavy metal Pb, and corresponding treatment measures must be taken. Soil types comprise 27 yellow, 23 paddy, and 46 limestone soils. The overall content of Pb in yellow soil is the highest, followed by limestone and paddy soils. In terms of land use, the content of grassland is the highest, followed by cropland (Fig. 3).

### 3.2 Feature bands of heavy metal

The “burr” after smoothing by Savitzky–Golay and RE is significantly reduced, and the spectral information is significantly improved (Fig. 2). The spectral curve of the soil shows an upward trend. Three significant absorption bands are observed at approximately 1400, 1900, and 2200 nm, and the degree of absorption is slightly different. This result is consistent with the study of Teng et al. (2016).

The spectral characteristics of the soil in the study area are the same, and the spectral curves of different soil types are slightly different (Fig. 4). The average absorbance of different soil types follows the order paddy soil > yellow soil > limestone soil (Fig. 4d). The average reflectance of soil samples follows the order limestone soil > yellow soil > paddy soil. Studies have shown that the use of

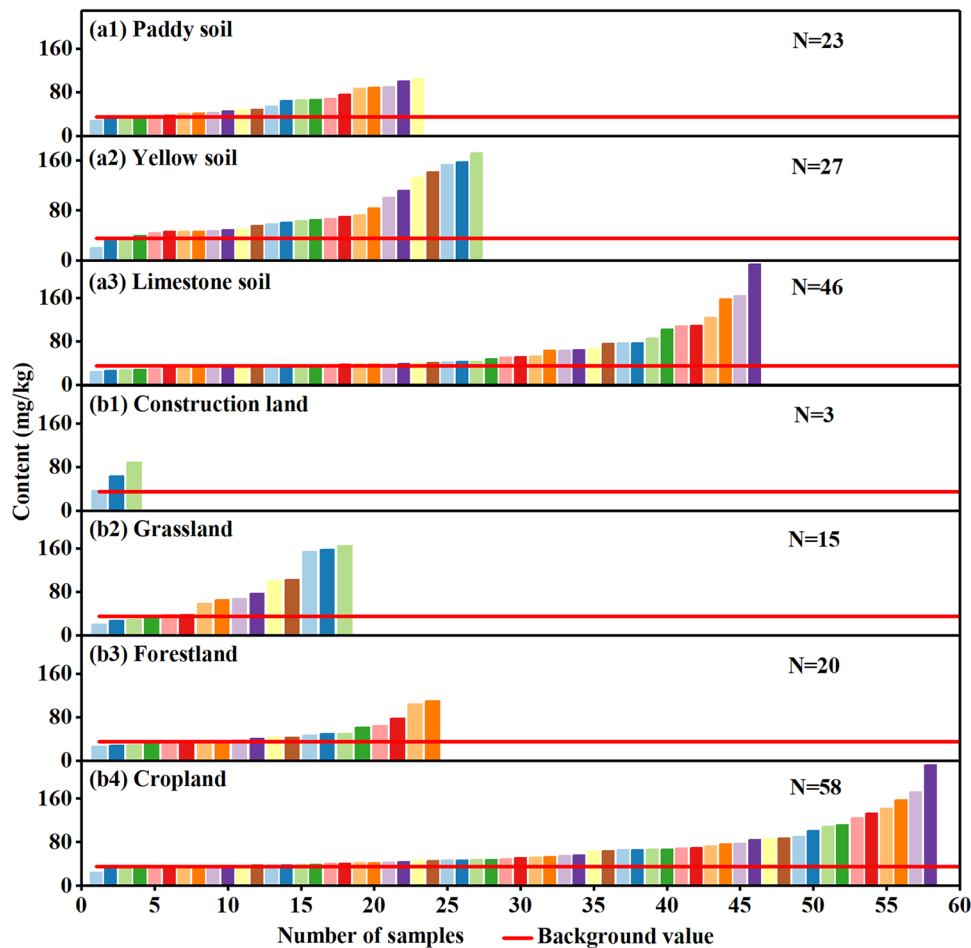


**Table 1** Descriptive statistics of soil Pb content (mg/kg)

Sample set	N	Max	Min	Mean	SD	Kurt	Skew	CV	Background value (pH < 6.5)	Excess number
Calibration set	64	221.27	24.16	64.40	39.51	3.68	1.86	0.61	35	79
Validation set	32	164.43	19.88	61.45	35.04	1.96	1.53	0.57		
Total set	96	221.27	19.88	63.42	38.11	3.22	1.77	0.60		

*N* number of samples, *SD* standard deviation, *Kurt* kurtosis, *Skew* skewness, *CV* coefficient of variation

**Fig. 3** Statistics of total Pb content in the soil. **a** Statistics of Pb content in different soil types and **b** statistics of Pb content in different land-use types

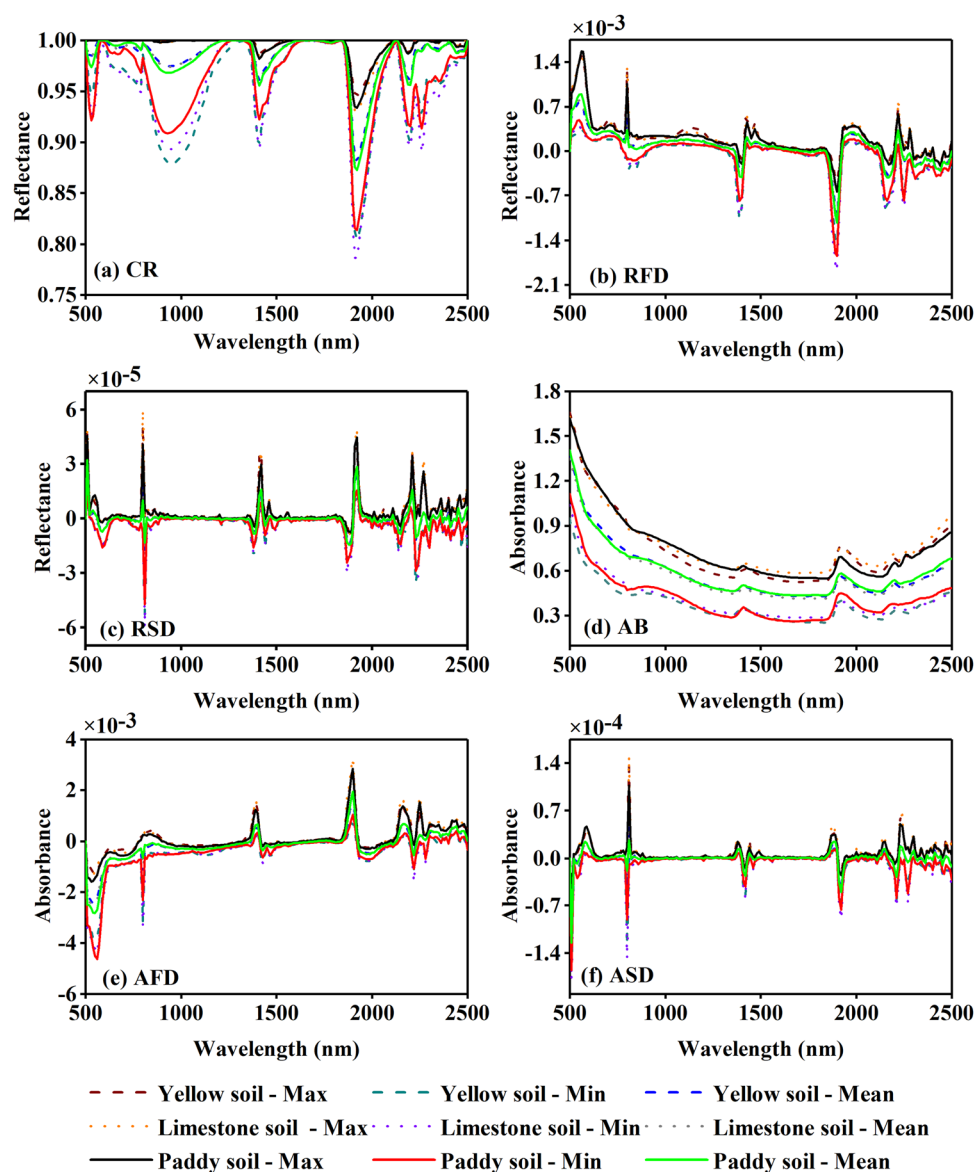


spectral information can indirectly predict heavy metal content (Song et al. 2015); thus, the screening of spectral feature bands is an important step in the establishment of a Pb content prediction model. The absorption bands of the spectral curves after CR are concentrated at approximately 530, 950, 1410, 1910, 2200, and 2260 nm (Fig. 4a). The absorption bands of the spectral curve after RFD are concentrated near 630, 810, 1390, 1440, 1900, 2160, 2250, 2310, and 2440 nm (Fig. 4b). The absorption bands of the spectral curve after RSD are concentrated near 590, 800, 1380, 1440, 1870, 2140, 2240, 2300, and 2440 nm (Fig. 4c). The absorption bands of the spectral curve after AB conversion are concentrated at 790, 1410, 1920, 2220,

and 2260 nm (Fig. 4d). The absorption bands of the spectral curve after AFD and ASD conversion are the same as that of RFD and RSD.

The water absorption bands are generally around 1400 and 1900 nm. The absorption bands of soil organic matter are at approximately 500, 600, 800, 2200, and 2300 nm. To avoid interference with modeling, the bands between 2400 and 2500 nm are not considered as feature bands due to the large noise (Wang et al. 2007; Cheng et al. 2017). Finally, combined with the six variations, the feature absorption bands of the soil spectrum are concentrated near the near-infrared bands of 1440, 1870, 2140, 2160, 2240, 2250, and 2260 nm.

**Fig. 4** Spectral curve transformation. **a** CR, **b** RFD, **c** RSD, **d** AB, **e** AFD and **f** ASD



The Pearson method was used to analyze the correlation between the Pb content and the spectral values of the feature absorption bands of the six transformed forms. There no significant correlation between the spectral values corresponding to the seven bands and the Pb content after CR (Table 2), So CR should not be considered as a variable when modeling. In addition, the correlation coefficient after AB conversion is improved compared with the original reflectance, and the overall correlation is strongest, indicating that the spectral data after transformation have the highest correlation with the Pb content. To improve the accuracy of the model and achieve better estimation results, this study selected the band that is significantly correlated at the 0.01 level as the spectral feature band of heavy metal Pb in soil. The feature bands of RE, RFD, AB, and AFD are at 1870, 2140, 2160, 2240, and

2260 nm. The feature bands of RSD and ASD are at 1440, 1870, 2140, 2240, 2250, and 2260 nm.

### 3.3 Hyperspectral estimation model and accuracy evaluation

#### 3.3.1 Modeling results

The spectral value of the spectral feature band and the Pb content were taken as the independent and dependent variables, respectively, to establish the models.  $R^2$  in six variants is greater than 0.5 (Table 3), indicating that the MLR model is reliable and can be used to estimate Pb content (Wang et al. 2014; Tao et al. 2018). The model with AB conversion has the best effect. The  $R^2$  of the calibration and validation sets are 0.649 and 0.799,

**Table 2** The correlation coefficients between soil heavy metal Pb content and spectral variables

Band (nm)	RE	RFD	RSD	AB	AFD	ASD	CR
1440	- 0.238*	- 0.168	- 0.611**	0.263*	0.124	0.648**	0.127
1870	- 0.348**	- 0.516**	- 0.479**	0.348**	0.575**	0.542**	0.159
2140	- 0.487**	- 0.698**	- 0.664**	0.493**	0.726**	0.715**	0.084
2160	- 0.560**	- 0.684**	0.088	0.569**	0.723**	0.091	0.110
2240	- 0.599**	- 0.637**	- 0.677*	0.610**	0.634**	0.689**	0.280
2250	- 0.622**	- 0.643**	- 0.302**	0.633**	0.652**	0.408**	0.220
2260	- 0.631**	- 0.646**	- 0.363**	0.643**	0.652**	0.379**	0.220

\*Significant correlation at the 0.05 level (both sides)

\*\*Significant correlation at the 0.01 level (both sides)

**Table 3** The MLR model of soil total Pb content

Spectral variable	Calibration set		Validation set		Total	
	R <sup>2</sup> c	RMSEc	R <sup>2</sup> v	RMSEv	R <sup>2</sup> t	RMSEt
RE	0.614	26.006	0.798	25.113	0.706	25.560
RFD	0.551	28.054	0.742	28.404	0.647	28.229
RSD	0.607	26.247	0.736	28.671	0.672	27.459
AB	0.649	24.805	0.799	25.034	0.724	24.920
AFD	0.596	26.606	0.724	29.363	0.660	27.985
ASD	0.660	24.418	0.723	29.434	0.692	26.926

respectively, and RMSE are 24.805 and 25.034, respectively. Although the calibration set of ASD has the largest R<sup>2</sup> and the smallest RMSE, the stability and accuracy of the verification set are far less than the model built by AB. Moreover, the accuracy of the model built by RE is much higher than that of several variables other than AB. The stability and accuracy of the model constructed with different spectral variables follow the order AB > RE > ASD > RSD > AFD > RFD.

Similar to the modeling results of the MLR model, the effect of the PCR model is the best in modeling the spectral variables after AB transformation, and the modeling effect with RE as variable is second (Table 4). The model constructed with AB as variable, regardless of whether it is the calibration set or the R<sup>2</sup> and RMSE of the validation set, reaches the maximum and minimum. The model effects constructed from different spectral variables follow the order AB > RE > ASD > RSD > RFD > AFD.

**Table 4** The PCR model of soil total Pb content

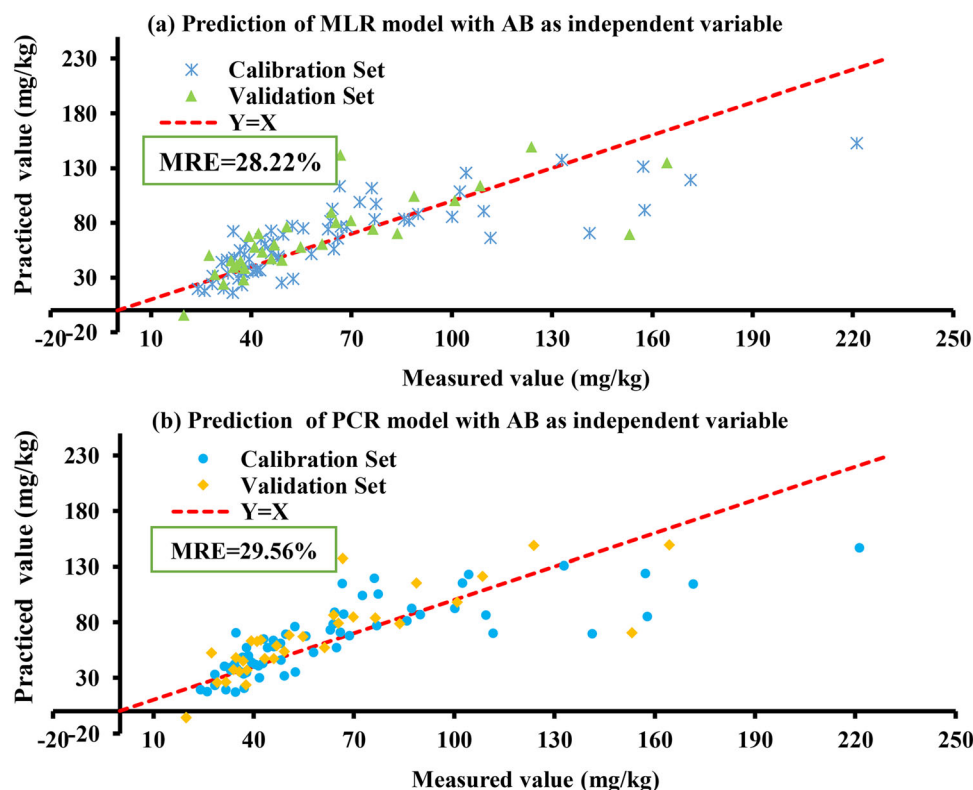
Spectral variable	Calibration set		Validation set		Total	
	R <sup>2</sup> c	RMSEc	R <sup>2</sup> v	RMSEv	R <sup>2</sup> t	RMSEt
RE	0.595	25.155	0.532	24.056	0.564	24.606
RFD	0.535	26.943	0.444	26.211	0.490	26.577
RSD	0.484	28.108	0.52	24.352	0.502	26.230
AB	0.629	24.067	0.533	24.039	0.581	24.053
AFD	0.582	25.547	0.372	27.866	0.477	26.707
ASD	0.535	26.954	0.507	24.697	0.521	25.826

### 3.3.2 Accuracy comparison

In summary, the MLR and PCR models are the most stable models constructed with AB as independent variables. Overall, the stability and fitting effects of the MLR model are superior to the PCR model. A scatter plot of the measured and predicted values of the MLR and PCR models built with the AB as an independent variable is created. The closer the sample point is to the 1:1 line (i.e., Y = X line), the closer the estimation result of the model is to the true value, and the stronger the estimation capability. Moreover, MRE is the reference indicator. The smaller the value of MRE, the stronger the model prediction capability.

The overall MRE of both models is less than 30%, which means that the estimation accuracy is higher than 70%, and all models have certain soil Pb metal estimation capability (Fig. 5). The prediction capability of the MLR model is relatively stronger than that of the PCR model.

**Fig. 5** Comparison of measured and predicted Pb content. **a** Prediction of MLR model with AB as independent variable and **b** prediction of PCR model with AB as independent variable



The sample point is closer to the 1:1 line, and the MRE value is low. The MRE is 28.22%, and the prediction accuracy reaches 71.78%.

### 3.3.3 Analysis of the applicability of models in different soil types

The above studies confirm that both models have certain reliability, but the applicability of these models in different soil types need further analysis due to a noticeable soil differentiation in the karst area. The MLR and PCR models are built by distinguishing three different soil types, namely, yellow, lime, and paddy soils, with the best predicted AB as an independent variable. The  $R^2$  of the MLR and PCR models of the three soil types are greater than 0.6, indicating that both models have certain applicability and can be used to predict Pb content in different soil types (Fig. 6). Overall, the prediction effects of the MLR and PCR models in different soil types are not much different. However, the  $R^2$  of the MLR model is larger than that of the PCR model, indicating that the prediction effect of the MLR model on different soil types is better than the PCR model, especially for the prediction of limestone soil. The  $R^2$ , RMSE, and MRE of the MLR model are better than those of the PCR model, which are 0.747%, 22.28%, and 25.37%, respectively. That is, the applicability of the PCR model is not as strong as the MLR model. However, the

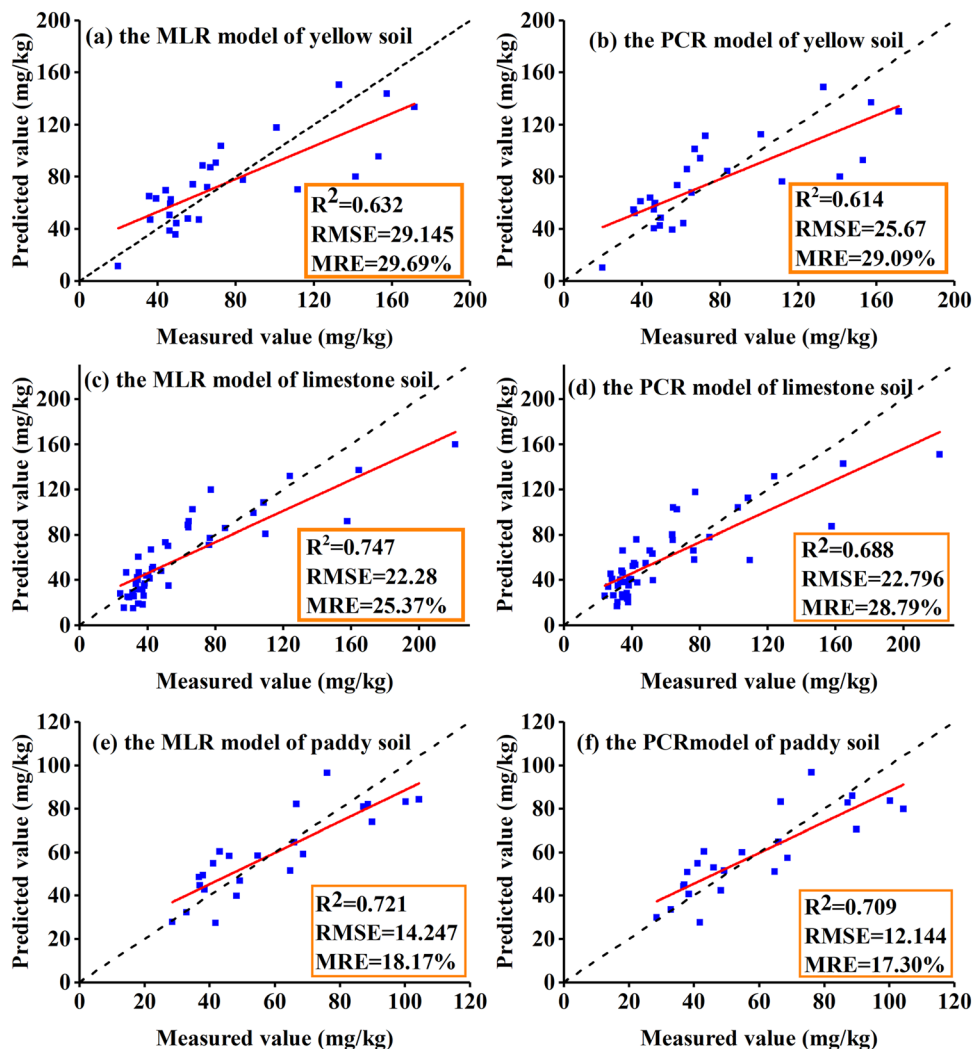
prediction accuracy of the PCR model slightly improves after distinguishing the soil type compared with the previous one, and the  $R^2$  increases from 0.581 to over 0.614, with a maximum value of 0.709 (Table 4). When using the PCR model for Pb content prediction, an improved effect can be achieved by distinguishing soil types for prediction.

## 4 Discussion

The MLR model can prevent the omission of effective independent variables, and the PCR model can effectively solve the collinearity problem between independent variables (Xu et al. 2013; Ye et al. 2017). These models are widely used in spectroscopy prediction research because they are simple, easy to develop, and low cost (Wang et al. 2017). In this study, the two methods are used to model heavy metal Pb content in a small watershed of a typical karst plateau, with the spectral feature band as an independent variable. The extraction of spectral information of heavy metals in soil is difficult, and the modeling accuracy is easily subjected to interference from other composition information because soil heavy metals are not the dominant factor controlling spectroscopy remote sensing information. Therefore, removing the bands that are susceptible to interference from other information is the key to establishing a hyperspectral model (Zhang et al. 2015). First,



**Fig. 6** Applicability of MLR and PCR models based on different soil types. **a** The MLR model of yellow soil, **b** the PCR model of yellow soil, **c** the MLR model of limestone soil, **d** the PCR model of limestone soil, **e** the MLR model of paddy soil and **f** the PCR model of paddy soil



several mathematical transformations are performed on the spectral reflectance to remove the interference of the noise on the spectral information to simplify the finding of the reflection valley and the absorption peak of the soil spectrum. Second, the bands that may interfere with information, such as moisture and organic matter, are removed, and Pearson correlation analysis is used to screen the spectrally sensitive bands of Pb content in the near-infrared, which is similar to the previous studies (Jiang et al. 2017; Zhang et al. 2017). Finally, a hyperspectral quantitative estimation model between spectral variables and soil heavy metal Pb is built. The  $R^2$ , RMSE, and MRE indicators are used for modeling and evaluation. The best stability and reliability are achieved when the models are built on the basis of AB; this finding is consistent with the results of the relevant analysis. Except for AB, the model effect created by other transformed spectral variables is not better than the model established by the original reflectivity; thus, whether spectral transformation is required should be determined in accordance with the actual situation (Wu et al. 2014).

Studies on heavy metals concentrated in the soil of karst areas are relatively few. The estimation accuracy of the models constructed in this study meets the requirements of soil heavy metal Pb, thus providing some ideas for research on similar areas. Moreover, the results of this study provide certain technical support for solving the problem of soil heavy metal pollution in karst areas. A large spatial heterogeneity is observed in soil heavy metal content owing to the complex geological background in a karst region. Future research will consider the effect of other factors on soil heavy metal content in karst areas and explore the influence of heavy metal pollution on human health through isotope tracing.

## 5 Conclusion

The average value of soil Pb content in the study area is 1.8 times the national background value, and the variation level is high, indicating that the soil in the Houzhai River

Watershed shows strong spatial heterogeneity. Moreover, heavy metal Pb pollution is serious, and corresponding treatment measures must be taken.

Soil Pb elements are sensitive in the near-infrared region and are concentrated near several bands of 1440, 1870, 2140, 2160, 2240, 2250, 2260, and 2350 nm. After mathematical transformation, except for CR, the other spectral variables are significantly correlated with soil Pb, and the correlation of AB is the most significant.

The MLR and PCR models are the most reliable models with AB as independent variable. The overall  $R^2$  and RMSE reach the maximum and minimum, respectively, which is consistent with the results of the correlation analysis. The stability, fitting effect, and prediction capability of the MLR model are better than the PCR model.

The MLR and PCR models have certain applicability in different soil types, and the applicability of the MLR model is stronger. When using the PCR model for prediction, distinguishing different soil types can achieve improved results.

In the present study, the model constructed using hyperspectral technology can predict heavy metal Pb content in the soil, which is crucial for large-area and large-scale rapid monitoring of soil heavy metal content. In future research, the use of airborne and spaceborne hyperspectral images will be attempted to seek fast and broad breakthroughs in the monitoring of heavy metals in soil.

**Acknowledgements** This research work was supported jointly by National Key Research Program of China (Nos. 2016YFC0502300 and 2016YFC0502102), Chinese Academy of Science, and Technology Services Network Program (No. KFJ-STS-ZDTP-036), 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 and 41571130042), Science and Technology Plan of Guizhou Province of China (No. 2017–2966).

#### Compliance with ethical standards

**Conflict of interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

## References

- Cheng XF, Song TT, Chen Y et al (2017) Retrieval and analysis of heavy metal content in soil based on measured spectra in the Lanping Zn–Pb mining area, western Yunnan Province. *Acta Petrol Mineral* 36(1):60–69
- China National Environmental Monitoring Center (1990) Chinese soil element background value. China Environmental Science Press, Beijing
- Gholizadeh A, Borůvka L, Saberioon MM et al (2015) Comparing different data preprocessing methods for monitoring soil heavy metals based on soil spectral features. *Soil Water Res* 10(4):218–227
- Guo YK, Cao XY, Xie Q et al (2015) Primary research on the inversion model of soil heavy metal based on simulation multi-spectral data. *Eng Surv Mapp* 24(12):7–16
- Jiang ZL, Yang YS, Sha JM (2017) Application of GWR model in Vis–NIR spectroscopy prediction of soil heavy metals. *Acta Geogr Sin* 72(3):533–543
- Li X, Liu L, Wang Y et al (2013a) Heavy metal contamination of urban soil in an old industrial city (Shenyang) in Northeast China. *Geoderma* 192:50–58
- Li YB, Luo GJ, Chen AY et al (2013b) A typical case study on evolution of karst rocky desertification in Houzhaihe, Puding County, central Guizhou Province, China. *Geogr Res* 32(5):828–838
- Li GW, Gao XH, Yang LY et al (2017) Estimating soil organic matter contents from different soil particle size using visible and near-infrared reflectance spectrum: a case study of the Huangshui watershed. *Chin J Soil Sci* 6:1360–1370
- Lu Q, Wang SJ, Bai XY et al (2019) Rapid inversion of heavy metal concentration in karst grain producing areas based on hyperspectral bands associated with soil components. *Microchem J* 148:404–411
- Ma WB, Tan K, Li HD et al (2016) Hyperspectral inversion of heavy metals in soil of a mining area using extreme learning machine. *J Ecol Rural Environ* 32(2):213–218
- Maliki AA, Bruce D, Owens G (2014) Prediction of lead concentration in soil using reflectance spectroscopy. *Environ Technol Innov* s1-2(1):8–15
- Peng J, Chi CM, Xiang HY et al (2014) Inversion of soil salt of content based on continuum-removal method. *Acta Pedol Sin* 3:459–469
- Ren HY, Zhuang DF, Singh AN et al (2009) Estimation of As and Cu contamination in agricultural soils around a mining area by reflectance spectroscopy: a case study. *Pedosphere* 19(6):719–726
- Shi Z (2014) Principle and method of soil surface Vis–NIR spectroscopy remote sensing. Science Press, Beijing, p 31
- Song L, Jian J, Tan DJ et al (2015) Estimate of heavy metals in soil and streams using combined geochemistry and field spectroscopy in Wan-sheng mining area, Chongqing, China. *Int J Appl Earth Obs Geoinf* 34(1):1–9
- Song B, Tian ML, Chen TB et al (2017) Characterization of Pb contents in typical soils of Guangxi and risk assessment based on literature date. *Acta Pedol Sin* 54(1):92–107
- Sun W, Zhang X, Sun X et al (2018) Predicting nickel concentration in soil using reflectance spectroscopy associated with organic matter and clay minerals. *Geoderma* 327:25–35
- Tan K, Ye YY, Du PJ et al (2014) Estimation of heavy metal concentrations in reclaimed mining soils using reflectance spectroscopy. *Spectrosc Spectr Anal* 34(12):3317
- Tan K, Wang HM, Chen LH et al (2020) Estimation of the spatial distribution of heavy metal in agricultural soils using airborne hyperspectral imaging and random forest. *J Hazard Mater* 382:120987
- Tao C, Chang Q, Jing L et al (2016) Identification of soil heavy metal sources and improvement in spatial mapping based on soil spectral information: a case study in northwest China. *Sci Total Environ* 565:155–164
- Tao C, Wang YJ, Zhou B et al (2018) Assessment and analysis of migrations of heavy metal lead and zinc in soil with hyperspectral inversion model. *Spectrosc Spectr Anal* 38(6):1850–1855

- Teng J, He ZW, Ni ZY et al (2016) Spectral response and inversion models for prediction of total copper content in soil of Xifanping mining area. *Spectrosc Spectr Anal* 36(11):3637–3642
- Tian SQ, Wang SJ, Bai XY et al (2019) Hyperspectral prediction model of metal content in soil based on the genetic ant colony algorithm. *Sustainability* 11(3197):1–21
- Tu YL, Zou B, Jiang XL et al (2018) Vis–NIR spectroscopy remote sensing based modeling of Cu content in mining soil. *Spectrosc Spectr Anal* 38(2):575–581
- Wang L, Lin QZ, Jia D et al (2007) Study on the prediction of soil heavy metal elements content based on reflectance spectra. *J Remote Sens* 11(6):906–913
- Wang J, Cui L, Gao W et al (2014) Prediction of low heavy metal concentrations in agricultural soils using visible and near-infrared reflectance spectroscopy. *Geoderma* 216(4):1–9
- Wang HB, Zhao ZQ, Yi L et al (2017) Leaf area index estimation of spring maize with canopy Vis–NIR spectroscopy data based on linear regression algorithm. *Spectrosc Spectr Anal* 37(5):1489–1496
- Wang F, Gao J, Zha Y (2018) Vis–NIR spectroscopy sensing of heavy metals in soil and vegetation: feasibility and challenges. *ISPRS J Photogramm Remote Sens* 136:73–84
- Wu MZ, Li XM, Sha JM (2014) Spectral inversion models for prediction of total chromium content in subtropical soil. *Spectrosc Spectr Anal* 34(6):1660–1666
- Xu MX, Wu SH, Zhou SL et al (2011) Vis–NIR spectroscopy reflectance models for retrieving heavy metal content: application in the archaeological soil. *J Infrared Millim Waves* 30(2):109–114
- Xu D, Sun L, Luo JS (2013) Wavelet denoising of Vis–NIR spectroscopy remote sensing imagery based on multiple linear regression. *Remote Sens Inf* 28(6):78–86
- Yao P, Zhang D, Zhang SR et al (2014) Modeling to predict lead and nickel contents in soil of the mid-and lower reaches of shiting river using RS and GIS. *J Agro-Environ Sci* 33(1):95–102
- Ye Q, Jiang X, Li X et al (2017) Comparison on inversion model of soil organic matter content based on Vis–NIR spectroscopy data. *Trans Chin Soc Agric Mach* 48(3):164–172
- Zhang XB, Wang SJ, Cao JH et al (2010) Features of water loss and soil erosion and some scientific problems on karst rocky desertification in Southwest China karst area. *Carsologica Sin* 29(3):274–279
- Zhang W, Gao XH, Yang Y et al (2014) Estimating heavy metal contents for topsoil based on spectral analysis: a case study of Yushu and Maduo counties in the three-river source region. *Soils* 6:1052–1060
- Zhang L, Wu Q, Liu S et al (2015) Research on monitoring of Pb content in the sewage irrigation area of Longkou city using Vis–NIR spectroscopy data. *Environ Eng* 33:114–122
- Zhang Q, Zhang H, Liu W et al (2017) Inversion of heavy metals content with Vis–NIR spectroscopy reflectance in soil of well-facilitated capital farmland construction area. *Trans Chin Soc Agric Eng* 33(12):230–239
- Zhang ZM, Zhou YC, Wang SJ et al (2018) Spatial distribution of stony desertification and key influencing factors on different sampling scales in small karst watersheds. *Int J Environ Res Public Health* 15(4):743