Assessment of heavy metal pollution in agricultural soil around a gold mining area in Yitong County, China, based on satellite hyperspectral imagery

Fuyu Wu,[a] Xue Wang,[b,c,*] Zhaoxian Liu,[d] Jianwei Ding,[d] Kun Tan,[a,b,c,*] and Yu Chen[a]

[a]China University of Mining and Technology, MNR Key Laboratory for Land Environment and Disaster Monitoring, Xuzhou, China

[b]East China Normal University, Key Laboratory of Geographic Information Science (Ministry of Education), Shanghai, China

[c]East China Normal University, Key Laboratory of Spatial-Temporal Big Data Analysis and Application of Natural Resources in Megacities (Ministry of Natural Resources), Shanghai, China

[d]The Second Surveying and Mapping Institute of Hebei, Shijiazhuang, China

Abstract. Soil is one of the essential natural resources that is at risk from heavy metal pollution. The traditional sampling method for soil heavy metal monitoring and assessment cannot meet the requirements for large-scale areas. The purpose of this study was to estimate the soil heavy metal concentrations based on Gaofen 5 (GF5) satellite hyperspectral imagery for the assessment of the heavy metal pollution in the study area and to analyze the scale effect under different resolutions. A total of 96 topsoil samples were collected in this work, and these samples were analyzed for the arsenic (As), cadmium (Cd), chromium (Cr), copper (Cu), nickel (Ni), lead (Pb), and zinc (Zn) contents. To solve the problem of the insignificant features caused by the complex imaging conditions of spaceborne hyperspectral satellite imagery, the binary weight symbiotic organisms search algorithm (BWSOS) was developed. After feature selection based on the BWSOS method, the heavy metal contents are inverted by the use of support vector machine regression. The experimental results show that the BWSOS feature selection method shows a good performance, with the R_p^2 values for As, Cd, Cr, Cu, Ni, Pb, and Zn being 0.67, 0.68, 0.73, 0.71, 0.66, 0.65, and 0.71, respectively. Based on the estimated heavy metal concentration maps, the geoaccumulation index (I_{geo}), the pollution index, and the potential ecological risk index were calculated to assess the heavy metal pollution status in the study area. The results showed that only As contamination is present at a significant level, but with a low level of potential risk for the whole study area. A comparison with the results obtained using HyMap airborne hyperspectral imagery showed that the GF5 satellite hyperspectral imagery can obtain consistent results for heavy metal pollution assessment. The airborne hyperspectral imagery can provide more fine details, whereas the spaceborne hyperspectral imagery is more suitable for large-scale pollution assessment at a low cost.
[DOI: 10.1117/1.JRS.15.042613]

Keywords: hyperspectral imagery; symbiotic organisms search; soil heavy metal estimation; heavy metal pollution; pollution index; scale effect.

Paper 210329SS received May 21, 2021; accepted for publication Oct. 8, 2021; published online Nov. 3, 2021.

1 Introduction

Black soil resources, which are widespread in the northeast of China, play an important role in cereal grain production in the country.[1–3] However, high-intensity industrialization and urbanization can increase the concentrations of heavy metals in soil.[4] Excessive heavy metal concentrations in soil impact the soil quality and inevitably pose a threat to the whole ecosystem.[5–7]

*Address all correspondence to Kun Tan, tankuncu@gmail.com

1931-3195/2021/$28.00 © 2021 SPIE

Thus, detailed investigation of the environmental impact of heavy metals in agricultural soil is of great importance.

Mining, mineral processing, and metallurgical extraction are the three principal steps of gold mining industries. However, these operations can lead to soil contamination by heavy metals such as arsenic (As), mercury (Hg), cadmium (Cd), chromium (Cr), lead (Pb), copper (Cu), and zinc (Zn).[8–10] Okang'Odumo et al.[11] analyzed the environmental influence of artisanal gold mining activity in Kenya and concluded that the study area was heavily polluted with Hg. Ono et al.[12] evaluated As bioaccessibility and assessed children's exposure to As-contaminated materials by sampling the soil around a gold mining area in Brazil. This research showed that the total As concentration was high in the gold mining area, together with a low bioaccessibility. However, as the soil in gold mining regions is often polluted by multiple heavy metals, the assessment of a single heavy metal cannot assess the soil heavy metal pollution status accurately.[13,14] Therefore, the geochemical accumulation index, the enrichment factor, and the potential ecological risk (PER) index are widely used in soil heavy metal pollution assessment. For example, Chen et al.[15] used these indices to describe the soil heavy metal pollution degree around the Haigou gold mines in Jilin, China. The results indicated that the study area represented a potentially strong risk to ecological health. Wu et al.[16] determined the characteristics of the soil heavy metal pollution and identified the sources of the heavy metal pollutants in the soil around Xiaoqinling gold mining region in Shaanxi, China, by collecting 133 soil samples. However, the high cost of the sampling and chemical testing limits the accurate assessment of soil heavy metal pollution in gold mining areas. Hyperspectral imaging technology, which has the characteristics of rich spectral features, provides a potential way to solve this problem.[17,18]

Many studies have demonstrated the significant correlation between soil spectra and heavy metal concentrations, which forms the basis of the hyperspectral estimation of soil heavy metal contents.[19] The high degree of correlation between soil spectra and heavy metal concentrations highlights the potential of applying remote sensing technology to assess the type and degree of soil pollution.[20–22] With the successful launch of hyperspectral satellites, such as Gaofen 5 (GF5) and ZY1-02D, a new solution for the low-cost and rapid mapping of soil heavy metal pollution in soil has been provided. However, the spectral features of hyperspectral imagery contain much redundant information, which impacts the prediction accuracy of the estimation models. Feature selection algorithms can be applied to hyperspectral inversion models to improve the predictive ability. For example, Shen et al.[23] analyzed the relationship between soil spectral reflectance and heavy metal concentration based on Pearson correlation coefficients and selected the bands above the 0.05 saliency level as the feature bands to build the estimation model. The competitive adaptive reweighted sampling technique (CARS), which is based on the importance level of each wavelength, has been widely applied in feature selection.[24] For example, Wei et al.[25] combined CARS and a gradient boosting regression tree to monitor soil heavy metal pollution. However, the bands with relatively small absolute regression coefficients are removed by force in CARS, which can cause an unstable performance and can cause the model to fall into local optimum solutions. The metaheuristic algorithms have been introduced into feature selection due to their excellent performance in the optimization process. For example, Shi et al.[26] applied a genetic algorithm (GA) to select informative spectral variables from soil and leaf spectra and achieved a high prediction precision for soil As content. Tian et al.[27] combined a GA and an ant colony algorithm for the feature selection and successfully estimated the heavy metal concentrations in soil. The symbiotic organisms search (SOS) algorithm, which offers a way to find the optimum solution for many tough engineering problems in a short span of time, was proposed by Cheng and Prayogo.[28] The SOS algorithm continuously searches for the optimal solution based on the mutualism phase, commensalism, phase, and parasitism phase in the search space and has been used to select the features for classification and pan-sharpening in the remote sensing field.[29,30] However, the SOS algorithm is a continuous optimization algorithm and is less effective in feature selection for regression with small sample sizes. In this paper, we propose a binary weight symbiotic organisms search algorithm (BWSOS) for the estimation of soil heavy metal concentrations.

The objectives of this study were: (i) to develop a suitable algorithm for feature selection; (ii) to verify the feasibility of estimating soil heavy metal concentrations using GF5 satellite hyperspectral data; (iii) to assess the heavy metal pollution degree in the agricultural soil around

a gold mining area based on the estimation results; and (iv) to compare the effect of the scale on the retrieval processing using airborne and spaceborne hyperspectral data.

2 Materials and Methods

2.1 *Study Area*

The study area is located in Yitong Manchu Autonomous County, Jilin province, China. The area has a humid monsoon climate with an average annual temperature of 5.5°C. The average annual rainfall and sunlight amount to about 651.7 mm and 2536 h, respectively. The gold mining area is located within a longitude of 125°26′E and a latitude of 43°16′N, and it covers an area of about 0.5 km^2. The location of the research area is shown in Fig. 1.

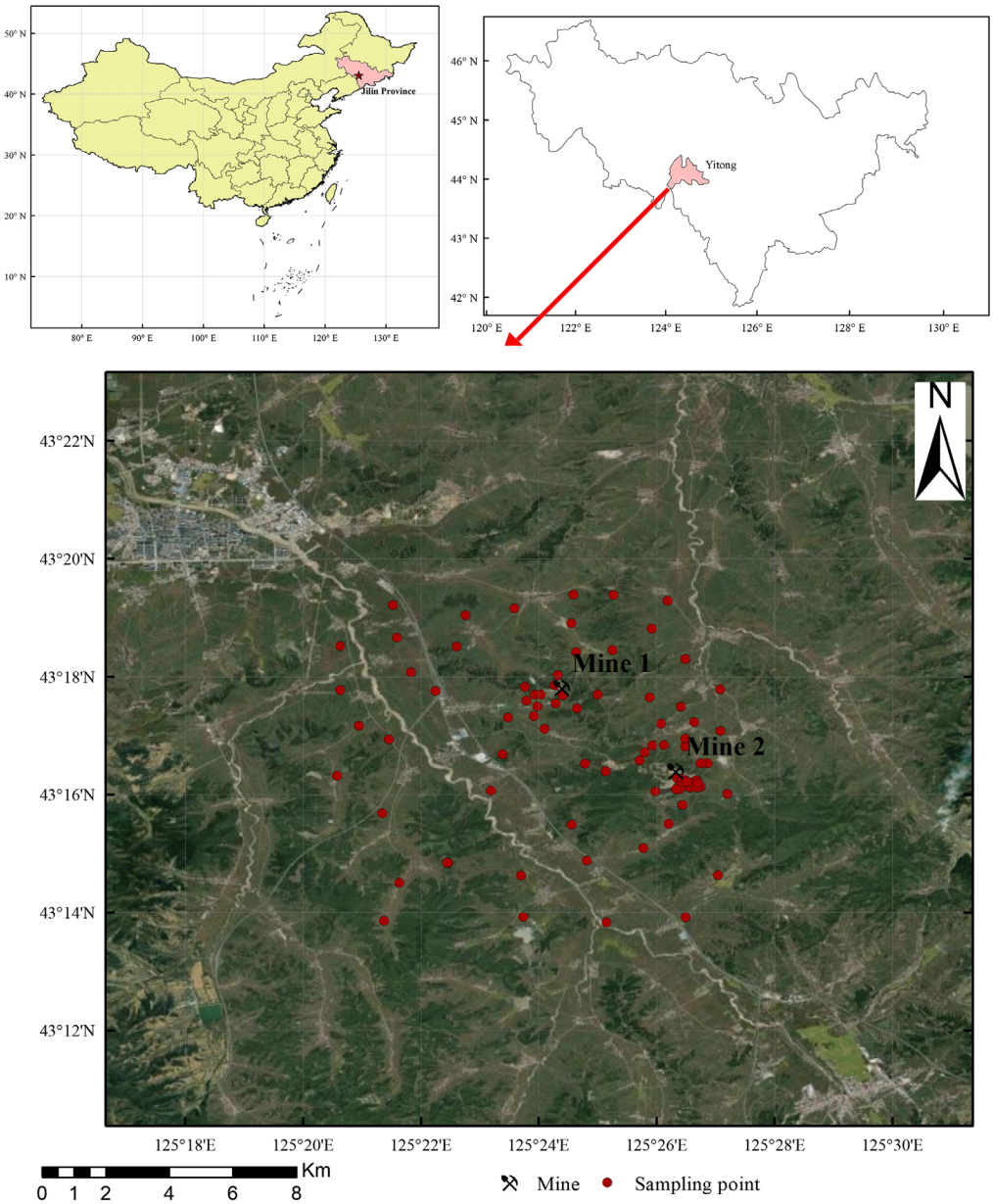

Fig. 1 Study area and sampling points.

2.2 *Datasets*

2.2.1 *Soil sample collection and testing*

A total of 96 soil samples were collected during May 1 to 5, 2019, in the study area. The sample locations are shown in Fig. 1. To reduce the influence of mixed pixels on the modeling, we chose large areas of bare soil during the sampling process. The location of each sampling site was confirmed using a global positioning system. The surface soil samples (0 to 20 cm) were collected at five points around each sampling site and fully mixed to reduce sampling errors.[31] Each sample was kept in a labeled and sealed polytetramethylene bag. The small stones and vegetation residues in the soil samples were removed, and all the soil samples were naturally air-dried and ground with an agate mortar to pass a 100-mesh sieve (0.15 mm) before the chemical analysis of the soil properties. A solution of HNO_3 and HCL with a ratio of 1:1 was then poured onto the soil samples for acid boiling digestion. Finally, the concentrations of the heavy metals were determined by inductively coupled plasma-mass spectrometry.

2.2.2 *Hyperspectral image data and preprocessing*

The hyperspectral data of the study area were obtained on February 1, 2019, by the visible-shortwave infrared advanced hyperspectral imager (AHSI) sensor carried onboard the GF5 satellite. The AHSI sensor has 330 spectral bands to characterize the solar reflective regime from 400 to 2500 nm.[32] The radiometric correction of the hyperspectral data was based on the calibration coefficients proposed by Tan et al.[33] The surface reflectance was obtained using the fast line-of-sight atmospheric analysis of hypercubes module, after the radiometric correction. Figure 2(a) shows the hyperspectral image after radiometric correction and atmospheric correction.

As this study was focused on the heavy metal concentrations in soil, the soil area was extracted using the support vector machine (SVM) classification method. Before classification, four kinds of categories were selected, i.e., bare soil, vegetation, water, and building. The training samples of each class were selected from the hyperspectral data based on visual interpretation. The preprocessed satellite image was classified by SVM, and the resulting classification map is shown in Fig. 2(b).

It was necessary to implement effective preprocessing methods to reduce the noise from the instrument and the environment, so the preprocessed data could better highlight the spectral

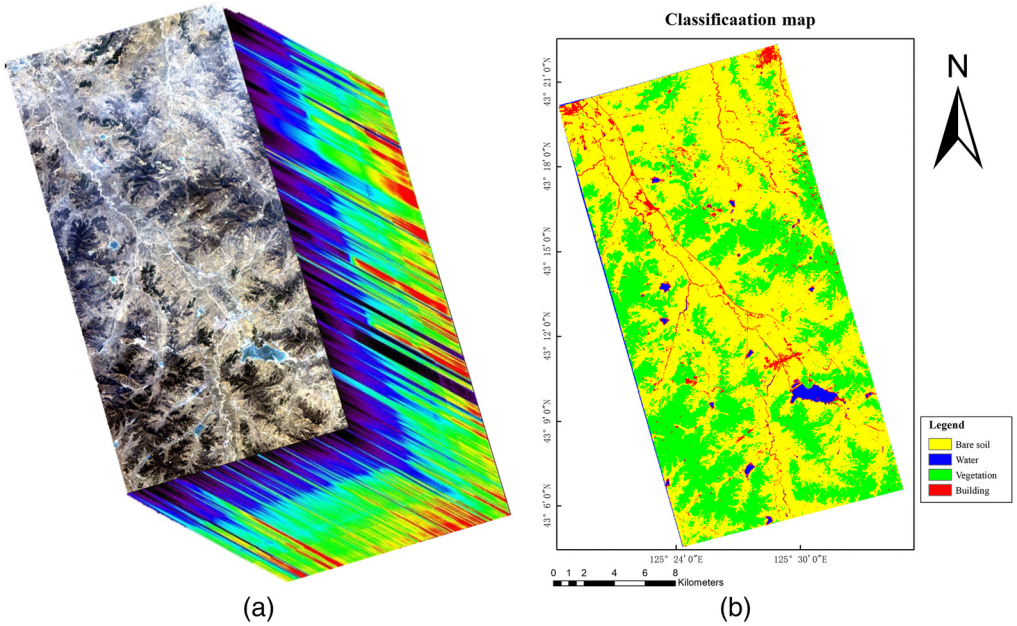

Fig. 2 (a) GF5 hyperspectral image after radiometric correction and atmospheric correction. (b) The classification map.

characteristics. The common preprocessing methods,[34–36] i.e., Savitzky–Golay (SG) smoothing, first derivative (FD) processing, second derivative (SD) processing, and standardization (ST), were conducted in this study.

2.3 *Binary Weight Symbiotic Organisms Search Algorithm*

The SOS algorithm is a metaheuristic optimization algorithm based on the natural relationships among organisms. The algorithm starts by a group of random organisms representing the candidate solutions and updates the organisms by the use of a mutualism phase, a commensalism phase, and a parasitism phase. In the process of the SOS algorithm, the mutualism phase and commensalism phase have a stronger local exploitation ability, and the parasitism phase improves the global exploration ability. However, with no augmentation of the auxiliary information, the algorithm is less effective for feature selection in a regression model with a small sample size. In this study, we introduced weight information into the mutualism phase and commensalism phase of the SOS algorithm to search for more effective features. The weights were determined by the coefficients in the partial least squares (PLS) model. The PLS model is widely used to establish the linear relationship between two variables. In the following, x represents the p-dimensional spectral data, and y represents the heavy metal concentration. If we suppose that W is the weight vector, T is a linear combination of x with W, and c is the regression coefficient vector of y against T by least squares, then the following equations apply:

$$T = xW, \tag{1}$$

$$Y = Tc + e = xWc + e, \tag{2}$$

$$b = Wc, \tag{3}$$

where e is the prediction error and b is the coefficient vector. The value of b_i in b represents the contribution of the i'th band to y. To compare the contribution of each band in the whole process, a normalized weight is defined as

$$w_i = \frac{|b_i|}{\sum_{i=1}^{p} |b_i|}. \tag{4}$$

The root-mean-square error (RMSE) is used as the fitness criterion of the BWSOS algorithm.

In the algorithm initialization phase, some organisms representing candidate solutions are initialized, and each dimension of the organisms is set to $\{0,1\}$, where 0 indicates that the feature of the corresponding dimension is not selected, and 1 indicates that the feature is selected.

The mutualism phase refers to a symbiotic relationship between two different organisms in which both benefit. In the BWSOS algorithm, X_i is the i'th organism of the ecosystem, and another organism X_j is selected randomly to interact with X_i. A new organism X_{ij} is formed by merging X_i and X_j, and the weight of each band selected in organism X_{ij} is calculated. $X_{i\text{new}}$ and $X_{j\text{new}}$ are generated separately according to the weights of X_{ij}. The rule for updating is that features with higher weights have a higher probability of being selected, whereas bands with lower weights have a higher probability of being discarded.

The commensalism phase defines a symbiotic relationship between two distinct organisms, in which one benefits and the other is unaffected. Organism X_j is selected randomly to interact with X_i. According to the update rule, $X_{i\text{new}}$ is generated based on the weight of each band selected in organism X_j.

The parasitism phase represents a symbiotic relationship between two distinct organisms, where one benefits and the other is harmed. $X_{i\text{para}}$ is created by duplicating organism X_i and changing the value of the randomly selected dimensions. Organism X_j is selected randomly to interact with X_i. The RMSE of organisms $X_{i\text{para}}$ and X_j is then evaluated to measure their fitness. If $X_{i\text{para}}$ has a better fitness value, it will kill organism X_j and assume its position in the ecosystem.

In all the phases, the new organisms in the ecosystem are updated only if their RMSE is less than their preinteraction RMSE value, and the decision to replace the best organism is made by comparing their RMSE values. The overall process is shown in Fig. 3.

2.4 *Heavy Metal Estimation Model*

2.4.1 *Support vector machine regression*

The support vector machine regression (SVR) model uses a cost function to measure the empirical risk and minimize the regression error and has been proved to be an effective model for nonlinear regression tasks with a small sample size.[37] Given a training dataset $\{(x_1, y_1), (x_2, y_2) \ldots (x_n, y_n)\}$, the estimation function in SVR is formulated as follows:

$$f(x) = w^T x + b, \tag{5}$$

where w and b are determined from the training dataset by minimizing the regression risk $R_{\text{reg}}(f)$:

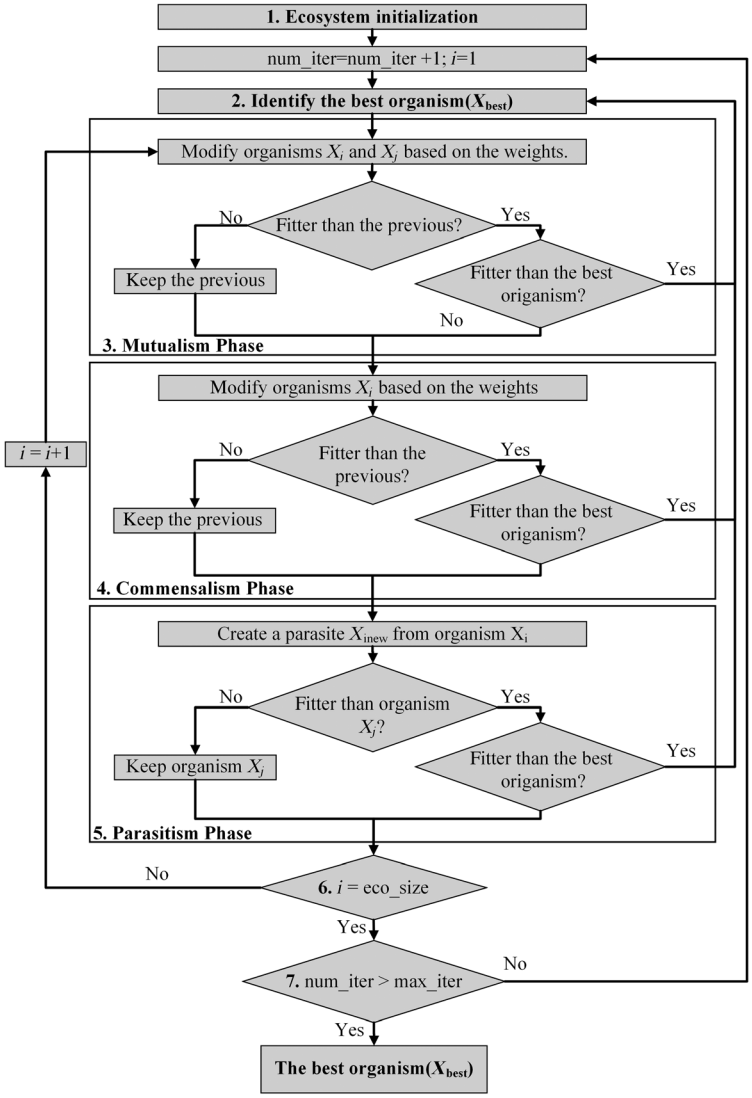

Fig. 3 General workflow of the proposed BWSOS method.

$$R_{\text{reg}}(f) = C\sum_{i=1}^{n}\phi(f(x_i) - y) + \frac{1}{2}(w \bullet w), \tag{6}$$

where C is a prespecified value and $\phi(\bullet)$ is a cost function:

$$\phi(f(x) - y) = \begin{cases} 0, & |y - f(x)| < \varepsilon \\ |y - f(x)| - \varepsilon, & \text{otherwise} \end{cases}. \tag{7}$$

This optimization problem can be solved by finding the Lagrangian, and the dual form of the optimization problem can be written as follows:

$$\max\ aa^* - \varepsilon\sum_{i=1}^{n}(a_i + a_i^*) + \sum_{i=1}^{n}(a_i^* - a_i)y_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(a_i^* - a_i)(a_j^* - a_j)x_i^T x_j, \tag{8}$$

which is subject to

$$\sum_{i=1}^{n}(a_i^* - a_i) = 0, \quad a_i, a_i^* \in [0, C], \tag{9}$$

where α and α^* are the corresponding Lagrange multipliers.

In this study, a radial basis function (RBF) kernel was selected to map the data to the high-dimensional feature space and to solve the problem of linear inseparability in the original space:[38]

$$K(x, x_i) = \exp(-\beta|x - x_i|^2). \tag{10}$$

Finally, the estimation function in Eq. (5) can be rewritten as

$$f(x) = \sum_{i=1}^{n}(a_i^* - a_i)k(x_i, x) + b. \tag{11}$$

To obtain the optimal regression model, two important parameters, i.e., the penalty coefficients C and β of the RBF kernel, are determined through the k-fold cross-validation approach.

2.4.2 *Evaluation of the calibration model*

To evaluate the performance of the model, the coefficient of determination (R^2), the RMSE, and the mean absolute error (MAE) are used in this paper. The R^2, RMSE, and MAE functions are represented as shown in Eqs. (12)–(14), respectively:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}, \tag{12}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{n}}, \tag{13}$$

$$\text{MAE} = \frac{\sum_{i=1}^{n}|y_i - \widehat{y_i}|}{n}, \tag{14}$$

where y_i is the measured value, $\widehat{y_i}$ is the predicted value, $\overline{y}$ is the average of the measured value, and n is the number of samples.

Based on the gradient of the heavy metal concentrations, all the samples were divided into training and test sets in a ratio of 2:1. The evaluation parameters of the training sets are denoted as R_c^2, RMSE_c, and MAE_c, and those of the test sets are denoted as R_p^2, RMSE_p, and MAE_p.

2.5 *Risk Assessment Model*

2.5.1 *Geoaccumulation index*

The influences of both the natural geological processes and human activities on heavy metal pollution are included in the geoaccumulation index (I_{geo}).[39] Thus, I_{geo} can not only reflect the natural change characteristics of the heavy metal distribution but it can also be used to judge the impact of human activities. The geoaccumulation index is defined as

$$I_{geo} = \log_2\left(\frac{C_n}{1.5B_n}\right), \tag{15}$$

where C_n represents the detected concentration of the heavy metal (mg/kg), and B_n is the reference value. In this study, the background content in Chinese soil (CNEMC, 1990) was chosen as the reference value (mg/kg) of the different heavy metals. As shown in Table 1, I_{geo} can be divided into seven levels.[40]

2.5.2 *Heavy metal pollution evaluation*

The single pollution index (PI) and the integrated Nemerow pollution index (NPI) were introduced to assess the soil pollution.[41] The PI and NPI are calculated as follows:

$$\mathrm{PI} = \frac{C_n}{B_n}, \tag{16}$$

$$\mathrm{NPI} = \sqrt{\frac{(\mathrm{MaxPI})^2 + (\mathrm{AvePI})^2}{2}}, \tag{17}$$

where MaxPI and AvePI are the maximum and average PI values, respectively.

The NPI considers both the average pollution status of various pollutants and the status of the most serious pollutant, which results in the NPI being an effective indicator for heavy metal pollution.

As shown in Tables 2 and 3,[42,43] according to the degree of contamination, the PI and NPI can be classified into four grades and five grades, respectively.

2.5.3 *Potential ecological risk index*

The PER index considers multielement coordination, the toxicity level, the pollution concentration, and environmental sensitivity, and is widely used in environmental risk assessment.[44] The PER index is calculated as

Table 1 I_{geo} contamination levels.

Level	Pollution state
Level 0($I_{geo} < 0$)	Uncontaminated
Level 1($0 \le I_{geo} < 1$)	Uncontaminated to moderately contaminated
Level 2($1 \le I_{geo} < 2$)	Moderately contaminated
Level 3($2 \le I_{geo} < 3$)	Moderately to highly contaminated
Level 4($3 \le I_{geo} < 4$)	Highly contaminated
Level 5($4 \le I_{geo} < 5$)	Highly to very highly contaminated
Level 6($I_{geo} \ge 5$)	Very highly contaminated

Table 2 PI grades.

Grade	Pollution state
Grade 1 (PI $\leq$ 1)	Unpolluted domain
Grade 2 (1 < PI $\leq$ 2)	Slightly polluted domain
Grade 3 (2 < PI $\leq$ 3)	Moderately polluted domain
Grade 4 (PI > 3)	Seriously polluted domain

Table 3 NPI grades.

Grade	Pollution state
Grade 1 (NPI $\leq$ 0.7)	Safety domain
Grade 2 (0.7 < NPI < 1)	Precaution domain
Grade 3 (1 < NPI $\leq$ 2)	Slightly polluted domain
Grade 4 (2 < NPI $\leq$ 3)	Moderately polluted domain
Grade 5 (NPI > 3)	Seriously polluted domain

$$C_j = C_n / B_n, \tag{18}$$

$$E_j = T_n * C_j, \tag{19}$$

$$\mathrm{PER} = \sum_{i=1}^{n} E_j, \tag{20}$$

where E_j indicates the PER index for each heavy metal. The toxic response factors (T_n) for As, Cd, Cr, Cu, Ni, Pb, and Zn are set to 10, 30, 2, 5, 6, 5, and 1, respectively.[45] Table 4 lists the different levels of PER, as classified by E_j and PER.

3 Results and Discussion

3.1 *Descriptive Statistical Analysis of the Heavy Metal Concentrations*

The heavy metal concentration statistics are listed in Table 5, which shows the max, min, mean, standard deviation (Std.), and coefficient of variation (C.V.). In this study area, we found that the mean concentrations of As, Cd, Cr, Cu, Ni, Pb, and Zn were 43.30, 0.11, 50.61, 15.61, 24.98, 11.73, and 57.53 mg/kg, respectively. Compared with the background contents (11.2, 0.097,

Table 4 Levels of PER.

Level	Pollution state
E_j < 40 or PER < 150	Low risk
40 $\leq E_j$ < 80 or 150 $\leq$ PER < 300	Moderate risk
80 $\leq E_j$ < 160 or 300 $\leq$ PER < 600	Considerable risk
160 $\leq E_j$ < 320 or PER $\geq$ 600	High risk
$E_j \geq$ 320	Very high risk

Table 5 Descriptive statistics of the heavy metal concentrations in the study area.

	As	Cd	Cr	Cu	Ni	Pb	Zn
Min	1.42	0.01	3.65	0.94	2.84	0.62	6.76
Max	445.5	0.42	106.7	24.81	60.90	16.64	109.1
Mean	43.30	0.11	50.61	15.61	24.98	11.73	57.53
Std.	73.69	0.054	13.34	3.68	7.69	2.53	12.32
C.V.	1.70	0.49	0.26	0.23	0.31	0.21	0.21

61.0, 22.6, 26.9, 26.0, and 74.2 mg/kg, respectively) and the soil environmental quality standards in China (GB 15618-2018) (40, 0.3, 150, 50, 60, 70, and 200 mg/kg, respectively). It can be seen that the mean concentration of As greatly exceeds the background value and the national standard value, and Cd is slightly higher than the background value. The mean concentrations of the other heavy metals are far below the background values. Meanwhile, the C.V. values of As and Cd are slightly higher than those of the other heavy metals, indicating that these two heavy metals are affected by the mining and mineral processing.

3.2 *Estimation and Accuracy Validation of the Soil Heavy Metal Concentrations*

The spectra of the sampling points, which were extracted from the GF5 hyperspectral imagery, were preprocessed by SG smoothing and derivative processing. The SVR model was then constructed using the feature bands extracted with the BWSOS method. CARS, SOS, GA, Pearson correlation coefficients, and the full bands were also selected as comparisons. The accuracies of the different feature selection methods are listed in Table 6, where the best results are highlighted in bold.

Due to the complexity of the imaging process of spaceborne hyperspectral data, the traditional feature selection methods, such as SOS, GA, and Pearson, have difficulty in extracting effective features. The R_c^2 of the GA for Cr, Cu, Ni, and Pb reaches almost one, but only in the accuracy evaluation of the training set and not in the test set. This illustrates that these features result in serious overfitting for the SVR model.

Compared with the traditional feature selection methods, the CARS and BWSOS methods with weighting information are more effective in feature selection. The CARS method shows a good performance for As, Cu, Pb, and Zn, with R_p^2 values of 0.6155, 0.5522, 0.6026, and 0.5803, respectively, indicating that the weighting information can significantly improve the effectiveness of the feature selection. For As with high C.V. values, the accuracies of the CARS method for the training and test sets are different, which indicates overfitting. In contrast, all the R^2 values of the BWSOS method reach more than 0.6, among which Cr, Cu, and Zn reach more than 0.7. The results for the training and test sets are also similar, which means that the model shows no overfitting or underfitting. In general, the models with high R^2 value have lower RMSE and MAE value. Table 6 shows that the RMSE and MAE of BWSOS are the smallest among all methods. The RMSE is influenced by the outliers. For As with high C.V. values, the RMSE value does not increase significantly compared with MAE, which indicates that the features selected by the BWSOS algorithm are more effective.

The main reasons for this are: (i) BWSOS takes full account of the inconsistency of the weights in the different feature combinations through the mutualism phase and commensalism phase; and (ii) the parasitic phase enhances the global search capability and prevents the model from falling into local optimum solutions.

Due to the low content of heavy metals in soil, it is not the main factor affecting soil spectrum. The correlation between heavy metals and spectra is extremely low. Through machine learning algorithm, the potential relationship can be deeply excavated to achieve high precision regression. By comparing the results of full-band, it can be found that the BWSOS algorithm can

Table 6 Regression results with different feature selection methods.

Element	Method	Calibration set			Validation set		
		R_c^2	RMSE$_c$	MAE$_c$	R_p^2	RMSE$_p$	MAE$_p$
As	FD + BWSOS	0.7893	32.9774	13.9166	**0.6781**	**43.9847**	**33.8941**
	FD + SOS	0.2993	60.1437	19.5301	0.1555	71.2479	44.0921
	FD + CARS	0.9734	11.7007	1.6336	0.6155	48.0688	36.8365
	FD + GA	0.2071	63.9821	23.6537	0.0481	75.6402	36.4621
	FD + Pearson	0.7641	34.8937	6.2387	0.2386	67.6488	41.1735
	FD + full_bands	0.5809	46.5086	7.6355	0.3538	62.3221	36.8456
Cd	ST + FD + BWSOS	0.7131	0.0314	0.0203	**0.6822**	**0.0254**	**0.0192**
	ST + FD + SOS	0.9997	0.0011	0.0011	0.0071	0.0452	0.0342
	ST + FD + CARS	0.6404	0.0351	0.0187	0.4871	0.0323	0.0257
	ST + FD + GA	0.0365	0.0575	0.0319	0.0334	0.0442	0.0312
	ST + FD + Pearson	0.2166	0.0522	0.0251	0.1175	0.0423	0.0304
	ST + FD + full_bands	0.9997	0.0011	0.0011	0.0061	0.0451	0.0343
Cr	FD + BWSOS	0.7567	6.4011	4.2286	**0.7321**	**7.0299**	**5.8546**
	FD + SOS	0.2973	10.8788	6.1251	0.0324	13.3589	9.5698
	FD + CARS	0.9399	3.1811	1.4113	0.4731	9.8581	8.0637
	FD + GA	0.9994	0.0994	0.0992	0.0014	13.5916	8.6455
	FD + Pearson	0.0527	12.6314	7.6505	0.0419	13.2935	8.3321
	FD + full_bands	0.0299	12.7823	7.4038	0.0133	13.4905	8.4043
Cu	FD + BWSOS	0.8085	1.7213	1.0133	**0.7156**	**1.7217**	**1.3863**
	FD + SOS	0.1327	3.6639	2.3891	0.0946	3.0887	2.4258
	FD + CARS	0.7496	1.9687	1.2568	0.5522	2.1723	1.7739
	FD + GA	0.9993	0.0993	0.0991	0.0565	3.1531	2.5021
	FD + Pearson	0.9993	0.1001	0.1001	0.3653	2.5861	2.0005
	FD + full_bands	0.4732	2.8555	1.4701	0.2929	2.7295	2.3248
Ni	ST + FD + BWSOS	0.6721	4.5153	2.4721	**0.6686**	**4.2811**	**3.4111**
	ST + FD + SOS	0.9998	0.1000	0.1000	0.0716	7.1661	4.7112
	ST + FD + CARS	0.7183	4.1848	2.3592	0.4544	5.4935	4.5369
	ST + FD + GA	0.9998	0.0995	0.0994	0.005	7.4396	4.6421
	ST + FD + Pearson	0.1787	7.1461	3.4713	0.1786	6.7406	4.2638
	ST + FD + full_bands	0.9998	0.0991	0.0987	0.1075	7.0262	4.8851

Table 6 (*Continued*).

Element	Method	Calibration set			Validation set		
		R_c^2	$RMSE_c$	MAE_c	R_p^2	$RMSE_p$	MAE_p
Pb	ST + FD + BWSOS	0.6701	1.4166	0.9743	**0.6541**	**1.5864**	**1.3789**
	ST + FD + SOS	0.1727	2.2432	1.5058	0.1677	2.4604	1.7915
	ST + FD + CARS	0.7769	1.1648	0.7628	0.6026	1.6346	1.2336
	ST + FD + GA	0.9983	0.0992	0.0991	0.0404	2.6419	1.8411
	ST + FD + Pearson	0.4908	1.7598	1.0328	0.2921	2.2692	1.5378
	ST + FD + full_bands	0.9984	0.0979	0.0971	0.0012	2.6953	1.8965
Zn	FD + BWSOS	0.7183	5.6778	4.2322	**0.7141**	**8.2349**	**6.7076**
	FD + SOS	0.1458	9.8867	6.8422	0.0155	15.2811	9.1068
	FD + CARS	0.7473	5.3774	2.7334	0.5803	9.9769	7.5178
	FD + GA	0.4697	7.7898	4.8271	0.1623	14.0961	8.7198
	FD + Pearson	0.4432	7.9821	4.7423	0.2364	13.4583	8.0066
	FD + full_bands	0.1617	9.7947	6.3628	0.0331	15.1448	9.3267

significantly reduce the redundant information and improve the regression accuracy. Table 7 shows the bands selected by the BWSOS method.

In summary, the results confirm that the band selection by BWSOS enables the SVR to learn the relationships between the spectra and the heavy metal concentrations correctly, obtaining a consistently excellent performance. The scatter plots between the measured and predicted values of the estimation models are shown in Fig. 4. The results clearly show that the predictions of

Table 7 Selection of feature bands statistics by BWSOS method.

Element	No.	Wavelengths (nm)
As	30	390, 394, 433, 445, 591, 664, 706, 771, 852, 993, 1006, 1030, 1080, 1097, 1249, 1266, 1299, 1308, 1485, 1536, 1569, 2015, 2041, 2049, 2108, 2260, 2361, 2386, 2420, 2462
Cd	47	390, 394, 403, 638, 677, 685, 702, 736, 788, 800, 809, 843, 852, 856, 860, 895, 899, 916, 950, 976, 1002, 1015, 1030, 1046, 1080, 1122, 1131, 1206, 1232, 1266, 1333, 1350, 1468, 1569, 1594, 1679, 2015, 2024, 2032, 2100, 2150, 2167, 2209, 2226, 2294, 2378, 2487
Cr	36	390, 398, 433, 510, 612, 647, 655, 715, 723, 745, 749, 753, 758, 762, 771, 852, 950, 976, 1002, 1029, 1072, 1131, 1148, 1198, 1519, 1679, 1704, 1738, 2074, 2083, 2142, 2159, 2193, 2209, 2268, 2336
Cu	45	403, 407, 501, 510, 522, 544, 612, 617, 702, 732, 736, 741, 762, 865, 869, 873, 912, 920, 942, 963, 967, 1006, 1046, 1072, 1206, 1215, 1316, 1325, 1485, 1494, 1527, 1645, 1763, 2032, 2041, 2074, 2176, 2193, 2226, 2243, 2260, 2361, 2412, 2454, 2471
Ni	18	390, 407, 411, 433, 535, 578, 621, 630, 647, 655, 706, 843, 873, 972, 1002, 2041, 2159, 2319
Pb	36	390, 394, 437, 458, 514, 578, 595, 600, 642, 698, 723, 766, 800, 852, 869, 929, 946, 980, 1002, 1024, 1038, 1105, 1122, 1148, 1206, 1215, 1257, 1274, 1519, 1999, 2032, 2193, 2209, 2361, 2395, 2412
Zn	21	420, 484, 561, 715, 719, 723, 728, 741, 848, 869, 907, 929, 1002, 1089, 1206, 1299, 1527, 1645, 1990, 2302, 2344

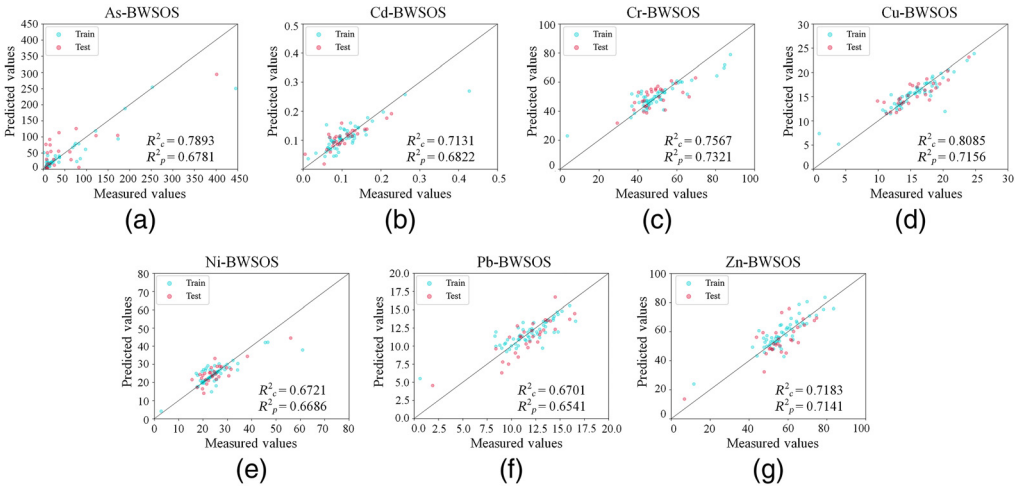

Fig. 4 Scatter plots of the predicted results for (a) As, (b) Cd, (c) Cr, (d) Cu, (e) Ni, (f) Pb, and (g) Zn.

these heavy metals are all around the prediction lines, indicating that the models show a good predictive performance.

3.3 *Heavy Metal Estimation with the GF5 Satellite Hyperspectral Data*

The trained models were applied to the soil information in hyperspectral imagery to obtain the heavy metal concentration maps. The heavy metal concentrations were divided into background values, low risk, medium risk, and high risk, based on the soil environmental quality standards in China (GB 15618-2018). The heavy metal concentration maps are shown in Fig. 5. The results show that there is high As and Cr pollution in the study area, while the other heavy metal concentrations are below the low risk level. It can be clearly seen from Fig. 5(a) that As pollution is serious in the study area, especially in the high-risk areas around the mining area. According to the field research, area A, where two gold mines are located, is a major source of pollution in the study area and is the most seriously polluted area. The transport of heavy metals in soil takes place via solute transport, which is influenced by convection. In the study area, As is transported in the direction of the river by rain and soil water and is adsorbed and accumulated during transport. The heavy metal concentration map suggests that the As from around the mining area is transported along the river. Area B is a contaminated area formed by the adsorption and accumulation of transported As. Area C is the main transportation road in the study area, on which large transportation vehicles loaded with cement, soil, and stones travel. The dust caused by the vehicles also causes As pollution in the study area. Figure 5(c) shows that Cr pollution exists in the urban area and distributed along the roads. The main sources of Cr pollution are likely to be municipal solid waste and dust from the transport vehicles.

3.4 *Pollution Assessment Based on the Estimated Concentration Maps*

The calculated I_{geo} values of the heavy metals are presented in Table 8. The I_{geo} values range from −16.95 to 3.91 (mean 0.61) for As, −15.62 to 2.62 (mean −1.11) for Cd, −15.36 to 1.19 (mean −1.29) for Cr, −11.68 to 1.67 (mean −1.25) for Cu, −9.49 to 2.25 (mean −0.66) for Ni, −13.13 to 0.16 (mean −1.97) for Pb, and −10.36 to 0.84 (mean −0.95) for Zn. The mean values of I_{geo} decrease in the order of As > Ni > Zn > Cd > Cu > Cr > Pb. Box plots of the geoaccumulation index values of the different heavy metals are shown in Fig. 6.

When considering both the analysis of the estimated concentration maps and the box plots, it is clear that the As pollution in the study area is concentrated near the mining area, while there is little to no pollution in the other areas. According to the statistics of the As geoaccumulation index, the risk intervention area is about 47%, of which 36% is level 2 (moderately contaminated) and 10% is level 3 (moderately to highly contaminated). Overall, the mean I_{geo} of As in

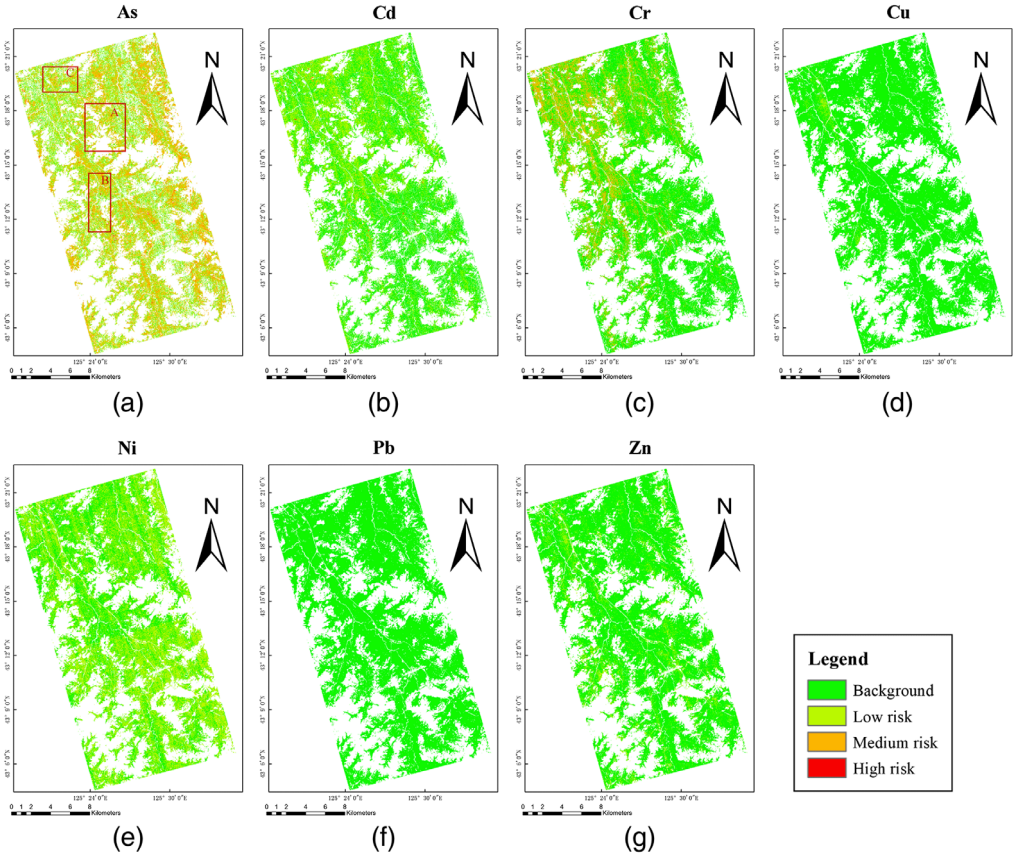

Fig. 5 The concentration maps for the different heavy metals. (a) As, (b) Cd, (c) Cr, (d) Cu, (e) Ni, (f) Pb, and (g) Zn.

Table 8 Geoaccumulation index statistics.

	As	Cd	Cr	Cu	Ni	Pb	Zn
Max	3.91	2.61	1.19	1.67	2.25	0.16	0.84
Min	−16.94	−15.62	−15.35	−11.68	−9.49	−13.13	−10.36
Mean	0.61	−1.11	−1.29	−1.25	−0.66	−1.97	−0.95

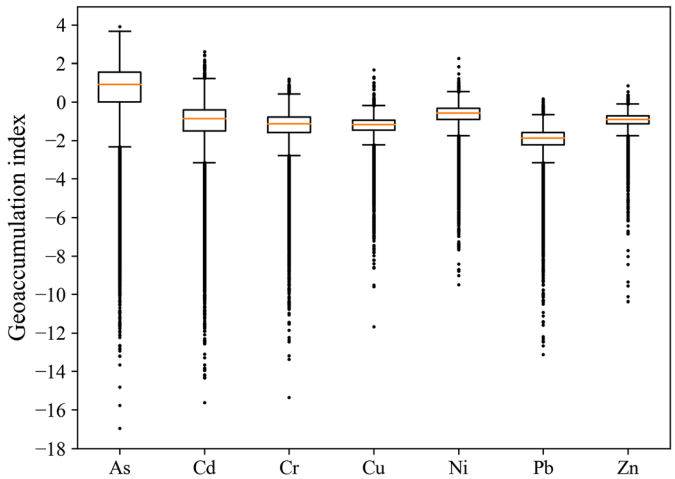

Fig. 6 Box plots of the geoaccumulation index values of the different heavy metals.

this study area varies from no pollution to low pollution, and the mean I_{geo} of the other heavy metals indicates no pollution.

The PI results, which were calculated according to the background concentrations of the heavy metals, indicate that As is in grade 4 (seriously polluted domain), and the other elements are in grade 1 (unpolluted domain) (Table 9 and Fig. 7). The PI results for all the heavy metals are in the descending order of As > Ni > Cd > Zn > Cr > Cu > Pb.

For As, about 47% of the samples are in the seriously polluted grade 4 (PI > 3). The PI results for the other heavy metals are lower (PI < 1), and mean values of 0.84, 0.68, 0.65, 0.99, 0.41, and 0.79 are obtained for Cd, Cr, Cu, Ni, Pb, and Zn, respectively. These data indicate that the soil around the mining is heavily contaminated with As. The NPI of all the soil samples varies from 0.41 to 16.16, with an average of 2.44, which shows that the study area is moderately polluted by heavy metals.

Table 10 and Fig. 8 show the single-risk indices of the individual heavy metals and the PER values in the surface soil, respectively. All the heavy metals show single-risk indices of below 40, which indicates a low pollution risk. The PER values suggest that the heavy metals in the surface soil pose a low ecological risk.

However, it is apparent from the box plots that more outliers are found in the high-risk areas. Therefore, the soil heavy metal pollution in the study area, and especially the mining area, still needs further treatment.

3.5 *Spatiotemporal and Scaling Effects Analysis for Arsenic*

To analyze the temporal evolution pattern of the main polluting element in the study area, the soil As contents for 2017 and 2019 were compared. As shown in Fig. 9, the correlation between the 2 years' contents is as high as 0.92, and the slope of the fitted straight line is near 1, which indicates that the contamination by As in the study area did not further increase. Nevertheless, government departments still need to carry out further pollution control and remediate the contaminated soil.

Table 9 PI and NPI statistics.

	As	Cd	Cr	Cu	Ni	Pb	Zn	NPI
Max	22.54	9.19	3.43	4.79	7.17	1.68	2.68	16.16
Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.41
Mean	3.15	0.84	0.68	0.65	0.99	0.41	0.79	2.44

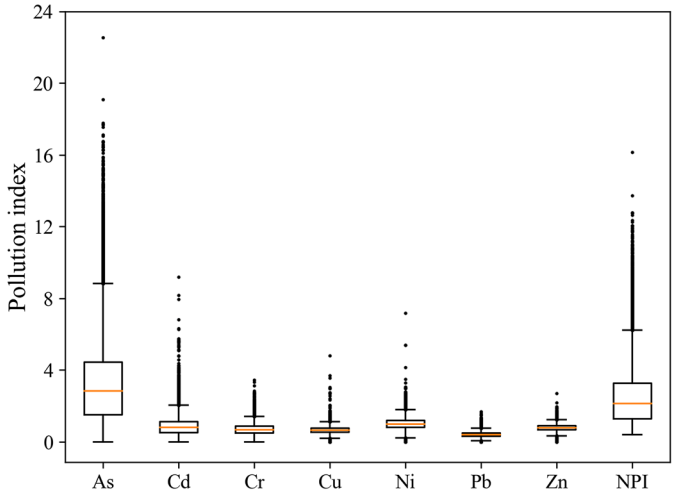

Fig. 7 Box plots of the PI and NPI values for the different heavy metals.

Table 10 PER index statistics.

	As	Cd	Cr	Cu	Ni	Pb	Zn	PER
Max	225.47	275.77	6.87	23.96	43.07	8.41	2.68	343.96
Min	0.00	0.00	0.00	0.00	0.01	0.00	0.00	8.11
Mean	31.52	25.28	1.37	3.29	5.99	2.04	0.79	70.32

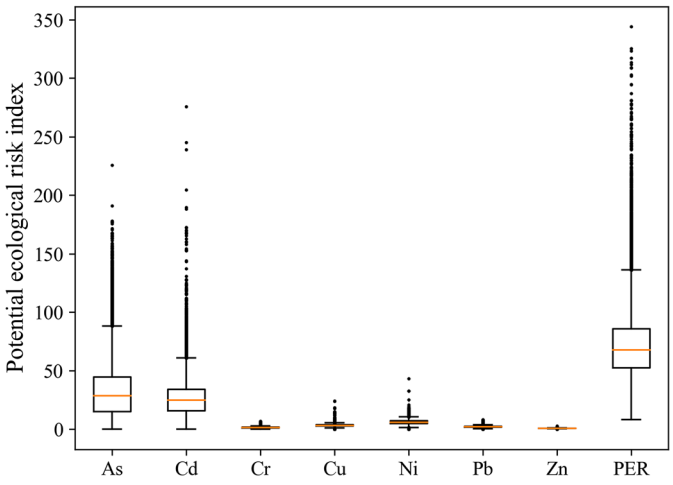

Fig. 8 Box plots of the PER index values for the different heavy metals.

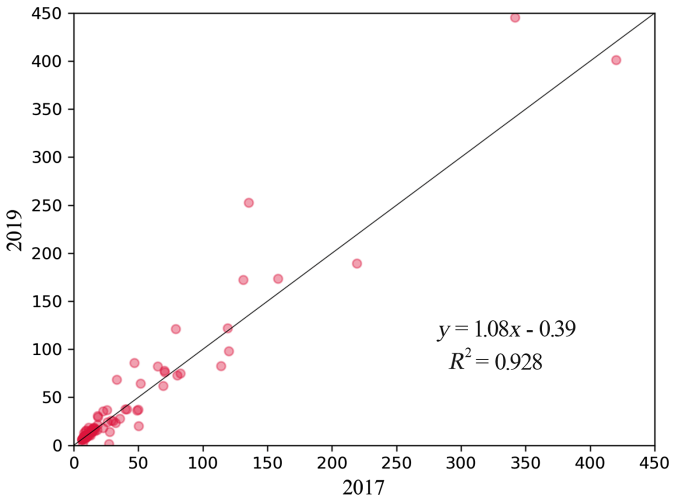

Fig. 9 Scatter plots of the soil arsenic contents in 2017 and 2019.

As the concentrations of As at the sampling sites were basically the same in the 2 years, we also compared the estimated results obtained in this study with the heavy metal concentration map obtained from HyMap hyperspectral imagery in 2017 by Ou et al.,[46] to analyze the consistency and variability of the estimated results at different spatial resolutions. The analysis was based on the overlapping areas of the HyMap data and GF5 data, as shown in Fig. 10. Although the HyMap and GF5 hyperspectral imagery were acquired in May 2017 and February 2020, respectively, the study area was both uncultivated and had large areas of bare soil on these dates, so the temporal gap would not affect the analysis of the scaling effects. The two estimated heavy metal concentration maps show the accumulation of As near the mine (area C), river (areas

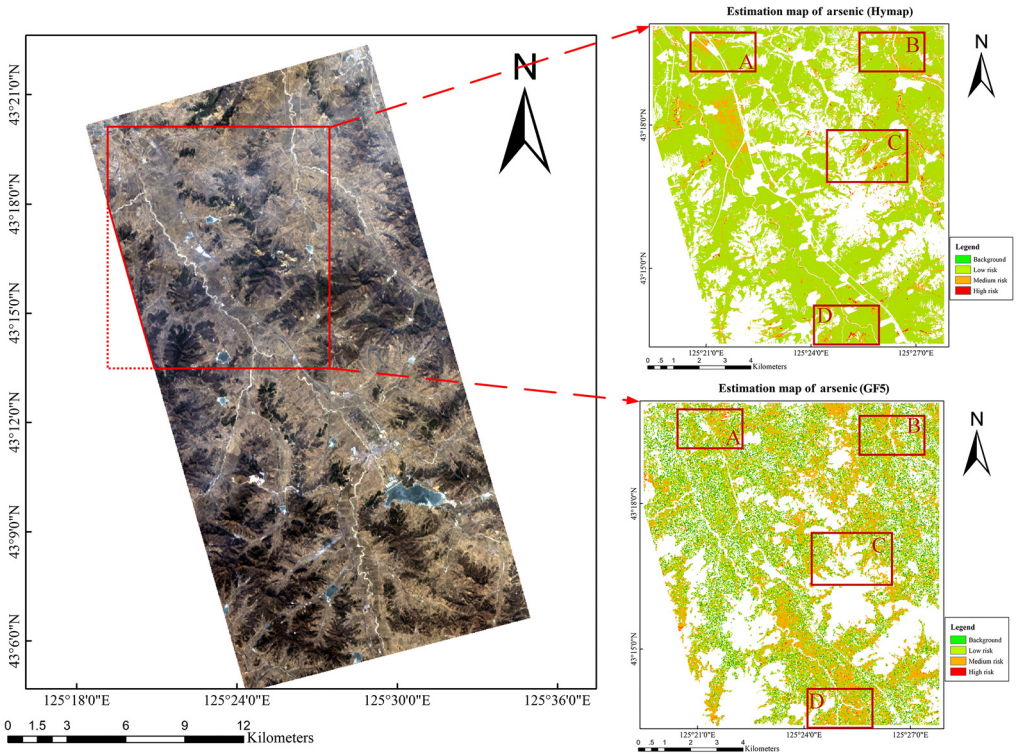

Fig. 10 Overlapping areas of the HyMap estimation results and the GF5 estimation results.

B and D), and road (area A). However, there are also some differences in the details, due to the different spatial resolutions. The 4.5-m spatial resolution of HyMap hyperspectral imagery shows more details in the estimated map. For example, in area C, the HyMap results show that pollution appears in the area around the gold mine but not in the cultivated land north of the mine. The 30-m spatial resolution of the GF5 map, limited by the low spatial resolution, cannot provide this level of detail.

To further analyze the effect of the scale on the environmental pollution assessment, the geoaccumulation index and PI values were assessed for As (Table 11). For I_{geo}, the mean values of the three methods are 0.45, 0.67, and 0.29, respectively, which all indicate that the study area is in the uncontaminated to moderately contaminated level. The mean value of the PI based on the sampling is 3.87, indicating that the study area is seriously polluted, whereas the results for GF5 and HyMap indicate that the study area is in the moderately polluted level. The reason for this is that the sampling points are more densely distributed in the mining area, which affects the overall data distribution. The assessment of heavy metals based on hyperspectral imagery can reduce the errors associated with the sampling point distribution and can provide more accurate results.

Violin plots are used to show the detailed data distribution characteristics shown in Fig. 11. As shown in Fig. 11(a), there are two peaks for the geoaccumulation index in the sampling

Table 11 Geoaccumulation index and PI statistics for the different data sources.

	I_{geo}			PI		
	GF5	HyMap	Samples	GF5	HyMap	Samples
Max	3.67	4.9	4.72	19.15	44.78	39.77
Min	−16.94	−9.94	−3.56	0.00	0.00	0.12
Mean	0.45	0.67	0.29	2.87	2.59	3.87

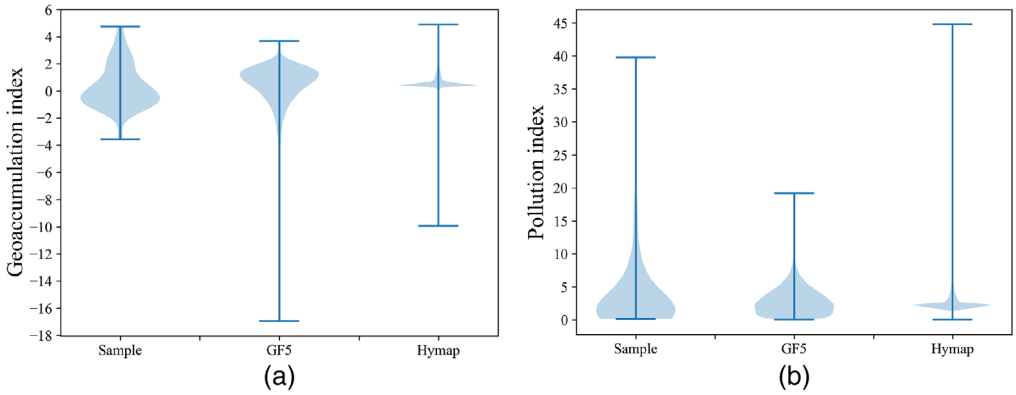

Fig. 11 Violin plots of the index comparisons. (a) Geoaccumulation index. (b) PI.

results, at about −0.5 and 2.3, respectively. The GF5 and HyMap assessments both have only one peak, and the GF5 peak is ∼1.6, indicating that the study area is moderately polluted. The peak value for HyMap is ∼0.2, indicating an unpolluted to moderately polluted level, and the HyMap assessment results are more concentrated. The PI results show that the data peaks for all three data types occur between 2 and 3, indicating that the study area is at the moderately polluted level [Fig. 11(b)]. However, the GF5 assessment results are more similarly distributed to the sampling results, whereas the HyMap results are concentrated and have a large number of outliers. It may be the case that the HyMap hyperspectral imagery, with a 4.5-m spatial resolution, provides a more precise estimation map, and its data distribution is more consistent with the actual distribution of the small-scale study area. Moreover, as the HyMap hyperspectral imagery was stitched together from multiple aerial hyperspectral images, it is possible that the variation in illumination during the imaging process causes differences in the spectra, which poses a challenge for the data modeling and leads to some outliers in the estimated heavy metal concentration.

The proportions of the different pollution levels obtained by I_{geo} and PI were calculated, and the results are shown in Tables 12 and 13. The results for GF5 and HyMap are slightly different, while the distribution for HyMap is more concentrated. The results for I_{geo} based on HyMap indicate that about 81% of the area is in level 1, while 30% and 34% of the GF5 results are distributed in level 1 and level 2, respectively. The main reason for this is that the low spatial resolution of GF5 causes higher estimates for some pixels, especially around the gold mining area. The distribution of the PI also confirms this conclusion.

In general, the scale effect of the different spatial resolutions affects the data distribution in the pollution assessment but has little impact on the final conclusions. Satellite hyperspectral imagery with a 30-m spatial resolution can provide accurate assessment results, especially at a large scale.

Table 12 The proportions of the different levels obtained by I_{geo}.

Level	Pollution state	GF5	HyMap	Samples
Level 0($I_{geo} < 0$)	Uncontaminated	0.2813	0.0142	0.5403
Level 1($0 \leq I_{geo} < 1$)	Uncontaminated to moderately contaminated	0.3011	0.8142	0.1725
Level 2($1 \leq I_{geo} < 2$)	Moderately contaminated	0.3462	0.1431	0.0919
Level 3($2 \leq I_{geo} < 3$)	Moderately to highly contaminated	0.0703	0.0242	0.1265
Level 4($3 \leq I_{geo} < 4$)	Highly contaminated	0.0011	0.0042	0.0459
Level 5($4 \leq I_{geo} < 5$)	Highly to very highly contaminated	0.0000	0.0001	0.0229
Level 6($I_{geo} \geq 5$)	Very highly contaminated	0.0000	0.0000	0.0000

Table 13 The proportions of the different pollution levels obtained by PI.

Level	Pollution state	GF5	HyMap	Samples
Grade 1 (PI $\leq$ 1)	Unpolluted domain	0.1806	0.0040	0.3678
Grade 2 (1 < PI $\leq$ 2)	Slightly polluted domain	0.2037	0.2887	0.2645
Grade 3 (2 < PI $\leq$ 3)	Moderately polluted domain	0.1981	0.5357	0.0804
Grade 4 (PI > 3)	Seriously polluted domain	0.4176	0.1716	0.2873

4 Conclusions

The BWSOS method has been proposed for feature selection for soil heavy metal concentration estimation. The BWSOS method introduces weighting information and takes full account of the inconsistency of the weights under different combinations of features, which is an approach that can extract effective features for soil heavy metal estimation in spaceborne hyperspectral data with insignificant features. Compared with other metaheuristic algorithms, BWSOS has fewer parameters, reducing the instability caused by the parameter settings. Based on the BWSOS and SVR model, we successfully applied GF5 satellite hyperspectral imagery to estimate the heavy metal concentrations in the agricultural soil around a gold mining area and produced spatial distribution maps for a large spatial area.

According to the estimated heavy metal concentration maps, the pollution levels and ecological risk levels were assessed for the whole study area. In the single element pollution assessment, the geoaccumulation index and PI of As showed that the study area is in level 1 (uncontaminated to moderately contaminated) and grade 4 (seriously polluted), respectively. In the comprehensive pollution assessment, the integrated NPI indicated that the study area is moderately polluted, and since As and Cr contamination in the study area is only concentrated around the roads, rivers, and mines, the study area was deemed to be at a low risk level overall in the potential risk assessment.

From the comparison with the test data from 2017, it was found that the concentration of As in the study area, which is seriously polluted, did not change. The spatial distributions obtained from the HyMap airborne hyperspectral imagery at a 4.5-m spatial resolution and the GF5 satellite hyperspectral imagery at a 30-m spatial resolution are essentially the same. However, the estimation map obtained by HyMap has finer distribution details due to its higher spatial resolution. In the pollution assessment, the final assessment results were consistent for all three approaches. However, the assessment results for sampling were more influenced by the sampling deployment. There were also differences in the data distribution between the assessment results of GF5 and HyMap. The HyMap assessment result distribution is more concentrated and has larger anomalous values, while the GF5 assessment result better reflects the pollution level of the study area. In general, GF5 satellite hyperspectral data can be used to obtain accurate heavy metal concentration maps and pollution assessment results. However, because of the low spatial resolution, it cannot provide fine details of the pollution distribution. Nevertheless, the advantages of the low cost and wide range mean that satellite hyperspectral imagery will likely play a greater role in soil environmental assessment in the future.

Acknowledgments

This research was supported in part by the National Natural Science Foundation of China (Grant Nos. 41871337 and 42171335). The authors would also like to thank Shanghai Institute of Technical Physics at the Chinese Academy of Sciences, Shanghai, China, for providing the advanced hyperspectral imager data used in the experiments.

References

1. X. Yang et al., "Black soil degradation by rainfall erosion in Jilin, China," *Land Degrad. Dev.* **14**(4), 409–420 (2003).

2. X. Xu et al., "Soil loss and conservation in the black soil region of Northeast China: a retrospective study," *Environ. Sci. Policy* **13**(8), 793–800 (2010).
3. Y. Ou et al., "Spatio-temporal patterns of soil organic carbon and pH in relation to environmental factors: a case study of the black soil region of Northeastern China," *Agric. Ecosyst. Environ.* **245**, 22–31 (2017).
4. M. Wang et al., "Differences in the risk assessment of soil heavy metals between newly built and original parks in Jiaozuo, Henan Province, China," *Sci. Tot. Environ.* **676**, 1–10 (2019).
5. B. J. Alloway, Ed., "Sources of heavy metals and metalloids in soils," in *Heavy Metals in Soils*, pp. 11–50, Springer (2013).
6. K. Lei et al., "Contamination and human health risk of lead in soils around lead/zinc smelting areas in China," *Environ. Sci. Pollut. Res.* **23**(13), 13128–13136 (2016).
7. R. Xiao et al., "Soil heavy metal contamination and health risks associated with artisanal gold mining in Tongguan, Shaanxi, China," *Ecotoxicol. Environ. Saf.* **141**, 17–24 (2017).
8. S. Guedron et al., "Mercury speciation in a tropical soil association; consequence of gold mining on Hg distribution in French Guiana," *Geoderma* **153**(3-4), 331–346 (2009).
9. D. Chakraborti et al., "Environmental arsenic contamination and its health effects in a historic gold mining area of the Mangalur greenstone belt of Northeastern Karnataka, India," *J. Hazardous Mater.* **262**, 1048–1055 (2013).
10. M. O. Fashola, V. M. Ngole-Jeme, and O. O. Babalola, "Heavy metal pollution from gold mines: environmental effects and bacterial strategies for resistance," *Int. J. Environ. Res. Public Health* **13**(11), 1047 (2016).
11. B. Okang'Odumo et al., "Impact of gold mining associated with mercury contamination in soil, biota sediments and tailings in Kenya," *Environ. Sci. Pollut. Res.* **21**(21), 12426–12435 (2014).
12. F. B. Ono et al., "Arsenic bioaccessibility in a gold mining area: a health risk assessment for children," *Environ. Geochem. Health* **34**(4), 457–465 (2012).
13. Q. Yang et al., "A review of soil heavy metal pollution from industrial and agricultural regions in China: pollution and risk assessment," *Sci. Tot. Environ.* **642**, 690–700 (2018).
14. A. Adewumi, "Contamination, sources and risk assessments of metals in media from Anka artisanal gold mining area, Northwest Nigeria," *Sci. Tot. Environ.* **718**, 137235 (2020).
15. M. Chen et al., "Heavy metal pollution in soil associated with a large-scale cyanidation gold mining region in southeast of Jilin, China," *Environ. Sci. Pollut. Res.* **24**(3), 3084–3096 (2017).
16. Y. Wu et al., "Heavy metals pollution and the identification of their sources in soil over Xiaoqinling gold-mining region, Shaanxi, China," *Environ. Earth Sci.* **64**(6), 1585–1592 (2011).
17. K. Tan et al., "Random forest-based estimation of heavy metal concentration in agricultural soils with hyperspectral sensor data," *Environ. Monit. Assess.* **191**(7), 446 (2019).
18. X. Jia et al., "VIRS based detection in combination with machine learning for mapping soil pollution," *Environ. Pollut.* **268**, 115845 (2020).
19. K. Tan et al., "An improved estimation model for soil heavy metal (loid) concentration retrieval in mining areas using reflectance spectroscopy," *J. Soils Sediments* **18**(5), 2008–2022 (2018).
20. K. Tan et al., "Estimation of arsenic contamination in reclaimed agricultural soils using reflectance spectroscopy and ANFIS model," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **7**(6), 2540–2546 (2014).
21. F. Wang, J. Gao, and Y. Zha, "Hyperspectral sensing of heavy metals in soil and vegetation: feasibility and challenges," *ISPRS J. Photogramm. Remote Sens.* **136**, 73–84 (2018).
22. A. Śliwińska, A. Smolinski, and P. Kucharski, "Simultaneous analysis of heavy metal concentration in soil samples," *Appl. Sci.* **9**(21), 4705 (2019).
23. Q. Shen et al., "Hyperspectral indirect inversion of heavy-metal copper in reclaimed soil of iron ore area," *Spectrochim. Acta Part A: Mol. Biomol. Spectrosc.* **222**, 117191 (2019).
24. H. Li et al., "Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration," *Anal. Chim. Acta* **648**(1), 77–84 (2009).

25. L. Wei et al., "An improved gradient boosting regression tree estimation model for soil heavy metal (arsenic) pollution monitoring using hyperspectral remote sensing," *Appl. Sci.* **9**(9), 1943 (2019).
26. T. Shi et al., "Improving the prediction of arsenic contents in agricultural soils by combining the reflectance spectroscopy of soils and rice plants," *Int. J. Appl. Earth Obs. Geoinf.* **52**, 95–103 (2016).
27. S. Tian et al., "Hyperspectral prediction model of metal content in soil based on the genetic ant colony algorithm," *Sustainability* **11**(11), 3197 (2019).
28. M.-Y. Cheng and D. Prayogo, "Symbiotic organisms search: a new metaheuristic optimization algorithm," *Comput. Struct.* **139**, 98–112 (2014).
29. Z. Jaffel and M. Farah, "A symbiotic organisms search algorithm for feature selection in satellite image classification," in *4th Int. Conf. Adv. Technol. for Signal and Image Process. (ATSIP)*, pp. 1–5 (2018).
30. C. S. Yilmaz et al., "Metaheuristic pansharpening based on symbiotic organisms search optimization," *ISPRS J. Photogramm. Remote Sens.* **158**, 167–187 (2019).
31. K. Tan et al., "Estimating the distribution trend of soil heavy metals in mining area from HyMap airborne hyperspectral imagery based on ensemble learning," *J. Hazardous Mater.* **401**, 123288 (2021).
32. Y.-N. Liu et al., "The advanced hyperspectral imager: aboard China's GaoFen-5 satellite," *IEEE Geosci. Remote Sens. Mag.* **7**(4), 23–32 (2019).
33. K. Tan et al., "Vicarious calibration for the AHSI instrument of Gaofen-5 with reference to the CRCS Dunhuang test site," *IEEE Trans. Geosci. Remote Sens.* **59**, 3409–3419 (2020).
34. X.-X. Qiao et al., "Hyperspectral estimation of soil organic matter based on different spectral preprocessing techniques," *Spectrosc. Lett.* **50**(3), 156–163 (2017).
35. S. Chen et al., "Rapid estimation of leaf nitrogen content in apple-trees based on canopy hyperspectral reflectance using multivariate methods," *Infrared Phys. Technol.* **111**, 103542 (2020).
36. Q. Zhou et al., "Non-destructive discrimination of the variety of sweet maize seeds based on hyperspectral image coupled with wavelength selection algorithm," *Infrared Phys. Technol.* **109**, 103418 (2020).
37. H. Yang, L. Chan, and I. King, "Support vector machine regression for volatile stock market prediction," in *Int. Conf. Intell. Data Eng. and Autom. Learn.*, pp. 391–396 (2002).
38. C. Axelsson et al., "Hyperspectral analysis of mangrove foliar chemistry using PLSR and support vector regression," *Int. J. Remote Sens.* **34**(5), 1724–1743 (2013).
39. G. Chevillotte et al., "Exposure method development for risk assessment to cosmetic products using a standard composition," *Food Chem. Toxicol.* **68**, 108–116 (2014).
40. E. Solgi et al., "Soil contamination of metals in the three industrial estates, Arak, Iran," *Bull. Environ. Contam. Toxicol.* **88**(4), 634–638 (2012).
41. P. Zhang et al., "Risk assessment and source analysis of soil heavy metal pollution from lower reaches of Yellow River irrigation in China," *Sci. Tot. Environ.* **633**, 1136–1147 (2018).
42. W. Ma et al., "Contamination source apportionment and health risk assessment of heavy metals in soil around municipal solid waste incinerator: a case study in North China," *Sci. Tot. Environ.* **631**, 348–357 (2018).
43. J. Cheng et al., "Assessment and mapping of environmental quality in agricultural soils of Zhejiang Province, China," *J. Environ. Sci.* **19**, 50–54 (2007).
44. L. Hakanson, "An ecological risk index for aquatic pollution control. A sedimentological approach," *Water Res.* **14**(8), 975–1001 (1980).
45. W. Wu et al., "Assessment of heavy metal pollution and human health risks in urban soils around an electronics manufacturing facility," *Sci. Tot. Environ.* **630**, 53–61 (2018).
46. D. Ou et al., "Semi-supervised DNN regression on airborne hyperspectral imagery for improved spatial soil properties prediction," *Geoderma* **385**, 114875 (2021).

Fuyu Wu is currently working toward his PhD in photogrammetry and remote sensing at China University of Mining and Technology (CUMT), Xuzhou, China. He received his BE degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2015. His research interests include the hyperspectral imagery processing and ecological monitoring.

Xue Wang is currently a postdoctoral researcher with East China Normal University, Shanghai, China. He received his BS degree in geographic information system and his PhD in photogrammetric and remote sensing from CUMT, Xuzhou, China, in 2014 and 2019, respectively. His research interests include the hyperspectral imagery processing, deep learning, and ecological monitoring.

Zhaoxian Liu is currently an engineer with the Second Surveying and Mapping Institute of Hebei, Shijiazhuang, China. He received his BS degree in geographic information system from North China University of Science and Technology, Tangshan, China, in 2010, and his master's degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2012.

Jianwei Ding is currently a professorate senior engineer with the Second Surveying and Mapping Institute of Hebei, Shijiazhuang, China. He received his BE degree in geomatics engineering from Wuhan University, Wuhan, China, in 2001, and his master's degree in geomatics engineering from CUMT, Jiangsu, China, in 2015.

Kun Tan is currently a professor with East China Normal University, Shanghai, China. He received his BS degree in information and computer science from Hunan Normal University, Hunan, China, in 2004, and his PhD in photogrammetric and remote sensing from CUMT, Jiangsu, China, in 2010. His research interests include hyperspectral image classification and detection, spectral unmixing, quantitative inversion of land surface parameters, and urban remote sensing.

Yu Chen is currently an adjunct professor with CUMT. She received her master's degree in photogrammetry and remote sensing from the CUMT, Xuzhou, China, in 2012, and her PhD in earth and planetary science from the University of Toulouse, Toulouse, France, in 2017. Her current research focuses on SAR interferometry, with particular emphasis on its application for geophysical studies.