

Article

Hyperspectral Inversion of Chromium Content in Soil Using Support Vector Machine Combined with Lab and Field Spectra

Yun Xue ^{1,2}, Bin Zou ^{2,3,*}, Yimin Wen ⁴, Yulong Tu ^{2,3} and Liwei Xiong ^{1,2}

¹ School of Municipal and Surveying Engineering, Hunan City University, Yiyang 413000, China; viviengis@csu.edu.cn (Y.X.); xinfang@csu.edu.cn (L.X.)

² Key Laboratory of Metallogenic Prediction of Nonferrous Metals and Geological Environment Monitoring (Central South University), Ministry of Education, Changsha 410083, China; 165011045@csu.edu.cn

³ School of Geosciences and Info-Physics, Central South University, Changsha 410083, China

⁴ School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin 541004, China; ymwen@guet.edu.cn

* Correspondence: 210010@csu.edu.cn

Received: 1 May 2020; Accepted: 28 May 2020; Published: 29 May 2020



Abstract: Chromium is not only an essential trace element for the growth and development of living organisms; it is also a heavy metal pollutant. Excessive chromium in farmland soil will not only cause harm to crops, but could also constitute a serious threat to human health through the cumulative effect of the food chain. The determination of heavy metals in tailings of farmland soil is an essential means of soil environmental protection and sustainable development. Hyperspectral remote sensing technology has good characteristics, e.g., high speed, macro, and high resolution, etc., and has gradually become a focus of research to determine heavy metal content in soil. However, due to the spectral variation caused by different environmental conditions, the direct application of the indoor spectrum to conduct field surveys is not effective. Soil components are complex, and the effect of linear regression of heavy metal content is not satisfactory. This study builds indoor and outdoor spectral conversion models to eliminate soil spectral differences caused by environmental conditions. Considering the complex effects of soil composition, we introduce a support vector machine model to retrieve chromium content that has advantages in solving problems such as small samples, non-linearity, and a large number of dimensions. Taking a mining area in Hunan, China as a test area, this study retrieved the chromium content in the soil using 12 combination models of three types of spectra (field spectrum, lab spectrum, and direct standardization (DS) spectrum), two regression methods (stepwise regression and support vector machine regression), and two factors (strong correlation factor and principal component factor). The results show that: (1) As far as the spectral types are concerned, the inversion accuracy of each combination of the field spectrum is generally lower than the accuracy of the corresponding combination of other spectral types, indicating that field environmental interference affects the modeling accuracy. Each combination of DS spectra has higher inversion accuracy than the corresponding combination of field spectra, indicating that DS spectra have a certain effect in eliminating soil spectral differences caused by environmental conditions. (2) The inversion accuracy of each spectrum type of SVR_SC (Support Vector Regression_Strong Correlation) is the highest for the combination of regression method and inversion factor. This indicates the feasibility and superiority of inversion of heavy metals in soil by a support vector machine. However, the inversion accuracy of each spectrum type of SVR_PC (Support Vector Regression_Principal Component) is generally lower than that of other combinations, which indicates that, to obtain superior inversion performance of SVR, the selection of characteristic factors is very important. (3) Through principal component regression analysis, it is found that the pre-processed spectrum is more stable for the inversion of Cr concentration.

The regression coefficients of the three types of differential spectra are roughly the same. The five statistically significant characteristic bands are mostly around 384–458 nm, 959–993 nm, 1373–1448 nm, 1970–2014 nm, and 2325–2400 nm. The research results provide a useful reference for the large-scale normalization monitoring of chromium-contaminated soil. They also provide theoretical and technical support for soil environmental protection and sustainable development.

Keywords: chromium content; support vector machine; hyperspectral remote sensing

1. Introduction

Soils contribute to basic human needs like food, clean water, and clean air, and they are a major carrier for biodiversity. Healthy soils and healthy land are the basic conditions for the successful implementation and realization of the sustainable development goals (SDGs) [1,2]. In the development and utilization of mineral resources, human activities such as mining, transportation, sewage treatment, and fertilization have posed a continuous threat to soil health, and inevitably bring many environmental problems and disasters [3]. Among them, heavy metal pollution of tailings farmland soil is one of the most serious problems in the ecological environment of mining areas [4]. Chromium is not only an essential trace element for the growth and development of living organisms; it is also a heavy metal pollutant. Excessive chromium in farmland soil will not only cause harm to crops, but can also cause a serious threat to human health through the cumulative effect of the food chain. Therefore, the determination of quantities of heavy metals in tailings of farmland soil is a necessary means of soil environmental protection and sustainable development [5,6].

Although traditional methods of heavy metal detection in soil are highly accurate, they have low efficiency and high cost, and are only suitable for small-scale monitoring. Hyperspectral remote sensing technology has the advantages of high efficiency, low cost, and suitability for macro-monitoring. It provides a possibility for the rapid prediction of heavy metal content in soil [7–13]. However, due to the low content of heavy metals in the soil, the response to the soil spectral curve is weak. It is difficult to estimate the content by directly analyzing the characteristic spectra of heavy metal elements in soil samples.

So far, a large number of studies have shown that the spectra obtained under laboratory conditions can quantitatively estimate the concentration of chromium in the soil to a certain extent, thanks to well-controlled measurement conditions. However, due to the pretreatment of soil samples, such as air-drying, grinding, and controlling the spectral measurement conditions in the laboratory, the spectra obtained in the laboratory and the spectra obtained in the field are inevitably disturbed. In fact, the soil spectra measured under field conditions are affected by many factors, such as soil moisture, soil particle size, soil surface conditions, solar radiation, ambient light, and temperature. This is significantly different from the corresponding factors that affect the spectra obtained by the laboratory. Comparing the hyperspectral inversion models of soil clay, calcium carbonate, and salt content based on laboratory spectra and field spectra, it is known that the inversion model of soil composition based on laboratory spectra is often difficult to directly apply to the field. Direct standardization (DS) algorithm is a common spectral correction method, which has been successfully used to estimate the content of soil organic matter and clay components [14–19].

In theory, the estimation of soil chromium concentration based on hyperspectral data depends on its correlation with soil spectral reflectance. In naturally contaminated soil samples, various soil properties (such as soil type, soil organic matter, and iron-containing oxide properties) strongly affect the spectral reflectance. In addition, multiple heavy metals may be present in the soil at the same time. Therefore, the spectral difference caused by chromium content is difficult to separate from complex factors. The linear model constructed by traditional precise mathematical methods is used to retrieve the nonlinear relationship between chromium content and the reflection spectra, which

affects its predictive ability. In contrast, machine learning methods continuously learn feedback errors during model correction and improve the complex relationship between independent and dependent variables, which is an effective way to solve nonlinear regression problems. A support vector machine is a new type of machine learning algorithm developed on the basis of statistical learning theory. The algorithm learns according to the principle of structural risk minimization. It is a method in which decision rules obtained from limited training samples can still obtain small errors for independent test sets. There are many successful cases in hyperspectral regression [8,20–26].

Due to the differences in soil spectrum caused by different indoor and outdoor environmental conditions, the direct application of an indoor spectrum to investigate field pollution is not effective. Soil composition is complex, and the linear regression effect of heavy metal content is not ideal. In this study, a soil indoor–outdoor spectral conversion model was constructed to perform spectral conversion of the outdoor spectrum of soil samples to eliminate soil spectral differences caused by environmental conditions. Considering the complex effects of soil components, a support vector machine model that has advantages in solving problems such as small samples, nonlinearity, and high dimensions, is introduced to perform the inversion of chromium content [27–30].

Taking a mining area in Hunan, China as our test case, a strong correlation factor and principal component factor of the field spectrum, DS spectrum, and lab spectrum were used as input variables, and the soil chromium content was used as the dependent variable. A support vector machine combined with indoor and outdoor spectra was used to retrieve the chromium content. The results are expected to provide a useful reference for the use of hyperspectral data to carry out large-scale normalization monitoring of chromium-contaminated soil and provide theoretical and technical support for soil environmental protection and sustainable development.

2. Materials and Methods

2.1. Soil Sample Collection and Spectral Determination

A typical heavy metal-contaminated mining area is located in Hunan Province, China. Long-term mining activities have caused serious heavy metal pollution to the surrounding paddy fields, vegetable fields, and other soils. Taking into account a variety of pollution diffusion factors in the study area, a strip sampling route was set up, and 46 samples of natural contaminated soil surface (0–20 cm) were collected by the grid sampling method (Figure 1). The land-use type of the sampling site was mainly agricultural, although a small number of samples were from construction land and mining soil. Based on removing impurities such as stones and grass, the PSR-3500 field portable spectrometer (spectral wavelength range 350–2500 nm) is used for field spectrum measurement. Before the spectrum measurement, it is necessary to measure the white board, perform calibration, obtain absolute reflectance, and then use a spectrometer to collect the soil spectrum. To ensure the validity of the spectrum, the average value of the 10 spectra of each soil sample point is taken as the field spectrum of the sample point. The soil samples were taken back to the laboratory for chromium content determination and lab spectrum measurement.

2.2. Field Spectral Conversion

Due to the variation in the soil spectra due to different environmental conditions, it is difficult to effectively apply the indoor spectrum directly to a large-scale survey of soil heavy metal pollution in the wild. To this end, a suitable conversion sample set is selected to construct a conversion model of the lab and field spectra [31–33]. The specific steps are as follows: (1) Remove the influence band of water vapor and the band with a small signal-to-noise ratio, and obtain the effective soil spectral data of the lab and field spectrum. (2) Using the Kennard–Stone (KS) algorithm to calculate the Euclidean distance between the spectra, select the sample of the transform set of the field spectrum and the lab spectrum. The specific implementation steps of the KS algorithm are: First, select the two samples with the largest distance between the two samples as the first and second transform-set samples. Then, calculate

the distance between the remaining samples and the selected samples. For each remaining sample, the shortest distance from the selected sample is selected. Then, select the sample corresponding to the longest of these shortest distances as the third transform-set sample. Repeat until the number of selected transformation set samples equals a predetermined number. (3) The DS algorithm is used to construct the transformation model of the field and lab spectra of the selected transformation set samples. (4) Use the constructed conversion model to perform a spectral conversion on the field spectrum to obtain the converted field spectrum data (DS spectrum).

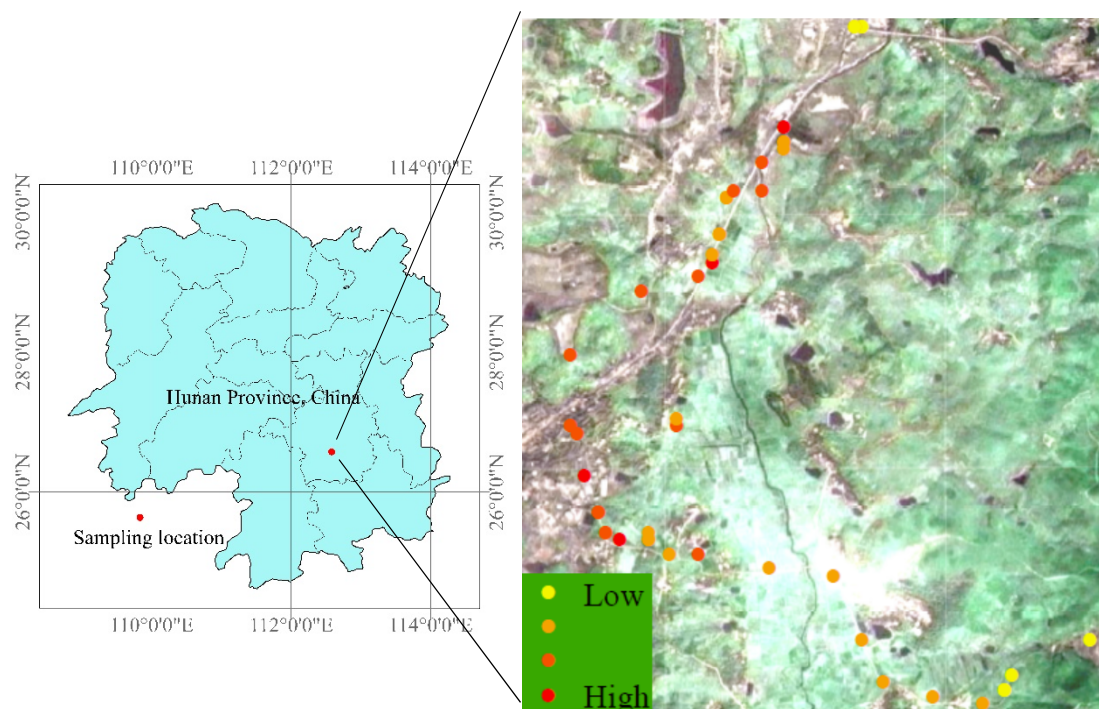


Figure 1. Sampling point location map.

2.3. Spectra Pretreatment

During the process of collecting and transmitting spectral signals, some noise will be generated, which makes the signal-to-noise of spectral data relatively low. The preprocessing of spectral data can reduce its impact. The original hyperspectra can be transformed to eliminate background noise, enhance differences, and highlight spectral features. In this study, the bands with low signal-to-noise ratios from 350–380 nm, 2410–2500 nm, and those affected by water vapor, 1880–1965 nm, were removed. Hyperspectral data have high dimensionality, with information redundancy and high correlation between adjacent bands. Each spectrum was resampled at 10 nm intervals to reduce the correlation between bands and improve data processing efficiency. First-order differential processing is performed on each spectrum to eliminate the spectral shift caused by water absorption to a certain extent, amplify weak spectral information, and improve multicollinearity.

2.4. Inversion Factor Selection

2.4.1. Strong Correlation Factor (SC)

Correlation analysis can reflect the strength of the linear relationship between the two variables. The Pearson correlation coefficient between the reflectance and heavy metal content in this study can be used to indicate the strong correlation band of the difference in soil heavy metal content. A band with a significance level of 0.05 during the t-test of the correlation coefficient is a strong correlation band.

2.4.2. Principal Component Factor (PC)

Through principal component transformation, this study transforms the pre-processed spectrum into linearly independent principal component variables. The principal component transform converts the pre-processed spectrum into a linearly independent principal component variable. To avoid the false rejection of effective weak signals in the regression modeling process, all of the extracted principal components are retained as model-independent variables to be used as another factor for the inversion of chromium content.

2.5. Model Building

2.5.1. Stepwise Regression (SR)

Stepwise regression is a type of multiple linear regression that can select the best-fitting combination of independent variables for dependent variable prediction with forward-adding and backward-deleting variables. This method introduces variables one by one into the model. F-tests are performed after each explanatory variable is introduced, and t-tests are performed on the selected explanatory variables one by one. When the originally-introduced explanatory variable becomes less significant due to the introduction of later explanatory variables, it is deleted. This ensures that only significant variables are included in the regression equation before each new variable is introduced.

2.5.2. Support Vector Regression (SVR)

Support vector regression has received increasing attention from researchers due to its good generalization performance. SVR is an algorithm that minimizes structural risk as a regression goal. The basic idea is to use a kernel function to transform the nonlinear regression in the original sample space into a linear regression in a high-dimensional space. The penalty coefficient C and the kernel parameter Γ are related to the prediction performance of the SVR. Selecting appropriate parameters can improve the generalization ability and prediction accuracy of the SVR. The Grey Wolf Optimizer (GWO) is an optimized search method, developed by the process of predation behavior of the gray wolf group. The algorithm has good self-organized learning, simple parameters, easy implementation, fast convergence speed, and strong global search ability. In recent years, GWO has received extensive attention from researchers, and has been successfully applied in many fields. In this study, we chose GWO to optimize SVR parameters.

The coefficient of determination (R^2), the root mean square error (RMSE) of the modeled samples, the root mean square error (RMSEp), and the relative analysis error (RPD) of the predictive samples are used as criteria for judging the predictive ability of the model. In general, the larger the R^2 and RPD, the smaller the RMSE and RMSEp, and the higher the model's accuracy.

3. Results

3.1. Descriptive Statistics of Soil Cr Concentrations

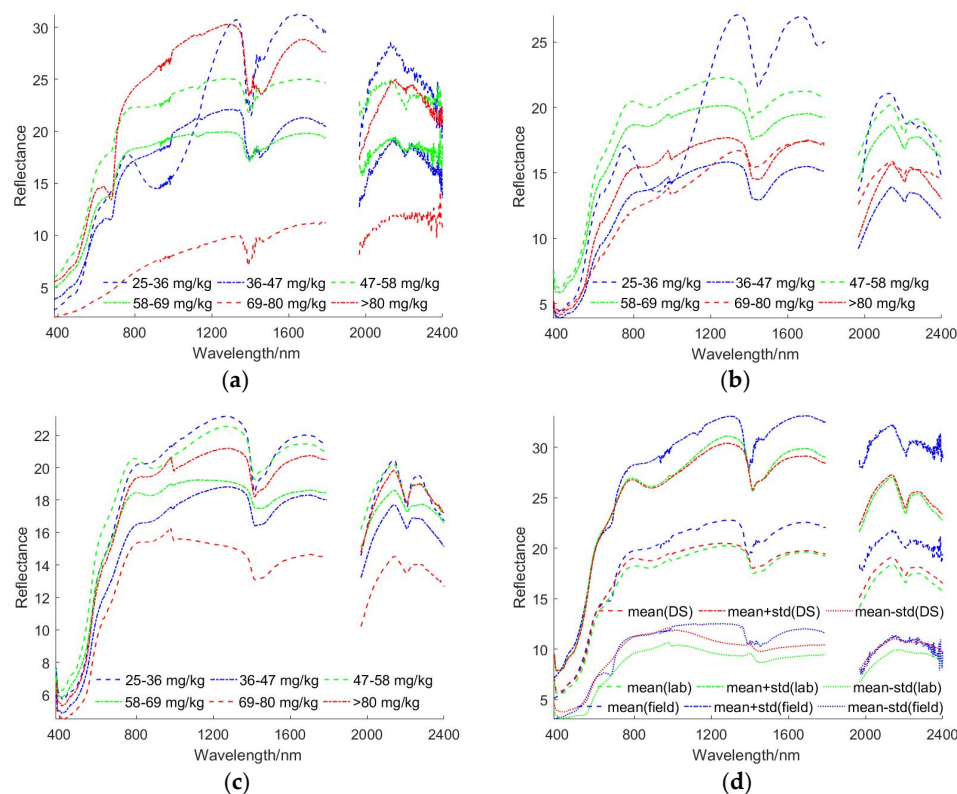
The chromium content in the soil in the study area is low, and the sample average value is lower than the first-class pollution standard (90 mg/kg) of the Chinese soil environmental quality standard (GB15618-1995) (Table 1). However, among the 46 sampling points, the chromium content of one sample point exceeded the standard. The standard deviation of the statistical value of soil chromium content is large, indicating that the pollution in the study area is quite different. The χ^2 -test ($\chi^2 = 2.38 < \chi_{0.05,3}^2, \chi_{0.05,3}^2 = 7.815$) showed that the chromium content data obeyed a normal distribution at a significant level of 0.05. Seventy percent of the 46 samples were randomly selected as training samples, and the remaining 30% were used as test samples.

Table 1. Descriptive statistics of chromium content in soil samples (mg/kg).

Minimum Value	Maximum Value	Mean	Standard Deviation	Skewness
25.47	92.08	60.93	12.70	−0.04

3.2. Spectral Characteristics of the Soil Samples

According to the amount of chromium content, all samples were divided into six groups at equal intervals. The average spectral reflectance of all samples in each group was calculated. The spectral curve is shown in Figure 2a–c. The field spectrum, lab spectrum, and DS spectrum curve shapes are roughly the same, and the curves are approximately parallel. The reflectance in the visible light band is lower than that in the near-infrared band, and the spectral difference is also slightly smaller than in the near-infrared band. The positions of the characteristic absorption bands are roughly the same, but the absorption depth is slightly different. Comparing Figure 2a–c, the direction of the field spectrum is chaotic and changes drastically, while the direction of the lab spectrum and the DS spectrum are more consistent, and the changes are more stable. As shown in Figure 2d–g, the mean and standard deviation of the lab spectrum, DS spectrum, and field spectrum are closer to each other after preprocessing. As shown in Figure 2h, the differences between the field spectrum and the lab spectrum and the difference between the field spectrum and the DS spectrum are both large, indicating that the field environment has a significant impact on the spectrum of Cr-contaminated soil. In the visible-light region, the three types of reflectance are almost the same, but the reflectance of the field spectrum in the near-infrared region is significantly higher than the reflectance of the lab spectrum and the DS spectrum. As shown in Figure 2i, the difference of the pre-processed spectra is significantly reduced, which indicates that pre-processing can effectively improve the quality of the spectral data participating in subsequent modeling.

**Figure 2.** Cont.

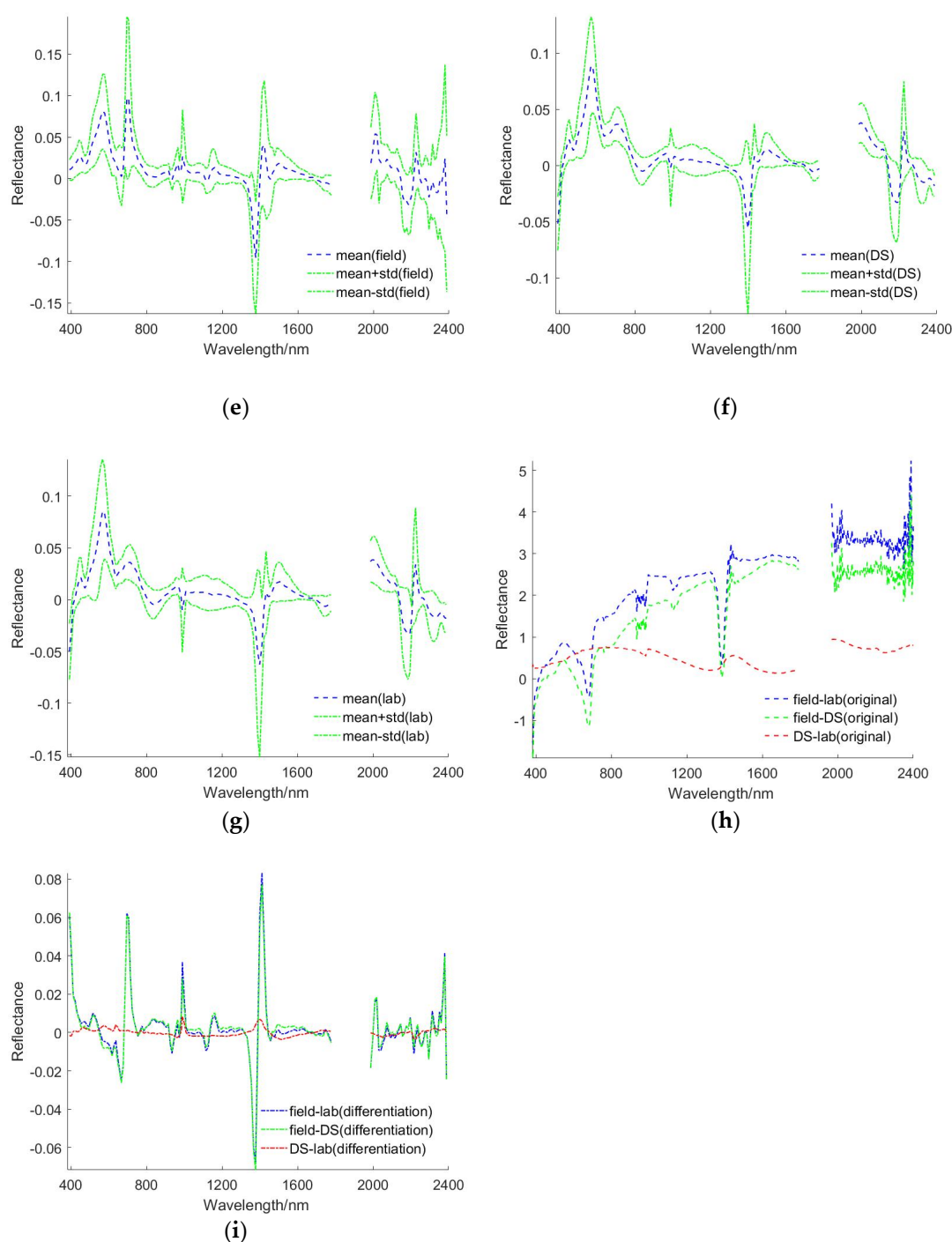


Figure 2. Spectral characteristics of the soil samples. (a) Grouped field spectrum curve of chromium content; (b) grouped lab spectrum curve of chromium content; (c) grouped direct standardization (DS) spectrum curve of chromium content; (d) the mean and standard deviation of the original spectrum; (e) the mean and standard deviation of field differential spectrum; (f) the mean and standard deviation of DS differential spectrum; (g) the mean and standard deviation of lab differential spectrum; (h) the spectral difference of the original spectrum; (i) the spectral difference of the differential spectrum.

3.3. Correlation Analysis

Table 2 and Figure 3 show the Pearson correlation coefficients of the Cr content and the three types of the spectrum (field spectrum, lab spectrum, and DS spectrum). After resampling, the number of bands of the three types of spectra is 192, and the number of strongly correlated bands of the field spectrum, lab spectrum, and DS spectrum are 65, 132, and 108, respectively. In particular, the number of strongly correlated bands with a correlation coefficient greater than 0.5 is 8, 44, and 30, respectively. The results show that the field spectrum has fewer strong correlation bands due to the large environmental impact, the lab spectrum has more strong correlation bands due to the small environmental impact, and the DS spectrum is centered on the number of strong correlation bands due to the indoor and outdoor environment. The strong correlation bands of the three types of spectra were used as the inversion factors of chromium content. A field_SC (strong correlation band of the field spectrum) with an absolute correlation coefficient greater than 0.5 appears near positions 1032, 1044, 1172, 1218, 1510, 1554, 1576, and 1746 nm. A lab_SC (strong correlation band of the lab spectrum) with an absolute correlation coefficient greater than 0.5 appears near 676–744, 973, 1218–1230, 1466–1641, 1736, 1987–2093, 2206, 2351–2388 nm. A DS_SC (strong correlation band of the DS spectrum) with an absolute correlation coefficient greater than 0.5 appears near 391, 400, 587–765, and 1987–2125 nm.

Table 2. Number of strong correlation factors for different spectral types.

Spectral Type	Number of Strong Correlation Factors	Number of Strong Correlation Factors with a Correlation Coefficient Greater than 0.5
Field spectrum	65	8
Lab spectrum	132	44
DS spectrum	108	30

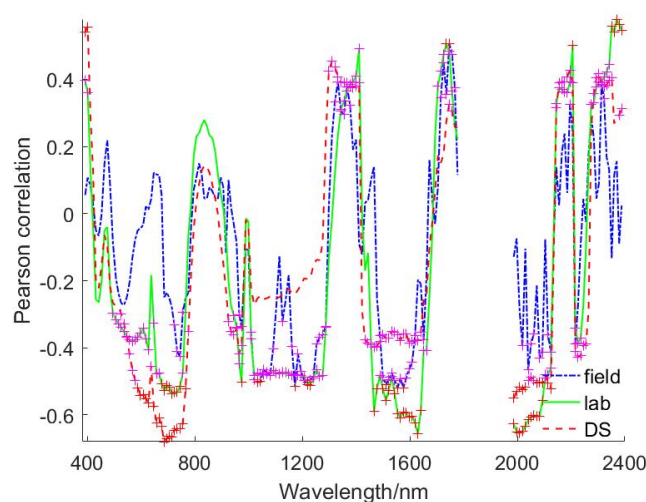


Figure 3. Correlation analysis between differential spectrum and chromium content. +Dots represent strong correlations.

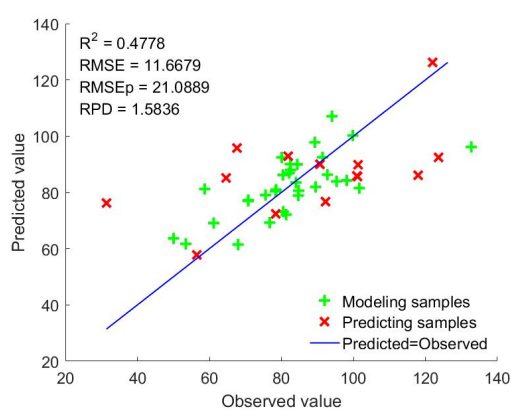
3.4. Inversion Modeling

Table 3 and Figure 4 are the results of inversion using a combination model of three types of spectra (field spectrum, lab spectrum, and DS spectrum), two regression methods (stepwise regression and support vector machine regression), and two factors (strong correlation factor and principal component factor). The abbreviations of the models are defined in Table 3. The results show that, in terms of the combination of regression methods and inversion factors, the inversion accuracy of each spectral type of SVR_SC is higher than that of the corresponding spectral types of SR_SC, SR_PC, and SVR_PC. We proved the feasibility and superiority of support vector machine for the inversion of

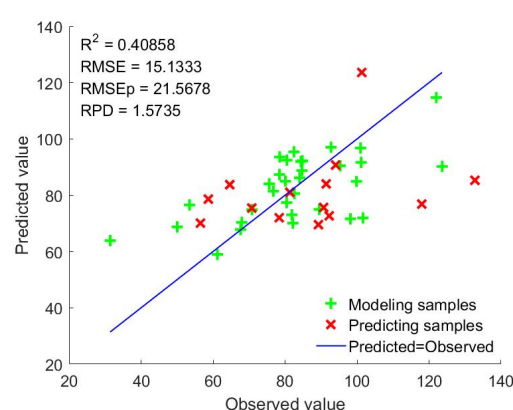
soil heavy metals, which provides a reference for selecting the inversion method of soil chromium content. However, the inversion accuracy of each spectrum type of SVR_PC is generally lower than that of other combinations, which indicates that, to obtain superior inversion performance of SVR, the selection of characteristic factors is very important. As far as the type of spectrum is concerned, the inversion accuracy of each combination of field spectra is generally lower than the accuracy of the corresponding combination. It shows that the field environmental interference caused some factors of the field spectrum to lose their original physical interpretation power. This resulted in the real effective factors not being recognized by the model during the modeling process, which ultimately affects the modeling accuracy. The inversion accuracy of each combination of the DS spectrum is higher than that of the corresponding combination of the field spectrum. The R-square of the DS_SVR_SC combination reaches 0.98, and the RMSE and RMSEP are 3.21 and 11.91, respectively. This shows that the DS spectrum has a certain role in eliminating soil spectral differences caused by environmental conditions.

Table 3. Inversion results.

Spectral Type	Inversion Method	Factor	Model Abbreviation	R ²	RMSE	RMSEP	RPD
field	SR	SC	field_SR_SC	0.48	11.67	21.09	1.58
		PC	field_SR_PC	0.41	15.13	21.57	1.57
	SVR	SC	field_SVR_SC	0.86	7.64	17.95	1.15
		PC	field_SVR_PC	0.55	19.62	15.54	21.95
lab	SR	SC	lab_SR_SC	0.56	14.27	13.07	1.25
		PC	lab_SR_PC	0.49	13.16	19.89	1.71
	SVR	SC	lab_SVR_SC	0.94	5.49	16.11	1.11
		PC	lab_SVR_PC	0.29	15.47	22.18	6.24
DS	SR	SC	DS_SR_SC	0.55	14.59	13.97	1.32
		PC	DS_SR_PC	0.59	12.41	25.65	1.57
	SVR	SC	DS_SVR_SC	0.98	3.21	11.91	1.06
		PC	DS_SVR_PC	0.49	12.94	18.29	1.34



(a)



(b)

Figure 4. Cont.

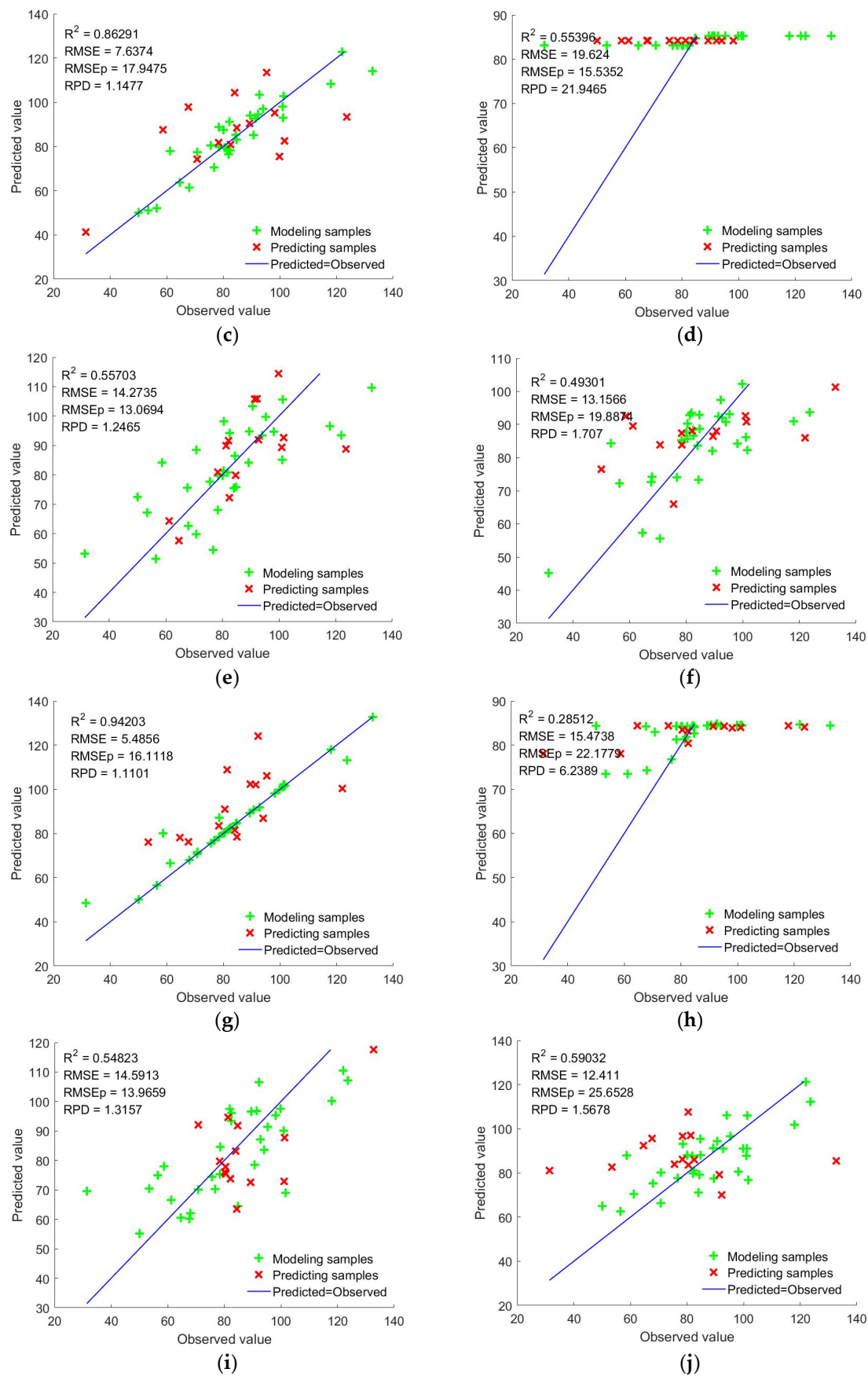


Figure 4. Cont.

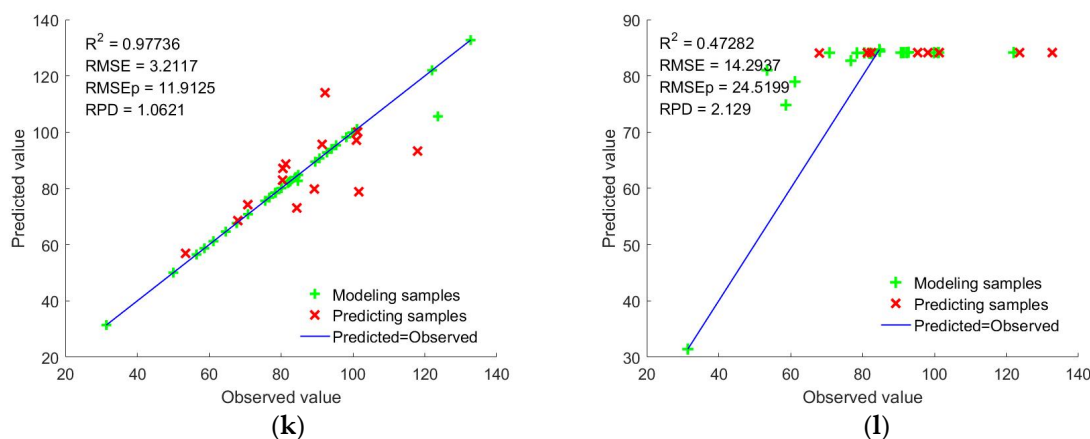


Figure 4. Scatter plot of the inversion of Cr content. (a) field_SR_SC; (b) field_SR_PC; (c) field_SVR_SC; (d) field_SVR_PC; (e) lab_SR_SC; (f) lab_SR_PC; (g) lab_SVR_SC; (h) lab_SVR_PC; (i) DS_SR_SC; (j) DS_SR_PC; (k) DS_SVR_SC; (l) DS_SVR_PC. SR: Stepwise Regression; SC: Strong Correlation; PC: Principal Component; SVR: Support Vector Regression.

4. Discussion

The principal component regression modeling method inversely transforms the principal component regression coefficients to obtain each band's contribution to the model, thereby identifying characteristic bands. Figure 5a–f shows that the standard deviation range of the regression coefficient of the pre-processed model is smaller than that of the original model, which indicates that the pre-processed model is more stable for the inversion of Cr concentration. The regression coefficients of the principal component regression model of the field spectrum have significant peaks around 934–990 nm, 1373–1437 nm, 1970 nm, 2008 nm, and 2332–2400 nm. The regression coefficients of the principal component regression model of the lab spectrum have obvious peaks around 390–400 nm and 969–993 nm. The regression coefficient of the principal component regression model of the DS spectrum has obvious peaks around 384–409 nm and 959–991 nm. Figure 5g shows that the regression coefficient curves of the three types of original spectra are inconsistent, the peaks are scattered, and the characteristic band is not obvious. Figure 5h shows that the regression coefficients of the three types of differential spectra are roughly the same, and the reflection peaks are mainly around 384–458 nm, 959–993 nm, 1373–1448 nm, 1970–2014 nm, and 2325–2400 nm. These five characteristic bands are of statistical significance.

Compared with previous studies, the characteristic bands of chromium in the soil identified in this study have certain similarities and differences. The similarities are as follows. A previous study found the characteristic bands of chromium in the soil to be 1430, 1439, 1440, 2226, 2228, and 2230 nm [34], which are basically within the characteristic band intervals found in this study. Most of the previous work focused on the study of Cu and Pb in the soil, and there has been relatively little research on chromium in the soil. The five statistically significant characteristic bands found in this study provide a basis for the future inversion of chromium content in the soil.

Healthy soils and healthy land are the basic conditions for successful implementation and realization of the SDGs [2,35]. The rapid and accurate determination of heavy metals in soil provides technical support for soil environmental protection and sustainable development. Following previous research on the hyperspectral inversion of heavy metal content, this study explored the possibility of quantitative inversion of chromium content based on soil spectra. Compared to the previous studies, its novelty lies in the use of a multi-source spectrum (field spectrum, lab spectrum, and DS spectrum), two types of regression methods (SR and SVR), and two types of factors (SC and PC) to invert soil chromium content. The performance of each combined model is compared to provide a reference for the selection of modeling spectrum types, modeling methods, and modeling factors for inversion of chromium content in soil based on hyperspectral data.

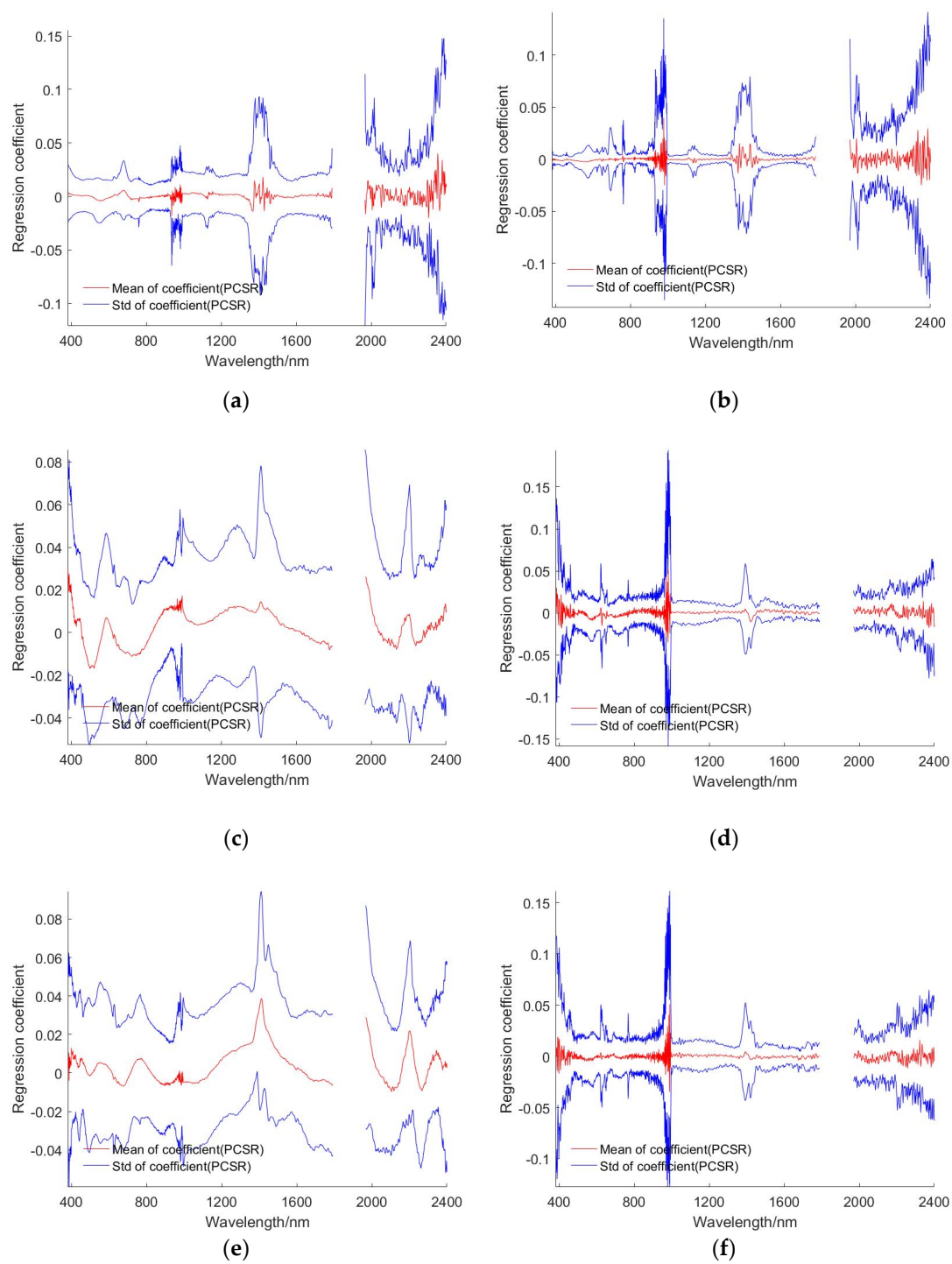


Figure 5. Cont.

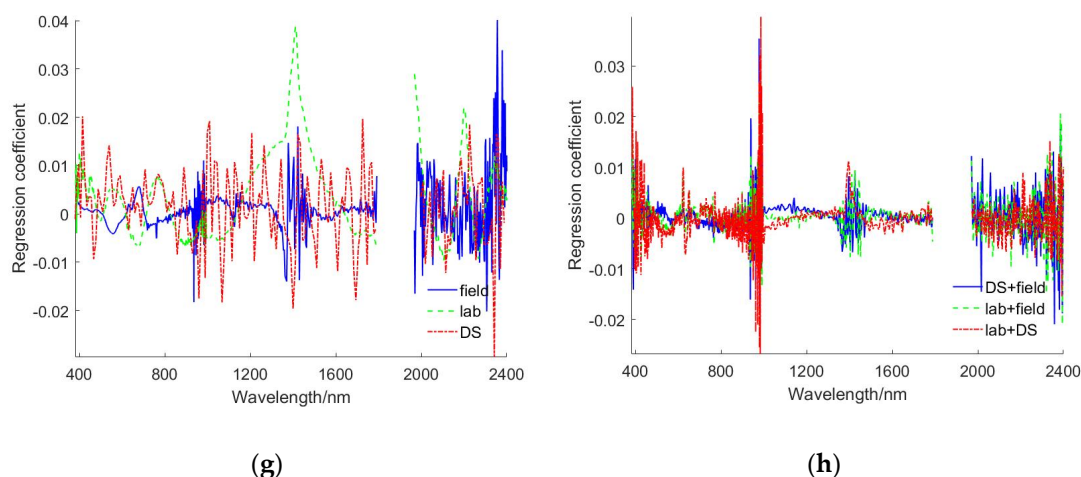


Figure 5. Feature band recognition. (a) Regression coefficients of the field spectrum; (b) regression coefficients of the field differential spectrum; (c) regression coefficients of the DS spectrum; (d) regression coefficients of the DS differential spectrum; (e) regression coefficients of the lab spectrum; (f) regression coefficients of the lab differential spectrum; (g) regression coefficients of the original spectrum; (h) regression coefficients of the differential spectrum.

Through the analysis of the differences among the lab spectrum, the DS spectrum, and the field spectrum, the correlation analysis of the three types of spectra and chromium content, and the inversion results of each combined model, it is shown that the different indoor and outdoor environmental conditions will cause variation in the soil spectrum. This will affect the accurate determination of heavy metal content, similar to that of previous studies [32,33]. In this study, 12 combined models were established to invert the chromium content, and the inversion accuracy of each spectral type of SVR_SC was the highest, indicating the feasibility and superiority of inversion of heavy metal content in soil by support vector machine. However, the inversion accuracy of each spectral type of SVR_PC is generally lower than other combinations, which indicates that the selection of characteristic factors is crucial for SVR to obtain superior inversion performance. In addition, many new methods of machine learning (e.g., random forest, extreme learning machine, convolutional neural network, etc.) can continuously feedback errors during the training of the model, and improve the complex relationship between the independent variable and the dependent variable, which is an effective method for solving nonlinear regression problems [20,21,36]. Therefore, stable and robust machine learning methods are further needed for high accuracy inversion of chromium content in soil based on hyperspectral data recently.

Finally, the ultimate goal of developing hyperspectral remote sensing technology is to achieve low-cost, high-precision, and high-efficiency quantitative estimation of soil heavy metal content. Although this study initially verified the theoretical feasibility of our method, there are still some problems that have not yet been explored and countered in practical applications, such as soil thickness, soil moisture, vegetation coverage, atmospheric absorption, and sunlight.

5. Conclusions

Taking a mining area in Hunan, China as a test area, this study retrieved the chromium content in soil using 12 combination models of three types of spectra (field spectrum, lab spectrum, and DS spectrum), two regression methods (stepwise regression and support vector machine regression), and two factors (strong correlation factor and principal component factor). The results show that: (1) The differences between the field spectrum and the lab spectrum and the difference between the field spectrum and the DS spectrum are both large, indicating that the field environment has a significant effect on the spectrum of Cr-contaminated soil. (2) The field spectrum has fewer strong correlation bands due to the large environmental impact, the lab spectrum has more strong correlation bands due to the small environmental impact, and the DS spectrum is centered on the number of strong

correlation bands due to the indoor and outdoor environment. (3) In terms of the combination of the regression method and inversion factor, the inversion accuracy of each spectral type of SVR_SC is higher than that of each corresponding spectral type of SR_SC, SR_PC, and SVR_PC. This shows the feasibility and superiority of support vector machine in the inversion of heavy metals in soil. However, the inversion accuracy of each spectrum type of SVR_PC is generally lower than that of other combinations, which indicates that, in order to obtain superior inversion performance of SVR, the selection of characteristic factors is very important. (4) As far as the type of spectrum is concerned, the inversion accuracy of each combination of the field spectrum is generally lower than the accuracy of the corresponding combination, which indicates that the field environmental interference has an impact on the accuracy of the model. Each combination of DS spectra has higher inversion accuracy than the corresponding combination of the field spectra, indicating that the DS spectra have a certain role in eliminating soil spectral differences caused by environmental conditions. (5) The standard deviation range of the regression coefficient of the pre-processed model is smaller than that of the original model, which indicates that the pre-processed model is more stable for the inversion of Cr concentration. The regression coefficient curves of the three types of original spectra are inconsistent, the peaks are scattered, and the characteristic band is not obvious. The regression coefficients of the three types of differential spectra are roughly the same, and the reflection peaks are mainly around 384–458 nm, 959–993 nm, 1373–1448 nm, 1970–2014 nm, and 2325–2400 nm. These five characteristic bands are of statistical significance.

Author Contributions: Provided fund support, Y.X.; provided technical guidance, B.Z.; conceived the methodology, Y.W. and Y.T.; participated in the revision, L.X. and B.Z.; made the data processing, the analysis of them, and the writing of the paper, Y.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was financially supported by Hunan Province Engineering & Technology Research Center for Rural Water Quality Safety (Grant No.2019TP2079), the research program of Hunan Province Science and Technology Department (Grant No. 2017SK2271), Open Research Fund Program of Key Laboratory of Metallogenic Prediction of Nonferrous Metals and Geological Environment Monitoring (Central South University), Ministry of Education (Grant No. 2018YSJS02).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Visser, S.M.; Keesstra, S.; Maas, G.; De Cleen, M.; Molenaar, C. Soil as a basis to create enabling conditions for transitions towards sustainable land management as a key to achieve the SDGs by 2030. *Sustainability* **2019**, *11*, 6792. [[CrossRef](#)]
2. Keesstra, S.; Bouma, J.; Wallinga, J.; Tuttonell, P.; Smith, P.; Cerda, A.; Montanarella, L.; Quinton, J.; Pachepsky, Y.; Van Der Putten, W.H.; et al. The significance of soils and soil science towards realization of the United Nations sustainable development goals. *Soil* **2016**, *2*, 111–128. [[CrossRef](#)]
3. Huang, Y.; Wang, L.; Wang, W.; Li, T.; He, Z.; Yang, X. Current status of agricultural soil pollution by heavy metals in China: A meta-analysis. *Sci. Total Environ.* **2019**, *651*, 3034–3042. [[CrossRef](#)] [[PubMed](#)]
4. Li, Z.; Ma, Z.; Van Der Kuip, T.J.; Yuan, Z.; Huang, L. A review of soil heavy metal pollution from mines in China: Pollution and health risk assessment. *Sci. Total Environ.* **2014**, *468*, 843–853. [[CrossRef](#)] [[PubMed](#)]
5. Bhuiyan, M.A.; Parvez, L.; Islam, M.; Dampare, S.B.; Suzuki, S. Heavy metal pollution of coal mine-affected agricultural soils in the northern part of Bangladesh. *J. Hazard. Mater.* **2010**, *173*, 384–392. [[CrossRef](#)] [[PubMed](#)]
6. Rodríguez, L.; Ruiz, E.; Alonso-Azcárate, J.; Rincon, J. Heavy metal distribution and chemical speciation in tailings and soils around a Pb–Zn mine in Spain. *J. Environ. Manag.* **2009**, *90*, 1106–1116. [[CrossRef](#)]
7. Shen, Q.; Xia, K.; Zhang, S.; Kong, C.; Hu, Q.; Yang, S. Hyperspectral indirect inversion of heavy-metal copper in reclaimed soil of iron ore area. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2019**, *222*, 117191. [[CrossRef](#)]
8. Tao, C.; Cui, W.; Wang, Y.; Zou, B.; Zou, Z. Soil heavy metal qualitative classification model based on hyperspectral measurements and transfer learning. *Spectrosc. Spectr. Anal.* **2019**, *39*, 2602–2607.
9. Tan, K.; Ma, W.; Wu, F.; Du, Q. Random forest-based estimation of heavy metal concentration in agricultural soils with hyperspectral sensor data. *Environ. Monit. Assess.* **2019**, *191*, 446. [[CrossRef](#)]

10. Liu, Z.; Lu, Y.; Peng, Y.; Zhao, L.; Wang, G.; Hu, Y. Estimation of soil heavy metal content using hyperspectral data. *Remote Sens.* **2019**, *11*, 1464. [\[CrossRef\]](#)
11. Shen, Q.; Zhang, S.; Ge, C.; Liu, H.; Zhou, Y.; Chen, Y.; Hu, Q.; Ye, H.; Huang, Y. Hyperspectral inversion of heavy metal content in soils reconstituted by mining wasteland. *Spectrosc. Spectr. Anal.* **2019**, *39*, 1214–1221.
12. Wang, F.; Gao, J.; Zha, Y. Hyperspectral sensing of heavy metals in soil and vegetation: Feasibility and challenges. *ISPRS J. Photogramm. Remote Sens.* **2018**, *136*, 73–84. [\[CrossRef\]](#)
13. Kumar, V.; Sharma, A.; Kaur, P.; Sidhu, G.P.S.; Bali, A.S.; Bhardwaj, R.; Thukral, A.K.; Cerda, A. Pollution assessment of heavy metals in soils of India and ecological risk assessment: A state-of-the-art. *Chemosphere* **2018**, *216*, 449–462. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Zou, B.; Jiang, X.; Feng, H.; Tu, Y.; Tao, C. Multisource spectral-integrated estimation of cadmium concentrations in soil using a direct standardization and Spiking algorithm. *Sci. Total Environ.* **2019**, *701*, 134890. [\[CrossRef\]](#)
15. Liu, C.; Li, T.; Wei, L.; Xu, Y.; Wu, J. Research on application of direct standardization algorithm in near-infrared spectrum calibration transfer of acid value and peroxide value of edible oil. *Spectrosc. Spectr. Anal.* **2017**, *37*, 3042–3050.
16. Xi, C.; Feng, Y.; Hu, C. Evaluation of piecewise direct standardization algorithm for near infrared quantitative model updating. *Chin. J. Anal. Chem.* **2014**, *42*, 1307–1313.
17. Zou, B.; Tu, Y.; Jiang, X.; Tao, C.; Zhou, M.; Xiong, L. Estimation of Cd content in soil using combined laboratory and field DS spectroscopy. *Spectrosc. Spectr. Anal.* **2019**, *39*, 3223–3231.
18. Wang, S.; Han, P.; Song, H.; Liang, G.; Cheng, X. Application of slope/bias and direct standardization algorithms to correct the effect of soil moisture for the prediction of soil organic matter content based on the near infrared spectroscopy. *Spectrosc. Spectr. Anal.* **2019**, *39*, 1986–1992.
19. Fan, P.; Li, X.; Lu, M.; Wu, N.; Liu, Y. Vis-NIR model transfer of total nitrogen between different soils. *Spectrosc. Spectr. Anal.* **2018**, *38*, 3210–3214.
20. Zhou, X.; Sun, J.; Tian, Y.; Lu, B.; Hang, Y.; Chen, Q. Development of deep learning method for lead content prediction of lettuce leaf using hyperspectral images. *Int. J. Remote Sens.* **2019**, *41*, 2263–2276. [\[CrossRef\]](#)
21. Zhou, X.; Sun, J.; Tian, Y.; Lu, B.; Hang, Y.; Chen, Q. Hyperspectral technique combined with deep learning algorithm for detection of compound heavy metals in lettuce. *Food Chem.* **2020**, *321*, 126503. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Tan, K.; Wang, H.; Chen, L.; Du, Q.; Du, P.; Pan, C. Estimation of the spatial distribution of heavy metal in agricultural soils using airborne hyperspectral imaging and random forest. *J. Hazard. Mater.* **2019**, *382*, 120987. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Tian, S.; Wang, S.; Bai, X.; Zhou, D.; Luo, G.; Wang, J.; Wang, M.; Lu, Q.; Yang, Y.; Hu, Z.; et al. Hyperspectral prediction model of metal content in soil based on the genetic ant colony algorithm. *Sustainability* **2019**, *11*, 3197. [\[CrossRef\]](#)
24. Liu, P.; Liu, Z.; Hu, Y.; Shi, Z.; Pan, Y.; Wang, L.; Wang, G. Integrating a hybrid back propagation neural network and particle swarm optimization for estimating soil heavy metal contents using hyperspectral data. *Sustainability* **2019**, *11*, 419. [\[CrossRef\]](#)
25. Qiu, L.; Wang, K.; Long, W.; Wang, K.; Hu, W.; Amable, G.S. A comparative assessment of the influences of human impacts on soil cd concentrations based on stepwise linear regression, classification and regression tree, and random forest models. *PLoS ONE* **2016**, *11*, e0151131. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Vega, F.; Andrade, M.L.; Covelo, E. Influence of soil properties on the sorption and retention of cadmium, copper and lead, separately and together, by 20 soil horizons: Comparison of linear regression and tree regression analyses. *J. Hazard. Mater.* **2010**, *174*, 522–533. [\[CrossRef\]](#)
27. Choubin, B.; Moradi, E.; Golshan, M.; Adamowski, J.; Hosseini, F.S.; Mosavi, A. An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. *Sci. Total Environ.* **2019**, *651*, 2087–2096. [\[CrossRef\]](#)
28. Liu, S.; Feng, Z.-K.; Feng, B.-F.; Min, Y.-W.; Cheng, C.; Zhou, J. Comparison of multiple linear regression, artificial neural network, extreme learning machine, and support vector machine in deriving operation rule of hydropower reservoir. *Water* **2019**, *11*, 88. [\[CrossRef\]](#)
29. Chen, W.; Pourghasemi, H.R.; Naghibi, S.A. A comparative study of landslide susceptibility maps produced using support vector machine with different kernel functions and entropy data mining models in China. *Bull. Eng. Geol. Environ.* **2017**, *77*, 647–664. [\[CrossRef\]](#)

30. Pham, B.T.; Bui, D.T.; Prakash, I. Bagging based support vector machines for spatial prediction of landslides. *Environ. Earth Sci.* **2018**, *77*, 146. [[CrossRef](#)]
31. Li, X.-Y.; Liu, Y.; Lv, M.-R.; Zou, Y.; Fan, P.-P. Calibration transfer of soil total carbon and total nitrogen between two different types of soils based on visible-near-infrared reflectance spectroscopy. *J. Spectrosc.* **2018**, 1–10. [[CrossRef](#)]
32. Ji, W.; Li, S.; Chen, S.; Shi, Z.; Rossel, R.A.V.; Mouazen, A.M. Prediction of soil attributes using the Chinese soil spectral library and standardized spectra recorded at field conditions. *Soil Tillage Res.* **2016**, *155*, 492–500. [[CrossRef](#)]
33. Chen, Y.; Qi, K.; Liu, Y.; He, J.-H.; Jiang, Q. Transferability of hyperspectral model for estimating soil organic matter concerned with soil moisture. *Guang Pu Xue Yu Guang Pu Fen Xi = Guang Pu* **2015**, *35*, 1705–1708. (In Chinese) [[PubMed](#)]
34. Keesstra, S.; Mol, G.; De Leeuw, J.; Okx, J.; Molenaar, C.; De Cleen, M.; Visser, S.M. Soil-related sustainable development goals: Four concepts to make land degradation neutrality and restoration work. *Land* **2018**, *7*, 133. [[CrossRef](#)]
35. Jiang, Z.; Yang, Y.; Sha, J. Application of GWR model in hyperspectral prediction of soil heavy metals. *Acta Geogr. Sin.* **2017**, *72*, 533–544.
36. Xin, Z.; Sun, J.; Yan, T.; Quansheng, C.; Xiaohong, W.; Yingying, H. A deep learning based regression method on hyperspectral data for rapid prediction of cadmium residue in lettuce leaves. *Chemometr. Intell. Lab. Syst.* **2020**, *200*, 103996. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).