plants

Article

Indirect Estimation of Heavy Metal Contamination in Rice Soil Using Spectral Techniques

Liang Zhong [1,2], Shengjie Yang [1,2], Yicheng Rong [1,2], Jiawei Qian [1,2], Lei Zhou [3], Jianlong Li [1,2,*] and Zhengguo Sun [4,*]

[1] State Key Laboratory of Pharmaceutical Biotechnology, School of Life Sciences, Nanjing University, Nanjing 210023, China; zhongliang1007@163.com (L.Z.)
[2] Department of Ecology, School of Life Sciences, Nanjing University, Nanjing 210023, China
[3] Livestock Development and Promotion Center, Linyi 276037, China
[4] College of Agro-Grassland Science, Nanjing Agricultural University, Nanjing 210095, China
* Correspondence: jlli2008@nju.edu.cn (J.L.); sunzg@njau.edu.cn (Z.S.)

Abstract: The rapid growth of industrialization and urbanization in China has led to an increase in soil heavy metal pollution, which poses a serious threat to ecosystem safety and human health. The advancement of spectral technology offers a way to rapidly and non-destructively monitor soil heavy metal content. In order to explore the potential of rice leaf spectra to indirectly estimate soil heavy metal content. We collected farmland soil samples and measured rice leaf spectra in Xushe Town, Yixing City, Jiangsu Province, China. In the laboratory, the heavy metals Cd and As were determined. In order to establish an estimation model between the pre-processed spectra and the soil heavy metals Cd and As content, a genetic algorithm (GA) was used to optimise the partial least squares regression (PLSR). The model's accuracy was evaluated and the best estimation model was obtained. The results showed that spectral pre-processing techniques can extract hidden information from the spectra. The first-order derivative of absorbance was more effective in extracting spectral sensitive information from rice leaf spectra. The GA-PLSR model selects only about 10% of the bands and has better accuracy in spectral modeling than the PLSR model. The spectral reflectance of rice leaves has the capacity to estimate Cd content in the soil (relative percent difference [RPD] = 2.09) and a good capacity to estimate As content in the soil (RPD = 2.97). Therefore, the content of the heavy metals Cd and As in the soil can be estimated indirectly from the spectral data of rice leaves. This study provides a reference for future remote sensing monitoring of soil heavy metal pollution in farmland that is quantitative, dynamic, and non-destructive over a large area.

Keywords: rice; soil-crop system; heavy metal contamination; spectral technique; genetic algorithm; indirect estimation

check for updates

Citation: Zhong, L.; Yang, S.; Rong, Y.; Qian, J.; Zhou, L.; Li, J.; Sun, Z. Indirect Estimation of Heavy Metal Contamination in Rice Soil Using Spectral Techniques. *Plants* **2024**, *13*, 831. https://doi.org/10.3390/plants13060831

Academic Editor: Youzhen Xiang

Received: 1 February 2024
Revised: 29 February 2024
Accepted: 9 March 2024
Published: 14 March 2024

Copyright: © 2024 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 (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Soil serves both as the basis for the growth of crops and as a vital natural resource for the sustenance and production of human beings [1]. However, soil environmental pollution has increased significantly in China due to rapid industrialization and urbanization [2]. Among various pollutants, soil heavy metal contamination stands out due to its slow migration, high toxicity, and irreversible nature [3]. Over time, heavy metals accumulate in the food chain and pose severe health risks when ingested and accumulated by humans [4,5]. According to the 2014 Chinese Soil Pollution Status Report, the overall pollution excess rate of soil in China is 16.1% [6]. This alarming statistic has profound implications for China's food security as heavy metal pollution in soil leads to an annual loss of approximately 12 million tons of grain crops [7]. As a result, both the government and scholars have shown widespread concern regarding soil heavy metal pollution [8,9]. Quantitative monitoring of heavy metal content plays an important role in understanding the extent and sources of heavy metal pollution in a region [10]. It also offers a theoretical foundation for the remediation and management of such pollution [11,12]. While traditional chemical analysis methods are highly accurate in detecting soil heavy metal content, they are time-consuming, labor-intensive, and costly. Consequently, they fail to support the demands of real-time and large-scale monitoring for efficient heavy metal content assessment [13]. Fortunately, the advancement of remote sensing technology has paved the way for nondestructive and rapid soil heavy metal monitoring using spectral remote sensing [14,15]. This technique, characterized by its high-resolution and ability to capture details of the object's spectral information, holds great potential in enabling fast and efficient monitoring of soil heavy metal content [16].

Spectral remote sensing applications for monitoring soil heavy metal contamination include two main approaches: direct and indirect. Direct monitoring focuses on the mechanism where soil heavy metal are adsorbed by soil organic matter, iron-manganese oxides and clay minerals [17]. These components affect soil spectral morphology and reflectance, leading to specific soil spectral absorption features. At present, more research has been conducted on direct monitoring [18–22]. The direct monitoring method is not without its limitations, although it results in high model accuracy and stable models. Firstly, it is a time-consuming and cumbersome process that requires field soil sampling and laboratory processing to obtain soil spectral data. In addition, due to soil drying, grinding, and sieving, the spectral features of heavy metals extracted from laboratory soil spectra often differ from those obtained from remote sensing images. This difference complicates the direct application of these models to large-scale soil pollution monitoring using aerospace images [23]. Therefore, a more convenient and practical approach for widespread application is to use spectral data from plant leaves or canopies for indirect estimation of soil heavy metal contamination [24]. The approach is based on the principle that heavy metals move from the soil to the plants, accumulating there [25]. Under the stress of heavy metals, the protein and chlorophyll content of the plants are affected, which leads to a discernible difference in the reflectance spectra [26,27]. Related research has made significant progress. For example, Shi et al. [28] developed a multivariate spectral vegetation index based on rice canopy spectra for the estimation of arsenic (As) in farmland soil. Zhong et al. [29] estimated leaf Cu content using leaf hyperspectra and then inverted the Cu content of other parts of wheat and soil using bioconcentration factors. Wang et al. [30] observed that wheat canopy Cu increased with increasing soil Cu concentration, accompanied by distinct variations in spectral reflectance, providing a foundation for indirect estimation of soil Cu conten. However, the feasibility and accuracy of indirect estimation of soil heavy metals from rice leaves are not yet clear.

In this study, we will explore the potential of rice leaf spectra for estimating soil heavy metals. Firstly, we collected soil samples from farmland and also measured the spectral data of rice leaves in Xushe Town, Yixing City, Jiangsu Province, China. Then, we processed the leaf spectra with various spectral transformations and screened the spectral feature bands using a genetic algorithm (GA). Next, we used partial least squares regression (PLSR) to model soil Cadmium (Cd) and As content for for the spectra after different pre-processing. Finally, we evaluated the accuracy of the models and obtained the best estimation models.

2. Results

2.1. Statistics of Soil Samples

Figure 1 shows the content of the soil Cd and As of the sampling sites in the study area. The average value of soil pH at the sampling site was 5.86. The Cd content was between 0.13 to 0.97 mg kg^{-1} with a average value of 0.29 mg kg^{-1}, and the high values were located in the eastern and central parts of the study area. The As content was between 3.23 to 9.32 mg kg^{-1} with a average value of 5.64 mg kg^{-1}, and the high values were mainly located in the central part of the study area. The correlation coefficient between Cd and As content at the sampling sites was 0.33.

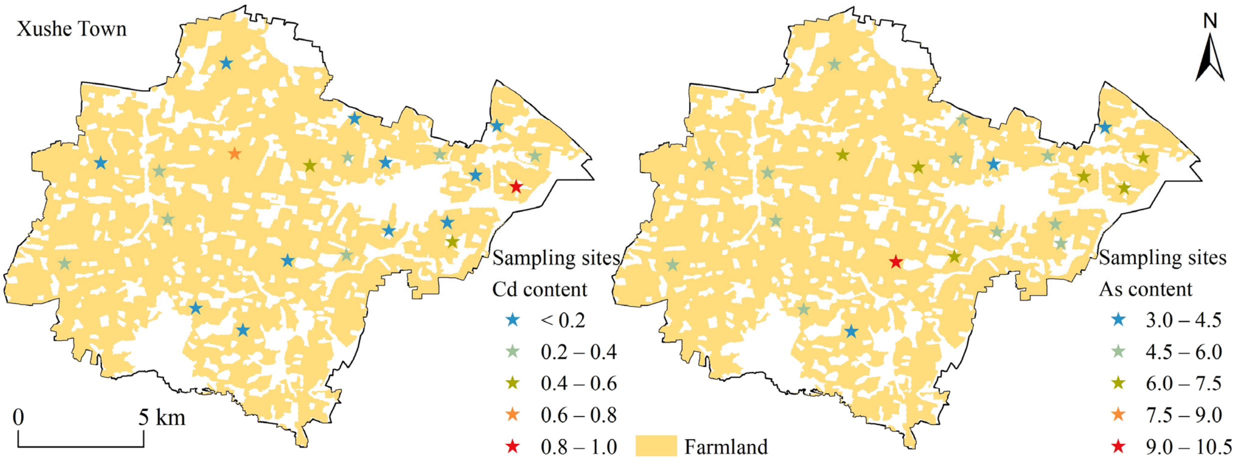

Figure 1. Distribution of Cd and As content at sampling sites in the study area.

2.2. Characterization of Spectral Curves of Rice Leaves

Figure 2 shows the spectra of rice leaves after different pre-processing. From the raw spectra (R) (Figure 2a), it can be seen that the spectral curve of rice leaves is significantly different in the 760–1120 nm band. It has a green light reflection peak at 550 nm, blue-violet light absorption valleys at 450 nm, and red light absorption valleys at 670 nm. This is due to the fact that chlorophyll absorbs weakly in the green band of light and strongly in the blue-violet and red bands of light. There is a high reflectance in the 760 nm to 1120 nm near-infrared band, which may be caused by multiple reflections within the leaf structure.

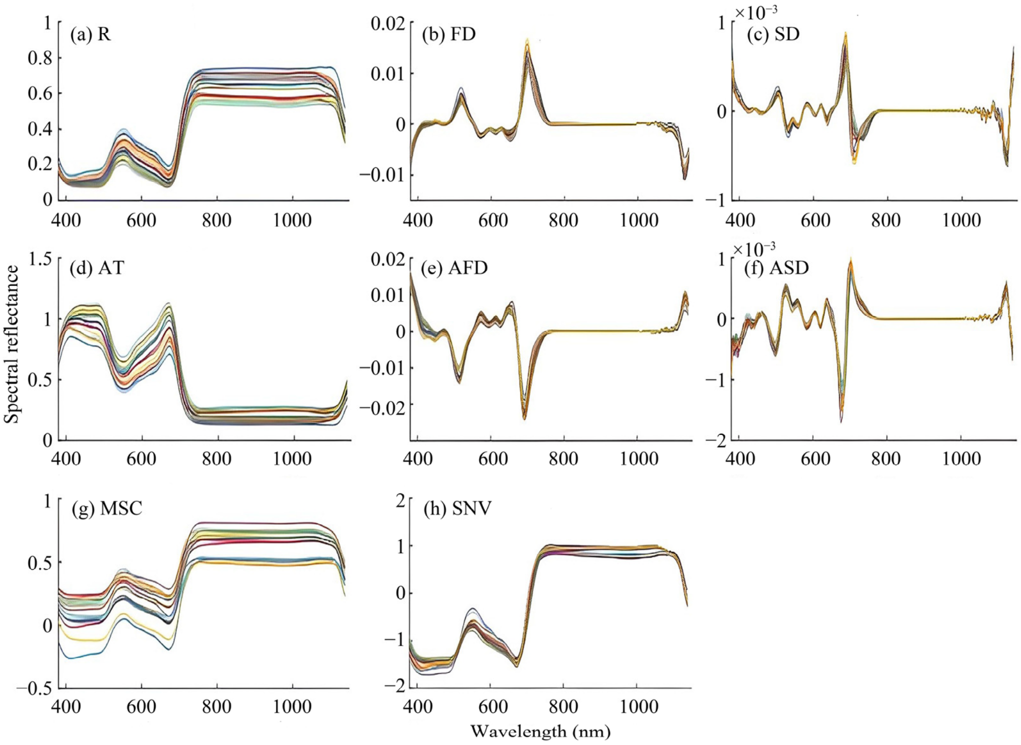

Figure 2. Characteristics of the spectral curves of rice leaves with different pre-processing.

Compared with R, the shape of the first-order derivative (FD) spectrum curve (Figure 2b) changed significantly, with a valley of absorption at 1129 nm and peaks of reflectance at 516 nm and 705 nm. The shape of the second-order derivative (SD) spectrum curve (Figure 2c) also changed significantly, with valleys of absorption at 712 nm and 1119 nm and peaks of reflectance at 503 nm and 686 nm. The absorbance transformation (AT) spectrum (Figure 2d) has high reflectance in the 400–510 nm band, low reflectance in the 730–1100 nm band, a valley of absorption at 553 nm, and a peaks of reflectance at 673 nm. The first-order derivative of absorbance (AFD) spectrum (Figure 2e) has valleys of absorption at 516 nm and 693 nm, and a reflection peak at 572 nm. The second-order derivative of absorbance (ASD) spectrum (Figure 2f) has valleys of absorption at 503 nm and 679 nm, and peaks of reflectance at 446 nm, 526 nm, 560 nm, 709 nm, and 1123 nm. The trend of the multiplicative scatter correction (MSC) spectrum curve (Figure 2g) is the same as that of the original spectrum, but the difference in the spectral curve of the 380–700 nm band is enlarged. The trend of the standard normal variate (SNV) spectrum curve (Figure 2h) is similar to that of the original spectrum, but the spectra are denser, indicating the ability of the SNV transform to reduce background noise.

2.3. Spectral Feature Bands Selected by GA

Rice leaf spectral feature bands were screened using GA as shown in Figure 3. Under different spectral pre-processing, GA selected 17–25 feature bands of soil Cd and 15–30 feature bands of soil As among 230 full bands. The soil Cd feature bands are mostly concentrated in 400–410 nm, 520–580 nm, 630–650 nm, 690–770 nm, 870–925 nm, and 970–1000 nm; the soil Cd feature bands are mostly concentrated in 400–440 nm, 815–860 nm, 890–920 nm, 970–990 nm, 1005–1030 nm, and 1065–1135 nm.

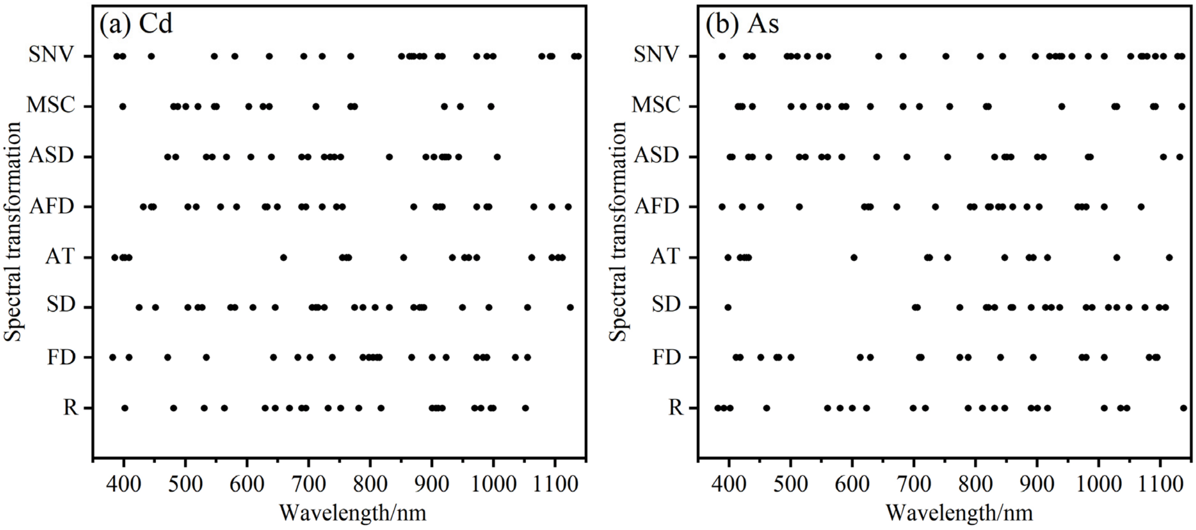

Figure 3. The feature bands of rice leaves spectral screened by GA.

2.4. Comparison of GA-PLSR and PLSR Modeling Results

Using the GA-PLSR and PLSR models, the modeling and analysis of rice leaf spectra under different pre-processing were carried out, and the cross-validation results are shown in Table 1. Compared with the PLSR model established directly using the full band, the PLSR model established by first screening the bands with GA has the same or reduced number of PCs, indicating that GA can select rice leaf spectral bands that are more meaningful to the PLSR model. Compared with the PLSR model, the R^2_{cv} value of estimating soil Cd content under different pre-processing spectra using the GA-PLSR model increased by 6.25~33.96%, and the RMSE_{cv} value decreased by 0.00~50.00%. The R^2_{cv} value of estimating soil As content increased by 14.29~53.19%, and the RMSE_{cv} value decreased by 3.33~69.64%. The results indicate that using GA for spectral wavelength selection before establishing a model for estimating heavy metal content in rice leaf spectra can improve model accuracy and stability.

2.5. Best Estimate Model

A cross-validation and an external validation were carried out on the GA-PLSR model to estimate the heavy metal content of the soil, and the results are presented in Table 2. The results of the soil Cd content estimation model show that, compared with R, the accuracy of each indicator was improved to different degrees in the cross-validation and external validation of the 7 transformed spectra. This indicates that the accuracy and stability of the Cd content estimation model have been improved after different spectral transformations. Among them, the RPD values for R, FD, SD, ASD, and MSC spectral preprocessing are all less than 1.50, suggesting poor accuracy in estimating soil Cd content. The RPD values of AT and SNV spectral preprocessing ranges between 1.50 and 2.00, indicating the possibility to discriminate between soil with high and low Cd content. The AFD spectral preprocessing has the highest model accuracy, with R^2_{cv}, RMSE$_{\mathrm{cv}}$, R^2_{ev}, RMSE$_{\mathrm{ev}}$, and RPD of 0.71, 0.07 mg kg^{-1}, 0.77, 0.06 mg kg^{-1}, and 2.09, respectively, indicating the ability to approximate soil Cd content estimation.

Table 1. Comparison of the accuracy of GA-PLSR and PLSR models for estimating soil Cd and As content in rice leaves.

Heavy Metal	Pre-Processing	Number of Bands	GA-PLSR			PLSR		
			PC	R^2_{cv}	RMSE$_{\mathrm{cv}}$/(mg kg^{-1})	PC	R^2_{cv}	RMSE$_{\mathrm{cv}}$/(mg kg^{-1})
Cd	R	22	2	0.34	0.16	2	0.32	0.18
	FD	21	3	0.46	0.15	3	0.39	0.15
	SD	25	1	0.42	0.15	2	0.35	0.17
	AT	17	2	0.53	0.15	2	0.45	0.15
	AFD	25	5	0.71	0.07	8	0.53	0.14
	ASD	22	2	0.47	0.13	4	0.38	0.17
	MSC	16	1	0.47	0.15	3	0.40	0.16
	SNV	26	2	0.52	0.15	2	0.45	0.15
As	R	21	2	0.50	1.18	3	0.41	1.27
	FD	20	1	0.52	1.15	2	0.42	1.22
	SD	21	2	0.55	1.16	3	0.44	1.25
	AT	15	2	0.56	1.16	2	0.49	1.20
	AFD	23	5	0.89	0.34	9	0.61	1.12
	ASD	23	2	0.72	0.82	6	0.47	1.44
	MSC	22	4	0.58	1.15	4	0.42	1.27
	SNV	30	2	0.70	0.93	6	0.51	1.17

Table 2. Accuracy of GA-PLSR models for estimating soil Cd and As content in rice leaves.

Heavy Metal	Pre-Processing	Cross-Validation		External Validation		
		R^2_{cv}	RMSE$_{\mathrm{cv}}$/(mg kg^{-1})	R^2_{ev}	RMSE$_{\mathrm{ev}}$/(mg kg^{-1})	RPD
Cd	R	0.34	0.16	0.41	0.11	1.30
	FD	0.46	0.15	0.52	0.10	1.44
	SD	0.42	0.15	0.47	0.10	1.37
	AT	0.53	0.15	0.59	0.09	1.56
	AFD	0.71	0.07	0.77	0.06	2.09
	ASD	0.47	0.13	0.53	0.09	1.46
	MSC	0.47	0.15	0.49	0.10	1.40
	SNV	0.52	0.15	0.62	0.08	1.62
As	R	0.50	1.18	0.57	0.65	1.52
	FD	0.52	1.15	0.68	0.56	1.77
	SD	0.55	1.16	0.66	0.58	1.71
	AT	0.56	1.16	0.64	0.60	1.66
	AFD	0.89	0.34	0.89	0.30	2.97
	ASD	0.72	0.82	0.71	0.53	1.86
	MSC	0.58	1.15	0.64	0.60	1.66
	SNV	0.70	0.93	0.76	0.48	2.06

The results of the model for estimating soil As content show that the cross-validation and external validation of the seven transformed spectra improve the accuracy of each index to varying degrees compared to R. This indicates that the different spectral transformations have improved the accuracy and stability of the As estimation models. Among them, the R, FD, SD, AT, ASD, and MSC spectral preprocessing have an RPD values between 1.50 and 2.00, indicating the possibility of distinguishing between soil with high and low As content. The RPD values of SNV spectral preprocessing is 2.06, indicating the ability to approximate soil As content estimation. The AFD spectral preprocessing has the highest level of model precision, with $R^2{}_{\mathrm{cv}}$, RMSE$_{\mathrm{cv}}$, $R^2{}_{\mathrm{ev}}$, RMSE$_{\mathrm{ev}}$, and RPD of 0.89, 0.34 mg kg^{-1}, 0.89, 0.30 mg kg^{-1}, and 2.97, respectively, indicating good ability to estimate soil As content.

3. Discussion

3.1. Effect of Spectral Pre-Processing and Feature Selection for Modeling Performance

Modelling results from different spectral preprocessing techniques show that most preprocessed spectra show varying degrees of accuracy improvement over the original spectra. This improvement can be attributed to the fact that external disturbances can introduce noise when collecting spectral data, making it difficult to accurately represent the spectral characteristics of objects [31]. Spectral pre-processing approaches efficiently reduce spectral noise and improve the information about the spectral features [32,33]. For indirect soil heavy metal content estimation using rice leaf spectra, AFD is the optimal spectral transformation method for both Cd and As estimation. This is because of the FD spectral transform, which can effectively extract and enhance the hidden information in the spectrum [34]. Yao et al. [35] found that derivative transformations were able to highlight the spectral features more compared with MSC, SNV and Continuum removal. Meanwhile, the absorbance transformation can further improve the inversion accuracy of As content, which is consistent with our ability to obtain a good estimation of soil As content using AFD spectra.

In this study, the number of feature bands selected by GA in the optimal inversion models for soil Cd and As contents were 25 and 23, respectively, only about 10% of the bands were used. Moreover, higher accuracy was achieved by utilizing GA for selecting spectral characteristic bands compared to PLSR modelling. The reason for this approach is that spectral data have properties of redundancy and collinearity, and direct modelling with PLSR is susceptible to being disturbed by significant amounts of redundancy information [36]. GA improves model quality and stability by successfully filtering feature bands from the full spectrum [37]. The results of Sun and Zhang [38], Sun et al. [39], and Zhong et al. [17], who also showed that GA-PLSR outperforms PLSR in estimating heavy metal concentration using soil spectral data, are consistent with this methodology. It indicates that the method applied to the spectral modelling of heavy metal content in agricultural soils. In addition, Zhang et al. [40] in the estimation of soil heavy metal Cd, the accuracy of R^2 was 0.88 by PLSR modelling using soil spectral features associated with organic matter extracted using GA. Wei et al. [41] in the estimation of soil heavy metal As, the accuracy of R^2 was 0.82 and 0.70 for Honghu City and Daye City in Hubei Province, China, after selecting the feature bands by using the stable competitive adaptive reweighting sampling algorithm coupled the successive projections algorithm followed by PLSR modelling. Bian et al. [42] in the estimation of a variety of soil heavy metals, PLSR and extreme learning machine models obtained the best accuracy for Cd (R^2 of 0.89) and As (R^2 of 0.86), respectively. These studies generally obtained good model accuracies, but they mainly used laboratory soil spectra for direct monitoring of soil heavy metal content. Our study used rice leaf spectra to indirectly estimate soil heavy metal Cd and As contents, which is of significant value in the future development of field soil heavy metal hyperspectral instrumentation and in the exploration of aerospace hyperspectral remote sensing for monitoring soil heavy metal contamination on a large scale.

3.2. Application and Perspectives of Spectral Techniques in Heavy Metal Inversion in Soil-Rice System

Spectral data are characterized by high spectral resolution, which can obtain fine spectral information about the object. Using spectral analysis technology, it can highlight the subtle differences between spectra, which is conducive to extracting spectral features and finally constructing models to invert the information of the object [23]. In the soil-crop system, heavy metals migrate from the soil and accumulate in the crop. When the concentration of heavy metals in the soil increases, the content of heavy metals in different parts of rice usually increases as well, resulting in a stress effect [43]. As the stress of heavy metals in rice increases, some of the cellular structures are damaged, resulting in a decrease in chlorophyll content, which is reflected in differences in leaf spectra [44,45]. Therefore, we successfully inverted the content of heavy metals (Cd and As) in soil by spectral transformation, characteristic analysis, and modeling of rice leaf spectra. This study can help to develop soil heavy metal monitoring instruments in the future directly through rice leaf spectra, and thereby improve the detection efficiency of soil heavy metals. Meanwhile, in the future, attempts can be made to expand this indirect monitoring approach to aerospace spectral remote sensing, further combining space-air-ground spectral remote sensing data. This is conducive to making full use of the characteristics and laws of ground-based spectral indirect monitoring, while giving full play to the characteristics of aerospace spectral remote sensing in terms of dynamics and wide range. This will be helpful to form a multi-scale monitoring and validation system for soil-crop heavy metal pollution [17].

3.3. Limitations and Future Work

In this study, we used rice leaf spectra to indirectly estimate the content of soil heavy metals Cd and As, but there are still some limitations that we would like to explore and improve in future work, such as: (1) We studied and tested our method based on spectral data at only one sampling time, and the stability and applicability of the method need to be further validated. In the future, the indirect estimation method of soil heavy metal content in this study can be further explored for application at different time and spatial scales. (2) We only selected two typical pollutant elements in the study area for our study, and the potential of other heavy metal elements in spectral monitoring needs to be further explored in subsequent studies. (3) The mechanisms of uptake, transport and accumulation of heavy metal elements in the soil-crop system can be further explored in the future, which will provide more basis for the indirect monitoring of soil heavy metal content using crop spectra.

4. Materials and Methods

4.1. Study Area

The study area is in the town of Xushe (31°18′–31°27′ N, 119°31′–119°44′ E), west of Yixing City, Jiangsu Province (Figure 4), and the total area is about 1.8×10^4 hm^2. The region features a subtropical monsoon climate with well-defined seasons, ample rainfall, and an average annual temperature of 16.0 °C, accompanied by precipitation of 1434.0 mm. The topography of the region is characterized by higher elevation in the western areas and lower elevation in the eastern areas, consisting mainly of plains and hills. With a cultivated area of 1.2×10^4 hm^2, Xushe Town is the largest agricultural town in Yixing City, primarily used for cultivating rice and wheat. Paddy, fluvo-aquic and yellow-brown soils are the three main soil types.

4.2. Data Collection and Processing

4.2.1. Soil Sampling and Data Determination

We collected 22 surface (0–20 cm) soil samples in September 2019 in the study area using a five-point mixing method (Figure 3). At the time of sampling, we determined and recorded the location of each sampling point using a hand-held GPS device. The soil samples were then returned to the laboratory in sealed bags. In the laboratory, all soil samples were dried, ground and sieved (0.15 mm pore size) and divided equally into two parts. A part of the soil samples was used for the measurement of the pH by the potentiometric method (NY/T 1377-2007) [46]. The other part was weighed at 0.2 g of soil sample, put into the bottom of the PTFE digestion tank, added 5 mL of nitric acid, 2 mL of hydrogen peroxide, and 2 mL of hydrofluoric acid, and microwaved digested for 15 min. After the digestion solution was clarified, it was fixed to 50 mL and filtered. Finally, the Cd and As contents were determined by inductively coupled plasma mass spectrometry (ICP-MS) [47].

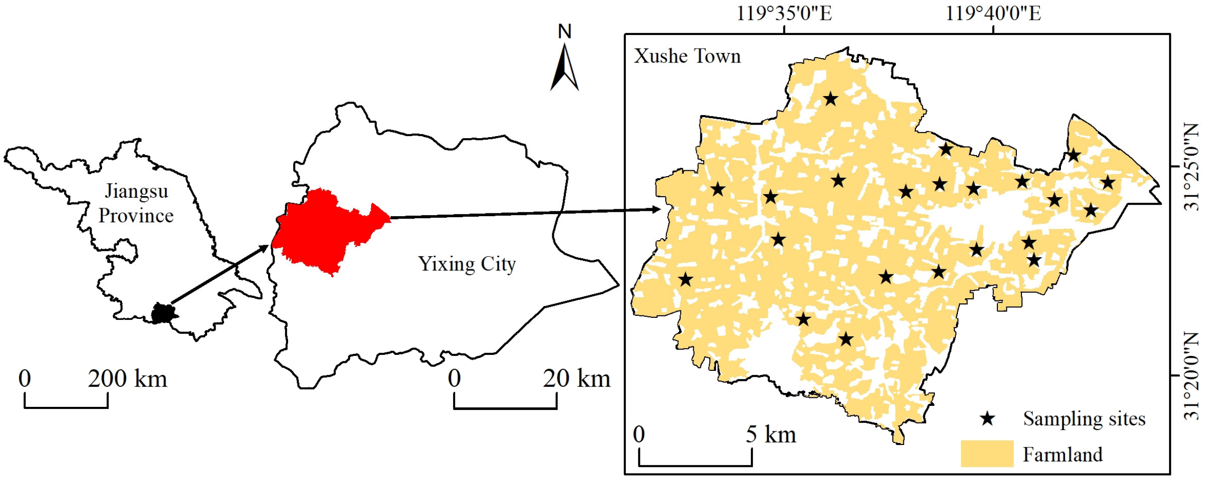

Figure 4. Location of the study area and distribution of sampling sites.

During the soil sampling process, we collected spectral data of rice leaves using a portable field spectrometer (UniSpec, PP systems, Haverhill, MA, USA) between 11:00 a.m. and 2:00 p.m. Beijing time. The spectrometer had a spectral range of 301–1145 nm and a spectral resolution of 3.3 nm. At each sampling site, five rice plants were randomly selected, and three fully expanded leaves from each rice were selected for spectral measurement under sunny and light wind conditions. White calibration was carried out before each spectral measurement and five measurements were repeated. For each sampling site, 75 spectral data points were collected from rice leaves and averaged to obtain the spectral data.

4.2.2. Spectral Pre-Processing

Firstly, the spectral data of the rice leaf are stripped of the noisy edge bands below 380 nm. At each sampling site, the bands between 380 and 1145 nm are selected for spectral data processing and analysis. Next, the leaf spectral data of rice is processed using Savitzky-Golay smoothing. The smoothed spectral data is referred to as raw spectrum R. Finally, based on the R spectra, the spectral pre-processing was carried out by applying mathematical transform methods such as AT, SNV, MSC, FD, AFD, SD, and ASD [48].

4.3. Research Methods

4.3.1. Genetic Algorithm

The GA is an evolutionary algorithm used for solving optimization problems. It simulates the mechanisms of genetics and natural selection, assessing individuals with superior fitness, selecting, crossing, and mutating them using genetic operators to generate individuals in the new generation of the population. In order to find the optimal solution, the iterative process is repeated until the convergence criteria have been met [49]. Simultaneously, GA can avoid the overfitting problem of general iterative methods, which may fall into the local minimum.

Prior to modeling, feature bands are selected using GA in order to reduce redundancies and optimize model performance [50]. In feature selection using GA, each band is treated as a gene, and a specific number of bands are designated as chromosomes. Next, a subset of the samples is taken to form the initial population. Then, crossing and mutation are used to simulate the genetic and evolutionary processes of random populations in nature, while the fitness function is used to assess the model's performance in predicting outcomes. After conducting tests, the GA parameters for population size, crossover probability, mutation probability, and genetic generation were set to 40, 0.5, 0.01, and 100, respectively. We repeated the process 10 times to minimize the influence of randomness. We used the root mean squared error of cross-validation ($\mathrm{RMSE_{cv}}$) for PLSR as the fitness criterion. As the individual's fitness increases, the $\mathrm{RMSE_{cv}}$ decreases.

4.3.2. Partial Least Squares Regression

The PLSR is one of the most commonly used methods for processing spectral data to estimate soil heavy metal content [51]. In this method, the independent variable and the dependent variable are projected onto a new coordinate system. The principal component, which has the strongest explanatory power, is extracted and used to construct a new linear model. This helps reduce collinearity and noise effects and makes the model more robust [52]. During the process of PLSR modeling, cross-validation is utilized to identify the number of most efficient principal components.

The spectra after different pre-processing were used to select the characteristic bands using GA and estimate the heavy metal content using PLSR. The 22 data samples are divided into two parts, with one sample selected from every 4 samples for validation. In all, 17 samples were used for modeling, and 5 samples were used for validating the accuracy of the model. The GA feature band selection and the PLSR modelling were done in R-Studio 3.5.3 (https://posit.co/products/open-source/rstudio/ (accessed on 6 November 2019)).

4.3.3. Model Assessment

This study used the coefficient of determination (R^2_{cv}) and the $\mathrm{RMSE_{cv}}$ for model cross-validation. The R^2_{ev}, $\mathrm{RMSE_{ev}}$, and relative percent difference (RPD) were chosen for model external validation. The closer R^2_{cv} and R^2_{ev} are to 1, the lower $\mathrm{RMSE_{cv}}$ and $\mathrm{RMSE_{ev}}$ are, and a higher RPD indicates a better model fit and accuracy. The five-layer interpretation method proposed by Williams et al. [53] was adopted for the evaluation criteria of RPD. If the RPD exceeds 3.00, the model has excellent ability to estimate. If the RPD ranges from 2.50 to 3.00, the model is considered to have good predictive performance. If the RPD ranges from 2.00 to 2.50, the model can be used for an approximate quantitative estimate. If the RPD ranges from 1.50 and 2.00, the model has the ability to discriminate between high and low values. If the RPD is less than 1.50, the model has a poor ability to estimate.

5. Conclusions

By combining spectral preprocessing, feature selection and modelling methods, this study fully explored the potential of rice leaf spectra for indirect estimation of soil heavy metals Cd and As, and the following conclusions were drawn:

(1) Spectral preprocessing technology enhances the modeling accuracy by revealing hidden information in the spectrum, leading to varying degrees of improvement compared to the original spectrum. When modeling rice leaf spectra, the most effective estimation models for soil Cd and As content are obtained through AFD spectral preprocessing. These results highlight the advantages of mathematical transformations, such as derivative transformation and absorbance, in extracting spectral sensitive information.

(2) The GA-PLSR model demonstrates superior performance compared to the PLSR model in modeling of rice leaf spectra. Specifically, compared to the PLSR model, GA-PLSR used only approximately 10% of the bands and enhanced the R^2_{cv} values for estimating soil Cd and As content by 0.00% to 50.00% and 3.33% to 69.64%, respectively, for the different preprocessed spectra. These findings illustrate that

incorporating a GA for spectral band selection before establishing a model for estimating soil heavy metal content can significantly enhance the accuracy and efficiency of the model.

(3) In the modeling of soil Cd content using rice leaf spectra, the best estimation model is the combination of AFD and GA-PLSR, with $R^2{}_{\mathrm{ev}}$, $\mathrm{RMSE}_{\mathrm{ev}}$, and RPD values of 0.77, 0.06 mg kg^{-1}, and 2.09, respectively, which has the ability to approximate estimation. The best estimation model for soil As content is also the combination of AFD and GA-PLSR, with $R^2{}_{\mathrm{ev}}$, $\mathrm{RMSE}_{\mathrm{ev}}$, and RPD values of 0.89, 0.30 mg kg^{-1}, and 2.97, respectively, which has good estimation ability.

Author Contributions: Conceptualization, L.Z. (Liang Zhong), J.L. and Z.S.; methodology, J.Q. and Y.R.; software, S.Y. and J.Q.; validation, Y.R., S.Y. and L.Z. (Lei Zhou); formal analysis, S.Y. and L.Z. (Lei Zhou); investigation, J.Q. and L.Z. (Lei Zhou); resources, J.L. and Z.S.; data curation, L.Z. (Liang Zhong) and J.L.; writing—original draft preparation, L.Z. (Liang Zhong); writing—review and editing, L.Z. (Liang Zhong), J.L. and Z.S.; visualization, L.Z. (Liang Zhong) and Y.R.; supervision, J.L. and Z.S.; project administration, J.L.; funding acquisition, J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the High-level international cooperation and innovation exchange forum (grant number: 202202), the National Key R&D Plan Project of China (grant number: 2018YFD0800201), and the project of Asia-Pacific network for global change research (grant number: ARCP2015-03CMY-Li).

Data Availability Statement: Data are contained within the article.

Acknowledgments: We are grateful to the anonymous reviewers for their comments and recommendations. We are grateful for the financial support from the fund to guarantee the smooth running of the study. We thank all the members of the research team for their help in this study.

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

References

1. Sanchez, P.; Ahamed, S.; Carre, F.; Hartemink, A.; Hempel, J.; Huising, J.; Lagacherie, P.; McBratney, A.; McKenzie, N.; Mendonca-Santos, M.; et al. Digital Soil Map of the World. *Science* **2009**, *325*, 680–681. [CrossRef]
2. Jiang, Y.; Huang, M.; Chen, X.; Wang, Z.; Xiao, L.; Xu, K.; Zhang, S.; Wang, M.; Xu, Z.; Shi, Z. Identification and risk prediction of potentially contaminated sites in the Yangtze River Delta. *Sci. Total Environ.* **2022**, *815*, 151982. [CrossRef]
3. Xia, F.; Zhao, Z.; Niu, X.; Wang, Z. Integrated pollution analysis, pollution area identification and source apportionment of heavy metal contamination in agricultural soil. *J. Hazard. Mater.* **2024**, *465*, 133215. [CrossRef]
4. Jiang, Y.; Chen, S.; Hu, B.; Zhou, Y.; Liang, Z.; Jia, X.; Huang, M.; Wei, J.; Shi, Z. A comprehensive framework for assessing the impact of potential agricultural pollution on grain security and human health in economically developed areas. *Environ. Pollut.* **2020**, *263*, 114653. [CrossRef]
5. Yuan, X.; Xue, N.; Han, Z. A meta-analysis of heavy metals pollution in farmland and urban soils in China over the past 20 years. *J. Environ. Sci.* **2021**, *101*, 217–226. [CrossRef]
6. Chen, W.; Yang, Y.; Xie, T.; Wang, M.; Peng, C.; Wang, R. Challenges and countermeasures for heavy metal pollution control in farmlands of china. *Acta Pedol. Sin.* **2018**, *55*, 261–272. [CrossRef]
7. Jiang, S.; Wang, F.; Li, Q.; Sun, H.; Wang, H.; Yao, Z. Environment and food safety: A novel integrative review. *Environ. Sci. Pollut. Res.* **2021**, *28*, 54511–54530. [CrossRef]
8. Yang, Q.; Li, Z.; Lu, X.; Duan, Q.; Huang, L.; Bi, J. A review of soil heavy metal pollution from industrial and agricultural regions in China: Pollution and risk assessment. *Sci. Total Environ.* **2018**, *642*, 690–700. [CrossRef]
9. Zhang, J.; Wang, X.; Li, J.; Luo, J.; Wang, X.; Ai, S.; Cheng, H.; Liu, Z. Bioavailability (BA)-based risk assessment of soil heavy metals in provinces of China through the predictive BA-models. *J. Hazard. Mater.* **2024**, *465*, 133327. [CrossRef]
10. Fei, X.; Lou, Z.; Xiao, R.; Ren, Z.; Lv, X. Source analysis and source-oriented risk assessment of heavy metal pollution in agricultural soils of different cultivated land qualities. *J. Clean. Prod.* **2022**, *341*, 130942. [CrossRef]
11. Ren, S.; Song, C.; Ye, S.; Cheng, C.; Gao, P. The spatiotemporal variation in heavy metals in China's farmland soil over the past 20 years: A meta-analysis. *Sci. Total Environ.* **2022**, *806*, 150322. [CrossRef]
12. Sorrentino, M.; Giordano, S.; Spagnuolo, V. Phytomonitoring and Phytoremediation of Environmental Pollutants. *Plants* **2023**, *13*, 366. [CrossRef]
13. Zhong, L.; Guo, X.; Xu, Z.; Ding, M. Soil properties: Their prediction and feature extraction from the LUCAS spectral library using deep convolutional neural networks. *Geoderma* **2021**, *402*, 115366. [CrossRef]

14. Shi, T.; Chen, Y.; Liu, Y.; Wu, G. Visible and near-infrared reflectance spectroscopy—An alternative for monitoring soil contamination by heavy metals. *J. Hazard. Mater.* **2014**, *265*, 166–176. [CrossRef] [PubMed]
15. 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]
16. Shi, T.; Guo, L.; Chen, Y.; Wang, W.; Shi, Z.; Li, Q.; Wu, G. Proximal and remote sensing techniques for mapping of soil contamination with heavy metals. *Appl. Spectrosc. Rev.* **2018**, *53*, 783–805. [CrossRef]
17. Zhong, L.; Chu, X.; Qian, J.; Li, J.L.; Sun, Z. Multi-Scale Stereoscopic Hyperspectral Remote Sensing Estimation of Heavy Metal Contamination in Wheat Soil over a Large Area of Farmland. *Agronomy* **2023**, *13*, 2396. [CrossRef]
18. 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* **2018**, *11*, 419. [CrossRef]
19. Pyo, J.; Hong, S.; Kwon, Y.; Kim, M.; Cho, K. Estimation of heavy metals using deep neural network with visible and infrared spectroscopy of soil. *Sci. Total Environ.* **2020**, *741*, 140162. [CrossRef]
20. Guo, F.; Xu, Z.; Ma, H.; Liu, X.; Tang, S.; Yang, Z.; Zhang, L.; Liu, F.; Peng, M.; Li, K. Estimating chromium concentration in arable soil based on the optimal principal components by hyperspectral data. *Ecol. Indic.* **2021**, *133*, 108400. [CrossRef]
21. Zhou, W.; Yang, H.; Xie, L.; Li, H.; Lu, H.; Zhao, Y.; Yue, T. Hyperspectral inversion of soil heavy metals in Three-River Source Region based on random forest model. *Catena* **2021**, *202*, 105222. [CrossRef]
22. Shi, T.; Fu, Z.; Miao, X.; Lin, F.; Ma, J.; Gu, S.; Li, L.; Wu, C.; Luo, Y. Would it be better for partition prediction of heavy metal concentration in soils based on the fusion of XRF and Vis-NIR data? *Sci. Total Environ.* **2024**, *908*, 168381. [CrossRef]
23. 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]
24. Zhou, W.; Zhang, J.; Zou, M.; Liu, X.; Du, X.; Wang, Q.; Liu, Y.; Li, J. Feasibility of using rice leaves hyperspectral data to estimate CaCl2-extractable concentrations of heavy metals in agricultural soil. *Sci. Rep.* **2019**, *9*, 16084. [CrossRef]
25. Zhang, L.; Liu, Z.; Song, Y.; Sui, J.; Hua, X. Advances in the Involvement of Metals and Metalloids in Plant Defense Response to External Stress. *Plants* **2023**, *13*, 313. [CrossRef]
26. Sanaeifar, A.; Yang, C.; De la Guardia, M.; Zhang, W.; Li, X.; He, Y. Proximal hyperspectral sensing of abiotic stresses in plants. *Sci. Total Environ.* **2023**, *861*, 160652. [CrossRef]
27. Falcioni, R.; Gonçalves, J.; De Oliveira, K.; De Oliveira, C.; Reis, A.; Crusiol, L.; Furlanetto, R.; Antunes, W.; Cezar, E.; De Oliveira, R.; et al. Chemometric Analysis for the Prediction of Biochemical Compounds in Leaves Using UV-VIS-NIR-SWIR Hyperspectroscopy. *Plants* **2023**, *12*, 3424. [CrossRef]
28. Shi, T.; Liu, H.; Chen, Y.; Wang, J.; Wu, G. Estimation of arsenic in agricultural soils using hyperspectral vegetation indices of rice. *J. Hazard. Mater.* **2016**, *308*, 243–252. [CrossRef]
29. Zhong, L.; Yang, S.; Chu, X.; Sun, Z.; Li, J. Inversion of heavy metal copper content in soil-wheat systems using hyperspectral techniques and enrichment characteristics. *Sci. Total Environ.* **2024**, *907*, 168104. [CrossRef]
30. Wang, G.; Wang, Q.; Su, Z.; Zhang, J. Predicting copper contamination in wheat canopy during the full growth period using hyperspectral data. *Environ. Sci. Pollut. Res.* **2020**, *27*, 39029–39040. [CrossRef]
31. Sun, T.; Li, Z.; Wang, Z.; Liu, Y.; Zhu, Z.; Zhao, Y.; Xie, W.; Cui, S.; Chen, G.; Yang, W.; et al. Monitoring of Nitrogen Concentration in Soybean Leaves at Multiple Spatial Vertical Scales Based on Spectral Parameters. *Plants* **2024**, *13*, 140. [CrossRef] [PubMed]
32. Tan, K.; Wang, H.; Zhang, Q.; Jia, X. An improved estimation model for soil heavy metal(loid) concentration retrieval in mining areas using reflectance spectroscopy. *J. Soils Sediments* **2018**, *18*, 2008–2022. [CrossRef]
33. Zhong, L.; Guo, X.; Ding, M.; Ye, Y.; Jiang, Y.; Zhu, Q.; Li, J. SHAP values accurately explain the difference in modeling accuracy of convolution neural network between soil full-spectrum and feature-spectrum. *Comput. Electron. Agric.* **2024**, *217*, 108627. [CrossRef]
34. Zhou, W.; Zhang, J.; Zou, M.; Liu, X.; Du, X.; Wang, Q.; Liu, Y.; Liu, Y.; Li, J. Prediction of cadmium concentration in brown rice before harvest by hyperspectral remote sensing. *Environ. Sci. Pollut. Res.* **2019**, *26*, 1848–1856. [CrossRef] [PubMed]
35. Yao, L.; Xu, M.; Liu, Y.; Niu, R.; Wu, X.; Song, Y. Estimating of heavy metal concentration in agricultural soils from hyperspectral satellite sensor imagery: Considering the sources and migration pathways of pollutants. *Ecol. Indic.* **2024**, *158*, 111416. [CrossRef]
36. Arif, M.; Qi, Y.; Dong, Z.; Wei, H. Rapid retrieval of cadmium and lead content from urban greenbelt zones using hyperspectral characteristic bands. *J. Clean. Prod.* **2022**, *374*, 133922. [CrossRef]
37. Hasan, U.; Jia, K.; Wang, L.; Wang, C.; Shen, Z.; Yu, W.; Sun, Y.; Jiang, H.; Zhang, Z.; Guo, J.; et al. Retrieval of Leaf Chlorophyll Contents (LCCs) in Litchi Based on Fractional Order Derivatives and VCPA-GA-ML Algorithms. *Plants* **2023**, *12*, 501. [CrossRef]
38. Sun, W.; Zhang, X. Estimating soil zinc concentrations using reflectance spectroscopy. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *58*, 126–133. [CrossRef]
39. Sun, W.; Liu, S.; Zhang, X.; Zhu, H. Performance of hyperspectral data in predicting and mapping zinc concentration in soil. *Sci. Total Environ.* **2022**, *824*, 153766. [CrossRef]
40. Zhang, X.; Sun, W.; Cen, Y.; Zhang, L.; Wang, N. Predicting cadmium concentration in soils using laboratory and field reflectance spectroscopy. *Sci. Total Environ.* **2019**, *650*, 321–334. [CrossRef]
41. Wei, L.; Pu, H.; Wang, Z.; Yuan, Z.; Yan, X.; Cao, L. Estimation of Soil Arsenic Content with Hyperspectral Remote Sensing. *Sensors* **2020**, *20*, 4056. [CrossRef]

42. Bian, Z.; Sun, L.; Tian, K.; Liu, B.; Huang, B.; Wu, L. Estimation of multi-media metal(loid)s around abandoned mineral processing plants using hyperspectral technology and extreme learning machine. *Environ. Sci. Pollut. Res.* **2023**, *30*, 19495–19512. [CrossRef] [PubMed]
43. Elazab, D.; Lambardi, M.; Capuana, M. In Vitro Culture Studies for the Mitigation of Heavy Metal Stress in Plants. *Plants* **2023**, *12*, 3387. [CrossRef] [PubMed]
44. Su, Z.; Wang, G.; Xu, L.; Zhang, J.; Liu, X. Effects of Cu stress on physiological, biochemical, and spectral properties of wheat at different growth stages. *Int. J. Agric. Biol. Eng.* **2019**, *12*, 147–153. [CrossRef]
45. Székely, Á.; Szalóki, T.; Jancsó, M.; Pauk, J.; Lantos, C. Temporal Changes of Leaf Spectral Properties and Rapid Chlorophyll—A Fluorescence under Natural Cold Stress in Rice Seedlings. *Plants* **2023**, *12*, 2415. [CrossRef]
46. *NY/T 1377-2007*; De-Termination of pH in Soil. Standards Press of China: Beijing, China, 2007.
47. Le, S.; Duan, Y. Determination of Heavy Metal Elements in Soil by ICP-MS. *Chin. J. Inorg. Anal. Chem.* **2015**, *5*, 16–19. [CrossRef]
48. Shi, Z.; Wang, Q.; Peng, J.; Ji, W.; Liu, H.; Li, X.; Viscarra Rossel, R.A. Development of a national VNIR soil-spectral library for soil classification and prediction of organic matter concentrations. *Sci. China Earth Sci.* **2014**, *57*, 1671–1680. [CrossRef]
49. Mitchell, M. *An Introduction to Genetic Algorithms*; MIT Press: Cambridge, MA, USA, 1998.
50. Du, X.; Li, X.; Liu, Y.; Zhou, W.; Li, J. Genetic algorithm optimized non-destructive prediction on property of mechanically injured peaches during postharvest storage by portable visible/shortwave near-infrared spectroscopy. *Sci. Hortic.* **2019**, *249*, 240–249. [CrossRef]
51. Nawar, S.; Mohamed, E.; Essam-Eldeen Sayed, S.; Mohamed, W.; Rebouh, N.; Hammam, A. Estimation of key potentially toxic elements in arid agricultural soils using vis-NIR spectroscopy with variable selection and PLSR algorithms. *Front. Environ. Sci.* **2023**, *11*, 1222871. [CrossRef]
52. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130. [CrossRef]
53. Williams, P.; Manley, M.; Antoniszyn, J. *Near Infrared Technology: Getting the Best Out of Light*; African Sun Media: Stellenbosch, South Africa, 2019.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.