

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/262146124>

Monitoring Arsenic Contamination in Agricultural Soils with Reflectance Spectroscopy of Rice Plants

ARTICLE in ENVIRONMENTAL SCIENCE & TECHNOLOGY · MAY 2014

Impact Factor: 5.33 · DOI: 10.1021/es405361n · Source: PubMed

CITATIONS

5

READS

110

6 AUTHORS, INCLUDING:



Tie Shi

Shenzhen University

13 PUBLICATIONS 50 CITATIONS

SEE PROFILE



Huizeng Liu

Hong Kong Baptist University

4 PUBLICATIONS 5 CITATIONS

SEE PROFILE



Teng Fei

Wuhan University

37 PUBLICATIONS 175 CITATIONS

SEE PROFILE



Guofeng Wu

Wuhan University

28 PUBLICATIONS 318 CITATIONS

SEE PROFILE

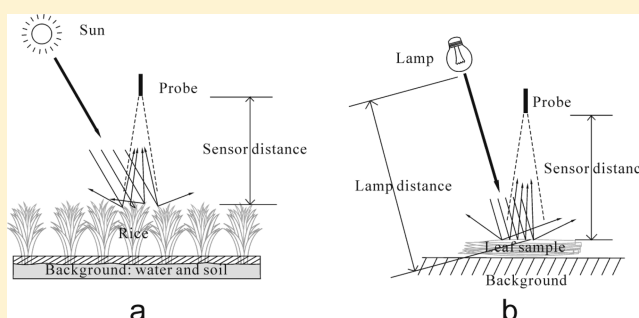
Monitoring Arsenic Contamination in Agricultural Soils with Reflectance Spectroscopy of Rice Plants

Tiezhu Shi,[†] Huizeng Liu,[†] Junjie Wang,[†] Yiyun Chen,[†] Teng Fei,[†] and Guofeng Wu^{*,‡,†}

[†]School of Resource and Environmental Science and Key Laboratory of Geographic Information System of the Ministry of Education, Wuhan University, 430079 Wuhan, China

[‡]Key Laboratory for Geo-Environment Monitoring of Coastal Zone of the National Administration of Surveying, Mapping and GeoInformation and Shenzhen Key Laboratory of Spatial Smart Sensing and Services and College of Life Sciences, Shenzhen University, 518060 Shenzhen, China

ABSTRACT: The objective of this study was to explore the feasibility and to investigate the mechanism for rapidly monitoring arsenic (As) contamination in agricultural soils with the reflectance spectra of rice plants. Several data pretreatment methods were applied to improve the prediction accuracy. The prediction of soil As contents was achieved by partial least-squares regression (PLSR) using laboratory and field spectra of rice plants, as well as linear regression employing normalized difference spectral index (NDSI) calculated from field spectra. For laboratory spectra, the optimal PLSR model for predicting soil As contents was achieved using Savitzky-Golay smoothing (SG), first derivative and mean center (MC) (root-mean-square error of prediction (RMSEP) = 14.7 mg kg⁻¹; $r = 0.64$; residual predictive deviation (RPD) = 1.31). For field spectra, the optimal PLSR model was also achieved using SG, first derivative and MC (RMSEP = 13.7 mg kg⁻¹; $r = 0.71$; RPD = 1.43). In addition, the NDSI with 812 and 782 nm obtained a prediction accuracy with $r = 0.68$, RMSEP = 13.7 mg kg⁻¹, and RPD = 1.36. These results indicated that it was feasible to monitor the As contamination in agricultural soils using the reflectance spectra of rice plants. The prediction mechanism might be the relationship between the As contents in soils and the chlorophyll-*a*/*-b* contents and cell structure in leaves or canopies of rice plants.



INTRODUCTION

Arsenic (As) is a poisonous pollutant to the environment,¹ and it is rated as “a known carcinogen” by the World Health Organization (WHO).² During the past 30 years, some anthropogenic activities, such as mining, smelting, animal manures, chemical fertilizers, As-bearing pesticides and irrigation with As-contaminated water, have caused the continuous increases of As content in the agricultural soils in China.³ The excessive As accumulation in agricultural soils not only hinders the growth and decreases the production of crops but also ends up in crops, reducing the quality of agricultural products.¹ Moreover, if As enters the human body through food chains, it might pose a serious health threat.^{4,5}

A reliable and environmentally friendly method is needed to rapidly assess the As contamination in agricultural soils for diagnosing suspected contaminated areas and controlling rehabilitation processes.⁶ The conventional method of assessing the soil contamination by heavy metals is based on regular field samplings and sequent chemical analyses (e.g., wet chemistry) followed by geostatistical interpolation,^{7,8} and it is costly and time-consuming. Moreover, such investigation can only provide limited information at specific locations and moments and could not describe the spatial and temporal dynamics of the As contamination in soils over large areas.⁹

Visible and near-infrared diffuse reflectance spectroscopy (VNIRS, 350–2500 nm) can serve as a potential alternative to the conventional method for monitoring the As contamination in agricultural soils. Compared with conventional methods, the practical advantages of VNIRS technique include that (1) the technique is fast, nondestructive, and cost-effective; (2) less or no hazardous chemical reagents are required; (3) several soil properties could be predicted simultaneously from one spectrum; (4) the technique may be used both in the laboratory and *in situ*; and (5) this technique has better spatial and temporal continuities.¹⁰

VNIRS techniques contain three types of spectral data: laboratory, field, and remotely sensed spectra.¹¹ Laboratory spectra are the spectral measurements under controlled conditions for exploring the feasibility, extracting the spectral features, and investigating the mechanism of predicting the As contents in soils.^{12–14} Field spectra are generally acquired *in situ* and could be of great benefit as they can be acquired fast and almost continuously.¹⁵ The application of field spectra,

Received: December 1, 2013

Revised: April 19, 2014

Accepted: May 7, 2014

Published: May 7, 2014

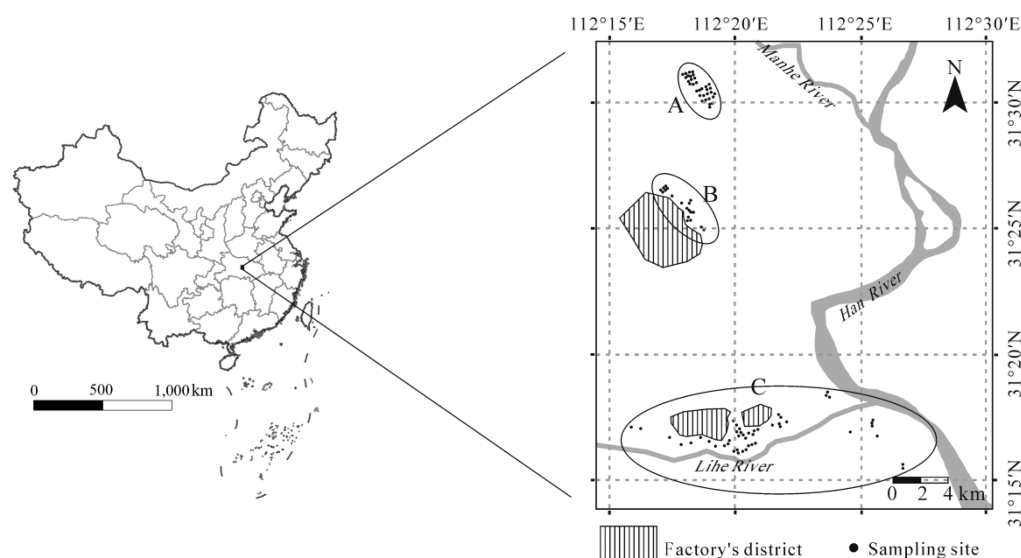


Figure 1. Maps showing the study area of Zhongxiang in China and the sampling sites.

however, requires some constraints to be taken into account, including pixel mixing, atmospheric effect, and illumination condition.^{15,16} Therefore, field spectra are cheap data for exploring the solutions to these constraints and for establishing the foundations for the applications of remotely sensed spectra. Remotely sensed spectra (e.g., hyperspectral images) are obtained using air- or space-borne sensors, which can quickly acquire spatially continuous spectral information on target objects. Therefore, remotely sensed spectra play a vital role in monitoring the As contamination in agricultural soils over large areas and long time periods.¹⁷ However, agricultural soils are generally covered by crops or their residues, which means that it is not always possible to acquire the remotely sensed spectra of agricultural soils directly.

For the regions covered by crops or their residues, the spatial distribution of heavy metal contamination in soils might be inferred from the diagnostic features of plant reflectance spectroscopy,¹⁸ because the plants growing in heavy-metal-polluted soils show some stresses, such as an increase of chlorophyll hydrolysis, a decrease of water contents, and destruction of cellular structure,¹⁹ which might change the remotely sensed reflectance spectra of plants. For examples, Kooistra et al.²⁰ found that the VNIRS of ryegrass (*Lolium perenne*) could be used to predict the Zn contamination levels in the soils of Rhine and Meuse river floodplains in The Netherlands, and Kooistra et al.¹⁸ observed that the Ni, Cd, Cu, Zn, and Pb contents in river floodplain soils were well correlated with the spectral reflectance of grasses (*Poa annua* and *Lolium perenne*) over the spectral region of 400–1350 nm.

There are 120 million hm^2 of agricultural lands in China, about 8.3% of which are contaminated by heavy metals.²¹ However, the spatial information on the As contamination in agricultural soils are very limited. Rice (*Oryza sativa* L.) is a staple food crop, and it occupies 24% of all agricultural lands in China.²² Therefore, it might be possible to assess the As contamination in some agricultural soils using the reflectance spectroscopy of rice plants. However, to the best of our knowledge, this hypothesis has not been tested so far. Therefore, the objective of this study was to explore the feasibility and investigate the mechanism for rapidly predicting

the As contents in agricultural soils with the reflectance spectra of rice plants at leaves and canopies levels.

MATERIALS AND METHODS

Study Area. Zhongxiang is situated in the Hubei Province of China (Figure 1), and its average annual temperature is 15.9 °C with a mean annual precipitation of 952.6 mm. Zhongxiang is one of the main production regions of rice in central China, with a rice planting area of about 4,5000 hm^2 . However, the heavy metal contamination, especially As contamination, is becoming increasingly serious in this region because of the rapid developments of phosphate mining and phosphorus chemical industries.

Data Collection. A total of 100 sampling sites were randomly selected from three regions (A, B, and C in Figure 1) in the Zhongxiang region. According to the soil classification system of China, the soil type of these sites belongs to paddy soil, and according to the FAO system, the dominant soil type of the sampling sites belongs to anthrosols for rice planting. The irrigation water of rice in region A was underground water and rainwater from reservoirs. Region B was near chemical factories, by which the irrigation water was polluted. Region C was near the Lihe River, and its irrigation water was river water, which was polluted by industrial waste discharged by local factories. The rice plant variety of our sample sites was Liangyou 287.

Field Spectrum Measurement. The field spectra of rice plants were measured on Aug. 8–12, 2012 during the booting stage. The field spectral measurements were taken between 10:00 and 14:00 under cloudless or near-cloudless weather conditions, using an ASD FieldSpec3 portable spectroradiometer (Analytical Spectral Devices Inc., USA). The spectroradiometer has a spectral range of 350–2500 nm, and its sampling interval is 1.4 in the 350–1000 nm range and 2 in the 1000–2500 nm range; and its spectral resolution is 3 nm in the 350–1000 nm range and 10 nm in the 1000–2500 nm range. The measured values are interpolated, and finally the spectroradiometer provides a spectrum of 2151 bands with a uniform spectral interval of 1 nm. A standardized white Spectralon panel with 100% reflectance (Labsphere, www.labsphere.com) was used to optimize signal and calibrate

accuracy and detector responses. The radiance of the white Spectralon panel was measured before and after each sample. The RS³ Spectral Acquisition Software (<http://www.asdi.com/products/spectroscopy-software/rs3>) was used to collect spectral data and control the instrument. The radiance of rice plants was obtained by taking 10 consecutive scans from a height of 1.0 m above the canopies of the rice plants using an optical probe with a 25° field of view. By dividing the mean radiance of 10 consecutive scans by the radiance over the Spectralon panel, the spectral reflectance of the rice plants at the canopies level was calculated.

Leaves and Soils Sampling. The samples of rice leaves and soils were collected at all of the sites in which the canopy spectra of rice were measured. At each sampling site, the geographical coordinate was recorded using a global positioning system receiver (Garmin Ltd., USA). The leaves on the top layer of the canopy were cut out, and about 1.0–1.5 kg of surface soils from a depth of 0–10 cm were collected after wiping off plants and their residues as well as stones. The collected leaves and soils samples were kept in sealed plastic bags for their spectral measurements and soil As content analyses in the laboratory.

Laboratory Spectrum Measurement. The laboratory spectra of 100 leaves samples were measured using the same spectroradiometer as that used in the field. The measurements were conducted in a dark room with a 50 W halogen lamp as light source, which was positioned 0.3 m away from leaves sample, with a 15° zenith angle; the sample was spread evenly on a black paperboard tray; and the optical probe was installed about 0.15 m above the leaves sample.²³ The correction of the white Spectralon panel was taken prior to the first scan and after every six samples. The spectral reflectance of each leaves sample was acquired using the same approach as that used in the field.

Soil Arsenic Content Measurement. The collected soil samples were air-dried for 3 days, ground in a mechanical agate grinder, and passed through a 100-mesh sieve (0.15 mm). The finely ground soil samples were digested by HF–HClO₄–HNO₃, and the digested samples were then used for the chemical analysis of As contents using a hydride generation atomic fluorescence spectrometry (HG-AFS) method.²⁴ Certified soil reference materials, GBW 07401, GBW 07402, and GBW 07407 (National Research Center for Certified Reference Materials of China), were used to verify the accuracy and precision of the HG-AFS method.

The index of geo-accumulation (I_{geo})^{25,26} was applied to assess the As contamination in the soils:

$$I_{\text{geo}} = \log_2 \frac{C_{\text{As}}}{1.5B_{\text{As}}} \quad (1)$$

where C_{As} is the measured As contents in the soils and B_{As} is the geochemical background value of As (13 mg kg^{−1}).²⁶ The constant of 1.5 was used to eliminate fluctuations caused by regional differences and anthropogenic influences.^{26,27} $I_{\text{geo}} \leq 0$ indicates practically uncontaminated, and $I_{\text{geo}} > 0$ means contaminated.²⁶

Soil Arsenic Content Prediction Using PLSR. *Data Pretreatment.* The presence of spectral outliers may affect model prediction accuracy. Therefore, in our study, a robust principal component analysis method²⁸ was carried out to detect spectral outliers. The remaining data set was randomly

separated into two data sets: two-thirds as the calibration data set and the remaining one-third as the validation data set.²⁹

The laboratory spectra of the leaves of the rice plants were resampled to 10 nm intervals (e.g., 350, 360, and 370 nm, etc.) using a Gaussian model,¹⁴ which takes the band center and the full width at half-maximum into account. This spectral resampling reduced the spectral variables from 2151 to 216, which increased calculation speed and reduced the overfitting of calibration models.^{30,31} For the field spectra of the canopies of the rice plants, the noise regions near 1400, 1900, and 2500 nm caused by atmospheric moisture were removed, and the spectral variables of field spectra were not reduced in order to avoid losing more information.

In order to enhance spectral features and further to establish a robust prediction model, several data pretreatment methods were performed for the laboratory and field spectra, including the Savitzky–Golay smoothing (SG) method, first and second derivatives of reflectance, multiplicative scatter correction (MSC), standard normal variate transformation (SNV), and mean centering (MC).³²

Model Calibration. The spectral data of the leaves and canopies of the rice plants were first processed by the aforementioned data pretreatment methods and their combinations. Then partial least-squares regression (PLSR) method was used to establish relations between the preprocessed spectra with soil As contents. Leave-one-out cross-validation³³ was used to determine the factors in PLSR models, and Akaike information criterion (AIC) was applied to preserve model parsimony.³⁴

The important wavelengths used in PLSR models were determined by the variable importance in projection (VIP; eq 2)³⁵ and PLSR coefficients (b -coefficients).³⁶ For a spectral variable, the VIP was calculated by

$$\text{VIP}_k(a) = K \sum_a w_{ak}^2 \left(\frac{\text{SSY}_a}{\text{SSY}_t} \right) \quad (2)$$

where $\text{VIP}_k(a)$ is the importance of the k th predictor variable based on a model with a factors, w_{ak} is the corresponding loading weight of the k th variable in the a th PLSR factor, SSY_a is the explained sum of squares of y by a PLSR model with a factor, SSY_t is the total sum of squares of y , and k is the total number of predictor variables.³³ The threshold for VIP was set to 1,³⁵ and the values of b -coefficients were based on their standard deviations.³⁷ The spectral wavelengths were considered as important if the VIP score and b -coefficient values met these two requirements.³³

Model Validation. The best calibrated PLSR models at the leaves and canopies levels were used to predict the soil As contents of the validation data set. The correlation coefficient (r) between the predicted and reference soil As contents, and the root-mean-square error of prediction (RMSEP) and residual predictive deviation (RPD) were used to evaluate the predictive ability of PLSR models. A relatively small RPD value indicates a less robust PLSR model, and the increase of the RPD value means an improvement of predictive power.³⁸

Soil Arsenic Content Prediction Using Normalized Difference Spectral Index. One of the greatest challenges in applying field spectra and remotely sensed spectra of plants to quantitatively analyze the As contents in soils is the mixed pixel problem. This problem appears when the background factors (e.g., water, soil, and shadow) affect the presence of green vegetation in the spectra, and it often makes the discrimination

of vegetation difficult. The normalization of reflectance values (e.g., normalized difference vegetation index (NDVI)) using a few wavelengths is effective in reducing the influence of background differences.³⁹ However, Inoue et al.⁴⁰ found that normalized difference spectral index (NDSI, eq 3) was more effective than NDVI since it used the merits of data richness and continuity of hyperspectra. In this study, multiple NDSI values were calculated from the field spectra which were processed by the data pretreatment methods adopted by the best calibrated PLSR model of the canopies level. And the NDSI value which had the highest correlation with soil As contents was then employed to establish the linear regression models for predicting the soil As contents in the soils.

$$\text{NDSI}(x, y) = (y - x)/(x + y) \quad (3)$$

where x and y are the reflectances at the wavelengths of i and j nm over the spectral region of 350–1200 nm.

RESULTS

Soil Arsenic Contents and Reflectance Spectra of Rice Plants. The percent mean standard error of the HG-AFS method for As determination was 2.7% ($n = 100$). The statistical descriptions for the As contents in soils at regions A, B and C are shown in Table 1. Considering I_{geo} values, 52% of

Table 1. Statistical Descriptions for the Arsenic Contents (mg kg^{-1}) of 100 Soil Samples at Three Sampling Regions (A, B, and C) in Zhongxiang (per, Percent Value of Contaminated Samples)

	no.	minimum	maximum	mean	per, %
total	100	10.3	133.4	26.6	52
A	33	10.3	21.3	15.2	3
B	20	18.1	133.4	54.3	95
C	47	13.1	34.8	21.3	66

100 soil samples were contaminated by As, while the remaining 48% are uncontaminated. Among the three regions, region B is the most polluted area, followed by regions C and A. Most of the contaminated regions were close to factories or irrigated using wastewater, which indicates that the industrial wastewater, exhaust gas, or waste residue produced by local chemical factories are the major causes of As contamination in agricultural soils in the Zhongxiang region.

Four samples in laboratory spectra and five samples in field spectra were recognized as outliers and were excluded from further investigations. The statistical descriptions of soil As contents for the whole data sets, calibration data sets, and validation data sets are shown in Table 2. The characteristic

statistics of the calibration and validation data sets are similar to those of the whole data sets, which means that the calibration and validation data sets can effectively represent the whole data sets.

The mean laboratory and field reflectance spectra for four groups of rice plant samples growing in soils with four arsenic gradients are shown in Figure 2. The laboratory reflectance of leaves of rice plants are lower at the region of 350–750 nm than other spectral regions, and they show two absorption peaks near 500 and 680 nm and one reflection peak near 550 nm (Figure 2a), in which the absorption peaks result from the strong absorption by the pigments of the leaves of the rice plants. The spectral curve moves up sharply near 730 nm, which is known as the red-edge position.⁴¹ The curve has a large fluctuation and high reflectance over 800–1300 nm, which are associated with the mesophyll cell structure of the leaves of the rice plants.⁴² Two prominent absorption peaks around 1450 and 1900 nm are related to water content.

The field reflectance spectra of the canopies of the rice plants show the same trend as the laboratory reflectance spectra (Figure 2b); however, they contain three noise regions near 1400, 1900, and 2500 nm caused by the absorption of moisture contents within the atmosphere, and they have lower reflectance compared with the laboratory spectra. The differences between laboratory and field spectra could be caused by different structures of the targets, background, illuminant, and atmospheric environment. By comparing the reflectance of rice plants growing in soils with different As contents, it can be found that the reflectances over 750–1350 nm generally decrease with the increasing soil As contents at the leaves and canopies levels (Figure 2).

Soil Arsenic Content Prediction Using PLSR. The prediction accuracy strongly depends on the data pretreatment methods applied to the reflectance spectra (Table 3). In terms of RMSECV and AIC, the factor number of PLSR was 5, and the first five principal components (PCs) explained about 96.78% of the variation of the spectral data. The best calibration result was achieved by using SG, first derivatives, and MC pretreatment methods together (RMSECV = 16.7 mg kg^{-1} and $r = 0.65$). According to the VIP scores and b -coefficients (Figure 3), the wavelengths centered near 540, 690, 730, 780, and 1170 nm were identified as the important wavelengths for PLSR modeling using laboratory spectra. The best calibrated PLSR model was employed to predict the As contents of the validation data set, and the validation result showed that there existed significantly moderate correlation between the reference and predicted As contents at a significance level of 0.05 ($r = 0.64$; $p < 0.05$), with a RMSEP of 14.7 mg kg^{-1} and RPD of 1.31 (Figure 5a; Table 3).

Table 2. Statistical Description of the Soil Arsenic Contents (mg kg^{-1}) for the Whole Data Sets, Calibration Data Sets, and Validation Data Sets^a

	laboratory spectra			field spectra		
	whole data set	calibration data set	validation data set	whole data set	calibration data set	validation data set
n	96	64	32	95	63	32
minimum	10.3	10.3	11.3	11.1	11.1	11.3
maximum	133.4	133.4	112.0	133.4	133.4	112.0
mean	25.3	25.6	24.8	27.5	28.6	25.5
median	19.7	19.7	19.2	20.3	20.4	20.2
SD	21.1	22.1	20.7	23.7	22.3	21.6

^aNote: n is the sample number; SD is standard deviation.

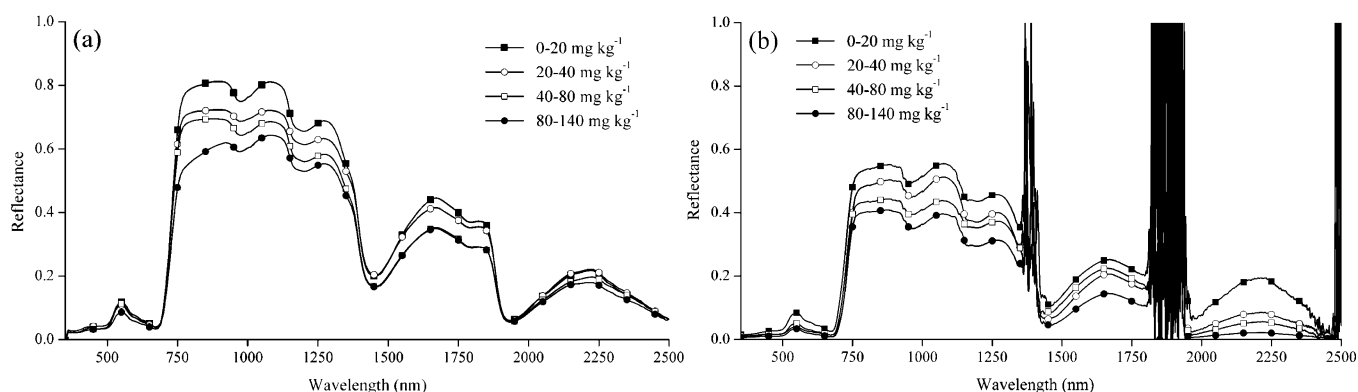


Figure 2. Mean laboratory (a) and field (b) reflectance spectra for four groups of rice plant samples growing in soils with four arsenic gradients.

Table 3. Calibration and Validation Results of PLSR Models for Predicting the Arsenic Contents in Soils with Multivariate Data Pretreatment Methods^a

data pretreatment methods		calibration data set			validation data set		
		factors	RMSECV(mg kg ⁻¹)	<i>r</i>	RMSEP(mg kg ⁻¹)	<i>r</i>	RPD
			(<i>n</i> = 64)		(<i>n</i> = 32)		
laboratory spectra (350–2500 nm, 10 nm interval)	none	5	18.1	0.57	14.1	0.68	1.36
	SG	5	18.2	0.56	14.1	0.68	1.36
	MC	5	17.7	0.59	13.7	0.71	1.40
	SNV + SG + MC	4	17.6	0.59	13.7	0.75	1.40
	MSC + SG + MC	4	17.6	0.60	13.7	0.77	1.40
	first + SG + MC	5	16.7	0.65	14.7	0.64	1.31
	second + SG + MC	5	17.5	0.61	16.5	0.49	1.16
			(<i>n</i> = 63)		(<i>n</i> = 32)		
field spectra (350–1200nm, 1 nm interval)	none	6	20.6	0.61	17.5	0.60	1.12
	SG	6	20.6	0.61	17.6	0.59	1.11
	MC	6	19.8	0.64	17.1	0.62	1.15
	SNV + SG + MC	4	19.3	0.66	17.3	0.51	1.13
	MSC + SG + MC	4	19.7	0.64	17.9	0.49	1.10
	first + SG + MC	5	18.5	0.69	13.7	0.71	1.43
	second + SG + MC	5	23.1	0.47	16.7	0.50	1.17

^aNote: *n* is the sample number, *r* is the correlation coefficient between the predicted and reference arsenic contents, RMSECV is the root-mean-square error of leave-one-out cross-validation, RMSEP is the root-mean-square error of prediction, RPD is the residual predictive deviation, SG is Savitzky–Golay smoothing, first and second are first and second derivatives of reflectance, MSC is the multiplicative scatter correction, SNV is the standard normal variate transformation, and MC is mean centering. The results of selected models are emphasized in bold.

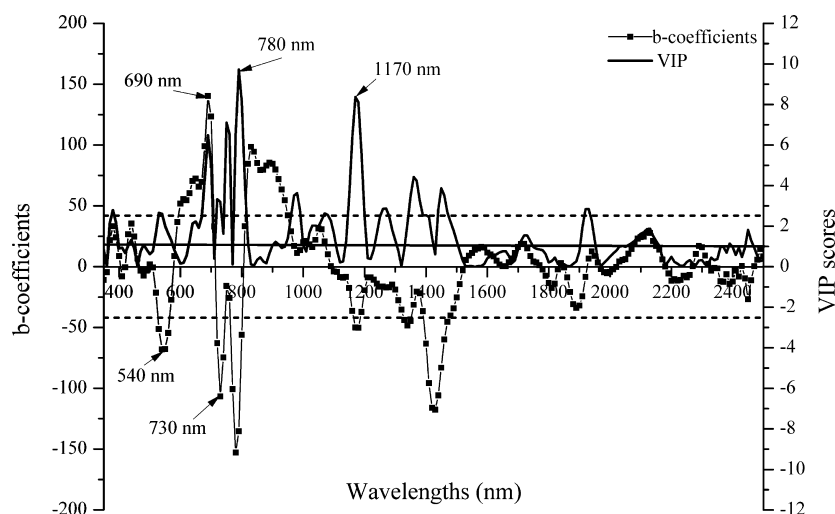


Figure 3. Variable importance projection (VIP) scores (black line) and *b*-coefficients (black squares) associated with the cross-validation of partial least-squares regression (PLSR) model for arsenic content prediction using laboratory spectroscopy. The threshold for VIP was set to 1 (black horizontal line), and the threshold for the *b*-coefficients was based on their standard deviation ($\sigma = 37$; horizontal dashed lines).

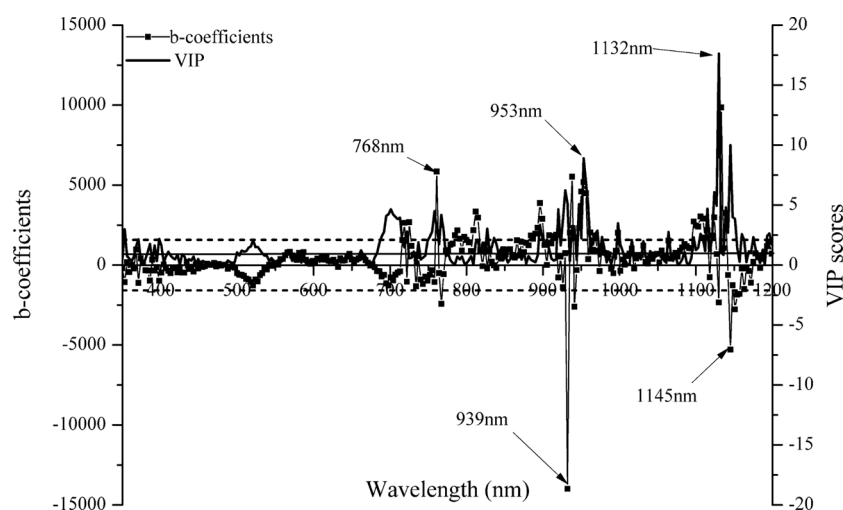


Figure 4. VIP scores (black line) and b -coefficients (black squares) associated with the cross-validation of the PLSR model for arsenic content prediction using field spectroscopy. The threshold for VIP was set to 1 (black horizontal line), and the threshold for the b -coefficients was based on their standard deviation ($\sigma = 1578$; horizontal dashed lines).

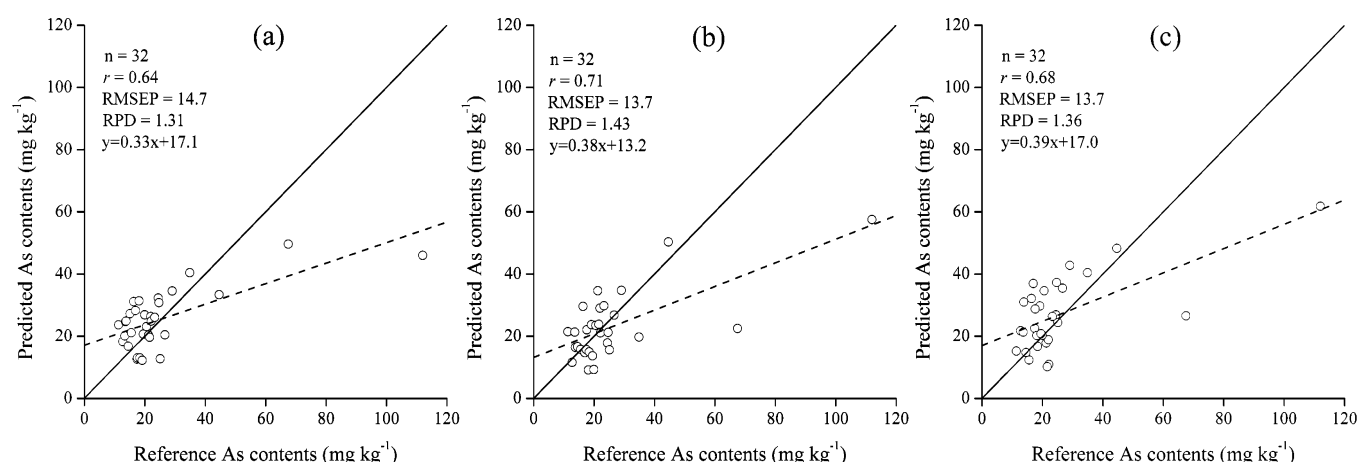


Figure 5. Scatter plots of the reference vs predicted soil arsenic contents (mg kg^{-1}) for laboratory (a) and field spectra (b) using PLSR, and field spectra using NDSI with 812 and 782 nm (c) (the solid line is the 1:1 line, and the dashed line is the regression line between the predicted and reference values).

For the field spectra, the informative region over 1201–2500 nm was removed due to there being no important wavelengths identified by the laboratory spectra of the leaves of the rice plants. The remaining spectral variables over 350–1200 nm were employed for PLSR model development. The factor number of PLSR was 5, and the first five PCs explained approximately 85.79% of the variation of the field spectral data. The best calibration result was also achieved by using SG, first derivatives, and MC pretreatment methods together (RMSECV = 18.5 and $r = 0.69$) for predicting the As contents in soils from the field spectra of rice plants. According to the VIP scores and b -coefficients (Figure 4), the wavelengths centered near 768, 939, 953, 1132, and 1145 nm were identified as the important wavelengths for PLSR modeling with the field spectra of rice plants. The best calibrated PLSR model was applied to predict the As contents of the validation data set, and the validation result also indicated that there was significantly moderate correlation between the reference and predicted As contents at a significance level of 0.05 ($r = 0.71$; $p < 0.05$), with a RMSEP of 13.7 mg kg^{-1} and RPD of 1.43 (Figure 5b; Table 3).

Soil Arsenic Content Prediction Using NDSI. The NDSI values with 812 and 782 nm calculated from field spectra could

explain 46.6% of the variation of the soil As contents in soils of the calibration data set using a linear regression model (eq 5). The regression model was employed to predict the soil As contents of the validation data set, and the validation result showed that there existed significantly moderate correlation between the reference and predicted As contents at a significance level of 0.05 ($r = 0.68$; $p < 0.05$), with a RMSEP of 13.6 mg kg^{-1} and RPD of 1.36 (Figure 5c).

$$\text{As}_{\text{contents}} = 1930 \times \text{NDSI}(\text{R}812, \text{R}782) + 16.5 \quad (5)$$

DISCUSSION

There are two types of As in the environment: inorganic As (AsH_3 , As(III), and As(V), etc.) and organic As (monomethylarsonic acid, dimethylarsinic acid).⁴³ In general, the toxicities of As species are based on their solubility and mobility, and they follow an order of $\text{AsH}_3 > \text{As(III)} > \text{As(V)} > \text{monomethylarsonic acid} > \text{dimethylarsinic acid}$.⁴⁴ In many parts of China, the agricultural soils are often dry land and paddyfield farming in rotation. In dry land (aerobic soils), the As species are present as As(V) bound to iron (Fe), aluminum

(Al), and manganese (Mn) (hydr)oxides, and thus the As contents in soil solution are low.⁴⁵ However, in a paddyfield (anoxic soils), the Fe (hydr)oxides are reduced to Fe^{2+} because of the depletion of oxygen, and As(V) is released into soil solution.⁴⁶ Furthermore, As(V) is reduced to As(III), which exhibits the dominant As species in flooded fields and is more mobile in soil solution than As(V).^{46,47} Therefore, compared with the crops growing in dry land (maize, soybean, and sesame, etc.), the rice plants growing in paddyfield soils may more easily accumulate and metabolize As contents in agricultural soils. Rice plants could better serve as an indicative plant sensitive to As contamination in agricultural soils among the field-crop plants.

This study indicated that the VNIRS of rice plants could be used to predict the As contents in agricultural soils, and the prediction accuracies were close to the application level ($\text{RPD} > 1.5$ ⁴⁸). Phosphate and As have similar physicochemical behavior in soils. Guo et al.⁴⁹ demonstrated that the phosphate with high contents in soils could hinder the As assimilation and accumulation in rice plants by relieving the symptom of As poisoning in rice plants. In Zhongxiang, because of the impacts of phosphate mining and phosphorus chemical industries, the As-contaminated soils also contained high phosphate contents, and the phosphate stress might be the key factor affecting the prediction accuracies in our study. We considered that good results might be achieved when the VNIRS of rice plants was employed to predict the As contents in agricultural soils without phosphate stress. This hypothesis needs further validation.

In this study, the wavelengths centered near 540, 690, 730, 780, and 1170 nm of laboratory spectra at the leaves level were statistically identified as the important wavelengths for predicting the As contents in agricultural soils. All identified important wavelengths were within the region of 350–1200 nm, which indicated that the spectral region of 1201–2500 nm might be uninformative for predicting the As contents in soils from the reflectance spectroscopy of rice plants. In order to explore the mechanism describing how the As contents in soils indirectly affect these wavelengths of rice plants, the relationships between the important wavelengths and the biochemical compositions of rice plants need be first figured out. The wavelengths centered near 540, 690, 730, and 780 nm were identified as the optimal wavelengths for predicting the chlorophyll-*a*/*-b* contents of plant leaves.⁵⁰ The reflectance near 1170 nm was caused by the reflection and dispersion of the mesophyll cell of leaves of the rice plants.⁴² Moreover, in a glass house study, Rahman et al.⁵¹ observed that the As contents in soils ($\text{Na}_2\text{HAsO}_4 \cdot 7\text{H}_2\text{O}$) had negative correlation with chlorophyll-*a* ($R^2 = 0.83$) and chlorophyll-*b* ($R^2 = 0.75$) contents in the leaves of the rice plants, and they also found that the increased As contents in soils caused an alternation of the chloroplast shape and cell structure in the leaves of the rice plants. Therefore, we deduced that the As contents in agricultural soils might be indirectly related to these important wavelengths through the chlorophyll-*a*/*-b* contents and the cell structure of the leaves of the rice plants.

In this study, we observed that the wavelengths around 768, 939, 953, 1132, and 1145 nm of the spectra at the canopies level were identified as the important wavelengths for predicting the As contents in agricultural soils. Liu et al.⁵² demonstrated that the spectral region of 750–950 nm of the canopies was significantly correlated with leaf area index ($r > 0.45$; $p < 0.05$) and chlorophyll density ($r > 0.6$; $p < 0.05$). The

wavelengths near 1132 and 1145 nm are related to the cellular structure of the rice plants.⁴² Therefore, the As contents in agricultural soils may relate these bands through the cell structure, leaf area index, and chlorophyll density of rice plants at the canopy level, because the As stress might inhibit the enzyme activity of chloroplasts and the growth of leaves,⁵³ which will alter the cell structure and decrease the leaf area index and chlorophyll density of rice plants.

We noticed that the important wavelengths extracted from laboratory and field spectra were different. Such difference might be caused by the differences of the measurement environments, the target structure, and the background. In the laboratory, the measurement environment was stable, and the background was almost undetectable since the target was tiled leaves, while the environment was unstable in the field, and the spectra inevitably contained information on the background, including soil, water, and sometimes duckweed. However, despite containing background information, the field spectra did not result in much lower prediction accuracies compared with the laboratory spectra. In field sampling, we observed the symptoms of reduced growth and tillering of rice plants growing in soils with higher As contents, which might also increase background information in field spectra. Therefore, the background information on soil and water might correlate with the As contents in soils, and they were informative and useful for improving the predictions of As contents in soils.

Our study also found that the NDSI with 812 and 782 nm calculated from the field spectra could be employed to predict the As contents in agricultural soils. Choe et al.¹⁷ proved that it was feasible to extend ground-derived spectral parameters to remotely sensed images, although their differences in observation scales could lead to the discrepancies in spectral parameter values between image and *in situ* spectral measurements. Therefore, it might be credible to extend the NDSI with 812 and 782 nm calculated from field spectra to hyperspectral images for predicting the spatial distribution of the As contents in soils in our study area.

Wu et al.⁵⁴ suggested that using hyperspectral images to monitor heavy metal contamination in soils could be conducted in winter or early spring, when vegetation coverage is low. Our study demonstrated that the As contamination in soils could be assessed using the VNIRS of rice plants. However, some limitations should be taken into considerations when this method is applied to other areas. In our study, the As contamination in agricultural soils was the main stress factor restricting the growth of rice plants. However, in other areas, except for heavy metal contamination, the stress factors might be diverse and complicated, such as water shortage, fertilizer deficiency, pests, diseases, and multiple metals (As, Cu, Pb, and Cd etc.).

In further studies, it would be meaningful to examine the usefulness of the reflectance spectra of other crops (e.g., wheat, maize, and soybean) to simultaneously predict various toxic heavy metals (including As, Cu, Pb, and Cd) in different soil types. The information on soil quality is an important part of precision agriculture, and how to quickly and accurately acquire soil quality information is a key problem precision agriculture is facing. Vegetation is the “hair” of soil, and its growth conditions may reflect the degree of soil quality to some extent. Our study provided a kind of thought of acquiring soil quality information through remote sensing technique when soils were covered by vegetation.

AUTHOR INFORMATION

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China (Grant No. 41171290).

REFERENCES

- (1) Naidu, R.; Smith, E.; Owens, G.; Bhattacharya, P. *Managing arsenic in the environment: From soil to human health*; CSIRO: Collingwood, Victoria, Australia, 2006.
- (2) WHO Arsenic and arsenic compounds. *Environmental health criteria* 224; World Health Organization: Geneva, Switzerland, 2001.
- (3) Huang, R. Q.; Gao, S. F.; Wang, W. L.; Staunton, S.; Wang, G. Soil arsenic availability and the transfer of soil arsenic to crops in suburban areas in Fujian Province, southeast China. *Sci. Total Environ.* **2006**, *368*, 531–541.
- (4) Sacco, L.; Baldassarre, A.; Masotti, A. Diet's role in the toxicity of inorganic arsenic (iAs): A journey from soil to children's mouth. *J. Geochem. Explor.* **2013**, *131*, 45–51.
- (5) Santra, S. C.; Samal, A. C.; Bhattacharya, P.; Banerjee, S.; Biswas, A.; Majumdar, J. Arsenic in foodchain and community health risk: a study in Gangetic West Bengal. *Procedia Environ. Sci.* **2013**, *18*, 2–13.
- (6) Schwartz, G.; Eshel, G.; Ben-Dor, E. Reflectance spectroscopy as a tool for monitoring contaminated soils. In *Soil contaminants*; Pascucci, S., Ed.; InTech: Rejeka, Croatia, 2011; 10.5772/23661.
- (7) Leenaers, H.; Okx, J. P.; Burrough, P. A. Employing elevation data for efficient mapping of soil pollution on floodplains. *Soil Use Manage.* **1990**, *6*, 105–114.
- (8) Steiger, B. V.; Webster, R.; Schulin, R.; Lehmann, R. Mapping heavy metals in polluted soil by disjunctive Kriging. *Environ. Pollut.* **1996**, *94*, 205–215.
- (9) Liu, Y. L.; Li, W.; Wu, G. F.; Xu, X. G. Feasibility of estimating heavy metal contaminations in floodplain soils using laboratory-based hyperspectral data—A case study along Le'an River, China. *Geo-spat. Inf. Sci.* **2011**, *14*, 10–16.
- (10) Viscarra Rossel, R. A.; Walvoort, D. J. J.; McBratney, A. B.; Janik, L. J.; Skjemstad, J. O. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* **2006**, *131*, 59–75.
- (11) Stevens, A.; van Wesemael, B.; Bartholomeus, H.; Rosillon, D.; Tychon, B.; Ben-Dor, E. Laboratory, field and airborne spectroscopy for monitoring organic carbon content in agricultural soils. *Geoderma* **2008**, *144*, 395–404.
- (12) Wu, Y. Z.; Chen, J.; Ji, J. F.; Gong, P.; Liao, Q. L.; Tian, Q. J.; Ma, H. R. A mechanism study of reflectance spectroscopy for investigating heavy metals in soils. *Soil Sci. Soc. Am. J.* **2007**, *71*, 918–926.
- (13) Ren, H. Y.; Zhuang, D. F.; Singh, A. N.; Pan, J. J.; Qiu, D. S.; Shi, R. H. Estimation of As and Cu contamination in agricultural soils around a mining area by reflectance spectroscopy: A case study. *Pedosphere* **2009**, *19*, 719–726.
- (14) Wu, Y. Z.; Chen, J.; Wu, X. M.; Tian, Q. J.; Ji, J. F.; Qin, Z. H. Possibilities of reflectance spectroscopy for the assessment of contaminant elements in suburban soils. *Appl. Geochem.* **2005**, *20*, 1051–1059.
- (15) Kooistra, L.; Wanders, J.; Eperma, G. F.; Leuven, R. S. E. W.; Wehrens, R.; Buydens, L. M. C. The potential of field spectroscopy for the assessment of sediment properties in river floodplains. *Anal. Chim. Acta* **2003**, *484*, 189–200.
- (16) Leone, A. P.; Sommer, S. Multivariate analysis of laboratory spectra for the assessment of soil development and soil degradation in the Southern Apennines (Italy). *Remote Sens. Environ.* **2000**, *72*, 346–359.
- (17) Choe, E.; van der Meer, F.; van Ruitenbeek, F.; van der Werff, H.; de Smeth, B.; Kim, K. W. Mapping of heavy metal pollution in stream sediments using combined geochemistry, field spectroscopy, and hyperspectral remote sensing: A case study of the Rodalquilar mining area, SE Spain. *Remote Sens. Environ.* **2008**, *112*, 3222–3233.
- (18) Kooistra, L.; Salas, E. A. L.; Clevers, J. G. P. W.; Wehrens, R.; Leuven, R. S. E. W.; Nienhuis, P. H.; Buydens, L. M. C. Exploring field vegetation reflectance as an indicator of soil contamination in river floodplains. *Environ. Pollut.* **2004**, *127*, 281–290.
- (19) Manios, T.; Stentiford, E. I.; Millner, P. A. The effect of heavy metals accumulation on the chlorophyll concentration of *Typha latifolia* plants, growing in a substrate containing sewage sludge compost and watered with metaliferous water. *Ecol. Eng.* **2003**, *20*, 65–74.
- (20) Kooistra, L.; Leuven, R. S. E. W.; Wehrens, R.; Nienhuis, P. H.; Buydens, L. M. C. A comparison of methods to relate grass reflectance to soil metal contamination. *Int. J. Remote Sens.* **2003**, *20*, 4995–5010.
- (21) Liu, Y. L.; Wen, C.; Liu, X. China's food security soiled by contamination. *Science* **2013**, *339*, 1382.
- (22) Liu, M. L.; Liu, X. N.; Ding, W. C.; Wu, L. Monitoring stress levels on rice with heavy metal pollution from hyperspectral reflectance data using wavelet-fractal analysis. *Int. J. Appl. Earth Obs.* **2011**, *13*, 246–255.
- (23) Zhai, Y. F.; Cui, L. J.; Zhou, X.; Gao, Y.; Fei, T.; Gao, W. X. Estimation of nitrogen, phosphorus, and potassium contents in the leaves of different plants using laboratory-based visible and near-infrared reflectance spectroscopy: comparison of partial least-square regression and support vector machine regression methods. *Int. J. Remote Sens.* **2013**, *34* (7), 2502–2518.
- (24) Guo, J. H.; Ma, H.; Wang, S. F. Determination of arsenic in national standard reference soil and stream sediment samples by atomic fluorescence spectrometry (in Chinese). *Rock Miner. Anal.* **2009**, *28*, 182–184.
- (25) Muller, G. Index of geoaccumulation in sediments of the Rhine River. *Geojournal* **1969**, *2*, 108–118.
- (26) Loska, K.; Wiechula, D.; Korus, I. Metal contamination of farming soils affected by industry. *Environ. Int.* **2004**, *30*, 159–165.
- (27) Li, J. J.; Ma, J. T.; Chu, X. J.; Wang, S. B.; Wu, H.; Wang, Z. H. Application of index of geo-accumulation and enrichment factor in safety assessment of heavy-metal contamination in soil of copper refining (in Chinese). *China Saf. Sci. J.* **2006**, *16*, 135–139.
- (28) Verboven, S.; Hubert, M. LIBRA: A MATLAB library for robust analysis. *Chemom. Intell. Lab. Syst.* **2005**, *75*, 127–136.
- (29) Moron, A.; Cozzolino, D. Determination of potentially mineralizable nitrogen and nitrogen in particulate organic matter fractions in soil by visible and near-infrared reflectance spectroscopy. *J. Agric. Sci.* **2004**, *142*, 335–343.
- (30) Kemper, T.; Sommer, S. Estimate of heavy metal contamination in soils after a mining accident using reflectance spectroscopy. *Environ. Sci. Technol.* **2002**, *36*, 2742–2747.
- (31) Otto, M. *Chemometrics—Statistics and computer application in analytical chemistry*; Wiley-VCH: Weinheim, Germany, 1999.
- (32) Shi, T. Z.; Chen, Y. Y.; Liu, Y. L.; Wu, G. F. Visible and near-infrared reflectance spectroscopy—An alternative for monitoring soil contamination by heavy metals. *J. Hazard. Mater.* **2014**, *265*, 166–176.
- (33) Gomez, C.; Lagacherie, P.; Coulouma, G. Continuum removal versus PLSR method for clay and calcium carbonate content estimation from laboratory and airborne hyperspectral measurements. *Geoderma* **2008**, *148*, 141–148.
- (34) Akaike, H. Information theory and extension of maximum likelihood principle. *Second international symposium on information theory*; Akademia Kiado: Budapest, Hungary, 1973.
- (35) Chong, I. G.; Jun, C. H. Performance of some variable selection methods when multicollinearity is present. *Chemom. Intell. Lab. Syst.* **2005**, *78*, 103–112.
- (36) Haaland, D. M.; Thomas, E. V. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Anal. Chem.* **1988**, *60*, 1193–1202.
- (37) Viscarra Rossel, R. A. Parles: Software for chemometric analysis of spectroscopic data. *Chemom. Intell. Lab. Syst.* **2008**, *90*, 72–83.

- (38) Moros, J.; Vallejuelo, S. F.-O.; Gredilla, A.; Diego, A.; Madariaga, J. M.; Garrigues, S.; Guardia, M. Use of reflectance infrared spectroscopy for monitoring the metal content of the estuarine sediments of the Nerbioi-Ibaizabal River (Metropolitan Bilbao, Bay of Biscay, Basque County). *Environ. Sci. Technol.* **2009**, *43*, 9314–9320.
- (39) Huete, A. R. A soil vegetation adjusted index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309.
- (40) Inoue, Y.; Penuelas, J.; Miyata, A.; Mano, M. Normalized difference spectral indices for estimating photosynthetic efficiency and capacity at a canopy scale derived from hyperspectral and CO₂ flux measurements in rice. *Remote Sens. Environ.* **2008**, No. 112, 156–172.
- (41) Horder, D. N. H.; Dockray, M.; Barber, J. The red edge of plant leaf reflectance. *Int. J. Remote Sens.* **1983**, *4*, 273–288.
- (42) Wu, C. S.; Tong, Q. X.; Zheng, L. F.; Liu, W. D. Correlation analysis between spectral data and chlorophyll of rice and maize (in Chinese). *J. Basic Sci. Eng.* **2000**, *8*, 31–36.
- (43) Duan, G. L.; Wang, L. H.; Chen, Y.; Xu, Y. X.; Meng, X. Y.; Zhu, Y. G. Health risk from consumption of rice with elevated arsenic and studies of arsenic metabolism in rice plants (in Chinese). *J. Agro-Environ. Sci.* **2007**, *26* (2), 430–435.
- (44) Abedin, M. J.; Cresser, M. S.; Meharg, A. A.; Feldmann, J.; Cotter-Howells, J. Arsenic accumulation and metabolism in rice. *Environ. Sci. Technol.* **2002**, *36* (5), 962–968.
- (45) Inskeep, W. P.; McDermott, T. R.; Fendorf, S. Arsenic(V)/(III) cycling in soils and natural waters: Chemical and microbiological processes. In *Environmetal chemistry of arsenic*; Frankenberger, W. T. J., Ed.; CRC Press, Marcel Dekker: New York, NY, USA, 2002; p 183.
- (46) Takahashi, Y.; Minamikawa, R.; Hattori, K. H.; Kurishima, K.; Kihou, N.; Yuita, K. Arsenic behavior in paddy fields during the cycle of flooded and non-flooded periods. *Environ. Sci. Technol.* **2004**, *38*, 1038–1044.
- (47) Garcia-Manyes, S.; Jimenez, G.; Padro, A.; Rubio, R.; Rauret, G. Arsenic speciation in contaminated soils. *Talanta* **2002**, *58*, 97–109.
- (48) Vohland, M.; Besold, J.; Hill, J.; Frund, H. C. Comparing different multivariate calibration methods for the determination of soil organic carbon pools with visible to near infrared spectroscopy. *Geoderma* **2011**, *166* (1), 198–205.
- (49) Guo, Z. H.; Meng, M.; Hou, Y. L. Effect of P and As couple stress on growth, and P and As absorption of different P-tolerant rice seedlings (in Chinese). *Chin. J. Appl. Environ. Biol.* **2009**, *15* (5), 596–601.
- (50) Blackburn, G. A. Quantifying chlorophylls and carotenoids at leaf and canopy scales: An evaluation of some hyperspectral approaches. *Remote Sens. Environ.* **1998**, *66*, 273–285.
- (51) Rahman, M. A.; Hasegawa, H.; Rahman, M. M.; Islam, M. N.; Miah, M. A. M.; Tasmen, A. Effect of arsenic on photosynthesis, growth and yield of five widely cultivated rice (*Oryza sativa* L.) varieties in Bangladesh. *Chemosphere* **2007**, *67*, 1072–1079.
- (52) Liu, W. D.; Xiang, Y. Q.; Zheng, L. F.; Tong, Q. X.; Wu, C. S. Relationships between rice LAI, CH.D and hyperspectra data (in Chinese). *J. Remote Sens.* **2000**, *4*, 279–283.
- (53) Marin, A. R.; Pezeshki, S. R.; Masschelen, P. H.; Choi, H. S. Effect of dimethylarsenic acid (DMAA) on growth, tissue arsenic, and photosynthesis of rice plants. *J. Plant Nutr.* **1993**, *16*, 865–880.
- (54) Wu, Y. Z.; Chen, J.; Ji, J. F.; Tian, Q. J.; Wu, X. M. Feasibility of reflectance spectroscopy for the assessment of soil mercury contamination. *Environ. Sci. Technol.* **2005**, *39*, 873–878.