Evaluation of MLSR and PLSR for estimating soil element contents using visible/near-infrared spectroscopy in apple orchards on the Jiaodong peninsula

Xiang Yu [a,b,*], Qing Liu [c], Yebao Wang [a,b], Xiangyang Liu [a,b], Xin Liu [a,**]

[a] Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Yantai 264000, China
[b] University of Chinese Academy of Sciences, Beijing 100049, China
[c] College of Resources and Environment, Qingdao Agricultural University, Qingdao 266109, China

ARTICLE INFO

ABSTRACT

Conventional methods for soil element content determination based on laboratory analyses are costly and time-consuming. A soil reflectance spectrum is an alternative approach for soil element content estimation with the advantage of being rapid, non-destructive, and cost effective. Visible/near-infrared spectra (350 nm to 2500 nm) were measured from 105 soil samples originating from 30 apple orchards on the Jiaodong peninsula. The Savitzky–Golay (FD-SG) technique for spectral data was implemented to reduce the signal noise. Logarithm of the reciprocal of reflectance ($\log R^{-1}$) and the first derivative transformation (DR) were used to accentuate the features and to prepare the data for use in quantitative estimation models. The SI (sum index), DI (difference index), PI (product index), RI (ratio index), and NDI (normalized difference index) were calculated to extract sensitive waveband combinations that are significantly related to soil element contents. Soil element contents were retrieved based on sensitive waveband combinations by multiple linear stepwise regression (MLSR) and partial least square (PLSR) models. The results showed that DR performed better than $\log R^{-1}$ in eliminating the interfering factors of soil particle size and spectral noise. The MLSR and PLSR calibration models based on PI performed better than those based on SI or DI did. The MLSR performed better than PLSR in estimating soil elemental content. The contents of total nitrogen (TN), arsenic (As), and mercury (Hg) could be estimated well using MLSR and PLSR calibration models developed with PI. The MLSR calibration model developed with PI performed well in estimating available potassium (A-K) content. However, the contents of available phosphorus (A-P), ammonium nitrogen (NH_4^+-N), nitric nitrogen (NO_3^--N), and soil organic matter (SOM) could not be estimated using MLSR or PLSR calibration models. These outcomes will provide the theoretical basis and technical support for estimations of soil element content using visible/near-infrared spectra. Although they were shown to be useful in apple orchards of the Jiaodong peninsula, these models and methods should be further tested in soil samples from other regions and countries to prove their validity.

1. Introduction

It is difficult to measure elements in soil. Traditional determination methods cannot meet the need for rapid and timely surveys, despite their precision, because they are time-consuming and difficult. The development of rapid and economical soil monitoring methods for measuring soil element contents is required for soil fertility diagnosis and precision farming. In recent years, several spectral technologies have been developed (Ben-Dor et al., 1997; Reeves et al., 2000; Sorensen and Dalsgaard, 2005; Sudduth and Hummel, 1993), using portable spectrometers (Barnes et al., 2003) or hyperspectral remote sensors (Gomez et al., 2008). Models for estimating soil elemental contents using visible/near-infrared (VIS–NIR) spectra benefit from interactions within soil element contents and soil reflectance (Gaffey et al., 1993). The models offer high prediction accuracy and explain more than 80% of the variability (Chang and Laird, 2002; McCarty et al., 2002).

Factors that influence the reflectance of soil include soil moisture, particle size, and mineral composition, especially the presence of iron (Bowers and Hanks, 1965; Gaffey et al., 1993; Lobell and Asner, 2002; McCarty et al., 2002). Pre-processing transformations have been applied in previous studies to transform soil spectral data and remove signal noise as well as, to accentuate features and to prepare the data for use in quantitative estimation models (Dunn et al., 2002; Kooistra et al., 2003; McCarty et al., 2002). Common pre-processing transformations include smoothing, averaging, normalization, scatter correction,

baseline correction, and derivative algorithms. The measured diffuse reflectance spectra (R) are generally transformed to absorbance (logarithm of the reciprocal of reflectance, $\log R^{-1}$) to study the linearization between the spectra and concentration (Stenberg et al., 2010).

The first derivative transformation (DR) rapidly identifies spectral characteristic positions of inflection including maximum and minimum values. For example, Chang and Laird (2002) used the partial least square (PLSR) method to estimate total soil carbon, based on the first derivative and smoothing techniques. Dunn et al. (2002) developed a prediction model for soil organic carbon using the PLSR method with the first derivative spectra in the 400–2500 nm region.

Several calibration techniques based on VIS–NIR spectra have been used in the development of models to estimate soil properties. Considering the advantage of avoiding collinearity, MLSR has been used to develop models that estimate soil properties (Sullivan et al., 2005; Zheng, 2010; B.P. Liu et al., 2007; H.J. Liu et al., 2007; Krishnan et al., 1980). Similarly, a partial least square regression (PLSR) model has also been successfully constructed to avoid collinearity (Van-Waes et al., 2005; Xu and Xie, 2012; Zhang, 2010; Zhang et al., 2011).

However, there are some problems with the aforementioned methods for realistic applications. The MLSR and PLSR models are commonly developed with a wide range of spectral bandwidth to obtain high prediction accuracy. It is unclear whether the elements in soil can be predicted via MLSR or PLSR methods. These methods are based on the characteristic spectral absorption of soil elements. It also uses correlations between the soil element and soil constituents, such as iron oxides and clay minerals (Bartholomeus et al., 2008). To overcome these disadvantages and to improve soil element content prediction via spectral analysis, some studies have selected sensitive wavebands related to soil element contents and have developed spectral index method models (Bartholomeus et al., 2008; Lu et al., 2007; Sha et al., 2003). For instance, Galvão and Vitorello (1998) and Bartholomeus et al. (2008) developed spectral indices (RI, ratio index) for estimating SOM content based on the absorbed apex. These types of models can effectively solve the transferability among various remote sensors based on specified spectral waveband information.

The objectives of the study were: (1) to explore the characteristic spectral bands of R, $\log R^{-1}$, and DR and develop sensitive waveband combinations for estimating soil element contents based on five spectral indices using SI (sum index), DI (difference index), PI (product index), RI (ratio index), and NDI (normalized difference index); (2) to develop calibration models to estimate soil element contents using MLSR and PLSR modeling methods based on sensitive waveband combinations of soil element contents; and (3) to compare the performance of published indices and these models for soil element estimation.

2. Material and methods

2.1. Study areas

The study area was Qixia County, which is located on the Jiaodong peninsula in Shandong province (37°05′–37°32′N, and 120°15′–121°33′E) (Fig. 1). This region is under a warm temperate monsoon type continental climate with mean annual rainfall of 664 mm; the mean annual temperature is 11.6 °C. The annual sunshine duration is 2631 h. The climate has four distinctive seasons with no severe cold in winter and no heat in summer. Its terrain is composed of gently hilly mountains mostly covered with brown loam soil that is slightly acidic, rich in mineral elements, and with good permeability. The unique geographical location and the superior climate soil conditions are suitable for the growth of apples. Qixia is an apple-growing region around the Bohai Gulf. It is an apple advantage area designated by the Ministry of Agriculture. Qixia is described as the first apple city of China. It boasts of pollution-free apple production and has been honored as the apple city (Zhu et al., 2009).

2.2. Field soil samples

Soil samples distributed in the soil layers at 20–40 cm depth were gathered in April 2009. Before sampling, the natural grass surfaces were removed. Each sample was mixed with multiple nearby samples and placed in a numbered polymer bag. Dry soil blocks were cracked after being air-dried. Intrusive bodies such as the roots of the plants and stone were removed roughly with grinding and sieving in the final steps. Eight soil elements were studied (The Committee of Agrochemistry of the Chinese Society of Soil Science, 1983). Soil organic matter (SOM) was measured by the $K_2Cr_2O_7$ oxidation method. An elemental analyzer (Elementar Vario Micro, Germany) was implemented to measure total nitrogen (TN) content. The NH_4^+-N and NO_3^--N contents were given by sequence flow analyzer (San++ SKALAR, Netherlands). Available phosphorus (A-P) was determined by molybdenum blue spectrophotometry. Available potassium (A-K) was determined by flame atomic absorption spectrometry. Concentrations of arsenic (As) and mercury (Hg) were determined by Agilent 7500 ICP-MS (Inductively coupled plasma mass spectrometry) after being digested by aqua regia. Elemental analyses of the sampling sites are showed in Table 1. The dataset (n = 105) was randomly divided into 53 samples (50%) for model development and 52 samples (50%) for model validation. The statistical values of the soil element contents for the calibration soil samples and the validation samples showed that the soil element contents had a larger range. Selection of calibration
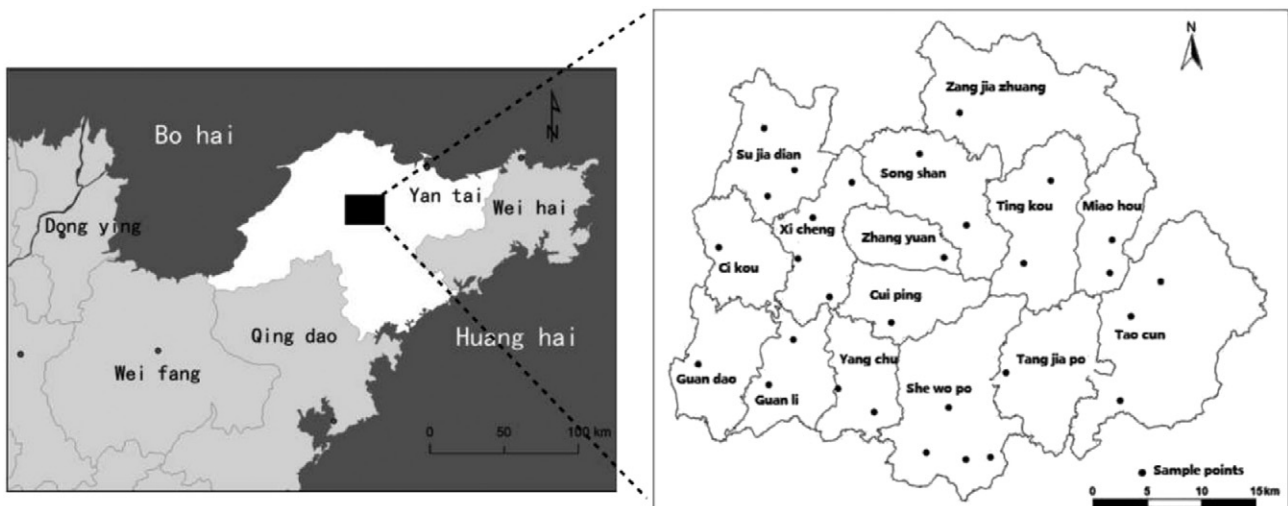

Fig. 1. Distribution of study objects collection plots.

Table 1
Basic statistics for soil element contents.

	N		Min		Max		Mean		SD	
	Total	Calibration	Total	Calibration	Total	Calibration	Total	Calibration	Total	Calibration
A-P (mg/kg)	105	53	2.15	4.36	337.52	337.52	75.62	95.61	64.49	61.09
A-K (mg/kg)	105	53	25.89	25.89	6600.24	1003.14	339.17	238.02	704.80	176.87
NH_4^+-N (mg/kg)	105	53	0.45	1.26	110.99	95.40	14.28	11.10	24.73	16.41
NO_3^--N (mg/kg)	105	53	1.32	2.42	126.56	126.56	16.04	23.24	20.68	18.16
TN (%)	105	53	0.06	0.06	0.39	0.25	0.15	0.14	0.058	0.04
SOM (%)	105	53	0.48	0.48	2.13	2.13	0.97	1.01	0.30	0.16
As (mg/kg)	105	53	2.79	5.03	20.93	19.77	10.59	10.50	3.23	3.09
Hg (mg/kg)	105	53	0.01	0.03	1.02	0.85	0.12	0.11	0.13	0.11

samples was used for modeling because the SD of the calibration samples was lower than that of total samples.

2.3. Measurement and preprocessing of soil visible/near-infrared spectra

Reflectance spectra of soil samples were measured with an ASD Field-Spec FR2500 Spectra-radiometer (Analytical Spectral Devices Inc., Boulder, CO, USA) with a spectral range of 350–2500 nm (Hatchell, 1999). A spectrum reflectance panel with 99% (Lab-sphere, Inc., North Sutton, NH, USA) was utilized as a reference standard to adjust and optimize the spectrometer for incoming irradiation. As long as the visible/near-infrared spectrum of one soil sample was measured, the optimization procedure occurred. Soils were scanned from below using a high intensity source probe (Analytical Spectral Devices Inc., Boulder, CO, USA). The sampling interval was 1.4 nm at 350–1050 nm and 2 nm at 1000–2500 nm. The spectral resolution was 3 nm at 350–1050 nm and 10 nm at 1000–2500 nm. The measurement conditions included the probe view angle (25°), light incident angle (45°), light source distance (30 cm) and probe distance (15 cm). The soil spectral curves of the four directions were gained by rotating the sample dishes three times 90° at a time to reduce the influence of the soil sample spectral anisotropy when measuring. Savitzky–Golay smoothing was used to reduce the signal noise. The $logR^{-1}$ and DR at each waveband were computed to remove the baseline effects (Ben-Dor et al., 1997; Duckworth, 1998) and to strengthen the spectral features (Lu et al., 2007; Schlerf et al., 2010)(Fig. 2). The performance of R, $logR^{-1}$, and DR were compared, and the characteristic spectral wavebands significantly related to soil element contents were explored based on correlation analysis.

2.4. Spectral indices and estimation models

Spectra indices as SI (Eq. (1)), DI (Eq. (2)), PI (Eq. (3)), RI (Eq. (4)), and NDI (Eq. (5)) were calculated based on characteristic spectral wavebands of R, $logR^{-1}$, and DR. The parameters were then correlated with soil elemental contents to select the best spectral index for estimating the soil elemental contents where R_i and R_j are the reflectance

at i nm (reference wavelength) and j nm (effective wavelength), respectively. A computer program based on MATLAB 6.0 software (the MathWorks, Inc., Natick, MA) was developed to assist with analyses. Sensitive waveband combinations significantly related to soil element contents were extracted.

$$SI_{(i,j)} = (R_i + R_j) \tag{1}$$

$$DI_{(i,j)} = (R_i - R_j) \tag{2}$$

$$PI_{(i,j)} = (R_i * R_j) \tag{3}$$

$$RI_{(i,j)} = (R_i/R_j) \tag{4}$$

$$NDI_{(i,j)} = (R_i - R_j)/(R_i + R_j). \tag{5}$$

In addition, the MLSR and PLSR regression methods (Geladi and Kowalski, 1986) were used to establish the relationship between soil element contents and spectra of SI, DI, PI, RI, and NDI. SPSS Version 17.0 for Windows (Statistical Package for the Social Sciences Inc., Chicago, USA) was used for statistical analysis. The coefficients of determination (R^2), root mean squared error (RMSE), and relative prediction deviation (RPD) were used to evaluate the models' performance. The ratio of performance to deviation (RPD), which was defined as the ratio of standard deviation (SD) to RMSE, was employed as a sign of model stability (Nduwamungu et al., 2009). Chang et al. (2001) defined 3 categories of models according to the RPD values: i) model in category A (RPD > 2) accurately predicts a given property; ii) model in category B (1.4 < RPD < 2.0) is an intermediate class; and iii) model in category C (RPD < 1.4) has no prediction ability.

3. Results

3.1. Soil visible/near-infrared spectra

The mean reflectance spectral data of calibration samples and validation samples (Fig. 2a) showed that the shapes of reflectance were
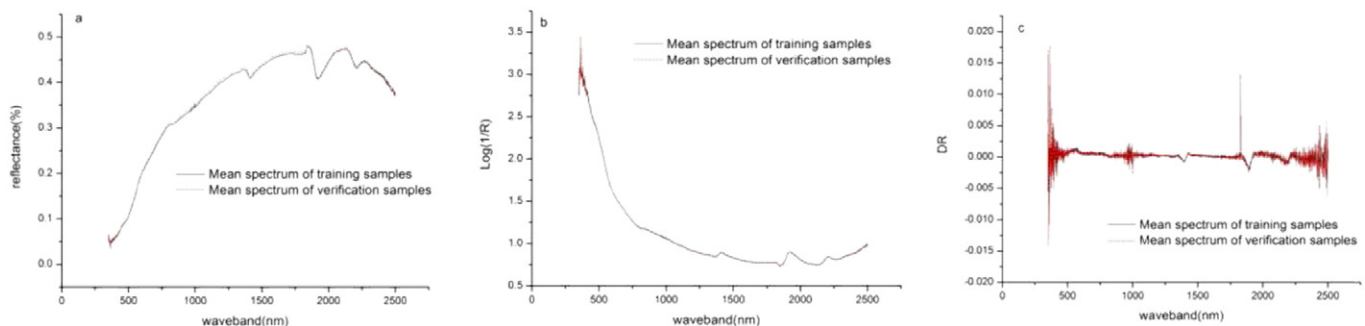
Fig. 2. Mean spectra of (R)(a), $logR^{-1}$(b), and DR (c) for calibration and verification samples.

similar to those of soil samples in other studies (Huang and Liu, 1995; Gomez et al., 2008; Zornoza et al., 2008). The reflectance profiles rose and shifted quickly toward the long-waveband direction from 350 nm to 760 nm. The absorption band at 450–480 nm and the reflection peak at 520–580 nm could be attributed to the existence of hematite, limonite, and goethite in the soil (Zhao, 2003). The reflectance changed gradually in the spectral range of 600–680 nm. The reflectance increased slightly in the 850–2350 nm range. The spectral curve had three absorption features at 1380, 1940, and 2250 nm, which indicate the presence of hydroxide (OH) in free water (1400 and 1900) and the Al–OH lattice in clay minerals (2200 nm) (Ben-Dor et al., 2002; Viscarra Rossel et al., 2006; Summers et al., 2011).

3.2. Relationship of soil element contents to spectral indices

Relationship of soil element contents to spectral formats as R, $\log R^{-1}$, and DR were calculated in the entire 350–2500 nm spectral region (Fig. 3, Table 2). Significant correlations between R and As occurred at 30 wavebands ($r > 0.195$, $p < 0.05$). No significant correlation was observed between R and any other soil element content. Similarly, significant correlations between $\log R^{-1}$ and As occurred at 25 wavebands. No significant correlation was observed between $\log R^{-1}$ and any other soil element content. However, significant correlations between DR and soil element contents occurred at more wavebands than at R or $\log R^{-1}$. The correlation coefficients (r) were calculated between SI, DI, PI, RI, and NDI and soil element contents including all possible pair combinations of wavebands that were significantly related to soil element contents (Fig.4, Table 3).

3.3. Performance of multiple linear stepwise regression and partial least square regression

Coefficients of determination (R^2) were constructed using MLSR and PLSR regression for SI, DI, PI, RI, and NDI against soil element contents including sensitive waveband combinations that were significantly related to soil element contents. The numbers of sensitive waveband combinations used in the regression analysis are listed in Table 3. However, no valid values for RI or NDI for DR have been gained as a result of the existence of zero (Table 3). Thus the MLSR and PLSR regression of RI and NDI against soil element contents was not performed. PLSR regression models of SI calculated based on R and $\log R^{-1}$ against As content performed poorly ($R^2 = 0.375$ and 0.232, respectively). The PLSR regression models of PI calculated based on R against As content still performed poorly ($R^2 = 0.252$). In addition, the MLSR regression models of SI calculated based on the two spectral formats (R or $\log R^{-1}$), and PI calculated based on R against As content were not useful. The information above indicated that R and $\log R^{-1}$ performed poorly in accentuating the features and in preparing the data for use in quantitative estimation models.

Table 2
Correlations (r) between spectral formats (R, $\log R^{-1}$, and DR) and soil element contents.

|	R			$\log R^{-1}$			DR		
	$/r_{min}/$	$/r_{max}/$	n	$/r_{min}/$	$/r_{max}/$	n	$/r_{min}/$	$/r_{max}/$	n
A-P	0	0.15	0	0	0.16	0	0	0.29*	85
A-K	0	0.17	0	0	0.18	0	0	0.31*	70
NH_4^+-N	0	0.16	0	0	0.15	0	0	0.32*	98
NO_3^--N	0	0.19	0	0	0.19	0	0	0.38*	99
TN	0	0.2	0	0	0.19	0	0	0.34*	112
SOM	0	0.18	0	0	0.19	0	0	0.34*	99
As	0	0.31*	30	0	0.31*	25	0	0.29*	99
Hg	0	0.13	0	0	0.12	0	0	0.33*	91

n = the number of sensitive wavebands.
* Significant correlation ($r > 0.195$, $p < 0.05$).

Fig. 5 showed the best results of MLSR and PLSR regression of SI, DI, and PI calculated based on DR against soil element contents. The optimal numbers of latent variables for PLSR calibration models used to estimate the soil element contents were cross-validated with SIMCA-P 11.5 (Table 4). The use of MLSR and PLSR algorithms to the three spectral indices based on DR resulted in the calibration models listed in Table 5. The results showed that the MLSR and PLSR calibration models based on different spectral indices exhibited significant differences. The performances of the MLSR calibration models with selected sensitive waveband combinations were better than that of PLSR calibration models. The MLSR and PLSR calibration models based on PI mainly performed better than those models based on SI or DI.

3.4. Testing soil element contents via estimation models

To test whether the calibration models described above were reliable and applicable to the estimation of soil elemental contents, independent datasets of 52 soil samples were used to validate their respective performances (Fig. 6). The performance of MLSR and PLSR models were evaluated with R^2, RMSE, and RPD when calibration samples and validation samples were concurrently considered. The results showed that MLSR and PLSR calibration models based on SI or DI performed poorly in estimating soil element contents. The MLSR calibration model based on PI performed poorly in estimating the A-P content ($R^2 = 0.63$, RMSE = 42.09, RPD = 1.62). However, MLSR calibration models based on PI performed well in estimating A-K contents ($R^2 = 0.97$, RMSE = 144.56, RPD = 5.75); the MLSR validation model was stable ($R^2 = 0.95$, RMSE = 111.29, RPD = 4.26). The A-K content could be evaluated well using MLSR calibration models developed with PI. As to the MLSR calibration models based on PI in estimating TN ($R^2 = 0.91$, RMSE = 0.01, RPD = 3.28), As ($R^2 = 0.79$, RMSE = 1.43, RPD = 2.14) and Hg ($R^2 = 0.99$, RMSE = 0.01, RPD = 10.34), the MLSR validation models were also stable with the R^2 values of 0.96, 0.78, and 0.99. The RMSE were 0.01, 1.62, and 0.01, and the RPD were 5.16, 2.08, and 10.78, respectively. The results indicated that the contents of A-K, TN,
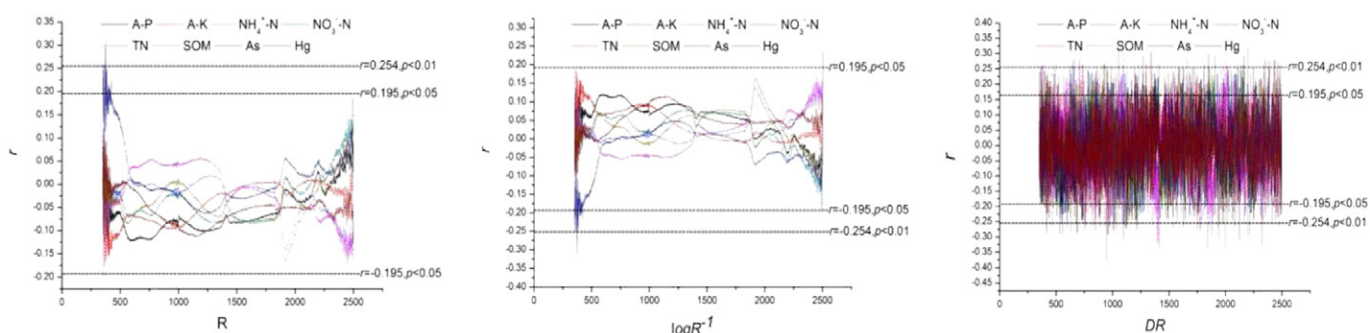
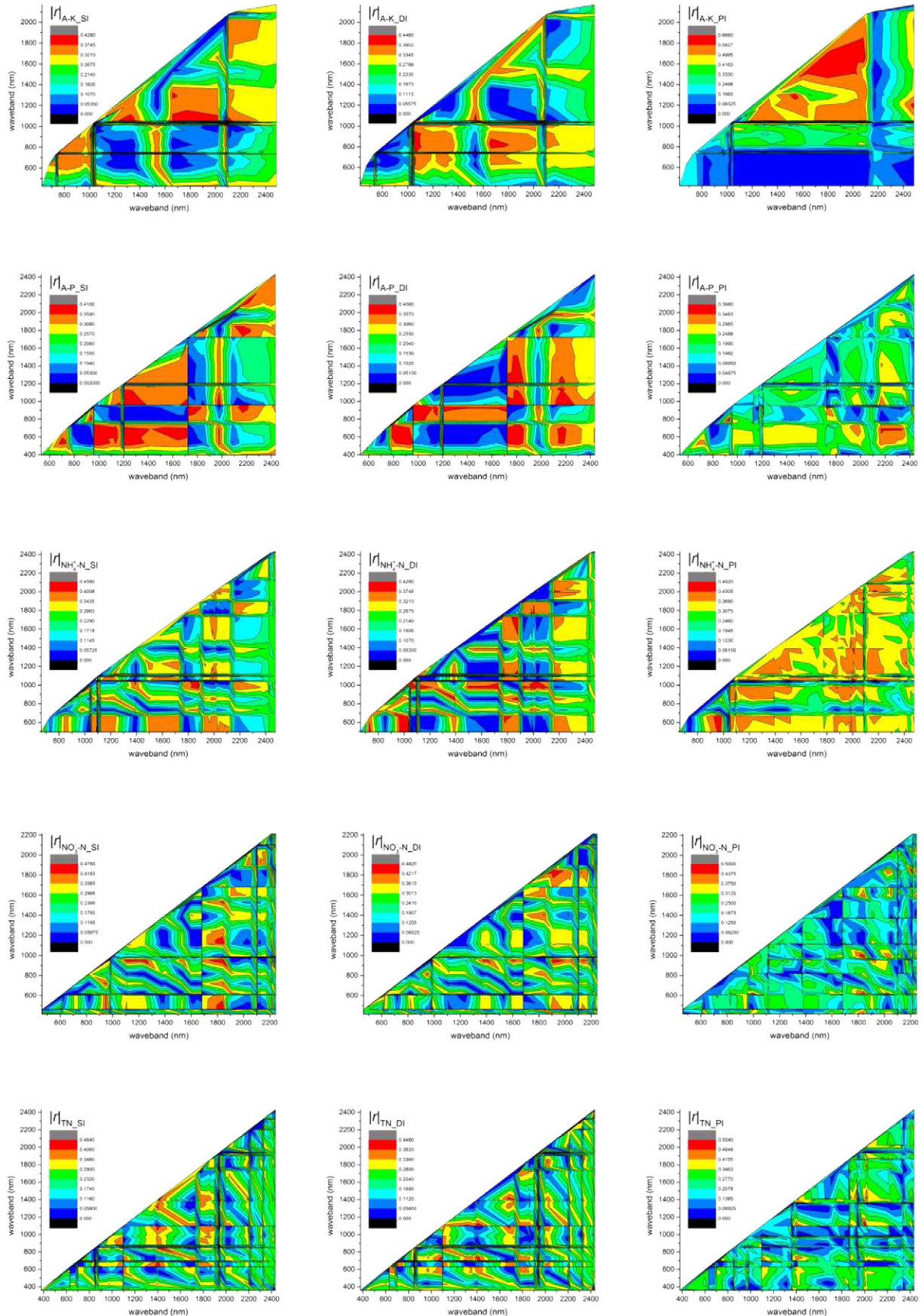**Fig. 3.** Correlations (r) between three spectral formats (R, $\log R^{-1}$, and DR) and soil element contents.

Fig. 4. Correlations of soil element contents to three spectral indices (SI, DI, and PI) calculated based on sensitive wavebands.
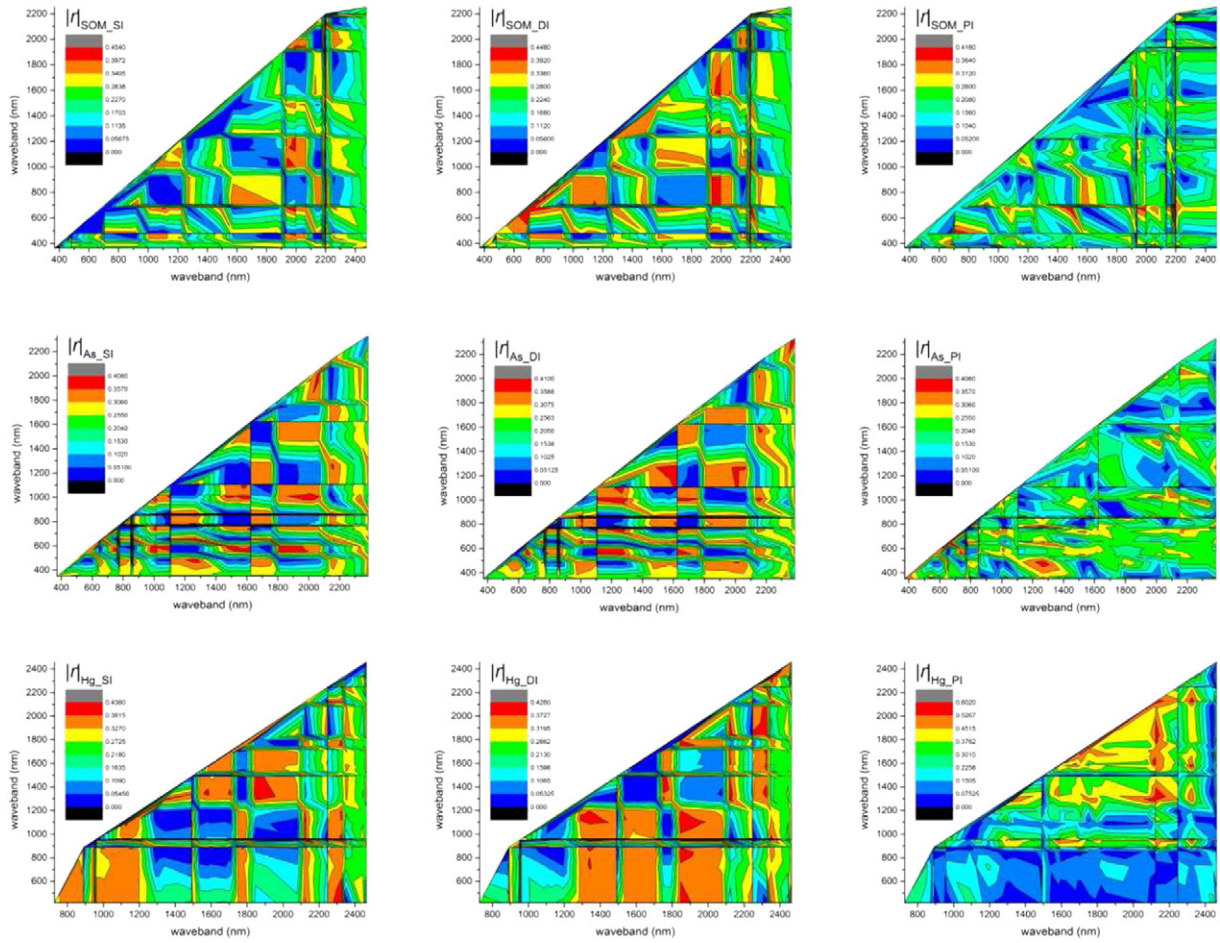
Fig. 4 (*continued*).

As, and Hg could be estimated well using MLSR calibration models developed with PI. Similarly, the PLSR calibration model based on PI performed well in estimating the contents of TN ($R^2 = 0.77$, RMSE = 0.02, RPD = 2.06), As ($R^2 = 0.85$, RMSE = 1.19, RPD = 2.56), and Hg ($R^2 = 0.98$, RMSE = 0.02, RPD = 7.69). The PLSR validation models remained stable with R^2 values of 0.92, 0.85, and 0.98. The RMSE values were 0.02, 1.31, and 0.02, and the RPD values were 3.56, 2.57, and 6.65, respectively. The TN, As, and Hg could be estimated well using PLSR calibration models developed with PI. The performance of the MLSR calibration model developed with PI was better than the PLSR calibration model in estimating TN content ($R^2 = 0.91$ and 0.77, respectively).

The performance of the PLSR calibration model developed with PI was better than the MLSR calibration models in estimating As content ($R^2 = 0.85$ and 0.79, respectively). The performance of the MLSR calibration model developed with PI was only slightly better than that of the PLSR calibration models in estimating Hg content ($R^2 = 0.99$ and 0.98, respectively). Moreover, the MLSR validation model developed with PI performed well in estimating A-K content ($R^2 = 0.95$, RPD = 4.26), but the PLSR validation model performed poorly ($R^2 = 0.71$, RPD = 1.59). Overall, the performances of the MLSR calibration models developed with PI were better than those of the PLSR calibration models developed with PI except for As. However, the A-P, NH_4^+-N.

Table 3
Correlations (r) between spectral indices and soil element contents.

		SI			DI			RI			PI			NDI		
		$/r_{min}/$	$/r_{max}/$	n	$/r_{min}/$	$/r_{max}/$	n	$/r_{min}/$	$/r_{max}/$	n	$/r_{min}/$	$/r_{max}/$	n	$/r_{min}/$	$/r_{max}/$	n
R	As	0	0.32*	148	0	0.19	0	0	0.18	0	0	0.33*	188	0	0.19	0
$logR^{-1}$	As	0	0.28*	15	0	0.28*	1	0	0.28*	1	0	0.28*	1	0	0.28*	1
DR	A-P	0	0.41*	187	0	0.41*	198	–	–	–	0	0.40*	60	–	–	–
	A-K	0	0.43*	148	0	0.45*	157	–	–	–	0	0.66*	111	–	–	–
	NH_4^+-N	0	0.46*	301	0	0.43*	294	–	–	–	0	0.50*	467	–	–	–
	NO_3^--N	0	0.48*	307	0	0.48*	323	–	–	–	0	0.50*	166	–	–	–
	TN	0	0.46*	490	0	0.45*	500	–	–	–	0	0.55*	434	–	–	–
	SOM	0	0.45*	285	0	0.45*	302	–	–	–	0	0.42*	158	–	–	–
	As	0	0.41*	260	0	0.41*	270	–	–	–	0	0.40*	166	–	–	–
	Hg	0	0.44*	294	0	0.44*	306	–	–	–	0	0.60*	278	–	–	–

n = the number of sensitive waveband combinations. – is null value.
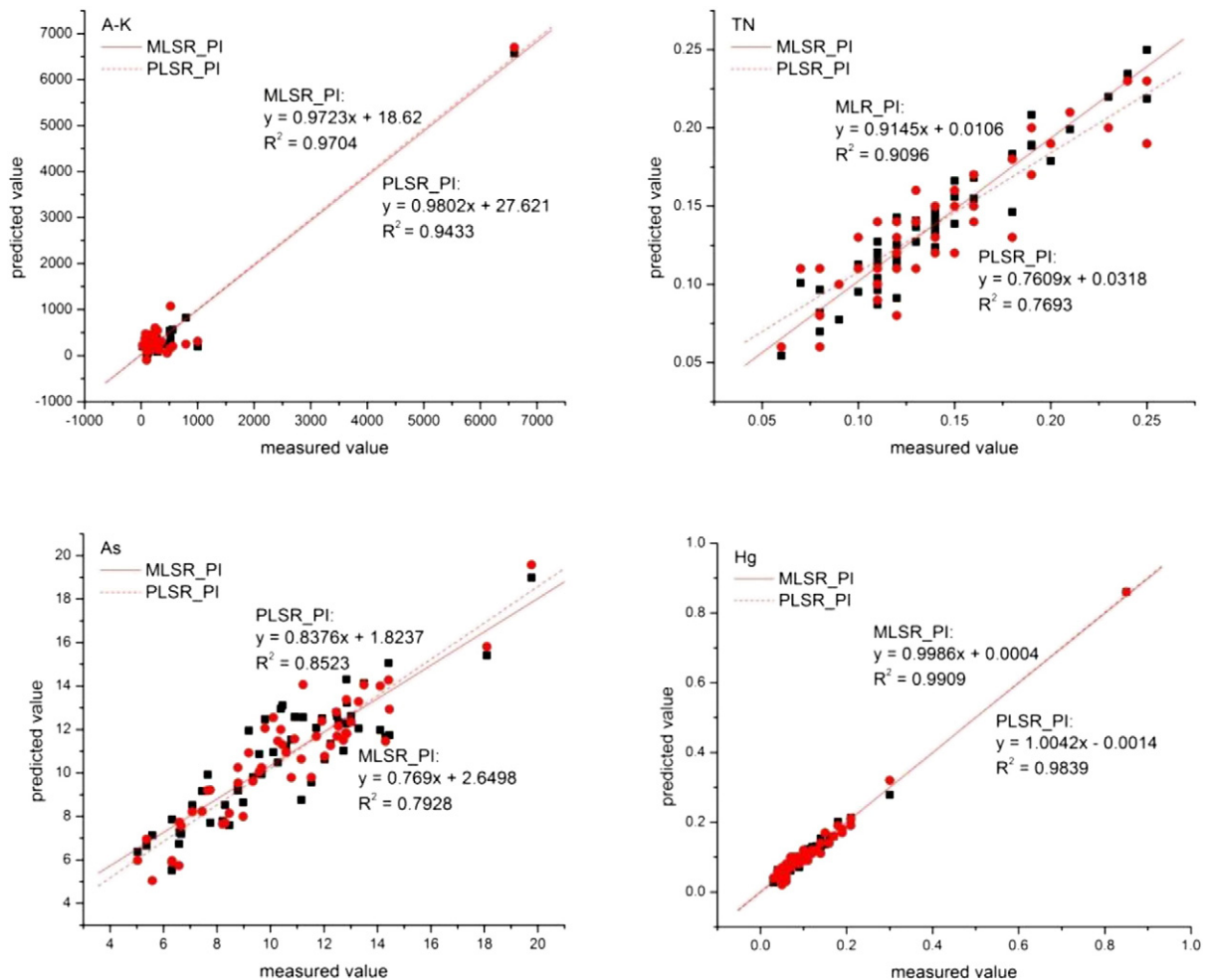
* Significant correlation ($r > 0.195$, $p < 0.05$).

Fig. 5. Comparison of observed and predicted contents of A-K, TN, As, and Hg based on MLSR and PLSR calibration models.

NO_3^--N and SOM could not be estimated using MLSR or PLSR calibration models.

4. Discussion

Soil reflectance spectral characteristics are closely related to the physical and chemical properties. Therefore, differences in spectral characteristics are attributable to the physical and chemical properties

Table 4
Cross-validation results for the PLS models of soil element contents with different spectral indices calculated based on DR.

		Number of factors	R^2	RMSE			Number of factors	R^2	RMSE
A-P	SI	2	0.67	40.06	TN	SI	3	0.63	0.03
	DI	1	0.57	45.47		DI	5	0.60	0.03
	PI	1	0.63	44.71		PI	4	0.77	0.02
A-K	SI	1	0.73	446.30	SOM	SI	2	0.80	0.17
	DI	1	0.72	449.49		DI	4	0.79	0.17
	PI	1	0.94	203.35		PI	2	0.82	0.16
NH_4^+-N	SI	2	0.45	13.22	As	SI	1	0.71	1.79
	DI	5	0.53	12.58		DI	2	0.69	1.76
	PI	3	0.67	10.16		PI	5	0.85	1.19
NO_3^--N	SI	1	0.70	14.63	Hg	SI	2	0.73	0.06
	DI	4	0.79	11.77		DI	3	0.75	0.06
	PI	5	0.93	6.75		PI	4	0.98	0.02

of different soil types (Xu, 2000). The technical analysis of soil spectra in the laboratory is popular because it is easily controlled and performed. Many studies have concentrated on sensitive spectral wavebands and the relationship with soil elements (Dematte et al., 2007; Lu et al., 2007). For instance, Lu et al. (2007) found that the SOM of black soil had a negative correlation with the reflectance spectra in the 545–738 nm range as well as the first derivative spectra in the 481–598 nm range. They also found a positive correlation with the first derivative spectra in the 816–932 nm and 1039–1415 nm ranges. The results of this study also showed that the first derivative spectral formats correlated well with soil element contents. These results are consistent with previous studies and suggest that the original spectral reflectance of the first derivative spectra could be the most sensitive for evaluating the soil element contents.

Selected spectral indices have been proposed to estimate soil nutrient parameters. He et al. (2006) developed a soil iron oxide index $SFOI = (Lg_{Red} - Lg_{Blue}) / (Lg_{Red} + Lg_{Blue})$ for estimating iron oxide content by analyzing the spectral absorption characteristics of iron oxide. Here red and blue were averaged values of reflectance in the 700–760 nm and 450–510 nm ranges, respectively. An organic matter identification index (OII, $R_{(507-516)}/R_{(450-750)}$) and the normalized first derivative spectra were also proposed for the retrieval of SOM content in black soil from northeast China (B.P. Liu et al., 2007; H.J. Liu et al., 2007; He et al., 2007). Thus, the spectral index technique is a potential method to evaluate soil nutrient status and was used here to analyze the correlations of the reflectance spectra and the first derivative spectra with soil elemental analysis.

Table 5
Prediction results for the MLR and PLS models of soil element contents with different spectral indices.

			SI			DI			PI		
			R^2	RMSE	RPD	R^2	RMSE	RPD	R^2	RMSE	RPD
A-P	Calibration	MLR	0.67	30.64	1.72[*]	0.67	39.94	1.71[*]	0.63	42.09	1.62[*]
		PLS	0.67	40.06	1.70[*]	0.57	45.47	1.50[*]	0.63	44.71	1.53[*]
	Validation	MLR	0.44	40.26	1.29	0.51	38.06	1.37	0.48	39.09	1.33
		PLS	0.49	38.92	1.33	0.51	37.67	1.38	0.39	44.32	1.17
A-K	Calibration	MLR	0.75	427.89	1.94[*]	0.72	446.63	1.86[*]	0.97	144.56	5.75[**]
		PLS	0.73	446.30	1.86[*]	0.72	449.49	1.85[*]	0.94	203.35	4.09[**]
	Validation	MLR	0.51	360.65	1.32	0.58	328.51	1.45[*]	0.95	111.29	4.26[**]
		PLS	0.37	413.07	1.15	0.42	388.00	1.22	0.71	298.36	1.59[*]
NH_4^+-N	Calibration	MLR	0.45	12.99	1.25	0.46	13.48	1.21	0.67	9.88	1.65[*]
		PLS	0.45	13.22	1.23	0.53	12.58	1.29	0.67	10.16	1.60[*]
	Validation	MLR	0.74	16.14	1.89[*]	0.78	15.01	2.03[**]	0.92	8.75	3.49[**]
		PLS	0.78	14.85	2.06[**]	0.81	14.27	2.14[**]	0.92	8.97	3.40[**]
NO_3^--N	Calibration	MLR	0.77	12.35	2.02[**]	0.77	12.57	1.98[**]	0.89	8.43	2.96[**]
		PLS	0.70	14.63	1.70[*]	0.79	11.77	2.12[**]	0.93	6.75	3.69[**]
	Validation	MLR	0.36	10.37	1.04	0.33	10.52	1.02	0.53	7.86	1.37
		PLS	0.30	11.29	0.95	0.39	9.66	1.11	0.73	5.69	1.89[*]
TN	Calibration	MLR	0.59	0.03	1.48[*]	0.46	0.03	1.28	0.91	0.01	3.28[**]
		PLS	0.63	0.03	1.59[*]	0.60	0.03	1.50[*]	0.77	0.02	2.06[**]
	Validation	MLR	0.90	0.03	2.31[**]	0.89	0.03	2.27[*]	0.96	0.01	5.16[**]
		PLS	0.90	0.03	2.30[**]	0.91	0.03	2.54[**]	0.92	0.02	3.56[**]
SOM	Calibration	MLR	0.79	0.17	2.11[**]	0.78	0.17	2.08[**]	0.90	0.11	3.10[**]
		PLS	0.80	0.17	2.13[**]	0.79	0.17	2.09[**]	0.82	0.16	2.26[**]
	Validation	MLR	0.65	0.14	1.61[*]	0.59	0.15	1.47[*]	0.75	0.12	1.85[*]
		PLS	0.64	0.15	1.52[*]	0.79	0.12	1.85[*]	0.61	0.15	1.49[*]
As	Calibration	MLR	0.64	1.90	1.61[*]	0.58	2.04	1.50[*]	0.79	1.43	2.14[**]
		PLS	0.71	1.79	1.71[*]	0.69	1.76	1.73[*]	0.85	1.19	2.56[**]
	Validation	MLR	0.68	1.97	1.71[*]	0.69	1.93	1.74[*]	0.78	1.62	2.08[**]
		PLS	0.60	2.23	1.51[*]	0.69	1.94	1.73[*]	0.85	1.31	2.57[**]
Hg	Calibration	MLR	0.66	0.07	1.67[*]	0.66	0.07	1.68[*]	0.99	0.01	10.34[**]
		PLS	0.73	0.06	1.87[*]	0.75	0.06	1.96[*]	0.98	0.02	7.69[**]
	Validation	MLR	0.74	0.07	1.92[*]	0.72	0.08	1.86[*]	0.99	0.01	10.78[**]
		PLS	0.76	0.07	2.00[**]	0.76	0.07	2.01[**]	0.98	0.02	6.65[**]

[*] Model in category B.
[**] Model in category A.

The MLSR and PLSR methods were also used to develop the calibration models for estimating soil element contents. The results showed that for both methods, the accuracy of prediction was improved when sensitive spectral waveband regions were used to calibrate the model. This suggests that these prediction models based on MLSR or PLSR methods were based more directly upon the characteristic spectral absorption of soil element contents than models calibrated with the full spectrum wavebands. In addition, the accuracies of the MLSR and PLSR models are typically influenced by soil texture and color (Dalal and Henry, 1986; Krishnan et al., 1980). A good pre-processing technique can be used with spectral data to improve the prediction accuracy (Bartholomeus et al., 2008; Kooistra et al., 2003). For instance, compared with published results (R^2 = 0.86, 0.72, 0.86, 0.90; Daniel et al., 2003; McCarty et al., 2002; Viscarra Rossel et al., 2006), the SOM calibration models developed here may potentially be more reliable and precise. This is because of the selection of the pre-processing method of DR and the application of sensitive waveband combinations for the model calibration rather than full spectrum wavebands (400–2500 nm).

A comparison of the performance of calibration models using MLSR and PLSR techniques indicates that MLSR is better than PLSR. In addition, DR performed better than $logR^{-1}$ in eliminating the interfering factors of soil particle size and spectral noise and was suitable for the ASD instrumentation used in this study. PI performed better than SI and DI in developing sensitive band combinations for estimating soil element contents. Thus, the results are repeatable when similar instruments are used, with the same measurement configuration and the same pre-processing methodology. However, these models and methods should be tested further with soils from other regions and countries to confirm their validity. In addition, MLSR and PLSR are based on linear algorithms modeling spectral responses, but soil visible/near-infrared spectra may be nonlinear (Fuller and Griffiths, 1978). Further study should be conducted to estimate soil element contents based on nonlinear regression technology techniques.

5. Conclusions

Based on comprehensive analysis of the relationship between soil element contents and corresponding reflectance spectra formats (R, $logR^{-1}$ and DR), MLSR and PLSR were applied to estimate soil element contents based on five spectral indices (SI, DI, PI, RI, and NDI). The DR performed better than $logR^{-1}$ in eliminating the interfering factors of soil particle size and spectral noise and was suitable for the ASD instrumentation used in this study. MLSR and PLSR calibration models based on PI performed better than those based on SI or DI. MLSR performed better than PLSR in estimating soil element contents. The contents of TN, As, and Hg could be estimated very well using MLSR and PLSR calibration models developed with PI. The MLSR calibration model alone developed with PI performed well in estimating A-K content. However, contents of A-P, NH_4^+-N, NO_3^--N, and SOM could not be estimated using MLSR or PLSR calibration models. These outcomes will provide the theoretical basis and technical support for estimations of soil element contents using visible/near-infrared spectra. These estimation models must undergo further testing and optimization prior to wider application for soil ecosystem observation.
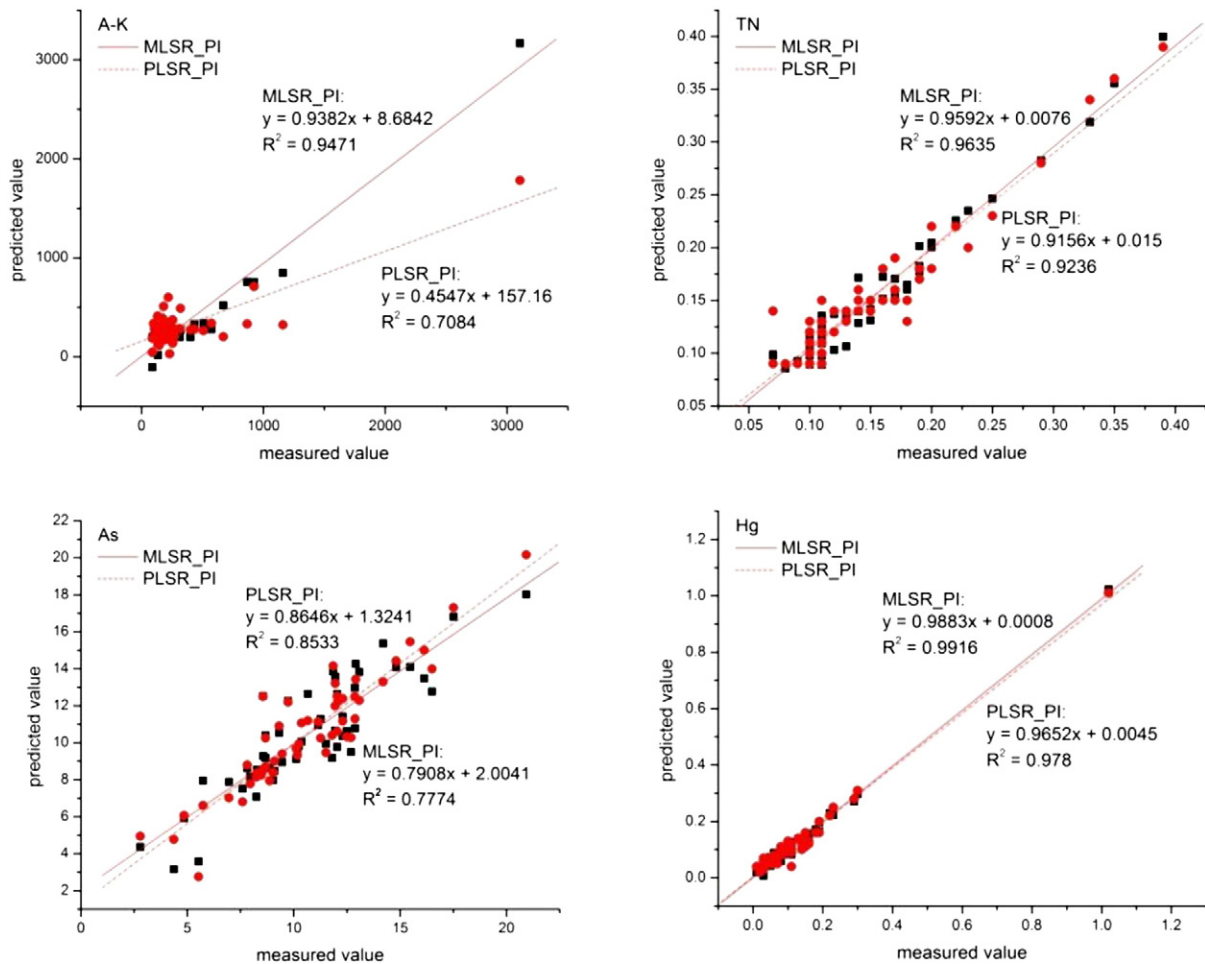
Fig. 6. Comparison of observed and predicted contents of A-K, TN, As, and Hg based on MLSR and PLSR validation models.

the Chinese Academy of Sciences, and ZR2011DL013 from the Shandong Province Natural Science Fund Committee.

References

Barnes, E.W., Sudduth, K.A., Hummel, J.W., Lesch, S.M., Corwin, D.L., Yang, C., Daughtry, C.S.T., Bausch, W.C., 2003. Remote and ground-based sensor techniques to map soil properties. Photogramm. Eng. Remote. Sens. 69, 619–630.

Bartholomeus, H.M., Schaepman, M.E., Kooistra, L., Stevens, A., 2008. Spectral reflectance based indices for soil organic carbon quantification. Geoderma 145 (1–2), 28–36.

Ben-Dor, E., Inbar, Y., Chen, Y., 1997. The reflectance spectra of organic matter in the visible near-infrared and short wave infrared region (400–2500) during a controlled decomposition process. Remote Sens. Environ. 61 (1), 1–15.

Ben-Dor, E., Patkin, K., Banin, A., Karnieli, A., 2002. Mapping of several soil properties using DAIS-7915 hyperspectral scanner data—a case study over clayey soils in Israel. Int. J. Remote Sens. 23 (6), 1043–1062.

Bowers, S.A., Hanks, R.J., 1965. Reflection of radiant energy from soils. Soil Sci. 100, 130–138.

Chang, C.W., Laird, D.A., 2002. Near-infrared reflectance spectroscopic analysis of soil C and N. Soil Sci. 167, 110–116.

Chang, C.W., Laird, D.W., Mausbach, M.J., Hurburgh, C.R., 2001. Near-infrared reflectance spectroscopy principal components regression analyses of soil properties. Soil Sci. Soc. Am. J. 6, 480–490.

Dalal, R.C., Henry, R.J., 1986. Simultaneous determination of moisture, organic carbon and total nitrogen by near infrared reflectance spectrophotometry. Soil Sci. Soc. Am. J. 50, 120–123.

Daniel, K.W., Tripathi, N.K., Honda, K., 2003. Artificial neural network analysis of laboratory and in situ spectra for the estimation of macronutrients in soils of Lopburi (Thailand). Aust. J. Soil Res. 41, 47–59.

Dematte, J.A.M., Galdos, M.V., Guimaraes, R.V., Genu, A.M., Nanni, M.R., Zullo-Junior, J., 2007. Quantification of tropical soil attributes from ETM+/LANDSAT-7 dat. Int. J. Remote Sens. 28 (17), 3813–3829.

Duckworth, J., 1998. Spectroscopic qualitative analysis. In: Workman Jr., J., Springsteen, A. (Eds.), Applied Spectroscopy. A Compact Reference for Practitioners. Academic Press, London, UK, pp. 93–163.

Dunn, B.W., Beecher, H.G., Batten, G.D., Ciavarella, S., 2002. The potential of nearinfrared reflectance spectroscopy for soil analysis-a case study from the riverine plain of south-eastern Australia. Aust. J. Exp. Agric. 42, 607–614.

Fuller, M.P., Griffiths, P.R., 1978. Diffuse reflectance measurements by infrared Fourier transform spectrometry. Anal. Chem. 50 (13), 1906–1910.

Gaffey, S.J., McFadden, L.A., Nash, D., Pieters, C.M., 1993. Ultraviolet, visible, and nearinfrared reflectance spectroscopy: laboratory spectra of geologic materials. In: Pieters, C.M., Englert, P.E. (Eds.), Remote Geochemical Analysis: Elemental and Mineralogical CompositionTopics in Remote Sensing Series 4. Cambridge University Press, Cambridge, United Kingdom, pp. 43–77.

Galvão, L.S., Vitorello, I., 1998. Variability of laboratory measured soil lines of soils from southeastern Brazil. Remote Sens. Environ. 63, 166–181.

Geladi, P., Kowalski, B.R., 1986. Partial least-squares regression: a tutorial. Anal. Chim. Acta 185, 1–17.

Gomez, C., Viscarra-Rossel, R.A., McBratney, A.B., 2008. Soil organic carbon prediction by hyperspectral remote sensing and field VIS–NIR spectroscopy: an Australian case study. Geoderma 146 (3–4), 403–411.

Hatchell, D.C., 1999. Analytical Spectral Devices, Inc. (ASD) Technical Guide. 3rd ed. Analytical Spectral Devices, Inc., Boulder.

He, T., Wang, J., Cheng, Y., Lin, Z.J., 2006. Study on spectral features of soil Fe₂O₃. Geogr. Inf. Sci. 22, 30–34 (in Chinese).

He, Y., Huang, M., Garcia, A., Hernandez, A., Song, H., 2007. Prediction of soil macronutrients content using near-infrared spectroscopy. Comput. Electron. Agric. 58 (2), 144–153.

Huang, Y., Liu, T., 1995. Spectral characteristics of main types of soils in South China and soil classification. Acta Pedol. Sin. 32 (1), 58–68.

Kooistra, L., Wanders, J., Epema, G.F., Leuven, R.S.E.W., Wehrens, R., Buydens, L.M.C., 2003. The potential of field spectroscopy for the assessment of sediment properties in river flood plains. Anal. Chim. Acta 484, 189–200.

Krishnan, P., Alexander, J.D., Butler, B.J., Hummel, J.W., 1980. Reflectance technique for predicting soil organic matter. Soil Sci. Am. J. 44, 1282–1285.

Liu, B.P., Qin, H.J., Luo, X., Cao, S.W., Wang, J.D., 2007a. Determination of four contents of feed stuff powder using near infrared spectroscopy by PLS-BP model. Spectrosc. Spectr. Anal. 27, 2005–2009 (in Chinese).

Liu, H.J., Zhang, B., Zhao, J., Zhang, X.Y., Song, K.S., Wang, Z.M., Duan, H.T., 2007b. Spectral models for prediction of organic matter in black soil. Acta Pedol. Sin. 44, 27–32 (in Chinese).

Lobell, D.B., Asner, G.P., 2002. Moisture effects on soil reflectance. Soil Sci. Soc. Am. J. 66, 722–727.

Lu, Y.L., Bai, Y.L., Yang, L.P., Wang, H.J., 2007. Prediction and validation of soil organic matter content based on hyperspectrum. Sci. Agric. Sin. 40 (9), 1989–1995.

McCarty, G.W., Reeves III, J.B., Reeves, V.B., Follett, R.F., Kimble, J.M., 2002. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. Soil Sci. Soc. Am. J. 66, 640–646.

Nduwamungu, C., Ziadi, N., Parent, L.E., Tremblay, G.F., Thuries, L., 2009. Opportunities for, and limitations of, near infrared reflectance spectroscopy applications in soil analysis: a review. Can. J. Soil Sci. 89 (5), 531–541.

Reeves III, J.B., McCarty, G.W., Meisinger, J.J., 2000. Near infrared reflectance spectroscopy for the determination of biological activity in agricultural soils. J. Near Infrared Spectrosc. 8, 161–170.

Schlerf, M., Atzberger, C., Joachim, H., Buddenbaum, H., Werner, W., Schuler, G., 2010. Retrieval of chlorophyll and nitrogen in Norway spruce (Picea abies L. Karst.) using imaging spectroscopy. Int. J. Appl. Earth Obs. Geoinf. 12, 17–26.

Sha, J.M., Chen, P.C., Chen, S.L., 2003. Characteristics analysis of soil spectrum response resulted from organic material. Res. Soil Water Conserv. 10, 21–24 (in Chinese).

Sorensen, L.K., Dalsgaard, S., 2005. Determination of clay and other soil properties by near infrared spectroscopy. Soil Sci. Soc. Am. J. l69, 159–167.

Stenberg, B., Viscarra-Rossel, R.A., Mouazen, A.M., Wetterlind, J., 2010. Visible and near infrared spectroscopy in soil science. Adv. Agron. 107, 163–215.

Sudduth, K.A., Hummel, J.W., 1993. Soil organic matter, CEC, and moisture sensing with a portable NIR spectrophotometer. Trans. Am. Soc. Agric. Eng. 36, 1571–1582.

Sullivan, D.G., Shaw, J.N., Richman, D., 2005. IKONOS imagery to estimate surface soil property variability in two Alabama physiographies. Soil Sci. Soc. Am. J. 69, 1789–1798.

Summers, D., Lewis, M., Ostendorf, B., Chittleborough, D., 2011. Visible near-infrared reflectance spectroscopy as a predictive indicator of soil properties. Ecol. Indic. 11 (1), 123–131.

The Committee of Agro-chemistry of the Chinese Society of Soil Science, 1983l. The Conventional Analysis Methods in Soil Agro-Chemistry. Science Press, Beijing.

Van-Waes, C., Mestdagh, I., Lootens, P., Carlier, L., 2005. Possibilities of near infrared reflectance spectroscopy for the prediction of organic carbon concentrations in grassland soils. J. Agric. Sci. Technol. 143, 487–492.

Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131, 59–75.

Xu, B.B., 2000. Study on spectral reflectance of soil section. Soil 6, 281–287 (in Chinese).

Xu, L.H., Xie, D.L., 2012. Prediction of total nitrogen and total phosphorus concentrations using hyper-spectral remote sensing. J. Agricult. Mech. Res. 2012 (4), 119–122.

Zhang, T.T., 2010. Partial Least Squares Modeling of Hyperspectral Remote Sensing for Mapping Agriculture Soil Properties. School of Earth Sciences, Jilin University, Jilin.

Zhang, J.J., Tian, Y.C., Yao, X., Cao, W.X., Ma, X.M., Zhu, Y., 2011. Estimating soil total nitrogen content based on hyperspectral analysis technology. J. Nat. Resour. 26 (5), 881–890.

Zhao, Y.S., 2003. The Principle and Method of Analysis of Remote Sensing. Application Science Press, Beijing, pp. 431–448 (chapter 14).

Zheng, G.H., 2010. Study on the Quantitative Estimating of Soil Properties with Hyper–Spectrum in Some Parts of Jiangsu Province. School of Geographic & Oceanographic Sciences, Nanjing University, Nanjing.

Zhu, X.C., Zhao, G.X., Dong, F., Wang, L., Lei, T., Zhan, B., 2009. Monitoring models for phosphorus content of apple flowers based on hyperspectrum. Chin. J. Appl. Ecol. 20 (10), 2424–2430 (in Chinese).

Zornoza, R., Guerrero, C., Mataix-Solera, J., Scow, K.M., Arcenegui, V., Mataix-Beneyto, J., 2008. Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. Soil Biol. Biochem. 40 (7), 1923–1930.