



## Original papers

# Assessment of important soil properties related to Chinese Soil Taxonomy based on vis–NIR reflectance spectroscopy



Dongyun Xu<sup>a</sup>, Wanzhu Ma<sup>b,\*</sup>, Songchao Chen<sup>c,d</sup>, Qingsong Jiang<sup>a,e</sup>, Kang He<sup>a</sup>, Zhou Shi<sup>a</sup>

<sup>a</sup> Institute of Applied Remote Sensing and Information Technology, College of Environmental and Resource Sciences, Zhejiang University, Hangzhou 310058, China

<sup>b</sup> Institute of Digital Agriculture, Zhejiang Academy of Agricultural Sciences, Hangzhou 310058, China

<sup>c</sup> INRA, Unité Infosol, 45075 Orléans, France

<sup>d</sup> UMR SAS, INRA, Agrocampus Ouest, 35000 Rennes, France

<sup>e</sup> College of Information Engineering, Tarim University, Alar 843300, China

## ARTICLE INFO

## Keywords:

vis–NIR spectroscopy  
Soil properties  
Chinese Soil Taxonomy  
PLSR  
Selectivity ratio

## ABSTRACT

As a rapid, inexpensive and accurate analysis technique, vis–NIR spectra has shown great advantages for determining a wide variety of soil properties, such as soil organic matter content, mineral composition, water content, particle size and color. Thus, this technique is becoming increasingly popular in soil science. We aim to assess the applicability of using vis–NIR spectra to estimate eighteen different soil properties that are important for Chinese Soil Taxonomy (CST). In this study, vis–NIR reflectance spectra were measured under laboratory conditions. First, partial least-squares regression (PLSR) was used to predict the following soil properties related to soil classification: coarse crumb, sand, silt, and clay contents, bulk density (BD), pH (H<sub>2</sub>O), pH (KCl), soil organic matter (SOM), total nitrogen (TN), total potassium (TK), and total phosphorus (TP) contents, cation exchange capacity (CEC), free iron (Fe<sub>2</sub>O<sub>3</sub>), soluble salts (salt), available phosphorus (AP), exchangeable aluminum (ExAl), aluminum saturation (AS) and base saturation (BS). Then, the important bands for modeling these soil properties were selected based on the selectivity ratio (SR). Finally, the spectral chromophores of the soil and the correlations of soil properties were analyzed. The results showed that (1) the prediction accuracy based on the PLSR algorithm was good for pH, SOM, TN, Fe<sub>2</sub>O<sub>3</sub>, salt, AS and BS (RPD > 2.0, R<sup>2</sup> between 0.70 and 0.90). For sand, silt, clay, BD, TP, TK, CEC, AP and ExAl, the PLSR model could achieve acceptable estimates (1.4 < RPD < 2.0, R<sup>2</sup> between 0.56 and 0.72), while for coarse crumb, the PLSR model was unable to make reliable predictions (RPD < 1.4, R<sup>2</sup> below 0.50). (2) As chromophore properties, SOM, TN, Fe<sub>2</sub>O<sub>3</sub>, clay and salt are and can be predicted by spectroscopy. Besides, BD, pH, TK, TP, CEC, AP, ExAl, AS and BS have significant correlations with chromophore properties and can also be predicted by vis–NIR spectroscopy. Therefore, except for coarse crumb, the soil properties important to CST can be quantitatively predicted by PLSR based on vis–NIR reflectance spectroscopy. This study is significant to CST, and it provides a fast and efficient method for soil classification.

## 1. Introduction

Soil classification is one of the signs of the development of soil science, the basis of soil survey mapping, as well as the language for communicating soil information. The Chinese Soil Taxonomy (CST) is a soil classification system based on diagnostic horizons and diagnostic characteristics (Gong et al., 2007). The diagnostic horizons and diagnostic characteristics were extracted from field surveys of soil properties and include chemical, physical and morphological parameters (Chinese Soil Taxonomy Research Group, Institute of Soil Science, Chinese Academy of Sciences, 2001). The soil properties are usually

described in detail by a professional pedologist and then validated and elaborated by laboratory wet chemistry analyses (Ben-Dor et al., 2008). However, wet chemistry analysis is time consuming and expensive and requires complicated sample treatment procedures, which are common problems for CST, as well as for the Brazilian and American systems of soil classification (Santos et al., 2006; Soil Survey Staff, 2003).

A variety of agricultural sensors have been applied to determine soil properties rapidly in recent decades (Gebber et al., 2010). Spectroscopy, in particular, has increased in popularity because it is rapid, timely, cost effective, nondestructive and straightforward (Li et al., 2015; Ji et al., 2016). The predictive capability of visible and

\* Corresponding author.

E-mail addresses: [xudongyun@zju.edu.cn](mailto:xudongyun@zju.edu.cn) (D. Xu), [xdy900815@126.com](mailto:xdy900815@126.com) (W. Ma), [songchao.chen@inra.fr](mailto:songchao.chen@inra.fr) (S. Chen), [qingsongjiang0827@126.com](mailto:qingsongjiang0827@126.com) (Q. Jiang), [21614116@zju.edu.cn](mailto:21614116@zju.edu.cn) (K. He), [shizhou@zju.edu.cn](mailto:shizhou@zju.edu.cn) (Z. Shi).

<https://doi.org/10.1016/j.compag.2017.11.029>

Received 16 August 2017; Received in revised form 25 October 2017; Accepted 21 November 2017  
0168-1699/ © 2017 Elsevier B.V. All rights reserved.

near-infrared (vis–NIR) spectroscopy for soil organic matter, clay minerals, soil texture, soil moisture and other soil properties has been fully demonstrated by many studies (Kuang et al., 2012; Stenberg et al., 2010; Xia et al., 2015). In addition, vis–NIR spectroscopy can predict several soil properties simultaneously. Early in 1995, Ben-Dor examined the capability of NIR spectroscopy for predicting six soil properties and concluded that NIR spectroscopy is a promising method for rapid and nondestructive analysis of soil properties (Ben-Dor and Banin, 1995). Chang et al., evaluated the ability of NIR spectroscopy to predict more than thirty soil properties and found that NIR spectroscopy can be used to rapidly estimate several soil properties simultaneously with acceptable accuracy (Chang et al., 2001). Cohen studied twenty-two soil properties of wetlands using vis–NIR reflectance spectroscopy (Cohen et al., 2005). vis–NIR reflectance spectroscopy have the potential to simultaneously predict various soil properties related to soil quality (Viscarra Rossel et al., 2006). Generally, vis–NIR spectroscopy is seen as an effective alternative to traditional chemical analysis, and it can be used to predict several soil properties at the same time (Bilgili et al., 2010; Viscarra Rossel and Webster, 2012; Ji et al., 2014). For soil classification, Dematte et al. (2004) evaluated soil types using vis–NIR reflectance spectroscopy and obtained favorable results for a number of soil classes. Mouazen et al. (2005) employed vis–NIR reflectance spectroscopy to discriminate soil texture classes. Later, Dematte et al. (2014) evaluated the potential of the vis–NIR spectral sensing to detect soil differences along a toposequence and showed a new perspective for pedological evaluation.

Many studies have been conducted on the use of vis–NIR spectroscopy to evaluate soil attributes. However, among these soil properties, few studies have focused on CST. Thus, the objective of this study was to evaluate the feasibility of vis–NIR spectroscopy to predict the soil properties that are essential to CST. We did so by (i) predicting eighteen soil properties by linear regression of the PLSR algorithm, (ii) analyzing the important characteristic bands of soil properties using the selectivity ratio (SR), and (iii) analyzing the soil spectral chromophores and correlations among soil properties to study whether these properties could be predicted by vis–NIR spectroscopy.

## 2. Materials and methods

### 2.1. Soil sampling and chemical analysis

Soil samples were collected from Zhejiang Province in the southeast of China, with latitudes ranging from 27°2' N to 31°5' N and longitudes from 118° E to 123° E (Fig. 1). In total, 146 sampling sites were selected by professional pedologists. At each sampling site, soil samples were collected from genetic soil horizons. A total of 591 soil samples were collected for analysis. The soil samples were air-dried, ground and sieved to less than 2 mm in the laboratory for further analysis. The determination methods for the physical and chemical soil properties are listed in Table 1.

### 2.2. Spectra measurements and data preprocessing

The spectra of soil samples were measured using a Fieldspec® ProFR vis–NIR spectrometer (Analytical Spectral Devices, Boulder, CO, USA). The spectral range is from 350 to 2500 nm. The instrument has a spectral resolution of 3 nm between 350 and 1000 nm and 10 nm between 1000 and 2500 nm. The sampling resolution of the spectra is 1 nm. Before each measurement, a Spectralon® panel with 99% reflectance was used to calibrate the spectrometer. The soil sample was spread on a culture dish (5 cm in diameter and 1 cm high) and flattened before scanning. Stones, roots, and voids were avoided. Then, the soil samples were measured using a high-intensity contact probe (Analytical Spectral Devices, Boulder, CO, USA) with its own light source. Three measurements were made at different random positions in the dish. At each measurement, the instrument made 10 internal scans to obtain a

satisfactory signal-to-noise ratio. A total of 30 spectra were averaged into one spectrum for each sample.

To eliminate the influence of noise, the spectral regions of 350–399 nm and 2451–2500 nm were removed. The reflectance spectra (R) were transformed to apparent absorbance ( $\log 1/R$ ). To further reduce noise and to enhance the signals, we used the Savitzky–Golay algorithm (Savitzky and Golay, 1964), with a window size of 11 and a polynomial of order 2. And then mean-centered before the multivariate modeling.

We used Rank–KS algorithm (Chen et al., 2016) to compile training data. First, the data were arranged in ascending order, according to soil properties, and then Kennard–Stone (KS) algorithm (Kennard and Stone, 1969) was applied for each n subsets of data. We tried n = 3, 6, 9, 12, and 15 and found that the best training data was obtained with n = 6. Additionally, the training data represented two-thirds of the whole data set.

### 2.3. Important bands selection and analysis

To find the important bands for prediction models, a variable selection method, SR, was used in this study. The selectivity ratio (SR) is defined as the ratio between the explained variance of each variable and the residual variance. A high value denotes a variable with good predictive performance (Rajalahti et al., 2009).

For further analysis of the spectroscopy, we studied the chromophore properties. A chromophore is a parameter or substance (chemical or physical) that significantly affects the shape and nature of a soil spectrum (Ben-Dor et al., 2008). The absorption features of the reflectance spectrum are attributed to specific chemical groups in various structural configurations, including overtone, combination modes, and electronic processes.

### 2.4. Model construction and evaluation

Partial least-squares regression (PLSR) (Wold et al., 1983) is one of the most popular algorithms used for spectral calibration and prediction among the multiple linear calibration algorithms. It has the advantages of eliminating the multiple mutual linear problems of the independent variables (Næsset et al., 2005; Wang et al., 2006).

Leave-one-out cross-validation was used to select the optimal PLSR calibration model, and twenty factors were tested in this study. To select the best latent variable number, two aspects should be considered: the RMSE should be small and the model should have the fewest number of factors (Viscarra Rossel et al., 2006).

The coefficient of determination ( $R^2$ ), root mean square error (RMSE) and ratio of prediction derivation (RPD) were used to compare the prediction accuracies.

$$R^2 = \frac{[\sum_{i=1}^N (\hat{y}_i - \bar{y}) (y_i - \bar{y})]^2}{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2 \sum_{i=1}^N (y_i - \bar{y})^2} \quad (1)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (2)$$

$$\text{RPD} = \text{SD} / \text{RMSE} \quad (3)$$

where  $y_i$  is the  $i$  th observed value and  $\hat{y}_i$  is the corresponding prediction;  $\bar{y}$  is the mean of observed values and  $\bar{\hat{y}}$  is the mean of the predictions, SD is the standard deviation of observed values, and N is the number of observations. Generally, the larger  $R^2$  and RPD are and the smaller the RMSE is, the better the model prediction.

According to Chang et al. (2001), the prediction ability for soil properties can be divided into three categories based on the RPD value. If  $\text{RPD} > 2.0$ , the model is Grade A, which indicates that the prediction is excellent. If the RPD is between 1.4 and 2.0, the model is Grade B, suggesting the model can achieve approximate quantitative predictions. If  $\text{RPD} < 1.4$ , the model is Grade C, indicating the model is very poor

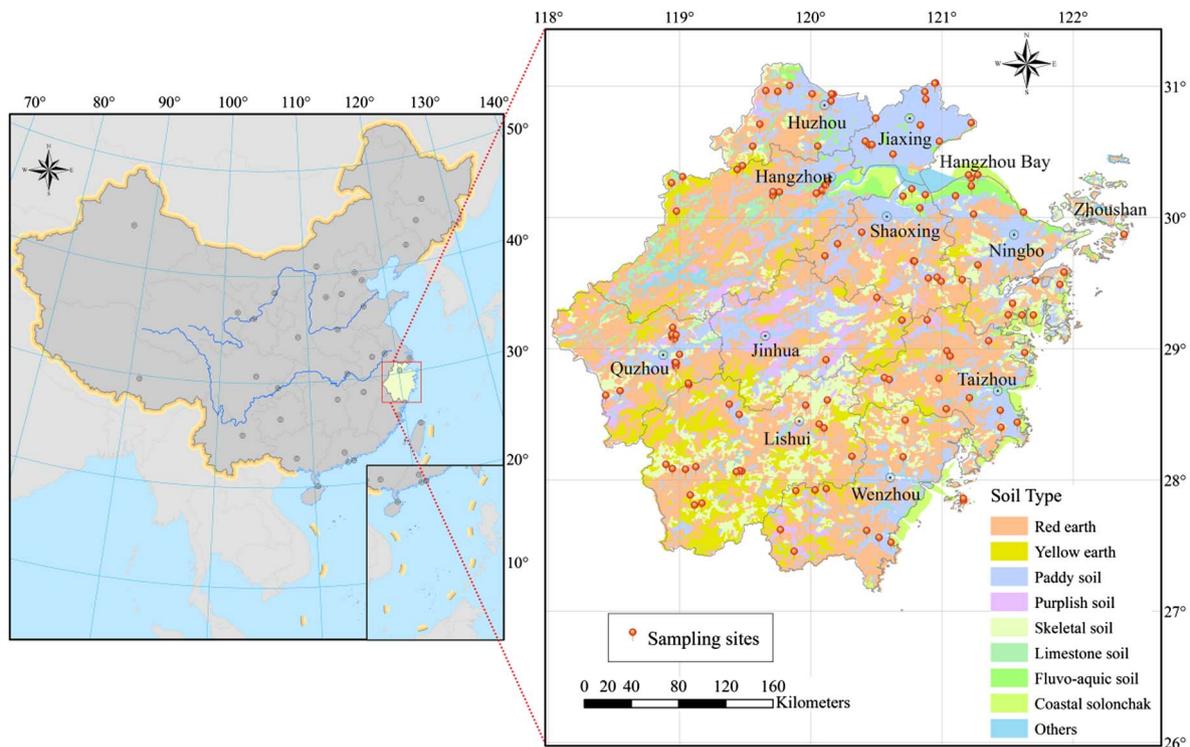


Fig. 1. Maps showing the location of the study site, the soil type of the study site and the positions of the sampling sites in this study.

Table 1  
Soil properties measured and analytical methods.

Soil property	Method
Particle-size distribution	Pipette method
BD	Cutting ring method
pH(H <sub>2</sub> O)	pH measured in 1:1 soil/water suspension, potential method
pH(KCl)	pH measured in 1:2 soil/KCl suspension, potential method
SOM	H <sub>2</sub> SO <sub>4</sub> –K <sub>2</sub> Cr <sub>2</sub> O <sub>7</sub> oxidation method
TN	Se–CuSO <sub>4</sub> –H <sub>2</sub> SO <sub>4</sub> digestion method–distillation
TP	NaOH–Mo–Sb colorimetric method
TK	NaOH–flame photometric method
CEC	CH <sub>3</sub> COONH <sub>4</sub> –EDTA exchange capacity at pH 7
Fe <sub>2</sub> O <sub>3</sub>	Sodium citrate–sodium hyposulfite–DCB extraction
Salt	Measurement of electrical conductivity in 1:5 soil:water extracts
AP	NaHCO <sub>3</sub> extraction–Mo–Sb colorimetric method NH <sub>4</sub> F and HCl extraction–Mo–Sb colorimetric method (acidic soil)
ExAl	KCl extraction
AS	ExAl/(ExCa + ExMg + ExK + ExAl)
BS	HCl exchange–neutralization titration method

and useless.

Mathematical pre-processing and chemometric analyses were carried out using R software (R Core Team, 2014), MatLab version 7 (The MathWorks Inc., Natick, MA, USA), and PLS\_Toolbox8.5 (Eigenvector Research Inc., Wenatchee, WA, USA).

### 3. Results

#### 3.1. Data preprocessing

Eighteen soil properties were determined by the methods described in Table 1. Some properties had strongly positively skewed distributions, and to stabilize their variances for the spectroscopic modeling, we

Table 2  
Statistical data of soil properties.<sup>a</sup>

Soil property	N	Max.	Min.	Mean	Med.	Std dev.	Skew.
Coarse crumb (%)	461	87.00	1.00	14.70	10.00	14.63	1.77
Log <sub>10</sub> (coarse crumb)		1.94	0.00	0.95	1.00	0.47	–0.33
Sand (%)	588	950.41	46.61	392.92	366.00	187.93	0.57
Sqrt (sand)		30.83	6.83	19.23	19.13	4.80	0.04
Silt (%)	588	775.82	27.40	431.60	452.73	140.29	–0.47
Clay (%)	588	666.06	9.57	175.48	159.35	102.55	1.09
Sqrt (clay)		25.81	3.09	12.67	12.62	3.88	0.14
BD (g cm <sup>–3</sup> )	588	1.73	0.25	1.23	1.22	0.17	–0.15
pH (H <sub>2</sub> O)	588	9.60	3.30	6.27	5.90	1.50	0.49
Log <sub>10</sub> [pH(H <sub>2</sub> O)]		0.98	0.52	0.79	0.77	0.10	0.16
pH (KCl)	588	8.20	3.10	5.13	4.70	1.39	0.61
Log <sub>10</sub> [pH(KCl)]		0.91	0.49	0.70	0.67	0.11	0.32
SOM (g kg <sup>–1</sup> )	588	141.72	0.82	16.60	10.67	16.84	2.82
Log <sub>10</sub> (SOM)		2.15	–0.09	1.05	1.03	0.38	0.10
TN (g kg <sup>–1</sup> )	591	6.70	0.01	0.88	0.61	0.79	2.30
Log <sub>10</sub> (TN)		0.83	–2.00	–0.20	–0.21	0.37	–0.24
TP (g kg <sup>–1</sup> )	590	7.49	0.01	0.51	0.43	0.57	7.88
Log <sub>10</sub> (TP)		0.87	–2.00	–0.41	–0.37	0.31	–0.14
TK (g kg <sup>–1</sup> )	591	46.00	2.70	18.64	18.20	6.94	0.63
CEC [cmol(+) kg <sup>–1</sup> ]	591	40.50	1.05	12.23	11.21	5.81	1.21
Sqrt (CEC)		6.36	1.02	3.40	3.35	0.80	0.40
Fe <sub>2</sub> O <sub>3</sub> (g kg <sup>–1</sup> )	588	79.70	1.80	22.57	20.00	13.07	1.37
Log <sub>10</sub> (Fe <sub>2</sub> O <sub>3</sub> )		1.90	0.26	1.28	1.30	0.25	–0.38
Salt (g kg <sup>–1</sup> )	343	17.47	0.00	0.63	0.14	1.93	6.53
Log <sub>10</sub> (salt)		1.24	–2.39	–0.78	–0.85	0.64	0.46
AP (mg kg <sup>–1</sup> )	281	124.89	0.06	13.99	4.72	22.74	2.72
Log <sub>10</sub> (AP)		2.10	–1.25	0.65	0.67	0.73	–0.26
ExAl [cmol(+) kg <sup>–1</sup> ]	419	13.70	0.00	2.04	0.68	2.49	1.42
Log <sub>10</sub> (ExAl)		1.14	–2.33	–0.19	–0.17	0.78	–0.26
AS (%)	420	107.00	0.02	26.08	9.57	29.53	0.87
Sqrt (AS)		10.34	0.15	4.08	3.09	3.07	0.39
BS (%)	439	99.85	5.42	62.99	74.07	33.02	–0.33

<sup>a</sup> Note: N = sample size, max. = maximum, min. = minimum, med. = median, std dev. = standard deviation, skew. = skewness.

**Table 3**  
Prediction accuracy based on PLSR.<sup>b</sup>

Soil properties	n	LVs	R <sup>2</sup>	RMSE	RPD	Grade
Log <sub>10</sub> (coarse crumb)	154	8	0.33	0.39	1.23	C
Sqrt (sand)	196	19	0.67	2.80	1.72	B
Silt	196	14	0.70	78.44	1.90	B
Sqrt (clay)	196	11	0.58	2.48	1.55	B
BD	196	17	0.63	0.11	1.47	B
Log <sub>10</sub> [pH(H <sub>2</sub> O)]	196	16	0.86	0.04	2.62	A
Log <sub>10</sub> [pH (KCl)]	196	20	0.82	0.05	2.34	A
Log <sub>10</sub> (SOM)	196	20	0.80	0.18	2.18	A
Log <sub>10</sub> (TN)	197	17	0.80	0.16	2.25	A
Log <sub>10</sub> (TP)	196	16	0.61	0.19	1.60	B
TK	197	17	0.72	3.80	1.68	B
Sqrt (CEC)	197	11	0.72	0.43	1.90	B
Log <sub>10</sub> (Fe <sub>2</sub> O <sub>3</sub> )	196	17	0.81	0.11	2.24	A
Log <sub>10</sub> (salt)	115	10	0.82	0.29	2.31	A
Log <sub>10</sub> (AP)	94	14	0.56	0.50	1.51	B
Log <sub>10</sub> (ExAl)	140	20	0.59	0.51	1.55	B
Sqrt (AS)	140	19	0.79	13.56	2.18	A
BS	146	20	0.78	15.40	2.15	A

<sup>b</sup> Note: n means the number of validation dataset.

transformed the data to approximate normal distributions by square root or logarithmic transformations. A statistical description of the original and transformed soil attributes is given in Table 2. After transformation, the eighteen soil properties exhibited clear normal distributions.

### 3.2. Prediction of soil properties using PLSR

The prediction accuracies for the eighteen soil properties using PLSR and the associated accuracy classes are shown in Table 3 and Fig. 2. Among these measured soil properties, pH, SOM, TN, Fe<sub>2</sub>O<sub>3</sub>, salt, AS and BS were the best predicted (RPD > 2.0 and R<sup>2</sup> were between 0.78 and 0.86, Grade A). Sand, silt, clay, BD, TP, TK, CEC, AP and ExAl were also estimated reasonably well but with less accuracy (RPD = 1.47–1.90, R<sup>2</sup> = 0.56–0.72, Grade B). The remaining property, coarse crumb, was classified as Grade C (RPD < 1.40, R<sup>2</sup> < 0.50), indicating that coarse crumb was poor and not useful.

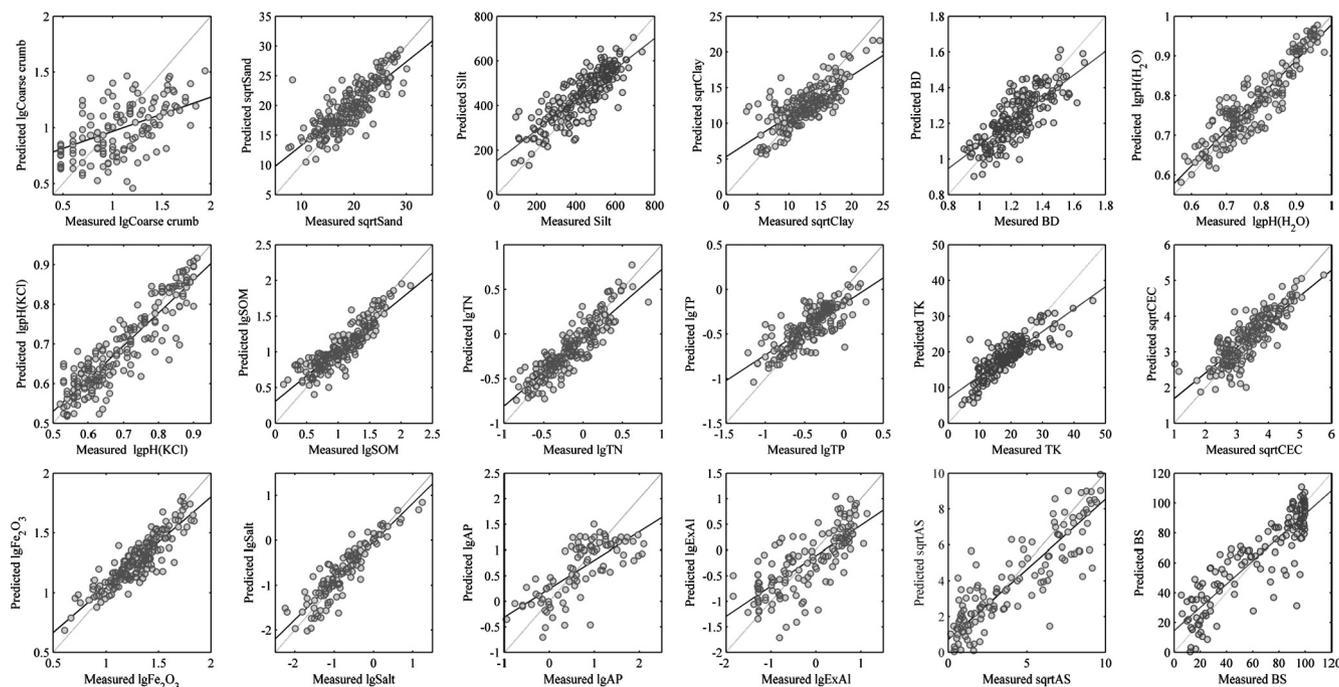


Fig. 2. PLSR models of each soil property.

### 3.3. Important bands for soil properties in PLSR models

The values of SR are calculated for individual variables and can therefore be plotted against the spectral data, as illustrated in Fig. 3.

#### 3.3.1. Physical soil properties

In this study, physical soil properties included granular structure, sand, silt, clay and BD. For coarse crumb, high SR values were mainly in the spectral regions of 720 nm and 1900–2400 nm (Fig. 3). For sand, there were small peaks and valleys of SR at 480, 920, 1910 and 2200 nm, indicating these bands may be used to explain the sand property. For silt, bands near 780 and 1080 nm had high SR values, with some small peaks near 2200 and 2390 nm. For clay, the SR values were extremely high, especially for the bands at 410–580 nm, 1410, 1900, 2200 and 2400 nm. For BD, high SR values were located in the bands at 580, 810, 1400, 1950, 2180 and 2280 nm. For physical soil properties, mainly related to soil texture, the significant bands included 480, 580, 1400–1410 nm, 1900–1950 nm, 2200 and 2350 nm in this study.

#### 3.3.2. Chemical soil properties

The SR values of SOM and TN had similar characteristics (Fig. 3). SOM is affected by the particle size distribution and specific chemical groups. The highest SR values for SOM were in the visible region of 580–800 nm, with small valleys in the bands of 1400, 1900 and 2200 nm. As a result, the significant bands of SOM and TN include 580–800 nm, 1400, 1900 and 2200 nm. Fig. 3 shows that the SR values of pH and clay exhibited similar patterns in the regions of 400–600 nm and above 1900 nm. Furthermore, high SR values for pH were located in the bands of 480, 780, 1120, 1910, 2200 and 2390 nm. Thus, these bands can be considered significant bands for pH. For TP and AP, the high SR scores were located at bands of 870, 1020, 1410, 1870, 2170 and 2230 nm. For TK, the high SR values were between 400 and 480 nm, and there were small peaks at 1870 and 2190 nm. As shown in Fig. 4, the high SR values of CEC were located in the bands at 680, 890, 1410, 1900, 2210 and 2400 nm. The SR values of Fe<sub>2</sub>O<sub>3</sub> were high in the visible region of 400–480 nm, and there were peaks at 900, 1400, 1900, 2200 and 2390 nm. Thus, the 400–480 nm, 1900 and 2200 nm bands can be considered characteristic in this study. For the salt

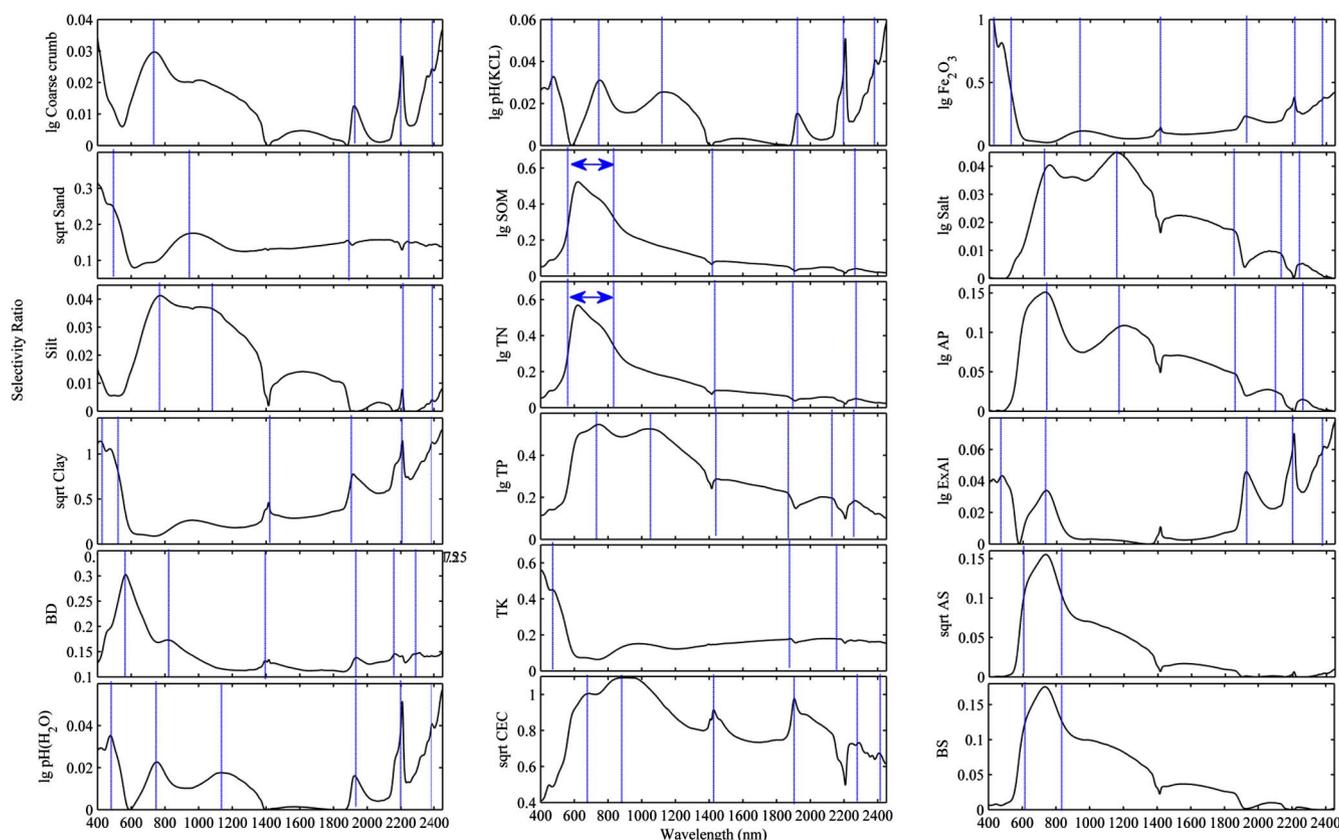


Fig. 3. Selectivity ratio of each soil property.

content, the high SR values were mainly located at 710–1190 nm, with small peaks at 1820, 2120 and 2230 nm. ExAl had the same trend as pH in terms of SR. For AS and BS, the SR values were almost the same, and the high SR values were located in the region of 600–800 nm. There were also small valleys at 1400 and 1900 nm, and small peaks at 2200 nm.

### 3.4. Spectral chromophores of soil

The chromophores in soil include minerals (clay, iron oxide, salt, and primary minerals, such as feldspar and carbonates), organic matter and water. Some of the functional groups and mechanisms acting as soil chromophores are shown in Fig. 4. In the visible region (400–780 nm) and at 880 nm, the functional groups are mainly related to iron oxide; the functional groups at approximately 825, 930, 1100, 1500, 2000 and 2200 nm are mainly related to SOM and TN; and the functional groups at approximately 1395, 1415, 2160 and 2300 nm are mainly related to clay minerals. Other functional groups are mainly related to soil water.

The spectral features of soil properties in the vis–NIR spectral region are related to the chromophores and the vibration modes of the functional groups. These soil properties, which are related to chromophores, including clay, SOM, TN,  $\text{Fe}_2\text{O}_3$  and salt in this instance, are able to be predicted by spectroscopy. However, not all properties have distinct spectral features. However, Ben-Dor and Banin (1995) noted that indirect spectral assignments may provide reasonable predictions for soil properties that do not have clear spectral features.

To identify these indirect spectral assignments, linear correlations between soil properties were analyzed (Fig. 5). Coarse crumb was significantly correlated with sand, silt, clay, pH, TK, CEC, AP and AS. However, its correlation with chromophore properties such as SOM, TN,  $\text{Fe}_2\text{O}_3$  and salt were low, possibly explaining the poor spectroscopic modeling accuracy. The correlations of sand and silt with clay, SOM, TN,  $\text{Fe}_2\text{O}_3$  and salt were significant ( $p < .01$ ). A high correlation also

existed between sand and silt; hence, the SR behavior of silt was similar to that of sand, as shown in Fig. 3. For BD, the correlations with clay, SOM and TN were significant ( $p < .01$ ), while the correlations with  $\text{Fe}_2\text{O}_3$  and salts were not significant. The correlations between pH and other soil properties were almost all significant. Thus, the prediction of pH by spectroscopy achieved good results. The correlation of TN with SOM was extremely high, up to 0.933\*\*, possibly explaining the similarity in SR values between TN and SOM. TP was positively correlated with SOM and TN at the 0.01 significance level but was correlated with  $\text{Fe}_2\text{O}_3$  and salt at only the 0.05 significance level. TK and AP had significant negative correlations with clay and  $\text{Fe}_2\text{O}_3$  ( $p < .01$ ), and their correlations with SOM were only at the 0.05 level. CEC had significant positive correlations with clay, SOM, TN and  $\text{Fe}_2\text{O}_3$  at the 0.01 level. ExAl, AS and BS all had significant correlations with clay,  $\text{Fe}_2\text{O}_3$  and salt at the 0.01 level. A significant correlation was present between ExAl and pH, and this could explain their similar SR values. These results agree with the results of Table 3 and indicate that soil properties that have significant correlations with properties related to chromophores can be predicted by spectroscopy, while properties that have little correlation with chromophore-related properties may not be able to be predicted by spectroscopy.

## 4. Discussion

The results show that vis–NIR spectroscopy can be used to predict most soil properties that are important for CST. Among these soil properties, pH, SOM, TN,  $\text{Fe}_2\text{O}_3$  and salt were the best predicted, with  $R^2 > 0.78$  and  $\text{RPD} > 2.0$ . The successful predictions of SOM and TN are mainly because carbon and nitrogen have direct spectral responses due to the overtones and combinations of N–H, C–H + C–H and C–H + C–C in the vis–NIR spectra. This phenomenon can also be seen in Fig. 3, where the high SR values of SOM are in the visible region of 580–800 nm, which is mainly affected by iron oxides and the color of

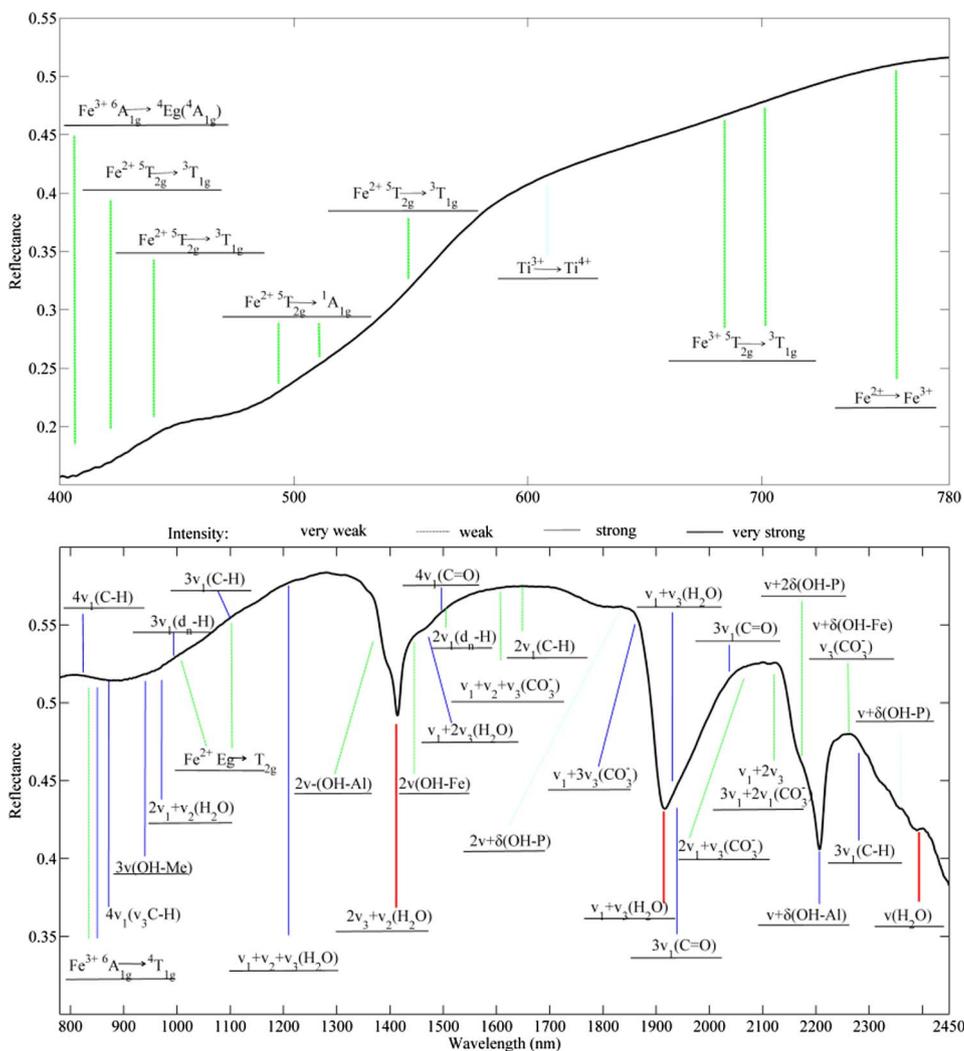


Fig. 4. Functional groups and mechanism in the soil chromophores (referred from Ben-Dor et al., 2008; Stenberg et al., 2010).

the SOM itself. Small peaks in the bands are also present around 1400, 1900, and 2200 nm, which are mainly affected by the functional groups of N–H, C–H, C–O and C=O (Stenberg et al., 2010). The significant bands for SOM in this study are identical to the findings of Shi et al. (2015) and Ji et al. (2016). The pH of soil is regulated by a variety of factors. Although there are no direct spectral responses to pH in the vis–NIR spectra, the prediction accuracy for pH is extremely high, which may be due to its relationship with the wavelengths of minerals (Viscarra Rossel and Behrens, 2010; Ji et al., 2014). In this study, the correlations between pH and other soil properties are significant (Fig. 5). Additionally, the SR graphs of pH and clay were similar in some regions. This similarity is inconsistent with the results of previous studies in which the spectra of pH were reportedly very similar to those of clay and CEC (Islam et al., 2003; Pirie et al., 2005). The electron transition of goethite, hematite and iron oxides in the vis–NIR spectra contributes to the successful predictions of  $\text{Fe}_2\text{O}_3$  (Kuang et al., 2012; Stenberg et al., 2010). The predictions of SOM,  $\text{Fe}_2\text{O}_3$ , pH and TN are comparable to previous studies (Shi et al., 2014; Xie et al., 2012; Abdi et al., 2012; Bilgili et al., 2010).

Sand, silt and clay are important attributes for soil texture. Among these factors, the clay content draws more attention because it has a large influence on the soil structure by promoting the formation of soil aggregates, and its swelling and shrinking properties can form cracks (Stenberg et al., 2010). Clay content is related to OH in water and  $\text{Mg}^+$ ,  $\text{Al}^+$ , and  $\text{Fe-OH}$  in the crystal lattice of minerals (Ben-Dor and Banin, 1995). The influence of mineralogy on vis–NIR spectra can be assumed to be a valuable feature for the prediction of the clay content because

clay particles mainly consist of clay minerals (Stenberg et al., 2010). For example, smectite has strong characteristic absorptions near 1400 nm, 1900 nm, and 2200 nm. Thus, these bands play an important role in predicting the clay content (Fig. 3). The prediction results for sand, silt and clay were compared to the results of Chang et al. (2001). BD was found to be significantly related to clay, SOM and TN ( $p < .01$ ); hence, it could be moderately well predicted, with an  $R^2$  of 0.63 and RPD of 1.47. For physical soil properties, mainly related to soil texture, the significant bands include 480, 580, 1400–1410 nm, 1900–1950 nm, 2200 and 2350 nm in this study, which are in accord with the results of Huang et al. (2009).

The CEC is important for the buffering capacity of soil and is related to the clay fraction and organic matter content (Stenberg et al., 2010). Additionally, the CEC is more directly related to the mineralogy and soil particle size distribution than to the clay content. Thus, the CEC can be well predicted by vis–NIR spectroscopy (Chang et al., 2001; Shepherd and Walsh, 2002; Brown et al., 2006; Viscarra Rossel et al., 2016). The predictions of TP and TK were moderately useful in this study. Abdi et al. (2012) reported similar prediction performances for soil TK and TP. AP was moderately well predicted, with an  $R^2$  of 0.56 and an RPD of 1.51 (Grade C), and its successful prediction may be due to the covariation with other soil properties that have direct spectral responses in the vis–NIR range (Stenberg et al., 2010). The prediction results for AS and BS were similar, and both achieved good results, with  $R^2$  values of 0.79 and 0.78, respectively, and RPD values of 2.18 and 2.15, respectively. The good results may be due to their significant relationships with clay, salt ( $P < .01$ ) and  $\text{Fe}_2\text{O}_3$  ( $P < .05$ ).

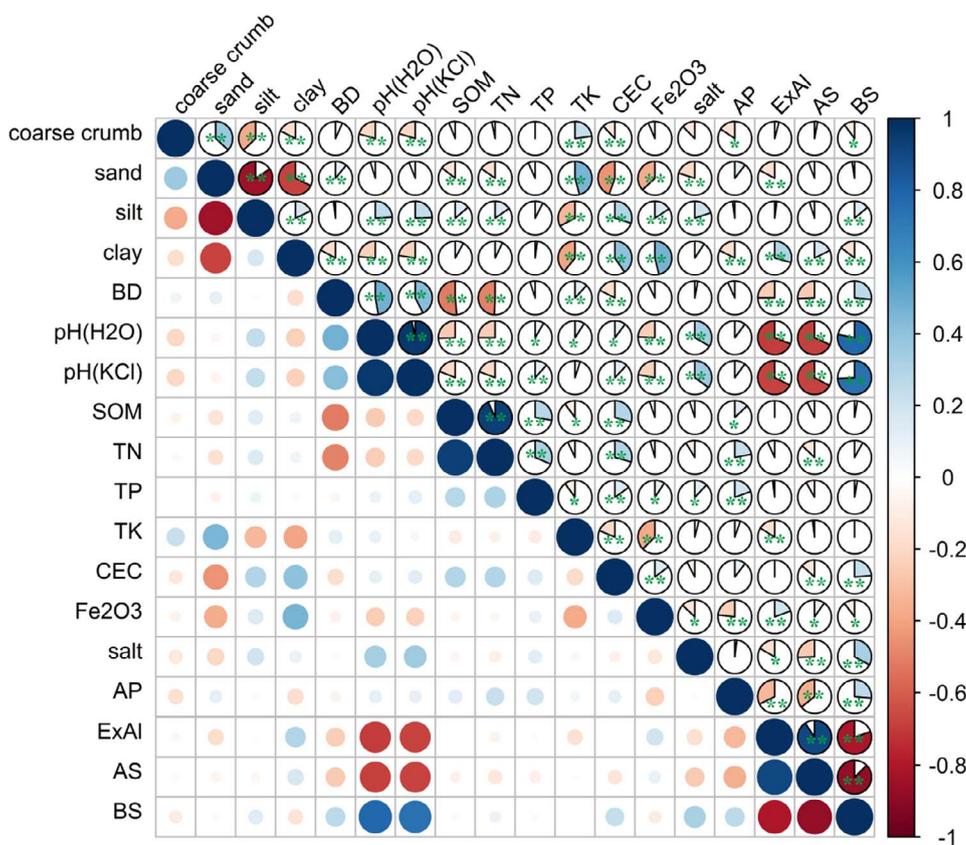


Fig. 5. Correlation between soil properties (\*\* indicates the correlation is significant at the 0.01 level, \* indicates the correlation is significant at the 0.05 level).

The successful prediction of these soil properties is significant for CST, as some of these properties are essential for determining diagnostic horizons and diagnostic characteristics. Additionally, this technique provides a reliable, efficient, rapid, nondestructive classification method for CST.

## 5. Conclusions

The prediction accuracy based on the PLSR algorithm was good for pH, SOM, TN, Fe<sub>2</sub>O<sub>3</sub>, salt, AS and BS (Grade A); moderate for sand, silt, clay, BD, TP, TK, CEC, AP and ExAl (Grade B); and poor for coarse crumb.

The SR graphs were generated to identify the significant bands for various soil properties. As shown in this study, the bands at approximately 1400, 1900 and 2200 nm were important for all properties. In addition to these regions, bands at approximately 480 nm were significant for sand, clay, pH, TK and ExAl. The bands at approximately 2400 nm were significant for granular structure, silt, clay, pH, CEC, Fe<sub>2</sub>O<sub>3</sub> and ExAl. The bands at approximately 580 nm were significant for BD. The significant bands for SOM and TN also included the 580–800 nm range. The bands at approximately 900 nm were considered to be characteristic for Fe<sub>2</sub>O<sub>3</sub>. The important bands for soluble salts were mainly located at 600–1400 nm. The important bands for AS and BS were located in the region of 600–800 nm.

The SOM, TN, Fe<sub>2</sub>O<sub>3</sub>, clay and soluble salts are chromophore properties, and these properties can be predicted by spectroscopy with the proper algorithms. Although sand, silt, BD, pH, TP, TK, CEC, AP, ExAl, AS and BS are not chromophore properties, they have significant correlations with chromophore properties ( $p < .01$  or  $p < .05$ ). Thus, these properties can also be predicted with the proper algorithms based on spectroscopy. However, coarse crumb is not a chromophore property and its correlations with chromophore properties are low, which may explain the poor prediction. The ability to accurately predict important soil properties related to CST indicates that vis-NIR spectroscopy

represents a reliable, efficient, rapid, nondestructive classification method for CST.

## Acknowledgments

We thank the support of the National Key Research and Development Program (2017YFD0700501).

## Conflict of interest

The authors declare no conflict of interest.

## Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.compag.2017.11.029>.

## References

- Abdi, D., Tremblay, G.F., Ziadi, N., Bélanger, G., Parent, L.-É., 2012. Predicting soil phosphorus-related properties using near-infrared reflectance spectroscopy. *Soil Sci. Soc. Am. J.* 76, 2318.
- Ben-Dor, E., Taylor, R.G., Hill, J., Dematté, J.A.M., Whiting, M.L., Chabrilat, S., Sommer, S., 2008. Imaging spectrometry for soil applications. *Adv. Agron* 97, 321–392.
- Ben-Dor, E., Banin, A., 1995. Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. *Soil Sci. Soc. Am. J.* 59, 364–372.
- Bilgili, A.V., van Es, H.M., Akbas, F., Durak, A., Hively, W.D., 2010. Visible-near infrared reflectance spectroscopy for assessment of soil properties in a semi-arid area of Turkey. *J. Arid Environ.* 74, 229–238.
- Brown, D.J., Shepherd, K.D., Walsh, M.G., Dwayne Mays, M., Reinsch, T.G., 2006. Global soil characterization with VNIR diffuse reflectance spectroscopy. *Geoderma* 132, 273–290.
- Chang, C.W., Laird, D.A., Mausbach, M.J., Hurburgh, C.R., 2001. Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. *Soil Sci. Soc. Am. J.* 65, 480–490.
- Chen, S.C., Peng, J., Ji, W.J., Zhou, Y., He, J.X., Shi, Z., 2016. Study on the characterization of VNIR-MIR spectra and prediction of soil organic matter in paddy soil. *Spectrosc. Spect. Anal.* 36, 1712–1716.
- Chinese Soil Taxonomy Research Group, Institute of Soil science, Chinese Academy of

- Sciences, Cooperative Research Group on Chinese Soil Taxonomy, 2001. Keys to Chinese Soil Taxonomy, third ed. Chinese Science and Technological University Press, Hefei.
- Cohen, M.J., Prenger, J.P., DeBusk, W.F., 2005. Visible-near infrared reflectance spectroscopy for rapid, nondestructive assessment of wetland soil quality. *J. Environ. Qual.* 34, 1422–1434.
- Dematté, J.A.M., da Silva Terra, F., 2014. Spectral pedology: a new perspective on evaluation of soils along pedogenetic alterations. *Geoderma* 217–218, 190–200.
- Dematte, J.A.M., Campos, R.C., Alves, M.C., Fiorio, P.R., Nanni, M.R., 2004. Visible-NIR reflectance: a new approach on soil evaluation. *Geoderma* 121, 95–112.
- Gebbers, R., Adamchuk, V.I., 2010. Precision agriculture and food security. *Science* 327, 828–831.
- Gong, Z.T., Zhang, G.L., Chen, Z.C., 2007. Pedogenesis and Soil Taxonomy. Science Press, Beijing.
- Huang, M.X., Cheng, X.L., Wang, K., Gong, J.H., Li, H.Y., Shi, Z., 2009. Coastal soil hyperspectral characteristics and soil sand content prediction. *Acta Pedol. Sin.* 5, 181–186.
- Islam, K., Singh, B., McBratney, A., 2003. Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared reflectance spectroscopy. *Aust. J. Soil Res.* 41, 1101–1114.
- Ji, W.J., Li, S., Chen, S.C., Shi, Z., Viscarra Rossel, R.A., Mouazen, A.M., 2016. Prediction of soil attributes using the Chinese soil spectral library and standardized spectra recorded at field conditions. *Soil Till. Res.* 155, 492–500.
- Ji, W.J., Shi, Z., Huang, J.Y., Li, S., 2014. In situ measurement of some soil properties in paddy soil using visible and near-infrared spectroscopy. *PLoS ONE*.
- Kennard, R.W., Stone, L.A., 1969. Computer aided design of experiments. *Technometrics* 11, 137–148.
- Kuang, B., Mahmood, H.S., Quraishi, M.Z., Hoogmoed, W.B., Mouazen, A.M., van Henten, E.J., 2012. Sensing Soil Properties in the Laboratory, in Situ, and on-line: a Review. In: Sparks, D.L., (Ed.), *Adv. Agron.* vol. 114. pp. 155–223.
- Li, S., Shi, Z., Chen, S.C., Ji, W.J., Zhou, L.Q., Yu, W., Webster, R., 2015. In situ measurements of organic carbon in soil profiles using Vis-NIR spectroscopy on the Qinghai-Tibet plateau. *Environ. Sci. Technol.* 49, 4980–4987.
- Mouazen, A.M., Karoui, R., De Baerdemaeker, J., Ramon, H., 2005. Classification of soil texture classes by using soil visual near infrared spectroscopy and factorial discriminant analysis techniques. *J. Near Infrared Spec.* 13, 231–240.
- Næsset, E., Bollandsas, O.M., Gobakken, T., 2005. Comparing regression methods in estimation of biophysical properties of forest stands from two different inventories using laser scanner data. *Remote Sens. Environ.* 94, 541–553.
- Pirie, A., Singh, B., Islam, K., 2005. Ultra-violet, visible, near-infrared, and mid-infrared diffuse reflectance spectroscopic techniques to predict several soil properties. *Aust. J. Soil Res.* 43, 713–721.
- R Core Team, 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.
- Rajalahti, T., Arneberg, R., Berven, F.S., Myhr, K.-M., Ulvik, R.J., Kvalheim, O.M., 2009. Biomarker discovery in mass spectral profiles by means of selectivity ratio plot. *Chemometr. Intell. Lab. Syst.* 95, 35–48.
- Santos, H.G., Jacomine, P.K.T., Anjos, L.H.C., 2006. Sistema Brasileiro de Classificac, ão de Solos, second ed. Embrapa Solos, Rio de Janeiro.
- Savitzky, A., Golay, M.J., 1964. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 36, 1627–1639.
- Shepherd, K.D., Walsh, M.G., 2002. Development of reflectance spectral libraries for characterization of soil properties. *Soil Sci. Soc. Am. J.* 66, 988–998.
- Shi, Z., Ji, W., Rossel, R.A.V., Chen, S., Zhou, Y., 2015. Prediction of soil organic matter using a spatially constrained local partial least squares regression and the Chinese vis-NIR spectral library. *Eur. J. Soil Sci.* 66, 679–687.
- Shi, Z., Wang, Q.L., Peng, J., Ji, W.J., Liu, H.J., Li, X., Viscarra Rossel, R.A., 2014. Development of a national VNIR soil-spectral library for soil classification and prediction of organic matter concentrations. *Sci. China Earth Sci.* 57, 1671–1680.
- Soil Survey Staff, 2003. Keys to Soil Taxonomy, 9th ed. United States Department of Agriculture, Washington.
- 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.
- Viscarra Rossel, R.A., Behrens, T., 2010. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma* 158, 46–54.
- Viscarra Rossel, R.A., Behrens, T., Ben-Dor, E., 2016. A global spectral library to characterize the worlds soil. *Earth-Sci. Rev.* 155, 198–230.
- 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.
- Viscarra Rossel, R.A., Webster, R., 2012. Predicting soil properties from the Australian soil visible-near infrared spectroscopic database. *Eur. J. Soil Sci.* 63, 848–860.
- Wang, H.W., Wu, Z.B., Meng, J., 2006. Partial Least Squares Regression: Linear and Nonlinear Methods. National Defence Industry Press, Beijing.
- Wold, S., Martens, H., Wold, H., 1983. The multivariate calibration problem in chemistry solved by the PLS method. *Lect. Notes Math.* 973, 286–293.
- Xia, F., Peng, J., Wang, Q.L., Zhou, L.Q., Shi, Z., 2015. Prediction of heavy metal content in soil of cultivated land: hyperspectral technology at provincial scale. *J. Infrared Millim. Waves* 34, 593–605.
- Xie, X.L., Pan, X.Z., Sun, B., 2012. Visible and near-infrared diffuse reflectance spectroscopy for prediction of soil properties near a copper smelter. *Pedosphere* 22, 351–366.