Visible and near-infrared spectroscopy with chemometrics are able to predict soil physical and chemical properties

Jinbao Liu[1,2] · Jiancang Xie[1] · Jichang Han[2,3] · Huanyuan Wang[2,3] · Jianhong Sun[3] · Rui Li[3] · Shaoxuan Li[1]

Abstract

Purpose Comparing to conventional laboratory methods, visible–near-infrared reflectance (vis–NIR) spectroscopy is a more practical and cost-effective approach for estimating soil physical and chemical properties.

Materials and methods This paper aims to build statistical machine learning models to investigate the efficiency of spectral data for comprehensive evaluation of the soil quality indicators. Seventeen physical and chemical properties were measured using standard methods as indicators of soil quality. Soil samples were scanned in the laboratory in the vis–NIR range (350–2500 nm), the calibration set of 31 samples and the validation set of 13 samples for cross-validation and independent validation; twenty-four preprocessing methods were tested to improve predictions, and a partial least squares regression (PLSR) was used to predict soil quality indicators.

Results and discussion Comparing model indices, the model constructed based on the PLSR machine learning method has a good predictive power ($R^2 > 0.9$, ratio of performance to deviation (RPD) > 3.0). For physical and chemical properties, the bulk density (BD, $R^2 = 0.97$, RPD = 5.90), soil organic matter (SOM, $R^2 = 0.98$, RPD = 8.56), pH ($R^2 = 0.95$, RPD = 4.40), and TN ($R^2 = 0.98$, RPD = 6.67) concentration were predicted. This indicates that the method is suitable for the prediction of these soil elements in this study area. For the heavy properties, except for Mn, Zn, Cd, and As, the other five heavy metal concentrations were well predicted. It can be seen that the prediction ability of the construction model is Hg, Cr, Pb, Ni, and Cu in order of superiority to inferiority. The results show that a combination of spectroscopic and chemometric techniques can be applied as a practical, rapid, low-cost, and quantitative approach for evaluating soil physical and chemical properties in Shaanxi, China.

Keywords Vis–NIR spectroscopy · Soil properties · Partial least squares regression · Spectral preprocessing · Chemometrics

Abbreviations

BD	Bulk density
SW	Gravimetric soil water
M	Soil organic matter
TN	Total nitrogen
NN	Nitrate nitrogen
AK	Available potassium
AP	Available phosphorus
SG	Savitzky-Golay
FD	First deviation
SD	Second deviation
SNV	Standard normal variate
MSC	Multiplicative scatter correction
NOR	Normalization
Max	Maximum
Min	Minimum
SD	Standard deviation
CV	Coefficient of variation

✉ Jiancang Xie
  jcxie@xaut.edu.cn

1   Institute of Water Resources and Hydro-electric Engineering, Xi'an University of Technology, Xi'an 710048, China

2   Key Laboratory of Degraded and Unused Land Consolidation Engineering, The Ministry of Land and Resources, Xi'an 710075, Shaanxi, China

3   Shaanxi Provincial Land Engineering Construction Group Co., Ltd., Xi'an 710075, Shaanxi, China

1 Introduction

Soil is a fundamental natural resource that consists of organic and inorganic mineral matter, water, and air, which people rely on for the production of food and energy (Palma et al. 2007; Cuffney 2010). Soil is an environmental filter for metals, nutrients, and other contaminants that may leach into the environment. The ability of a soil to support any of these functions depends on its structure, composition, and chemical, biological, and physical properties (Karlen et al. 1997). Evaluation of soil quality usually utilizes a combination of physical and chemical indicators as a minimum data set (Bo et al. 2010; Rossel et al. 2016). Compared with conventional analytical methods, vis–NIR is nondestructive, fast, and cost-effective (Shepherd and Walsh 2002; Rossel et al. 2006; Hong et al. 2017).

Soil spectral properties are affected by soil minerals, water content, organic matter, and texture. The transitional abilities of different soil components are different, and when NIR radiation interacts with a soil sample, it is the overtones (Hbroge et al. 2004; Rossel et al. 2008). Therefore, the different absorption bands in the soil absorption spectrum allow quantitative analysis of the soil component content. In the visible region (400–780 nm), molecules of various components of the soil produce electron absorption spectra. Electronic excitations are the main process as the energy of the radiation is high and this region contains useful information in organic and inorganic materials in the soil (Mortimore et al. 2004; Schoell et al. 2005). For absorptions in the NIR region (780–2500 nm), the absorption of soil is mainly caused by fundamental frequency vibration of C–H, N–H, O–H, and frequency-doubling vibration absorption (Bo et al. 2010; Rodionov et al. 2014). Water has a strong influence on vis–NIR spectra of soil, the absorption bands around 1400–1900 nm, and other parts of the vis–NIR region.

Several literatures have shown the capacity of the combined use of vis–NIR spectroscopic and chemometrics to predict soil physical and chemical properties in the most recent 20 years (Rossel et al. 2006; Mouazen et al. 2010; Dotto et al. 2017). For example, soil bulk density and soil water content with spectroscopy have been discussed with satisfied predictive results (Quraishi and Mouazen 2013; Al-Asadi and Mouazen 2014; Liu et al. 2014; Morellos et al. 2016). Regarding the chemical properties, several researchers have demonstrated accurate predictions of soil chemical content. Hong et al. (2017) and Jiang et al. (2016) demonstrated new possibilities to estimate SOM using vis–NIR spectroscopy with multivariate modeling techniques. Besides, combined vis–NIR spectroscopy with partial least squares regression (PLSR) has been applied in total nitrogen (TN), moisture content, and pH (Buondonno et al. 2012; Kuang and Mouazen 2013; Morellos et al. 2016; Filippi et al. 2018). In some agricultural areas, the vis–NIR spectroscopy-based approach has also been widely used in monitoring agricultural pollution, especially for heavy metal pollution (Angelopoulou et al. 2017; Todorova et al. 2018).

In this study, we collected 44 soil samples from common parent materials and textures in the three regions of Shaanxi, China. Soil samples were scanned in the laboratory in the range of 350–2500 nm; twenty-four preprocessing methods and correlation analysis were tested to improve predictions, and a partial least squares regression was used to predict soil quality indicators. Our goal is to establish a quantitative model of soil quality and evaluate the ability of soil spectra for predicting soil physical and chemical properties in Shaanxi Province.

2 Materials and methods

2.1 Study area and soil sampling

The study site was located in the middle of Guanzhong Plain, Shaanxi Province, China (Fig. 1). The terrain is high in the northwest and low in the southeast. It descends from north to south and is divided into three roads. It consists of four land types: the frontal alluvial fan, the loess plateau, the flood plain, and the alluvial terrace. The annual average temperature is 12.9 °C, and the average annual precipitation is between 552.6 and 663.9 mm.

We collected 44 soil samples in February 2013. The soil type is earth–cumulic–orthic anthrosols, and the sampling depth was the thickness of the tillage layer, usually 0–30 cm. These soil samples consist of different physical and chemical properties, and these substances affect the reflectance and absorption spectral characteristics of soil. In order to establish a more accurate spectral quantitative model for the study area, seventeen soil property indicators were measured including soil organic matter (SOM); total nitrogen (TN), bulk density (BD), soil water (SW), total nitrogen (TN), nitrate nitrogen (NN), available potassium (AK), available phosphorus (AP), Cr, Mn, Ni, Cu, Zn, As, Cd, Hg, and lead (Pb).

2.2 Chemical analysis and spectral measurement

2.2.1 Soil physical measurements

Soil samples were air-dried and crushed, and crop residues, root material, and stones were removed so the soil could pass through the 2-mm sieve. Bulk weight was analyzed by collecting a known volume of soil using a metal ring pressed into the soil (intact core), and determining the weight after drying (Ito et al. 2002). The moisture content was determined gravimetrically after heating the samples at 105 °C for 24 h (Ruberto et al. 2010).
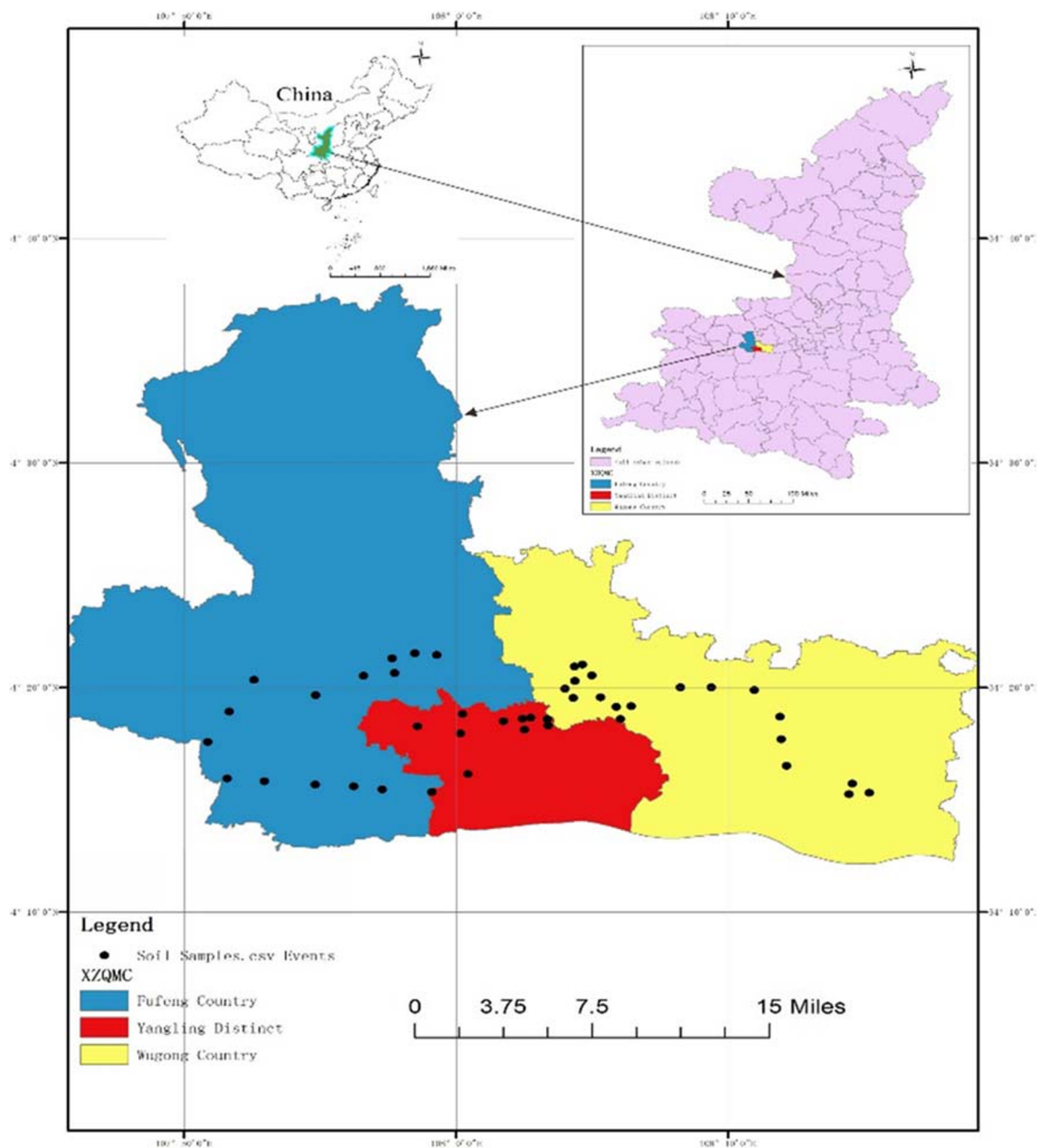
Fig. 1 Study area and sampling point distribution

2.2.2 Soil chemical measurements

SOM was measured by the combustion method using a LECO® FP2000 analyzer (Laboratory Equipment Corporation, St Joseph, MI, USA). Soil pH was measured in deionized water (Zornoza et al. 2008). TN was measured by the Kjeldahl method (Houba 1997). NN was determined by a UV spectrophotometer (Gross et al. 2010). AP was measured by the NaHCO₃ method (Yaseen and Malhi 2009). AK was measured by the acetic acid-flame photometric method (Kataoka et al. 1991). The heavy metal content was determined by inductively coupled plasma mass spectrometry (ICP-MS, Agilent 7700), and included Cr, Mn, Ni, Cu, Zn, As, Cd, Hg, and lead (Pb) (Gajek et al. 2013).

2.2.3 Soil reflectance measurements

Diffuse reflectance spectra of soil samples were measured using a portable spectroradiometer (Fieldspec 4, Analytical Spectral Devices, Inc.) with a spectral range of 350–2500 nm. The spectroradiometer had a bandwidth of 1.4 nm at 350–1000 nm and 1.1 nm at 1001–2500 nm. Measurements were made with a Hi-Brite Contact Probe that uses halogen bulb color temperature (2901 ± 10 K) for illumination. The contact probe measures a spot

Table 1 Statistical analysis of the attributes of soil properties

Soil quality indicators	Mean	Max	Min	SD	CV
BD (g cm^{-3})	1.23	1.61	1.02	0.133	0.11
SOM (g kg^{-1})	18.39	26.93	11.83	3.60	0.20
SW (%)	0.18	0.38	0.09	0.04	0.25
pH (unitless)	8.12	8.38	7.47	0.18	0.02
TN (g kg^{-1})	1.35	1.60	0.61	0.16	0.12
NN (mg kg^{-1})	4.68	13.75	0.88	2.82	0.60
Ak (mg kg^{-1})	142.6	368.9	80.1	56.85	0.40
AP (mg kg^{-1})	26.0	105.1	0.8	17.15	0.66
Cr (mg kg^{-1})	17.95	19.99	16.06	0.93	0.05
Mn (mg kg^{-1})	274	310	249	13.58	0.05
Ni (mg kg^{-1})	12.77	14.15	14.15	0.68	0.05
Cu (mg kg^{-1})	7.29	8.37	6.37	0.45	0.06
Zn (mg kg^{-1})	15.81	19.79	12.71	1.88	0.12
As (mg kg^{-1})	7.51	13.40	4.78	1.87	0.25
Cd (mg kg^{-1})	0.40	0.51	0.32	0.04	0.11
Hg (mg kg^{-1})	12.58	23.92	6.49	5.17	0.41
Pb (mg kg^{-1})	21.05	24.98	19.28	1.16	0.05

SOM soil organic matter, *TN* total nitrogen, *BD* bulk density, *SW* soil water, *TN* total nitrogen, *NN* nitrate nitrogen, *AK* available potassium, *AP* available phosphorus, *Max* maximum, *Min* minimum, *SD* standard deviation, *CV* coefficient of variation

size of 10 nm. The sensor was calibrated with a white reference panel once every ten measurements. For every sample, the Petri dish was rotated 4 times, each time by 90°, and 40 spectra were averaged to minimize noise and to maximize the signal-to-noise ratio, then averaged in order to form a representative spectrum. Spectra were recorded with a sampling resolution of 1 nm, and thus, each spectrum comprised 2151 reflectance channels in the range of 350–2500 nm.

2.3 Spectral preprocessing

Spectral preprocessing with mathematics was commonly used to correct measurements and noisy spectra. In addition to chemical properties, soil structural properties also influence soil spectra, such as light scattering effects. To enhance the more chemically relevant peaks in the spectra and reduce the effects

baseline shifts and overall curvature, various spectral preprocessing approaches have been studied. For example, Savitzky–Golay (SG) smoothing with the first and second derivatives with a second-order polynomial and window size of 9 wavelengths were used to reduce baseline variation and increase resolutions of spectral peak features (Peng et al. 2014; Fu et al. 2018). Standard normal variable transformation (SNV) transform was used to reduce the particle size effect and curvilinear trend of the spectrum (Dhanoa et al. 1995). Multiplication scattering correction (MSC) and normalization (NOR) removes additive or multiplicative signal effects (Burger and Geladi 2007). Logarithmic transformation (LOG (1/R)) not only enhances the difference in visible spectrum but also reduces the influence of multiplicative factors under light conditions (Liu et al. 2018). In this paper, prior to building the predictive models, twenty-four spectral preprocessing methods were explored, in order to improve the final soil physical and chemical predictions performances. The treatments included SG smoothing with the first and second derivatives with a second-order polynomial and a window size of 9 wavelengths alone and in conjunction with SNV, MSC, NOR, and LOG (1/R).

2.4 Prediction model building and testing

Partial least squares regression (PLSR) is a primary statistical method, and it combines features from principal component analysis and multiple regression in order to reduce complex spectral matrix (Ergon 2014). It is particularly useful when we need to predict a set of dependent variables from a very large set of independent variables. Because it can handle the large dimensionality and collinearity of data produced by vis–NIR spectroscopy, this method has been widely applied in the field of chemometrics (Qiao et al. 2015). To train and then test our models, we selected at random 70% of the soil samples to represent the calibration set (33 samples) to build PLSR models, and the remaining 30% was used for the validation set (13 samples).

Fig. 2 Reflectance curves of soil samples with different content
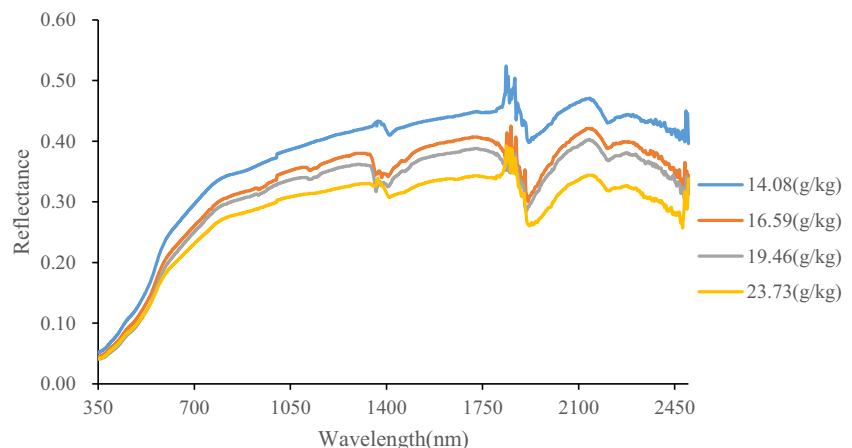
For each PLSR model, several performance metrics were used for model evaluation. The coefficient of determination (R^2) and the root mean squared error (RMSE) give indications of the validity of the prediction models. The standard error (SE) measures the precision of the prediction models. The ratio of performance to deviation (RPD) was calculated as the standard deviation (SD) of the validation set divided by the RMSE (Sarathjith et al. 2014; Gholizadeh et al. 2017). Equations (1)–(4) are as follows:

$$R^2 = \frac{\sum_{i=1}^{n}\left(x_i-\overline{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i-\overline{y}_i\right)^2} \tag{1}$$

$$RMSE = \sqrt{\sum_{i=1}^{N}\frac{\left(\widehat{y}_i-y_i\right)^2}{N}} \tag{2}$$

$$SE = \sqrt{\sum_{i=1}^{N}\frac{\left(\widehat{y}_i-\overline{y}\right)^2}{(N-1)}} \tag{3}$$

$$RPD = \frac{SD}{RMSE} \tag{4}$$

where $\widehat{y}_i$ indicates the values estimated by the model, y_i indicates the observed values, and N is the number of observations of the variable to be modeled. The RPD was evaluated with the following criteria: RPD > 3 represents an excellent prediction; 2 < RPD < 3 has limited predictive ability; RPD < 2 has no predictive ability (Prs et al. 2012; Sun et al. 2018).

3 Results and discussion

3.1 Soil properties

The statistics of reference values of soil property concentrations are summarized in Table 1. The research areas' soil texture is silt. The pH values ranged from 7.47 to 8.38, neutral to weak alkaline. The mean value of pH indicators shows that the soil is alkaline in the study area. The BD values ranged from 1.02 to 1.61. The content of SOM ranged from 11.83 to 26.93 g kg⁻¹. In addition, the ranges of TN, NN, AK, and AP concentrations were 0.61–1.60 mg kg⁻¹, 0.88–13.75 mg kg⁻¹, 80.1–368.9 mg kg⁻¹, and 0.8–105.1 mg kg⁻¹, respectively. According to the Soil Environment Quality standards in China, the Hg, Pb, and Mn concentrations showed more skewed and irregular distributions. This is probably caused by irrational use of mercury, which is usually contained in fertilizers and pesticides. In addition, compared with Ni, Cu, As, and Cd, the Hg concentrations presented a higher standard deviation (SD = 5.17)

Table 2 The linear correlation coefficients among soil properties ($P < 0.01$)

Properties	BD	SOM	Moisture content	pH	TN	NN	Ak	AP	Cr	Mn	Ni	Cu	Zn	As	Cd	Hg	Pb
SOM (g kg⁻¹)	-0.33	1															
SW (g g⁻¹)	0.14	0.09	1														
pH (unitless)	0.08	-0.38	0.24	1													
TN (g kg⁻¹)	0.18	0.19	0.02	-0.05	1												
NN (mg kg⁻¹)	0.04	0.08	0.10	-0.03	0.01	1											
Ak (mg kg⁻¹)	0.08	-0.03	0.16	-0.01	0.12	-0.06	1										
AP (mg kg⁻¹)	-0.09	0.23	-0.09	-0.26	0.21	0.03	0.11	1									
Cr (mg kg⁻¹)	0.25	-0.35	0.01	0.36	-0.04	0.01	-0.13	-0.18	1								
Mn (mg kg⁻¹)	0.22	-0.45	0.07	0.39	0.06	0.03	-0.12	-0.26	0.88	1							
Ni (mg kg⁻¹)	0.23	-0.29	0.02	0.35	0.04	0.09	-0.15	-0.24	0.94	0.89	1						
Cu (mg kg⁻¹)	0.17	-0.20	0.05	0.25	0.02	0.16	0.01	-0.26	0.78	0.76	0.83	1					
Zn (mg kg⁻¹)	0.12	0.45	-0.05	0.45	-0.04	-0.04	-0.19	-0.17	0.62	0.68	0.62	0.61	1				
As (mg kg⁻¹)	-0.04	-0.05	-0.05	-0.04	-0.27	-0.13	-0.03	-0.04	-0.02	-0.05	-0.10	-0.03	0.11	1			
Cd (mg kg⁻¹)	-0.06	0.24	-0.11	-0.07	0.11	0.06	0.06	-0.17	0.19	0.30	0.31	0.55	0.13	0.21	1		
Hg (mg kg⁻¹)	0.10	-0.56	-0.04	0.50	-0.02	-0.06	-0.17	-0.12	0.59	0.76	0.57	0.50	0.83	0.24	0.14	1	
Pb (mg kg⁻¹)	0.10	-0.22	-0.08	0.20	0.01	-0.08	-0.22	-0.05	0.64	0.75	0.59	0.68	0.55	0.13	0.46	0.68	1

Table 3 The results of calibration and validation using the preprocessing in the soil properties

Indicators	Optimum preprocessing	Pc	Calibration			Validation			
			R^2	RMSE	SE	R^2	SE	RMSE	RPD
BD	log(1/R) + SG + MSC + FD	11	0.96	0.03	0.03	0.97	0.02	0.02	5.90
SOM	SG + MSC + FD	8	0.97	0.49	0.50	0.98	0.46	0.45	8.56
SW	log(1/R) + SG + SNV + FD	6	0.87	0.02	0.02	0.80	0.01	0.01	2.09
pH	R + SG + SNV + FD	8	0.95	0.03	0.03	0.95	0.05	0.05	4.40
TN	R + SG + NOR + SD	9	0.96	0.03	0.03	0.98	0.02	0.02	6.67
NN	log(1/R) + SG + FD	5	0.87	1.09	1.11	0.90	0.64	0.62	3.07
Ak	R + SG + SNV + SD	5	0.84	2.44	2.48	0.89	1.44	1.39	2.91
AP	log(1/R) + SG + MSC + SD	6	0.82	5.82	5.92	0.85	5.98	5.75	3.58
Cr	log(1/R) + SG + MSC + FD	8	0.93	0.23	0.24	0.97	0.10	0.10	5.96
Mn	log(1/R) + SG + SNV + FD	7	0.88	4.22	4.29	0.80	3.08	2.96	2.09
Ni	R + SG + MSC + SD	7	0.94	0.14	0.15	0.93	0.13	0.13	3.74
Cu	R + SG + NOR + FD	6	0.91	0.11	0.11	0.92	0.08	0.08	3.38
Zn	log(1/R) + SG + MSC + FD	5	0.80	0.75	0.76	0.91	0.33	0.31	3.13
As	log(1/R) + SG + MSC + SD	6	0.89	0.62	0.63	0.87	0.65	0.62	2.58
Cd	log(1/R) + SG + NOR + FD	9	0.94	0.01	0.01	0.97	0.03	0.03	2.77
Hg	log(1/R) + SG + SNV + FD	8	0.98	0.61	0.62	0.99	0.13	0.12	8.59
Pb	R + SG + SNV + SD	6	0.93	0.33	0.33	0.97	0.10	0.10	5.57

SG Savitzky–Golay, *SNV* standard normal variate, *MSC* multiplicative scatter correction, *NOR* normalization, *Pc* principal component score

and a high coefficient of variation (CV = 41%), thereby exceeding the critical value. The contamination may result from sewage irrigation. This indicated that the selected properties were spatially variable within the study area, and the wide range of physical and chemical properties provides vis–NIR calibration and validation accuracy (Abdi et al. 2012; Xu et al. 2018).

3.2 Soil spectral analysis

Because of the transition energy levels of different soil components, there are different absorption bands in the soil absorption spectrum curve. The calculated mean reflectance spectrums for soil samples with different SOM content are shown in Fig. 2. In general, the soil samples have similar spectral patterns. In particular, the spectral reflectance increases with wavelength and has three prominent absorptions near 1400 nm, 1900 nm, and 2100 nm; this is because of soil moisture was strongly absorbed. The first absorption region at approximately 1400 nm is the first overtone of the O–H stretch and C–H combination of aromatic structures from lignin (Vasques et al. 2008). The second absorption at approximately 1900 nm is due to the combination of O–H and H–O–H. The third absorption at approximately 2200 nm results from O–H stretches and the Al/Fe-OH bend (Bendor 2002;

Fig. 3 Scatter plots of the measured and vis–NIR predicted soil bulk density and gravimetric soil water. Note: BD, bulk density; SW, gravimetric soil water
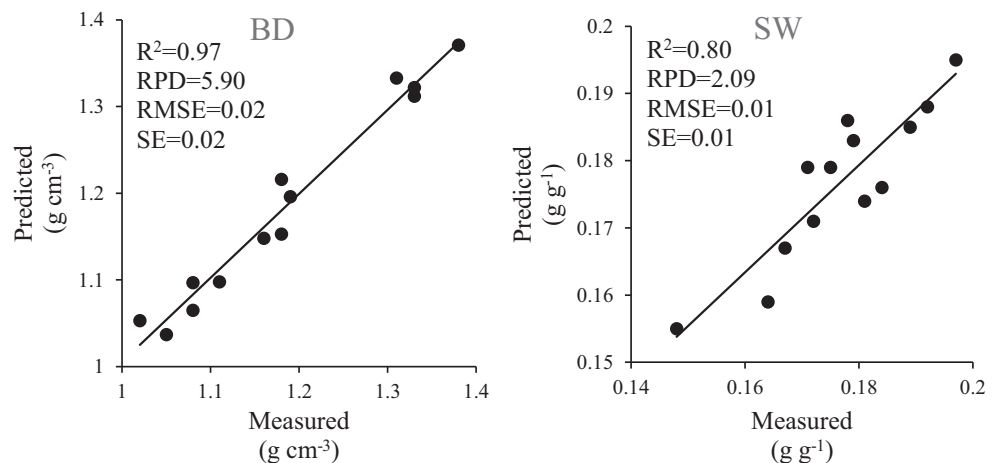
Table 4 Best results of calibration and validation in the soil chemical properties
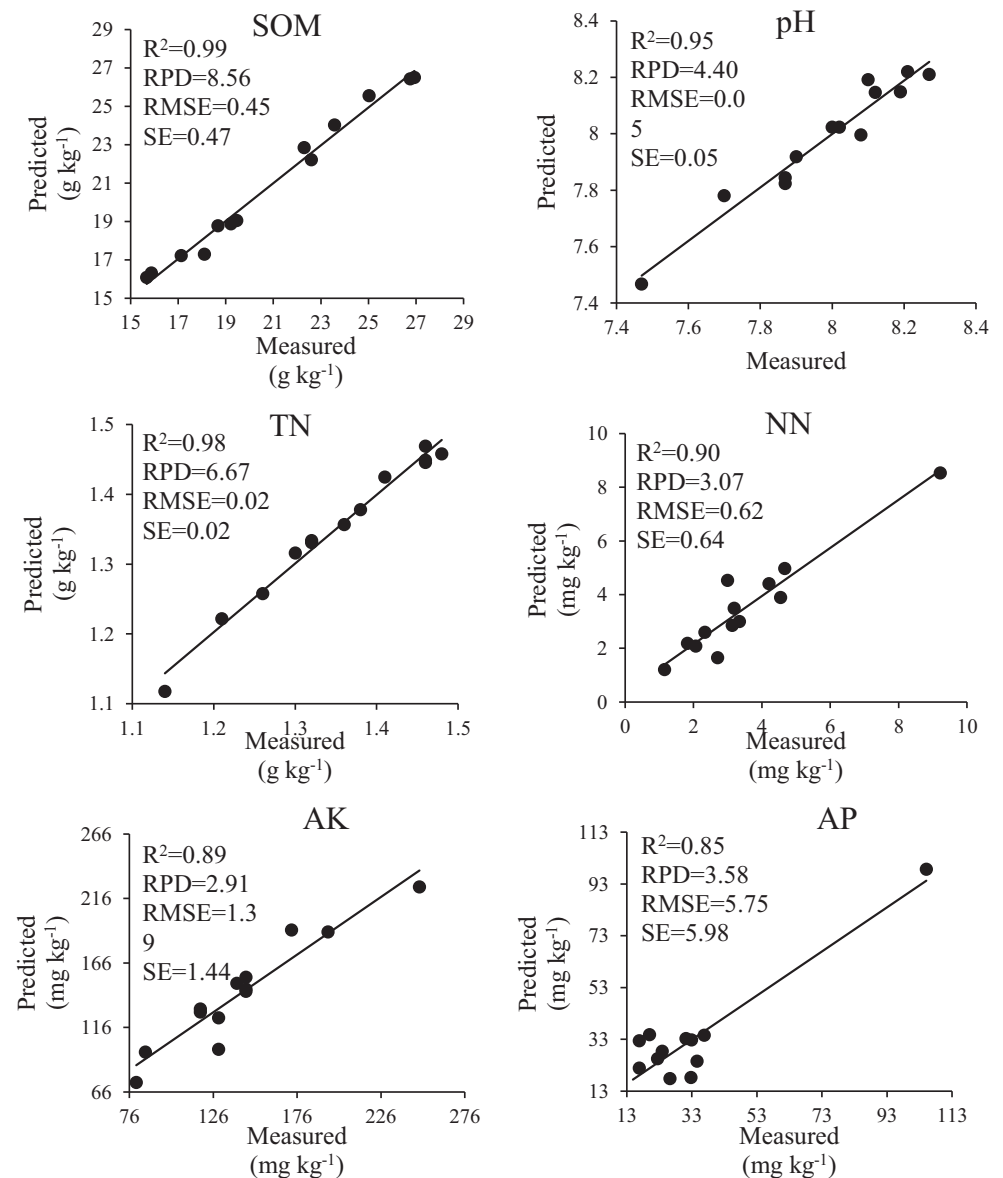
Indicators	Optimum preprocessing	Calibration			Validation			
		R^2	RMSE	SE	R^2	SE	RMSE	RPD
SOM	SG + MSC + FD	0.97	0.49	0.50	0.98	0.46	0.45	8.56
pH	R + SG + SNV + FD	0.95	0.03	0.03	0.95	0.05	0.05	4.40
TN	R + SG + NOR + SD	0.96	0.03	0.03	0.98	0.02	0.02	6.67

SG Savitzky–Golay, *SNV* standard normal variate, *MSC* multiplicative scatter correction, *NOR* normalization, *Pc* principal component score

Ding et al. 2013). Between 350 and 760 nm, the reflectance curves showed an increasing trend. The absorption regions between 350 and 950 nm were caused by the iron oxides and organic matter, and the absorptions between 1100 and 1550 nm result from clay minerals and water; between 1700 and 1800 nm, they were caused by carbon, and absorptions

between 2100 and 2500 nm result from kaolinite, smectite, carbonates, and organic compounds (Bishop et al. 2008; Nawar et al. 2016). Despite the similarity, it is also clear to see that the samples with higher SOM content tend to have lower reflectance values. This is because as the SOM content increases, the soil color gradually becomes deeper. When light
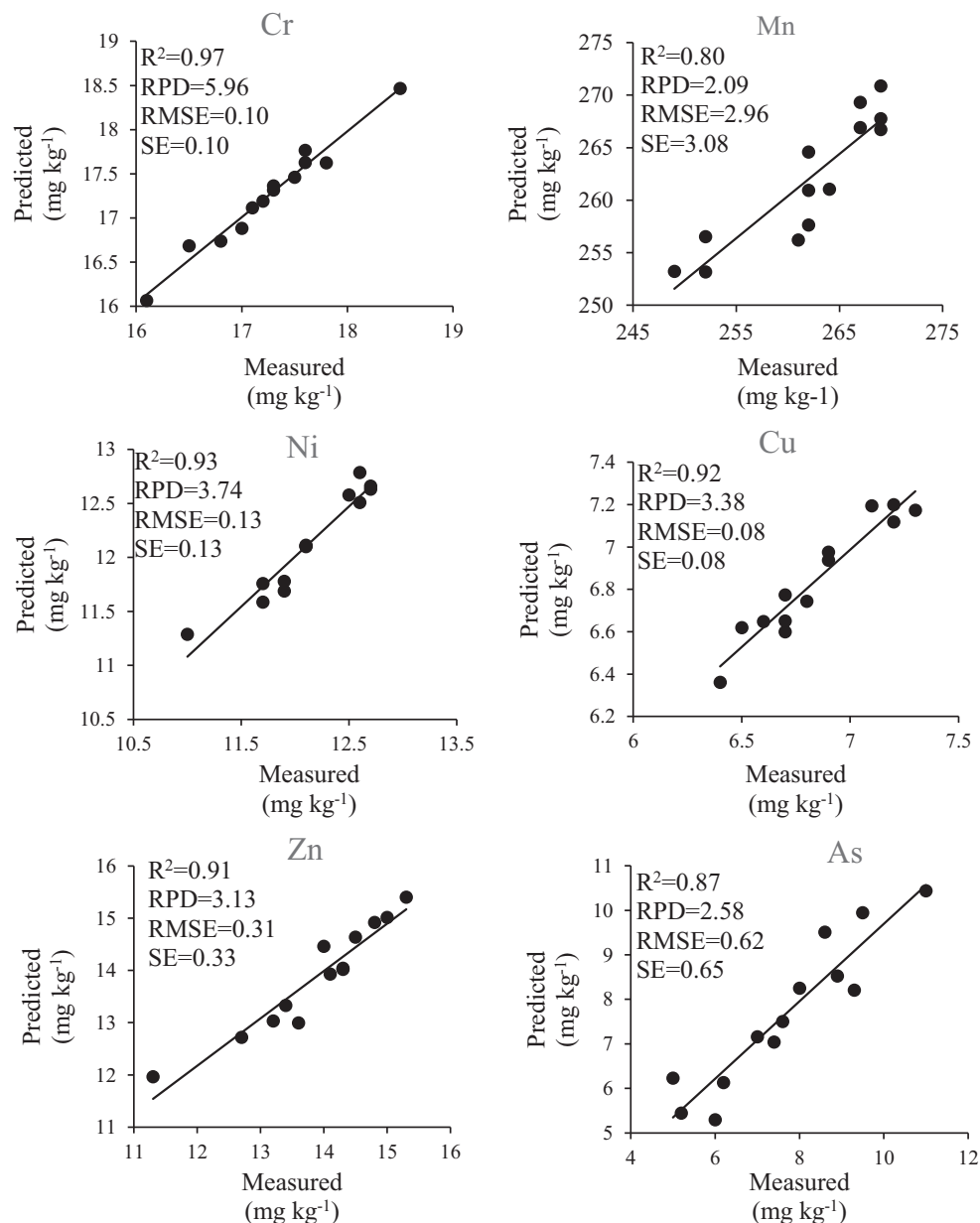

Fig. 4 Scatter plots of the measured and vis–NIR predicted soil chemical contents. Note: SOM, soil organic matter; TN, total nitrogen; NN, nitrate nitrogen; AK, available potassium; AP, available phosphorus

Table 5 Calibration and validation statistics of the best results of five heavy metals

Indicators	Optimum preprocessing	Calibration			Validation			
		R^2	RMSE	SE	R^2	SE	RMSE	RPD
Cr	$\log(1/R) + SG + MSC + FD$	0.93	0.23	0.24	0.97	0.10	0.10	5.96
Ni	$R + SG + MSC + SD$	0.94	0.14	0.15	0.93	0.13	0.13	3.74
Cu	$R + SG + NOR + FD$	0.91	0.11	0.11	0.92	0.08	0.08	3.38
Hg	$\log(1/R) + SG + SNV + FD$	0.98	0.61	0.62	0.99	0.13	0.12	8.59
Pb	$R + SG + SNV + SD$	0.93	0.33	0.33	0.97	0.10	0.10	5.57

SG Savitzky–Golay, *SNV* standard normal variate, *MSC* multiplicative scatter correction, *NOR* normalization, *Pc* principal component score

Fig. 5 Scatter plots of the measured and vis–NIR predicted soil heavy metal contents
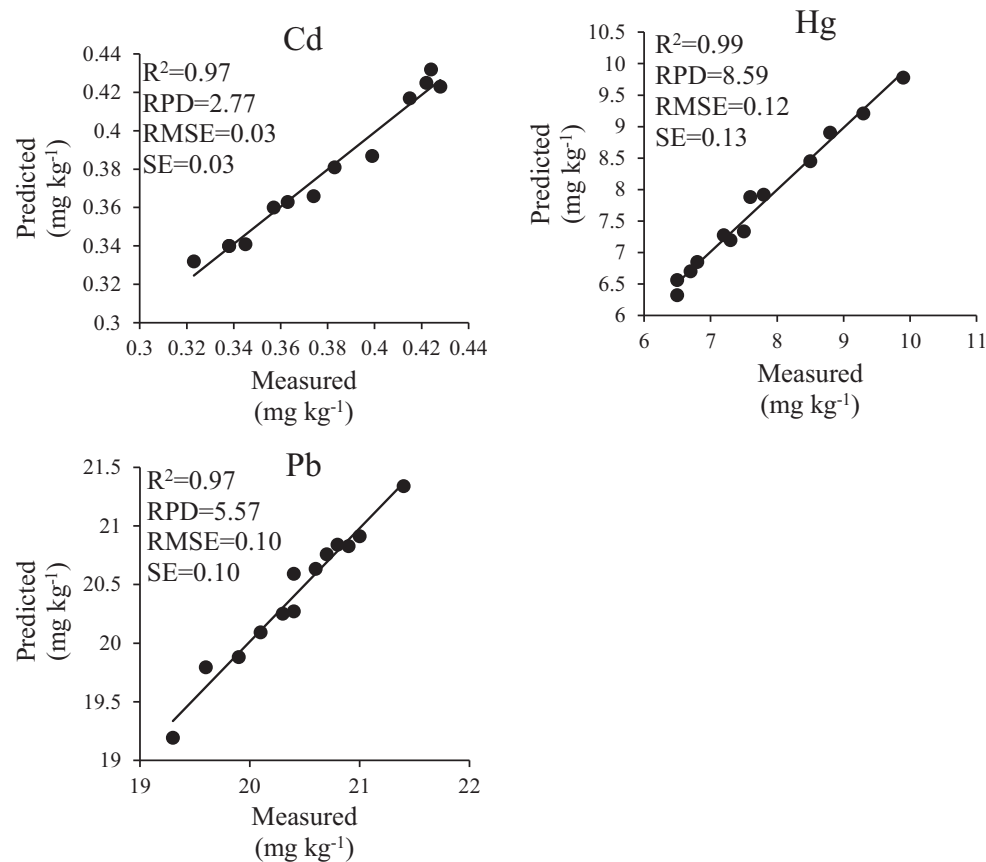
Fig. 5 continued.

hits the surface of the soil, the incident light is absorbed by the soil, causing the soil reflectivity to gradually decrease.

3.3 Correlation analysis

Soil is the product of weathering and soil formation of the surface rocks of the crust. Generally, the chemical composition of the soil is relatively stable, and the content level and variation range of the elements are low. Therefore, the content of soil chemical elements is highly comparable. In studying the correlation between elements, it could be inferred whether the sources are the same. Various elements in the soil have different migration and enrichment trends, showing that the correlation of each element content is different. The Pearson correlation coefficient was used to analyze relationships between soil elements (Table 2).

From the results, we founded that strong correlations were observed between SOM with Zn ($r = -0.45$), which was mainly caused by zinc accumulation in soil with high humus content. The zinc content of soil increases with the increase of humus and organic matter content (Wang et al. 2014). Also, the correlations between heavy metal elements are significant. For example, strong correlations were observed between Cr and Ni ($r = 0.94$), Mn ($r = 0.88$), Pb ($r = 0.64$), Cu ($r = 0.62$),

and Hg ($r = 0.59$). The correlation coefficients of Ni and Cu, Zn, Hg, and Pb were 0.83, 0.57, and 0.59, respectively; the correlation coefficients of Cu and Zn, Cd, Hg, and Pb were 0.61, 0.55, 0.50, and 0.68, respectively. The correlation coefficients of Zn and Hg and Pb were 0.83 and 0.55, respectively. The correlation coefficient of Hg and Pb was 0.68. The Cu, Zn, and Pb elements are sulfur-philic elements, and the similar elements have a certain commonality in epigenetic geochemistry, and thus have significant correlation. It shows that these heavy metals are more likely to come from the same source of pollution and have a strong companion effect.

3.4 Model performance

After S–G nine-point smoothing, transformation of NOR, MSC, and SNV, the first deviation (FD), second deviation (SD), and LOG ($1/R$) combination PLSR were used to establish the prediction model. At the same time, in order to avoid over-fitting, the dimension of the input variables was kept as small as possible.

The model performances for different soil indicators are summarized in Table 3, and accuracies were reported on the testing dataset. For physical indictors BD and SW, the best result was achieved for BD, the best preprocessing was

$\log(1/R) + SG + MSC + FD$, the accuracies were $R^2 = 0.97$, RPD = 5.90, RMSE = 0.02, and SE = 0.02, which can be considered as an excellent model. For the SW, the $R^2 < 0.90$, RPD = 2.09 < 3, which cannot be considered an excellent model (Fig. 3). The predictions of BD and SW have been obtained, as observed in previous studies (Quraishi and Mouazen 2013; Al-Asadi and Mouazen 2014; Liu et al. 2014). The BD could be affected by good soil structure that leads to overall good soil quality (Figueroa 2003).

Calibration and validation model results show that different preprocessing methods have significant effects on soil physical and chemical properties. For chemical properties, compared with performance metrics R^2, RMSE, SE, and RPD, best results of calibration and validation in the soil chemical properties were achieved for SOM ($R^2 = 0.98$, RPD = 8.56, RMSE = 0.45), pH ($R^2 = 0.95$, RPD = 4.40, RMSE = 0.05), and TN ($R^2 = 0.98$, RPD = 6.67, RMSE = 0.02) (Table 4). The three best optimum preprocessing models include SG + MSC + FD, R + SG + SNV + FD, and R + SG + NOR + SD. The performances of SOM, pH, and TN prediction were better than AK ($R^2 = 0.89$, RPD = 2.91 < 3.0, RMSE = 1.39), AP ($R^2 = 0.85 < 0.90$, RPD = 3.58, RMSE = 5.7495), and NN ($R^2 = 0.90$, RPD = 3.07, RMSE = 0.62) (Fig. 4), resulting from the overtone and combinations of O–H, C–H (Rossel et al. 2006; Vasques et al. 2008). Therefore, further research needs to improve the calibration and validation accuracy of soil properties, especially AK, AP, and NN.

Regarding the heavy properties, calibration and validation statistics of the best results of five heavy metals are presented in Table 5. Models for five heavy metals include Hg ($R^2 = 0.99$, RPD = 8.59, RMSE = 0.12), Cr ($R^2 = 0.97$, RPD = 5.96, RMSE = 0.10), Ni ($R^2 = 0.93$, RPD = 3.74, RMSE = 0.13), Pb ($R^2 = 0.97$, RPD = 5.57, RMSE = 0.10), and Cu ($R^2 = 0.92$, RPD = 3.38, RMSE = 0.08). Models for As ($R^2 = 0.87$, RPD = 2.58), Mn ($R^2 = 0.80$, RPD = 2.09), and Cd (RPD = 2.77) had $R^2 < 0.9$ and RPD < 3.0; for the model of Zn, although $R^2 = 0.91 > 0.90$, RPD = 3.13 > 3.0, the offset had too much deviation (Fig. 5). According to the results, Mn, Zn, and As were not as good as the others, and the essential reason is that these heavy metals in soil are spectrally featureless and are limited in vis–NIR spectra (Wu et al. 2007; Wang et al. 2014).

4 Conclusions

Vis–NIR spectroscopy and chemometric analysis have been demonstrated to be reliable tools for estimating soil physical and chemical properties. In this study, based on collected soil spectral data and chemical indicators, we have established a prediction model of seventeen soil elements and analyzed the potential of vis–NIR spectroscopy to predict the contents of soil properties using the PLSR technique.

Our results show that accurate calibration and validation models in the soil chemical properties were achieved, including BD, SOM, pH, and TN. In addition, Hg, Cr, Ni, Pb, and Cu concentrations were well predicted. Moreover, vis–NIR spectroscopy is as an effective, nondestructive, fast, and cost-effective tool for accurate predictions of soil properties, and combined with methodology can be used for rapid acquisition of spectra, multi-parameters, and assessment of soil physical and chemical properties in the center of Shaanxi Province. Although the predictions were less accurate than the laboratory, we are optimistic of the potential of vis–NIR and chemometric methods in the prediction of soil properties.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Research involving human participants and/or animals This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

Abdi D, Tremblay GF, Ziadi N, Bélanger G, Parent L (2012) Predicting soil phosphorus-related properties using near-infrared reflectance spectroscopy. Sci Soil Sci Soc Am J 76(6):2318

Al-Asadi RA, Mouazen AM (2014) Combining frequency domain reflectometry and visible and near infrared spectroscopy for assessment of soil bulk density. Soil Tillage Res 135:60–70

Angelopoulou T, Dimitrakos A, Terzopoulou E, Zalidis G, Theocharis J, Stafilov T, Zouboulis A (2017) Reflectance spectroscopy (Vis-NIR) for assessing soil heavy metals concentrations determined by two different analytical protocols, based on iso 11466 and iso 14869-1. Water Air Soil Pollut 228(11):436

Bendor E (2002) Quantitative remote sensing of soil properties. Adv Agron 75(02):173–243

Bishop JL, Lane MD, Dyar MD, Brown AJ (2008) Reflectance and emission spectroscopy study of four groups of phyllosilicates: smectites, kaolinite-serpentines, chlorites and micas. Clay Miner 43(1): 35–54

Bo S, Rossel RAV, Mouazen AM, Wetterlind J (2010) Chapter five – visible and near infrared spectroscopy in soil science. Adv Agron 107(107):163–215

Buondonno A, Amenta P, Viscarra-Rossel RA, Leone AP (2012) Prediction of soil properties with PLSR and Vis-NIR spectroscopy: application to Mediterranean soils from southern Italy. Curr Anal Chem 8(2)

Burger J, Geladi P (2007) Spectral pre-treatments of hyperspectral near infrared images: analysis of diffuse reflectance scattering. J Near Infrared Spectrosc 15(2):29

Cuffney TF (2010) Input, movement and exchange of organic matter within a subtropical coastal black water river-flood plain system. Freshw Biol 19(3):305–320

Dhanoa MS, Lister SJ, Sanderson R, Barnes RJ (1995) The link between multiplicative scatter correction (msc) and standard normal variate (snv) transformations of nir spectra. Near Infrared Spectrosc 2(1): 43–47

Ding J, Yao Y, Wang F (2013) Quantitative remote sensing of soil salinization in arid regions based on three dimensional spectrum Eigen spaces. Acta Pedol Sin 50(5):853–861

Dotto AC, Dalmolin RSD, Grunwald S, Ten Caten A, Pereira Filho W (2017) Two preprocessing techniques to reduce model covariables in soil property predictions by Vis-NIR spectroscopy. Soil Tillage Res 172:59–68

Ergon R (2014) Principal component regression (PCR) and partial least squares regression (PLSR). In: Granato, Ares G (eds) Mathematical and statistical methods in food science and technology. Wiley, Blackwell, Chichester, pp 121–142

Figueroa D (2003) The role of relict vegetation in maintaining physical, chemical, and biological properties in an abandoned stipa -grass agroecosystem. Arid Land Res Manag 17(2):103–111

Filippi P, Cattle SR, Bishop TFA, Jones EJ, Minasny B (2018) Combining ancillary soil data with VisNIR spectra to improve predictions of organic and inorganic carbon content of soils. MethodsX 5:551–560

Fu C, Gan S, Yuan X, Xiong H, Tian A (2018) Determination of soil salt content using a probability neural network model based on particle swarm optimization in areas affected and non-affected by human activities. Remote Sens 10:1387

Gajek R, Barley F, She J (2013) Determination of essential and toxic metals in blood by ICP-MS with calibration in synthetic matrix. Anal Methods 5(9):2193–2202

Gholizadeh A, Carmon N, Klement A, Ben-Dor E, Borůvka L (2017) Agricultural soil spectral response and properties assessment: effects of measurement protocol and data mining technique. Remote Sens 9:1078

Gross A, Boyd CE, Seo J (2010) Evaluation of the ultraviolet spectrophotometric method for the measurement of total nitrogen in water. J World Aquacult Soc 30(3):388–393

Hbroge N, Gthomsen A, Hgreve M (2004) Prediction of topsoil organic matter and clay content from measurements of spectral reflectance and electrical conductivity. Acta Agric Scand 54(4):232–240

Hong YS, Yu L, Chen YY, Liu Y (2017) Prediction of soil organic matter by Vis-NIR spectroscopy using normalized soil moisture index as a proxy of soil moisture. Remote Sens 10(1):28

Houba VJG (1997) Methods of soil analysis-chemical methods. Sci Hortic 70(4):342–343

Ito K, Kato T, Ona T (2002) Non-destructive method for the quantification of the average particle diameter of latex as water-based emulsions by near-infrared Fourier transform Raman spectroscopy. J Raman Spectrosc 33(6):466–470

Jiang Q, Chen Y, Guo L, Fei T, Qi K (2016) Estimating soil organic carbon of cropland soil at different levels of soil moisture using VIS-NIR spectroscopy. Remote Sens 8:755

Karlen DL, Mausbach MJ, Doran JW, Cline RG, Harris RF, Schuman GE (1997) Soil quality: a concept, definition, and framework for evaluation (a guest editorial). Soil Sci Soc Am J 61(1):4–10

Kataoka H, Ueno Y, Makita M (1991) Analysis of o-phosphoamino acids in proteins by gas chromatography with flame photometric detection. Agric Biol Chem 55(6):1587–1592

Kuang B, Mouazen AM (2013) Non-biased prediction of soil organic carbon and total nitrogen with Vis–NIR spectroscopy, as affected by soil moisture content and texture. Biosyst Eng 114(3):249–258

Liu YL, Jiang QH, Shi TZ, Fei T, Wang JJ, Liu GL, Chen YY (2014) Prediction of total nitrogen in cropland soil at different levels of soil moisture with Vis/NIR spectroscopy. Acta Agric Scand Sect B 64(3):267–281

Liu JB, Han JC, Zhang Y, Wang HY, Kong H, Shi L (2018) Prediction of soil organic carbon with different parent materials development using visible-near infrared spectroscopy. Spectrochim Acta Part A 204:33–39

Morellos A, Pantazi XE, Moshou D, Alexandridis T, Whetton R, Tziotzios G, Wiebensohn J, Bill R, Mouazenb AM (2016) Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using Vis-NIR spectroscopy. Biosyst Eng 152:104–116

Mortimore JL, Marshall LJR, Almond MJ, Hollins P, Matthews W (2004) Analysis of red and yellow ochre samples from clearwell caves and Çatalhöyük by vibrational spectroscopy and other techniques. Spectrochim Acta A Mol Biomol Spectrosc 60(5):1179–1188

Mouazen AM, Kuang B, Baerdemaeker JD, Ramon H (2010) Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. Geoderma 158(1):23–31

Nawar S, Buddenbaum H, Hill J, Kozak J, Mouazen AM (2016) Estimating the soil clay content and organic matter by means of different calibration methods of Vis-NIR diffuse reflectance spectroscopy. Soil Tillage Res 155:510–522

Palma LD, Ferrantelli P, Merli C, Petrucci E, Pitzolu I (2007) Influence of soil organic matter on copper extraction from contaminated soil. J Soil Contam 16(3):13

Peng XT, Shi TZ, Song AH, Chen YY, Gao WX (2014) Estimating soil organic carbon using VIS/NIR spectroscopy with SVMR and SPA methods. Remote Sens 6:2699–2717

Prs V, Marchão RL, Brunet D, Becquer T (2012) The potential of nir spectroscopy to predict soil texture and mineralogy in cerrado latosols. Eur J Soil Sci 63(5):743–753

Qiao T, Ren J, Craigie C, Zabalza J, Maltin C, Marshall S (2015) Quantitative prediction of beef quality using visible and NIR spectroscopy with large data samples under industry conditions. J Appl Spectrosc 82(1):137–144

Quraishi MZ, Mouazen AM (2013) A prototype sensor for the assessment of soil bulk density. Soil Tillage Res 134(8):97–110

Rodionov A, Pätzold S, Welp G, Pallares RC, Damerow L, Amelung W (2014) Sensing of soil organic carbon using visible and near-infrared spectroscopy at variable moisture and surface roughness. Sci Soil Sci Soc Am J 78(3):949

Rossel RAV, Walvoort DJJ, Mcbratney AB, Janik LJ, Skjemstad JO (2006) Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131(1):59–75

Rossel RAV, Jeon YS, Odeh IOA, Mcbratney AB (2008) Using a legacy soil sample to develop a mid-ir spectral library. Soil Res 46(1):1–16

Rossel RAV, Behrens T, Ben-Dor E, Brown DJ, Demattê JAM, Shepherd KD (2016) A global spectral library to characterize the world's soil. Earth Sci Rev 155:198–230

Ruberto L, Dias R, Balbo AL, Vazquez SC, Hernandez EA, Cormack WPM (2010) Influence of nutrients addition and bioaugmentation

on the hydrocarbon biodegradation of a chronically contaminated Antarctic soil. J Appl Microbiol 106(4):1101–1110

Sarathjith MC, Das BS, Wani SP, Sahrawat KL (2014) Dependency measures for assessing the covariation of spectrally active and inactive soil properties in diffuse reflectance spectroscopy. Soil Sci Soc Am J 78(5):1522–1530

Schoell A, Zou Y, Huebner D, Urquhart SG, Schmidt T, Fink R et al (2005) A comparison of fine structures in high-resolution x-ray-absorption spectra of various condensed organic molecules. J Chem Phys 123(4):45

Shepherd KD, Walsh MG (2002) Development of reflectance spectral libraries for characterization of soil properties. Soil Sci Soc Am J 66(3):988–998

Sun WC, Xia Z, Sun XJ, Sun YL, Yi C (2018) Predicting nickel concentration in soil using reflectance spectroscopy associated with organic matter and clay minerals. Geoderma 327:25–35

Todorova M, Atanassova S, Sitaula B, Apturachim D, Valkova P, Dermendgieva D (2018) Application of nirs as a rapid and alternative method for prediction of heavy metals content in soil. Transbound Emerg Dis 65(1):S32–S37

Vasques GM, Grunwald S, Sickman JO (2008) Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. Geoderma 146(1):14–25

Wang JJ, Cui LJ, Gao WX, Shi TZ, Chen YY, Gao Y (2014) Prediction of low heavy metal concentrations in agricultural soils using visible and near-infrared reflectance spectroscopy. Geoderma 216(4):1–9

Wu YZ, Chen J, Ji JF, Gong P, Liao QL, Tian QJ, Ma HR (2007) A mechanism study of reflectance spectroscopy for investigating heavy metals in soils. Soil Sci Soc Am J 71(3):918–926

Xu SX, Zhao YC, Wang MY, Shi XZ (2018) Comparison of multivariate methods for estimating selected soil properties from intact soil cores of paddy fields by Vis-NIR spectroscopy. Geoderma 310:29–43

Yaseen M, Malhi SS (2009) Differential growth performance of 15 wheat genotypes for grain yield and phosphorus uptake on a low phosphorus soil without and with applied phosphorus fertilizer. J Plant Nutr 32(6):1015–1043

Zornoza R, Guerrero C, Mataix-Solera J, Scow KM, 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