

# ESTIMATION OF SOIL NICKEL USING VISIBLE AND NEAR-INFRARED SPECTRASCOPY BASED ON SPECTRAL CLASSIFICATION

Weichao Sun<sup>1,2</sup>, Xia Zhang<sup>1</sup>, Honglei Lin<sup>1,2</sup>

1. Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, No. 20A, Datun Road, Chaoyang District, Beijing, 100101, China. Emails: [sunwc@radi.ac.cn](mailto:sunwc@radi.ac.cn); [zhangxia@radi.ac.cn](mailto:zhangxia@radi.ac.cn);
2. University of Chinese Academy of Sciences, No. 19A, Yuquan Street, Shijingshan District, Beijing 100049, China. Email: [linhoml@163.com](mailto:linhoml@163.com)

**KEY WORDS:** Soil nickel; VNIRS; Spectral classification; PLSR

**ABSTRACT:** Soil heavy metal pollution is an increasingly severe environmental problem. Efficient investigation of soil contamination with heavy metals is the premise of soil remediation. Visible and near-infrared reflectance spectroscopy (VNIRS) has been widely used in the assessment of soil contaminant elements due to its rapidity and convenience. The relation between spectrally active soil characteristics (organic matter, iron oxide and clay minerals) and heavy metals is the mechanism of reflectance spectroscopy for assessment of soil heavy metals. With different spectrally active soil characteristics, soil reflectance spectra exhibit distinctive curve forms. In consideration of various soil reflectance curve forms, spectral clustering was applied to explore the potential of classification in soil nickel (Ni) estimation. In this study, 74 soil samples were collected from Hunan Province, China and their reflectance spectra were employed to estimate Ni concentration by partial least square regression (PLSR). Savitzky-Golay (SG) smoothing method was used to reduce the noise caused by spectral measurement, and the spectra were clustered into two groups, group 1 and group 2, by k-means method based on squared Euclidean distance. Root mean square error of prediction (RMSEP), coefficient of determination ( $R^2$ ) and residual prediction deviation (RPD) were adopted to evaluate the prediction accuracy. The result shows that the accuracy in group 1 was improved by clustering with  $R^2$  increased from 0.32 to 0.59, RPD increased from 1.21 to 1.56, while the prediction accuracy in group 2 was decreased with RPD dropped from 1.21 to 0.76. Preliminary analysis shows that the decrement for soil samples in group 2 was caused by the weak absorption feature of clay minerals. Future research may focus on the effect of soil constituents on absorption of Ni on soil. The result indicates the significance of considering soil reflectance curve forms in soil Nickel estimation using VNIRS.

## 1. INTRODUCTION

Soil provides fundamental natural resources for the survival of plants, animals and human race. Soil heavy metal pollution has become a serious environment issue and anthropogenic activities contribute significantly to environment contamination (Tchounwou et al., 2012). Due to the resistance to decomposition, Soil heavy metals can accumulate in the food chain. Chronic lower level intakes of toxic elements have adverse effects on human race and other animals. It was found that soil nickel (Ni) exposure was significantly associated with liver and lung cancer (Chen et al., 2015a). Additionally, soil Ni contamination may threaten food security as well. Consequently, rapid investigation of soil Ni pollution on its pollution level and spatial distribution is essential to human health and economic development.

The conventional method for investigating soil contaminant elements is based on field samplings and subsequent chemical analyses in the laboratory, followed by geo-statistical interpolation (Steiger et al., 1996). However, this method is time-consuming and expensive resulting from extensive soil samplings and chemical analyses. In addition, the conventional investigation can only provide information at limited points and can't

describe the dynamic evolution of contaminant elements over large areas due to its spatial and temporal limitations (Liu et al., 2010).

Soil reflectance is the comprehensive behaviors of its physical and chemical properties. Soil visible and near-infrared reflectance spectroscopy (VNIRS) has been widely employed in soil science for more than two decades (Viscarra Rossel et al., 2006). Compared with the conventional method, VNIRS is cost-effective, non-destructive and has better spatial and temporal continuities (Bellon-Maurel et al., 2010). Studies have found that the inter-correlation between contaminant elements and spectrally active soil components is the mechanism of estimating contaminant elements with VNIRS (Kooistra et al., 2001; Wu et al., 2007).

Soils were developed from different parent materials at different climatic zones with diverse precipitation and terrain, which result in a variety of distinct types with different composition of soil physical and chemical properties. Soil constituents, including organic matter, soil moisture, particle size, iron oxides and soil mineralogy, were further modified by the intensity and mode of cultivations. Soil reflectance is a cumulative property which derives from inherent spectral behavior of the heterogeneous combination of mineral, organic and fluid matter that comprises mineral soils (Stoner and Baumgardner, 1981). Consequently, soil reflectance curves may exhibit distinct curve forms for different soil types.

Stoner and Baumgardner (1981) collected 485 soil samples from 39 states of the United States and Brazil and identified five distinct soil reflectance curve forms from reflectance spectra ranging from 520nm to 2320nm by curve shape and the presence or absence of absorption bands. (Huang and Liu, 1995) collected 33 samples of main types of soils in southern China and classified the spectra into three categories, including flat type, slope type and steep type, according to soil spectral characters in the region of 360nm to 2500nm. The steep type is similar to the iron-affected curve form identified by Stoner. Soil reflectance spectra classification was adopted in interpreting soil diffuse reflectance spectra and estimating soil properties including organic carbon, clay content and organic matter concentration, which turns out to be effective in improving estimation accuracy (Rossel and Behrens, 2010; Shi et al., 2014b). However, soil reflectance spectral classification is rarely reported being employed in assessment of soil contamination elements with the VNIRS.

In this study, spectral clustering based on squared Euclidean distance was employed to explore the possibility of spectral classification in prediction of soil nickel with the VNIRS using PLSR. Soil nickel was used as an example in this study due to its relatively continuity in concentration.

## **2. MATERIALS AND SAMPLING**

### **2.1 Study Area and Sampling**

The study was conducted in Qingjiang Village which is located in Chenzhou City, Hunan Province, southern China. Qingjiang lead/zinc mine, one of the six biggest mines in the basin of the Dong River, is situated in Qingjiang Village. Soil sampling was carried out in Qingjiang Village and 83 topsoil samples (depth 0-20cm) were collected. The soil samples were collected in the area of 113° 17' 17.76"E to 113° 17' 36.07"E and 25°45'39.29"N to 25°46'1.13"N at the altitude of 335m to 451m..

The collected topsoil samples were air-dried at 20°C and ground into fine particles. Concentration of Ni for soil samples were determined using acid digestion method and measured by flame atomic absorption spectrometry.

### **2.2 Spectral measurement and pre-processing**

Soil spectra were measured by a portable spectrometer PSR-3500 (Spectral Evolution Inc., USA) which covers a spectral range of 350-2500 nm and offers spectral resolutions of 3.5 nm at 700 nm, 10 nm at 1500 nm and 7 nm at

2100 nm. Spectral intervals were 1.5 nm at 700 nm, 3.8 nm at 1500 nm and 2.5 nm at 2100 nm. A 50 W halogen lamp at an angle of 30° from nadir was mounted at 60cm above the center of the samples as stable light source in a dark room. Every sample was measured 5 times, and the average spectral was calculated for further processing.

In general, it is believed that wavelengths in the fringe of visible and near-infrared (VNIR) region for the spectrometer have relatively low signal-to-noise ratio. Wavelengths at the intervals of 350nm to 399nm and 2400nm to 2500nm were therefore removed so as to reduce noise. Among the 83 soil reflectance spectra, there were 9 curves with reflectance more than or close to 100% in several bands due to measurement error. The 9 reflectance spectra were therefore removed as well.

The remaining 74 soil reflectance spectra are shown in figure 2. It's evident that wavelengths in the VNIR region of 400-2400nm are affected by noises produced in the phase of spectral measurement. Savitzky-Golay (SG) smoothing is an effective spectral pre-processing method and has been used in spectral noise reduction, especially for the random noise. Therefore, a quadratic polynomial based on 14 points was employed to smooth the spectra so as to reduce the effect of noises.

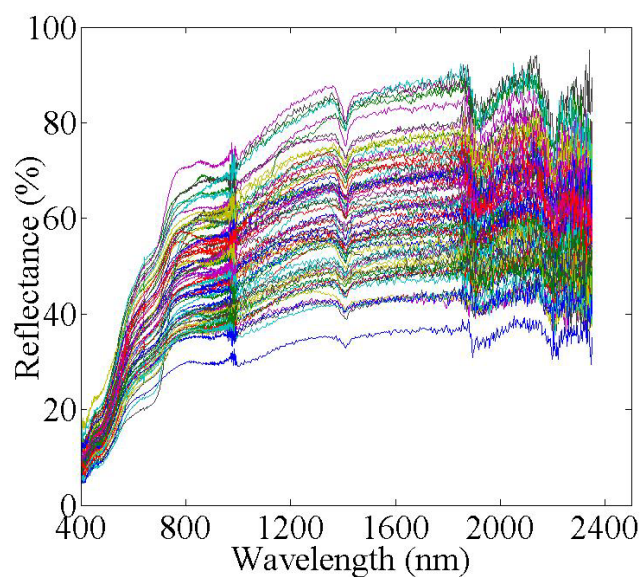


Figure 1. Reflectance Spectra of the 74 Soil Samples

### 2.3 Soil Reflectance Spectral Classification

For dry and ground soil, soil reflectance in VNIR region is related to organic matter, iron oxide, clay mineral and parent material. Studies indicate that iron oxide is the main factor that influence soil reflectance spectrum in southern China (Huang and Liu, 1995). It was found that reflectance curve forms of the four main soil groups in southern China including latosol, red earth, yellow earth and paddy soil over range of 360-2500nm can be categorized into two distinctive reflectance curve patterns based on curve shape and character of spectral absorption bands (Wu and Wang, 1991).

It can be seen from figure 2 that spectra with high reflectance have strong absorption at 1400 nm, 1900 nm and 2200 nm, and the low reflectance spectra have weak absorption at the three bands. Therefore, the number of classification was determined as two. The 74 soil reflectance spectra were then classified into two categories by k-means clustering method according to squared Euclidean distance.

### 2.4 Model Construction and Validation

Partial least square regression (PLSR) is a mainstream method in quantitatively deriving information from

reflectance spectra and the most frequently used method in estimation of soil contaminant elements with VNIRS. PLSR has the ability to handle data that are highly collinear and situations in which the number of variables considerably exceeds the number of available samples (Shao et al., 2010). Therefore, PLSR was adopted in model calibration.

Leave-one-out cross-validation was used to determine the number of components. To select the optimal PLSR components (PCs), the root mean square error of validation (RMSECV) was calculated. The model with the lowest RMSECV was selected. RMSECV is expressed as equation (1):

$$RMSECV = \sqrt{\sum (C_m - C_p)^2 / N_c} \quad (1)$$

Where  $C_m$  is the value measured by the chemical analysis,  $C_p$  is the value predicted by the model,  $N_c$  is the number of samples in the calibration set.

With reference to previous studies (Chen et al., 2015b; Shi et al., 2014a), three parameters, including coefficient of determination ( $R^2$ ), root mean square error of prediction (RMSEP) and residual prediction deviation (RPD), were adopted to evaluate the prediction accuracy. The RMSEP is defined as follows:

$$RMSEP = \sqrt{\sum (C_m - C_p)^2 / N_p} \quad (2)$$

Where  $C_m$  and  $C_p$  are the same as equation (1),  $N_p$  is the number of sample in the validation set.  $R^2 \leq 0.5$  or  $RPD \leq 1.4$  indicates useless models; medium quality models with  $0.5 < R^2 \leq 0.8$  or  $1.4 < RPD \leq 2.0$ ;  $R^2 > 0.8$  or  $RPD > 2.0$  indicate excellent models.

### 3. RESULT

The 74 soil spectra were then classified into two groups, group 1 and group 2, by k-means clustering method based on squared Euclidean distance. There are 38 soil reflectance spectra in group1 and 36 spectra in group 2. The two groups are present in figure 2.

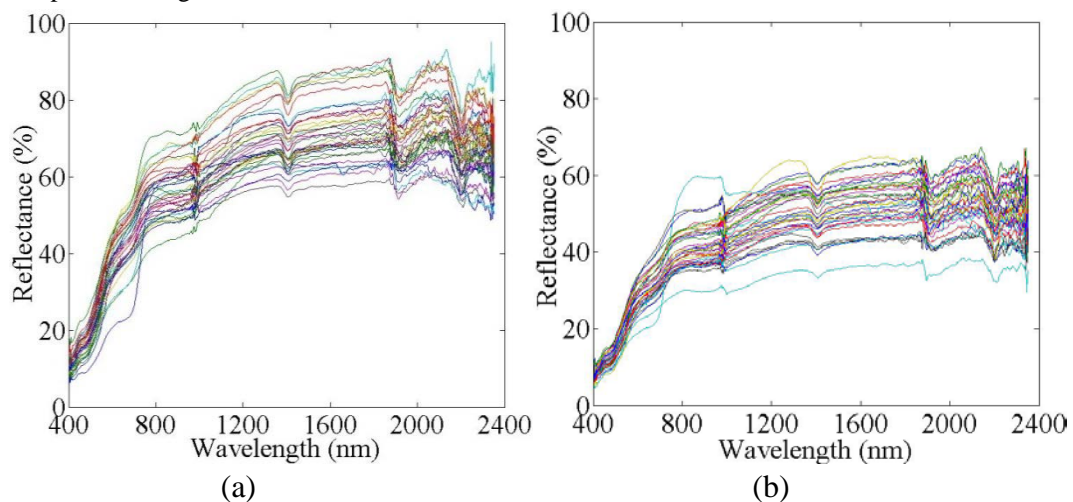


Figure 2. Result of k-means clustering based on squared Euclidean distance. (a) is group1, and (b) is group2.

Soil spectral values and soil Ni concentrations in the calibration set of group 1 and group 2 were used to build PLSR models, respectively. For comparison, the 74 soil reflectance spectra without classifications were divided into calibration set and validation set, and then used to construct estimation model.

The results of the estimation for group 1 and group 2 were present in table 1. Coefficients of determination ( $R^2$ ) less than 0.20 are not listed in the tables. It can be seen from Table 1 that the prediction accuracy of Ni has been largely improved by classification in group 1. RPD and  $R^2$  of Ni increased from 1.21 to 1.56 and 0.32 to 0.59. However, the accuracy in group 2 reveals nearly the contrary result that the accuracy of Ni decreased dramatically with classification.

Table 1. Prediction Result of Soil Ni by PLSR Models

	PCs	RMSEP	RPD	$R^2$
<b>Samples without classification</b>	3	6.89	1.21	0.32
<b>Samples in group 1</b>	8	7.55	1.55	0.59
<b>Samples in group 2</b>	4	3.95	0.76	

It can be seen from figure 2 that spectra in group 1 exhibit high reflectance with strong absorption at bands related to spectrally active soil characteristics, and spectra show low reflectance with weak absorption at the bands in group 2. The mechanism of soil contaminant elements estimation using VNIRS is based on the correlation among spectrally active soil characteristics, including organic matter, iron oxide and clay minerals, and the harmful elements. The weak absorption at wavelengths of spectrally active soil characteristic, especially at 1400 nm, 1900 nm and 2200 nm which are caused by clay minerals. The presence of clay minerals provides more sites for the adsorption of Ni on soil. The decrement of prediction accuracy of Ni in group 2 may relate to the weak absorption feature of clay minerals.

#### 4. CONCLUSION AND DISCUSSION

Reflectance of dry soil is a cumulative behavior of soil physical and chemical properties. In addition to different mode and intensity of land use, environment pollutions have profound impact on soil components as well. Consequently, it's essential to classify soil VNIRS before putting the spectra and concentrations of contaminant elements into regression analysis. In consideration of different soil reflectance curve forms in the VNIR region, soil spectral clustering based on squared Euclidean distance was adopted to explore the possibility of spectral classification in predicting soil nickel with VNIRS. The result shows that the accuracy in group 1 was improved, while the accuracy in group 2 was decreased due to the classification. The mechanism of soil Ni estimation using VNIRS relies on the correlation between Ni and soil spectrally active constituents mainly including organic matter, iron oxide and clay minerals. Preliminary analysis shows that the decrement for soil samples in group 2 was caused by the weak absorption feature of clay minerals. Future research may focus on the effect of soil constituents on absorption of Ni on soil. A clearly understanding of the effect will promote the estimation of soil Nickel using VNIRS.

#### REFERENCE

- Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.-M., McBratney, A., 2010. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *Trac-Trends in Analytical Chemistry* 29(9), 1073-1081.
- Chen, K., Liao, Q.L., Ma, Z.W., Jin, Y., Hua, M., Bi, J., Huang, L., 2015a. Association of soil arsenic and nickel exposure with cancer mortality rates, a town-scale ecological study in Suzhou, China. *Environmental Science and Pollution Research* 22(7), 5395-5404.
- Chen, T., Chang, Q., Clevers, J.G., Kooistra, L., 2015b. Rapid identification of soil cadmium pollution risk at regional scale based on visible and near-infrared spectroscopy. *Environmental pollution* 206, 217-226.
- Huang, Y., Liu, T., 1995. Spectral characteristic of main types of soils in southern China and soil classification

- (in Chinese). *ACTA PEDOLOGICA SINICA* 32(01), 58-68.
- Kooistra, L., Wehrens, R., Leuven, R.S.E.W., Buydens, L.M.C., 2001. Possibilities of visible–near-infrared spectroscopy for the assessment of soil contamination in river floodplains. *Analytica Chimica Acta* 446(1–2), 97-105.
- Liu, Y., Chen, H., Wu, G., Wu, X., 2010. Feasibility of estimating heavy metal concentrations in *Phragmites australis* using laboratory-based hyperspectral data-A case study along Le'an River, China. *International Journal of Applied Earth Observation and Geoinformation* 12, S166-S170.
- Rossel, R.A.V., Behrens, T., 2010. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma* 158(1-2), 46-54.
- Shao, X., Bian, X., Liu, J., Zhang, M., Cai, W., 2010. Multivariate calibration methods in near infrared spectroscopic analysis. *Analytical Methods* 2(11), 1662-1666.
- Shi, T., Chen, Y., Liu, Y., Wu, G., 2014a. Visible and near-infrared reflectance spectroscopy-an alternative for monitoring soil contamination by heavy metals. *Journal of hazardous materials* 265, 166-176.
- Shi, Z., Wang, Q., Peng, J., Ji, W., Liu, H., Li, X., Viscarra Rossel, R.A., 2014b. Development of a national VNIR soil-spectral library for soil classification and prediction of organic matter concentrations. *Science China-Earth Sciences* 57(7), 1671-1680.
- Steiger, B.V., Webster, R., Schulin, R., Lehmann, R., 1996. Mapping heavy metals in polluted soil by disjunctive kriging. *Environmental pollution* 94(2), 205-215.
- Stoner, E.R., Baumgardner, M.F., 1981. CHARACTERISTIC VARIATIONS IN REFLECTANCE OF SURFACE SOILS. *Soil Science Society of America Journal* 45(6), 1161-1165.
- Tchounwou, P.B., Yedjou, C.G., Patlolla, A.K., Sutton, D.J., 2012. Heavy metal toxicity and the environment. *Exs* 101, 133-164.
- 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(1-2), 59-75.
- Wu, H., Wang, R., 1991. Soil spectral characteristic and its quantitative nalysis in soil classification (in Chinese). *ACTA PEDOLOGICA SINICA* 28(02), 177-185.
- Wu, Y., Chen, J., Ji, J., Gong, P., Liao, Q., Tian, Q., Ma, H., 2007. A Mechanism Study of Reflectance Spectroscopy for Investigating Heavy Metals in Soils. *Soil Science Society of America Journal* 71(3), 918.