Supplementary Material

Hyperspectral-based inversion of heavy metal content in the soil of coal mining areas

Le Hou ¹, Xinju Li ¹⁺, Fang Li ¹,²⁺

¹College of Resources and Environment, Shandong Agricultural University, Taian, 271000, China.
²College of Economics and Management, Shandong Agricultural University, Taian, 271000, China.
*Corresponding author
Xinju Li
Email address: lxj0911@126.com
Tel: (+86) 13505388080
Fang Li
Email address: 906709947@qq.com
Tel: (+86) 1836616621

7 Pages; 0 Table; 7 Figures

Physical basis of hyperspectral inversion

Based on the continuous spectrum of soil hyperspectral resolution obtained in the visible, near-infrared, mid-infrared, and thermal infrared ranges, combined with ancillary analysis of physicochemical properties data, sensitivity bands and diagnostic features were identified, thus establishing a soil concentration information inversion prediction model (Wu et al., 2007).

Although there is no direct characteristic bands in vis-NIR region for heavy metal, the heavy metal contents are closely related to soil properties that have direct responses on vis-NIR spectroscopy, such as soil organic matter, clay minerals and iron-manganese oxide (He et al., 2015). Cd mainly reflects characteristic soil reflectance spectra through the influence of iron oxide components; Cu content inversion mainly involves the sensitive band of clay minerals in soil; the inversion of Ni content is related to the characteristic band of organic matter; the inversion of Pb content requires the simultaneous consideration of the characteristic bands of iron oxides and clay minerals; the inversion of Cr content needs to take into account the
effects of many factors, such as soil organic matter, moisture, and so on (Ming-Xing et al., 2011).

**Operation and effect of pretreatment method**

Proper pretreatment method can effectively highlight the spectral characteristic band of soil, and make the reflection peak and absorption valley of the spectral curve more prominent. Pretreatment can enhance the correlation between soil spectra and heavy metal content, thereby improving the prediction accuracy of the model.

[Savitzky-Golay](https://en.wikipedia.org/wiki/Savitzky%E2%80%93Golay_smoothing) is a smoothing method used to de-noise spectral curves and eliminate errors potentially caused by baseline drift, tilt, etc (Savitzky and Golay, 1964). The smoothing method is used to reduce random noise and improve the signal-to-noise ratio of the data (Gorry, 1990). SG smoothing refers to selecting multiple points before and after the smoothed point for fitting or averaging to determine the optimal value of the point. All the transformations were implemented using MATLAB2014a software. The specific process is as follows: SG smoothing takes 2m+1 consecutive points as a window in the spectral range. Polynomials are used in a least-squares fitting routine of the spectral window (independent variable is point $i$, $i=0, \pm 1, \pm 2, ..., \pm m$) to obtain the corresponding polynomial coefficient. The polynomial coefficient in turn is used to calculate the smoothing value of the central wavelength point ($i = 0$) in the window. The window moves over the full-wave spectral range to calculate the SG smoothing spectrum of the original spectrum.

Supplemental Fig. S1 The original spectrum
Multiplicative scatter correction is used to separate the physical light scattering information and the chemical light absorption information in the scattering medium spectrum, and then to eliminate the difference of the physical scattering information between different spectra, that is, to correct the scattering information in different sample spectra to the same level (Macdougall et al., 1985). The spectral data obtained after the correction can effectively eliminate the influence of scattering and enhance the spectral absorption information related to the content of the components. Multiplicative scatter correction is a common spectral data preprocessing technique used to eliminate interference information from spectral data and to enhance useful absorption information in spectral images (Wang et al., 2014). The goal of MSC is to obtain ideal spectral data by correcting every scattering phenomenon in the spectrum. All the transformations were implemented using MATLAB2014a software. The specific process is as follows:

Calculating the average spectrum that needs to be corrected,

$$\bar{A} = \frac{\sum_{j=1}^{M} A_j}{M}$$  \hspace{1cm} (1)

Use the average spectral data to calculate the regression coefficient,

$$A_j = \bar{A}m_j + b_j$$  \hspace{1cm} (2)

Calculating MSC corrected spectra by using regression coefficients,

$$\tilde{A}_j = \frac{(A_j-b_j)}{m_j}$$  \hspace{1cm} (3)

$A_j (j=1,2, \ldots, M)$ is the original measured spectrum of each sample, $\bar{A}$ is the average
spectrum of all samples, $m_j$ and $b_j$ are regression coefficients, $\bar{A}_j$ is the corrected spectrum after each sample passes through the MSC.

Supplemental Fig. S3 The original spectrum

Supplemental Fig. S4 After the MSC pretreatment

Supplemental Fig. S5 After the first differential pretreatment
Hyperspectral data acquisition

It is difficult to obtain reflectance spectral data in the field. The composition of farmland soil is complex, and its organic composition, temperature and pH will affect the adsorption characteristics of heavy metals. Reflectance spectral data can be affected by many factors, such as surface conditions (soil moisture, vegetation cover, etc.), atmospheric absorption and light conditions, etc. (Ji et al., 2015). Current technology cannot fully support to measure spectral data in the field.

If this data was to be obtained directly in the field, the spectrum curve will be disorderly. The reflection peak, absorption valley and characteristic absorption band would not be clear in the spectral curve. And the correlation between the spectral data and heavy metal content of soils would be weakened (Conforti et al., 2015). If this data was to be obtained directly in the field, the ground would have to be free of plants/organic debris. Best data is obtained from or on bare, flat soil.

Characteristic band information

The most significant influence of organic matter on soil spectral curve occurs at 770nm. 850nm is the characteristic absorption band of iron oxides. 2200nm is the characteristic absorption band of clay minerals. Their absorption features around 770, 850 and 2200 nm can be distinguished and the depth of absorption valley or absorption peak increases as concentration of heavy metals increases.
Supplemental Fig. S7 Characteristic spectral bands of heavy metals with different concentrations

When the concentration of Ni increased, the correlation between Ni and OM enhanced, and the depth of absorption peak at characteristic band 770nm increased. Iron oxides play an important role in the prediction of soil heavy metals using reflectance spectra (Wu et al., 2006). Due to the low Cd content, the correlation between Cd and iron oxides is lower than other heavy metals. The spectral feature band is not prominent, which affects the accuracy of modeling. After SG or a multiplicative scatter correction (MSC) processing, the spectral absorption characteristic is more obvious.

References


Ji, W., R.A.V. Rossel, and Z. Shi. 2015. Accounting for the effects of water and the
environment on proximally sensed vis–NIR soil spectra and their calibrations.

Macdougall, D., H. Martens, and P. Geladi. 1985. Linearization and
Spectrosc. 39(3): 491–500.

Zhu. 2011. Hyperspectral reflectance models for retrieving heavy metal
content: application in the archaeological soil. J. Infrared Millim. Waves 30(30):
109–114.

Savitzky, A., and M.J.E. Golay. 1964. Smoothing and Differentiation of Data by

regions on near infrared spectroscopy calibration results. Guang Pu Xue Yu
Guang Pu Fen XI 34(9): 2387–2390.

reflectance spectroscopy for measuring soil heavy metal content as a quick

Study of Reflectance Spectroscopy for Investigating Heavy Metals in Soils. Soil