Application of PLSR in Spectral Retrieval Model of Cadmium Content

To cite this article: Jinbao Liu and Dan Yang 2018 IOP Conf. Ser.: Earth Environ. Sci. 170 032157

View the article online for updates and enhancements.

Related content
- Study on Hyperspectral Characteristics and Estimation Model of Soil Mercury Content
  Jinbao Liu, Zhenyu Dong, Zenghui Sun et al.
- AS TO CADMIUM IN THE SUN
  C. E. St. John and Charlotte E. Moore
- Thermoreflectance study. II. Inter-band transitions in Pb1−x Cd x Se alloys
  A M Abdula, M I Baleva and M H Maksimov
Application of PLSR in Spectral Retrieval Model of Cadmium Content

Jinbao Liu1, 2, 3, 4, * and Dan Yang3

1 Institute of Land Engineering and Technology, Shaanxi Provincial Land Engineering Construction Group Co., Ltd., Xi'an 710075, China
2 Key Laboratory of Degraded and Unused Land Consolidation Engineering, the Ministry of Land and Resources, Xi'an 710075, China
3 Shanxi Provincial Land Engineering Construction Group Co., Ltd, Xi'an 710075, China
4 Shaanxi Provincial Land Consolidation Engineering Technology Research Center, Xi'an 710075, China

*Corresponding author e-mail: liujinbao1212@163.com

Abstract. In this study, 44 soil samples collected from the country of Fufeng, Yangling, and Wugong of Shaanxi Province. Using the ASD FieldSpec HR (350–2500 nm) spectroscopy collected. The original reflectance spectra were processed by NOR, MSC and SNV. The first deviation, the second deviation and the reciprocal logarithm of reflectivity were used respectively. The optimal hyperspectral estimation model of Cd was established by partial least squares regression. Chemical analysis showed that there was a serious Cd pollution phenomenon in the study area, and the Cd content approached the critical value. The results of the model show that: (1) The reflectance spectra of the reflectance spectra significantly increase the signal-to-noise ratio of the reflectance spectrum after being transformed by NOR, MSC and SNV. Combined with the differential transformation, it helps to improve the information of heavy metal elements in soils. Significantly improve model stability and predictive power. (2) The modeling accuracy of the optimal model for Cd elemental spectra established by PLSR method is 0.5451, 0.9912 and 0.6182, respectively. (3) The optimal estimation model established by different elements with different treatment methods has better stability and higher accuracy, and can quickly predict the Cd concentration in heavy metals in this area.

1. Introduction

As the most precious natural resource on which mankind depends, soils have a buffer and purifying effect on environmental pollutants [1, 2]. In recent years, with the continuous development of industrialization and urbanization, the content of heavy metals in soil rapidly increases. Heavy metals are easily enriched and corrosive, seriously affecting crop yields and indirectly causing serious harm to humans [3, 4]. At present, the determination of heavy metal content in soils is mainly done by field fixed-point sampling, and the content of different heavy metal elements is obtained based on laboratory chemical experiments. The traditional method of determination of heavy metals is costly...
and inefficient. Different elements require different chemical treatments and can’t meet the needs of monitoring heavy metal pollution in soils in large areas. Hyperspectral technology is widely used in the prediction of heavy metal content due to its advantages of abundant information, saving time and not destroying the structure of samples.

2. Materials and methods

2.1. Soil sample collection
The study area is located in Fufeng, Yangling, and Wugong, Shaanxi Province. Soil types are loamy soil. Soil samples were collected using the "S" patterning method. Reflectance spectra were measured using a high-density reflection probe to remove soil contaminants, for a total of 44 soil samples. Samples were air-dried, to impurities after mixing 200g soil samples and 100 mesh sieve for determination of indoor heavy metal content. Heavy metals were assayed by inductively coupled plasma-atomic emission spectrometry-mass spectrometry (ICP-MS, Agilent 7700), including cadmium (Cd). Heavy metal elements of statistical characteristics in Table 1.

| Samples | Quantity | Max (mg kg⁻¹) | Min (mg kg⁻¹) | Mean (mg kg⁻¹) | Standard (mg kg⁻¹) | CV         |
|---------|----------|---------------|---------------|----------------|-------------------|------------|
| Cd      | 44       | 0.513         | 0.323         | 0.398          | 0.0419            | 0.105260773|

2.2. Spectral data determination
Soil reflectance spectroscopy was performed in the field using a high-density reflection probe equipped with an ASD Field Spec HR spectrometer. Spectrometer wavelength range of 350 ~ 2500 nm, the sampling bandwidth of 1.3 nm (350~1000 nm) and 2 nm (1000~2500 nm), resampling interval of 1 nm. High-density reflection probe can effectively avoid the impact of soil stray light, but also be able to eliminate the impact of the weather. The front 2 cm of vision area can avoid the stone in the soil, crop roots.

2.3. Spectral data processing

2.3.1. Breakpoint repair. Soil samples will introduce different degrees of errors during the collection, processing and analysis, which will affect the later data analysis and modeling accuracy. The Mahalanobis distance is based on multivariate normal distribution, taking covariance, mean and variance factors into account, and can comprehensively reflect the comprehensive index of soil samples [5, 6]. Therefore, this study used Mahalanobis distance method to detect outliers of soil properties and spectral data.

2.3.2. Spectral differential transformation. First-order differential, second-order differential, and reciprocal logarithm transformation of the original reflectance spectra were performed. In addition to directly analyzing the spectral reflectance of the soil, three transformations were performed to find the response regions of different heavy metal elements. First-order and second-order differential transformations can increase the correlation between reflectivity and heavy metal elements while eliminating or limiting the effects of a partially linear or near-linear background. The first-order differential the formula is:

\[
\rho'(\lambda_i) = \frac{\rho(\lambda_{i+1}) - \rho(\lambda_{i-1})}{\lambda_{i}}
\]  

(1)

\[
\rho''(\lambda_i) = \frac{\rho'(\lambda_{i+1}) - \rho'(\lambda_{i-1})}{\lambda_{i}}
\]  

(2)
2.4. Modeling and inspection

2.4.1. Model validation. Based on Mahalanobis distance, TQ Analyst randomly divides all datasets into modeling sets and verification sets, and uses PLS to establish the prediction model. The model results were validated by the coefficient of determination $R^2$ and root mean square error RMSE. Calculated as follows:

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

(3)

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2}
\]

(4)

3. Results and analysis

3.1. Correlation analysis
In order to further analyze the correlation between heavy metal elements and spectral reflectance, the first order differential, second order differential, and logarithmic conversion of reflectance spectra were performed to analyze the correlations between heavy metal elements and heavy metal elements. Compared with the original reflection spectrum, the correlation was significantly improved after the differential transformation of the reflection spectrum. From the differential transformation method, the reflection spectrum of the first-order differential best, second-order differential. Among them, Cd of first order differential of reflection spectrum were -0.54, reaching a significant correlation. After the soil reflectance spectra are differentially transformed, the absorption characteristics can be remarkably highlighted, and the correlation between heavy metal elements and reflectivity can be enhanced.

![Figure 1. Correlation coefficient reflection spectra and their different transformations](image)

3.2. Establishment and Test of Soil PH Prediction Model
The original reflectance spectrum was used as the control group. The spectral data were preprocessed by NOR, MSC and SNV. The first derivative, the second derivative and the logarithm transform of
reflectivity were all Savitzky-Golay smoothing. The PLSR was used to establish the corresponding Heavy metal estimation models were tested using the determination coefficient and root mean square error. The larger the coefficient of modeling decision, the smaller the root mean square error is, indicating that the stability of the estimation model is better. The larger the prediction coefficient is, the smaller the root mean square error is, indicating that the model prediction ability is stronger. At the same time, in order to avoid over-fitting the model, the smaller the dimension of the model's independent variables, the better. Table 3 lists all the heavy metal elements modeling results, Figure 2 is the best modeling results of heavy metals and the predicted value of the scatter plot.

![Figure 2. The Regression Coefficient of Model Based Partial Least Squares](image)

The lowest accuracy of cadmium modeling and prediction was $R^2_c = 0.5451$, $R^2_v = 0.4579$, RMSEC = 0.0328 and RMSEP = 0.0437. Although the modeling and prediction coefficients of cadmium are not high, the root mean square error is the smallest, which may be related to the less cadmium in the soil in this study. According to the modeling and forecasting results, it can be concluded that the modeling effect is good and the forecasting effect is not necessarily the best.

| Elements | Pre-treatment | Calibration | Validation |
|----------|---------------|-------------|-------------|
| Cd       |               | $R^2_c$     | $R^2_v$     | RMSEC       | RMSEP       |
| S+C      | 0.5403        | 0.329       | 3           | 0.1689      | 0.0535      |
| C+FD     | 0.5153        | 0.335       | 1           | 0.3959      | 0.0449      |
| C+SD     | 0.5358        | 0.330       | 1           | 0.4578      | 0.0437      |
| C+LOG    | 0.1869        | 0.384       | 1           | 0.1865      | 0.0501      |
| NOR+S    | 0.5707        | 0.321       | 2           | 0.2682      | 0.0520      |
| NOR+FD   | 0.5295        | 0.332       | 1           | 0.4213      | 0.0443      |
| **NOR+SD** | **0.5451** | **0.0328** | **1**       | **0.4579**  | **0.0437**  |
| NOR+LOG  | 0.5564        | 0.325       | 2           | 0.2037      | 0.0535      |
| MSC+S    | 0.4242        | 0.354       | 1           | 0.0041      | 0.0537      |
| MSC+FD   | 0.9971        | 0.00299     | 10          | 0.1781      | 0.0627      |
| MSC+SD   | 0.5541        | 0.326       | 1           | 0.4289      | 0.0446      |
| MSC+LOG  | 0.4123        | 0.356       | 1           | 0.0265      | 0.0533      |
| SNV+S    | 0.6378        | 0.301       | 4           | 0.0791      | 0.0536      |
| SNV+FD   | 0.5349        | 0.330       | 1           | 0.3701      | 0.0455      |
| SNV+SD   | 0.5502        | 0.327       | 1           | 0.4263      | 0.0447      |
| SNV+LOG  | 0.5882        | 0.316       | 4           | 0.0959      | 0.0540      |
4. Conclusion
In the present study, the original reflectance spectrum was processed by NOR, MSC and SNV. The first order differential, the second order differential and the reciprocal logarithm of reflectivity were used respectively. Combined with partial least squares regression, the optimal hyperspectral estimation model of Cd. By comparing the effects of different pre-treatment methods on the establishment of soil heavy metal spectral inversion models, the following conclusions can be drawn:

(1) The cadmium content in this study area is seriously exceeded, and the Cd content is close to the critical value, posing a serious threat to the normal growth of plants and animals in the area.

(2) The first, second derivative and reciprocal logarithm transformation of the reflectance spectra after the treatment of NOR, MSC and SNV respectively effectively reduce the influence of external factors such as soil particle size and surface scattering on the spectrum. At the same time, differential transformation can help to improve the correlation between heavy metal elements and reflectance spectra in soil, and the combination band with higher correlation can significantly improve the stability and prediction ability of the model.

(3) The spectral optimal model established by PLSR has good stability and high precision, which can quickly predict the Cd content of heavy metal elements. According to the modeling and forecasting results, it can be concluded that the modeling effect is good and the forecasting effect is not necessarily the best.

Acknowledgments
This work was financially supported by Shaanxi Province Land Engineering Construction Group Research Project "Moisture Characteristics and Transport Law of Mixed Media in feldspathic sandstone and aeolian sandy soil in Mu Us Sandy Land (DJNY2017-23). We greatfully acknowledge the Institute of Land Engineering and Technology, Shanxi Provincial Land Engineering Construction Group, Xi'an, China.

References
[1] WU Y Z, CHEN J, WU X M, et al. Possibilities of reflectance spectroscopy for the assessment of contaminant elements in suburban soils [J]. Applied Geochemistry, 2005, 20: 1051-1059.
[2] AL- ABBAS A H, SWAIN P H, BAUMGARDNER M F. Relating organic matter and clay content to the multispectral radiance of soils[J]. Soil Science, 1972, 114(6): 477 - 485.
[3] KOOISTRA L, WEHRENS R, LEUVEN R S E W, et al. Possibilities of visible-near-infrared spectroscopy for the assessment of soil contamination in river floodplains [J]. Analytica Chimica Acta, 2001, 446: 97-105.
[4] GRZEGORZ S, GREGORY W M, TOMASZ I S, et al. Near and mid-infrared diffuse reflectance spectroscopy for measuring soil metal content[J]. Journal of Environment Quality, 2004, 33: 2056-2069.
[5] Han CL, Liu JB, Fu XM, et al. 2017. Hyperspectrum Characteristics and Modeling of Organic Matter Content in Soils Developed from Trachytic Tephra. Chinese Journal of Soil Science, 2017, 47(1): 39 - 44.
[6] Savitsky, A., & Golay, J. E. (1964). Smoothing and differentiation of data by simplified least square procedure. Analytical Chemistry, 36, 1627–1639.