Study on Quantitative Retrieval of Soil Organic Matter Based on Bare Soil Spectrum

Jinliang Wang*, Qiming Qin*, Heng Dong*, Chao Chen*, Qingye Meng*
*Institute of Remote Sensing and Geographic Information System, Peking University,
Beijing, 100871, China
Corresponding author: Jinliang Wang. Tel: +86-010-62754430. Fax: +86-010-62751962.
E-mail address: findlay.wjl@gmail.com

Abstract. As we know, soil organic matter (SOM) is very important to the plant, especially for the crop. How to quickly access the information of SOM has become a major problem faced by many researchers. Now hyperspectral remote sensing technology became more and more useful in the retrieval of SOM. In this paper, we used different forms of bands combination, and partial least square regression to get the SOM retrieve model. The result shows that using the 2nd derivative of reflectance model can get the best SOM retrieval result.

Keywords: Soil Organic Matter; Hyperspectral Data; Regression Analysis; Partial Least Square

1.   Introduction

Soil is a foundation material of agricultural activity, and the data of some important factors related with production are assurance of agriculture. Soil organic matter (SOM) is the most important constituent part of soil and can supply the nutrient element which plant needed. How to quickly access the information of soil nutrients has become a major problem faced by the researchers. Traditional chemical detection methods are accurate, but time-consuming, laborious and expensive.

In recent years, the interaction between SOM content and soil spectral data forms the basis for SOM spectral models. Recent studies discovered the relationships between absorption feature parameters in the spectral region of 0.4–2.5mm wavelength and soil properties. Thus many forms of indices and models were built to retrieve the SOM. At the same time, some papers found that partial least squares regression (PLSR) is the most common visible and near-infrared calibration method. PLSR has many advantages, such as its simplicity, robustness,

predictability, precision, and clearly quantitative explanations. Using the PLSR method to get the retrieve model becomes more and more popular.

Partial least squares regression (PLS regression) is a statistical method that bears some relation to principal components regression; instead of finding hyperplanes of minimum variance between the response and independent variables, it finds a linear regression model by projecting the predicted variables and the observable variables to a new space. Because both the X and Y data are projected to new spaces, the PLS family of methods are known as bilinear factor models.

PLS is used to find the fundamental relations between two matrices (X and Y), i.e. a latent variable approach to modeling the covariance structures in these two spaces. A PLS model will try to find the multidimensional direction in the X space that explains the maximum multidimensional variance direction in the Y space. PLS regression is particularly suited when the matrix of predictors has more variables than observations, and when there is multicollinearity among X values. By contrast, standard regression will fail in these cases.[6]

High spectral resolution spectrum is in the range of 10-2λ, it can display how substances affect the spectral. The high spectral have more subtle information. Hyperspectral remote sensing technology can get data quickly, and also can monitor a region many times. Retrieval of SOM with Hyperspectral remote sensing attracts a large number of researchers engaged in [1-2].

Krishnan [3] found that four types of soil reflectance spectra had no spectral absorption features related to SOM in the range of 800–2400 nm wavelength and the correlation coefficients between the SOM content and spectral reflectance in 624 nm and 564 nm was higher.

Bocheng [7] found that spectral characteristics of brown-humid soil were studied to estimate the SOM content, indicating that the SOM content was significantly correlated to the original reflectance, logarithm for reciprocal reflectance in 447 nm, and the first differentiation of logarithm for reciprocal reflectance in the bands of 516 nm and 615 nm.

Our research focuses on different forms to retrieve the SOM using PLSR methods. The detail information of the module is as below:

2. Methodology

In our study, band combination、deviation of arch、spectral data、1st /2nd derivative of reflectance data and the forms of reflectance transformation are used for SOM inversion.

First, the measured soil spectral data was smoothed by 5 nm window, and the bands absorbed by water vapour are erased.

Then we analyzed the correlations between measured soil spectral data、other forms spectral data and SOM in each wavelength. Ten most relevant bands were selected as alternative band for inversion models. After evaluating the accuracy of different retrieval models, the most accurate stable quantitative inversion models of

SOM were selected. Finally, the best bands and model for SOM inversion were gotten using partial least square regression analysis.

Soil spectral data were measured by ASD FieldSpec Pro spectrometer. The measured spectral range is 350-2500 nm, spectral resolution of is 2 nm. 45 soil samples representative of soil with different organic matter content were collected at YuCheng Observation Site which is in Shan Dong Province China. YuCheng mainly crop is wheat.

The soil there is mainly fluvo-aquic soil. We chose 45 different bare soil sites located around the test area. Soil samples were tested in the laboratory using potassium dichromate-volumetric method. The maximum of SOM test results' is 18.35 g/kg, minimum is 10.00 g/kg. The distribution of result basically shows normal distribution. More details are as follows:
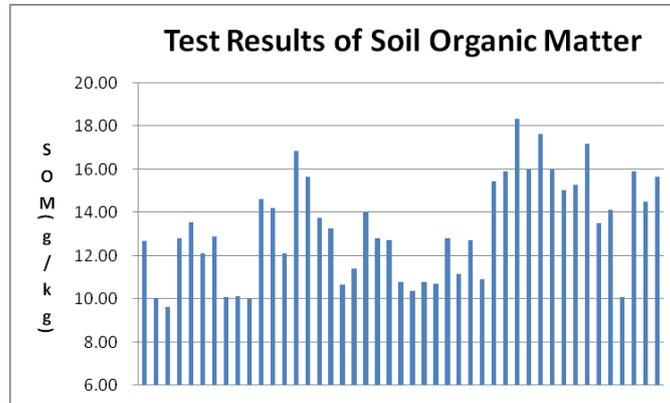

Figure 1. Test Results of SOM

First we smooth the soil reflectance collected from experiment using 5nm window. Second we delete the bands absorbed by the water vapour:1350-1460，1796-1970，2470-2500nm respective.

Then we compare two older models to predict the SOM. The first one used bands combination [3] (Eq.1), and the second one used deviation of arch [4] (Eq.2). Both models did the linear regression with the collected SOM directly, and the two models used the logarithm soil spectrum at 560 and 620 nm.

The models' equations and the goodness are as follows :

$$SOM = K \cdot \frac{d \log(1/r_{0.62})/dr_{0.62}}{d \log(1/r_{0.56})/dr_{0.56}} \qquad \text{Eq. 1}$$

$$\log SOM = 0.1271 - 1.083 \log(\frac{r_{0.55} + r_{0.65}}{2} - r_{0.60}) \qquad \text{Eq. 2}$$
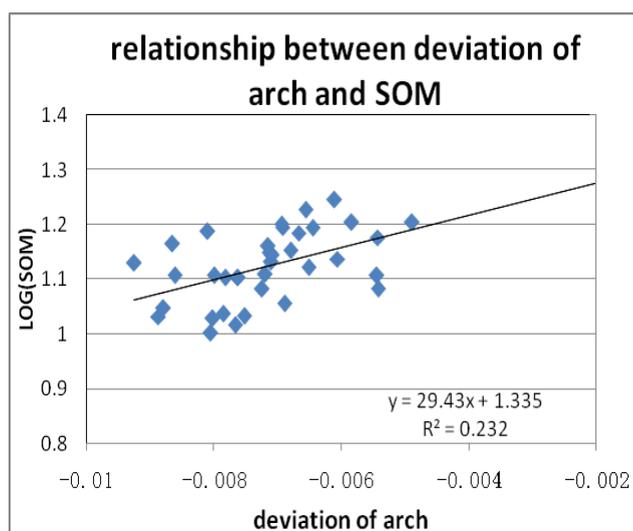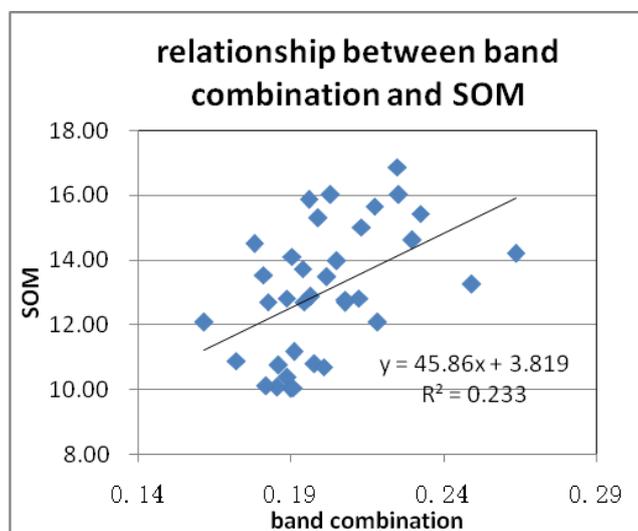
Figure 2. relationship between band combination、deviation of arch and SOM

In the second way, we use four different forms of reflectance transformation、four forms of 1st derivative of reflectance data、four forms of 2nd derivative of reflectance data to do the linear regression with SOM band by band. The four different forms of reflectance transformation include the multiplicative inverse of the reflectance、the logarithm of the reflectance、the multiplicative inverse of

logarithm reflectance、the square root of the reflectance. We also use the same four transformation using the 1st/2nd derivative of reflectance to do the linear regression with SOM.

Using the above method, we get 10 most sensitive bands with SOM in each transformation form. Moreover in our research, we find 10 stable sensitive bands with SOM when using 2nd derivative of reflectance data to do the linear regression. The 10 most sensitive bands are the same using the four forms of 2nd derivative of reflectance data. These 10 bands are 779，783，588，659，631，449，397，399，708，774nm.

Finally, the partial least square regression was used to get the model for predicting the SOM using the best 10 stable sensitive 2nd derivative of reflectance bands.

3.   Preliminary results

The goodness of the 17 model forms of fit statistics is of the table below.

Table 1. Goodness of the fit statistics

	R2	Std. deviation	MSE	RMSE
bands combination	0.233	/	0.00035	0.0190
deviation of arch	0.232	/	0.00329	0.0573
soil reflectance	0.070	2.315	5.119	2.263
logarithm of the reflectance	0.065	2.321	5.146	2.269
multiplicative inverse of the reflectance	0.059	2.328	5.181	2.276
multiplicative inverse of logarithm reflectance	0.068	2.317	5.131	2.265
square root of the reflectance	0.068	2.317	5.132	2.265
1st derivative of the reflectance	0.543	1.622	2.514	1.586
logarithm of the1st derivative of reflectance	0.606	1.506	2.166	1.472
multiplicative inverse of the1st derivative of reflectance	0.554	1.603	2.457	1.567

multiplicative in-verse's logarithm of the 1st derivative of reflectance	0.545	1.619	2.506	1.583
square root of the 1st derivative of ref-lectance	0.533	1.640	2.571	1.604
2nd derivative of the reflectance	0.623	1.474	2.075	1.440
logarithm of the2nd derivative of reflectance	0.583	1.549	2.293	1.514
multiplicative in-verse of the2nd de-rivative of reflec-tance	0.532	1.641	2.574	1.604
multiplicative in-verse's logarithm of the 2nd derivative of reflectance	0.616	1.486	2.111	1.453
square root of the 2nd derivative of ref-lectance	0.607	1.505	2.164	1.471

R2 stands for the square of relationship. MSE stands for mean square error. RMSE stands for root-mean-square error.

$$MSE = \frac{1}{n}\sum_{1}^{n}(y - y_i)^2 \qquad \text{Eq. 3}$$

$$RMSE = \sqrt{MSE} \qquad \text{Eq. 4}$$

Where n present the number of the samples. y present the measure value. y_i present the model estimated value

The inversion model's error results show that the relationship using forms of 1st/2nd derivative of reflectance data is much better than forms of reflectance data. The relationship using forms of 2nd derivative of reflectance data is better than the forms of 1st derivative of reflectance data because the 2nd derivative of reflectance could better represent the soil spectrum, and is more stable than the 1st derivative of reflectance. The 2nd derivative of reflectance could be impacted by the original soil spectrum less.

The band combination and deviation of arch depends on the reflectance too much. And the sensitive bands are not stable. Sometimes the optimum wavelength is shifting, affect the model's stability.

In our research, the optimal bands which have the best fitness (R2=0.623) are the 2nd derivative of reflectance data at 779，783，588，659，631，449，397，399，708，774nm, and these optimal wavelengths for predicting the SOM are stable in different four transformation.

From the selected bands, we could find that 10 optimal bands located at visible and near infrared bands. This result is consistent with the previous research and papers.

The best model is below (R2 is 0. 623, Std. deviation is 1.474, RMSE is 1.440):

$$SOM = 13.58 + 89047.55 \times \ddot{r}_{779} + 51168.05 \times \ddot{r}_{783} - 72821.24 \times \ddot{r}_{588} + 67157.73 \times \ddot{r}_{659} + 116794.21 \times \ddot{r}_{631}$$
$$+ 90038.99 \times \ddot{r}_{449} + 23213.71 \times \ddot{r}_{397} - 31211.24 \times \ddot{r}_{399} + 105574.84 \times \ddot{r}_{708} - 115153.54 \times \ddot{r}_{774}$$

4.　Conclusion

In this study, we compare different forms to predict the SOM, and find 10 stable bands (779，783，588，659，631，449，397，399，708，774nm) using 2nd derivative of reflectance data to retrieve the SOM. The partial least square method can get the inversion model quickly and stable.

Soil spectral reflectance is the reflectance of soil inherent physical and chemical characteristics. Different types of soil have different physical and chemical properties, thus the soil spectrum is different. So the SOM reversion model's spatial and temporal applicability is the most needed problem to be investigated.

The further study will focus on the improvement and validation of this method considering more soil parameters (soil water content、different type of soil、mineral types of clay and the compositions of SOM) using different kinds of soil types.

Acknowledgment

References

[1]　　　Curran，P.J., G.M. Foody, K.Ya. Kondratyev, et al.，Remote Sensing of soil and Vegetation in the USSR, 1990, Taylor & Francis, New York

[2]   Baumgardner M F, Kristof S, Johannsen C J, et al. Effects of Organic Matter on the Multispectral Properties of Soils [J].Proc .Indiana Acad.Sci,1970(79):413~422.

[3]   KRISHNAN P,ALEXANDER J D,BUTLER B J,et al.Reflectance technique for predicting soil organic matter[J].Soil Sci.Soc.Am.J.,1980,44:1 282-1 285.

[4]   Zhou Qing , Study on the Hyperspectral Prediction Models of soil organic Matter and its Difference among soil types [D], Zhejiang University,2004

[5]   ZHANG Juan-Juan , YU Hua , QIAO Hong-Bo. Soil organic matter content estimation based on hyperspectral properties[J]. Chinese Journal of Eco-Agriculture, May 2012, 20(5): 566□ 572

[6]   Introduction   of   Partial   Least   Squares   Regression   From   Wikipedia http://en.wikipedia.org/wiki/Partial_least_squares_regression.

[7]   X. Bocheng, X. Xuzhang, W. Jihua, W. Guodong Spectral characteristics of brown-humid soil and estimation of soil organic matter content by soil reflectivity. Chinese Journal of Soil Science, 35 (4) (2004), pp. 391–395