

Study on Hyperspectral Estimation Model of Total Nitrogen Content in Soil of Shaanxi Province

Jinbao Liu^{1,3,4,5,*}, Zhenyu Dong², Xi Chen^{1,3,4,5}

¹Institute of Land Engineering and Technology, Shaanxi Provincial Land Engineering Construction Group Co., Ltd., Xi'an 710075, China

²College of Forestry, Shenyang Agricultural University, Shenyang, 110866, China

³Key Laboratory of Degraded and Unused Land Consolidation Engineering, the Ministry of Land and Resources, Xi'an 710075, China

⁴Shaanxi Provincial Land Engineering Construction Group Co., Ltd, Xi'an 710075, China

⁵Shaanxi Provincial Land Consolidation Engineering Technology Research Center, Xi'an 710075, China

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

Abstract. The development of hyperspectral remote sensing technology has been widely used in soil nutrient prediction. The soil is the representative soil type in Shaanxi Province. In this study, the soil total nitrogen content in Shaanxi soil was used as the research target, and the soil samples were measured by reflectance spectroscopy using ASD method. Pre-treatment, the first order differential, second order differential and reflectance logarithmic transformation of the reflected spectrum after pre-treatment, and the hyperspectral estimation model is established by using the least squares regression method and the principal component regression method. The results show that the correlation between the reflectance spectrum and the total nitrogen content of the soil is significantly improved. The correlation coefficient between the original reflectance and soil total nitrogen content is in the range of 350 ~ 2500nm. The correlation coefficient of soil total nitrogen content and first deviation of reflectance is more than 0.5 at 142nm, 1963nm, 2204nm and 2307nm, the second deviation has a significant positive correlation at 1114nm, 1470nm, 1967nm, 2372nm and 2402nm, respectively. After the reciprocal logarithmic transformation of the reflectance with the total nitrogen content of the correlation analysis found that the effect is not obvious. $R_c^2 = 0.7102$, $RMSEC = 0.0788$; $R_v^2 = 0.8480$, $RMSEP = 0.0663$, which can achieve the rapid prediction of the total nitrogen content in the region. The results show that the principal component regression model is the best.

1. Introduction

Nitrogen is an important nutrient element in the soil and is an important indicator of soil fertility and nitrogen nutrition. Traditional soil nitrogen determination methods include potassium dichromate - sulfuric acid digestion, perchloric acid - sulfuric acid digestion, semi - micro Kay's method and diffusion absorption method. Although the accuracy is high, the results are stable, but because of its



long measurement time, more supplies, it is difficult to obtain real-time nitrogen content in the field and other shortcomings affect the measurement efficiency. Hyperspectral remote sensing has the advantages of high spectral resolution, strong continuity of band, fine spectral information and so on, and has wide application prospect in rapid soil nitrogen content acquisition.

Studies at home and abroad show that the hyperspectral spectrum has made great progress in estimating the nitrogen content of different soil types. Lee and other studies have shown that soil nitrogen absorption band at 510nm, due to the lack of soil type of pride differences (Lee et al, 2001). Used the partial least squares method to establish the soil total nitrogen content estimation model based on spectral analysis. The results show that the correlation coefficient between the predicted value and the chemical measured value is above 0.90(Chang et al, 2002). Based on the partial least squares method, BP neural network and characteristic spectral index, the estimation model of total nitrogen content of five main soil types in middle and east China was established (Zhang et al, 2011). Used the re-sampling multiple stepwise regression model and the re-sampling partial least squares regression model to establish the estimation model of soil total nitrogen in Panjin wetland (Wang LW et al, 2016). Used the partial least squares regression method to establish the soil total nitrogen prediction model for 33 paddy soils in the Three Gorges Reservoir area (Xu et al, 2013). Using the method of multiple linear regression and partial least squares regression to establish the model of soil total nitrogen hyperspectral prediction in Jinghe county, Xinjiang (Li et al, 2017). Used the correlation analysis combined with the partial least squares regression method to establish the soil total nitrogen inversion model of the Huangmushan soil in Wuyi County, Shaanxi Province (Liu et al, 2015). Typical red soils in Fuzhou as the research goal, the use of stepwise multiple linear regression to establish the soil nitrogen hyperspectral inversion optimization model (Wu et al, 2013). Based on the principal component regression and the spectral characteristics of stepwise regression, the optimal prediction model of soil total nitrogen content was established (Lu Y L et al, 2010.)

2. Materials and methods

2.1. Soil sample collection

In this study, the main soil types were soil type in Fufeng County, Yangling County and Wugong County, Shaanxi Province. The soil samples were collected according to the "S" -shaped sampling method. The sampling depth was the thickness of the tillage layer, usually 0-30cm. A total of 44 soil samples were sampled. After the sample was air-dried, the samples were weighed and dried in 0.149mm 1.0 ~ 3.0g, semi-micro Kai-Kai method for the determination of total nitrogen, the statistical results shown in Table 1:

Table 1. Descriptive statistics of TN in soil samples

Indicator	Samples	Maximum	Minimum	Average	Standard error	Range
N	44	1.60	0.61	1.35	0.16	0.12

2.2. Spectral data determination

Contact probe for soil spectrum measurement by the external stray light interference less, you can get more accurate data. In this study, soil reflectance spectroscopy was performed in the field using a high density reflective probe equipped with an ASD Field Spec HR spectrometer. The wavelength range of the spectrometer is 350 ~ 2500 nm, the sampling bandwidth is 1.3 nm (350 ~ 1000 nm) and 2 nm (1000 ~ 2500 nm), and the sampling interval is 1 nm. High-density reflective probes can effectively avoid the effects of soil stray light and eliminate the effects of weather. 2cm front view area can avoid the soil in the stone tablets, crop roots, etc., see the parameters shown in Table 2.

Table 2. The technical spaces of high intensity contact probe

Length	10"(25.4 cm)
Quality	1.5 lbs
Voltage	12 ~ 18 VDC,6.5W
Type	Halogen bulb/1500h
Temperature	2901+/-10%K
Spot size	10mm

2.3. Spectral data processing

2.3.1. Data smooth. Use the ViewSpePro to remove the jump spectrum curve and calculate the average as the actual reflection spectrum of the soil sample. In order to eliminate the influence of scattering between soil samples, the average reflectance spectra (Normalization, NOR), Multiplication Scatter Correlation (MSC), Standard Normal Variation (SNV) treatment.

2.3.2. Spectral differential transformation. The original spectral spectrum was optimized by standard normal variable transformation, and the chromatogram information was purified. Differential transformations can improve the band resolution and sensitivity, greatly reducing the noise generated in different test backgrounds (Pu et al, 2000). In this study, the second deviation transformation of the first deviation and the reciprocal logarithm of the reflectance was given as follows: differential, second order differential, reflectance reciprocal logarithm, reflectance reciprocal logarithm logarithm,

$$\rho'(\lambda_i) = [\rho(\lambda_{i+1}) - \rho(\lambda_{i-1})] / \Delta\lambda \quad (1)$$

$$\rho''(\lambda_i) = [\rho'(\lambda_{i+1}) - \rho'(\lambda_{i-1})] / \Delta\lambda \quad (2)$$

$$\log\left(\frac{1}{\rho(\lambda_i)}\right) = -\log \rho(\lambda_i) \quad (3)$$

2.4. Modeling and inspection

2.4.1. Methods. Partial least squares regression (PLSR) is a new method of multivariate statistical data analysis, which solves the problem that the number of samples is less than the number of variables, and reduces the height between the variables. Linear correlation problem, PLSR in the spectral dimensionality while taking into account the role of the target variable matrix, effectively combined with the regression.

Principal component regression (PCR) is a data compression technique in multivariate statistics. The high correlation band is attributed to independent variables. The regression equation is established by selecting a small number of new variables and the internal test is used to prevent the model from over Fitting the phenomenon. In this study, the abnormal samples were removed by Mahalanobis distance method, 70% was selected as the modeling group and 30% as the verification group. The prediction model was validated by using the one-way cross validation. The model results are verified by the decision factor R² and the root mean square error RMSE.

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (5)$$

3. Results and analysis

3.1. Soil TN spectroscopy

Figure 1 shows the spectral curves of the total nitrogen content in 44 soil samples, and the trend of each spectral curve is approximately the same. Compared with different total nitrogen content of the spectral curve, the higher the total nitrogen content, the lower the spectral reflectance. In the range of 350 ~ 568nm, the reflectivity increases rapidly with the increase of wavelength, and the curve is steep. In the range of 535 ~ 781nm, the curve growth trend is slightly slowed down. 781 ~ 1350nm and 1495 ~ 797nm range, the reflectivity gradually increased, but the trend is relatively slow.

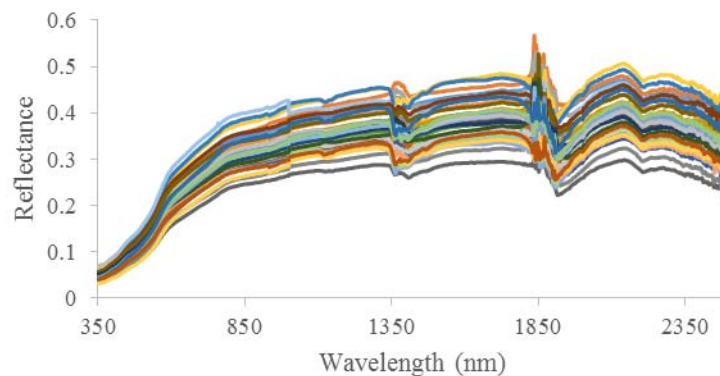


Figure 1. The characteristics of spectral curves of different TN content

3.2. Correlation analysis of soil TN and reflectance

The correlation between soil total nitrogen content and spectral reflectance, first deviation, second deviation, and logarithmic logarithm of reflectance is shown in Fig2. Compared with the original reflection spectrum, the reflection spectrum of the differential transformation, the correlation significantly improved. The original reflectance was negatively correlated with soil total nitrogen content in the range of 350-2500 nm, reaching a maximum at -0.23 in 2010 nm. There was a positive correlation between the total nitrogen and the first deviation of reflectance at 1428nm, 1982nm, 1963nm, 2204nm and 2307nm. The correlation coefficient was the highest at 1982 nm, which was 0.56. The correlation coefficients of soil total nitrogen and reflectance second deviation at 1114nm, 1470nm, 1967nm, 2372nm and 2402nm are all 0.5 or more, and the maximum value is 0.61 at 1967nm. After the reciprocal logarithmic transformation of the reflectance and the correlation analysis of the total nitrogen content, the effect is not obvious, only three water absorption peak correlation coefficient is larger.

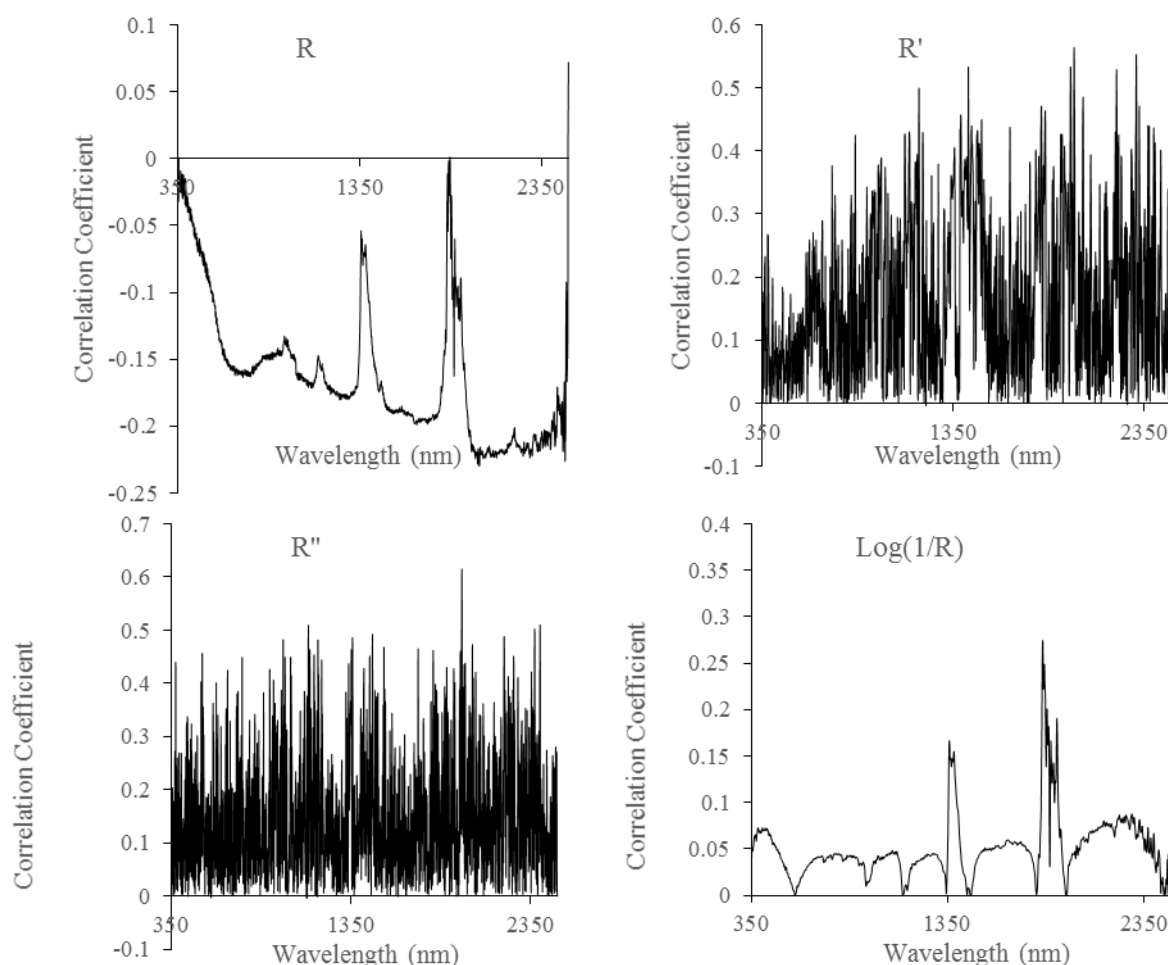


Figure 2. The Correlation of TN content based on different transformations

3.3. Establishment and Test of Soil TN Prediction Model

The partial least squares regression model and the principal component regression model were established with the original reflectance (Table 3), first and second deviation and reflectance reciprocal logarithmic correlation coefficient as independent variables and total nitrogen content as dependent variables. Comparing the results of the two modeling methods, we can conclude that the two models based on the reciprocal logarithm of reflectivity are the worst. In the PLSR model modeling group, the second-order differential model based on the reflectivity is the best, $R_c^2 = 0.7882$, $RMSEC = 0.0689$, but its prediction effect is poor, $R_v^2 = 0.4559$, $RMSEP = 0.112$. In contrast, based on the model established by 13 transformations, it can be concluded that the model with the second order differential after SNV transformation is the best, $R_c^2 = 0.6639$, $RMSEC = 0.0837$; $R_v^2 = 0.7443$, $RMSEP = 0.0920$. In the model group, the model was established with the first order differential after the MSC transformation, and $R_c^2 = 0.7427$ and $RMSEC = 0.0750$. In the prediction group, the best pre-treatment method is the best for the model obtained by combining the first order differential with the reflectance, and $R_v^2 = 0.8480$ and $RMSEP = 0.0663$. It is found that the correlation coefficient of the model is 0.7 or more and $RMSEC < 0.08$ after the first order differential transformation of the reflectivity in the modeling group, regardless of the pre-treatment. The correlation coefficient in the prediction group was above 0.79, $RMSEP < 0.09$. Modeling effect of a good model, the forecast is not necessarily the best. Therefore, the principal component regression model established by combining the first order derivative with NOR conversion is the best, $R_c^2 = 0.7102$, $RMSEC = 0.0788$; $R_v^2 = 0.8480$, $RMSEP = 0.0663$.

Based on the above-obtained optimal processing method, the measured and predicted values of the model established by partial least squares regression and principal component regression are shown in Fig3. After comparison, it can be concluded that the principal component regression model is better for the first order differential transformation after NOR pre-treatment of the reflectivity. The coefficient of decision and the coefficient of prediction are 0.7102 and 0.8480, respectively, which are larger than the coefficient of decision ($R_c^2 = 0.6639$, $R_v^2 = 0.7443$) of the partial least squares regression model based on SNV transform combined with second order differential. (RMSEC = 0.0837, RMSEP = 0.0920), and the mean square error of the model is 0.0788 and 0.0663, which are less than the optimal partial least squares model.

Table 3. The result of regression based on different pre-treatment methods

Method	Pre-treatment	Calibration		Validation		
		R_c^2	RMSEC	Pc	R_v^2	RMSEP
PLSR	S+C	0.4915	0.0975	7	0.4135	0.112
	C+FD	0.6136	0.0884	4	0.5882	0.0896
	C+SD	0.7882	0.0689	4	0.4559	0.112
	C+LOG	0.4913	0.0975	10	0.4177	0.116
	NOR+FD	0.5311	0.0949	5	0.6337	0.0992
	NOR+SD	0.6081	0.0889	7	0.6470	0.0981
	NOR+LOG	0.4356	0.101	9	0.4874	0.105
	MSC+FD	0.5096	0.0964	1	0.5901	0.107
	MSC+SD	0.5262	0.0952	7	0.2021	0.128
	MSC+LOG	0.4216	0.102	6	0.6666	0.0920
	SNV+FD	0.5249	0.0953	6	0.6370	0.0997
	SNV+SD	0.6639	0.0837	8	0.7443	0.0920
	SNV+LOG	0.4148	0.102	4	0.6165	0.0958
Method	Pre-treatment	Calibration		Validation		
		R_c^2	RMSEC	Pc	R_v^2	RMSEP
PCR	S+C	0.6010	0.0895	10	0.6150	0.0965
	C+FD	0.7237	0.0773	10	0.8292	0.0736
	C+SD	0.5967	0.0876	10	0.7106	0.0919
	C+LOG	0.4560	0.0971	10	0.4923	0.108
	NOR+FD	0.7102	0.0788	10	0.8480	0.0663
	NOR+SD	0.5997	0.0873	10	0.7076	0.0892
	NOR+LOG	0.5818	0.0887	10	0.5132	0.106
	MSC+FD	0.7427	0.0750	10	0.7962	0.0831
	MSC+SD	0.5932	0.0878	10	0.7215	0.0910
	MSC+LOG	0.5866	0.0884	10	0.6111	0.0973
	SNV+FD	0.7421	0.0751	10	0.8005	0.0824
	SNV+SD	0.5942	0.0878	10	0.7218	0.0909
	SNV+LOG	0.5873	0.0883	10	0.6106	0.0973

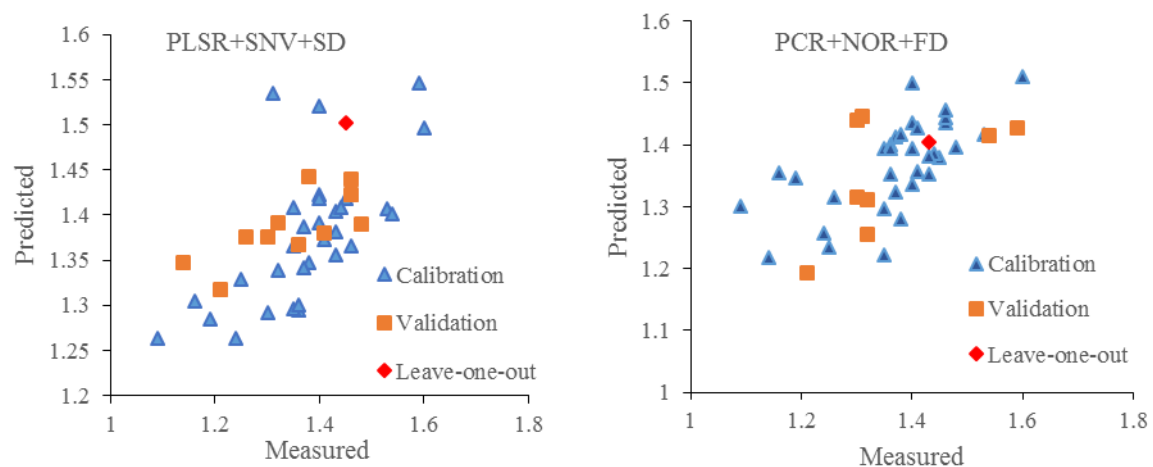


Figure 3. Comparison between measured and predicted of soil TN content

4. Conclusion

In this study, the total nitrogen content of soil in Shaanxi soil was hyperspectral inversion, and the original reflection spectra were pre-treated by NOR, MSC and SNV. The first deviation, second deviation, reflectance reciprocal logarithmic transformation combined with partial least. The hyperspectral inversion model of soil total nitrogen in Shaanxi soil was established by regression and principal component regression. By comparing the effects of different pre-treatment methods on the establishment of soil heavy metal spectral inversion model, the following conclusions are obtained:

Compared with the original reflection spectrum, the reflection spectrum of the differential transformation, the correlation significantly improved. There was a positive correlation between soil total nitrogen and first order differential of reflectance, and the correlation coefficient was the highest at 1982 nm, which was 0.56. There was a significant positive correlation between soil total nitrogen and second order differential of reflectance, the maximum value was 0.61 at 1967 nm.

The principal component regression model is better for the first deviation transformation after NOR pre-treatment of the reflectivity. The coefficient of decision and the coefficient of prediction are 0.7102 and 0.8480, respectively, which are larger than the coefficient of decision ($R_c^2 = 0.6639$, $R_v^2 = 0.7443$) of the partial least squares regression model based on SNV transform combined with second order differential. (RMSEC = 0.0837, RMSEP = 0.0920), and the mean square error of the model is 0.0788 and 0.0663, which are less than the optimal partial least squares model. Therefore, the principal component regression model established by NOR pre-treatment with reflectivity is the best. The model has good stability and high prediction accuracy, and can realize the rapid determination of soil total nitrogen content in this area.

Acknowledgments

This work was financially supported by National Key Technology Support Program fund. We greatly acknowledge the Institute of Land Engineering and Technology, Shaanxi Provincial Land Engineering Construction Group, Xi'an, China.

References

- [1] Lee W S, Mylavarapu R, S, Choe J S, et al. Study on soil properties and spectral characteristics in Florida[R]. American Society for Aerospace Education Paper, 2001.
- [2] Chang C, David Laird A. Near-infrared reflectance spectroscopic analysis of soil C and N [J]. Soil Science, 2002, 167:110 - 116.
- [3] Zhang J J, Tian Y C, Yao X, et al. Estimating Soil Total Nitrogen Content Based on Hyperspectral Analysis Technology. *Journal of Natural Resources*, 2011, 26 (5): 881 - 890.

- [4] Wang L W, Wei Y X. Estimating the total nitrogen and total phosphorus content of wetland soils using hyperspectral models. *Acta Ecologica Sinica*, 2016, 36(16): 5116 - 5125.
- [5] Xu L H, Xie D T, Wei C F, et al. Prediction of Total Nitrogen and Total Phosphorus Concentrations in Purple Soil Using Hyperspectral Data. *Spectroscopy and Spectral Analysis*, 2013, 33(3): 723 -727.
- [6] Li Y, Wang R H, Guan Y L. Prediction Analysis of Soil Total Nitrogen Content based on Hyperspectral. *Remote Sensing Technology and Application*, 2017, 32 (1): 173 - 179.
- [7] Liu X Y, Wang L, Chang Q R, et al. Prediction of total nitrogen and alkali hydrolysable bitrogen content in loess using hyperspectral data based on correlation analysis and partial least squares regression. *Chinese Journal of Applied Ecology*, 2015, 26(7): 2107 - 2114.
- [8] Wu M Z, Li X M, Sha J M. Spectral Inversion Models for Prediction of Red Soil Total Nitrogen Content in Subtropical Region. *Spectroscopy and Spectral Analysis*, 2013, 33(11): 3111 - 3115.
- [9] Lu Y L, Bai Y L, Wang L, et al. Determination for total nitrogen content in black soil using hyperspectral data. *Transactions of the CASE*, 2010, 26(1): 256 - 261.