IOP Conf. Series: Materials Science and Engineering **274** (2017) 012020 doi:10.1088/1757-899X/274/1/012020

Prediction of Soil pH Hyperspectral Spectrum in Guanzhong Area of Shaanxi Province Based on PLS

Jinbao Liu*, Yang Zhang, Huanyuan Wang, Jie Cheng, Wei Tong and Jing Wei

[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]Shaanxi 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. The soil pH of Fufeng County, Yangling County and Wugong County in Shaanxi Province was studied. The spectral reflectance was measured by ASD Field Spec HR portable terrain spectrum, and its spectral characteristics were analyzed. The first deviation of the original spectral reflectance of the soil, the second deviation, the logarithm of the reciprocal logarithm, the first order differential of the reciprocal logarithm and the second order differential of the reciprocal logarithm were used to establish the soil pH Spectral prediction model. The results showed that the correlation between the reflectance spectra after SNV pre-treatment and the soil pH was significantly improved. The optimal prediction model of soil pH established by partial least squares method was a prediction model based on the first order differential of the reciprocal logarithm of spectral reflectance. The principal component factor was 10, the decision coefficient Rc2 = 0.9959, the model root means square error RMSEC = 0.0076, the correction deviation SEC = 0.0077; the verification decision coefficient Rv2 = 0.9893, the predicted root mean square error RMSEP = 0.0157, The deviation of SEP = 0.0160, the model was stable, the fitting ability and the prediction ability were high, and the soil pH can be measured quickly.

1. Introduction

Soil pH was an important indicator of soil pH, soil alkalization often occurs in arid and semi-arid areas. The higher the pH value, the stronger the soil compaction, the worse the permeability (Li et al,1998; Guan and Liu,2001), the resulting plant can't grow normally, the ecological environment was fragile and so on, which was leading to the study area soil fertility degradation important reason One (Jiang et al,2007). In order to control soil alkalization more effectively, rational development and utilization of soil saline soils, the need for continuous monitoring of soil pH changes. The traditional soil pH was obtained by field sampling, and the soil pH was obtained by chemical analysis in the room. Although the precision was high, it was time-consuming and laborious to adapt to the monitoring of large area saline soil. With the continuous development of indoor hyperspectral technology, rapid prediction can be achieved by establishing a prediction model of soil pH and soil reflectance (Wang et al,2009; Qu et al,2009). The data acquisition method was simple and real-time monitoring whether the soil has been alkalized, and gradually become an important means to obtain soil pH index in the field, which will provide strong support for large area monitoring and evaluation of soil quality.

In recent years, some domestic and foreign scholars have applied spectroscopy on soil chemical properties in different regions of the quantitative research (Cloutis et al,1996; Dalai et al,1986; Reeves et l, 2002; Udelhoven et al,2003), salinization and other uses hyperspectral quantitative model of the first-order differential method to achieve the western Jilin Province Salt Alkaline soil effective prediction (Liu et al,2008). The used partial least squares method to realize the rapid inversion of soil pH in the Songnen Plain (Ma,2014). Multiple statistical analysis combined with multiple regression method to establish the Xinjiang Kaidu River Basin soil salinization hyperspectral model, obtained a good prediction effect (Li et al,2012). The PLSR-BP composite model to predict soil pH in Qitai County, Xinjiang (Wang et al,2014). Based on the above findings, researchers selected in the salinization of soil hyperspectral data pre-processing methods were more single modeling, methods were very different, but few studies of high-alkali soil spectral characteristics of Shaanxi Province The On the basis of the above studies, the soil was basified to OFF in Shaanxi Province of research objectives, and sampling carried out field distribution chamber spectroscopy, determining a correlation between soil pH and different spectral reflection characteristics transform differentiation binding using Partial least squares regression method was used to establish the quantitative hyperspectral model of alkaline soil in the study area. By comparing the six prediction models, the best prediction model of soil pH in Guan Zhong area of Shaanxi Province was established to realize the rapid detection of soil pH.

2. Materials and methods

2.1. Soil sample collection

The study area was located in Fufeng County, Yangling County, Wugong County, Shaanxi Province, the soil type was mainly soil. The soil samples were collected according to the "S" -shaped sampling method. The sampling depth was the thickness of the tillage layer, usually 0 - 30 cm. A total of 44 soil samples were collected. The samples were dried by natural drying and passed through a 2-mm hole. Each soil sample was divided into two parts, one for indoor pH meter determination; the other for spectral reflectance measurement, soil pH measurement results of the statistical characteristics in Table 1:

Table 1. Descriptive statistics of SOM in soil samples

indicator	Samples	Maximum	Minimum	Average	Standard error	Range
pH	44	8.38	7.47	8.12	0.18	7.47-8.38

2.2. Spectral data determination

Determination of Spectral Reflectance of Soil Samples in Fuping Analysis and Testing Center, Shanxi Institute of Land Engineering Technology. The sampling range was 350 ~ 2500 nm, the sampling bandwidth was 1.3 cm (350 ~ 1000 nm), 2 nm (1000 ~ 2000 nm), and the resampling interval was 1 nm. With a diameter of 8cm, 2 cm deep black dish filled with 2 mm hole sieve soil samples, with a straight ruler flattened surface, the spectral reflectance measurement (Han et al,2017). The petri dish was rotated twice during the measurement and rotated once after each measurement to determine the different positions on the same soil sample. Each side of a soil sample was subjected to one side of the whiteboard calibration.

2.3. Spectral data processing

2.3.1. breakpoint repair, Use View Spec Pro software to remove the anomalous spectrum and obtain the average reflectance as the spectral reflectance of the soil sample. Due to the differences in the energy response of the spectrometer, there was a breakpoint at 1000 nm. In this study, the average spectral curve was corrected by Splice Correlation. The effect was as follows:

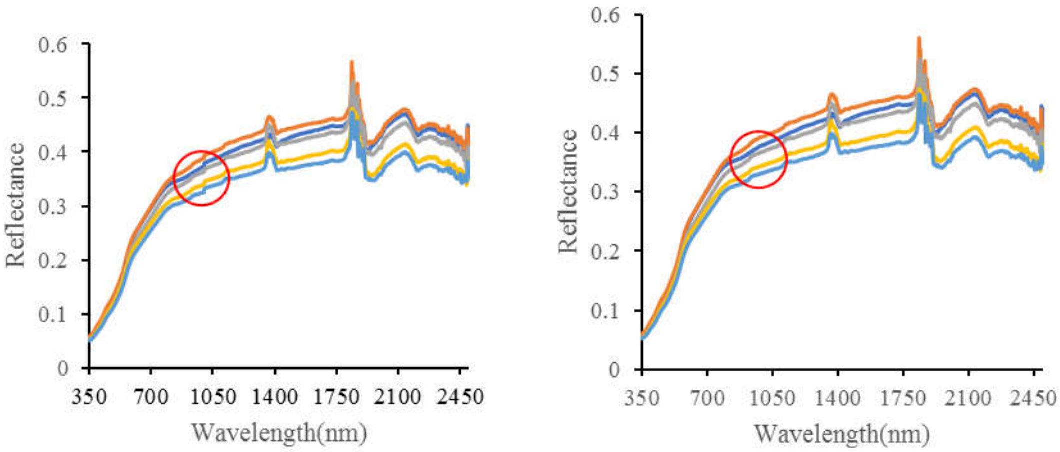

Figure 1. The repair of breakpoint 1000nm

2.3.2. Data was smooth, In order to eliminate the influence of solid particles, surface scattering and optical path variation on the reflection spectrum at different angles, the standard normal variable transformation of the spectral curve was carried out, and the Savitzky - Golay convolution was carried out by combining the Unscrambler 9.7. 2, smooth points: 9) (Savitaky and Goaly,1964), smooth spectral curve to remove the larger noise range, to retain the 400 ~ 2450nm band spectral information.

2.3.3. 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,

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

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

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

$$\log'\left(\frac{1}{\rho(\lambda_i)}\right) = -\log\frac{\rho'(\lambda_i)}{\log\rho(\lambda_i)} \tag{4}$$

$$\log''\left(\frac{1}{\rho(\lambda_i)}\right) = -\frac{\rho''(\lambda_i)\rho(\lambda_i)-\left[\rho'(\lambda_i)\right]^2}{\log\left[\rho(\lambda_i)\right]^2} \tag{5}$$

2.4. Modeling and inspection

2.4.1. Methods, 70% of soil samples were selected for modeling and 30% for model validation. Based on the soil pH of the model data as the dependent variable, the spectral reflectance and its mathematical transformation data as the independent variables, the spectral prediction model of soil pH was established by multiple linear regression, partial least squares regression and principal component regression. Using the ASD spectrometer ViewSpecPro software for five differential transformations, excel 2013 correlation analysis of soil pH and spectral reflectance, The Unscrambler 9.7 for data smoothing (Savitzky - Golay, quadratic polynomial, nine - point smoothing) and the establishment of partial least squares Regression model.

2.4.2. Model validation, the evaluation parameters used in the test include: correlation coefficient of calibration (Rc), prediction coefficient Rc^2, correlation coefficient of validation (Rv), prediction coefficient Rv^2, standard error of calibration (RMS), root mean square error of prediction (RMSEP), and the standard error of prediction (BEP), the standard error of prediction (SEP), the root mean square error of prediction (RMSEP) (Standard error of prediction, SEP), predictive bias (Bias). The Coefficient of determination (R^2) and the root mean square error (RMSE)were used to validate different modeling sets and verification sets. The larger the R^2, the smaller the RMSE, SEC, SEP, Bias, the more stable the model, the better the prediction. Calculated as follows:

$$R^2 = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \overline{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}_i\right)^2} \tag{6}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{7}$$

$$SEC = \sqrt{\frac{\sum_{i=1}^{m}\left(y_i - \hat{y}_i\right)^2}{m-1-k}} \tag{8}$$

$$SEP = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i - Bias\right)^2}{n-1}} \tag{9}$$

$$R_c = \sqrt{1-\frac{SEC^2\times(m-k-1)}{SD\times(m-1)}} \tag{10}$$

$$R_v = \sqrt{1 - \frac{SEP^2 \times (n-k-1)}{SD \times (n-1)}} \quad (11)$$

Where was the mean value of the observed values; m was the number of samples for the calibration set; n was the number of samples to be verified; SD was the standard deviation of the chemical reference; K was the calibration model in the dimension of the independent variable.

3. Results and analysis

3.1. Characteristics of soil pH spectroscopy

The soil reflectance spectral information contains spectral characteristics such as soil type, soil texture and soil moisture. Spectral measurements in the chamber not only enhance the intensity of the reflected light in laboratory spectral measurements, but also do not cause changes in the characteristic position. Soil samples were naturally dried and ground, and the effects of soil moisture and texture on the soil spectrum were eliminated. The reflectance characteristics of soil samples mainly reflected the characteristics of soil pH change. The reflectance spectra of the selected soil samples were shown in Fig. Each of the curves in Figure 1 represents a certain value of soil pH. In the range of 780 ~ 1790 nm, the reflectivity curve of most samples tends to be gentle. In the range of 1960 ~ 2500 nm, the reflectivity curve increases rapidly and then decrease. Due to the effect of OH- in soil minerals, there were three absorption bands with different absorption intensities in the vicinity of 1400 nm, 1900 nm and 2400 nm.

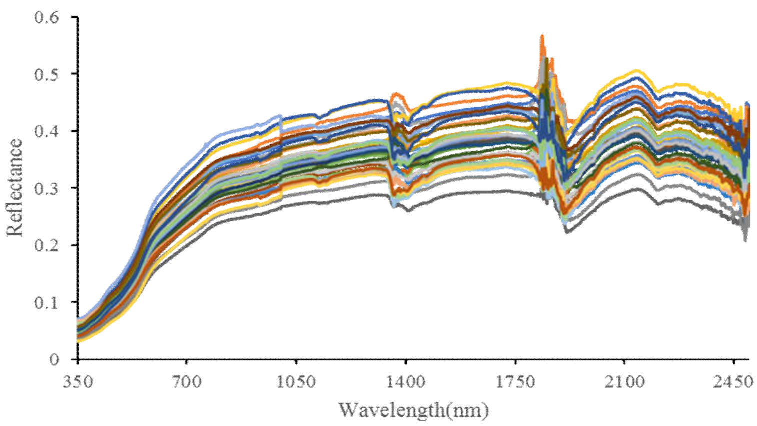

Figure 2. Reflectance curves of soil different in soil PH

3.2. Correlation analysis of soil pH and reflectance

Fig.3 shows the correlation between soil pH value and spectral reflectance and its five transformations in the range of 350 ~ 2500 nm, and the distribution of correlation coefficient at each wavelength was obtained. At both significant levels of 0.01 and 0.05. The maximum correlation coefficient between the original reflectance and soil pH was -0.1872, and the maximum correlation coefficient of the first order differential transformation with soil pH reached the maximum at 1825 nm, which was 0.5453. The correlation between the original spectrum and the soil pH was analyzed by the second order differential treatment. The maximum value of the correlation coefficient reached the maximum at 1824 nm, which was 0.5924.

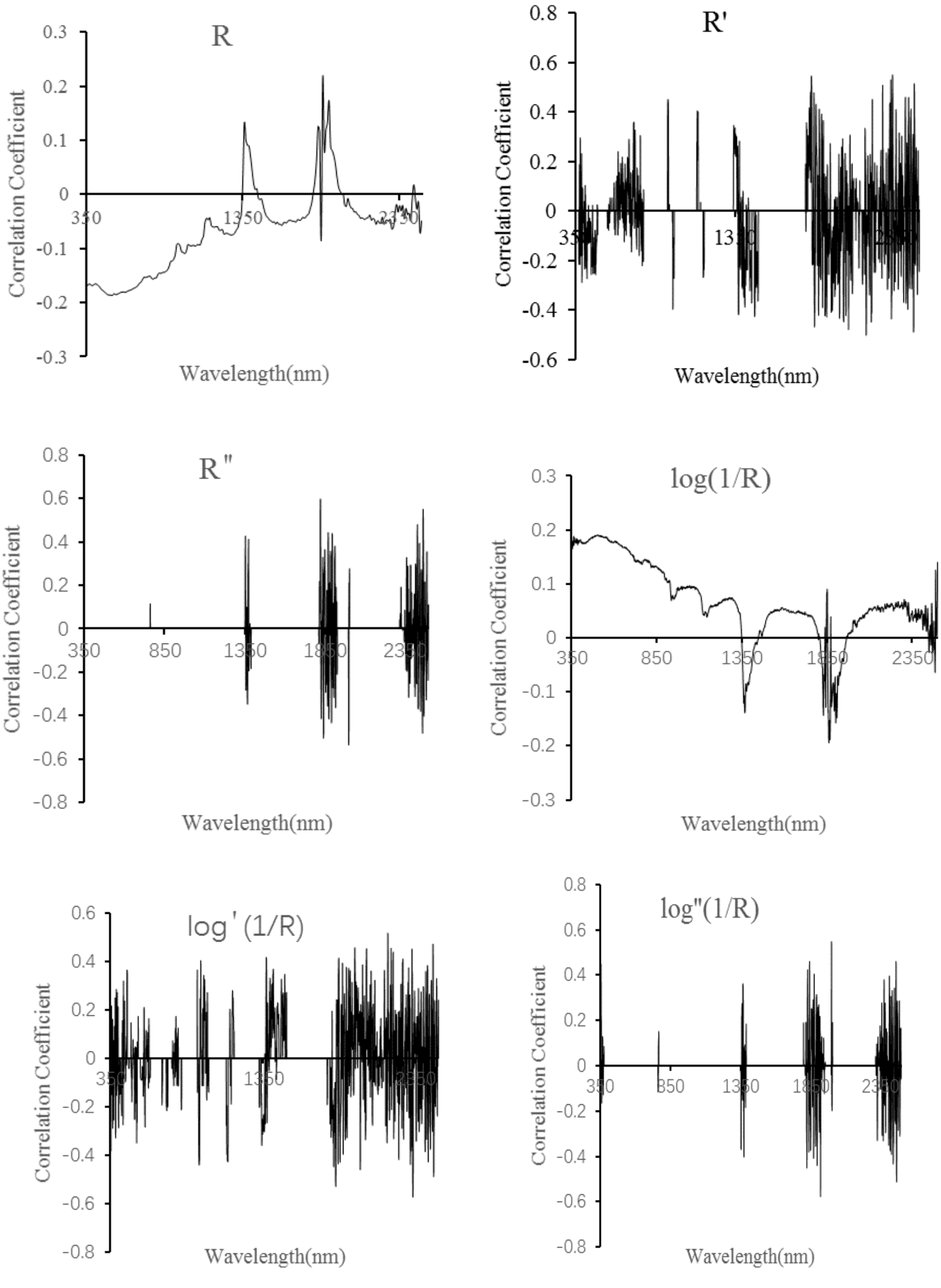

Figure 3. correlation analysis of SOC content and conversion of reflectivity different forms

The reciprocal logarithmic transformation of the original spectral reflectance reveals that the absolute value of the correlation between soil pH and each band does not change much, with a maximum of 0.1894. The maximum correlation coefficient between the spectral reflectance and the soil pH after the first order transformation of the original spectral reflectance was -0.5733 at 2331 nm. The correlation coefficients between the soil pH and the logarithm of the logarithm of the spectral reflectance were found to be extremely significant and reached the maximum at 1921 nm, which was -05581 (Table 2).

Table 2. Correlation analysis of different spectrum transform with soil pH

Types of spectral	Maximum correlation wave (nm)	Correlation coefficient
ρ	367	-0.1872
ρ'	1825	0.5453
ρ''	1824	0.5924
log(1/ρ)	519	0.1894
log'(1/ρ)	2331	-0.5733
log"(1/ρ)	1921	-0.5581

3.3. Establishment and Test of Soil PH Prediction Model

31 soil samples were randomly selected for soil pH model establishment. For the normal transformation of the soil reflectance spectra obtained from 31 samples, the first order differential, second order differential, reflectance reciprocal logarithm, reciprocal logarithmic differential, reciprocal logarithmic second order differential Processing, plus the original reflectivity to dry after a total of six data processing methods, as the prediction of the input set of modeling. The effects of different pre-treatment methods on the modeling results were analyzed in order to determine the best prediction model and data pre-processing method.

Because of the large number of spectral bands, there was a high degree of correlation between bands, which inevitably leads to multiple collinearity problems between independent variables. If the traditional multi-step linear regression modeling was used, the model was too complex and unstable The Partial least squares regression was based on the analysis of multiple regression and principal component analysis, which solves this problem. Not only to achieve the data compression, but also a good solution to the number of samples less than the number of variables. (RMSC), modeling decision coefficient (Rc^2) and calibration standard error (SEC) were used as the main evaluation indexes, and the modeling decision was made by using the partial least squares regression method to predict the soil pH. The higher the coefficient Rc^2 and the prediction coefficient (Rv^2), the higher the model stability and the fitting line. The smaller the predicted mean square error (RMSEP), the smaller the calibration standard error (SEP), the smaller the prediction bias (Bias), the higher the accuracy of the model and the better the prediction. In order to avoid the over-fitting of the model, when the optimal model was selected, the forecasting model with less number of principal components should be selected as far as possible. In Table 3, Pc represents the number of principal components used in modeling.

Table 3. PLSR Prediction Model Based on Different Transformations

transformation	Component Number(Pc)	Rc	Rc^2	SEC	RMSEC	Bias
R	12	0.9966	0.9933	0.0098	0.0096	0.0000
R'	9	0.9970	0.9940	0.0092	0.0091	0.0000
R''	11	0.9956	0.9913	0.0111	0.0110	0.0000
log(1/R)	13	0.9969	0.9939	0.0093	0.0092	0.0000
log'(1/R)	10	0.9979	0.9959	0.0077	0.0076	0.0000
log"(1/R)	12	0.9973	0.9945	0.0088	0.0087	0.0000

Table 4. PLSR Validation Model Based on Different Transformation

Transformation	Rv	Rv^2	SEP	RMSEP	Bias
R	0.9893	0.9782	0.0176	0.173	0.0002
R'	0.9938	0.9876	0.0133	0.0131	0.0005
R''	0.9893	0.9787	0.0174	0.0171	0.0001
Log (1/R)	0.9894	0.9787	0.0174	0.0171	-0.0001
Log (1/R) $'$	0.9910	0.9893	0.0160	0.0157	0.0000
Log (1/R) $''$	0.9949	0.9899	0.0119	0.0118	-0.0003

In order to verify the accuracy of the prediction model, the remaining 13 spectral data as a validation set. Based on the measured value of soil pH as the abscissa and the soil pH prediction value as the ordinate, the soil pH prediction map was established, and the prediction ability of the model was verified. It can be seen from Table 3 and Table 4 that the above six kinds of partial least squares regression models have better prediction effect on soil pH value, and the model established by the first order differential transformation of reciprocal logarithm was superior to other Five models were used to establish the model. The RMSEEC and Rc^2 of the model were 0.0076 and 0.9959, and the stability and fit of the model were high. In terms of predictive power, verify that the RMSEP and Rv^2 of the sampleswere0.0157 and 0.9893, respectively. The order of the spectrum modeling is: log '(1 / R)> log "(1 / R)> R'> log (1 / R)> R> R" in six different ways of pre-processing modeling, (1 / R)> R '> log' (1 / R)> log (1 / R)> R "> R was the order of the prediction ability of the model. It was found that the model with the best modeling effect was not necessarily optimal, such as the reciprocal logarithm logarithmic first order differential model was better than the reflectance reciprocal logarithmic second order differential modeling model, but the prediction ability But there was no reflection of the reciprocal logarithmic second order differential model effect.

The most modeling and prediction models were selected to establish the regression model coefficients and the prediction results (Fig4, Fig5). The larger the model coefficients in the regression equation, the higher the importance of the corresponding band in the model, and the greater the contribution of the region to the model (Figure 4). After the logarithmic first order differential transformation, the data points of the model and the verification samplewerealmost completely on the straight line of 1: 1, and the fitting performance and prediction ability of the model were the best, which was shown in Table 3 and Table 4 The results were consistent. Therefore, when establishing the soil pH prediction model, it was very important to carry out the first order differential transformation of the reciprocal logarithm of the reflectance.

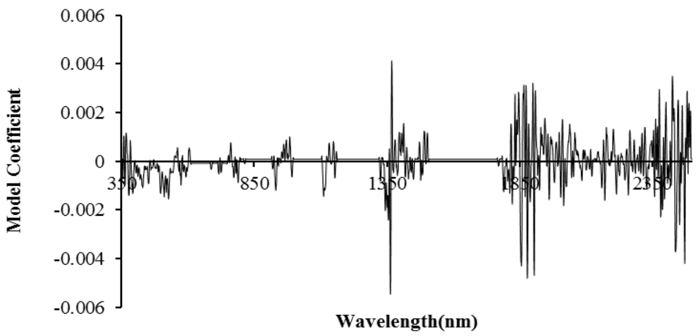

Figure 4. The Regression Coefficient of Model Based Partial Least Squares

Therefore, the partial least squares regression prediction model based on the first order differential of the spectral reflectance reciprocal logarithm was more accurate and the prediction effect was better: the principal component factor Pc = 10, the prediction coefficient Rv^2 = 0.9893, the predicted root mean square Error RMSEP = 0.0157, prediction correction deviation SEP = 0.0160, prediction bias difference Bias = 0.0000 (Figure 5).

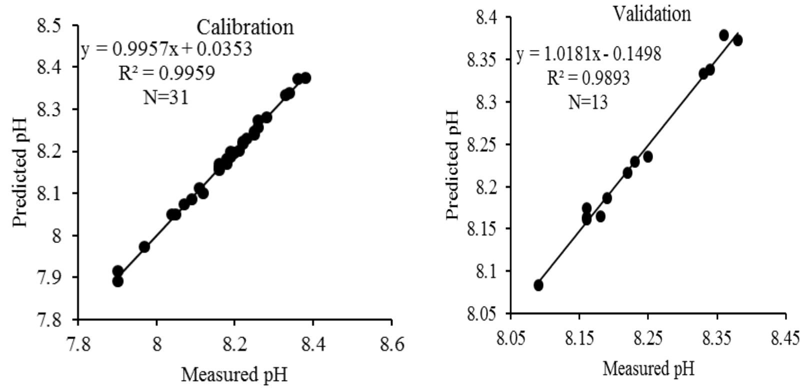

Figure 5. The result of model prediction

4. Conclusion

The change of soil pH and reflectance was significantly improved by standard normal transformation and differential treatment of the original spectral reflectance of soil. The absolute value of the correlation between the soil pH and the individual bands did not change much after the logarithmic transformation of the original spectrum. The original spectral reflectance was transformed by the reciprocal logarithmic first order transformation, and the reciprocal logarithmic second order differential transformation has a significant increase in soil pH.

From the perspective of prediction accuracy and data pre-treatment of model, the model of soil pH in the Guan Zhong area of Shaanxi Province was predicted by hyperspectral spectrum in the visible-near infrared band, and the model accuracy was established based on the first order differential of the reciprocal logarithm of reflectance Higher predictive effect was better: the principal component factor Pc = 10, the prediction decision coefficient Rv^2 = 0.9893, the predicted root mean square error RMSEP = 0.0157, the predicted correction deviation SEP = 0.0160, the prediction bias difference Bias = 0.0000, which can be achieved Rapid determination of soil ph.

Acknowledgments

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

References

[1] Li QS, Qiu SW, DENG W. Study on the Secondary Saline-Alkalization of Land in Songnen Plain. Scientia Geographica Sincia, 1998, 18 (3): 268 - 272.
[2] Guan TX, Liu HJ. A Review of Regional Soil Salinization Monitoring by Remote Sensing. Remote Sensing Technology And Application, 2001, 16 (1): 40 - 44.
[3] Jiang HN, Tahiplot tiyip, Xu YC Study on Remote Sensing Monitoring of Soil Salinizaton in Arid Areas. Arid Zone Research, 2007, 24 (2): 168 - 173.
[4] Wang FF, Wu SX, Qiao M, et al. Invesigation and Analysis on the Salinization Degree of Cultivated Land in Xinjiang Based on 3s Technology. Arid Zone Research, 2009, 26 (3) : 195 -201.
[5] Qu YH, Duan XL, Gao HY, et al. Quantitative Retrieval of Soil Salinity Using Hyperspectral Data in the Region of Inner Mongolia Hetao Irrigation District. Spectroscopy and Spectral Analysis, 2009, 25 (5): 1 362-1366.
[6] Liu HJ, Zhang B, Wang ZM, et al. Soil Saline-Alkalization Evaluation Basing on Spectral Reflectance Characteristics. Infrared Millim Waves, 2008, 27 (2): 138 - 142.
[7] Ma C. Quantitatve Retrieval of Soil Salt Content in the Songnen Plain Based on HJ1A-HIS

Images. Arid Zone Research, 2014, 31 (2): 226 - 230.
[8] Li XG, Li HP, Ren YX, et al. Analysis on the Characteristics of the Oasis Soil Salinization and Soil Spectrum in the Lower Reaches of Kaidu River Basin, Xinjiang. Chinese Journal of Soil Science, 2012,43 (1): 166 – 170.
[9] Wang KL, Xiong HG, Zhang F. PLSR-BP Complex Model-Based Hyper-Spectrum Retrieval of Oasis Soil pH. Arid Zone Research, 2014, 31 (6): 1005 - 1009.
[10] CLOUTIS E A. Hyperspectral geological remote sensing: Evaluation of analytical techniques [J]. International Journal Remote Sensing, 1996, 17 (12): 2215 - 2242.
[11] 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.
[12] Pu RL, Gong P. 2000. Hyperspectral Remote Sensing and Applications. Beijing: Higher Education Press.
[13] Savitsky, A., & Golay, J. E. (1964). Smoothing and differentiation of data by simplified least square procedure. Analytical Chemistry, 36, 1627–1639.
[14] Dalai, R. C., & Henry, R. J. (1986). Simultaneous determination of moisture, organic carbon, and total nitrogen by near infrared reflectance spectrophotometry. Soil Science Society of America Journal, 50, 120 – 123.
[15] Reeves, J., III, McCarty, G., & Mimmo, T. (2002). The potential of diffuse reflectance spectroscopy for the determination of carbon inventories in soils. Environmental Pollution, 116,277 – 284.
[16] Udelhoven, T., Emmerling, C., & Jarmer, T. (2003). Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial least-square regression: a feasibility study. Plant and Soil, 251, 319 – 329.