



HAL
open science

Modelling and Predicting of Soil Electrical Conductivity and PH from Semi-arid Grassland Using VIS-NIR Spectroscopy Technology

Shangbin Lei, Nisha Bao, Shanjun Liu, Xiaocui Liu

► **To cite this version:**

Shangbin Lei, Nisha Bao, Shanjun Liu, Xiaocui Liu. Modelling and Predicting of Soil Electrical Conductivity and PH from Semi-arid Grassland Using VIS-NIR Spectroscopy Technology. 10th International Conference on Computer and Computing Technologies in Agriculture (CCTA), Oct 2016, Dongying, China. pp.442-453, 10.1007/978-3-030-06155-5_45 . hal-02180012

HAL Id: hal-02180012

<https://inria.hal.science/hal-02180012v1>

Submitted on 12 Jul 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Modelling and predicting of soil electrical conductivity and pH from semi-arid grassland using VIS-NIR spectroscopy technology

ShangBin Lei^{1,2}, NiSha Bao¹ (✉), ShanJun Liu¹, XiaoCui Liu^{1,3}

¹ Institute for Geo-informatics & Digital Mine Research, Northeast University, Shenyang, 110819

baonisha@hotmail.com

² CCCC (Tianjin) Eco-Environmental Protection Design & Research Institute Co., Ltd.

³ Land source surveying and planning of Liaoyang branch, Liaoyang, China

Abstract . The electrical conductivity (EC) and pH value are key indicators for soil physical and chemical properties, which can reflect the level of soil acid and alkali, furthermore, influence the vegetation growth. The spectroscopy technique can estimate and evaluate electrical conductivity and pH value rapidly and efficiently, which can provide useful information on the real-time soil management in the semi-arid rangeland or grassland. We picked the semi-arid grassland of northern China covering an area about 200 km² as the target research area, given that it is highly sensitive to grazing and mining affect. Soil samples were collected from 72 sampling sites in this area, which covered grazing exclusion, over grazing and grassland restoration area. The SVC HR-1024 spectroradiometer was used to acquire soil spectrum. This study aims to indicate the spectral characteristic for soil EC and pH, and propose a predicting modeling method with optimal input spectral region and transformation by comparing the support vector machine (SVM) regression method and partial least squares (PLSR) regression modeling method. Our results showed that: 1) once EC value is larger than 0.10 μs/m, the soil spectral reflectance decreases with increasing of EC value. The absorption depth, width and area at 1900nm reduce with increasing of EC value as well; 2) There are positive correlation between EC, pH value and soil spectral reflectance. The highest correlation coefficient value of 0.7 between pH and reflectance is recorded at visible region around 500nm; 3) The SVM modeling method produce the higher prediction accuracy (RPD=2.18, RMSE =0.035, R² =0.78 for EC, RPD>3, RMSE =0.349, R² =0.91 for pH) for EC, tive correlation between PLSR methods in soil EC and pH prediction. This study indicated that it was possible to use the spectroradiometer technology to predict EC and pH value for the soil from semi-arid grassland, which would provide the basis for soil acid and alkali detecting using hyper-spectral remote sensing technology.

Key words : Soil electrical conductivity and pH · semi-arid grassland · spectral characteristic · spectral modeling and predicting

1 Introduction

Soil salinity is world-known as a major threat to agriculture, especially in arid and semi-arid regions^[1]. Soil electrical and pH can serve as one of indicators to detect soil salinization problem in arid and semi-arid soil^[2]. It is vital to assess the extent and degree of severity in the early stage identification of soil salinization process in semi-arid areas^[3]. Conventional methods has the limitation in requiring large amounts of labor and chemicals for acquiring soil properties. Reflectivity can respond to soil changes in real time, quickly and efficiently, and can provide useful information^[4]. Based on the hypersepctral techonology, soil siface map was produced for assessing soil salinity^[5].

Soil constituents have unique absorption features over the entire visible (350 – 700 nm) and near

infrared (700 – 2500 nm) region due to overtones related to stretching and bending vibrations in molecular bonds such as C – C, C – H, N – H and O – H^[6]. The NIR and SWIR regions have potential for estimation of soil salinity levels. It is found that the salt terminal molecules have obvious absorption characteristics at 1450nm, 1900nm and 2200nm^[7]. Many multivariable techniques have been able to use spectral features to construct predictive inversion models, which is possible to be applicable to non-surveyed areas having similar geographical conditions. The linear models, such as PLSR and principal component analysis, have been widely applied to build models for soil organic carbon prediction. Farifeh et al indicated that the PLSR permit reliable estimations of soil EC at experimental level from the island of Texel in the northwest of the Netherlands^[3]. A support vector machine (SVM) is a kernel-based learning method from statistical learning theory, which can be learning in high-dimensional feature space with fewer training data. This method has attracted the attention of researchers and has been widely applied in hyperspectral data analysis^[8, 9].

The semi-arid rangeland of north-east China, occupies over 20% of the total grassland in China. The main grassland in this region has been shaped by drought, grazing and fire^[10], and more recently, but more dramatically by mining activities^[11]. In the region of the range, the national soil VIS-NIR library better than local scale models^[12]. The local calibrations of soil spectroscopic models of field sampling sets may be more accurate than national calibration^[13]. Thus, the specific applicable EC and pH prediction model should be developed in local semi-arid rangeland environment. This study aims to indicate the spectral characteristic for soil EC and pH, and propose a predicting modeling method with optimal input spectral region and transformation by comparing the robust regression method, support vector machine (SVM) regression method and partial least squares (PLSR) regression modeling method.

2. Materials and Methods

2.1 Study area and sample set

The study area is located in the central of Hulunber stepped meadow, land utilization patterns including open-pit mine, mine rehabilitated land, lake, farmland, fenced pasture, grazing pasture. The main soil type is chestnut soil. A stratified random sampling method was adopted to set up 72 sample points in the study area. (Figure 1). The strata were defined based upon land use/cover map derived from the high resolution imagery of 2015, topography map and mine rehabilitation plan. The soil samples were collected at August 2015. Four surface soil samples, 10 cm deep and 5 cm wide, were collected at each sampling point and compressed in a 20 cm box, and then transported and processed in the lab. All datasets were selected to represent the full range in soil organic matter. The dominant pretreatment of soils include drying and sieving for VIS-NIR analysis in the laboratory, which ensure a constant particle size effect on spectra .

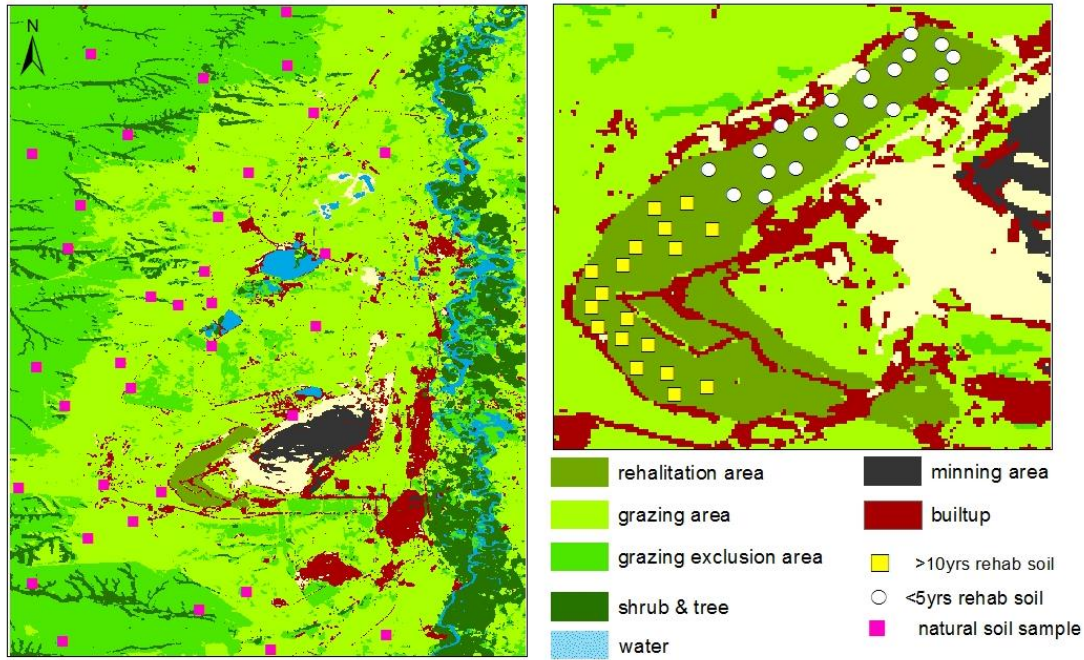


Figure 1 The sampling location in the land use map

Electrical conductivity and pH are the two indicators of soil samples testing. Currently extraction method is commonly used in testing the electrical conductivity of soil. The absolute content and relative content of all kinds of salinity in water extract are greatly influenced by the ratio of soil and water. During the process of analysis and testing, the ratio of saturated soil paste and water soil is frequently 1:5 in the water extract .The electrical conductivity of the soil water sample, whose ratio of water and soil is 1:5, is made by taking 30g dried soil sample through 2mm-sieve, adding 150ml carbon dioxide free water, oscillating with stopper for 3 min, air exhausting and filtering, finally water extract is gained, and the electrical conductivity of it is measured with the use of Mettler SevenMulti pH/electrical conductivity comprehensive test instrument . Seven Multi pH electrical conductivity comprehensive test instrument is used to test the pH of mixed supernatant with 10g soil sample and 25ml distilled water. The statistical results of electrical conductivity (EC) and pH data of 72 soil samples is listed in table 1.

Table 1 Electrical conductivity and pH of the soil

Statistical index	Sample number				Mean	Maximum	Minimum	Standard deviation	Variation coefficient
	Sample number	Pastoral areas	Fence	Artificial grassland					
EC (ms/cm)	72	25	10	37	0.19	0.73	0.04	0.13	0.67
pH	72	25	10	37	7.68	10.13	5.42	1.29	0.17

2.2 Laboratory spectral measurement

In our study, the measurements was taken in the lab under the same environmental condition using the field spectroradiometer, which make sure that all the spectrum of the soil samples are standard. The soil spectrum test with the SVC-HR-1024 spectrometer is located at the vertical 30cm above the surface of the

soil sample, using a 4 degree field angle lens. and the area of soil spectrum was a circle with a diameter of 2 cm. The first step of the spectral test is the whiteboard measurement. Then each soil sample is measured four times. The circular aluminum box is rotated by 90 ° between each measurement. Each angle is repeatedly measured 5 times, the sample was measured 20 times in total, and the average value was taken as the actual reflection spectrum data of the soil sample.

2.3 Spectral data preprocessing

Spectral preprocessing with mathematical functions is commonly used to correct for non-linearity, sample variations, and noisy spectra. Additionally, it can highlight spectra characteristic and improve the accuracy of prediction models. In this study, the preprocessing method include logarithm of the reciprocal of reflectance ($\log 1/R$)^[14] and the first derivative of reflectance (DR) at each waveband remove baseline effects^[6]. The first derivative of absorbance (R') is computed as shown in Eq. (1).

$$R' = \frac{(R_{n+1} - R_n)}{(\lambda_{n+1} - \lambda_n)} \quad (1)$$

The spectral data were processed by continuum removal, and the absorption characteristics curve of the spectral curve was obtained. In Figure 2, the horizontal coordinate is the wavelength, and the vertical coordinate is the spectral reflectivity. The absorption characteristics of a spectrum curve is composed of spectral absorption low point m and two shoulders (S_1 and S_2) absorbed by spectrum, the connection of S_1 and S_2 is called the non-absorption baseline. Different objects have different indicators to extract spectral absorption characteristics, in this paper, absorption depth D (depth), absorption width W (width), total absorption area A (area), symmetry λ (symmetry) are the characteristics indicators of soil spectrum.

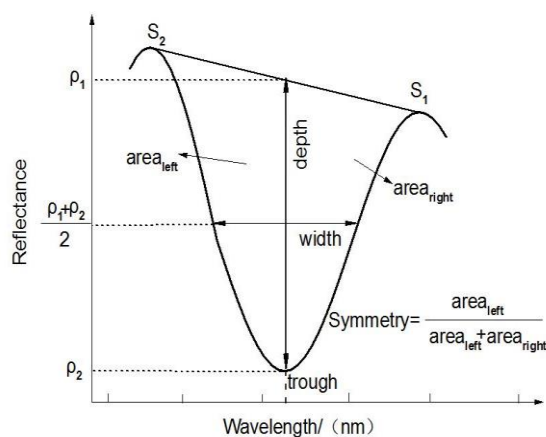


Fig.2 Shape parameters of a spectrum

The Pearson product-moment correlation coefficient is selected to indicate the relationship between soil EC, pH value and reflectance band by band. The Pearson correlation coefficient (Pearson's r) is a measure of the linear correlation between two variables, the correlation coefficient is represented by r , and the larger the absolute value of r , the stronger the correlation. It is computed by two variables X and Y , as shown in Eq. (2), where X is reflectance spectra and transformation results of a soil sample in a certain band, Y is PH value or EC value of a soil sample, i is number of soil samples.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2)$$

2.4 Modeling and predicting Methods

The spectral data sets are used to establish models for predicting EC and pH. In order to verify the predictive ability of the established model, the data set is divided into calibration / validation. (2/1) with 48 samples for calibration and a test set with 24 samples for verification. The RMSE (root mean squared error) value and adjusted coefficients of determination (R^2) value of predictions were calculated to estimate the model. The residual prediction deviation (RPD) (the ratio of standard deviation to RMEcv) was used as rubrics for evaluating the stability and accuracy of multivariable models^[15]. The best prediction models are characterized by a RPD of >2.0 with R^2 of 0.80-1.0. The RPD value is most useful when the validation set is independent of the calibration set.

PLSR is a regression technique to find the best function by minimizing the RMSE value. In this study, PLSR is used to predict EC and pH measured in the lab using the spectra matrix. The programs used for the PLSR calculations were those from the PLSR toolbox, with R package version 3.2.1.

SVM has the advantage in solving small sample, nonlinear and high-dimensional pattern recognition, and can be extended to other machine learning model by function fitting. Low dimensional vector sets can be projected to high-dimensional spaces based on the kernel function in SVM. The SVM was applied by using the R package of “e1071 package”, an R interface to library for support vector machines (LIBSVM). The optimal parameter of C, e and kernel-specific in SVM was chosen by using leave one-out cross-validation.

3. Result and Analysis

3.1 The characteristic soil spectrum of EC and pH

Once EC value greater than 0.1 μ m/s, the lower reflectance observed for the soil was linked to the higher EC, which absorbs electromagnetic energy (Figure 3). Apart from the highest pH value of 9.5, the lower reflectance observed for the soil was responded to the lower pH value in the visible spectrum region (Figure 4). The absorption feature at 1900nm was selected to analyze the soil spectrum characteristic with different EC and pH level. As shown in the figure 5 and figure 6, the absorption width, depth and area was increasing corresponding to the increasing of EC value. The largest width at 1900nm absorption was recorded for the soil spectrum from pH value of 9.5.

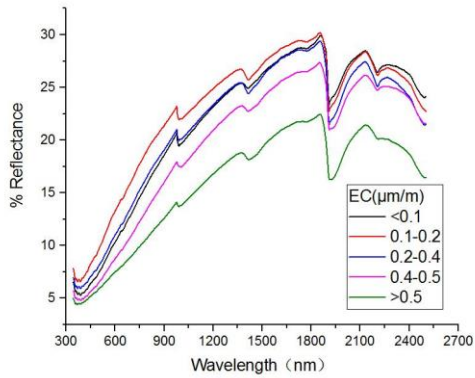


Fig.3 Hyperspectral reflectance curves of soil with different conductivities

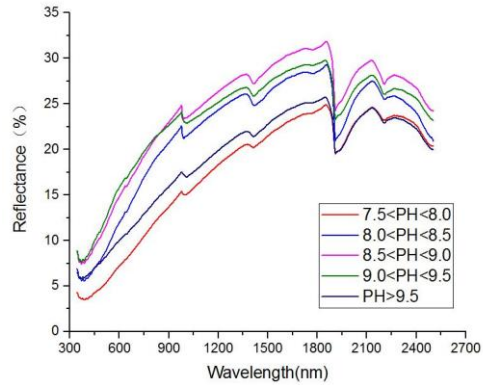


Fig.4 Hyperspectral reflectance curves of soil with different pH

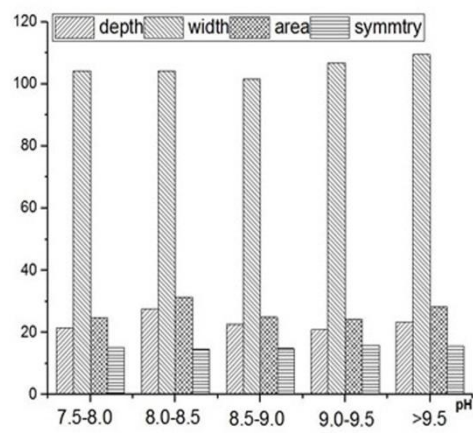
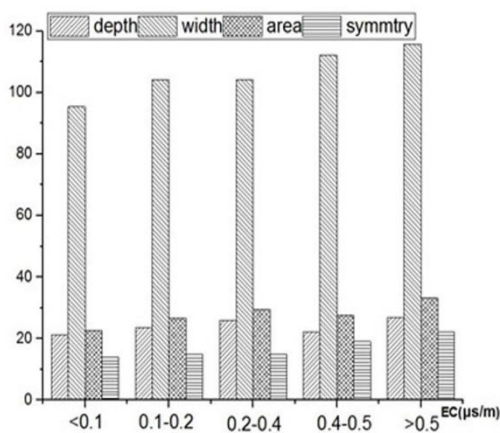


Fig.5 Statistics of characteristics at 1900nm absorption bands with different conductivity and pH values

In order to indicate the sensitive spectrum region between EC, pH and reflectance, the correlation coefficient was calculated with different spectrum including raw, first derivative and LOG transformation. For the raw spectrum, the correlation for the EC and reflectance was relatively lower than that for pH value and reflectance (Figure 6). There is a positive correlation from the 400nm-1200nm region between EC and reflectance. The highest correlation coefficient value was recorded around 2200nm. The correlation coefficient value was decreasing between pH and reflectance from 500nm-2500nm. There is the highest correlation coefficient value of 0.70 at the visible band of 500nm. For the first derivative spectrum, the sensitive region with the higher correlation coefficient value was similar to the raw spectrum (Figure 7). Although the correlation coefficient value was fluctuated dramatically from the LOG transformed spectrum, the correlation coefficient value was higher than that from raw and derivative spectrum. The highest correlation coefficient value was 0.50 for EC at 962 nm, and 0.78 for pH at 407nm (Figure 8).

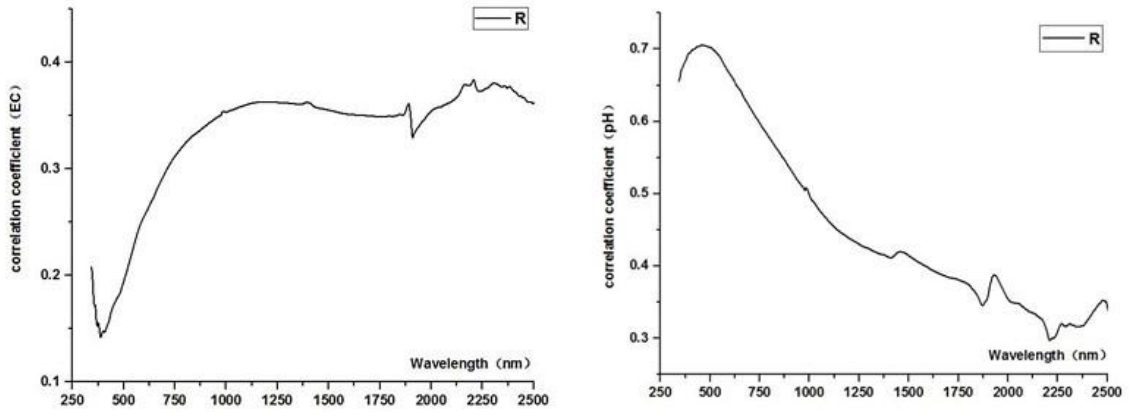


Fig.6 Pearson's Correlation coefficient between measured EC、pH and reflectance at different bands

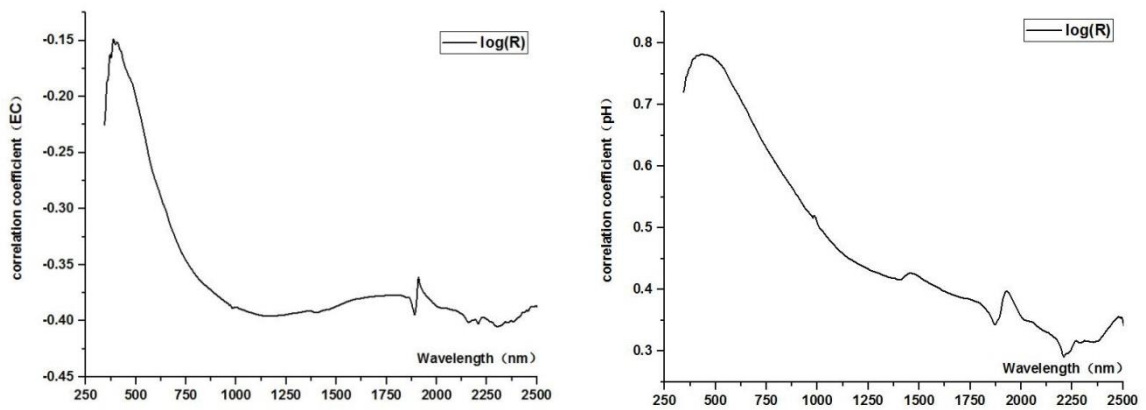


Fig.7 Pearson's Correlation coefficient between measured EC、pH and log reflectance value at different bands

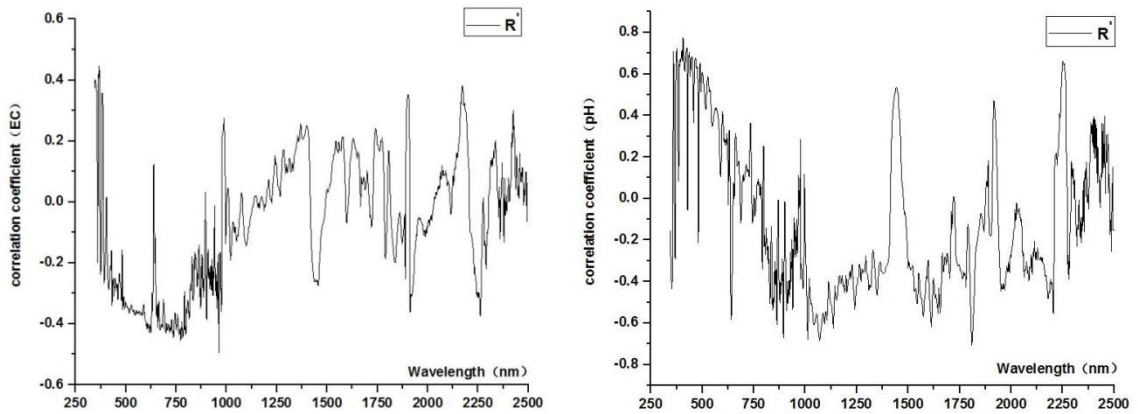


Fig.8 Correlation coefficient between measured EC、pH and differential reflectance at different bands

3.2 The validation of modeling results

The optimal input spectral region for SVM and PLSR modeling was selected according to relationship between soil EC and reflectance from the Pearson product-moment correlation coefficient. The 714nm-2500nm for raw spectrum with Pearson correlated coefficient larger than 0.3, 480nm-817nm with Pearson correlated coefficient larger than 0.3 for derivative spectrum and 645-2500nm with Pearson correlated coefficient greater than 0.3 was selected as the input spectral region.

3.2.1 SVM

Both preprocessing and input spectral region had a greater impact on modeling methods. As shown in the table 2 and 3, for SVM modeling, the raw spectrum, LOG and first derivative transformation can provide the good model performance in cross validation. For individual input spectral region, the highest correlation is higher than 0.3 between EC (pH) and reflectance. SVM model with EC in the sensitive band by LOG transformation has the highest predictive power with RPD =2.7 for cross validation ($R^2 = 0.86$, RMSE = 0.034) and RPD = 2.18 for external validation ($R^2 = 0.78$, RMSE =0.035)(Figure 9). SVM model with pH in the sensitive band by raw spectrum has the highest predictive power with RPD >3 for cross validation ($R^2 = 0.95$, RMSE = 0.275) and RPD >3 for external validation ($R^2 = 0.91$, RMSE =0.349) (Figure 10).

Table2 Prediction of EC in different spectral transformations of SVM modeling

Spectral transformation	Modeling band/nm	Cross-validation			External-validation		
		R ²	RMSE	RPD	R ²	RMSE	RPD
R	714-2500	0.82	0.03	2.28	0.682	0.041	1.75
R'	480-817	0.89	0.018	2.45	0.51	0.037	1.43
log(R)	645-2500	0.86	0.034	2.7	0.78	0.035	2.18

Table 3 Prediction of PH in different spectral transformations of SVM modeling

Spectral transformation	Modeling band/nm	Cross-validation			External-validation		
		R ²	RMSE	RPD	R ²	RMSE	RPD
R	350-2500	0.95	0.275	>3.00	0.91	0.349	>3.00
R'	350-2500	0.93	0.285	>3.00	0.90	0.326	>3.00
log(R)	350-2500	0.92	0.335	>3.00	0.91	0.349	>3.00

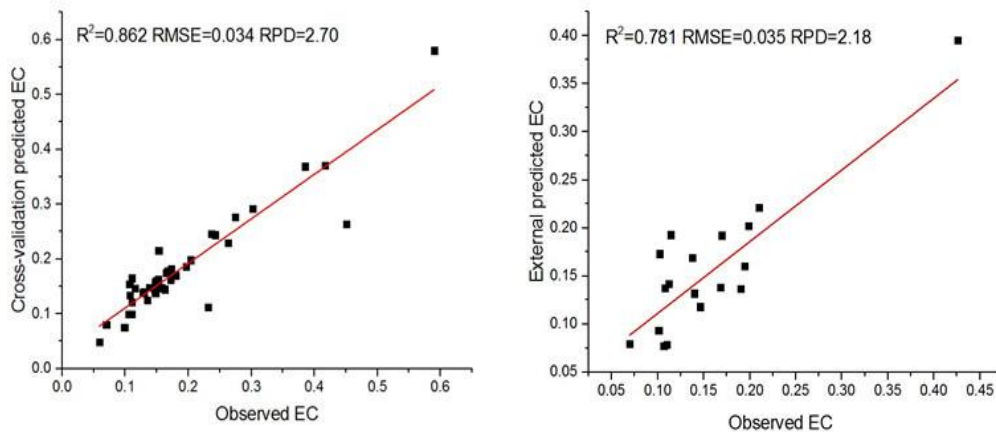


Fig.9 The fitting results of prediction values and the actual values of EC with SVM

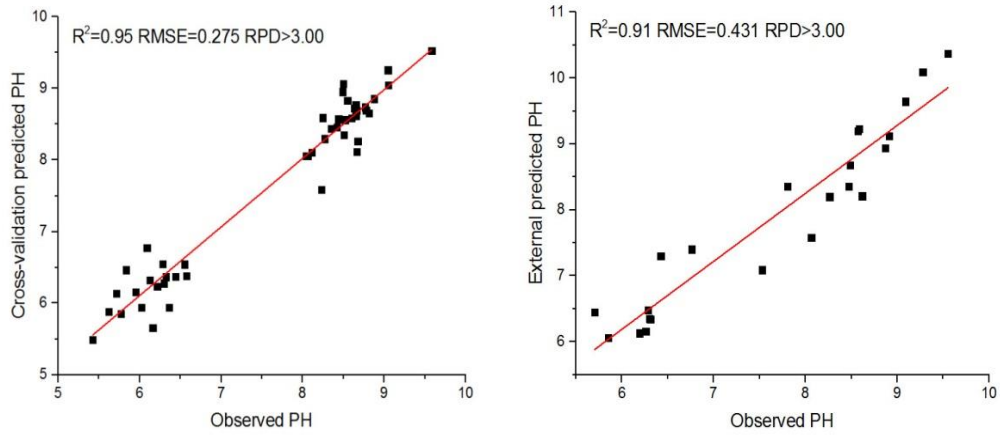


Fig.10 The fitting results of prediction values and the actual values of pH with SVM

3.2.2 PLSR

As shown in the table 4 and 5, for PLSR modeling, the raw spectrum, LOG and first derivative transformation can provide the good model performance in cross validation. For individual input spectral region, the highest correlation is higher than 0.3 between EC (pH) and reflectance. PLS model with EC in the sensitive band by raw spectrum has the highest predictive power with RPD >3 for cross validation ($R^2 = 0.99$, RMSE = 0.001). However, for the external validation, the modeling method should be further improved according to the RPD value less 2 ($R^2 = 0.69$, RMSE = 0.039) (Figure 11). PLS model with pH in the sensitive band by raw spectrum has the highest predictive power with RPD >3 for cross validation ($R^2 = 0.99$, RMSE = 0.119) and RPD = 2.53 for external validation ($R^2 = 0.89$, RMSE = 0.497) (Figure 12).

Table 4 Prediction of EC in different spectral transformations of PLSR modeling

Spectral transformation	Modeling band/nm	Cross-validation			External-validation		
		R ²	RMSE	RPD	R ²	RMSE	RPD
R	714-2500	0.99	0.001	>3.00	0.69	0.039	1.42
R'	480-817	0.92	0.021	>3.00	0.47	0.044	1.38
log(R)	645-2500	0.98	0.010	>3.00	0.65	0.040	1.34

Table 5 Prediction of pH in different spectral transformations of PLSR modeling

Spectral transformation	Modeling band/nm	Cross-validation			External-validation		
		R ²	RMSE	RPD	R ²	RMSE	RPD
R	350-2500	0.99	0.119	>3.00	0.89	0.497	2.53
R'	350-2500	0.99	0.095	>3.00	0.78	0.648	2.00
log(R)	350-2500	0.94	0.297	>3.00	0.85	0.566	2.44

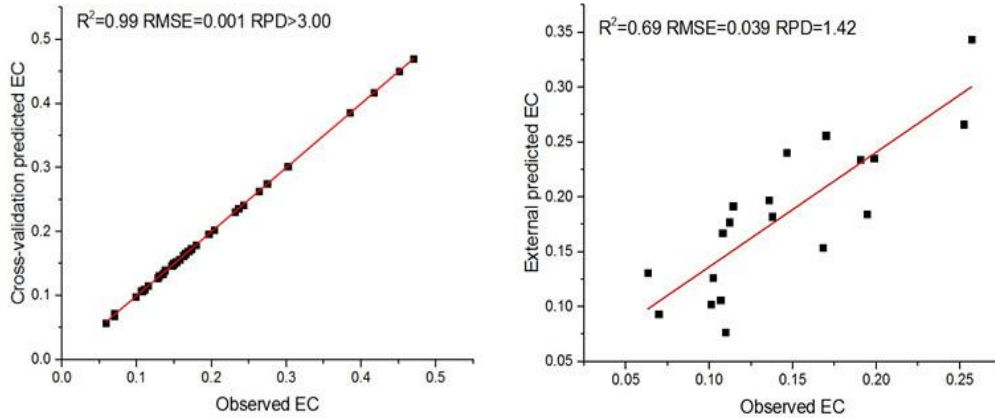


Fig.11 The fitting results of prediction values and the actual values of EC with PLSR

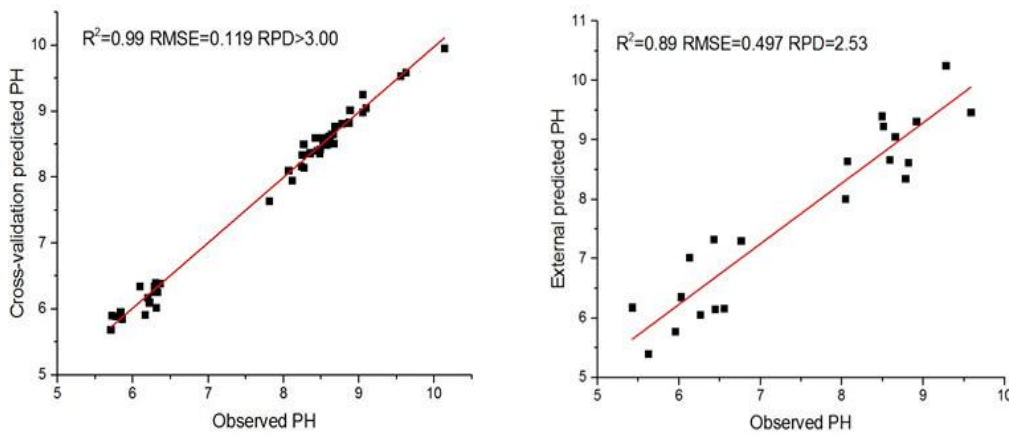


Fig.12 The fitting results of prediction values and the actual values of pH with PLSR

4 Discussion and Conclusion

In this study, the spectral characteristics of the visible and near-infrared spectra of soil samples from the semi-arid grassland are described. Nonlinear Support Vector Machine (SVM) and Linear Partial Least Square Regression (PLSR) were compared to obtain the best prediction model of EC and pH.

Spectral characteristics know the laws of surface soil changes under different vegetation cover conditions. The soil with higher EC value was recorded with lower reflectance. The absorption feature of depth and width at 1900nm can be used to classify the different level of EC, which increases corresponding to the increasing of EC and pH value. Generally, there is higher correlation for pH with reflectance rather than EC in the 350nm-2500nm. The higher correlation coefficients for EC was detected in the visible spectral coverage. Melendez-Pastor^[2] found that the highest correlation coefficient value was 0.70 for mean EC of 1.75 μ m/s at the visible band of 500nm.

Both SVM model and PLS model performed well in predicting pH and EC with RPD value larger than 2. The SVM modeling results for EC with RMSE 0.035 μ m/S were slightly higher than PLS model 0.039 μ m/S. One possible reason is that PLSR-SVM deals with the non-linear part of the spectral data, whereas PLSR can only deal with the linear part of the correlation between the EC and the spectral data^[3]. There were three key factors in the utilization of field spectroscopy predictions, as proposed by Milton et

al^[16]. Our study provided the answer to one question, the optimum spectral region for predicting EC from semi-arid grasslands. It was found that preprocessing method and input spectral region had a great impact on the different modeling approach. According to the optimum spectral region (714nm-2500nm) by LOG transformation from Pearson's r analysis as input for the SVM model, the results of cross validation and internal validation of EC prediction have achieved very high accuracy. Generally, the derivative and LOG transformation could improve the modeling accuracy. However, for pH predicting, the raw spectrum was more effective than other transformation both in PLSR modeling and SVM modeling.

Acknowledgements

We acknowledge the National Natural Science Foundation of China for young researchers (41401233), the Fundamental Research Funds for the Central Universities (N160102001).

References

1. Scudiero, E., Skaggs, T. H., Corwin, D. L.: Comparative regional-scale soil salinity assessment with near-ground apparent electrical conductivity and remote sensing canopy reflectance. *Ecological Indicators*. 70(November),276-284(2016)
2. Melendez-Pastor, I., Navarro-Pedreño, J., Gómez, I., Koch, M.: Identifying optimal spectral bands to assess soil properties with VNIR radiometry in semi-arid soils. *Geoderma*. 147(3),126-132(2008)
3. Farifteh, J., Meer, F. V. D., Atzberger, C., Carranza, E. J. M. :Quantitative analysis of salt-affected soil reflectance spectra: a comparison of two adaptive methods (PLSR and ANN). *Remote Sensing of Environment*. 110(1), 59-78 (2007)
4. Demattê J. A. M., Campos, R. C., Alves, M. C., Fiorio, P. R., Nanni, M. R. Visible–nir reflectance: a new approach on soil evaluation. *Geoderma*. 121(1), 95-112(2004)
5. Ben-Dor, E., Patkin, K., Banin, A., Karnieli, A. Mapping of several soil properties using dais-7915 hyperspectral scanner data - a case study over clayey soils in israel. *International Journal of Remote Sensing*. 23(6), 1043-1062(2002)
6. Ben-Dor, E., Chabrillat, S., Demattê J. A. M., Taylor, G. R., Hill, J., Whiting, M. L., et al. Using imaging spectroscopy to study soil properties. *Remote Sensing of Environment*. 113(1), S38-S55(2009)
7. Metternicht, G.I., Zinck, J.A. Remote sensing of soil salinity: potentials and constraints. *Remote Sensing of Environment*. 85(1), 1-20(2003)
8. Chauchard, F., Cogdill, R., Roussel, S., Roger, J.M., Bellon-Maurel, V. Application of ls-svm to non-linear phenomena in nir spectroscopy: development of a robust and portable sensor for acidity prediction in grapes. *Chemometrics and Intelligent Laboratory Systems*. 71(2), 141-150(2004)
9. Bao, N., Wu, L., Ye, B., Yang, K., Zhou, W. Assessing soil organic matter of reclaimed soil from a large surface coal mine using a field spectroradiometer in laboratory. *Geoderma*. 288, 47-55(2017)
10. Wang, X., Li, F., Gao, R., Luo, Y., & Liu, T. Predicted npp spatiotemporal variations in a semiarid steppe watershed for historical and trending climates. *Journal of Arid Environments*. 104(104), 67-79 (2014)
11. He, P., Yang, L., Xu, X., Zhao, S., Chen, F., & Li, S., et al. Temporal and spatial variation of soil available potassium in china (1990–2012). *Field Crops Research*. 173, 49-56(2015)

12. Gogé F, Gomez, C., Jolivet, C., & Joffre, R. Which strategy is best to predict soil properties of a local site from a national Vis-NIR database?. *Geoderma*. 213(213), 1-9(2014)
13. Wetterlind, J., Stenberg, B. Near-infrared spectroscopy for within-field soil characterization: small local calibrations compared with national libraries spiked with local samples. *European Journal of Soil Science*. 61(6), 823-843 (2010)
14. Stenberg, B., Rossel, R. A. V., Mouazen, A. M., & Wetterlind, J. Chapter five - visible and near infrared spectroscopy in soil science. *Advances in Agronomy*. 107(107), 163-215(2010)
15. Chang, C. C., & Lin, C. J. LIBSVM: A library for support vector machines. *ACM*. (2011)
16. Milton, E. J., Fox, N. P., & Schaepman, M. E. Progress in field spectroscopy. *Remote Sensing of Environment*. 113(Suppl. 1), S92–S109(2006)