A Comparative Study of MLR and PLSR for Spectral Detection of Salt Content in Sandy Loam Soil: Postprint

Authors: Wang Tao, Caili Yu

Date: 2018-11-14T00:00:00+00:00

Abstract

To rapidly and effectively detect changes in salt content of typical soils (sandy loam) in the southern Xinjiang region, near-infrared hyperspectral and electrical conductivity data of saline soils were measured using a spectrometer and conductivity meter in the jujube planting area of Alar City, southern Xinjiang. Based on 7 different spectral preprocessing methods and 2 feature wavelength selection algorithms, soil salinity monitoring models were established using Multiple Linear Regression (MLR) and Partial Least Squares Regression (PLSR), respectively. The results showed that among the 7 preprocessing methods, normalization, multiplicative scatter correction, standard normal variate, and first derivative could effectively improve the prediction accuracy of soil salinity models. For the Multiple Linear Regression model based on Stepwise Multiple Regression (SMR) wavelength selection method (SMLR), Rval2 was greater than 0.9489, RPD was greater than 6.2949, and RMSEP was less than 0.4356; for the Multiple Linear Regression model based on Successive Projections Algorithm (SPA) (SPA-MLR), Rval2 was greater than 0.9568, RPD was greater than 6.9221, and RMSEP was less than 0.3616, with prediction results superior to the Partial Least Squares Regression (PLSR) model. Among these, the prediction accuracy of SMLR and SPA-MLR based on normalization preprocessing was the most ideal, with Rval2=0.9792, RPD=9.9078, RMSEP=0.2876 and Rval2=0.9805, RPD=10.50, RMSEP=0.2783, respectively, and fewer characteristic wavelengths were selected. This indicates that normalization is a more effective spectral preprocessing method, and Multiple Linear Regression (MLR) is more suitable for establishing prediction models of salt content in typical sandy loam soils in southern Xinjiang.

Full Text

A Comparison of Salt Content in Sandy Soil Between the MLR Model and PLSR Model

WANG Tao[1,2], YU Cai-li[1,2], YAO Na[1,2], WANG Fei[1,2], BAI Tie-cheng[1,3]※

[1] College of Information Engineering, Tarim University, Aral 843300, Xinjiang, China

[2] South Xinjiang Agricultural Informatization Research Center, Xinjiang Aral 843300, Xinjiang, China

[3] Gembloux Agro-Bio Tech, University of Liège, Gembloux 25030, Belgium

Abstract

To rapidly and effectively monitor typical soil salt content (sandy loam soil) in South Xinjiang, China, and to improve the precision of soil salt content estimation models by removing noise from soil hyperspectral absorbance, this study investigated the inversion relationship between soil spectrum and electrical conductivity (EC) using multiple spectral pretreatment methods. Multiple linear regression (MLR) and partial least squares regression (PLSR) modeling were then applied to establish salt content models based on hyperspectral analysis techniques. The effective and predictive capacities of different models were validated. This research took a typical arid area in South Xinjiang as the study region, obtaining hyperspectral data and EC measurements using a near-infrared spectrometer (Zolix Gaia Sorter) and conductivity meter (DDS-307). One hundred forty-two soil samples at 0–20 cm depth were collected, and these samples were highly representative of the EC values. Seven pretreatment methods were used to process the original spectral data: vector normalization (VN), multiplicative scatter correction (MSC), standard normal variate (SNV), moving-average (MA) smoothing, Savitzky-Golay (SG) smoothing, first derivative (1-Der), and second derivative (2-Der). Characteristic wavelengths were then extracted using stepwise multiple regression (SMR) and successive projections algorithm (SPA), which served as input variables for MLR and PLSR modeling. The results showed that the optimal pretreatment methods were VN, MSC, SNV, and 1-Der. According to different pretreatments, in the stepwise multiple linear regression (SMLR) prediction model, R_{val}^2 was greater than 0.95, RPD was greater than 6.2, and RMSEP was less than 0.44. In the multiple linear regression model based on the successive projections algorithm (SPA-MLR), R_{val}^2 was greater than 0.96, RPD was greater than 6.9, and RMSEP was less than 0.36, which were better than those in SMLR. In the partial least squares regression prediction model, R_{val}^2 was greater than 0.88, RPD was greater than 4.4, RMSEP was less than 0.62, and in the partial least squares regression model based on the successive projections algorithm (SPA-PLSR), R_{val}^2 was greater than 0.59, RPD was greater than 2.4, and RMSEP was less than 1.1, which were less than those in SMLR and SPA-MLR. The best predictions of SMLR and

SPA-MLR after vector normalization (VN) were as follows: RMSEP = 0.2876, R_{val}^2 = 0.9792, RPD = 9.9078 and RMSEP = 0.2783, R_{val}^2 = 0.9805, RPD = 11.50, which had fewer characteristic wavelengths. Therefore, VN is the most effective pretreatment method. It is more suitable to establish the prediction model of soil conductivity in jujube orchards using MLR with PLSR. This could be an important consideration for future research on how to choose the right algorithm to analyze soil salt content with spectral data.

Keywords: soil electrical conductivity; multiple linear regression; partial least squares regression; hyperspectral

1 Materials and Methods

1.1 Data Collection

The study area was located in a typical arid region of South Xinjiang. One hundred forty-two soil samples were collected from 0–20 cm depth using a soil auger. The original soil spectra were measured using a Zolix Gaia Sorter near-infrared spectrometer with a spectral range of 400–1000 nm. Soil electrical conductivity was measured using a DDS-307 conductivity meter after preparing soil-water extracts (1:5 ratio). The samples covered a wide range of salinity levels, providing good representation for model calibration and validation.

1.2 Spectral Pretreatment Methods

Seven pretreatment methods were applied to the original spectral data to eliminate noise and enhance spectral features: vector normalization (VN), multiplicative scatter correction (MSC), standard normal variate (SNV), moving-average smoothing (MA), Savitzky-Golay smoothing (SG), first derivative (1-Der), and second derivative (2-Der). These methods effectively reduced baseline drift, scattering effects, and random noise while preserving the characteristic absorption information related to soil salinity.

1.3 Characteristic Wavelength Selection

Two methods were employed to select informative wavelengths: stepwise multiple regression (SMR) and successive projections algorithm (SPA). SMR selects variables based on statistical significance, while SPA chooses wavelengths with minimal redundancy and maximal information content. The selected wavelengths served as input variables for subsequent modeling.

1.4 Model Development

Four modeling approaches were compared: stepwise multiple linear regression (SMLR), SPA-MLR, partial least squares regression (PLSR), and SPA-PLSR. Model performance was evaluated using three metrics: coefficient of determination for validation (R_{val}^2), ratio of performance to deviation (RPD), and root mean square error of prediction (RMSEP). Models with $Rval^2 > 0.9$, RPD > 3, and low RMSEP were considered to have excellent predictive capability.

2 Results and Discussion

2.1 Model Performance with Different Pretreatments

Table 2 presents the performance of the SMLR model under different pretreatment conditions. The VN-pretreated spectra yielded the best results with RMSEP = 0.2876, $Rval^2$ = 0.9792, and RPD = 9.9078. MSC and SNV also performed well, with $Rval^2$ values exceeding 0.95 and RPD values above 6.5. The derivative methods (1-Der and 2-Der) improved model performance compared to raw spectra but were slightly inferior to VN, MSC, and SNV.

2.2 Comparison of Modeling Methods

Figure 2 illustrates the prediction models for soil salt content based on VN-pretreated spectra. The SPA-MLR model demonstrated superior performance compared to SMLR, with fewer selected wavelengths and higher accuracy. Similarly, SPA-PLSR showed improvement over conventional PLSR, though the enhancement was less pronounced than in the MLR case.

[Figure 2: see original paper]

Table 4 provides a comprehensive comparison of all four modeling approaches. The SPA-MLR model achieved the highest precision: RMSEP = 0.2783, $Rval^2$ = 0.9805, and RPD = 11.50, using only a small subset of wavelengths. This represents a significant improvement over the full-spectrum PLSR model (RMSEP = 0.3616, $Rval^2$ = 0.9568, RPD = 6.9221) and demonstrates that careful wavelength selection is crucial for building robust and parsimonious models.

The results indicate that VN is the most effective pretreatment method for soil salinity estimation, as it normalizes spectral vectors and reduces multiplicative scattering effects. The combination of SPA with MLR provides an optimal balance between model simplicity and predictive accuracy, making it particularly suitable for field applications where rapid assessment is required.

2.3 Practical Implications

For monitoring soil salinity in jujube orchards of arid regions, the SPA-MLR model based on VN-pretreated spectra offers the best solution. The model' s high RPD value (>11) indicates excellent predictive capability, while the small number of selected wavelengths facilitates instrument development and field deployment. Future research should focus on validating these models across different soil types and environmental conditions.

References

[1] PANG G, TAO W, JIE L, et al. Quantitative model based on field-derived spectral characteristics to estimate soil salinity in Minqin County, China[J]. Soil Science Society of America Journal, 2014, 78(2): 546-555.

[2] ZHAO Z, TASHPOLAT T, ZHANG F, et al. Spectral characteristics of soil salt content in typical oasis of Tarim River' s middle reaches[J]. Journal of Natural Disasters, 2012, 21(5): 72-78.

[3] WU M, LI X, SHA J. Spectral inversion models for prediction of red soil total nitrogen content in subtropical region (Fuzhou)[J]. Spectroscopy & Spectral Analysis, 2013, 33(11): 3111-3115.

[4] NAGY A, PETERRICZU U, GALYAB, et al. Spectral estimation of soil water content in visible and near infra-red range[J]. 2014, 3(3): 163-171.

[5] ZHANG P, LI Y. Study on the comparisons of the establishment of two mathematical modeling methods for soil organic matter content based on spectral reflectance[J]. Spectroscopy & Spectral Analysis, 2016, 36(3): 903-910.

[6] ZHANG X, HUANG B, JI J, et al. Quantitative prediction of soil salinity content with visible-near infrared hyperspectral in northeast China[J]. Spectroscopy & Spectral Analysis, 2012, 32(8): 2075-2079.

[7] YANG A, DING J, LI Y, et al. Apparent electrical conductivity and measured spectral for monitoring soil salt content in Arid Lakeside Wetland[J]. Journal of Desert Research, 2016, 36(5): 1365-1373.

[8] GOLDSHLGER N, CHUDNOVSKY A, BEN-BINYAMIN R. Predicting salinity in tomato using soil reflectance spectra[J]. International Journal of Remote Sensing, 2013, 34(17): 6079-6093.

[9] NETO O R, TEIXEIRA A, LEAO R, et al. Hyperspectral remote sensing for detecting soil salinization using ProSpecTIR-VS aerial imagery and sensor simulation[J]. Remote Sensing, 2017, 9(1): 1.

[10] PENG J, WANG J, XIANG H, et al. Comparative study on hyperspectral inversion accuracy of soil salt content and electrical conductivity[J]. Spectroscopy & Spectral Analysis, 2014, 34(2): 510-514.

[11] QU Y, DUAN X, GAO H, et al. Quantitative retrieval of soil salinity using hyperspectral data in the region of Inner Mongolia hetao irrigation district[J]. Spectroscopy & Spectral Analysis, 2009, 29(5): 1362-1366.

[12] DING J, WU M, LIU H, et al. Study on the soil salinization monitoring based on synthetical hyperspectral index[J]. Spectroscopy & Spectral Analysis, 2012, 32(7): 1918-1922.

[13] DUAN P, XIONG H, LI R, et al. A quantitative analysis of the reflectance of the saline soil under different disturbance extent[J]. Spectroscopy & Spectral Analysis, 2017, 37(2): 571-576.

[14] LI Y, QIAO J, DONG T, et al. Hyperspectral inversion of soil water and salt content in soils with different textures[J]. Chinese Journal of Applied Ecology, 2016, 27(12): 3807-3815.

[15] LIU Y, PAN X, WANG C, et al. Prediction of coastal saline soil salinity based on vis-nir reflectance spectroscopy[J]. Acta Pedologica Sinica, 2012, 49(4): 824-829.

[16] ZHAN X Y, ZHAO N, LIN Z Z, et al. Effect of algorithms for calibration set selection on quantitatively determining asiaticoside content in centella total glucosides by near infrared spectroscopy[J]. Spectroscopy & Spectral Analysis, 2014, 34(12): 3267-3272.

[17] NI Z, HU C, FENG F. Progress and effect of spectral data pretreatment in NIR analytical technique[J]. Chinese Journal of Pharmaceutical Analysis, 2008, 28(5): 824-829.

[18] XIA J, LI P, LI X, et al. Effect of different pretreatment method of non-destructive measure vitamin C content of umbilical orange with near-infrared spectroscopy[J]. Transactions of the Chinese Society for Agricultural Machinery, 2007, 38(6): 107-111.

[19] LIU Y, PAN X, WANG C, et al. Predicting soil salinity based on spectral symmetry under wet soil condition[J]. Spectroscopy & Spectral Analysis, 2013, 33(10): 2771-2776.

[20] CHEN W, DING J, TAN J, et al. Analysis vegetation of hyperspectral reflectance and shallow soil water content in arid area[J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, (9): 1-16.

[21] HOU Y, TIYIP T, SAWUT M, et al. Estimation model of desert soil organic matter content using hyperspectral data[J]. Transactions of the Chinese Society of Agricultural Engineering, 2014, 30(16): 113-120.

[22] PENG X, HU D, ZENG W, et al. Estimating soil moisture from hyperspectral in saline soil based on EPO-PLS regression[J]. Transactions of the Chinese Society for Agricultural Machinery, 2016, 32(11): 167-173.

[23] ZENG W Z, HUANG J S, XU C, et al. Hyperspectral reflectance models for soil salt content by filtering methods and waveband selection[J]. Ecological Chemistry & Engineering S, 2016, 23(1): 117-130.

[24] FARIFTEH J, MEER F V D, ATZBERGER C, et al. Quantitative analysis of salt-affected soil reflectance spectra: A comparison of two adaptive methods (PLSR and ANN)[J]. Remote Sensing of Environment, 2007, 110(1): 59-78.

Note: Figure translations are in progress. See original paper for figures.