Estimation of Soil Total Nitrogen and Soil Moisture Based on NIRS Technology

Xiaofei An, Minzan Li, Lihua Zheng

Estimation of Soil Total Nitrogen and Soil Moisture based on NIRS Technology

Xiaofei An[1], Minzan Li[1] and Lihua Zheng[1]

[1] Key Laboratory of Modern Precision Agriculture System Integration Research,
China Agricultural University, Ministry of Education, Beijing, China
Email: limz@cau.edu.cn

Abstract. Estimation model between soil moisture content and the near infrared reflectance was established by the linear regression method and the models between soil total nitrogen content and the near infrared reflectance were also established by the BP neural network method and Support Vector Machine (SVM) method. Forty-eight soil samples were collected from China Agricultural University Experimental Farm. After the soil samples were taken into the laboratory, NIR absorbance spectra were rapidly measured under the original conditions by the FT-NIR (Fourier Transform Near Infrared Spectrum) analyzer. At the same time the soil moisture (SM) and soil total nitrogen (TN) were measured by the laboratory analysis methods. The results of the study showed that a linear regression method achieved an excellent regress effect for soil moisture. The correlation coefficient of the calibration (R_C) was 0.88, and the correlation coefficient of the validation (R_V) was 0.85. The model was passed F test and t test. For soil total nitrogen, the model effect of BP neural network was better than that of SVM method, and the correlation coefficient of the calibration (R_C) coefficient and the validation (R_V) was 0.92 and 0.88, respectively. Both RMSE and PMSE were low. The results provided an important reference for the development of a portable detector.

Keywords: Soil total nitrogen, Soil moisture, BP neural network, Support vector machine

1   Introduction

Precision agriculture requires a clear understanding of the spatial variability of soil characteristics and the nutrient status in real time. Spectroscopy, as a rapid, non-destructive, simple, and green measure and analysis technique, plays an important role in the determination of soil nutrients [1]. Research based on real-time spectral analysis for determination of soil nutrients in precision agriculture can provide accurate scientific data and theoretical basis for the development of a low-cost and intelligent measure instrument.

Many scholars have researched on the relation between the soil parameters and the spectral characteristics in the 20th century. Alabhas et al (1972) found that soil organic matter in the near infrared area had characteristics related with several functional groups in the organic compound[2]; Li et al (1999) analyzed soil parameters, such as NO3-N, SOM, EC and pH based on NIRS, and then determined

sensitive wavebands of the parameter[3]; Sun Jianying (2006) measured soil data of NIRS by a FT-NIR analyzer at the wavelength range of 830-2500 nm and several forecasting models for soil moisture, soil organic matter and soil total nitrogen were established[4];Zheng Lihua (2010) established the forecasting models for soil total nitrogen and soil organic matter by WT (Wavelet Transform) and SVM method[5]; Gao Hongzhi (2011) selected nine wavelengths to establish the forecasting models of soil total nitrogen and soil organic matter based on actual experience [6].

Remarkable progress has been made in the forecasting of soil moisture, soil organic matter and soil total nitrogen based on NIRS[7-14].However, above studies provided the dried and crushed soil samples in the experiment so that the results obtained could not be directly used to real-time evaluate soil parameters. Hence, it was important to reveal the feasibility and possibility of analyzing soil parameters with NIR spectroscopic techniques using raw soil samples and provide a reference for the development of a portable detector.

2 Materials and Methods

2.1 Experiments and collection

Forty-eight soil samples were collected from the experimental field of China Agricultural University located in the suburb of Beijing. Soil samples were collected from different experimental areas and depth for the different levels. A double-layer sealing plastic bag and kraft paper bag were adopted to avoid soil moisture loss. The NIRS of the forty-eight soil samples were measured in the laboratory by a MATRIX-I type of FT-NIR analyzer with a rotating sample pool produced by the Germany Bruker Optical Company. About 15 grams of soil sample could be directly placed into the rotating sample pool by a forty-eight mm diameter of quartz glass. The spectral measure range was 12798~3599 cm-1(781~2779nm), resolution 4 cm-1, wavenumber accuracy 0.1 cm-1, scan times 32. A spectral curve could be obtained after taking a small amount of soil sample into the rotating sample pool and setting the relevant parameters. Fig.1 shows several spectral curves of soil samples in the range of 850~2450 nm.

Soil moisture contents and soil total nitrogen contents were measured by physical method and chemical method respectively in the laboratory. Soil moisture contents were measured by the drying and weighing method with an electric heater, and soil total nitrogen contents were measured by the semi-micro kjeldahl determination with a FOSS2300 Kjeldahl nitrogen instrument.
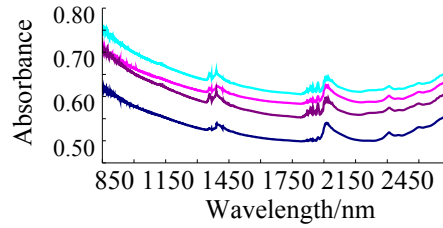
Fig. 1. Spectral curves of soil samples

2.2 Modeling

Single linear regression analysis is a mathematical statistical method to deal with the correlation between variables [14]. A linear regression method can be used for the obvious correlation between soil moisture contents and the absorbance of the soil samples. But the correlation between the soil total nitrogen content and the absorbance was lower than 0.3 in full spectral range. In addition, several papers demonstrated that there was weak correlation and nonlinear relation between soil total nitrogen contents and the near infrared spectra [4-9]. So that it was a better attempt to estimate the soil total nitrogen content by a nonlinear prediction model. In our study, soil total nitrogen contents were predicted by BP neural network method and support vector machine method respectively.

BP neural network is back-propagation network using nonlinear differentiable function to serve as weight value train [13,14]. According to the theorem of BP neural network, single hidden layer BP network can be appropriated for any continuous function in a closed interval. A three layers BP network can complete the mapping from uncertain dimension to determined dimension. This network has strong nonlinear mapping ability, which is very suitable for nonlinear function approximation. In the training period, the error is inverted propagation.

Support vector machine was proposed in 1992 by Vapnik et al [15].One of the most important characters in SVM method is to generate hypotheses for the sparse matrix of dual representation, according to the general theory of reinforcement learning. This makes it become a very effective algorithm. Another important characteristic is that the corresponding optimization problem becomes convex problem and no local minimum, because of the introduction of kernel function. SVM has the advantages of simple structure and the solution of the small sample, nonlinear, high dimension and local minima, compared with the traditional BP neural network.

3 Results and discussion

3.1 Data preprocessing

Statistic characteristics of soil moisture contents and soil total nitrogen contents are shown in Fig.2. In the box graph, the first column is the soil moisture content distribution and the second column is the soil total nitrogen content distribution. It was observed from Table 1 that the standard deviation of the soil moisture content was larger than that of the soil total nitrogen content.
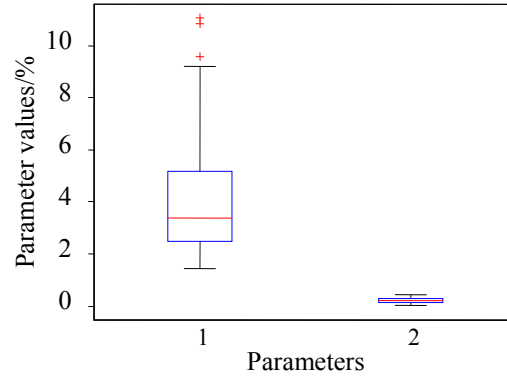

Fig.2. Statistic characteristics of soil samples

Table 1. The measured values of two soil parameters in the laboratory

Parameters	The number of samples	Min /%	Max /%	Average /%	Mean/ %	Standard deviation
SM	48	1.44	11.05	4.18	3.34	2.54
TN	48	0.04	0.45	0.22	0.21	0.10

The soil spectral reflectance needed to be smoothed since the white noise interference could be weakened after smoothing processing in a certain degree. The first deviation could remove the baseline wander or flat background interference effects and improve the original spectral resolution and spectral profile transformation. The forecasting model of soil moisture content was established by the original absorbance spectral data, while the forecasting model of soil total nitrogen content was established by using the data processed by the first deviation of the spectral absorbance. The first deviation of the spectral absorbance was calculated using the Equation (1) [15,16] ：

$$y_i = \frac{1}{6h}(-11x_i + 18x_{i-1} - 9x_{i-2} + 2x_{i-3}) \tag{1}$$

Where: y_i is the first derivative, x_i is the NIR absorbance, h is the sampling interval.

3.2 Estimation of soil moisture

The forecasting model of soil moisture content was established by the linear regression method with the original absorbance spectral data. All the forty-eight soil samples were divided into two parts: calibration set (including 36 soil samples) and validation set (including 12 soil samples). Fig.3 shows the comparisons between measured data and predicted data of soil moisture. The model formula is shown in Equation (2). The correlation coefficient of the calibration (R_C) was 0.88, and the correlation coefficient of the validation (Rv) was 0.85. The model was passed the F test and t test, which illustrated that the model was significant and the parameters were significant. It was proved that the wavelength of 1450 nm could be taken as an initiative light source selection for a portable detector.

$$y_i = 43.72 \times x_i - 24.61 \qquad (2)$$

Where: y_i is prediction of soil moisture content; x_i is absorbance of soil at the wavelength of 1450 nm.
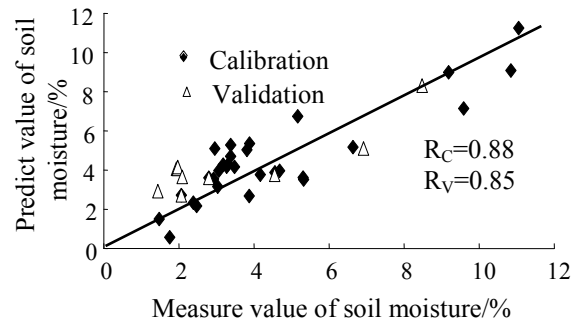

Fig.3. Comparisons between measured data and predicted data of soil moisture

3.3 Estimation of soil total nitrogen by SVM

The correlation between soil moisture and absorbance at 1450 nm was observed, while the correlation coefficients between absorbance and soil total nitrogen were less than 0.3 in full spectral range. It showed difficult for use linear regression method to complete the prediction of soil total nitrogen. SVM and BP neural network were respectively chosen to establish the soil total nitrogen forecasting model.

There was no unified model to determine parameters of SVM. The parameters could be chosen by virtue of experience, experiment or a large range of search methods. After a lot of attempts, the parameters were determined: the kernel function was RBF, penalty parameter C was 11.314 and kernel parameter g was 2.0. Table 2

shows all the parameters. All the first deviation data was divided into two parts: training set and validation set. The data were used for calibration and validation by the SVM method after normalization.

Table 2. Parameters of SVM

Parameters	Parameter values
kernel function	RBF
penalty parameter C	11.314
kernel parameter g	2.0
MSE	0.0124
input variants	6
output variant	1

Fig.4 shows the comparisons between measured value and predicted value of soil total nitrogen by SVM. The correlation coefficient of the calibration (R_C) was 0.89; the correlation coefficient of the validation (R_V) was 0.78. The result of calibration and validation demonstrated that the model to evaluate TN is feasible. The results showed could be taken as theoretic basis for the analysis of real-time TN.
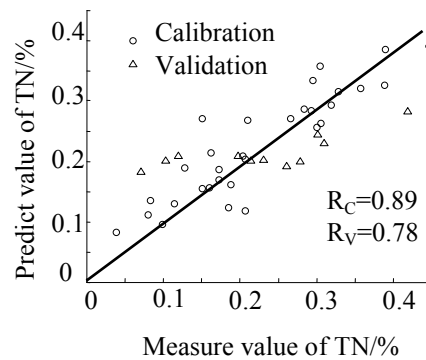

Fig.4. Comparisons between measure value and predict value of TN by SVM

3.4 Estimation of TN by BP neural network

BP neural network can avoid adverse effects of abnormal data caused by equipment, human factor and etc. It makes the results meet practical prediction accuracy. A three layers BP neural network was established, and S type functions were selected as transfer function of each layer. The input neurons were 6 and output neuron was 1. The neurons in hidden layer were 27, error value of 0.005. Other parameters such as

momentum factor, step length and initial weights value were set in a certain range of random by the program. The input neurons were the first deviation of absorbance of soil samples at six wavelengths (940, 1050, 1100, 1200, 1300, 1550 nm), and the output neuron stood for the soil total nitrogen.
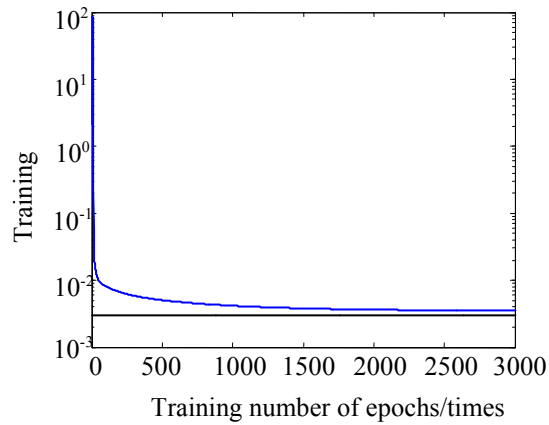

Fig.5. Bp NN training curve

Table 3. Bp neural network parameters

Parameters of BP NN	Parameter values
input neurons	6
output neurons	1
neurons in hidden layer	27
network structure	single hidden layer
transfer function in input layer	tansig
transfer function in output layer	purelin
training function	trainlm
error	0.00343

All the forty-eight soil samples were divided into two parts: calibration set (including 36 soil samples) and validation set (including 12 soil samples). The correlation coefficient of the calibration (R_C) was 0.92, and the correlation coefficient of the validation (Rv) was 0.88.

The forecasting effect of BP neural network was better than that of SVM method. It could provide a model reference for the development of a portable detector. Fig.6 shows the comparisons between measured dataset and predicted dataset of soil total nitrogen by BP neural network.
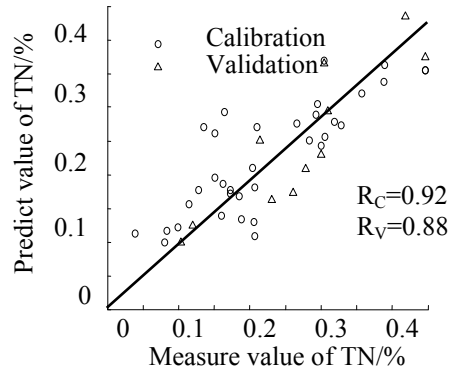
Fig.6. Comparisons between measured dataset and predicted dataset
of TN by BP neural network

4 Conclusions

The forecasting models of soil moisture content and soil total nitrogen were established by the linear regression method and nonlinear regression methods respectively. The following conclusions were obtained.

1. The soil moisture content could be predicted by a linear regression method, and the correlation coefficient of the calibration (R_C) and the validation (Rv) were 0.88 and 0.85, respectively.
2. The first deviation of absorbance data could effectively reduce the system error of the network model. The forecasting model of soil total nitrogen content established by the BP neural network was better than that of SVM method, and the correlation coefficient of the calibration (R_C) and the validation (R_V) were 0.92 and 0.88, respectively.
3. The models were robust. The accuracy of the model would not be affected when the number of the samples was many enough even no abnormal data were eliminated. All the seven wavelengths chosen in the model could be found the corresponding single wavelength LED in the market. Thus, it provided an excellent model reference for the development of a portable detector.

References

1. Li, M.Z.: Evaluating Soil Parameters with Visible Spectroscopy. J. Transactions of the CSAE. 19(5), 36--42 (2003)

2. Al-Abbas, A.H., Swain, P.H., Baumgardner, M.F.: Relating Organic Matter and Clay Content to the Multispectral Radiance of soils. J. Soil Science. 114(6), 477--485 (1973)

3. Li, M.Z., Shibusawa, S., Sasao, A.: Spectroscopic Approach to Soil Parameters Sensing. In: Nir 99 International conference on Agricultural Engineering, pp. 98--103.Beijing (1999)

4. Sun, J.Y., Li, M.Z., Zheng, L.H.: Real-time Analysis of Soil Moisture, Soil Organic Matter, and Soil Total Nitrogen with NIR Spectra. J. Spectroscopy and Spectral Analysis. 26(3), 426--429 (2006)

5. Zheng, L.H., Li, M.Z., An, X.F.: Forecasting Soil Parameters Based on NIR and SVM. J.Journal of Agricultural Engineering. 26(2):81--87 (2010)

6. Gao, H.Z, Lu, Q.P.: Near Infrared Spectral Analysis and Measuring System for Primary Nutrient of Soil. J.Spectroscopy and Spectral Analysis. 31(5),1245--1249 (2011)

7. Yu, F.J., Min, S.G., Huang, X.T. : Near Infrared Spectrum Analysis of Soil Organic Matter and Nitrogen. J.Analysis Laboratory. 11(3), 47--51(2002)

8. Zheng, L.H., Li, M.Z., Sun, J.Y.: Evaluation of Soil Fertility with Spectrophotometer and Spectroradiometer. J.Infrared and Photoelectronic Ima gers and Detector Devices. Proceedings of SPIE Volume (5881)138—146 (2 001)

9. Chen, P.F., Liu, L.Y., Wang, J.H.: Near Infrared Spectroscopy Real-time Measuring Soil Total Nitrogen and Phosphorus Content. J.Spectroscopy and Spectral Analysis .57 (2), 295--298 (2008)

10. Yuan, S.L., Ma, T.Y., Song, T.: Total Nitrogen in Soil and the Total Phosphorus Content in the Near Infrared Spectrum of Real-time Detection Methods. J.Journal of Agricultural Machinery. 9 (S1), 150--153 (2009)

11. Barthes, B G, Didier, B, Edmond, H.: Determining the Distributions of Soil Carbon and Nitrogen in Particle Size Fractions using Near-infrared Reflectance Spectrum of Bulk Soil Samples. J.Soil Biology & Biochemistry.40 (6), 1533--1537 (2008)

12. Song, H.Y.: Based on the OSC and PLS Soil Organic Matter of Near Infrared Spectrometry. J. Journal of Agricultural Machinery. 38 (12), 113--115 (2007)

13. Sun, J.Y, Li, M.Z., Zheng, L.H.:.Based on the North of the Near Infrared Spectrum Chao Soil Parameter Real-time Analysis. J. Spectroscopy and Spectral Analysis .26 (5), 426 --429 (2006)

14. Zheng, L.H, Li, M.Z., Pan, L.: Based on the Soil of Near Infrared Spectral Technology Parameters of the BP Neural Network Forecast. J.Spectroscopy and Spectral Analysis.28 (5), 1160--1164 (2008)

15. Li, M.Z.: Spectral Analysis Technology and Application. China Science Press. Beijing (2008)

16. Lu, W.Z.: Modem Near-Infrared Spectroscopy Analytical Technology.China Petrochemical Press. Beijing. (2006).