Near-infrared spectroscopy technology for Soil nutrients detection based on LS-SVM

Yandan Qiao[1], Shujuan Zhang[2,*]

1 Center for Popularization of Agricultural Machinery Technology of Shanxi Province

2 College of Engineering , Shanxi Agricultural University,Taigu,Shanxi, 030801, China

zsujuan@263.net

Abstract: The detection method of the soil nutrients (organic matter and available N, P, K) were analyzed based on the near infrared spectroscopy technology in order to decision-making for precision fertilization. 54 samples with 7m×7m was collected using DGPS receiver positioning in a soybean field. The soil organic matter, available nitrogen (N), available phosphorus (P), available potassium (K) content was determined, the near-infrared diffuse reflectance spectrum of the soil samples were obtained by FieldSpec3 spectrometer. 54 samples were randomly divided into 40 prediction sets and 14 validation sets. After smoothing, the eight principal components of original spectra were extracted by principal component analysis (PCA). Prediction model of soil organic matter, available nitrogen (N), available phosphorus (P), potassium (K) were respectively established with the eight principal component as input and soil nutrients by measured as the output, and the 14 validation samples were predicted. The results showed that the soil organic matter, available nitrogen (N), available phosphorus (P), potassium (K) prediction model were set up with principal component analysis and LS-SVM, which the correlation coefficients between the prediction value and measurement value were 0.8708, 0.7206, 0.8421 and 0.6858, the relative errors of the LS-SVM prediction was smaller and those mean values were 1.09%, 1.06%, 4.08% and 0.69%. The method of soil organic matter content prediction is feasible.

Keywords: Near infrared spectroscopy, Soil nutrient, Detection, PCA, LS-SVM

1   Introduction

Soil nutrients are fast nondestructive measurement for agriculture information collection (LUO et al.,2006). Recently, with widely applications of near-infrared spectroscopy, it has become a focus of many domestic scholars that soil nutrient information were gained by near infrared spectroscopy(Ben-Dor E et al.,1995).The spectral characteristics of the soil was analyzed of mechanism(Wu Yun Zhao et al.,2003). The nitrogen, calcium and magnesium of dry, after sifting the soil was studied using near infrared spectroscopy(Lee et al.,2003); soil constituents were

studied by near infrared spectroscopy(Chang et al.,2001). Yu Fei Jian found that NIRS and total nitrogen, organic matter, nitrogen of after sifting the soil has a good correlation. The Near-infrared spectral characteristics of the soil after sifting treatment were studied by He Yong et al. The exchange capacity of Soil organic matter, cation and soil moisture were studied using a portable NIRS spectrometer in the laboratory and field (Sudduth et al.,1993).

Near infrared reflectance(NIR) spectroscopy(NIRS) is a physical non-destructive, rapid, reproducible and low-cost approach that characterizes materials according to their reflectance in the wavelength range between 800 and 2500 nm. The analysis of NIR spectra relies on calibration, which in general is a multivariate regression procedure that that expresses a given property, determined using a conventional method, as a function of absorbance at all or selected wavelengths of the NIR region. The calibration equation can then be used to predict that property on new samples from their NIR spectra only, the acquisition of which is time- and cost-effective (<1 min per sample, no consumables required). The application of NIRS to soil has been mentioned from the 1960sand it has been used extensively to determine soil content in carbon and nitrogen(Bernard et al.,2011).

Accurate NIRS determination of C and N have been extensively reported, especially for rather homogeneous textural sets (Dalal and Henry,1986;Morra et al.,1991;Brunet et a.,2007).Several studied have also reported NIRS determination of C/N(Chang and Larid,2002;Ludwig et al.,2002)The main objective of this study was to investigate the potential of NIR reflectance spectroscopy as a rapid tool for the measurement of soil nutrients.

2 Materials and methods

2.1 Soil Samples

The soil was sampled to 7m ×7m grid using DGPS navigation system in a soybean field of Shanxi Agricultural University campus on November 9, 2009. Depth of samples was 20cm and the number of samples was 54. The soil type and texture is cinnamon and sandy loam.

2.2 Measurement of spectral characteristics of the soil

The test device was composed of computer, spectrometer, halogen light, white board and other components correct in the spectral measurements of soil. Spectrum was obtained using FieldSpec3 spectrometer (Analytical Spectral Device company, USA), with sampling collection interval of 1nm, spectrum range of 350 ~ 2500nm, and the resolution of 3.5nm. 54 soil samples were air-dried, hand ground, and samples of particles of 2mm were got. The glass Petri dishes that diameter was 90mm and

thickness was 15mm were selected to place the soil.Spectrometer was placed in soil samples and the distance between spectrometer and soil surface was 100nm and test angle was 45 °, field view probe was 10°. Sample spectrum was collected with diffuse reflection mode. Each soil sample was six repeated and each sample was scanned 10 times. The spectrum was processed for average processing by ASD View Spec pro software, and the absorbance was converted according to log [1/R]. The spectral data was exported in ASCII format, and processed was use of ASD View Spec Pro, Unscramble V9.8 and MATLAB2009.

2.3   Measurement of soil nutrients

The soil nutrients of soil organic matter, nitrogen, available phosphorus and potassium were determined by routine laboratory analysis. The measured were divided into two steps: firstly, the elements of analysis were extracted from the isolated from the soil; then those elements were quantitative tested. The soil was dried, sieved, extraction, those element were obtained using the following chemical analysis methods:

1)Soil organic matter content was determined by potassium dichromate volumetric method - external heating method.

2)Soil nitrogen content was determined using 1mol / L NaOH diffusion.

3)Soil available phosphorus content was measured using 0.5mol / L $NaHCO_3$ extraction - molybdenum blue method.

4)Soil potassium was measured using 1mol/L NH_4OA_C extraction - flame photometry.

The statistical parameters of chemical measurement results of soil organic matter (SOM), available nitrogen (N), phosphorus (P), available potassium (K) were shown in Table 1.

Table 1.   Soil chemical constituents statistical parameters

Nutrients	Sample number	Minimum	Maximum	Mean	Standard Deviation	Coefficient of variation(%)
N(mg/kg)	54	30.63	40.86	35.77	1.89	5.28
P(mg/kg)	54	3.23	16.34	8.04	2.69	33.40
K(mg/kg)	54	209.45	287.55	235.56	15.30	6.50
SOM(g/kg)	54	19.00	29.47	24.53	1.61	6.55

3   Soil available N, P, K and organic matter spectrum

3.1   Data Preprocessing

In order to remove the effects from the high-frequency random noise, baseline drift, uneven samples and light scattering, spectra was processed using average smoothing, and smoothing window size was 9. The soil near infrared spectra of pretreatment was shown in Figure 1.
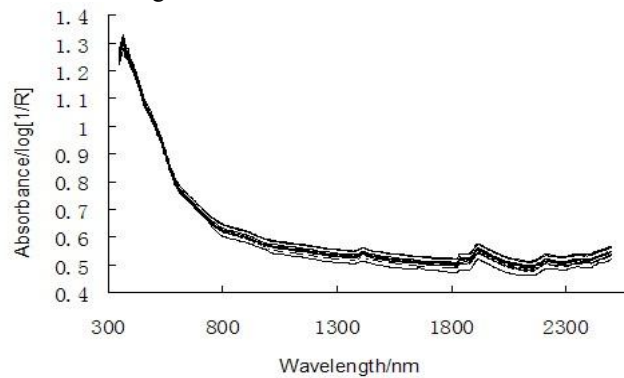

Fig. 1.   Absorbance graph of soil near-infrared spectroscopy

3.2 Principal component analysis of soil nutrients and the least squares support vector machine Modeling

3.2.1 Principal components are extracted based on principal component analysis

The 54 samples were randomly divided into test set of 40 and validation set of 14. The spectral bands of samples had total 2151 points from 350 ~ 2500nm, those computationally intensive and the spectral information of samples in some regions was very weak. The correlation lack between nature and the spectral bands. Principal component analysis(PCA) takes a dimension reduction method, The original number of indicators transform into a few new indicators using linear algebra and related knowledge, both of these new indicators unrelated to each other to avoid overlap and duplication of information, but also a comprehensive reflection of the original number of indicators. The sets of validation and prediction were analyzed by PCA, the credibility of its main components was shown in Table 2.

Table 2.   Accumulative reliabilities of the first 6 PCs

PC	PC01	PC02	PC03	PC04	PC05	PC06
Cumulative reliability prediction set (%)	96.829	99.180	99.624	99.823	99.882	99.919
Accumulated credibility validation set (%)	96.087	98.865	99.551	99.780	99.859	99.906

The explanation of original variables was collected through the accumulated credibility. The accumulated credibility of the first six principal components was to 99.9% in prediction set from Table 2, and 99.9% of the original variable wavelength was interpreted; The accumulated credibility of the first six principal components was to 99.906% in validation set, and 99.9% of the original variable wavelength was interpreted, Least squares vector machine model was established that the six principal components were selected as input.

3.2.2 The predictive models of soil organic matter based on least squares vector machines (LS-SVM)

The LS-SVM model was established with 6 principal components as an input layer and measured values of soil organic matter as output layer. Hyper-parameters gam and the RBF kernel parameter sig2 were obtained by grid search method with cross-validation, The optimal combination of two parameters were obtained that gam was 1251.8, and sig2 was 174.6. The set of 14 forecasts was predicted and correlation analysis between predicted and measured values was shown in Figure 2. The error analysis results was shown in Table 3.
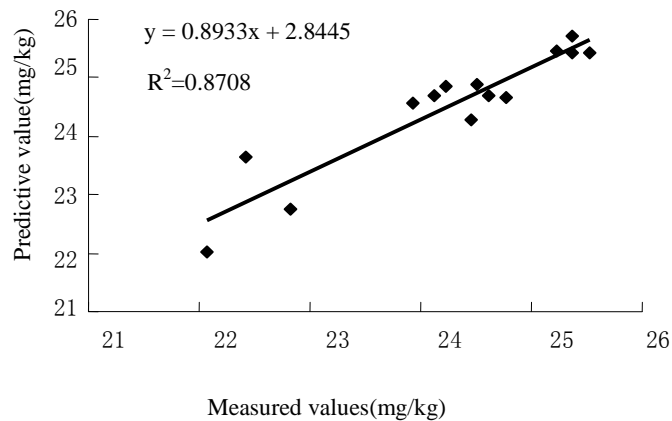

Fig. 2. Relationship between chemical analysis values and prediction values of SOM

Table 3. Prediction results for SOM by LS-SVM model

Sample No.	Measured values	Predictive value	Absolute error	Relative error
1	22.07	22.02	-0.05	-0.25
2	23.93	24.56	0.63	2.63

3	25.53	25.42	-0.11	-0.43
4	25.22	25.47	0.25	0.99
5	25.37	25.70	0.33	1.29
6	24.24	24.84	0.60	2.50
7	25.37	25.44	0.07	0.27
8	24.62	24.69	0.08	0.31
9	24.46	24.29	-0.17	-0.72
10	22.42	23.64	1.22	5.48
11	22.83	22.75	-0.08	-0.38
12	24.77	24.67	-0.10	-0.41
13	24.50	24.90	0.40	1.62
14	24.12	24.69	0.57	2.33

The coefficient of determination of prediction set between the predicted value and the measured values was 0.8708. The maximum absolute error of soil organic matter was 1.23mg/kg, and the average absolute of relative error was1.09%.

3.2.3 The predictive models of Soil available N based on least squares vector machines (LS-SVM)

The LS-SVM model was established with 6 principal components as an input layer and measured values of soil organic matter as output layer. Hyper-parameters gam and the RBF kernel parameter sig2 were obtained by grid search method with cross-validation, The optimal combination of two parameters were obtained that gam was 13.5, and sig2 was 17.0. The set of 14 forecasts was predicted and correlation analysis between predicted and measured values was shown in Figure 3. The error analysis results was shown in Table 4.
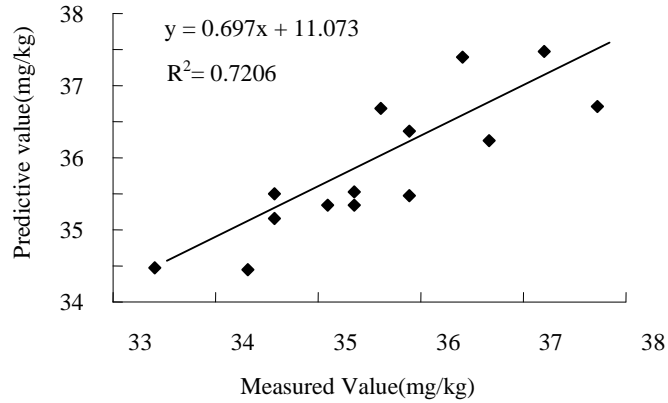
y = 0.697x + 11.073

R^2 = 0.7206

Fig. 3. Relationship between chemical analysis values and prediction values of available N

The coefficient of determination of prediction set between the predicted value and the measured values was 0.7206. The maximum absolute error of soil organic matter was 1.92mg/kg, and the average absolute of relative error was 1.06%.

3.2.4 The predictive models of Soil available P based on least squares vector machines (LS-SVM)

The LS-SVM model was established that 6 principal components as an input layer and measured values of soil organic matter as output layer. Hyper-parameters gam and the RBF kernel parameter sig2 were obtained by grid search method with cross-validation, The optimal combination of two parameters were obtained that gam was 47.7, and sig2 was 153.3. The set of 14 forecasts was predicted and correlation analysis between predicted and measured values was shown in Figure 4. The error analysis results was shown in Table 5

Table 4. Prediction results for available N by LS-SVM model

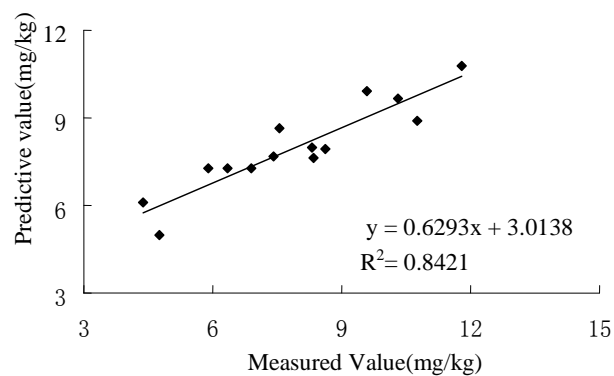| Sample No.	Measured values	Predictive value	Absolute error	Relative error
1	33.40	34.47	1.07	3.20
2	34.57	35.16	0.59	1.70
3	34.57	35.49	1.92	5.55
4	36.67	36.24	-0.43	-1.17
5	35.36	35.35	-0.01	-0.02
6	35.36	35.54	-0.82	-2.32
7	35.62	36.68	1.06	2.97
8	35.10	35.34	0.24	0.68
9	35.88	35.47	-0.41	-1.16
10	34.31	34.45	0.14	0.42
11	35.88	36.38	1.50	4.17
12	37.72	36.71	-1.01	-2.67
13	37.20	37.48	0.28	0.75
14	36.41	37.40	0.99	2.73

.

$y = 0.6293x + 3.0138$

$R^2 = 0.8421$

Fig. 4. Relationship between chemical analysis values and prediction values of available P

Table 5. Prediction results for available P by LS-SVM model

Sample No.	Measured values	Predictive value	Absolute error	Relative error
1	8.36	7.62	-0.74	-8.85
2	5.89	7.29	1.40	23.73
3	4.37	6.08	2.71	62.10
4	4.75	5.00	0.25	5.31

5	7.41	7.68	-1.73	-23.34
6	8.61	8.93	-0.68	-7.85
7	8.30	8.00	-0.30	-3.56
8	6.33	7.25	0.92	14.60
9	7.54	8.66	1.12	14.82
10	11.78	10.76	-2.02	-17.19
11	9.59	9.90	1.31	13.61
12	6.90	7.27	0.37	5.30
13	10.32	9.66	-0.66	-6.37
14	10.77	8.90	-1.87	-17.41

The coefficient of determination of prediction set between the predicted value and the measured values was 0.8421. The maximum absolute error of soil organic matter was 1.71mg/kg, and the average absolute of relative error was 4.08%.

3.2.5 The predictive models of Soil available K based on least squares vector machines (LS-SVM)

The LS-SVM model was established with 6 principal components as an input layer and measured values of soil organic matter as output layer. Hyper-parameters gam and the RBF kernel parameter sig2 were obtained by grid search method with cross-validation, The optimal combination of two parameters were obtained that gam was 615.9, and sig2 was 153.3. The set of 14 forecasts was predicted and correlation analysis between predicted and measured values was shown in Figure 5. The error analysis results was shown in Table 6.
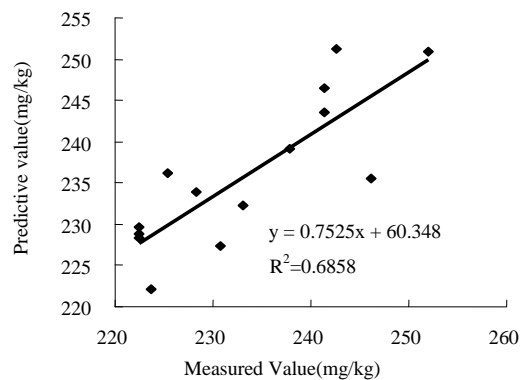
Table 6. Prediction results for available K by LS-SVM model

Sample No.	Measured values	Predictive value	Absolute error	Relative error
1	225.43	236.15	10.72	4.75
2	223.65	222.17	-1.48	-0.66
3	222.47	228.34	5.87	2.64
4	233.12	232.23	-0.89	-0.38
5	230.75	227.41	-3.34	-1.45
6	242.58	251.20	-14.36	-5.92
7	228.38	233.93	5.55	2.43
8	222.47	229.64	7.17	3.22
9	222.47	228.81	6.34	2.85
10	241.40	243.60	10.20	4.23
11	246.13	235.61	-10.52	-4.27
12	237.85	239.10	1.25	0.52
13	252.05	250.98	-1.07	-0.43
14	241.4	246.43	5.03	2.08

The coefficient of determination of prediction set between the predicted value and the measured values is 0.6858. The maximum absolute error of soil organic matter was 10.72mg/kg, and the average absolute of relative error was 0.69%.

There are been a few attempts to predict variables related to the soil by NIR reflectance spectroscopy. Good predictions(0.81<R2<0.90) were obtained for exchangeable calcium and magnesium, water soluble carbon, water holding capacity and urease activity (ZORNOZA,et al,.2008). The prediction of K, P and Na was classified as good(R2pre=0.68-0.74 and RPD=1.77–1.94), where quantitative predictions were considered possible. It is recommended to adopt BPNN-LVs modeling technique for higher accuracy measurement of the selected soil properties with vis–NIR spectroscopy, in comparison with PCR, PLS and BPNN-PCs modeling techniques(A.M.Mouazen,et al,.2010).

4   Results and discussion

(1) The least squares support vector machine model of soil organic matter, available nitrogen, phosphorus, available potassium were established using principal component analysis, and predicted results are verified that the coefficient of determination of soil organic matter was 0.8708 and the average absolute of relative error was 1.09%. It's proved that the method is feasible to predict the soil organic matter content.

(2) The coefficient of determination of prediction model of soil available nitrogen and available potassium were 0.7206 and 0.6858 by Near-infrared spectroscopy, and the average of relative error were 1.06% and 0.69%. The predicted error of those models was small, but the coefficient of determination was not high. The coefficient of determination of soil phosphorus was 0.8421 and the average of relative error was 4.08%, the correlation between measured and predicted was high and the relative error was large.The prediction for the content of soil available nitrogen, phosphorus and potassium by Near-infrared spectroscopy is feasible and practical.

References

Ben-Dor,E.,Banin,A.:Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties [J]. Soil Science of American Journal,1995,59:364-372.

Bernard,G.,Barthe,D.,Brunet,B.R.,et al.:Near infrared reflectance spectroscopy (NIRS) could be used for characterization of soil nematode community[J].Soil Biology & Biochemistry,2011,43:1649-1659

Brunet, D., Barthes,B.G.,Chotte,J.L.,Feller,C.:Determination of carbon and nitrogen contents in Alfisols, Oxisols and Ultisols from Africa and Brazil using NIRS analysis: Effects of sample grinding and set heterogeneity[J],Geoderma,2007,139, 106–117.

Chang,C.,Laird,D.A.,Mausbach,M.J., et al.: Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties[J]. Soil Science of American Journal, 2001, 65: 480-490.

Chang, C.W., Laird, D.A. :Near-infrared reflectance spectroscopic analysis of soil C and N. Soil Science[J] ,2002,167, 110–116.

Dalal, R.C.,Henry, R.J, et al.:Simultaneous determination of moisture,organic carbon and total nitrogen by near infrared reflectance spectrophotometry[J]. Soil Science Society of America Journal,1986,50:120–123.

He,Y.,Song,H.Y.,Annia,G.P., et al. :A new approach to predict N,P,Kand OM content in a loamy mixed soil by using near infrared reflectance spectroscopy[J]. Lecture Notes in Computer Science,2005,3644:859-867.

Lee,W.S.,Sanchez,J.F., Mylavarapu,R,S,,et al.: Estimating chemical properties of Florida soil using spectral reflectance[J]. Transactions of the ASAE，2003,46(5):1443-1453.

Li, W.,Zhang,S.H.,Zhang,Q., et al.: Rapid prediction of available N,P and K content in soil using near-infrared reflectance spectroscopy[J]. Transactions of the Chinese Society of Agricultural Engineering, 2007,23(1):55-59.

Ludwig, B., Khanna, P.K., Bauhus, J., Hopmans, P. Near infrared spectroscopy of forest soils to determine chemical and biological properties related to soil sustainability[J]. Forest Ecology and Management.2002,171:121–132.

Luo Xiwen; Zang Ying; Zhou Zhiyan. Research progress in farming information acquisition technique for precision agriculture[J]. Transactions of the Chinese Society of Agricultural Engineering, 2006,22(1):167-173.

Morra, M.J., Hall, M.H., Freeborn, L.L. Carbon and nitrogen analysis of soil fractions using near-infrared reflectance spectroscopy[J]. Soil Science Society of America Journal , 1991,55:288–291.

Mouaze,A.M.,Kuang,B., Baerdemaeker J.D.,et al.: Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy[J].Geoderma,2010,158:23-31.

SHA Jin ming; CHEN Peng cheng; CHEN Song lin. Characteristics Analysis of Soil Spectrum Response Resulted From Organic Material[J]. Research of Soil and Water Conservation, 2003,10(2):21-24.

Sudduth K A, Hummel J. Soil organic matter, CEC and moisture sensing with a portable NIR spectrophotometer[J]. Transactions of the ASAE,1993,36(6):1571-1582.

R,Zornoza,C.Guerrero,J.Mataix-Solera,et al. Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils[J].Soil Biology & Biochemistry,2008,40:1923-1930.

WU Yun zhao ,TIAN Qing jiu,JI Jun feng, et al. Soil Remote Sensing Research Theory Method and Application[J]. Remote Sensing Information, 2003,(1):40-47.

YU Fei jian; Min Shun geng; Ju Xiao tang, et al. Determination the content of nitrogen and organic substance in dry soil by using near infrared diffusion reflectance spectroscopy[J]. Chinese Journal of Analysis Laboratory, 2002,21(3):49-51.

Zhu Dengsheng, Wu Di ,Song Haiyan, et al. Determination of organic matter contents and pH values of soil using near infrared spectroscopy[J]. Transactions of the Chinese Society of Agricultural Engineering, 2008,24(6):196-199.