

Application of Slope/Bias and Direct Standardization Algorithms to Correct the Effect of Soil Moisture for the Prediction of Soil Organic Matter Content Based on the Near Infrared Spectroscopy

WANG Shi-fang^{1,2,3}, HAN Ping^{1,3*}, SONG Hai-yan^{2*}, LIANG Gang^{1,3}, CHENG Xu²

1. Beijing Research Center for Agriculture Standards and Testing, Beijing 100097, China
2. College of Engineering, Shanxi Agricultural University, Taigu 030801, China
3. Beijing Municipal Key Laboratory of Agriculture Environment Monitoring, Beijing 100097, China

Abstract Soil moisture has strong absorption in near infrared spectroscopy (NIRS) and causes interference in the prediction of the soil organic matter (SOM) content. In this paper, 41 dry soil samples were used to establish the SOM calibration model by PLSR, and 9 samples were used as the prediction set. All soil samples were rewetted to four different moisture contents (5%, 10%, 15% and 17%). The slope/bias (S/B) and direct standardization (DS) algorithms were used to correct SOM prediction results and whole-spectra obtained by different moisture content, eliminating the differences caused by soil moisture. The results showed that the bias reduced and prediction performances of the model were improved, with R_p higher than 0.89 and RMSEP lower than 0.885%. The study indicated that S/B and DS algorithm corrections could effectively remove the influence of soil moisture in NIRS and improve the accuracy of SOM predictions.

Keywords Slope/bias algorithm; Direct standardization algorithm; Soil moisture; Soil organic matter; Near infrared spectroscopy

中图分类号: O657.3 文献标识码: A DOI: 10.3964/j.issn.1000-0593(2019)06-1986-07

Introduction

Near infrared spectroscopy (NIRS) for soil samples has many research applications, mainly the analysis of soil moisture and soil organic matter (SOM) in recent studies. Both soil moisture and SOM have a significant influence on the reflectance spectra^[1]. However, several parameters are known to affect SOM detection and the precision of NIRS analyses, such as soil moisture^[2-5], soil matrix, surface roughness, voids, temperature, sample homogeneity, and particle size, which may mask or alter the absorption features of SOM. Soil moisture is one of the most important sources of error, especially because soil moisture shows strong absorption at two

characteristic bands (1 930 and 1 450 nm) in the NIRS region. The absorption bands of water and SOM overlap, so soil moisture may mask the characteristic band of SOM and hinder the detection of SOM content.

Environmental factors, such as the amount of water, ambient light, and temperature, cause differences in NIRS detection. To eliminate the influence of environmental factors and improve the accuracy of detection, model transfer has been applied to NIRS analyses. Model transfer algorithms mainly include the transfer of standard samples and non-standard samples. Model transfer algorithms of standard samples, such as external parameter orthogonalization (EPO)^[6], direct standardization (DS)^[2,6-7], piecewise direct standardization (PDS)^[7-14], slope/bias (S/B) correction^[15-16], and or-

Received: 2018-05-02; accepted: 2018-10-29

Foundation item: The National Key Research and Development Program of China(2016YFD0701804), the National Natural Science Foundation of China(41201294), the Collaborative Innovation Center Project of Beijing Academy of Agricultural and Forestry Sciences

Biography: WANG Shi-fang, (1989—), Master, Beijing Research Center for Agriculture Standards and Testing

e-mail: wangshifang1302@126.com * Corresponding authors e-mail: yybao@163.com; hanping1016@163.com

thogonal signal correction (OSC)^[15], can eliminate the influence of environmental factors. In previous studies, the effectiveness of these five approaches, i. e., EPO, DS, global moisture modelling (GMM), S/B, and selective wavelength modelling (SWM), was compared as applied to correct the effect of soil moisture to improve the prediction precision of soil organic carbon (OC) and inorganic carbon (IC)^[17]. The PDS algorithm was applied to achieve calibration model transfer between the master and slave instruments, and the Rank-Kennard-Stone-PDS method was applied to predict the alkaloid and glycoside content in tobacco^[18]. The DS algorithm was used to correct differences among instruments, and the genetic algorithm-DS method was used to improve the model precision and simplify NIRS model development as applied to aviation fuel density^[19]. Likewise, finite impulse response (FIR), which is the model transfer algorithm of non-standard samples, can also effectively eliminate the influence of environmental factors. Target soil samples that were calibrated by FIR algorithm were predicted using partial least squares regression (PLSR), and higher prediction accuracy was obtained^[20]. These algorithms were employed to standardize the prediction results and spectra obtained in different test environments.

The aims of this study were to (1) describe how the S/B and DS algorithms remove the effects of soil moisture from spectra recorded under different moisture content conditions, so that predictions and whole-spectra calibrations derived with dry soil samples might be used with spectra of different moisture contents to predict SOM content, (2) demonstrate the improvements from the S/B and DS algorithms using bias, and (3) compare SOM prediction accuracies using the S/B and DS algorithms as measured by PLSR.

1 Materials and methods

1.1 Sample

In this study, 50 original loam soil samples were collected from Taigu, Guandishan, and Youyu of Jinzhong city, Shanxi Province. The soil samples were air-dried for several days and divided into two groups. One group was used for the detection of SOM content. The other group was used for preparation of soil samples with different moisture conditions. The SOM content was tested according to the potassium dichromate method (National standard method of China, GB9834—88). The SOM contents ranged from 0.402% to 7.920%, with 25 analyses from 0.402% to 2.000%, 16 analyses from 2.000% to 3.000%, and 9 analyses from 3.000% to 7.920%, with the mean value 2.220% and the standard deviation 1.22% respectively.

1.2 Sample preparation and NIR spectroscopy measurement

The soil samples were sieved through 10 mesh and then divided into five parts of each sample. One part was taken as a control group. All soil samples were rewetted to four different moisture contents (5%, 10%, 15% and 17%) according to the following equation

$$\theta_m = \frac{Q_w}{Q_s} \times 100 \quad (1)$$

where θ_m is the mass moisture content (%); Q_w is the quantity of purified water (g); and Q_s is the quantity of oven-dry soil samples (g). Each part was weighed to 500 g and placed in a 1 000 mL beaker. Purified water (25, 50, 75 and 85 mL) was added into the beaker and stirred for several minutes to mix adequately and then sealed in plastic film. During sealing, the beaker was shaken to ensure uniform moisture diffusion. The original soil was processed to get 250 soil samples. The NIR spectra of all 250 soil samples were scanned in the same experimental environment 3 days later.

Soil samples were filled into a 12.5 cm diameter Petri dish. They were placed in the detection instrument to stabilize to a constant temperature of $(20 \pm 0.5) ^\circ\text{C}$. A FieldSpec3 spectrophotometer (ASD Company, USA) used to acquire visible (Vis) NIR spectra with a wavelength range of 2~500 nm. To avoid interference of background light, the measurements were performed in a dark space. Each reflectance spectrum was scanned 3 times, and each sample was scanned in triplicate. The spectral resolution was set at 3 nm in the range 350~1 000 nm and 10 nm in the range 1 000~2 500 nm. A breakpoint correction at 1 000 nm was applied using the ViewSpecPro Software. For each soil sample, the average of 9 spectra was analyzed as the spectrum for soil sample. The near infrared spectrum almost includes all the information of the hydrogen group in the organic matter, which contains the information of the structure and the composition state of the molecules. The information of SOM in near infrared spectrum is very rich, so the spectrum range 780~2 500 nm is selected to analyze in the research.

1.3 Slope/bias algorithm

The S/B algorithm is a model transfer algorithm based on the correction of predicted results. For example, the SOM of soil samples with moisture content of 5% was predicted using a SOM model of dry soil samples. The SOM calibration model K with dry soil samples was established using PLSR with the NIR spectrometer results and randomly selected n soil samples with moisture content of 5% to compose the spectra matrix X_1 for the standard sample set. The standard sample set X_1 predicted using calibration model K and the SOM matrix C_1 with moisture content of 5% were then obtained. The relation of the true value matrix C_0 and C_1 with moisture content of 5% is assumed as follows

$$C_0 = \text{Slope} \times C_1 + \text{Bias} \quad (2)$$

where: C_0 is the true value in moisture content of 5%, and C_1 is the predicted value of standard sample set X_1 spectra using SOM calibration model K . C_0 and C_1 are fitted by a linear regression equation, which is solved based on the principle of minimum sum of squares of residuals. The corresponding slope and bias are then obtained according to the Eq. (3) and Eq. (4).

$$\text{Slope} = \frac{\sum (C_{0,i} - \bar{C}_0)(C_{1,i} - \bar{C}_1)}{\sum (C_{1,i} - \bar{C}_1)^2} \quad (3)$$

$$\text{Bias} = \bar{C}_0 - \text{Slope} \times \bar{C}_1 \quad (4)$$

The spectra matrix X_2 of unknown samples with moisture content of 5% is collected and predicted using calibration model K . SOM prediction matrix C_2 is then obtained. SOM correction matrix C_3 is obtained by Eq. (5).

$$C_3 = \text{Slope} \times C_2 + \text{Bias} \quad (5)$$

where: C_3 is the correction from the slope/bias algorithm, and C_2 is the predicted value of unknown samples in set X_2 spectra using SOM calibration model K .

1.4 Direct standardization algorithm

The DS algorithm is a multivariate model transfer algorithm based on the correction of the whole spectrum. It uses the entire set of wavelength data to calculate the transition matrix of a specific wavelength. The transfer process is complicated. The DS algorithm proposed by Wang et al.^[21] was applied with the objective of establishing a mathematical relationship between the data obtained under different conditions, such as soil moisture and temperature, etc. For the DS algorithm, data obtained with representative samples called transfer samples (n Trans) are used to calculate a transformation matrix F , as presented below

$$T_m = T_s \times F + E \quad (6)$$

where: T_m ($n \times m$) is the matrix obtained with the original soil moisture content (dry soil samples), and T_s ($n \times m$) is the matrix obtained with different soil moisture contents (5%, 10%, 15% and 17%), respectively. F is a matrix of size ($m \times m$), where m is the number of wavelengths for the spectra. E is the residual error matrix.

In the DS method, the transformation F matrix is calculated by the least square method and is estimated as

$$F = T_s^{-1} \times T_m \quad (7)$$

where: T_s^{-1} is the inverse of T_s . However, the number of spectral variables is greater than the number of transfer samples, and T_s is a rectangular matrix.

An unknown sample matrix X_s measured with different soil moisture contents can be standardized by multiplying the F matrix obtained from Eq. (7) to calculate X_m according to $X_m = X_s \times F + E$, to match the spectra matrix measured with dry soil moisture.

1.5 Model evaluation standard

In this paper, PLSR as a quantitative analysis was introduced to establish the SOM prediction model. A multivariate calibration model between SOM and spectra was developed using PLSR with n -fold cross validation. PLSR quantitative analysis was done in MATLAB R2015a (The MathWorks, Inc., Natick, MA, USA). The performance of models was evaluated by the following parameters: the correlation coefficient of calibration (R_c), the standard error of calibration (SEC), the root mean standard error of calibration (RMSEC), the correlation coefficient of prediction (R_p), the standard error of prediction (SEP), and the root mean standard error of prediction (RMSEP). Generally, a model with high R_c and R_p and low SEC, RMSEC, SEP, and RMSEP is considered a good model.

2 Results and discussion

2.1 Original spectroscopy

The mean absorbance spectra of 50 samples in different soil moisture content are shown in Fig. 1. There are three characteristic peaks in the NIRS range: 1 930 and 1 410 nm with strong absorption, and 2 210 nm with weak absorption. The peaks near 1 930 and 1 410 nm increased with increasing soil moisture content. Here, 1 930 and 1 410 nm are two characteristic absorption bands for water molecules in the NIRS region. This indicates that moisture had strong absorption in the NIRS region. Therefore, soil moisture caused interference for the NIRS detection of SOM.

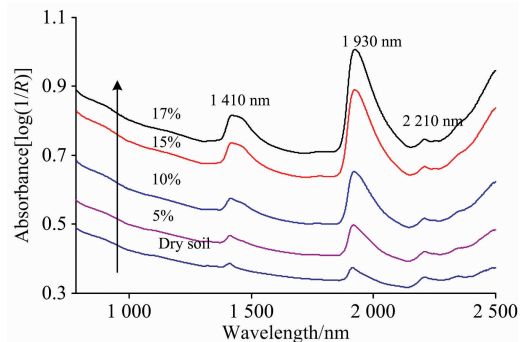


Fig. 1 The mean absorbance spectra of soil samples under different moisture contents

2.2 SOM calibration model

In total, 41 dry soil samples (calibration data set) were used to establish the prediction model for SOM detection by PLSR in the range 780~2 500 nm, and the remaining 9 samples (SOM content: 0.825%, 1.071%, 1.428%, 1.740%, 1.941%, 1.969%, 2.320%, 3.123% and 4.239%) were used as the prediction set. The PLSR calibration model gave the best prediction performance for SOM, as measured by five principal components, R_c of 0.962, SEC of 0.358%, RMSEC

of 0.383%, R_p of 0.961, SEP of 0.597%, and RMSEP of 0.600%.

2.3 Slope/bias model transfer

41 samples with moisture content of 5%, 10%, 15% and 17% were predicted by using the established SOM calibration model. The true value matrix and predicted value matrix were fitted by linear regression, and the slope and bias were calculated. The linear relationship of the true values and predicted values is shown in Table 1.

The spectra of the remaining 9 samples with moisture

content of 5%, 10%, 15% and 17% were predicted using the SOM calibration model established. The SOM corrected values were calculated according to the linear relationship by the S/B algorithm and Eq. (5). The corrected results are shown in Table 1. The corrected values for the S/B algorithm were close to the true value. Compared with uncorrected values, bias showed an obvious decrease. The results show that the prediction accuracies of SOM content were improved using the S/B algorithm correction.

Table 1 The prediction results of soil organic matter content for original and correction models by slope/bias algorithm

True value	Dry	5%	5%S/B	10%	10%S/B	15%	15%S/B	17%	17%S/B
0.825	0.340	0.419	0.391	0.785	0.362	1.121	0.454	1.472	0.529
1.071	0.457	0.674	0.647	1.332	0.892	1.476	0.771	1.608	0.649
1.428	1.662	1.779	1.757	2.418	1.944	3.178	2.291	3.687	2.474
1.696	1.387	1.981	1.961	2.399	1.926	3.254	2.359	3.361	2.188
1.740	1.963	1.733	1.711	2.728	2.244	2.783	1.938	3.030	1.897
1.941	3.063	3.017	3.002	3.531	3.023	4.559	3.524	4.611	3.286
2.320	2.522	3.053	3.038	4.014	3.490	4.561	3.526	5.108	3.722
3.123	3.760	4.079	4.069	4.568	4.026	5.200	4.097	5.793	4.324
4.239	5.088	5.280	5.277	5.604	5.030	5.742	4.580	6.218	4.696
Bias	0.206	0.404	0.385	1.000	0.506	1.499	0.573	1.834	0.598

The linear relationship: 5%: $T=1.005\ 1C_{5\%}-0.030\ 6$
 10%: $T=0.968\ 5C_{10\%}-0.397\ 9$
 15%: $T=0.892\ 8C_{15\%}-0.546\ 4$
 17%: $T=0.878\ 1C_{17\%}-0.763\ 3$

T is the correction values from the slope/bias algorithm, and $C_{5\%}$, $C_{10\%}$, $C_{15\%}$, $C_{17\%}$ is the predicted values of unknown samples using soil organic matter calibration model.

2.4 Direct standardization algorithm

Totally, 41 soil samples were selected to establish the transformation model by the DS algorithm, and the remaining 9 samples were used to verify the adaptability of the algorithm. Transformation matrices $F_{5\%}$, $F_{10\%}$, $F_{15\%}$ and $F_{17\%}$ between NIR spectra with dry soil samples and soil samples with moisture content of 5%, 10%, 15% and 17% were calculated according to Eq. (7). Correction spectra for the 9 soil samples were calculated according to transformation matrix and $X_m = X_s \times F + E$.

The original and DS correction absorbance spectra of the random 3 soil samples (SOM content: 0.825%, 2.320%, 4.239%) with varied moisture content in the NIR region are shown in Fig. 2. The changes in the correction spectra by the DS algorithm are consistent with values before the correction. The corrected spectra are close to the spectra with the dry soil samples, and the height and width of the characteristic peak are also similar. The difference between the maximum and the minimum of the absorbance by DS correction was 0.05 ~

0.10. The absorbance of correction spectra decreased, and the characteristic absorption of water molecules became weaker.

The original and correction spectra of 9 soil samples with varied moisture content were predicted using the SOM calibration model established. The results are shown in Table 2. The corrected values by the DS algorithm are close to the true values. Compared with uncorrected values, the bias obviously decreased. Compared with corrected values by the S/B algorithm, the bias with moisture content 5% increased, and the bias with moisture contents 10%, 15% and 17% obviously improved. The results show that the DS algorithm can eliminate the influence of moisture content for detection of SOM in NIR spectra.

2.5 SOM prediction by PLSR

The corrected results for S/B algorithm and the corrected spectra for DS algorithm were predicted by using the SOM calibration model. The predicted results are shown in Table 3.

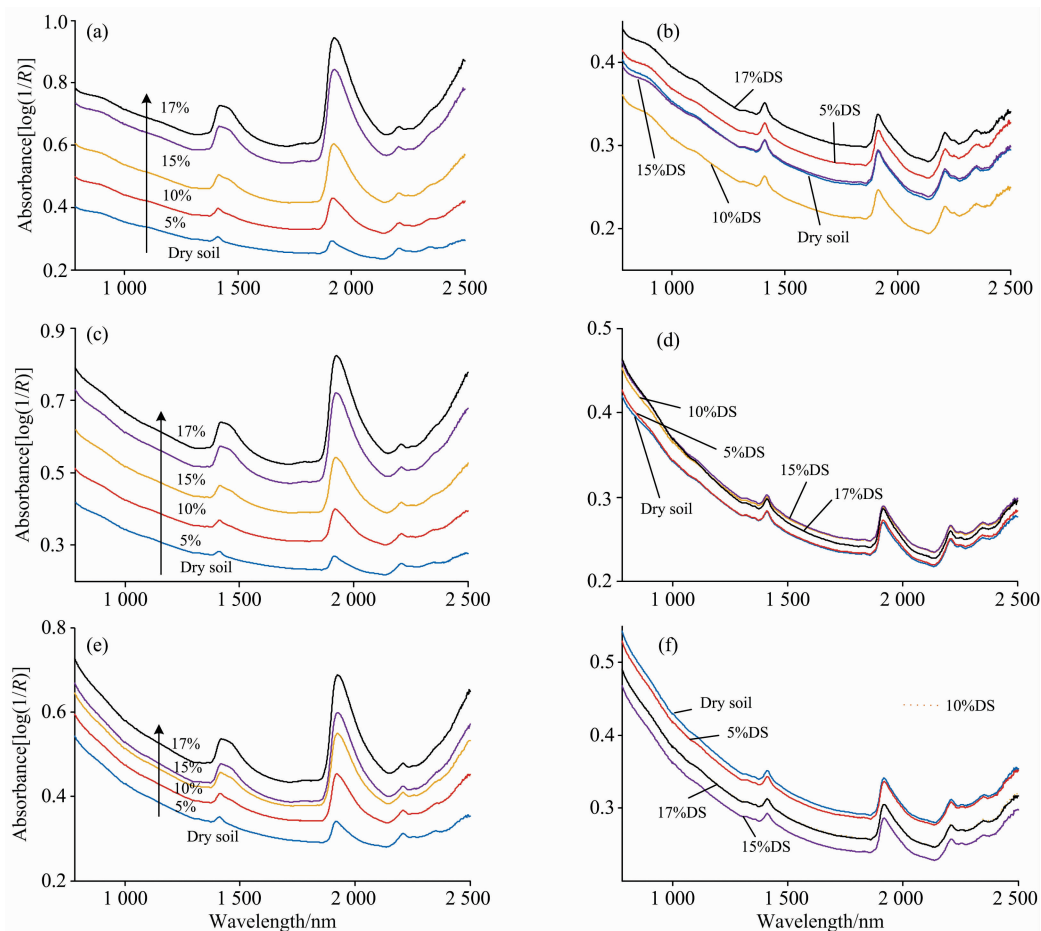


Fig. 2 Comparison of the original and direct standardization correction absorbance spectra of soil samples (organic matter content: 0.825%, 2.320%, 4.239%) with different moisture contents

(a): 0.825% (original spectra); (b): 0.825% (correction spectra); (c): 2.320% (original spectra);
 (d): 2.320% (correction spectra); (e): 4.239% (original spectra); (f): 4.239% (correction spectra)

Table 2 The prediction results of soil organic matter content for original and correction models by direct standardization algorithm

True value	Dry	5%	5%DS	10%	10%DS	15%	15%DS	17%	17%DS
0.825	0.340	0.419	0.510	0.785	0.212	1.121	0.242	1.472	0.593
1.071	0.457	0.674	0.606	1.332	0.084	1.476	0.783	1.608	0.624
1.428	1.662	1.779	2.145	2.418	1.946	3.178	1.972	3.687	2.438
1.696	1.387	1.981	2.352	2.399	1.779	3.254	1.964	3.361	1.720
1.740	1.963	1.733	2.056	2.728	2.105	2.783	1.992	3.030	1.753
1.941	3.063	3.017	3.043	3.531	2.869	4.559	3.205	4.611	3.339
2.320	2.522	3.053	2.928	4.014	3.367	4.561	3.470	5.108	3.509
3.123	3.760	4.079	4.013	4.568	3.525	5.200	4.010	5.793	4.613
4.239	5.088	5.280	5.229	5.604	4.430	5.742	4.462	6.218	4.559
Bias	0.206	0.404	0.500	1.000	0.215	1.499	0.413	1.834	0.529

Table 3 shows the statistical parameters of the nine-sample prediction set for both the original SOM values and corrected values from the S/B and DS algorithms. The S/B algorithm is a model transfer algorithm based on the correction of predicted results, while DS algorithm is a multivariate model transfer algorithm based on the correction of the whole-spec-

trum. Therefore, R_p values from the S/B algorithm were unchanged, and R_p values from the DS algorithm were changed, but over 0.895. For moisture content of 5%, the prediction performance of the S/B correction was better than DS correction, and the difference between SEP and RMSEP was only 0.086%. However, for moisture contents of 10%, 15% and

17%, compared with uncorrected values, the prediction performance of the S/B and DS algorithms both improved. The prediction performance of the DS correction was better than that of the S/B correction, SEP and RMSEP values were lower, and the difference between SEP and RMSEP was only

0.031%, 0.096% and 0.123%, respectively. The results indicate that the S/B and DS algorithms can effectively eliminate the effect of soil moisture and improve the accuracy of the SOM model prediction.

Table 3 The statistical parameters of partial least squares regression prediction model for soil organic matter content. The model prediction performances of the original and correction values by the slope/bias and direct standardization algorithms under different moisture content are compared

Calibration	PC	R_c	SEC	RMSEC	Prediction	R_p	SEP	RMSEP
dry	5	0.962	0.358	0.383	dry	0.961	0.597	0.600
					5%	0.974	0.587	0.685
					5%S/B	0.974	0.594	0.680
					5%DS	0.964	0.565	0.724
					10%	0.959	0.871	1.148
					10%S/B	0.959	0.573	0.730
					10%DS	0.911	0.691	0.660
					15%	0.900	1.408	1.676
					15%S/B	0.900	0.692	0.848
					15%DS	0.918	0.620	0.716
					17%	0.907	1.744	2.005
					17%S/B	0.907	0.728	0.885
					17%DS	0.895	0.757	0.880

3 Conclusions

This implementation of the S/B and DS algorithms to account for the effects of moisture is encouraging. The results showed that the predicted values and spectroscopic predictions of SOM can be improved significantly using the S/B and DS algorithms under different moisture contents, respectively. And prediction performances of the model were improved

using the S/B and DS algorithm corrections, with R_p higher than 0.89 and RMSEP lower than 0.885%. The study indicated that S/B and DS algorithm corrections could effectively remove the influence of soil moisture in NIRS analyses and improve the accuracy of SOM predictions.

Acknowledgements

We would like to thank the Beijing Research Center for Agriculture Standards and Testing, and Shanxi Agricultural University for technical support.

References

- [1] Lobell D B, Asner G P. Soil Science Society of America Journal, 2002, 66(3): 722.
- [2] Ji W, Viscarra Rossel R A, Shi Z. European Journal of Soil Science, 2015, 66(3): 555.
- [3] Song Haiyan, Cheng Xu. Spectroscopy and Spectral Analysis, 2014, 34(5): 1240.
- [4] Wang Shifang, Cheng Xu, Song Haiyan. Spectroscopy and Spectral Analysis, 2016, 36(10): 3249.
- [5] Minasny B, Mcbratney A B, Bellon-Maurel V, et al. Geoderma, 2011, 167(167): 118.
- [6] Ji W, Li S, Chen S, et al. Soil & Tillage Research, 2016, 155: 492.
- [7] Milanez K D T M, Silva A C, Paz J E M, et al. Microchemical Journal, 2016, 124(7): 121.
- [8] Ji W, Viscarra Rossel R A, Shi Z. European Journal of Soil Science, 2015, 66(4): 670.
- [9] Wang J X, Xing Z N, Qu J. Spectroscopy-Springfield then Eugene then Duluth, 2013, 28(6): 36.
- [10] Alves J C, Poppi R J. Spectrochimica Acta. Part A: Molecular and Biomolecular Spectroscopy, 2013, 103(4): 311.
- [11] Griffiths M L, Svozil D, Worsfold P, et al. Journal of Analytical Atomic Spectrometry, 2006, 21(10): 1045.
- [12] Huang Chengwei, Dai Liankui, Dong Xuefeng. Spectroscopy and Spectral Analysis, 2011, 31(5): 1279.
- [13] Tian G Y, Chu X L, Yuan H F, et al. Chinese Journal of Analytical Chemistry, 2006, 34(07): 927.
- [14] Xi C C, Feng Y C, Hu C Q. Chinese Journal of Analytical Chemistry, 2014, 42(9): 1307.
- [15] Liu X, Han L J, Yang Z L. Journal of Dairy Science, 2011, 94(11): 5599.
- [16] Ji Nayu, Li Ming, Lü Wenbo, et al. Spectroscopy and Spectral Analysis, 2017, 37(1): 227.

- [17] Wijewardane N K, Ge Y, Morgan C L S. *European Journal of Soil Science*, 2016, 67(5): 605.
- [18] Liang C, Yuan H F, Zhao Z, et al. *Chemometrics & Intelligent Laboratory Systems*, 2016, 153: 51.
- [19] Wang J X, Qu J, Li H, et al. *Liquid Fuels Technology*, 2012, 30(19): 1975.
- [20] Song H Y, Qin G. *Spectroscopy and Spectral Analysis*, 2015, 35(12): 3360.
- [21] Wang Y, Veltkamp D J, Kowalski B R. *Analytical Chemistry*, 2002, 63(23): 530.

S/B 和 DS 算法校正土壤水分对土壤有机质近红外光谱预测的影响

王世芳^{1,2,3}, 韩 平^{1,3*}, 宋海燕^{2*}, 梁 刚^{1,3}, 程 旭²

1. 北京农业质量标准与检测技术研究中心, 北京 100097
2. 山西农业大学工学院, 山西 太谷 030801
3. 农产品产地环境监测北京市重点实验室, 北京 100097

摘 要 土壤水分对近红外光谱表现出强烈的吸收和对土壤有机质含量的预测造成干扰。研究选择 41 个样本作为校正集和 9 个样本作为预测集, 所有样本做不同含水率(5%, 10%, 15% 和 17%)的处理。采用 S/B 和 DS 算法分别对预测结果和全光谱进行校正, 消除土壤水分的影响。结果得出预测结果偏差减小和模型预测性能得到改善, R_p 高于 0.89 和 RMSEP 低于 0.885%。研究表明 S/B 和 DS 算法能有效消除土壤水分的影响和提高土壤有机质预测的准确性。

关键词 S/B 算法; DS 算法; 土壤水分; 土壤有机质; 近红外光谱

(收稿日期: 2018-05-02, 修订日期: 2018-10-29)

* 通讯联系人