

Prediction of Soil Organic Matter Content Using VIS/NIR Soil Sensor

Yubing Wang, Cuiping Lu, Liusan Wang, Liangtu Song,
Rujing Wang, Yunjian Ge

Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei 230031, China
Tel.: +86-551-65591195, fax: +86-551-65592420
E-mail: ybwang@iim.ac.cn

Received: 22 January 2014 / Accepted: 7 March 2014 / Published: 30 April 2014

Abstract: In present work, the potential of visible/near-infrared (VIS/NIR) soil sensor to predict organic matter (OM) contents in soils was studied using VIS/NIR soil sensor. Different preprocessed methods were studied to improve the correlations of spectra with organic matter contents in soil samples, and three regression methods, including direct linear regression, principal component regression (PCR) and back propagation-neutral network (BP-ANN), were employed to construct the prediction models in calibration stage, whose performances were evaluated in the validation stage. The prediction results indicated that PCR for the preprocessed spectra by MSC together with S-G polynomial filter, and BP-ANN for the first ten principal components of the spectra by MSC pretreatment could provide appropriate results for the estimation of OM contents in unknown soil samples. Copyright © 2014 IFSA Publishing, S. L.

Keywords: Soil OM content, Near-infrared spectra, Principal component regression, BP neutral network.

1. Introduction

Though promoted widespread increases of grain production, traditional fertilization produced serious problems in ecological environment meanwhile. Lack of useful and accurate fertility information of soils could lead to improper compositions and low efficiencies in fertilization [1, 2]. With the enhanced consciousness on environment conservation and the demand of economic benefits, precision fertilization became a hot issue in the research of agricultural science, which required to take organic matter, nitrogen, phosphorus, potassium and many crop growth-promoted elements and organic fertilizer to a scientific formula, according to different region and different soil types, as well as gains and losses of soil nutrients and crop types, and thereby led to a destination scientific application of fertilizers.

As an important part of modern precision agriculture, precision fertilizer would be an effective way to improve the soil and environment quality, reduce water and soil pollution and finally enhance crop yield and quality [3-5].

The accurate perception of soil composition is a precondition for precision fertilization. Most routine analyses of soil composition were carried out in chemical laboratories, combined with traditional sampling method, which would be cost prohibitive and time-consuming. Visible and near-infrared (VIS/NIR) spectroscopy emerged as a rapid and nondestructive analytical method in agriculture, which correlated the diffusely reflected VIS/NIR radiation with chemical and physical properties in soil [6]. When a sample was illuminated with light, certain bonds within the molecules vibrate with the varying electric field, and

the vibration bonds absorbed optical energy and caused less light to be reflected off the sample. The predominant absorbers in NIR region were C-H, N-H and O-H functional groups, which provided an advisable tool for quantifying the content of carbon, nitrogen and water, respectively [7, 8].

The studies on soil characteristics with the use of VIS/NIR spectroscopy began earlier in foreign countries, and good results were obtained by foreign soil scientists since 1960s, especially during the last two decades, with the rapid development of chemistry stoichiometric. Soil organic matter is an accumulation of partially decayed and resynthesized plant and animal residues [9]. Soil OM content is a significant variable in modern soil management, which is tightly related to several key parameters, such as soil moisture infiltration and retention, soil tilts, rooting depth, soil-applied herbicide activity, nitrogen release and other aspects of nutrient cycling [10]. At the beginning, soil color was used to estimate soil organic matter content by researchers [11, 12]. Single red LEDs with specific wavelengths in VIS and NIR region were then adopted in the design of soil sensors, and high correlations with OM content were obtained [13-15]. Numerous studies were also implemented to predict soil OM content using VIS/NIR spectrometer sensors, which were proved to be feasible, and different methods were studied and compared to improve the accuracy and efficiency of the prediction model [16, 17].

In this paper, different multivariate statistical methods were employed to correlate the VIS/NIR reflectance spectra with soil OM content in our soil sensor. The prediction models of soil OM content by different linear/nonlinear regression methods were studied and compared.

2. Materials and Methods

2.1. Field Description and Sampling Strategy

The soil samples were collected in Shuangshan Farm, Qiqihar City, Heilongjiang Province, located at around 48.829°N and 125.477°E, where the overall flat fields were characterized by black soils in north-east China. The normal grid method was adopted to sample the soil, which divided the whole field into several grids with the same sizes and then collected one soil sample from each grid to analyze soil properties over whole field. The 156 soil samples in present work were collected from five sub-farms in Shuangshan Farm, at a depth of about 10 cm and with a grid interval of about thirty meters.

All the samples were carefully handled to avoid external contamination. After the rejection of weed and small pieces of rocks, the soil samples were then naturally air-dried, crushed, and passed through a 2 mm sieve sieved (American recommended soil test procedures, provided by Missouri Agriculture Experimental Station SB 1001 and the U.S.

Department of Agriculture cooperating, 1998). The preprocessed soil samples were divided into two portions by four-way division: A and B. The samples of group A were placed in properly closed bags and taken to the soil science laboratory in Anhui Geological Research Institute for the analysis of OM content with colorimeter method, as reported in many literatures. The OM content of total 156 samples selected from Shuangshan Farm varied from 50.33 to 69.75 g/kg, with mean value being 59.833 g/kg and standard deviation being 5.5047. The samples of group B were taken to spectral measurements using VIS/NIR soil sensor, which would be described in depth in the next section.

2.2. Experimental Instruments and Information Acquisition

Spectral reflectance of soil samples were obtained in laboratory with the use of Veris VIS/NIR soil sensor (Veris Technologies, Inc., Salina, KS, USA). As a kind of tractor-mounted sensor to collect real-time soil information, the spectrophotometers and optical system of Veris VIS/NIR soil sensor were built into a shank, mounted on a toolbar and then pulled by a tractor. The measurement were made through a sapphire window mounted to the bottom of the shank. A tungsten halogen bulb was used to illuminate the soil, and the reflected light was transmitted to the spectrometer by optical fibers. An charge-coupled device (CCD) array spectrometer (USB 4000, Ocean Optics, Inc., Dunedin, FL, USA) and an InGaAs photodiode-array spectrometer TG-COOLED II, Hamamatsu Photonics, K. K., Japan) were used at the receiving ends of the optical fibers, which measured spectra from 350 to 1022 nm and 1100 to 2225 m, respectively.

Approximately 20 spectra per second were acquired from spectrometer and immediately transferred to a personal computer. Then, the reflectance data for each sample were averaged and transformed to absorbance spectra for further analysis. Total 156 samples were put in a Petri dish and the surface were flushed beforehand. For each sample, four spectra were obtained and averaged, i.e., the sample was rotated by 90° for the next scan. The averaged absorbance soil spectra from VIS/NIR soil sensor were given in Fig. 1(a), from which we could find that two large absorption peaks existed, located at around 1420 and 1930 nm, respectively, both of which were coincident with characteristic absorption peak of H₂O.

3. Results and Discussions

3.1. Analysis and Pretreatment of Data

All the data analysis and preprocessing procedures in present work were compiled

with Matlab. Generally, the spectra data collected from VIS/NIR soil sensor were easily influenced by individual differences (such as the size of samples, the intensity of light source, the condition of measurement and etc.), baseline variations and substantial noises. Therefore, appropriate preprocessing methods should be applied

to minimize the effects of usefulness and then to increase the correlation between spectra and experimental measured values. The frequently-used pretreatment methods included: normalization, first and second derivatives, multiplicative scatter correlation (MSC), standard normal variate (SNV), detrending and any combination thereof [18, 19].

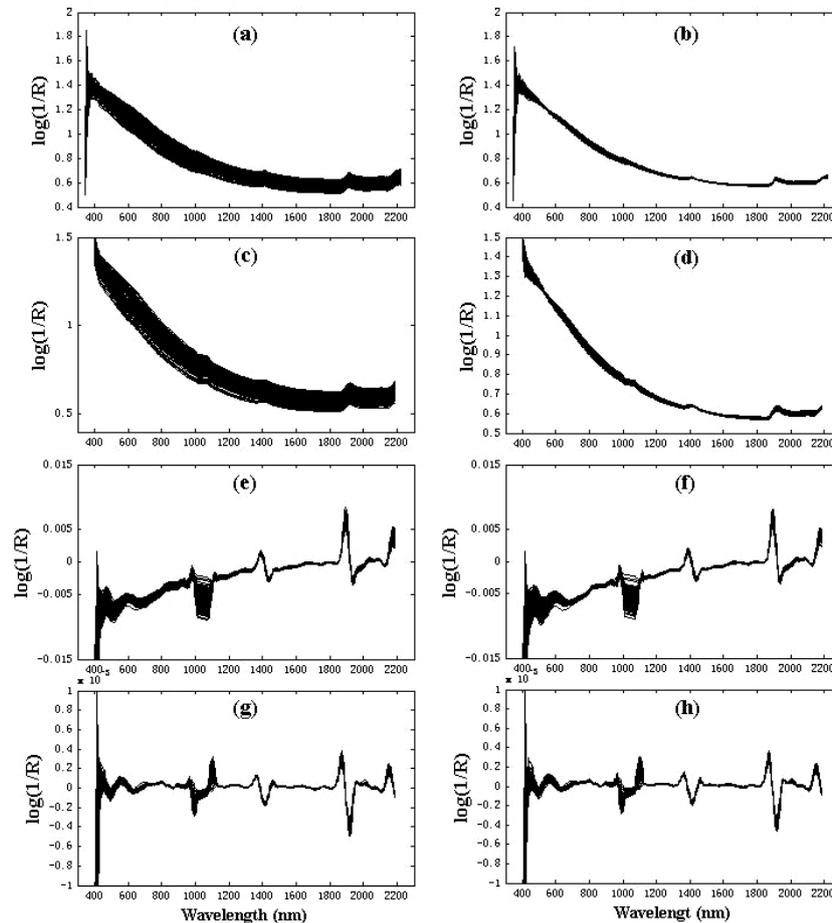


Fig. 1. Original and preprocessed spectra of 155 soil samples: (a) original spectra; (b) preprocessed spectra with MSC; (c) preprocessed spectra with S-G polynomial smoothing; (d) preprocessed spectra with MSC and S-G polynomial smoothing; (e) preprocessed spectra with S-G polynomial 1-st derivative; (f) preprocessed spectra with MSC and S-G polynomial 1-st derivative; (g) preprocessed spectra with S-G polynomial 2-st derivative; and (h) preprocessed spectra with MSC and S-G polynomial 2-st derivative.

The pretreatment methods employed in present work were MSC, together with first and second derivatives by Savitzky-Golay (S-G) polynomial fitting. MSC transformations could remove the slope variations from spectra caused by scatter and variations of particle size. To remove high frequency noise, S-G polynomial smoothing filter digitally smoothed a given spectrum data by approximating it within a specified data window by a polynomial of a specified order that best matched the data in the window on a least-square basis, in which a filter with a polynomial of order 3 and a window of 11 data points was used in present work.

As also displayed in Fig. 1, compared with original spectra, preprocessed methods (including

MSC, S-G polynomial fitting and their combinations) could smooth the spectra, reduce the influence of noise effectively, and then enhance the resolution and characteristics of the spectra. The correlation coefficients (R) calculated from the different spectra (raw and preprocessed spectra) with organic matter contents of 156 soil samples were presented in Fig. 2, from which we could observe that pretreatment improved relativity significantly, in comparison with the relationship of OM contents with original spectra. The wavelengths with high absolute value of correlation coefficients were adopted in the next section for the construction of prediction model with direct linear regression method.

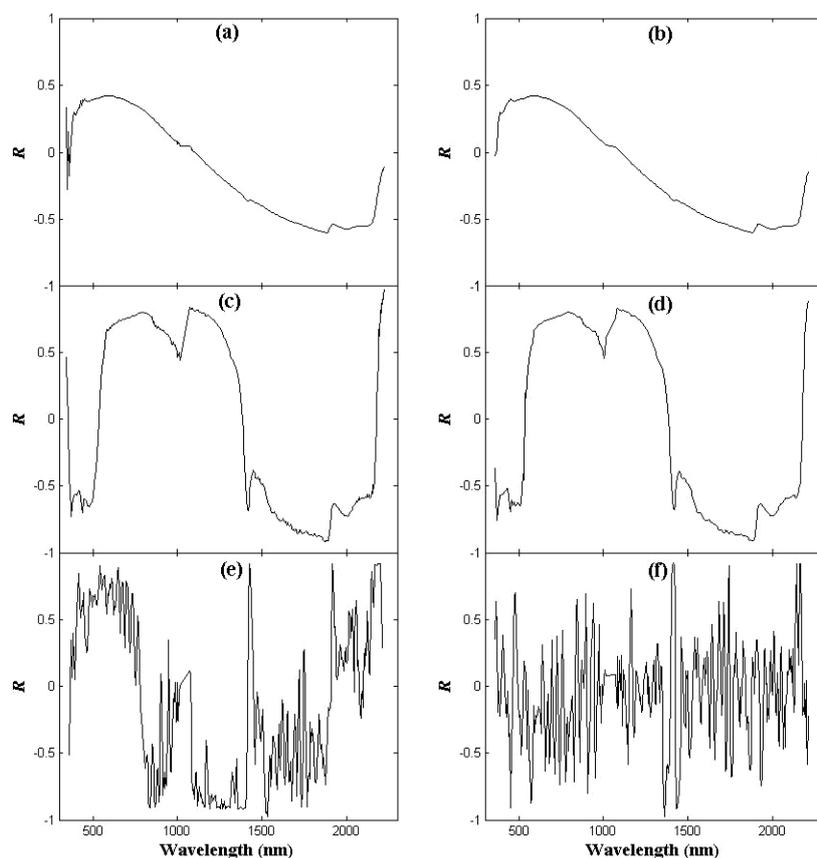


Fig. 2. Calculated correlation coefficients for raw and preprocessed spectra: (a) raw spectra; (b) preprocessed spectra with S-G polynomial smoothing; (c) preprocessed spectra with MSC; (d) preprocessed spectra with MSC and S-G polynomial smoothing; (e) preprocessed spectra with MSC and S-G polynomial 1-st derivative; and (f) preprocessed spectra with MSC and S-G polynomial 2-st derivative.

3.2. Establishment of Prediction Models

VIS/NIR spectroscopy was basically an empirical method, whose complete procedure consists of two stages: (i) calibration stage where the spectra from the training set were matched with the results from laboratory analysis and then prediction model was developed; and (ii) validation stage in which the prediction model was validated. For a practical application, total 156 soil samples were randomly divided into calibration set and validation set, in which 116 soil samples in calibration set were elected to establish the prediction models and the rest of 40 samples were used to evaluate the prediction ability of models.

The selections of wavelength as variables played a vital role in the establishment of prediction models, since multicollinearity always existed in spectral variables, which led to the increase of computational complexity, and moreover, the reduction of accuracy and reliability of prediction models. In present work, the prediction models with three different regression methods: direct linear regressions with high-correlation wavelengths, principle components regressions (PCR) and back propagation artificial neural networks (BP-ANN) were constructed in the calibration stage, which would be evaluated by

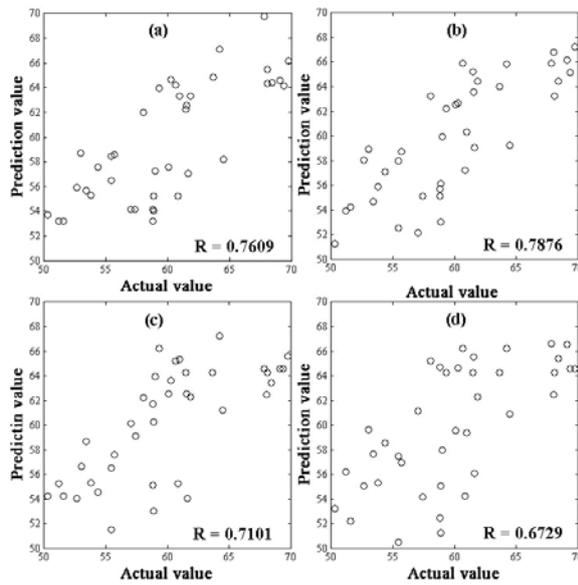
the prediction results of 40 soil samples in the validation stage.

3.2.1. Direct Linear Regressions

To accelerate the speed of calculation and enhance the precision of predicted models effectively, the wavelength variables with high relativity were chosen. As shown in Fig. 2, the wavelengths with absolute correlation coefficient values being above 0.75 for the spectra with different preprocessing methods, including MSC together with S-G polynomial fitting, were adopted in the construction of prediction model. Using established models, the comparison results between the prediction results and actual results of 40 samples in validation set were shown in Fig. 3. Statistical parameters including determination coefficient (R) and root mean square error of prediction (RMSEP) were also calculated and presented in Table 1 for direct comparisons. Summarized briefly, the prediction results with direct linear regression method were not very ideal on the whole, with a highest determination coefficient being 0.7876 for the preprocessed spectra with MSC and S-G polynomial smoothing.

Table 1. Prediction statics data of OM contents in soils for different regression and preprocessed methods.

| Regression method | Preprocessed method | RMSE | R |
|--------------------------------|----------------------------------|--------|--------|
| Direct linear method | MSC | 3.6134 | 0.7609 |
| | MSC + SM | 3.4006 | 0.7876 |
| | MSC + 1 st derivative | 3.9111 | 0.7101 |
| | MSC + 2 st derivative | 4.2520 | 0.6729 |
| Principle component regression | MSC | 3.1236 | 0.8561 |
| | MSC + SM | 3.1361 | 0.8827 |
| | MSC + 1 st derivative | 3.5110 | 0.7733 |
| | MSC + 2 st derivative | 3.9651 | 0.7028 |
| BP-ANN | N = 5 | 3.0059 | 0.8566 |
| | N = 6 | 2.8804 | 0.8845 |

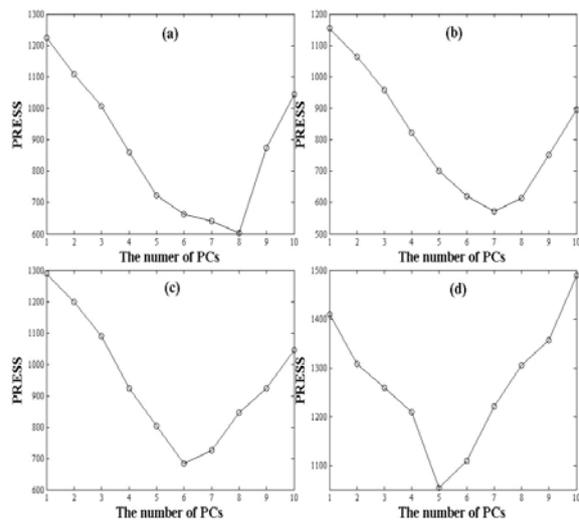
**Fig. 3.** Correlation between the prediction and actual results obtained from direct linear regression for different preprocessed spectra: (a) preprocessed spectra with MSC; (b) preprocessed spectra with MSC and S-G polynomial smoothing; (c) preprocessed spectra with MSC and S-G polynomial 1-st derivative; and (d) preprocessed spectra with MSC and S-G polynomial 2-st derivative.

3.2.2. Principle Component Regressions

Principle components analysis (PCA) occupied a crucial position in chemometrics, with the ultimate goals of reducing the data dimensions and transformation to new variables, which were the linear combination of original variables and meanwhile retained as much characteristic information of original ones. After compressed using PCA, the optimum numbers of principal components (PCs) with different preprocessed spectra were determined with the employment of leave-out-one cross validation (LOOCV) techniques, in which each sample in training set was omitted and predicted by the prediction model developed by the remaining samples [20].

The prediction residual error sum of squares (PRESS) in LOOCV for different numbers of PCs

were calculated and shown in Fig. 4. The options of optimum PCs were of crucial influence on the prediction performance, since several useful information were missed for less number of PCs while several useless noise were included for more PCs. As shown in Fig. 4, the optimum numbers of PCs were selected when the curves of PRESS versus PCs were least or turned to be stable, and used in the construction of prediction model, by which the prediction values compared with actual values were displayed in Fig. 5. The prediction performance of the model evaluated by statistic parameters, including determined coefficient (R), root mean square error of prediction (RMSEP) were investigated, as also shown in Table 1, from which we could find that with the preprocessed methods include MSC and S-G polynomial filter for smooth and first-order derivative were better than those with MSC or S-G filter for smooth and first-order derivative individually.

**Fig. 4.** The value of PRESS at different number of PCs in the leave-one-out validation for the results from different preprocessed spectra: (a) preprocessed spectra with MSC; (b) preprocessed spectra with MSC and S-G polynomial smoothing; (c) preprocessed spectra with MSC and S-G polynomial 1-st derivative; and (d) preprocessed spectra with MSC and S-G polynomial 2-st derivative.

3.2.2. Back Propagation Artificial Neural Network Regressions

Due to the nonlinearities caused by instrument noises, base drift, operation of experiments, and other factors, prediction model developed with the nonlinear regression methods were also discussed in present work. As a widely used technique, artificial neural network algorithm (ANN) has gained several successful results to overcome nonlinearities of prediction model, in which back propagation artificial neural network algorithm (BP-ANN) was of most popular. BP-ANN, known as a 'feed-forward back-propagation network', in which

the learning process was training by the iterative comparisons between the forward propagation of operating signal and back propagation of error signal, and finally lead to a minimum of global error.

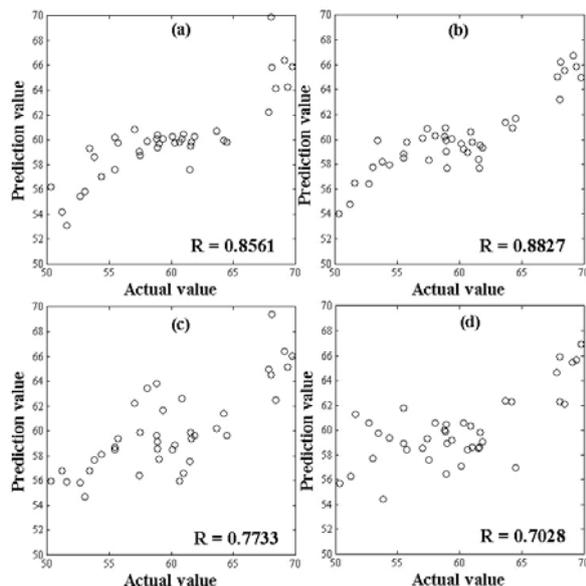


Fig. 5. Correlation between the prediction and actual results obtained from principal component regressions for different preprocessed spectra: (a) preprocessed spectra with MSC; (b) preprocessed spectra with MSC and s-Golay polynomial smoothing; (c) preprocessed spectra with MSC and s-Golay polynomial 1-st derivative; and (d) preprocessed spectra with MSC and s-Golay polynomial 2-st derivative.

In present work, a BP-ANN consisting of three layers (input layer, hidden layer and output layer) was constructed. The *logsig* function was used for the transfer function between the input layer and the hidden layer, while the *purelin* function was adopted between the hidden and output layers. At first, the normalized spectra were compressed by principle component analysis, in which the first ten components were used as the input variables, and the Levenberg-Marquardt optimization algorithm was adopted in the training of network. The nodes of hidden layer (N) were chosen as 5 and 6 in the development of regressions, as the results shown in Fig. 6, from which we could observe that the regression coefficients were 0.84699 and 0.9371, respectively.

3.3. Comparison and Discussions

Based on the above-described data, BP-ANN could provide a most accurate prediction results among three regression methods, but of high statistical randomness at the same time; while the results by direct linear regression were relatively poorer to a certain degree. To summarized briefly, principle component regressions with the

preprocessed spectra including MSC and Savitzky-Golay (S-G) polynomial filter, and BP-ANN for the first ten components of normalized spectra both could given favorable prediction results, which implied VIS/NIR spectroscopy sensor was a ideal tool for the estimation of OM content in soils.

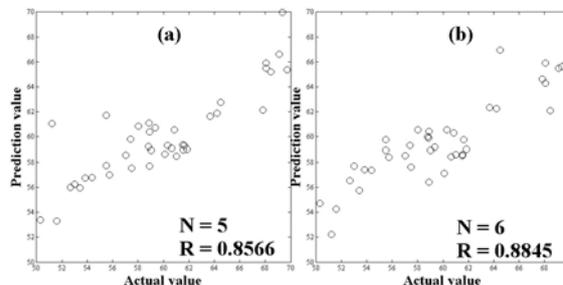


Fig. 6. Correlation between the prediction and actual results obtained from BP-ANN regressions for different preprocessed spectra: (a) preprocessed spectra with MSC; (b) preprocessed spectra with MSC and s-Golay polynomial smoothing; (c) preprocessed spectra with MSC and s-Golay polynomial 1-st derivative; and (d) preprocessed spectra with MSC and s-Golay polynomial 2-st derivative.

4. Conclusions

In conclusions, the potential of VIS/NIR soil sensor to predict soil OM contents were studied in present work. Different preprocessed methods were adopted to improve the correlation between soil OM contents with VIS/NIR spectra, which presented that the preprocessed methods including MSC and Savitzky-Golay (S-G) polynomial filter for smooth and first-order derivative could provide preferable results. The prediction models developed by principle component regressions with the preprocessed spectra including MSC and S-G polynomial filter for smooth and first-order derivative, and BP-ANN for the first ten components of normalized spectra both could given favorable results, which implied VIS/NIR spectroscopy sensor was a ideal tool for the estimation of organic matter content of soils.

Acknowledgements

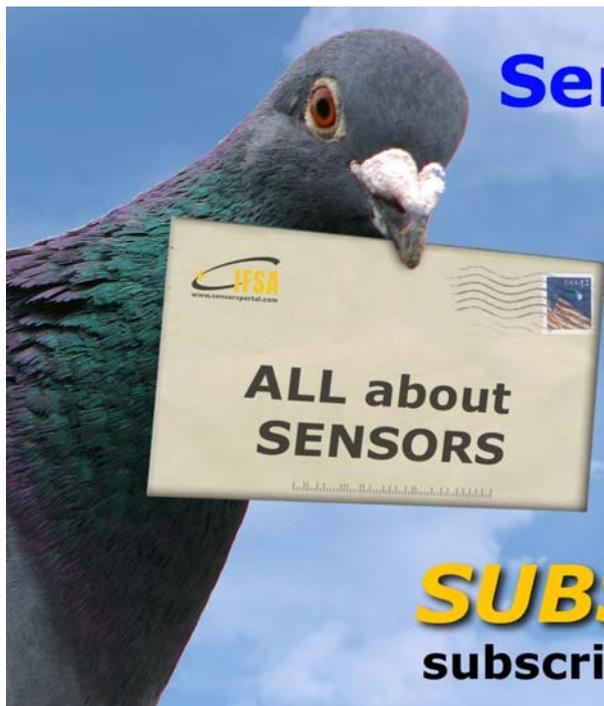
This work was supported by the Modern Agriculture Demonstration Project of China Academy of Sciences (Grand No. CXJQ1201) and National Science Foundation of China (Grand No. 61301060).

References

- [1]. M. H. Wang, Development of precision agriculture and innovation of engineering technologies,

- Transactions of the American Society of Agricultural Engineerings*, 15, 1, 1999, pp. 1-8.
- [2]. C. J. Zhao, Strategy thinking on precision agriculture of China, *Agriculture Network Information*, 4, 2010, pp. 5-8.
- [3]. W. S. Lee, V. Alchanatis, C. Yang, D. Moshou, C. Li, Sensing technologies for precision specialty crop production, *Computers and Electronics in Agriculture*, 74, 2010, pp. 2-33.
- [4]. V. I. Adamchuk, J. W. Hummel, M. T. Morgan, S. K. Upadhyaya, On-the-go soil sensors for precision agriculture, *Computers and Electronics in Agriculture*, 44, 2004, pp. 71-91.
- [5]. X. Xi, X. Y. Xie, Z. Hai, Application of theory and technology of wireless sensor network system for soil environment monitoring, *Sensors & Transducers*, Vol. Special Issue, 2013, pp. 78-84.
- [6]. M. Blanco, L. Villarroya, NIR spectroscopy: a rapid-response analytical tool, *Trends in Analytical Chemistry*, 21, 4, 2002, pp. 240-250.
- [7]. E. B. Franklin, Theory and principles of near-infrared spectroscopy, *Spectroscopy Europe*, 14, 1, 2002, pp. 14-18.
- [8]. H. W. Siesler, Y. Ozaki, S. Kawata, H. M. Heise, Near-infrared spectroscopy: principles, instruments, applications, *John Wiley & Sons*, 2008.
- [9]. H. L. Bohn, B. L. McNeal, G. A. O'Connor, Soil Chemistry, *John Wiley & Sons*, New York, 1979.
- [10]. K. L. Fleming, D. F. Heermann, D. G. Westfall, Evaluating soil color with farmer input and apparent soil electrical conductivity for management zone delineation, *Agronomy Journal*, 96, 6, 2004, pp. 1581-1587.
- [11]. N. R. Page, Estimation of organic matter in Atlantic coastal plain soils with a color difference matter meter, *Agronomy Journal*, 66, 1974, pp. 652-653.
- [12]. J. L. Shonk, L. D. Gaultney, D. G. Schulze, G. E. Van Scoyoc, Spectroscopic sensing of soil organic matter content, *Transactions of THE ASAE*, 34, 5, 1991, pp. 1978-1984.
- [13]. J. H. Sung, Development of non-destructive grouping system for soil organic matter using VIS and NIR spectral reflectance, *Agricultural and Biological Engineering*, 6, 2005, pp. 15-21.
- [14]. D. F. Roberts, V. I. Adamchuk, J. F. Shanahan, R. B. Ferguson, J. S. Schepers, Estimation of surface soil organic matter using a ground-base active sensor and aerial imagery, *Precision Agriculture*, 12, 2011, pp. 82-102.
- [15]. G. Kweon, C. Maxton, Soil organic matter sensing with an on-the-go optical sensor, *Biosystem Engineering*, 115, 2013, pp. 66-81.
- [16]. Y. He, M. Huang, A. García, A. Hernández, H. Y. Song, Prediction of soil macronutrients content using near-infrared spectroscopy, *Computers and Electronics in Agriculture*, 58, 2007, pp. 144-153.
- [17]. M. Z. Li, L. Pan, L. H. Zheng, X. F. An, Development of a portable SOM detector based on NIR diffuse reflection, *Spectroscopy and Spectral Analysis*, 30, 4, 2010, pp. 1146-1150.
- [18]. A. Rinan, F. van den Berg, S. B. Engelsen, Review of the most common pre-processing techniques for near-infrared spectra, *Trends in Analytical Chemistry*, 28, 10, 2009, pp. 1201-1222.
- [19]. A. Savitzky, M. J. E. Golay, Smoothing and differentiation of data by simplified least squares procedures, *Analytical Chemistry*, 36, 8, 1964, pp. 1627-1639.
- [20]. H. Abdi, L. J. Williams, Principal component analysis, *Wiley Interdisciplinary Reviews: Computational Statistics*, 2, 2010, pp. 433-459.

2014 Copyright ©, International Frequency Sensor Association (IFSA) Publishing, S. L. All rights reserved.
(<http://www.sensorsportal.com>)



**Sensors Industry
News**

**FREE Monthly
IFSA Newsletter**

ISSN 1726-6017

SUBSCRIBE NOW
subscribe@sensorsportal.com