

Detection of Soil Total Nitrogen by Vis-SWNIR Spectroscopy

Yaoze Feng¹, Xiaoyu Li^{1,2*}, Wei Wang¹, Changju Liu¹,

¹ College of Engineering, Huazhong Agricultural University, Wuhan, China

² State Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau, Chinese Academy of Sciences and Ministry of Water Resources, Yangling, China
{yzfeng, lcj94007}@webmail.hzau.edu.cn; {lixiaoyu, wangwei}@mail.hzau.edu.cn

Abstract. Measurement of soil total nitrogen (STN) is urgently important concerning requirement of precise and quantitative fertilizer application. To overcome the shortage of routine chemical detection, vis- short wavelength near infrared spectroscopy (Vis-SWNIRS) was employed as an accurate, cheap and timely alternative. Kennard-Stone algorithm was utilized for sample set partitioning, where 22 of the 32 samples were selected as calibration set, and the remaining 10 were included in validation set. No outliers were discovered under criterion based on spectral Mahalanobis distance and Dixon testing. Partial least squares regression (PLSR) was used to build STN detection model based on Vis-SWNIRS with cross validation by leave-one-out method. As a result, 5 principal components turned to be optimal considering model performance, where correlation coefficients for calibration and validation were 0.9724 and 0.8691, respectively, with PRESS (Prediction Residual Error Sum of Square) of 0.0684. It is feasible to detect STN by Vis-SWNIR spectroscopy.

Keywords: Soil; Total Nitrogen; Vis- Short Wavelength Near Infrared Spectroscopy; Data Set Assignment; Outlier Detection.

1 Introduction

Soil nitrogen content is one of the most important factors that affect modern agriculture, and it is attributing a lot to improvement of people's living standard [1]. However, it is also among the plant nutrients which are extremely hard to manage effectively. If ill treated, its negative impacts towards environment, quality of agricultural products as well as human health will be tremendous. For example, if too much nitrogen is employed, there will appear either eutrophication of surface waters or underground water contamination to some extent. Adversely, we can easily find poor growth of the crops and eventually lower outputs if nitrogen is inadequately applied. As a result, it seems increasing urgent to determine soil total nitrogen especially under circumstances of global food shortage and safety problems. But current methods mainly rely on chemical test which is time-consuming and expensive although high accuracy is its privilege. A fast and cheap but meanwhile precise method is more promising and actually in great demand in practical application.

Near infrared spectroscopy has gained fast development in recent years and is applied successfully in many fields, such as agriculture, food, medicine, petrochemical industry, and so on [2-6]. The use of near infrared spectroscopy into detection of soil total nitrogen (STN) starts from Dalal and Henry's research, whose results demonstrated the feasibility of this technology in calibrating spectra data with chemical reference of STN, and 1702nm, 1870nm, 2052nm were recognized as featured wavelengths [7]. Chang et al. [8] Employed independent spectra to determine STN other than previous methods that relied on the close relationship between levels of nitrogen and carbon content. Further, plenty of work was conducted to study different factors that may affect determination of STN by NIRS. For example, Reeve and McCarty [9] discovered that a wider range of soil types would lead to poorer performance of established calibration models, while Wu studied the influence from soil particle size and water moisture content [10, 11]. What is more, Bao et al. [12] optimized the illuminating angle and height. Several attempts were made to understand the feasibility of NIRS on determination of total nitrogen of undisturbed soil samples. For instance, Sun et al. [13] studied black soil in Northeast China, and found that the accuracy of prediction can be improved by enlarging variance of STN values. Besides, Michel reached satisfactory results in establishing NIRS-STN model by utilizing first and second derivative [14]. However, very rare research considered the significance of outlier detection and data set partitioning, which also have strong impact on model performance.

The purpose of this work is to evaluate Vis-NIRS as a tool for sensing TN of paddy soil in Southern China, and to assess feasibility of establishing short wavelength calibrations for STN. Specifically, this research addressed the following objectives: (i) to assign the soil samples into two groups, that is, calibration set and validation set, by using Kennard-Stone algorithm; (ii) to detect potential outliers in the calibration set; (iii) to qualify STN content after establishment of STN-Vis-NIRS calibration model.

2 Materials and Methods

2.1 Soil Sample Collection and preparation

Soil samples used in this study were collected from paddy experimental plot in Huazhong Agricultural University. To ensure variance of levels of STN, an S-type sampling route was employed to collect the surface layer soil (0-20cm). Totally, 32 samples were gained. They were packed into plastic bags and put into a climate incubator for preservation at 4°C. Then the soil samples were oven-dried at 105°C for 24 hours following National Standards of China (GB7172-1987) before both spectral scanning and chemical experiment.

2.2. Vis-shortwave near infrared spectroscopy measurement

Samples were placed into aluminum specimen boxes with diameter of 55mm and height of 33mm. A fiber-type spectrometer usb4000-uvvis/nir256 (Ocean Optics, USA) was employed to scan the soil spectra, where wavelength range was set as 400-900nm, with an increment of 0.21nm. During spectral measurement, to cover the utmost area, vertical distance between the fiber-optics probe and sample surface was determined as 25mm, since the illuminating angle of the probe was 12.5°. Each spectrum was obtained by taking the average of two individual spectrum of the same sample measured at different rotation angles, while each of the individual spectrum was the outcome of 32 scanning to the sample. Fig.1 shows the soil spectra acquired.

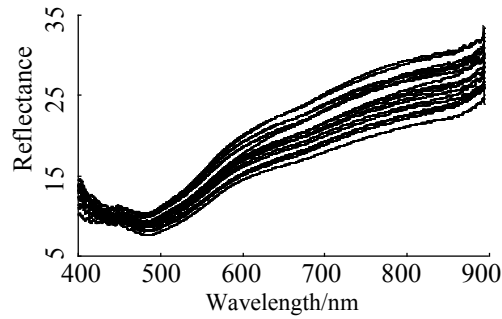


Fig. 1. Soil spectra

2.3. STN determination by laboratory analysis

Before chemical experiment, the dried samples were smashed and passed a 2mm-diameter sieve. An automatic nitrogen determination apparatus ATN-300 (Hongji, China) was utilized, and its principle abides with National Standards of China GB7173-1987. A detailed description of the STN values was given in Tab.1.

Tab. 1 Statistics of soil total nitrogen content

	Num.	Min. (%)	Max. (%)	Mean (%)	Std. (%)
Calibration set	22	0.07	0.98	0.52	0.24
Validation set	10	0.14	1.12	0.52	0.25

3. Vis-NIR spectroscopy prediction

3.1. Data set partitioning

Data set assignment was realized through employing Kennard-Stone algorithm [15], which was based on Euclidian distance between samples. Given each row represents a sample and each column stands for reflectance values for all samples at one specific wavelength in the spectral matrix, to include a certain number of samples, say, Num, into calibration set, the following calculation can be performed:

(i) To calculate spectral Euclidian distance between each two samples, where the distance between sample i and j is denoted as $E(i, j)$. And the two samples that make the largest distance are picked out. For easy expression, the two selected samples are named after a and b , respectively;

(ii) To determine the distance between each of the remaining samples with sample a and b , respectively, denoted as $E(m, a)$ and $E(m, b)$ where $m \neq a, b$;

(iii) For each sample not chosen, the minimum value among the distance between this sample and the selected samples is deemed to be a corresponding reference, that is $D_m = \min(E(m, a), E(m, b), \dots)$, while the N th selected sample is the one whose reference is larger than that of other samples $N = \text{index}(\max(D_k))$;

(iv) To repeat step (iii) until the desired Num samples are chosen.

In this study, 22 samples were categorized into calibration set while the other 10 were included in prediction set by performing Kennard-Stone algorithm

3.2. Outlier detection

Spectral Mahalanobis distance (M-distance) of a sample is the distance between the spectrum of this sample and the average spectrum of the sample set, and it is a frequently used method for outlier detection [16, 17]. The definition of spectral Mahalanobis distance of the i th sample can be described as:

$$D_i^2 = (t_i - \bar{t}) \text{Cov}^{-1}(t) (t_i - \bar{t})' \quad (1)$$

Where t_i is the score vector of the i th sample, and \bar{t} and Cov are the average score and covariance of all samples in calibration set, respectively.

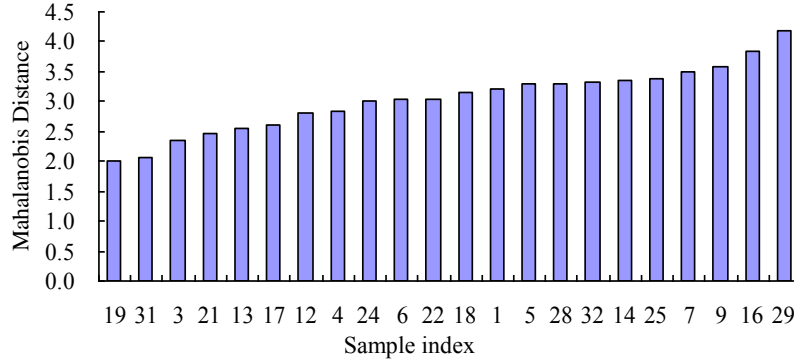


Fig. 2. M-distance of samples in calibration set

Dixon testing [18] was utilized to discriminate outliers. Firstly, according to the number of samples as well as the reference form given in National Standards (GB/T 4883-2008), statistical value of D_{22} and \hat{D}_{22} were calculated. Secondly, \hat{D}_{22} and $\max(D_{22}, \hat{D}_{22})$ were compared with the threshold of low-end and high-end values at defined discrimination level, respectively. Specifically, if \hat{D}_{22} and $\max(D_{22}, \hat{D}_{22})$ are smaller than their corresponding thresholds, the samples tested are not outliers, and vice versa. In this study, no outlier was found by employing this method.

3.3. Partial least squares regression and prediction

In partial least squares regression, concentration matrix of n components of m samples, $Y = (y_{ij})_{n \times m}$, as well as absorbance matrix of n samples at p wavelengths, $X = (x_{ij})_{n \times p}$, are decomposed into the form of eigenvectors, say,

$$Y = UQ + F \quad (2)$$

$$X = TP + E \quad (3)$$

Where $U_{n \times d}$ and $T_{n \times d}$ are characteristic factor matrix for component concentration and absorbance, respectively, and d is the number of representative components; $Q_{n \times m}$ and $P_{d \times p}$ is loading matrix of concentration and absorbance, respectively; $E_{n \times m}$ and $F_{n \times p}$ are residual matrix of concentration and absorbance, respectively [19].

The regression model built can be expressed as:

$$U = TB + E_d \quad (4)$$

Where E_d is composed of random error, while B is diagonal matrix of regression coefficient.

Thus, for unknown samples, only with their spectral value, we can obtain the chemical concentration by calculating

$$y=x(UX)'BQ \quad (5)$$

In this study, we employed “autoscale” as preprocessing method and leave-one-out cross validation was performed to test the stability of the established model. In this study, 5 principal components (PCs) were selected, because as Fig. 3 indicated, when the number of PCs were smaller than 5, there appeared sharper decreasing of RMSEC than that when more than 5 PCs were utilized. Finally correlation coefficients for calibration and validation were achieved as 0.9724 and 0.8691 (refer to Fig. 4), respectively, with PRESS (Prediction Residual Error Sum of Square) of 0.0684. However, there seemed to be some difference between calibration result and validation output. This may probably due to sample acquisition method, because we collected the samples randomly and no adjustment was implemented to keep STN levels in great range. What is more, a small sample set (only 22 samples are included in calibration set) would also take the responsibility. However, a low PRESS indicated that the established model is acceptable.

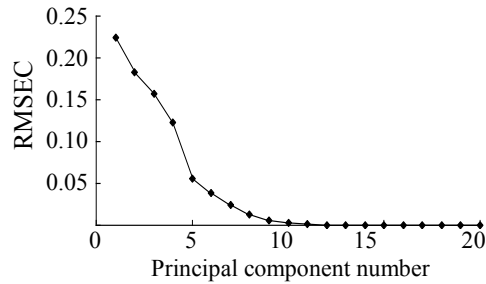


Fig. 3. Determination of optimal number of PCs

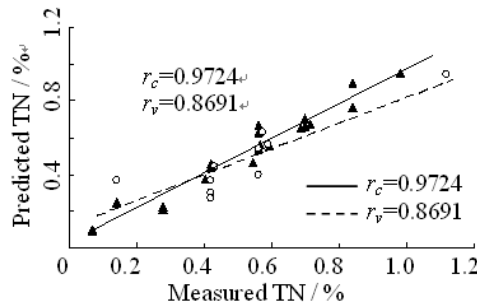


Fig. 4. Predicted vs. measured STN

Additionally, as Fig. 5 illustrated, there was no sample that had too much influence on model performance, that is, no outlier was discovered, which is consistent with Mahalanobis distance-Dixon method.

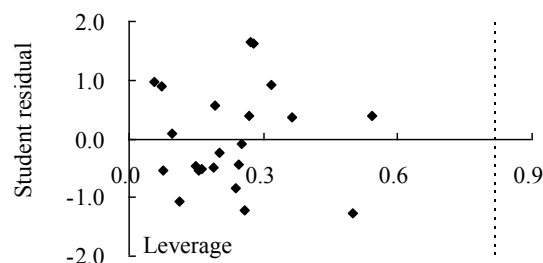


Fig. 5 Leverage vs. student residual plot

4. Conclusions

Visible shortwave near infrared spectroscopy (Vis-SWNIRS) is promising in sensing soil total nitrogen. In this study, data set assignment method and outlier detection criterion were utilized to enhance both precision and stability of the calibration model. 5 PCs were chosen as optimal rank to establish the calibration model, and correlation coefficients for calibration and validation were gained of 0.9724 and 0.8691, respectively. Additionally, though based on small sample set, this model is stable since PRESS was as low as 0.0684.

5. Acknowledgment

This research is funded by State Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau of China (funding number: 10501-226). Also, many thanks are acknowledged to Dr Wang Shanqin and Mr Li Shuo from College of Resources and Environment for their kind shear of the spectrometer and patient guidance on its operation procedure. Also, we are grateful to Mr Wu Wenqing and Mr. Wang Anping for their assistance during chemical experiment.

6. References

- [1] Zhou, Q.X., “*Healthy Soil: Soil Health Quality and Safety of Agricultural Products*”, Science Press, Beijing, China, 2005
- [2] Cozzolino, D., Kwiatkowski, M.J., Parker, M., et al, “Prediction of phenolic compounds in red wine fermentations by visible and near infrared spectroscopy”, *ANALYTICA CHIMICA ACTA*, Elsevier Science BV, Amsterdam, Netherlands, 2004, 73-80
- [3] Kim, S., Lim, Y.T., Soltész, E.G., et al. “Near-infrared fluorescent type II quantum dots for sentinel lymph node mapping”, *Nature Biotechnology*, Nature Publishing Group, New York, USA, 2004, pp. 93-97

- [4] Loo, C., Lowery, A., Halas, N.J., et al, "Immunotargeted nanoshells for integrated cancer imaging and therapy", *AMER CHEMICAL SOC*, Washington DC, USA, 2005, pp.709-711
- [5] Lee, Y., Chung, H., Kim, N., "Spectral Range Optimization for the Near-Infrared Quantitative Analysis of Petrochemical and Petroleum Products: Naphtha and Gasoline", *Applied spectroscopy*, 2006, pp. 892-897
- [6] Sun D.W. "*Infrared Spectroscopy for Food Quality Analysis and Control*", Elsevier, San Diego, California, USA 2009
- [7] Dalal, R.C., Henry, R. J., "Simultaneous determination of moisture, organic carbon and total nitrogen by near-infrared reflectance spectrophotometry", *Soil Sci. Soc. Am. J.*, The Soil Science Society of America, 1986, pp.120-123
- [8] Chang, C.W., and Laird, A., "Near-infrared reflectance spectroscopic analysis of soil C and N", *Soil Science*, Wolters Kluwer, 2002, pp.110-116
- [9] Reeves, J. and Maccarty, G., "The potential of NIRS as a tool of spatial mapping of soil compositions". *Journal of Near Infrared Spectroscopy*, IM Publications, West Sussex, United Kingdom, 1999, pp. 153-158
- [10] Wu, C.Y., Jacobson, A.R., Laba, M., et al., "Accounting for surface roughness effects in the near-infrared reflectance sensing of soils", *Geoderma*, 2009a, pp. 171-180
- [11] Wu, C.Y., Jacobson, A.R., Laba, M., et al., "Alleviating moisture content effects on the visible near-Infrared diffuse-reflectance sensing of soils", *Soil Science*, Wolters Kluwer, 2009b, pp. 456-465
- [12] Bao, Y.D., He, Y., Fang, H., et al., "Spectral Characterization and N Content Prediction of Soil with Different Particle Size and Moisture Content", *Spectroscopy and spectral analysis*, Peking University Press, Beijing, China, 2007, pp. 62-65
- [13] Sun, J.Y., Li, M.Z., Tang, N., et al., "Spectral characteristics and their correlation with soil parameters of black soil in Northeast China", *Spectroscopy and spectral analysis*, Peking University Press, Beijing, China, 2007, pp. 1502-1505
- [14] Michel, K., Terhoeven-Urselmans, T., Nitschke, R., et al. "Use of near- and mid-infrared spectroscopy to distinguish carbon and nitrogen originating from char and forest-floor material in soils", *Journal of Plant Nutrition and Soil Science-Zeitschrift Fur Pflanzenernahrung Und Bodenkunde*, 2009, pp. 63-70
- [15] Kennard, R.W., and Stone, L., "Computer aided design of experiments", *Technometrics*, American Statistical Association, Alexandria, USA, 1969, pp.137-148
- [16] Richard, G., Whitfield, Mark, E.G., et al., "Near-Infrared Spectrum Qualification via Mahalanobis Distance Determination", *Applied. Spectroscopy*, 1987. pp. 1204-1213
- [17] Min, S.G., Li, N., Zhang, M.X., "Outlier diagnosis and calibration model optimization for near infrared spectroscopy analysis", *Spectroscopy and spectral analysis*, Peking University Press, Beijing, China, 2004, pp. 1205-1209
- [18] Dean, R. B. and Dixon W. J., "Simplified Statistics for Small Numbers of Observations", *Anal. Chem.*, 1951, pp. 636-638
- [19] Yan, Y.L., "*Near Infrared Spectroscopy Fundamentals and Applications*", China Light Industry Press, Beijing, China, 2005