Tropentag 2023
September 20-22, 2023

Conference on International Research on Food Security, Natural Resource Management and Rural Development
organised by the Leibniz Centre for Agricultural Landscape Research (ZALF), Germany in cooperation with Humboldt-Universität zu Berlin, Germany

Visible and near-infrared spectroscopy for agricultural soil analysis using alternative data preprocessing and wavelength selection

Natchanon Santasup[a,b,c], Parichat Theanjumpol[b], Chanchai Sangchayosawat[a] and Nuttapon Khongdee[d]

[a] Department of Plant and Soil Sciences, Faculty of Agriculture, Chiang Mai University, Chiang Mai 50200, Thailand
[b] Postharvest Technology Research Center, Faculty of Agriculture, Chiang Mai University, Chiang Mai 50200, Thailand
[c] Graduate School of Chiang Mai University under the CMU Presidential Scholarship
[d] Department of Highland Agriculture and Natural Resources, Faculty of Agriculture, Chiang Mai University, Chiang Mai, 50200, Thailand

Abstract

Visible and near-infrared (VIS/NIR) spectroscopy has been extensively utilized to predict soil properties due to its rapidity, affordability, and environmental friendliness. However, the accuracy of predictions varied due to the regions, soil pedological characteristics, and particularly site-specific practices. Therefore, soil-specific predictive models should be developed to increase the accuracy of the model. This study aimed to evaluate the effect of data preprocessing and wavelength selection on the prediction of organic matter (OM), total carbon (TC), and total nitrogen (TN) in agricultural soil using NIRs. A total of 148 soil samples were randomly collected from different agricultural areas in northern Thailand for soil chemical components i.e. OM, TC, and TN analysis. The Walkley-Black method was used to analyze OM, while TC, and TN were analyzed by dry combustion technique. Soil samples were then scanned for VIS-NIR (400-2500 nm of wavelength), NIR (700-2500 nm of wavelength), and long-wave near-infrared (LWNIR) (1100-2500 nm of wavelength). Five data preprocessing techniques were tested, including Smoothing (SMO), Saviizki-Golay derivatives (SGD), Multiplicative scatter correction (MSC), Mean centering (MC), and Standard normal variate (SNV). Data preprocessing techniques were combined with partial least squares regression (PLSR) and principal component regression (PCR). The performance of the prediction was evaluated by the coefficient of determination (R^2) and the root mean square (RMSE). For a result, the best prediction was obtained with the combination of SMO preprocessing and the PLSR model in 400-2500 nm of wavelength. The R^2_P values of OM, TC, and TN were 0.83, 0.81, and 0.84 and $RMSE_p$ were 0.76, 0.47, and 0.04, respectively. This study demonstrated that the model could be used as an alternative method for determining OM, TC and TN in agricultural soil. However, large-sample populations and improved model algorithms could further improve prediction.

Keywords: Chemometrics, Prediction, Model, and Soil property

Introduction

Spectroscopic techniques are widely recognized as physical methods of characterization. They involve the investigation of the interaction between electromagnetic waves in the ultraviolet, visible, and infrared wavelengths and a given material. Moreover, the integration of these techniques with multivariate data analysis has demonstrated their efficacy in developing quantitative and classification models across various fields, such as food technology (Sun *et al*., 2021), petroleum engineering (He *et al*., 2020) and soil science (Meng *et al*., 2021). Recently, visible and near-infrared spectroscopy has become more popular for soil characterization. Barra *et al*. (2021) revealed that a wide range of soil properties can be predicted with high accuracies using this technique ($R^2 > 0.80$). This includes soil organic carbon/matter (SOC/SOM), inorganic carbon (IC), total carbon (TC) and soil texture. Nevertheless, the accuracy of predictions varied greatly due to the regions, soil pedological characteristics, and particularly site-specific practices. Therefore, soil-specific predictive models should be developed to increase the accuracy of the model for this technique. So, this study aimed to evaluate the effect of data preprocessing and wavelength selection on the prediction of organic matter (OM), total carbon (TC), and total nitrogen (TN) in agricultural soil using VIS-NIR spectroscopy.

Material and Methods

A total of 148 topsoil (0-15 cm) samples were collected by a composite method using an auger from various agricultural fields, then grass and plant debris on the surface were removed. All the soil samples were naturally air-dried in the laboratory and passed through a 0.5 mm mesh sieve before soil analysis and spectral data acquisition. The mentioned soil properties in this study were analyzed as follows: soil organic matter (SOM) was analyzed as total oxidized carbon and measured using the wet oxidation method (Walkley and Black, 1934), total carbon (TC) and total nitrogen (TN) were analyzed by combustion technique (Soil Health Institute, 2023). After that, soil samples were measured for NIR spectral data using the commercial NIR (400-2500 nm). Three wavelength ranges were separated thus: visible-near infrared: VIS-NIR (400-2500 nm), near-infrared: NIR (700-2500 nm), and long-wave near-infrared: LWNIR (1100-2500 nm). Five data pre-processing techniques were tested, including Saviizki-Golay smoothing, first derivatives, multiplicative scatter correction (MSC), mean centering (MC), and standard normal variate (SNV), partial least squares regression (PLSR) and principal component regression (PCR) were used for model development with 70% of soil samples were used for calibration and the rest 30% for validation. The performance of the prediction was evaluated by the coefficient of determination (R^2) and the root mean square (RMSE).

Results and Discussion

The average range of organic matter, total carbon, and total nitrogen in the 148 soil samples utilized for this investigation is presented in Table 1. Upon scanning the soil sample, a spectrum was acquired, as depicted in Figure 1.

Table 1 Statistics of soil chemical constituents

Soil Chemical Constituent	Range	Mean	S.D.
Soil organic matter (%)	0.60-6.40	3.04	1.68
Total carbon (%)	0.35-3.72	1.77	0.98
Total nitrogen (%)	0.03-0.32	0.15	0.08

Soil spectra were mainly dominated by a combination of fundamental vibrational bands for C-H, N-H, and O-H bonds and weak overtones. According to the data presented in Figure 1, it was observed that the light absorption range of the soil samples exhibited six distinct peaks. These peaks were identified at approximately 1186, 1426, 1510, 1900, 2134, and 2160 nm of wavelength, which correspond to the absorption range associated with specific chemical bonds, namely C-H (second overtone), O-H (first overtone), N-H (first overtone), C=O (second overtone) bonds, and N-H+C=O structure, respectively (Osbrone *et al*., 1993). These peaks are the main component of soil organic matter, total carbon and total nitrogen in the soil. Despite the fact that water absorbance features were observed around 1400 nm, they were also shown to be associated with aliphatic C-H. Additionally, the peak observed at 1900 nm was found to be associated with amide N-H. (Rodiguez-Perez *et al*., 2021). Moreover, the spectral shape around 2200 nm was associated with groups such as phenolic O-H, amide N-H, amine N-H, and aliphatic C-H (Moron and Cozzolino, 2003).

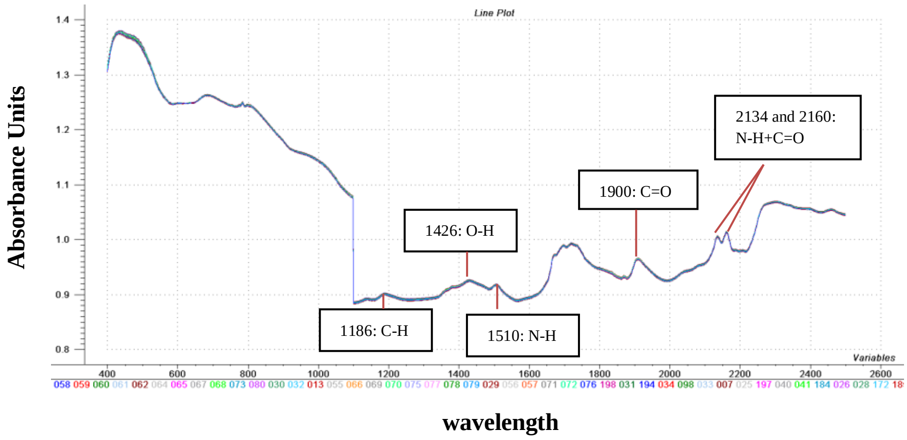

Figure 1 Raw NIR spectra of soil sample

Table 2. Model parameters and statistical indices for prediction of SOM using PLS and PCR regression with different data preprocessing (400-2500 nm)

Parameter	Spectral pre-processing	Calibration		Validation	
		R^2	RMSE	R^2	RMSE
SOM	-	0.68	0.670	0.57	1.210
	Smoothing	**0.86**	**0.608**	**0.83**	**0.758**
	1 st derivative	0.93	0.444	0.67	1.058
	MSC	0.94	0.397	0.61	1.157
	Mean centering	0.68	0.907	0.57	1.210
	SNV	0.94	0.397	0.61	1.157
Total carbon	-	0.68	0.590	0.57	0.806
	Smoothing	**0.82**	**0.397**	**0.81**	**0.468**
	1 st derivative	0.92	0.258	0.67	0.615
	MSC	0.95	0.193	0.63	0.656
	Mean centering	0.83	0.385	0.55	0.721
	SNV	0.90	0.295	0.55	0.725
Total nitrogen	-	0.66	0.053	0.60	0.059
	Smoothing	**0.85**	**0.031**	**0.84**	**0.037**
	1 st derivative	0.92	0.022	0.67	0.053
	MSC	0.85	0.031	0.53	0.063
	Mean centering	0.88	0.028	0.66	0.053
	SNV	0.85	0.031	0.53	0.063

For model development, the accuracy of PLSR was better than PCR in all conditions because PCR concentrates the reflectance data only in terms of the statistical properties. Therefore, the most relevant information may be included in the first few PCs and later used as regression factors during calibration. At the same time, PLSR balances the two objectives of explaining the response and predictor variation (thus, calibrations and predictions are more robust) and it performs decomposition and regression in a single step (Xu *et al*., 2018). Moving to wavelength selection, VIS-NIR has the highest prediction accuracy; this maybe caused by more peaks at the VIS (400-700) range related to SOM, TC, and TN. The result of this study was confirmed by Recena *et al.*

(2019) that the wavelengths around 488, 548, 660, and 772 nm were a significant peak for SOM prediction. For spectral preprocessing methods, Saviizki-Golay smoothing was the suitable method for predicting SOM, TC, and TN in this study. However, the suitable preprocessing method was varied due to soil sample, spectrometer, and conditions for spectral data acquisition (Barra *et al.*, 2021). So, in this study, the best prediction model for SOM TC and TN was obtained from the PLSR model with VIS-NIR wavelength and Saviizki-Golay smoothing, and the R^2 and RMSE of calibration and validation were shown in Table 2.

Conclusions

This research showed significant relationships between measured soil properties (SOM, TN, and TC) and VIS=NIR absorbance spectra in agricultural soil. The models developed using PLSR with VIS-NIR techniques could be suitable as useful tools to predict SOM, TN, and TC concentrations of agricultural soil.

Acknowledgements:

Financial support for this research from the Graduate School of Chiang Mai University under the CMU Presidential Scholarship

References

Barra, I., Haefele, S. M., Sakrabani, R., & Kebede, F. (2021). Soil spectroscopy with the use of chemometrics, machine learning and pre-processing techniques in soil diagnosis: Recent advances–A review. TrAC Trends in Analytical Chemistry, 135, 116166. doi:https://doi.org/10.1016/j.trac.2020.116166

He, K., Zhong, M., Li, Z., & Liu, J. (2020). Near-infrared spectroscopy for the concurrent quality prediction and status monitoring of gasoline blending. Control Engineering Practice, 101, 104478. doi:https://doi.org/10.1016/j.conengprac.2020.104478

Health, S. (2023). Soil total carbon and nitrogen by dry combustion. In. Morris: Soil Health Institute.

Meng, Y., Zhang, Y., Li, C., Zhao, J., Wang, Z., Wang, C., & Li, Y. (2021). Prediction of the Carbon Content of Six Tree Species from Visible-Near-Infrared Spectroscopy. Forests, 12(9), 1233.

Moron, A., & Cozzolino, D. (2003). Exploring the Use of near Infrared Reflectance Spectroscopy to Study Physical Properties and Microelements in Soils. Journal of Near Infrared Spectroscopy, 11(2), 145-154.

Osborne, B. G., Fearn, T., Hindle, P. H., & Hindle, P. T. (1993). Practical NIR Spectroscopy with Applications in Food and Beverage Analysis (2 ed.): Longman Scientific & Technical.

Recena, R., Fernández-Cabanás, V. M., & Delgado, A. (2019). Soil fertility assessment by Vis-NIR spectroscopy: Predicting soil functioning rather than availability indices. Geoderma, 337, 368-374. doi:https://doi.org/10.1016/j.geoderma.2018.09.049

Rodríguez-Pérez, J. R., Marcelo, V., Pereira-Obaya, D., García-Fernández, M., & Sanz-Ablanedo, E. (2021). Estimating Soil Properties and Nutrients by Visible and Infrared Diffuse Reflectance Spectroscopy to Characterize Vineyards. Agronomy, 11(10), 1895.

Sun, X., Li, H., Yi, Y., Hua, H., Guan, Y., & Chen, C. (2021). Rapid detection and quantification of adulteration in Chinese hawthorn fruits powder by near-infrared spectroscopy combined with chemometrics. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 250, 119346. doi:https://doi.org/10.1016/j.saa.2020.119346

Walkley, A., & Black, I. A. (1934). An Examination of the Degtjareff Method for Determining Soil Organic Matter and a Proposed Modification of the Chromic Acid Titration Method. Soil Science, 37(1): 29-38.

Xu, S., Zhao, Y., Wang, M., & Shi, X. (2018). Comparison of multivariate methods for estimating selected soil properties from intact soil cores of paddy fields by Vis–NIR spectroscopy. Geoderma, 310, 29-43. doi:10.1016/j.geoderma.2017.09.013