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

Natchanon Santasup[1,2,4], Parichat Theanjumpol[2], Chanchai Sangchayosawat[1] and Nuttapon Khongdee [3]

[1]Department of Plant and Soil Sciences, Chiang Mai University, Thailand
[2]Postharvest Technology Research Center, Chiang Mai University, Thailand
[3]Department of Highland Agriculture and Natural Resources, Chiang Mai University, Thailand
[4] Graduate school of Chiang Mai University under the CMU Presidential Scholarship, Thailand

Introduction

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 model's accuracy.

Objective

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

Materials and Methods

1 Soil sample collection

A total of 148 topsoil (0-15 cm) samples were collected by a composite method using an auger from various agricultural fields

2 Spectral data acquisition

- Measured for spectral data using the commercial NIR (FOSS NIR 6500)
- VIS-NIR (400-2500 nm)
- Near-infrared: NIR (700-2500 nm)
- long-wave near-infrared: LWNIR (1100-2500 nm)

3 Soil properties analysis

- Soil organic matter (SOM) : Walkley-Black method
- Total carbon (TC) and total nitrogen (TN) : Dry combustion technique

4 Data preprocessing

- Saviizki-Golay smoothing
- 1st derivatives
- Multiplicative scatter correction (MSC)
- Mean centering (MC)
- Standard normal variate (SNV)

5 Model development

- Multivariate calibration tools: partial least squares regression (PLS) and principal component regression (PCR)
- Calibration : Validation; 70%:30%

6 Model performance assessment

The performance of the prediction was evaluated by the coefficient of determination (R^2) and the root mean square error (RMSE)

Results

Table 1. Model parameters and statistical indices for prediction of SOM using PLS and PCR regression with different data preprocessing (400-2500 nm)

Model	Pre-processing	Calibration		Validation	
		R^2	RMSE	R^2	RMSE
PLS	-	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
PCR	-	0.57	1.043	0.61	1.159
	Smoothing	0.62	0.984	0.64	1.103
	1 st derivative	0.61	0.999	0.56	1.231
	MSC	0.57	1.048	0.56	1.232
	Mean centering	0.61	1.001	0.58	1.205
	SNV	0.59	1.028	0.55	1.240

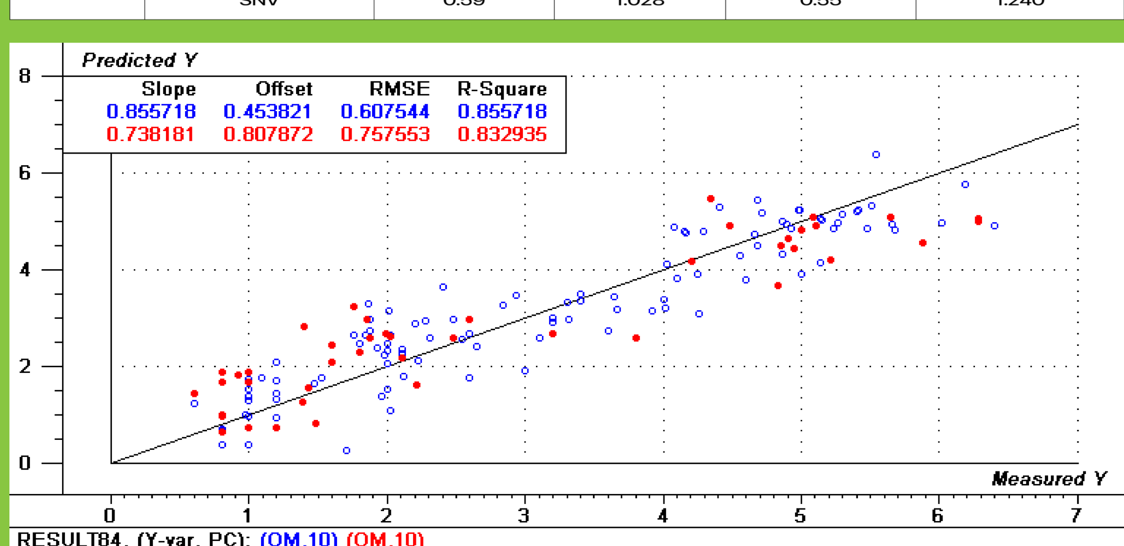

Figure 1: Correlation between measured and predicted values of SOM using PLS regression with smoothing preprocessing (400-2500 nm)

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

Model	Pre-processing	Calibration		Validation	
		R^2	RMSE	R^2	RMSE
PLS	-	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
PCR	-	0.61	0.582	0.58	0.700
	Smoothing	0.62	0.572	0.65	0.640
	1 st derivative	0.62	0.573	0.56	0.712
	MSC	0.59	0.597	0.55	0.721
	Mean centering	0.61	0.582	0.58	0.700
	SNV	0.59	0.597	0.55	0.721

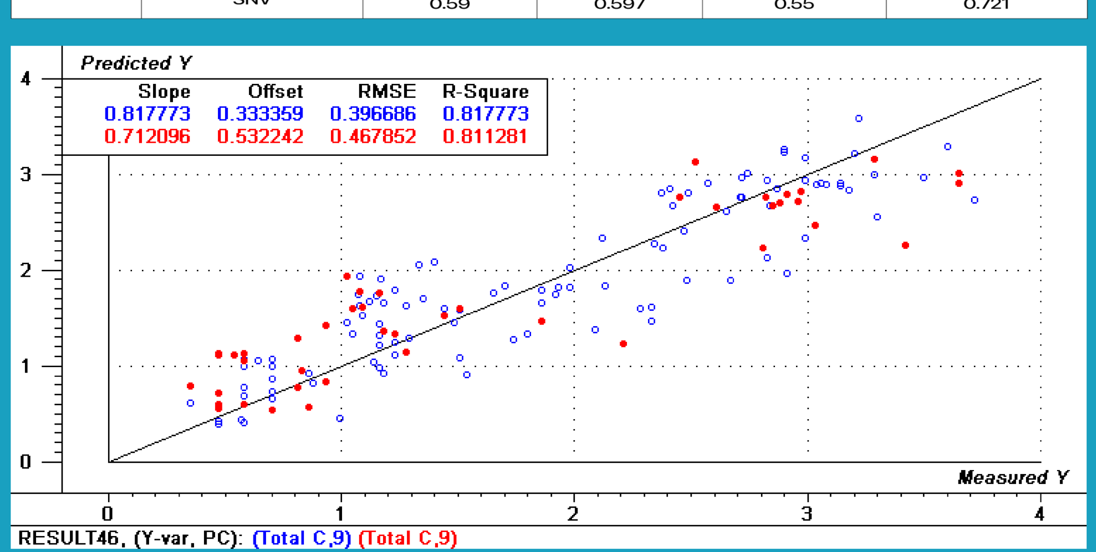

Figure 2: Correlation between measured and predicted values of TC using PLS regression with smoothing preprocessing (400-2500 nm)

Table 3. Model parameters and statistical indices for prediction of TN using PLS and PCR regression with different data preprocessing (400-2500 nm)

Model	Pre-processing	Calibration		Validation	
		R^2	RMSE	R^2	RMSE
PLS	-	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
PCR	-	0.60	0.050	0.58	0.060
	Smoothing	0.62	0.049	0.65	0.054
	1 st derivative	0.61	0.050	0.56	0.061
	MSC	0.58	0.051	0.55	0.061
	Mean centering	0.60	0.050	0.58	0.060
	SNV	0.58	0.051	0.55	0.062

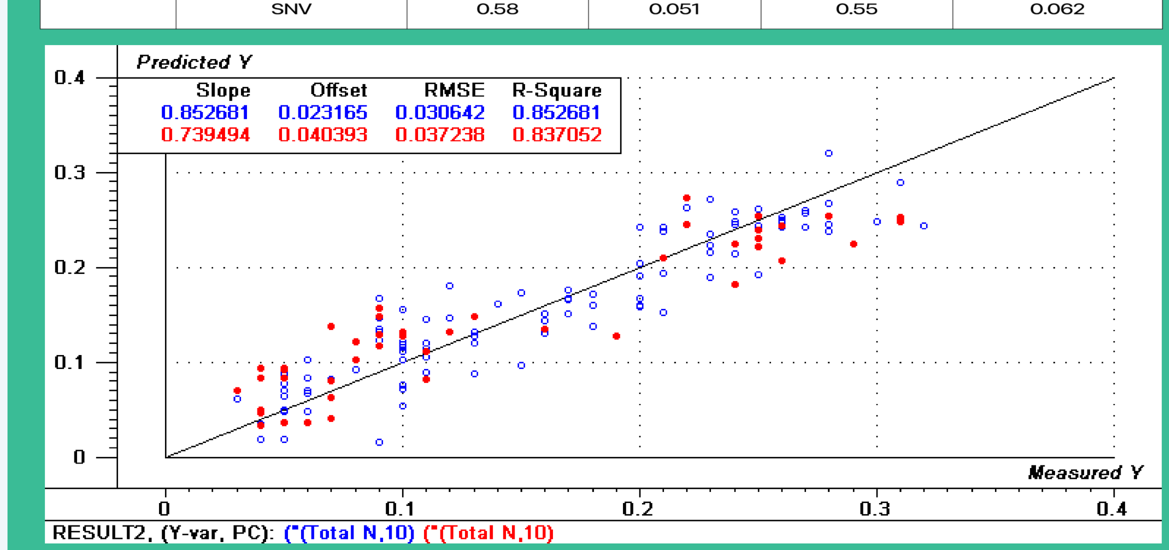

Figure 3: Correlation between measured and predicted values of TN using PLS regression with smoothing preprocessing (400-2500 nm)

Conclusion

Significant relationships were found 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 these soil.

Acknowledgments

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