Rapid Prediction of Soil Quality Indices using NIRS

-by Amitakshar, Arghyadip, Kuntal, Soumyajyoti

Table of Contents

PROJECT INFORMATION (ABSTRACT)

- This aim of this study is to apply near infrared spectroscopy (NIRS) in determining soil macro nutrients namely N, P and K.
- Near infrared spectrum were enhanced using de-trending (DT) method.
- Prediction models, used to predict N, P and K, were established using principal component regression (PCR) algorithm followed by leverage validation.
- The results showed that NIRS method can determine all three quality indices with good accuracy and robustness. Maximum correlation coefficient (r) for N, P, K prediction were achieved using DT correction method with r = 0.86 for N prediction, r = 0.90 for both P and K prediction.

Introduction

A short introduction to the project

- Origin of the Idea
- Near Infrared Spectroscopy (NIRS)
- What we did

Origin of the Idea

- In precision farming practices, it is important to monitor soil quality condition and health.
- Plants can grow ideally on healthy soil from which it has physical and chemical properties that are suitable with plant growth.
- Soil chemical properties are usually related to macro nutrients needed by plants like nitrogen, phosphorus and potassium.
- The above can added by fertilization practice.
- Fertilization and excessive use of fertilizer will cause pollution to the environment because it can cause artificial nutrient deposits that are not utilized by plants.

Near Infrared Spectroscopy (NIRS)

- Near infrared technology has been widely used and become most promising methods of analysis in many field areas including in soil and agriculture science due to its advantages.
- It has simple sample preparation, rapid, and environmental friendly process and the potential ability to determine multiple parameters simultaneously. The prediction model performance are sufficiently robust and accurate
- Numerous studies and publications on the application of near infrared spectroscopy (NIRS), shows that NIRS was feasible to be applied as a rapid and non-destructive tools for quality attributes prediction in agricultural sectors.

What we did

- Based on advantages and excellence of NIRS performance, we performed a study to apply the NIRS in predicting soil quality indices (N, P and K). These macro nutrients are crucial to be monitored in precision farming practices to ensure plant growth optimally.
- We attempted to apply spectra enhancement method namely de-trending (DT) and compare reduction result obtained from raw original spectra data.
- The prediction models were developed based on near infrared spectroscopic data and actual reference data using principal component regression (PCR).

Materials and Methods

The methods and materials used in the project

- Collecting soil samples
- Near Infrared Spectra acquisition
- Spectra data enhancement
- Soil macronutrients prediction
- Principal Component Analysis (PCA)
- Principal Component Regression (PCR)
- Steps of PCR
- Prediction model performance

Collecting soil samples

- ➔ A bulk of soil samples (nearly 100g) was collected from five different soil locations in Aceh Besar.
- ➔ A total of 40 soil samples were collected and stored for two days to equilibrate before spectra acquisition and further chemical analysis.
- ➔ Soil samples were then sorted from rocks and other substances to ensure uniformity.

Near Infrared Spectra acquisition

Spectral datasets of soil samples were acquired using a benchtop Fourier transform near infrared spectroscopy (FTIR, Thermo Nicolet Antaris II MDS). Soil sample was packed in a sample cup which was fixed on a quartz window through which a halogen lamp irradiated soil samples from down to up. Also the cup was spinning around slowly in order to obtain the averaged spectrum of each soil sample. Background spectra correction was taken once every 10 sample acquisitions. Spectral data was collected and recorded as absorbance in the presence of energies in wavelength range from 1000 to 2500 nm for the total 40 bulk soil samples. Each reading spectra data contained 1557 wavelength variables and as an average of 64 successive data acquisition.

Spectra data enhancement

- ➔ To minimize light scatter and interference effects, the collected spectra data were enhanced and corrected using de-trending (DT).
- ➔ Prediction models for N, P and K were developed using raw original spectra and de-trending spectra data.
- ➔ The enhanced data provided a better correlation for the model.
- ➔ One can also use other commonly used spectra correction methods like smoothing, normalization, transformation and spectra derivation to correct and enhance data.

Soil macronutrients prediction

- ➔ Soil quality indices, in form of soil macronutrients were predicted using principal component regression (PCR) and validated using leverage validation method.
- ➔ PCR Models were generated using NIR data as independent variable (X) and actual soil fertility properties: N, P and K as dependent variables (Y).
- ➔ The actual soil fertility properties were measured using conventional chemical and laboratory methods.

Principal Component Analysis (PCA)

- ➔ Principal component analysis (PCA) is the process of computing the principal components and using them to perform a change of basis on the data, sometimes using only the first few principal components and ignoring the rest.
- ➔ It is commonly used for dimensionality reduction by projecting each data point onto only the first few principal components to obtain lower-dimensional data while preserving as much of the data's variation as possible.

Principal Component Regression (PCR)

- ➔ In PCR, instead of regressing the dependent variable on the explanatory variables directly, the principal components of the explanatory variables are used as regressors.
- ➔ Often the principal components with higher variances are selected as regressors. However, for the purpose of predicting the outcome, the principal components with low variances may also be important, in some cases even more important.
- ➔ One major use of PCR lies in overcoming the multicollinearity problem which arises when two or more of the explanatory variables are close to being collinear.

Steps of PCR

- ★ Perform PCA on the observed data matrix for the explanatory variables to obtain the principal components, and then select a subset of the principal components for further use.
- ★ Now regress the observed vector of outcomes on the selected principal components as covariates, using ordinary least squares regression to get a vector of estimated regression coefficients
- ★ Now transform this vector back to the scale of the actual covariates, using the selected PCA loadings to get the final PCR estimate for estimating the regression coefficients characterizing the original model.

Prediction Model performance

- ➔ Prediction performances were quantified and judged for their accuracies and robustness using several statistical indicators: coefficient of determination (R^2), correlation coefficient (r), root mean square error (RMSE) and the residual predictive deviation (RPD) index.
- ➔ A good and robust model should have high R^2 and r, low value of RMSE and few number latent variables of PCR.

$$R^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y}^2)}, \; r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}$$

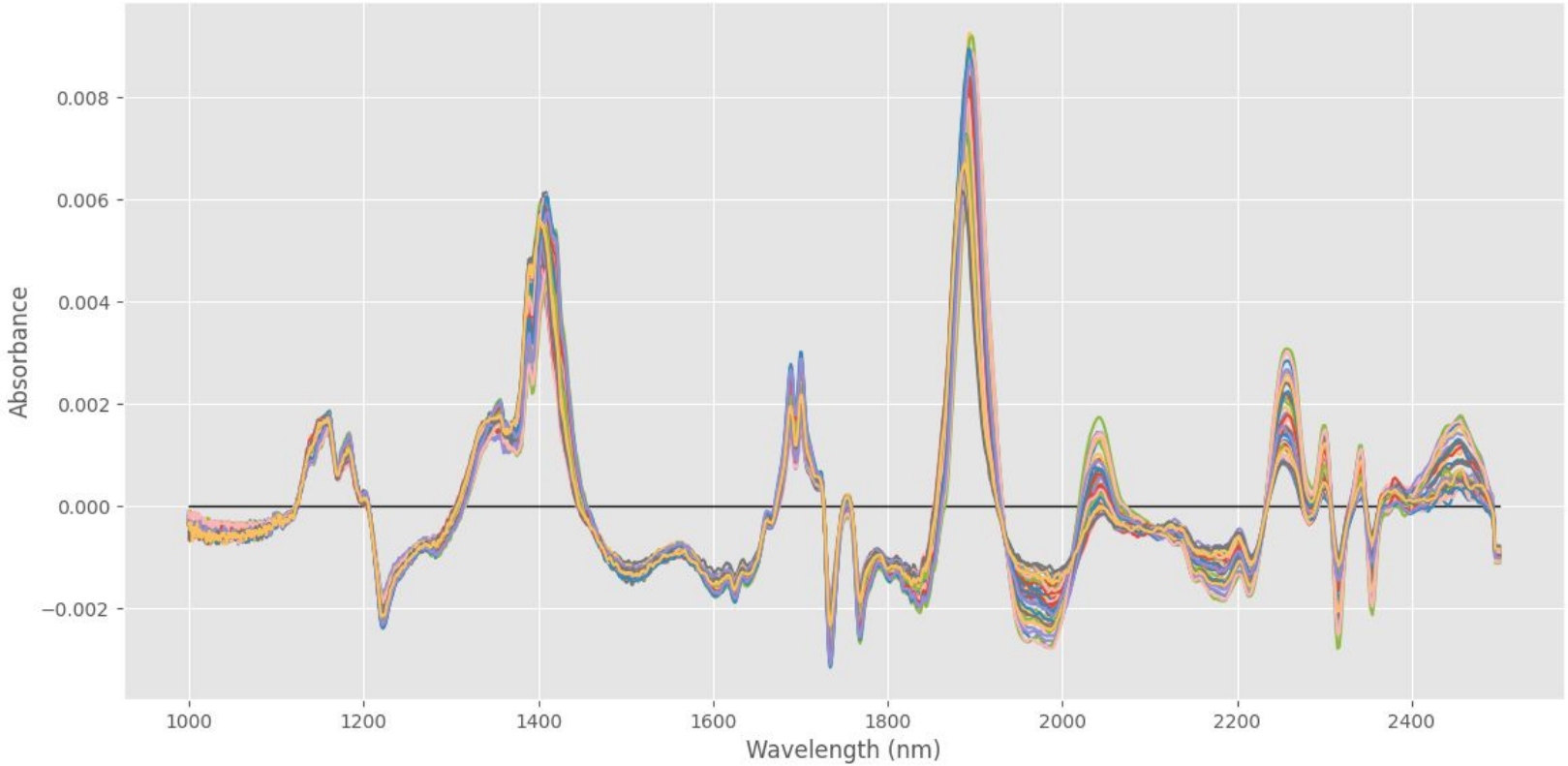
$$RMSE = \sqrt{\frac{1}{n} \sum (\hat{y}_i - y_i)^2}, \; RPD = \frac{\sigma_y}{RMSE}$$

Result and related discussion

The result obtained from PCR and related discussion

- Spectra of soil sample
- Validation & its importance
- Soil macronutrients prediction
- Observations
- Alternate models and macronutrients

Spectra Data of soil samples after DT enhancement



Spectra of soil sample

- The plot show several peaks which represent the vibration of molecular bonds of C-C, O-H, N-H, etc.
- Original spectra data contains interference due to noise from light scattering, so the spectra data is enhanced using de-trending enhancement method. The de-trending pre-treatment method tends to remove nonlinear trends in spectroscopic data and also reduce amplification due to light scattering and offset due to additive chemical effects
- Spectra correction and enhancement clearly will enhance spectra appearance and remove some noises due to light scattering and interfering effects. Now we can proceed with further analysis.

Validation & its Importance

- Overfitting is a kind of modeling error that occurs when a function is too closely fit to a limited set of data points.
- To avoid overfitting we divide the data into a calibration and a validation set. The calibration set is used to regress the model and the validation set is used to check the accuracy or robustness of the model obtained.
- Leverage validation has been used in this project. One can also use Cross-Validation, another well known validation technique.

Soil macronutrients prediction

- Macronutrient contents on soil samples were predicted using principal component regression (PCR) by regressing the near infrared spectra data (X-variables) and actual reference N, P, K data obtained using laboratory method (Y-variables).
- The process is done once with the original data and once using the DT enhanced data.
- The findings are presented in the following table:

Macro nutrients	Spectra Correction	Statistical Indicators				
		R^2	r	RMSE	RPD	LVs
N	Raw	0.68	0.74	1.31	1.31	7
	DT	0.77	0.86	0.07	2.05	6
P	Raw	0.78	0.85	5.27	2.17	8
	DT	0.84	0.90	4.76	2.41	6
K	Raw	0.80	0.85	0.21	2.45	7
	DT	0.84	0.90	0.17	3.03	7

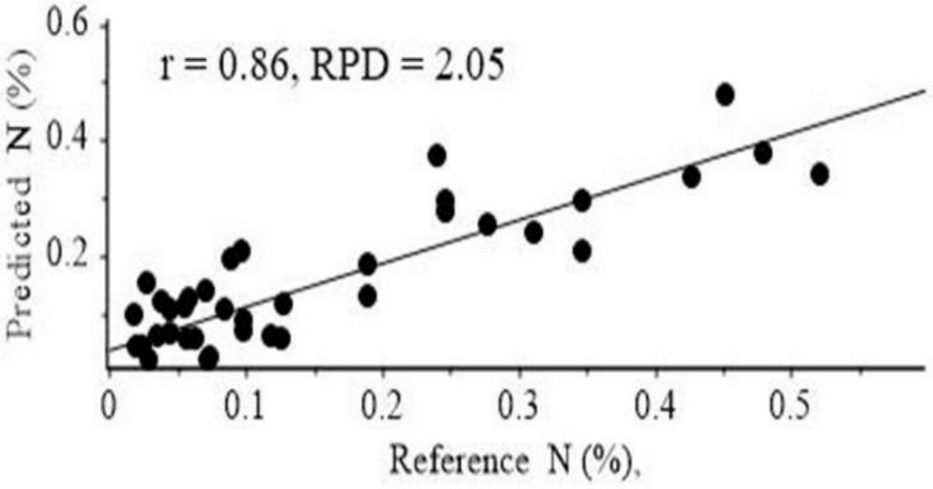
N: nitrogen content. P: phosphorus content, K: potassium content, DT: de-trending, R^2: coefficient of determination, r: coefficient of correlation, RMSE: root mean square error, RPD: residual predictive deviation index, LVs: latent variables

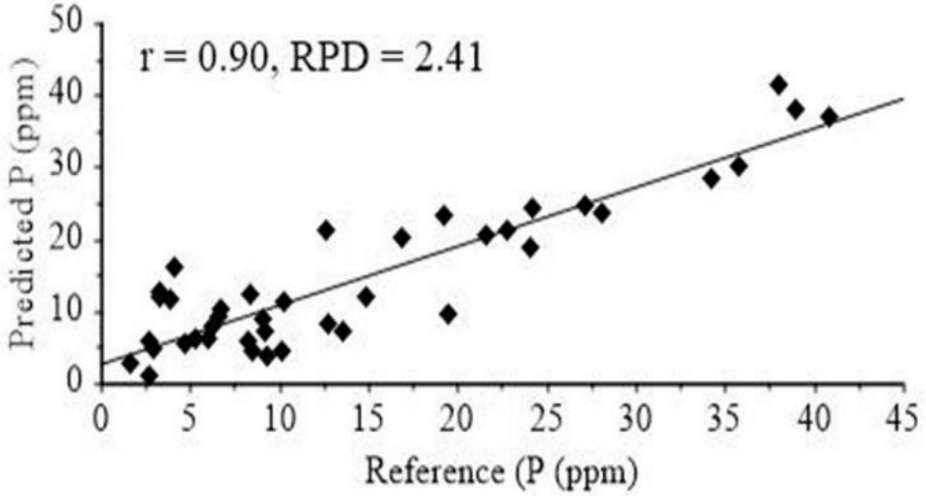
Observations

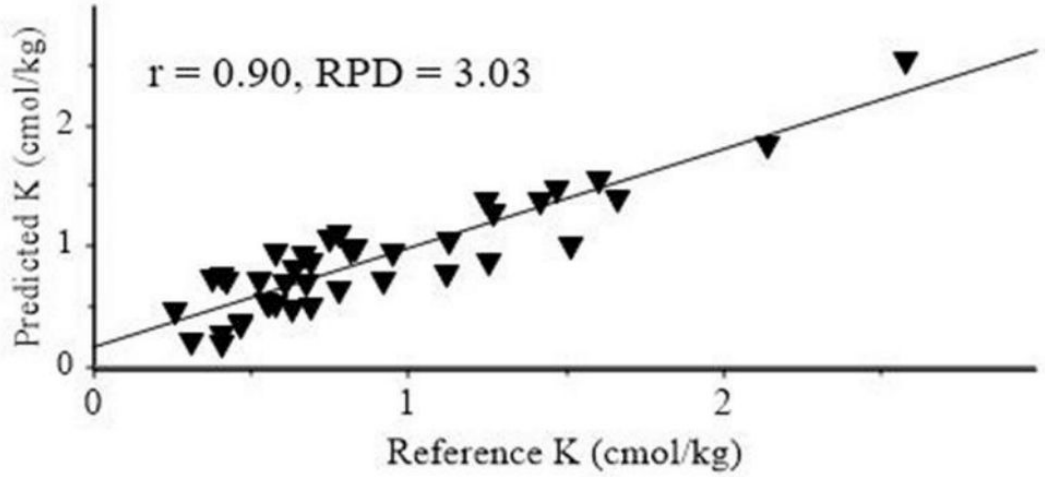
- ➔ At first, we attempted to predict N, P and K using raw original spectra data. As shown in the table, the maximum correlation coefficient achieved was 0.85 for P and K predictions
- ➔ Moreover, using the raw original spectra data, the maximum RPD index was 2.45 for K prediction which categorized as good prediction performance, while for P and N prediction, the RPD index were 2.17 and 1.31, which were categorized as sufficient and coarse prediction performance respectively.
- ➔ The maximum latent variables (LVs) required to establish the model was 8 for P prediction model, while for N and K, maximum LVs required was 7.

Observations

- ➔ Prediction model performances were improved when the models were developed using de-trending spectra data.
- ➔ The correlation coefficients were significantly increased for all three parameters. The maximum correlation coefficient was 0.90 for P and K predictions, while for N prediction, the r coefficient was 0.86.
- ➔ Moreover, the RPD index was also improved when the models were established using DT spectrum. The highest RPD index was achieved for K prediction (3.03) which categorized as excellent prediction performance.
- ➔ Scatter plot derived between actual reference soil macronutrients and predicted ones in the following slides:


r = 0.86, RPD = 2.05
Predicted N (%)
Reference N (%),
0.6
0.4
0.2
0
0.1
0.2
0.3
0.4
0.5


r = 0.90, RPD = 2.41
Predicted P (ppm)
Reference (P (ppm)
50
40
30
20
10
0
5
10
15
20
25
30
35
40
45


r = 0.90, RPD = 3.03
Predicted K (cmol/kg)
Reference K (cmol/kg)
2
1
0
0
1
2

Observations

- ➔ It is obvious that spectra enhancement significantly improved prediction accuracy and robustness for all those three soil macronutrients.
- ➔ As shown in the scatter plots, it is clear that NIRS can be used to predict N, P and K contents of soil samples.

Alternate models and macronutrients

So can we only use PCR and only to predict N, P and K values of the soil? No. We can also use partial least square regression (PLSR) to predict macronutrient contents on soil samples. In fact it might provide better results in some cases. And as for the case of macronutrients we can also predict Ca, Mg and even pH of the soil using the same method. We just need to measure them using conventional laboratory method and we can then regress them as the dependent variable (Y) to generate the required PCR Model or even PLSR Model.

Conclusion

Final statements to the project

- Final Statement

Final Statement

So based on achieved prediction results, we may conclude that near infrared spectroscopy (NIRS) can be applied in precision farming practises and employed as a rapid and environmental friendly method used to predict soil quality indices. This method can be used to monitor soil condition and take further actions required to maintain soil quality. Also spectra enhancement can be used to improve prediction accuracy and robustness for all macro nutrient parameters of soil samples. Achieved present study shows that NIRS technology was feasible to be used to monitor and rapidly determine the N, P and K contents of soil samples with good prediction accuracy and robustness.

Acknowledgement

We are very thankful to our Professor Rituparna Sen for giving us the chance to do this project and also helping us. We are also thankful to our friends for helping us. And ofcourse we are thankful to our group members who worked hard to make this happen.

Sources

- Original Article: https://doi.org/10.1088/1755-1315/365/1/012043
- Supplementary Article: https://doi.org/10.1016/j.dib.2020.105469
- Data available with the Article
- PCR: https://rpubs.com/esobolewska/pcr-step-by-step
- PCR in Python: https://nirpyresearch.com/principal-component-regression-python/
- PCA, PCR: Wikipedia

THANK YOU