Theme 2 | Advances in soil mapping and monitoring

Predictive Modeling of Soil Properties (Total Nitrogen and pH) Using MIR Spectroscopy.

Reda Mokere[1] *, Mohamed Ghassan [1] & Issam Barra [1]

1 *: Center of Excellence in Soil and Fertilizer Research in Africa CESFRA, College of Agriculture and Environmental Sciences CAES, Mohammed VI Polytechnic University, Benguerir, Morocco*
**: Reda.mokere@um6p.ma*

BACKGROUND

Mid-infrared (MIR) spectroscopy has emerged as a non-destructive, eco-friendly, and cost-efficient tool for soil analysis, offering high resolution and wide absorbance range capabilities. Integrating MIR data with machine learning, especially Partial Least Square Regression (PLSR), has shown excellent predictive capabilities for various soil properties such Total Nitrogen (TN) and pH. However, the accuracy of these predictions can be affected by factors like moisture, soil particle size, and baseline drift, necessitating effective data preprocessing.

AIM OF THE PROJECT

This study evaluates the impact of various preprocessing techniques and variable selection methods, such as SNV and Savitzky-Golay filters in combination with variable selection techniques, on improving the prediction accuracy of Total Nitrogen and pH.

MATERIAL AND METHODS

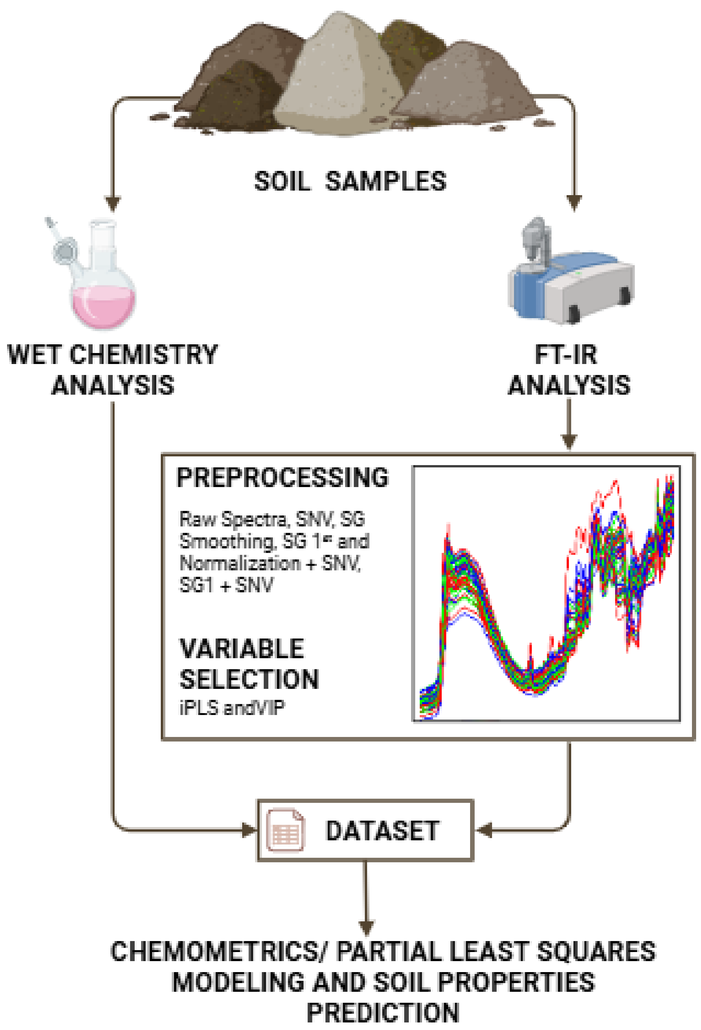

Fig.1: *Workflow for evaluating the impact of preprocessing and its combination with variable selection on PLSR models for estimating TN and pH.*

RESULTS

FTIR SPECTRA OF SOILS

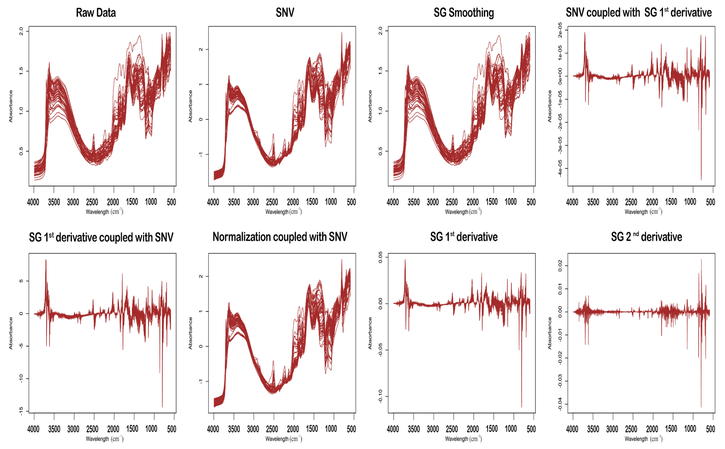

Fig. 2: *FTIR spectra of the soil samples before preprocessing and after applying different statistical pretreatment method*

SOIL PROPERTIES PREDICTION:

TN and pH were predicted using raw data and data subjected to various preprocessing techniques to identify the optimal preprocessing method for improving prediction accuracy. Then, the best preprocessing method was selected to apply the variable selection methods (VIP and iPLS).

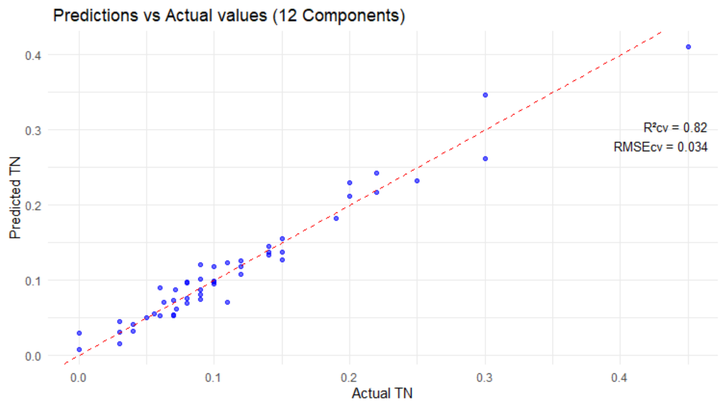

Fig. 3: *The prediction of TN using Raw data (Unprocesseced data)*

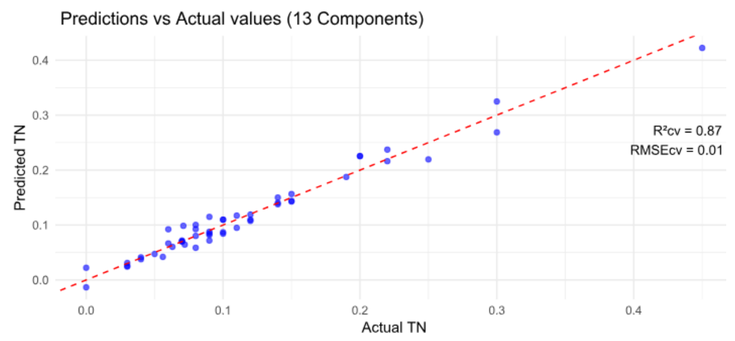

Fig. 4: *The prediction of TN using Savitzky-Golay 1st derivative*

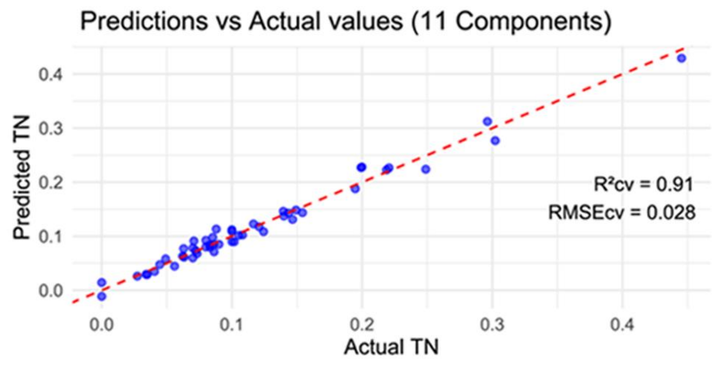

Fig. 5: *The prediction of TN using Savitzky-Golay 1st derivative Combined with variable in proection (VIP)*

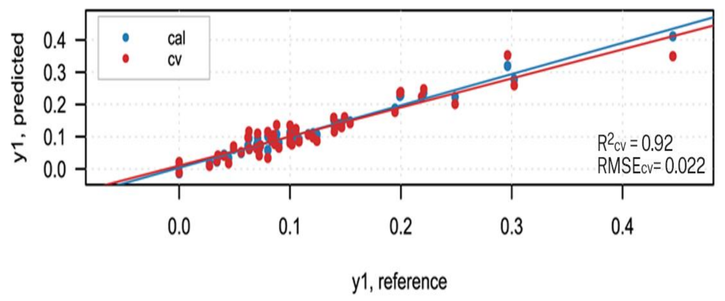

Fig. 6: *The prediction of TN using Savitzky-Golay 1st derivative and interval-PLS (iPLS).*

Table 1: *Table of merit of the prediction of TN and pH using MIR data and preprocessing data with the application of iPLS and VIP.*

	Raw Data		SNV		SG Smoothing		NOR+SNV		SG1	
	R2	RMSEcv	R2	RMSEcv	R2	RMSEcv	R2	RMSEcv	R2	RMSEcv
TN	0.82	0.03	0.8	0.03	0.81	0.03	0.83	0.03	0.87	0.01
pH	0.66	0.36	0.69	0.35	0.61	0.39	0.73	0.2	0.71	0.33

	SG1+SNV			SG1+iPLS				SG1+VIP		
	R2	RMSEcv		R2	RMSEcv	NI		R2	RMSEcv	TRC
TN	0.83	0.03		**0.92**	**0.02**	8		0.91	0.02	2.5
pH	0.66	0.37		**0.93**	**0.15**	73		0.86	0.23	3

TN: Total Nitrogen,, SNV: standard normal variate, NOR: Normalization, SG S: Savitzky–GolaySmoothing, SG1: Savitzky–Golay 1st Derivative, VIP: Variable Importance in Projection, iPLS: Intervals Partial least squares.

CONCLUSION

The Savitzky-Golay first derivative has proven to be the most efficient preprocessing method for improving the prediction of TN and pH using a Partial Least Squares Regression (PLSR) model. Due to these findings, SG1 was chosen as the optimal preprocessing method to be combined with VIP and iPLS approaches for variable selection. iPLS demonstrated excellent improvements for TN and pH, while VIP also yielded excellent results but iPLS performed better. Combining variable selection methods with adequate preprocessing significantly improved the prediction of the studied soil properties.

Acknowledgement

We gratefully acknowledge the support and guidance of the Center of Excellence in Soil and Fertilizer Research in Africa and Mohammed VI Polytechnic University, which were essential to this research.