

Comparative Analysis of Raw and Pre-processed Vis-NIR and MIR Spectra for Soil Property Estimation

Yasas Gamagedara¹, Gary Feng², Mary Love Tagert¹, Vitor S. Martins¹, Nuwan K. Wijewardane¹

¹Department of Agricultural & Biological Engineering, Mississippi State University, Starkville, MS, USA

²United States Department of Agriculture – Agricultural Research Service, Starkville, MS, USA

INTRODUCTION & OBJECTIVES

- High-throughput, cost-effective soil analysis is increasingly important for precision agriculture & soil health assessment
- Diffuse reflectance spectroscopy in the Visible–Near-Infrared (Vis-NIR) and Mid-Infrared (MIR) regions provides rapid, non-destructive soil property estimation
- Spectral pre-processing can enhance signal quality & improve model performance
- Limited comparison exists using large, harmonized datasets across both Vis-NIR & MIR domains

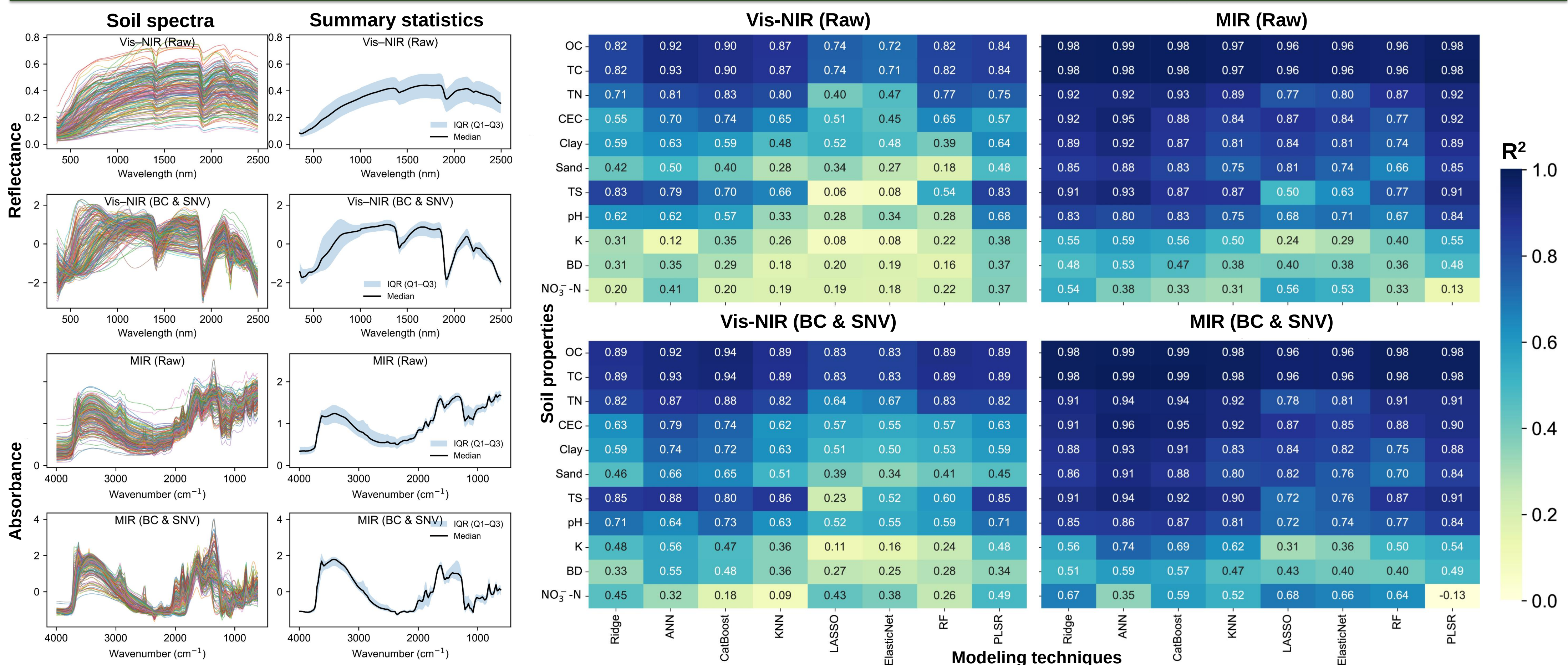
Objectives

- Compare raw vs. pre-processed Vis-NIR & MIR spectra for predicting 11 key soil properties
- Evaluate eight machine learning models for both spectral domains
- Identify which pre-processing techniques & spectral regions yield the best performance across soil properties

METHOD

- Dataset sourced from USDA-NRCS Kellogg Soil Survey Lab (KSSL)
- Selected 8,304 soil samples common to both MIR & Vis-NIR
- Soil properties (11): OC, TC, TN, CEC, clay, sand, TS, pH, K, BD, NO_3^- -N
- Spectrometers (2)
 - MIR (4000–600 cm^{-1}): Bruker Vertex 70 FTIR (HTS-XT)
 - Vis-NIR (350–2500 nm): ASD LabSpec
- Spectral pre-processing techniques (2)
 - Baseline correction (BC)
 - Standard normal variate (SNV)
- Machine learning models (8): Ridge regression (Ridge), Artificial neural network (ANN), Categorical gradient boosting (CatBoost), k-nearest neighbors regression (KNN), Least absolute shrinkage & selection operator (LASSO), Elastic net regularized regression (ElasticNet), Random forest regressor (RF), Partial least squares regression (PLSR)

RESULTS



DISCUSSION

- MIR provides higher accuracy than Vis-NIR for all soil properties
- Pre-processing significantly improves model performance, especially in Vis-NIR
- OC, TC, TN, and CEC show excellent predictability across domains
- BD and NO_3^- -N remain difficult, requiring further methodological improvements
- ANN & CatBoost consistently outperform other algorithms

FUTURE WORK

Integrate additional global spectral libraries (OSSS, LUCAS, AfSIS) to improve model generalization

CONCLUSION

- Pre-processing is essential for maximizing prediction accuracy in both Vis-NIR & MIR spectral domains
- MIR remains the most robust and stable method for soil property estimation
- ANN & CatBoost demonstrate strong generalizability & should be prioritized for operational workflows
- This study supports the strategic expansion of MIR spectroscopy for national-scale soil monitoring & sustainable agriculture programs

REFERENCE

Seybold, C. A., et al, (2019). Application of Mid-Infrared Spectroscopy in Soil Survey. Soil Science Society of America Journal, 83(6), 1746-1759