Spring 4-2023

Spectroscopic Sensor Data Fusion to Improve the Prediction of Soil Nutrient Contents

Bidhan Ghimire

SPECTROSCOPIC SENSOR DATA FUSION TO IMPROVE THE PREDICTION OF SOIL NUTRIENT CONTENTS

by

Bidhan Ghimire

A THESIS

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfilment of Requirements

For the Degree of Master of Science

Major: Agricultural and Biological Systems Engineering

Under the Supervision of Professor Yufeng Ge

Lincoln, Nebraska

April, 2023

SPECTROSCOPIC SENSOR DATA FUSION TO
IMPROVE THE PREDICTION OF SOIL NUTRIENT CONTENTS

Bidhan Ghimire, M.S.
University of Nebraska, 2023
Advisor: Yufeng Ge

This study aims to advance the understanding and application of spectroscopic sensor data fusion for improving soil nutrient content predictions. In addition to presenting an extensive review of studies on the spectroscopic sensor data fusion, a research investigation was conducted to assess the effectiveness of five fusion algorithms in predicting three primary nutrients (Nitrogen, Phosphorous, and Potassium) and two secondary nutrients (Calcium and Magnesium) in soil using Visible and Near-Infrared, Mid-Infrared, and X-ray Fluorescence data. Among the five fusion algorithms, one was a low-level fusion involving data concatenation. Two were mid-level fusions, incorporating feature extraction by applying (i) Principal Component reduction and (ii) Partial Least Squares reduction. The other two were high-level fusions, namely (i) Simple Averaging and (ii) Granger Ramanathan Averaging. The results indicate that Low-Level Fusion may not be suitable for inherently incompatible data. Mid-level fusion improved the R^2 by 0.1-18%, RMSE by 0-8%, and RPIQ by 0-11.5%, while high-level fusion enhanced the R^2 by 0-12.5%, RMSE by 0-12%, and RPIQ by 2.3-13.4%, depending on the nutrients and fusion algorithms. Despite these improvements, predictions were only satisfactory for primary nutrients, and none of the algorithms could notably enhance predictions for Phosphorous. The study also finds that fusion algorithms do not significantly improve

bias. The study provided evidence on improvement in prediction accuracy with data fusion which can aid in delineating management zones for precision agriculture. It also encourages further research on novel approaches of sensor fusion and algorithms that can effectively handle non-linearity introduced due to fusion of data.

Keywords: Soil spectroscopy, Sensor Fusion, Nutrient Prediction, VisNIR, MIR, XRF

ACKNOWLEDGMENTS

I want to extend my sincere appreciation to Dr. Yufeng Ge, my advisor, for his support, mentorship, and patience throughout my studies and research. Dr. Ge has been an exceptional inspiration to me, providing motivation during moments of difficulty and guiding me through challenging times.

I want to thank my thesis committee member and advisor, Dr. Stephen Scott, and my committee member Dr. Trenton Franz for their helpful ideas and advice. I would also like to thank Dr. Mohammad Omar Faruk Murad, Dr. Frank (Geng) Bai and Dr. Nuwan Wijewardane for their valuable suggestions throughout the period of the research.

I also acknowledge USDA-NIFA (2018-67007-28529) for providing funding for this research.

I want to thank my friends who helped me with my research and provided personal support during my stay in Lincoln. I sincerely appreciate Husein, Sabiha, Kanti, Shiva, Junxiao, Shubham, Sadia, Zhaocheng, Gautam, and Pascal. I would also like to express my gratitude to friends from the Nepali community, including Nikee, Nawaraj, Sagar, and everyone at East Campus Villas. Life in the States would have been completely different without Anita, Asu, and Suraj here and I would like to thank them. I would also like to thank Ashmita for being there for me.

I finally thank my parents, my elder brother, and Ramesh *Mama* for their unconditional love and support throughout my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

1.1 SPECTROSCOPIC SENSOR FUSION

Agriculture today mainly faces two challenges: (1) the increasing demand for food from a growing population and (2) the increasing need for environmental conservation practices (Nawar et al., 2022). These growing needs for higher productivity and sustainability are intricately twisted together and difficult to achieve simultaneously. Nevertheless, the scientific community is responding to these challenges with an approach called Precision Agriculture (PA). PA is a farming management approach involving better management of farm inputs by doing the right management practice at the right place and time and in the right form (Mulla, 2013). It improves the efficiency of agricultural inputs to maximize production and, at the same time, minimizes environmental impacts. While conventional management practice uses a whole-field approach, PA considers the within-field variabilities to customize the rate of production inputs (such as seeding and fertilization) and management practices (such as weeding) site specifically (Nawar et al., 2017). These treatments of areas with different variations or site-specific management are facilitated by delineating homogeneous sub-field regions with similar attributes of interest as separate management zones (MZs) (Castrignanò et al., 2018).

Proper delineation of management zones is crucial for site-specific management as they are often used to create recommendation maps for agricultural inputs. Management factors, the size of the field, and natural variability within the field determine how management zones are defined (N. Zhang et al., 2002). Soil characterization is necessary

to comprehend the natural variability within a field for delineating management zones. Since soil is a complex material with significant spatial and temporal variations, efficient techniques are needed for this characterization process. Although the traditional laboratory analysis has contributed much to our understanding of the unique characteristics of soil in the past, it is costly, time-consuming, and will not be able to meet these requirements (Viscarra Rossel et al., 2006). However, the advancement of information technology, sensor technology, geospatial analysis, and chemometric modeling has made it increasingly possible to identify and assess the spatial and temporal variability of soils economically and has given rise to soil spectroscopy.

Further, soil spectroscopy is environmentally friendly, non-destructive, and does not require expensive and time-consuming sample preprocessing. In general, soil spectroscopy involves the collection of spectra and the creation of multivariate statistical models to predict soil properties. As a result, we can quantitatively infer multiple soil properties with a single scan, which helps reduce costs and improve efficiency.

While the general concept of management zones has been established with an approach to maximize the efficiency of agriculture production and minimize environmental impact, not all management zones have the same objective. Further, they may not have the same size or are managed by the same management practices everywhere. Therefore, there is a trade-off between price and optimal temporal and spatial resolution depending on the management objective and size of MZs. Furthermore, multiple sensors are available, which one may use according to their need. These sensors can provide highly accurate and robust measurements efficiently at a massive scale.

Nevertheless, a single sensor may not be able to characterize all the properties of interest, and a combination of sensors can be employed in an approach called sensor fusion. Furthermore, the accuracy of a single sensor might be low because sensors are sensitive to moisture content, density, temperature and other factors which can and interfere with the signal from a specific property of interest. (Adamchuk. et al., 2011; Kuang et al., 2012). Therefore, sensor fusion has the potential to provide complementary information on specific property. This can improve prediction and provide operational benefits for agricultural applications (J. Wang et al., 2022).

Sensor fusion is an approach of combining results from multiple sensors to have a better understanding of the environment to make better decisions. An analogous method of sensor fusion is using multiple senses, experiences, and reasoning abilities by humans to improve the chances of survival. Sensor data fusion has long been used in other fields of science. The main goal of sensor data fusion is to utilize the synergy among sensor data to deduce more inferences than a single sensor (Y. Zhang et al., 2020). While individual sensors may completely fail but their combination can enhance the predictive relationship from their complementary effect (Ji et al., 2019). Not only for better accuracies but a combination of sensors can also improve the prediction's robustness, i.e., reducing uncertainty and increasing the system's reliability (Viscarra Rossel et al., 2011). Robustness is advantageous when faulty measurements from a single sensor can create issues. Furthermore, when independent measurements are made on the same system by multiple sensors, it is easier to identify and address errors or uncertainties as compared to a single measurement. Different sensors are less likely to be affected by a similar error

during the measurement. Thus, a combination of sensors also helps increase confidence in the measurement or prediction. Other operational benefits of the sensor fusion approach include extended attribute coverage and increased dimensionality of the measurement space.

The challenging task in sensor fusion is combining the data to utilize synergy among sensor data. It is easy to assume that a combination of sensor data will improve the output, but it is not always the case. Many times, the fusion of sensors produces results below the individual sensors. This type of fusion is called catastrophic sensor fusion. Theoretically, data fusion can help remove interferences by each sensor, but caution should be applied as data fusion can introduce the accumulation of prediction errors. For example, biases in different data sources may accumulate when data is fused from multiple sources. Another issue with sensor fusion is related to the costs. Sometimes, the benefit associated with sensor fusion falls short against the cost of using a single sensor

CHAPTER 2 LITERATURE REVIEW

2.1 INTRODUCTION

This paper will delve into different fusion levels and review approaches that scientists have adopted for data fusion. There are various types of sensor fusion, for instance, combining data from a single source taken at different times or combining data from multiple sensors or a mix of them. In this paper, we will use sensor fusion and data fusion as synonyms, as they are practically the same in the context of this paper. Generally considered, there are three levels of data fusion: Low-level, Mid-level, and High-level (Borràs et al., 2015; Elmenreich, 2002; Silvestri et al., 2013). Nevertheless, sometimes researchers apply hybrid approaches, mixing the different levels of fusion, such as fusing output from low-level fusion with latent variables extracted from data. Figure 1 is a flow diagram showing the fusion of data from an array of sensors at the low, medium, and high levels. In this chapter, multiple tables are presented for various degrees of spectroscopy data fusion. To improve the readability of these tables, numerous abbreviations are used, all of which are defined in Table 1.

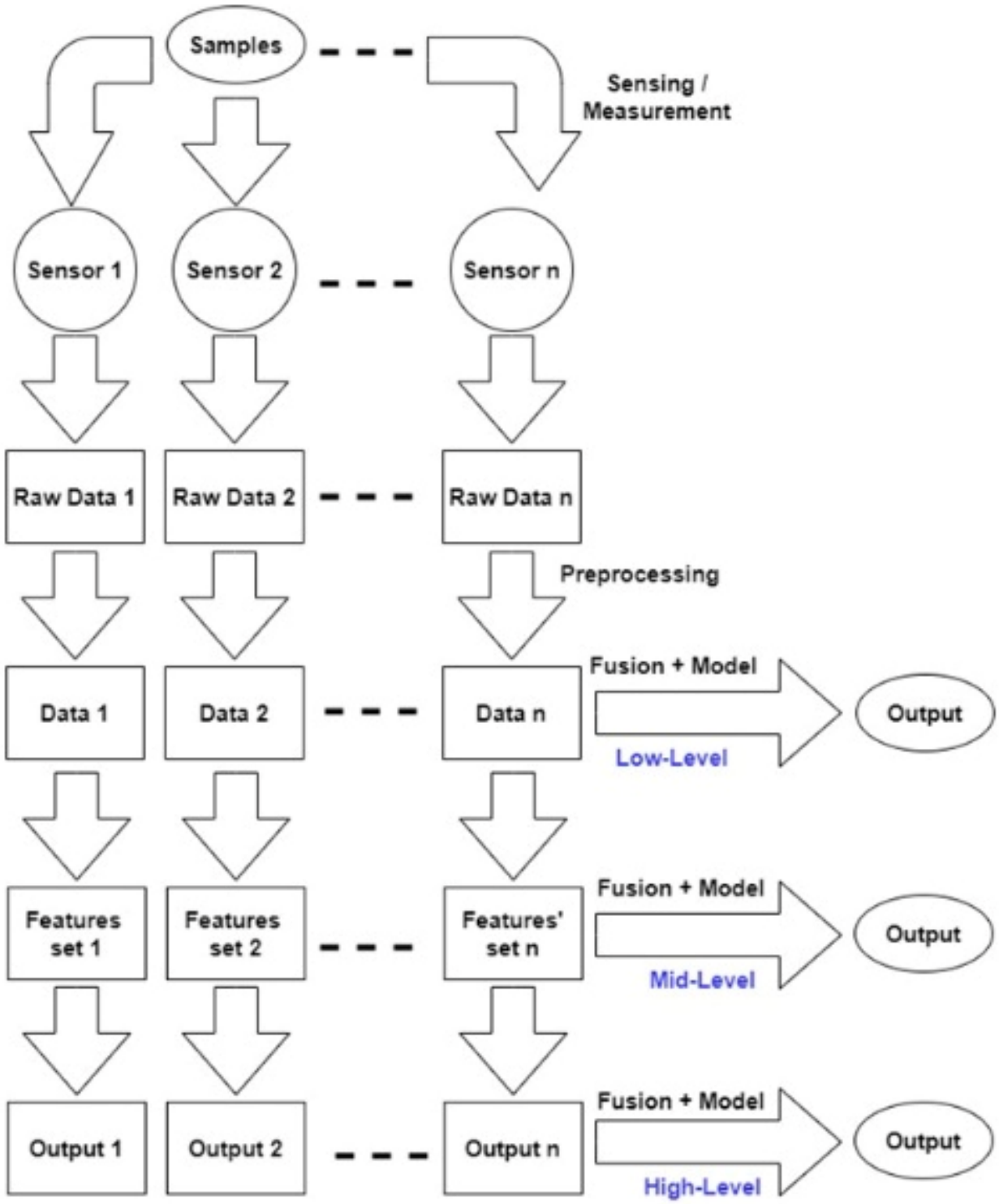

FIGURE 1 SCHEMATIC OF SENSOR FUSION AT THE LOW, MEDIUM, AND HIGH LEVEL

TABLE 1 LIST OF ACRONYMS

Acronym	Meaning
ANN	Artificial Neural Network
CART	Classification And Regression Tree
CEC	Cation Exchange Capacity
CNN	Convolutional Neural Network
DEM	Digital Elevation Model
ECa	Apparent Electrical Conductivity
EMI	Electromagnetic Induction
GA-PLSR	Genetic Partial Least Squares Regression
GNSS	Global Navigation Satellite System
INLA-SPDE	Integrated Nested Laplace Approximation with Spatial Partial Differential Equation
LBC	Lime Buffer Capacity
LIBS	Laser Induced Breakdown Spectroscopy
LWR	Locally Weighted Regression
MARS	Multivariate Adaptive Regression Splines
MIR	Mid Infrared
MRVBF	Multi-Resolution Valley Bottom Flatness
MZs	Management Zones
OC	Organic Carbon
OM	Organic Matter

PA	Precision Agriculture
PCR	Principal Component Regression
PLSR	Partial Least Squares Regression
QRF	Quantile Random Forest
RF	Random Forest
SMLR	Stepwise Multiple Linear Regression
SOM	Soil Organic Matter
SVR	Support Vector Regression
TC	Total Carbon
TN	Total Nitrogen
TWI	Topographic Witness Index
VisNIR	Visible Near Infrared
XRF	X-ray Fluorescence

2.2 LOW-LEVEL FUSION

This type of fusion is also called signal-level fusion and is the most straightforward fusion technique used to combine similar data from various sources. Although it is conceptually simple, we might have to apply caution while applying this approach. One significant drawback of this approach is the dominance of one data source over another, particularly when a source has a higher number of features or variables. If the dominating data source's features correlate poorly, the results can significantly deteriorate.

Additionally, this fusion level is associated with a large volume of data, presenting another challenge.

Concatenation of spectra is one of the most common approaches to low-level fusion. It increases the number of dimensions in data while keeping the observations unchanged. In this approach, data from different sources are generally normalized to make them compatible before they are combined (Comino et al., 2018). Nevertheless, sometimes, there are cases where normalization is not desirable, and a higher level of fusion are preferred (Barbedo, 2022). Several of the studies presented in this section will also be discussed in Sections 2.1.2 and 2.1.3. Therefore, this section will mostly discuss the low-level fusion aspects of these studies.

Studies have found that the concatenation of data helps improve the accuracy of results. For example, Wang et al. (2015) investigated modeling TC and TN with VisNIR and XRF using Penalized Spline Regression and Random Forest (RF) as modeling techniques. They reported that the synthesis of both spectra produced the best result compared to any of them individually. Similarly, Aldabaa et al. (2015) predicted soil salinity with the combination of VisNIR, XRF, and Landsat spectral data and reported similar results. Moreover, for some studies, the results were not only better than the individual data, but concatenation showed results superior to high-level fusion (J. Wang et al., 2022). For instance, (Wan et al., 2020) concatenated XRF and VisNIR data to predict CEC and reported improvement over a single dataset. Similarly, (Filippi et al.,

2020; Ji et al., 2019; N. Li et al., 2018) also reported improvement in results for most of the properties predicted.

It is not always the case that the concatenation improves the result. Studies have also shown no improvement or even poorer results with concatenation. Fontenelli et al. (2021) reported no improvement in results when they combined VisNIR, MIR, and XRF for various soil properties. (Comino et al., 2018) used data from NIR spectroscopy and X-ray fluorescence and compared the results with models using a single source. These authors reported that low-level fusion might not be the best option to improve the predictive potential of the sensors if they are not sensitive enough. Ng et al. (2019) also fused visNIR and MIR spectra and employed Partial Least Squares Regression (PLSR) and Cubist methods for chemometric modeling. They reported no significant improvement compared to the models with only MIR spectra and reasoned it to be due to the high predictive accuracy of the MIR spectra with PLSR and Cubist models. Further, (D. Xu et al., 2019) reported no improvement or even weaker predictions when they fused VisNIR, MIR, XRF, and Laser-induced breakdown spectroscopy **(**LIBS) spectra to predict Soil organic matter (SOM), Total Nitrogen (TN), Phosphorous (P), Potassium (K), and pH. Some studies reported mixed responses on the concatenation type of fusion. Rodrigues Tavares et al. (2020) reported they saw improvement for properties like clay, pH, and P while no improvement for others. Similarly, the predictions were better for only two of the five properties in another study by Javadi et al. (2021).

Studies have shown that some processing of data is sometimes required before fusing datasets. For example, Kriging is a geostatistical method used to determine a variable's value based on observations around it. It assumes that observations of a variable are spatially correlated, and this correlation will decrease with an increase in distance. Kriging, along with concatenation, can be an approach for low-level, mid-level, or high-level fusion according to the application. For example, Li et al. (2018) used kriging with low-level fusion technique where they fused (interpolated) elevation, gamma-ray, and electrical conductivity data by ordinary kriging and modeled soil Cation Exchange Capacity (CEC) using Bayesian Inference. They reported that the results modeled with the combined data were the most accurate. Pixel fusion is also another approach of low-level fusion, which is generally employed on image data. This fusion is performed on a pixel-by-pixel basis. It produces a fused image in which information associated with a pixel is derived from a set of pixels in source images to improve the performance of image processing tasks such as segmentation (Dong et al., 2009).

Table 2 summarizes the studies that have implemented concatenation as a fusion approach. Although there was an improvement in results for most of the studies, some showed a decline in performance with the concatenation of data, and some showed a mixed performance. For this approach, the most common data type fused was spectroscopic data, and the most common property of interest is CEC. Furthermore, the most common chemometric modeling approaches we observed were PLSR and Support Vector Regression (SVR).

TABLE 2 LOW-LEVEL FUSION WITH CONCATENATION APPROACH

Research	Property	Data Sources	Modeling Algorithm	Improvement over best individual sensor
(Tavares et al., 2020)	OM, CEC, pH, base saturation(V), P, K, Ca	VisNIR, XRF	PLSR, SVR	Yes, for Clay, pH, P
(Comino et al., 2018)	N, P, K, Ca, Mg, Mn, Zn, and B	NIR, XRF	PLSR	No
(Ng et al., 2019)	sand, clay, TC, OC, CEC, and pH	VisNIR, MIR	PLSR, Cubist	No
(D. Xu et al., 2019)	SOM, TN, AN, AP, pH	VisNIR, MIR, XRF, LIBS	GA-PLSR	No
(D. Wang et al., 2015)	TC and TN	VisNIR and XRF	PSR, RF	Yes

(J. Wang et al., 2022)	CEC and pH	EMI, gamma-ray, Sentinel-2, EM, depth	QRF	Yes
(Wan et al., 2020)	CEC	VisNIR and XRF	PLSR, SVR	Yes
(N. Li et al., 2018)	CEC	DEM, gamma-ray, EM induction	Bayesian inference (INLA-SPDE)	Yes
(Filippi et al., 2020)	Soil sodicity	Landsat, DEM, Slope, TWI, MrVBF,	RF	Yes
(Ji et al., 2019)	SOM, pH, LBC, P, K, Ca, Mg, Al	GNSS data, ECa (EMI), gamma-ray, VisNIR,	PLSR, SVR, RF, MARS, CART	Yes
(Javadi et al., 2021)	pH, OC, P, Mg, Calcium, Na	VisNIR, XRF	PLS, LWR, CNN	Yes, for OC and Mg
(Fontenelli et al., 2021)	Clay, sand, Silt, P, K, Ca, Mg, CEC, SOM,	VisNIR, MIR, XRF	PLSR	No

2.3 MID-LEVEL FUSION

This type of fusion is also referred to as feature-level fusion, which involves extracting salient features independently from each source and fusing similar elements to build a model. These features are thought to be free from background noises and can improve the model performance. The results are more interpretable since the contribution of each data set can be analyzed compared to those of low-level fusions.

One of the most common mid-level fusion approaches is to extract latent variables from the raw sensor data. Usually, researchers either extract PC scores or latent factors from partial least squares discriminant analysis (PLS-DA) separately from each data source (Biancolillo et al., 2014) and combine them afterward. The selection of the optimal combination of extracted features and preprocessing techniques is challenging. Generally, researchers select the number of PC components for each data source using two approaches. The first is choosing the number of PC components that would explain a particular amount of variance in the data, for example, 95% or 99%. The second approach is through the predictive performance during the cross-validation. In addition, researchers may apply multiple data preprocessing techniques separately which can result in distinct variations of preprocessed data. Consequently, multiple sets of features are derived from a single data source leading to several combinations of fused data matrices. Another approach to feature extraction is through PLS reduction. In PLS reduction, the predictor variables are projected into a new space of lower dimensionality such that there is maximum covariance between predictor and response variables. This is done by

creating latent variables, which are linear combination of original variables (Geladi, Paul, and Kowalski, 1985).

We found that almost an equal number of studies reported improvement and no improvement in results. Some studies showed better results than individual sensors and low-level fusion. Studies like (Comino et al., 2018) reported that they fused PC components from two sources and observed no decrease in the predictive performance of the best individual model for any properties, and it was better than low-level fusion. Even when the results were comparable, Ratio of Performance to Deviation (RPD) was better. Further, they suggested that a rapid screening to detect nutrient deficiency was feasible with this kind of fusion. Similarly, a study by (X. Xu et al., 2019) found improvement when they applied two approaches of mid-level data fusion to predict SOM: a) fuse together the PC scores of preprocessed LIBS and MIR spectra, b) fuse the latent variables of preprocessed spectra acquired by PLS Reduction. The result was also better than the low-level data fusion they applied. Nevertheless, while some studies showed better performance than individual sensors, they had lower performance than low-level fusion. A study by Zhao et al. (2022) combined VisNIR, MIR, and XRF data to predict grain-size distribution, clay-mineral properties, and geochemical ratios and reported that it was better than individual sensors but was inferior to low-level fusion.

Studies showed PC fusion had similar or slightly lower performance compared to individual data. A study by Mahmood et al. (2012) observed poor performance with the PC fusion, whereas research by Xing et al. (2021) found similar or slightly lower performance. The study by Xing et al. (2021) employed a lower-level fusion which had

better results than both mid-level fusion and individual spectra. Some studies also showed mixed performance results. In the study conducted by Javadi et al. (2021), they employed the PC approach but with two different modeling techniques: PC fusion with PLS (PC-PLS) and PC fusion with CNN (PC-CNN). In general, PC-PLS was lowest among the PC-CNN approach, low-level fusion, and even individual sensors in terms of prediction accuracy. A study by Tsimpouris et al. (2021) fused PC scores from different variations of data from the same sensor rather than different sensors and showed improvement in results. These studies regarding the PC-based and PLS-based mid-level sensor fusion is summarized in Table 3**.**

Researchers also employ another method of mid-level data fusion: Outer product analysis (OPA). OPA emphasizes the co-evolution of spectral regions by containing the product of all combinations of the elements of the vector (Barros et al., 2008). In simple terms, absorbances from different spectral ranges are brought together into one common spectral domain, where these absorbances become more intense (Terra et al., 2019). A slightly modified method is also employed sometimes, which involves the outer product of the PC of the spectra instead of the original spectral data to reduce the computational load. The output matrix of an OPA between two datasets, obtained by multiplying each element of the first data with each element of the second, will be a 3-dimensional matrix with the first, second, and third dimensions representing the number of samples and signals from the first and second datasets, respectively. One would apply a multiway method of unfolding the matrix into two dimensions to analyze it. For OPA, we observed it is common to employ methods to reduce the dimensions of the outer product matrix to

reduce the computational load. For example, some extracted the PC scores of the spectra before multiplying like Cécillon et al. (2012) and some extracted the PCs after they found the outer product matrix (Veselá et al., 2007). Others sample the matrix by selecting a representative value for a certain number of variables after OPA.

A higher proportion of studies reported poor performance with OPA fusion. For example, a study by (Veselá et al., 2007) used the OPA approach to fuse NIR and MIR data and found results similar to those without fusion. Similarly, studies by (Vohland et al., 2022; Zhao et al., 2022) also found that results with OPA were no better than the individual sensors. They reported lower performance compared to a fusion with concatenation. (Javadi & Mouazen, 2021) had also conducted another study and found similar results. Nevertheless, some studies argued in favor of OPA even when they observed no improvement in predictive performance directly. A study combined the NIR and MIR spectra to discriminate five biochemical properties and reported that the performance was better than the NIR model but similar to the MIR model (Bellon-Maurel & McBratney, 2011). They suggested using a fused model because MIR spectra alone can be questionable for soil samples with contrasting particle sizes and moisture content. In another study, (Ng et al., 2019) used CNN with OPA and observed slightly better prediction than the simple concatenation but no significant improvement. Despite this, they concluded their study with the importance of OPA in utilizing information in several domains (ranges) to predict a parameter of interest and identifying relationships between different spectral variables in various domains concerning those parameters. Some studies showed improvement directly in prediction. Terra et al. (2019) conducted a study

investigating the fusion of vis-NIR and MIR spectra to predict soil organic C using the OPA approach and reported that it produced significantly better results than any single source. Similarly, Xu et al. (2020) and (H. Xu et al., 2020) found improved results when using OPA to fuse data. We have summarized these studies in Table 4.

2.4 HIGH-LEVEL FUSION

This type of fusion is also called decision-level fusion. The outputs of different models are fused to improve the prediction. The crucial task in this type of fusion is to find suitable models for the individual datasets so that their combination produces even better predictions. One of the significant advantages of this approach is that the result will not be worsened by any of the inefficient data compared to other levels of fusion. This is mainly due to individual datasets being treated independently. However, with high-level fusion, although we may not get poor performance compared to individual sensors, we may lose information if we fail to capture correlation among the datasets. Model averaging is the most common approach to decision-level fusion, and among the different averaging approaches, simple averaging is the simplest. Other model averaging approaches that have been used are Bates-Granger Averaging (BGA), Granger-Ramanathan averaging (GRA), and Bayesian model averaging.

a. Simple Averaging (SA)

SA is also called Equal Weights Averaging, and the final prediction of the model after the simple averaging can be given by;

TABLE 3 MID-LEVEL SENSOR FUSION WITH FEATURE EXTRACTION APPROACH, PROPERTIES ARE OF SOIL IF NOT OTHERWISE STATED

Research	Property	Data sources	Modeling	Improvement over best individual sensor
(Comino et al., 2018)	N, P, K, Ca, Mg, Mn, Zn,	NIR, XRF	PLSR	Yes
(X. Xu et al., 2019)	SOM	LIBS and MIR	PLSR, SVR, ANN	Yes
(D. Xu et al., 2019)	SOM, TN, N, P, pH	VisNIR, MIR, XRF, LIBS	GA-PLSR	No
(Javadi et al., 2021)	pH, OC, P, Mg, Ca, Na	VisNIR, XRF	PLS, CNN	No (PLS) Yes, except P and Ca (CNN)
(Zhao et al., 2022)	grain-size distribution, clay-mineral properties	VisNIR, MIR, XRF	RF, MLR	Yes
(Mahmood et al., 2012)	soil texture, EC, pH, TOC, TN, and C/N	VisNIR, ECa	SMLR	No, except for sand content
(Xing et al., 2021)	SOM	MIR, Raman spectra	PLSR	No
(Tsimpouris et al., 2021)	particle size distribution, pH, CEC, OC, CaCO3,	VisNIR	ANN	Yes

TABLE 4 MID-LEVEL SENSOR FUSION WITH OPA APPROACH, PROPERTIES ARE OF SOIL IF NOT OTHERWISE STATED

Research	Property	Data sources	Modeling	Improvement over best individual sensor
(Zhao et al., 2022)	grain-size distribution, clay-mineral properties	Vis-NIR, MIR, XRF	RF, PLSR	No
(H. Xu et al., 2020)	Soil classification*	Vis-NIR, MIR	SVR	Yes
(Dongyun Xu et al., 2019)	Chromium	XRF, VisNIR	PLSR	Yes
(Javadi & Mouazen, 2021)	pH, OC, Mg, Ca	VisNIR, XRF	PLS	No
(Terra et al., 2019)	C	VisNIR, MIR	SVR	Yes
(Ng et al., 2019)	sand, clay, TC, OC, CEC, and pH	VisNIR, MIR	CNN	No
(Cécillon et al., 2012)	SOM Composition	NIR, MIR	PCT	No
(Veselá et al., 2007)	fat, nitrogen, and moisture in cocoa powder	NIR, MIR	PLS	No

$y = \frac{1}{n} \sum_{i=1}^{n} y_i$

y_i is the predicted output from the i^{th} source, n is the number of data sources, and y is the final predicted output (J. Wang et al., 2022).

In (J. Wang et al., 2022), we observed that, for CEC, the fusion results with simple averaging were better than any of the single sources, as shown by the Lin's concordance correlation coefficient (LCCC), RPD, and RMSE. Nevertheless, the opposite was true for pH, where fusion with simple averaging always produced poorer results than the single source that was best among them. Mostly, the prediction models were built using PLSR.

b. Bates Granger Averaging (BGA):

This approach is one of the earliest works in combined forecasting and uses diagonal elements of the estimated mean squared prediction error (MSPE) matrix to evaluate the combination of weights (Bates & Granger, 1969; J. Wang et al., 2022). The final prediction of the output is given by,

$y = \sum_{i=1}^{n} w_i y_i$

where y_i is the predicted output from the i^{th} source and

$$w_i = \frac{\widehat{\sigma}^2(i)}{\sum_{j=1}^{n} \widehat{\sigma}^2(j)}$$

where, w_i and $\widehat{\sigma}^2(i)$ are the weight and estimated MSPE of the single data source, respectively.

Wang et al. (2022) applied this approach with fuse remote and proximal sensing data and observed prediction accuracy no better than a single source.

c. Granger-Ramanathan Averaging (GRA):

Granger Ramanathan is a regression-based combination method whereby the final prediction is determined as follows:

$$y = b + \sum_{i=1}^{n} a_i y_i$$

where y is the final prediction, n is the number of data sources, and y_i is the output of i^{th} data source (Granger & Ramanathan, 1984; J. Wang et al., 2022). In this method, first, the predictions from training set for each data source are used to fit a Multiple Linear Regression (MLR) model to determine the value of b and a_i's. Next, predictions from the test sets, along with previously determined b and a_i's , are used to calculate the final predictions.

A study by O'Rourke et al. (2016a) showed improvement for almost 22 out of 45 soil properties over the independent sensor when they used this method to fuse VisNIR, MIR, and XRF spectra. Further, they reported that this approach resulted in good predictions for many trace elements in their dataset of unpolluted soils, and the method is suitable for the characterization of a full suite of soil geochemistry. Further, Wang et al. (2022) also found that this approach produced better results from a single source or other combinations. Tavares et al. (2020) also observed that it produced better results than single sensors or low-level fusion approaches. Finally, Xu et al. (2020) also employed Granger Ramanathan averaging and were able to improve the predictions. However, their results showed that predictive accuracy was lower than the OPA approach they applied. These studies have been summarized in Table 5.

2.5 HYBRID FUSION

Researchers may also apply a hybrid data fusion strategy to combine the different fusion levels. Loiseau et al. (2019) combined mid-level fusion with a low-level approach. They extracted features from some of the datasets and combined them with other raw datasets. With this type of fusion, they reported that results were better than a benchmark value. In the study by Li et al.(2021), the authors observed mixed results with an application of hybrid fusion. Better results were observed when combining extracted features of MIR and XRF but not in the case of extracted features of VisNIR and XRF. Taghizadeh-Mehrjardi et al. (2022) also applied hybrid fusion and found better results. While they implemented a low-level fusion to merge various datasets, essentially by concatenating them, they combined the outcomes from multiple modeling algorithms using different model averaging techniques, which is a high-level fusion, for the final prediction. These studies have been summarized in Table 5.

TABLE 5 HIGH-LEVEL FUSION AND HYBRID FUSION

Research	Fusion approach	Property	Data sources	Modeling	Improvement over best individual sensor
(Dongyun Xu et al., 2019)	GRA	Chromium	XRF, VisNIR	PLSR	Yes
(O'Rourke et al., 2016),	GRA	42 soil properties	VisNIR, MIR, XRF	Cubist	Yes, for 22 properties
(J. Wang et al., 2022)	SA, GRA, Bates-Granger	CEC and pH	EM, gamma ray, Sentinel-2, DEM and LiDAR	QRF	Yes, with GRA
(Tavares et al., 2020)	GRA	OM, CEC, pH, base saturation(V), P, K, Ca and Mg	VisNIR, XRF	PLSR, SVR	Yes, 8 of 9

(Taghizadeh-Mehrjardi et al., 2022)	Hybrid	SOM, CCE, Gypsum, Clay, Silt, Sand, EC, pH	36 environmental covariates	kNN, GP, SVR, LASSO, ANN, DNN, RF, ANFIS, XGB	Yes
(F. Li et al., 2021)	Hybrid	As, Cr, Cu, Ni, Zn, Pb, and Cd.	XRF, NIR. MIR	PLS	Yes, except for Pb and Cd
(Loiseau et al., 2019)	Hybrid	Clay content	Sentinel-2, MODIS, and PROBA-V NDVI, Land use map, Global Soil Map	QRF	Yes

2.6 DISCUSSION

Data fusion, which initially emerged in the military domain, has been employed in various fields for some time, including computer vision and robotics. However, its application in soil spectroscopic analysis is relatively recent. During our review study on data fusion in soil science, we found that the studies have been increasing in recent years due to advancement in sensor technology, increase in computational capacity, and development of new algorithms. This review only focused on the hard fusion of the data, not on the soft fusion, which would have involved human-based data expressed preferably in natural language. Moreover, we primarily focused on proximal soil sensing while also reviewing a few studies with remote sensing. Most of the research being carried out in the field of sensor fusion for proximal soil sensing (PSS) mainly focuses on low-level fusion or, specifically, concatenation. The studies started to become scanty while we moved up with the fusion level. Even many studies that employed low-level fusion, i.e., concatenation, do not even mention sensor fusion in the study. Concatenating datasets is the first approach that comes to mind while fusing data from different sources, and researchers do not attempt other approaches. It may be because researchers get satisfactory results with this fusion, and they do not have to explore other approaches. Even when they fail to combine their original data sources constructively, they may add additional data sources to try a different combination of data sources that works better. Again, they do not opt for higher levels of fusion. The next most common approach to data fusion we observed was PC reduction of data sources and fusing them. Researchers have already applied PC reduction in many studies to reduce multicollinearity among variables or reduce noise even with a single data source. Therefore, while some

researchers used this approach to achieve effective synergy among sensors, other researchers used this approach with a mindset of preprocessing rather than data fusion.

It makes complete sense to opt for the best method of data collection, best sensors, best preprocessing, best modeling, or any other best practices while carrying out the study to get better performance. However, it is also recommended that researchers look for the best data fusion practices that may help to achieve better performance in less time, effort, and cost. Further, spectroscopy is a non-destructive method, and we can use multiple sensors to collect data from the same sample once it has been collected in the case of off-site data collection. Further, soil spectroscopies require little sample preparation. Data focused on different studies from the past can be combined to carry out new research if we can find synergy among the sensors. In remote sensing, a plethora of data is already collected from multiple sources, and no extra work is required in terms of data collection. It is left up to the researchers to explore these existing datasets and find the best ways to use them through sensor fusion.

One other interesting observation was regarding the naming of the level of fusion in the studies. While we have categorized the studies based on the nomenclature in the specific studies, they are sometimes inconsistent in labeling it. For example, some studies specified the fusion of spectra after they applied preprocessing, such as variable selection, as a higher-level fusion. However, to maintain consistency, we have still categorized those fusions and all other similar fusions that did not transform the data sources into a latent space different from the original as low-level.

In recent years, astonishing achievements have been made in other scientific fields, including self-driving cars, natural language processing, and virtual assistants, to name a few. These advances can be associated with the application of deep learning in these fields. Nevertheless, we observed very little research involving deep learning in sensor fusion for proximal soil sensing. As deep learning is a data-hungry approach, there are only a few studies on it, even with a single type of sensor, and these studies were mainly possible due to large soil libraries like those of KSSL (National Cooperative Soil Survey, 2013; Wijewardane et al., 2016) or LUCAS(Statistical office of the European Union, 2017) . If only data from other complementary sensors could be added to these already existing large soil libraries, we could explore the possibility of sensor fusion with a deep learning approach. The reason for this discussion is that we observed that modeling algorithms have a direct effect on fusion efficiencies. Some algorithms that seem to work perfectly fine with individual sensors deteriorate with the addition of data. One evident approach would be exploring a better way to fuse data, but we would also encourage exploring the algorithms that would handle these fusions more swiftly. The reasons for the aforementioned poor performance can be an increase in dimensional space, an increase in collinearity among variables, and non-linearity that may have been introduced due to additional data sources. An algorithm that can handle these would surely benefit the fusion approach. Nonlinear modeling algorithms like Support Vector Regression (SVR), RF and Artificial Neural Networks (ANN) are among the algorithms that offer such capabilities. A topic that we were unable to cover again due to the non-availability of studies was cost-benefit analysis. It would be interesting to see the dimension of an

additional cost associated with adding sensors introduced in the studies and analyzed with improved accuracy and robustness of results.

Spectroscopic data fusion is a niche part of sensor data fusion. While some researchers combined spectroscopic data with other data, some combined only data from different spectroscopies. Spectroscopic data are sufficient for predicting many of the soil properties and different spectroscopy work on different parts of the electromagnetic spectrum. Due to this, they can predict some soil properties with high reliability but not others. So, it makes sense to fuse different kinds of sensors, and finding the most optimum fusion combination is essential. While it started with a focus on key visible or near-infrared bands, electromagnetic wavelength in use ranges from ultraviolet to microwave portions of the spectrum today (Mulla, 2013). VisNIR and MIR are the spectroscopies with the most significant number of literatures because of their usefulness in predicting major soil properties. Although MIR has better predictive capability than VisNIR for soil geochemistry, VisNIR is cheaper. So, both are of research interest to investigators for their unique reasons. However, VisNIR shows less reliability in predicting macronutrients (except for N) and other primary micronutrients. The spectroscopy that becomes helpful in determining these elemental concentrations is XRF. Due to this and other reasons, the popularity of XRF sensors, especially portable XRF (pXRF), has been growing the fastest within the soil science community (Mancini et al., 2022).

Further, even if properties are not active on one spectroscopy, the prediction can still be favorable if they are highly correlated to active properties (Ng et al., 2022). Metals can

interact with the spectrally active components of soil (Song et al., 2012), and studies have supported the use of XRF in predicting spectrally active components of soil due to their significant correlations with XRF-sensed elements (Wang et al., 2015). Conversely, metals have no direct spectral response on VisNIR and MIR spectra (Stenberg et al., 2010). However, these can be indirectly predicted due to their correlation with properties like SOM and texture. Further, many soil laboratories now have these three instruments: VisNIR, MIR, and XRF (O'Rourke et al., 2016). Hence, a fusion of these instruments would help predict nutrients to delineate MZs and several studies have already been conducted to assess this synergy. Nevertheless, we found no research that conducted this study with an approach to delineate MZs. Further, they have not compared the different fusion levels, and we did not come across a study that used deep learning as a fusion approach. Lastly, we could not find any consensus among the available research for any optimum approach for fusion of these three spectroscopies.

2.7 CONCLUSION AND FUTURE WORKS

Scientists have put precision agriculture forward as an approach to address two main challenges of agriculture, i.e., increasing demand for food from a growing population and the increasing need for environmental conservation practices today. First, site-specific management is a technique of precision agriculture that requires delineation of MZs. Proper delineation of MZs requires in-field soil data at high spatial (and in some applications temporal) resolution that traditional laboratory analysis cannot afford.

Thanks to scientific advancements, it is now possible to use soil spectroscopy to meet the soil measurement requirements of precision agriculture. A combination of sensors or sensor fusion may be suitable according to specific management practices, cost, and management zone size. After reviewing the studies that combined or intended to combine soil sensors data, further work is recommended on combining VisNIR, MIR, and XRF. While their combination has enormous potential to characterize a full suite of geochemical properties, we recommend studying their combination, focusing on soil nutrients to delineate MZs for nutrient management. We also recommend exploring deep learning techniques for sensor fusion.

CHAPTER 3: RESEARCH METHODOLOGIES

3.1 SOIL SAMPLING SITES

Soil samples are from different projects named ARS, Rogers Farm, Havelock, and NEKS. Three of them, ARS, Rogers Farm, and Havelock, each had one field where samples were collected, and all of these sites were in Nebraska. In the NEKS project, the soil samples were collected from six fields in Kansas. The intention behind including samples from the different projects was to improve the generalization of the results. All the Nebraskan sites are croplands presented in research fields/farms owned by the University of Nebraska-Lincoln. Four of the fields from NEKS are cropland, whereas two of them are pastureland, and all of them are privately owned.

Table 6 shows the summary of the soil samples and brief information about the soil sampling sites. The field in Havelock project was in Havelock Farm (40^{o} 51’ 43’’ N, 96^{o} 36’ 47’’ W)(Wijewardane et al., 2019). The size of the field was 12.5 ha (340×405 m^{2}). The predominant soil types in the area were Crete silty clay loam (56.4%) and Crete silt loam (24.8%). One hundred forty-three surface samples (0-10 cm) were collected from this site bases on an equally spaced grid (40 m × 40 m). The field in the ARS study was in Eastern Nebraska Research and Extension Center (ENREC), UNL (41° 09' 41" N, 096° 27' 49" W). The size of the field was 8.8 ha (440×200m). The major soil types were Yutan silty clay loam (34.7%), Tomek silt loam (28.7%), and Filbert silt loam (32.2%). One hundred thirty-one surface soil samples were collected during this study, and sampling was done based on a grid of (30 m × 30 m). The field in Rogers Farm project is

on Rogers Memorial Farm (40°51'9.31"N, 96°28'7.89"W) of UNL. The major soil type was Aksarben silty clay loam. The samples were collected at two depths, i.e., from 0-10 cm and 10-20 cm, compared to only surface samples in Havelock & ARS sites. In total, 254 samples were collected from 127 locations at two depths. The sampling was conducted according to the existing cover crop x nitrogen fertilization rate treatments, and samples were collected from three locations within each plot (or treatment). The name of the six fields in the NEKS study are KS-02-E04 (37° 47' 03"N,98° 10' 37"W), KS-10-SE01 (37° 47' 32" N, 98° 02' 22" W), KS-12-SE01 (37° 54' 54" N, 98° 06' 38" W), KS-15-SE01 (37° 57' 34" N, 98° 02' 32" W), KS-28-S01 (37° 55' 11" N, 98° 05' 03" W), and KS-29-S02(37° 50' 26"N, 97° 58' 39"W). The sampling was done on a random basis on this site. The major soil types were Crete silty clay loam, Shellabarger-Nalim complex, Shellabarger sandy loam, Nalim loam, Saltcreek-Funmar-Farnum complex, and Albion-Shellabarger sandy loams. In NEKS study, nine cores, each of length 100 cm, were extracted from each of these fields. Each of these cores was further subsampled into ten samples, each of 10 cm in length such that the first sample corresponds to a depth of 0-10 cm, the second sample corresponds to a depth of 10-20 cm, and so on. Nevertheless, the total number of samples taken from these sites in this study is only 189.

TABLE 6 SUMMARY OF SOIL SAMPLING SITES, *THE TOTAL NUMBER OF SAMPLES INCLUDED IN THIS STUDY IS LESS THAN THE TOTAL NUMBER OF SAMPLES COLLECTED IN NEKS SITES

Projects/Sites	Location	Site Type	Soil Type/Slope	Sampling depth (cm)	No. of samples
Rogers Farm (NE)	40° 51' 16" N, 96° 28' 06" W	Cropland	Aksarben silty clay loam, Slopes: 2-6 % (72.5%) and 6-11% (27.5%)	0-10 & 10-20	254
ARS (NE)	41° 09' 41" N, 96° 27' 49" W	Cropland	Yutan silty clay loam, terrace, 2-6% slopes, eroded (34.7%), Tomek silt loam, 0-2% slopes (28.7%), and Filbert silt loam, 0-1% slopes (30.2%)	0-10	131
Havelock (NE)	40°51' 43" N, 96° 36' 47" W	Cropland	Crete silty clay loam, 1-3% slopes (58.6%) and Crete silt loam, 0-1% slopes (24.8%)	0-10	143

NEKS (KS)	37° 47' 03"N, 98° 10' 37"W	Cropland	Shellabarger-Nalim complex, 1-3% slopes	0-100 (increasing depth of 10)	189*
	37° 47' 32" N, 98° 02' 22" W	Cropland	Shellabarger-Nalim complex, 1-3% slopes		
	37° 54' 54" N, 98° 06' 38" W	Cropland	Shellabarger sandy loam, 1-3% slopes		
	37° 57' 34" N, 98° 02' 32" W	Cropland	Nalim loam, 0-1% slopes		
	37° 55' 11" N, 98° 05' 03" W	Pasture	Saltcreek-Funmar-Farnum complex, 1-3% slopes (79.2%) and Nalim loam, 0-1% slopes (20.8%)		
	37° 50' 26"N, 97° 58' 39"W	Pasture	Albion-Shellabarger sandy loams, 1-3% slopes		

3.2 DATASET

Laboratory data for soil samples were obtained from wet chemistry analysis at Ward Lab (Ward Laboratories Inc, Kearney, NE, USA). The samples were analyzed for phosphorus using the Mehlich III method, for nitrogen (NO3-N) using the LACHAT method, and for Mg, K, and Ca using the ammonium acetate extraction method. Figure 2 shows summary statistics and distribution of three primary nutrients: N, P, and K, and two secondary nutrients: Ca and Mg. None of the nutrients in the sample were normally distributed.

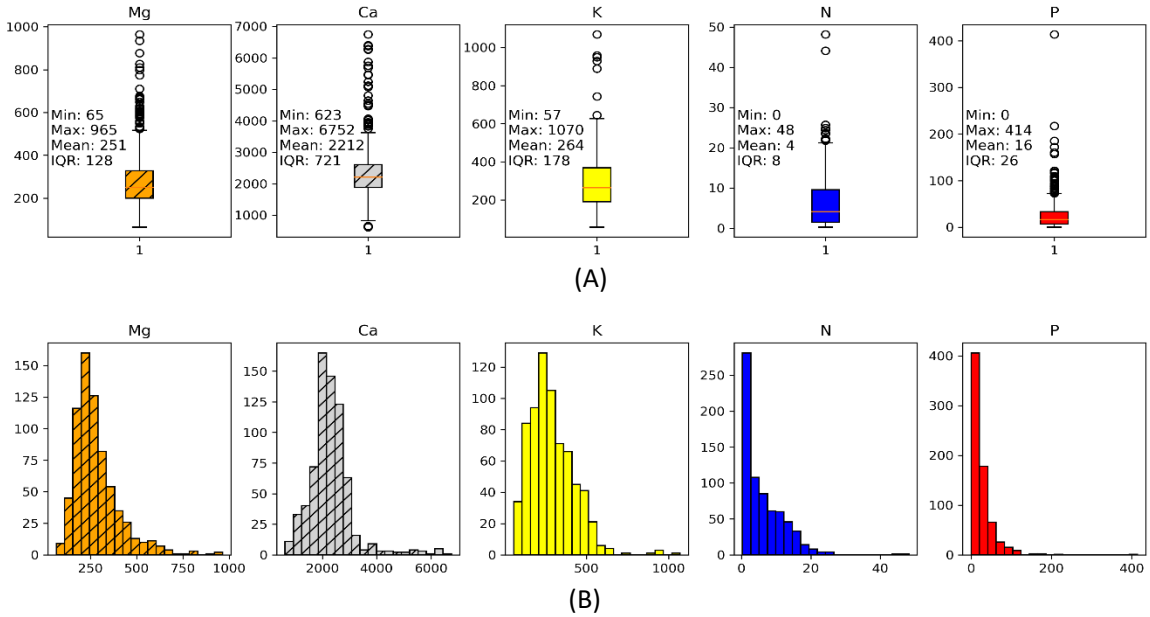

FIGURE 2 BOX PLOT WITH SUMMARY STATISTICS AND DISTRIBUTION OF THE FIVE NUTRIENTS

For VisNIR spectra, soil samples were scanned using the SR-3500 Spectrometer (Spectral Evolution, Haverhill, Massachusetts, USA) which has a spectral range of 350 nm to 2500 nm and samples at 1 nm interval. Figure 3 illustrates the different steps involved in the VisNIR spectral data acquisition process. Air-dried and ground (<2mm)

samples were placed in a sample holder (Figure 3C) with a clear fused silica window at the bottom and scanned using Spectral Evolution's Mug-light Accessory (Figure 3B).

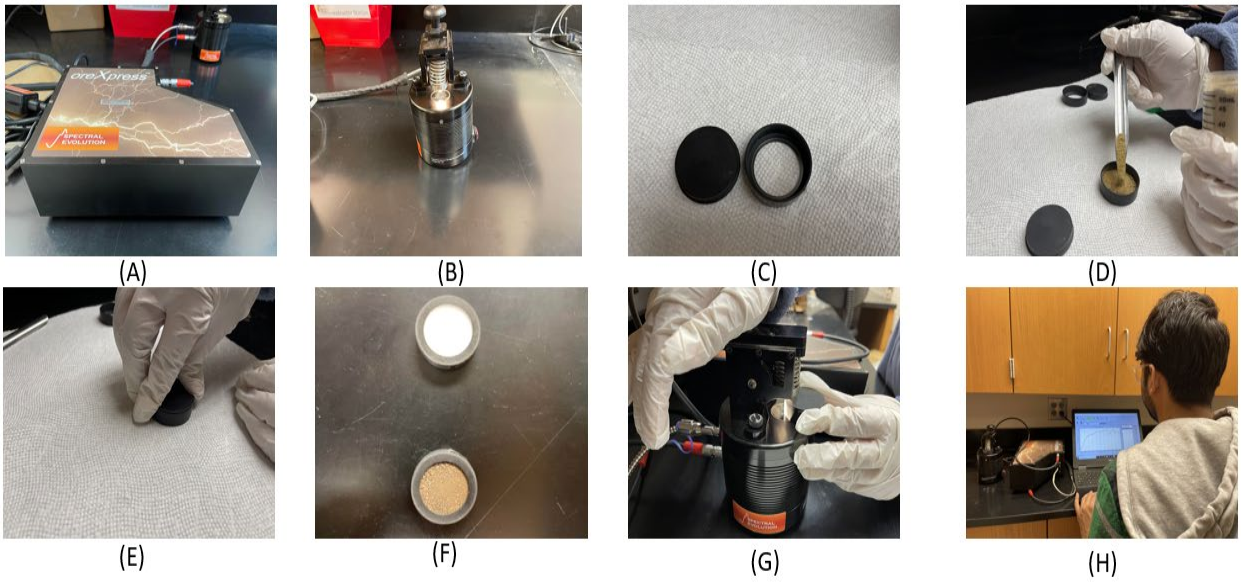

FIGURE 3 VISNIR SPECTRAL DATA ACQUISITION PROCESS - (A) SR-3500 SPECTROMETER, (B) MUG-LIGHT ACCESSORY, (C) EMPTY SAMPLE HOLDER, (D) POURING SAMPLE INTO SAMPLE HOLDER, (E) PLACING SAMPLE HOLDER CAP, (F) TOP: WHITE REFERENCE, BOTTOM: SAMPLE HOLDER FILLED WITH SOIL

Each spectrum consisted of 50 co-added instantaneous internal scans. A standard Spectralon panel (99% White, Figure 3F TOP) served as a white reference to convert radiometric digital numbers to reflectance. Spectral Evolution's DARWin™ SP Spectral Data Acquisition software was employed for spectral data acquisition. Initially, all spectral data were stored in a format used by the software, which was later converted into CSV format using the Spectrolab library in R studio (R Core Team, 2021) for further analysis.

MIR spectra data acquisition involved further processing of soil samples before they were used for the scanning. The air-dried and ground samples (<2mm) were further grinded into the fine ground (<200 µm) samples using Retsch MM200 Mill (Retsch GmbH,

Germany). Figure 4 shows the different steps involved in the grinding process. Initially, the samples that were air-dried and ground to a size of less than 2mm were loaded into a 25 mL milling cup along with five steel balls and covered with a cap (Figure 4A & 4B). Two sets of these cups were then placed between two rubber supports of a U-shaped bracket that held the milling cup closed, and a black wheel was rotated to clamp the filling cup (Figure 4C) firmly. The silver locking nut was rotated until it was flushed with the U-shaped bracket, and a sickle wrench was used to tighten it against the bracket (Figure 4D). The mill was then operated for 15 minutes at 25 Hz, and the resulting samples were stored in glass vials.

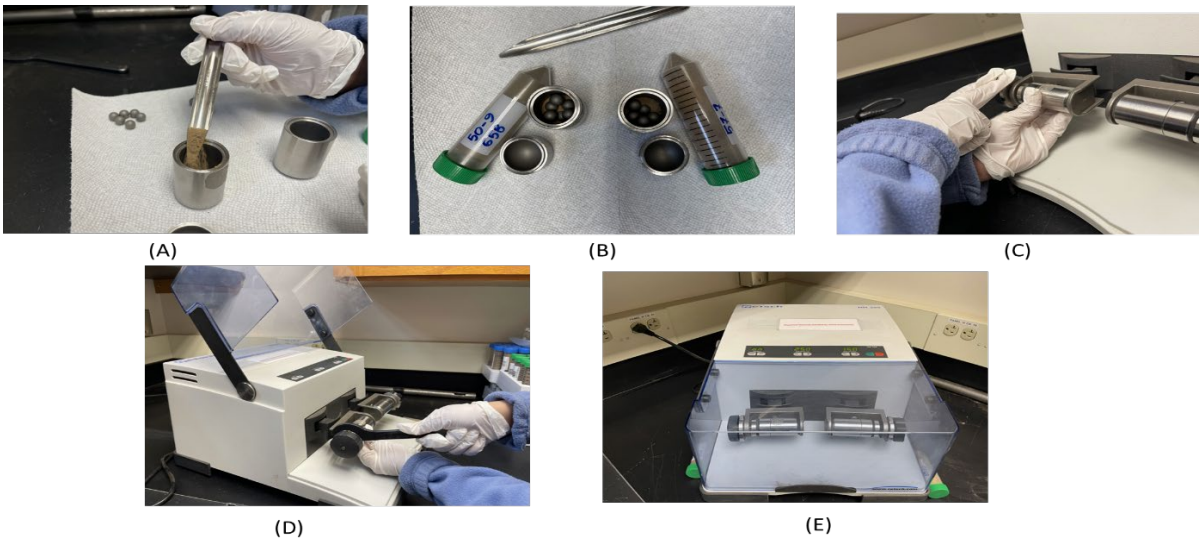

FIGURE 4 SAMPLE GRINDING PROCESS (A) POURING SAMPLE INTO MILLING CUP (B) MILLING CUPS FILLED WITH SOIL SAMPLE AND STEEL BALLS (C) PLACING MILLING CUPS (D) TIGHTENING SILVER LOCKING NUT WITH A SICKLE WRENCH (E) RETSCH MM200 MILL LOADED WITH MILLING CUPS

For MIR spectra, scanning of samples was performed using a Bruker Alpha II Spectrometer (Bruker Optics, Billerica, MA, USA). Samples from Rogers Memorial Farm were scanned using a Front-reflectance module, while all other samples were

scanned using a DRIFT (diffuse reflectance infrared Fourier Transform) module. The sample loading process differed slightly for these two accessories. Figure 5 shows the sample loading process for the Front-reflectance module, whereas Figure 6 shows the sample loading process for the DRIFT module. For Front reflectance, the sample holder was placed on an aluminum plate with silver magnets facing upwards. A scoop of the sample was then placed in the hole of the sample holder (Figure 5B) and pressed with a cylindrical metal press rod to compress the soil mound flat (Figure 5C). Loose soil was then removed from the interior of the sample holder by gently flipping and tapping it. The sample holder was then inserted into the scanning device (Figure 5D). For the DRIFT module, a small stainless steel sample holder was placed inside another plastic holder ((Figure 6A). A scoop of soil sample was then poured into the sample holder and leveled to flush with the sample holder (Figure 6B). After that, the soil was pressed gently with a stainless-steel rod with a flat end (Figure 6C) and inserted in the DRIFT module for scanning (Figure 6D). Each sample was measured two times to obtain spectra, with 128 co-added scans at a spectral resolution of 4 cm^{-1}. Prior to analyzing each new sample, background spectra were collected on a gold mirror. Initially, all spectrums were saved in the format used by the OPUS software of the FT-IR spectrometer. Afterward, they were transformed into CSV format using Spectragryph (F. Menges "Spectragryph - optical spectroscopy software", Version 1.2.16, 2022) for the purpose of analysis and processing.

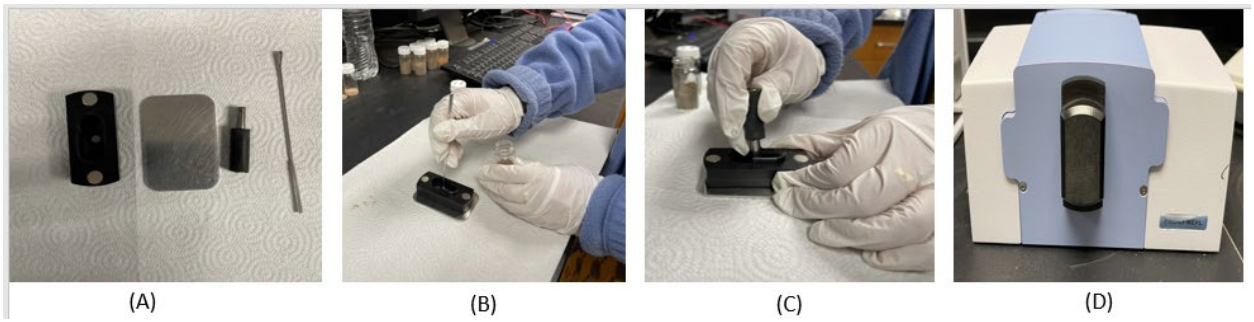

FIGURE 5 SAMPLE PREPARATION AND LOADING: FRONT REFLECTANCE MODULE (A) LEFT TO RIGHT: SAMPLE HOLDER, ALUMINUM PLATE, CYLINDRICAL METAL PRESS ROD, A STEEL SCOOP (B) POURING SOIL SAMPLE (C) COMPRESSING SOIL MOUND FLAT (D) FRONT REFLECTANCE MODULE LOADED WITH SAMPLE

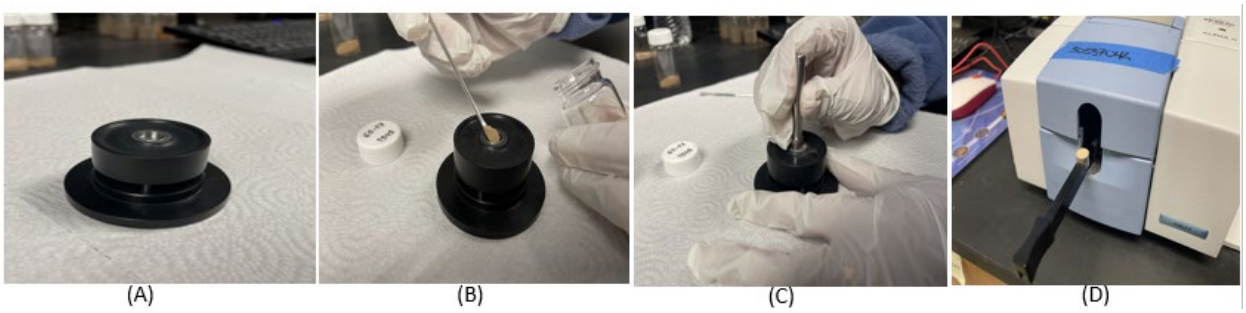

FIGURE 6 SAMPLE PREPARATION AND LOADING: DRIFT MODULE (A) A STAINLESS-STEEL SAMPLE HOLDER INSERTED INTO A PLASTIC HOLDER (B) POURING SOIL SAMPLE (C) PRESSING SOIL SAMPLE (D) SAMPLE HOLDER INSERTED IN THE DRIFT MODULE

XRF spectra were collected using Niton™ XL5 Handheld XRF Analyzer (Thermo Fisher Scientific, Tewksbury, Massachusetts, USA). The instrument was operated inside a Portable Test Stand (Thermo Fisher Scientific, Tewksbury, Massachusetts, USA) to safeguard against any scattered radiation. The Niton XL5 XRF Analyzer utilizes a silver (Ag) anode X-ray tube with a voltage range of 6-50 kV, a maximum power output of 5W, and a dynamically adjustable current of up to 500 µA. Ag anode tubes are effective in exciting elements with lower atomic numbers (Goodale et al., 2012; Silva et al., 2021) due to the strong L lines that are produced in addition to the K lines. Further, the Niton

XL5 XRF Analyzer is equipped with four filters - Main (Al @ 40kV), High (Mo @ 50kV), Low (Cu @ 20 kV), and Light (No Filter @ 6.2 kV) – that are designed to be used with a specific range of X-ray tube voltages. These filters can modify the X-ray beam produced, enabling the analyzer to optimize the energy range of the X-rays produced for different elements and improve the accuracy, sensitivity, and precision of the analysis. Like VisNIR spectra data acquisition, there was no further processing of samples done, and air-dried and ground soil (<2mm) was used. The samples were packed in double open-ended sample cups of size 32 × 24 mm (SC4331, Premier Lab Supply Inc., St. Lucie, FL, USA) and were wrapped by Propylene X-ray Film (TF240255, Premier Lab Supply Inc., St. Lucie, FL, USA) of thickness 4 μm. Prior to filling the cup with soil, a cotton ball was placed in the cup to keep the soil securely in place. This is the most commonly reported way to analyze soil samples using a portable XRF instrument (Ravansari et al., 2020; Silva et al., 2021; Weindorf et al., 2014, 2018). Figure 7 demonstrates the steps involved in preparing the sample to load into an XRF analyzer. Each sample was scanned for 180 s in the mining mode. The XRF analyzers have factory-installed calibration for element detection (ORourke et al., 2016). So, the output consists of a geochemical profile of 42 elemental concentrations and four raw spectra corresponding to the four filters, which were accessed and converted to CSV file using Niton™ Data Transfer (Thermo Fisher Scientific, Tewksbury, Massachusetts, USA). Figure 8 shows sample VisNIR, MIR and XRF(Main) spectra.

FIGURE 7 SAMPLE PREPARATION FOR XRF SAMPLE A. POURING SAMPLE INTO PLASTIC CUPS, B. SEALING WITH POLYPROPYLENE FILM, C. AFFIXING THE CAP

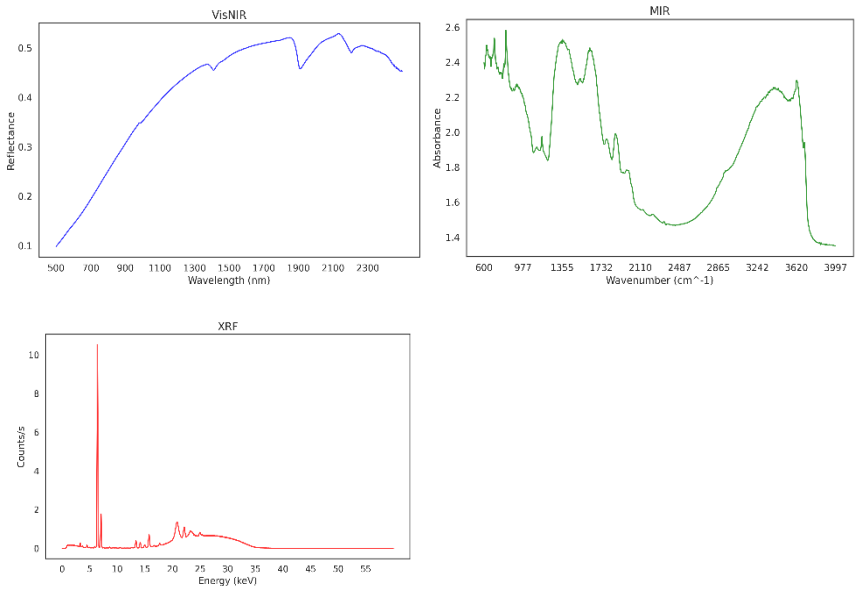

FIGURE 8 SAMPLE VISNIR, MIR AND XRF SPECTRA

3.3 PREPROCESSING

Raw spectra were processed to detect outliers using the MT library (Lin, 2022) in R studio (R Core Team, 2021) based on Principal Components (PC) reduction. On separate analyses of XRF, VisNIR, and MIR data, 16 outliers were found in XRF Data, 13 in VisNIR data, and no outlier in MIR data. Outliers were detected using the Mahalanobis distance of PC1 and PC2 (Figure 9 and Figure 10). For VisNIR and MIR, the cumulative

variance explained by PC1 and PC2 was more than 95%, while 12 PC components were required to explain 95% of the variance for XRF data. Therefore, outlier detection for XRF was conducted using the Mahalanobis distance of the first 12 PC components. Additionally, points lying on the border of the confidence ellipse were not removed. Visual inspection of plots of raw spectra was also done to aid the outlier analysis. These 29 outliers were removed before further processing. After outlier removal, all the other processing was done using Python programming language. For VisNIR spectra, the spectra ranging from 350 to 499 nm were omitted from the spectra as they have a low signal-to-noise ratio (Wijewardane, Ge, & Morgan, 2016). After that, the VisNIR data was converted into absorbance (A = log (1/R)). Both MIR and VisNIR spectra were preprocessed with an averaging window of 10 bands to enhance the signal-to-noise ratio of the spectra, reduce the dimensionality of data for effective computation, and avoid model overfitting (Wijewardane et al., 2018). No preprocessing was performed on either spectral data or elemental concentration data from XRF. There were 202 and 167 predictor variables in preprocessed spectra of VisNIR and MIR respectively. Similarly, there were 2919,1449,3561,454 and 42 predictor variables in Main, Low, High, Light, and elemental concentration data of XRF respectively.

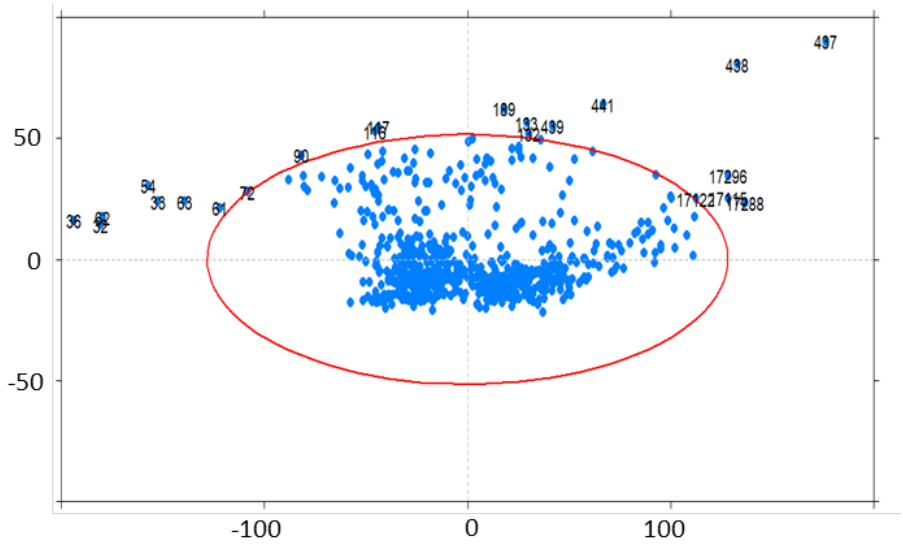

FIGURE 9 PC SCORE PLOT OF VISNIR SPECTRA

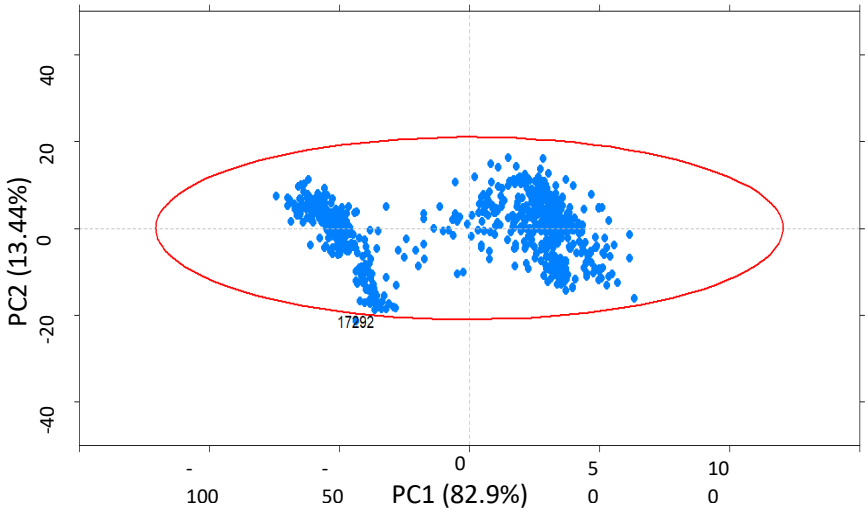

FIGURE 10 PC SCORE PLOT OF MIR SPECTRA

3.4 MODELING

PLSR was used for chemometric modeling as it is the most commonly used and a *de facto* standard method in soil spectroscopy (Wijewardane et al., 2018). PLSR, like PC

analysis, decreases the number of predictor variables to a number of synthetic variables called "latent variables", while taking the response variable into account (Helland, 2005). Afterward, a linear model is established between the latent variables and the response variable. PLSR modeling is less resource-intensive and easier to interpret. An optimization function was defined for the purpose of parameter tuning as different models were built based on the varying number of latent variables, from 1 to 30. Then, the model with the number of latent variables that gave the first minimum of root mean squared error of cross-validation (RMSECV) was selected using 10-fold cross-validation. The data were randomly split into training and testing data in the split 70:30. For each type of fusion, 100 models were trained, and all these 100 models were trained and tested based on 100 separate random splits. Model performances were evaluated by calculating R^2 (coefficient of determination between predicted value and ground truth), Root Mean Square Error of Prediction ($RMSE_p$), ratio of performance to inter-quartile range (RPIQ), and bias. RPIQ was used instead of RPD because none of the properties of interest were normally distributed (Wijewardane, Ge, & Morgan, 2016).

$$\boldsymbol{RMSE_p} = \sqrt{\frac{\mathbf{1}}{\boldsymbol{n}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}}(\hat{\boldsymbol{y}} - \boldsymbol{y})^{\mathbf{2}}} \quad \textbf{(1)}$$

$$RPIQ = \frac{IQR}{RMSE} \quad \textbf{(2)}$$

$$bias = \bar{y} - \frac{1}{n}\sum_{i=1}^{n} \hat{y} \quad (3)$$

where $\hat{y}$ is the vector of the predictions, y is the vector of the ground truth, n is the total number of samples, and IQR is the interquartile range of the ground truth data (Terra et al., 2019).

3.5 METHODS AND LEVELS OF FUSION

Five different fusion algorithms were assessed to determine their ability to enhance predictive performance. Before undergoing fusion, data from the three instruments were utilized separately in constructing models. To streamline the fusion process, only one data type was selected from the five available within XRF. A review of several XRF spectroscopy studies on soil revealed that Elemental Concentration Data (ECD) is predominantly employed for modeling purposes. Consequently, models were built for each element with all XRF data types separately, and the outcomes were compared. For Phosphorus, Calcium, and Inorganic Nitrate, the results with ECD were markedly superior to those obtained from other XRF data types. In the case of potassium, the performance of ECD & Main spectra surpassed that of the others. However, none of the data types yielded significantly better results than others for Magnesium. Therefore, ECD was employed in the fusion algorithms alongside VisNIR and MIR. Fusion of data was performed at three levels: namely, Low-level, Mid-level, and High-level.

3.5.1 LOW-LEVEL FUSION

Concatenation of data is one of the most common approaches to low-level fusion. This fusion approach showed improvement in several papers that were reviewed. It increases the number of dimensions in data while keeping the observations unchanged. Data were standardized before they were concatenated together to make them compatible. Standardization was performed by subtracting the mean from data and then dividing by standard deviation. After that, VisNIR, MIR, and XRF were concatenated, and this whole concatenated data was used for model development and evaluation.

3.5.2 MID-LEVEL FUSION

This type of fusion is also referred to as feature-level fusion, which involves extracting salient features independently from each source and fusing similar elements to build a model. These features are thought to be free from background noises and improve the model's performance. The two most common approaches of feature extraction were employed to extract features from the data: PC reduction and PLS reduction. PC reduction is an unsupervised approach to feature reduction, whereas PLS Reduction is a supervised approach to feature extraction. Only training data was used to fit both the PC reduction algorithm and PLS reduction algorithm, and the fitted models were used to transform both training and test data into latent space and extract latent variables. These latent variables were then fused together. For PC reduction fusion, first 40 PC components were extracted from each of the data type to fuse. For PLSR, distinct number of latent variables were chosen for each data type depending on the PLS model fitted on

training data for each data type. Various models were constructed using 1 to 30 latent variables. The model with the lowest RMSECV, determined through 10-fold cross-validation, was selected. Table 7 shows the latent variables that were extracted from PLS reduction for fusion.

TABLE 7 NUMBER OF LATENT VARIABLES SELECTED FOR PLS REDUCTION FUSION

	No. of latent variables		
Nutrient	VisNIR	MIR	XRF
Magnesium	19-20	11-20	2-9
Calcium	18-20	11-20	7-20
Nitrogen	14-19	7-9	4-6
Phosphorous	16-19	6	7-10
Potassium	13-15	16-19	3-5

3.5.3 HIGH-LEVEL FUSION:

This type of fusion is also called decision-level fusion. The outputs of different models are fused together to improve the prediction. The two most common approaches of high-level fusion, Simple Averaging (SA) and Granger Ramanathan Averaging (GRA), were implemented. These two approaches have shown improvements in predictive performance in other studies. In Simple Averaging, predictions from all models are given equal weights, whereas, in Granger-Ramanathan averaging, Ordinary Least Squares (OLS) estimation is used to derive the optimal weight for each model.

3.6 MEAN SEPARATION TEST

A mean separation test was conducted to identify differences in the performance of various fusion algorithms. The 100 iterations performed during the evaluation were

treated as repetitions, and four performance metrics— R^2, RMSE, RPIQ, and bias—were assessed through this test. To further analyze the results and compare the performance of the algorithms, multiple comparison tests were conducted using Tukey's Honest Significant Difference (HSD) test.

CHAPTER 4: RESULTS AND DISCUSSION

Distinct data sources exhibited superior performance for each of the five elements when models were constructed using these sources. Figure 11 reveals that models utilizing VisNIR data yielded higher R^2 values for Nitrogen and Magnesium. Meanwhile, models employing MIR data demonstrated improved R^2 values for Potassium and Calcium, and those incorporating Elemental Concentration data from XRF showed better R^2 values for Phosphorous. The same pattern was observed for other performance metrics: RMSE, RPIQ, and bias.

The mean separation test provided a more detailed perspective when compared to the box plots, as it did not entirely agree with the latter due to the differences in statistical significance (see Appendix for Mean Separation Test results). Although the mean R^2 values of models using VisNIR data were higher than those employing MIR data for Magnesium, there was not enough evidence to conclude their significant superiority (Table A1). Nevertheless, VisNIR-based models exhibited significantly better R^2 values compared to MIR and XRF data models for Nitrogen. For Potassium and Calcium, models with MIR data displayed significantly better R^2 values than those with VisNIR and XRF data. Models incorporating XRF data demonstrated significantly better R^2 values compared to MIR and VisNIR data models for Phosphorous.

RMSE and RPIQ followed a similar pattern, apart from Nitrogen and Potassium (Table A2 and Table A3). For Nitrogen, although these performance metrics were higher for

VisNIR models than MIR models, the difference was not statistically significant. Similarly, for Potassium, these performance metrics were better for MIR models compared to the other two, but the difference was not significant when compared to VisNIR models. Regarding the bias performance metrics, the MIR models exhibited superior performance for all elements, apart from phosphorus, where the XRF model outperformed. However, none of the models demonstrated statistically significant superiority.

4.1 LOW-LEVEL FUSION:

Low-level fusion did not enhance the prediction performance compared to individual models, except for some modest improvements observed in specific cases. The box plot (Figure 11) illustrates that Low-Level Fusion improved the prediction of Magnesium and Potassium in terms of R^2, but only Magnesium concerning RMSE, RPIQ, and bias. However, there was insufficient evidence to confirm that these enhancements were statistically significant (See Appendix for mean separation test results).

4.2 MID-LEVEL FUSION:

Mid-level fusion showed improvements compared to the individual models. Mid-level fusion improved the R^2 by 0.1-18%, RMSE by 0-8%, and RPIQ by 0-11.5%, depending on the nutrients and fusion algorithm (Table A1, Table A2 and Table A3). The box plot analysis reveals that the performance of PLSR fusion methods is generally superior to that of PC-Fused models regarding R2, RMSE, RPIQ, and bias metrics for all nutrients, except for Potassium, where the latter performs better (Figure 11). According to the mean

separation test, PLSR fusion models demonstrate significantly enhanced performance compared to PC Fusion models for Magnesium and Nitrogen (See Appendix for mean separation test results). However, the analysis fails to confirm the statistically significant superiority of PLSR fusion for Calcium and Phosphorous, as well as the superiority of PC fusion models for Calcium. This observation remains consistent across all performance metrics. Further, with respect to bias, there was no significant difference between the fusion algorithms.

Overall, when comparing all models, Mid-level fusion shows better predictive performance for all nutrients. PC Fusion ranks first, followed by PLSR Fusion as the second-best approach for Potassium, while PLSR achieves the best performance for the remaining four nutrients in terms of R^2 and bias. This trend is also observed in terms of RMSE and RPIQ, except for Phosphorous, where XRF models outperform others in these metrics. Mean separation test provides evidence for PC Fusion's improved performance for Potassium (18%), while PLSR fusion's performance is only significant for two of the four nutrients, Magnesium (8%) and Calcium (3.3%), in terms of R^2 compared to individual models (see Appendix for mean separation test results). RPIQ (8% improvement for K, 11.5% for Mg, and 14 % for Ca) follows the same trend as R^2, but for RMSE, only PC Fusion improved performance for Potassium (5% improvement) compared to individual models.

4.3 HIGH-LEVEL FUSION:

High-level fusion improved the R^2 by 0-12.5%, RMSE by 0-12%, and RPIQ by 2.3-13.4% depending on nutrients and fusion algorithm (Table A1, Table A2 and Table A3). In high-level fusion comparisons, Simple Averaging (SA) outperformed Granger Ramanathan Fusion (GR) for Inorganic Nitrate and Potassium based on R^2, RMSE, and RPIQ metrics. However, GRA was superior for Magnesium and demonstrated better results for Phosphorus in R^2 and RMSE, while SA excelled in RPIQ. GR also surpassed SA for Calcium in RMSE and RPIQ and for Phosphorus in R^2 and RMSE. The Mean separation test revealed that only SA significantly exceeded GR for Potassium in terms of R^2 and RMSE.

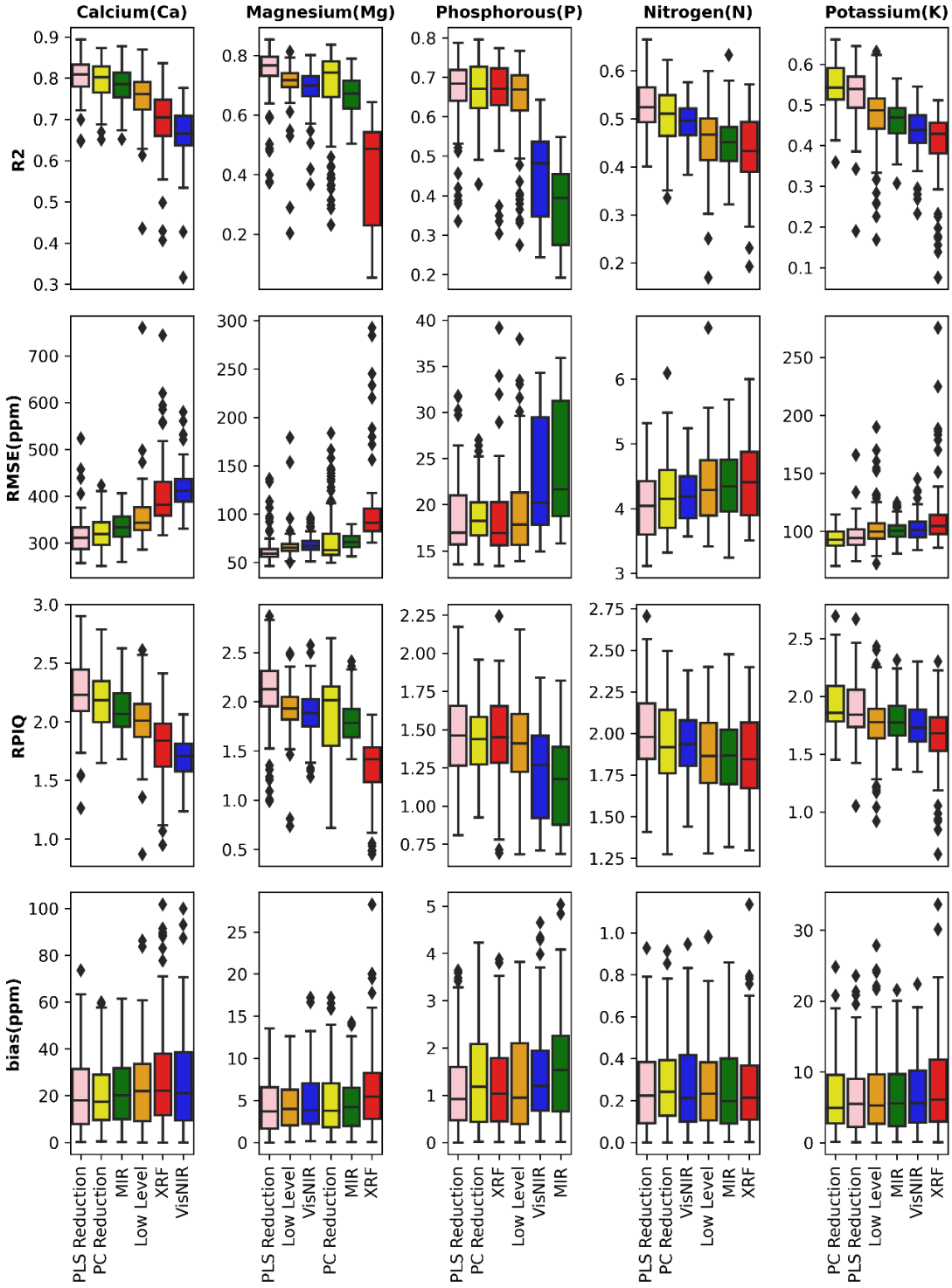

FIGURE 11 BOX PLOT COMPARING PERFORMANCE METRICS OF MID-LEVEL FUSION, LOW-LEVEL FUSION, AND INDIVIDUAL MODELS

4.4 OVERALL COMPARISON:

The comparison of all fusion algorithms was divided into two sections based on the nutrients being evaluated. One group included the primary nutrients Nitrogen (N), Phosphorus (P), and Potassium (K), while the other group included the secondary nutrients Calcium (Ca) and Magnesium (Mg). Figure 12 shows the comparison of performance metrics of all models for Ca and Mg. For Calcium and Magnesium, the top three models with the best performance in terms of R^2, RMSE, and RPIQ were two high-level fusion models and the PLS Reduction Fusion. The same was not true for bias, though, as VisNIR had the lowest bias for Ca, and MIR had the lowest bias for Mg. GRA was found to be the best model for both Magnesium and Calcium in terms of RPIQ and RMSE and for Magnesium in terms of R^2. However, for Calcium, SA was found to be the best model. It is worth noting that the R^2 between GRA and SA for Calcium was almost the same. Additionally, through the mean separation test, GRA and SA were found to be significantly better than Low-Level Fusion or Individual models in terms of R^2 and RPIQ. However, in terms of RMSE, there was not enough evidence to support this claim compared to VisNIR models. GRA improved the predictions of Mg by 10% and Ca by 4.35%, while SA enhanced the prediction of Mg by 6.7% and Ca by 4.6% in terms of R^2 compared to individual models. There was no significant difference between models in terms of bias.

Different fusion algorithms had a higher predictive performance for the primary nutrients. Figure 13 shows the comparison of performance metrics of all models for N, P, and K. SA demonstrated the best performance across all performance metrics for N, and

PC reduction fusion proved to be the best-performing algorithm for K. Regarding Phosphorus, PLS reduction exhibited best performance in terms of R^2, while SA

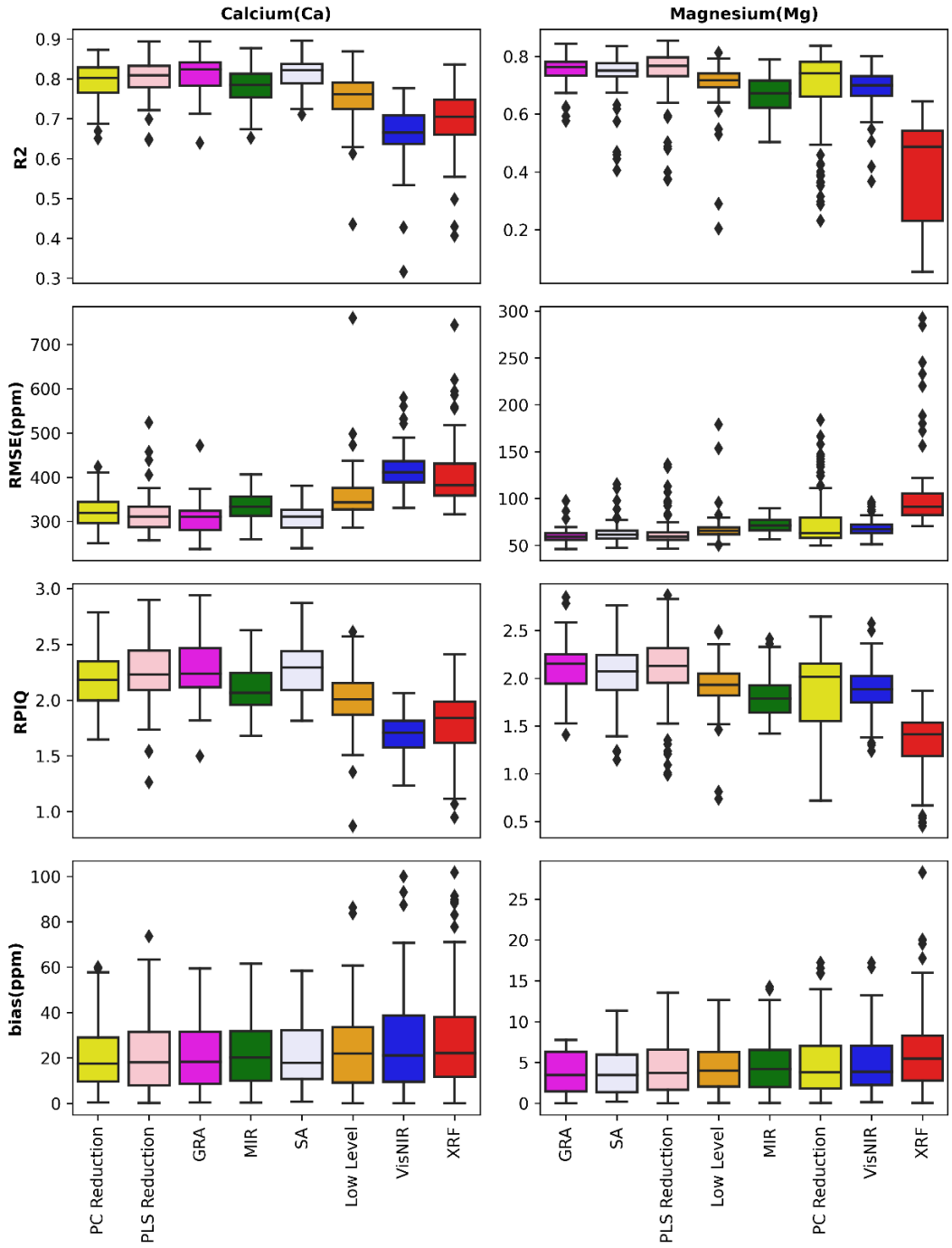

FIGURE 12 COMPARISON OF PERFORMANCE METRICS OF MODELS FOR CALCIUM AND MAGNESIUM

outperformed other models in terms of RPIQ. Bias was lowest for GRA, and RMSE was lowest for XRF among all models. Improvement in R^2 (17.5%), RMSE (7%), and RPIQ (8%) by PC Reduction Fusion for Potassium was established through mean separation test. Similarly, the Mean Separation Test showed significant improvement for R^2 (10%) and RPIQ (8%) for Nitrogen with SA. None of the fusion algorithms could significantly improve any of the performance metrics for Phosphorous.

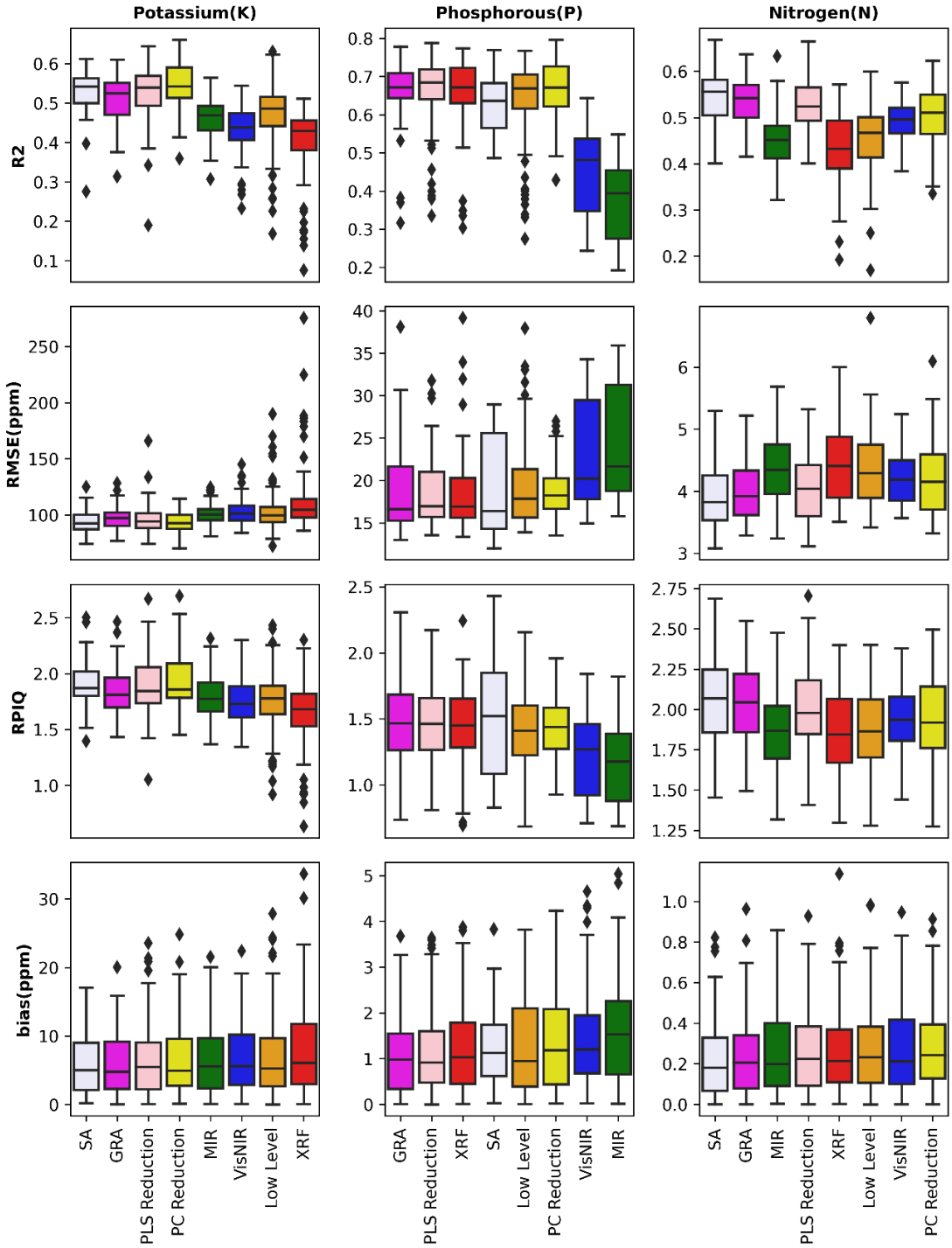

FIGURE 13 COMPARISON OF PERFORMANCE METRICS OF THE MODEL FOR PRIMARY NUTRIENTS

4.5 DISCUSSION:

The results of this study shed light on the effectiveness of data fusion algorithms in enhancing predictive performance. MIR exhibited good performance in predicting Ca, while VisNIR demonstrated good performance for Mg. Although metals do not have a direct spectral response in VisNIR and MIR spectra, they can be indirectly predicted due to their correlation with other spectrally active components. This could explain the accurate predictions of metals such as Mg and Ca. Among all the nutrients tested in this research, the better predictive performance for Ca may also be attributed to its higher concentration in the soil samples compared to others. It has been noted in the literature that K is one of the more challenging secondary properties to measure accurately, which was observed in this study as well. Interestingly, the predictive performance for Nitrogen was better with VisNIR compared to MIR. Among these nutrients, XRF performed best for Calcium, which could be attributed to its highest atomic number (Z) value among these nutrients. However, this trend did not hold for other nutrients. Therefore, the response from XRF to these nutrients can be considered a combination of direct interactions and indirect relationships. The moderate and poor performance of XRF in predicting these nutrients can be attributed to spectral overlap and reduced fluorescence emission from low atomic number (Z) elements at low concentrations.

Table 9 presents a comprehensive comparison of all the models based on means of performance metrics with a focus on whether the difference is significant or not. Low-Level fusion could not improve the predictions for any of the nutrients. This inability to improve predictions can be attributed to the substantial differences in the electromagnetic

spectrum ranges of VisNIR and MIR compared to XRF wavelengths, which may not effectively complement each other when concatenated directly. Mid-level fusion generally improved the predictive performance for all nutrients, raising all R^2 values above 0.5. Transforming the data into a latent space reduced noise, captured relevant features, and simplified data representation. This process can be viewed as a means to increase compatibility between two data sources, which, in turn, led to improvement. Furthermore, the superior performance of PLSR over PCA can be attributed to PLSR taking advantage of the correlation between spectra and soil properties to produce latent factors directly related to the soil attributes. PLSR reduction fusion is consistently better for all nutrients compared to other algorithms. In fact, PLSR reduction fusion is the only algorithm that never had R^2 lower than individual models for any of the nutrients. It can be observed that none of the fusion algorithms had any significant impact on the bias. This could be due to the different datasets having different systemic biases, which might coexist or be compounded in the fused models. Further, the fusion algorithm focuses on reducing noise and capturing relevant features, improving other metrics but not addressing the systematic errors in the data. Further, none of the algorithms could improve predictions significantly for P. This could be due to the lack of synergy among spectra concerning P. In general, SA, GRA, and PLS Reduction Fusion are the three algorithms that improved the model predictions for all nutrients. Despite the improvement in prediction by fusion algorithms, the best obtained prediction results for primary nutrients are satisfactory only. While five different algorithms were implemented for fusion, only a single base algorithm, PLSR, was used for modeling. PLSR can handle multicollinearity as a linear method. However, the process of fusing data could have

potentially introduced further non-linearity. In such scenarios, non-linear methods like SVR and RF could be beneficial, as they can provide better predictive performance where linear methods fall short.

Moreover, sensor fusion can also be employed in the case of field-based applications for analysis of field moist soils. For field applications, moisture is the most important factor that influences the soil reflectance spectra (Ge et al., 2014; Wijewardane et al., 2016). Since MIR and VisNIR spectra exhibit distinct reactions to changes in field conditions, such as soil moisture, they may provide complementary information. Combining these spectra can potentially help mitigate the impact of soil moisture content variations. Often, field applications involves combination of data from not only spectrometers but other types of sensors too. (Wijewardane et al., 2020) developed a VisNIR multi-sensing penetrometer for automated vertical soil sensing which incorporated data from various sensor types, including spectrometers, load cells, and ultrasonic depth sensors. Further, integrating remote sensing data, like that from Landsat, can boost the ability of soil scientists to map soil characteristics across vast terrains with increased precision, leading to better soil management practices (Ge et al., 2011).

TABLE 8 COMPARISON OF MEAN PERFORMANCE METRICS OF FUSION ALGORITHMS TO INDIVIDUAL MODELS. THE TICK MARK DENOTES A SIGNIFICANT HIGHER MEAN THAN BEST SINGLE SENSOR MODEL

R2

A	Mg	Ca	N	P	K
Low Level	✗	✗	✗	✗	✗
PC Reduction	✗	✗	✗	✗	✗
PLS Reduction	✓	✓	✓	✗	✓
SA	✓	✓	✓	✗	✓
GRA	✓	✓	✓	✗	✓

RMSE

B	Mg	Ca	N	P	K
Low Level	✗	✗	✗	✗	✗
PC Reduction	✗	✗	✗	✗	✓
PLS Reduction	✗	✓	✗	✗	✗
SA	✗	✓	✓	✗	✓
GRA	✗	✓	✗	✗	✗

RPIQ

C	Mg	Ca	N	P	K
Low Level	✗	✗	✗	✗	✗
PC Reduction	✗	✗	✗	✗	✓
PLS Reduction	✓	✓	✗	✗	✓
SA	✓	✓	✓	✗	✓
GRA	✓	✓	✓	✗	✗

bias

D	Mg	Ca	N	P	K
Low Level	✗	✗	✗	✗	✗
PC Reduction	✗	✗	✗	✗	✗
PLS Reduction	✗	✗	✗	✗	✗
SA	✗	✗	✗	✗	✗
GRA	✗	✗	✗	✗	✗

CHAPTER 5: CONCLUSION

An in-depth study was conducted to compare and evaluate the performance of five distinct fusion algorithms in predicting three primary nutrients—Nitrogen (N), Phosphorous (P), and Potassium (K)—and two secondary nutrients—Calcium (Ca) and Magnesium (Mg). The three primary data sources fused together were Visible Near Infrared (VisNIR) spectra, Mid-Infrared (MIR) spectra, and X-ray Fluorescence (XRF) data. The major conclusions drawn from this study are as follows:

a. Low-Level Fusion may not be suitable for fusing data that are inherently compatible, as no significant improvement in predictive performance was observed for any of the nutrients.
b. Mid-level fusion improved the R^2 by 0.1-18%, RMSE by 0-8%, and RPIQ by 0-11.5%, depending on the nutrients and fusion algorithm. Partial Least Squares (PLS) reduction fusion is a superior fusion algorithm compared to Principal Component (PC) reduction fusion algorithm, as it can produce spectral vectors directly related to the soil attributes.
c. PLS reduction fusion is consistently better for all nutrients compared to other algorithms. In fact, PLS reduction fusion is the only algorithm that never had R^2 lower than individual models for any of the nutrients.
d. High-level fusion improved the R^2 by 0-12.5%, RMSE by 0-12%, and RPIQ by 2.3-13.4% depending on nutrients and fusion algorithm. High-level fusion

algorithms were able to improve the predictive performance for secondary nutrients significantly.

e. Although fusion algorithms improved predictive performance overall, the predictions regarding primary nutrients were satisfactory only. None of the algorithms could significantly improve predictions for Phosphorous.

f. None of the fusion algorithms significantly improved bias. Fusion algorithms focused on reducing noise and extracting features.

g. Researchers could consider exploring novel fusion algorithms and use of non-linear methods, such as SVR and RF, as based models to address the non-linearity introduced by the fusion of data effectively.

APPENDIX

TABLE A 1 MEAN SEPARATION TEST OF R2 VALUES

i. Magnesium			
Fusion	**Mean**		
GRA	0.754		A
PLS Red.	0.745	B	A
SA	0.735	B	A
Low Level	0.704	B	C
VisNIR	0.689		C
PC Red.	0.678		C
MIR	0.668		C
XRF	0.419		D

ii. Calcium			
Fusion	**Mean**		
SA	0.815		A
GRA	0.813		A
PLS Red.	0.805		A
PC Red.	0.794	B	A
MIR	0.779	B	
Low Level	0.755		C
XRF	0.698		D
VisNIR	0.663		E

iii. Nitrogen			
Fusion	**Mean**		
SA	0.542		A
GRA	0.534		A
PLS Red.	0.526	B	A
PC Red.	0.503	B	C
VisNIR	0.494		C
Low Level	0.459		D
MIR	0.446	E	D
XRF	0.431	E	

iv. Phosphorous		
Fusion	**Mean**	
PLS Red.	0.663	A
PC Red.	0.663	A
XRF	0.663	A
GRA	0.658	A
Low Level	0.642	A
SA	0.626	A
VisNIR	0.452	B
MIR	0.377	C

v. Potassium			
Fusion	**Mean**		
PC Red.	0.547		A
SA	0.531	B	A
PLS Red.	0.528	B	A
GRA	0.511	B	
Low Level	0.469		C
MIR	0.465		C
VisNIR	0.434		D
XRF	0.405		E

TABLE A 2 MEAN SEPARATION TEST OF RMSE VALUES

i. Magnesium				
Fusion	**Mean**			
XRF	103		A	
PC Red.	77		B	
MIR	72	C	B	
VisNIR	69	C	B	D
Low Level	68	C	B	D
SA	64	C		D
PLS Red.	63			D
GRA	60			D

ii. Calcium			
Fusion	**Mean**		
VisNIR	417		A
XRF	400		A
Low Level	357		B
MIR	335		C
PC Red.	323	D	C
PLS Red.	316	D	
SA	308	D	
GRA	306	D	

iii. Nitrogen			
Fusion	**Mean**		
XRF	4.45		A
MIR	4.37		A
Low Level	4.36		A
PC Red.	4.22	B	A
VisNIR	4.22	B	A
PLS Red.	4.07	B	C
GRA	4.00	B	C
SA	3.95		C

iv. Phosphorous		
Fusion	**Mean**	
MIR	24.0	A
VisNIR	22.6	A
Low Level	19.4	B
SA	18.8	B
PC Red.	18.7	B
GRA	18.6	B
PLS Red.	18.5	B
XRF	18.4	B

v. Potassium				
Fusion	**Mean**			
XRF	112		A	
Low Level	104		B	
VisNIR	103	C	B	
MIR	101	C	B	D
GRA	97	C	E	D
PLS Red.	96		E	D
SA	94		E	
PC Red.	93		E	

TABLE A 3 MEAN SEPARATION TEST OF RPIQ VALUES

i. Magnesium			
Fusion Type	**Mean**		
GRA	2.13		A
PLS Red.	2.09		A
SA	2.04	B	A
Low Level	1.92	B	C
VisNIR	1.88		C
PC Red.	1.84		C
MIR	1.79		C
XRF	1.34		D

ii. Calcium			
Fusion Type	**Mean**		
GRA	2.30		A
SA	2.28		A
PLS Red.	2.25		A
PC Red.	2.19	B	A
MIR	2.11	B	C
Low Level	2.00		C
XRF	1.79		D
VisNIR	1.70		D

iii. Nitrogen				
Fusion Type	**Mean**			
SA	2.08		A	
GRA	2.05	B	A	
PLS Red.	2.02	B	A	C
PC Red.	1.94	B	D	C
VisNIR	1.93		D	C
Low Level	1.88		D	
MIR	1.87		D	
XRF	1.85		D	

iv. Phosphorous		
Fusion Type	**Mean**	
SA	1.50	A
GRA	1.47	A
XRF	1.47	A
PLS Red.	1.46	A
PC Red.	1.43	A
Low Level	1.41	A
VisNIR	1.23	B
MIR	1.15	B

v. Potassium			
Fusion Type	**Mean**		
PC Red.	1.93		A
SA	1.90		A
PLS Red.	1.89		A
GRA	1.83	B	A
MIR	1.79	B	
Low Level	1.76	B	
VisNIR	1.74	B	C
XRF	1.65		C

TABLE A 4 MEAN SEPARATION TEST OF ABSOLUTE BIAS VALUES

i. Magnesium

Fusion Type	Mean	
XRF	6.5004	A
VisNIR	4.8815	B
PC Red.	4.8759	B
MIR	4.6256	B
Low Level	4.4333	B
PLS Red.	4.1725	B
SA	4.0052	B
GRA	3.7586	B

ii. Calcium

Fusion Type	Mean		
XRF	28.7646		A
VisNIR	26.1835	B	A
Low Level	23.4048	B	A
SA	23.0791	B	A
MIR	22.3492	B	A
GRA	22.192	B	A
PLS Red.	21.4126	B	A
PC Red.	20.5209	B	

iii. Nitrogen

Fusion Type	Mean	
PC Red.	0.2857	A
VisNIR	0.2791	A
Low Level	0.2775	A
XRF	0.2691	A
PLS Red.	0.2628	A
MIR	0.2621	A
GRA	0.2389	A
SA	0.2279	A

iv. Phosphorous

Fusion Type	Mean		
MIR	1.6302		A
VisNIR	1.4693	B	A
PC Red.	1.3949	B	A
Low Level	1.3011	B	A
SA	1.244	B	A
XRF	1.2153	B	A
PLS Red.	1.1859	B	
GRA	1.113	B	

v. Potassium

Fusion Type	Mean	
XRF	7.74	A
Low Level	7.1222	A
VisNIR	6.893	A
MIR	6.6373	A
PC Red.	6.3973	A
PLS Red.	6.3477	A
GRA	5.9131	A
SA	5.859	A

REFERENCES

1. Adamchuk., V. i., Viscarra Rossel, R. A., Sudduth, K. A., & LAmmmers, P. S. (2011). Sensor Fusion for Precision Agriculture. *Sensor Fusion - Foundation and Applications*. https://doi.org/10.5772/19983
2. Aldabaa, A. A. A., Weindorf, D. C., Chakraborty, S., Sharma, A., & Li, B. (2015). Combination of proximal and remote sensing methods for rapid soil salinity quantification. *Geoderma*, *239*, 34–46. https://doi.org/10.1016/J.GEODERMA.2014.09.011
3. Barbedo, J. G. A. (2022). Data Fusion in Agriculture: Resolving Ambiguities and Closing Data Gaps. *Sensors*, *22*(6). https://doi.org/10.3390/s22062285
4. Barros, A. S., Pinto, R., Bouveresse, D. J. R., & Rutledge, D. N. (2008). Principal component transform - Outer product analysis in the PCA context. *Chemometrics and Intelligent Laboratory Systems*, *93*(1), 43–48. https://doi.org/10.1016/J.CHEMOLAB.2008.03.009
5. Bates, J. M., & Granger, C. W. J. (1969). The Combination of Forecasts. In *Essays in Econometrics* (pp. 391–410). https://doi.org/10.1017/cbo9780511753961.021
6. Bellon-Maurel, V., & McBratney, A. (2011). Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils - Critical review and research perspectives. In *Soil Biology and Biochemistry* (Vol. 43, Issue 7, pp. 1398–1410). Pergamon. https://doi.org/10.1016/j.soilbio.2011.02.019

7. Biancolillo, A., Bucci, R., Magrì, A. L., Magrì, A. D., & Marini, F. (2014). Data-fusion for multiplatform characterization of an italian craft beer aimed at its authentication. *Analytica Chimica Acta*, *820*, 23–31. https://doi.org/10.1016/J.ACA.2014.02.024
8. Borràs, E., Ferré, J., Boqué, R., Mestres, M., Aceña, L., & Busto, O. (2015). Data fusion methodologies for food and beverage authentication and quality assessment – A review. *Analytica Chimica Acta*, *891*, 1–14. https://doi.org/10.1016/J.ACA.2015.04.042
9. Castrignanò, A., Buttafuoco, G., Quarto, R., Parisi, D., Viscarra Rossel, R. A., Terribile, F., Langella, G., & Venezia, A. (2018). A geostatistical sensor data fusion approach for delineating homogeneous management zones in Precision Agriculture. *Catena*, *167*(January), 293–304. https://doi.org/10.1016/j.catena.2018.05.011
10. Cécillon, L., Certini, G., Lange, H., Forte, C., & Strand, L. T. (2012). Spectral fingerprinting of soil organic matter composition. *Organic Geochemistry*, *46*, 127–136. https://doi.org/10.1016/J.ORGGEOCHEM.2012.02.006
11. Comino, F., Ayora-Cañada, M. J., Aranda, V., Díaz, A., & Domínguez-Vidal, A. (2018). Near-infrared spectroscopy and X-ray fluorescence data fusion for olive leaf analysis and crop nutritional status determination. *Talanta*, *188*, 676–684. https://doi.org/10.1016/J.TALANTA.2018.06.058
12. Dong, J., Zhuang, D., Huang, Y., & Fu, J. (2009). Advances in multi-sensor data fusion: Algorithms and applications. *Sensors*, *9*(10), 7771–7784. https://doi.org/10.3390/s91007771

13. Elmenreich, W. (2002). *An Introduction to Sensor Fusion SWILT-Swarm Intelligence Layer to Control Autonomous Agents View project Synthetic Data for NILM View project An Introduction to Sensor Fusion*. https://www.researchgate.net/publication/267771481

14. Filippi, P., Jones, E. J., & Bishop, T. F. A. (2020). Catchment-scale 3D mapping of depth to soil sodicity constraints through combining public and on-farm soil databases – A potential tool for on-farm management. *Geoderma*, *374*. https://doi.org/10.1016/J.GEODERMA.2020.114396

15. Fontenelli, J. V., Adamchuk, V. I., Ferreira, M. M. C., Amaral, L. R., Guimarães, C. C. B., Demattê, J. A. M., & Magalhães, P. S. G. (2021). Evaluating the synergy of three soil spectrometers for improving the prediction and mapping of soil properties in a high anthropic management area: A case of study from Southeast Brazil. *Geoderma*, *402*, 115347. https://doi.org/10.1016/J.GEODERMA.2021.115347

16. Ge, Y., Morgan, C. L. S., & Ackerson, J. P. (2014). VisNIR spectra of dried ground soils predict properties of soils scanned moist and intact. *Geoderma*, *221–222*, 61–69. https://doi.org/10.1016/j.geoderma.2014.01.011

17. Ge, Y., Thomasson, J. A., & Sui, R. (2011). Remote sensing of soil properties in precision agriculture: A review. *Frontiers of Earth Science*, *5*(3), 229–238. https://doi.org/10.1007/s11707-011-0175-0

18. Geladi, Paul and Kowalski, B. R. (1985). PARTIAL LEAST-SQUARES REGRESSION: A TUTORIAL. *Journal of Optoelectronics and Advanced Materials*, *10*(9), 2303–2306.

19. Goodale, N., Bailey, D. G., Jones, G. T., Prescott, C., Scholz, E., Stagliano, N., & Lewis, C. (2012). PXRF: A study of inter-instrument performance. *Journal of Archaeological Science*, *39*(4), 875–883. https://doi.org/10.1016/j.jas.2011.10.014

20. Granger, C. W. J., & Ramanathan, R. (1984). Improved Methods of Combining Forecasts. *Journal of Forecasting*, *3*, 197–204.

21. Helland, I. (2005). PARTIAL LEAST SQUARES REGRESSION. *Encyclopedia of Statistical Sciences*, *9*, 1–5. https://doi.org/10.1002/0471667196.ess6004

22. Javadi, S. H., & Mouazen, A. M. (2021). Data fusion of xrf and vis-nir using outer product analysis, granger–ramanathan, and least squares for prediction of key soil attributes. *Remote Sensing*, *13*(11). https://doi.org/10.3390/rs13112023

23. Javadi, S. H., Munnaf, M. A., & Mouazen, A. M. (2021). Fusion of Vis-NIR and XRF spectra for estimation of key soil attributes. *Geoderma*, *385*, 114851. https://doi.org/10.1016/j.geoderma.2020.114851

24. Ji, W., Adamchuk, V. I., Chen, S., Mat Su, A. S., Ismail, A., Gan, Q., Shi, Z., & Biswas, A. (2019). Simultaneous measurement of multiple soil properties through proximal sensor data fusion: A case study. *Geoderma*, *341*, 111–128. https://doi.org/10.1016/J.GEODERMA.2019.01.006

25. Kuang, B., Mahmood, H. S., Quraishi, M. Z., Hoogmoed, W. B., Mouazen, A. M., & van Henten, E. J. (2012). Sensing soil properties in the laboratory, in situ, and on-line. A review. In *Advances in Agronomy* (1st ed., Vol. 114). Elsevier Inc. https://doi.org/10.1016/B978-0-12-394275-3.00003-1

26. Li, F., Xu, L., You, T., & Lu, A. (2021). Measurement of potentially toxic elements in the soil through NIR, MIR, and XRF spectral data fusion. *Computers*

and Electronics in Agriculture, *187*, 106257. https://doi.org/10.1016/J.COMPAG.2021.106257

27. Li, N., Zare, E., Huang, J., & Triantafilis, J. (2018). Mapping Soil Cation-Exchange Capacity using Bayesian Modeling and Proximal Sensors at the Field Scale. *Soil Science Society of America Journal*, *82*(5), 1203–1216. https://doi.org/10.2136/sssaj2017.10.0356
28. Lin, W. (2022). *mt package* (2.0-1.19). https://cran.r-project.org/web/packages/mt/index.html
29. Loiseau, T., Chen, S., Mulder, V. L., Román Dobarco, M., Richer-de-Forges, A. C., Lehmann, S., Bourennane, H., Saby, N. P. A., Martin, M. P., Vaudour, E., Gomez, C., Lagacherie, P., & Arrouays, D. (2019). Satellite data integration for soil clay content modelling at a national scale. *International Journal of Applied Earth Observation and Geoinformation*, *82*. https://doi.org/10.1016/J.JAG.2019.101905
30. Mahmood, H. S., Hoogmoed, W. B., & van Henten, E. J. (2012). Sensor data fusion to predict multiple soil properties. *Precision Agriculture*, *13*(6), 628–645. https://doi.org/10.1007/s11119-012-9280-7
31. Mancini, M., Andrade, R., Teixeira, A. F. dos S., Silva, S. H. G., Weindorf, D. C., Chakraborty, S., Guilherme, L. R. G., & Curi, N. (2022). Proximal sensor data fusion for Brazilian soil properties prediction: Exchangeable/available macronutrients, aluminum, and potential acidity. *Geoderma Regional*, *30*, e00573. https://doi.org/10.1016/J.GEODRS.2022.E00573
32. Mulla, D. J. (2013). Twenty five years of remote sensing in precision agriculture:

Key advances and remaining knowledge gaps. In *Biosystems Engineering* (Vol. 114, Issue 4, pp. 358–371). Academic Press. https://doi.org/10.1016/j.biosystemseng.2012.08.009

33. National Cooperative Soil Survey. (2013). *National Cooperative Soil Characterization Database*. http://ncsslabdatamart.sc.egov.usda.gov/
34. Nawar, S., Corstanje, R., Halcro, G., Mulla, D., & Mouazen, A. M. (2017). Delineation of Soil Management Zones for Variable-Rate Fertilization: A Review. *Advances in Agronomy*, *143*, 175–245. https://doi.org/10.1016/BS.AGRON.2017.01.003
35. Nawar, S., Richard, F., Kassim, A. M., Tekin, Y., & Mouazen, A. M. (2022). Fusion of Gamma-rays and portable X-ray fluorescence spectral data to measure extractable potassium in soils. *Soil and Tillage Research*, *223*, 105472. https://doi.org/10.1016/J.STILL.2022.105472
36. Ng, W., Minasny, B., Jeon, S. H., & McBratney, A. (2022). Mid-infrared spectroscopy for accurate measurement of an extensive set of soil properties for assessing soil functions. *Soil Security*, *6*, 100043. https://doi.org/10.1016/J.SOISEC.2022.100043
37. Ng, W., Minasny, B., Montazerolghaem, M., Padarian, J., Ferguson, R., Bailey, S., & McBratney, A. B. (2019). Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra. *Geoderma*, *352*(January), 251–267. https://doi.org/10.1016/j.geoderma.2019.06.016
38. O'Rourke, S. M., Minasny, B., Holden, N. M., & McBratney, A. B. (2016).

Synergistic Use of Vis-NIR, MIR, and XRF Spectroscopy for the Determination of Soil Geochemistry. *Soil Science Society of America Journal*, *80*(4), 888–899. https://doi.org/10.2136/sssaj2015.10.0361

39. ORourke, S. M., Holden, N. M., McBratney, A. B., & Minasny, B. (2016). An assessment of model averaging to improve predictive power of portable vis-NIR and XRF for the determination of agronomic soil properties. *Geoderma*, *279*, 31–44.
40. Ravansari, R., Wilson, S. C., & Tighe, M. (2020). Portable X-ray fluorescence for environmental assessment of soils: Not just a point and shoot method. *Environment International*, *134*(July 2019), 105250. https://doi.org/10.1016/j.envint.2019.105250
41. Silva, S. H. G., Ribeiro, B. T., Guerra, M. B. B., de Carvalho, H. W. P., Lopes, G., Carvalho, G. S., Guilherme, L. R. G., Resende, M., Mancini, M., Curi, N., Rafael, R. B. A., Cardelli, V., Cocco, S., Corti, G., Chakraborty, S., Li, B., & Weindorf, D. C. (2021). pXRF in tropical soils: Methodology, applications, achievements and challenges. *Advances in Agronomy*, *167*, 1–62. https://doi.org/10.1016/bs.agron.2020.12.001
42. Silvestri, M., Bertacchini, L., Durante, C., Marchetti, A., Salvatore, E., & Cocchi, M. (2013). Application of data fusion techniques to direct geographical traceability indicators. *Analytica Chimica Acta*, *769*, 1–9. https://doi.org/10.1016/J.ACA.2013.01.024
43. Song, Y., Li, F., Yang, Z., Ayoko, G. A., Frost, R. L., & Ji, J. (2012). Diffuse reflectance spectroscopy for monitoring potentially toxic elements in the

agricultural soils of Changjiang River Delta, China. *Applied Clay Science*, *64*, 75–83. https://doi.org/10.1016/J.CLAY.2011.09.010

44. Statistical office of the European Union. (2017). *Database - Land cover/use statistics - Eurostat*. https://ec.europa.eu/eurostat/web/lucas/data/database
45. Stenberg, B., Viscarra Rossel, R. A., Mouazen, A. M., & Wetterlind, J. (2010). Visible and Near Infrared Spectroscopy in Soil Science. *Advances in Agronomy*, *107*(C), 163–215. https://doi.org/10.1016/S0065-2113(10)07005-7
46. Taghizadeh-Mehrjardi, R., Khademi, H., Khayamim, F., Zeraatpisheh, M., Heung, B., & Scholten, T. (2022). A Comparison of Model Averaging Techniques to Predict the Spatial Distribution of Soil Properties. *Remote Sensing*, *14*(3), 472. https://doi.org/10.3390/rs14030472
47. Tavares, T. R., Molin, J. P., Hamed Javadi, S., Pereira De Carvalho, H. W., & Mouazen, A. M. (2021). Combined Use of Vis-NIR and XRF Sensors for Tropical Soil Fertility Analysis: Assessing Different Data Fusion Approaches. *Sensors*, *21(1)*, 148. https://doi.org/10.3390/s21010148
48. Terra, F. S., Viscarra Rossel, R. A., & Demattê, J. A. M. (2019). Spectral fusion by Outer Product Analysis (OPA) to improve predictions of soil organic C. *Geoderma*, *335*, 35–46. https://doi.org/10.1016/J.GEODERMA.2018.08.005
49. Tsimpouris, E., Tsakiridis, N. L., & Theocharis, J. B. (2021). Using autoencoders to compress soil VNIR–SWIR spectra for more robust prediction of soil properties. *Geoderma*, *393*. https://doi.org/10.1016/J.GEODERMA.2021.114967
50. Veselá, A., Barros, A. S., Synytsya, A., Delgadillo, I., Čopíková, J., & Coimbra, M. A. (2007). Infrared spectroscopy and outer product analysis for quantification

of fat, nitrogen, and moisture of cocoa powder. *Analytica Chimica Acta*, *601*(1), 77–86. https://doi.org/10.1016/J.ACA.2007.08.039

51. Viscarra Rossel, R. A., Adamchuk, V. I., Sudduth, K. A., McKenzie, N. J., & Lobsey, C. (2011). Proximal Soil Sensing: An Effective Approach for Soil Measurements in Space and Time. In *Advances in Agronomy* (Vol. 113). Elsevier Inc. https://doi.org/10.1016/B978-0-12-386473-4.00005-1
52. Viscarra Rossel, R. A., Walvoort, D. J. J., McBratney, A. B., Janik, L. J., & Skjemstad, J. O. (2006). Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma*, *131*(1–2), 59–75. https://doi.org/10.1016/j.geoderma.2005.03.007
53. Vohland, M., Ludwig, B., Seidel, M., & Hutengs, C. (2022). Quantification of soil organic carbon at regional scale: Benefits of fusing vis-NIR and MIR diffuse reflectance data are greater for in situ than for laboratory-based modelling approaches. *Geoderma*, *405*(February 2021), 115426. https://doi.org/10.1016/j.geoderma.2021.115426
54. Wan, M., Hu, W., Qu, M., Li, W., Zhang, C., Kang, J., Hong, Y., Chen, Y., & Huang, B. (2020). Rapid estimation of soil cation exchange capacity through sensor data fusion of portable XRF spectrometry and Vis-NIR spectroscopy. *Geoderma*, *363*. https://doi.org/10.1016/J.GEODERMA.2019.114163
55. Wang, D., Chakraborty, S., Weindorf, D. C., Li, B., Sharma, A., Paul, S., & Ali, M. N. (2015). Synthesized use of VisNIR DRS and PXRF for soil characterization: Total carbon and total nitrogen. *Geoderma*, *243–244*, 157–167.
56. Wang, J., Zhao, X., Deuss, K. E., Cohen, D. R., & Triantafilis, J. (2022).

Proximal and remote sensor data fusion for 3D imaging of infertile and acidic soil. *Geoderma*, *424*, 115972. https://doi.org/10.1016/j.geoderma.2022.115972

57. Weindorf, D. C., Bakr, N., & Zhu, Y. (2014). Advances in portable X-ray fluorescence (PXRF) for environmental, pedological, and agronomic applications. In *Advances in Agronomy* (Vol. 128). Elsevier. https://doi.org/10.1016/B978-0-12-802139-2.00001-9
58. Weindorf, D. C., Chakraborty, S., Li, B., Deb, S., Singh, A., & Kusi, N. Y. (2018). Compost salinity assessment via portable X-ray fluorescence (PXRF) spectrometry. *Waste Management*, *78*, 158–163. https://doi.org/10.1016/j.wasman.2018.05.044
59. Wijewardane, N. K., Ge, Y., & Morgan, C. L. S. (2016). Moisture insensitive prediction of soil properties from VNIR reflectance spectra based on external parameter orthogonalization. *Geoderma*, *267*, 92–101. https://doi.org/10.1016/J.GEODERMA.2015.12.014
60. Wijewardane, N. K., Ge, Y., Wills, S., & Libohova, Z. (2018). Predicting Physical and Chemical Properties of US Soils with a Mid-Infrared Reflectance Spectral Library. *Soil Science Society of America Journal*, *82*(3), 722–731. https://doi.org/10.2136/sssaj2017.10.0361
61. Wijewardane, N. K., Ge, Y., Wills, S., & Loecke, T. (2016). Prediction of Soil Carbon in the Conterminous United States: Visible and Near Infrared Reflectance Spectroscopy Analysis of the Rapid Carbon Assessment Project. *Soil Science Society of America Journal*, *80*(4), 973–982. https://doi.org/10.2136/sssaj2016.02.0052

62. Wijewardane, N. K., Hetrick, S., Ackerson, J., Morgan, C. L. S., & Ge, Y. (2020). VisNIR integrated multi-sensing penetrometer for in situ high-resolution vertical soil sensing. *Soil and Tillage Research*, *199*, 104604. https://doi.org/10.1016/j.still.2020.104604

63. Wijewardane, N. K., Wang, L., Zhan, Y., Franz, T., Yu, H., Zhou, Y., Shi, Y., & Ge, Y. (2019). Mapping infield variability of soil properties to support precision agriculture using UAV, multi-depth EC, and aerial hyperspectral imagery. *Proximal Soil Sensing*, 15–20.

64. Xing, Z., Du, C., Shen, Y., Ma, F., & Zhou, J. (2021). A method combining FTIR-ATR and Raman spectroscopy to determine soil organic matter: Improvement of prediction accuracy using competitive adaptive reweighted sampling (CARS). *Computers and Electronics in Agriculture*, *191*, 106549. https://doi.org/10.1016/J.COMPAG.2021.106549

65. Xu, D., Chao, R., Li, S., Chen, S., Jiang, Q., Zhou, L., & Shi, Z. (2019). Multi-sensor fusion for the determination of several soil properties in the Yangtze River Delta, China. *European Journal of Soil Science*, *70*, 162–173. https://doi.org/10.1111/ejss.12729

66. Xu, Dongyun, Chen, S., Xu, H., Wang, N., Zhou, Y., & Shi, Z. (2020). Data fusion for the measurement of potentially toxic elements in soil using portable spectrometers. *Environmental Pollution*, *263*, 114649. https://doi.org/10.1016/J.ENVPOL.2020.114649

67. Xu, H., Xu, D., Chen, S., Ma, W., & Shi, Z. (2020). Rapid Determination of Soil Class Based on Visible-Near Infrared, Mid-Infrared Spectroscopy and Data

Fusion. *Remote Sensing 2020, Vol. 12, Page 1512*, *12*(9), 1512. https://doi.org/10.3390/RS12091512

68. Xu, X., Du, C., Ma, F., Shen, Y., Wu, K., Liang, D., & Zhou, J. (2019). Detection of soil organic matter from laser-induced breakdown spectroscopy (LIBS) and mid-infrared spectroscopy (FTIR-ATR) coupled with multivariate techniques. *Geoderma*, *355*. https://doi.org/10.1016/J.GEODERMA.2019.113905
69. Zhang, N., Wang, M., & Wang, N. (2002). Precision agriculture - A worldwide overview. *Computers and Electronics in Agriculture*, *36*(2–3), 113–132. https://doi.org/10.1016/S0168-1699(02)00096-0
70. Zhang, Y., Hartemink, A. E., & Alfred Hartemink, C. E. (2020). Data fusion of vis–NIR and PXRF spectra to predict soil physical and chemical properties. *European Journal of Soil Science*, *71*(3), 316–333. https://doi.org/10.1111/EJSS.12875
71. Zhao, L., Hong, H., Algeo, T. J., Liu, C., & Lu, A. (2022). Fusion of visible near-infrared and mid-infrared data for modelling key soil-forming processes in loess soils. *European Journal of Soil Science*, *73*(1), e13208. https://doi.org/10.1111/EJSS.13208