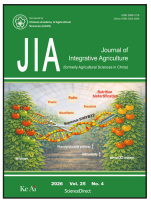
Journal of Integrative Agriculture 2026, 25(4): 1687–1699

Available online at ScienceDirect

Journal of Integrative Agriculture



RESEARCH ARTICLE

Improved selected soil properties predictions using MIR and pXRF sensor fusion

Junwei Wang[1, 2], Qi Zou[1, 2], Huimin Yuan[1, 2#]

[1] *College of Resources and Environmental Sciences/National Academy of Agriculture Green Development/Key Laboratory of Plant–Soil Interactions, Ministry of Education/China Agricultural University, Beijing 100193, China*

[2] *National Observation and Research Station of Agriculture Green Development, Quzhou 057250, China*

Highlights

- The accurate prediction of soil available nutrients and micronutrients could be achieved by the fusion of MIR and pXRF.
- LASSO-selected feature bands improved fusion performance over direct concatenation.
- GRA improved the prediction accuracy of most soil properties compared with the best single sensor models.
- The prediction accuracy of SO-PLS fusion might be influenced by the sequence in which data from different sensors are defined.

Abstract

The timely and accurate assessment of soil nutrient information is essential for ensuring global food security and sustainable agricultural development. This study evaluated the individual and fusion performance of mid-infrared (MIR) and portable X-ray fluorescence (pXRF) spectroscopy for predicting selected soil properties. Four sensor fusion strategies were implemented: direct concatenation (DC), feature-level fusion using stability competitive adaptive reweighted sampling (sCARS) and least absolute shrinkage and selection operator (LASSO) algorithms (sCARS-C and LASSO-C), multi-block fusion *via* sequential orthogonal partial least squares (SO-PLS), and Granger-Ramanathan model averaging (GRA) method to enhance prediction accuracy for 13 soil properties. The findings revealed that single sensor models using either MIR or pXRF provided accurate estimations for soil organic matter (SOM), total nitrogen (TN), available phosphorus (AP), calcium (Ca), iron (Fe), manganese (Mn), and pH, but showed limitations for total potassium (TK), magnesium (Mg), copper (Cu), zinc (Zn), available potassium (AK), and total phosphorus (TP). The DC model significantly improved predictions for Mg (Rp^2=0.76, RMSEp=358.76 mg kg^{-1}, RPDp=2.03) and TK (Rp^2=0.75, RMSEp=775.96 mg kg^{-1}, RPDp=2.00). The LASSO-C model demonstrated superior prediction accuracy compared to the DC model for AP, AK, TP, Zn, Mn, and Cu, achieving optimal results for AP (Rp^2=0.89, RMSEp=21.37 mg kg^{-1}, RPDp=3.01) and Zn (Rp^2=0.80, RMSEp=9.88 mg kg^{-1}, RPDp=2.32). This enhancement is attributed to LASSO's effective selection of feature information from the complete MIR and pXRF spectra. The GRA models achieved the highest prediction accuracy for TP, pH, AK, and Cu, with Rp^2 values of 0.80, 0.82, 0.82, and 0.65, RMSEp values of 129.21 mg kg^{-1}, 0.13, 48.38 mg kg^{-1}, and 3.87 mg kg^{-1}, and RPDp values of 2.23, 2.34, 2.37, and 1.67, respectively. For single-sensor applications, MIR spectra are recommended for predicting SOM, TN, and Ca (Rp^2≥0.88, RPDp≥2.87), while pXRF is more cost-effective for measuring Ca, Fe, and Mn (Rp^2≥0.80, RPDp≥2.22). This research demonstrates the effectiveness of MIR and pXRF sensor fusion in enhancing soil nutrient assessment accuracy, particularly for available nutrients and micronutrients.

Keywords: soil available nutrients, micronutrients, multi-block fusion, feature-level fusion, model averaging

1. Introduction

Timely access to soil information is essential for soil nutrient management, particularly regarding soil available nutrients and micronutrients information (Mukherjee 2022). There is an increasing demand for affordable access to high-quality soil nutrient data with adequate spatial and temporal

Received 8 February 2025; Received in revised form 26 April 2025; Accepted 29 August 2025; Available online 26 September 2025
Junwei Wang, E-mail: wangjw272@cau.edu.cn; #Correspondence Huimin Yuan, E-mail: hmyuan@cau.edu.cn

© 2026 CAAS. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/). Peer review under responsibility of Editorial Board of *Journal of Integrative Agriculture*.
doi: 10.1016/j.jia.2025.09.028

resolutions to enhance agricultural productivity and address environmental challenges (Lobsey *et al.* 2017). Proximal sensors enable rapid and cost-efficient estimation of soil nutrients (Andrade *et al.* 2019; Xu *et al.* 2019; Shi *et al.* 2022). Specifically, mid-infrared (MIR) diffuse reflectance and portable X-ray fluorescence (pXRF) spectroscopy represent promising tools for proximal soil sensing applications, as both techniques facilitate soil analysis with minimal sample preparation and provide insights into various soil compositions (Barra *et al.* 2021; Tavares *et al.* 2023).

MIR and pXRF spectroscopy have been extensively applied in estimating soil properties, including soil organic matter (SOM), total nitrogen (TN), clay minerals, iron oxides, and soil elemental composition (Soriano-Disla *et al.* 2014; Andrade *et al.* 2019; Barra *et al.* 2021; Tavares *et al.* 2023). Furthermore, pXRF provides characteristic elemental information (e.g., Mg, P, K, and Ca) for the measured soil. Most previous studies have employed these optical techniques independently to predict soil properties. Although single sensor techniques have demonstrated potential in assessing certain soil properties, a single sensor is unlikely to fully characterize soil properties with sufficient accuracy (Tavares *et al.* 2021). For instance, soil available nutrients, cation exchange capacity (CEC), and pH lack distinct absorption bands in the MIR spectral region and can only be indirectly predicted through models based on spectrally active components (e.g., SOM, carbonates, and clay), often yielding poor results (McBride 2021; Karami *et al.* 2024). Regarding the prediction of soil micronutrients from MIR spectra, despite their well-defined spectral peaks, identifying them in bulk soil spectra remains challenging due to their low intensities compared to other soil constituents (Karami *et al.* 2024). Conversely, pXRF devices cannot directly quantify light elements such as carbon, hydrogen, and oxygen, thus cannot directly measure SOM or TN. Additionally, matrix effects, resulting from complex sample compositions or co-presence of other elements, can compromise the accuracy of low atomic number element quantification by affecting X-ray absorption and scattering (Antonangelo and Zhang 2024). Consequently, the limited information from a single spectrum restricts the predictive accuracy of soil properties. Integrating complementary information from multiple sensors has emerged as a viable solution to enhance soil property estimation reliability.

Various sensor fusion strategies have been developed and proven effective in improving soil properties estimation accuracy (Xu *et al.* 2019, 2020; Gholizadeh *et al.* 2021; Javadi *et al.* 2021; Shrestha *et al.* 2022). The direct concatenation fusion method achieves the complementarity of different spectral data types and moderately improves soil property prediction performance in several studies (Xu *et al.* 2019; Li *et al.* 2024). However, direct concatenation of spectral data from different sensors typically results in data redundancy and high computational costs, as each sensor generates thousands of spectral variables. Therefore, variable selection is performed to extract feature bands from each sensor before sensor fusion, thereby reducing data redundancy and improving fusion model accuracy. The least absolute shrinkage and selection operator (LASSO) algorithm employs linear regression with a normal function constraining regression coefficient absolute values within a threshold, then compresses and eliminates variables with correlations below the threshold by optimizing the objective function (Wu *et al.* 2022). The feature bands extracted by LASSO have proven effective in predicting soil nutrients when fusing visible-near infrared (Vis-NIR) and MIR sensors (Li *et al.* 2024). Moreover, stability competitive adaptive reweighted sampling (sCARS) algorithm has demonstrated efficiency in variable selection (Li *et al.* 2022). Limited research has focused on applying spectral variable selection techniques in sensor fusion for soil property estimation. Furthermore, sCARS and LASSO have not yet been applied to the feature-level fusion of soil MIR and pXRF.

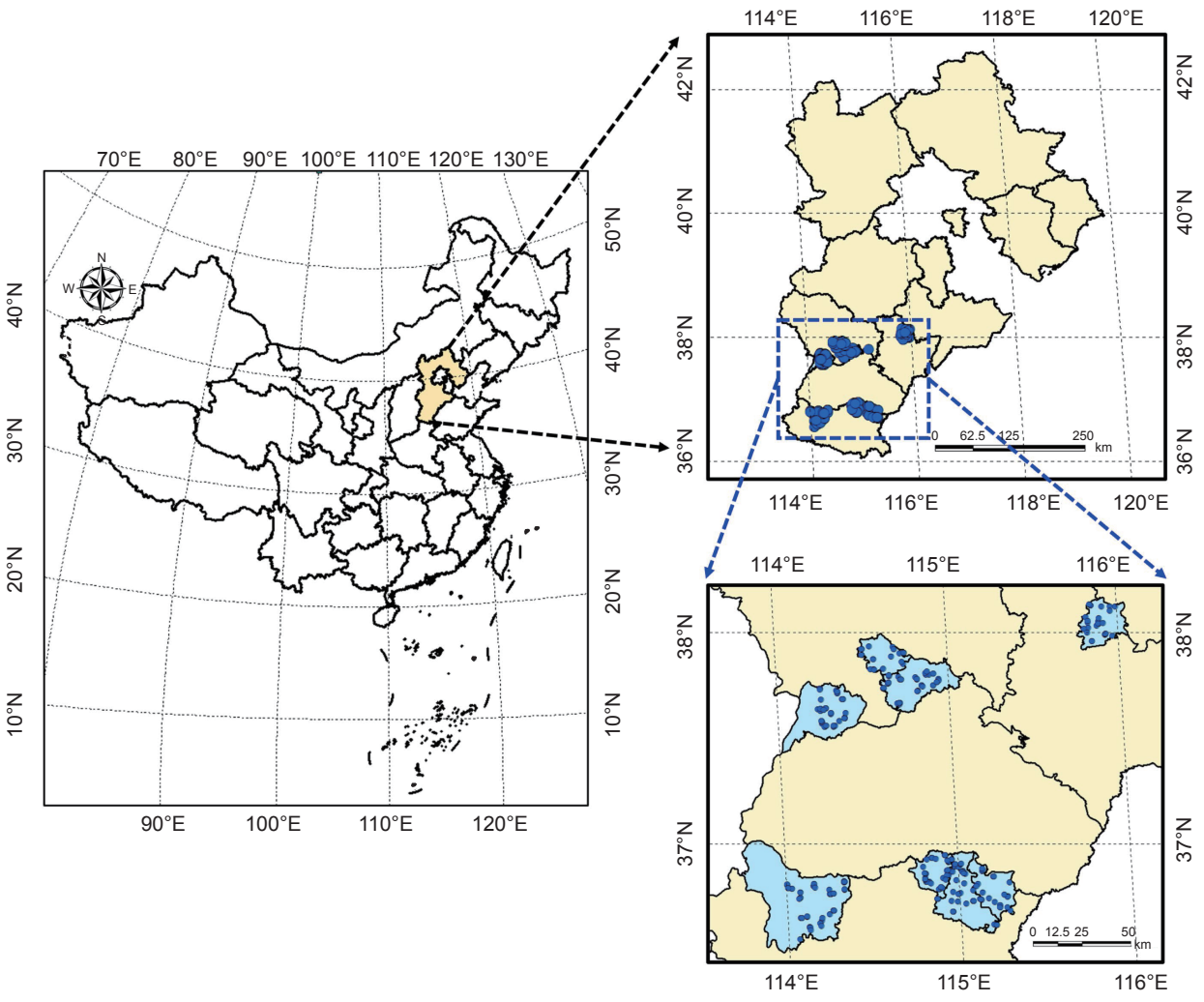
The multi-block fusion strategy, which combines data blocks from multiple sensors into a single model, has been employed to enhance the prediction performance of soil properties (Kandpal *et al.* 2022). The multi-block fusion based on sequential orthogonal partial least squares (SO-PLS) method has demonstrated effectiveness in soil analysis by sequentially identifying complementary information across different spectral data blocks and fully utilizing valid information from multiple sensor (Song *et al.* 2024). Model averaging represents another promising approach for improving soil spectroscopic prediction models. The Granger-Ramanathan model averaging (GRA) algorithm is preferred due to its operational simplicity and efficiency (Xu *et al.* 2020; Zhang *et al.* 2024). However, model averaging has limitations, as the random distribution of prediction errors from single sensor models constrains the useful information obtained, resulting in modest improvements compared to the best single sensor model (Xue *et al.* 2023). While research demonstrates that various sensor fusion strategies can enhance soil property predictions, no universal sensor fusion method exists for all datasets (Hong *et al.* 2022). Thus, investigating different sensor fusion methods remains essential for improving MIR and pXRF estimation of multiple soil properties.

Against this background, we hypothesized that the combined data from MIR and pXRF would provide complementary characterization of soil properties. The specific objectives of this study were to: (1) verify whether MIR or pXRF data could be used for the rapid and reliable estimation of 13 soil properties (namely, SOM, TN, available phosphorus (AP), available potassium (AK), pH, total phosphorus (TP), total potassium (TK), Ca, Mg, Fe, Mn, Cu, and Zn), with particular focus on soil available nutrients and micronutrients, and (2) explore four sensor fusion strategies to improve the prediction accuracy of soil properties and determine the most suitable approach for each target property.

2. Materials and methods

2.1. Study area, soil sampling and laboratory analyses

The study area encompassed eight counties in Hebei Province, China (Fig. 1). The region experiences a temperate continental monsoon climate, with January average temperatures below 3°C, July average temperatures



审图号 : GS 京 (2025) 2497 号

Fig. 1 Geographical location of the study area and distribution of soil sampling points.

between 18°C and 27°C, and mean annual precipitation of 484.5 mm. Fluvo-aquic soil predominates in the region. Major crops in the region include fruits, cereals, cotton, vegetables, and nuts. A total of 300 surface soil samples (0–30 cm) were collected from the study area. Sampling point coordinates were recorded using a handheld GPS device. After sampling, soil samples were placed in labeled plastic bags and transported to the laboratory for chemical and spectral analyses. The soil samples were air-dried at room temperature, ground, and sieved through a 2 mm mesh to remove plant roots and debris before analysis. Standard laboratory methods according to Bao (1981) were employed to determine soil SOM, TN, pH, AP, and AK contents. SOM content was measured using the potassium dichromate oxidation colorimetric method. The semi-micro automated Kjeldahl method was used to determine soil TN content. Soil pH was measured in a 1:1 soil-to-water suspension. The $NaHCO_3$ method was used for soil AP content determination. Soil AK content was analyzed using the NH_4OAc extraction method. The total concentrations of eight soil elements (TP, TK, Ca, Mg, Fe, Mn, Cu, and Zn) were then measured using a liquid chromatography inductively coupled plasma mass spectrometer (ICP-MS) (Agilent Technologies, USA).

2.2. MIR and pXRF spectra measurement

For MIR analysis, soil samples were further ground, sieved through a 0.15 mm sieve, and oven-dried at 105°C for 2 h (Wang *et al.* 2022). The samples were stored in a desiccator (Nalgene, Thermo Fisher Scientific, Waltham, MA, USA) until analysis. MIR spectra were collected using a diffuse reflectance infrared Fourier transform (DRIFT) Tensor-spectrometer equipped with a high throughput screening (HTS-XT) extension unit and a liquid N_2 cooled mercury cadmium telluride (MCT) detector (Bruker Optics, Karlsruhe, Germany). The spectral range extended from 4,000 to 600 cm^{-1} with a resolution of 4 cm^{-1}, yielding 2,380 data points. Samples were placed in special aluminum microplates containing 24 wells, with surfaces smoothed to minimize scattering effects. Each sample was analyzed in duplicate wells, scanned 32 times per well, and the spectra were averaged (Li *et al.* 2022; Wang *et al.* 2022).

Soil pXRF spectra were obtained using a Tracer 5i (Bruker

Optics, Karlsruhe, Germany). This portable instrument is equipped with a 4 W Rh X-ray tube and a Peltier-cooled silicon drift detector containing 2,048 channels. The instrumental conditions employed were: X-ray tube voltage and current set at 35 keV and 30 μA, respectively, and a scanning time of 120 s under atmospheric pressure without filters (Zou *et al.* 2025). A cellulose pellet served as a blank sample to verify contaminant-free operation of the equipment and was scanned every 30 samples. Samples of 10–15 g were placed in polyethylene cups (31 mm in diameter and 23 mm in height), which were covered at the top with a 4 μm X-ray-transparent polypropylene film. Each sample was scanned in triplicate, with the cup repositioning after each scan. The average values were calculated and analyzed.

2.3. Modeling workflow for sensor fusion

The MIR and pXRF spectra exhibit inconsistent energy levels when measured by different sensors. Direct fusion of these spectra may result in data with small spectral magnitude being treated as noise, leading to failed spectral information fusion (Li *et al.* 2024). Therefore, standard normal variate (SNV) preprocessing was applied to both MIR and pXRF spectra to eliminate dimensional effects. Following the modeling approach of previous researchers (Karami *et al.* 2024; Shi *et al.* 2024), the dataset was randomly divided into calibration (70%) and validation (30%) subsets. A ten-times repeated random splitting strategy was implemented to enhance model robustness and minimize random partitioning effects. Single sensor models and sensor fusion models were then established to predict thirteen soil properties. The models were developed using the calibration set and evaluated using the validation set. Model accuracy comparisons were based on mean performance across ten repeated random splits. The process flowchart appears in Fig. 2.

2.4. Single sensor models

For single sensor models, partial least squares regression (PLSR) algorithm was conducted on both MIR and pXRF data to establish predictive models for the soil properties, designated as MIR-PLSR and pXRF-PLSR, respectively. PLSR, one of the multiple linear calibration algorithms, is widely utilized for spectral calibration and prediction (Soriano-Disla *et al.* 2014). Similar to principal component analysis, PLSR reduces predictor variables to multiple synthetic variables, termed "latent variables (LVs)", while accounting for the response variable. A linear model is subsequently fitted between the latent and response variables. A detailed description of the PLSR procedure is available in Wold *et al.* (2001). Ten-fold cross validation was employed to determine the minimum root mean square error of cross-validation (RMSEcv) and identify the optimal number of LVs for the PLSR model.

2.5. Sensor fusion strategies

Direct concatenation Following SNV preprocessing, the complete spectral data from MIR and pXRF were concatenated to create a new data matrix for PLSR model input, generating the direct concatenation (DC) model. The fused spectral dimension equaled the sum of the two single-spectral dimensions, maintaining constant observations.

Multi-block fusion The sequential orthogonalized partial least squares (SO-PLS) algorithm was utilized for multi-block data fusion. SO-PLS, a multi-block PLS method, effectively handles large sets of collinear variables (data blocks) and remains unaffected by block variance, making it particularly suitable for spectral data (Song *et al.* 2024). The core process of SO-PLS relies on matrix orthogonalization, enabling sequential extraction of complementary information from different blocks or sensor data. For a detailed description of the SO-PLS methodology, readers are referred to Kandpal *et al.* (2022). The full MIR spectral dataset was designated as the first block (X_1) and the full pXRF spectral dataset as the second block (X_2).

Feature-level fusion Feature-level fusion extracted spectral feature bands related to soil properties using sCARS or LASSO from MIR and pXRF data as model inputs, rather than utilizing entire spectra. The selected feature bands were combined to form a new spectral matrix for predictive modeling, distinguishing this approach from the DC fusion method. sCARS is an established spectral feature variable selection technique introduced by Zheng *et al.* (2012). Based on the significance of each spectral variable (defined as the index of stability — the absolute value of the regression coefficient divided by its standard deviation), sCARS selects M variable subsets from M Monte Carlo sampling runs through an iterative and competitive process. During each run, sCARS employs an exponentially decreasing function and adaptive reweighted sampling to identify the optimal subset while minimizing the RMSEcv (Wang *et al.* 2022). This study utilized the following parameters: Monte Carlo sampling times (M)=500, Monte Carlo sampling ratio (r)=0.7, and number of calculation loops (N)=100. LASSO functions as a variable selection method based on a linear regression model. This technique filters full-spectrum data for features, reduces spectral dimensionality, and enhances model prediction accuracy while effectively preventing overfitting (Li *et al.* 2024). The implementation of LASSO utilized the "glmnet" package in R, employing ten-fold cross validation to determine the optimal penalty parameter lambda corresponding to the minimum mean square error (MSE) value.

Granger-Ramanathan model averaging The GRA method was implemented to combine prediction results from MIR and pXRF single sensor models. This algorithm employs ordinary least squares to integrate different prediction outcomes (Granger and Ramanathan 1984). The method fits a multiple linear regression model where measured values are regressed against the corresponding predictions derived from the single sensor models, as expressed in eq. (1):

$$Y=W_o+(C_{pXRF}\times X_{pXRF})+(C_{MIR}\times X_{MIR}) \quad (1)$$

where Y represents the vector with the soil property of interest (measured element contents), X_{pXRF} and X_{MIR} denote the independent variables (prediction outcomes of pXRF and MIR), W_o represents the intercept, and C_{pXRF} and C_{MIR} indicate the coefficients of pXRF and MIR outcomes, respectively.

2.6. Model performance evaluation

The prediction performance assessment utilized the

determination coefficient (R^2) (eq. 2), the root mean square error of prediction (RMSE) (eq. 3), and the residual prediction deviation (RPD) (eq. 4). These evaluation parameters are defined by the following equations:

$$R^2=1-\frac{\sum_{i=1}^{n}(x_i-y_i)^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2} \quad (2)$$

$$\text{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-y_i)^2} \quad (3)$$

$$\text{RPD}=\frac{\text{SD}}{\text{RMSEp}} \quad (4)$$

where x_i and y_i are the observed and predicted value at point i, $\overline{x}$ is the mean of the observed values, n is the number of samples, and SD is the standard deviation of the measured values (Wang *et al.* 2022).

R^2 quantifies the extent to which the dependent variable is explained. RMSE measures prediction accuracy using the same units as the data. RPD, calculated as the ratio of measured values' standard deviation to RMSE, determines the quality of derived models (Briedis *et al.* 2020). Models achieving R^2 and RPD values exceeding 0.75 or 2.0, respectively, demonstrate high accuracy with good predictive capability. Models with R^2 values between 0.65 and 0.75 and RPD values between 1.4 and 2.0 exhibit satisfactory accuracy with moderate predictive performance. Models with R^2 and RPD values below these thresholds indicate poor performance without any predictive ability (Chang *et al.* 2001;

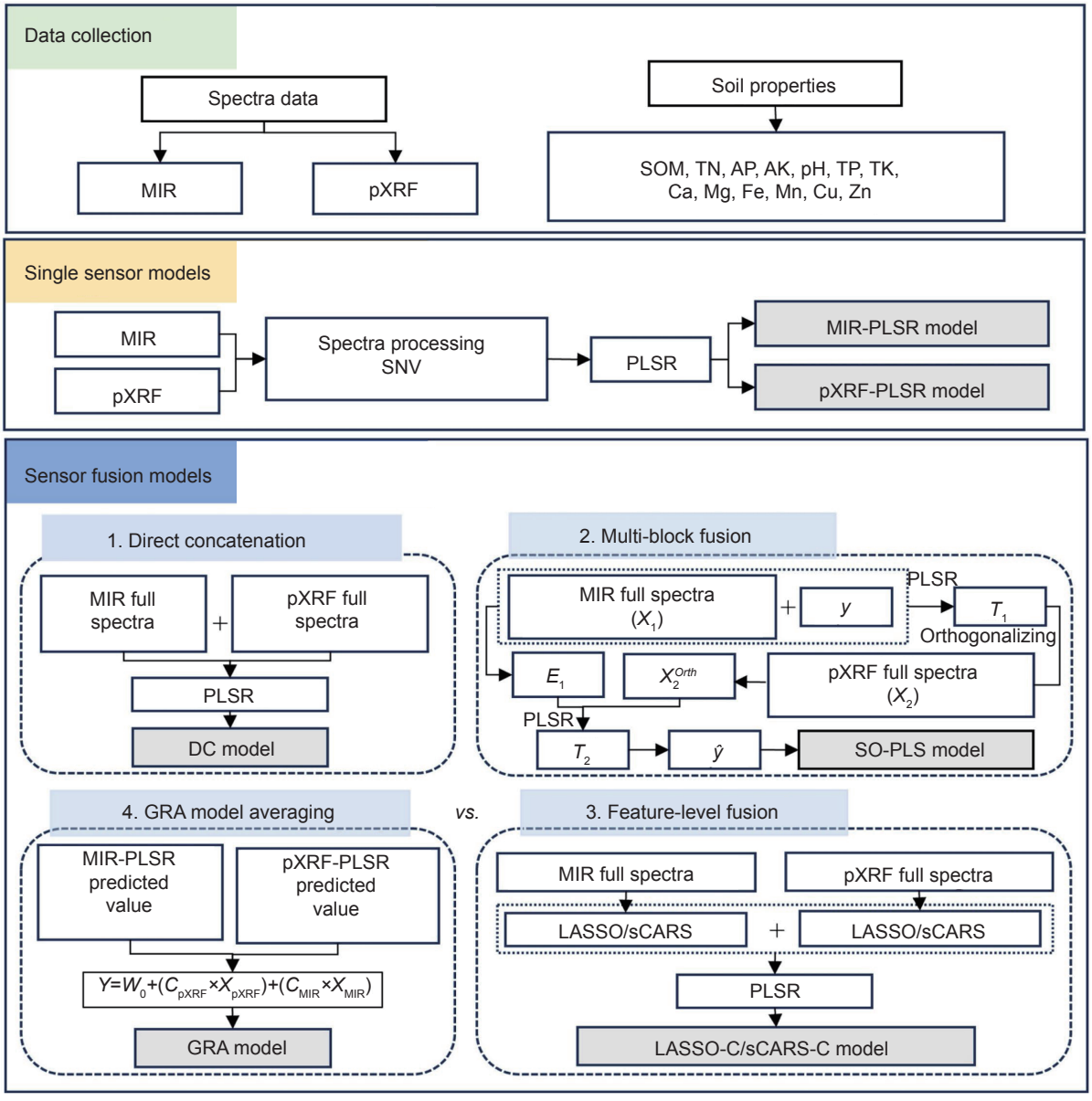


Fig. 2 Schematic diagram of four sensor fusion and modelling approaches using mid-infrared (MIR) and portable X-ray fluorescence (pXRF) spectra. SOM, soil organic matter; TN, total nitrogen; AP, available phosphorus; AK, available potassium; TP, total phosphorus; TK, total potassium; PLSR, partial least squares regression; SNV, standard normal variate; DC, direct concatenation; SO-PLS, sequential orthogonal partial least squares; GRA, Granger-Ramanathan model averaging; sCARS, stability competitive adaptive reweighted sampling; LASSO, least absolute shrinkage and selection operator algorithms.

Terhoeven-Urselmans *et al.* 2010; Johnson *et al.* 2019).

2.7. Data processing and statistics

The spectral preprocessing, sample set partitioning, sCARS variable selection, PLSR, and SO-PLS analyses were performed using Matlab version 2023a (The MathWorks, Natick, MA, USA) on the Windows 11 operating system. The SNV and PLSR implementations utilized code developed by Li *et al.* (2018), available at http://www.libpls.net/. Sample set partitioning was executed using the "randsel.m" function from the IOS package (available at https://github.com/zmzhang/IOS). GRA code was obtained from https://github.com/lpoz/GRA. SO-PLS code was obtained from https://www.chem.uniroma1.it/romechemometrics/research/algorithms/so-pls/.

3. Results

3.1. Soil properties and interpretation of spectral features

The descriptive statistics of the selected soil properties (SOM, TN, AP, AK, pH, TP, TK, Ca, Mg, Fe, Mn, Cu, and Zn) concentrations for the study area demonstrate data variability (Table 1). The variations in soil properties were categorized into three groups based on their coefficients of variation (CV) values: AP and AK (>50%), TN, SOM, Ca, Mg, Cu, Zn, and TP (20–50%), and Mn, Fe, TK, and pH (<20%). Low variability in soil properties typically results in unsuccessful regressions. The skewness of soil properties ranged from moderately negative (skewness=–1.32) to highly positive (skewness=4.59). SOM, Fe, and Mn exhibited flat-topped distributions (kurtosis<0), while the remaining ten indicators showed sharp-peaked distributions (kurtosis>0). Soil pH varied from 6.40 to 8.51, with most samples showing pH values above 7.50, indicating alkaline or strongly alkaline conditions. The soluble salts in saline-alkali soils primarily consist of calcium, magnesium, and potassium, typically present in higher concentrations.

Fig. 3-A and B present the raw MIR and pXRF spectra of soil samples. The MIR spectra of soil reveal mineral and organic information. The MIR spectra covered the wavenumber ranges from 4,000 to 600 cm^{-1}, with absorbance

Table 1 Descriptive statistics of the soil properties measured by chemical analysis

Soil properties[1)]	Number	Minimum	Maximum	Mean	Median	Standard deviation	Kurtosis	Skewness	Coefficient of variation (%)
SOM (g kg^{-1})	295	3.93	36.77	17.71	17.13	7.47	–0.57	0.35	42.18
TN (g kg^{-1})	280	0.29	3.42	1.17	1.08	0.52	0.97	0.89	44.44
pH	283	6.40	8.51	7.90	7.94	0.29	3.76	–1.32	3.67
AP (mg kg^{-1})	276	1.38	500.50	39.78	24.28	54.75	30.46	4.59	137.63
AK (mg kg^{-1})	275	32.00	786.00	159.24	125.00	104.79	4.99	1.75	65.81
TP (mg kg^{-1})	233	406.18	1,717.96	834.10	798.63	253.98	2.73	1.14	30.45
TK (mg kg^{-1})	255	10,729.24	20,023.27	15,504.69	15,629.28	1,410.29	0.41	–0.14	9.10
Ca (mg kg^{-1})	211	1,757.79	35,716.36	16,231.16	16,779.59	6,180.86	0.12	–0.26	38.08
Mg (mg kg^{-1})	255	495.18	5,150.57	1,724.43	1,600.03	727.18	1.63	1.00	42.17
Fe (mg kg^{-1})	217	19,301.76	36,353.93	27,196.03	27,077.73	3,516.54	–0.26	0.09	12.93
Mn (mg kg^{-1})	209	421.22	751.48	554.05	546.29	72.74	–0.36	0.40	13.13
Zn (mg kg^{-1})	258	31.53	207.03	73.14	70.59	19.96	10.77	2.27	27.29
Cu (mg kg^{-1})	258	0.65	31.54	12.35	12.15	5.62	7.52	1.48	45.51

[1)] SOM, soil organic matter; TN, total nitrogen; AP, available phosphorus; AK, available potassium; TP, total phosphorus; TK, total potassium.

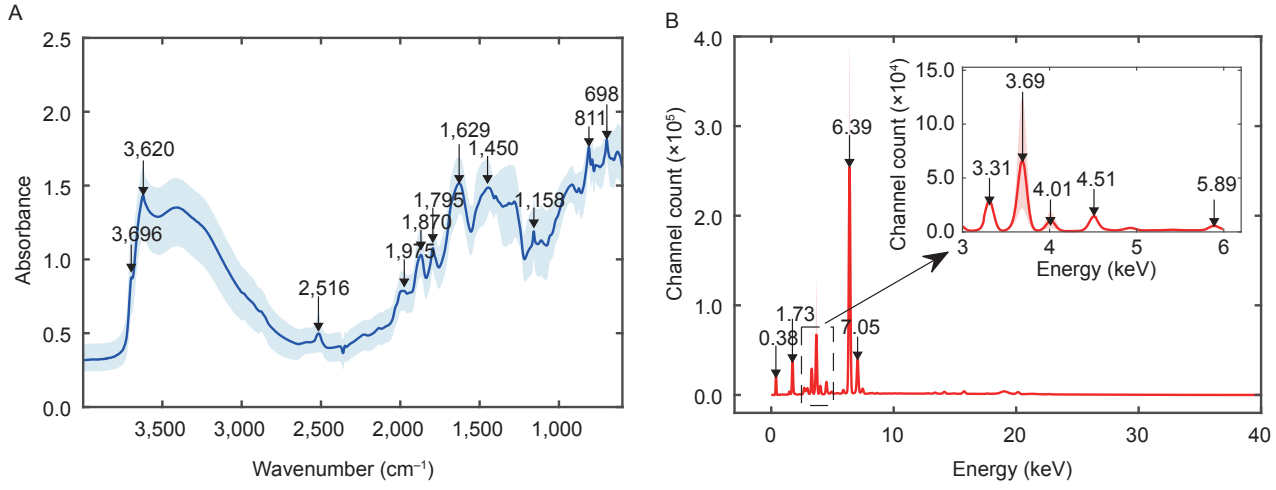


Fig. 3 Raw mid-infrared (MIR) (A) and portable X-ray fluorescence (pXRF) (B) spectra of soil samples.

ranges from 0.24 to 1.94. The averaged MIR spectra display several identifiable peaks (Fig. 3-A). The sharp absorption peaks at about 3,696 and 3,620 cm^{-1} were generated by O-H stretching vibrations. These absorption peaks indicate the presence of clay minerals (Ni *et al.* 2019). A broad band spanning 3,500–3,000 cm^{-1} corresponds to the stretching vibration of O-H from water, alcohols, carboxylic acids, phenols, and the stretching vibration of N-H from amides (Peltre *et al.* 2014). The twin peaks at 2,920 and 2,850 cm^{-1} are superimposed on the broad stretching vibration of O-H band, typically attributed to the stretching vibration of C-H of aliphatic methyl and methylene (Madhavan *et al.* 2016). The peak near 1,629 cm^{-1} results from vibrations of C=C, C=O, and O-H bonds. Furthermore, characteristic peaks corresponding to Si-O vibrations appear at 1,975, 1,870, 1,158, 811, and 698 cm^{-1}. Distinct characteristic absorption peaks from carbonates occur in the bands at 2,516, 1,795, and 1,450 cm^{-1}, due to the high carbonate content in this soil type (Ni *et al.* 2019).

In pXRF analysis, the count rates represent the emitted spectral intensity at each photon energy, forming the foundation for quantitative analysis and the built-in algorithm (Gholizadeh *et al.* 2021). The raw pXRF spectra (Fig. 3-B) exhibited a prominent peak at 6.40 keV and a minor peak around 7.05 keV for Fe. For Ca, two peaks were observed at 3.69 and 4.01 keV. For K, a minor peak appeared at 3.31 keV. For Mn, a small peak was detected around 5.89 keV, along with minor peaks for Sc (0.38 keV), Al (1.49 keV), Si (1.73 keV), Ar (2.96 keV), Ti (4.51, 4.92 keV) and Ni (7.46 keV), which were beyond the scope of this study. The emission lines of P and Mg were too weak to be detected in this region, likely due to the soil matrix complexity generating high background noise that obscured their peak signals.

3.2. Predicting soil properties using single sensor models

As shown in Table 2, MIR demonstrated good predictive performance for Ca, SOM, TN, Fe, and pH (Rp^2≥0.78, RPDp≥2.13). For AK, Mn, and AP, the model generated moderate predictions (0.70≤Rp^2≤0.75, 1.83≤RPDp≤1.98). However, the prediction performance was poor for TK, Mg, Zn, and Cu. Similarly, the pXRF-PLSR model provided good predictions for Ca, Fe, AP, and Mn (Rp^2≥0.80, RPDp≥2.22) (Table 3) and yielded moderate predictions for TP, TN, Zn, SOM, Mg, AK, and pH (0.67≤Rp^2≤0.75, 1.76≤RPDp≤1.99). However, the prediction accuracy remained poor for TK and Cu. Neither single sensor model of MIR or pXRF achieved good or moderate predictions for TP, AK, Zn, Mg, TK, and Cu (RPDp<2.00).

Table 2 Prediction accuracy of different soil properties using the PLSR model with MIR spectra

Soil properties[1)]	Calibration set			Validation set		
	Rc^2	RMSEc	RPDc	Rp^2	RMSEp	RPDp
Ca (mg kg^{-1})	0.87	2,090.39	2.78	0.89	2,275.94	2.99
SOM (g kg^{-1})	0.86	2.74	2.71	0.88	2.58	2.95
TN (g kg^{-1})	0.88	0.18	2.92	0.88	0.18	2.87
Fe (mg kg^{-1})	0.76	1,653.45	2.06	0.80	1,661.17	2.24
pH	0.87	0.10	2.78	0.78	0.14	2.13
AK (mg kg^{-1})	0.84	39.17	2.57	0.75	57.38	1.98
Mn (mg kg^{-1})	0.78	34.43	2.13	0.73	36.90	1.94
AP (mg kg^{-1})	0.79	22.60	2.27	0.70	34.80	1.83
TP (mg kg^{-1})	0.50	168.91	1.41	0.53	194.86	1.46
TK (mg kg^{-1})	0.36	1,067.22	1.25	0.43	1,165.98	1.33
Mg (mg kg^{-1})	0.26	622.39	1.16	0.35	587.29	1.25
Zn (mg kg^{-1})	0.34	15.09	1.24	0.27	19.32	1.17
Cu (mg kg^{-1})	0.25	4.52	1.15	0.24	5.59	1.14

[1)] SOM, soil organic matter; TN, total nitrogen; AP, available phosphorus; AK, available potassium; TP, total phosphorus; TK, total potassium.

Table 3 Prediction accuracy of different soil properties using the PLSR model with pXRF spectra

Soil properties[1)]	Calibration set			Validation set		
	Rc^2	RMSEc	RPDc	Rp^2	RMSEp	RPDp
Ca (mg kg^{-1})	0.83	2,409.04	2.41	0.88	2,381.40	2.86
Fe (mg kg^{-1})	0.81	1,484.86	2.30	0.84	1,467.64	2.53
AP (mg kg^{-1})	0.92	14.20	3.60	0.81	27.98	2.29
Mn (mg kg^{-1})	0.84	28.85	2.53	0.80	32.30	2.22
TP (mg kg^{-1})	0.86	87.38	2.81	0.75	143.22	1.99
TN (g kg^{-1})	0.88	0.18	2.96	0.74	0.26	1.96
Zn (mg kg^{-1})	0.78	8.68	2.18	0.74	11.47	1.98
SOM (g kg^{-1})	0.87	2.63	2.86	0.71	4.12	1.84
Mg (mg kg^{-1})	0.68	411.24	1.75	0.70	397.87	1.83
AK (mg kg^{-1})	0.72	53.10	1.88	0.69	63.10	1.80
pH	0.72	0.15	1.90	0.67	0.17	1.76
TK (mg kg^{-1})	0.54	908.83	1.48	0.61	975.76	1.60
Cu (mg kg^{-1})	0.60	3.22	1.73	0.60	4.06	1.58

[1)] SOM, soil organic matter; TN, total nitrogen; AP, available phosphorus; AK, available potassium; TP, total phosphorus; TK, total potassium.

3.3. Predicting soil properties using sensor fusion methods

Predictions using direct concatenation models Full spectra of MIR and pXRF were concatenated into a single dataset, upon which PLSR was implemented to develop prediction models for various soil properties. Notably, compared to the single sensor models, the DC models' prediction accuracy for AK, TK, Mg, and Zn improved significantly to good levels (0.75≤Rp^2≤0.80, 2.00≤RPDp≤2.26) (Table 4). Additionally, DC models demonstrated good predictions for Ca, SOM, TN, Fe, AP, AK, pH, Mg, Mn, Zn, and TK (0.75≤Rp^2≤0.89, 2.00≤RPDp≤3.05).

Predictions using multi-block fusion models In SO-PLS, the MIR full spectra constituted the first block, and the pXRF full spectra formed the second block. The SO-PLS model demonstrated well in predicting SOM, TN, AP, AK, Ca, Fe, Mn, Zn, and pH (0.75≤Rp^2≤0.89, 2.03≤RPDp≤3.03). Compared to the optimal predictions of single sensor models, the SO-PLS model substantially enhanced the prediction accuracy for AP, AK, TK, and Zn.

Predictions using feature-level fusion models Two feature selection methods were utilized in feature-level fusion, with results indicating LASSO-C superior to sCARS-C (Appendix A). Consequently, LASSO-C, identified as the optimal feature-level fusion method, was compared with other

Table 4 Prediction accuracy of different soil properties using four sensor fusion strategies[1)]

Soil properties[2)]	DC			SO-PLS			LASSO-C			GRA		
	Rp^2	RMSEp	RPDp	Rp^2	RMSEp	RPD	Rp^2	RMSEp	RPDp	Rp^2	RMSEp	RPD
Ca (mg kg^{-1})	0.89	2,238.18	3.05	0.89	2,249.54	3.03	0.88	2,369.12	2.90	0.90	2,184.13	3.14
SOM (g kg^{-1})	0.88	2.59	2.94	0.88	2.63	2.89	0.88	2.68	2.86	0.87	2.76	2.79
AP (mg kg^{-1})	0.84	25.33	2.53	0.88	22.15	2.90	0.89	21.37	3.01	0.86	24.89	2.59
TN (g kg^{-1})	0.87	0.18	2.84	0.87	0.19	2.74	0.86	0.19	2.70	0.88	0.18	2.87
Fe (mg kg^{-1})	0.86	1,401.58	2.65	0.85	1,451.80	2.56	0.84	1,486.97	2.53	0.85	1,423.94	2.63
AK (mg kg^{-1})	0.80	50.19	2.26	0.78	53.27	2.13	0.82	48.47	2.37	0.82	48.38	2.37
Mn (mg kg^{-1})	0.75	35.37	2.03	0.79	34.00	2.19	0.82	31.67	2.37	0.81	31.22	2.32
pH	0.79	0.14	2.20	0.78	0.14	2.14	0.75	0.15	2.02	0.82	0.13	2.34
Zn (mg kg^{-1})	0.75	11.18	2.03	0.75	11.18	2.03	0.80	9.88	2.32	0.78	10.65	2.14
TP (mg kg^{-1})	0.67	161.81	1.77	0.73	143.98	1.98	0.80	128.47	2.24	0.80	129.21	2.23
Mg (mg kg^{-1})	0.76	358.76	2.03	0.69	400.73	1.80	0.70	397.32	1.85	0.74	372.68	1.96
TK (mg kg^{-1})	0.75	775.96	2.00	0.72	818.83	1.90	0.73	810.04	1.93	0.73	829.01	1.89
Cu (mg kg^{-1})	0.48	4.61	1.39	0.52	4.43	1.45	0.60	4.06	1.59	0.65	3.87	1.67

[1)] DC, direct concatenation; SO-PLS, sequential orthogonalized partial least square; LASSO-C, fusion of spectral features selected by the least absolute shrinkage and selection operator; GRA, Granger-Ramanathan model averaging.
[2)] SOM, soil organic matter; AP, available phosphorus; TN, total nitrogen; AK, available potassium; TP, total phosphorus; TK, total potassium.

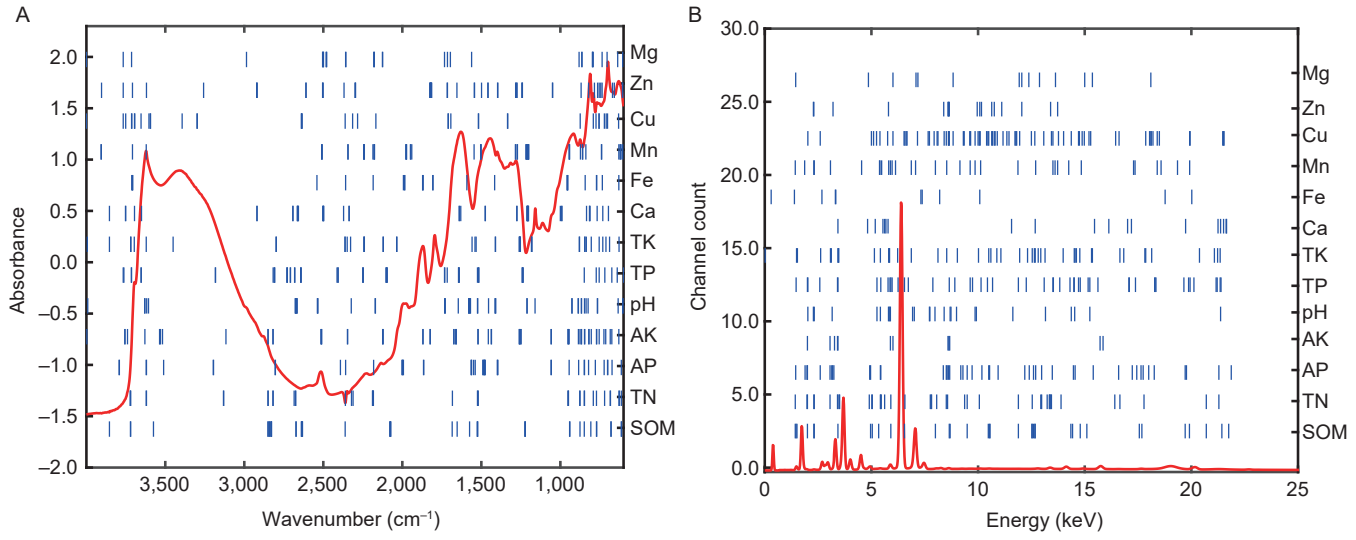


Fig. 4 Distribution of spectral variables selected by the LASSO algorithm from MIR (A) and pXRF (B) spectra. TK, total potassium; TP, total phosphorus; AK, available potassium; AP, available phosphorus; TN, total nitrogen; SOM, soil organic matter. The red curves represent the mean spectra after SNV preprocessing.

sensor fusion strategies. The full MIR spectral comprises 2,380 variables, and the full pXRF spectral contains 2,048 variables. Applying the LASSO algorithm as a variable selection method significantly reduced the variable number in MIR and pXRF spectra for soil property prediction. Fig. 4 illustrates the distribution of selected spectral features in the optimal feature-level fusion model (LASSO-C), which achieved the best performance across ten repeated random sample splits. The spectral bands selected by LASSO for SOM, TN, AP, AK, pH, TP, TK, Ca, Mg, Fe, Mn, Cu, and Zn numbered 46, 41, 48, 66, 51, 56, 45, 47, 34, 26, 56, 38, and 56 in the MIR range, respectively, while in the pXRF range they numbered 36, 44, 46, 12, 27, 51, 48, 19, 14, 11, 34, 75, and 18, respectively.

The LASSO-C models demonstrated good predictions for AP, Ca, SOM, TN, Fe, AK, Mn, Zn, TP, and pH (Rp^2≥0.75, RPDp≥2.02). Additionally, the LASSO-C models further improved the prediction accuracy of AP, AK, TP, Zn, Mn, and Cu compared to DC models, with Rp^2 improvements ranging from 0.02 to 0.15. Utilizing the "cor" function from the stats package in R, Pearson correlation coefficients (r) between soil properties (AP, AK, Zn, TP, Mn, Cu) and soil spectra (MIR or pXRF) were calculated, with LASSO-selected feature variables highlighted in Fig. 5 for detailed analysis within the LASSO-C model. For AP, AK, TP, Zn, and Cu, the r values for pXRF exceeded those of MIR, with more negative correlations observed in the MIR region. For Mn, the r ranges were comparable in both regions, spanning –0.81 to 0.73. LASSO effectively identified the characteristic emission lines of TP, Mn, and Zn in pXRF, specifically P-Kα at 2.02 keV (r=0.60), Mn-Kα at 5.89 keV (r=0.35) and Zn-Kα at 8.63 keV (r=0.67). The significant MIR bands selected

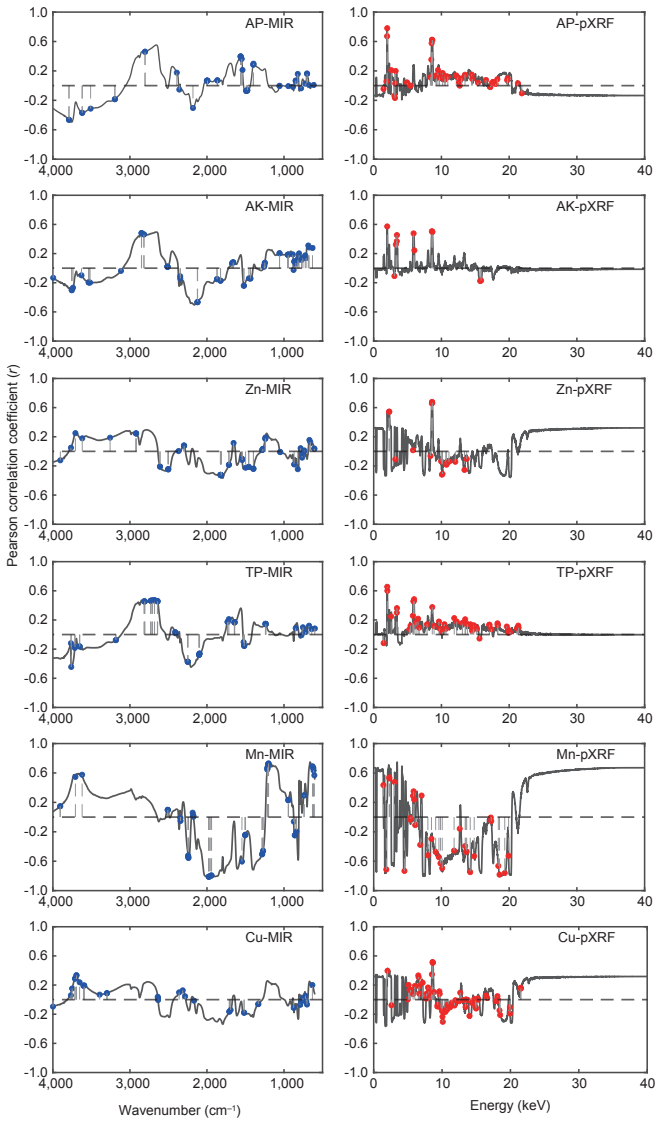


Fig. 5 Pearson correlation coefficients (*r*) between available phosphorus (AP), available potassium (AK), total phosphorus (TP), manganese (Mn), zinc (Zn), and copper (Cu) with soil MIR and pXRF spectra. The marked points represent the feature variables selected by the LASSO algorithm.

by LASSO from the MIR region, including 1,217–1,200 cm^{-1} (r=0.70), 3,620 cm^{-1} (r=0.58), and 740 cm^{-1} (r=0.30), contributed to Mn prediction. For AK, LASSO selected 94 variables from the MIR and pXRF spectral regions, with 50 variables showing r values above 0. Thus, the LASSO-based feature fusion model (LASSO-C) efficiently identified feature variables for AP, AK, TP, Mn, and Zn. For Cu, LASSO failed to extract relevant emission lines from pXRF. Fe ions in soil absorb Cu X-rays, thereby reducing its measured concentration in soil (Martín Peinado *et al.* 2010).

Predictions using GRA models Compared to other fusion strategies, GRA notably improved Cu prediction performance, achieving a moderate level (Rp^2=0.65, RMSEp=3.87 mg kg^{-1}, RPDp=1.67). Furthermore, except for SOM and TN where comparable accuracy was achieved, the GRA models outperformed the single sensor models in predicting the remaining 11 soil properties. Overall, the GRA models exhibited good predictive performance for the selected soil properties, except for Mg, TK, and Cu (0.78≤Rp^2≤0.90, 2.14≤RPDp≤3.14).

3.4. Best strategy for predicting soil properties

Fig. 6 displays the scatter plots of measured versus predicted values for the optimal strategies of selected soil properties, based on one of ten repeated random sample splits. Generally, accurate predictions were achieved for all soil properties except Cu, which reached a reasonable level. However, the principle of parsimony warrants consideration when selecting the optimal approach-specifically, model predictors should be limited to those whose elimination would significantly increase estimation error. Since a primary objective of proximal sensor usage is to improve the time and cost efficiency of collecting high-resolution soil information, sensor fusion should be recommended over single sensor use only when it provides significant model accuracy improvement. Therefore, sensor fusion studies should report the relative improvement in estimation accuracy compared with the use of an optimal single spectrometer (Greenberg *et al.* 2023). Consequently, MIR-PLSR model was recommended for SOM, TN, and Ca prediction, while pXRF-PLSR model was preferred for Fe and Mn. The LASSO-C model was selected for AP and Zn, the DC model for Mg and TK, and the GRA model for TP, pH, AK, and Cu.

4. Discussion

4.1. Prediction of soil properties from MIR and pXRF

In this study, MIR models demonstrated good prediction performance for SOM, TN, Ca, Fe, and pH (Rp^2≥0.78, RPDp≥2.13). The MIR region of soil spectra exhibits clearly distinguishable and well-defined absorption bands produced by minerals and organic matter (Dangal *et al.* 2019). The findings align with previous research indicating that MIR outperforms pXRF for properties related to soil organic composition (Xu *et al.* 2019; Greenberg *et al.* 2023). Xu *et al.* (2019) demonstrated that the GA-PLSR model of MIR predicted SOM and TN well (RPIQ>2.5), while the GA-PLSR model of pXRF was less effective. Similarly, Greenberg *et al.* (2023) reported superior validation performance of MIR for SOC and TN content prediction compared to pXRF spectra in a German arable field dataset. Furthermore, the MIR model achieved good predictions for Ca and Fe with R^2 values of 0.89 and 0.80, and RPD values of 2.99 and 2.24, respectively. These results correspond with Kandpal *et al.* (2022), who reported high prediction accuracy for Ca using MIR (R^2=0.92, RPD=3.66). Ji *et al.* (2016) achieved moderate accuracy in estimating soil Fe using MIR. Additionally, MIR models effectively predicted soil pH, primarily due to the covariation between soil pH and spectrally active components such as SOC and clay (Soriano-Disla *et al.* 2014).

Among the analyzed soil properties, pXRF models provided satisfactory predictions for Ca, Fe, Mn, and AP (Rp^2≥0.80, RPDp≥2.22). pXRF spectra enable comprehensive characterization of soil elemental composition, thereby complementing MIR. High prediction accuracy for Ca

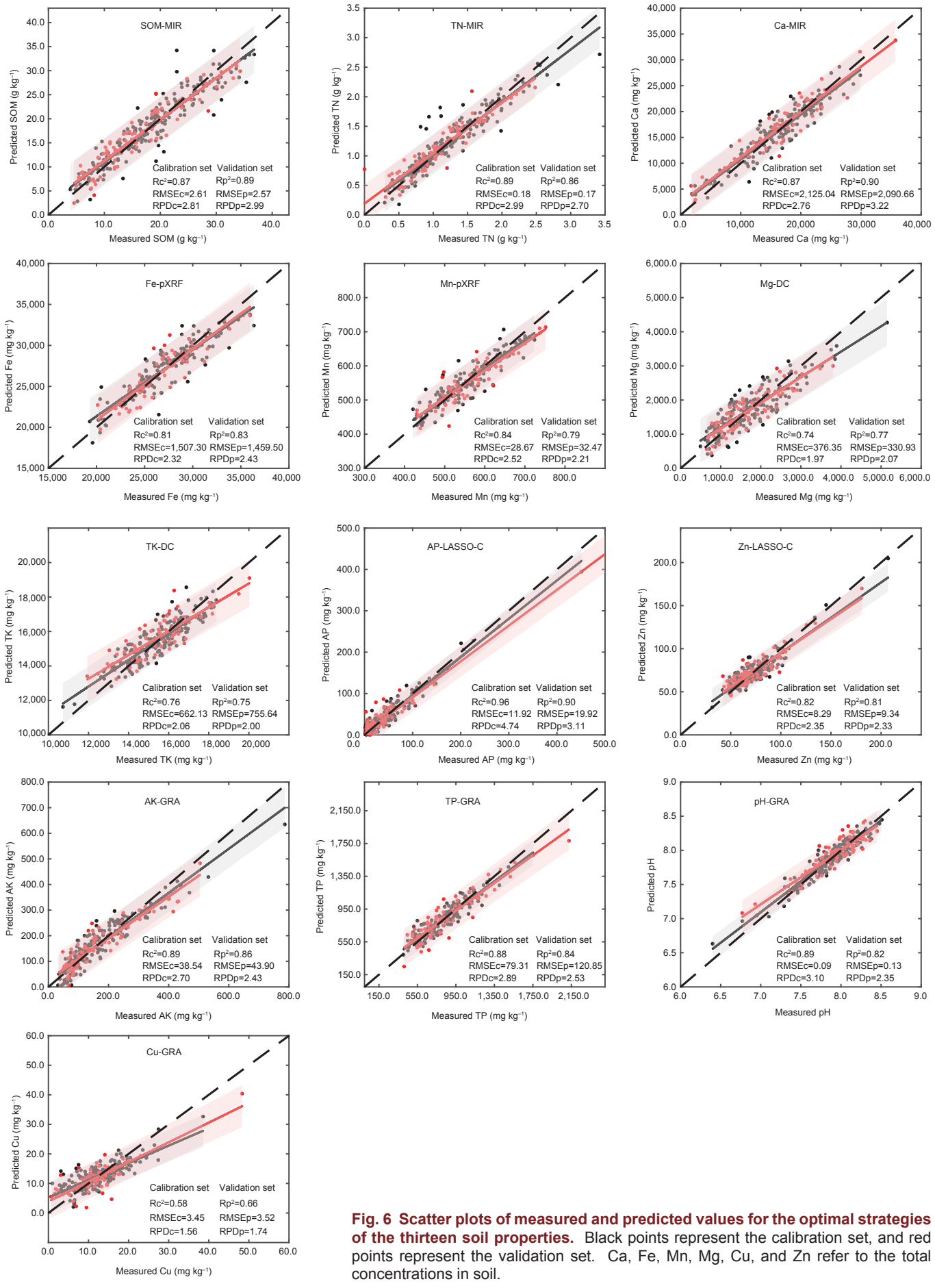


Fig. 6 Scatter plots of measured and predicted values for the optimal strategies of the thirteen soil properties. Black points represent the calibration set, and red points represent the validation set. Ca, Fe, Mn, Mg, Cu, and Zn refer to the total concentrations in soil.

using pXRF has been widely reported (O'Rourke *et al.* 2016a; Andrade *et al.* 2019), largely due to Ca's relatively high fluorescence emission intensity. The current study identified two Ca peaks at 3.69 and 4.01 keV. Greenberg *et al.* (2023) corroborated these findings, noting a strong Ca peak between 3 and 5 keV, with their pXRF model achieving good prediction accuracy (R^2=0.71, RMSEp=4,664.30 mg kg^{-1}, RPD=1.81). The accurate quantification of Fe and Mn resulted from their characteristic peaks in the pXRF spectra. Fe exhibited a prominent peak at 6.40 keV and a minor peak at approximately 7.05 keV, while Mn showed a small peak near 5.89 keV. Gholizadeh *et al.* (2021) demonstrated reliable Mn assessment using single pXRF data (R^2=0.80 and RMSEp=488.00 mg kg^{-1}). The successful P prediction in this study was likely attributed to its correlation with other elements. According to Tavares *et al.* (2023), while few studies showed satisfactory AP prediction, most successful cases occurred in tropical soils. These findings suggest that combining MIR and pXRF sensors expands the range of quantifiable soil properties.

4.2. Effectiveness of multi-block fusion strategy

Compared to single sensor models, multi-block fusion models significantly enhanced the prediction accuracy of AP, AK, TP, and TK, as demonstrated by Rp^2 values across ten repeated random splits (Appendices B and C). In multi-block fusion, SO-PLS identifies complementary information in different data blocks and integrates the blocks from two sensors to create a comprehensive dataset (Song *et al.* 2024). Specifically, SO-PLS utilizes the valid information from different sensor data to generate accurate estimation results. Currently, the SO-PLS model has been widely applied in food chemical component analysis (Biancolillo and Næs 2021) and has recently been adopted for soil sensor fusion analysis. Greenberg *et al.* (2023) demonstrated that SO-PLS or VIP-SO-PLS were effectiveness for fusing MIR and XRF to quantify pH, SOC, TN, P, K, Mg, Ca, and MC, with notable improvements in pH, P, and Ca prediction. Additionally, when fusing Vis-NIR and pXRF to predict soil TN, combining CARS with the SO-PLS fusion method and geographically stratified modeling strategies proved effective in improving soil TN estimation accuracy (RMSEp=0.18 g kg^{-1}, LCCC=0.86, RPIQ=3.19) (Song *et al.* 2024). While this study demonstrated the significance of the SO-PLS algorithm, its improvement in soil element analysis models was limited, showing decreased prediction accuracy for SOM, TN, TP, Mn, Mg, and Cu. This limitation may relate to first block selection. Further research is needed to explore the potential of the SO-PLS algorithm in soil property estimation using MIR and pXRF sensor fusion.

4.3. The importance of feature selection for sensor fusion

The present study revealed that feature-level fusion models achieved higher prediction accuracy for AP, AK, TP, Zn, Mn, and Cu compared to DC models (Appendices B and C). This finding confirms the significance of variable selection, in which the DC model concatenated full spectral of MIR and pXRF, while the feature-level fusion model extracted feature bands from them before concatenation. The feature bands selected by LASSO for AP, AK, TP, Zn, Mn, and Cu comprised only 1.09–2.77% of the full MIR and 0.54–3.66% of the full pXRF, respectively. The improved prediction performance of feature-level fusion models stems from LASSO's efficient extraction of relevant property information from MIR and pXRF spectra. Notably, the LASSO-C model achieved optimal prediction accuracy among the four fusion strategies for AP, AK, Zn, and Mn. These results align with previous research demonstrating the effectiveness of variable selection in improving sensor fusion estimation. LASSO effectively filters features from full-spectrum data and enhances model prediction accuracy through dimensionality reduction (Li *et al.* 2019; Ning *et al.* 2022). When applied to variable selection from individual sensors, the LASSO algorithm substantially reduced the Vis-NIR and MIR spectral variables used in modeling six soil nutrients. Consequently, the LASSO-C fusion method produced superior predictions for TN, TP, AN, AP, and AK compared to single sensor models (Li *et al.* 2024). Variable selection techniques thus prove crucial for effective feature-level sensor fusion, as different selection principles distinctly influence accuracy enhancement and model simplification.

4.4. The benefits of GRA strategy

For Ca, AK, TP, pH, and Cu, the GRA models demonstrated the highest prediction accuracy among the four fusion strategies (Appendices B and C). GRA, as a model averaging algorithm, integrates multiple models into a single prediction, ensuring no individual model dominates the final predicted value, thereby reducing the probability of anomalous predictions. Consequently, previous research has shown that GRA provides reliable prediction results and has been extensively utilized in multi-sensor fusion as an effective approach (O'Rourke *et al.* 2016b; Shrestha *et al.* 2022). For example, Javadi *et al.* (2021) evaluated various data fusion methods using pXRF and Vis-NIR spectra to predict key soil fertility properties, including pH, OC, Mg, and Ca. Their results indicated that the GRA model enhanced the prediction performance across all soil properties. The GRA model also improved the prediction accuracy of As, Cd, Cr, and Ni compared to single sensor models (Vis-NIR or pXRF) (Xu *et al.* 2020). Based on both relative improvements in RMSE and RPIQ, model averaging using Vis-NIR and XRF proved suitable for predicting soil pH, soil C, soil texture, CEC, ex-K, ex-Ca, and total elements K, Mg, Co, Cr, and Mn (O'Rourke *et al.* 2016b).

5. Conclusion

In the present study, single sensor models of MIR or pXRF could not provide accurate estimation for soil micronutrients (e.g., Mg, Cu, and Zn), TK, AK, and TP. Therefore, different sensor fusion strategies were implemented to address this limitation. The DC model proved optimal for predicting Mg and TK, LASSO-C for AP and Zn, and GRA for pH, TP, AK, and Cu. For SOM, TN, and Ca, MIR spectroscopy remained the preferred method, while pXRF demonstrated superior performance for Ca, Fe, and Mn. Additionally, the superior

accuracy of feature-level fusion models for AP, AK, TP, Zn, Mn and Cu compared to DC models validated the effectiveness of feature band selection. This study established that the fusion of MIR and pXRF represents a viable approach for accurate estimation of soil micronutrients and available nutrients. However, the dataset's limited geographic diversity and soil types may restrict the models' generalizability to other regions or soil conditions. Future research should evaluate these fusion strategies under real-field scenarios.

Acknowledgements

This paper was supported by the National Key Research and Development Program of China (2023YFD1900104).

Declaration of competing interest

The authors declare that they have no conflict of interest.

Declaration of generative AI and AI-assisted technologies in the writing process

The authors declare that they did not use AI in the preparation and writing of this manuscript.

Appendices associated with this paper are available at https://doi.org/10.1016/j.jia.2025.09.028

References

Andrade R, Faria W M, Silva S H G, Chakraborty S, Weindorf D C, Mesquita L F, Guilherme L R G, Curi N. 2019. Prediction of soil fertility via portable X-ray fluorescence (pXRF) spectrometry and soil texture in the Brazilian Coastal Plains. *Geoderma*, **357**, 113960.

Antonangelo J, Zhang H. 2024. Assessment of portable X-ray fluorescence (pXRF) for plant-available nutrient prediction in biochar-amended soils. *Scientific Reports*, **14**, 20377.

Bao S D. 1981. *Soil and Agricultural Chemistry Analysis*. China Agricultural Press, Beijing. pp.129–159. (in Chinese)

Barra I, Haefele S M, Sakrabani R, Kebede F. 2021. Soil spectroscopy with the use of chemometrics, machine learning and pre-processing techniques in soil diagnosis: Recent advances — A review. *TrAC Trends in Analytical Chemistry*, **135**, 116166.

Biancolillo A, Næs T. 2019. Chapter 6 — The sequential and orthogonalized PLS regression for multiblock regression: theory, examples, and extensions. in: Cocchi M, ed., *Data Handling in Science and Technology*, *Data Fusion Methodology and Applications.* Elsevier, Oxford, UK. pp. 157–177.

Briedis C, Baldock J, de Moraes Sá J C, Dos Santos J B, Milori D M B P. 2020. Strategies to improve the prediction of bulk soil and fraction organic carbon in Brazilian samples by using an Australian national mid-infrared spectral library. *Geoderma*, **373**, 114401.

Chang C, Laird D A, Mausbach M J, Hurburgh C R. 2001. Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. *Soil Science Society of America Journal*, **65**, 480–490.

Dangal S, Sanderman J, Wills S, Ramirez-Lopez L. 2019. Accurate and precise prediction of soil properties from a large mid-infrared spectral library. *Soil Systems*, **3**, 11.

Gholizadeh A, Coblinski J A, Saberioon M, Ben-Dor E, Drábek O, Demattê J A M, Borůvka L, Němeček K, Chabrillat S, Dajčl J. 2021. Vis-NIR and XRF data fusion and feature selection to estimate potentially toxic elements in soil. *Sensors* (Basel), **21**, 2386.

Granger C W, Ramanathan R. 1984. Improved methods of combining forecasts. *Journal of Forecasting*, **3**, 197–204.

Greenberg I, Vohland M, Seidel M, Hutengs C, Bezard R, Ludwig B. 2023. Evaluation of mid-infrared and X-ray fluorescence data fusion approaches for prediction of soil properties at the field scale. *Sensors* (Basel), **23**, 662.

Hong Y, Munnaf M A, Guerrero A, Chen S, Liu Y, Shi Z, Mouazen A M. 2022. Fusion of visible-to-near-infrared and mid-infrared spectroscopy to estimate soil organic carbon. *Soil & Tillage Research*, **217**, 105284.

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.

Ji W J, Adamchuk V I, Biswas A, Dhawale N M, Sudarsan B, Zhang Y K, Viscarra Rossel R A, Shi Z. 2016. Assessment of soil properties in situ using a prototype portable MIR spectrometer in two agricultural fields. *Biosystems Engineering*, **152**, 14–27.

Johnson J M, Vandamme E, Senthilkumar K, Sila A, Shepherd, K D, Saito K. 2019. Near-infrared, mid-infrared or combined diffuse reflectance spectroscopy for assessing soil fertility in rice fields in sub-Saharan Africa. *Geoderma*, **354**, 113840.

Kandpal L M, Munnaf M A, Cruz C, Mouazen A M. 2022. Spectra fusion of mid-infrared (MIR) and X-ray fluorescence (XRF) spectroscopy for estimation of selected soil fertility attributes. *Sensors* (Basel), **22**, 3459.

Karami A, Moosavi A A, Pourghasemi H R, Ronaghi A, Ghasemi-Fasaei R, Vidal E, Lado M. 2024. Proximal sensing approach for characterization of calcareous soils using multiblock data analysis. *Geoderma Regional*, **36**, e00752.

Li H, Wang J W, Zhang J X, Liu T Q, Acquah G E, Yuan H M. 2022. Combining variable selection and multiple linear regression for soil organic matter and total nitrogen estimation by DRIFT-MIR spectroscopy. *Agronomy*, **12**, 638.

Li H D, Xu Q S, Liang Y Z. 2018. libPLS: An integrated library for partial least squares regression and linear discriminant analysis. *Chemometrics and Intelligent Laboratory Systems*, **176**, 34–43.

Li X L, Pan W J, Li D C, Gao W, Zeng R, Zheng G H, Cai K, Zeng Y T, Jiang C Y. 2024. Can fusion of Vis-NIR and MIR spectra at three levels improve the prediction accuracy of soil nutrients? *Geoderma*, **441**, 116754.

Li Y Q, Pan T H, Li H R, Zou X B. 2019. NIR spectral feature selection using lasso method and its application in the classification analysis. *Spectroscopy and Spectral Analysis*, **39**, 3809–3815. (in Chinese)

Lobsey C R, Viscarra Rossel R A, Roudier P, Hedley C B. 2017. Rs-local data-mines information from spectral libraries to improve local calibrations. *European Journal of Soil Science*, **68**, 840–852.

Madhavan D B, Kitching M, Mendham D S, Weston C J, Baker T G. 2016. Mid-infrared spectroscopy for rapid assessment of soil properties after land use change from pastures to Eucalyptus globulus plantations. *Journal of Environmental Management*, **175**, 67–75.

Martín Peinado F, Morales Ruano S, Bagur González M G, Estepa Molina C. 2010. A rapid field procedure for screening trace elements in polluted soil using portable X-ray fluorescence (PXRF). *Geoderma*, **159**, 76–82.

McBride M B. 2021. Estimating soil chemical properties by diffuse reflectance spectroscopy: Promise versus reality. *European Journal of Soil Science*, **73**, e13192.

Mukherjee S. 2022. Soil fertility and nutrient management. In: *Current Topics in Soil Science*. Springer, Cham. pp. 241–248.

Ni L S, Fang N F, Shi Z H, Tan W F. 2019. Mid-infrared spectroscopy tracing of channel erosion in highly erosive catchments on the Chinese Loess Plateau. *Science of the Total Environment*, **687**, 309–318.

Ning H W, Wang J W, Jiang H, Chen Q S. 2022. Quantitative detection of zearalenone in wheat grains based on near-infrared spectroscopy. *Spectrochimica Acta* (Part A: Molecular and Biomolecular Spectroscopy), **280**, 121545.

O'Rourke S M, Minasny B, Holden N M, McBratney A B. 2016a. Synergistic use of Vis-NIR, MIR, and XRF spectroscopy for the determination of soil geochemistry. *Soil Science Society of America Journal*, **80**, 888–899.

O'Rourke S M, Stockmann U, Holden N M, McBratney A B, Minasny B. 2016b. 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.

Peltre C, Bruun S, Du C W, Thomsen I K, Jensen L S. 2014. Assessing soil constituents and labile soil organic carbon by mid-infrared photoacoustic spectroscopy. *Soil Biology and Biochemistry*, **77**, 41–50.

Shi T R, Fu Z C, Miao X H, Lin F F, Ma J Y, Gu S Y, Li L, Wu C F, Luo Y M. 2024, Would it be better for partition prediction of heavy metal concentration in soils based on the fusion of XRF and Vis-NIR data? *Science of the Total Environment*, **908**,168381.

Shi X Y, Song J H, Wang H J, Lv X, Zhu YQ, Zhang W X, Bu W Q, Zeng L Y. 2022. Improving soil organic matter estimation accuracy by combining optimal spectral preprocessing and feature selection methods based on pXRF and Vis-NIR data fusion. *Geoderma*, **430**, 116301.

Shrestha G, Calvelo-Pereira R, Roudier P, Martin A P, Turnbull R E, Kereszturi G, Jeyakumar P, Anderson C W N. 2022. Quantification of multiple soil trace elements by combining portable X-ray fluorescence and reflectance spectroscopy. *Geoderma*, **409**, 115649.

Song J, Shi X, Wang H, Lv X, Zhang W, Wang J, Li T, Li W. 2024. Combination of feature selection and geographical stratification increases the soil total nitrogen estimation accuracy based on Vis-NIR and pXRF spectral fusion. *Computers and Electronics in Agriculture*, **218**, 108636.

Soriano-Disla J M, Janik L J, Viscarra Rossel R A, MacDonald L M, McLaughlin M J. 2014. The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. *Applied Spectroscopy Reviews*, **49**, 139–186.

Tavares T R, Minasny B, McBratney A, Cherubin M R, Marques G T, Ragagnin M M, Alves E E N, Padarian J, Lavres J, de Carvalho H W P. 2023. Estimating plant-available nutrients with XRF sensors: Towards a versatile analysis tool for soil condition assessment. *Geoderma*, **439**, 116701.

Tavares T R, Molin J P, Nunes L C, Wei M C F, Krug F J, de Carvalho H W P, Mouazen A M. 2021. Multi-sensor approach for tropical soil fertility analysis: Comparison of individual and combined performance of VNIR, XRF, and LIBS spectroscopies. *Agronomy*, **11**, 1028.

Terhoeven-Urselmans T, Vagen T, Spaargaren O, Shepherd K D. 2010. Prediction of soil fertility properties from a globally distributed soil mid-infrared spectral library. *Soil Science Society of America Journal*, **74**, 1792–1799.

Wang J W, Liu T Q, Zhang J X, Yuan H M, Acquah G E. 2022. Spectral variable selection for estimation of soil organic carbon content using mid-infrared spectroscopy. *European Journal of Soil Science*, **73**, e13267.

Wold S, Sjöström M, Eriksson L. 2001. PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, **58**, 109–130.

Wu M H, Lin N, Li G J, Liu H L, Li D L. 2022. Hyperspectral estimation of petroleum hydrocarbon content in soil using ensemble learning method and LASSO feature extraction. *Environmental Pollutants and Bioavailability*, **34**, 308–320.

Xu D D, Zhao R, Li S, Chen S C, 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.

Xu D Y, Chen S C, Xu H Y, 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.

Xue J, Zhang X L, Chen S C, Lu R, Wang Z, Wang N, Hong Y S, Chen X Y, Xiao Y, Ma Y X, Shi Z. 2023. The validity domain of sensor fusion in sensing soil quality indicators. *Geoderma*, **438**, 116657.

Zhang X L, Xue J, Chen S C, Zhuo Z Q, Wang Z, Chen X Y, Xiao Y, Shi Z. 2024. Improving model performance in mapping cropland soil organic matter using time-series remote sensing data. *Journal of Integrative Agriculture*, **23**, 2820–2841.

Zheng K Y, Li Q Q, Wang J J, Geng J P, Cao P, Sui T, Wang X, Du Y P. 2012. Stability competitive adaptive reweighted sampling (sCARS) and its applications to multivariate calibration of NIR spectra. *Chemometrics and Intelligent Laboratory Systems*, **112**, 48–54.

Zou Q, Wang J W, Yuan H M. 2025. The accurate estimation of soil available nutrients achieved by feature selection coupled with preprocessing based on MIR and pXRF fusion. *European Journal of Agronomy,* **168**, 127633.

Executive Editor-in-Chief Jinshui Wu
Managing Editor Lujuan Sun