



RESEARCH ARTICLE

Improving soil organic carbon prediction using visible and near infrared spectroscopy: A comparative evaluation of spectral preprocessing and modelling approaches under semi-arid conditions

C G Kusuma^{1*}, J Saralakumari¹, S Dharumarajan², A Sathish¹ & Umashankar N³

¹Department of Soil Science and Agricultural Chemistry, College of Agriculture, University of Agricultural Sciences (UAS), Gandhi Krishi Vignana Kendra (GKVK), Bengaluru 560 065, Karnataka, India

²ICAR-National Bureau of Soil Survey and Land Use Planning, Regional Centre, Bangalore 560 024, Karnataka, India

³Department of Agricultural Microbiology, College of Agriculture, University of Agricultural Sciences (UAS), Gandhi Krishi Vignana Kendra (GKVK), Bengaluru 560 065, Karnataka, India

*Correspondence email - kusumacg4@gmail.com

Received: 21 April 2025; Accepted: 06 July 2026; Available online: Version 1.0: 01 August 2026

Cite this article: Kusuma CG, Saralakumari J, Dharumarajan S, Sathish A, Umashankar N. Improving soil organic carbon prediction using visible and near infrared spectroscopy: A comparative evaluation of spectral preprocessing and modelling approaches under semi-arid conditions. Plant Science Today. 2026; 13(sp5): 1-14. <https://doi.org/10.14719/pst.15142>

Abstract

Accurate estimation of soil organic carbon (SOC) is essential for soil quality assessment, carbon management and sustainable agriculture, particularly in semi-arid regions characterised by high spatial variability. Visible and near infrared (Vis–NIR) spectroscopy has emerged as a rapid and non-destructive technique for SOC prediction; however, its performance is strongly influenced by spectral preprocessing and modelling approaches. Despite growing interest, systematic evaluation of preprocessing-model interactions for SOC prediction under semi-arid conditions in India remains limited. This study aimed to systematically evaluate the effect of different preprocessing techniques and modelling methods on SOC prediction using Vis–NIR spectral data from a semi-arid region of southern India. A total of 300 surface soil samples were collected and analysed for SOC using the Walkley–Black method. Spectral data (400–2400 nm) were preprocessed using raw spectra, Savitzky–Golay first and second derivatives (SG-1D and SG-2D), multiplicative scatter correction (MSC) and standard normal variate (SNV). Partial least squares regression (PLSR) and random forest (RF) models were developed using calibration (75 %) and validation (25 %) datasets. Results showed that spectral preprocessing significantly influenced model performance, with SG-1D and MSC providing consistent improvements by enhancing spectral features and reducing scattering effects. The RF model outperformed PLSR, with the best performance achieved using RF combined with MSC and SG-1D (validation $R^2 = 0.50$, RMSE = 0.27 %, RPIQ ≈ 1.66 –1.67). The findings indicate that optimal SOC prediction depends on the combined selection of preprocessing techniques and modelling approaches. This study provides a practical and reliable framework for improving soil spectroscopy applications in digital soil mapping and precision agriculture.

Keywords: machine learning models; prediction; semi-arid; soil organic carbon; spectral preprocessing; visible and near-infrared spectroscopy

Introduction

Soil organic carbon (SOC) is a fundamental component of terrestrial ecosystems and plays a crucial role in regulating soil fertility, ecosystem productivity and global carbon cycling (1). It represents one of the largest carbon pools in terrestrial environments and serves as a key indicator of soil quality, influencing nutrient availability, water retention and biological activity (2). Even small changes in SOC stocks can significantly affect atmospheric carbon dioxide concentrations and climate change processes, underscoring their importance for sustainable environmental monitoring and land management (3). In agricultural systems, SOC directly contributes to crop productivity and resilience, making its accurate assessment essential for precision agriculture and soil health management. Despite its significance, accurate estimation of SOC remains

challenging due to its high spatial variability and complex interactions with soil-forming factors such as climate, topography, parent material and land management (4). Conventional laboratory methods for SOC determination, including the Walkley-Black wet oxidation method and dry combustion techniques, are labour-intensive, time-consuming and require extensive sample preparation. Additionally, these methods involve the use of hazardous chemicals and are not well suited for large-scale or high-resolution soil monitoring (5). Consequently, there is an increasing demand for rapid, cost-effective and environmentally sustainable alternatives for SOC estimation.

Visible and near infrared (Vis–NIR) spectroscopy has emerged as a promising technique for rapid and non-destructive assessment of soil properties (6, 7). This technique measures soil

reflectance across a broad spectral range (approximately 400–2500 nm), capturing information related to organic matter, minerals and other soil constituents. Numerous studies have demonstrated the capability of Vis–NIR spectroscopy to predict a wide range of soil properties, including SOC, texture, moisture, pH and mineral composition, with moderate to high accuracy depending on dataset characteristics (6, 8, 9, 10). The method offers several advantages, such as minimal sample preparation, rapid data acquisition and suitability for large datasets, making it highly applicable in digital soil mapping and precision agriculture frameworks (11). The relationship between SOC and spectral reflectance is primarily governed by the absorption characteristics of organic compounds and their influence on soil colour and albedo. Increased SOC content generally leads to reduced reflectance due to its darker colour and strong absorption features in the Vis–NIR regions (12, 13). However, this relationship is often complex due to the influence of other soil properties such as mineral composition, moisture content and texture, which can capture nonlinear interactions between spectral signals and SOC (14). Therefore, extracting meaningful information from spectral data requires robust calibration models and appropriate spectral preprocessing techniques.

Spectral preprocessing plays a crucial role in improving the performance of predictive models by enhancing relevant spectral features and reducing noise and unwanted variability (15). Raw Vis–NIR spectra are often affected by baseline shifts, scattering effects and instrumental noise, which can obscure subtle absorption features associated with SOC (16). These distortions reduce the signal-to-noise ratio and limit the ability of models to capture meaningful relationships between spectral data and soil properties. Therefore, appropriate preprocessing is essential to improve spectral quality, enhance feature extraction and ultimately increase model accuracy. To address these issues, a wide range of preprocessing techniques are commonly applied. Scatter-correction methods include continuum removal, normalisation, standard normal variate (SNV) and multiplicative scatter correction (MSC), while spectral-derivative approaches include Savitzky-Golay (SG) and Norris-Williams derivatives (17). Previous studies have demonstrated that preprocessing can significantly improve SOC prediction accuracy by enhancing spectral sensitivity and reducing physical interferences (18, 19). However, the effectiveness of preprocessing techniques is often dataset-dependent and their interaction with different modelling approaches remains insufficiently explored under consistent experimental conditions (20, 21).

In addition to preprocessing, the choice of modelling approach plays a crucial role in determining prediction accuracy. Spectral data are typically related to laboratory-measured soil properties using multivariate statistical or machine learning techniques. A variety of statistical and machine learning models have been applied for SOC prediction using Vis–NIR spectroscopy, reflecting the need to capture both linear and complex nonlinear relationships between spectral data and soil properties. Among these, partial least squares regression (PLSR) is widely used due to its ability to handle high-dimensional and collinear spectral data (22), whereas machine learning approaches such as random forest (RF) can capture nonlinear relationships and interactions among variables (23). The performance of these models often varies depending on data characteristics and preprocessing strategies (7, 19). Although numerous studies have explored preprocessing

techniques and modelling approaches, many have evaluated them independently or under differing experimental conditions, making direct comparisons difficult. Some studies have examined multiple preprocessing-model combinations within a single dataset (24), but such studies are relatively limited and often focus on a broader range of soil properties rather than SOC specifically. Semi-arid regions such as those in southern India present unique challenges for SOC prediction using Vis–NIR spectroscopy. These regions are characterised by high spatial variability in SOC driven by heterogeneous land use, low organic matter inputs and strong seasonality in soil moisture, all of which introduce significant spectral variability that complicates model calibration and generalisation (19). Despite these challenges, studies specifically addressing SOC prediction under semi-arid agroecological conditions in southern India remain limited, highlighting the need for region-specific systematic evaluation.

To address these gaps, this study examined the effects of different spectral preprocessing techniques and modelling approaches on SOC prediction using Vis–NIR spectroscopy under semi-arid conditions and compared the performance of linear and nonlinear models within a consistent experimental framework. Unlike previous studies that evaluated preprocessing and modelling approaches independently or across diverse soil properties, this study focused specifically on SOC under semi-arid agroecological conditions in southern India, providing region-specific insights into preprocessing–model interactions. The specific objectives of this study were: (i) to assess the effect of different spectral preprocessing techniques (raw spectra, SG derivatives, MSC and SNV) on SOC prediction accuracy and (ii) to compare the performance of linear (PLSR) and nonlinear (RF) models for SOC prediction using Vis–NIR spectral data. The outcomes of this study are expected to contribute to the development of efficient and reliable approaches for SOC estimation, with potential applications in digital soil mapping, precision agriculture and soil carbon monitoring.

Materials and Methods

Study area and soil sampling

Representative surface soil samples ($n = 300$) were collected from the Nagathi Basapura sub-watershed, located in Harapanahalli Taluk of Vijayanagar district, Karnataka, India (Fig. 1). The study area is characterised by red sandy loam soils developed over granite and gneiss parent materials, typical of the semi-arid agroecological regions of southern India. The climate is semi-arid with an average annual rainfall of approximately 717 mm, predominantly received during the southwest monsoon. Land use in the region is dominated by a mix of rainfed and irrigated agriculture, with crops such as millets, pulses and oilseeds forming the major cropping systems. To capture the spatial variability of soil properties across different landforms, a random sampling strategy was adopted. Surface soil samples were collected from a depth of 0–30 cm, representing the plough layer. This depth was selected as it represents the biologically active zone most influenced by land use and management practices, where SOC dynamics are most pronounced (4). At each sampling location, soils were collected using a hand auger, placed in labelled plastic bags and transported to the laboratory. In the laboratory, the samples were air-dried, gently crushed using a wooden mallet and sieved through a 2 mm mesh to obtain a uniform particle size. These

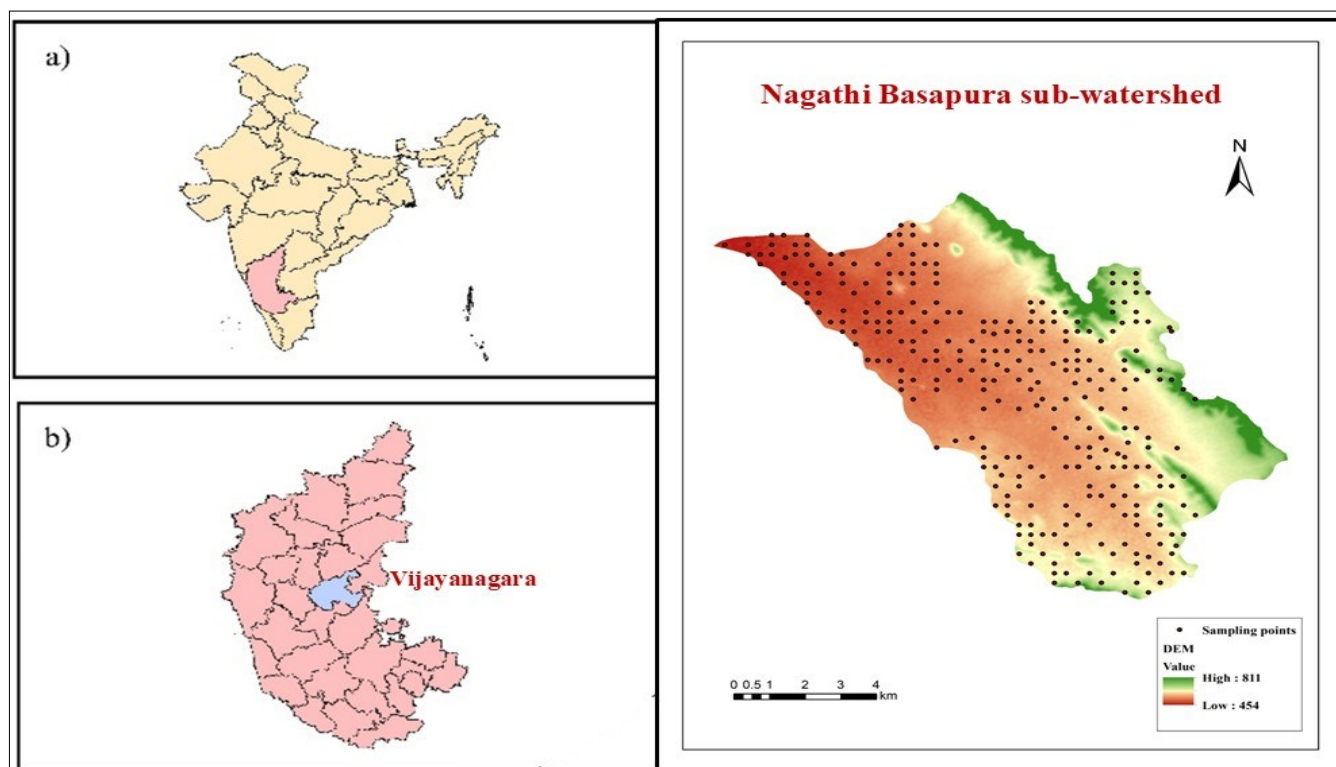


Fig. 1. Location map of the study area showing the Nagathi Basapura sub-watershed, Vijayanagar district, Karnataka, India, with the distribution of soil sampling locations and spatial variability of soil organic carbon [Source: Prepared using ArcGIS 10.7.1].

prepared samples were subsequently used for both spectral measurements and chemical analysis of SOC.

Soil spectral measurements and laboratory analysis

Spectral reflectance measurements were carried out using an SVC HR-1024i spectroradiometer (Spectra Vista Corporation, USA) at the National Bureau of Soil Survey and Land Use Planning (NBSS & LUP), Regional Centre, Bengaluru. The spectral reflectance data were acquired under laboratory conditions across the wavelength range of 350–2500 nm. The instrument provides a spectral bandwidth of 1.5 nm in the 350–1000 nm region, 3.8 nm in the 1000–1900 nm region and 2.5 nm in the 1900–2500 nm region. The spectral resolution [full width at half maximum (FWHM)] is approximately 2.8 nm at 700 nm, 8 nm at 1500 nm and 6 nm at 2100 nm. Prior to spectral acquisition, soil samples were placed in Petri dishes (15 cm diameter, 2 cm depth) to ensure a uniform surface. A controlled illumination setup using a halogen light source was employed to minimise external interference. Calibration was performed using a Spectralon® white reference panel (99 % reflectance) after every five samples to ensure measurement consistency. The integration time was set to 1 ms per scan and dark current correction was applied prior to measurements to account for detector noise. Each sample was scanned three times and the average reflectance spectrum was used for further analysis to reduce random noise. The instrument is equipped with low-noise silicon and InGaAs detector arrays and provides automatic dark current correction and spectrum averaging, ensuring reliable spectral acquisition across the 350–2500 nm wavelength range. This approach ensured high spectral reproducibility and minimised instrumental and environmental noise. The measured reflectance spectra were converted to absorbance values using the transformation $\log(1/R)$, where R represents reflectance. The absorbance spectra were subsequently subjected to spectral preprocessing and model development. For chemical analysis, a portion of each sample was

finely ground and sieved through a 0.25 mm mesh and SOC content was determined using the Walkley-Black wet oxidation method (25).

Preprocessing of soil spectral data

To enhance the quality of spectral data and improve the robustness of predictive models, several preprocessing techniques were applied and evaluated to minimise noise, correct scattering effects and emphasise relevant spectral features associated with SOC. Raw reflectance spectra often contain instrumental noise, baseline shifts and multiplicative scattering effects arising from variations in particle size, surface roughness and illumination conditions (16, 18). These factors can obscure subtle absorption features and negatively impact model performance. Prior to preprocessing, spectral data were truncated to the 400–2400 nm range, excluding noisy edge regions (350–399 nm and 2451–2500 nm), which are typically associated with low detector sensitivity and high noise levels. Spectral data were resampled to a uniform wavelength interval of 1 nm using linear interpolation prior to preprocessing, resulting in 2001 spectral bands for each sample. All preprocessing steps were implemented in R using the *prospectr* package and were applied in the following sequence: (i) spectral truncation to 400–2400 nm, (ii) resampling to uniform wavelength intervals and (iii) application of the respective preprocessing technique (first-derivative SG (SG-1D), second-derivative SG (SG-2D), MSC, or SNV). The following preprocessing techniques were evaluated:

Savitzky-Golay smoothing and derivative transformations

The SG filtering approach is widely used in Vis-NIR spectroscopy for simultaneous noise reduction and feature enhancement. This method applies a moving window across the spectrum and fits a least-squares polynomial to the data points within the window, allowing the central point to be replaced by a smoothed or differentiated value (26). Unlike simple averaging methods, SG filtering preserves the shape, width and position of spectral

absorption features, which are critical for chemical interpretation. In this study, both SG-1D and SG-2D transformations were computed using an SG filter with an 11-point window and a second-order polynomial. The SG-1D transformation is effective in correcting baseline offsets and enhancing the resolution of overlapping absorption bands. The SG-2D transformation further removes linear trends and enhances narrow and weak absorption features, improving sensitivity to subtle chemical variations. However, higher-order derivatives may amplify noise if not properly smoothed, making SG filtering essential (27, 28). The SG derivatives enhance diagnostically relevant absorption features linked to organic functional groups, thereby improving SOC prediction (24).

Multiplicative scatter correction

The MSC corrects multiplicative and additive scattering effects caused by variations in particle size, sample packing and surface roughness (29). Each spectrum is adjusted relative to a reference spectrum (typically the mean spectrum) using linear regression, which removes scaling differences and baseline offsets (28). This improves comparability among spectra and enhances model performance. This transformation enhances the linear relationship between spectral data and target variables by reducing variability unrelated to chemical composition.

Standard normal variate

The SNV is a widely used technique that corrects scatter-related distortions by centring and scaling each individual spectrum. The mean reflectance is subtracted and divided by the standard deviation, resulting in spectra with zero mean and unit variance (30). This method reduces multiplicative interferences caused by particle size variation and improves spectral consistency (27, 31). By standardising spectral responses, SNV improves the comparability of samples and supports more stable model calibration.

Modelling framework

To evaluate the predictive capability of Vis-NIR spectral data for SOC, both linear (PLSR) and nonlinear (RF) models were developed using raw and preprocessed spectral datasets.

Dataset partitioning

The dataset consisting of 300 soil samples was divided into calibration (75 %) and validation (25 %) subsets. To ensure a uniform distribution of SOC values in both subsets, samples were sorted based on their SOC values and every fourth sample was assigned to the validation set, while the remaining samples were allocated to the calibration set. This approach ensured that both datasets represented the full range of SOC variability.

Partial least squares regression

The PLSR was employed to model the relationship between spectral reflectance and SOC. The PLSR is a widely used multivariate technique in soil spectroscopy that transforms highly collinear spectral variables into a reduced set of orthogonal latent variables while maximising covariance with the response variable (SOC) (22). This makes it particularly suitable for high-dimensional hyperspectral data, where the number of predictors far exceeds the number of observations. The PLSR effectively addresses multicollinearity and noise in spectral datasets by extracting the most relevant information from the predictor space (32). However, it primarily captures linear relationships between spectral variables and the target property, which may limit its predictive performance when complex nonlinear

interactions are present. In this study, PLSR models were developed using the calibration dataset, with spectral bands as predictor variables and SOC as the response variable. The optimal number of latent variables was determined using leave-one-out cross-validation (LOO-CV) to minimise prediction error and avoid overfitting. A maximum of 30 components was evaluated and the final model was constructed using 15 components ($n_{comp} = 15$) based on cross-validation performance. Model predictions were generated for both calibration and validation datasets, allowing evaluation of both model fitting and generalisation capability. The PLSR modelling was implemented using the *pls* package in R.

Random forest

The RF was employed as a nonlinear ensemble learning approach to model the relationship between spectral data and SOC (23). The RF constructs many decision trees using bootstrap aggregation (bagging) and combines their predictions to improve accuracy and reduce variance (33). Unlike linear models, RF can capture complex nonlinear relationships and interactions among spectral variables, which are common in soil spectroscopy due to the combined influence of organic matter, minerals, moisture and texture on spectral response. Additionally, RF is robust to noise, less sensitive to overfitting and performs well with high-dimensional datasets. In RF, each tree is built using a random subset of predictor variables at each split, controlled by the parameter *mtry*, which promotes model diversity and reduces correlation among trees. The final prediction is obtained by averaging the outputs of all trees. The RF model was trained using 1000 trees ($n_{tree} = 1000$), as this number was found to provide stable out-of-bag error estimates with no significant improvement observed beyond this threshold. The number of variables randomly selected at each split was set to $mtry = 20$, which approximates the commonly recommended default of the square root of the total number of predictors for regression tasks, balancing model diversity and predictive accuracy (33). A minimum node size of 10 was selected to prevent overfitting to noise in the high-dimensional spectral dataset. Model predictions were generated for both calibration and validation datasets to evaluate model performance. RF modelling was implemented using the RF package in R (34).

Model evaluation and validation

The performance of the developed models was evaluated using both calibration and validation datasets. Model accuracy and predictive capability were assessed using the following statistical metrics: Coefficient of determination (R^2): Indicates the proportion of variance in SOC explained by the model. Root mean square error: Measures the average deviation between observed and predicted SOC values. Ratio of performance to interquartile distance (RPIQ): Evaluates model robustness relative to data variability. The RPIQ values were interpreted using established thresholds: $RPIQ < 1.4$ indicates poor prediction, $1.4-2.0$ indicates fair prediction suitable for approximate quantitative analysis and $RPIQ > 2.0$ indicates good to excellent prediction capability (11).

The mathematical expressions of these metrics are given below:

$$R^2 = 1 - [\sum (p_i - o_i)^2 / \sum (o_i - \bar{o})^2] \quad \text{Eq. (1)}$$

$$RMSE = \sqrt{(1/n) \sum (o_i - p_i)^2} \quad \text{Eq. (2)}$$

$$RPIQ = (Q_3 - Q_1) / RMSE \quad \text{Eq. (3)}$$

where p_i and o_i represent predicted and observed SOC

values, respectively; $\bar{o}$ is the mean of observed values; Q_3 and Q_1 denote the third and first quartiles; and n is the total number of samples.

Higher R^2 and RPIQ values, along with lower RMSE, indicate better model performance. Model evaluation was conducted separately for each preprocessing technique and modelling approach to enable systematic comparison (11). All metrics were

computed using the *ithir* package in R, which provides functions such as *goof* and *RPIQ* for model assessment. The overall methodological workflow adopted in this study is illustrated in Fig. 2.

Results and Discussion

Variability and distribution of soil organic carbon

The descriptive statistics of SOC in the study area (Table 1) indicate substantial variability across the sampled soils. The SOC values

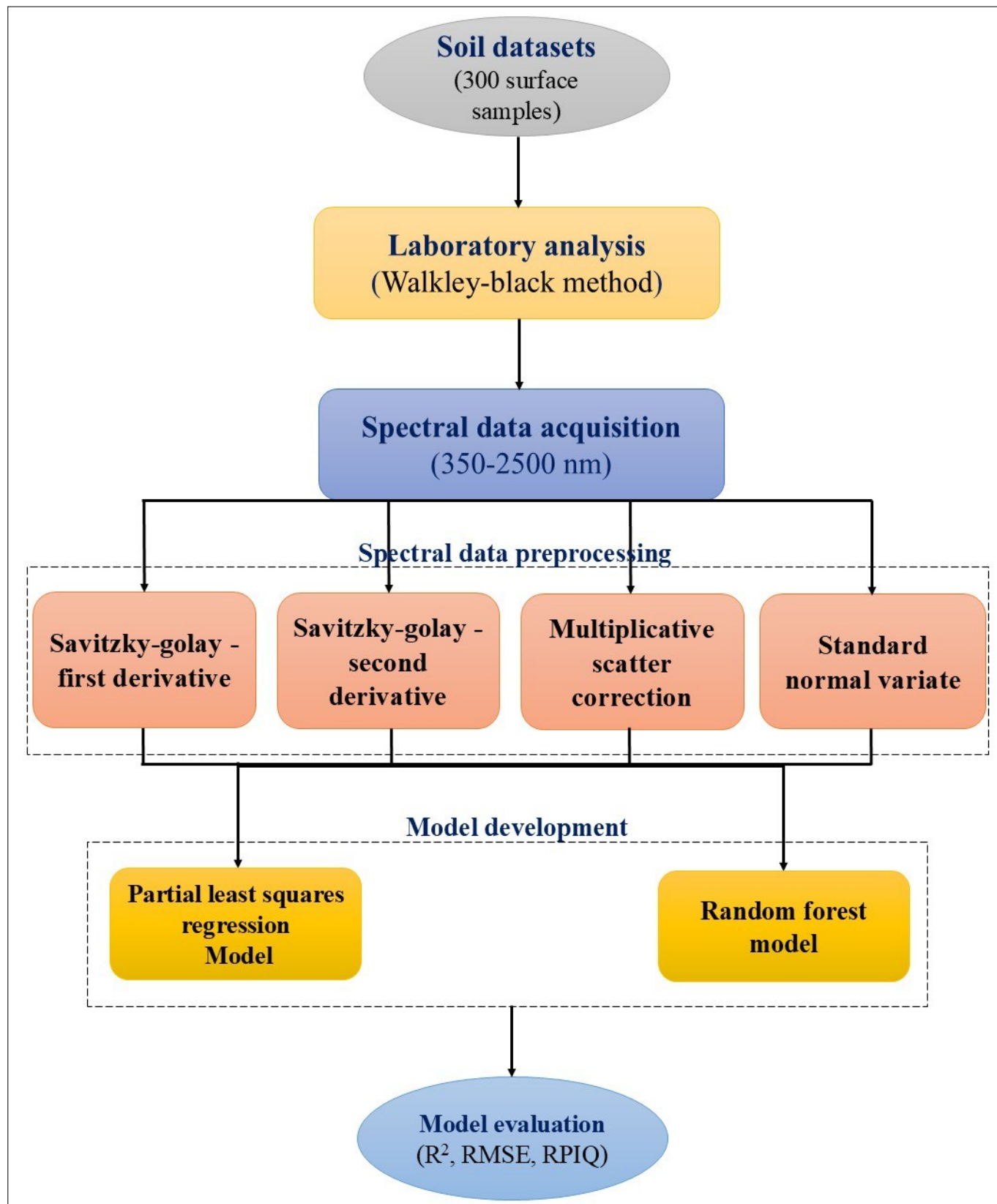


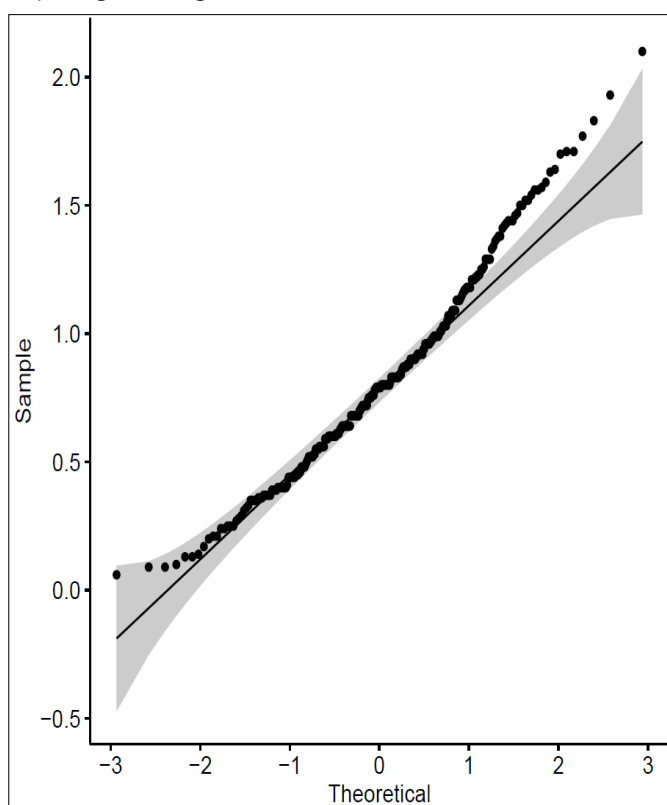
Fig. 2. Flowchart of the overall methodology for soil organic carbon prediction using visible and near infrared spectroscopy. R^2 : coefficient of determination; RMSE: root mean square error; RPIQ: ratio of performance to interquartile distance.

Table 1. Descriptive statistics of soil organic carbon (SOC) in the study area (n = 300)

Sl. No.	Statistic	SOC
1	Maximum (%)	1.93
2	Minimum (%)	0.06
3	Mean (%)	0.81
4	Standard deviation (%)	0.37
5	Coefficient of variation (%)	45.56
6	Skewness	0.50
7	Kurtosis	0.01

ranged from 0.06 % to 1.93 %, with a mean of 0.81 %, reflecting generally low to moderate carbon levels typical of semi-arid agroecosystems. The coefficient of variation (CV) of 45.56 % suggests moderate to high spatial variability in SOC distribution. Such variability is commonly observed in semi-arid regions due to heterogeneity in land use, management practices and soil-forming factors. The distribution of SOC exhibited slight positive skewness (0.50) and near-normal kurtosis (0.01), indicating a moderately skewed but approximately normal distribution. This is further supported by the Q-Q plot (Fig. 3), where most data points closely follow the theoretical normal line with minor deviations at the tails. Such distribution characteristics are favourable for model development, as extreme outliers are limited and the data approximate normality.

The observed variability is consistent with previous studies reporting heterogeneous SOC distribution under semi-arid

**Fig. 3.** Quantile–quantile plot comparing sample soil organic carbon values (y-axis) with the theoretical normal distribution (x-axis).**Table 2.** Performance of partial least squares regression models for soil organic carbon (SOC) prediction using different spectral preprocessing techniques

Preprocessing technique	Calibration			Validation		
	R ²	RMSE (%)	RPIQ	R ²	RMSE (%)	RPIQ
Raw	0.64	0.22	1.98	0.32	0.38	1.29
SG-1D	0.73	0.19	2.29	0.36	0.33	1.34
SG-2D	0.82	0.16	2.80	0.30	0.38	1.19
MSC	0.70	0.20	2.17	0.32	0.34	1.32
SNV	0.67	0.22	2.05	0.34	0.33	1.38

R²: coefficient of determination; RMSE: root mean square error; RPIQ: ratio of performance to interquartile distance; SG-1D: Savitzky–Golay first derivative; SG-2D: Savitzky–Golay second derivative; MSC: multiplicative scatter correction; SNV: standard normal variate.

conditions, where spatial differences in vegetation, soil texture and moisture availability strongly regulate carbon dynamics (19, 35). For instance, previous researchers reported CV values exceeding 40 % for SOC in semi-arid landscapes, consistent with the 45.56 % observed in the present study, reflecting the inherent spatial heterogeneity driven by land use and soil-forming factors (35). Similar variability ranges have also been reported in Vis–NIR-based SOC studies, highlighting the inherent complexity of modelling SOC across diverse soil environments (15, 36). From a modelling perspective, this level of variability is advantageous because it provides sufficient contrast for calibration, enabling models to capture meaningful relationships between spectral signatures and SOC. This variability is particularly important for evaluating the effectiveness of different spectral preprocessing techniques and modelling approaches, as it ensures that model performance is tested across a wide range of SOC conditions. However, high variability can also pose challenges for model generalisation, particularly when spectral responses are influenced by multiple interacting soil properties such as mineralogy and moisture (37, 38).

Effect of spectral preprocessing techniques on soil organic carbon prediction

Calibration performance was significantly improved following preprocessing across different spectral preprocessing techniques (Tables 2 and 3), highlighting the critical role of preprocessing in improving predictive accuracy. Overall, preprocessing enhanced model performance compared to raw spectra; however, the extent of improvement depended on both the type of preprocessing technique and the modelling approach.

For the PLSR model, calibration performance improved markedly with preprocessing. The calibration and validation performance of the PLSR models developed using different spectral preprocessing techniques is presented in Table 2. The highest calibration accuracy was observed for SG-2D (R² = 0.82, RMSE = 0.16 %, RPIQ = 2.80), followed by SG-1D (R² = 0.73, RMSE = 0.19 %, RPIQ = 2.29) and MSC (R² = 0.70, RMSE = 0.20 %, RPIQ = 2.17), compared to raw spectra (R² = 0.64, RMSE = 0.22 %, RPIQ = 1.98). However, this improvement in calibration was not consistently reflected in validation results. Validation performance remained moderate, with SG-1D (R² = 0.36, RMSE = 0.33 %, RPIQ = 1.34) and SNV (R² = 0.34, RMSE = 0.33 %, RPIQ = 1.38) showing relatively better generalisation, while SG-2D exhibited reduced validation accuracy (R² = 0.30, RMSE = 0.38 %, RPIQ = 1.19). This discrepancy between calibration and validation performance for SG-2D can be attributed to overfitting induced by noise amplification. Second-order derivatives inherently amplify high-frequency spectral noise and when PLSR maximises covariance between these amplified features and SOC during calibration, it captures noise-specific patterns that do not generalise to independent validation samples. Statistically, this manifests as an

Table 3. Performance of random forest models for soil organic carbon (SOC) prediction using different spectral preprocessing techniques

Preprocessing technique	Calibration			Validation		
	R ²	RMSE (%)	RPIQ	R ²	RMSE (%)	RPIQ
Raw	0.88	0.17	2.59	0.30	0.31	1.45
SG-1D	0.92	0.14	3.07	0.50	0.27	1.66
SG-2D	0.94	0.14	3.08	0.47	0.29	1.57
MSC	0.89	0.16	2.79	0.50	0.27	1.67
SNV	0.88	0.17	2.58	0.32	0.31	1.46

R²: coefficient of determination; RMSE: root mean square error; RPIQ: ratio of performance to interquartile distance; SG-1D: Savitzky–Golay first -derivative; SG-2D: Savitzky–Golay second-derivative; MSC: multiplicative scatter correction; SNV: standard normal variate.

inflated calibration R² with a disproportionate increase in validation RMSE, a classic signature of overfitting in high-dimensional regression models (39).

A more pronounced effect of preprocessing was observed in the RF model. Although calibration accuracy remained consistently high across all preprocessing techniques (R² = 0.88–0.94), validation performance varied considerably. The highest validation accuracy was achieved using MSC and SG-1D (R² = 0.50, RMSE = 0.27 %, RPIQ ≈ 1.66–1.67), followed by SG-2D (R² = 0.47, RMSE = 0.29 %, RPIQ = 1.57). In contrast, raw spectra and SNV resulted in lower validation performance (R² ≈ 0.30–0.32; RMSE ≈ 0.31 %; RPIQ ≈ 1.45–1.46). According to the RPIQ interpretation thresholds adopted in this study, most preprocessing–model combinations achieved fair predictive performance (RPIQ 1.4–2.0), with the best RF models (MSC and SG-1D) approaching the upper boundary of this range. These results indicate that while the models provide meaningful SOC predictions, their accuracy is moderate and should be interpreted with appropriate caution for practical applications. These results indicate that appropriate preprocessing significantly enhances the generalisation capability of RF models (17, 40).

The observed trends can be explained by the functional characteristics of preprocessing techniques. Scatter-correction methods such as MSC effectively reduce variability caused by particle size, surface roughness and illumination differences, thereby improving spectral consistency (29, 41). In contrast, derivative-based methods enhance subtle absorption features associated with SOC by removing baseline effects and resolving overlapping bands. The SG-1D approach provided an optimal balance between feature enhancement and noise control, whereas SG-2D likely amplified high-frequency noise, leading to reduced prediction stability (26, 42). Previous studies reported that SG-1D preprocessing improved PLSR validation R² from 0.41 to 0.53 in semi-arid soils, while MSC yielded comparable improvements, consistent with the present findings (19). Similarly, previous studies demonstrated that scatter-correction methods outperformed derivative approaches in

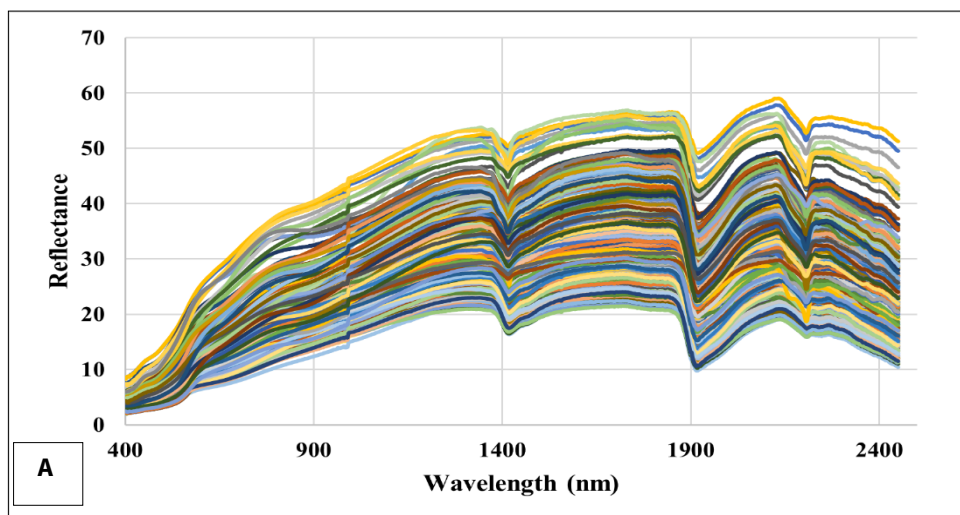
validation across diverse soil datasets, supporting the superior generalisation observed with MSC and SG-1D in this study (17).

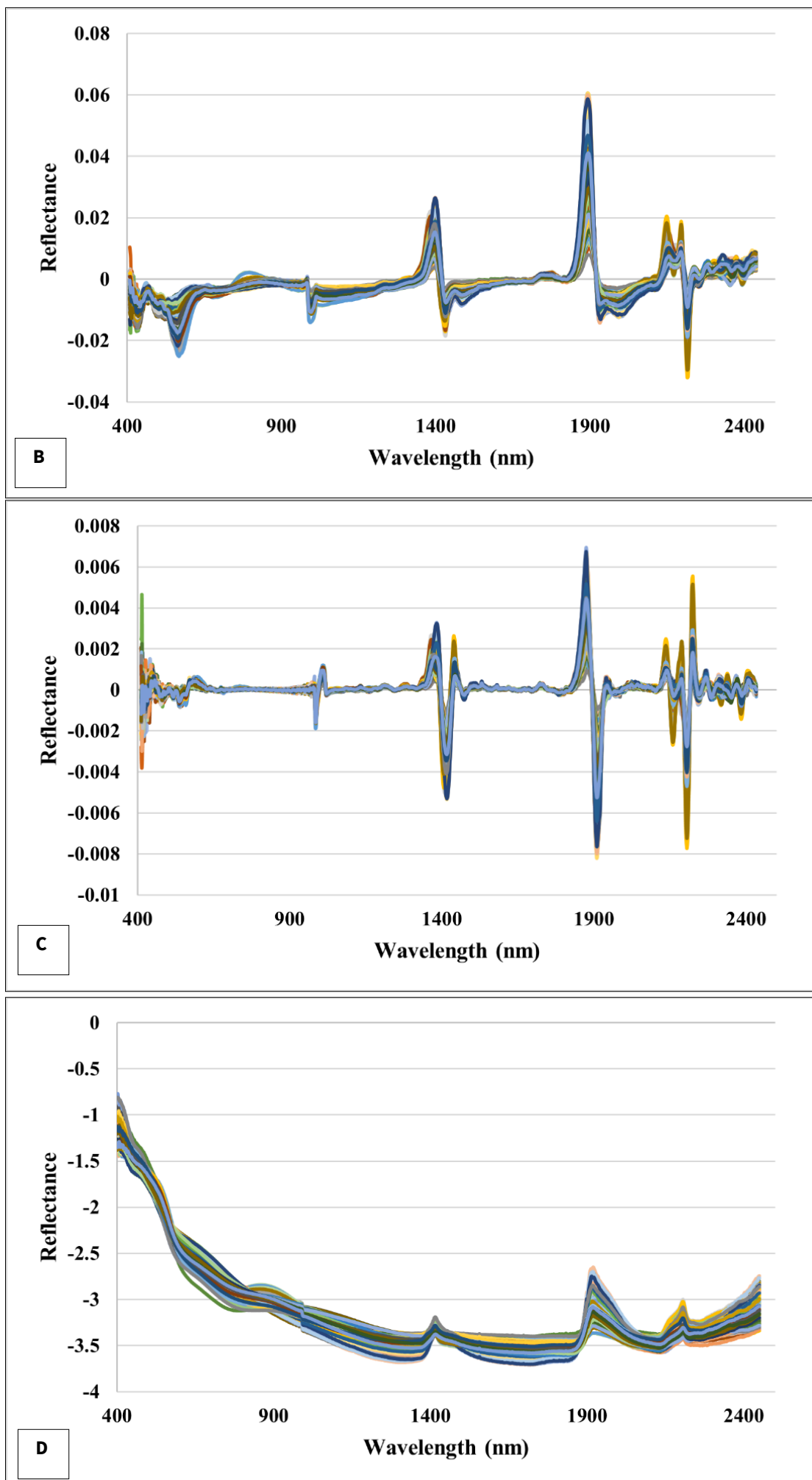
These effects are also evident in the spectral behaviour illustrated in Fig. 4. Variable importance analysis of the RF model trained with SG-1D preprocessing revealed that the spectral bands most influential for SOC prediction were concentrated in the NIR region, particularly around 1400 nm and 1900 nm, corresponding to O–H overtone and combination absorption bands characteristic of soil organic matter (Fig. 5). Moderate importance was also observed in the 400–700 nm visible range, associated with SOC-related soil colour and organic matter absorption, while the 2000–2200 nm region showed elevated importance consistent with C–H combination bands of organic compounds (12, 13, 24). These findings confirm the mechanistic basis of the RF model and are consistent with diagnostically relevant spectral regions widely reported in Vis–NIR SOC studies.

The moderate validation performance observed across all preprocessing–model combinations (maximum R² = 0.50) despite extensive preprocessing warrants discussion. Several factors likely contributed to this limitation. First, SOC in semi-arid soils is present at generally low concentrations (mean = 0.81 %), which reduces the spectral contrast available for model calibration. Second, the spectral signal of SOC is confounded by co-varying soil properties such as iron oxides, clay mineralogy and moisture content, which were not explicitly accounted for in the modelling framework. Third, the absence of external validation samples limits assessment of true model transferability. These factors collectively constrain prediction accuracy and highlight the challenges of SOC spectroscopy under semi-arid conditions (14, 37).

Comparison of modelling approaches

A comparison of modelling approaches across different preprocessing techniques (Tables 2 and 3) reveals clear differences in predictive performance between PLSR and RF models. While both models benefited from preprocessing, their ability to generalise to





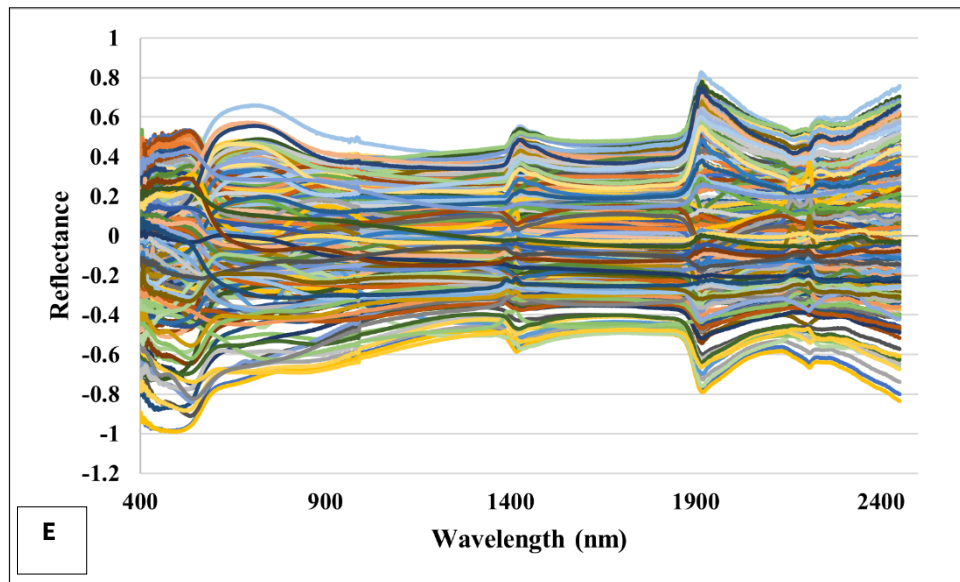


Fig. 4. Effect of different spectral preprocessing techniques on soil reflectance curves (400–2400 nm): (a) raw spectra; (b) Savitzky–Golay first derivative; (c) Savitzky–Golay second derivative; (d) multiplicative scatter correction (MSC); (e) standard normal variate.

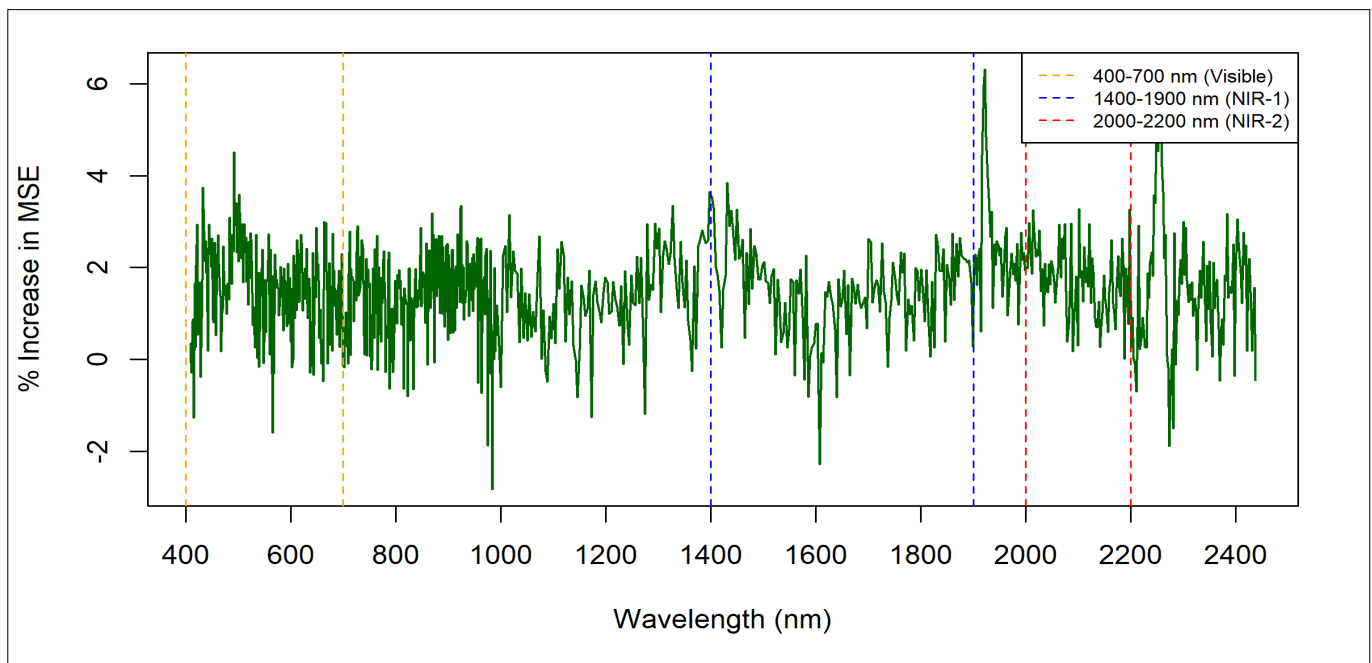


Fig. 5. Variable importance of spectral bands (expressed as % increase in mean square error) for soil organic carbon prediction using the random forest model. Dashed lines indicate key diagnostic spectral regions: orange (400–700 nm, visible), blue (1400–1900 nm, NIR-1) and red (2000–2200 nm, NIR-2).

independent data varied considerably.

The PLSR model exhibited moderate predictive performance, with validation R^2 values ranging from 0.30 to 0.36 across preprocessing techniques. Although preprocessing significantly improved calibration accuracy, this did not translate into comparable improvements in validation, indicating limited generalisation capability. This behaviour reflects the inherent limitation of PLSR in capturing complex nonlinear relationships between spectral variables and SOC (43). It should be noted that even the best PLSR validation performance ($R^2 = 0.36$) falls within the poor to fair RPIQ range (1.19–1.38), indicating that PLSR alone provides insufficient predictive accuracy for reliable quantitative SOC estimation in this semi-arid dataset.

The calibration and validation performance of the RF models under different preprocessing techniques is summarised in Table 3. The RF model in contrast to PLSR model demonstrated superior

performance, with higher validation R^2 values (up to 0.50), lower RMSE and higher RPIQ across preprocessing techniques. The best performance was achieved with MSC and SG-1D preprocessing, indicating that RF benefits from preprocessing methods that enhance spectral features while maintaining stability. The superior performance of RF can be attributed to its ensemble nature, which reduces prediction variance through averaging across many decorrelated decision trees and its ability to model nonlinear relationships and higher-order interactions among spectral variables that are common in heterogeneous semi-arid soil systems (44, 45). The bagging mechanism inherent to RF also provides implicit regularisation, reducing sensitivity to noise in individual spectral bands. This suggests that SOC–spectral relationships in the study area are predominantly nonlinear, influenced by complex interactions among organic matter, mineral composition and soil physical properties.

The differences in model behaviour are further supported by the observed versus predicted plots (Fig. 6 and 7). The PLSR model (Fig. 6) shows greater dispersion of validation points around the 1 : 1 line, indicating reduced prediction accuracy. In contrast, the RF model (Fig. 7) exhibits tighter clustering of validation points along the 1 : 1 line, reflecting better agreement between observed and predicted values and improved generalisation. These findings are consistent with previous researchers, who reported RF validation R^2

of 0.52–0.58 for SOC prediction using Vis-NIR spectroscopy (23) and RF outperformed PLSR across multiple preprocessing combinations in diverse soil datasets (24). These findings are consistent with previous studies demonstrating that ensemble machine learning approaches outperform linear models in complex spectroscopic datasets due to their ability to capture nonlinear patterns and reduce overfitting (46, 47). Despite the overall superiority of RF over PLSR, it is important to acknowledge that the maximum validation

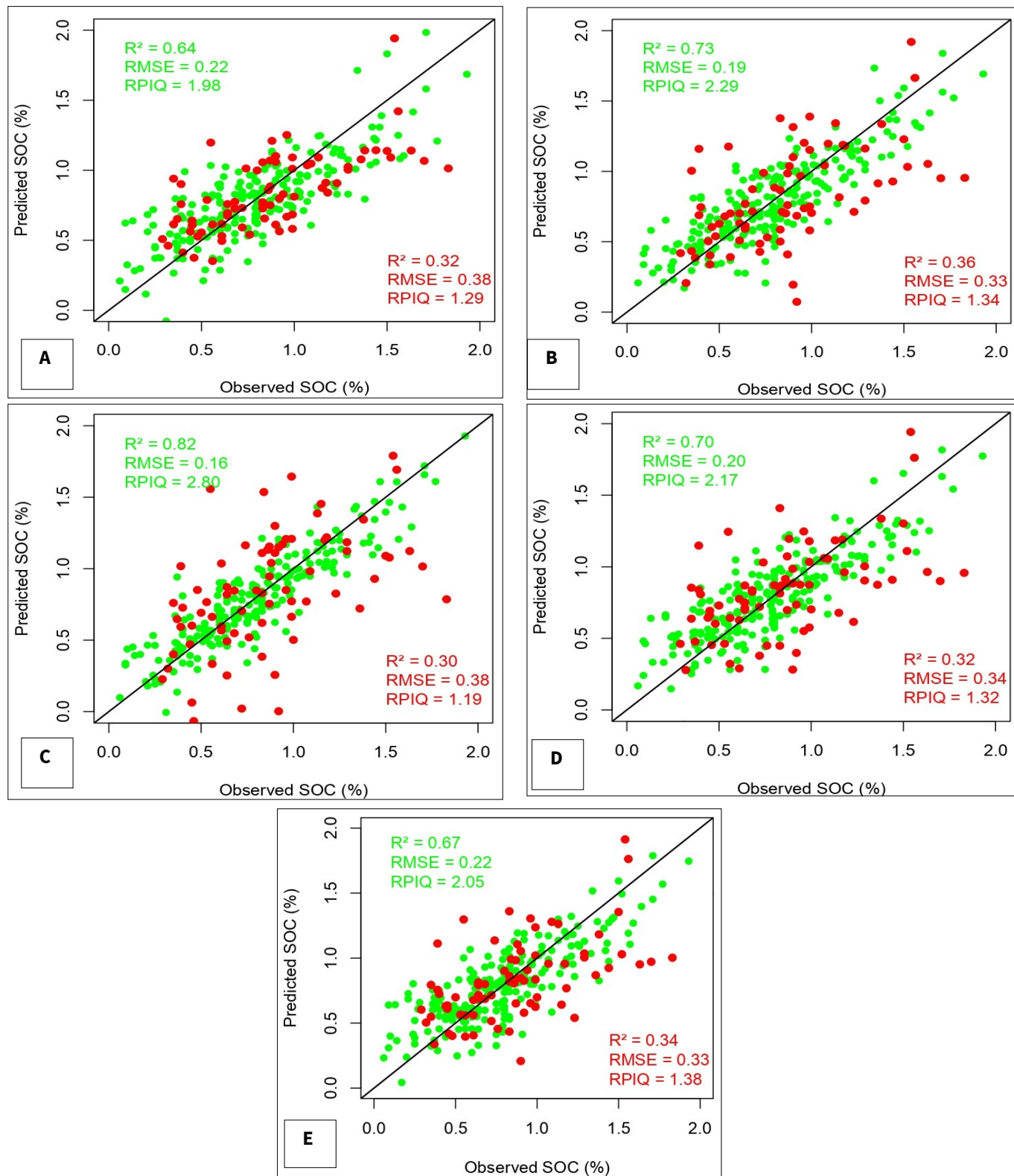


Fig. 6. Observed versus predicted soil organic carbon values for calibration and validation datasets using Partial least squares regression models, illustrating model performance under different preprocessing techniques: (a) raw spectra; (b) Savitzky-Golay first-derivative; (c) Savitzky-Golay second-derivative; (d) multiplicative scatter correction; (e) standard normal variate. Green dots represent calibration samples and red dots represent validation samples.

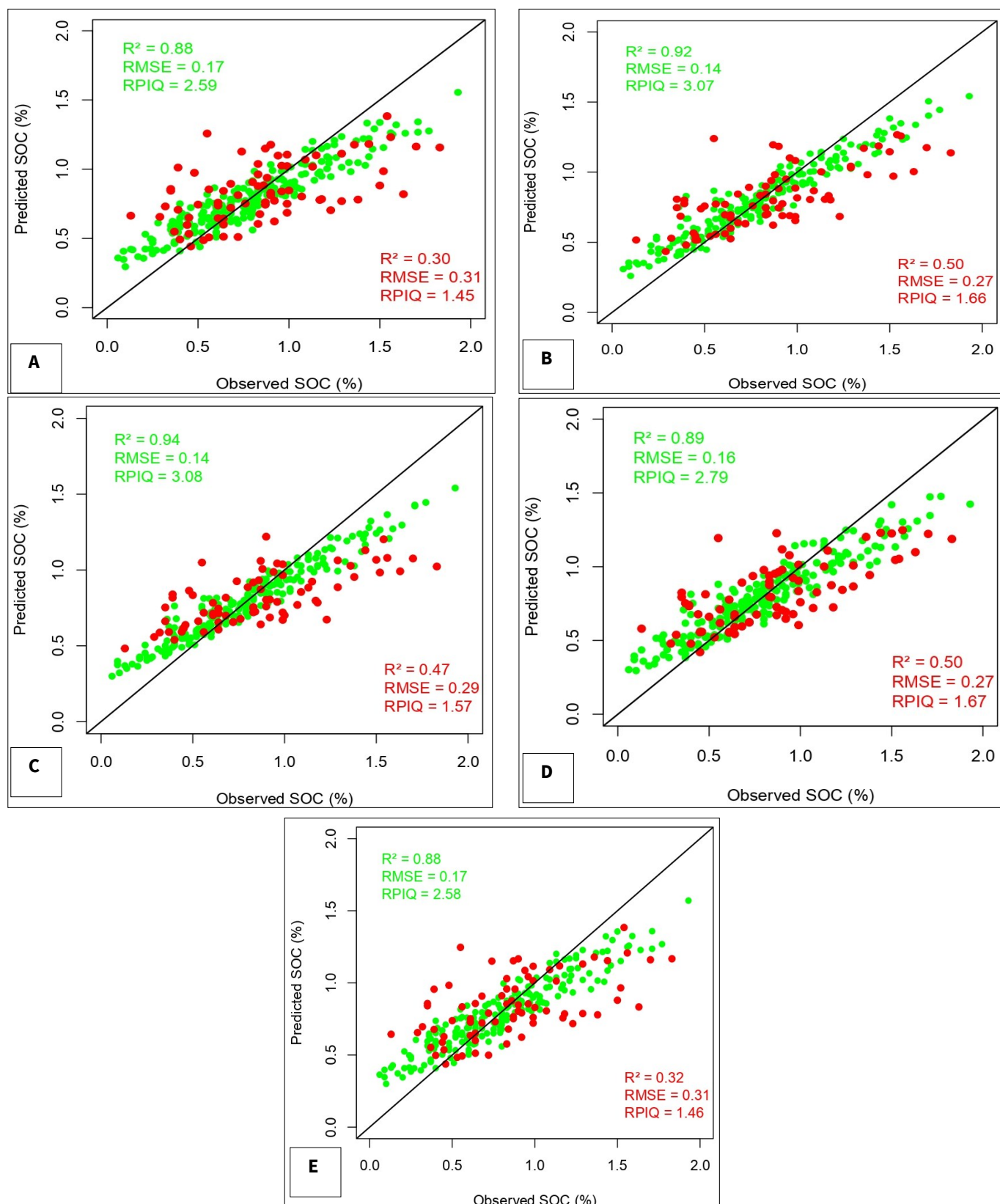


Fig. 7. Observed versus predicted soil organic carbon values for calibration and validation datasets using random forest models, illustrating model performance under different preprocessing techniques: (a) raw spectra; (b) Savitzky-Golay first-derivative; (c) Savitzky-Golay second-derivative; (d) multiplicative scatter correction; (e) standard normal variate. Green dots represent calibration samples and red dots represent validation samples.

R^2 of 0.50 achieved in this study reflects moderate predictive performance. This level of accuracy, while consistent with findings from comparable semi-arid soil spectroscopy studies, indicates that the developed models are not yet sufficiently robust for standalone quantitative SOC mapping without further validation and refinement. The inclusion of additional soil covariates, larger and more diverse training datasets and advanced machine learning

algorithms such as support vector machine (SVM), Cubist, or XGBoost may further improve predictive performance in future studies.

Importantly, these results directly address the second objective of this study by demonstrating that model choice significantly influences SOC prediction accuracy, with nonlinear approaches providing clear advantages under heterogeneous semi-

arid conditions. The results indicate that model performance is strongly influenced by both preprocessing and modelling approach. The RF model, when combined with appropriate preprocessing techniques, provides a more robust and reliable framework for SOC prediction. Based on the combined evaluation, the RF model coupled with MSC and SG-1D preprocessing emerged as the most effective approach, achieving the highest validation accuracy and lowest prediction error. This highlights the importance of integrating suitable preprocessing strategies with robust modelling techniques to improve SOC prediction under semi-arid conditions.

Limitations and future perspectives

Although this study demonstrated the significance of spectral preprocessing and modelling approach selection for SOC prediction under semi-arid conditions, certain limitations should be acknowledged. The analysis was based on a dataset of 300 samples from a specific semi-arid region, which may limit the applicability of the findings to other soil types and agro-ecological conditions. Soil spectral responses are influenced by multiple factors, including mineralogy, moisture conditions and land management practices, which can vary spatially and affect model performance. Furthermore, important soil properties such as texture, soil moisture, iron oxides and clay mineral composition, which are known to strongly influence Vis-NIR spectral behaviour and co-vary with SOC, were not incorporated into the modelling framework. Their exclusion may have contributed to the moderate validation performance observed and represents an important avenue for future investigation. Additionally, all model validation was conducted within the same sampling population using a held-out subset and no independent external validation dataset was used. This limits the assessment of true model transferability across different soil populations, regions, or measurement conditions. From a practical standpoint, the computational requirements and deployment feasibility of the developed models for field-scale SOC monitoring were not evaluated. Future work should assess the scalability and operational efficiency of these approaches for integration into real-time or large-scale digital soil mapping frameworks.

Future research should explicitly focus on the following directions: (i) expanding training datasets across diverse agroecological regions of India to improve model generalisability; (ii) incorporating additional soil covariates such as texture, iron oxide content and clay mineralogy into the modelling framework; (iii) evaluating advanced machine learning algorithms including SVM, Cubist, XGBoost and deep learning approaches for comparative assessment; (iv) conducting external validation using independent soil datasets from similar semi-arid regions; (v) conducting detailed explainable AI analyses (e.g., partial dependence analysis) to better interpret important spectral regions; and (vi) assessing the feasibility of deploying optimised preprocessing-model combinations in field-scale and proximal sensing platforms for real-time SOC monitoring.

Conclusion

This study systematically evaluated the influence of spectral preprocessing techniques and modelling approaches on soil organic carbon (SOC) prediction using visible and near infrared (Vis-NIR) spectroscopy under semi-arid conditions. The results demonstrate that both preprocessing and model selection significantly influence prediction accuracy, with their interactions playing a critical role in overall model performance. Among preprocessing techniques, first-

derivative Savitzky-Golay (SG-1D) and multiplicative scatter correction (MSC) consistently improved performance, whereas second-derivative SG (SG-2D), despite high calibration accuracy [$R^2 = 0.82$ for partial least squares regression (PLSR)], showed reduced validation performance ($R^2 = 0.30$), indicating overfitting due to noise amplification. The RF model outperformed PLSR, achieving the best results when combined with MSC and SG-1D (validation $R^2 = 0.50$, root mean square error (RMSE) = 0.27 %, ratio of performance to interquartile distance (RPIQ) ≈ 1.66 –1.67), while PLSR showed only moderate validation performance ($R^2 \leq 0.36$). The moderate validation accuracy achieved across all models highlights the challenges of SOC prediction under semi-arid conditions and underscores the need for future studies incorporating external validation, additional soil covariates and advanced machine learning algorithms to develop more robust and transferable prediction frameworks. Overall, integrating balanced preprocessing techniques with robust nonlinear models provides a reliable framework for improving SOC prediction, with important applications in digital soil mapping, precision agriculture and soil carbon monitoring.

Acknowledgements

The authors express their heartfelt gratitude to the Department of Soil Science and Agricultural Chemistry, the REWARD programme and the Karnataka Watershed Development Department for facilitating soil and spectral data collection. They also acknowledge the Centre of Excellence on Watershed Management, University of Agricultural Sciences, Bengaluru for providing the PhD fellowship under the REWARD programme.

Authors' contributions

CGK contributed to the preparation of original draft, methodology, investigation, formal analysis, software, validation and visualisation. JS contributed to conceptualisation, writing – review and editing and supervision. SD contributed to conceptualisation, data curation, software, visualisation and supervision. AS contributed to writing – review and editing and supervision. UN contributed to writing – review and editing. All authors have read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest: Authors do not have any conflict of interests to declare.

Ethical issues: None

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

During the preparation of this work, the authors used ChatGPT in order to improve the language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

1. Ramesh T, Bolan NS, Kirkham MB, Wijesekara H, Kanchikerimath M, Rao CS, et al. Soil organic carbon dynamics: Impact of land use changes and management practices: A review. *Adv Agron.* 2019;156:1–107. <https://doi.org/10.1016/bs.agron.2019.02.001>

2. Dharumarajan S, Adhikari K, Chakraborty R, Kalaiselvi B, Vasundhara R, Lalitha M, et al. Prediction and mapping of soil organic carbon stock via large datasets coupled with pedotransfer functions. *Earth Sci Inform.* 2025;18(3):314. <https://doi.org/10.1007/s12145-025-01822-z>
3. Dharumarajan S, Gomez C, Kusuma CG, Vasundhara R, Kalaiselvi B, Lalitha M, et al. Prediction of soil organic carbon stock along layers and profiles using Vis-NIR laboratory spectroscopy. *Catena.* 2025;257:109150. <https://doi.org/10.1016/j.catena.2025.109150>
4. McBratney AB, Stockmann U, Angers DA, Minasny B, Field DJ. Challenges for soil organic carbon research. In: *Soil carbon.* Cham: Springer; 2014. p. 3–16. https://doi.org/10.1007/978-3-319-04084-4_1
5. Ng WK, Maxfield PJ, Crew AP, Teixeira DL, Bevan T, Bell MJ. Comparison of soil organic carbon measurement methods. *Agronomy.* 2025;15(8):1826. <https://doi.org/10.3390/agronomy15081826>
6. Stenberg B, Viscarra Rossel RA, Mouazen AM, Wetterlind J. Visible and near infrared spectroscopy in soil science. *Adv Agron.* 2010;107:163–215. [https://doi.org/10.1016/S0065-2113\(10\)07005-7](https://doi.org/10.1016/S0065-2113(10)07005-7)
7. Viscarra Rossel RA, Behrens T, Ben-Dor E, Chabrilat S, Demattê JAM, Ge Y, et al. Diffuse reflectance spectroscopy for estimating soil properties: A technology for the 21st century. *Eur J Soil Sci.* 2022;73(4):e13271. <https://doi.org/10.1111/ejss.13271>
8. Kusuma CG, Dharumarajan S, Vasundhara R, Gomez C, Manjunatha MH, Hegde R. Predicting soil nutrient classes using Vis-NIR spectroscopy to support sustainable farming decisions. *Land Degrad Dev.* 2025;36(15):5313–23. <https://doi.org/10.1002/ldr.70007>
9. Mozaffari H, Moosavi AA, Ostovari Y, Nematollahi MA, Rezaei M. Developing spectrotransfer functions (STFs) to predict basic soil properties. *Geoderma.* 2022;428:116174. <https://doi.org/10.1016/j.geoderma.2022.116174>
10. Sweta K, Dharumarajan S, Kalaiselvi B, Gomez C, Lalitha M, Das B, et al. Reviewing the trajectory of Vis-NIR and MIR spectroscopic soil studies in India. *J Indian Soc Soil Sci.* 2024;72(1):23–34. <https://doi.org/10.5958/0974-0228.2024.00026.3>
11. Kusuma CG, Bhoomika SA, Dharumarajan S. Prediction of soil nutrients using visible-near-infrared reflectance spectroscopy. In: *Remote sensing of soils.* Elsevier; 2024. p. 493–502. <https://doi.org/10.1016/B978-0-443-18773-5.00001-6>
12. Chinilin AV, Vindeker GV, Savin IY. Vis-NIR spectroscopy for soil organic carbon assessment: A meta-analysis. *Eurasian Soil Sci.* 2023;56(11):1605–17. <https://doi.org/10.1134/S1064229323601841>
13. Stevens A, Nocita M, Tóth G, Montanarella L, van Wesemael B. Prediction of soil organic carbon at European scale using spectroscopy. *PLoS One.* 2013;8(6):e66409. <https://doi.org/10.1371/journal.pone.0066409>
14. Sharififar A, Francos N, Ng W, Minasny B, Gholizadeh A, Karunaratne S, et al. Navigating challenges of spectral soil sensing. *Soil Environ Health.* 2025;3(3):100160. <https://doi.org/10.1016/j.seh.2025.100160>
15. Ghosh AK, Das BS, Reddy N. VIS-NIR spectroscopy for SOC estimation with preprocessing techniques. *Geoderma Reg.* 2020;23:e00349. <https://doi.org/10.1016/j.geodrs.2020.e00349>
16. Reyes J, Liefß M. Spectral data processing for field-scale SOC monitoring. *Sensors.* 2024;24(3):849. <https://doi.org/10.3390/s24030849>
17. Dotto AC, Dalmolin RSD, ten Caten A, Grunwald S. Preprocessing for SOC prediction using Vis-NIR spectra. *Geoderma.* 2018;314:262–74. <https://doi.org/10.1016/j.geoderma.2017.11.006>
18. Eslamifar M, Tavakoli H, Thiessen E, Kock R, Correa J, Hartung E. Spectral preprocessing improves soil prediction accuracy. *Discover Appl Sci.* 2025;7(8):896. <https://doi.org/10.1007/s42452-025-07580-3>
19. Wang Y, Yang S, Yan X, Yang C, Feng M, Xiao L, et al. Evaluation of preprocessing and regression models for SOC estimation. *J Soils Sediments.* 2023;23(2):634–45. <https://doi.org/10.1007/s11368-022-03337-2>
20. Wu M, Huang Y, Zhao X, Jin J, Ruan Y. Effects of spectral processing on soil organic matter prediction. *J Soils Sediments.* 2024;24(2):914–27. <https://doi.org/10.1007/s11368-023-03691-9>
21. Zhang Z, Ding J, Zhu C, Wang J. Signal preprocessing and band selection for SOM prediction. *Spectrochim Acta A.* 2020;240:118553. <https://doi.org/10.1016/j.saa.2020.118553>
22. Nocita M, Stevens A, Toth G, Panagos P, van Wesemael B, Montanarella L. SOC prediction using local PLSR. *Soil Biol Biochem.* 2014;68:337–47. <http://dx.doi.org/10.1016/j.soilbio.2013.10.022>
23. Karthikeyan K, Kumar N, Naitam RK, Tiwary P. SOC prediction using Random Forest model. *J Sustain For.* 2025;44(10):1269–85. <https://doi.org/10.1080/10549811.2025.2576759>
24. Vestergaard RJ, Vasava HB, Aspinall D, Chen S, Gillespie A, Adamchuk V, et al. Optimized preprocessing and modeling for soil prediction. *Sensors.* 2021;21(20):6745. <https://doi.org/10.3390/s21206745>
25. Walkley A, Black IA. Examination of Degtjareff method for soil organic matter. *Soil Sci.* 1934;37(1):29–38.
26. Savitzky A, Golay MJ. Smoothing and differentiation of data. *Anal Chem.* 1964;36(8):1627–39. <https://doi.org/10.1021/ac60214a047>
27. Duckworth J. Mathematical data preprocessing. In: *Near-infrared spectroscopy in agriculture.* 2004. p. 113–32. <https://doi.org/10.2134/agronmonogr44.c6>
28. Rinnan Å, Van den Berg F, Engelsen SB. Review of preprocessing techniques for NIR spectra. *TrAC Trends Anal Chem.* 2009;28(10):1201–22. <https://doi.org/10.1016/j.trac.2009.07.007>
29. Maleki MR, Mouazen AM, Ramon H, De Baerdemaeker J. Multiplicative scatter correction in NIR. *Biosyst Eng.* 2007;96(3):427–33. <https://doi.org/10.1016/j.biosystemseng.2006.11.014>
30. Barnes RJ, Dhanoa MS, Lister SJ. Standard normal variate transformation. *Appl Spectrosc.* 1989;43(5):772–7. <https://doi.org/10.1366/0003702894202201>
31. Mechram S, Ayu IW, Farni Y. SNV preprocessing for soil nitrogen prediction. *IOP Conf Ser Earth Environ Sci.* 2024;1290:012026. <https://doi.org/10.1088/1755-1315/1290/1/012026>
32. Wold S, Sjöström M, Eriksson L. PLS regression: basic tool of chemometrics. *Chemom Intell Lab Syst.* 2001;58(2):109–30. [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1)
33. Breiman L. Random forests. *Mach Learn.* 2001;45(1):5–32. <https://doi.org/10.1023/A:1010933404324>
34. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation; 2016.
35. Odebiri O, Mutanga O, Odindi J, Slotow R, Mafongoya P, Lottering R, et al. Mapping SOC in arid landscapes. *Geoderma Reg.* 2024;37:e00817. <https://doi.org/10.1016/j.geodrs.2024.e00817>
36. Miloš B, Bensa A. Prediction of SOC using VIS-NIR spectroscopy. *Eurasian J Soil Sci.* 2017;6(4):365–73. <https://doi.org/10.18393/ejss.319208>
37. Liu S, Shen H, Chen S, Zhao X, Biswas A, Jia X, et al. Forest SOC estimation using spectroscopy. *Geoderma.* 2019;348:37–44. <https://doi.org/10.1016/j.geoderma.2019.04.003>
38. Castaldi F, Stenberg B, Liebisch F, Metzger K, Ben-Dor E, Knadel M, et al. SOC estimation using spectral libraries. *Smart Agric Technol.* 2025;101353. <https://doi.org/10.1016/j.atech.2025.101353>
39. Heil K, Schmidhalter U. Evaluation of NIR preprocessing methods. *Sensors.* 2021;21(4):1423. <https://doi.org/10.3390/s21041423>

40. Minu S, Shetty A, Gopal B. Review of preprocessing techniques in hyperspectral soil analysis. *Cogent Geosci.* 2016;2(1):1145878. <https://doi.org/10.1080/23312041.2016.1145878>
41. Yang J, Dong H, Zhang H, Wang J, Dai H, Zhou L, et al. Enhancing SOC estimation using hyperspectral data. *Int J Remote Sens.* 2025;46(13):4941–58. <https://doi.org/10.1080/01431161.2025.2506158>
42. Dotto AC, Dalmolin RSD, Grunwald S, ten Caten A, Pereira Filho W. Preprocessing to reduce model covariables. *Soil Tillage Res.* 2017;172:59–68. <https://doi.org/10.1016/j.still.2017.05.008>
43. Shin SK, Lee SJ, Park JH. Soil property prediction using Vis-NIR and ML: A review. *Sensors.* 2025;25(16):5045. <https://doi.org/10.3390/s25165045>
44. Zhou Y, Biswas A, Hong Y, Chen S, Hu B, Shi Z, et al. Soil profile analysis using hyperspectral imaging. *Geoderma.* 2024;450:117036. <https://doi.org/10.1016/j.geoderma.2024.117036>
45. He J, Matsui T, Tanaka TS. Soil property estimation using FTIR and ML. *Soil Sci Plant Nutr.* 2025;71(6):670–83. <https://doi.org/10.1080/00380768.2025.2558584>
46. Canero FM, Rodriguez-Galiano V, Aragonés D. Machine learning for soil spectroscopy. *Heliyon.* 2024;10(9):e30228. <https://doi.org/10.1016/j.heliyon.2024.e30228>
47. Wu X, Wu K, Hao S, Yu E, Zhao J, Li Y. Machine learning ensemble for soil evolution. *Sci Rep.* 2025;15(1):24332. <https://doi.org/10.1038/s41598-025-10608-8>

Additional information

Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonpublishing.com/journals/index.php/PST/open_access_policy

Publisher's Note: Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc. See https://horizonpublishing.com/journals/index.php/PST/indexing_abstracting

Copyright: © The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited (<https://creativecommons.org/licenses/by/4.0/>)

Publisher information: Plant Science Today is published by HORIZON e-Publishing Group with support from Empirion Publishers Private Limited, Thiruvananthapuram, India.