

Spectroscopic Prediction of Soil Iron Oxides, Organic Carbon, and Calcium Using VIS-NIR and PLSR in Northern Jordan

Abdulla Al-Rawabdeh ^{1*}, Abdel Rahman Alsaleh ^{1,2}, Muheeb Awawdeh ¹

¹Laboratory of Applied Geoinformatics, Department of Earth and Environment Sciences, Yarmouk University, Irbid, Jordan.

²Department of Civil and Environmental Engineering, Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates

Received on 22 October 2025; Accepted on 31 December 2025

Abstract

Soil spectroscopy has not yet been applied in Jordan as a proxy for assessing soil properties. In this study, forty surface soil samples were collected using a cluster random sampling method and analysed for three key soil properties —free iron oxides (Fed), soil organic carbon (SOC), and calcium (Ca), owing to their critical roles in soil fertility and known spectral responsiveness in the VIS-NIR range. Spectroscopy data (325 to 1075 nm) were acquired using a field spectroradiometer, and several pre-processing techniques were applied to enhance correlation accuracy and minimize noise. The analyzed soils were classified under the USDA soil texture system as clay, silty clay, silty clay loam, and clay loam. Geochemical results showed: Fe_d% (0.4- 1.28, mean 0.89), SOC% (1.09- 3.96, mean 1.69), and Ca% (0.08- 10.65, mean 3.19). Partial least squares regression (PLSR) analysis was employed to predict these soil properties from spectral reflectance data, yielding a high coefficient of determination (R²) for both model calibration and validation. The spectral signals of Fe_d, SOC, and Ca were found to be significant across different wavelengths of the VIS–NIR region, confirming their spectral activity. The performance of the PLSR models was validated, and the spectral signals of Fe_d, SOC, and Ca were found to be significant across various wavelengths within the visible and near-infrared (VIS–NIR) region (325–1075 nm). The applied spectral pre-treatment techniques and characteristic variable selection methods notably enhanced model performance and predictive accuracy. The findings demonstrate that VIS–NIR spectroscopic analysis can serve as a reliable and efficient alternative to traditional laboratory methods for determining key soil properties in northern Jordan. This study represents an important step toward establishing a national spectroscopic library of soil materials in Jordan.

© 2026 Jordan Journal of Earth and Environmental Sciences. All rights reserved

Keywords: soil; field spectroscopy; partial least squares regression; iron oxides; soil organic carbon; Jordan.

1. Introduction

Soil forms the uppermost layer of the Earth's surface and develops gradually over time (Plummer et al., 2016). Traditional soil survey methods rely on field sampling at multiple locations in the field, followed by laboratory analyses to determine soil properties, and interpolation techniques to map these properties across the study area. However, such methods are costly, time-consuming, and often yield low- accuracy interpolation results because they fail to fully capture spatial variations in soil characteristics (Zhang et al., 2017).

Modern techniques, including remote sensing and geographical information systems (GIS), have been developed to estimate soil properties indirectly and present the results in digital format. Remote sensing techniques can accurately estimate several physical and chemical soil properties- such as texture, organic matter (OM), moisture content, salinity, and iron content (Kumar et al., 2013), while significantly reducing cost and time compared to traditional mapping. Nevertheless, unlike conventional approaches that analyze deeper soil horizons, remote sensing typically examines only the upper few centimeters of the surface soil layer (Fatholouloumi et al., 2020).

Because different materials exhibit unique reflectance

and absorption behaviors at specific wavelengths, they can be distinguished by their spectral reflectance signatures. Soil spectral reflectance is influenced by both the chemical and physical properties (Kumar et al., 2013; Pereira et al., 2017). Several external factors, such as land cover, vegetation, and land use can affect soil reflectance and should be considered to avoid misinterpretation of results (Kumar et al., 2013). In dryland environments, where such influences are minimized, soil properties, such as moisture content and organic carbon, can be more reliably estimated from reflectance analysis under controlled laboratory conditions (Pereira et al., 2017).

In recent years, numerous studies have employed hyperspectral and multispectral remote sensing techniques to investigate soil properties. For example, Camargo et al. (2015) analyzed the spatial variability of clay, iron oxide, and adsorbed phosphate in Brazilian Oxisols using diffuse reflectance spectroscopy with a PerkinElmer Lambda 950 spectrophotometer. Using partial least squares regression (PLSR), they achieved high correlations between VIS–NIR spectral reflectance (380–2500 nm) and the measured soil attributes. Similarly, Demattê et al. (2015) investigated UV-VIS-NIR Spectroscopy as an alternative soil mapping technique, obtaining promising calibration and validation accuracy that confirmed its suitability for soil attribute estimation in Brazil. The study area was divided into two

* Corresponding author e-mail: abd_rawabdeh@yu.edu.jo

subareas: Soil samples from subarea 1 were used for model calibration, while those from subarea 2 were reserved for validation. The resulting models demonstrated high predictive accuracy, confirming that remote sensing provides a reliable and efficient alternative to conventional soil mapping methods. Forkuor et al. (2017) combined remote sensing and laboratory analysis to assess six soil characteristics, which are cation exchange capacity (CEC), soil organic carbon (SOC), silt, clay, nitrogen, and sand across a 580 km² watershed in southwestern Burkina Faso. Their study demonstrated that Landsat 8 and RapidEye satellite data, particularly in the SWIR and NIR bands, were highly predictive of soil variability.

Yu et al. (2018) generated geochemical soil property maps in China using hyperspectral imagery from the space-borne HJ-1A satellite, supplemented by field spectra collected with an SVC HR1024 spectroradiometer and validated through laboratory analysis. The authors mapped organic carbon, phosphorus, nitrogen, and potassium using a stepwise regression model, achieving high predictive performance. In northern Jordan, Landsat-8 and Sentinel-2 imagery were used combined with support vector machine (SVM) classification, to analyze spatial variations in soil surface colour. They observed strong correlations between red soil distribution, precipitation gradients, and geomorphological features. This approach proved effective for quantifying soil colour variations related to climatic and geomorphological factors.

Previous spectroscopic studies conducted in northern Jordan have primarily focused on qualitative or semi-quantitative relationships between soil spectral characteristics and surface properties such as soil color and iron oxide presence, using instruments such as the ASD FieldSpec® 3 Hi-Res spectroradiometer (e.g., Sahwan et al., 2021). While these studies demonstrated the sensitivity of visible-range spectroscopic data to surface soil attributes, they did not develop quantitative predictive models linking field-acquired spectra to laboratory-measured geochemical soil properties. Moreover, their findings revealed a strong relationship between visible-range spectral data and free iron oxides, whereas no significant correlation with soil organic carbon was reported.

The present study advances beyond spectral-color associations by integrating field spectroscopy with laboratory analyses and Partial Least Squares Regression (PLSR) to quantitatively predict free iron oxides (Fed), soil organic carbon (SOC), and calcium (Ca). The primary goal of this research is to evaluate soil spectral reflectance as a proxy for laboratory-based soil property estimation by modeling the relationships between VIS-NIR spectra and key geochemical parameters. Previous large-scale soil mapping efforts in Jordan, such as the National Soil Map and Land Use Project (NSMLUP, 1989–1995; MoA, 1995), relied mainly on USDA soil classification and produced limited-resolution maps (1:250,000) covering selected regions (Ababsa, 2013). Consequently, the adoption of advanced remote sensing-based techniques is essential for generating accurate, high-resolution digital soil maps.

Therefore, the main objective of this research is to assess the feasibility of using VIS-NIR field spectroscopy as a quantitative alternative to conventional laboratory soil analyses in northern Jordan. By developing and validating PLSR-based predictive models for key soil fertility indicators—free iron oxides, soil organic carbon, and calcium—this study provides a methodological framework to support future digital soil mapping initiatives. Rather than presenting a comprehensive national soil spectral library, the outcomes of this research represent a foundational contribution toward such an initiative, demonstrating the potential structure, performance, and limitations of spectroscopy-based soil characterization under Jordanian environmental conditions.

2. Materials and Methods

2.1 Soils in Jordan

The spatial distribution of soils in Jordan is primarily controlled by climate and topography. The country's climate ranges from arid and desert conditions in the eastern and southern regions to sub-humid Mediterranean conditions in the northern and western highlands. Jordan's geological diversity contributes significantly to its soil variability, as the landscape comprises various lithological units, including chalk, limestone, sandstone, chert, marl, and assorted Pleistocene and Holocene deposits such as alluvial, eolian, and lava flows materials (Bender et al., 1974). The contrasting physical and chemical characteristics of these parent materials result in a wide range of soil properties across the country (Bender et al., 1974). A substantial portion of Jordan's soils developed on carbonate-rich and basaltic rocks, which are common in the northern and eastern regions (Ababsa, 2013).

Several classification systems have been applied to categorize Jordanian soils. Table 1 summarizes the classifications proposed by Moorman (1959) and their corresponding equivalences in the Food and Agriculture Organization of the United Nations (FAO, 2006) and the United States Department of Agriculture (USDA, 1999) classification systems. The USDA Soil Taxonomy is particularly relevant, as it is internationally recognized for its hierarchical structure and practical applicability. This system classifies soils progressively from orders (the broadest level, e.g., Alfisols, Mollisols) down to suborders, great groups, subgroups, and soil series, the latter representing local-scale soil types with specific physical and chemical properties (Cordova et al., 2005).

Table 1. Comparison of soil classifications in Jordan as proposed by Moorman (1959) and their equivalents according to the FAO (2006) and USDA (1999) systems (after Cordova et al., 2005).

Moorman, 1959	FAO, 2006	USDA, 1990
Red Mediterranean soils (Mediterranean woodlands and forest-steppe soils)	Luvic calcisols Vertic luvisols Calcic luvisols Chromic luvisols Chromic cambisols	Xerochrepts, Xerothents, Chromoxererts
Yellow Mediterranean soils (soils in transition to steppe)	Calcic calixerolls Calcic phaeozems Luvic phaeozems	Calsixerolls Xerochrepts
Yellow soils (steppe and desert soils)	Luvic xerosols Calcic fluvisols Calcic regosols	

2.2 Study Area

The study area is situated within the administrative boundaries of the Greater Irbid Municipality in northern Jordan (Figure 1) and covers an approximate area of 185 km². Geographically, it extends between 32°26'00.17" N – 32°39'44.87" N and 35°51'17.52" E – 35°58'40.49" E. The region experiences a semi-arid Mediterranean climate, with most of the annual rainfall occurring between October and April (Al Qudah et al., 2015). The average temperature ranges from 8.0-15 °C in winter to 22-31 °C in summer (Al-Bakri et al., 2012).

According to the National Soil Map and Land Use Project (NSMLUP, 1989–1995), the dominant soil types in the area include calcixerollic, lithic, typic xerochrepts, and typic and entic chromoxererts (Ababsa, 2013). Geomorphologically, the study area lies within the Houran Plains, characterized by extensive flat terrain with thick red, clay-rich soils developed over basaltic and carbonate parent materials (Al Qudah et al., 2015). These soils play a vital role in regional agriculture and hydrological processes, making the area suitable for assessing soil spectral variability and geochemical composition.

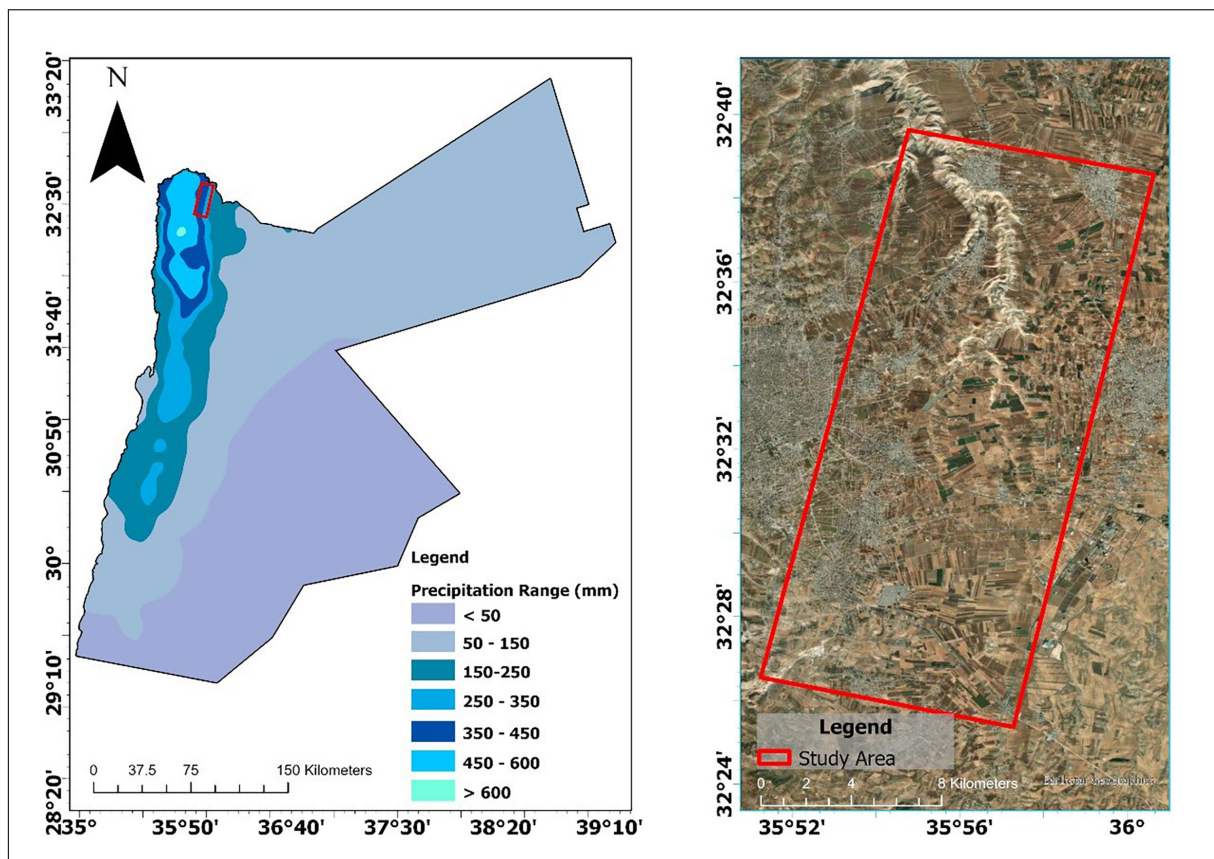


Figure 1. Map showing annual precipitation and the location of the study area in northern Jordan (derived from Landsat-8 imagery).

Most of the study area is covered by soil deposits; however, several rock formations are also exposed, belonging to the Balqa and Ajloun Groups (Figure 2). The Balqa group consists of six formations that are predominantly composed of chert beds and fossiliferous chalk, which is notably rich in vertebrate remains (Alnimrat et al., 2022). In contrast, only the Wadi As Sir Formation of the Ajloun Group is exposed within the study area. This formation is characterized by bedded chert, chalk, limestone, phosphorite, dolomite, and porcelanite (Abed, 2017).

These lithological units serve as the parent materials for the region's soils. The nature of the parent material exerts a significant influence on the concentration and type of nutrient elements within the soil—an effect that is more pronounced in young soils and diminishes as soils age and undergo weathering. Understanding the mineralogical composition of these rocks is therefore essential for interpreting their role in soil genesis and the region's geochemical variability (Singh and Schulze, 2015).

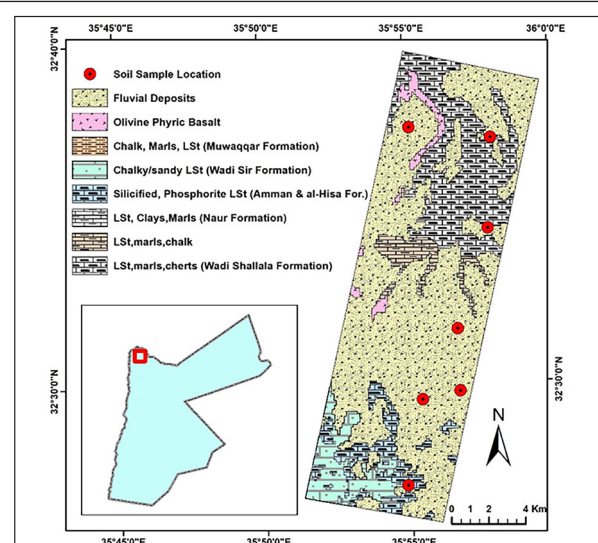


Figure 2. Simplified geological map of the formations exposed within the study area, northern Jordan, modified after Moh'd (1997) and Abdelhamid (1993).

2.3 Data Collection

Soil samples and corresponding spectra data were collected from the field using an ASD Handheld 2 Spectroradiometer (Figure 3). The spectral reflectance measurements were conducted prior to plowing season to ensure undisturbed soil surface conditions and reliable calibration. Following each spectral acquisition, approximately 2 kg of surface soil was collected from the same sampling point and sealed in labeled plastic bags for subsequent physical and geochemical analyses. The ASD Handheld 2 Spectroradiometer (Figure 4) is a portable non-destructive field instrument capable of rapid measurements within the visible to near-infrared (VIS–NIR) range of 325–1075 nm, covering 751 spectral bands. To maintain data accuracy, the instrument was calibrated every 10 minutes using a white reference panel. Several precautionary steps were taken to minimize external noise and measurement error. Field operators wore dark clothing to reduce light scattering, and the spectroradiometer sensor was maintained at a fixed height of approximately 2.0 m above the soil surface, oriented toward the sun to minimize shadow effects and ensure stable illumination during measurements. Measurements were also conducted away from electrical lines and reflective surfaces to prevent external electromagnetic and optical interference.

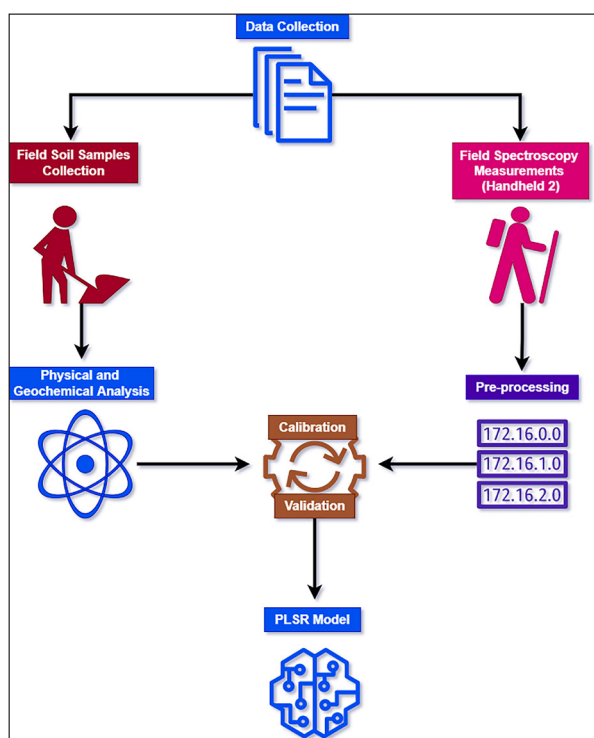


Figure 3. Flowchart illustrating the methodological framework adopted in this study, including field data collection, laboratory analysis, and statistical modeling procedures.

The soil sampling locations (Figure 5) were carefully selected to represent the diversity of soil types across the study area and to minimize the influence of vegetation on the soil’s spectral signal. Sites with minimal vegetation cover were prioritized to avoid interference from plant material during spectral measurements. Two temporal constraints guided data collection: 1) measurements were conducted during periods of lowest vegetation cover, and 2) sampling occurred only when solar radiation was

sufficient, following the recommendations of CropsScan™ (2001). All field measurements were carried out around noon, when the sun was near its zenith and cloud cover was minimal, to ensure consistent illumination conditions. A clustered random sampling strategy was adopted to ensure representative coverage of soil variability across the study area. Seven sampling clusters were selected based on differences in dominant soil types, land-use patterns, and geomorphological settings, as identified from existing soil and geological maps, and field reconnaissance. Within each cluster, approximately five to six surface soil samples were collected at randomly selected locations, resulting in a total of forty samples. This sampling design was chosen to capture both local-scale variabilities within clusters and broader spatial heterogeneity across the study area, while minimizing the influence of vegetation cover on the soil spectral signal.



Figure 4. Calibration process of the ASD Handheld 2 Spectroradiometer using a white reference panel to ensure measurement accuracy during field data acquisition.

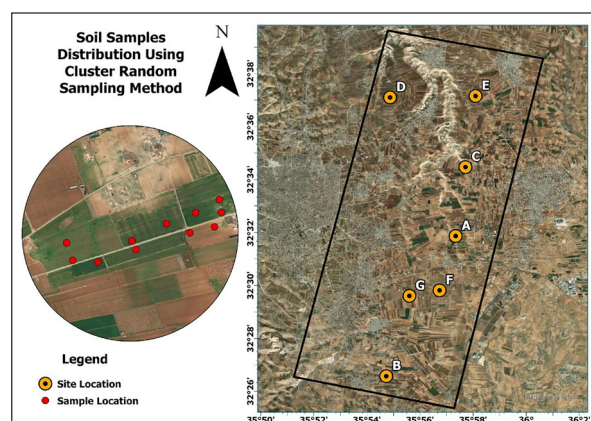


Figure 5. Spatial distribution of soil sampling clusters across the study area, illustrating the locations of the seven sampling sites. Site A is shown as an example of the randomly collected soil samples within a cluster.

2.4 Laboratory Analysis

Laboratory analyses were conducted to quantify the key soil properties targeted in this study and to provide reference measurements for calibrating and validating the spectroscopic models. The collected soil samples were subjected to standardized chemical analysis procedures to determine soil organic carbon (SOC), total calcium (Ca),

and free iron oxides (Fed). These parameters were selected due to their relevance to soil fertility, pedogenic processes, and their known spectral responsiveness in the VIS–NIR region. The analytical methods applied for each parameter are described in detail in the following subsections.

2.4.1 Soil Organic Carbon (SOC)

Soil organic carbon (SOC) plays a vital role in determining soil physical structure, influencing aggregation, porosity, and overall fertility (Sparks et al., 1996). Soil organic matter exists at various stages of decomposition, encompassing microbial biomass as well as plant and animal residues. The Walkley and Black (1934) wet oxidation method, as described in Sparks et al. (1996), was used to quantify SOC in the collected samples. The absorbance of calibration standards and sample solutions was measured at 600 nm using a Thermo Electron Corporation spectrophotometer housed in the Department of Earth and Environmental Sciences, Yarmouk University.

2.4.2 Calcium Content (Ca)

Calcium is a fundamental element for plant growth, contributing to cellular structure and fruit firmness, and indirectly influencing chlorophyll production through its interaction with iron (Fan et al., 2016). The sodium carbonate fusion method (Sparks et al., 1996) was employed to decompose the soil samples and estimate total calcium (Ca) and iron (Fe) contents. The digested samples were subsequently analyzed using a PerkinElmer AAS 200 atomic absorption spectrophotometer available in the Department of Chemistry, Yarmouk University.

2.4.3 Free Iron Content (Fed)

The citrate–bicarbonate–dithionite (CBD) extraction method was applied to determine free iron oxides (Fed) in the soil samples, as recommended for large sample sets (Sparks et al., 1996). The iron concentration in the extracted CBD solution was quantified using a PerkinElmer AAS 200 atomic absorption spectrophotometer at the Department of Chemistry, Yarmouk University.

2.5 Statistical analysis

The statistical analysis was conducted using Unscrambler® X v10.4 to develop a robust regression model that predicts unknown geochemical values for soil samples based on their spectroscopic attributes. Unscrambler® X is a commercial software package designed for multivariate data analysis, particularly useful in analytical applications such as near-infrared (NIR) spectroscopy, where it supports both calibration and predictive model development for real-time material analysis. In this study, two primary datasets were employed: 1) Response matrix (Y): geochemical data obtained from laboratory analyses, and 2) Predictor matrix (X): spectral reflectance data derived from field spectroscopy measurements. The analytical workflow comprised five main stages: 1) pre-processing of the spectral data to enhance signal quality and remove noise; 2) application of partial least squares (PLS) regression to model relationships between spectral and geochemical variables; 3) model calibration using reference data to establish the predictive relationship; 4) model validation to evaluate accuracy and generalizability; 5) model testing to assess the

final predictive performance.

In addition to the coefficient of determination (R^2), model performance was evaluated using the root mean square error of calibration (RMSEC), root mean square error of cross-validation (RMSECV), residual prediction deviation (RPD), and bias. RMSEC and RMSECV were used to quantify calibration accuracy and prediction uncertainty, respectively, while RPD—defined as the ratio between the standard deviation of the reference data and RMSECV—was used as an indicator of model robustness. According to commonly accepted criteria, RPD values greater than 2 indicate good predictive performance, whereas values between 1.4 and 2 suggest moderate predictive capability.

Model bias was calculated as the mean difference between predicted and measured values to assess systematic over- or underestimation. Bias values close to zero indicate unbiased predictions.

The optimal number of latent variables (PLSR factors) was determined by minimizing RMSECV and avoiding overfitting, following the guidelines provided in Unscrambler® X. Only factors that contributed significantly to reducing prediction error were retained in the final models.

The inclusion of these diagnostic metrics demonstrates that the developed PLSR models for free iron oxides (Fed), soil organic carbon (SOC), and calcium (Ca) exhibit satisfactory calibration accuracy, acceptable prediction uncertainty, and stable generalization performance based on their VIS–NIR spectral signatures.

2.5.1 Pre-processing of the spectroscopic data

Spectral pre-processing was applied to the raw VIS–NIR reflectance data to reduce noise, correct for instrumental and environmental effects, and enhance the correlation between spectral variables and laboratory-measured soil properties. All transformations were performed using Unscrambler® X v10.4 following recommended procedures for field-acquired soil spectra to optimize the relationship between the spectral data matrix and the laboratory-measured contents of free iron oxides (Fed), soil organic carbon (SOC), and calcium (Ca).

A series of transformations was applied sequentially: Baseline correction was first applied to adjust for spectral offsets, caused by variations in illumination intensity and sensor positioning. The minimum reflectance value within each spectrum was used as a reference, ensuring spectral normalization and improving inter-sample comparability.

Correlation Optimized Warping (COW) was subsequently employed to correct minor wavelength shifts along the spectral axis (x) resulting from sensor instability and field measurement conditions. Default segment length and slack parameters recommended by Unscrambler® X were used, which improved spectral alignment and reduced misregistration between samples.

A multiplicative Scatter Correction (MSC) was applied to minimize scattering effects related to differences in soil particle size, surface roughness, and moisture conditions (Unscrambler® X, 2014). MSC improved model linearity

by reducing multiplicative and additive scattering, leading to improved calibration stability, particularly for Fed and Ca models. While SOC data were excluded from this stage due to their specific response characteristics.

A deresolve transformation was then implemented to reduce spectral resolution and suppress high-frequency noise while preserving diagnostically relevant spectral features. This step reduced collinearity among adjacent wavelengths and improved model robustness without loss of predictive information.

Finally, Orthogonal Signal Correction (OSC) was applied to remove spectral variance uncorrelated to the response variables. OSC effectively filtered non-informative components from the X-matrix, resulting in improved signal-to-noise ratios and enhanced prediction accuracy during cross-validation.

The combined application of these pre-processing steps resulted in higher coefficients of determination (R^2) and reduced prediction errors in PLSR models compared to those, developed using raw spectra, confirming their effectiveness in improving model performance.

2.5.2 Model Calibration

The PLSR model calibration was conducted using diagnostic parameters such as score plots, loading plots, and regression coefficients to evaluate model structure and performance. The PLSR approach organizes samples along model components based on the variation within the predictor matrix (X)—representing the spectral data—and the response matrix (Y)—representing the laboratory-measured soil properties. PLS loadings describe how each variable in the X- and Y-spaces is contributing to the model components, providing insight into the correlation structure between spectral variables and soil attributes. To identify potential outliers, the X–Y relationship plot (t-scores from X versus u-scores from Y) was examined to detect samples that deviated significantly from the general data trend. Each variable was weighted by its regression coefficient, reflecting its importance in predicting a specific Y-response. The regression coefficients also served as indicators of model interpretability and predictive strength. Ideally, the predicted versus reference (measured) plot should exhibit a linear relationship with a slope near 1 and a correlation coefficient (R^2) close to unity, indicating high model accuracy. In a predicted-versus-reference plot, the predicted values should show a straight-line relationship with the measured values, ideally with a slope of 1 and a correlation of 1 or near 1. These diagnostic parameters were employed to iteratively exclude outlier samples, irrelevant variables, and spectral noise, ultimately producing a stable and optimized PLSR model with maximum predictive accuracy (Unscrambler® X, 2014).

2.5.3 Model Validation

Model validation is a critical step in assessing the reliability and predictive performance of a regression model that has been calibrated using empirical data. Validation estimates the level of uncertainty associated with future predictions and determines how effectively the model can

predict new observations of the same type as those used in its development. A model is considered valid when its prediction uncertainty is minimal, and its performance remains stable across independent datasets (Unscrambler® X, 2014).

In cases where a large number of samples are available (typically more than 50), independent validation or test-set validation can be applied. However, in this study, the dataset consisted of 33 samples, which was insufficient for reliable independent testing. Therefore, a cross-validation approach was adopted as a more suitable alternative. Cross-validation systematically partitions the dataset into subsets, iteratively using one subset for testing and the remainder for model calibration. This approach provides a robust measure of model accuracy and generalization capability, particularly when working with limited sample sizes.

2.5.4 Partial Least Squares Regression (PLS Regression)

The PLSR method, also known as Projection to Latent Structures (PLS), models the predictor (X) and response (Y) matrices simultaneously to extract latent variables from X that best predict the corresponding variables in Y. These latent variables, referred to as factors, are analogous to the principal components in Principal Component Analysis (PCA) but are specifically optimized to maximize the covariance between X and Y (Unscrambler® X, 2014).

Unlike Principal Component Regression (PCR), which first performs PCA on the X-matrix and then regresses the resulting component scores (T) against the Y data, PLSR integrates these steps. This simultaneous modeling approach allows PLSR to capture the most relevant spectral information while minimizing residual error, often achieving high predictive accuracy with fewer factors than PCR. Consequently, PLSR provides an efficient and robust framework for establishing quantitative relationships between soil spectral reflectance data and their corresponding laboratory-measured properties.

3. Results and Discussion

This section presents and discusses the results obtained from both laboratory analyses and spectroscopic modeling. It is organized into three main parts: first, the spatial variability and descriptive statistics of the laboratory-measured soil properties are reported; second, the performance of the spectral pre-processing and PLSR models is evaluated using calibration and validation metrics; and finally, the predictive results are interpreted in relation to soil spectral behaviour and compared with findings from previous studies. The discussion highlights the strengths and limitations of the developed models and their implications for soil spectroscopy applications in semi-arid environments.

3.1 Laboratory Analysis

Soil organic carbon (SOC) is primarily derived from the biological decomposition of plant residues and other organic materials within the soil (Jenkinson, 1966). The upper soil horizon (<30 cm) typically contains the highest SOC concentrations, as it is directly influenced by climatic factors and human activities such as cultivation, grazing, and fertilization (Batjes, 2006).

The laboratory analyses revealed substantial variation

in SOC content across the study area (Table 2). SOC values ranged from a minimum of 1.0% in sample no. 31 (Site E) to a maximum of 3.9% in samples no. 14 and 20 (Sites B and C), with a mean of 1.69%. This considerable variation likely reflects spatial differences in organic matter inputs, biological activity, and local microclimatic conditions.

The coefficient of variation (CV) for SOC was 36.1%, indicating high variability among the samples. Such variation suggests that SOC distribution within the study area is influenced by both land-use practices and environmental heterogeneity, consistent with findings from similar semi-arid Mediterranean landscapes.

Table 2. Descriptive statistics of soil properties analyzed in the study.

Soil parameter	n	Min	Max	Mean	SD	CV
SOC %	40	1.09	3.96	1.69	0.61	36.17
Ca %	40	0.08	10.65	3.19	2.22	69.83
Fed %	40	0.40	1.28	0.89	0.22	24.92

n, number of samples; SD, standard deviation; CV, coefficient of variation (=Mean/SD* 100%)

The results of the sodium carbonate fusion analysis are summarized in Table 2. The total calcium (Ca) content exhibited substantial variability, ranging from a minimum of 0.08% in sample no. 25 (Site D) to a maximum of 10.6% in sample no. 9 (Site A), with a mean value of 3.19%. The coefficient of variation (CV) was 69.83%, indicating an extremely high degree of dispersion among the samples. This pronounced variability likely reflects differences in carbonate content, soil management practices, and microenvironmental conditions across the study sites.

In contrast, the total iron (Fe) content across sites A to E was relatively uniform. This consistency suggests that iron concentration in the soils is primarily governed by parent material composition and pedogenic processes, which appear to be similar across the sampled locations.

Generally, free iron in soils originates from secondary iron oxides, either absorbed or precipitated onto mineral surfaces or bound in iron–organic matter complexes. Iron occurs in two oxidation states—ferrous (Fe^{2+}) and ferric (Fe^{3+})—both of which play essential roles in plant metabolism, respiration, and photosynthesis (Jones, 2020). The relative proportions of these oxidation states are strongly influenced by soil pH and pore-water content.

The analysis of free iron oxides (Fed) revealed a variation with values ranging from 0.40% (sample no. 13, Site B) to 1.28% (sample no. 28, Site D) (Table 2). These differences likely reflect localized variations in drainage conditions, oxidation–reduction dynamics, and organic matter associations, which collectively influence iron mobility and availability in the soil matrix.

3.2 Pre-processing and Correlation models

The relationships between the hyperspectral reflectance data and the three selected soil properties—free iron oxides (Fed), soil organic carbon (SOC), and calcium (Ca)—were examined through Partial Least Squares (PLS) regression analysis. After applying the pre-processing procedures described in Section 2.5.1, a robust and stable PLSR model

was developed for each soil property.

Model validation was subsequently performed to assess model stability and to evaluate the predictive confidence for estimating unknown soil samples. The validation results confirmed that the applied preprocessing transformations effectively enhanced the correlation between spectral signatures and laboratory-measured soil parameters, yielding reliable predictive models for Fed, SOC, and Ca.

3.2.1 Pre-processing of the spectroscopic data

Several spectral transformation techniques were applied to the raw hyperspectral data to enhance signal quality, remove noise, and achieve optimal correlations between spectral reflectance and the measured soil properties (Figure 6a–d).

Baseline correction was first applied to adjust for spectral offsets and normalize reflectance intensity (Figure 6a). The Correlation Optimized Warping (COW) method was then used to realign the spectra and correct minor wavelength shifts along the x-axis (Figure 6b). Subsequently, Multiplicative Scatter Correction (MSC) was implemented to minimize scattering effects caused by differences in particle size and surface roughness (Figure 6c). Finally, the Deresolve transformation was applied to reduce excessive spectral resolution and eliminate high-frequency noise, thereby improving data smoothness and model interpretability (Figure 6d). These pre-processing steps collectively enhanced the stability and accuracy of the subsequent PLSR models developed for Fed, SOC, and Ca prediction.

3.2.2 Correlation between soil properties and reflectance data (PLS regression)

Three stable Partial Least Squares Regression (PLSR) models were developed for predicting free iron oxides (Fed), soil organic carbon (SOC), and calcium (Ca) content, each demonstrating credible validation accuracy and strong predictive confidence.

For Fed, the PLSR model achieved coefficients of determination (R^2) of 0.82 for calibration and 0.85 for cross-validation, using only three latent factors (Figure 7c–d). The first two PLSR factors explained 88% of the variance in the X-matrix (spectral data) and 82% in the Y-matrix (Fed concentrations) (Figure 7a). The score plot showed a clear, consistent distribution of samples across the first two factors, confirming the model's robustness.

The regression coefficient plot identified the spectral bands most influential in predicting Fed (Figure 7b). The important X-variables (spectral bands) were characterized by strong positive or negative regression coefficients, distributed intermittently along the wavelength axis. According to Liu (2019), soil iron oxides exhibit maximum positive correlations at 585, 747, 1153, 1380, and 2187 nm. In this study, the critical wavelengths associated with iron oxides were distributed primarily across the visible and near-infrared (VIS–NIR) region (325–1075 nm). This pattern aligns with findings by Heller and Ben-Dor (2020), who noted that both soil organic matter (SOM) and Fed are spectrally active within the 400–1000 nm range, often

exhibiting overlapping effects.

To enhance model precision, spectral wavelengths that overlapped significantly with SOM reflectance

characteristics were excluded from the set of important variables, thereby isolating the spectral contribution of Fed and improving model interpretability.

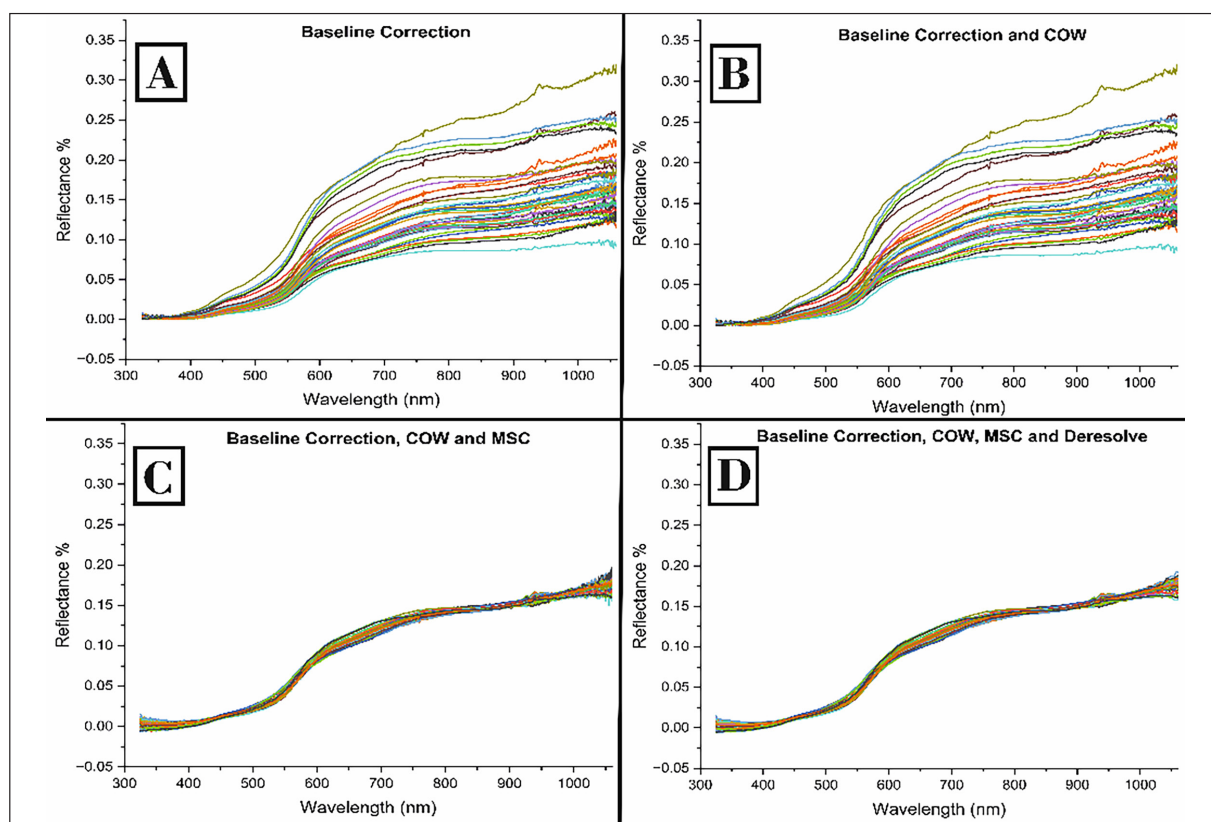


Figure 6. Pre-processing transformations applied to the spectroscopic data: (a) baseline correction to adjust spectral offset, (b) correlation optimized warping (COW) to correct wavelength shifts, (c) multiplicative scatter correction (MSC) to remove scattering effects, and (d) deresolve transformation to reduce high-frequency noise and improve signal smoothness.

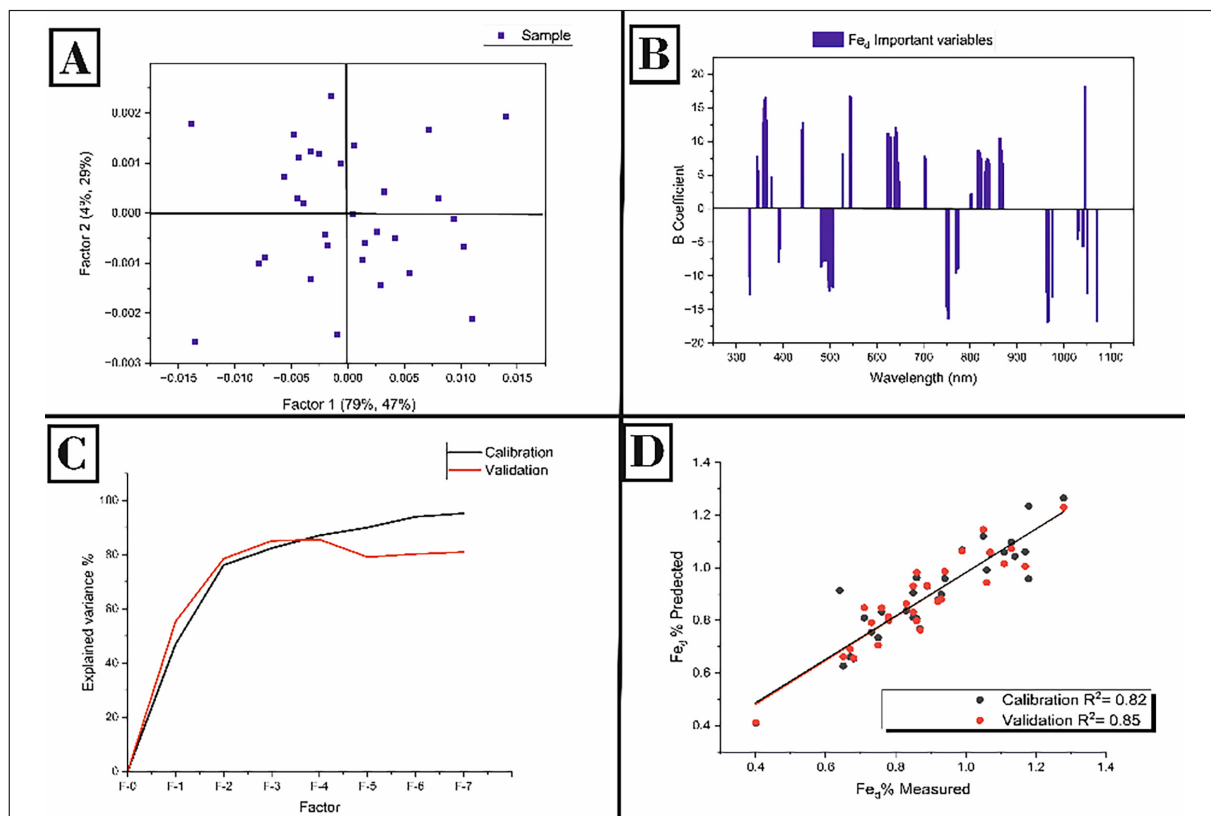


Figure 7. Partial Least Squares Regression (PLSR) model for free iron oxides (Fed) developed using 32 soil samples.

The strong predictive performance of the Fed model is consistent with previous studies that report high sensitivity of visible wavelengths to iron oxide absorption features (e.g., Sahwan et al., 2021; Heller and Ben-Dor, 2020). The relatively high RPD values obtained for Fed indicate robust model reliability, reflecting the pronounced spectral activity of iron oxides in the VIS–NIR region.

The PLSR model for soil organic carbon (SOC) demonstrated strong predictive performance, achieving a coefficient of determination (R^2) of 0.83 for calibration and 0.72 for cross-validation, with four latent factors employed (Figure 8b–c). The score plot illustrates the distribution of samples along the first two factors, where Factor 1 accounted for 100% of the variance in the X-matrix (spectral data) and

36% of the variation in the Y-matrix (SOC), while Factor 2 explained an additional 34% of the variance in Y (Figure 8b).

According to Xu et al. (2020), SOC is highly correlated with spectral features in the visible and near-infrared (VIS–NIR) regions, particularly at wavelengths around 400, 530, 610, 800, 1000, and 1100 nm. In the present study, the identified spectral bands for SOC prediction (Figure 8b) closely align with these ranges. The selected wavelengths are associated with variations in chromophores and the optical properties of humic acids (Xu et al., 2020). Moreover, the significant wavelengths observed in the NIR region correspond to vibrational overtones of C–H, O–H, and N–H bonds, which are characteristic of organic carbon compounds.

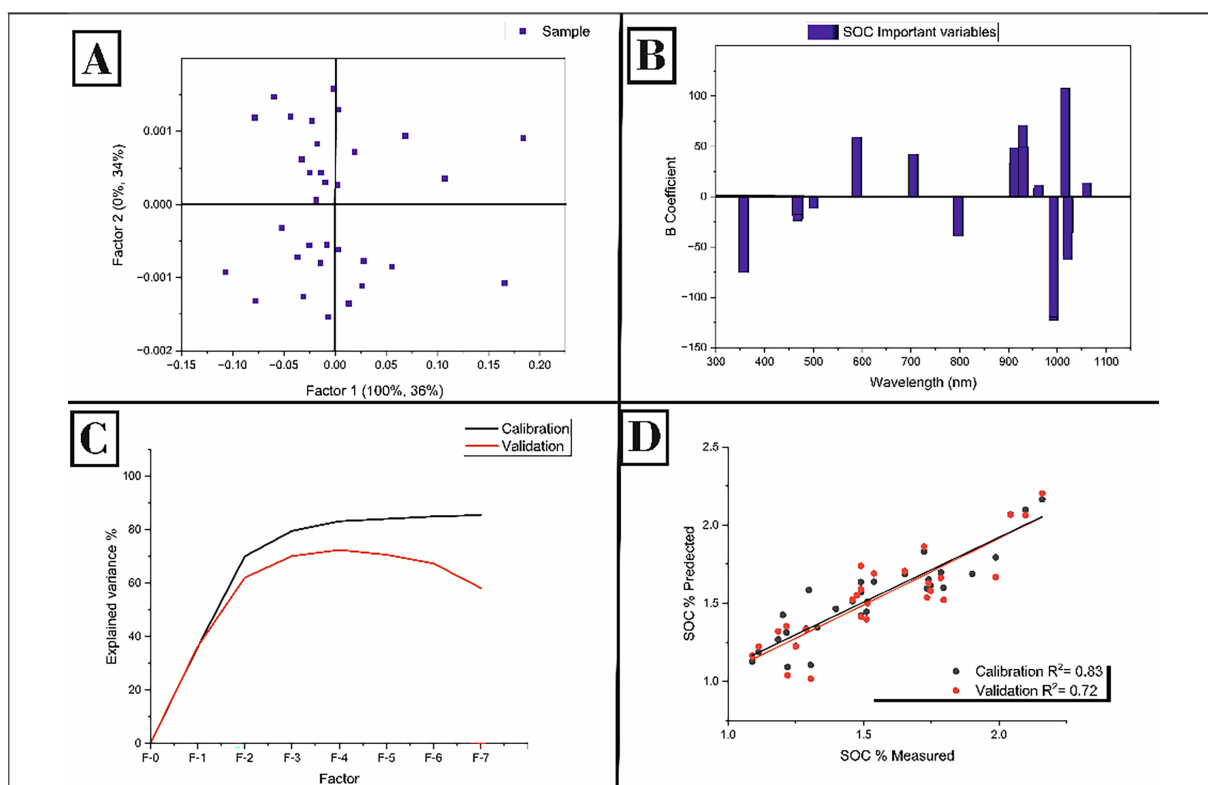


Figure 8. Partial Least Squares Regression (PLSR) model for soil organic carbon (SOC) developed using 32 soil samples. The model achieved an R^2 of 0.83 for calibration and 0.72 for cross-validation, utilizing four latent factors to explain the spectral–SOC relationship.

The SOC model exhibited lower validation accuracy compared to Fed and Ca, which can be attributed to the complex and indirect spectral behavior of organic carbon. SOC absorption features often overlap with those of iron oxides and soil moisture, reducing model specificity, particularly in semi-arid soils with relatively low organic matter content. Similar limitations have been reported in other VIS–NIR studies conducted in arid and semi-arid environments (Xu et al., 2020; Gomez, 2008)

The PLSR model developed for calcium (Ca) achieved a coefficient of determination (R^2) of 0.87 for calibration and 0.79 for validation (Figure 9a, c–d). The model utilized three latent factors, which explained 93% of the variance in the X-matrix (spectral data) and 58% of the variance in the Y-matrix (calcium content). The score and regression coefficient plots revealed a clear distribution of samples across the first two factors, as well as the key spectral variables contributing to Ca prediction (Figure 9b).

The correlation between calcium and VIS–NIR spectral reflectance is not yet fully understood. However, in this study, the most significant spectral bands correlated with Ca were identified within the visible to near-infrared range after data pre-processing, showing both positive and negative relationships (Figure 9b). According to Gomez (2008), important wavelengths for estimating CaCO_3 content in soils are typically observed near 500, 600, and 1430 nm, which are closely associated with soil colour—an attribute linked to calcium carbonate concentration. Gomez (2008) identified additional spectral features within the VIS–NIR range, notably in the visible wavelengths (approximately 500–650 nm) related to soil color variations and in the near-infrared region around 1400 nm, which are indirectly influenced by carbonate minerals. The consistency between these spectral regions and those identified in the present study further supports the applicability of VIS–NIR spectroscopy for estimating calcium-related soil properties.

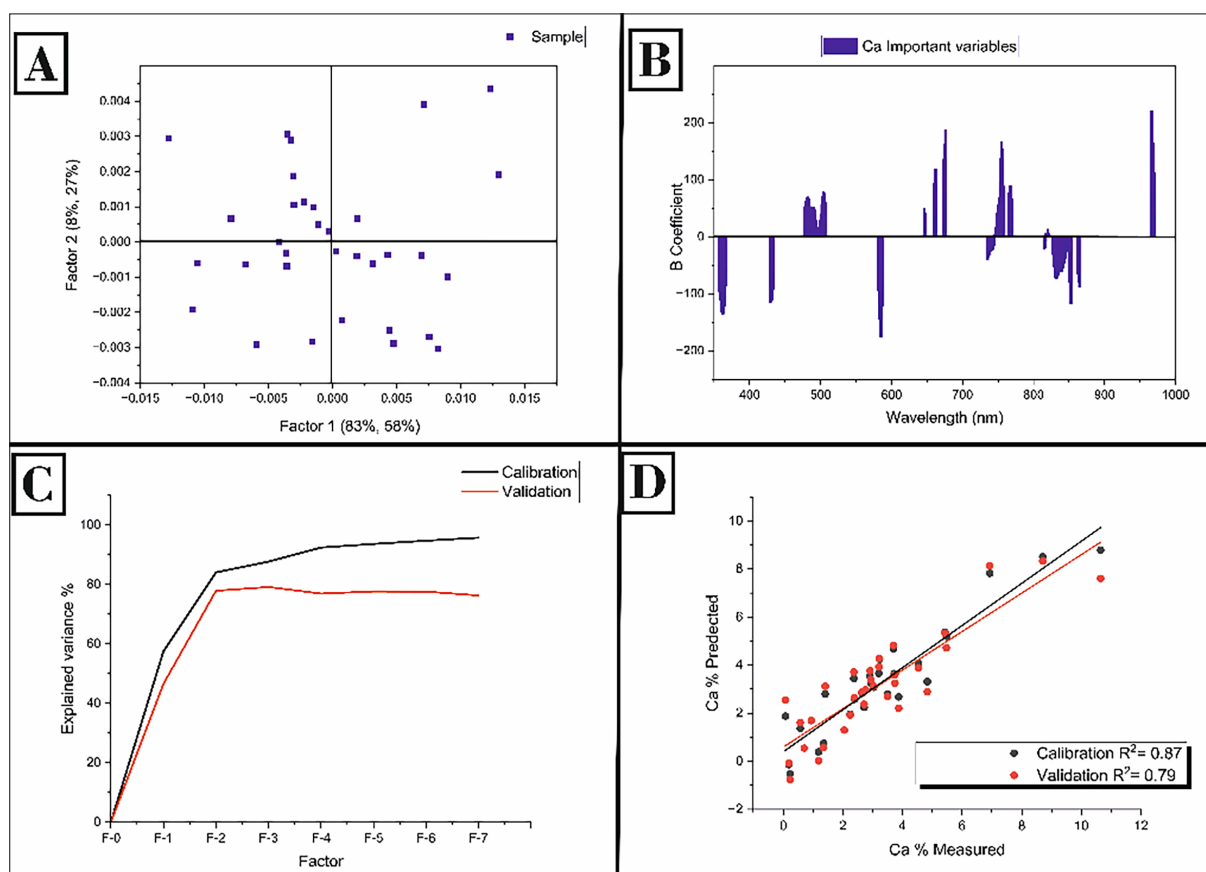


Figure 9. Partial Least Squares Regression (PLSR) model for calcium (Ca) developed using 32 soil samples. The model achieved R^2 values of 0.87 for calibration and 0.79 for validation, with three latent factors explaining 93% of the variance in X and 58% in Y.

The Ca model showed strong calibration performance, likely due to the influence of calcium carbonate on soil brightness and colour, which indirectly affects VIS–NIR reflectance. However, the moderate validation performance suggests that Ca prediction remains sensitive to soil texture and mineralogical variability.

4. Conclusions and Recommendations

This study demonstrates the robustness and efficiency of VIS–NIR field spectroscopy combined with Partial Least Squares Regression (PLSR) for characterizing selected soil properties under semi-arid Mediterranean conditions in northern Jordan. Based on correlations within the 325–1075 nm spectral range, three PLSR models were successfully developed to predict free iron oxides (Fed), soil organic carbon (SOC), and calcium (Ca), showing satisfactory calibration and validation performance.

The results revealed distinct, intermittent spectral responses associated with each soil property across the VIS–NIR region, confirming that Fed, SOC, and Ca are spectrally active and can be reliably modeled using multivariate regression. These findings highlight the potential of soil spectroscopy as a rapid and effective alternative to conventional laboratory analyses for soil property estimation.

Importantly, this research should be viewed as a preliminary but essential step toward the development of a soil spectral database in Jordan. Although the present dataset is spatially limited, it establishes a validated methodological foundation for future large-scale sampling,

model generalization, and integration with remote sensing platforms. Expanding the approach to include a wider range of soil types and geographic regions will be necessary to support the development of standardized spectral calibration models and, ultimately, a comprehensive national soil spectral library.

Furthermore, integrating satellite-based and unmanned aerial vehicle (UAV) platforms with soil spectroscopy has the potential to significantly enhance data acquisition efficiency and spatial coverage. Such integration would enable the generation of high-resolution digital soil maps, supporting precision agriculture and sustainable land management. The availability of up-to-date spectroscopy-based soil information would assist farmers and decision-makers in optimizing crop selection, irrigation practices, and fertilization strategies, thereby improving agricultural productivity and resource efficiency in Jordan.

References

- Ababsa, M.; Kohlmayer, C. Atlas of Jordan: History, Territories and Society; Presses de l'Ifpo: Beirut, Lebanon, 2013.
- Abdelhamid, G. Geological Sheet of Irbid; Natural Resources Authority: Amman, Jordan, 1993.
- Abed, A.M. An overview of the geology and evolution of Wadi Mujib. *Jordan Journal of Natural History* 2017, 4, 6–28.
- Al Qudah, K.; Abu-Jaber, N.; Jaradat, R.; Awawdeh, M. Artificial rainfall tests, soil moisture profiles, and geoelectrical investigations for the estimation of recharge rates in a semi-arid area (Jordanian Yarmouk River Basin). *Environmental Earth Sciences* 2015, 73, 6677–6689. <https://doi.org/10.1007/s12665-015-4094-3>

- Al-Bakri, J.; Saoub, H.; Nickling, W.; Suleiman, A.; Salahat, M.; Khresat, S.; Kandakji, T. Remote sensing indices for monitoring land degradation in a semiarid to arid basin in Jordan. *Proceedings of SPIE* 2012, 8538. <https://doi.org/10.1117/12.974438>
- Alnimrat, R.M.; Alqudah, M.; Al-Shereideh, S.; Alsaleh, A.R.S. Geochemical and petrological analyses of the phosphorite deposits (Amman/Al Hisa Formation) from northwest Jordan. *Arabian Journal of Geosciences* 2022, 15, 1211. <https://doi.org/10.1007/s12517-022-10238-5>
- Batjes, N.H. Soil carbon stocks of Jordan and projected changes upon improved management of croplands. *Geoderma* 2006, 132, 361–371. <https://doi.org/10.1016/j.geoderma.2005.05.009>
- Bender, F. *Geology of Jordan*; Gebrüder Borntraeger: Berlin, Germany, 1974.
- Camargo, L.A.; Júnior, J.M.; Barrón, V.; Alleoni, L.R.F.; Barbosa, R.S.; Pereira, G.T. Mapping clay, iron oxide and adsorbed phosphate in Oxisols using diffuse reflectance spectroscopy. *Geoderma* 2015, 251, 124–132. <https://doi.org/10.1016/j.geoderma.2015.03.017>
- Cordova, C.E.; Foley, C.; Nowell, A.; Bisson, M. Landforms, sediments, soil development, and prehistoric site settings on the Madaba–Dhiban Plateau, Jordan. *Geoarchaeology* 2005, 20, 29–56. <https://doi.org/10.1002/gea.20041>
- CropScan™. *Multispectral Radiometer (MSR): User's Manual and Technical Reference*; CropScan™: Rochester, MN, USA, 2001.
- Demattê, J.A.M.; Alves, M.R.; Gallo, B.C.; Fongaro, C.T.; Souza, A.B.; Romero, D.J.; Sato, M.V. Hyperspectral remote sensing as an alternative to estimate soil attributes. *Revista Ciência Agronômica* 2015, 46, 223–232. <https://doi.org/10.5935/1806-6690.20150026>
- Fan, S.-S.; Chang, F.-H.; Hsueh, H.-T.; Ko, T.-H. Measurement of total free iron in soils by H₂S chemisorption. *Journal of Analytical Methods in Chemistry* 2016, 2016, 1–7. <https://doi.org/10.1155/2016/8015452>
- FAO; ISSS; ISRIC. *World Reference Base for Soil Resources*; FAO: Rome, Italy, 2006.
- Fatholoulumi, S.; Vaezi, A.R.; Alavipanah, S.K.; Ghorbani, A.; Saurette, D.; Biswas, A. Improved digital soil mapping using multitemporal satellite data fusion. *Science of the Total Environment* 2020, 721, 137703. <https://doi.org/10.1016/j.scitotenv.2020.137703>
- Forkuor, G.; Hounkpatin, O.K.L.; Welp, G.; Thiel, M. High-resolution mapping of soil properties using remote sensing variables. *PLOS ONE* 2017, 12, e0170478. <https://doi.org/10.1371/journal.pone.0170478>
- Gomez, C.; Lagacherie, P.; Coulouma, G. Continuum removal versus PLSR for clay and CaCO₃ estimation. *Geoderma* 2008, 148, 141–148. <https://doi.org/10.1016/j.geoderma.2008.09.003>
- Heller Pearlshstien, D.; Ben-Dor, E. Effect of organic matter on the spectral signature of iron oxides. *Remote Sensing* 2020, 12, 1960. <https://doi.org/10.3390/rs12121960>
- Jenkinson, D.S. Studies on the decomposition of plant material in soil. *Journal of Soil Science* 1966, 17, 280–302. <https://doi.org/10.1111/j.1365-2389.1966.tb01462.x>
- Jones, J.D. Iron availability and management considerations. *Crops & Soils* 2020, 53, 32–37. <https://doi.org/10.1002/crso.20015>
- Kumar, N.S.; Anuncia, S.M.; Prabu, M. Application of satellite remote sensing for soil fertilization assessment. *International Journal of Online Engineering* 2013, 9. <https://doi.org/10.3991/ijoe.v9i2.2436>
- Liu, J.; Sun, Z. Hyperspectral characteristics and estimation of soil iron oxide content. *IOP Conf. Ser.: Earth Environ. Sci.* 2019, 300, 022147. <https://doi.org/10.1088/1755-1315/300/2/022147>
- Ministry of Agriculture (MoA). *The Soils of Jordan*; Ministry of Agriculture: Amman, Jordan, 1995.
- Moh'd, B. *Geological Sheet of Irbid*; Natural Resources Authority: Amman, Jordan,
- Moorman, F. *Report on the Soils of East Jordan*; FAO: Rome, Italy, 1959.
- Pereira, P.; Brevik, E.; Muñoz-Rojas, M.; Miller, B. *Soil Mapping and Process Modeling for Sustainable Land Use Management*; Elsevier, 2017. <https://doi.org/10.1016/C2016-0-04317-6>
- Plummer, C.C.; Carlson, D.; Hammersley, L. *Physical Geology*; McGraw-Hill, 2016.
- Sahwan, W.; Lucke, B.; Kappas, M.; Bäuml, R. Soil surface colors using Landsat-8 and Sentinel-2. *European Journal of Remote Sensing* 2018, 51, 850–862. <https://doi.org/10.1080/22797254.2018.1512114>
- Sahwan, W.; Lucke, B.; Sprafke, T.; Vanselow, K.A.; Bäuml, R. Spectral features and iron oxides in soils. *European Journal of Soil Science* 2021, 72, 80–97. <https://doi.org/10.1111/ejss.12947>
- Singh, B.; Schulze, D.G. Soil minerals and plant nutrition. *Nature Education Knowledge* 2015, 6, 1.
- Sparks, D.L. et al. *Methods of Soil Analysis, Part 3*; SSSA, 1996. <https://doi.org/10.2136/sssabookser5.3>
- The Unscrambler® X. *User Manual v10.3*; CAMO Software, 2014.
- USDA. *Soil Taxonomy*; USDA, 1999.
- Walkley, A.; Black, I.A. Determination of soil organic matter. *Soil Science* 1934, 37, 29–38. <https://doi.org/10.1097/00010694-193401000-00003>
- Xu, L. et al. Estimation of organic carbon using Vis–NIR spectroscopy. *Remote Sensing* 2020, 12, 3394. <https://doi.org/10.3390/rs12203394>
- Yu, H. et al. Prediction of soil properties using hyperspectral data. *Archives of Agronomy and Soil Science* 2018, 64, 546–559. <https://doi.org/10.1080/03650340.2017.1369526>
- Zhang, G.; Liu, F.; Song, X. Digital soil mapping: Progress and prospects. *Journal of Integrative Agriculture* 2017, 16, 2871–2885. [https://doi.org/10.1016/S2095-3119\(17\)61762-4](https://doi.org/10.1016/S2095-3119(17)61762-4)