soil systems

MDPI

Article

Validation-Aware VNIR Spectroscopy Reveals Target-Specific Transferability of SOC, SIC, and Gypsum Predictions in Dryland Agricultural Soils

Wajdi Allagui [1,2,3,*], Hayfa Zayani [4], Luca Bernardini [5], Pascal Pichelin [4], Nadhem Brahim [3], Roland Bol [6], Hamouda Aichi [7], Youssef Fouad [4] and Wolfgang Wanek [1,*]

1 Division of Terrestrial Ecosystem Research, Centre for Microbiology and Environmental Systems Science, University of Vienna, 1030 Vienna, Austria
2 Doctoral School in Microbiology and Environmental Science, University of Vienna, 1030 Vienna, Austria
3 Laboratory of Plants, Soils and Environment (LR21ES01), Faculty of Sciences of Tunis, University of Tunis El Manar, El Manar II, Tunis 2092, Tunisia
4 UMR SAS, Institut Agro Rennes-Angers, INRAE, 65 Rue de Saint-Brieuc, 35042 Rennes, France
5 Institute of Soil Research, Department of Ecosystem Management, Climate and Biodiversity, BOKU University, 1190 Vienna, Austria
6 Institute of Bio- and Geosciences, Agrosphere Institute (IBG-3), Forschungszentrum Jülich GmbH, 52428 Jülich, Germany
7 Laboratory of Agricultural Production Systems and Sustainable Development, Higher School of Agriculture of Mograne, University of Carthage, Zaghouan 1121, Tunisia
* Correspondence: allaguiw94@univie.ac.at or wajdi.allagui@fst.utm.tn (W.A.); wolfgang.wanek@univie.ac.at (W.W.)

Abstract

Hyper-arid agricultural soils combine low organic carbon, high calcium carbonate contents, salinity, and frequent gypsum accumulation, making rapid quantification of soil organic carbon (SOC), soil inorganic carbon (SIC), and gypsum valuable but analytically demanding. We evaluated visible–near-infrared (VNIR) spectroscopy for predicting these three properties in 216 samples from 18 plots, nested within six oasis sites spanning two oasis management systems, three topographic positions, and six soil depths, with particular emphasis on how validation design affects apparent model accuracy. Partial least squares regression (PLSR) and random forest (RF), single- and multi-target workflows, and global versus stratified calibrations were compared using plot-blocked GroupKFold, site-blocked GroupKFold, random KFold, and leave-one-out cross-validation (LOOCV). Under continuum-removed spectra, RF outperformed PLSR for SIC, whereas the advantage for SOC and gypsum was not robust to plot or site-level resampling. Predictability was strongly target-specific: gypsum was predicted most reliably (R^2 = 0.78), SOC moderately (R^2 = 0.44), and SIC least reliably (R^2 = 0.30). KFold and LOOCV generally yielded more optimistic estimates than plot-blocked GroupKFold, inflating SIC R^2 by 0.15–0.20 for the global model and by up to 0.64 for topography-stratified models (pooled across strata). Site-blocked validation further showed that spatial transferability was strongly target-specific: SOC performance remained similar between plot-blocked and site-blocked validation (R^2 = 0.41 and 0.43, respectively), gypsum performance declined moderately (R^2 = 0.78 to 0.64), whereas SIC performance decreased markedly from R^2 = 0.30 to −0.28, indicating poorer performance than a mean-based prediction under oasis-level blocking. VNIR spectroscopy therefore shows strong potential for gypsum, moderate utility for SOC, but limited transferability for SIC. Validation design, together with the spatial scale at which independence is imposed, critically determines how soil-spectroscopy performance should be interpreted for dryland monitoring.

Check for updates

Academic Editor: Yu (Frank) Yang

Received: 27 August 2026
Revised: 21 October 2026
Accepted: 29 October 2026
Published: 2 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Keywords: VNIR spectroscopy; soil organic carbon; soil inorganic carbon; gypsum; dryland soils; continuum removal; GroupKFold validation; random forest; multi-target modelling; model transferability

1. Introduction

Arid and semi-arid regions cover more than 40% of the terrestrial land surface and are characterized by chronic water limitation, high evaporative demand, sparse vegetation cover, and distinctive soil biogeochemical cycles [1–3]. In these systems, soil carbon occurs not only as soil organic carbon (SOC), which is often constrained by low plant inputs and physical degradation, but also as soil inorganic carbon (SIC), mainly as pedogenic and lithogenic carbonates such as calcium carbonate ($CaCO_3$) [4–7]. Many dryland soils also accumulate gypsum ($CaSO_4 \cdot 2H_2O$), particularly where shallow saline groundwater, evaporative concentration, topographic redistribution, or gypsiferous parent materials favour sulfate precipitation [8,9]. SOC, SIC, and gypsum are therefore central but functionally distinct constituents of dryland soils. SOC supports fertility, aggregation, nutrient cycling, and biological activity [4,9]; SIC represents a major carbon pool and contributes to soil buffering [7,10]; and gypsum strongly affects soil structure, water retention, hydraulic behaviour, salinity, crusting, and agricultural suitability [8,11].

The spatial distributions of SOC, SIC, and gypsum are highly heterogeneous in cultivated dryland landscapes [12,13]. Management influences organic matter inputs, irrigation, fertilization, tillage, and soil disturbance, whereas topography regulates runoff, erosion–deposition processes, groundwater depth, salt redistribution, and carbonate and gypsum accumulation. Soil depth adds a strong vertical gradient because SOC commonly decreases with depth, whereas SIC and gypsum may accumulate in subsurface horizons [14–16]. These interacting controls generate pronounced spatial variation over short distances, particularly in oasis and hyper-arid agricultural systems [17]. Rapid quantification of SOC, SIC, and gypsum is therefore important for soil monitoring, carbon-storage assessment, land evaluation, and agricultural management [18–20]. However, conventional laboratory analyses are labour-intensive, destructive, costly, and dependent on chemical reagents. Although variables such as pH, electrical conductivity, or soluble ions can be measured comparatively rapidly, specific quantification of SOC, SIC, and gypsum generally requires more demanding analytical procedures, limiting high-density monitoring in large, remote, or resource-limited dryland regions [21].

Visible–near-infrared spectroscopy (VNIR; approximately 400–2500 nm) provides a rapid, non-destructive, and cost-effective alternative for estimating soil properties from diffuse reflectance spectra [22,23]. VNIR spectra integrate absorption and scattering features associated with organic matter, clay minerals, carbonates, water, gypsum, iron oxides, particle size, and surface roughness. Relevant absorptions arise mainly from overtones and combination bands of molecular vibrations, including O–H and H–O–H features associated with water, structural hydroxyl groups, and hydrated minerals; C–H, N–H, and C=O features associated with organic matter; and CO_3-related features associated with carbonates [24,25]. Gypsum has relatively diagnostic hydration-related absorption bands near 1000, 1200, 1400, 1600, 1740, 1900, and 2200 nm [14,25,26], whereas carbonate absorptions commonly occur near 2300–2350 nm [14,24,27]. By contrast, SOC signals are broader and partly indirect because organic matter affects both specific molecular absorptions and overall spectral shape, colour, and scattering behaviour [18,19]. The spectral predictability of SOC, SIC, and gypsum may therefore differ substantially, particularly where strong mineral signals overlap or mask weaker SOC features.

Previous studies have shown that VNIR spectroscopy can predict SOC across diverse soils, although performance depends strongly on SOC range, mineralogy, texture, land use, calibration design, and spectral preprocessing [19,23,28]. VNIR has also been used to estimate carbonates and gypsum in arid and semi-arid soils, although gypsum has been studied rarely. Khayamim et al. [14] found that gypsum could be predicted more accurately than carbonates because of its stronger and more distinctive absorption features, whereas carbonate prediction was more affected by overlap with other soil constituents. That study also showed that full-spectrum PLSR could outperform continuum removal, indicating that enhancement of individual absorption features is not universally advantageous. Other proximal-sensing studies similarly demonstrate that predictions of gypsum, carbonate, salinity, SOC, and texture depend on spectral range, preprocessing, sensor type, model choice, mineralogical context, and calibration-sample distribution [10,29].

Spectral preprocessing and model choice are therefore important methodological decisions. Raw reflectance preserves full spectral shape and baseline information, whereas smoothing and scatter correction can reduce noise and effects associated with particle size, surface roughness, and illumination geometry [20]. Continuum removal can enhance diagnostic mineral bands but may also remove broadband information relevant to SOC and multivariate prediction. Likewise, partial least squares regression (PLSR) is widely used for high-dimensional, collinear spectral data [19,30], whereas random forest (RF) can capture nonlinear relationships and interactions among wavelengths [28,31]. Multi-target workflows may additionally exploit shared spectral information among SOC, SIC, and gypsum, but could perform poorly if relationships among targets vary among sites [32].

A further challenge is the strong environmental structure of dryland agricultural soils. Management, topography, and soil depth generate distinct pedological gradients, and stratified calibration may improve prediction if it reduces meaningful heterogeneity. Conversely, stratification reduces sample size and calibration range and may therefore weaken robustness and transferability [33]. Validation design is consequently critical. Random KFold cross-validation and leave-one-out cross-validation (LOOCV) can overestimate transferability when samples from the same plot or related soil profiles occur in both calibration and validation subsets [19,28]. This is particularly relevant in spatially structured datasets because samples from the same plot may share soil-forming history, mineralogy, management, and depth structure. GroupKFold avoids this overlap by assigning all samples from the same plot to the same fold, thereby providing a stricter test of prediction at held-out plots [34,35].

Despite substantial progress in soil VNIR spectroscopy, limited attention has been given to the simultaneous prediction of SOC, SIC, and gypsum in structured dryland agroecosystems, particularly to how validation design affects the interpretation of model performance. We therefore evaluated VNIR spectroscopy for predicting SOC, SIC, and gypsum in hyper-arid agricultural soils sampled across oasis management systems, topographic positions, and soil depths [36]. We addressed three questions: (i) How transferable are VNIR predictions of SOC, SIC, and gypsum, and how do these targets differ in predictability? (ii) How do spectral preprocessing, model type, and single- versus multi-target workflows affect prediction performance? (iii) How strongly do random KFold, LOOCV, plot-blocked GroupKFold, and site-blocked GroupKFold differ in their assessment of model transferability?

We hypothesized that prediction performance would be strongly target-specific, with gypsum showing the highest predictability because of its diagnostic hydration-related absorption features, SOC intermediate predictability because of its broader and partly indirect spectral response, and SIC weaker transferability because carbonate signals overlap with other mineral features and reflect site-specific carbonate redistribution. We further

expected preprocessing and workflow effects to be target-dependent, with continuum removal benefiting properties characterized by diagnostic absorption bands, such as gypsum, while full-spectrum approaches retain useful information for broader spectral responses such as SOC. Finally, we hypothesized that GroupKFold would yield more conservative and realistic performance estimates than random KFold or LOOCV because it prevents observations from the same plot from occurring in both calibration and validation sets; we further expected stricter oasis-level blocking to reveal any loss of transferability at broader spatial scales.

2. Materials and Methods

2.1. Study Area and Sampling Design

The study was conducted in the Djerid region of southwestern Tunisia, within the oasis landscapes of Tozeur, Deguache, and Nefta. The region is hyper-arid and lies between the hypersaline depressions of Chott El Gharsa to the north and Chott El Djerid to the south. Mean annual precipitation is approximately 80 mm, and the area is characterized by high evaporative demand, shallow saline groundwater near the chotts, and Quaternary deposits frequently enriched in gypsum. These conditions generate pronounced gradients in salinity, gypsum accumulation, carbonate redistribution, and groundwater depth.

The sampling design followed Allagui et al. [36]. Briefly, three study areas were selected within the Djerid region, with one traditional and one modern oasis system sampled in each area. Traditional oases are long-established, densely planted, structurally diverse systems with multilayered vegetation and greater reliance on organic inputs, whereas modern oases are more recently established, regularly planted date-palm systems with lower crop diversity and greater reliance on mineral fertilizers and regular irrigation (Figure S1). To capture topographic variation, one upslope, one midslope, and one downslope plot (10 × 10 m) was selected within each oasis system. Upslope plots were located farther from the Chott margin and had deeper groundwater, midslope plots represented transitional conditions, and downslope plots occurred closer to the hypersaline Chott margin, where groundwater was shallower and salinity effects were stronger.

The study was conducted across three study areas, each comprising two contrasting oasis management systems (traditional and modern), resulting in a total of six oasis sites. Within each oasis, three plots were established along a topographic gradient representing upslope, midslope, and downslope positions. The experimental design therefore comprised 18 primary sampling plots (3 study areas × 2 oasis management systems × 3 topographic positions). Plots within each oasis were separated by approximately 1–3 km, whereas distances among oases ranged from 10 to 50 km. The spatial distribution of the study areas, oasis sites, and sampling plots is presented in Figure S6, while the contrasting structural characteristics of the traditional and modern oasis systems are illustrated in Figure S1.

Within each plot, soil samples were collected from two microsite positions: beneath palm canopies and from adjacent interspaces. At each microsite, soil was sampled at six depth intervals: 0–5, 5–10, 10–30, 30–60, 60–90, and 90–120 cm. This hierarchical sampling design yielded a total of 216 soil samples (18 plots × 2 microsites × 6 soil depths). Following collection, soil samples were air-dried and sieved to <2 mm prior to laboratory analyses and visible–near-infrared (VNIR) spectroscopic measurements.

2.2. Laboratory Analyses

Soil total carbon (STC), soil organic carbon (SOC), and soil inorganic carbon (SIC) were determined as described in Allagui et al. (2026) [36]. Briefly, soil samples were oven-dried at 105 °C, homogenized, coarsely ground with a mortar and pestle, and subsequently finely ground in a ball mill at 25 Hz for 5 min (Retsch MM200, Haan, Germany). STC was

measured on untreated finely ground soil. SOC was measured on a separate aliquot after carbonate removal by pretreatment with 2 M HCl for 60 min, followed by drying at 105 °C. Carbon concentrations were determined using an elemental analyzer (EA-Isolink, Thermo Scientific, Vienna, Austria) coupled via a ConFlo IV interface to a Delta V Advantage isotope ratio mass spectrometer (Thermo Scientific). SIC was calculated by difference:

$$SIC = STC - SOC$$

where SIC is soil inorganic carbon, STC is soil total carbon, and SOC is soil organic carbon.

Gypsum ($CaSO_4 \cdot 2H_2O$) was quantified using a dissolution–ion chromatography assay, following Allagui et al. [36]. Briefly, 50 mg of finely ground, oven-dried soil was extracted with 25 mL ultrapure water for 18 h, using an extraction ratio sufficient to dissolve the gypsum present. Extracts were centrifuged at 10,000 rpm for 10 min, and dissolved Ca^{2+} and SO_4^{2-} were measured by high-performance ion chromatography using a Dionex ICS5000 system (Thermo Scientific). Cations were separated on a Dionex IonPac CS16 column and anions on a Dionex IonPac AS11-HC column, using CERS 500 and AERS 500 suppressors (Thermo Scientific), respectively. Because dissolution of $CaSO_4 \cdot 2H_2O$ releases Ca^{2+} and SO_4^{2-} in a 1:1 molar ratio, gypsum content was calculated from the lower molar amount of the two ions, assuming the limiting ion represented the amount of dissolved gypsum.

Soil pH was measured electrometrically in a 1:5 soil-to-water suspension using an inoLab pH 7110 pH meter (WTW-Xylem Analytics, Weilheim, Germany). Electrical conductivity (EC) was measured in a 1:5 soil-to-water extract using a C862 conductivity meter (Consort, Turnhout, Belgium). Soil texture was determined using the simplified sedimentation method of Kettler et al. [37], in which sand is separated by sieving and the remaining fine fraction is partitioned into silt and clay according to particle settling times in water. These supporting variables were used to characterize the dataset and interpret spectral model behaviour but were not included as predictors in the VNIR models. One pH value (sample Np1S10) was missing; correlations involving pH therefore used pairwise-complete observations ($n = 215$).

2.3. VNIR Spectral Acquisition

Diffuse reflectance spectra were acquired in a dark room using an ASD FieldSpec® 3 spectroradiometer (Analytical Ltd., London, UK) covering 350–2500 nm, with spectral sampling intervals of 1.4 nm from 350 to 1000 nm and 2 nm from 1000 to 2500 nm. Soil samples were placed in 5 cm diameter Petri dishes to a depth of approximately 1 cm and levelled before measurement. For each sample, four replicate spectra were recorded with a standard contact probe equipped with an internal light source (colour temperature ca. 3400 K), rotating the dish by 90° between scans to reduce effects of particle orientation and surface heterogeneity. Each replicate spectrum was averaged from 30 internal scans. A Spectralon® white reference panel (Labsphere Inc., North Sutton, NH, USA) was measured before sample acquisition and approximately every 15–20 min thereafter. The four replicate spectra per sample were averaged.

2.4. Spectral Preprocessing

Spectra were exported at 1 nm intervals by the instrument software, which interpolated the native sampling intervals of 1.4 nm (350–1000 nm) and 2 nm (1000–2500 nm). The 350–400 nm region was excluded due to low signal-to-noise, resulting in 2101 bands from 400 to 2500 nm for all preprocessing scenarios. No additional resampling was performed. Continuum removal was applied independently to each spectrum using an upper convex hull over 400–2500 nm, followed by linear interpolation between hull vertices and division of measured reflectance by the interpolated continuum.

The analytical workflow comprising spectral scenarios, calibration paths, model families, validation schemes, and performance metrics is summarized in Figure 1. Three spectral predictor matrices were evaluated, and their mean spectra are shown in Figure S5. First, raw reflectance spectra (RAW) were retained after wavelength trimming. Second, spectra were processed using a Savitzky–Golay first derivative (window length 11, polynomial order 3) followed by multiplicative scatter correction (RAW + SG + MSC) to emphasize local spectral structure while reducing additive and multiplicative scatter effects. Third, continuum removal (CR) was applied to enhance absorption features relative to the spectral continuum. For each spectrum, an upper convex hull was fitted to define the continuum, and continuum-removed reflectance was calculated as:

$$R_{CR}(\lambda) = \frac{R(\lambda)}{C(\lambda)} \quad (1)$$

where $R_{CR}(\lambda)$ is the continuum-removed reflectance at wavelength λ, $R(\lambda)$ the measured reflectance, and $C(\lambda)$ the continuum reflectance at wavelength λ. RAW, RAW + SG + MSC, and CR spectra were compared under GroupKFold validation for SOC, SIC, and gypsum. The main model interpretation was based on the preprocessing–model combinations retained from this validation-aware comparison, while the full preprocessing comparison is reported in the Supplementary Material.

2.5. Validation Strategy

Four validation strategies were compared. Plot-blocked GroupKFold was used as the primary validation because it tests transferability to held-out plots while respecting the hierarchical structure of the dataset. The grouping variable was plot, with 18 plot levels, and all samples from a given plot were assigned to the same fold so that no plot contributed simultaneously to calibration and validation. With 18 plots distributed across six folds, three complete plots (36 samples) were withheld in each plot-blocked fold, while the remaining 15 plots (180 samples) formed the calibration set. Five shuffled folds were used for random KFold (random_state = 42), and one held-out sample per fold for leave-one-out cross-validation (LOOCV). Blocking cross-validation according to the dependence structure of the data follows recommendations for spatially and hierarchically structured datasets [35].

In addition to plot-blocked GroupKFold, site-blocked cross-validation was performed by withholding all three plots from one oasis per fold, yielding six folds with 180 calibration and 36 validation samples. Plot-blocked validation assessed transfer to unsampled plots within the same oasis, whereas site-blocked validation assessed transfer to a completely held-out oasis within the Djerid study region; the six oasis sites were separated by approximately 10–50 km.

Random K-fold cross-validation and leave-one-out cross-validation (LOOCV) were used as supplementary benchmarks of internal predictive performance. In random K-fold cross-validation, individual samples were randomly assigned to five shuffled folds without accounting for the hierarchical structure of the dataset. Because partitioning ignored plot grouping and depth/management stratification, samples from the same plot could occur simultaneously in the calibration and validation subsets, creating the main risk of hierarchical information leakage. In LOOCV, each individual sample was withheld once for validation, while all remaining samples, including those originating from the same plot, remained in the calibration set.

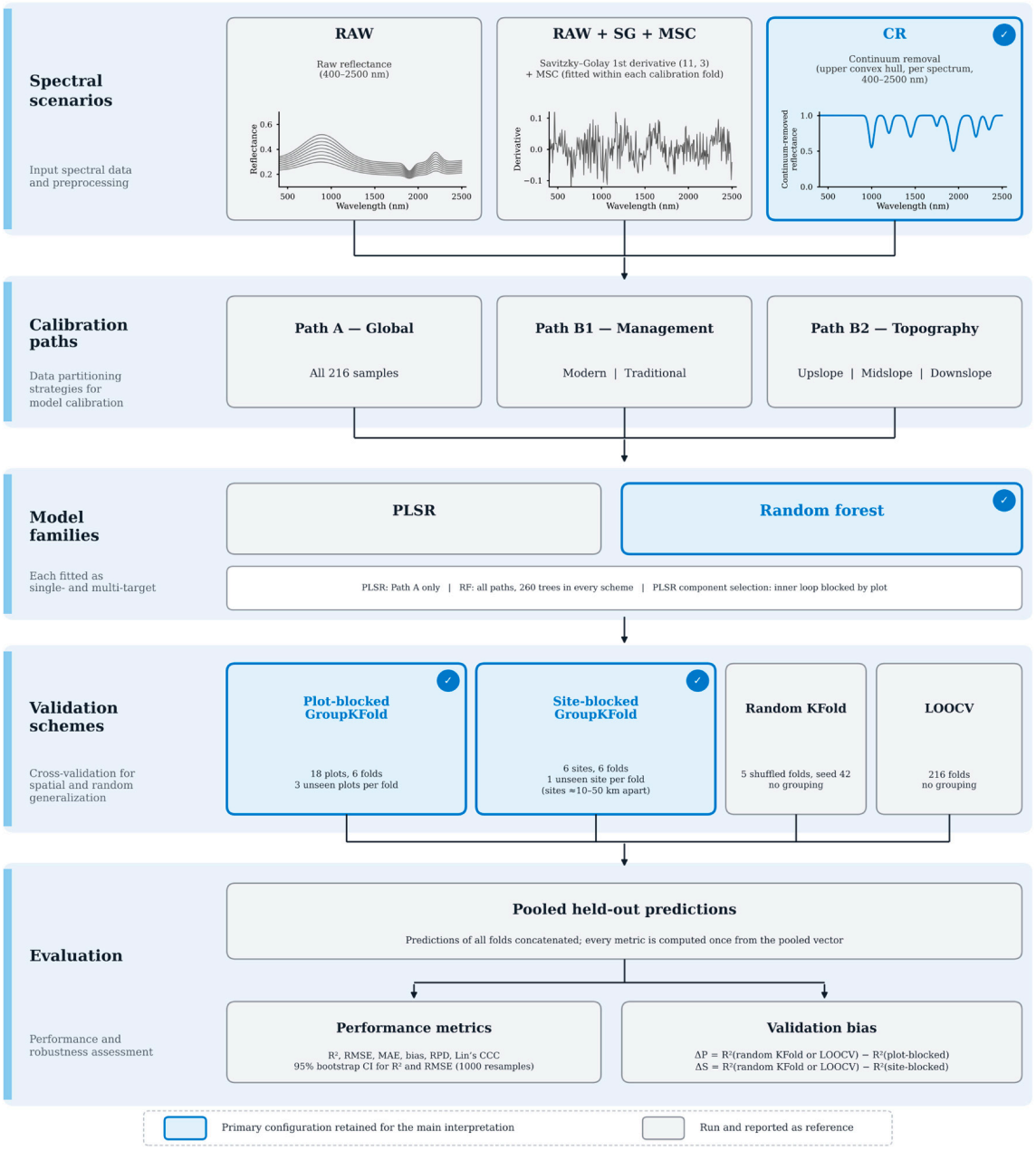


Figure 1. Validation-aware workflow for predicting soil organic carbon (SOC), soil inorganic carbon (SIC), and gypsum from visible–near-infrared spectra. The dataset comprised 216 dryland soil samples from 18 plots in six sites, measured at 400–2500 nm (2101 wavelength variables). Three spectral scenarios were evaluated: untransformed reflectance (RAW), Savitzky–Golay first-derivative pre-processing with multiplicative scatter correction (RAW + SG + MSC), and continuum removal (CR). Models were fitted using three calibration paths: Path A, global; Path B1, management-stratified; and Path B2, topography-stratified. Partial least squares regression (PLSR) was evaluated for Path A only, whereas random forest (RF) was evaluated for all paths; both single- and multi-target workflows were tested. Performance was assessed using plot-blocked GroupKFold (six folds, three plots held out per fold), site-blocked GroupKFold (six folds, one site held out per fold), random five-fold KFold, and leave-one-out cross-validation (LOOCV). Held-out predictions were pooled across folds before calculating performance metrics. ΔP and ΔS quantify the optimism of non-grouped validation relative to plot-blocked and site-blocked GroupKFold, respectively. Blue shading and checkmarks indicate the spectral scenario, model family and grouped validation schemes used for the main-text results (CR spectra, random forest, plot-blocked and site-blocked GroupKFold); other combinations are reported in the Supplementary Material. Spectra are schematic.

Neither validation scheme was considered a realistic assessment of model transferability to spatially independent observations. Instead, both were used as internal reference scenarios to quantify the extent to which predictive performance may be inflated by spatial and hierarchical information leakage.

Under GroupKFold, the global path used 180 training and 36 validation samples per fold. Stratified models were evaluated using the same held-out partitions as the global models. Plot-blocked partitions were defined for the full dataset and subsequently restricted to each stratum, preserving the original fold assignments. Each fold withheld plots representing the same topographic position in three oases (e.g., fold 1: Degache3, Machtla3, and Tozeur3; fold 4: Bouhlel3, Lahwar3, and Nefta3; Table S6). Each management stratum comprised 108 samples from nine plots and yielded six non-empty folds, with 84–96 samples used for calibration and 12–24 for validation per fold. Each topographic stratum comprised 72 samples from six plots and yielded two non-empty folds, each allocating three plots (36 samples) to validation and the remaining three plots (36 samples), from the other three oases to calibration. Consequently, under plot-blocked validation, topographic models were calibrated using substantially fewer samples than the global model (36 versus 180) and evaluated on plots from oases excluded from calibration. Under site-blocked validation, each management stratum yielded three non-empty folds, with 72 calibration samples and 36 validation samples per fold, whereas each topographic stratum yielded six folds, with 60 calibration samples and 12 validation samples per fold. For all models and validation schemes, held-out predictions were pooled across folds, and each performance metric was calculated once from the pooled predictions rather than averaged across folds. This also avoids undefined fold-level R^2 for LOOCV and unstable averaging across small validation folds.

All Random Forest models used 260 trees. PLSR components were selected by nested cross-validation using inner folds blocked by plot (identical to the outer grouping under plot-blocked validation and nested within the oasis-blocked outer folds under site-blocked validation); under site-blocked validation, the maximum was limited to four components to reduce extrapolation. Under site-blocked validation, each outer calibration set comprised five oases (15 plots, 180 samples), and the inner component search used five-fold GroupKFold by plot within this calibration set only, so no information from the held-out oasis entered component selection; an oasis-level inner split would have provided only five groups. Multi-target Random Forest responses were standardized within each calibration fold and back-transformed before evaluation, using a native multi-output forest with shared tree structures [38,39]. Uncertainty in R^2 and RMSE was estimated from 1000 bootstrap resamples of pooled held-out predictions, using the 2.5th and 97.5th percentiles. Bootstrap resampling was applied to individual held-out observations; it therefore does not preserve plot- or oasis-level dependence, and these intervals are approximate and may underestimate uncertainty. As a sensitivity analysis, we additionally performed cluster bootstrap resampling at the plot or oasis level, sampling entire clusters with replacement and retaining all observations within each selected cluster. Model comparisons used a paired bootstrap, with identical resampled clusters and observations applied to both models. The 95% intervals for differences in R^2 and RMSE were defined by the 2.5th and 97.5th percentiles of the bootstrap distributions. These cluster-level and paired intervals are conditional on the saved predictions and do not account for variability arising from model refitting (Tables S7 and S8). Differences between models were considered supported only when the corresponding paired interval excluded zero.

2.6. Performance Metrics

Model performance was evaluated using the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), bias, ratio of performance to deviation (RPD), and Lin's concordance correlation coefficient (CCC). All metrics were calculated from pooled held-out predictions for each validation scheme.

For n observations, with observed values y_i, predicted values $\hat{y}_i$, mean observed value $\bar{y}$ and mean predicted value $\bar{y}$, R^2 was calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (2)$$

Negative R^2 values were retained and indicate performance worse than predicting the mean observed value. To quantify optimism from non-grouped validation, ΔR^2 was calculated as:

$$\Delta R^2 = R^2_{\text{KFold or LOOCV}} - R^2_{\text{GroupKFold}} \quad (3)$$

Positive ΔR^2 therefore indicates higher apparent performance under non-grouped than plot-blocked GroupKFold validation. ΔP and ΔS denote ΔR^2 relative to plot-blocked and site-blocked GroupKFold, respectively.

RMSE, MAE, and bias were calculated as:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (4)$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (5)$$

$$\text{Bias} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i) \quad (6)$$

Positive bias indicates overprediction and negative bias underprediction.

RPD was calculated as:

$$\text{RPD} = \frac{\text{SD}(y)}{\text{RMSE}} \quad (7)$$

where SD(y) is the standard deviation of the pooled observed values.

Lin's CCC was calculated as:

$$\text{CCC} = \frac{2\rho\sigma_y\sigma_{\hat{y}}}{\sigma_y^2 + \sigma_{\hat{y}}^2 + \left(\mu_y - \mu_{\hat{y}}\right)^2} \quad (8)$$

where ρ is the Pearson correlation coefficient between observed and predicted values, σ_y and $\sigma_{\hat{y}}$ are their standard deviations, and $\bar{y}$ and $\bar{y}$ their means. CCC evaluates both precision and agreement with the 1:1 relationship.

2.7. Computational Implementation and Software

All spectral preprocessing, model fitting, validation, performance evaluation, wavelength-importance analysis, and figure/table generation were implemented in Python 3.14.0 using custom scripts. Data handling used pandas 3.0.2 [40], numerical operations NumPy 2.4.4 [41], signal processing SciPy 1.17.1 [42], and chemometric and machine-learning workflows scikit-learn 1.8.0 [43]. Figures were generated with Matplotlib 3.10.8 [44] and seaborn 0.13.2 [45], and SHAP 0.51.0 was used for post hoc interpretation of RF models and wavelength importance [46]. Random forest fitting and random KFold partitioning used a fixed random seed (random_state = 42) to ensure reproducibility.

Sampling position and depth were repeated within plots and were not used as predictors or stratification variables. Because plot-blocked validation kept all observations

from each plot within the same fold, samples from the same plot could not be split between calibration and validation. Mixed-effects models included depth, topographic position, and management as fixed main effects, without interaction terms, and plot as a random intercept to account for non-independence of repeated observations within plots. Models were fitted by restricted maximum likelihood, with factor significance assessed by an omnibus test for each fixed factor. PCA was performed on all 216 continuum-removed spectra after median imputation, mean-centring, and unit-variance scaling, and the first two components were retained for descriptive visualization. SHAP values were derived from Random Forest models fitted to the full dataset using the best-performing RF configuration (single- or multi-target) for each target and preprocessing variant (Section 3.5); this analysis was used for post hoc wavelength interpretation rather than predictive-performance estimation. Factor significance was evaluated using joint Type III Wald tests. Sampling position (beneath palm canopies or in interspaces; coded S and Sb in the sample identifiers) was not included as a fixed effect in the mixed-effects models; the plot-level random intercept accounts for within-plot dependence among the 12 samples, while variation between the two microsites within each plot is captured by the residual term. Response variables were analyzed without transformation. Plot was included as a random intercept to account for repeated observations within plots (12 samples per plot). The limited number of oases (six) and study areas (three) constrained reliable estimation of variance components at these levels. Adding oasis or study area as nested random intercepts did not alter the statistical significance of the tested factors (Table S9). Residual diagnostics indicated right-skewed residual distributions and heterogeneous residual variances for SOC and gypsum (Figure S8); consequently, the associated p values should be regarded as approximate. Log-transforming gypsum as $\log(1 + \text{gypsum})$ changed the topographic effect from non-significant to nominally significant ($p = 0.21$ for the untransformed response versus $p = 0.04$ for the transformed response), while the significance of the depth and management effects remained unchanged (Table S9). For each target, SHAP values were computed for all 216 samples using a Random Forest model refitted to the complete dataset. The single- or multi-target configuration was selected based on the highest plot-blocked cross-validated R^2 among models using continuum-removed or RAW + SG + MSC spectra (Section 3.5).

Savitzky–Golay first-derivative preprocessing was implemented with scipy.signal.savgol_filter using a window length of 11 wavelength points, polynomial order 3, and derivative order 1 [47]. Multiplicative scatter correction was performed within each calibration fold by regressing individual spectra against the mean calibration spectrum and correcting additive and multiplicative scatter effects. Continuum removal was implemented by fitting an upper convex hull to each reflectance spectrum and dividing measured reflectance by the corresponding continuum reflectance.

For PLSR, the number of latent variables was selected independently within each outer calibration fold (by nested cross-validation over candidate values from 1 to 25, with the inner loop blocked by plot (under site-blocked validation the outer folds were blocked by oasis, whereas the inner folds remained blocked by plots) and RMSE minimised using calibration data only; under site-blocked validation the maximum was limited to four components). The same procedure was applied to single-target and multi-target (PLS2) models. RF models were fitted with 260 trees in every validation scheme, using squared-error splitting, unrestricted tree depth, min_samples_leaf = 1, and max_features = "sqrt" [31]. Multi-target RF models used the native multi-output implementation in scikit-learn, with the response variables standardized using calibration-fold statistics and predictions back-transformed to mass-percent units before evaluation, whereas single-target RF models fitted one independent forest per target.

3. Results

3.1. Soil Property Variability and Spectral Structure

The dataset showed substantial variation in SOC, SIC, and gypsum (Table S4). SOC concentrations were generally low, ranging from 0.04 to 4.20% (mean 0.74%), whereas SIC varied over a narrower range of 0.21–3.80% (mean 1.65%). Gypsum showed the greatest variability and strong positive skew, ranging from 0 to 69.83% (mean 5.95%). Soils were predominantly sandy (50–95%; mean 77.4%), with lower silt (1–25%; mean 9.1%) and clay contents (2–38%; mean 13.2%). Soil pH ranged from 7.3 to 8.6 (mean 7.9), and EC from 0.31 to 3.57 dS m^{-1} (mean 1.50).

SOC, SIC, and gypsum showed contrasting distributions across topography, depth, and management (Figure 2; Figure S2). SOC declined strongly with depth across most topography–management combinations, with highest concentrations in surface soils and markedly lower values below 30 cm. SIC showed substantial overlap among groups, with only a modest tendency towards higher concentrations downslope. Gypsum showed the strongest environmental differentiation, with high concentrations occurring mainly in traditional oases, particularly at downslope positions and in the 60–90 and 90–120 cm depth intervals. Overall, depth most strongly structured SOC, whereas management and topographic position were more evident for gypsum; SIC showed comparatively weak categorical separation.

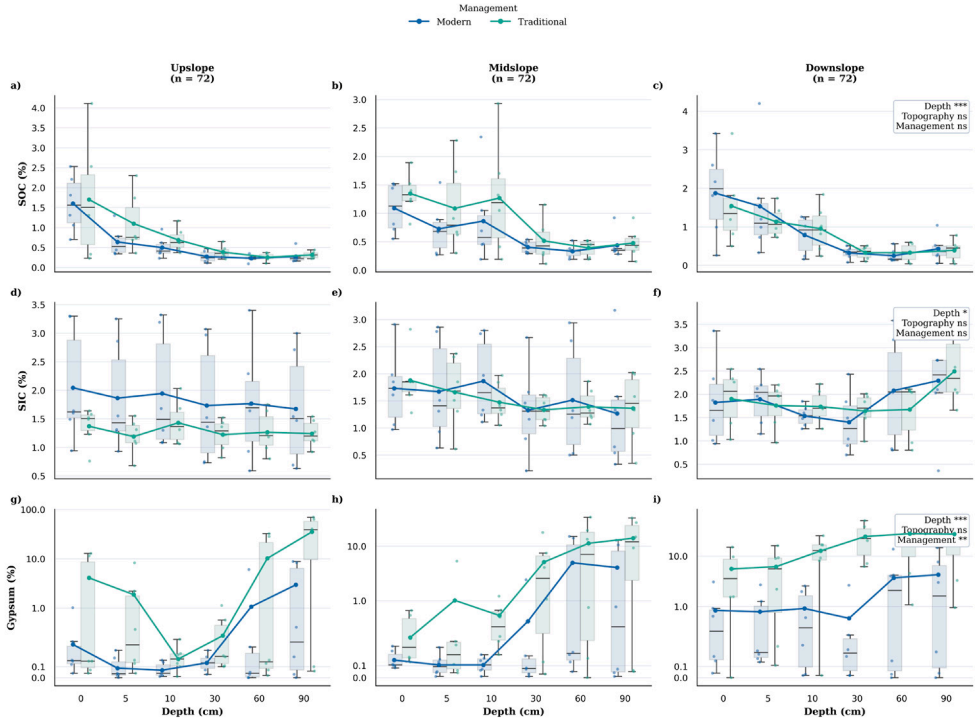


Figure 2. Distribution of SOC, SIC, and gypsum across topography, soil depth, and oasis management. Panels show the three target soil properties by topographic position, (**a**–**c**) SOC, (**d**–**f**) SIC, and (**g**–**i**) gypsum at upslope (**a**,**d**,**g**), midslope (**b**,**e**,**h**), and downslope (**c**,**f**,**i**) positions, with soil depth on the x-axis and oasis management distinguished within each depth class (Modern, blue; Traditional, green). Row annotations report *p* values for the main effects of soil depth, topography, and oasis management from linear mixed-effects models, with plot as a random intercept and no interaction terms; *p* values are approximate (Figure S8; Table S9). ***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$; ns, not significant ($p \geq 0.05$). Gypsum is shown on a symmetrical logarithmic y-axis to accommodate its strongly right-skewed distribution and zero values. Abbreviations: SOC, soil organic carbon; SIC, soil inorganic carbon.

Principal component analysis of the VNIR spectra showed partial but no distinct separation among management systems or topographic positions, while soil depth formed a secondary spectral gradient (Figure 3). The first two principal components explained 45.6% and 17.8% of the spectral variance. Correlations among SOC, SIC, gypsum, EC, pH, and texture variables were generally moderate rather than redundant (Figure S3).

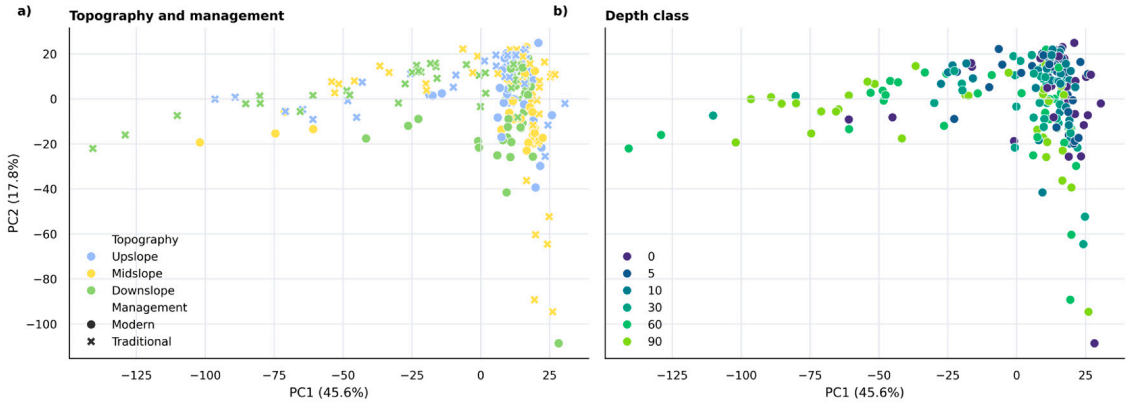


Figure 3. Principal component analysis of VNIR spectra. Principal component analysis was performed on all 216 continuum-removed VNIR spectra after median imputation, mean-centring, and unit-variance scaling. (**a**), Samples coloured by topographic position and shaped by oasis management type. (**b**), The same ordination coloured by soil depth. The ordination shows partial spectral structuring by environmental factors but no distinct separation among management systems or topographic positions. Abbreviations: VNIR, visible–near-infrared; PC, principal component.

3.2. Effects of Spectral Preprocessing and Prediction Workflow

The three spectral variants raw reflectance (RAW), continuum removal (CR), and Savitzky–Golay first-derivative preprocessing with multiplicative scatter correction (RAW + SG + MSC) showed distinct mean spectral patterns (Figure S5).

Preprocessing effects differed markedly among the target properties (Table S1). Gypsum was predicted reliably across all three spectral variants. The highest plot-blocked performance was obtained with RAW single-target PLSR ($R^2 = 0.80$, RMSE = 5.61%, RPD = 2.23), while CR-based RF models performed nearly as well ($R^2 = 0.78$ for both single- and multi-target workflows). RF achieved $R^2 = 0.74$ with RAW spectra and $R^2 = 0.77$–0.78 with RAW + SG + MSC preprocessing. The advantage of single target over multi-target PLSR under RAW spectra was supported by the paired bootstrap interval, whereas the remaining differences among these configurations fell within sampling uncertainty. Overall, gypsum prediction was comparatively robust to both spectral preprocessing and model choice.

SOC benefited more clearly from preprocessing. RAW spectra yielded weak to moderate performance, with PLSR reaching $R^2 = 0.27$–0.36 and RF reaching $R^2 = 0.14$–0.19. Performance increased with CR and especially RAW + SG + MSC, with the highest SOC prediction obtained by multi-target RF using RAW + SG + MSC ($R^2 = 0.47$, RMSE = 0.52%, RPD = 1.38). CR-based multi-target RF was only slightly weaker ($R^2 = 0.44$), a difference well within the bootstrap intervals of both estimates, indicating that both preprocessing approaches substantially improved SOC prediction relative to RAW RF models.

SIC was the least predictable and most preprocessing-sensitive target. RF models based on RAW spectra performed poorly ($R^2 < 0$), whereas RAW + SG + MSC increased RF performance to $R^2 = 0.27$–0.31. For SIC, the highest R^2 was obtained with single-target RF using RAW + SG + MSC preprocessing ($R^2 = 0.31$), followed closely by CR-based RF in the single-target and multi-target workflows ($R^2 = 0.30$ and 0.28; RMSE = 0.61% and 0.62%; RPD = 1.19 and 1.18; bootstrap intervals for $R^2 = 0.15$–0.42 and 0.14–0.40, respectively). The

paired bootstrap provided no evidence of a performance difference between CR and RAW + SG + MSC (ΔR^2 [CR − RAW + SG + MSC] = −0.01; 95% CI under plot-level resampling: −0.08 to 0.08; Table S8). Overall predictive performance remained modest. CR spectra were used for the main-text analyses (Figure 1) to define a consistent reporting workflow; this choice does not imply that CR was the best-performing preprocessing method for every target.

Multi-target modelling did not provide a consistent advantage across targets. Multi-target RF achieved slightly higher R^2 values for SOC than single-target RF with RAW + SG + MSC (0.47 versus 0.46) and CR preprocessing (0.44 versus 0.41). For SIC, however, multi-target RF had lower R^2 values across all three preprocessing methods (e.g., 0.28 versus 0.30 with CR), while gypsum performance remained essentially unchanged (Table S1). For CR spectra, the paired bootstrap intervals for differences between multi-target and single-target RF included zero for all three properties under plot-level resampling, providing no clear evidence of a performance difference between the two workflows (Table S8). Overall, preprocessing and workflow effects were therefore target-specific rather than universally beneficial.

3.3. Global Model Performance and Validation Effects

For the global modelling path using CR spectra, RF achieved higher R^2 values than PLSR for all three target properties under plot-blocked GroupKFold validation (Table 1). However, paired bootstrap comparisons based on plot-level resampling supported this advantage only for SIC (ΔR^2 [RF − PLSR] = 0.72; 95% CI: 0.29–1.73). The corresponding intervals included zero for SOC (ΔR^2 = 0.12; 95% CI: −0.06 to 0.25) and gypsum (ΔR^2 = 0.06; 95% CI: −0.01 to 0.19), providing no clear evidence of an RF advantage for these properties (Table S8). RF performance was highest for gypsum (R^2 = 0.78), intermediate for SOC (R^2 = 0.44), and weakest for SIC (R^2 = 0.30), whereas PLSR produced negative R^2 values for SIC. Observed versus predicted values of the best-performing workflow for each target are shown in Figure S7. The higher RF values persisted despite optimization of the number of PLSR latent variables by inner cross-validation.

Table 1. Plot-blocked GroupKFold validation performance of global PLSR and RF models for SOC, SIC, and gypsum using single- and multi-target workflows. Metrics were calculated from pooled held-out predictions. RMSE, MAE, and bias are expressed as % dry soil mass for all three target properties. Negative R^2 values indicate performance worse than prediction of the held-out mean. Abbreviations: SOC, soil organic carbon; SIC, soil inorganic carbon; PLSR, partial least squares regression; RF, random forest; R^2, coefficient of determination; RMSE, root mean square error; MAE, mean absolute error; CCC, concordance correlation coefficient; RPD, ratio of performance to deviation.

Target	Model	R^2 (95% CI)	RMSE (95% CI)	MAE	Bias	CCC	RPD
SOC	Direct PLSR	0.29 (−0.01, 0.47)	0.60 (0.50, 0.71)	0.40	−0.02	0.55	1.19
	Multi-target PLSR	0.19 (−0.15, 0.39)	0.64 (0.54, 0.75)	0.43	−0.06	0.55	1.11
	Direct RF	0.41 (0.20, 0.53)	0.55 (0.46, 0.65)	0.36	0.07	0.64	1.30
	Multi-target RF	0.44 (0.26, 0.56)	0.53 (0.44, 0.63)	0.35	0.05	0.65	1.33
SIC	Direct PLSR	−0.42 (−0.78, −0.17)	0.87 (0.79, 0.97)	0.67	−0.14	0.17	0.84
	Multi-target PLSR	−0.19 (−0.45, 0.01)	0.79 (0.72, 0.87)	0.62	−0.10	0.19	0.92
	Direct RF	0.30 (0.15, 0.42)	0.61 (0.54, 0.69)	0.46	−0.05	0.47	1.19
	Multi-target RF	0.28 (0.14, 0.40)	0.62 (0.54, 0.69)	0.47	−0.04	0.45	1.18
Gypsum	Direct PLSR	0.73 (0.58, 0.82)	6.55 (5.22, 7.95)	3.98	−0.04	0.85	1.91
	Multi-target PLSR	0.75 (0.59, 0.84)	6.25 (4.77, 7.67)	3.51	−0.03	0.86	2.00
	Direct RF	0.78 (0.64, 0.87)	5.81 (4.48, 6.89)	2.78	0.19	0.88	2.15
	Multi-target RF	0.78 (0.66, 0.86)	5.82 (4.54, 7.13)	2.94	0.09	0.88	2.15

Validation strategy strongly affected apparent model performance (Figure 4; Table S3). For the best global RF models, R^2 increased from 0.44 under plot-blocked GroupKFold to 0.51 under KFold and 0.50 under LOOCV for SOC, and from 0.30 to 0.44 and 0.48 for SIC. Gypsum showed the same pattern but a smaller relative effect, increasing from $R^2 = 0.78$ to 0.88. Other performance metrics showed corresponding improvements under non-grouped validation (Table S3), confirming that random KFold and LOOCV generally yielded more optimistic estimates than plot-blocked GroupKFold.

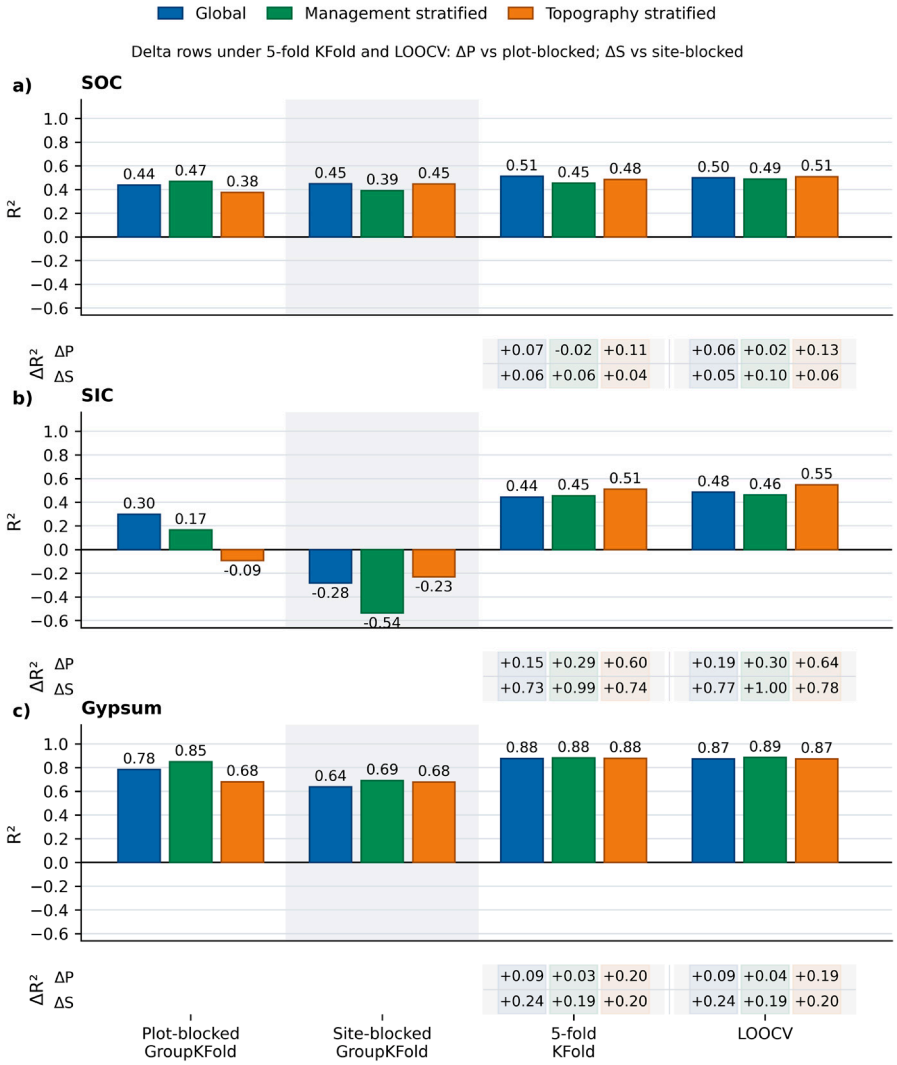


Figure 4. Effect of validation strategy on Random Forest (RF) performance across global, management-stratified, and topography-stratified modelling pathways using continuum-removed spectra. Bars represent R^2 values derived from pooled held-out predictions under plot-blocked GroupKFold, site-blocked GroupKFold, five-fold random KFold, and leave-one-out cross-validation (LOOCV); the shaded column indicates site-blocked validation. The rows below each panel report the increase in R^2 obtained with random KFold and LOOCV relative to plot-blocked GroupKFold (ΔP) and site-blocked GroupKFold (ΔS). Non-grouped validation generally produced higher apparent predictive performance, with the largest inflation observed for SIC, for which R^2 was negative across all modelling pathways under site-blocked validation. Panels: (**a**) SOC, (**b**) SIC, and (**c**) gypsum.

Relative to site-blocked GroupKFold validation, the inflation in predictive performance produced by non-grouped validation was substantially greater (Figure 4). For SIC, random K-fold cross-validation increased R^2 by 0.73 for the global model and by 0.99 for the management-stratified model relative to site-blocked validation, compared with increases of

0.15 and 0.29, respectively, relative to plot-blocked validation. For SOC, the corresponding increases were small under both grouped reference schemes ($\Delta R^2 \leq 0.13$), whereas for gypsum they were moderate ($\Delta R^2 = 0.09–0.24$). These results demonstrate that the apparent optimism introduced by random resampling depends strongly on the spatial scale of the grouped validation reference and is most pronounced for the target exhibiting the lowest spatial transferability.

The validation effect was particularly pronounced for SIC. For the global multi-target RF model, moving from plot-blocked GroupKFold to LOOCV increased R^2 from 0.28 to 0.48, reduced RMSE from 0.62 to 0.52%, and increased CCC from 0.45 to 0.63 and RPD from 1.18 to 1.39 (Table S3). A similar discrepancy occurred in topography stratified SIC models: KFold and LOOCV yielded positive R^2 values, whereas all three topographic strata showed negative R^2 under plot-blocked GroupKFold (Figure S4; Table 2). Thus, within-plot predictive structure did not translate into reliable prediction of held-out plots.

Table 2. Plot-blocked GroupKFold performance of RF models within management and topographic strata. Separate RF models were fitted for modern and traditional oasis systems and for upslope, midslope, and downslope positions using single- and multi-target workflows. Metrics were calculated from pooled held-out predictions within each stratum. RMSE, MAE, and bias are expressed as % dry soil mass for SOC, SIC, and gypsum. Negative R^2 values indicate performance worse than prediction of the held-out mean. Abbreviations: RF, random forest; SOC, soil organic carbon; SIC, soil inorganic carbon; R^2, coefficient of determination; RMSE, root mean square error; MAE, mean absolute error; CCC, concordance correlation coefficient; RPD, ratio of performance to deviation.

Stratification	Stratum	Target	Model	R^2 (95% CI)	RMSE (95% CI)	MAE	Bias	CCC	RPD
Management	Modern	SOC	Direct RF	0.41 (0.13, 0.60)	0.54 (0.39, 0.70)	0.34	0.02	0.62	1.30
			Multi-target RF	0.40 (0.08, 0.59)	0.55 (0.40, 0.70)	0.34	0.02	0.60	1.29
	Modern	SIC	Direct RF	0.27 (0.02, 0.49)	0.75 (0.61, 0.90)	0.54	−0.13	0.48	1.17
			Multi-target RF	0.26 (0.02, 0.47)	0.75 (0.61, 0.90)	0.56	−0.10	0.46	1.16
	Modern	Gypsum	Direct RF	0.82 (0.66, 0.90)	1.41 (0.95, 1.80)	0.70	−0.06	0.89	2.39
			Multi-target RF	0.80 (0.65, 0.88)	1.52 (1.03, 1.95)	0.80	−0.12	0.87	2.22
	Traditional	SOC	Direct RF	0.53 (0.29, 0.65)	0.49 (0.39, 0.59)	0.35	0.08	0.69	1.45
			Multi-target RF	0.53 (0.33, 0.64)	0.49 (0.40, 0.59)	0.35	0.06	0.68	1.45
	Traditional	SIC	Direct RF	−0.17 (−0.33, −0.06)	0.57 (0.48, 0.65)	0.45	−0.07	0.01	0.92
			Multi-target RF	−0.20 (−0.35, −0.11)	0.57 (0.48, 0.66)	0.45	−0.07	−0.08	0.91
	Traditional	Gypsum	Direct RF	0.83 (0.71, 0.89)	6.70 (5.10, 8.26)	4.12	0.12	0.90	2.41
			Multi-target RF	0.82 (0.71, 0.88)	6.90 (5.27, 8.41)	4.65	0.10	0.89	2.34
Topography	Upslope	SOC	Direct RF	0.34 (−0.04, 0.53)	0.58 (0.41, 0.75)	0.39	0.10	0.56	1.23
			Multi-target RF	0.40 (0.08, 0.55)	0.55 (0.39, 0.72)	0.38	0.09	0.57	1.29
	Upslope	SIC	Direct RF	−0.22 (−0.44, −0.10)	0.82 (0.67, 0.96)	0.62	−0.16	0.02	0.91
			Multi-target RF	−0.18 (−0.40, −0.06)	0.81 (0.65, 0.93)	0.61	−0.16	0.04	0.92
	Upslope	Gypsum	Direct RF	0.71 (0.48, 0.83)	7.14 (3.83, 9.91)	3.12	−0.07	0.80	1.86
			Multi-target RF	0.72 (0.60, 0.80)	7.07 (3.78, 10.10)	3.16	−0.41	0.79	1.88
	Midslope	SOC	Direct RF	−0.19 (−1.01, 0.19)	0.63 (0.51, 0.74)	0.45	0.20	0.38	0.92
			Multi-target RF	0.24 (−0.14, 0.42)	0.51 (0.39, 0.61)	0.38	0.12	0.47	1.15
	Midslope	SIC	Direct RF	−0.14 (−0.40, 0.00)	0.74 (0.63, 0.82)	0.61	−0.16	0.09	0.94
			Multi-target RF	−0.09 (−0.24, 0.00)	0.72 (0.62, 0.82)	0.58	−0.13	0.06	0.96
	Midslope	Gypsum	Direct RF	0.32 (−0.10, 0.55)	6.21 (4.13, 8.18)	3.50	−0.05	0.50	1.21
			Multi-target RF	0.29 (−0.11, 0.53)	6.35 (4.05, 8.30)	3.54	−0.27	0.46	1.19
	Downslope	SOC	Direct RF	0.33 (−0.12, 0.58)	0.67 (0.48, 0.86)	0.47	0.11	0.57	1.22
			Multi-target RF	0.41 (0.10, 0.61)	0.63 (0.45, 0.81)	0.44	0.02	0.58	1.30
	Downslope	SIC	Direct RF	−0.24 (−0.77, 0.09)	0.79 (0.65, 0.94)	0.58	0.15	0.06	0.90
			Multi-target RF	−0.14 (−0.65, 0.15)	0.75 (0.61, 0.88)	0.58	0.14	0.10	0.94
	Downslope	Gypsum	Direct RF	0.72 (0.64, 0.79)	7.77 (5.66, 9.80)	5.06	−0.99	0.80	1.88
			Multi-target RF	0.70 (0.63, 0.76)	7.97 (5.71, 10.17)	5.10	−1.50	0.78	1.83

Site-blocked validation revealed pronounced target-specific differences in spatial transferability (Table 3; Figure 5). SOC performance remained largely unchanged relative to plot-blocked validation (direct RF: R^2 = 0.41 vs. 0.43; RMSE = 0.55% vs. 0.54%), with overlapping bootstrap intervals, indicating that SOC performance remained similar under oasis-level blocking within the Djerid study region. Gypsum performance declined moderately under site-blocked validation, with R^2 decreasing from 0.78 to 0.64, RMSE increasing from 5.81% to 7.53%, and RPD decreasing from 2.15 to 1.66. Prediction bias also shifted from +0.19% under plot-blocked validation to −0.47% under site-blocked validation, indicating reduced transferability and a possible influence of oasis-specific concentration differences.

Table 3. Effect of the blocking level on Random Forest prediction of SOC, SIC and gypsum under continuum-removed spectra. Plot-blocked GroupKFold holds out three complete plots per fold (18 plots across 6 folds; 36 validation samples per fold); site-blocked GroupKFold holds out all three plots of an oasis together (6 oases, 6 folds). Metrics were calculated from pooled held-out predictions; values in parentheses are 95% bootstrap confidence intervals from 1000 resamples of the pooled predictions. RMSE, MAE and bias are expressed as % dry soil mass. Negative R^2 values indicate performance worse than prediction of the held-out mean.

Target	Model	Validation	R^2 (95% CI)	RMSE (95% CI)	MAE	Bias	CCC	RPD
SOC	Direct RF	Plot-blocked GroupKFold	0.41 (0.20, 0.53)	0.55 (0.46, 0.65)	0.36	0.07	0.64	1.3
		Site-blocked GroupKFold	0.43 (0.24, 0.56)	0.54 (0.45, 0.63)	0.37	0.1	0.64	1.32
	Multi-target RF	Plot-blocked GroupKFold	0.44 (0.26, 0.56)	0.53 (0.44, 0.63)	0.35	0.05	0.65	1.33
		Site-blocked GroupKFold	0.45 (0.26, 0.56)	0.53 (0.44, 0.63)	0.35	0.05	0.64	1.35
SIC	Direct RF	Plot-blocked GroupKFold	0.30 (0.15, 0.42)	0.61 (0.54, 0.69)	0.46	−0.05	0.47	1.19
		Site-blocked GroupKFold	−0.28 (−0.47, −0.15)	0.82 (0.75, 0.90)	0.64	0.03	−0.1	0.88
	Multi-target RF	Plot-blocked GroupKFold	0.28 (0.14, 0.40)	0.62 (0.54, 0.69)	0.47	−0.04	0.45	1.18
		Site-blocked GroupKFold	−0.30 (−0.48, −0.16)	0.83 (0.75, 0.91)	0.64	0.04	−0.11	0.88
Gypsum	Direct RF	Plot-blocked GroupKFold	0.78 (0.64, 0.87)	5.81 (4.48, 6.89)	2.78	0.19	0.88	2.15
		Site-blocked GroupKFold	0.64 (0.49, 0.73)	7.53 (5.81, 8.94)	3.66	−0.47	0.76	1.66
	Multi-target RF	Plot-blocked GroupKFold	0.78 (0.66, 0.86)	5.82 (4.54, 7.13)	2.94	0.09	0.88	2.15
		Site-blocked GroupKFold	0.63 (0.46, 0.72)	7.63 (5.95, 9.06)	3.86	−0.48	0.75	1.64

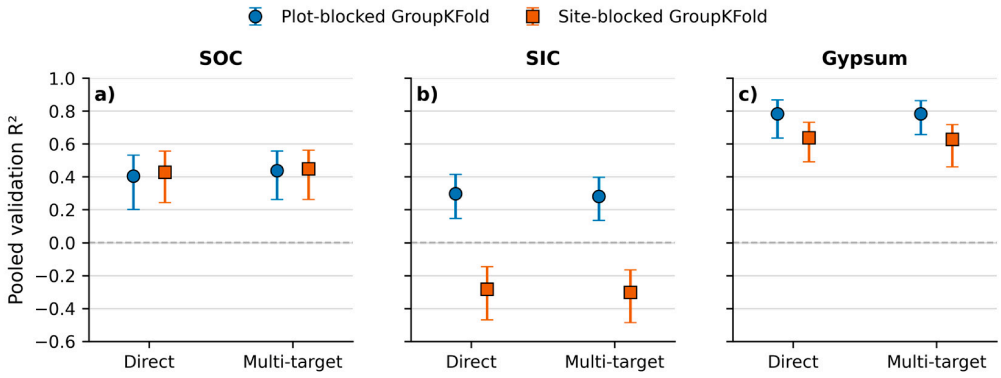


Figure 5. Effect of spatial blocking level on Random Forest prediction performance for SOC, SIC, and gypsum using continuum-removed spectra. Symbols represent R^2 values calculated from pooled held-out predictions under plot-blocked GroupKFold (18 plots; six folds) and site-blocked GroupKFold (six oases; six folds). Error bars indicate 95% bootstrap confidence intervals based on 1000 resamples of individual samples; these are probably too narrow, and intervals from plot and oasis level resampling are given in Table S7.

In contrast, SIC performance deteriorated markedly under site-blocked validation, with R^2 decreasing from 0.30 to −0.28 and CCC from 0.47 to −0.10. The non-overlapping sample-level bootstrap intervals for R^2 (0.15–0.42 under plot-blocked validation versus −0.47 to −0.15 under site-blocked validation support a substantial loss of predictive transferability across oases. Resampling whole plots still separated the two estimates (0.02 to 0.46 versus

−0.80 to −0.08) whereas oasis-level resampling gave overlapping intervals (−0.17 to 0.42 versus −1.64 to −0.01; Table S7), so this contrast should be regarded as indicative given the small number of oases. PLSR remained unstable under site-blocked validation, and the corresponding results are reported in Table S5.

3.4. Effects of Management and Topography Stratification

Stratification effects differed strongly among target properties (Table 2). Gypsum showed the most robust performance. Management-stratified models reached R^2 = 0.82 (single-target) and 0.80 (multi-target) in modern oases and R^2 = 0.83 and0.82 in traditional oases. Topography-stratified models also remained predictive, for upslope (R^2 = 0.71–0.72) and downslope (R^2 = 0.70–0.72) positions, but not for midslope (R^2 = 0.29–0.32, 95% CI −0.11 to 0.55). Thus, gypsum prediction remained transferable across held-out plots under both stratification approaches, with the strongest performance under management stratification.

SOC performance varied among subsets. The highest R^2 was obtained for the traditional oasis subset (R^2 = 0.53 for both single- and multi-target RF; RMSE = 0.49%; RPD = 1.45), whereas the midslope subset performed poorly and with very wide uncertainty (single-target RF R^2 = −0.19, 95% CI −1.01 to 0.19; multi-target RF R^2 = 0.24, 95% CI −0.14 to 0.42). Modern, upslope, and downslope models showed lower but consistently positive performance (Table 2).

SIC showed the opposite response. Its best plot-blocked GroupKFold performance was obtained with the global RF model (R^2 = 0.30 and 0.28 for single and multi-target workflows; Table 1), whereas stratification generally reduced predictive performance. Management-stratified SIC remained positive in modern oases (R^2 = 0.26–0.27) but was negative in traditional oases (R^2 = −0.20 to −0.17). All topography-stratified SIC models had negative R^2 under plot-blocked GroupKFold (−0.24 to −0.09), despite positive performance under KFold and LOOCV (Figure S4; Table S2). Thus, stratification improved or maintained prediction only for selected target–stratum combinations: it was most effective for gypsum, beneficial for SOC in specific subsets, but did not improve transferable SIC prediction.

Stratified models were evaluated using the same held-out partitions as the global models. Because each topographic stratum comprised only six plots across two outer folds and each management stratum nine plots, performance estimates were inherently uncertain. In addition, stratified models were calibrated using fewer samples per fold than the global model (36 for topographic strata and 84–96 for management strata, compared with 180). Furthermore, topographic models were evaluated on plots from oases excluded from calibration (Section 2.5). Consequently, performance differences between stratified and global models reflect differences in calibration-set size and transfer conditions, in addition to stratification. For example, the 95% bootstrap interval for the midslope SOC model ranged from −1.01 to 0.19. Consequently, differences between global and stratified models were within sampling uncertainty and were interpreted as indicative rather than conclusive.

3.5. Model Interpretation Using SHAP

SHAP analysis was used to identify wavelengths contributing most strongly to RF predictions. Mean absolute SHAP values quantify the overall importance of individual wavelengths across samples.

Wavelength-importance patterns differed among target properties but showed substantial overlap between SIC and gypsum (Figure 6). For SOC, the five highest-ranked wavelengths in the CR model clustered at 1842–1848 nm. SIC importance was concentrated mainly at 2314–2327 nm, whereas gypsum also showed strong contributions around 2324–2328 nm, together with additional wavelengths near 1926 and 2390 nm. Thus, SOC

relied on a distinct spectral region, while SIC and gypsum partly shared wavelength regions around 2320–2330 nm.

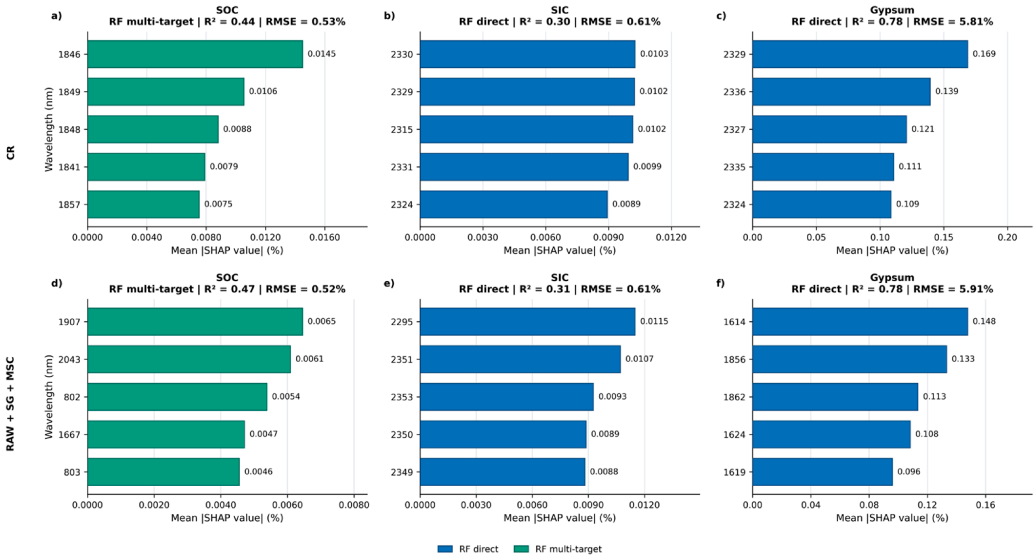


Figure 6. Wavelength importance for RF prediction of SOC, SIC, and gypsum based on mean absolute SHAP values. The five most important wavelengths are shown for the best-performing RF configuration within each target and preprocessing variant. (**a**–**c**), continuum-removed spectra (CR); (**d**–**f**), Savitzky–Golay first-derivative preprocessing with multiplicative scatter correction (RAW + SG + MSC). Bar colours distinguish single-target and multi-target RF models, and R^2 and RMSE are reported in each panel. Higher mean absolute SHAP values indicate greater contributions of individual wavelengths to model predictions. Abbreviations: SHAP, Shapley additive explanations; RF, random forest; SOC, soil organic carbon; SIC, soil inorganic carbon; CR, continuum removal; SG, Savitzky–Golay; MSC, multiplicative scatter correction.

The overlap between SIC and gypsum is consistent with their partly overlapping mineral spectral signatures and may contribute to the lower transferability of SIC relative to gypsum. However, SHAP identifies wavelengths used by the RF models rather than isolated chemical absorption mechanisms, and importance at correlated wavelengths should therefore not be interpreted as direct evidence of specific molecular or mineral features.

4. Discussion

4.1. Validation Design Determines the Interpretation of Model Performance

A central result of this study is that model performance depended strongly on validation strategy. Random KFold and LOOCV generally yielded higher apparent predictability than plot-blocked GroupKFold, with the largest discrepancy for SIC. For the global multi-target RF model, SIC R^2 increased from 0.28 under plot-blocked GroupKFold to 0.44 under KFold and 0.48 under LOOCV, accompanied by lower RMSE and higher CCC and RPD. Thus, non-grouped validation improved not only R^2 but the overall apparent accuracy and agreement of the model.

This difference is methodologically important because random KFold and LOOCV allow samples from the same plot to occur in both calibration and validation subsets. Such samples may share parent material, hydrological history, management, salinity, depth structure, and mineralogical background, thereby increasing similarity between training and validation data. Plot-blocked GroupKFold avoids this overlap by withholding entire plots and therefore provides a stricter test of transferability beyond the calibration plots [35,48]. Reliance only on KFold or LOOCV would consequently have overstated model transferability, particularly for SIC. These results support our hypothesis that grouped validation

provides more conservative and realistic estimates of predictive performance in hierarchically structured dryland datasets.

4.2. Target-Specific Predictability of SOC, SIC, and Gypsum

Prediction performance was strongly target-specific, following the order gypsum > SOC > SIC. Gypsum was the most predictable property, likely because of its diagnostic hydration-related absorption features and the wide concentration range represented in the dataset. RF models maintained high performance under plot-blocked GroupKFold, particularly after management stratification. This agrees with previous work showing more accurate VNIR prediction of gypsum than carbonates in arid and semi-arid soils, attributed to the stronger and more distinctive spectral features of gypsum [14]. Our results extend these findings to a stricter plot-blocked validation framework.

Beyond relative goodness-of-fit metrics, the practical utility of the models should be evaluated in relation to the concentration ranges encountered in the field. For SOC, the plot-blocked RMSE of 0.53% was small relative to the observed concentration range (0.04–4.20%), but it remained comparable in magnitude to the mean SOC content (0.74%) and to concentrations commonly observed in deeper soil horizons. Consequently, SOC predictions appear useful for distinguishing relatively carbon-rich surface soils from carbon-poor subsoils but are less suitable for resolving subtle differences within the low-concentration range that predominates in these soils. For gypsum, an RMSE of 5.81% across an observed range of 0–69.83%, together with an RPD of 2.15, indicates sufficient predictive performance for broad classification of soils according to gypsum content. Such discrimination may be useful for applications such as reclamation assessment and irrigation planning, although the model is less suited to precise quantification in highly gypsiferous horizons. In contrast, SIC predictions showed limited practical utility: an RMSE of 0.61% relative to an observed range of 0.21–3.80% and an RPD of 1.19 suggest performance near the lower limit of screening-level application. Moreover, the marked deterioration in SIC performance under site-blocked validation indicates that even this modest predictive capability is unlikely to transfer reliably to unsampled oases without site-specific or locally representative calibration data.

SOC showed moderate predictability, consistent with the low SOC concentrations (0.04–4.20%) and strong mineral background of these sandy, alkaline, hyper-arid soils. At low SOC contents, organic-matter-related VNIR signals may be weak relative to variation associated with carbonates, gypsum, salinity, and texture. Higher SOC prediction performance has been reported under other calibration conditions. For example, Marques et al. (2020) [49] obtained $R^2 = 0.74$ in gypsiferous agricultural soils using PLSR, while broader studies show strong dependence of SOC prediction on concentration range, mineralogy, preprocessing, and validation design [18,23,50]. The moderate performance observed here therefore likely reflects both the low SOC range and the more stringent plot-blocked validation.

SIC was the least transferable target. SIC comprises lithogenic and pedogenic carbonate pools influenced by parent material, hydrology, pH, evaporation, irrigation, dissolution–precipitation processes, and biological CO_2 inputs. Its relatively narrow concentration range (0.21–3.80%) and weak separation among management and topographic groups further constrained calibration. In addition, carbonate absorptions in the VNIR/SWIR region can overlap with clay-mineral and gypsum-related features. Previous studies similarly indicate lower robustness of carbonate than gypsum prediction and mineralogical interference around the 2300–2350 nm carbonate region [14,24,51]. Our results further show that moderate SIC performance under random validation does not necessarily translate into reliable prediction across held-out plots.

4.3. Preprocessing and Multi-Target Workflows Revealed Target-Specific Spectral Information

No single preprocessing strategy was optimal across all targets. Gypsum prediction was robust with RAW, RAW + SG + MSC, and CR spectra, with similar RF performance after continuum removal and RAW + SG + MSC ($R^2 = 0.78$) and the highest overall R^2 obtained with single-target PLSR on RAW spectra ($R^2 = 0.80$; Table S1). Nevertheless, RAW and RAW + SG + MSC retained substantial predictive information, consistent with previous studies showing that full-spectrum approaches can perform as well as or better than feature-specific continuum removal for gypsum and carbonates [14,51].

SOC and SIC were more sensitive to preprocessing. For SOC, RF performance improved markedly from RAW spectra to CR and RAW + SG + MSC, with the latter giving the highest performance. SIC showed a similar improvement relative to RAW, with RAW + SG + MSC and CR giving nearly identical RF performance ($R^2 = 0.31$ and 0.30), a difference not supported by the paired bootstrap (Table S8). These results indicate that both absorption enhancement and broader preprocessing of spectral structure can improve prediction, with their relative benefits depending on the target property.

Multi-target workflows likewise provided target-specific rather than uniform benefits. Joint prediction slightly improved global RF performance for SOC, lowered it for SIC under all three preprocessing approaches, and had little effect on gypsum; for CR spectra, none of these differences was distinguished from zero under plot-level resampling (Table S8). Shared spectra information among the three properties may therefore be useful for some targets, although relationships among targets may not remain stable across plots. The SHAP analysis supports partial spectral overlap: SOC importance was concentrated around 1842–1848 nm, whereas SIC and gypsum shared important wavelengths in the 2320–2330 nm region. Gypsum additionally showed contributions around 1926 and 2390 nm, consistent with hydration-related spectral features [18,20,25,26,52]. This broader spectral information may partly explain the greater robustness of gypsum prediction, whereas the overlap between SIC and gypsum may contribute to weaker SIC transferability. However, SHAP identifies predictive wavelength associations rather than isolated chemical absorption mechanisms.

4.4. Stratification Improves Prediction Only When It Matches Soil Structure

Environmental stratification did not universally improve prediction. Management stratification maintained strong gypsum performance and supported SOC prediction, whereas topographic stratification did not improve transferable SIC prediction. Stratification can therefore be beneficial when it reduces meaningful heterogeneity, but smaller calibration subsets may also reduce robustness [1,23,53], and the comparison between global and stratified models confounds stratification with calibration-set size (36–96 versus 180 samples).

The soil-property distributions partly explain these contrasting responses. SOC was structured mainly by depth, whereas gypsum showed clear management-related differentiation, with high concentrations concentrated in traditional oases and in deeper, downslope soils. This likely contributed to the strong performance of management-stratified gypsum models.

By contrast, SIC was only weakly separated among management and topographic classes. The global RF model therefore outperformed the stratified models under plot-blocked GroupKFold, while all topography-stratified SIC models produced negative R^2. Their positive performance under KFold and LOOCV suggests that topography captured local structure that did not transfer to held-out plots. Reduced calibration size may have further limited SIC model stability because of its narrow concentration range and weak group separation [35,54].

4.5. Implications for VNIR Monitoring of Dryland Soil Carbon and Gypsum

VNIR spectroscopy showed strongly target-dependent potential for dryland soil monitoring. Gypsum was predicted with high accuracy and good transferability, making it the most promising target for rapid screening in these hyper-arid oasis soils, although intervals from oasis-level resampling were wide (Table S7). SOC showed moderate predictability and may be useful for identifying broad spatial gradients, although low SOC concentrations and strong mineral-background effects limit precision. SIC remained the least transferable target: global RF models achieved only moderate performance, and stratification did not overcome this limitation.

Our gypsum results are consistent with the generally strong VNIR predictability reported for gypsum in arid soils, whereas SIC performance was more conservative than in several carbonate or SIC calibration studies [55,56]. This likely reflects the narrow SIC range, strong gypsum background, heterogeneity among plots, and stricter plot-blocked validation. More broadly, VNIR prediction of SOC and SIC is known to depend on preprocessing, model choice, mineralogy, soil depth, and calibration design [18,23,57]. Our results further show that validation design itself can alter the apparent predictive performance and therefore the practical interpretation of soil-spectroscopy models.

The contrast between plot- and site-blocked validation revealed clear target-specific differences in transferability. SOC performance remained similar across blocking levels, suggesting relatively consistent spectral–SOC relationships among oases, in agreement with the established potential of Vis–NIR spectroscopy for SOC estimation [58]. Gypsum prediction declined moderately under oasis blocking, indicating that the model captured a genuine spectral signal but was partly influenced by local concentration ranges. In contrast, SIC performance deteriorated strongly when entire oases were withheld, suggesting greater dependence on local mineralogical conditions. Carbonate prediction can be affected by spectral interference and masking from gypsum and other minerals, particularly in arid soils [10,14]. Thus, transferability depended more on the stability of the spectral–target relationship than on geographic distance alone.

From a practical perspective, SOC predictions were suitable for distinguishing carbon-rich from carbon-poor soils but less reliable at very low concentrations. Gypsum showed useful predictive performance (RMSE = 5.81%; RPD = 2.15), whereas SIC showed limited utility (RMSE = 0.61%; RPD = 1.19). Although RPD is commonly used to assess soil-spectroscopy models, fixed thresholds should be interpreted cautiously [58–60].

Finally, the results emphasize the importance of validation design. Random cross-validation can overestimate predictive performance when spatially related samples occur in both calibration and validation sets [35]. For this spatially clustered dataset, plot- and site-blocked validation therefore provided a more relevant assessment of transferability, while recognizing that spatial blocking may be overly conservative for some probability-based map-accuracy applications [61].

Overall, our hypotheses were supported to different degrees. The first hypothesis was fully supported because model performance and predictability were strongly target-specific: gypsum was predicted most reliably, SOC showed moderate performance, and SIC showed the weakest transferability. The second hypothesis was partially supported, as the effects of spectral preprocessing and modelling workflow varied among target properties. RAW spectra provided the highest predictive performance for gypsum, whereas RAW + SG + MSC yielded the best performance for SOC. For SIC, RAW + SG + MSC and continuum-removed (CR) spectra produced comparable results. However, several of these differences, as well as most differences between single- and multi-target modelling workflows, were within the uncertainty indicated by the bootstrap intervals, suggesting that the apparent performance advantages were not clearly distinguished by their uncertainty intervals. The

third hypothesis was strongly supported. Grouped validation generally yielded more conservative estimates of predictive performance than random K-fold cross-validation and LOOCV, with the largest discrepancies observed for SIC. Moreover, the contrast between plot-blocked and site-blocked validation demonstrated that model transferability depends on the spatial scale at which independence is imposed, and that the relevant validation scale varies among target properties.

5. Conclusions

VNIR spectroscopy showed clear target-specific potential in hyper-arid agricultural soils: gypsum was predicted robustly, SOC with moderate accuracy, whereas SIC remained poorly transferable across held-out plots and, especially, across held-out oases. More importantly, model performance depended strongly on validation design; random resampling generally produced more optimistic estimates of model transferability than grouped validation. Plot- and site-level blocking further revealed that transferability was strongly target-dependent: SOC performance remained similar under oasis-level blocking within the Djerid study region, gypsum prediction declined moderately, whereas SIC prediction deteriorated markedly under inter-oasis transfer.

These findings emphasize that reliable soil-spectroscopy applications require validation schemes that reflect the spatial and hierarchical structure of the intended prediction domain. For dryland monitoring, VNIR already appears well suited to rapid gypsum assessment and potentially useful for SOC screening, whereas transferable SIC prediction will require broader calibration datasets that better capture carbonate variability, mineralogical interference, and spatial heterogeneity. Expanding calibration across dryland regions and integrating spectrally complementary information may therefore be key to developing more robust multi-property monitoring frameworks.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/soilsystems10110140/s1. Figure S1: Conceptual scheme of traditional and modern oasis systems in southern Tunisia; Figure S2: Distribution of SOC, SIC, and gypsum across topography and oasis management; Figure S3: Correlation heatmap of target soil properties and supporting soil variables; Figure S4: RF model performance within management and topographic strata across validation schemes; Figure S5: Mean VNIR spectra of the 216 soil samples under the three spectral preprocessing scenarios; Figure S6: Study area and sampling design; Figure S7: Observed versus predicted values for SOC, SIC, and gypsum under the selected continuum-removed Random Forest workflow; Figure S8: Residual diagnostics of the linear mixed-effects models used for Figure 2; Table S1: Plot-blocked GroupK-Fold performance of global models under different spectral preprocessing and modelling work-flows; Table S2: Non-grouped validation performance of stratified RF models for SOC and SIC; Table S3: Effects of validation strategy on global RF model performance for SOC, SIC, and gypsum; Table S4: Descriptive statistics of soil properties in the dataset; Table S5: Effect of the blocking level on PLSR and Random Forest performance under continuum-removed spectra; Table S6: Fold de-sign of grouped validation; Table S7: Sensitivity of R^2 confidence intervals to the bootstrap resampling unit; Table S8: Paired bootstrap differences in R^2 between models evaluated on the same held-out samples; Table S9: Sensitivity of the linear mixed-effects models used for Figure 2 to the random-effects structure and to a $\log(1 + x)$ transformation of gypsum; Dataset S1: Reference SOC, SIC, and gypsum values, plot identifiers, and raw Vis-NIR reflectance spectra (400–2500 nm) of the 216 soil samples, with a readme sheet describing all data columns.

Author Contributions: Conceptualization, W.A., H.A. and W.W.; methodology, W.A., Y.F. and W.W.; software, W.A.; validation, W.A., H.Z. and W.W.; formal analysis, W.A.; investigation, W.A., H.Z., L.B. and P.P.; resources, W.A., Y.F., H.A., N.B. and W.W.; data curation, W.A.; writing—original draft preparation, W.A.; writing—review and editing, W.A., H.Z., L.B., P.P., N.B., R.B., H.A., Y.F. and W.W.; visualization, W.A.; supervision, W.W., H.A. and N.B.; project administration, W.W.; funding acquisition, W.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by project No. KoEF193, "Organic Carbon Dynamics in Sustainable Climate-Resistant Soils of Arid Tunisian Oases (OCDAT-Oases)", funded through the OeAD-KoEF programme Kooperation Entwicklungsforschung/Cooperation Development Research (KoEF/CDR) of the Austrian Federal Ministry of Education, Science and Research (BMBWF). Open Access Funding was provided by the University of Vienna.

Data Availability Statement: The supplementary tables and figures supporting the conclusions of this article are provided as Supplementary Material, and the raw Vis-NIR reflectance spectra (400–2500 nm) including reference SOC, SIC and gypsum values, plot identifiers used for grouped cross-validation, together with a readme file describing all data columns, are available from the corresponding author upon reasonable request.

Acknowledgments: We thank the Tunisian partners and the local oasis farmers for their support during field sampling, and L'Institut Agro Rennes-Angers/INRAE, Rennes, France, for access to the VNIR/NIRS spectroscopy facilities).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Aichi, H.; Fouad, Y.; Lili Chabaane, Z.; Sanaa, M.; Walter, C. Prediction Accuracy of Local and Regional Soil Total Carbon Models, Calibrated Based on Visible-near Infrared Spectra, in the Djerid Arid Region. *J. Near Infrared Spectrosc.* **2018**, *26*, 322–334. [CrossRef]
2. Reynolds, J.F.; Smith, D.M.S.; Lambin, E.F.; Turner, B.L.; Mortimore, M.; Batterbury, S.P.J.; Downing, T.E.; Dowlatabadi, H.; Fernández, R.J.; Herrick, J.E.; et al. Global Desertification: Building a Science for Dryland Development. *Science* **2007**, *316*, 847–851. [CrossRef] [PubMed]
3. Wang, X.; Dou, X.; Zhang, X.; Liu, H.; Li, H.; Meng, X. Development of Soil Spectral Allocation Models Considering the Effect of Soil Moisture. *Soil Tillage Res.* **2019**, *195*, 104374. [CrossRef]
4. Lal, R. Carbon Sequestration. *Phil. Trans. R. Soc. B* **2008**, *363*, 815–830. [CrossRef] [PubMed]
5. Naorem, A.; Jayaraman, S.; Dang, Y.P.; Dalal, R.C.; Sinha, N.K.; Rao, C.S.; Patra, A.K. Soil Constraints in an Arid Environment—Challenges, Prospects, and Implications. *Agronomy* **2023**, *13*, 220. [CrossRef]
6. Naorem, A.; Jayaraman, S.; Dalal, R.C.; Patra, A.; Rao, C.S.; Lal, R. Soil Inorganic Carbon as a Potential Sink in Carbon Storage in Dryland Soils—A Review. *Agriculture* **2022**, *12*, 1256. [CrossRef]
7. Raza, S.; Zamanian, K.; Ullah, S.; Kuzyakov, Y.; Virto, I.; Zhou, J. Inorganic Carbon Losses by Soil Acidification Jeopardize Global Efforts on Carbon Sequestration and Climate Change Mitigation. *J. Clean. Prod.* **2021**, *315*, 128036. [CrossRef]
8. Hamdi-Aissa, B.; Valles, V.; Aventurier, A.; Ribolzi, O. Soils and Brine Geochemistry and Mineralogy of Hyperarid Desert Playa, Ouargla Basin, Algerian Sahara. *Arid. Land Res. Manag.* **2004**, *18*, 103–126. [CrossRef]
9. Lal, R. Soil Carbon Sequestration Impacts on Global Climate Change and Food Security. *Science* **2004**, *304*, 1623–1627. [CrossRef] [PubMed]
10. Naimi, S.; Ayoubi, S.; Di Raimo, L.A.D.L.; Demattе, J.A.M. Quantification of Some Intrinsic Soil Properties Using Proximal Sensing in Arid Lands: Application of Vis-NIR, MIR, and pXRF Spectroscopy. *Geoderma Reg.* **2022**, *28*, e00484. [CrossRef]
11. Herrero, J.; Porta, J. The Terminology and the Concepts of Gypsum-Rich Soils. *Geoderma* **2000**, *96*, 47–61. [CrossRef]
12. Shen, Q.; Zhang, S.; Xia, K. Spectral Heterogeneity Analysis and Soil Organic Matter Inversion across Differences in Soil Types and Organic Matter Content in Dryland Farmland in China. *Sustainability* **2023**, *15*, 16310. [CrossRef]
13. Volkan Bilgili, A.; Van Es, H.M.; Akbas, F.; Durak, A.; Hively, W.D. Visible-near Infrared Reflectance Spectroscopy for Assessment of Soil Properties in a Semi-Arid Area of Turkey. *J. Arid. Environ.* **2010**, *74*, 229–238. [CrossRef]
14. Khayamim, F.; Wetterlind, J.; Khademi, H.; Robertson, A.H.J.; Cano, A.F.; Stenberg, B. Using Visible and near Infrared Spectroscopy to Estimate Carbonates and Gypsum in Soils in Arid and Subhumid Regions of Isfahan, Iran. *J. Near Infrared Spectrosc.* **2015**, *23*, 155–165. [CrossRef]
15. Laudicina, V.A.; Dazzi, C.; Delgado, A.; Barros, H.; Scalenghe, R. Relief and Calcium from Gypsum as Key Factors for Net Inorganic Carbon Accumulation in Soils of a Semiarid Mediterranean Environment. *Geoderma* **2021**, *398*, 115115. [CrossRef]
16. Raheb, A.; Asgari Lajayer, B.; Senapathi, V. The Effect of Short-Term Plants Cultivation on Soil Organic/Inorganic Carbon Storage in Newly Formed Soils. *Sci. Rep.* **2023**, *13*, 18500. [CrossRef] [PubMed]
17. Gozukara, G.; Hartemink, A.E.; Zhang, Y. Factors Driving Inorganic Carbon Levels in the Soils of the Conterminous USA. *CATENA* **2025**, *252*, 108841. [CrossRef]
18. Stenberg, B.; Viscarra Rossel, R.A.; Mouazen, A.M.; Wetterlind, J. Visible and Near Infrared Spectroscopy in Soil Science. In *Advances in Agronomy*; Elsevier: Amsterdam, The Netherlands, 2010; Volume 107, pp. 163–215.

19. Viscarra Rossel, R.A.; Cattle, S.R.; Ortega, A.; Fouad, Y. In Situ Measurements of Soil Colour, Mineral Composition and Clay Content by Vis–NIR Spectroscopy. *Geoderma* **2009**, *150*, 253–266. [CrossRef]
20. Viscarra Rossel, R.A.; Walvoort, D.J.J.; McBratney, A.B.; Janik, L.J.; Skjemstad, J.O. Visible, near Infrared, Mid Infrared or Combined Diffuse Reflectance Spectroscopy for Simultaneous Assessment of Various Soil Properties. *Geoderma* **2006**, *131*, 59–75. [CrossRef]
21. Swan, T.; Jang, H.J.; Huang, Y.-C.; Fidelis, C.; Yinil, D.; Bala, B.; Das, B.S.; Field, D. Comparative Analysis of Vis-NIR and MIR Spectroscopy for Predicting Soil Properties and Identifying Minerals at Smallholder Cocoa Farms across Papua New Guinea. *Soil Adv.* **2026**, *5*, 100094. [CrossRef]
22. Ben-Dor, E.; Chabrillat, S.; Demattê, J.A.M.; Taylor, G.R.; Hill, J.; Whiting, M.L.; Sommer, S. Using Imaging Spectroscopy to Study Soil Properties. *Remote Sens. Environ.* **2009**, *113*, S38–S55. [CrossRef]
23. Nocita, M.; Stevens, A.; Van Wesemael, B.; Aitkenhead, M.; Bachmann, M.; Barthès, B.; Ben Dor, E.; Brown, D.J.; Clairotte, M.; Csorba, A.; et al. Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring. In *Advances in Agronomy*; Elsevier: Amsterdam, The Netherlands, 2015; Volume 132, pp. 139–159.
24. Clark Roger, N. *Spectroscopy of Rocks and Minerals, and Principles of Spectroscopy*; USGS: Reston, VA, USA, 1999; Volume 3, pp. 3–58.
25. Hunt, G.R. Spectral signatures of particulate minerals in the visible and near infrared. *Geophysics* **1977**, *42*, 501–513. [CrossRef]
26. Zhang, J.; Cao, W.; Lu, Y.; Gázquez, F.; Krijgsman, W.; Zeng, X.; Zhong, Y.; Liu, W.; Liu, Q. A Novel Approach of Semi-Quantifying Gypsum in Sedimentary Rocks by Visible and Near-Infrared Diffuse Reflectance Spectroscopy. *Geochem. Geophys. Geosyst.* **2025**, *26*, e2024GC012118. [CrossRef]
27. Clark, R.N. Chapter 2. Spectroscopy of rocks and minerals, and principles of spectroscopy. In *Infrared Spectroscopy in Geochemistry, Exploration Geochemistry and Remote Sensing*; King, P.L., Ramsey, M.S., Swayze, G.A., Eds.; Mineralogical Association of Canada: Quebec, QC, Canada, 2004; pp. 17–55.
28. Stevens, A.; Nocita, M.; Tóth, G.; Montanarella, L.; Van Wesemael, B. Prediction of Soil Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy. *PLoS ONE* **2013**, *8*, e66409. [CrossRef] [PubMed]
29. Weindorf, D.C.; Chakraborty, S.; Herrero, J.; Li, B.; Castañeda, C.; Choudhury, A. Simultaneous Assessment of Key Properties of Arid Soil by Combined PXRF and V Is–NIR Data. *Eur. J. Soil Sci.* **2016**, *67*, 173–183. [CrossRef]
30. Wold, S.; Sjöström, M.; Eriksson, L. PLS-Regression: A Basic Tool of Chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130. [CrossRef]
31. Breiman, L.; Cutler, A.; Liaw, A.; Wiener, M. randomForest: Breiman and Cutlers Random Forests for Classification and Regression 2002, 4.7-1.2. Available online: https://cran.r-project.org/web/packages/randomForest/randomForest.pdf (accessed on 15 May 2026).
32. Gozukara, G.; Hartemink, A.E.; Huang, J.; Demattê, J.A.M. Prediction Accuracy of pXRF, MIR, and Vis-NIR Spectra for Soil Properties—A Review. *Soil Sci. Soc. Am. J.* **2025**, *89*, e70028. [CrossRef]
33. Grunwald, S.; Thompson, J.A.; Boettinger, J.L. Digital Soil Mapping and Modeling at Continental Scales: Finding Solutions for Global Issues. *Soil Sci. Soc. Am. J.* **2011**, *75*, 1201–1213. [CrossRef]
34. Meyer, H.; Reudenbach, C.; Hengl, T.; Katurji, M.; Nauss, T. Improving Performance of Spatio-Temporal Machine Learning Models Using Forward Feature Selection and Target-Oriented Validation. *Environ. Model. Softw.* **2018**, *101*, 1–9. [CrossRef]
35. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillera-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schröder, B.; Thuiller, W.; et al. Cross-validation Strategies for Data with Temporal, Spatial, Hierarchical, or Phylogenetic Structure. *Ecography* **2017**, *40*, 913–929. [CrossRef]
36. Allagui, W.; Brahim, N.; Allani, M.; Zougari, B.; Ibrahim, H.; Bol, R.; Aichi, H.; Wanek, W. Oasis Management and Topography Interactively Shape Soil Inorganic Carbon Dynamics in Hyper-Arid Soils. *Geoderma* **2026**, *471*, 117865. [CrossRef]
37. Kettler, T.A.; Doran, J.W.; Gilbert, T.L. Simplified Method for Soil Particle-Size Determination to Accompany Soil-Quality Analyses. *Soil Sci. Soc. Am. J.* **2001**, *65*, 849–852. [CrossRef]
38. Kocev, D.; Vens, C.; Struyf, J.; Džeroski, S. Tree Ensembles for Predicting Structured Outputs. *Pattern Recognit.* **2013**, *46*, 817–833. [CrossRef]
39. Segal, M.; Xiao, Y. Multivariate Random Forests. *WIREs Data Min. Knowl.* **2011**, *1*, 80–87. [CrossRef]
40. McKinney, W. Data Structures for Statistical Computing in Python. In Proceedings of the 9th Python in Science Conference, Austin, TX, USA, 28 June–3 July 2010; pp. 56–61.
41. Harris, C.R.; Millman, K.J.; Van Der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array Programming with NumPy. *Nature* **2020**, *585*, 357–362. [CrossRef] [PubMed]
42. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [CrossRef] [PubMed]
43. Pedregosa, F.; Varoquaux, G.; Org, N.; Gramfort, A.; Michel, V.; Fr, L.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
44. Hunter, J.D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [CrossRef]
45. Waskom, M. Seaborn: Statistical Data Visualization. *JOSS* **2021**, *6*, 3021. [CrossRef]

46. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
47. Savitzky, A.; Golay, M.J.E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal*. *Chem*. **1964**, *36*, 1627–1639. [CrossRef]
48. Chen, H.; Song, Q.; Tang, G.; Feng, Q.; Lin, L. The Combined Optimization of Savitzky-Golay Smoothing and Multiplicative Scatter Correction for FT-NIR PLS Models. *ISRN Spectrosc*. **2013**, *2013*, 642190. [CrossRef]
49. Marques, M.; Álvarez, A.; Carral, P.; Esparza, I.; Sastre, B.; Bienes, R. Estimating Soil Organic Carbon in Agricultural Gypsiferous Soils by Diffuse Reflectance Spectroscopy. *Water* **2020**, *12*, 261. [CrossRef]
50. Dotto, A.C.; Dalmolin, R.S.D.; Ten Caten, A.; Grunwald, S. A Systematic Study on the Application of Scatter-Corrective and Spectral-Derivative Preprocessing for Multivariate Prediction of Soil Organic Carbon by Vis-NIR Spectra. *Geoderma* **2018**, *314*, 262–274. [CrossRef]
51. Gomez, C.; Lagacherie, P.; Coulouma, G. Continuum Removal versus PLSR Method for Clay and Calcium Carbonate Content Estimation from Laboratory and Airborne Hyperspectral Measurements. *Geoderma* **2008**, *148*, 141–148. [CrossRef]
52. Yeşilbaş, M.; Vu, T.H.; Hodyss, R.; Poch, O.; Schmitt, B.; Choukroun, M.; Johnson, P.V.; Bishop, J.L. Geochemical Transformations of Gypsum Under Multiple Environmental Settings and Implications for Ca-Sulfate Detection on Mars. *ACS Earth Space Chem*. **2025**, *9*, 433–444. [CrossRef] [PubMed]
53. Guerrero, C.; Wetterlind, J.; Stenberg, B.; Mouazen, A.M.; Gabarrón-Galeote, M.A.; Ruiz-Sinoga, J.D.; Zornoza, R.; Viscarra Rossel, R.A. Do We Really Need Large Spectral Libraries for Local Scale SOC Assessment with NIR Spectroscopy? *Soil Tillage Res*. **2016**, *155*, 501–509. [CrossRef]
54. Chen, S.; Xu, H.; Xu, D.; Ji, W.; Li, S.; Yang, M.; Hu, B.; Zhou, Y.; Wang, N.; Arrouays, D.; et al. Evaluating Validation Strategies on the Performance of Soil Property Prediction from Regional to Continental Spectral Data. *Geoderma* **2021**, *400*, 115159. [CrossRef]
55. Bai, Z.; Chen, S.; Hong, Y.; Hu, B.; Luo, D.; Peng, J.; Shi, Z. Estimation of Soil Inorganic Carbon with Visible Near-Infrared Spectroscopy Coupling of Variable Selection and Deep Learning in Arid Region of China. *Geoderma* **2023**, *437*, 116589. [CrossRef]
56. Wang, Y.; Yin, K.; Hu, B.; Hong, Y.; Chen, S.; Liu, J.; Yang, L.; Peng, J.; Shi, Z. Ensemble and Transfer Learning of Soil Inorganic Carbon with Visible Near-Infrared Spectra. *Geoderma* **2025**, *456*, 117257. [CrossRef]
57. Viscarra Rossel, R.A.; Behrens, T.; Ben-Dor, E.; Chabrillat, S.; Demattê, J.A.M.; Ge, Y.; Gomez, C.; Guerrero, C.; Peng, Y.; Ramirez-Lopez, L.; et al. Diffuse Reflectance Spectroscopy for Estimating Soil Properties: A Technology for the 21st Century. *Eur. J. Soil Sci*. **2022**, *73*, e13271. [CrossRef]
58. Bellon-Maurel, V.; Fernandez-Ahumada, E.; Palagos, B.; Roger, J.-M.; McBratney, A. Critical Review of Chemometric Indicators Commonly Used for Assessing the Quality of the Prediction of Soil Attributes by NIR Spectroscopy. *TrAC Trends Anal*. *Chem*. **2010**, *29*, 1073–1081. [CrossRef]
59. Chang, C.-W.; Laird, D.A.; Mausbach, M.J.; Hurburgh, C.R. Near-Infrared Reflectance Spectroscopy–Principal Components Regression Analyses of Soil Properties. *Soil Sci*. *Soc*. *Am*. *J*. **2001**, *65*, 480–490. [CrossRef]
60. Bellon-Maurel, V.; McBratney, A. Near-Infrared (NIR) and Mid-Infrared (MIR) Spectroscopic Techniques for Assessing the Amount of Carbon Stock in Soils—Critical Review and Research Perspectives. *Soil Biol*. *Biochem*. **2011**, *43*, 1398–1410. [CrossRef]
61. Wadoux, A.M.J.-C.; Heuvelink, G.B.M.; De Bruin, S.; Brus, D.J. Spatial Cross-Validation Is Not the Right Way to Evaluate Map Accuracy. *Ecol*. *Model*. **2021**, *457*, 109692. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.