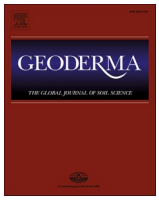
Geoderma 472 (2026) 117894

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma



Internal soil standard as a tool to assess and correct spectral variability between laboratories for a practical quantitative utilization

Marmar Sabetizade [a,*], Bas Van Wesemael [a], Vincent Baeten [b], Fabio Castaldi [c], Sabine Chabrillat [d,e], José A.M. Demattê [f], Asa Gholizadeh [g], Stephan Haefele [h], Konstantinos Karyotis [i], Aleš Klement [g], Robert Milewski [d], Simone Pascucci [j], Louis Paternostre [b], Stefano Pignatti [j], Jose Lucas Safanelli [k], Jonathan Sanderman [k], Thomas Schmid [l], Nikolaos Tziolas [m], Eyal Ben-Dor [n]

[a] *Earth and Life Institute, Université catholique de Louvain, Belgium*
[b] *Walloon Agricultural Research Centre (CRA-W), Quality and Authentication of Products Unit, Knowledge and Valorization of Agricultural Products Department, Chaussée de Namur, 24, 5030 Gembloux, Belgium*
[c] *Institute of BioEconomy, National Research Council of Italy (CNR), Via Giovanni Caproni 8, 50145 Firenze, Italy*
[d] *Helmholtz Center for Geosciences (GFZ), Section Remote Sensing and Geoinformatics, 14473 Potsdam, Germany*
[e] *Leibniz University Hannover (LUH), Institute of Earth System Sciences, Section Soil Science, 30419 Hannover, Germany*
[f] *Department of Soil Science, Luiz de Queiroz College of Agriculture, University of São Paulo, Brazil*
[g] *Department of Soil Science and Soil Protection, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague 16500 Prague, Czech Republic*
[h] *Soil Health and Management Department, Rothamsted Research, Harpenden, UK*
[i] *Aristotle University of Thessaloniki, School of Agriculture, Thessaloniki, Greece*
[j] *Institute of Methodologies for Environmental Analysis, National Research Council of Italy (CNR), C.da S. Loja s.n.c., 85050 Tito Scalo, Potenza, Italy*
[k] *Woodwell Climate Research Center, Falmouth MA, USA*
[l] *Department of Environment, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain*
[m] *Department of Soil, Water, and Ecosystem Sciences, University of Florida, Gainesville 32611, FL, USA*
[n] *Porter School of Environment and Earth Sciences, Faculty of Exact Sciences, Tel Aviv University, 69978 Ramat Aviv, Tel Aviv, Israel*

ARTICLE INFO

Handling Editor: Dr. B Heung

Keywords:
Internal Soil Standard (ISS)
Spectral variability correction
Ring trial
Quality control criterion
Soil carbon analysis
VNIR–SWIR harmonization

ABSTRACT

Soil spectroscopy is increasingly used to provide accurate and affordable prediction of various soil properties. However, variability in instrument characteristics and operational protocols still hinders the integration and harmonization of Visible-Near InfraRed- Short Wave InfraRed (VNIR–SWIR) Soil Spectral Libraries (SSLs). This study examines the use of 99% pure silica sand, Lucky Bay (LB), as an Internal Soil Standard (ISS), to reduce spectral variability and correct the systematic errors across laboratories. A ring trial test using 60 samples across 11 laboratories was conducted to assess the effect of the ISS correction on spectral consistency and model performance for Soil Organic Carbon (SOC) and clay content predictions. Spectral correction using the ISS reduced the Standard Deviation (SD) in reflectance by 13–70%, effectively reducing dissimilarity across instruments. Partial Least Squares Regression (PLSR) modeling using mean spectra of all instruments showed that a standard protocol resulted in prediction accuracy (Ratio of Performance to InterQuartile range (RPIQ) = 1.50) comparable to a frequently used reference instrument (Foss XDS, RPIQ = 1.57). While models built on individual datasets performed well, combining non-corrected spectra reduced prediction performance (RPIQ) by 11%, which was decreased to 8% after ISS correction. Therefore, we concluded that this approach is particularly useful for increasing interoperability of soil spectral datasets acquired across different instruments and laboratory conditions, which is key in the development of large-scale SSLs. Furthermore, reflectance variation at 1700 nm of the ISS was found as a practical Quality Assurance and Quality Control (QA/QC) indicator, offering a real-time baseline for evaluating instrument performance.

* Corresponding author.
E-mail address: Marmar.sabetizadeh@uclouvain.be (M. Sabetizade).

https://doi.org/10.1016/j.geoderma.2026.117894
Received 12 November 2025; Received in revised form 3 June 2026; Accepted 8 June 2026

1. Introduction

Soil spectroscopy is a rapid, cost-efficient, non-destructive, and reproducible analytical method, increasingly favored over more costly soil analysis techniques (Kopačková and Ben-Dor, 2016; Poppiel et al., 2022). Diffuse reflectance spectroscopy in soil and environmental sciences is advancing rapidly and has facilitated the generation of quantitative data for soils (e.g., Demattê et al., 2025; Safanelli et al., 2023; Viscarra Rossel et al., 2022). Representative Soil Spectral Libraries (SSLs) and precisely measured reference data are essential for the successful application of Visible-Near InfraRed- Short Wave InfraRed (VNIR–SWIR) spectroscopy in laboratory settings (Shepherd et al., 2022). From local to extensive national or continental databases, SSLs have been established worldwide, with corresponding laboratory measurements of several attributes (Orgiazzi et al., 2018; Ramirez-Lopez et al., 2019; Safanelli et al., 2025; Tóth et al., 2013; Viscarra Rossel et al., 2016).

To expand an SSL from regional to national and international scale and to maintain consistency with current and future measurements, a reliable and uniform procedure is needed (Romero et al., 2018; Viscarra Rossel and Bouma, 2016). Given the high sensitivity of soil spectral data to systematic and non-systematic effects in measurements, adherence to established standards is critical to facilitate merging and harmonization among SSLs worldwide (Ben-Dor et al., 2015). However, despite ongoing efforts to harmonize spectral methods, the lack of standardized protocols for reliable reflectance measurements in both laboratory and field settings remains challenging (Ben-Dor et al., 2024b; Mészáros et al., 2025). This need for standardization was emphasized by Viscarra Rossel et al. (2016), who suggested a first protocol for the Global Soil Spectral Library (GLOSIS), later recognized by the Global Soil Partnership (GSP) of the Food and Agriculture Organization (FAO) of the United Nations (UN) and its Global Soil Laboratory Network initiative on soil spectroscopy (FAO GLOSOLAN, 2023). To that end, an action was taken by the Institute of Electrical and Electronics Engineers (IEEE) Standards Association (SA), which in 2020 initiated a special working group (P4005) to develop an ISO standard and protocol to measure soil spectra in laboratory and field conditions (Ben-Dor et al., 2024a). The P4005 protocol utilizes a common Internal Soil Standard (ISS) using the Lucky Bay (LB) beach sand from southwest Australia (Ben-Dor et al., 2015). By addressing systematic effects, it offers a reliable method for harmonizing SSLs and improving the accuracy of predictive models based on diverse SSLs.

The first worldwide SSL, comprising spectra from various origins, measured with different devices and protocols under unknown maintenance circumstances, was created by Viscarra Rossel et al. (2016). They used wavelet transformation to correct noise effects and enhance the performance of models for soil characterization. Padarian et al. (2019) presented a transfer learning procedure to "localize" and refine a generic soil spectral model using fewer local data. Whether mathematical corrections alone can fully harmonize spectral data collected through diverse approaches remains uncertain. Furthermore, one of the most significant points in analytical chemistry is the necessity of following best practices prior to algorithmic interventions (Willis, 1972).

Pimstein et al. (2011) highlighted significant differences in diffuse reflectance spectroscopy measurements among three different instruments across the VNIR–SWIR regions (400–2500 nm). These authors suggested a standardization technique using the ISS concept. This method involves measuring the same sample under benchmark conditions, and under the user's specific conditions; observed differences would be attributed to the difference in measurement conditions (e.g., apparatus, sample preparation, and lighting). The ISS should be homogeneous with grain sizes and shapes similar to soil particles, stable over time and space, and spectrally featureless across the 400–2500 nm spectral range (Pimstein et al., 2011). Up till now, the ISS method uses beach sand samples from LB and Wylie Bay (WB) in southwest Australia, assessed at the Commonwealth Scientific and Industrial Research Organization (CSIRO) in Perth using a carefully calibrated benchmark setup spectrometer. The LB and WB sands are rich in quartz (99% and 90%, respectively), and spectrally featureless and stable because they do not introduce spectral artifacts during the standardization process (Francos et al., 2021; Notesco et al., 2019). A Correction Factor (CF) derived from ISS measurements under both local laboratory and CSIRO ISS conditions is applied to all of the user's spectral readings to correct most systematic effects. This assumes alignment with the benchmark setup, enabling merging of SSLs collected under different conditions. Sharing standard samples across laboratories worldwide could facilitate the development of calibration transfer across various labs. Additionally, the ISS method provides a CF and enables monitoring of the spectrometer's stability and performance over both short and long periods (Ben-Dor et al., 2015). Another important use of the ISS is to monitor the condition of the laboratory where the measurements are carried out. A stable ISS can provide insight into the stability of the measurement system and its potential deterioration before any soil spectral acquisition is taken.

Ben-Dor et al. (2015) proposed the ISS procedure to reduce uncertainty in VNIR–SWIR soil spectral measurements caused by varying environmental and instrumental conditions. Testing five soil samples using four Analytical Spectral Device (ASD) spectrometers across different setups and locations, they demonstrated its effectiveness in minimizing systematic errors. Kopačková and Ben-Dor (2016) expanded this work by analyzing additional soil samples with two spectrometer models (ASD and Spectral Evolutions), further validating the ISS technique's ability to align SSLs. Gholizadeh et al. (2017) compared ISS and non-ISS methods, showing improved SSL stability by correcting systematic variations. Romero et al. (2018) applied ISS to Brazilian tropical soils using three spectrometers, showing reduced spectral variability and improving clay content predictions. Demattê et al. (2019) described variances in soil chemical analysis between four laboratories using four types of spectral instruments. They concluded that applying the protocol optimized laboratory accuracy. Francos et al. (2022) found that effectively combining SSLs is possible when using an ISS, such as LB. However, most of the larger-scale SSLs (spanning nations or continents) still lack standardized ISS practices. Notably, Tziolas et al. (2019) created the first semi-continental SSL using ISS, showcasing its potential for harmonizing diverse datasets. Despite its success, ISS remains underutilized in large-scale SSLs.

Nevertheless, variation in spectral responses across different VNIR–SWIR instruments has rarely been quantified, and its implications for soil property predictions when using spectra from different instruments or from a single instrument over time are not well understood. Additionally, there is a lack of studies providing spectral indicators to detect the signal variability either between instruments or over time for a single instrument. To address these gaps, we conducted a ring trial experiment, leveraged from the Soil Spectroscopy for Global Good initiative (https://soilspectroscopy.org/) with three objectives: i) to quantify the effect of ISS correction on soil spectral variability across laboratories worldwide, ii) to determine the impact of systematic and overall spectral variability on the prediction accuracy of Soil Organic Carbon (SOC) and clay content, and iii) to propose a practical and straightforward criterion for Quality Assurance and Quality Control (QA/QC) applicable across laboratories.

Following Ben-Dor et al. (2015), we used LB, the recommended and more appropriate white sand, as an ISS and the CF to investigate the applicability and robustness of the ISS correction. We will investigate to what extent the ISS can be used as a QA/QC criterion. For this criterion to be practical, it must be applicable in real time and feasible in all laboratories, requiring a straightforward method. Consequently, the reflectance at 1700 nm was adopted as a quality index, based on the IEEE P4005 (Ben-Dor et al., 2024b) quality indicator recommendations. The wavelength of 1700 nm was selected for spectral analysis interpretation due to its stability in soil reflectance measurements, making it

a reference for assessing albedo variations, as it does not coincide with significant soil absorbance features (Ben-Dor et al., 2024b; Karyotis et al., 2023). This work examines the impact of spectral variation on the prediction accuracy of SOC and clay content using Partial Least Squares Regression (PLSR) models, with a particular focus on defining an acceptable range of variability based on the reflectance at 1700 nm in the ISS framework.

2. Material and methods

A ring trial experiment involves distributing an identical set of samples to various laboratories to be analyzed using standardized methods. The purposes of this approach are to assess the consistency and accuracy of analytical results among different participating laboratories and to support QC by checking on the precision and variability within each lab, as outlined by Safanelli et al. (2023). By identifying and quantifying biases and variances, it is possible to develop strategies that enhance uniformity and to share effective practices. Even though test programs have been common practice in soil science (FAO, 2023), soil spectroscopy in the solar domain has rarely been included in routine lab ring trials, with only a few instances documented in scholarly works (Ben-Dor et al., 2015; Gholizadeh et al., 2021; Pimstein et al., 2011). In this section, we detail the formation of an expansive network for soil spectroscopy ring trials, focusing on evaluating VNIR–SWIR instruments and laboratories.

2.1. Soil samples

Sixty highly diverse USA soil samples were examined at the USDA National Soil Survey Center, Kellogg Soil Survey Laboratory (USDA NSSC KSSL) for process quality control of various instruments. These samples were selected because they often have an extreme value of a particular property. To ensure consistency across laboratories, all samples were centrally prepared, they were air-dried and sieved (< 2 mm) to the Fine Earth (FE) fraction at Woodwell Climate Research Center. These samples were thoroughly homogenized and then split with different-sized riffle boxes into 50 g lots. This standardized preparation step was critical to minimize variability introduced prior to spectral acquisition. Following the preparation and wet chemistry analysis, the sets of these 60 samples were shipped to the 11 organizations participating in the trial (Table 1).

In this research, the focus was on two soil properties: SOC content (g kg^{-1}) and clay content (%). These properties were selected because of their distinct relationships with absorbance patterns in the VNIR–SWIR spectra (Dangal et al., 2019) and because of their role in soil functions and services. The analysis of SOC was carried out using the equation $SOC = TC - 0.12CaCO_3$, where Total Carbon (TC) was determined through dry combustion, and $CaCO_3$ was measured using a manometer, which measured the exhaust CO_2 from acid interaction with the soil outlined by Soil Survey Staff (2022). The measurement of clay content was conducted using the pipette method of particle size analysis, as outlined by the Soil Survey Staff in 2022.

2.2. Measurements and instruments

Twelve instruments, representing different organizations, were utilized to obtain VNIR–SWIR spectra for the 60 ring trial samples. These instruments were from three manufacturers, namely ASD (Boulder, CO, USA), Spectral Evolution (Spectral Evolution, Inc., USA), and Foss XDS Rapid Content Analyzer (FOSS NIR Systems Inc., Laurel, MD, USA), comprising various models and internal optics. All spectrometers covered the 350–2500 nm ranges, while the resolution among their detectors varies from 2 to 8 nm. Each laboratory was instructed to warm up its equipment for at least 60 or 90 min, depending on the spectrometer type. Each sample was subjected to three replicates, with 25 co-added scans per replicate (Table 1). Spectral range was 350–2500 nm for ASD and Spectral Evolution instruments and 400–2500 nm for Foss XDS. Output spectral resolution was 1 nm for ASD and Spectral Evolution instruments and 2 nm for Foss XDS.

Table 1
Instruments' metadata information provided by the ring trial participants.

ID	Name	Acronym	Instrument	Measurement setup
1	Tel Aviv University, Remote Sensing Lab, Israel.	TAU	ASD FieldSpec® 3	Contact probe[a]
2	Aristotle University of Thessaloniki, School of Agriculture, Laboratory of Remote Sensing, Spectroscopy and GIS (AUTH-RSL), Greece.	AUTH	Spectral Evolution PSR + 3500	Contact probe
3	UCLouvain, Earth and Life Institute, Belgium.	UCL	ASD FieldSpec® 3	Contact probe
4	Czech University of Life Sciences, Czech Republic.	CZU	ASD FieldSpec® 3	Contact probe
5	Helmholtz Centre for Geosciences, Spectroscopy lab, Germany.	GFZ	ASD FieldSpec® 3	Black box[b]
6	Rothamsted Research spectral laboratory, United Kingdom.	RTMS	ASD FieldSpec® 4	Contact probe
7	University of São Paulo, Remote Sensing Applied to Soils, Brazil.	SPU	ASD FieldSpec® 3	Contact probe
8	Consiglio Nazionale delle Ricerche – Istituto di Metodologie per l'Analisi Ambientale, Italy.	CNR_IMAA	ASD FieldSpec® 4	Contact probe
9	National Research Council of Italy (CNR), Institute of BioEconomy, Italy.	CNR_IBE	Spectral Evolution RS-5400	Contact probe
10	Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas, Spain.	CIEMAT	ASD FieldSpec® 3	Contact probe
11	Centre Wallon de Recherches Agronomiques, Belgium.	CRAW	ASD FieldSpec® 4	Quartz-bottom cell[c]
12	Centre Wallon de Recherches Agronomiques, Belgium.	XDS	FOSS XDS	Quartz-bottom cell

[a] - Contact probe: An optic assembly that touches the soil samples through a sapphire window. It maintains a stable viewing geometry and nadir tungsten halogen radiation and is a commercial product of ASD and Spectral Evolution.
[b] - Black box: A closed chamber isolated from artificial light and sunlight, with a fixed viewing and illumination geometry (Ben-Dor et al., 2015; Gholizadeh et al., 2021).
[c] - Quartz-bottom cell: A rectangular sample cell with a quartz window, used in FOSS systems, where the sample is measured from below (TA-0379, Terrace Assembly).

All participating partners followed the laboratory standard protocol (version 10) provided by IEEE P4005 (Ben-Dor et al., 2024b) and employed LB as ISS. According to this protocol, the white reference and ISS spectra were measured at the beginning and end of each batch, consisting of five soil samples split into triplicates and scanned to ensure consistency. The measurement setups included the contact probe (10 instruments), the black box method (1 instrument), and the quartz-bottom cell (1 instrument). The Foss XDS instrument minimizes sources of error due to its controlled measurement environment, stable light sources, sample holders, and reliable probe performance (Gholizadeh et al., 2021). Therefore, the measurements made by the FOSS instrument, which used quartz-bottom sample cells, were considered as a reference to compare the results obtained from other instruments. To address variability in resolution, all spectra were resampled at 1 nm intervals and subsequently truncated to the 450–2350 nm range (Karyotis et al., 2023).

Spectra from individual instruments were corrected to the ISS measured by CSIRO as a mother spectrum. A CF between LB spectra at each laboratory and CSIRO was applied according to Ben-Dor et al. (2015; equation (1)). To achieve this, the mean reflectance of ISS spectra (LB_{mean}) measured before and after each batch of samples was computed. These means were then applied in the following equation to compute the CF for each wavelength (Eq. (1). The correction was executed by multiplying this CF by the sample's spectra, aiming to harmonize the spectra acquired from different instruments.

$$CF = 1 - ((LB_{mean} - ISS_{CSIRO})/LB_{mean}) \quad (1)$$

where LB_{mean} is the mean reflectance of LB spectra measured before and after each batch of samples at each institute setups, and ISS_{CSIRO} is the reflectance of the ISS reference (LB) measured at CSIRO setup.

2.3. Statistical analysis

Initially, a soil with an extreme organic carbon value (491 g kg^{-1} SOC) was removed from the dataset as an outlier for modeling (n = 59), while all 60 samples were retained for variability tests. This step was taken to ensure that the dataset was limited to mineral soils with less than 180 g kg^{-1} SOC. Then, an analysis was conducted on SOC and clay data to assess their normality. Skewness, kurtosis, and the Kolmogorov-Smirnov normality test were employed for this evaluation. SOC data showed a skewness value greater than 3 and were therefore transformed using the Box-Cox method to achieve a normal Probability Distribution Function (PDF). Prediction performance metrics were calculated after restoring the SOC data to its original measurement unit. A statistical summary of the soil samples from the ring trial is provided in Table 2.

2.4. Variability tests

Exploratory data analysis was conducted to assess the spectral variance both within each instrument and among the instruments. To achieve this, the range of mean and Standard Deviation (SD) of reflectance values across all wavelengths (450–2350 nm) for each sample among 11 instruments was compared. Also, the average decrease in SD of reflectance values for all 60 ring trial samples before and after ISS correction (Eq. (1) was calculated among all instruments. This analysis aimed to determine whether this correction could potentially reduce the variability among different instruments. Prior to inferential testing, the normality of the paired differences, calculated as values after correction minus values before correction, was assessed for each instrument using Shapiro-Wilk tests. When the normality assumption was met, paired t-tests were used to assess whether the mean of the paired differences differed significantly from zero. When normality was violated, Wilcoxon signed-rank tests were applied instead as non-parametric alternative to assess whether the median of the paired differences differed from zero.

Principal Component Analysis (PCA) was performed on both the non-corrected and corrected datasets to identify the major patterns of variance and to demonstrate the dimensionality of the spectral data. The first two Principal Components (PC1 and PC2) were selected for detailed analysis due to their high explanatory power in capturing the majority of variance within the dataset. The scores of these components were plotted against each other to visually assess the spread. Normality assumptions were verified using the Shapiro-Wilk test, then a paired *t*-test analysis between average distances to the centroid in PCA space was conducted to assess clustering of the spectra corresponding to each instrument, both before and after ISS correction. Spectral correlations between each individual instrument and the reference instrument were computed at each wavelength by aligning paired observations by sample, ensuring strict one-to-one correspondence across datasets. Since the assumptions required for Pearson correlation were not consistently met across instruments and wavelengths, Spearman rank correlation coefficients were then calculated independently for each wavelength using the aligned paired reflectance values.

2.5. Modelling

The Foss XDS instrument, used in large SSLs such as the European Union Land Use/Cover Area frame Statistical survey (LUCAS), was considered the reference (Orgiazzi et al., 2018). Foss XDS spectrometers are widely used for VNIR–SWIR measurements due to their high precision, stability, strong repeatability, and standardization capabilities. The widespread adoption of Foss XDS instruments in soil science further underscores their reliability for large-scale projects like LUCAS (Castaldi et al., 2018; Stevens et al., 2013). A distinct PLSR model was developed using the spectra of the Foss XDS instrument, with model optimization performed using internal cross-validation to evaluate its predictive capability and to determine the extent to which the reference instrument could provide reliable results.

In the next step, soil spectra obtained from 11 instruments (excluding the Foss XDS) were averaged to create a mean spectral dataset of 59 ring trial samples (after removing one outlier sample). Then, this mean dataset was used to build models for predicting the selected soil properties, i.e., SOC and clay contents. The PLSR model was used as a standard chemometric algorithm to compare the impact of spectral variation. The spectra were both centered and scaled as a preprocessing step in the model. The modeling was performed on the entire ring trial dataset, and the optimal number of components was determined through a grid search process aiming to minimize the Root Mean Square Error (RMSE) using internal 5-fold cross-validation, repeated 30 times.

To introduce controlled variability in the spectral data, different coefficients (ranging from 0.1 to 2 in 0.1 intervals), known as Z-scores, were applied to all ranges of wavelengths for each ring trial sample (59 samples). This systematic approach was selected as we aimed to translate the controlled variability in the spectra back into variability observed in the ISS spectra, enabling us to introduce an index for QA/QC (variation is shown as an example in sample 55, Fig. S1). These Z-scores were multiplied by the SD of the mean dataset and added to the mean spectra, generating twenty additional spectral datasets. These twenty new datasets were used as new inputs for the model developed on the mean spectra. This allowed for an assessment of how well the models perform when dealing with the spectra that exhibit increased variability.

To assess the impact of ISS correction on model performance when building a SSL (merging datasets), models were constructed using individual instruments and mixed datasets composed of samples from multiple laboratories. Five distinct groups of 59 soil samples were generated, each composed of samples randomly selected from different laboratories to ensure diversity and generalizability. For each group, PLSR models were trained, both before and after ISS correction, resulting in models for comparison. ANOVA assumptions were verified by

Table 2
Statistical summary of SOC and clay content in mineral soil samples.

Soil property	n	Min	Mean	SD*	Median	Max	Skewness	Kurtosis	KS p-value**
Organic carbon (g kg^{-1})	59	0	17.76	25.38	11.00	162.41	3.90	17.84	0.002
Clay (%)	59	0	23.18	11.50	21.31	43.46	0.09	−0.80	0.19

* SD: Standard Deviation; **KS: Kolmogorov-Smirnov normality test.

testing normality of residuals (Shapiro-Wilk test) and homogeneity of variances (Levene's test). Statistical analysis was performed to determine whether there were significant differences in Ratio of Performance to InterQuartile range (RPIQ) values between groups, using a Tukey's Honestly Significant Difference (HSD) post hoc test when assumptions were met.

The evaluation included various error metric parameters, such as Coefficient of Determination (R-squared), RMSE, Ratio of Performance to Deviation (RPD), RPIQ, Bias, and Concordance Correlation Coefficient (CCC), to assess the prediction performance. Of these, RPIQ was utilized as an index to assess the accuracy of predictions and to establish a criterion for the acceptable variability in the spectral datasets (Eqs. (2) to (7)). For this purpose, the evaluation focused on the decrease in RPIQ as the variability of spectra escalated (implemented by increases in Z-scores to generate new spectral datasets).

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

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \tag{3}$$

$$RPD = \frac{SD}{RMSE} \tag{4}$$

$$RPIQ = \frac{IQ}{RMSE} \tag{5}$$

$$Bias = \sum_{i=1}^{n} \frac{y_i - \hat{y}_i}{n} \tag{6}$$

$$CCC = \rho . \frac{2}{\frac{\sigma_o}{\sigma_p} + \frac{\sigma_p}{\sigma_o} + \left(\frac{\mu_o - \mu_p}{\sigma_o \sigma_p}\right)^2} \tag{7}$$

where $\hat{y}$ = predicted value, $\bar{y}$ = mean observed value, y = observed values, n = number of samples with i = 1, 2, …, n, SD the Standard Deviation, IQ the InterQuartile range (quartiles 0.25 and 0.75) of the observed values, μ_o and μ_p are the means of the observed and predicted values, and σ_o and σ_p are the standard deviations of the observed and predicted values, respectively.

2.6. Defined criteria

The IEEE P4005 protocol (Ben-Dor et al., 2024b) suggests using the 1700 nm wavelength to assess the quality of spectral variations between measurements. This is because this wavelength does not correspond to any active chromophore in soils and is influenced solely by the real part of the complex refractive index, which, in dry soils, is affected by the orientation of the radiation relative to the soil surface structure. Accordingly, the average reflectance at 1700 nm for all samples was determined, then differences in reflectance were computed between the mean spectral dataset and each of the twenty supplementary spectral datasets. This yielded a single value representing the delta mean reflectance at 1700 nm for each spectral dataset (presented as Z-scores). The objective was to quantify the accuracy loss of the prediction models in proportion to the increase in delta reflectance at 1700 nm in the ring trial samples by examining these reflectance values alongside the percentage of decrease in RPIQ.

Considering the consistent scanning of the ISS interspersed among the ring trial samples, we hypothesized that the variation of reflectance at 1700 nm in the ISS would mirror the variation observed in ring trial samples. The relation between the mean reflectance at 1700 nm of ring trial samples and the ISS (based on their differences from the Foss XDS reflectance) was used to check this hypothesis (Fig. S2). Consequently, the delta mean reflectance at 1700 nm among the ring trial samples was substituted with the delta mean reflectance at 1700 nm of the ISS, aiming to establish a consistent and reproducible standard to assess variations of spectra. With this objective in mind, we proceeded to quantify the relationship between the percentage of decrease in RPIQ and the changes in delta mean reflectance at 1700 nm within the ISS. This method has the potential to be convenient and practical for on-the-fly spectrum measurements. The methodological steps are shown in a flowchart in Fig. 1 and schematic representation of the analyses carried out are shown in Fig. S7.

All analyses were carried out using the caret package (v. 6.0–94) in RStudio (2024.09.0 + 375). Data normalization, transformation, and the modeling pipeline were implemented in the R software (v. 4.2.3) using multiple packages, namely, pls (v. 2.8–3), BestNormalize (v. 1.9.1), and caret (v. 6.0–94).

3. Results

3.1. Spectral variability (dissimilarity)

Both the baseline alignment and the spectral shape consistency of sample 9 with high SOC content and sample 55 with high clay content, as extreme values among ring trial samples, show notable improvements after correction (Fig. 2). In the spectral plot, before correction, reflectance values varied considerably among instruments. For all ring trial samples, the means of reflectance across all instruments and wavelengths (450–2350 nm) range between 0.1 and 0.3, indicating consistent variability in reflectance measurements from different instruments. These values were reduced to 0.06 to 0.24 after correction, and a similar trend was observed in SD of reflectance across all instruments and wavelengths (450–2350 nm) per sample (Fig. S3 and S4). This improves the match between the spectra of all soils measured using different instruments, indicating that the correction effectively harmonized the data. Moreover, the average decrease in SD of reflectance values across all instruments per wavelength (450–2350 nm) ranged between 13.68 and 69.28% (median of 44.23%) for 60 samples of the ring trial. The average decrease in SD of reflectance values was 31.29% for sample 9 and 20.12% for sample 55 (data are not shown).

Additionally, spectral shapes have become more consistent post-correction. This is especially apparent in the regions between 900 and 1400 nm, with more than 45% average decrease in SD across all samples, and between 1800 and 1900 nm and between 2050 and 2250 nm, with more than 42% decrease (Fig. S5 and S6). The correction decreased discrepancies, leading to more uniform spectral patterns among the different instruments (Fig. 2). Normality of the paired differences was assessed using Shapiro-Wilk tests. For the seven instruments for which the paired differences were normally distributed, paired t-tests were applied and mean paired differences are reported. For the four instruments showing non-normal paired differences, Wilcoxon signed-rank tests were applied instead and median paired differences are reported. All tests showed significant differences before and after correction for all instruments, providing evidence of consistent changes in the spectra after ISS correction (Table S1).

3.2. PCA

The PCA revealed differences in the clustering and spread of spectral data points between non-corrected and corrected spectra across various instruments (Fig. 3). In the non-corrected spectra, data points exhibited a broader dispersion along both PC1 and PC2, indicating substantial variability likely due to instrumental differences and measurement conditions. Following the correction, the data points were more closely clustered and aligned, particularly along PC1. The paired *t*-test analysis between average distances of PCs to the centroid was conducted. The assumption of normality for differences was satisfied (Shapiro-Wilk W = 0.99, p = 0.90). The test revealed a p-value less than 2.2×10^{-16} demonstrating substantial improvement in spectral clustering.

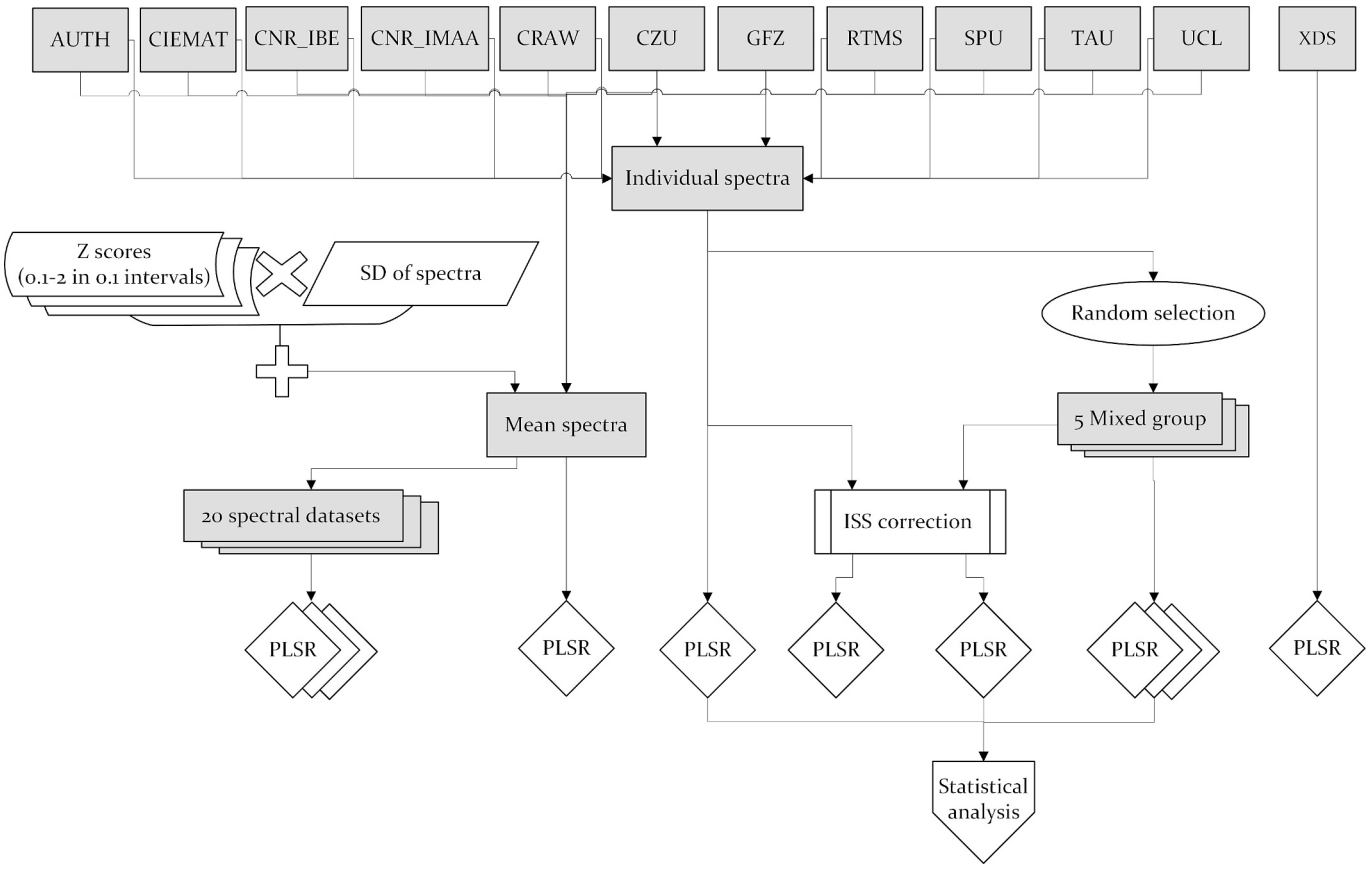


Fig. 1. Flowchart of the methods and approaches used in this study. Acronyms of institutes are same as Table 1 and PLSR: Partial Least Square Regression; SD: Standard Deviation; ISS: Internal Soil Standard.

Moreover, a 7.44% decrease in average Euclidean distance to the PCA centroid after correction suggests that the correction was both statistically and practically effective (Table S2). It reduced inter-instrument variability, thereby enhancing the comparability of the spectral data (Fig. 3).

3.3. Correlation across the spectra

Despite the variability observed, the Spearman rank correlations between each instrument and the reference instrument remain relatively high across all wavelengths. The "CRAW" exhibits consistently high rank correlations near 1.0 across almost the whole wavelength range. Three instruments, including "AUTH", "SPU", and "GFZ", show more noticeable declines in correlation, particularly around 700 nm. For the majority of the instruments, the spectral correction generally does not change the rank correlations. However, for one instrument, "GFZ", the correction resulted in a slight decline in correlation for all spectra. This indicates that the correction may have introduced some distortions or ineffectively addressed the pre-existing issue.

3.4. Model performance

Using PLSR on non-corrected spectra from the Foss XDS instrument as reference, SOC demonstrated higher internal cross-validation performance with an RPIQ of 1.57 and Lin's CCC of 0.86, while clay content presented a higher RPIQ of 1.99 but a lower CCC of 0.68 (Fig. 5). Also, there is a slight negative bias, suggesting that the model tends to underestimate SOC levels, while for clay, there is a near-zero bias. The results were used to evaluate the effect of averaging ring trial spectra and correction on prediction accuracy, allowing assessment of the predictive capability and determination of the extent to which a single reference instrument could provide accurate results.

The prediction accuracy of the models developed using the mean spectra from all instruments before correction (henceforth referred to as the base models) showed only minor differences when compared to the models developed using spectra from the Foss XDS instrument (Table 3). The RPIQ for SOC showed a slight decrease of 4.5%, from the FOSS instrument's value decreasing to 1.5. Conversely, the RPIQ for clay increased by 3%, reaching 2.05. For both soil properties, the bias remained almost unchanged. This finding aligns with the high correlation (greater than 0.9) between individual spectral datasets and the reference spectral dataset. As expected, when the base model was run using mean spectra + Z-scores multiplied by the SD, the prediction accuracy declined for both SOC and clay (Table 3). Regarding SOC, the RPIQ of the base model decreased by 13.2% for the mean + 1 SD and by 27.2% for the mean + 2 SD). For clay, the RPIQ of the base model dropped by 3.6% for the mean + 1 SD and by 10.6% for the mean + 2 SD. Considering that the bias of the base clay model is zero, the variability in spectra led to a significant bias for the mean + 1 SD (−1.25) and mean + 2 SD (−2.5) (Table 3).

3.5. Effect of ISS correction in merging multi-instrument spectral data

The prediction accuracies of each individual instrument were compared with that of the five randomly mixed groups using ANOVA followed by Tukey HSD test, both before and after ISS correction (Fig. 5). ANOVA assumptions were satisfied (Shapiro-Wilk test for normality: $p = 0.98$; Levene's test for homogeneity of variances: $p = 0.33$). The results showed that individual instrument models (RPIQ = 1.59) significantly outperformed the mixed, non-corrected groups (RPIQ = 1.31) but were not statistically different from the mixed, corrected groups (RPIQ = 1.39). Moreover, the difference between the RPIQ of mixed non-

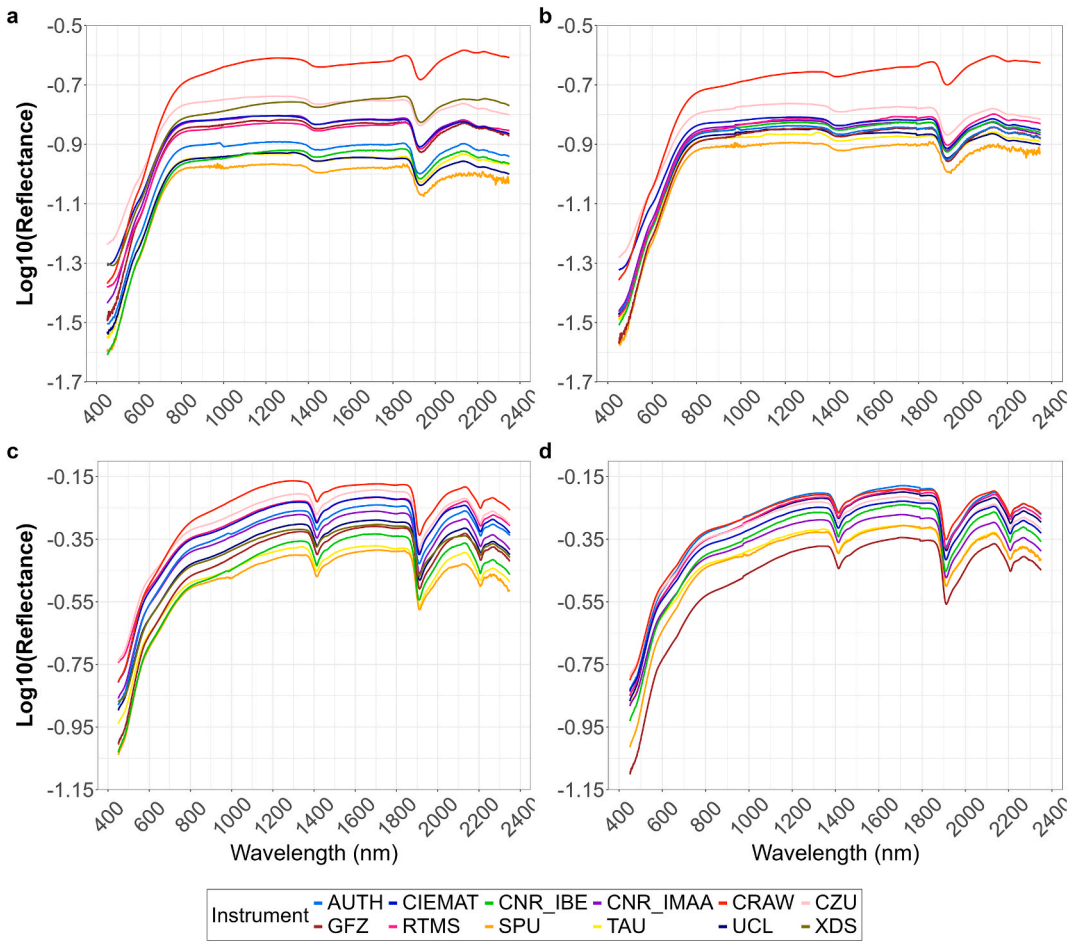


Fig. 2. Spectral variation between instruments for a) sample 9 with high SOC content (162.4 g kg^{-1}) before correction and b) after correction, c) sample 55 with high clay content (43.46%) before correction, and d) after correction. See the measurements and instruments section for the abbreviations of the spectrometers.

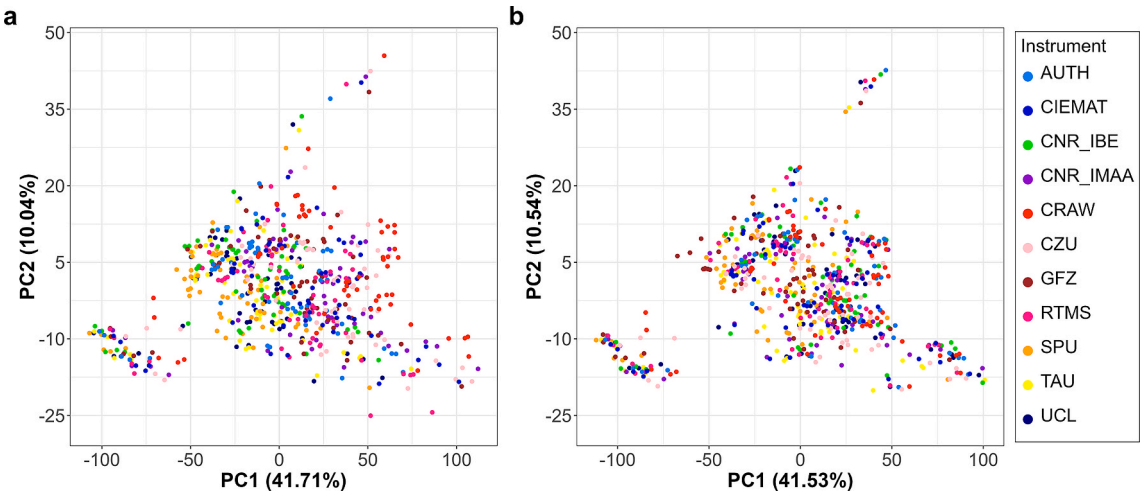


Fig. 3. Ring trial samples from 11 different instruments plotted on the Principal Component (PC) space of spectra a) before and b) after spectral correction.

corrected groups and of corrected groups was not significant, even though the corrected group showed a higher range.

3.6. Correlation between delta mean reflectance variation and RPIQ decrease

As the Z-scores increase, both the median and variability of delta reflectance at 1700 nm increases, indicating a positive correlation, with greater spread reflected by larger InterQuartile Ranges (IQR) (Fig. S7).

Fig. 6 (a and b) illustrates the relationship between the delta mean reflectance at 1700 nm and the percentage decrease in RPIQ (relative to the base model) for the SOC and clay models. The loss in RPIQ is positively correlated with an increase in the delta mean reflectance at 1700 nm, which in turn corresponds to Z-scores ranging from 0 to 2 in 0.1

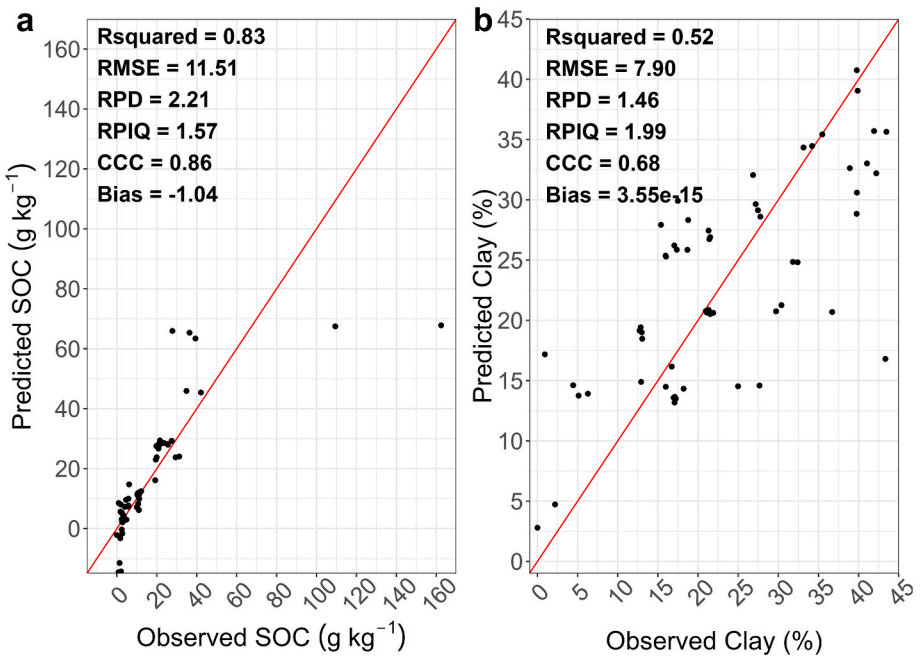


Fig. 4. Partial Least Squares Regression (PLSR) model performance using Foss XDS reference instrument spectra (n = 59 mineral soil samples). a) Soil Organic Carbon (SOC) and b) clay. R-squared: Coefficient of Determination; RMSE: Root Mean Square Error; RPD: Ratio of Performance to Deviation; RPIQ: Ratio of Performance to InterQuartile range; CCC: Concordance Correlation Coefficient.

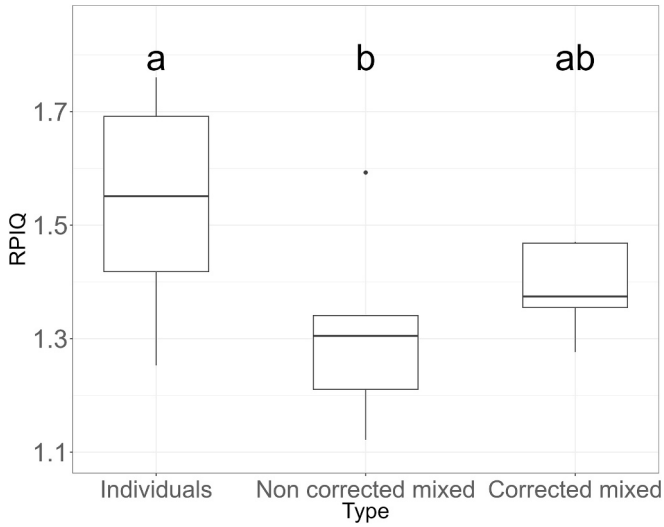


Fig. 5. Comparison of Ratio of Performance to InterQuartile range (RPIQ) for Soil Organic Carbon (SOC) prediction models built on a) individual instrument datasets (n = 11), b) five randomly mixed multi-instrument groups without ISS correction, and c) the same five mixed groups with ISS correction. Letters above boxes indicate significant differences. Groups sharing at least one letter are not significantly different, whereas groups with different letters differ significantly according to Tukey post-hoc test (α = 0.05).

increments. The decrease in RPIQ shows a linear trend for SOC and a mildly exponential trend for clay. Generally, the maximum decrease in RPIQ for a Z-score of 2 occurs at a delta mean reflectance of 0.145. However, the clay plot demonstrates a lower maximum decrease in RPIQ, about 11%, compared to 27% for SOC (Fig. 6, a and b).

3.7. Variability in the ISS reflectance at 1700 nm and prediction accuracy

We observed that the reflectance variance at 1700 nm in the ISS is correlated with that of the ring trial samples (R^2 = 0.81) since the ISS was consistently scanned among them. In fact, the delta mean reflectance at 1700 nm for ISS across different instruments was consistently higher than that of the ring trial samples (Fig. S2). To create a uniform and repeatable baseline for evaluating spectral variation, we replaced the delta mean reflectance at 1700 nm of ring trial samples with those of the ISS. To this end, we plotted the delta mean reflectance at 1700 nm of ISS against the percentage decrease in RPIQ for both the SOC and clay models (Fig. 6, c and d). For the same delta reflectance at 1700 nm in the ISS, the decrease of RPIQ percentage for clay was less than for SOC. For example, at a delta reflectance of 0.1, the RPIQ decreased by 13% for SOC, and a 3.6% decrease was observed for clay.

Meanwhile, by running the base model with spectral data collected from each participant's instrument, both before and after corrections, and calculating the decrease in RPIQ percentage compared to the base model's RPIQ (Fig. 6), we can understand the changes between the expected and real loss of accuracy. To validate the QA/QC framework established in Fig. 6, we applied the base model to individual instrument datasets and compared observed accuracy loss to predicted values (Fig. 7). We observed the decrease in RPIQ (%) of the base model for SOC, using both non-corrected and corrected spectra for each instrument (Fig. 7). When we compare this plot with the previous one (Fig. 6, c and d), it becomes evident that the results do not conform to the

Table 3
PLSR model performances for SOC and clay using mean spectra, mean spectra + 1 SD, and mean spectra + 2 SD.

Soil property	Type of spectra	RMSE	R^2	RPD	RPIQ	CCC	Bias
SOC (g kg^{-1})	Mean spectra	12.04	0.81	2.11	1.50	0.85	−1.01
	Mean spectra + SD	13.87	0.84	1.83	1.30	0.79	−6.04
	Mean spectra + 2SD	16.54	0.85	1.53	1.09	0.70	−9.49
Clay (%)	Mean spectra	7.69	0.55	1.5	2.05	0.71	0
	Mean spectra + SD	7.98	0.53	1.44	1.97	0.70	−1.25
	Mean spectra + 2SD	8.61	0.49	1.34	1.83	0.70	−2.5

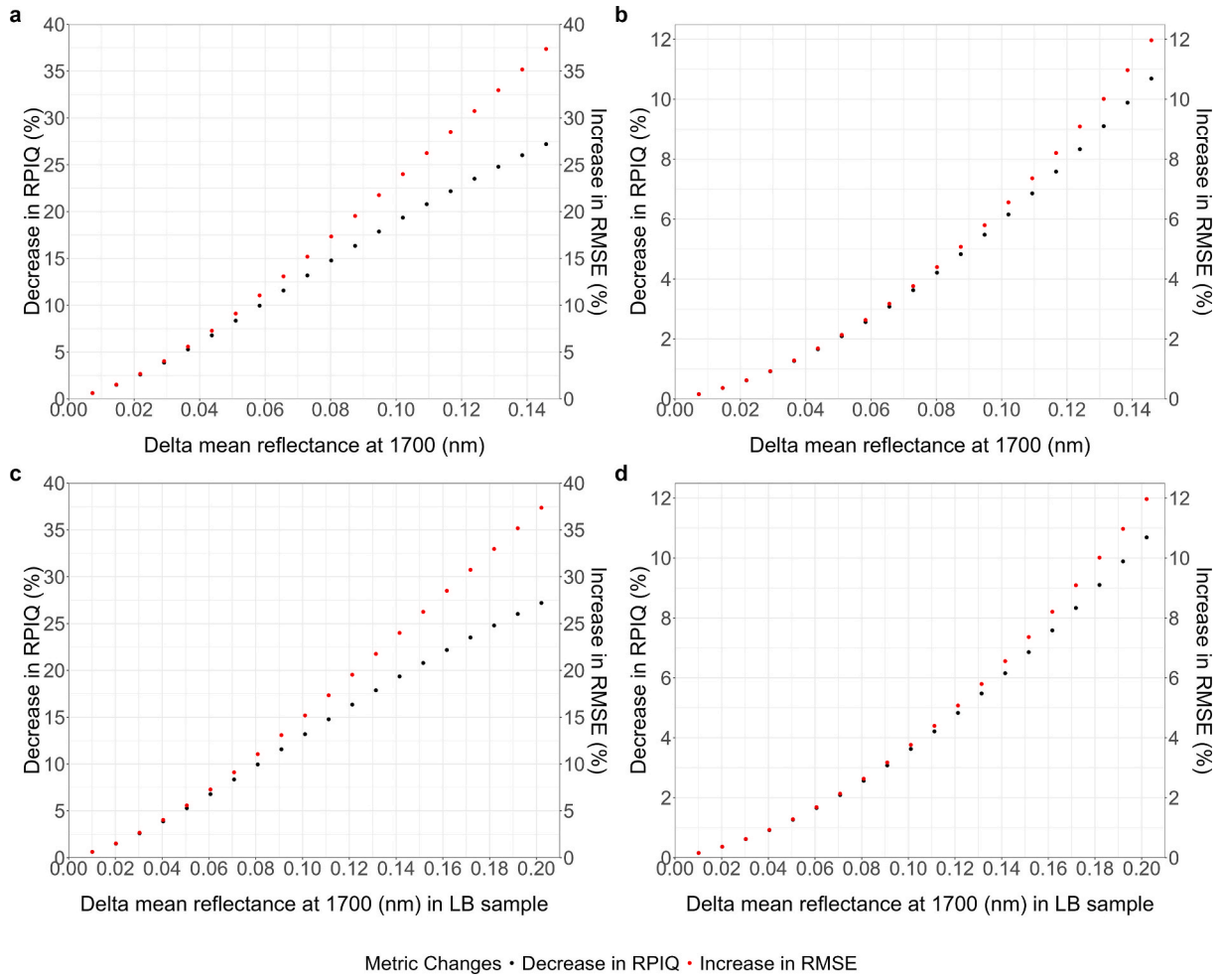


Fig. 6. Quality Assurance/Quality Control (QA/QC) criteria relating spectral variability to model performance loss. Relationship between delta mean reflectance at 1700 nm and prediction accuracy degradation for a, c) Soil Organic Carbon (SOC) and b, d) clay content. Top panels (a and b) use ring trial samples. Bottom panels (c and d) provide practical QA/QC thresholds using Internal Soil Standard (ISS) Lucky Bay (LB) measurements, which correlate with ring trial variability ($R^2 = 0.81$). Black dots denote decreases in Ratio of Performance to InterQuartile Range (RPIQ), and red dots denote increases in Root Mean Square Error (RMSE).

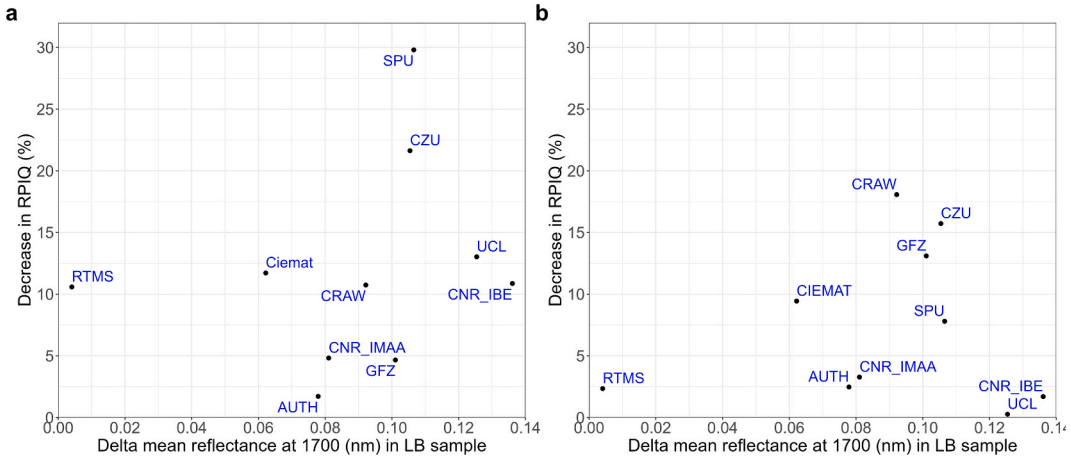


Fig. 7. Comparison of Internal Soil Standard (ISS) reflectance variability (absolute delta mean reflectance at 1700 nm in Lucky Bay (LB) samples) to observed prediction accuracy loss relative to the base model for Soil Organic Carbon (SOC), using a) non-corrected and b) ISS-corrected spectra. Each point represents one of the 11 participating instruments (acronyms in Table 1).

expected patterns predicted by the controlled variability hypothesis. Interestingly, an increase in delta mean reflectance at 1700 nm does not necessarily correspond to a larger decrease in RPIQ. However, the corrections applied to the spectral datasets have resulted in a more harmonized grouping of different instruments and an improved RPIQ. For "CRAW" and "GFZ"'s corrected spectra, there was a noticeable

decline in the RPIQ of the base model, which can be related to different measurement setups used by these two institutes. These results agree with the correlation analysis of spectra with the reference instrument.

3.8. Practical use of the ISS reflectance at 1700 nm for QA/QC

The reflectance measurements from 400 to 2400 nm of the mean ISS spectra of individual instruments and the mean ISS spectra of all instruments are shown in Fig. 8. The CSIRO standard spectra (ISS), which have been used as the reference spectra in several earlier publications, are also included in this figure (Ben-Dor et al., 2015; Gholizadeh et al., 2021; Safanelli et al., 2023).

Reflectance measurements demonstrate the distinct spectral profiles characteristic of each instrument. Notably, "CRAW" exhibits a different shape and higher reflectance values, particularly between 850 and 1450 nm, likely due to the different operational settings, such as the use of quartz-bottom sample cells.

The benchmark CSIRO line, while somewhat aligned with the tested instruments, shows more similarity to those with higher reflectance values. In contrast, instruments such as "AUTH", "SPU", "UCL", "CNR_IBE", and Foss XDS demonstrate much lower reflectance across the spectra. This may be attributed to differences in sensor sensitivities related to instrument type (Table 1) or to sample handling methods, for example, using quartz-bottom sample cells in the case of Foss XDS. Also, variation from LB CSIRO spectra might be related to variation in the white reference characteristics (material and aging) (Futerman et al., 2025).

These variations underscore the importance of standardizing instrument responses to ensure comparability of spectral data across different platforms, which is essential for accurate spectral analysis and application. The plot, zoomed in around 1700 nm (Fig. 8), explicitly shows distinct reflectance values among the instruments, within a range of 0.3. Evaluating the reflectance variability at 1700 nm, depending on the chosen reference measurement (CSIRO or Foss XDS), can help determine whether correction is necessary.

4. Discussion

This study provides insights into the challenges and solutions related to spectral variability among different spectrometers. The objective was to determine when spectral correction, between laboratory measurement and between laboratories, is necessary and to assess the potential loss in prediction accuracy if no correction is applied. To ensure applicability, the QA/QC approach was designed to be both practical and suitable for real-time implementation across various laboratory environments. Consequently, the proposed method is intentionally simple to promote ease of use and consistency in routine soil spectral analysis.

4.1. Spectral variability: insights from PCA and correlation analyses

The effectiveness of the ISS correction in reducing instrument differences is evidenced by the improved baseline alignment, the reduced range and SD of mean reflectance, and the enhanced consistency in spectral shapes after correction (Fig. 2, S3, and S4). Additionally, the decrease in average SD of reflectance across instruments further confirms that the correction process effectively harmonized inter-instrument variability. The statistical validation using paired t-tests reinforces the role of correction in reducing systematic errors in SSLs. Ge et al. (2011) also noted that following their mathematical adjustments, the variations in soil spectra decreased. Comparably, Ben-Dor et al. (2015) and Pimstein et al. (2011) claimed that while the ISS technique significantly reduces the discrepancies between sensors and geometries, it does not change the shape of spectral signatures. Similar to Kopačková and Ben-Dor (2016), correction using ISS brings the reflectance of soil samples closer to that of the reference instrument (Foss XDS) and groups the spectra of the same soil within different instruments to a closer agreement. Less pronounced results from the ISS correction are shown in sample 55 (Fig. 2), which can be related to the soil's clay content. Romero et al. (2018) used ISS on soils with three distinct clay classes and produced comparable results. In their investigation, the samples with higher sand content also showed less spectral alignment. At the "TAU" and "GFZ" laboratories with different conditions ("GFZ" wet and dry data and "TAU" data), the spectral shape and absorption properties of soil samples followed a similar curve both before and after the ISS correction. However, there was still a discrepancy in the reflectance and albedo intensity values (Chabrillat et al., 2019). In order to ensure that spectral data from various instruments may be combined into greater datasets without significantly adding biases or losing accuracy, the correction is essential and can only be achieved using better standardization procedures.

The PCA results provide additional evidence of the correction's effectiveness. The outcome of the paired *t*-test on Euclidean distance to the PCs centroid and the reduction in average Euclidean distance highlight not only statistical significance in spectral clustering before and after correction but also practical effectiveness, reflecting improved structure of the spectral dataset. The observed improvements in data clustering in PCA space suggest a considerable reduction in measurement biases and inter-instrument variability, while the variance between spectral datasets is maintained (Fig. 3). This is in accordance with the work of Chabrillat et al. (2019), where better clustering and less bias were observed in PCA analysis after correction using ISS, and three different measurement conditions ("GFZ" wet data, "GFZ" dry data, and

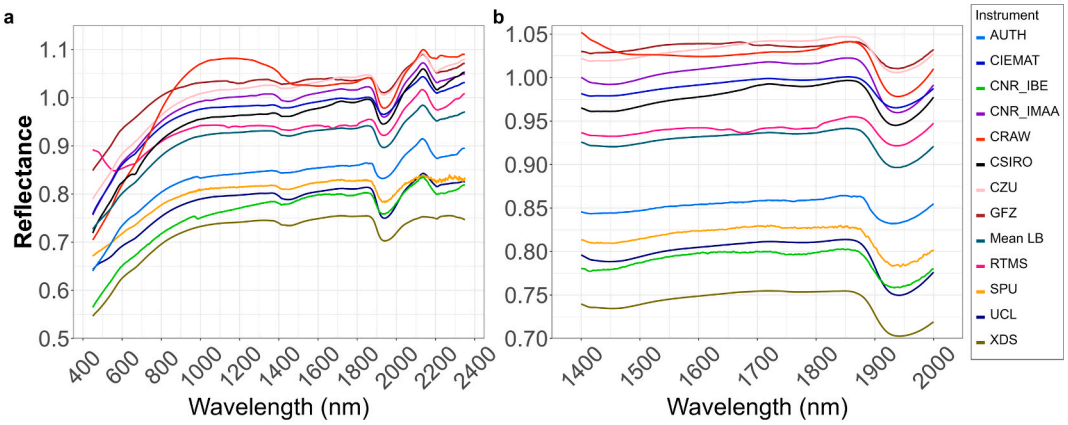


Fig. 8. Inter-instrument spectral variability demonstrated by Internal Soil Standard (ISS) Lucky Bay (LB) sand measurements. a) Full spectral range (400–2400 nm) showing mean ISS reflectance from each of the 11 participating instruments plus the CSIRO benchmark reference (acronyms in Table 1). b) Detailed view of the 1400–2000 nm region highlighting reflectance at 1700 nm.

"TAU" data) were more aligned. They reported that before corrections, different laboratory setups and conditions introduced additional sources of variance that were difficult to differentiate from soil type variation on spectral data.

While all instruments had generally high correlations, the slight decrease in correlation after correction in some instruments suggests that the procedure may not have uniformly addressed all pre-existing issues such as different measurement setups, instrument conditions or sample preparation methods. This was observed by Bartsch et al. (2024) as well. Of 30 soil properties, 21 increased performance using the protocol, and 9 showed slight decreases. This outcome highlights potential limitations of the correction method, particularly when applied to specific measurement configurations. This emphasizes the importance of ongoing assessment and potential improvement of correction methods to guarantee their efficacy with a range of instruments and measurements. Standardization should also account for consistent sample preparation, instrument setup, calibration, and maintenance procedures to ensure reliable and comparable spectra. Nonetheless, the use of ISS, although not always perfectly harmonized with the data sets, provides a marker to assess the stability of individual laboratories or protocols. In this context, ISS should be viewed as a complementary tool that enables monitoring of instrument performance and detection of systematic errors.

The result obtained from PCA and the Spearman rank correlation analysis suggest that the ISS method performs effectively for spectral correction across different instruments. However, the strict implementation of a single protocol, even without ISS correction, results in a very similar reflectance in the same samples measured by different instruments.

4.2. Model performance

A PLSR model was used as a basic chemometric approach using the mean spectral data. As anticipated, the mean spectra from all individual instruments generally offer acceptable prediction accuracy comparable to that of the FOSS XDS instrument, as indicated by similar error metrics. The controlled environment, fixed light sources, and reliability of the probe measurements are unique advantages of Foss XDS, which reduce the source of error and might be a reason for a slightly better result (Gholizadeh et al., 2021) (Figs. 4 and 5).

The similar prediction accuracy of mean spectra to the FOSS XDS instrument (Fig. 4 and Table 3) suggests that when similar samples are measured using the same standard procedure and protocol, and the spectrometer and lamp are stable, the systematic effects are minimal, resulting in relatively good predictions. This remains valid even when spectral responses for the same soil sample differ due to the unique characteristics of each instrument (Fig. 2). The minimal differences in prediction accuracy between individual instrument models, before and after correction, as well as their similarity to the Foss XDS model performance, further confirm this (Tables S3 and S4). This can be explained by the fact that a standard measurement protocol would result in smaller systematic errors and reproducible prediction performance (Francos et al., 2022; Romero et al., 2018). It may also be related to the smoothing effect of averaging, which in turn reduces the variability between the spectra of different instruments.

Regardless of the variation, clay was the more constant soil attribute with the least amount of performance variation when the spectra varied (Table 3 and Fig. 6, a and b). This is likely because absorption features related to clay are associated with specific and relatively stable spectral regions, whereas SOC influences broader and more variable spectral features. As a result, SOC was more sensitive to spectral variability. Chromophores associated with SOC vary more due to natural heterogeneity, leading to greater uncertainty in spectral data calibration. These overall performance results have a resemblance to those in prior studies that assessed the spectral range for soil property monitoring (Francos et al., 2022; Gholizadeh et al., 2021; Safanelli et al., 2023).

4.3. Effect of ISS correction in merging multi-instrument spectral data

Since building models within each laboratory's dataset may not reveal substantial differences due to the uniform measurement protocol, five models were developed using mixed datasets incorporating samples from multiple laboratories. This approach enabled an assessment of how the correction influences model accuracies when applied to heterogeneous, inter-laboratory datasets. The accuracy of predictions generated using merged SSLs may be impacted by variability in the reflectance measurements from different instruments, which could result in inconsistent estimations. As expected, models built from individual instrument datasets performed very similarly (regardless of spectral variability, Tables S3 and S4) and significantly better than those using mixed, non-corrected data (Fig. 5, Table S5). This confirms that systematic variability between instruments reduces model accuracy when combining spectral data. However, no significant difference was observed between the individual models and the mixed, corrected models, suggesting that the ISS correction effectively reduced inter-instrument variability and improved model accuracy to a level comparable with individual instrument performance. This supports the use of ISS when combining spectral data from multiple datasets for building SSLs or when a systematic error is assumed to occur amongst otherwise stable instruments.

4.4. Applying reflectance variation at 1700 nm as a QA/QC in laboratory soil spectroscopy

The impact of spectral correction on data stability is further demonstrated by the analysis of reflectance variation at 1700 nm (Figs. 1 to 3). Given the absence of a strong soil absorbance feature at this wavelength, the IEEE P4005 protocol (Ben-Dor et al., 2024b) recommends using reflectance at 1700 nm to quickly determine variations in the albedo. As shown in Figs. 2 (a to d) and 10, this area does not have a significant absorbance feature in most natural soil samples, consistent with previous studies (Ben-Dor et al., 2024b, 2015; Gholizadeh et al., 2021). The relationship between the percentage drop in RPIQ and the delta mean reflectance shows how spectral variability affects model performance. The smaller decline in RPIQ for clay indicates that prediction of clay content is less sensitive to the negative effects of spectral variability (Figs. 7 and 8). This difference in stability between attributes could explain why models for clay content tend to be more robust and reproducible, whereas SOC models might suffer from higher uncertainty. This also suggests that ISS correction, or any standardization technique, may be more critical for SOC data than for clay.

The lower delta mean reflectance at 1700 nm for the ring trial compared to ISS samples indicates that using ISS reflectance at this wavelength provides a cautious basis for comparison (Fig. S2). This minimizes the potential loss of accuracy, supporting the practical use of ISS reflectance at 1700 nm as a QA/QC indicator.

By accounting for the variability at 1700 nm reflectance and its impact on prediction accuracy (Fig. 6), this approach establishes a structured framework for assessing whether spectral measurements need correction or replication. Fig. 6 (c and d) allows users to check the mean reflectance at 1700 nm of ISS in their measurement setup and compare it to a reference measurement (ISS-CSIRO or Foss XDS) (Fig. 8). Users can then determine whether the potential percentage decrease in the model accuracy is acceptable or if the measurements need to be repeated, based on their objectives. This is a noteworthy accomplishment since it offers a practical and real-time application for assessment and reduction of spectral variability. It enables consistent soil spectra measurement across multiple laboratories with different devices, thereby improving the overall quality of the predictions, especially for building SSLs or incorporating new datasets (Fig. 8).

When the delta mean reflectance at 1700 nm of ISS replications remains constant, the comparison between the expected and actual loss in accuracy across instruments suggests that these observed differences are

not solely due to the systematic measurement effects and cannot be corrected using the ISS reference spectra (Fig. 6). Nonetheless, the average increase in RPIQ of running the base model on instruments with the same measurement setup after ISS correction (i.e., excluding "CRAW" and "GFZ" data, with negative effect) was 46% for the SOC model. This highlights the importance of spectral correction for reducing variability across soil spectrometers and improving the accuracy of spectral data.

5. Conclusions

This study focuses on the use of an internal standard (ISS) to harmonize soil spectral data and reduce spectral variability across different instruments and laboratories. The findings indicate that ISS correction reduces discrepancies and improves the consistency of spectral data. While some limitations about measurement setups remain, the correction helps data alignment and prediction accuracy, particularly when combining spectra from multiple instruments. The results also highlight that the need for correction may vary depending on the property being predicted, and it should be considered prior to merging datasets and developing models.

Modeling results further confirmed that prediction accuracy is comparable across instruments, when standard protocols are strictly followed even without correction. However, ISS correction shows additional benefit when merging multi-instrument datasets in the context of large-scale SSLs development. The ISS correction did lead to a more consistent grouping of instruments, improving prediction accuracy for most, but accuracy loss cannot be entirely attributed to systematic measurement effects or be fully corrected using the ISS. Furthermore, the reflectance at 1700 nm of the ISS, devoid of major soil absorbance features, can serve as a criterion for a fast and simple quality check of inter-instrument variation and the variability of a single instrument over time. This approach offers a practical and real-time framework for assessing the need for spectral data correction or replication, ensuring the stability of soil spectral data. Overall, the findings advocate for the broader adoption of standard protocols and ISS-based correction in soil spectroscopy to ensure consistency and reduce variability across devices and laboratory conditions, thereby improving the accuracy of soil property predictions.

CRediT authorship contribution statement

Marmar Sabetizade: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Bas Van Wesemael:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization. **Vincent Baeten:** Writing – review & editing, Resources. **Fabio Castaldi:** Writing – review & editing, Resources. **Sabine Chabrillat:** Writing – review & editing, Resources. **José A.M. Demattê:** Writing – review & editing, Resources. **Asa Gholizadeh:** Writing – review & editing, Resources. **Stephan Haefele:** Writing – review & editing, Resources. **Konstantinos Karyotis:** Resources, Project administration, Investigation, Data curation. **Aleš Klement:** Investigation. **Robert Milewski:** Writing – review & editing, Investigation. **Simone Pascucci:** Writing – review & editing, Resources. **Louis Paternostre:** Writing – review & editing, Resources. **Stefano Pignatti:** Writing – review & editing, Resources. **Jose Lucas Safanelli:** Writing – review & editing, Validation, Software, Methodology. **Jonathan Sanderman:** Writing – review & editing, Resources, Project administration, Methodology, Funding acquisition. **Thomas Schmid:** Writing – review & editing, Resources. **Nikolaos Tziolas:** Writing – review & editing, Resources. **Eyal Ben-Dor:** Writing – review & editing, Resources, Project administration, Methodology, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Funding for this work came from USDA NIFA Award 2020-67021-32467 (Soil Spectroscopy for the Global Good). JLS and JS were partially supported by the Lawrence Livermore National Laboratory (LLNL) LDRD Program (24-SI-002). The authors would like to thank the staff at the NSSC-KSSL for facilitating this research and Matthew Jacques for preparing and shipping samples to all participants.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.geoderma.2026.117894.

Data availability

Data will be made available on request.

References

Bartsch, B.d.A., Rosin, N.A., dos Santos, U.J., Coblinski, J.A., Pelegrino, M.H.P., Rosas, J. T.F., Poppiel, R.R., Ortiz, E.B., Kochinki, V.C.V., Gallo, P., Ben-Dor, E., Falcioni, R., Nanni, M.R., Gonçalves, J.V.F., Oliveira, C.A.d., Vedana, N.G.V., Furlanetto, R.H., Demattê, J.A.M., 2024. A step forward in hybrid soil laboratory analysis: merging chemometric corrections, protocols and data-driven methods. Remote Sens. 16. https://doi.org/10.3390/rs16234543.

Ben-Dor, E., Chabrillat, S., Karyotis, K., Schmid, T., 2024a. P4005 standards and protocols for soil spectroscopy in laboratory and field conditions [Technical committees]. IEEE Geosci. Remote Sens. Mag. 12, 358–363. https://doi.org/10.1109/MGRS.2024.3469069.

Ben-Dor, E., Efrati, B., Amir, O., Francos, N., Shepherd, J., Khosravi, V., Gholizadeh, A., Klement, A., Borůvka, L., 2024b. A standard and protocol for in-situ measurement of surface soil reflectance. Geoderma 447. https://doi.org/10.1016/j.geoderma.2024.116920.

Ben-Dor, E., Ong, C., Lau, I.C., 2015. Reflectance measurements of soils in the laboratory: Standards and protocols. Geoderma 245–246, 112–124. https://doi.org/10.1016/j.geoderma.2015.01.002.

Castaldi, F., Chabrillat, S., Chartin, C., Genot, V., Jones, A.R., van Wesemael, B., 2018. Estimation of soil organic carbon in arable soil in Belgium and Luxembourg with the LUCAS topsoil database. Eur. J. Soil Sci. 69, 592–603. https://doi.org/10.1111/ejss.12553.

Chabrillat, S., Gholizadeh, A., Neumann, C., Berger, D., Milewski, R., Ogen, Y., Ben-Dor, E., 2019. Preparing a soil spectral library using the Internal Soil Standard (ISS) method: Influence of extreme different humidity laboratory conditions. Geoderma 355. https://doi.org/10.1016/j.geoderma.2019.07.013.

Dangal, S., Sanderman, J., Wills, S., Ramirez-Lopez, L., 2019. Accurate and Precise Prediction of Soil Properties from a Large Mid-Infrared Spectral Library. Soil Syst. 3, 11. https://doi.org/10.3390/soilsystems3010011.

Demattê, J.A.M., Dotto, A.C., Bedin, L.G., Sayão, V.M., Souza, A.B.e., 2019. Soil analytical quality control by traditional and spectroscopy techniques: Constructing the future of a hybrid laboratory for low environmental impact. Geoderma 337, 111–121. https://doi.org/10.1016/j.geoderma.2018.09.010.

Demattê, J.A.M., Minasny, B., Hartemink, A.E., 2025. Vital for sustainable agriculture: pedological knowledge and mapping. Eur. J. Soil Sci. 76, 1–14. https://doi.org/10.1111/ejss.70040.

FAO, 2023. Report of the Eleventh Plenary Assembly of the Global Soil Partnership, 12–14 July 2023: Global Soil Laboratory Network (GLOSOLAN) and its initiative on soil spectroscopy (GLOSOLAN-Spec). *Food and Agriculture* Organization of the United Nations, Rome. Available at: https://www.fao.org/3/cc6481en/cc6481en.pdf (Accessed 3 November 2025).

Francos, N., Gholizadeh, A., Demattê, J.A.M., Ben-Dor, E., 2022. Effect of the internal soil standard on the spectral assessment of clay content. Geoderma 420. https://doi.org/10.1016/j.geoderma.2022.115873.

Francos, N., Notesco, G., Ben-Dor, E., 2021. Estimation of the Relative Abundance of Quartz to Clay Minerals using the Visible–Near-Infrared–Shortwave-Infrared Spectral Region. Appl. Spectrosc. 75, 882–892. https://doi.org/10.1177/0003702821998302.

Futerman, S.I., Lugassi, R., Shepherd, J.E., Karyotis, K., Chabrillat, S., Schmid, T., Dematte, J., Ben-Dor, E., 2025. EXPRESS: Mitigating Spectral Errors from White Reference Degradation in Soil Spectroscopy. Appl. Spectrosc., 00037028251396975 https://doi.org/10.1177/00037028251396975.

Ge, Y., Morgan, C.L.S., Grunwald, S., Brown, D.J., Sarkhot, D.V., 2011. Comparison of soil reflectance spectra and calibration models obtained using multiple spectrometers. Geoderma 161, 202–211. https://doi.org/10.1016/j.geoderma.2010.12.020.

Gholizadeh, A., Carmon, N., Klement, A., Ben-Dor, E., Borůvka, L., 2017. Agricultural soil spectral response and properties assessment: Effects of measurement protocol and data mining technique. Remote Sens. 9, 1–14. https://doi.org/10.3390/rs9101078.

Gholizadeh, A., Neumann, C., Chabrillat, S., van Wesemael, B., Castaldi, F., Borůvka, L., Sanderman, J., Klement, A., Hohmann, C., 2021. Soil organic carbon estimation using VNIR–SWIR spectroscopy: the effect of multiple sensors and scanning conditions. Soil Tillage Res. 211, 1–24. https://doi.org/10.1016/j.still.2021.105017.

Karyotis, K., Chabrillat, S., Ben-Dor, E., 2023. P4005: the IEEE SA standard and protocol scheme for soil spectral measurement in both laboratory and field. Int. Geosci. Remote Sens. Symp. 2023-July, 2493–2495. https://doi.org/10.1109/IGARSS52108.2023.10282263.

Kopačková, V., Ben-Dor, E., 2016. Normalizing reflectance from different spectrometers and protocols with an internal soil standard. Int. J. Remote Sens. 37, 1276–1290. https://doi.org/10.1080/01431161.2016.1148291.

Mészáros, J., Kovács, Z., László, P., Vass-Meyndt, S., Koós, S., Pirkó, B., Szűcs-Vásárhelyi, N., Bakacsi, Z., Laborczi, A., Balog, K., Pásztor, L., 2025. Vis-NIR soil spectral library of the Hungarian Soil Degradation Observation System. Sci. Data 12, 1–9. https://doi.org/10.1038/s41597-025-04667-9.

Notesco, G., Weksler, S., Ben-Dor, E., 2019. Mineral classification of soils using hyperspectral longwave infrared (LWIR) Ground-based data. Remote Sens. 11. https://doi.org/10.3390/rs11121429.

Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A., Fernández-Ugalde, O., 2018. LUCAS Soil, the largest expandable soil dataset for Europe: a review. Eur. J. Soil Sci. 69, 140–153. https://doi.org/10.1111/ejss.12499.

Padarian, J., Minasny, B., McBratney, A.B., 2019. Transfer learning to localise a continental soil vis-NIR calibration model. Geoderma 340, 279–288. https://doi.org/10.1016/j.geoderma.2019.01.009.

Pimstein, A., Notesco, G., Ben-Dor, E., 2011. Performance of three Identical Spectrometers in Retrieving Soil Reflectance under Laboratory Conditions. Soil Sci. Soc. Am. J. 75, 746–759. https://doi.org/10.2136/sssaj2010.0174.

Poppiel, R.R., Paiva, A.F. da S., Demattê, J.A.M., 2022. Bridging the gap between soil spectroscopy and traditional laboratory: Insights for routine implementation. Geoderma 425. https://doi.org/10.1016/j.geoderma.2022.116029.

Ramirez-Lopez, L., Wadoux, A.M.J.C., Franceschini, M.H.D., Terra, F.S., Marques, K.P.P., Sayão, V.M., Demattê, J.A.M., 2019. Robust soil mapping at the farm scale with vis–NIR spectroscopy. Eur. J. Soil Sci. 70, 378–393. https://doi.org/10.1111/ejss.12752.

Romero, D.J., Ben-Dor, E., Demattê, J.A.M., Souza, A.B.e., Vicente, L.E., Tavares, T.R., Martello, M., Strabeli, T.F., da Silva Barros, P.P., Fiorio, P.R., Gallo, B.C., Sato, M.V., Eitelwein, M.T., 2018. Internal soil standard method for the Brazilian soil spectral library: Performance and proximate analysis. Geoderma 312, 95–103. https://doi.org/10.1016/j.geoderma.2017.09.014.

Safanelli, J.L., Hengl, T., Parente, L.L., Minarik, R., Bloom, D.E., Todd-Brown, K., Gholizadeh, A., de Sousa Mendes, W., Sanderman, J., 2025. Open Soil Spectral Library (OSSL): Building reproducible soil calibration models through open development and community engagement. PLoS One 20, 1–50. https://doi.org/10.1371/journal.pone.0296545.

Safanelli, J.L., Sanderman, J., Bloom, D., Todd-Brown, K., Parente, L.L., Hengl, T., Adam, S., Albinet, F., Ben-Dor, E., Boot, C.M., Bridson, J.H., Chabrillat, S., Deiss, L., Demattê, J.A.M., Scott Demyan, M., Dercon, G., Doetterl, S., van Egmond, F., Ferguson, R., Garrett, L.G., Haddix, M.L., Haefele, S.M., Heiling, M., Hernandez-Allica, J., Huang, J., Jastrow, J.D., Karyotis, K., Machmuller, M.B., Khesuoe, M., Margenot, A., Matamala, R., Miesel, J.R., Mouazen, A.M., Nagel, P., Patel, S., Qaswar, M., Ramakhanna, S., Resch, C., Robertson, J., Roudier, P., Sabetizade, M., Shabtai, I., Sherif, F., Sinha, N., Six, J., Summerauer, L., Thomas, C.L., Toloza, A., Tomczyk-Wójtowicz, B., Tsakiridis, N.L., van Wesemael, B., Woodings, F., Zalidis, G. C., Żelazny, W.R., 2023. An interlaboratory comparison of mid-infrared spectra acquisition: Instruments and procedures matter. Geoderma 440. https://doi.org/10.1016/j.geoderma.2023.116724.

Shepherd, K.D., Ferguson, R., Hoover, D., van Egmond, F., Sanderman, J., Ge, Y., 2022. A global soil spectral calibration library and estimation service. Soil Secur. 7, 100061. https://doi.org/10.1016/j.soisec.2022.100061.

Soil Survey Staff, 2022. Kellogg Soil Survey Laboratory Methods Manual. Soil Survey Investigations Report No. 42, Version 6.0 Part 1: Current Methods. U.S. Department of Agriculture, Natural Resources Conservation Service.

Stevens, A., Nocita, M., Tóth, G., Montanarella, L., van Wesemael, B., 2013. Prediction of Soil Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy. PLoS One 8. https://doi.org/10.1371/journal.pone.0066409.

Tóth, G., Jones, A., Montanarella, L., 2013. LUCAS Topsoil Survey: Methodology. Data, and Results, Publications Office of the European Union. https://doi.org/10.2788/97922.

Tziolas, N., Tsakiridis, N., Ben-Dor, E., Theocharis, J., Zalidis, G., 2019. A memory-based learning approach utilizing combined spectral sources and geographical proximity for improved VIS-NIR-SWIR soil properties estimation. Geoderma 340, 11–24. https://doi.org/10.1016/j.geoderma.2018.12.044.

Viscarra Rossel, R.A., Behrens, T., Ben-Dor, E., Brown, D.J., Demattê, J.A.M., Shepherd, K.D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aïchi, H., Barthès, B.G., Bartholomeus, H.M., Bayer, A.D., Bernoux, M., Böttcher, K., Brodský, L., Du, C.W., Chappell, A., Fouad, Y., Genot, V., Gomez, C., Grunwald, S., Gubler, A., Guerrero, C., Hedley, C.B., Knadel, M., Morrás, H.J.M., Nocita, M., Ramirez-Lopez, L., Roudier, P., Campos, E.M.R., Sanborn, P., Sellitto, V.M., Sudduth, K.A., Rawlins, B.G., Walter, C., Winowiecki, L.A., Hong, S.Y., Ji, W., 2016. A global spectral library to characterize the world's soil. Earth-Science Rev. 155, 198–230. https://doi.org/10.1016/j.earscirev.2016.01.012.

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., Shi, Z., Stenberg, B., Webster, R., Winowiecki, L., Shen, Z., 2022. Diffuse reflectance spectroscopy for estimating soil properties: a technology for the 21st century. Eur. J. Soil Sci. 73, 1–9. https://doi.org/10.1111/ejss.13271.

Viscarra Rossel, R.A., Bouma, J., 2016. Soil sensing: a new paradigm for agriculture. Agr. Syst. 148, 71–74. https://doi.org/10.1016/j.agsy.2016.07.001.

Willis, D.E., 1972. Internal standard method calculations. Chromatographia 5, 42–43. https://doi.org/10.1007/BF02270655.