scientific reports

Check for updates

OPEN

Long-term stability of soil spectral libraries with chemical and spectral insights

Jonti Evan Shepherd✉, Ori Kanner, Or Amir, Bar Efrati & Eyal Ben-Dor

Soil spectral libraries (SSLs) are physical soil samples that are stored under different conditions by many users for decades. Yet the long-term stability of soil properties under these stored conditions remains largely unexplored. This study investigates the chemical and spectral stability of the Israeli legacy SSL, established in 1987, stored under uncontrolled indoor conditions for 34–37 years. Ninety-one Mediterranean soils from this collection were reanalyzed for soil organic matter (SOM), calcium carbonate ($CaCO_3$), using identical protocols and spectroscopic methods in 1987 and 2024 (chemical) and 2004 and 2024 (spectral). Results demonstrate minimal changes in SOM and $CaCO_3$, supported by strong linear correlations between historical and contemporary datasets (R^2 of 0.925 and 0.962 for SOM and $CaCO_3$ respectively). Spectroscopic analysis showed superior precision and reliability compared to wet chemistry. Additionally, spectral stability over time was confirmed using the modified average spectral difference stability (mASDS) metric, Principal Component Analysis (PCA) and partial least squares regression (PLSR), underscoring the robustness of spectroscopic approaches. Spectral modeling of the chemical data from both years revealed outliers which we assume emerged from analytical accuracy differences and not from spectroscopy errors. This study highlights that Mediterranean soils stored under simple un-controlled conditions maintain their physical and chemical integrity, enabling reliable longitudinal studies. These findings advocate for broader SSL archiving efforts to support soil health monitoring, climate change studies, and sustainable land management practices by utilizing old collections of stored soils that can be measured spectrally to enrich SSLs worldwide. Future research should focus on other climatic regions and soil types to generalize these findings and address possible microbial activity impacts during storage. This work underscores SSLs as critical resources for soil science, offering insights into temporal soil dynamics and facilitating global soil monitoring efforts.

Soil Spectral Libraries (SSLs) are carefully curated collections of physical soil samples systematically gathered from the field, brought into the laboratory, and thoroughly analysed, recording chemical and physical properties as well as spectral measurements[1]. These samples are prepared for analysis by air drying, gently crushing, and sieving through a > 2 mm sieve to ensure consistent measurement conditions[2]. Then the soil undergoes chemical, physical and spectral measurements preferably under well-accepted standards and protocols (e.g., ISO 11277:2020)[3]. After the initial analysis, samples are stored in sealed compartments under uncontrolled room conditions, for future chemical or spectral evaluations, especially when new samples are added to update the SSL. Each sample's data includes geographical coordinates, soil chemical attributes such as SOM, and their respective reflectance measurements, along with any relevant metadata available. A critical issue for SSLs is whether prolonged storage impacts the stability of sample properties—chemically or spectrally—over time[4]. This is especially relevant as SSLs are often updated with new soil samples or new measurement methods that have been applied to re-utilize existing samples to monitor soil attributes across timeframes. Any changes in stored samples due to environmental factors (temperature, humidity, and light exposure) at the storage facility may require re-measurement to uphold data accuracy and consistency within the SSL[5]. Previous studies have examined storage effects on soils outside SSLs, primarily focusing on methods to limit microbial activity[6]. Certain soil properties, such as mineral content and chemical composition, have shown relative stability under controlled storage[7], while microbial activity was identified as the main factor that could alter soil characteristics[6]. The microbiome may influence soil properties, indirectly affecting attributes such as cation exchange capacity (CEC), pH, electrical conductivity (EC), base saturation, and aggregation, which play critical roles in soil functionality and overall health[5]. Storage protocols, such as ISO 10,381, recommend specific low-temperature,

The Remote Sensing Laboratory, Porter School of Environment and Earth Science, Faculty of Exact Science, Tel Aviv University, Tel Aviv, Israel. ✉email: jontis@mail.tau.ac.il

dry conditions for soils from extreme climates to preserve their physicochemical and biological characteristics[8]. Similarly, guidelines by[9] request caution, that air drying may alter soil chemistry and aggregation, indicating the urgent need for careful consideration of storage methods.

The first SSL was delivered in 1980 [10] and then many others evolved either at local, regional, and even global scales, which introduces the importance of SSLs in soil science[11]. The SSLs are used to generate proximal models to enable soil analyses in the most convenient, environmentally friendly, and cost-effective ways[12]. These libraries are heavily utilized to generate soil maps based on remote sensing data from different domains[13]. Despite the importance of SSLs[14] widespread use of soil spectroscopy[15] and understanding that spectroscopy can provide estimates strongly correlated with wet-chemistry analyses[16], no consideration was drawn to long term storage effects of SSLs. It must be mentioned that stability of soils under laboratory condition has been examined over shorter periods of time 5–10 years, with limited focus on the longer-term storage of soils from different regions[17]. Robertson recognized this gap[18], who was advocating for comprehensive research on the long-term effects. It is worth noting that long-term storage research is particularly important for SSLs, as many facilities have samples that are 10–30 years old. These SSLs are stored under a variety of conditions, often determined by local practices that may incur minimal costs or lack sophistication. Limited research (if any) has specifically addressed the specific stability of SSLs-stored soils, leaving questions about the effects of extended storage conditions on sample integrity.

To evaluate the stability of soil properties, this study focuses on soil organic matter (SOM), a dynamic property that reflects biological activity, and calcium carbonate ($CaCO_3$), a relatively chemically stable mineral. These contrasting characteristics provide a balanced perspective on the resilience of organic and inorganic soil components over time[19]. The objective of this study is to evaluate the stability of a selected SSL over an extended period, specifically focusing on samples measured both spectrally and chemically with a 20–38-year interval. The hypothesis is that, under common and practical uncontrolled indoor storage (air-drying and sealing), key soil properties (SOM, $CaCO_3$) remain stable over a multi-decade period. This study will contribute critical insights into the long-term preservation of soil spectral and selected chemical properties within SSLs, inferring the most optimal practices for management and storage of other SSLs.

Materials and methods
Soil samples
Ninety-one soils from Israel, known as the legacy of the Israeli Soil Spectral Library (SSL)[11] were used for this research. The legacy SSL was established in 1987, and the information on sample locations, soil attributes, and spectral measurements is reported in multiple publications[20–24]. In general, the soils were sampled from the upper 0–5 cm over Mediterranean climate regions. The soils were brought to the laboratory, air-dried, gently crushed, and sieved to particles smaller than 2 mm. The < 2 mm fraction has been stored in plastic containers under indoor conditions. Annual room temperatures varied seasonally, ranging from 10 °C in the winter (for approximately 2 months) to 40 °C in the summer (for about 4 months), with intermediate temperatures (20–30 °C) for 6 months. Relative humidity ranged from 40% in winter to 70% in summer. This storage cycle has continued for 38 years.

Chemical analyses
To evaluate the stability of soil organic matter (SOM) and calcium carbonate ($CaCO_3$) respectively over time, two primary comparison strategies were employed: histograms to visualize the distribution of values for each period (1987 and 2024), and linear regression to assess the strength of relationships between the two datasets[25]. Scatter plots with fitted regression lines were used to identify trends and quantify stability, with the coefficient of determination (R^2) serving as the primary metric.

Following this comparative analysis, SOM and $CaCO_3$ were measured using standardized protocols to ensure consistency:

- *SOM*: The loss on ignition method[26] was used in both data sets. Three replicates of 5–8 g of precisely weighed soil were heated at 105 °C for 24 h to account for hygroscopic water. After weighing, the samples were heated at 405 °C for 8 h. Once cooled in a desiccator, the samples were weighed again. SOM was calculated based on the weight loss between 105 °C and 405 °C, accounting for the dry soil weight after hygroscopic water removal. Three replications were used for each soil sample.
- *$CaCO_3$*: The calcimeter method[26] was used in both data sets. Three replicates of 3–5 g of soil (precisely weighed) were reacted with 5% HCl in a calcimeter, where the resulting CO_2 gas was captured and measured relative to an analytical $CaCO_3$ standard powder measured under the same conditions. The $CaCO_3$ content was calculated relative to the pure $CaCO_3$ reaction with HCl.

Spectral measurement
An ASD FieldSpec 3 and an ASD FieldSpec 4 Hi-Res spectrometer were used in 2004 and 2024 respectively to measure soil reflectance. The same protocol was utilized for each year as follows: Each sample was measured three times using a contact probe device with a white reference (Labsphere*) for calibration and reference. Reflectance was calculated relative to the white reference (Labsphere*) considered as 100% reflectance, and the three replicate measurements were averaged before and after every nine soil sample measurements. Spectral measurements were conducted in both 2004 and 2024 using the ASD spectrometer models mentioned above, using different (clean) white references each year. The same soil preparation protocol was applied in both years. In 2004, there was no procedure available at that time to harmonize the spectral data with the Internal Soil Standard (ISS) method (e.g., LB)[27], so no ISS correction was applied to the 2024 measurements. Detector alignment corrections were applied to the two detectors encompassing the full spectral range following the ASD protocol[27].

Spectral differences

The modified Average Spectral Difference Stability (mASDS) was used to evaluate the consistency of the spectrometer across repeated measurements of the same sample for the 2004 and 2024 datasets[21,28]. The mASDS quantifies variability in spectral reflectance values over a specified wavelength range[28]. First, the mean reflectance at each wavelength was calculated for each spectrum Eq. 1. Next, the absolute differences between each spectrum and the mean spectrum were computed to determine spectral differences Eq. 2. The mASDS was then derived by averaging these absolute differences across all measurements (n) and wavelengths (m) Eq. 3. To facilitate comparison, the mASDS was normalized by the mean reflectance value of the sample and expressed as a percentage Eq. 4. This approach provides a robust metric for evaluating spectral stability, ensuring repeatability and reliability under controlled conditions[21,28].

$$S(\lambda) = \frac{1}{n} \sum_{i=1}^{n} S_i(\lambda) \tag{1}$$

This equation calculates the mean spectral reflectance (S(λ)) for a given wavelength (λ) across all n measurements of the same sample. Si(λ): Spectral reflectance of the i-th measurement at wavelength λ. n: Total number of measurements.

$$\Delta Si(\lambda) = |Si(\lambda) - S(\lambda)| \tag{2}$$

This equation calculates the absolute difference (ΔSi(λ)) between the spectral reflectance of the i-th measurement (Si(λ)) and the mean spectral reflectance (Si(λ)) for a given wavelength(λ).

$$mASDS = \frac{1}{n \cdot m} \sum_{i=1}^{n} \sum_{\lambda=1}^{m} \Delta S_i(\lambda) \tag{3}$$

The modified Average Spectral Difference Stability (mASDS) quantifies the average absolute difference in spectral reflectance across all measurements (n) and wavelengths (m). m: Total number of wavelengths analyzed in the spectral range.

$$Stability(\%) = \left(\frac{mASDS}{S(\lambda)} \right) \times 100 \tag{4}$$

This equation expresses mASDS as a percentage relative to the mean reflectance value, providing a relative measure of spectral stability.

Spectra modelling

We established a predictive modeling framework designed to evaluate the alignment and compatibility between the 1987/2004 (chemical and spectral databases respectively) and 2004/2024 SSL dataset. The modeling approaches were developed to ensure robustness in handling $CaCO_3$ and SOM content predictions across datasets collected with a significant temporal gap. To explore trends and relationships, we tested a range of methodologies, including regression and machine learning techniques, aiming to identify a framework that balances reliability and generalization. Principal Component Analysis (PCA) was used to assess spectral stability by analyzing variance across the 2004 and 2024 datasets. The spectral data were standardized prior to PCA to ensure comparability. The first two principal components (PC1 and PC2), which accounted for over 99% of the variance, were analyzed to determine the consistency of spectral patterns over time[29]. Based on its capability to effectively balance model complexity with prediction accuracy, Partial Least Squares Regression (PLSR) emerged as the preferred technique for its suitability in handling high-dimensional spectral data but also for its simplicity in hyperparameter tuning. PLSR was applied to model the relationship between spectral data and measured SOM and $CaCO_3$. Dimensionality reduction techniques and preprocessing steps, such as signal smoothing and data transformation, were employed to improve the model's robustness. Leave-one-out cross-validation (LOOCV) was used to minimize bias, with model performance evaluated using metrics such as the coefficient of determination (R^2) and the Ratio of Performance to Interquartile Range (RPIQ)[30]. Feature importance was calculated by averaging the PLSR coefficients across all folds of the LOOCV. This approach provided a consistent measure of the dominant spectral regions contributing to the prediction of soil properties. Outliers in the chemical and spectral datasets were identified based on deviations from the 1:1 line in scatter plots of predicted versus measured values and through PCA analysis to detect anomalies in spectral patterns. These outliers were reviewed to determine whether they arose from measurement errors or variability in the soil samples themselves[29,31]. All statistical analyses were performed using Python libraries, including scikit-learn (v1.2.2), NumPy (v1.23.5), and pandas (v1.5.3). Data visualization was conducted using Matplotlib (v3.6.2).

Results and discussion
Results

In Fig. 1) we provide the general information about the soil population utilized for this research, aligning with the USDA soil orders. This indicates the soil population is highly representative. Figure 2 presents histograms of SOM and $CaCO_3$ for each dataset, demonstrating a broad and even spread across their respective ranges, indicating that the selected soil samples adequately represent the variability of these properties within the

studied population. The chemical ranges for 1987 and 2024 appear consistent, suggesting no significant changes over the 38-year span. To evaluate this further, two scatter plots comparing the 1987 and 2024 datasets for SOM and $CaCO_3$ in Fig. 3 reveal a strong linear relationship between the two ($R^2 = 0.925$ and $R^2 = 0.962$ respectively), confirming the stability of chemical values over time. However, deviations from the 1:1 line and differences in statistical values, such as the coefficient of determination (R^2), standard error of prediction (SEP), and residuals, suggest that these discrepancies are likely due to measurement precision and uncertainties in the wet chemistry analyses of both datasets.

To compare spectral stability over time (2004–2024), twelve representative samples were selected, varying by both USDA order and soil properties. Their spectral pairs (2004 vs. 2024) are shown in Fig. 4 revealing a high correlation quantitatively, with the mASDS metric confirming minimal differences (0.002 and 0.066) indicating high spectral stability of the accuracy (spectral average (AVG)) between years further confirmed by Shepherd[32], which highlighted similar results boasting the significance of long-term SSLs. The high stability observed holds for both albedo (real part of the complex refractive index) and spectral features (imaginary part of the complex refractive index), which are directly influenced by the chemical properties of the scanned samples, such as the molecular interactions (e.g., C-H, O-H, and C-O bonds) associated with soil organic matter and calcium carbonate. Spectral measurements consistently demonstrated better precision and reliability than wet chemistry analyses, highlighting their (spectral) superior performance. To further support this observation, Principal Component Analysis (PCA) was applied to validate the consistency of spectral patterns across datasets. The first two principal components (PC1 and PC2) loadings from PCA, as shown in Fig. 5, highlight each feature's contribution to the directions of maximum variance. The raw spectral data were standardized prior to PCA, ensuring a fair comparison across datasets. The similarity in PC1 loadings confirms that the dominant patterns of variance are highly consistent. Spectral peaks or variations, which often indicate regions of higher variance, were well captured. The variance percentages explained by PC1 (97.3% for 2024 and 97.1% for 2004) reinforce this finding, showing that PC1 captures nearly all the key variance in both datasets. Additionally, PC2, capturing approximately 2% of the variance, reflects consistent secondary variance patterns, further confirming the spectral alignment between datasets as seen in Fig. 5.

Partial least squares regression (PLSR) was used to assess the relationship between chemical and spectral data for both datasets, with model performance evaluated using leave-one-out cross-validation (LOOCV) to ensure consistency. For $CaCO_3$, the model used the 2024/2024 dataset which achieved an R^2 of 0.57 with an RPIQ (Ratio of Performance to Interquartile Range)[33] of 2.26, closely matching the 1987 /2004 dataset, which had an R^2 of 0.60 and an RPIQ of 2.24. Similarly, for SOM, the model of the 2024 dataset produced an R^2 of 0.59 and an RPIQ of 2.50, while the 1987 /2024 dataset performed slightly better, with an R^2 of 0.67 and an RPIQ of 2.58 as seen in Fig. 6. Scatter plots of predicted versus actual values seen in Fig. 6, further illustrate how trends in predictive performance were consistent across both datasets. The outliers in each scatter plot are different, and as we have already demonstrated, the spectral measurements were of high quality. This leads us to assume that the outliers evolved from variations or uncertainties in the chemical measurements each year. Similar observations were noted by Williams[34] when measuring food products Each wavelengths contribution was identified based on the coefficients of the partial least squares regression (PLSR) model, which highlights the spectral regions most influential in predicting soil properties seen in Fig. 7[35]. The dominant spectral regions contributing to soil property predictions were identified, highlighting consistent features across datasets. The resulting mean
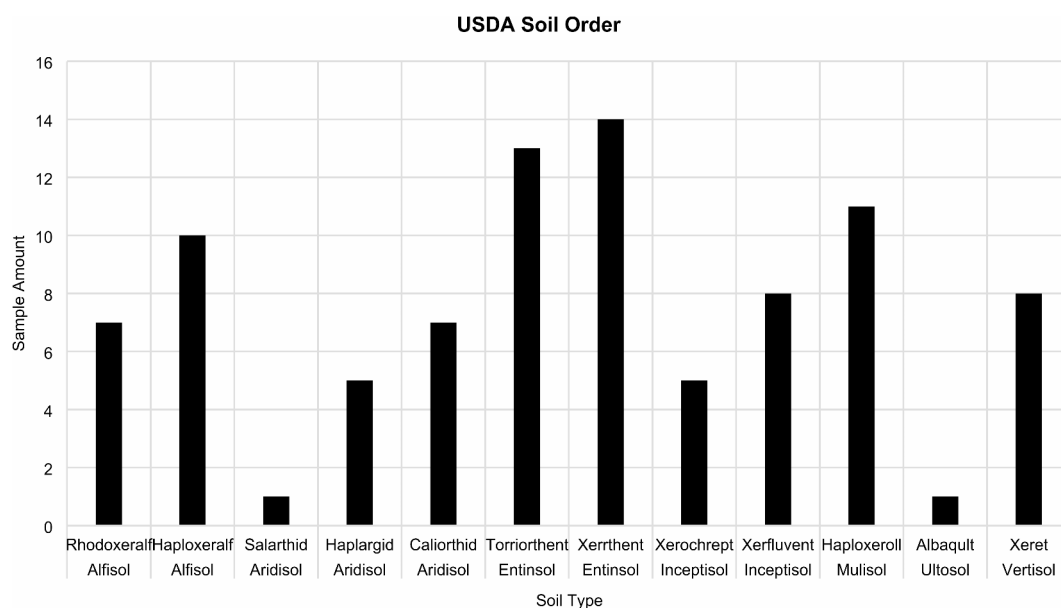

Fig. 1. Representation of Soil Population Across USDA Soil Orders. This figure illustrates the soil population used in the study, represented in the lower list, mapped against USDA soil orders in the upper list. It highlights the diversity and representativeness of the sampled soils within the research context.
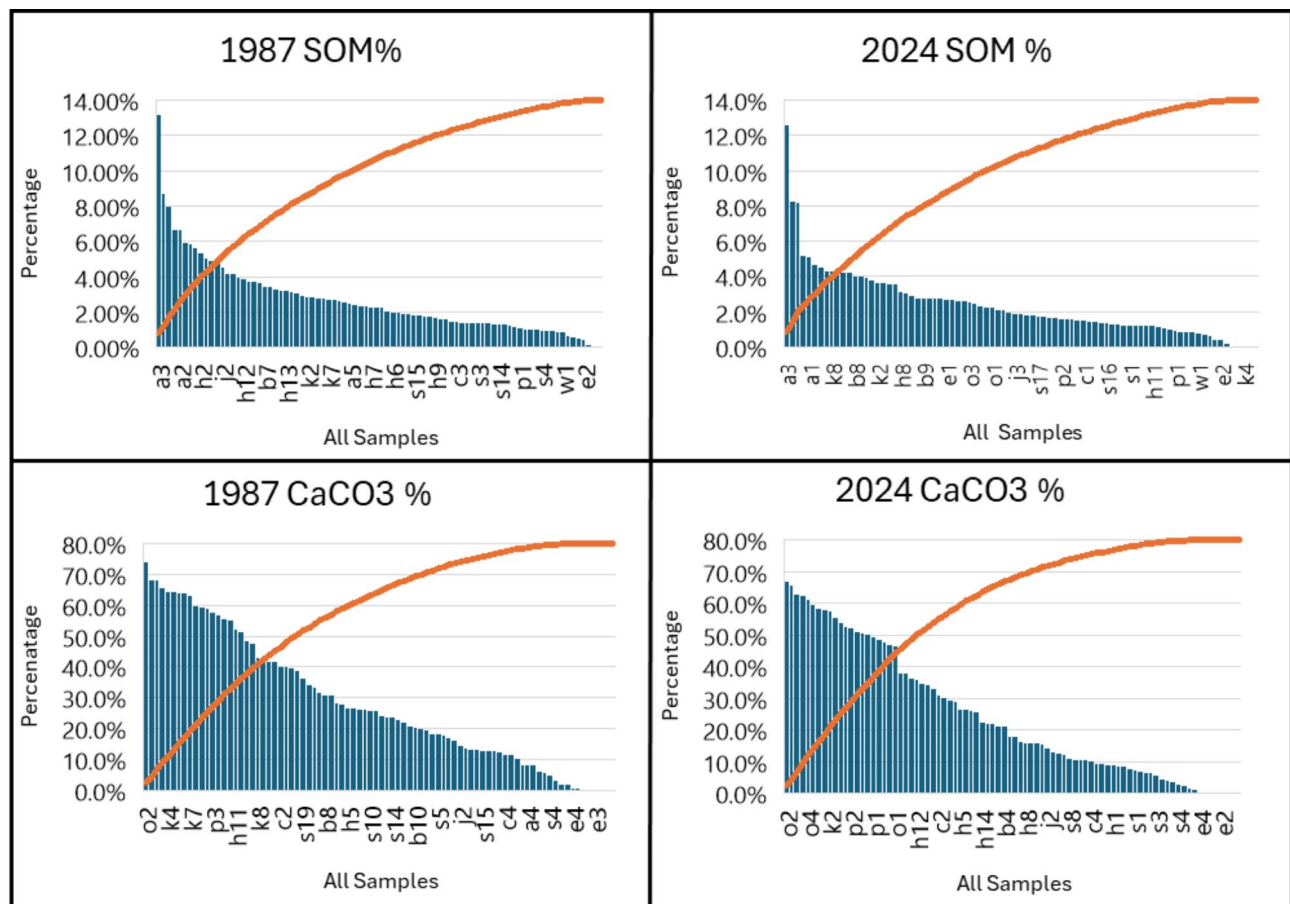
Fig. 2. Histograms of Soil Organic Matter (SOM) and Calcium Carbonate (CaCO₃) Percentages. Histograms showing the distribution of SOM and CaCO₃ percentages for the datasets from 1987 and 2024. The figure highlights the consistency in the chemical ranges of the two datasets over a 37-year storage period.
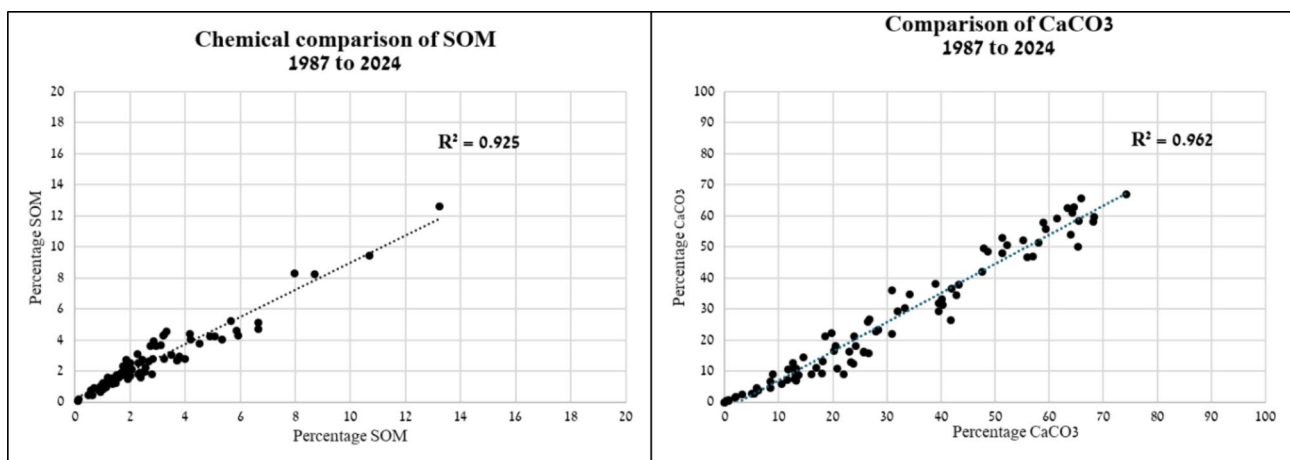

Fig. 3. Scatter plots comparing SOM (left) and CaCO₃ (right) between 1987 and 2024. Scatter plots illustrating the linear chemical relationship between SOM and CaCO₃ measurements from 1987 and 2024.

coefficients were used to identify the dominant spectral regions, providing a more reliable measure of feature importance. The higher feature importance values for CaCO₃ compared to SOM can be attributed to the distinct and well-defined spectral absorption bands associated with calcium carbonate, particularly in regions such as 2300–2350 nm and 1400–1450 nm. In contrast, SOM, comprising a mix of diverse organic compounds, exhibits more diffuse and complex spectral features, resulting in comparatively lower feature importance values. For

Fig. 4. Spectral stability across diverse soil samples over time. 12 Representative soil order samples with spectral comparisons over time (1987 vs. 2024) for selected soil samples (A3, B3, C3, E3, EC2, H3, K3, J3 O3, P3, S3, W1).
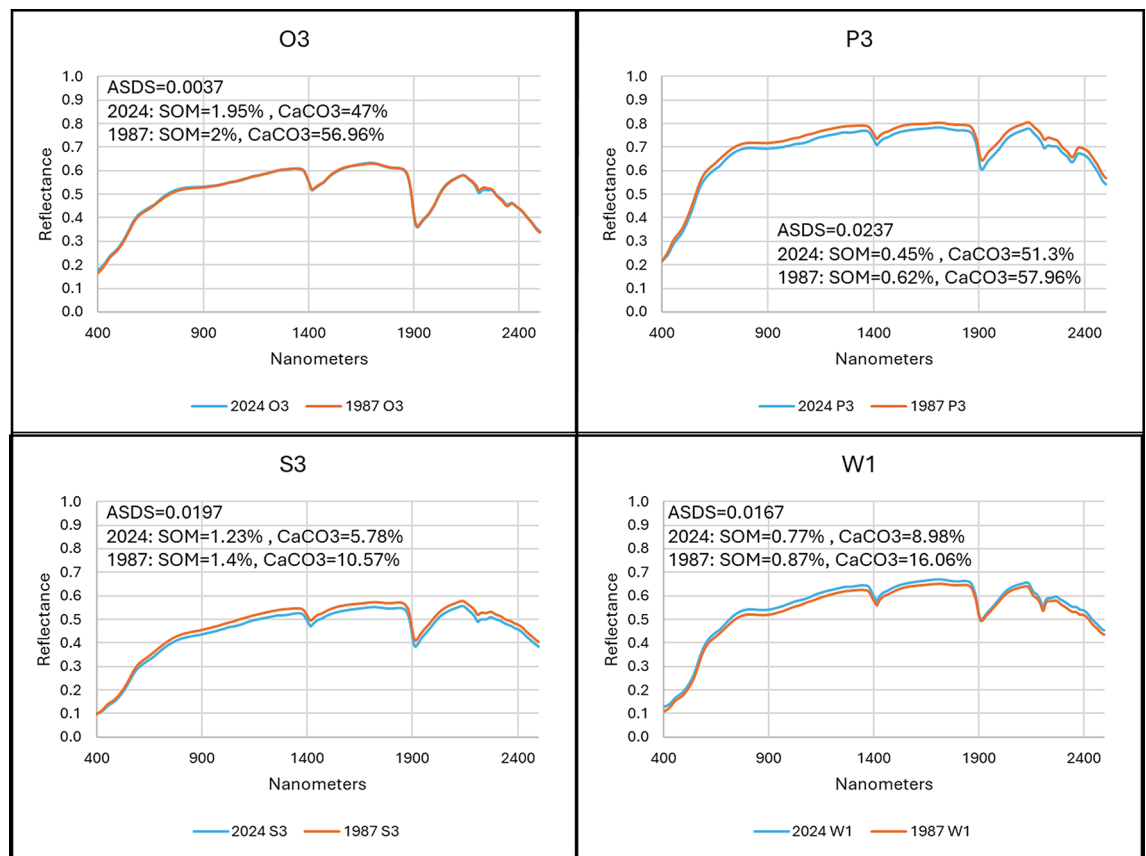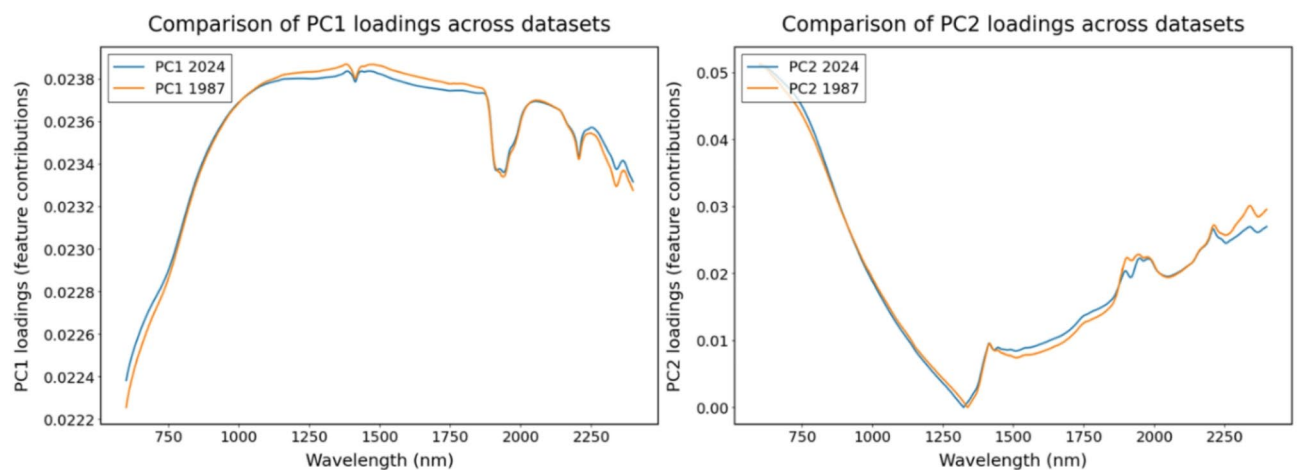
Figure 4. (continued)

Fig. 5. Principal component analysis (PCA) loadings for spectral data. The figure compares PC1 (left) and PC2 (right) loadings for 2024 and legacy (2004) datasets. PC1 explains ~97% variance in both datasets, while PC2 accounts for ~2%. This highlights consistent variance structures and stable feature contributions over time.

SOM, dominant bands in both years included 2200–2350 nm, 2000–2100 nm, 1800–1850 nm, 1400–1550 nm and 600–700 nm, while for $CaCO_3$, significant regions in both years included 2300–2350 nm, 2010–2050 nm, 1400–1450 nm, and 700–800 nm Fig. 7. Absorptions in SOM at 2200–2350 nm, 2000–2100 nm, and 1800–1850 nm regions may not solely reflect functional groups of organic compounds (e.g., C–H, O–H, and N–H bonds) but could also be influenced by interactions between organic matter and reactive minerals, such as 2:1 clay minerals, which are prevalent in mineral soils with low organic matter content. The 1400–1550 nm region reflects fundamental O–H stretching and H–O–H bending vibrations, which may be linked to hygroscopic
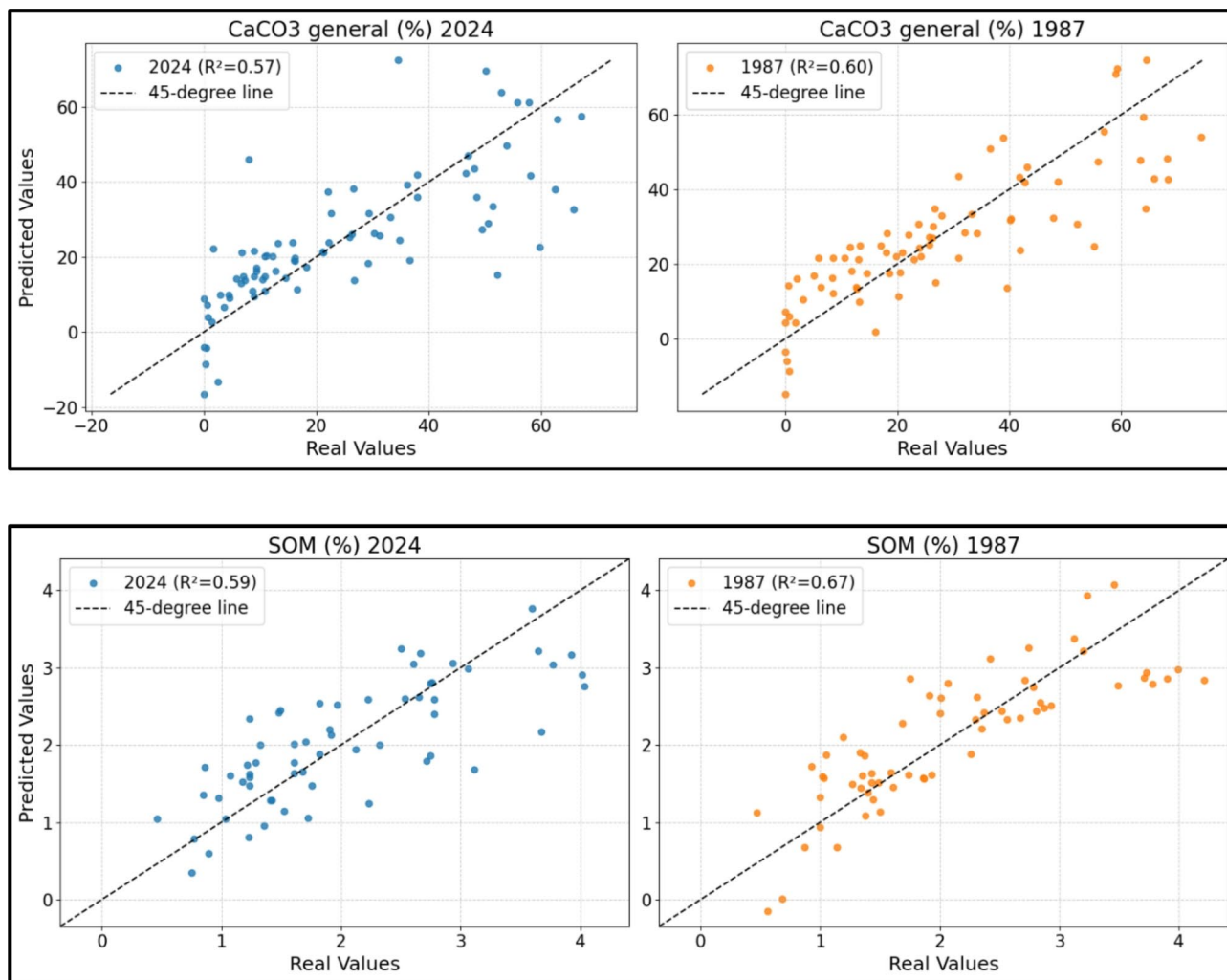
Fig. 6. Scatter plots for predictive modeling. Scatter Plots for $CaCO_3$ (top) and SOM (bottom), 2024 Dataset (left, blue), 1987 Dataset (right, orange). All plots reveal the predictive stability of the data.
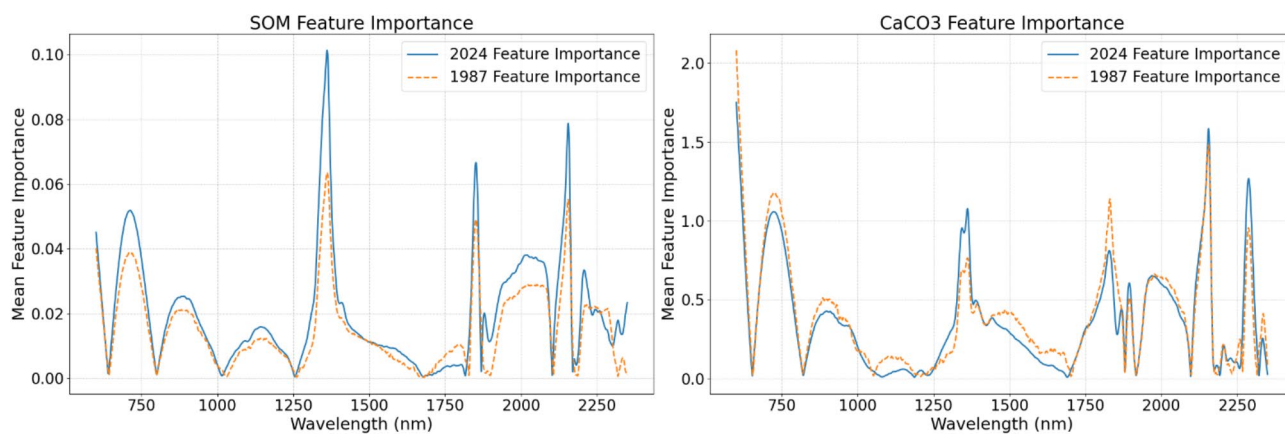

Fig. 7. Spectral feature importance for prediction analysis.. Plots for SOM (%) (left) and $CaCO_3$ general (%) (right), showing the contribution of wavelengths (x-axis) to predictions (y-axis). Solid lines represent 2024 data; dashed orange lines represent 1987 data. Peaks indicate key wavelengths for predictions

8

water bound within organic matter and humic substances, as well as hydroxyl groups (OH-Al) present in the octahedral layers of clay minerals such as phyllosilicates. The visible region at 600–700 associated with the SOM spectral slope[36]. Absorption in $CaCO_3$, at 2300–2350 nm and 2010–2050 nm region captures strong absorption from asymmetric stretching and overtone combinations of C-O bonds in carbonate ions (CO_3^{2-}) while the 1400–1450 nm range is attributed to O-H stretching modes of weakly bound water in hydrated carbonate minerals[37]. The 700–800 nm region is tied to overall albedo reflectance on the bright $CaCO_3$'s material[37]. These spectral features were consistent across both datasets, reflecting the correct underlying properties that should be seen as spectrally important for both $CaCO_3$3 and SOM[38], revealing stability of predictive patterns over time[39]. The results demonstrate strong spectral similarity and alignment between the datasets, reinforcing the reliability of spectroscopy for long-term soil property prediction[34,40,41].

Discussion

The results demonstrate that Mediterranean soil samples stored under common, practical, uncontrolled, indoor conditions show minimal changes in key properties, even after decades. This stability is maintained despite outdoor environmental temperature fluctuations ranging from 10 °C in winter to 40 °C in summer. While this study focused on soil organic matter (SOM) and calcium carbonate ($CaCO_3$), past studies by Edwards and Turner[27] have suggested that other soil attributes may also remain stable under similar conditions. SOM, a dynamic soil property under natural conditions, and $CaCO_3$, a relatively stable mineral property, were chosen to represent contrasting characteristics. The observed variability in wet chemistry results likely stem from analytical precision and accuracy rather than changes in soil properties, as supported by the higher consistency and precision observed in the spectroscopic measurements. Using a similar spectrometer model and identical protocols over the decades further underscores the reliability of spectroscopy, which once calibrated with wet chemistry, the spectrometers are increasingly recognized for higher repeatability over time than repeating wet chemistry measurements, which can have higher procedural variability.

These findings highlight the critical need for standardized protocols in both spectral and chemical analyses to ensure reproducibility across laboratories and studies. Strong correlations between historical and contemporary datasets ($R^2 = 0.925$ for SOM and $R^2 = 0.962$ for $CaCO_3$) indicate that these properties, along with spectroscopic characteristics, have remained remarkably stable over approximately 3 decades. This stability suggests that, when soil samples are stored appropriately (e.g., air-dried and sealed), they can serve as reliable archives for longitudinal studies, enabling accurate assessments of temporal changes.

The minimal changes observed also support the potential use of archived soil samples as benchmarks in studying pedogenic processes and environmental changes, such as those driven by climate change. Moreover, countless soil samples stored in laboratories after field projects could be incorporated into new soil spectral libraries (SSLs) if spectral information is measured. For example, soils previously analyzed in Digital Soil Mapping projects could be spectrally scanned and added to SSLs, significantly increasing the sample base of global SSLs. This approach assumes that properties of archived soils remain in a condition similar to when they were sampled, presenting a valuable opportunity to expand SSL archives.

The results further emphasize the importance of harmonizing spectral measurements through common protocols, which is crucial for merging SSLs developed in different laboratories. Re-analyzing SSLs with and without newly added data remains feasible and ensures comparability. While this study maintained consistent methodologies across decades, the findings also stress the need for standardized chemical analysis protocols to match the precision and accuracy currently achieved in spectroscopy.

The implications of these findings are broad. Long-term soil sample storage under common and practical uncontrolled indoor conditions offers cost-effective archiving without the need for specialized facilities, preserving valuable environmental data for future research. Archived samples can validate historical datasets, inform studies on carbon cycling, soil health, ecosystem sustainability, and climate change impacts. Researchers can use these samples to monitor changes in soil properties over decades, while farmers and agronomists can apply these insights to improve soil management practices. Furthermore, environmental agencies can utilize these findings to guide policy decisions on soil conservation and land use.

This study supports the use of existing global and local SSLs, such as ISRIC and LUCAS, as benchmarks for monitoring $CaCO_3$, SOM, and other soil attributes over time. Although Mediterranean soils exhibit remarkable stability, further research is needed to examine the storage effects on soils from other regions, such as tropical or organic-rich soils, where microbial activity and organic matter dynamics may remain active in storage conditions.

The findings also have broader applications in disciplines such as geology and archaeology, highlighting the utility of well-preserved soil archives for historical and environmental assessments. Overall, this study provides critical evidence that Mediterranean soils stored under simple and uncontrolled conditions can maintain their physical and chemical integrity over long periods. These findings encourage more widespread archiving of soil samples and support their use as "fossil benchmarks" for studying soil formation, contamination, and climate change processes. By demonstrating the feasibility of long-term storage, this research lays a foundation for future environmental assessments, monitoring programs, and the development of more robust soil management strategies.

Conclusion

This study demonstrated that both soil chemical attributes (SOM and $CaCO_3$) and spectral properties remained stable over 2–3 decades, confirming that Mediterranean soils can be effectively preserved under common and practical room-temperature, uncontrolled storage conditions. This stability suggests that other soil properties,

particularly those with lower sensitivity to environmental changes, may also remain unchanged over similar periods.

The findings highlight that soil samples stored under common and practical uncontrolled indoor conditions that have been air-dried and sealed can serve as reliable archives for longitudinal studies, enabling accurate assessments of temporal changes. Spectroscopic analysis showed higher precision and reproducibility than wet chemistry when calibrated properly, underscoring its reliability as a complementary tool for analyzing archived soils. Standardizing chemical and spectroscopic methods are strongly recommended to enhance reproducibility across SSLs.

These results underscore the value of long-term soil sample storage under simple conditions, offering cost-effective archiving without the need for specialized facilities. Such archived samples can validate historical datasets and inform research on carbon cycling, soil health, ecosystem sustainability, and climate change impacts. Revisiting older SSLs on a global scale could provide valuable benchmarks for evaluating temporal changes in soil properties, particularly in other climatic conditions such as tropical regions.

In conclusion, this study provides critical evidence that Mediterranean soils stored under common and practical indoor uncontrolled conditions maintain their integrity over decades. This underscores the long-term value of SSLs as essential resources for monitoring soil processes, supporting environmental research, and informing soil and land management policies.

Data availability

References

1. Chabrillat, S. et al. Preparing a soil spectral library using the internal soil standard (ISS) method: Influence of extreme different humidity laboratory conditions. *Geoderma* **355**, 113855 (2019).
2. Tan, K. H. Soil sampling, preparation, and analysis. *Soil. Sampl. Prep. Anal.* https://doi.org/10.1201/9781482274769 (2005).
3. ISO 11277:2020. (en), Soil quality: Determination of particle size distribution in mineral soil material—method by sieving and sedimentation. https://www.iso.org/obp/ui/en/#iso:std:iso:11277:ed-3:v1:en
4. Forster, J. C. Soil sampling, handling, storage and analysis. *Methods Appl. Soil. Microbiol. Biochem.* 49–121. https://doi.org/10.1016/B978-012513840-6/50018-5 (1995).
5. Rubin, B. E. R. et al. Investigating the impact of storage conditions on microbial community composition in soil samples. *PLoS One* **8**, e70460 (2013).
6. Taylor, J. P., Wilson, B., Mills, M. S. & Burns, R. G. Comparison of microbial numbers and enzymatic activities in surface soils and subsoils using various techniques. *Soil. Biol. Biochem.* **34**, 387–401 (2002).
7. Cotrufo, M. F., Ranalli, M. G., Haddix, M. L., Six, J. & Lugato, E. Soil carbon storage informed by particulate and mineral-associated organic matter. *Nat. Geosci.* **12**, 989–994 (2019).
8. Pavlovska, M., Prekrasna, I., Parnikoza, I. & Dykyi, E. Soil sample preservation strategy affects the microbial community structure. *Microbes Environ.* **36**, 1–7 (2021).
9. van Schuylenborgh, J. The effect of air-drying of soil samples upon some physical soil properties. *Neth. J. Agric. Sci.* **2**, 50–57 (1954).
10. Stoner, E. R., Baumgardner, M. F., Weismiller, R. A., Biehl, L. L. & Robinson, B. F. Extension of Laboratory-measured soil spectra to field conditions. *Soil Sci. Soc. Am. J.* **44**, 572–574 (1980).
11. Viscarra Rossel, R. A. et al. A global spectral library to characterize the world's soil. *Earth Sci. Rev.* **155**, 198–230 (2016).
12. Wadoux, A. M. J. C., Minasny, B. & McBratney, A. B. Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth Sci. Rev.* **210**, 103359 (2020).
13. Rizzo, R., Demattê, J. A. M., Lepsch, I. F., Gallo, B. C. & Fongaro, C. T. Digital soil mapping at local scale using a multi-depth Vis–NIR spectral library and terrain attributes. *Geoderma* **274**, 18–27 (2016).
14. Rossel, R. A. V., Jeon, Y. S., Odeh, I. O. A. & McBratney, A. B. Using a legacy soil sample to develop a mid-IR spectral library. *Soil. Res.* **46**, 1–16 (2008).
15. Ben-Dor, E. et al. Using imaging spectroscopy to study soil properties. *Remote Sens. Environ.* **113**, S38–S55 (2009).
16. Nocita, M. et al. Soil spectroscopy: An alternative to wet chemistry for soil monitoring. *Adv. Agron.* **132**, 139–159 (2015).
17. Pavlovska, M., Prekrasna, I., Parnikoza, I. & Dykyi, E. Soil sample preservation strategy affects the microbial community structure. *Microbes Environ.* **36**, ME20134 (2021).
18. Robertson, P. K. Soil classification using the cone penetration test. *Can. Geotech. J.* **27**, 151–158 (1990).
19. Rowley, M. C., Grand, S., Verrecchia & É. P. Calcium-mediated stabilisation of soil organic carbon. *Biogeochemistry* **137**, 27–49 (2018).
20. Ben-Dor, E., Heller, D. & Chudnovsky, A. A. Novel method of classifying soil profiles in the field using optical means. *Soil Sci. Soc. Am. J.* **72**, 1113–1123 (2008).
21. Ben-Dor, E., Granot, A. & Notesco, G. A simple apparatus to measure soil spectral information in the field under stable conditions. *Geoderma* **306**, 73–80 (2017).
22. Gholizadeh, A., Saberioon, M., Carmon, N., Boruvka, L. & Ben-Dor, E. Examining the performance of PARACUDA-II data-mining engine versus selected techniques to model soil carbon from reflectance spectra. *Remote Sens. (Basel),* **10**(8), 1172 (2018).
23. Falcioni, R. et al. Non-invasive assessment, classification, and prediction of biophysical parameters using reflectance hyperspectroscopy. *Plants* **12**, 2526 (2023).
24. Ben-Dor, E., Patkin, K., Banin, A. & Karnieli, A. Mapping of several soil properties using DAIS-7915 hyperspectral scanner data: A case study over clayey soils in Israel. *Int. J. Remote Sens.* **23**, 1043–1062 (2002).
25. Mozaffari, H., Moosavi, A. A. & Nematollahi, M. A. Predicting saturated and near-saturated hydraulic conductivity using artificial neural networks and multiple linear regression in calcareous soils. *PLoS One.* **19**, e0296933 (2024).
26. Nogueira Martins, R., Santos, F. L. D., De Moura, F., Araújo, G., De Arruda Viana, L. & Fim Rosas, J. T. Accuracy assessments of stochastic and deterministic interpolation methods in estimating soil attributes Spatial variability. *Commun. Soil. Sci. Plant. Anal.* **50**, 2570–2578 (2019).

27. Ben Dor, E., Ong, C. & Lau, I. C. Reflectance measurements of soils in the laboratory: Standards and protocols. *Geoderma* **245–246**, 112–124 (2015).
28. Ben-Dor, E., Kindel, B. & Goetz, A. F. H. Quality assessment of several methods to recover surface reflectance using synthetic imaging spectroscopy data. *Remote Sens. Environ.* **90**, 389–404 (2004).
29. Godoy, J. L., Vega, J. R. & Marchetti, J. L. Relationships between PCA and PLS-regression. *Chemometr. Intell. Lab. Syst.* **130**, 182–191 (2014).
30. Yan, Q., Yang, C. & Wan, Z. A. Comparative Regression Analysis between principal component and partial least squares methods for flight load calculation. *Appl. Sci.* **13**, 8428 (2023).
31. Zeng, R. et al. How similar is similar, or what is the best measure of soil spectral and physiochemical similarity? *PLoS One* **16**, e0247028 (2021).
32. Shepherd, K. D. et al. A global soil spectral calibration library and Estimation service. *Soil. Secur.* **7**, 100061 (2022).
33. Wan, X., Wang, W., Liu, J. & Tong, T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med. Res. Methodol.* **14**, 1–13 (2014).
34. Williams, P. C. & Stevenson, S. G. Near-infrared reflectance analysis: Food industry applications. *Trends Food Sci. Technol.* **1**, 44–48 (1990).
35. Norris, F. A. et al. Field studies of Ethoprop movement and degradation in two Florida soils. *J. Contam. Hydrol.* **8**, 299–315 (1991).
36. Ben-Dor, E., Inbar, Y. & Chen, Y. The reflectance spectra of organic matter in the visible near-infrared and short wave infrared region (400–2500 nm) during a controlled decomposition process. *Remote Sens. Environ.* **61**, 1–15 (1997).
37. van der Meer, F., Yang, H. & Lang, H. Imaging Spectrometry And Geological Applications. 201–218 (2002). https://doi.org/10.1007/978-0-306-47578-8_7
38. Gomez, C. & Coulouma, G. Importance of the Spatial extent for using soil properties estimated by laboratory VNIR/SWIR spectroscopy: Examples of the clay and calcium carbonate content. *Geoderma* **330**, 244–253 (2018).
39. Adeline, K. R. M., Gomez, C., Gorretta, N. & Roger, J. M. Predictive ability of soil properties to spectral degradation from laboratory Vis-NIR spectroscopy data. *Geoderma* **288**, 143–153 (2017).
40. Ben-Dor, E. & Banin, A. Near-Infrared reflectance analysis of carbonate concentration in soils **44**, 1064–1069. http://dx.doi.org/10.1366/0003702904086821 (1990).
41. Lagacherie, P., Baret, F., Feret, J. B., Netto, M., Robbez-Masson, J. M. & J. & Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements. *Remote Sens. Environ.* **112**, 825–835 (2008).

Acknowledgements

Author contributions

J.E.S. wrote the main manuscript, preformed the spectral measurements and research and prepared the figures.O.K. preformed the modeling.O.A. preformed the chemical analysis.B.E. Preformed the chemical analysis.E.B.D. raised the idea, gave supervision, insight and leadership, preformed original measurements of the data from 1987, as well as writing and editing.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to J.E.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.