Geoderma 460 (2025) 117434

Contents lists available at ScienceDirect

Geoderma

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

Discussion paper

Proficiency tests for soil analysis services via proximal sensing

Tiago R. Tavares [a,*], José A.M. Demattê [b,c], Abdul M. Mouazen [d], Eyal Ben-Dor [e], Adriano A. Anselmi [f], Victor Monseff de A. Campos [g], Budiman Minasny [h], Sérgio H. Godinho Silva [i], Sabine Chabrillat [j,k], Konstantinos Karyotis [l], Heitor Cantarella [m], Cristiano A. Andrade [n], Maurício R. Cherubin [b,c], Hudson W.P. de Carvalho [a], José Lavres [a]

[a] Center for Nuclear Energy in Agriculture (CENA), University of São Paulo (USP), Piracicaba, São Paulo, Brazil
[b] Department of Soil Science, "Luiz de Queiroz" College of Agriculture, USP, Piracicaba, São Paulo, Brazil
[c] Center for Carbon Research in Tropical Agriculture (CCARBON), USP, Piracicaba, São Paulo, Brazil
[d] Department of Environment, Faculty of Bioscience Engineering, Ghent University, Gent, Belgium
[e] Department of Geography, Porter School of Environmental and Earth Sciences, Faculty of Exact Science, Tel Aviv University, Tel Aviv, Israel
[f] Bayer Crop Science, São Paulo, SP, Brazil
[g] 3R Ribersolo, Agricultural Analysis Laboratory, Ribeirão Preto, SP. Brazil
[h] School of Life and Environmental Sciences & Sydney Institute of Agriculture, The University of Sydney, NSW, Australia
[i] Department of Soil Science, Federal University of Lavras (UFLA), Lavras, Minas Gerais, Brazil
[j] Helmholtz Centre for Geosciences (GFZ), Section Remote Sensing and Geoinformatic, Telegrafenberg, Potsdam, Germany
[k] Leibniz University Hannover (LUH), Institute of Earth System Science, Hannover, Germany
[l] School of Agriculture, Faculty of Agriculture, Forestry, and Natural Environment, Aristotle University of Thessaloniki, Thessaloniki, Greece
[m] Soils and Environmental Resources Center, Agronomic Institute of Campinas (IAC), Campinas, São Paulo, Brazil
[n] Embrapa Meio Ambiente, Jaguariúna, São Paulo, Brazil

ARTICLE INFO

Handling Editor: Naoise Nunan

Keywords
Analytical chemistry
Proximal soil sensing
Soil security
Soil health
Hybrid laboratory, standard and protocol

ABSTRACT

Soil Spectral Libraries (SSL) and spectroscopy-based methods using varied regions of the electromagnetic spectrum for direct soil analysis are rapidly gaining traction in agricultural and environmental applications. However, the absence of standardized tests, specifically designed for these new methods, prompts concerns about performance consistency and transparency in the marketplace. This discussion paper underscores the critical factors that influence predictive performance and stresses the need for proficiency tests tailored to sensor-based techniques. It also highlights the significance of error metrics for accurate interpretation. Establishing robust proficiency tests is crucial for ensuring reliable soil analysis services and promoting the adoption of best practices in this technology.

1. Introduction

The term soil spectral libraries (SSL) refers to collections of soil samples that have undergone spectral and physical/chemical analyses, along with their metadata (e.g., geographic location and edaphic information, sensors' instrumental conditions, and reference analysis details). In many cases, the samples are also stored in physical containers, allowing for potential reanalysis and additional sensor readings. Spectroanalytical techniques for proximal soil sensing—such as Diffuse Reflectance Spectroscopy in the Visible and Near-Infrared (Vis-NIR) and Mid-Infrared (MIR), X-ray Fluorescence Spectroscopy (XRF), and Laser-Induced Breakdown Spectroscopy (LIBS)—combined with predictive models based on SSLs are revolutionizing soil analysis by providing a fast, cost-effective, and sustainable alternative to gather extensive soil information and accelerate soil analysis (Ng et al., 2022a; Peng et al., 2025; Shepherd et al., 2022). This approach supports the growing demand for accurate, timely, and spatially detailed soil data, which is essential for monitoring soil changes over time and tackling global soil health and security challenges (Evangelista et al., 2024).

The portability of spectroanalytical sensors, their ease of use, and reagent-free operation, along with simple sample preparation requirements, offer innovative and practical solutions for soil analysis and monitoring. Spectroscopy-based sensing methods can significantly reduce the use of chemical reagents, making analyses more sustainable,

* Corresponding author.
E-mail address: tiagosrt@usp.br (T.R. Tavares).

https://doi.org/10.1016/j.geoderma.2025.117434

Available online 16 July 2025

economical, and environmentally friendly (Bartsch et al., 2024). In addition, these technologies enable (i) extending diagnostics to agricultural frontiers, far from laboratory infrastructure, through online cloud-based services (Demattê et al., 2022); and (ii) performing real-time analyses during on-site inspections (Tavares et al., 2023). However, despite the urgency of monitoring soil conditions using advanced technologies (Biswas, 2024), their adoption depends on users' confidence in the accuracy and reliability of the results, requiring comprehensive validation over time. Concerns about the quality of inferences via sensor-based methods (McBride, 2022) emphasize the need for rigorous and independent validation procedures and networks.

While laboratory wet chemistry soil analysis follows established standards, such as ISO 17025 and United States Department of Agriculture (USDA) soil testing protocols, there are still no formal international guidelines for spectroscopy-based methods as alternatives for soil testing—although initiatives are underway (Ben Dor et al., 2024; Stenberg et al., 2024). However, none of these efforts currently addresses proficiency testing—specifically for sensor-based methods. The absence of a methodology for proficiency testing in sensor-based estimation services —similar to those established for traditional soil testing (e.g., Quaggio et al., 1994)—allows service providers to overlook the implementation of best practices in proximal soil spectroscopy. Therefore, without international guidelines for the standardization of sensor-based soil analysis and proficiency testing, public concerns and restrictions on the reliability of these methods are raised.

Undeniably, the robust development of sensor-based inference systems requires mastering multidisciplinary skills, including data science for predictive modeling, spectroscopy for spectral acquisition, soil science for understanding agronomic complexities, and analytical chemistry for high-quality reference data. Therefore, recognizing (i) the potential of optical sensor-based estimation services; (ii) the technical challenges of developing reliable and reproducible spectral data and predictive models; and (iii) the global expansion of these techniques, this discussion paper advocates for the implementation of regional proficiency testing, based on international guidelines, as a crucial next step to strengthen public confidence in sensor-based methods using proximal spectroscopy.

Additionally, this document highlights key factors that influence the performance of spectral sensor-based methods, which must be properly addressed by both commercial and scientific service providers, as well as core aspects of proficiency testing that should guide the development of regional procedures. Although service providers may use different sensing techniques, proficiency tests remain independent of these methodological variations, focusing on (i) evaluating their predictive results; (ii) comparing them against analytical tolerance limits; and (iii) issuing public quality certificates.

2. Key factors influencing spectral sensor-based methods

Each spectral technique has underlying physical and/or chemical principles that explain the relationship between the soil spectrum and its properties. These principles determine the strength of the correlation between a technique and a specific soil attribute and the predictive model's potential for generalization. For instance, some techniques, such as near-infrared spectroscopy (NIR, 750–2500 nm) for plant-available nutrients in soils (e.g., available P, K, Ca, and Mg), show limited generalization due to variable attribute-spectra relationships (Stenberg et al., 2010). Plant-available nutrients in soils lack such signatures in NIR spectrum, which is primarily related to molecular vibrations related to molecular vibrations associated with functional groups containing a relatively heavy atom (C, N, O, and S) bonded to a hydrogen (Pasquini, 2018). Successful regressions of these nutrients are often attributed to their covariations with primary parameters that exhibit direct spectral responses. Understanding these principles is essential for developing robust modeling strategies grounded in cause-effect relationships—a concept emphasized in Minasny et al. (2024), which advocated for soil science-informed machine learning.

Research related to the different techniques used for proximal soil sensing is still evolving, although it has been built on a robust foundation established over the past decades. For instance, the inference of spectrally active attributes, such as soil organic carbon or clay content using diffuse reflectance spectroscopy, has already demonstrated robustness at regional (Demattê et al., 2019), national (Seidel et al., 2019), and continental scales (Padarian et al., 2019; Viscarra Rossel et al., 2016). The same applies for soil texture predictions from local to national and multinational scales using XRF technique (Chatterjee et al., 2021; Mancini et al., 2024; Silva et al., 2020; Zhu et al., 2011). Beyond the relationship between spectral features and soil attributes, several practical factors critically influence the reliability and stability of sensor-based methods. The following factors should be carefully considered by sensor-based service providers, who must be up to date with advances in best practices for each of them:

1. **Sample preparation procedures:** Drying and sieving reduce micro-heterogeneities and physical matrix effects, improving measurement consistency (Murray and Cowe, 2015). Spectroscopy is highly dependent on the particle-size distribution of the soil sample and may, therefore, be sensitive to micro aggregation during surface preparation for spectral measurements (Ben-Dor, 2002). Suitable preparation of the soil for measurement is a crucial step to obtaining reliable and reproducible results. Furthermore, since any spectroscopy technique is influenced by the water content in the soil analyzed, a suitable residual moisture state monitoring of the soil samples and soil standards used during spectral measurements is important to prevent minor variations that could subtly degrade SSLs performance (Chabrillat et al., 2019). The drying process must be standardized—especially for soils with divergent water retention (e.g., clayey vs. sandy)—and precautions should be taken to prevent moisture absorption from the air, ensuring consistent residual moisture levels and reliable spectra.
2. **Sensor technique(s) employed:** Each technique has its own potential and limitations for predicting different soil attributes, which must be carefully considered. The range of successfully inferred attributes can increase with the combination of techniques; for example, Vis-NIR spectroscopy has potential for texture and soil organic carbon (SOC) (Barra et al., 2021), while XRF or LIBS sensors have shown potential for inferring plant-available nutrients (Piikki et al., 2016). In addition, combining complementary sensors through data fusion approaches can further boost predictive accuracy of a single attribute, as observed for SOC when combining Vis-NIR and XRF techniques (Shi et al., 2023; Teixeira et al., 2022).
3. **Hardware specifications:** In general, the type and stability of the radiation source, as well as the range, resolution, and sensitivity of the detector, directly affect signal quality; however, each technique has specific requirements. For instance, in diffuse reflectance spectroscopy (DRS) sensors, a stable and uniform radiation output across the required spectral range is essential. In contrast, in XRF sensors, radiation sources using Rh-anode X-ray tubes may perform differently from Ag-anode tubes; likewise, voltage and current settings also influence performance (Tavares et al., 2020).
   Technological advances that optimize these components can further enhance signal quality and, consequently, model performance. Additionally, any of the aforementioned radiation sources may degrade over time; therefore, their output should be routinely monitored and inspected to ensure consistent performance.
4. **Spectral data acquisition procedures:** Each sensor user must follow a spectral data acquisition protocol or establish its own one to standardize the intrinsic aspects of each technique that are fundamental for obtaining temporally consistent and reproducible spectral measurements. Adopting international protocols is strongly recommended to ensure best practices, enhance comparability, and facilitate data exchange. Protocols and standardizations for laboratory

measurements using DRS sensors have been discussed by Ben Dor et al. (2015), and a formal standardization and protocol for DRS sensors is currently under development (Ben Dor et al., 2024). For XRF analysis of soils, some procedures have been proposed (Silva et al., 2021; USEPA, 2007; Weindorf and Chakraborty, 2016). However, standardization efforts for soil analysis using other techniques, such as XRF and LIBS sensors, have yet to be formally established, and are strongly encouraged to be addressed by the scientific community.

In general, the best practices for spectral-analytic analysis include adjusting sensor-sample geometry, instrumental conditions (e.g., integration time for DRS sensors, and voltage, current, and scanning time for XRF sensors), choosing an optimal number of replicates or a spectral acquisition strategy that covers enough sample amount (e.g., schemes to rotate the sample during the reading), choosing an appropriate warm-up time for radiation source stabilization (for DRS sensors), procedures for verifying the stability of the source and detector, and controlling any variable that may influence spectral data, such as laboratory humidity and temperature (Chabrillat et al., 2019).

Regarding sample measurement representativeness, it is important to maximize the area probed during spectral acquisition, as limited measurement points can lead to spectra that poorly represent the sample's overall composition. Rotating or moving the sample relative to the optical beam in NIR sensors, for instance, has been shown to enhance representativeness and improve SOC model's performance (Fonseca et al., 2022).

Incorporating periodic measurements of an Internal Soil Standard (ISS) into routine analyses protocol is also important to track possible fluctuations due to changes in operation, technical incidents, and changes in the environment. An ISS can be also used to minimize systematic effects caused by divergences in laboratory conditions or measurement protocols (Ben Dor et al., 2015), as is proposed by the IEEE-SA Working Group P4005 Standards and Protocols in Soil Spectroscopy (Ben Dor et al., 2024). This concept is summarized in Fig. 1. The construction of SSL using an ISS for data standardization is key for the collaborative development of large-scale soil spectral libraries.

5. **Reference analysis quality:** The quality of the reference analysis used during the modeling process directly impacts the error of the sensor-based method, so it is always essential to use high-quality references. This need has already been evidenced in a recent regional study (Demattê et al., 2019), where analytical variability significantly affected the accuracy and reliability of spectral soil analysis.

   It is worth highlighting that conventional wet chemistry analysis is susceptible to errors, particularly due to human error during the various steps into the established methodologies, reagent quality, and variation in equipment settings, as observed by Paiva et al. (2022) when compared 38 wet chemistry laboratories. Spectral service providers should be aware of the crucial importance of using good reference analyses, with known uncertainties, for calibrating their models.

   Metrics to evaluate wet chemistry data are essential, such as understanding the error of the reference method through replicate analyses. Models that rely on wet chemistry data from different laboratories and/or analytical methods (e.g., texture determined by the pipette or densimeter method) may suffer from reduced accuracy, even if the spectral measurements themselves are precise.

6. **Modeling strategy:** Soil complexity requires advanced models that address matrix effects and capture specific contexts regarding the relationships between spectra and soil attributes (Ng et al., 2022b). Furthermore, the methods considered for spectral data preparation (e.g., averaging, smoothing, derivatives, and variable selection) are important in the modeling process to predict soil attributes, since they can remove physical effects in the spectra and improve predictions (Rinnan et al., 2009).

   After modeling, it is crucial to test the model with an external validation set (i.e., samples not used in calibration) that reflects the target region. This ensures independent validation and provides an estimate of uncertainty relative to the reference method. Additionally, establishing a robust outlier detection procedure and continuous inputs of data into the library are highly recommended, in agreement with Poppiel et al. (2022).

   Another essential step in the model validation process is to investigate and explain the spectral regions utilized by the model. Since random noise can sometimes be correlated with the analyte, it is crucial to ensure that the model's reliance on specific spectral features is based on meaningful spectral properties rather than spurious correlations. When using models with limited interpretability (e.g., deep learning techniques), it is recommended it is essential to perform post-modeling analysis, such as Shapley value-based methods, to assess the contribution and relevance of different spectral regions, enhancing trust and understanding (Wadoux, 2023).

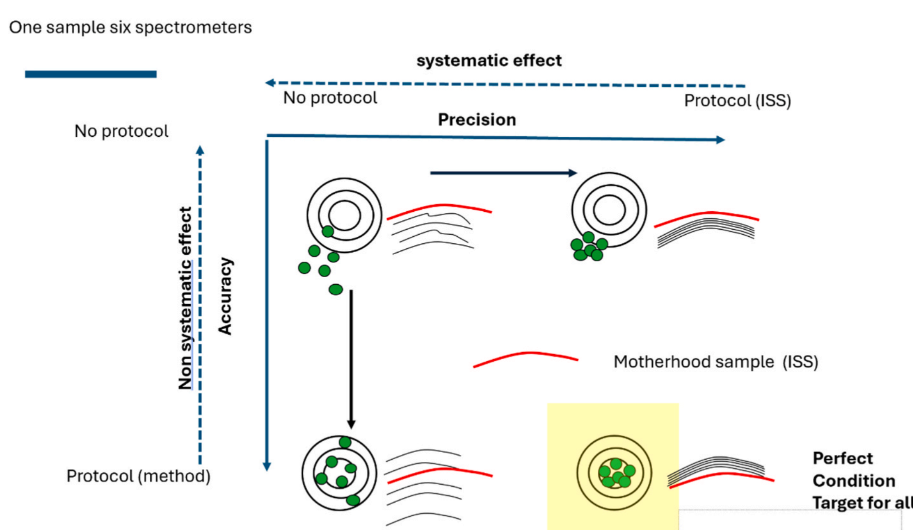

Fig. 1. Illustration of the variation in Vis-NIR spectral measurements from six different spectrometers analyzing the same soil sample. The observed discrepancies reflect issues of precision (random/non-systematic error) and accuracy (systematic error). The figure also demonstrates how the application of a shared internal standard—agreed upon by all six laboratories—can correct for both types of error, resulting in harmonized measurements with improved accuracy and precision.

7. **Geographical scope of the SSL:** Calibration sets must represent the target regional scope in terms of both spectral and soil variations. While local libraries may require a few dozen to hundred samples to represent a specific area (Lucà et al., 2017; Ramirez-Lopez et al., 2014; Wetterlind and Stenberg, 2010), regional or national libraries demand thousands of samples to capture the full extent of soil variability. To cover the target regional scope, soil sampling location selection can be done using auxiliary data from remote sensing imagery and/or soil property legacy maps (Castaldi et al., 2019). When dealing with the geographical challenge, the extra challenge to achieve a balanced dataset (i.e., same number of samples with low, medium, and high contents) will naturally occur and need to be managed, since the natural frequency distribution of soil attribute contents is unbalanced, tending to concentrate on low values (e.g., nutrient contents in tropical soils) or high values (e.g., sand content in agricultural soils).
   Accordingly, one important aspect is ensuring that geographical coordinates are provided for every sample within the SSL. Including this information in the SSL allows not only for the evaluation of key soil-forming factors such as climate, topography, and parent material, but also for mapping soil types (classes) and attributes. This, in turn, facilitates the context-specific use of SSL data for any given area or use transfer learning strategies to improve local predictions (Viscarra Rossel et al., 2024).
8. **Documentation:** Although it does not directly affect model performance, documenting spectral metadata and modeling details is essential for transparency and data sharing in spectral services. This documentation should include sensor instrumental conditions, details of the spectral data acquisition protocol, the wet chemistry method used as a reference in modeling, and details of the modeling process—such as modeling approach, validation procedures, performance metrics, and descriptive statistics of the modeled attribute (e.g., number, range, and distribution of sample contents used in calibration and validation), along with the geographic location of the samples.

All sensor users, including service providers, should consider these eight keys factors when building and maintaining their SSL. We highly recommend that service providers be transparent in reporting the procedures used as well as metadata of instrumental conditions and reference analysis, such as details of sensor configurations and wet chemistry analytical protocols. Additionally, providing quantitative outputs accompanied by uncertainty estimates enhances transparency and fosters greater trust in the technology.

Information on the content of the analyte in the calibration set and, when available, additional sample data (e.g., location, parent material, biome, and land use) is also important to identify potential model limitations and define the geographical scope of applicability. Transparency from service providers, combined with independent auditing via proficiency testing, is key to enhancing the credibility and adoption of this technology. Key aspects of proficiency testing and its importance are discussed in the following sections.

Finally, it is important to address that field analyses using spectroscopy require even more meticulous attention due to external factors such as moisture, aggregate size, noise due to vibrations in case of on-the-go measurement, radiation source, soil-to-sensor geometry, vibration and foreign materials (e.g., straw, stones) (Mouazen et al., 2009). Extrapolating laboratory-calibrated models to *in situ* measurements demands separate strategies to address these external influences (Kuang et al., 2012). These considerations should be incorporated alongside the factors mentioned above; further details are discussed by Francos and Ben-Dor (2022).

3. Core aspects for proficiency testing

Proficiency tests adapted for spectroscopy-based methods are essential to ensure consistent performance and build user confidence (Nogueira, 2001). This concept mirrors the accreditation process for analytical laboratories, where their operational performance is regularly evaluated and benchmarked against other laboratories (see (Quaggio et al., 1994). However, proficiency tests for optical sensor-based technologies must be tailored to address specific aspects inherent to analytical inference using predictive models and SSL, such as its geographical scope.

In general, a proficiency test is based on sending homogeneous samples (previously defined by specific criteria) to several participating laboratories, which return their results to the test provider, where the results are used to obtain the assigned value (target value, usually defined by a measure of average behavior) and the acceptance limits for the performance classification of the participants (Thompson et al., 2006). A key advantage of sensor-based methods is that they are non-destructive, allowing the exact same sample to be analyzed by multiple laboratories, reducing the effects of micro-heterogeneities and increasing the robustness of comparisons. Additionally, some soil attributes, such as texture and organic carbon, maintain consistent content in the sample when properly stored. This feature enables the reuse of the same samples in proficiency tests conducted over consecutive years (Shepherd et al., 2025), facilitating the evaluation of temporal trends in model performance. Furthermore, this reduces maintenance costs for test providers, as the need for frequent sample updates is minimized.

It is important to note that ISO 17043 establishes general requirements for the competence and impartiality of proficiency testing providers, ensuring the consistent operation of proficiency testing schemes. The guidelines outlined below, focusing on the technical aspects of sensor-based methods, should be considered in alignment with these international standards. Regional proficiency testing providers can be either public or private bodies recognized by an accreditation mechanism (e.g., International Laboratory Accreditation Cooperation, ILAC) to guarantee that the tests are conducted in a robust, standardized, and impartial manner.

Below, we highlight core aspects for the development of regional proficiency tests.

3.1. Delimiting a target region

Using spectral libraries to predict soil properties may perform well in one region but fail in others if the library does not adequately represent the pedological and environmental variability of the target area. Delimiting the coverage area of the proficiency test by carrying out regional evaluations (e.g., following country divisions) makes it possible to deal with this aspect and identify regions with high and low predictive performance. Therefore, we recommend that proficiency tests be regional, so that their samples represent the variability of soils (matrices) in the region of interest.

3.2. Sample selection and number of samples

Robust performance assessments require a selection of representative data sets that capture the variability of the soil matrix (i.e., physical and chemical constituents) for the area in question, covering its full range of the analyte. Regions with greater soil variability demand more samples, as well as require advanced predictive modeling strategies. While determining the exact number of samples is challenging, pedogenic maps — which defines homogeneous areas based on soil-forming factors — can effectively guide a proper sample selection that minimally covers the region of interest (Styc et al., 2025).

As an initial recommendation, we suggest selecting samples that are as varied as possible in terms of soil types, parent materials (as they drastically influence soil matrix), depth of sampling, management practices, land uses, and geographical position. It is essential to avoid using too few samples (e.g., a few dozen), which may overestimate performance if the soil matrix is similar to the spectral library or underestimate it if the matrices differ significantly.

3.3. Determining the target value of the samples provided in the proficiency test

The homogeneous samples provided during proficiency testing should have a known target value for the evaluated attribute, along with its associated uncertainty. A rigorous approach to defining accurate and reliable target values with well-characterized uncertainties is crucial for the success of proficiency tests. This value can be determined either by procedures that use interlaboratory analysis results, as performed by Demattê et al. (2019), or by an independent and impartial laboratory with a recognized high standard of quality.

Obtaining the target value and its associated uncertainty from multiple certified laboratories on the market is a practical approach to mitigate natural systematic errors inherent in wet chemistry soil data (e.g., influenced by small variations within batches, reagents, and ambient temperature and humidity, etc.). In this context, it is essential to rigorously quantify the uncertainty of the target value—employing methods such as linear mixed-effects models or probability distribution functions (van Leeuwen et al., 2022)—to distinguish true soil variability from measurement errors.

3.4. Quality indicators of the model determination

During the model development process, it is essential to characterize the quality indicators of the prediction results in both the calibration and validation sets (Dardenne, 2010). However, for evaluating prediction performance within the context of proficiency testing, we recommend prioritizing metrics that provide a clear understanding of the model's accuracy profile on the selected sample set (Feinberg, 2007)—for example, calculated from the deviation of each sample's prediction from the target value.

Relying solely on metrics, such as coefficient of determination (R2), ratio of performance deviations (RPD), concordance correlation coefficient (CCC), and ratio of performance to interquartile range (RPIQ) can

Table 1
Comparison of metrics to measure the performance of models based in spectroanalytical sensors to predict soil attributes.

Metric	Definition	Strengths	Weaknesses
Root Mean Square Error (RMSE)	Measures the average magnitude of prediction error, giving higher weight to larger errors.	Widely used, provides an intuitive measure of error magnitude.	Sensitive to large errors, may not differentiate between systematic and random errors. Does not allow comparisons with reference analyzes with different unit and magnitude.
Coefficient of Determination (R^2)	Indicates the proportion of variance explained by the model; closer to 1 means better fit.	Useful for assessing overall model fit and comparing different models.	Can be misleading if the dataset has low variability; high R^2 does not always mean good predictive performance.
Ratio of Performance to Deviation (RPD)	Ratio of standard deviation of reference data to the model RMSE. Suggested by Williams (1987).	Accounts for variability in the dataset, making it useful for performance comparison between models, even using different units.	Can be biased if the dataset is too homogeneous and lacks variability; low RPD does not always mean poor predictions; the same measure as R^2.
Concordance Correlation Coefficient (CCC)	Measures the reproducibility of continuous measurements (Lin, 1989). It measures how much the fitted linear relationship deviates from the concordance line (accuracy) and how far each observation deviates from the fitted line (precision).	Accounts for both correlation and bias, making it more informative than Pearson's correlation alone.	Sensitive to data dispersion; can be misleading if the data range is too narrow. Requires normally distributed data.
Ratio of Performance to Interquartile Range (RPIQ)	Ratio of interquartile range of reference data to the model RMSE. Suggested by Bellon-Maurel et al. (2010).	Similar to RPD, however, more robust to outliers, providing a more reliable assessment in skewed datasets.	Can also be biased if the dataset is too homogeneous and lacks variability; low RPIQ does not always mean poor predictions.
Bias	Measures systematic error; a low absolute bias indicates a consistent model extrapolation.	Identifies systematic model bias, which is crucial for adjusting predictions.	Does not capture random error, only systematic bias. In addition, can produce a low value if the distribution of the residues (errors) are not random across the property predicted values.
Maximum Error	Largest absolute difference between observed and predicted values (in modulus); it informs the model's worst prediction.	Provides the model's worst prediction, useful for applications where extreme errors are critical.	It can be influenced by a single extreme outlier, which may not be representative of the model's overall performance.
Error Percentiles (e.g., 75th, 90th)	Value below which a certain percentage of the errors (in modulus) are contained. For example, 90 % of the samples had an error lower than that indicated by the 90th percentile.	It offers a more stable view of the distribution of errors, reducing the influence of outliers; it is useful for determining acceptable error limits in practical analyses.	It may not reflect the degree of extreme error and is inadequate for detecting serious faults; should be evaluated together with the average and maximum error.
Z-Score	Ratio of the deviation between prediction and the standard deviation for proficiency testing. Detailed by ISO 13528.	Internationally accepted in quality control programs and aligned with standards such as ISO. Easy to calculate and interpret.	It does not differentiate between precision and accuracy, preventing the investigation of causal factors. Sensitive to a poorly defined standard deviation, which may result in misinterpretations.

be misleading, as these metrics are influenced by the dataset's range (Wadoux and Minasny, 2024). Therefore, we suggest prioritizing descriptive statistics of the absolute error (i.e., the modulus of predicted minus measured values) as the primary quality indicators, along with bias and root mean squared error (RMSE). Table 1 presents a comparative overview of these metrics, outlining their definitions, strengths, and limitations.

Descriptive statistics of absolute error, such as the maximum error and selected percentiles (e.g., 75th, 90th), provide valuable insights into error behavior and model stability across the assessed sample set. For example, a small gap between the mean and the maximum error indicates higher consistency, while larger gaps suggest greater variability. These values can also be compared with maximum acceptable error thresholds, allowing us to understand what percentage of the sample has acceptable error for a given analytical tolerance, helping to understand whether the sensor-based method is suitable for a specific application (e.g., for attributes where the method is not yet widely accepted) (see Section below).

It is worth noting that the main metric used in proficiency testing schemes for chemical measurement—though still rarely applied by the proximal soil sensing community—is the z-score (Z), which compares the error obtained with the standard deviation for proficiency testing (S_P), using the following Eq. (1):

$$Z = \frac{(X_P - X_R)}{S_P} \tag{1}$$

where X_P is the predicted value and X_R is the reference value. Each sample, Z is interpreted as excellent ($|Z| \leq 1$), good ($1 < |Z| \leq 2$), questionable ($2 < |Z| \leq 3$) or unsatisfactory ($|Z| > 3$). This metric could be used to rank and classify the confidence level of service providers.

It is also important to highlight that, unlike conventional analytical methods that often assume a uniform uncertainty across samples, multivariate models produce an individual uncertainty estimate for each sample. Reporting these uncertainties alongside the predicted values is essential for attributing confidence to sensor-based results. Recent efforts in the literature aim to quantify this uncertainty (Omondiagbe et al., 2024), which should be encouraged as a good practice in both model evaluation and service reporting.

Similar to what was described above regarding absolute error, descriptive statistics of Z-scores (e.g., mean, percentiles, and maximum values) can also be used to assess model quality across the entire sample set. However, we recommend that the S_P be defined using the fitness-for-purpose criterion based on analytical tolerance (see ISO 13528), rather than being calculated from the results of that specific test round. The document published by (Analytical Methods Committee, 2015) outlines the main disadvantages of using an S_P derived from participant results, including:

(i) It allows most of the service providers to obtain a satisfactory score on most occasions, regardless of whether their uncertainties are sufficiently small to meet customer requirements;
(ii) It is unconsistent, since the observed standard deviation varies from round to round. This prevents individual participants from tracking performance over time, identifying trends, and determining whether corrective actions—such as changes in equipment or procedures—have been effective.

In contrast, the fitness-for-purpose criterion defines a target S_P based on the analytical tolerance (accuracy) required for decision-making (see Equation (2).

$$S_P = f \times AT \tag{2}$$

where f is an empirical factor (e.g., 0.5 for high-precision analysis and 1.0 for less rigorous applications), and AT is the analytical tolerance, i.e., the maximum acceptable error threshold.

3.5. Defining maximum error thresholds (analytical tolerance)

Establishing maximum acceptable error thresholds, referred to as analytical tolerance (AT), for specific applications helps users determine whether a sensor-based estimation service is suitable for a given application. This is a fitness-for-purpose validation strategy (Feinberg, 2007) that complements the information obtained from the model's quality indicators.

The AT value is not arbitrary but rather depends on the objectives of the analytical procedure. These thresholds must be based on technical rationales that consider local attribute variations and the level of accuracy required for the application. For example, the AT for mapping clay content for soil fertility management may be broader than that required for pedological classification (e.g., 6 g kg^{-1} and 10 g kg^{-1}, respectively).

Once AT for each application is defined, it can be used for SP calculation and then the Z-score, as described above. In addition, a suggested approach to evaluate the model consistency could be a classification by comparing descriptive statistics of the absolute error with the AT. For instance, when evaluating soil texture for fertilizer recommendations, if the AT for predicting clay content is set at 'X,' model consistency can be interpreted by comparing the descriptive statistics of the absolute error (e.g., 50th percentile, 75th percentile, 90th percentile, and maximum error, all observed in modulus) to this 'X' value, as follows:

- Inconsistent models: all errors (50th percentile, 75th percentile, 90th percentile, and maximum error) above the specified limit.
- Little consistent models: only 50th percentile below the limit.
- Reasonably consistent models: 50th percentile and 75th percentile below the limit.
- Consistent models: 50th percentile, 75th, and 90th percentiles below the limit.
- Very consistent models: all error metrics (50th percentile, 75th percentile, 90th percentile, and maximum error) below the limit.

This evaluation approach would consider samples with the most limiting errors, rather than relying solely on average values (e.g., RMSE). Those interpretations of consistency can be adjusted on the application. The key aspect is that this metric allows us to understand how many samples fall within or exceed the established limit, which helps dealing with the classic dilemma of model interpretation: "predict the average samples well or predict the extreme samples best". Fig. 2 illustrates the idea behind the suggested interpretation of consistency.

Although promising, this type of evaluation approach is still rarely applied in soil spectroscopy. Establishing clear AT values and promoting their adoption by the community would be essential steps toward making sensor-based analysis outputs more interpretable and supporting its broader adoption.

4. The importance of regional proficiency tests

The absence of standardized regional proficiency testing for sensor-based soil analysis threatens the reliability and credibility of these emerging technologies. Without independent certification of performance, service providers may lack proper quality control. Proficiency testing ensures measurements stay within acceptable error margins, enabling laboratories and service providers to detect and correct systematic errors. Failing to implement such measures risks undetected biases, compromising the validity of soil assessments.

In addition to checking performance, laboratories can also use proficiency testing as an opportunity to understand their uncertainties, improve internal protocols, and seek external support when needed. The network of participating laboratories and proficiency test providers can serve as a valuable source of information, for example through annual meetings for technical discussion. In this sense, proficiency testing

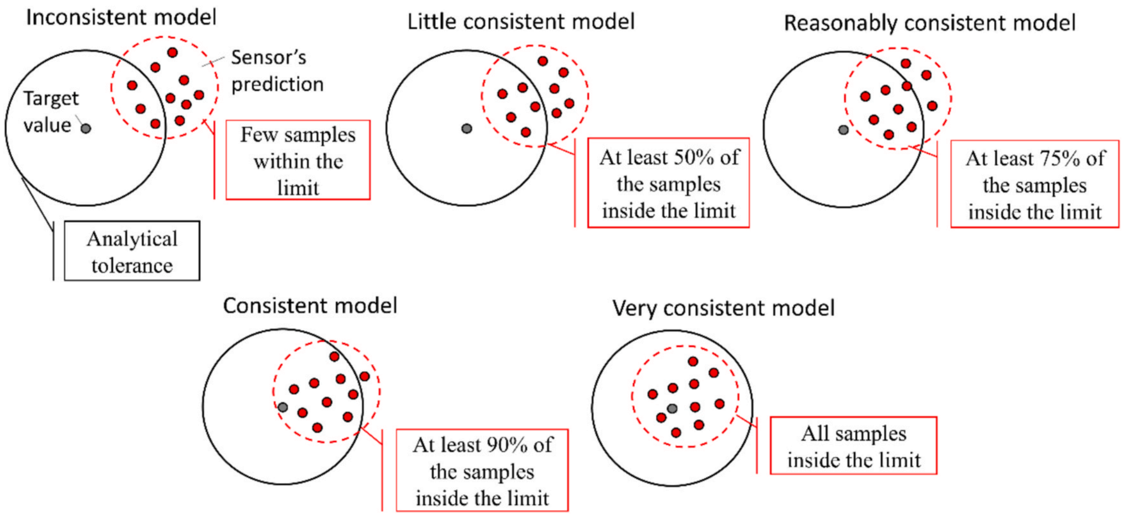

Fig. 2. Illustration of the interpretation of the model's consistency based on a comparison of the error of the predicted values with the analytical tolerance (AT). The reference value is represented with gray dots, the predicted values with red dots and the AT with a black circle.

should not be seen merely as a compliance system, but rather as a constructive and collaborative environment that fosters continuous improvement and shared learning, which can also be supported by scientists developing joint research to overcome some possible performance issues.

If end users lose confidence in a sensing technology due to poor performance, it hinders adoption and creates skepticism toward future innovations, even those proven accurate and reliable. For this reason, it is also important to engage researchers, service providers, and sensor manufacturers in the implementation of proficiency testing. Their technical and financial contributions can strengthen the development of spectral libraries, standard samples, and training initiatives, helping to raise the overall quality and trustworthiness of the sensor-based analyses. Therefore, implementing regional proficiency tests is crucial for sensor-based soil analysis to achieve the performance demonstrated in decades of scientific research, fostering stakeholder confidence.

5. Final considerations

Sensor-based soil analysis using spectroanalytical techniques has the potential to optimize and expand soil characterization and monitoring, offering faster, cost-effective, environmentally friendly and more sustainable alternatives to traditional methods. The global significance of this approach has been increasingly recognized, with a major milestone being the establishment of the Global Soil Analysis by Soil Spectroscopy (GLOSOLAN-Spec) in 2020 under the Food and Agriculture Organization (FAO). This initiative aims to advance the development and adoption of soil spectroscopy by fostering international collaboration (Peng et al., 2025).

Collaboration among researchers, policymakers, industry, and end users of the results will be essential in refining spectral libraries, improving calibration models, and developing global benchmarking systems. Cross-institutional partnerships will further facilitate data sharing and enhance predictive accuracy, ultimately improving the reliability of sensor-based methods (Safanelli et al., 2025).

To address the current lack of standardized procedures, Ben Dor et al. (2024) proposed an international protocol for soil analysis using sensors, titled Standardization and Protocol Development for Soil Spectral Reflectance in Laboratory and Field Settings, under the umbrella of the IEEE Standards Association (SA)—Working Group P4005, focusing on DRS sensors. For most other sensors with increasing adoption (e.g, XRF and LIBS), standardization procedures like this one are still needed. Establishing official standards and protocols through recognized legal Institutions is a critical step toward enhancing data quality and building credibility and strengthening user trust. Once these methods receive formal recognition, implementing proficiency testing will be an essential next step to ensure consistency and analytical accuracy across sensor-based service providers.

Although these measures will enhance the credibility of spectral soil analysis, challenges persist, particularly in defining tolerable error thresholds for each different application. A concerted effort and collaboration among the scientific community, regulatory bodies, and industry stakeholders is crucial to ensure that these technologies provide reliable, high-quality soil information to support decision making for sustainable land management.

Advances in proficiency testing for sensor-based soil analysis will also be fundamental to expanding the use of sensors and unlocking large-scale soil health and/or carbon farming programs (e.g., Cherubin et al., 2024) that require high-resolution data, either spatially or temporally. The cost of soil monitoring (using traditional wet chemistry methods) is one of the main barriers to these programs, since large amounts of samples are needed. Therefore, cost-effective methods, such as sensor-based soil analysis technologies that are accepted by international standards, have tremendous potential in the global marketplace over the coming decades.

Finally, to fully realize the potential of sensor-based soil analysis, it is important to maintain long-term engagement with end users, particularly farmers, by delivering reliable results for actionable, spatially explicit outputs such as soil maps, change detection, and diagnostic insights. Delivering such spatial products, rather than isolated point-based data, increases relevance in real-world decision-making. In parallel, integrating data from laboratory, field-based, and remote sensing platforms will further expand the applicability and impact of sensor-based methods. This multidimensional approach not only enhances the quality of insights delivered to users but also supports broader acceptance, enabling informed and site-specific soil management decisions.

Funding

Tiago R. Tavares was funded by São Paulo Research Foundation (FAPESP), grant number 2020/16670-9. Jose Demattê, Jose Lavres, Maurício Cherubin, and Sérgio Henrique Godinho Silva thank the CNPq for the research fellowship (grant numbers 307190/2021-8; 303718/2020-0; 302249/2025-7; 307532/2022-4.

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.

Acknowledgments

We would like to thank Prof. Celio Pasquini, who kindly read and helped polish the text. We also thank the support from the Center for Carbon Research in Tropical Agriculture (CCARBON) at the University of São Paulo, sponsored by the São Paulo Research Foundation (FAPESP) under grant 2021/10573-4.

References

Analytical Methods Committee, A., 2015. Fitness for purpose: the key feature in analytical proficiency testing. Anal. Methods 7, 7404–7405. https://doi.org/10.1039/C5AY90052B.

Barra, I., Haefele, S.M., Sakrabani, R., Kebede, F., 2021. Soil spectroscopy with the use of chemometrics, machine learning and pre-processing techniques in soil diagnosis: recent advances–a review. TrAC Trends Anal. Chem. 135, 116166. https://doi.org/10.1016/j.trac.2020.116166.

Bartsch, B.D., Rosin, N.A., dos Santos, U.J., Coblinski, J.A., Pelegrino, M.H.P., Rosas, J. T., Poppiel, R.R., Ortiz, E.B., Kochinki, V.C., Gallo, P., Dor, E.B., Falcioni, R., Nanni, M.R., Gonçalves, J.V., Oliveira, C.A., Vedana, N.G., 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. https://doi.org/10.3390/rs16234543.

Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.-M., McBratney, A., 2010. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. TrAC Trends Anal. Chem. 29, 1073–1081. https://doi.org/10.1016/j.trac.2010.05.006.

Ben-Dor, E., 2002. Quantitative remote sensing of soil properties. Adv. Agron. 75, 173–243. https://doi.org/10.1016/S0065-2113(02)75005-0.

Ben Dor, E., Karyotis, K., Chabrillat, S., Schmid, T., 2024. Standardization and protocol development for soil spectral reflectance in laboratory and field settings: an overview of IEEE SA P4005 activity. In: IGARSS 2024–2024 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 268–270. https://doi.org/10.1109/IGARSS53475.2024.10641689.

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.

Biswas, A., 2024. Monitor soil health using advanced technologies. Nature 631, 740. https://doi.org/10.1038/d41586-024-02396-4.

Castaldi, F., Chabrillat, S., van Wesemael, B., 2019. Sampling strategies for soil property mapping using multispectral sentinel-2 and hyperspectral EnMAP satellite data. Remote Sens. https://doi.org/10.3390/rs11030309.

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, 113855. https://doi.org/10.1016/j.geoderma.2019.07.013.

Chatterjee, S., Hartemink, A.E., Triantafilis, J., Desai, A.R., Soldat, D., Zhu, J., Townsend, P.A., Zhang, Y., Huang, J., 2021. Characterization of field-scale soil variation using a stepwise multi-sensor fusion approach and a cost-benefit analysis. Catena 201, 105190. https://doi.org/10.1016/j.catena.2021.105190.

Cherubin, M.R., Pinheiro Junior, C.R., Alves, L.A., Bayer, C., Cerri, C.E.P., Barioni, L.G., Peper, A., Anselmi, A.A., 2024. Carbon farming initiative: a national-scale public-private partnership to promote regenerative agriculture in Brazil. Exp. Agric. 60, e28. https://doi.org/10.1017/S0014479724000255.

Dardenne, P., 2010. Some considerations about NIR spectroscopy: closing speech at NIR-2009. NIR News 21, 8–14. https://doi.org/10.1255/nirn.1165.

Demattê, J.A.M., Dotto, A.C., Bedin, L.G., Sayão, V.M., Souza, A.B.E., 2019a. 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., Dotto, A.C., Paiva, A.F.S., Sato, M.V., Dalmolin, R.S.D., Maria do Socorro, B., da Silva, E.B., Nanni, M.R., ten Caten, A., Noronha, N.C., 2019b. The Brazilian Soil Spectral Library (BSSL): a general view, application and challenges. Geoderma 354, 113793.

Demattê, J.A.M., Paiva, A.F., Poppiel, R.R., Rosin, N.A., Ruiz, L.F., Mello, F.A., Minasny, B., Grunwald, S., Ge, Y., Ben Dor, E., Gholizadeh, A., Gomez, C., Chabrillat, S., Francos, N., Ayoubi, S., Fiantis, D., Biney, J.K., Wang, C., Belal, A., Naimi, S., Hafshejani, N.A., Bellinaso, H., Moura-Bueno, J.M., Silvero, N.E.Q., 2022. The Brazilian soil spectral service (BraSpecS): a user-friendly system for global soil spectra communication. Remote Sens. https://doi.org/10.3390/rs14030740.

Evangelista, S.J., Field, D.J., McBratney, A.B., Minasny, B., Ng, W., Padarian, J., Román Dobarco, M., Wadoux, A.-M.-J.-C., 2024. Soil security—strategizing a sustainable future for soil. Adv. Agron. 183, 1–70. https://doi.org/10.1016/bs.agron.2023.10.001.

Feinberg, M., 2007. Validation of analytical methods based on accuracy profiles. J. Chromatogr. A 1158, 174–183. https://doi.org/10.1016/j.chroma.2007.02.021.

da Fonseca, A.A., Pasquini, C., Cristina Costa, D., Mercês Barros Soares, E., 2022. Effect of the sample measurement representativeness on soil carbon determination using near-infrared compact spectrophotometers. Geoderma 409, 115636. https://doi.org/10.1016/j.geoderma.2021.115636.

Francos, N., Ben-Dor, E., 2022. A transfer function to predict soil surface reflectance from laboratory soil spectral libraries. Geoderma 405, 115432. https://doi.org/10.1016/j.geoderma.2021.115432.

Kuang, B., Mahmood, H.S., Quraishi, M.Z., Hoogmoed, W.B., Mouazen, A.M., van Henten, E.J., 2012. Chapter four - sensing soil properties in the laboratory, in situ, and on-line: a review. In: Sparks, D.L.B.T.-A. (Ed.), Advances in Agronomy. Academic Press, pp. 155–223. https://doi.org/10.1016/B978-0-12-394275-3.00003-1.

Lin, L.-I.-K., 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics 45, 255. https://doi.org/10.2307/2532051.

Lucà, F., Conforti, M., Castrignanò, A., Matteucci, G., Buttafuoco, G., 2017. Effect of calibration set size on prediction at local scale of soil carbon by Vis-NIR spectroscopy. Geoderma 288, 175–183. https://doi.org/10.1016/j.geoderma.2016.11.015.

Mancini, M., Andrade, R., Silva, S.H.G., Rafael, R.B.A., Mukhopadhyay, S., Li, B., Chakraborty, S., Guilherme, L.R.G., Acree, A., Weindorf, D.C., Curi, N., 2024. Multinational prediction of soil organic carbon and texture via proximal sensors. Soil Sci. Soc. Am. J. 88, 8–26. https://doi.org/10.1002/saj2.20593.

McBride, M.B., 2022. Estimating soil chemical properties by diffuse reflectance spectroscopy: promise versus reality. Eur. J. Soil Sci. 73, e13192. https://doi.org/10.1111/ejss.13192.

Minasny, B., Bandai, T., Ghezzehei, T.A., Huang, Y.-C., Ma, Y., McBratney, A.B., Ng, W., Norouzi, S., Padarian, J., Rudiyanto, Sharififar, A., Styc, Q., Widyastuti, M., 2024. Soil science-informed machine learning. Geoderma 452, 117094. https://doi.org/10.1016/j.geoderma.2024.117094.

Mouazen, A.M., Maleki, M.R., Cockx, L., Van Meirvenne, M., Van Holm, L.H.J., Merckx, R., De Baerdemaeker, J., Ramon, H., 2009. Optimum three-point linkage set up for improving the quality of soil spectra and the accuracy of soil phosphorus measured using an on-line visible and near infrared sensor. Soil Tillage Res. 103, 144–152. https://doi.org/10.1016/j.still.2008.10.006.

Murray, I., Cowe, I., 2015. Sample preparation. In: Near-Infrared Spectroscopy in Agriculture, Agronomy Monographs. pp. 75–112. https://doi.org/10.2134/agronmonogr44.c5.

Ng, W., Minasny, B., Jeon, S.H., McBratney, A., 2022a. Mid-infrared spectroscopy for accurate measurement of an extensive set of soil properties for assessing soil functions. Soil Secur. 6, 100043. https://doi.org/10.1016/j.soisec.2022.100043.

Ng, W., Minasny, B., Jones, E., McBratney, A., 2022b. To spike or to localize? Strategies to improve the prediction of local soil properties using regional spectral library. Geoderma 406, 115501. https://doi.org/10.1016/j.geoderma.2021.115501.

Nogueira, J., 2001. EPTIS: the new European database of proficiency testing schemes for analytical laboratories. TrAC Trends Anal. Chem. 20, 457–461. https://doi.org/10.1016/S0165-9936(01)00091-7.

Omondiagbe, O.P., Roudier, P., Lilburne, L., Ma, Y., McNeill, S., 2024. Quantifying uncertainty in the prediction of soil properties using mid-infrared spectra. Geoderma 448, 116954. https://doi.org/10.1016/j.geoderma.2024.116954.

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.

Paiva, A.F. da S., Poppiel, R.R., Rosin, N.A., Greschuk, L.T., Rosas, J.T.F., Demattê, J.A. M., 2022. The Brazilian program of soil analysis via spectroscopy (ProBASE): combining spectroscopy and wet laboratories to understand new technologies. Geoderma 421, 115905. https://doi.org/10.1016/j.geoderma.2022.115905.

Pasquini, C., 2018. Near infrared spectroscopy: a mature analytical technique with new perspectives – a review. Anal. Chim. Acta 1026, 8–36. https://doi.org/10.1016/j.aca.2018.04.004.

Peng, Y., Ben-Dor, E., Biswas, A., Chabrillat, S., Demattê, J.A.M., Ge, Y., Gholizadeh, A., Gomez, C., Guerrero, C., Herrick, J., Maynard, J.J., Mouazen, A.M., Ma, Y., McBratney, A.B., Minasny, B., Ramirez-Lopez, L., Robertson, A.H.J., Viscarra Rossel, R.A., Shi, Z., Stenberg, B., Wadoux, A.-M.-J.-C., Winowiecki, L.A., Zhang, G., 2025. Spectroscopic solutions for generating new global soil information. Innov, 100839. https://doi.org/10.1016/j.xinn.2025.100839.

Piikki, K., Söderström, M., Eriksson, J., Muturi John, J., Ireri Muthee, P., Wetterlind, J., Lund, E., 2016. Performance evaluation of proximal sensors for soil assessment in smallholder farms in Embu county, Kenya. Sensors 16, 1950. https://doi.org/10.3390/s16111950.

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, 116029. https://doi.org/10.1016/j.geoderma.2022.116029.

Quaggio, J.A., Cantarella, H., van Raij, B., 1994. Evolution of the analytical quality of soil testing laboratories integrated in a sample exchange program. Commun. Soil Sci. Plant Anal. 25, 1007–1014. https://doi.org/10.1080/00103629409369094.

Ramirez-Lopez, L., Schmidt, K., Behrens, T., van Wesemael, B., Demattê, J.A.M., Scholten, T., 2014. Sampling optimal calibration sets in soil infrared spectroscopy. Geoderma 226–227, 140–150. https://doi.org/10.1016/j.geoderma.2014.02.002.

Rinnan, Å., van den Berg, F., Engelsen, S.B., 2009. Review of the most common pre-processing techniques for near-infrared spectra. TrAC Trends Anal. Chem. 28, 1201–1222. https://doi.org/10.1016/j.trac.2009.07.007.

Safanelli, J.L., Hengl, T., Parente, L.L., Minarik, R., Bloom, D.E., Todd-Brown, K., Gholizadeh, A., Mendes, W. de S., Sanderman, J., 2025. Open Soil Spectral Library (OSSL): building reproducible soil calibration models through open development and community engagement. PLoS One 20, e0296545. https://doi.org/10.1371/journal.pone.0296545.

Seidel, M., Hutengs, C., Ludwig, B., Thiele-Bruhn, S., Vohland, M., 2019. Strategies for the efficient estimation of soil organic carbon at the field scale with vis-NIR spectroscopy: spectral libraries and spiking vs. local calibrations. Geoderma 354, 113856. https://doi.org/10.1016/j.geoderma.2019.07.014.

Shepherd, J.E., Kanner, O., Amir, O., Efrati, B., Ben-Dor, E., 2025. Long-term stability of soil spectral libraries with chemical and spectral insights. Sci. Rep. 15, 9068. https://doi.org/10.1038/s41598-025-93792-x.

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.

Shi, X., Song, J., Wang, H., Lv, X., Zhu, Y., Zhang, W., Bu, W., Zeng, L., 2023. Improving soil organic matter estimation accuracy by combining optimal spectral preprocessing and feature selection methods based on pXRF and vis-NIR data fusion. Geoderma 430, 116301. https://doi.org/10.1016/j.geoderma.2022.116301.

Silva, S.H.G., Ribeiro, B.T., Guerra, M.B.B., de Carvalho, H.W.P., Lopes, G., Carvalho, G. S., Guilherme, L.R.G., Resende, M., Mancini, M., Curi, N., Rafael, R.B.A., Cardelli, V., Cocco, S., Corti, G., Chakraborty, S., Li, B., Weindorf, D.C., 2021. pXRF in tropical soils: methodology, applications, achievements and challenges. Adv. Agron. 167, 1–62. https://doi.org/10.1016/bs.agron.2020.12.001.

Silva, S.H.G., Weindorf, D.C., Pinto, L.C., Faria, W.M., Acerbi Junior, F.W., Gomide, L.R., de Mello, J.M., de Pádua Junior, A.L., de Souza, I.A., Teixeira, A.F. dos S., Guilherme, L.R.G., Curi, N., 2020. Soil texture prediction in tropical soils: a portable X-ray fluorescence spectrometry approach. Geoderma 362, 114136. https://doi.org/10.1016/j.geoderma.2019.114136.

Stenberg, B., Koganti, T., Castaldi, F., Metzger, K., Buttafuoco, G., van Egmond, F., Cayuela, J.A., Borůvka, L., Debaene, G., Liebisch, F., Sandén, T., 2024. D5.1 ProbeField: best practice protocol for field spectroscopy and assessment by soil spectral library based calibrations. https://doi.org/10.5281/zenodo.14150972.

Stenberg, B., Viscarra Rossel, R.A., Mouazen, A.M., Wetterlind, J., 2010. Visible and near infrared spectroscopy in soil science. Adv. Agron. 107, 163–215. https://doi.org/10.1016/S0065-2113(10)07005-7.

Styc, Q., Pachon, J., Ng, W., Padarian, J., McBratney, A., 2025. Creating soil districts for Australia based on pedogenon mapping. Geoderma 454, 117164. https://doi.org/10.1016/j.geoderma.2025.117164.

Tavares, T.R., Minasny, B., McBratney, A., Cherubin, M.R., Marques, G.T., Ragagnin, M. M., Alves, E.E.N., Padarian, J., Lavres, J., de Carvalho, H.W.P., 2023. Estimating plant-available nutrients with XRF sensors: towards a versatile analysis tool for soil condition assessment. Geoderma 439, 116701. https://doi.org/10.1016/j.geoderma.2023.116701.

Tavares, T.R., Molin, J.P., Nunes, L.C., Alves, E.E.N., Melquiades, F.L., de Carvalho, H.W. P., Mouazen, A.M., de Carvalho, H.W.P., Mouazen, A.M., 2020. Effect of X-ray tube configuration on measurement of key soil fertility attributes with XRF. Remote Sens. 12, 963. https://doi.org/10.3390/rs12060963.

Teixeira, A.F. dos S., Andrade, R., Mancini, M., Silva, S.H.G., Weindorf, D.C., Chakraborty, S., Guilherme, L.R.G., Curi, N., 2022. Proximal sensor data fusion for tropical soil property prediction: soil fertility properties. J. S. Am. Earth Sci. 116, 103873. https://doi.org/10.1016/j.jsames.2022.103873.

Thompson, M., Ellison, S.L.R., Wood, R., 2006. The International Harmonized Protocol for the proficiency testing of analytical chemistry laboratories (IUPAC technical report). Pure Appl. Chem. Pure Appl. Chem. 78, 145–196. https://doi.org/10.1351/pac200678010145.

USEPA, 2007. Method 6200: Field portable X-ray Fluorescence Spectrometry for the Determination of Elemental Concentrations in Soil and Sediment.

van Leeuwen, C.C.E., Mulder, V.L., Batjes, N.H., Heuvelink, G.B.M., 2022. Statistical modelling of measurement error in wet chemistry soil data. Eur. J. Soil Sci. 73. https://doi.org/10.1111/ejss.13137.

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-Sci. Rev. 155, 198–230. https://doi.org/10.1016/j.earscirev.2016.01.012.

Viscarra Rossel, R.A., Shen, Z., Ramirez Lopez, L., Behrens, T., Shi, Z., Wetterlind, J., Sudduth, K.A., Stenberg, B., Guerrero, C., Gholizadeh, A., Ben-Dor, E., St Luce, M., Orellano, C., 2024. An imperative for soil spectroscopic modelling is to think global but fit local with transfer learning. Earth-Sci. Rev. 254, 104797. https://doi.org/10.1016/j.earscirev.2024.104797.

Wadoux, A.-M.-J.-C., 2023. Interpretable spectroscopic modelling of soil with machine learning. Eur. J. Soil Sci. 74, e13370. https://doi.org/10.1111/ejss.13370.

Wadoux, A.-M.-J.-C., Minasny, B., 2024. Some limitations of the concordance correlation coefficient to characterise model accuracy. Ecol. Inform. 83, 102820. https://doi.org/10.1016/j.ecoinf.2024.102820.

Weindorf, D.C., Chakraborty, S., 2016. Portable X-ray fluorescence spectrometry analysis of soils. Methods Soil Anal. 1, 1384–1392. https://doi.org/10.2136/methods-soil.2015.0033.

Wetterlind, J., Stenberg, B., 2010. Near-infrared spectroscopy for within-field soil characterization: small local calibrations compared with national libraries spiked with local samples. Eur. J. Soil Sci. 61, 823–843. https://doi.org/10.1111/j.1365-2389.2010.01283.x.

Williams, P.C., 1987. Variables affecting near-infrared reflectance spectroscopic analysis. In: Williams, P., Norris, K. (Eds.), Near-Infrared Technology in the Agricultural and Food Industries. American Association of Cereal Chemists, Saint Paul, MN, pp. 143–167.

Zhu, Y., Weindorf, D.C., Zhang, W., 2011. Characterizing soils using a portable X-ray fluorescence spectrometer: 1. Soil texture. Geoderma 167–168, 167–177. https://doi.org/10.1016/j.geoderma.2011.08.010.