Simplifying lab analysis for mapping texture and OM content via sensor-based inference: a case study showing maps of ECa, crop yield, and traditional methods

Marcello Franchi[1]; Enzo Casciello[2]; João Vítor Fiolo Pozzuto[3]; Tiago Rodrigues Tavares[4]; Luiza M. A. L. da Silva[5]; Eudocio Rafael Otavio da Silva[6]; Hudson Wallace Pereira de Carvalho[7]

[1] Eng. Agrônomo, GeoMec - Pesquisa e Consultoria, Sorriso - MT. mra.franchi@gmail.com;
[2] Eng. Agrônomo, Fazenda Serra Branca, Balsas - Maranhão. enzo@sb.agr.br;
[3] Eng. Agrícola, Lab. de Instrumentação Nuclear, CENA, USP, Piracicaba – SP;
[4] Eng. Agrônomo, Pesquisador Pós-Doc, CCARBON, ESALQ, USP, Piracicaba - SP;
[5] Eng. Agrícola e Ambiental, 3R-Ribersolo, Ribeirão Preto – SP;
[6] Eng. Agrícola e Ambiental, Pesquisador Pós-Doc, CCARBON, ESALQ, USP, Piracicaba - SP;
[7] Químico, Professor - Lab. de Instrumentação Nuclear, CENA, USP, Piracicaba – SP

A paper from the Proceedings of the

17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture

13-15 July 2026

Porto Alegre, Brazil

Abstract.

Precision agriculture hinges on soil information at spatial resolutions that capture within-field variability, yet conventional laboratory workflows often constrain sampling density due to cost and turnaround time. In this study, we evaluated a laboratory-based sensor inference service for mapping soil organic matter (OM) and texture (clay) using visible and near-infrared spectroscopy (vis-NIR) and X-ray fluorescence (XRF) spectroscopy. We further evaluated whether the resulting maps are consistent with those derived from traditional laboratory analyses and with apparent electrical conductivity (ECa), an on-the-go sensor frequently related to textural variability in Brazilian soils. Texture was predicted from XRF data, whereas OM was predicted using vis-NIR+XRF data fusion. Two calibration strategies were compared: (i) a "global" approach using a generalist model trained on a national spectral library and (ii) a "local" approach combining field-specific samples with selected national samples. Local modeling consistently improved accuracy compared with global models. For OM, the local model achieved $R^2 = 0.91$, RMSE = 2.7 g kg^{-1}, and bias = −0.4 g kg^{-1}, while the global model achieved $R^2 = 0.81$, RMSE = 6.9 g kg^{-1}, and bias = 4.7 g kg^{-1}. For clay, the

The authors are solely responsible for the content of this paper, which is not a refereed publication. Citation of this work should state that it is from the Proceedings of the 17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture.Franchi, M.; Casciello, E.; Pozzuto, J.V.F.; Tavares, T.R.; Silva, L.M.A.L.; Silva, E.R.O.; Carvalho, H.W.P. (2026). Simplifying lab analysis for mapping texture and OM content via sensor-based inference: a case study showing maps of ECa, crop yield, and traditional methods. In Proceedings of the 17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture (unpaginated, online). Monticello, IL: International Society of Precision Agriculture and Brazilian Congress on Precision and Digital Agriculture.

local model achieved $R^2 = 0.67$, RMSE = 52.3 g kg^{-1}, and bias = −13.1 g kg^{-1}, whereas the global model achieved $R^2 = 0.63$, RMSE = 94.8 g kg^{-1}, and bias = −76.6 g kg^{-1}. For sand, the local model achieved $R^2 = 0.72$, RMSE = 52.5 g kg^{-1}, and bias = -15.3 g kg^{-1}, while the global model achieved $R^2 = 0.68$, RMSE = 69.2 g kg^{-1}, and bias = 45.5 g kg^{-1}. Despite their higher bias, global models reproduced spatial patterns similar to those derived from laboratory analyses, especially for clay, whereas the global OM model showed weaker spatial agreement due to the overestimation of a specific group of samples. Apparent electrical conductivity (ECa) showed only weak-to-moderate agreement with clay spatial variability, likely because its readings were influenced by site-specific management or other factors beyond texture. Soybean yield showed low and non-significant relationships with the evaluated soil properties. Overall, these results provide field-scale evidence that laboratory-based proximal sensing, particularly when coupled with local spiking and multisensor fusion, can monitor OM and texture variability while maintaining consistency at a lower analytical cost than conventional workflows and without the use of chemical reagents.

Keywords.

Predictive modeling, fertility diagnostics, proximal sensing, geostatistics

Introduction

Precision agriculture depends on soil information collected at spatial resolutions that represent within-field variability. However, conventional laboratory analysis often limits sampling density because it is costly, time-consuming, and dependent on chemical reagents (Molin & Tavares, 2019). These constraints are particularly relevant in intensive agricultural systems, such as those with two crop seasons per year, where decisions must be made within short operational windows. Sensor-based soil analysis, conducted both in the lab and/or in the field, has therefore emerged as a practical complementation to increase analytical throughput, reduce costs, and support spatially explicit soil management.

Among proximal sensing techniques (Viscarra Rossel et al., 2011), visible and near-infrared spectroscopy (vis-NIR) and X-ray fluorescence (XRF) spectroscopy are particularly promising for soil characterization (Javadi et al., 2021; Xu et al., 2019). Vis-NIR is primarily sensitive to organic and mineral absorption features associated with soil organic matter, clay minerals, water, and iron oxides (Ben-Dor et al., 2015), whereas XRF provides elemental information from the inorganic soil matrix, including elements linked to texture and mineral composition. Because these techniques capture different but complementary soil properties, their combined use can improve the inference of complex attributes such as organic matter (OM) (Shi et al., 2023), while XRF alone can provide useful information for texture prediction (Tavares et al., 2020). This complementarity makes vis-NIR and XRF suitable for laboratory-based sensor inference services aimed at mapping key soil attributes with lower cost and faster turnaround than conventional methods. Apart from the potential number of characterizable properties, these techniques are particularly interesting because they require minimal sample preparation for proper utilization, namely drying and sieving (< 2mm) (Ben-Dor et al., 2015; Tavares et al., 2019), and can be applied together in analytical laboratories as sensor-based inference services (Tavares et al., 2025).

The development of soil spectral libraries has expanded the potential of these techniques beyond small experimental datasets. Large spectral libraries enable the calibration of generalist models that predict soil attributes across broad geographic and pedological conditions (Viscarra Rossel et al., 2016). In Brazil, a national-scale library developed in a lab-based environment has shown potential for predicting attributes such as OM and texture (Demattê et al., 2019), supporting their use in commercial analytical workflows and digital soil mapping. Nevertheless, broad-scale models may not fully represent local soil variability, especially when field-specific mineralogical, textural, or management conditions differ from those included in the calibration set. Under these conditions, local modeling strategies can improve prediction accuracy (Ng et al., 2022).

One practical approach to improve local predictions is by enriching the global spectral library, or part of it, with local samples. This strategy, often referred to as local spiking (Guerrero et al., 2014), in theory preserves the advantage of large spectral datasets while increasing the representativeness of the calibration set for the target field. In this context, global models can provide a scalable first approximation, whereas localized models can reduce bias and improve the spatial reliability of predictions for farm-level mapping.

Despite growing evidence of the predictive capacity of vis-NIR and XRF models, most studies evaluate sensor-based inference primarily using non-specific metrics such as the coefficient of determination (R^2) and the ratio of performance to interquartile distance (RPIQ). For precision agriculture (PA) applications, however, the spatial consistency of sensor-derived maps is equally important. Comparing maps generated from sensor-based predictions with those obtained from conventional laboratory analysis, apparent electrical conductivity (ECa), and crop yield can provide a more practical assessment of whether sensor inference captures meaningful field variability.

In this context, the present study aimed to evaluate whether sensor-based inference (using XRF and Vis-NIR libraries) for soil texture and organic matter achieves mapping performance comparable to conventional methods (wet chemistry). Specifically, outputs for OM and texture prediction, derived from global and local modeling strategies, were evaluated on an independent local dataset against traditional wet-chemistry tests. In addition, the resulting maps were also assessed against conventional laboratory maps, as well as ECa and soybean yield maps.

Material and Methods

National spectral library for texture and OM prediction

Two national spectral libraries were used in this study. Clay prediction was performed using an XRF spectral library composed of approximately 4,437 Brazilian soil samples with reference texture values. OM prediction was performed using a larger national library composed of approximately 12,800 soil samples with both vis-NIR and XRF spectra available.

Under laboratory conditions, XRF data were acquired using a portable energy-dispersive spectrometer (S1 TITAN Model 800, Bruker AXS, USA) equipped with a Rh X-ray tube and a Silicon Drift Detector (Bruker, USA), and vis-NIR measurements were performed using a benchtop spectrometer (Model SM-3500, Spectral Evolution, USA), covering 350–2500 nm at a resolution of < 8.0 nm. Vis-NIR spectra were reduced to the 380–2480 nm range and converted from reflectance to absorbance using log(1/R). XRF spectra were organized as intensity channels and transformed to logarithmic intensity.

For OM prediction, vis-NIR and XRF spectra were standardized to the minimum and maximum values from the national calibration database, then concatenated into a single predictor matrix. This preprocessing step normalized both sensors to a comparable intensity scale and enabled the combined use of organic/mineral information from vis-NIR and elemental information from XRF. For texture prediction, only XRF spectra were used due to their direct relationship to elemental composition (Tavares et al., 2020).

Reference analysis of OM was performed from organic carbon measured by the Walkley–Black method, following the standard procedure adopted by the commercial soil laboratory. Soil texture was determined using the pipette method, and clay content was the primary textural variable evaluated in this study. These laboratory results were used as reference values for evaluating the sensor-based predictions.

The global clay model was calibrated using XRF spectra (n = 4,437), whereas the global OM model was calibrated using the concatenated vis-NIR + XRF matrix (n = 12,829). Cubist regression was used as the generalist modeling algorithm because it is well-suited to high-dimensional spectral datasets and can produce rule-based local regressions for groups of samples with similar spectral behavior (Minasny and McBratney, 2008). Thus, the global models represented a generalist prediction strategy based on broad-scale spectral information from the national libraries.

Calibration of local models for texture and OM prediction

The farm evaluated in this study comprises a commercial agricultural field located in the municipality of Balsas, Maranhao State, Brazil, covering 3500 ha used for grain production. A total of 99 soil samples were selected for XRF scanning and reference texture analysis, whereas 121 soil samples were selected for vis-NIR+XRF spectral acquisition and reference OM analysis. These datasets were used to build the local calibration databases for texture and OM prediction, respectively.

The local spectra were processed using the same transformations and standardization parameters applied to the national spectral libraries, ensuring compatibility between local and national spectral data. The local databases were then combined with their respective national libraries to identify national samples that were spectrally similar to the field samples.

Spectral similarity was assessed in the principal component space of the spectral data. Mahalanobis distances were calculated among samples in this reduced spectral space, and one spectrally similar national sample was selected for each local sample. The selected national samples were then added to the corresponding local calibration set. This procedure generated locally enriched calibration libraries composed of field-specific samples and spectrally similar samples selected from the national databases.

The final enriched local library for texture prediction contained 198 samples, including 99 local samples and 99 selected national samples. For OM prediction, the final enriched local library contained 242 samples, including 121 local samples and 121 selected national samples.

Local independent set for model validation and spatial mapping

An independent sampling campaign was conducted before the 2025/2026 crop season to validate the global and local predictive models for soil organic matter (OM) and texture. A total of 598 soil samples were collected using a georeferenced sampling grid of 223 × 223m, designed to represent the spatial variability of the field. Samples were collected from both the surface layer, at 0–10 cm depth, and the subsurface layer, at 20–40 cm depth.

The predictive performance of both global and local models was evaluated by comparing sensor-based predictions with conventional laboratory reference values. Model performance was assessed using the coefficient of determination (R^2), the ratio of performance to interquartile distance (RPIQ), the systematic deviation from the reference value (bias), and the root mean squared error (RMSE) (Belon-Maurel et al., 2010). In addition, to better characterize the prediction error profile, the 50th, 90th, and 95th percentiles of the absolute error were calculated, together with the maximum observed error, as recommended by Tavares et al. (2025). Absolute error was calculated as the modulus of the difference between measured and predicted values.

Sensor-based predictions of texture and OM obtained from both global and local models, as well as the conventional laboratory reference values, were used for spatial map comparison. Only surface soil data, corresponding to the 0–10 cm layer, were used for mapping assessment. Spatial interpolation was performed following the recommendations of Oliver and Webster (2014), using ordinary kriging (OK) or universal kriging (UK) depending on the exhibition of trends in the coordinates (x + y). The nugget effect, partial sill (sill minus nugget), and range of the variogram were calculated in gstat package (Pebesma, 2004), being fitted for three distinct models: spherical, Gaussian, and exponential. The optimal model was identified via k-fold cross-validation, selecting the configuration that yielded the lowest root mean squared error (RMSE) and the highest coefficient of determination (R^2). For texture and OM, validation and kriging used leave-one-out cross-validation (LOOCV) on all available data; for yield and ECa, a 20-fold cross-validation with a maximum of 200 neighbors was used for computational viability. The resulting maps were compared to evaluate whether sensor-based predictions reproduced the main spatial patterns observed with conventional laboratory analysis. Regarding the geostatistical parameters, the spatial structure of the data was gathered following Bejarano et al. (2026), calculating both spatial dependence (SD) and Moran's bivariate index. SD was quantified by the ratio between partial sill and the sill (partial sill + nugget). The classes of spatial dependence followed Konopatzki et al. (2012): <20 (Very Low - VL); [20–40) (Low - L); [40–60) (Moderate - M); [60–80) (High - H); and ≥ 80 (Very High - VH). Moran's index depends on the neighborhood order and directionality of the data; therefore, to provide a representative assessment, the spatial weights matrix was computed using the average across all directions (omnidirectional approach).

CEa and yield data

ECa data were collected during 16 days of field expedition (from 25/02/2026 to 13/03/2026) using the tractor-driven Veris-3100 Soil Electrical Conductivity Sensor (Veris®, Salina, KS, USA). Data acquisition was performed with a pass-to-pass spacing of 30 m and a sampling frequency of 1 Hz, corresponding to one measurement per second and totaling approximately 32 points ha^{-1}. The ECa dataset was filtered to remove inconsistent observations, including inliers, outliers, and points associated with equipment maneuvers or acquisition artifacts.

Soybean yield data were recorded during harvest using a John Deere GS4 yield monitor, totaling approximately 326 observations ha^{-1}. The yield monitor was calibrated using the real weight measured on the farm. Yield data were filtered to remove inconsistent values, including errors associated with harvester maneuvers, abrupt changes in grain flow, positioning errors, and outliers. After filtering, the yield and ECa dataset were interpolated, using the same procedure described in the above section, and used as an ancillary spatial layer to evaluate whether OM and texture maps derived from sensor-based predictions were consistent with crop performance patterns observed in the field.

RESULTS AND DISCUSSION

The descriptive statistical analysis of the calibration and validation set of the national and local libraries is presented in Table 2. The validation subset showed means and coefficients of variation (CVs) consistent with the local training library for both clay and OM; although, the range of concentrations in the local calibration set was slightly lower than that in the validation set for both clay and organic matter. This was because the samples for the local calibration set were selected at random from a wide geographical area of the farm. Conversely, the global training subset encompassed a significantly broader and certainly more heterogeneous pedological domain, which is explained by national-scale soil variability rather than farm-scale variability. The calibration subset is usually selected to be broader than the validation set, ensuring that predictions are made solely within a range to which the algorithm was already exposed. The rationale is to avoid analyzing the potentiality of an algorithm or, in this case, a library, based on extrapolation errors caused by extrapolating beyond the calibration range rather than the true efficacy of the technique (Stenberg et al., 2010).

Table 2. Descriptive statistics of soil attributes for the calibration and validation set of both national and local libraries.

	Clay		OM	
	National	**Local**	**National**	**Local**
	-------------g kg-1-------------			
	---------------- Calibration set----------------			
n	4,437	99	12,829	121
Minimum	16.0	341	2.0	14
Maximum	770.0	656	73.0	42
Mean	268.0	512	23.9	25
1_{st} Quartile	116.0	451	15.0	18
3^{rd} Quartile	410.0	579	31.0	30
SD	184.3	82	10.7	8
CV (%)	68.8	16.1	44.6	31.0
	---------------- Independent validation set----------------			
n	626		598	
Minimum	165.0		14.0	
Maximum	793.0		45.0	
Mean	547.3		25.8	
1^{st} Quartile	487.3		17.0	
3_{rd} Quartile	602.8		32.8	
SD	82.8		8.3	
CV (%)	15.1		32.0	

The performance of the different calibration scenarios (global vs. local) for sensor-based inferences is shown in Fig. 2 and Table 3. Overall, local libraries provided satisfactory predictions for clay using pXRF and for organic matter (OM) using pXRF–VNIR fusion, with RPIQ > 2.2 and R^2 > 0.67. Compared with global models, local models generally reduced RMSE by approximately 50% and showed lower bias, indicating a better representation of the farm-specific variability. Although global models achieved R^2 values similar to those of local models, their higher bias and lower RPIQ suggest that, despite capturing the general linear relationship with the reference analyses, they were less accurate for local prediction. This pattern is also evident in the scatterplots, where predictions for clay and OM followed a linear trend close to the 1:1 line, but global models showed greater systematic deviation. For OM, the global model also overestimated a specific group of samples, a behavior not observed in the local model, likely because the global training subset did not include samples representative of those local conditions.

The performance for clay prediction via pXRF aligns with previous tropical soil characterizations (Silva et al., 2020; Gozukara et al., 2025). Guzukara et al. (2025) reported a mean clay prediction accuracy (R^2) across various articles of 0.73 ± 0.17, which aligns with the 0.63 and 0.67 found in this study. For predicting OM using the fusion of VNIR and pXRF, Tavares et al. (2020) reported R^2 values between 0.72 and 0.85 for tropical soils, which align with the accuracy found in this study (0.81 and 0.91), especially given the higher sampling density used here. Xu et al. (2019) reported R^2 values of approximately 0.58 and 0.76, depending on the fusion strategy, which were lower than those reported above, possibly due to differences in soil between China and Brazil. The better performance of the local libraries can be explained by the inherent limitations of directly applying broad-scale national (or global) soil spectral libraries to local datasets (Viscarra Rossel et al., 2024). Consequently, to achieve accurate local predictions using a global library, it can be recommended to employ localization techniques such as transfer learning or "spiking" (enriching the global dataset with representative subset of local samples) to improve accuracy (Bellinaso et al., 2021; Ng et al., 2022). Furthermore, the discrepancy between the training subsets (Table 2) explains the differences in model performance. The global library's exposure to a wider range of pedological conditions, many of which are unrepresentative of the specific farm soil conditions, ultimately limited its predictive capabilities.

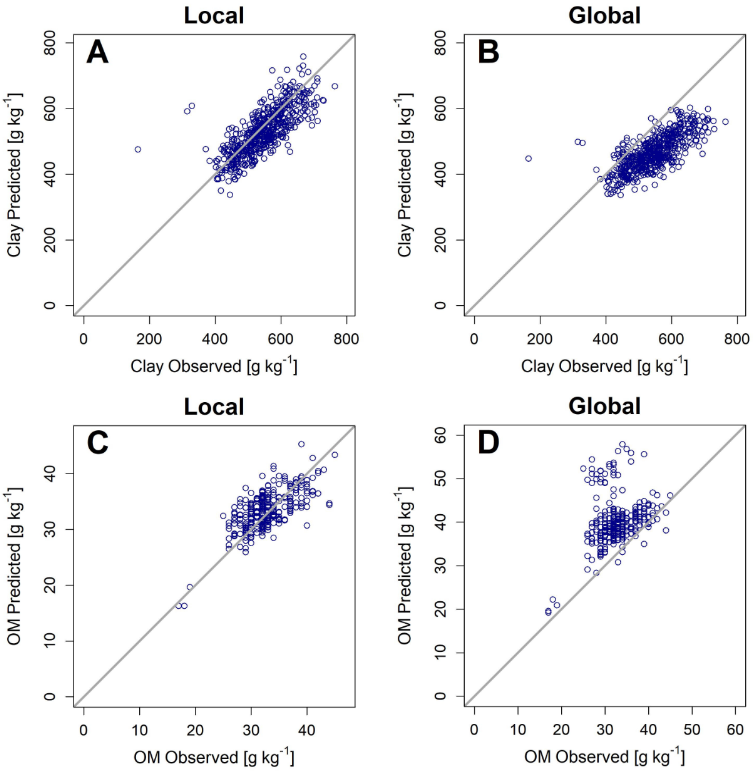

Fig 2. Scatterplot of the predicted and observed values for the exclusively external validation subset. A) Clay prediction using local XRF library; B) Clay prediction using global XRF library; C) Organic Matter prediction using local XRF+VNIR library; D) Organic Matter prediction using local XRF+VNIR library.

Table 3. Performance of clay and organic matter prediction models for local and global libraries.

	Clay [$g\ kg^{-1}$]		Organic Matter [$g\ kg^{-1}$]	
	Local	Global	Local	Global
Mean Error	43	87.2	2.2	4.9
Error 90%	83	142	4.4	9.9
Error 95%	98	164.9	5.1	16.6
Error max.	641	575	11	27.4
RMSE	52.3	94.8	2.73	7
RPIQ	2.2	1.2	5.8	2.3
bias	-13.1	-76.6	-0.4	4.7
R^2	0.67	0.63	0.91	0.81

The descriptive statistics for the ECa and Yield data (Table 4) indicate that the soybean yield was consistent across the measured points, with a moderate CV (15%) and first (Q1) and third quartiles (Q3) close to the mean, with respectively values of 51.4, 61.4 and 56,3 t ha^{-1}, indicating that most of the field was close to the mean. At the same time, the minimum (21.8 t ha^{-1}) and maximum (107.2 t ha^{-1}) yields showed a wide productive range at the field's extremes, indicating heterogeneity that could be explained, among other factors, by soil variability, which will be further discussed. The ECa showed similar behavior, with Q1 (4.8 ms m^{-1}) and Q3 (6.5 ms m^{-1}) close to the mean (5.7 ms m^{-1}) but a minimum (2.6 ms m^{-1}) and a maximum (11.8 ms m^{-1}), indicating field heterogeneity that may also be explained by soil variability.

Table 4. Descriptive statistics of soybean yield and apparent electrical conductivity (ECa).

	Yield [t ha^{-1}]	ECa [ms m^{-1}]
n (samples)	460,416	94,163
Minimum	21.8	2.6
Q1	51.4	4.8
Mean	56.3	5.7
Q3	61.4	6.5
Maximum	107.2	11.8
Standard Deviation	8.6	1.3
Coefficient of Variance (%)	15.2	21.9

Regarding the correlation between soil properties, both statistical (Spearman's correlation - Fig. 3) and geostatistical (Moran's bivariate index - Fig. 4) correlations agreed with each other. Regarding Spearman's correlation, the local, global, and laboratory approaches were strongly correlated, with the local library approach closer to the laboratory than the global one, corroborating external validation modeling results. Clay exhibited the strongest relationships across the estimation methods (Fig. 3). The local library prediction for clay showed a high correlation with the laboratory data (r = 0.81), outperforming the global library prediction (r = 0.76), indicating that accounting for site-specific soil variance improved the predictive relationship. This was particularly true for OM, which showed a strong correlation with the laboratory (r = 0.68), whereas the global approach showed a weaker correlation (r = 0.39).

The bivariate Moran's index (Fig. 4) confirmed that the predictive relationships observed in Spearman's correlation consistently translate into spatial dependence. For the clay fraction, the spatial correlation between the local library and the laboratory data was high (I = 0.87), slightly surpassing that of the global library (I = 0.83), indicating that the local calibration better predicts site-specific variance. For the OM, this was especially true, with the local library surpassing the global one (I = 0.61 versus I = 0.23), probably due to the overestimated group of samples, which disrupted the spatial correlation.

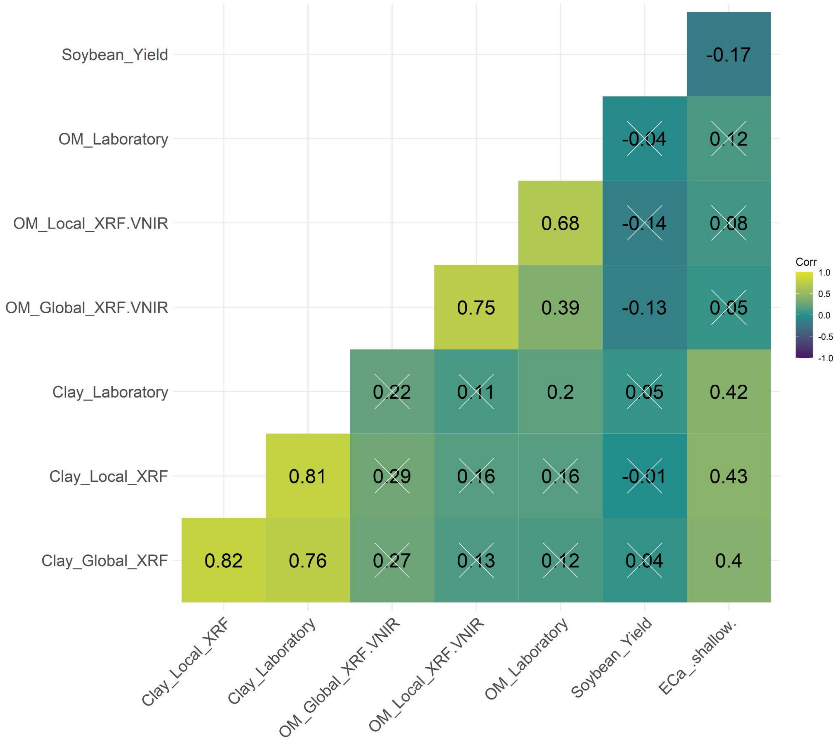

Fig 3. Spearman's correlation coefficient matrix of predicted and measured soil properties, Yield, and ECa. Cells marked with 'X' indicate that the Spearman correlation was not significant at the 5% level.

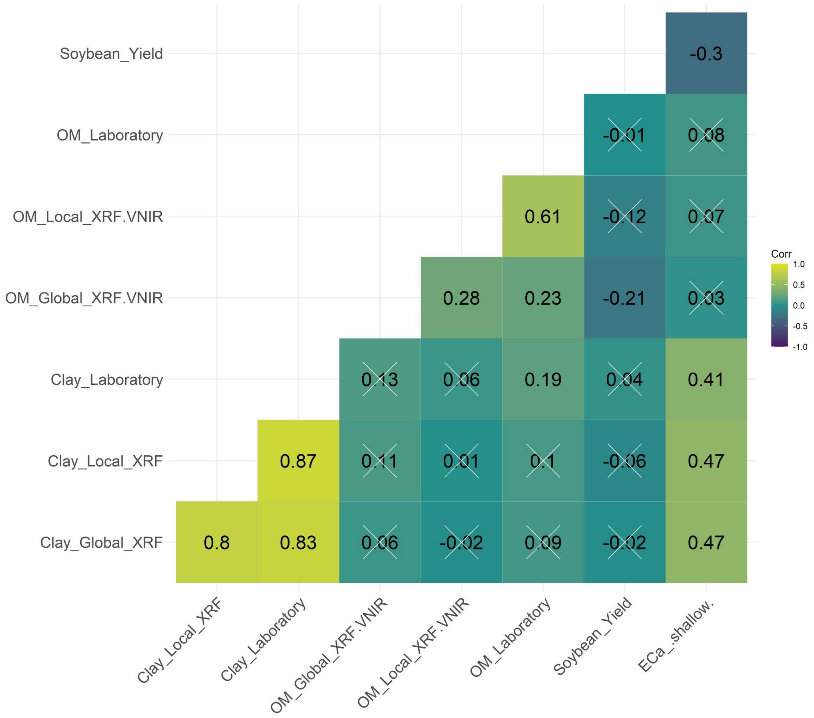

Fig 4. Bivariate Moran's Index matrix (mean between both directions) of predicted and measured soil properties, Yield, and ECa. Cells marked with 'X' indicate that the Spearman correlation was not significant at the 5% level.

Regarding the mapping capability, the spatial results corroborate the previously disclosed models, both for soil properties and across libraries. The local library inference maps were much more similar than those from the global library (Fig. 5). This was particularly true for the OM mapping. The local

OM model successfully captured a moderate spatial structure (R^2 = 0.34, Moran's I = 0.37), whereas the global model largely failed to represent the field's variability (R^2 = 0.05, very low spatial dependence). This discrepancy in OM global map is likely due to an overestimated group of samples, which could have interfered with the interpolation of the results. Clay mapping with the global library, however, showed spatial distribution consistent with the laboratory version, supported by strong spatial dependence (Moran's I = 0.55) and moderate model accuracy (R^2 = 0.59). This indicates the potential of global libraries in sensor-based inference. Unlike the texture and OM maps, which were interpolated for the entire study area, the extensive data in the Yield and ECa maps (approximately 460,000 and 100,000 observations, Table 4) preclude calculation of variograms. Rather, interpolation was performed for each of the seven individual fields; thus, the geostatistical parameters for yield and ECa are omitted from Table 5, which, in turn, presents all geostatistical parameters from the texture and OM mapping tests.

Although with generally good relationships between ECa and textural components, especially the clay fraction (Corwin & Lesch, 2005), this trend was not seen visually across maps (Fig. 2). The ECa spatial variability (map J, Fig.2) did not follow the visual patterns of clay (maps A-C, Fig. 2); Rather, it showed the presence of striated and geometric visual patterns, suggesting that ECa readings were dominated by anthropogenic and site-specific management practices, which alter soil density and water microdynamics, ultimately masked the mineralogical and textural influence on the proximal sensor. The anthropogenic factor is commonly seen in ECa mapping (Corwin & Lesch, 2005). It is important to note that because the interpolation was performed for each individual field rather than the entire area, the field edges were more pronounced, which may further reduce the relationship. Moreover, the Spearman correlation (Fig. 3) and the Moran’s index (Fig. 4) indicated a moderate relationship between ECa and clay (0.40 - 0.43 and 0.41 - 0.47, respectively), corroborating the discussion that the mineralogical relation of the ECa is present, but its spatial distribution is partially confounded by other factors (Sanches et al., 2018), possibly management-induced noise.

The yield map (map K, Fig. 2) showed one field with a contrasting pattern among the others, with very low soybean yield (~26-50 t ha^{-1}) and an anthropogenic characteristic, totally different from the neighboring fields. The texture (map C, Fig. 2), OM (map I, Fig. 2), slope, fertility, and compaction (data not shown) of that field are similar to those of the adjacent fields. Also, the variety chosen was the same. Possibly, the low yield was due to management-specific reasons, especially the post-emergence herbicide used. Apart from this field, the yield did not follow the studied soil properties (clay and OM) patterns, having a more pronounced chance in neighbor harvesting rows than along the area, and, when not considering the low-yield field previously discussed, it appeared to have a low range of variability (spatially). The quantitative results corroborate these findings, showing low (and non-significant) Spearman correlations (Fig. 3) and Moran’s bivariate Index (Fig. 4) between yield and the other studied properties. While soil matrix properties establish the long-term attainable yield potential of an area, seasonal crop performance is highly sensitive to immediate, dynamic interactions like localized microclimate shifts, weed pressure, or chemical management footprints (Pusch et al., 2022; Sudduth et al., 2005). Although not specific to this study, these findings indicate that soil-based sensors alone sometimes are insufficient to predict yield variability, suggesting that incorporating vegetative information (such as vegetation indices) could improve prediction.

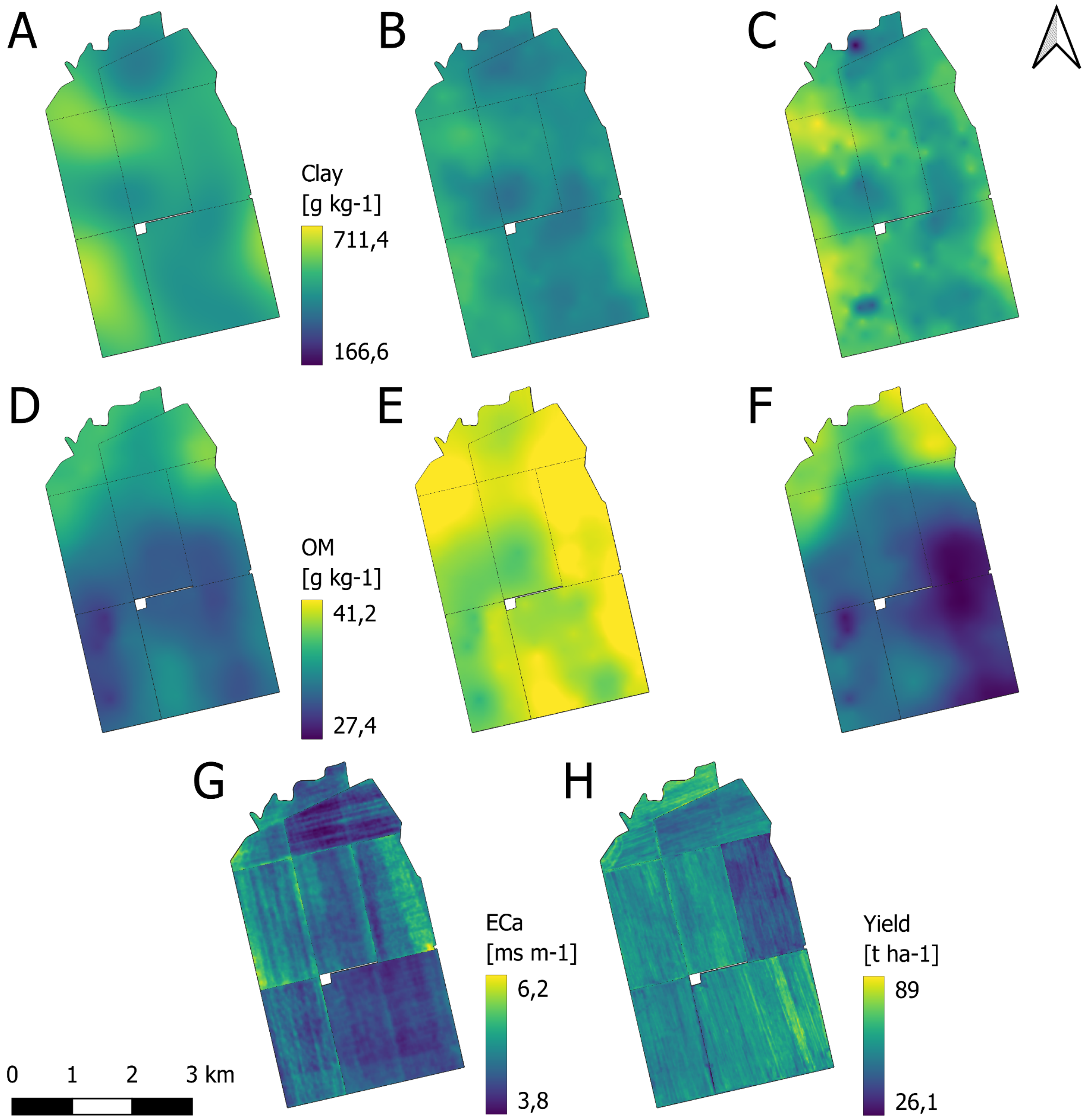

Fig 5. Spatially interpolated maps: A) Clay with local library (lib); B) Clay with global lib; C) Clay by wet chemistry; D) Organic Matter (OM) with local library (lib); E) OM with global lib; F) OM by wet chemistry; G) apparent Electrical Conductivity (ECa); H) Yield.

Table 5. Geostatistical parameters of the interpolated soil properties (Clay and Organic Matter) maps.

	------------- Clay -------------			------- Organic Matter-------		
	Local	Global	Lab	Local	Global	Lab
Model	Gau	Sph	Exp	Exp	Exp	Exp
R^2	0.64	0.59	0.71	0.34	0.05	0.57
RMSE	42.93	34.04	39.06	3.01	5.36	2.79
Range	1573	1803	1016	2698	609	8399
SD.	71.1 (VH)	80.7 (VH)	100 (VH)	51.1 (M)	17.6 (VL)	88.0 (VH)
Moran's I	0.59	0.55	0.63	0.37	0.1	0.58
Cluster / Random	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster

SD., Spatial dependence; Cluster or Random based on the Moran's bivariate Index p-value (α = 0.05).

Soil analysis using sensor-based methods and spectral libraries is the key to meeting the data needs of agricultural and environmental monitoring. There is a demand in the market for sensor-based inference services; for example, we highlight the recent protocols from VERRA (e.g., Verra's Verified Carbon Standard - VM0042 and VT014, Social Carbon Standard's SCM0005), which suggest the complementary use of sensors to expand the spatial coverage and frequency of measurements for Monitoring, Reporting, and Verification (MRV) of soil carbon stocks. In addition, there is a growing number of Brazilian laboratories offering sensor-based analysis services alongside traditional methods. This adoption, still in its early stages, aims to reduce reagent use and shorten analysis time. Despite these demands, sensor-based inferences are still viewed with some skepticism by users, especially because the method lacks regulation/standardization and compliance, and certification mechanisms (e.g., proficiency tests), as discussed by Tavares et al. (2025).

In this context, the results of this study show that sensor-based inference services using local models provided more accurate predictions than global models. Local calibration reduced RMSE, increased RPIQ, and minimized bias for both clay and OM, indicating a better representation of farm-specific variability. Nevertheless, although global models were more biased, they are precise and still reproduced the main spatial variability patterns observed in laboratory-derived maps. For example, clay RMSE was 52.3 and 94.8 g kg^{-1} for the local and global models, respectively, but their spatial agreement with the laboratory map remained high, with bivariate Moran's I values of 0.87 and 0.83. The main exception was the global OM model, which showed weaker spatial agreement with the laboratory map, likely due to the overestimation of a specific group of samples.

Finally, even when global models are less accurate than local calibrations, their performance may still be useful for decision-making if the sensor-based error remains below the analytical tolerance required for a given application (Tavares et al., 2026). Therefore, future studies should define application-specific error thresholds for sensor-based soil analysis, allowing different sensor strategies, including global and locally calibrated models, to be evaluated more objectively according to their intended use.

CONCLUSION: This case study demonstrated that laboratory-based sensor inference can support the mapping of clay and OM variability under field conditions. Local spectral libraries provided the best predictive performance, reducing RMSE and bias compared with global models for both attributes. However, despite their higher bias, global models were still able to reproduce spatial patterns similar to those obtained from conventional laboratory analyses, especially for clay. The main limitation was observed for the global OM model, which overestimated a group of samples and showed weaker spatial agreement with the laboratory map. Overall, these results indicate that proximal sensing, particularly when combined with local calibration, can increase the spatial density of soil information while reducing analytical cost, turnaround time, and reagent use compared with conventional laboratory workflows.

ACKNOWLEDGMENTS: The authors thank 3r Ribersolo laboratory for providing the samples and reference analyses.

REFERENCES:

ANDRADE, R. FARIA, W.M.; SILVA, S.H.G. CHAKRABORTY, S.; WEINDORF, D.C.; MESQUITA, L.F. GUILHERME, L.R.G. CURI, N. Prediction of soil fertility via portable x-ray fluorescence (pxrf) spectrometry and soil texture in the brazilian coastal plains. **Geoderma**, 357, p.113960, 2020.

BEJARANO, L. D.; OLIVEIRA, A. L. G.; POZZUTO, J. V. F.; SÁNCHEZ, D. C.; AMARAL, L. R. Performance of interpolation methods in digital soil mapping: the influence of data characteristics. **Precision Agriculture**. v. 27(1), 10. 2026

BELLON-MAUREL, V.; FERNANDEZ-AHUMADA, E.; PALAGOS, B.; ROGER, J.M. MCBRATNEY, A. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. **TrAC Trends in Analytical Chemistry**, v.29, n.9, p.1073-1081, 2010.

BEN-DOR, E., ONG, C., & LAU, I. C. Reflectance measurements of soils in the laboratory: Standards and protocols. **Geoderma**, 245–246, p. 112–124, 2015.

CORWIN, D. L.; LESCH, S. M. Apparent soil electrical conductivity measurements in agriculture. **Computers and Electronics in Agriculture**, 46, 1–3, p. 11–43, 2005.

DEMATTÊ, J. A. M. et al. The Brazilian Soil Spectral Library (BSSL): A general view, application and challenges. **Geoderma**, v. 354, p. 113793, 15 nov. 2019.

GOZUKARA, G.; HARTEMINK, A. E.; HUANG, J.; DEMATTÊ, J. A. M. Prediction accuracy of pXRF, MIR, and Vis-NIR spectra for soil properties—A review. **Soil Science Society of America Journal**, 89, n. 2, p. e70028, 2025.

GUERRERO, C., STENBERG, B., WETTERLIND, J., VISCARRA ROSSEL, R.A., MAESTRE, F.T., MOUAZEN, A.M., ZORNOZA, R., RUIZ-SINOGA, J.D. AND KUANG, B. Assessment of soil organic carbon at local scale with spiked NIR calibrations: effects of selection and extra-weighting on the spiking subset. **European Journal of Soil Science**, 65(2), pp.248-263, 2014.

JAVADI, S. H.; MUNNAF, M. A.; MOUAZEN, A. M. Fusion of Vis-NIR and XRF spectra for estimation of key soil attributes. **Geoderma**, v. 385, p. 114851, 2021.

KONOPATZKI, M. R. S., SOUZA, E. G., NÓBREGA, L. H. P., URIBE-OPAZO, M. A., & SUSZEK, G. Spatial variability of yield and other parameters associated with Pear trees. **Engenharia Agrícola**, 32(2), p. 381–392, 2012.

KRUG, F. J.; ROCHA, F. R. P. Métodos de preparo de amostras para análise elementar. **EditSBQ, Sociedade Brasileira de Química, São Paulo, 572p**, 2016.

LIMA, T.M.; WEINDORF, D.C.; CURI, N.; GUILHERME, L.R.; LANA, R.M.; RIBEIRO, B.T. Elemental analysis of Cerrado Agricultural soils via portable X-ray fluorescence spectrometry: Inferences for soil fertility assessment. **Geoderma**, v.353, p.264–272, 2019.

MALDANER, L. F., MOLIN, J. P., & SPEKKEN, M. Methodology to filter out outliers in high spatial density data to improve maps reliability. Scientia Agricola, 79(1), e20200178, 2022.

NG, W.; MINASNY, B.; JONES, E.; MCBRATNEY, A. To spike or to localize? Strategies to improve the prediction of local soil properties using regional spectral library. **Geoderma**, 406, 2022.

MINASNY, B. AND MCBRATNEY, A.B. Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. **Chemometrics and intelligent laboratory systems**, 94(1), pp.72-79, 2008.

MOLIN, J. P.; TAVARES, T. R. Sensor systems for mapping soil fertility attributes: challenges, advances, and perspectives in brazilian tropical soils. **Engenharia Agrícola**, v. 39, p. 126-147, 2019.

MONDIA, J.P., GOH, F., BRYNGELSON, P.A., MACPHEE, J.M., ALI, A.S., WEISKOPF, A. AND LANAN, M. Using X-ray fluorescence to measure inorganics in biopharmaceutical raw materials. **Analytical Methods**, 7(8), p. 3545-3550, 2015.

OLIVER, M. A., & WEBSTER, R. A tutorial guide to geostatistics: Computing and modelling variograms and kriging. **CATENA**, 113, p. 56–69, 2014

PEBESMA, E. J. Multivariable geostatistics in S: the gstat package. **Computers & Geosciences**, 30(7), p. 683–691, 2004.

PUSCH, M.; SAMUEL-ROSA, A.; OLIVEIRA, A. L. G.; MAGALHÃES, P. S. G.; AMARAL, L. R. DO. Improving soil property maps for precision agriculture in the presence of outliers using covariates. **Precision Agriculture**, v. 23(5), p. 1575–1603, 2022.

SANCHES, G. M.; MAGALHÃES, P. S. G.; REMACRE, A. Z.; FRANCO, H. C. J. Potential of apparent soil electrical conductivity to describe the soil pH and improve lime application in a clayey soil. **Soil and Tillage Research**, 175, p. 217–225, 2018.

SCHAEFER, C. E. G. R.; SIQUEIRA, R. G.; PEREIRA, L. F.; GOMES, L. DE C.; ALMEIDA, P. H. A.; FRANCELINO, M. R.; FIRMINO, F. H. T.; SOUZA, J. J. L. L. DE; KER, J. C.; FERNANDES-FILHO, E. I. Brazilian Latossolos (Ferralsols, Oxisols) from different biomes: a multiproxy study on the spatial variability of the most weathered tropical soils in South America. **Geoderma Regional**, 43, p. e01012, 2025.

SHI, X., SONG, J., WANG, H., LV, X., ZHU, Y., ZHANG, W., BU, W. AND ZENG, L. 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, p.116301, 2023.

SILVA, S.H.G.; RIBEIRO, B.T.; GUERRA, M.B.B.; CARVALHO, H.W.P.; LOPES, G.; CARVALHO, G.S.; GUILHERME, L.R.G.; RESENDE, M.; MANCINI, M.; CURI, N. RAFAEL, R.B.A. pXRF in tropical soils: Methodology, applications, achievements and challenges. **Advances in Agronomy**, 167, p.1-62, 2021.

SILVA, S. H. G.; WEINDORF, D. C.; PINTO, L. C.; FARIA, W. M.; ACERBI JUNIOR, F. W.; GOMIDE, L. R.; MELLO, J. M. DE; PÁDUA JUNIOR, A. L. DE; SOUZA, I. A. DE; TEIXEIRA, A. F. DOS S.; GUILHERME, L. R. G.; CURI, N. Soil texture prediction in tropical soils: A portable X-ray fluorescence spectrometry approach. **Geoderma**, 362, p. 114136, 2020.

STENBERG, B.; VISCARRA ROSSEL, R.A.; MOUAZEN, A.M.; WETTERLIND, J. Visible and near infrared spectroscopy in soil science. **Advances in Agronomy**, v.107, p.163-215, 2010.

SUDDUTH, K. A.; KITCHEN, N. R.; WIEBOLD, W. J.; BATCHELOR, W. D.; BOLLERO, G. A.; BULLOCK, D. G.; CLAY, D. E.; PALM, H. L.; PIERCE, F. J.; SCHULER, R. T.; THELEN, K. D. Relating apparent electrical conductivity to soil properties across the north-central USA. **Computers and Electronics in Agriculture**, 46, n. 1- 3 SPEC. ISS., p. 263–283, 2005.

TAVARES, T. R., DEMATTÊ, J. A. M., MOUAZEN, A. M., BEN-DOR, E., ANSELMI, A. A., CAMPOS, V. M. de A., et al. Proficiency tests for soil analysis services via proximal sensing. **Geoderma**, 460, 117434, 2025

TAVARES, T. R.; MOLIN, J.P.; NUNES, L.C.; ALVES, E.E.N.; MELQUIADES, F.L.; CARVALHO, H.W.P.; MOUAZEN, A.M. effect of X-ray tube configuration on measurement of key soil fertility attributes with XRF. **Remote Sensing**, v.12, n.6, p.963, 2020.

TAVARES, T. R.; NUNES, L. C.; ALVES, E. E. N.; ALMEIDA, E. de; MALDANER, L. F.; KRUG, F. J.; CARVALHO, H. W. P. de; MOLIN, J. P. Simplifying Sample Preparation for Soil Fertility Analysis by X-ray Fluorescence Spectrometry. **Sensors**, v. 19, n. 23, 2019.

TAVARES, T.R., DEMATTÊ, J.A., BEDUM, G.V., GELAIN, M.S., RAGAGNIN, M.M., REZENDE, V.P., MOUAZEN, A.M., MOLIN, J.P., LAVRES, J. AND DE CARVALHO, H.W.P. Rapid in situ analysis with portable XRF sensor: a pioneering study proposing fitness-for-purpose approach for performance assessment in tropical soils. **Soil Security**, p.100228, 2026.

TEIXEIRA, A.F.D.S.; WEINDORF, D.C.; SILVA, S.H.G.; GUILHERME, L.R.G.; CURI, N. Portable X-ray fluorescence(pXRF) spectrometry applied to the prediction of chemical attributes in Inceptisols under different land uses. **Ciênc. Agrotec**, v.42, p.501–512,2018.

VAN RAIJ, B.; ANDRADE, J.C.; CANTARELLA, H.; QUAGGIO, J.A. Análise química para avaliação de solos tropicais. **IAC**: Campinas, Brasil, p.285,2001.

VISCARRA ROSSEL, R.A.; ADAMCHUK, V.I., SUDDUTH, K.A., MCKENZIE, N.J., LOBSEY, C. Proximal Soil Sensing: An Effective Approach for Soil Measurements in Space and Time. Adv. Agron. 113, p. 243–291, 2011.

VISCARRA ROSSEL, R. A. et al. A global spectral library to characterize the world's soil. **Earth-Science Reviews**, 155, p. 198–230, 2016.

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. An imperative for soil spectroscopic modelling is to think global but fit local with transfer learning. **Earth-Science Reviews**, 254, p. 104797, 2024.

XU, D.; ZHAO, R.; LI, S. CHEN, S.; JIANG, Q.; ZHOU, L.; SHI, Z. Multi-sensor fusion for the determination of several soil properties in the Yangtze River Delta, China. **European Journal of Soil Science**, v. 70, n. 1, p. 162–173, 2019.