JAAS

PAPER

View Article Online
View Journal | View Issue

Cite this: *J. Anal. At. Spectrom.*, 2024, **39**, 2903

Assessing the performance of handheld LIBS for predicting soil organic carbon and texture in European soils†

Alex Wangeci, *[a] Maria Knadel, [b] Olga De Pascale,[c] Mogens H. Greve[a] and Giorgio S. Senesi *[c]

Laser-induced breakdown spectroscopy (LIBS) has contributed to the advanced and rapid determination of soil properties including soil organic carbon (SOC) and texture. Recent developments of commercial handheld LIBS (hLIBS) have allowed the use of the technique directly in the field. However, to date, the performance of hLIBS on different types of soils covering wide geographical distributions has not been evaluated. In this study, a total of 305 soil samples covering a continental scale were used to assess the repeatability and reproducibility of LIBS data acquired using a commercially available hLIBS instrument. Furthermore, the performance of the prediction models for SOC and texture was evaluated based on the prediction error. The repeatability and reproducibility of LIBS data were evaluated based on the relative standard deviation (RSD) for measurements performed under similar and different environmental conditions (temperature and humidity). First, the RSD of the signal ratios and the predicted values for soil properties under investigation were calculated. Then, the prediction accuracy of the various soil properties was compared based on the standardized root mean error of prediction (SRMSEP) and the ratio of performance to interquartile distance (RPIQ). The signal ratios assessed using the C, Si, Ca, and K LIBS emission lines achieved a repeatability of 4–9% and a reproducibility of 7–10%, whereas the repeatability and reproducibility for predicting SOC and texture were <25%. The prediction of sand content exhibited the lowest error (SRMSEP = 0.14) followed by clay and silt (SRMSEP = 0.15), and then SOC (SRMSEP = 0.16). The results of this work underscore the promising potential of hLIBS for large-scale SOC and texture determination, with the opportunity to enhance the prediction accuracy by integrating soil mineralogy information for soil classification before applying the modeling process.

Received 13th August 2024
Accepted 5th September 2024

DOI: 10.1039/d4ja00292j

rsc.li/jaas

1. Introduction

Soil organic carbon (SOC) and texture are important soil properties exerting a critical influence on sustainability and soil management for improved crop productivity. On a global scale, approximately 1.5×10^{15} kg of organic C is stored in the top meter of the soil layer, representing approximately three times the C in the aboveground biomass and two times atmospheric C (as CO_2).[1,2] The SOC content in the upper 10 to 20 cm is often used as an indicator of soil quality and productivity for fertilization and soil management recommendations.[3] Soil texture is an important physical parameter that has direct implications in soil management, such as fertilizer application, tillage, irrigation, and other agronomic interventions.[4] Unlike texture, which is a relatively stable soil property *i.e.*, it does not change over time, SOC has been shown to fluctuate especially depending on human activities.[5] Due to the increasing need of accurate agronomic interventions and related fundamental research, these critical soil quality indicators are continually tested involving field sampling and laboratory analyses that are costly, lengthy, and unfriendly to the environment. Moreover, due to the increased demand for spatially resolved soil information, it is challenging to scale laboratory tests to achieve a reasonable turn-around-time.

Nowadays, various spectroscopic techniques showing the potential to bridge the gap between cost, accuracy, and speed are used for soil analyses.[6,7] Advances of digital soil mapping have led to increased demand for remote and adaptable field sensors to monitor multiple soil properties, including SOC and texture at local, regional, national, continental, and even global scales.[8] Furthermore, the commercial availability of portable, mobile, and handheld analytical instruments has made it possible to perform field measurements directly in the field by

[a]*Department of Agroecology, Aarhus University, Blichers Allé 20, 8830 Tjele, Denmark. E-mail: aw@agro.au.dk*

[b]*GeoPark Vestjylland, Skæreum Møllevej 4, 7570 Vemb, Denmark*

[c]*Consiglio Nazionale delle Ricerche (CNR), Istituto per la Scienza e Tecnologia dei Plasmi (ISTP), Sede di Bari Via Amendola, 122/D, 70126 Bari, Italy. E-mail: giorgio.senesi@cnr.it*

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4ja00292j

"taking the laboratory to the sample". Although laboratory analytical methods are more accurate as compared to proximal on-site analysis, they often cannot fully respond to the growing demand for soil analysis, especially at national and continental scales.

Among various spectroscopic techniques available nowadays, laser-induced breakdown spectroscopy (LIBS) has shown to be a very promising analytical technique for the measurement of the elemental composition, nutrients, and heavy metal contaminants in agricultural samples, including fertilizers, plants, and soils.[9,10] Furthermore, in recent years, the commercial availability of handheld (h) LIBS has offered greater flexibility and widened the range of LIBS applications also for soil analysis.[11] Although the accuracy of the hLIBS is subject to various instrumental and sample factors, and may not match that of laboratory or benchtop LIBS,[12] its portability allows direct field and rapid acquisition of preliminary data that can then be used as a pre-screening step before an in-depth and detailed laboratory analysis.[13] However, several studies have reported comparable performance of portable and laboratory LIBS setups. For instance, Wainner *et al.*[14] reported that the analysis of lead was similar regardless of the instrument used (portable or laboratory LIBS setup). Several studies have assessed the quantification of soil carbon[15–18] using portable LIBS systems while others have focused on soil nutrients and heavy metals.[14,19–22] To the best of our knowledge, hLIBS has been used in only one study[12] to investigate texture and other properties of a local dataset of soil samples in a field in Germany.

The major well-known challenge in the quantification of soil properties by LIBS is ascribed to the so-called matrix effects, which consist in changes of the intensity of emission lines of specific elements in the analyzed samples due to the variation of physical properties, *e.g.*, surface roughness due to different particle sizes, and/or chemical composition,[18] so that the sensitivity, repeatability and precision of the analysis are negatively affected.[23] Although most matrix effects related to the sample can be controlled by optimized preparation steps such as drying, sieving, milling, and pelletizing, environmental factors (*e.g.*, temperature and humidity) are difficult to control, but can be monitored. The stability of LIBS analysis, which is mostly evaluated by its repeatability, is critical as it influences data accuracy and precision.[24] Generally, the precision of soil analysis is impacted by the complex interaction between the soil and the laser, which is related to both laser features and the physical and chemical properties of the analyzed soil.[25] In particular, the largest contributor to the low repeatability of the LIBS signal is ascribed to the inherent soil heterogeneity.[14]

To enhance the performance of hLIBS analysis, it is essential to improve its ruggedness, stability, and reliability for different soil types.[18] However, the repeatability and reproducibility of LIBS, in terms of the variation of predicted values for soil properties, have not been extensively studied, as most previous studies have only focused on the signal intensity and variability for selected emission lines.[16,26–28] In particular, Ebinger *et al.*[26] investigated the reproducibility of the calibration curve for measurements carried out over 30 days using the ratio between the C 193.09 nm emission line and the sum of Al 199.05 nm and Si 212.41 nm emission lines. The authors achieved a coefficient of determination (R^2) (calibration) of 0.99 and concluded that there was no significant difference between the slopes of the calibration curves during the 30 day period. Cunat *et al.*[20] reported a precision and accuracy of respectively 8% and 14% for the analysis of Pb in homogeneous soils using a portable LIBS. Xu *et al.*[28] used selected emission lines from C, N, K, Ca, Mg, and Si to investigate the effect of the shot layer and number of shots on LIBS data quality based on the relative standard deviation (RSD) of the selected spectral lines, which ranged from 30% to 45%. To improve the overall robustness and accuracy of LIBS analysis, normalization of the spectra coupled with calibration and prediction modeling, commonly referred to as multivariate data analysis, should be applied. In particular, partial least squares regression (PLSR)[29] is a widely applied multivariate data analysis approach for treating LIBS spectral data, especially in investigations where large datasets are involved.[30–32] This method considers the whole spectrum of a heterogeneous material such as soil; thus it takes into account matrix effects, thus improving the accuracy and repeatability of quantitative analysis.[27] In general, the influence of matrix effects on the quantification of soil properties by LIBS is more apparent when a wide range of different soil types are analyzed, which is due to the corresponding complexity of their physical and chemical properties.[33–35]

Based on the issues considered above, this work aimed to evaluate the performance and robustness of the hLIBS technique in addressing the existing lack of large-scale studies on the prediction of key soil properties. In particular, the main objectives of this study were to: (a) assess the repeatability and reproducibility of the LIBS data achieved using a commercial hLIBS instrument in predicting the SOC content and texture of a wide range of soil types sampled at a continental scale from 21 European countries, and covering a wide geographical and textural distribution under similar and variable environmental conditions; and (b) develop and evaluate the performance of prediction models on the basis of the prediction accuracy achieved for SOC and texture.

2. Materials and methods

2.1. Soil samples and analysis

The topsoil samples (0–20 cm) used in this work were provided from the Land Use/Cover Area frame Survey (LUCAS) database that consists of a large collection of soils sampled from 21 European countries during the 2015 sampling campaign (Fig. S1†). A selection of 896 samples used in a previous study[36] was implemented by a conditioned Latin hypercube sampling (cLHS) optimized-stratified strategy (in *R*) that uses the texture, the coordinates, and the number of points per country as criteria to assign a chance of selection to each sample.[37] Then, the 305 samples used in this work were selected based on criteria that excluded samples featuring a *P* concentration (extractable *P*) below the reference method detection limit (<10 mg kg^{-1}), a pH > 6.2, and a SOC > 20% or below the reference method detection limit (<0.2%).

A detailed description of the sampling methodology can be found in the study by Orgiazzi *et al.*[38] and Jones *et al.*[39] The list of countries and the number of samples selected per country are reported in Table S1.† To ensure uniformity in terms of geographical coverage between the calibration and validation sample sets, the 305 samples were first sorted according to the country of origin, and then every third sample, one was selected for inclusion in the validation set so that it comprised 102 samples, whereas the 203 remaining samples were used for calibration.

All samples were air-dried, sieved to <2 mm, and stored away from light to avoid any physical or chemical changes over time. The soil texture (clay, silt, and sand percentages) was determined by laser diffraction according to the International Organization for Standardization (ISO) procedure number 13320:2009.[40] The total carbon content was determined using the dry combustion method and an elemental analyzer.[41] The inorganic carbon (carbonates) was determined by acid washing and titration.[42] For soils with no carbonates, the total carbon measured is equal to SOC. For soil samples where carbonates were present, the inorganic C (carbonates) content was subtracted from the total C content to ensure that only SOC was considered.[38]

2.2. Handheld LIBS analysis

The hLIBS analysis was conducted on pelletized samples obtained by pressing about 1 g of each sample placed in a 14 mm diameter cup using an automatic press (FOSS Analytical A/S, Denmark) at 1948 kg cm^{-2} for 30 s. A commercial hLIBS instrument (SciAps, Z-903 Woburn, MA, USA) consisting of a neodymium-doped yttrium aluminum garnet (Nd:YAG) operating at a wavelength of 1064 nm and a repetition rate of 50 Hz was used. The laser pulse delivered 5–6 mJ of 1 ns pulse duration with a nominal 100 μm beam size. The delay time was fixed at 650 ns over a 3 ms integration time. The spectrometer covered a wavelength range between 187 nm and 950 nm, with an approximate spectral resolution of 0.1 nm (187–400 nm) and 0.3 nm (400–950 nm), resulting in 5039 measured data points. The unit was equipped with an *XYZ* stage that enabled beam-rastering, *i.e.*, a better averaging for heterogeneous materials such as soil. Two cleaning pulses prior to each laser pulse/shot were applied (two laser shots on the soil pellet before spectra acquisition) for the removal of surface contaminants. During measurement, continuous argon purging at a pressure of approximately 12 psi was used for signal enhancement.

To reduce the possible effects of sample inhomogeneity, three points were analyzed on each sample. On each point, 12 locations in a 3 × 4 rectangle were measured in less than 10 s, *i.e.*, in approximately 30 s for each sample (Fig. S2†). At each location of the rectangle, four spectra were acquired, resulting in a total of 48 spectra per point analysis (three per sample) and 144 spectra per each sample. These spectra were then averaged, yielding 36 spectra for each sample.

2.3. Handheld LIBS soil maps

Soil maps can be a quick way to determine the relative concentration and distribution of specific elements and, indirectly, of soil properties. For instance, the relative distribution of SOC can be determined based on the abundance of related emission lines, such as carbon and calcium (which is associated with inorganic C). This can potentially be applied as a fast way to distinguish between organic and mineral soils. The emission lines of C at 193.09 nm and Ca at 422.67 nm were selected to develop the soil maps of two representative and arbitrarily selected soil samples based on the measured SOC content, *i.e.*, a soil sample that had a SOC content < 0.5%, classified as a low SOC sample, and a soil sample that had a SOC content > 15%, classified as a high SOC sample. The microscale mapping of the C and Ca elemental distribution on the soil pellet surface was achieved using the Geochem Pro application mounted on the hLIBS instrument. A single laser shot at a fixed laser firing rate of 10 Hz was used on a 16 × 16 grid (256 locations) that covered an area of approximately 2 mm^2 at 12.5 μm steps. The maps were developed by integrating the peak area of the specific element for each physical location measured. The normalized peak area was color-coded from low concentration (blue) to high concentration (red). This procedure allowed the development of "heat maps" that represent the distributions of C and Ca concentrations in the form of relative abundance.

2.4. Data analysis/modeling

The SciAps proprietary Profile Builder software was used for the initial data processing prior to spectra exportation for data analysis. The analysis of all multivariate data was carried out using MatLab R2021a (MathWorks, Inc., Natick, MA, USA) and PLS Toolbox 8.7 software (Eigenvector Research Inc.).

Different preprocessing methods including standard normal variate (SNV),[43] automatic Whittaker filter baseline correction,[44] and multiplicative scatter correction (MSC)[45] were applied. The assessment of the various preprocessing combinations was implemented in PLS Toolbox using the model builder function which allows simultaneous comparison of performance of different preprocessing methods *e.g.*, using the RMSECV. The preprocessing that resulted in the lowest RMSECV was selected for the prediction models. In this case, SNV and mean-centering preprocessing were applied for all the investigated soil properties. This was followed by partial least squares regression (PLSR) to correlate LIBS data with clay, silt, sand, and SOC reference data.[29] Additionally, feature selection using a variable selection algorithm implemented by applying interval partial least squares regression (iPLSR) was used for the prediction of soil properties. Finally, we compared the linear methods with a nonlinear modelling approach applied through an artificial neural network (ANN). For simplicity, a feedforward neural network architecture with 2 nodes in the first layer was applied for all the soil properties under investigation.

2.5. Repeatability and reproducibility assessment

A total of 101 samples, selected in terms of geographical distribution, were used to evaluate the repeatability of measurements and the robustness and accuracy of the model, the RSD of the signal ratio of each measured element and R^2 and root mean square error of cross-validation (RMSECV) were determined.

First, the signal ratio variation for C, Si, Ca, and K was determined by calculating the RSD of measurements performed on the same day and on different days. The four emission lines were selected based on the relationship with the soil properties under investigation (soil texture and SOC). For instance, there is a direct relationship between the C emission line and SOC while Si and Ca are indirectly related to soil texture since clay mineralogy is mainly composed of phyllosilicates ($Al_2Si_2O_5$) and sand is mainly composed of quartz (SiO_2) and limestone ($CaCO_3$). The K emission line was selected due to its relative abundance in the soil and well-resolved emission lines in the soil spectrum.

The ratios were calculated by selecting two emission lines of each element and dividing the higher intensity line by the lower intensity one (Table 1). Then, cross-validation models were compared for different measurements conducted on the same day and on different days. The procedure also involved the calculation of the RSD of the predicted value of each soil property for all the samples (eqn (1)).

$$\mathrm{RSD} = \frac{\mathrm{SD}}{\mathrm{mean}} \times 100 \quad (1)$$

where SD is the standard deviation of the signal ratio or predicted values for three measurements performed on each of the samples (on the same day and on three different days).

The repeatability was evaluated by performing three measurements on the same pellet under the same conditions on the same day. The average spectrum of each measurement for all samples was used to calculate the signal ratios. The spectra from individual measurements were also used to develop cross-validation models.

The reproducibility was assessed by conducting measurements at specified intervals over five days under varying environmental conditions of temperature and humidity (Table S2†). The first three measurements (days 1, 14, and 22) were performed on the same pellet while the last two were performed on a different pellet of the same sample (days 40 and 54). Furthermore, the effect of soil surface heterogeneity on the prediction accuracy was assessed by using different pellets of the same sample on days 40 and 54. Finally, a test for statistically significant difference (ANOVA) was performed for all achieved predictions of both same-day and different-day measurements.

2.6. Prediction models

The entire dataset of 305 samples was used for developing prediction models for clay, silt, sand, and SOC. The dataset was divided into 203 calibration and 102 validation samples. Cross-validation models were used for the assessment of repeatability by applying venetian blinds (20 samples per blind). On the other hand, the manual split of the samples into calibration and validation samples was applied for the prediction models (independent validation) of the soil properties under investigation.

The prediction accuracy was assessed using the root mean square error of prediction (RMSEP) (eqn (2)), followed by the calculated standardized root mean square error of prediction (S) RMSEP (eqn (3)), which was used to compare the accuracy across the soil properties. The SRMSEP value provides a scaleless measure for comparing different models across different soil properties. In particular, the lower the SRMSEP is, the more accurate the model is. Additionally, we used the ratio of performance to interquartile distance (RPIQ). The RPIQ considers the interquartile range enabling a comparison of performance across soil properties and between studies where samples with different concentration ranges are used. It represents the population spread better, regardless of the distribution.[46] The higher the RPIQ, the better the prediction model (eqn (4)).

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

$$\mathrm{SRMSEP} = \frac{\mathrm{RMSEP}}{\mathrm{range}} \quad (3)$$

$$\mathrm{RPIQ} = \frac{Q_3 - Q_1}{\mathrm{RMSEP}} \quad (4)$$

where y_i and $\hat{y}_i$ are, respectively, the i_{th} measured value and the corresponding predicted value of the soil properties, N is the total number of samples; and range is the difference between the minimum and maximum content of the soil properties considered, and Q_1 and Q_3 are the first and third quartiles. The difference between the two represents the interquartile range.

Table 1 Emission lines of C, Si, Ca, and K used for calculating signal ratios

Element	Higher intensity emission line (nm)	Lower intensity emission line (nm)
Carbon (C)	247.82	193.01
Silicon (Si)	288.16	212.36
Calcium (Ca)	393.41	422.67
Potassium (K)	766.62	770.12

3. Results and discussion

3.1. Exploratory data analysis

The descriptive statistics for clay, silt, sand, and SOC in % are shown in Table 2. As the variability in the dataset has an implication on model performance,[47] the coefficient of variation (CV) was used to verify that the variability of soil properties was covered in both the calibration and validation datasets. Silt had the lowest variability (CV = 35%) in the validation set, followed by sand and clay, while the highest variability (CV = 75.22%) was shown by SOC.

As expected, a strong negative correlation was exhibited between silt and sand and clay and sand, whereas a moderate positive correlation between clay and silt and a weak negative correlation were shown between SOC and clay and SOC and sand, and a weak positive correlation was observed between SOC and silt (Fig. 1). We expected a moderate to strong positive correlation between SOC and clay due to the feasible association

Table 2 Descriptive statistics for clay, silt, sand, and SOC in %

Soil property	Dataset	Average	Min.	Max.	SD[c]	CV[d]	Q_1[e]	Q_3[f]
Clay	Full ($n = 305$)	14	1	42	7	52	8	18
	Cal.[a] ($n = 203$)	14	3	42	7	51	9	19
	Val.[b] ($n = 102$)	13	1	31	7	53	8	17
Silt	Full	44	3	69	14	31	35	56
	Cal.	45	15	69	13	29	36	57
	Val.	43	3	68	15	35	31	53
Sand	Full	42	7	96	19	45	27	57
	Cal.	40	7	82	18	45	26	55
	Val.	45	9	96	20	44	31	62
SOC	Full	5.39	0.26	17.90	3.84	71.09	2.59	7.08
	Cal.	5.20	0.29	17.44	3.54	68.14	2.79	6.79
	Val.	5.79	0.26	17.90	4.35	75.22	2.32	8.10

[a] Calibration data set. [b] Validation data set. [c] Standard deviation. [d] Coefficient of variation (SD/mean). [e] First quartile. [f] Third quartile.

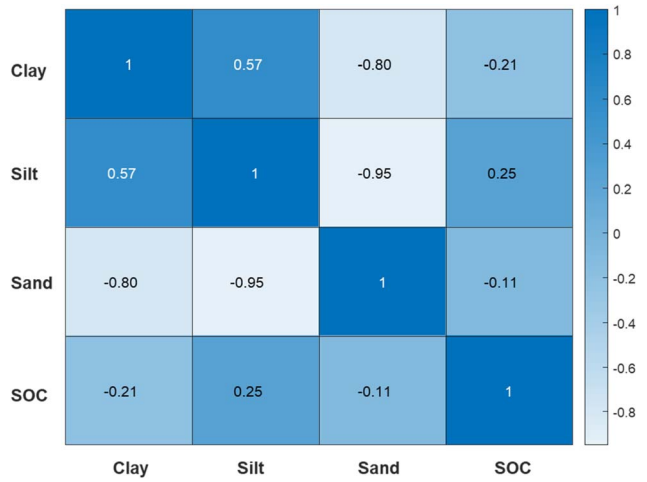

Fig. 1 Matrix of Pearson's correlation coefficients of investigated soil properties (clay, silt, sand, and SOC). Significant values are presented by a color gradient, ranging from light blue (negative correlations) to dark blue (positive correlations).

of organic matter with clay,[48] and/or between SOC and sand, as sandy soils frequently receive high amounts of organic fertilizers that increase the content of SOC.[49] A possible explanation of the discrepancy between our data and literature data might be ascribed to the vast geographical distribution of the soil samples examined that included land uses other than agricultural use.

Fig. 2 shows a representative soil LIBS spectrum covering the 187 nm to 950 nm wavelength range. The spectrum was characterized by several emission lines typical of soil, including those of Fe, Al, Ca, Mg, Si, Na, K, Li, Ti, and C. Some emission lines could be identified univocally, but peak overlap often occurred, thus making difficult the identification of emission lines at specific wavelengths.

3.2. Soil maps

Soil maps can distinguish soils based on the distribution of the intensity of specific LIBS emission lines, *e.g.*, soils containing higher SOC contents showed a higher intensity of C and Ca emission lines compared to that of soils with lower SOC contents (Fig. 3). These results suggested that it would be possible to rapidly classify soil samples based on the emission intensity of selected elements. In the case studied, the relationship between SOC content and C and Ca LIBS emission lines could be used to roughly differentiate between mineral and organic soils.

3.3. Repeatability and reproducibility assessment

3.3.1. Variation of signal ratios. The median RSD of C, Si, Ca, and K signal ratios evaluated for three measurements performed on the same day and on three different days for each sample was used as an indicator of the stability, *i.e.*, repeatability and reproducibility of the measurements (Fig. 4).

The median signal ratio RSD for measurements performed on the same day was approximately 9% for C, 6% for Si and Ca, and 4% for K. For measurements performed on three different days, the approximate median signal ratio RSD was 10% for C and Si, and 7% for Ca and K. Although potential outliers were present for the calculated Si and K signal ratios, which could be partly attributed to sample heterogeneity especially for the larger sand particles, a 4–9% variation in signal ratio was achieved for measurements performed on the same day (repeatability) and a 7–10% variation for measurement performed on three different days (reproducibility).

In a study by Xu *et al.*,[28] the emission lines of C, N, K, Ca, Mg, and Si were used to evaluate the effect of the shot layer and the number of shots on the quality of LIBS measurements. The approximate median RSD values referred to the shot layer were 35% for C, 30% for N, and 45% for K, Ca, Mg and Si, and the signal intensity variability was found to decrease with an increasing number of shots. Thus, even the median repeatability and reproducibility up to 10% in signal intensity ratio achieved in the present study could be considered satisfactory, *i.e.*, the measurements could be considered quite stable, especially because very different soils covering a wide geographical distribution were used, and the measurements were performed

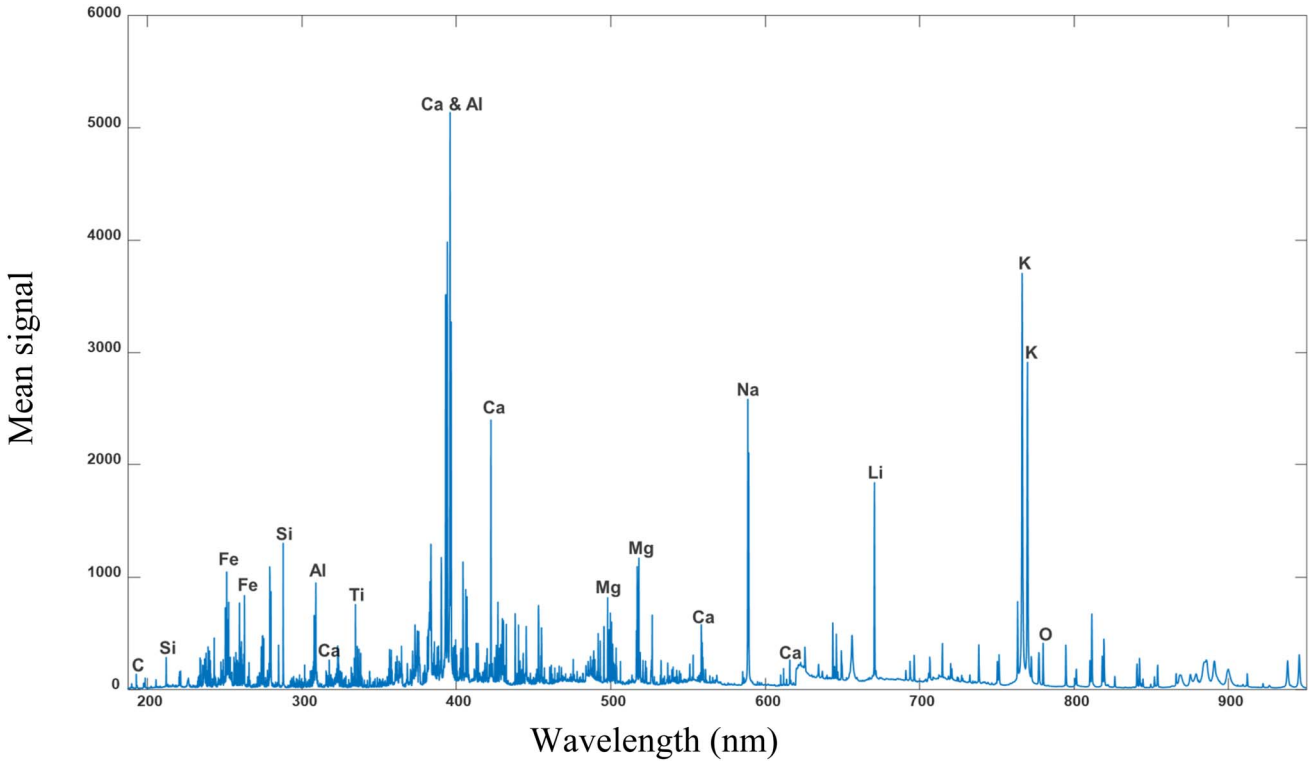

Fig. 2 A representative soil LIBS spectrum in the 187 to 950 nm range with the relevant emission lines identified.

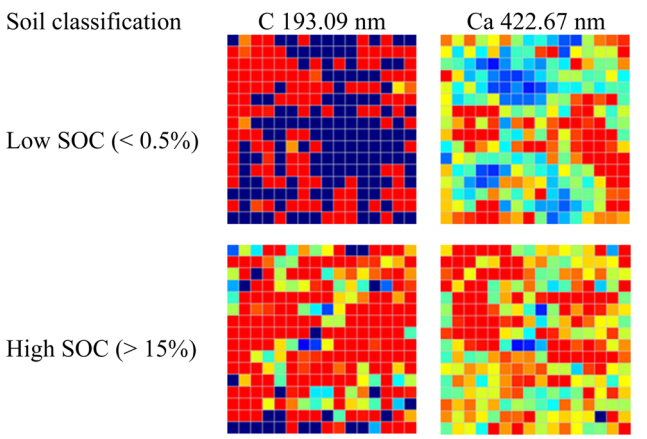

Fig. 3 Distribution of emission line intensity of C and Ca on the pellet surface (16 × 16 grid) of two soils classified as low and high SOC content. The emission intensity for C and Ca is shown in a color gradient that varies from blue (low relative abundance) to red (high relative abundance).

on three different days under different environmental conditions (Table S2†).

3.3.2. Comparing signal ratios for different-day measurements. The signal ratios of C, Si, Ca, and K were also evaluated for measurements performed on five different days that included three measurements performed on three different days on one set of sample pellets (set A) and two measurements performed on two different days on a different set of the same sample (set B) (Fig. 5).

The median signal ratio ranged from 1.7 to 1.9 for C, from 4.8 to 5.3 for Si, and from 1.3 to 1.4 for Ca and K. The highest variation in the signal ratio was observed for Si, while Ca and K showed the lowest variation (Fig. 5). The relatively higher variation of the Si signal ratio might be attributed to interferences caused by the nonuniform distribution of Si in the samples which might also be related to the large sand particles that are not uniformly distributed on the sample pellet.

In this work, the signal ratio variability was used as a measure of the shot-to-shot repeatability and reproducibility of the LIBS signal. The signal variability is influenced by several factors including the physical and chemical properties of the sample. Physical properties include surface homogeneity/heterogeneity, which in the case of soil, largely depends on texture.[50] As soils are highly heterogeneous, the difference in particle sizes, even after pelletization, influences plasma formation and, in turn, signal intensity.[51] Increasing the number of shots has been suggested as a strategy to improve the robustness of the mean LIBS spectrum and thus increase the prediction accuracy.[28] Although in this study the influence of the shot layer has not been evaluated as such, soil samples of different texture are expected to exhibit differences in shot layer depth due to differences in ablation characteristics.[50,52] These effects might thus result in a nonstoichiometric ablation of the sample thus increasing signal variability.[53] It is therefore realistic to assume that for LIBS soil measurements on pelletized samples, the spectra acquired from different soils would simulate those acquired at varying depths from the soil pellet surface.

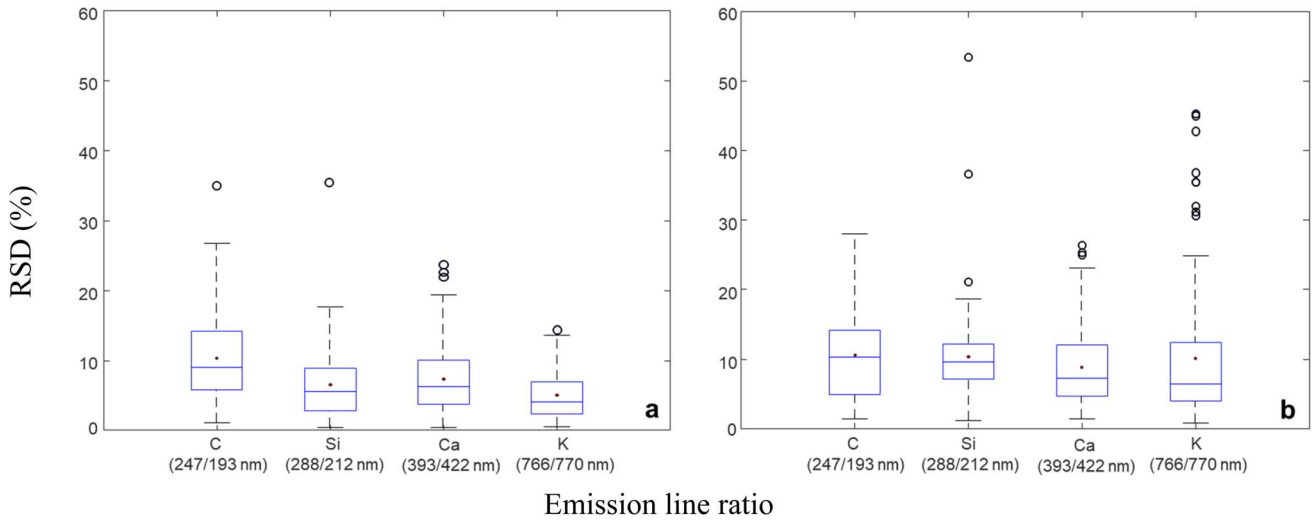

Fig. 4 Boxplots of the RSD distribution for C, Si, Ca, and K signal ratios among 101 soil samples for three measurements performed on the same day (a) and on three different days (b). The empty circles indicate potential outliers for the selected emission lines.

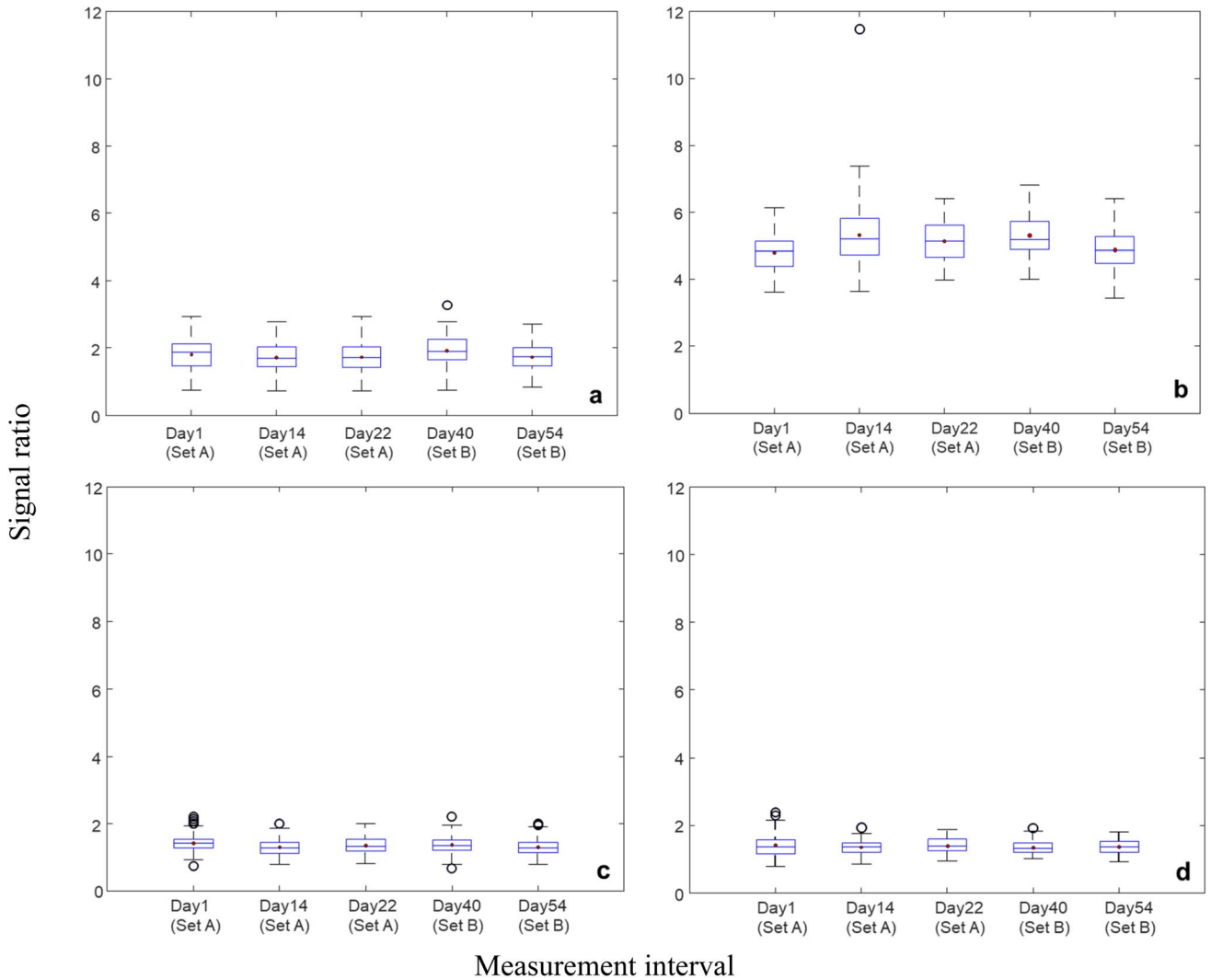

Fig. 5 Boxplots of the signal ratios for 101 soil samples referred to C 247/C 193 nm (a), Si 288/Si 212 nm (b), Ca 396/Ca 422 nm (c), and K 766/K 770 nm (d) for measurements performed on five different days and on two sets of sample pellets (set A and set B). The empty circles indicate potential outliers for selected emission lines.

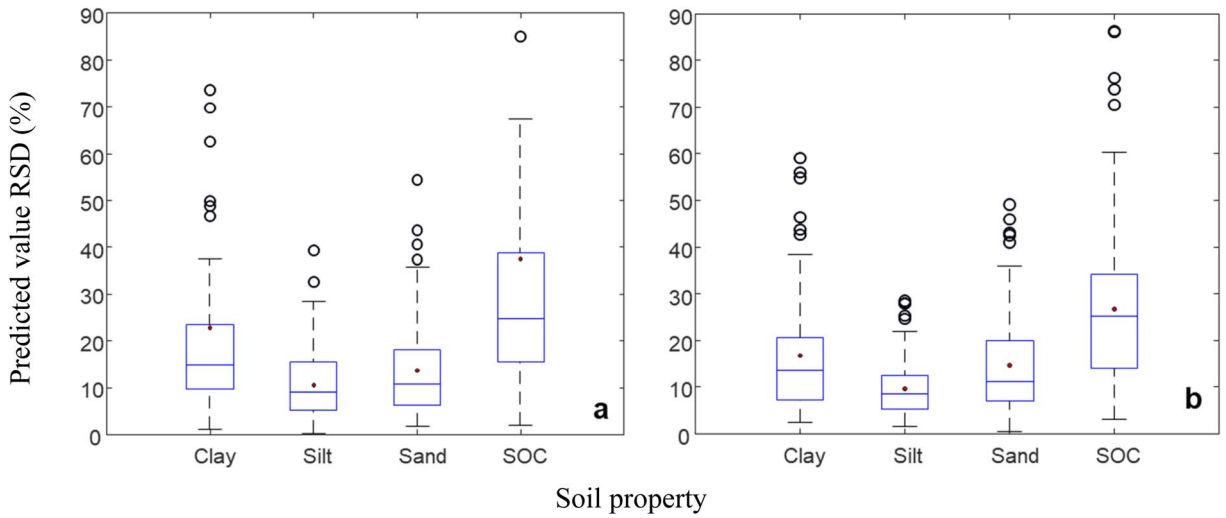

Fig. 6 Boxplots of the RSD distribution for the predicted values of clay, silt, sand, and SOC in 101 soil samples for measurements performed on the same day (a) and on three different days (b). The empty circles indicate potential outliers.

3.3.3. Prediction of soil properties. The variation of the predicted values for clay, silt, sand, and SOC for three measurements performed on the same day and on three different days was evaluated using the distribution of RSD for the measured 101 samples (Fig. 6). The prediction variability evaluated by the median RSD of three measurements per sample showed the same trend for measurements performed either on the same day or on three different days. In both cases, the median RSD of predicted values ranged from 10% to 25% for all the investigated soil properties. In particular, the SOC content showed the highest median RSD followed by clay, sand, and silt. These results appeared satisfactory if considering the wide range of geographical and soil properties distribution of the samples used.

A mean RSD of 2.04% was achieved for LIBS quantification of Ag in four different soil types using back-propagation neural network (BPNN).[54] In another study where a portable LIBS system and univariate method of prediction were used, RSD values of 7.69% and 12.98% were achieved in the analysis of total Cr and Cr VI in soil.[55] The higher median RSD of predicted values achieved for SOC and texture in this study, however, would be ascribed to the involvement of several emission lines contributing to the variation of the soil properties. For instance, in clay the variation is influenced by elements associated with its mineralogy (*e.g.*, Al, Si, Mg, and Fe), thus its prediction must consider more than one element emission line to account for the "matrix effect". Differently, when predicting the content of a single element such as Ag or Cr, the model needs to consider only the Ag or Cr emission lines, so that the negative impact of matrix effects on the model performance is strongly reduced.

3.3.4. Cross-validation models for same-day measurements. To evaluate the repeatability of the analytical method, the average spectrum of each of the three measurements performed on the same day was used to develop cross-validation models for each investigated soil properties, and the root mean square of cross validation (RMSECV) and R^2 of cross validation (cv) values were used to compare the individual measurements (Table 3).

Although the RMSECV values of the individual measurements 1, 2, and 3 differed among each property in model accuracy achieved for the same day, no statistically significant difference ($p > 0.05$) existed among the corresponding individual measurements for each investigated soil property. Measurement 3, however, exhibited a slight underperformance

Table 3 Cross validation results developed from the individual spectrum of three measurements and their average spectrum acquired on the same day

	Measurement 1		Measurement 2		Measurement 3		Average (3 measurements)	
Property	RMSECV[a] (%)	R^2 cv[b]	RMSECV (%)	R^2 cv	RMSECV (%)	R^2 cv	RMSECV (%)	R^2 cv
Clay	5	0.54	5	0.55	5	0.36	5	0.51
Silt	10	0.53	9	0.61	11	0.47	10	0.58
Sand	13	0.58	12	0.62	14	0.47	12	0.6
SOC	2.63	0.64	2.28	0.73	3.18	0.48	2.22	0.74

[a] Root mean square error of cross-validation. [b] R^2 of cross-validation.

Table 4 Cross validation results for measurements performed on five different days for set A and set B sample pellets

	Sample set A day 1		Sample set A day 14		Sample set A day 22		Sample set B day 40		Sample set B day 54	
Property	RMSECV[a] (%)	R^2 cv[b]	RMSECV (%)	R^2 cv	RMSECV (%)	R^2 cv	RMSECV (%)	R^2 cv	RMSECV (%)	R^2 cv
Clay	5	0.53	5	0.42	5	0.54	5	0.53	5	0.49
Silt	11	0.46	11	0.47	10	0.53	9	0.6	10	0.56
Sand	13	0.53	14	0.52	13	0.58	12	0.62	12	0.62
SOC	2.52	0.66	2.85	0.59	2.63	0.64	2.61	0.65	2.64	0.64

[a] Root mean square error of cross-validation. [b] R^2 of cross-validation.

compared to measurements 1 and 2, as shown by the lower R^2 for all investigated soil properties and the higher RMSECV for silt, sand, and SOC. These results might be attributed to the surface heterogeneity of the sample pellet analyzed as only a relatively small area of it (approximately 2 mm^2) compared to the entire pellet surface (approximately 154 mm^2) was ablated during each measurement (Fig. S2†).

Furthermore, the slight reduction of RMSECV and SOC and the slight increase of R^2 achieved for sand and SOC if the average spectrum from the three measurements was used to develop the cross-validation models might be attributed to the increased number of shots, which has been shown to better account for matrix effects, thus improving predictions.[28]

3.3.5. Cross-validation models for different-day measurements. To evaluate the reproducibility of the analytical method, cross-validation models for measurements performed on five different days were developed and compared based on RMSECV and R^2 (Table 4).

Also in this experiment, although the RMSECV values of the individual measurements 1, 2, and 3 performed on sample sets A and B differed among each property in model accuracy achieved on different days, no statistically significant difference ($p > 0.05$) existed among the corresponding individual measurements for each investigated soil property. The slight underperformance of day 14 measurements for silt and sand might be attributed to the higher moisture recorded on that day as compared to other days (Table S2†). Moisture is known to affect LIBS signal intensity depending on soil types.[56] This effect was also slightly visible for the Si signal ratio, which was higher on day 14, as compared to the other days (Fig. 5b). This effect on day 14, however, was not observed for the other selected emission lines.

In a study by Ebinger *et al.*,[26] the reproducibility of C quantification in soil was assessed based on the ratio of the C 193.09 nm emission line to the sum of Al 199 nm and Si 212 nm lines using 6 samples for 30 different days to develop the calibration curve, and 12 validation samples. The correlation achieved for the validation set was 0.95 and no significant difference of the calibration slopes was measured for each of the 30 days. The lower performance achieved in this study might be due to the much higher number and distribution of samples used and the highly variable SOC contents. Furthermore, Ebinger *et al.*[26] used dry combustion as the reference method, but it is not clear if they have discussed total C or organic C.

In particular, the model accuracy achieved for clay was relatively similar for all measurement days, which suggested a stable clay model that might be ascribed to the homogeneity of the fine clay particles, as compared to that of the larger silt and sand particles that are unevenly distributed on the pellet surface.

3.4. Prediction and performance of LIBS modeling

To assess the overall performance of hLIBS, prediction models were developed for clay, silt, sand, and SOC. The PLSR results are shown in Table 5 while the iPLSR and ANN results are shown in the ESI (Table S3).†

PLSR is the most applied data analysis method for spectral data analysis.[57,58] We therefore focused the discussion of the performance of the prediction models on the PLSR models. Furthermore, there was not a remarkable improvement, in terms of accuracy, when iPLSR or ANN was applied.

The sand prediction model showed the lowest prediction error (SRMSEP = 0.14), followed by clay and silt (SRMSEP = 0.15), and SOC (SRMSEP = 0.16). All the prediction models of investigated soil properties featured an approximate R^2 pred. of 0.6, which indicated a moderate correlation between the predicted and measured value of the corresponding property. In this work, the risk of overfitting was reduced by using a relatively small number of latent variables (3 to 5 latent variables/

Table 5 Cross-validation and prediction results for clay, silt, sand, and SOC in %

Property	RMSECV[a] (%)	R^2 cv[b]	RMSEP[c] (%)	R^2 pred.[d]	LV[e]	SRMSEP[f]	RPIQ[g]
Clay	6	0.36	4	0.58	3	0.15	2.0
Silt	9	0.55	10	0.55	5	0.15	2.2
Sand	12	0.55	12	0.62	5	0.14	2.5
SOC	2.17	0.62	2.77	0.63	4	0.16	2.1

[a] Root mean square error of cross-validation. [b] R^2 of cross-validation. [c] Root mean square error of prediction. [d] R^2 of prediction. [e] Number of factors (latent variables) applied in the model. [f] Standardized root mean square error of prediction. [g] Ratio of performance to interquartile distance.

components) in the model of each investigated soil property. Similar to the SRMSEP results, the assessment of model performance using the RPIQ showed the prediction of sand to be superior (higher RPIQ) compared to other investigated soil properties.

The lowest number of latent variables was used for the clay model, which achieved an RMSEP of 4% and R^2 pred. of 0.58. The slight underprediction of samples featuring a clay content > 20% (Fig. 7a) might be due a possible nonlinearity in the dataset. The clay results of this study are worse than those achieved by Erler *et al.*,[12] who obtained an R^2 (prediction) of 0.9 and RMSEP of 3.09, for clay content with a range of 28.84%, thus an SRMSEP of 0.11, using a hLIBS and PLSR (variable selection) for predicting the clay content in soils from a field in Germany. The different prediction performance might be reasonably attributed to the different dataset scale used in terms of geographical distribution and soil properties investigated. To improve the prediction accuracy, information on soil types and geological similarities (*e.g.*, clay mineralogy) should be used to classify samples before modeling[59] and ensure that soils of similar geological origin are considered in the calibration and validation datasets.

The prediction models for silt and sand, which were quite similar in terms of dominant mineralogy (quartz), were comparable, in terms of SRMSEP. The prediction accuracy of silt in this work was lower (SRMSEP = 0.15) than that achieved by Erler *et al.*[12] (SRMSEP = 0.11), which confirmed the findings of previous studies performed using benchtop and laboratory LIBS setups that have reported a poor silt prediction model compared to those achieved for clay and sand.[60,61] Silt and sand textural fractions are characterized by large soil-grain sizes, thus a means to improve their prediction using LIBS would be the milling of the soil sample before pelletization and measurement, which would promote sample homogeneity and reduce physical and chemical matrix effects. However, the inclusion of a milling step would affect seriously the accurate distinct quantification of textural fractions, and limit the benefits of using hLIBS as a faster analysis technique in the field.

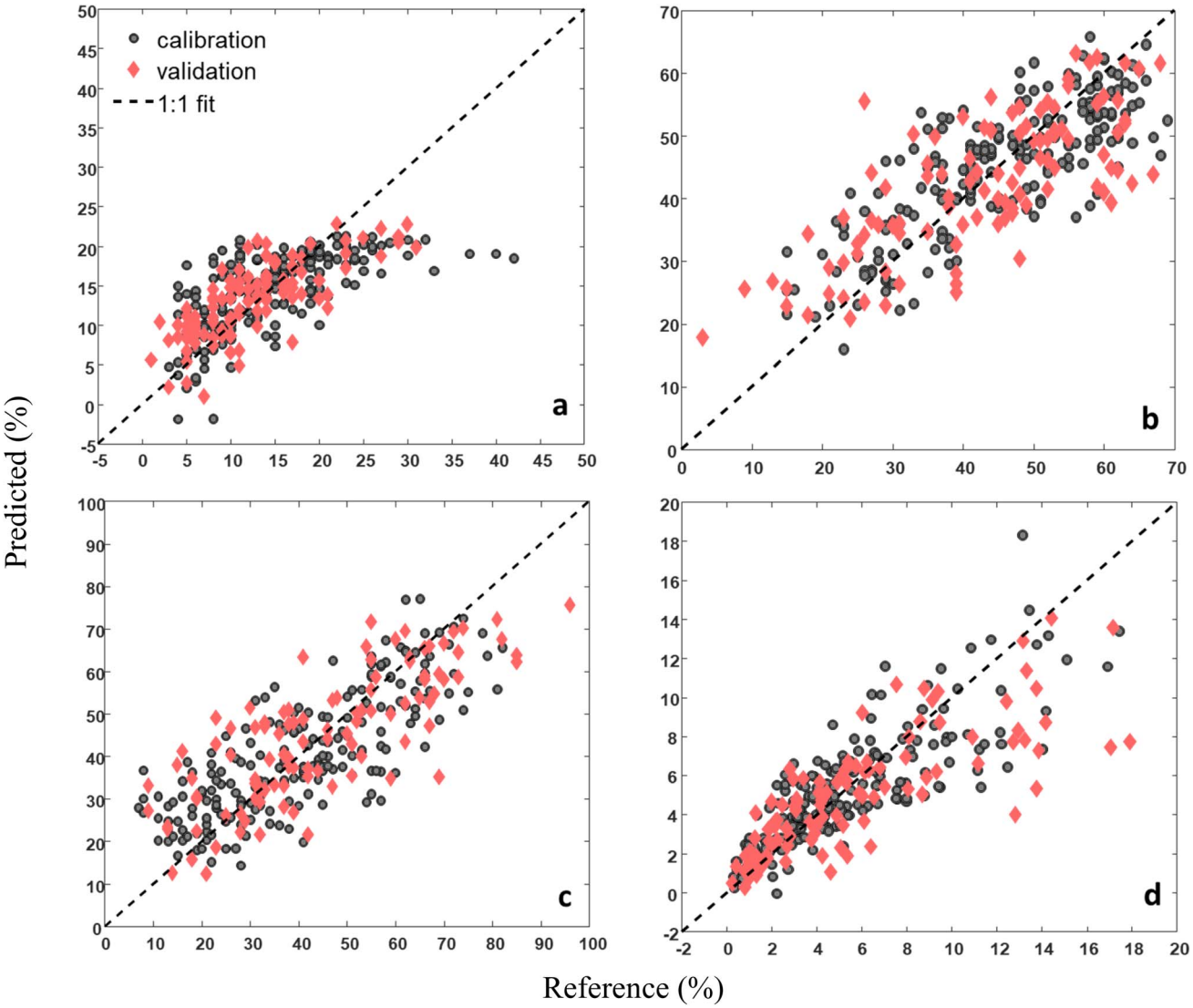

Fig. 7 PLSR plots for clay (a), silt (b), sand (c), and SOC (d).

The high prediction error of the SOC model might be ascribed to the higher variation of the validation set (CV = 75%) compared to that of the calibration set (CV = 68%) (Table 2). In particular, a substantial underprediction was found for samples with a SOC content > 10% (Fig. 7d), which indicated a possible nonlinearity in the corresponding dataset. For example, Knadel *et al.*[60] reported a high prediction error for silt that was attributed to a higher variability of the silt content in the validation dataset compared to that of the calibration dataset.

Approximately 40% of the samples used in this study contained carbonates; thus another factor that might impact the prediction of SOC is the accuracy of reference SOC values, which depend on the accurate extraction of carbonates from the soil sample (especially alkaline soils) by acid treatment and their subsequent determination by titration. As the value of SOC measured by the dry combustion method should be corrected by subtracting inorganic C (*i.e.*, carbonates) from total C to ensure that only SOC content is considered,[38] any error-prone step, such as carbonate extraction and titration, would increase the uncertainty of the SOC reference values so diminishing SOC prediction accuracy.

Several previous studies have dealt with the application of handheld LIBS for C analysis in soils. For instance, da Silva *et al.*[15] predicted the total C content in six Brazilian soils achieving a cross-validation correlation coefficient of 0.91. Glumac *et al.*[16] used a portable LIBS system to predict the SOC content in six soils from US agricultural farms achieving an R^2 of 0.94. In a comparison of three field-based methods, Izaurralde *et al.*[17] achieved an R^2 of 0.92 for predictions of the C content in soils from two fields in USA and Mexico using a portable LIBS system. The three studies mentioned above used the C emission line intensity to develop calibrations for predicting C in soil, whereas a different approach was used in this study, which involved the selection of various emission lines explanatory for the variation of the investigated soil properties and the subsequent use of PLSR as a multivariate data analysis method.

Overall, the prediction accuracy evaluated from the calculated SRMSEP was comparable for both textural fractions and SOC. For clay, silt and sand, this result might be explained by the obvious relationship existing between the various textural fractions, *e.g.*, the more the clay content, the less the sand content in a soil. Although a weak correlation was found between texture, especially clay, and SOC, the regression vector plots for the individual models suggested that the same emission lines could explain the variability of each soil property studied (Fig. 8). In particular, results of previous studies,[62,63] the LIBS database embedded in the SciAps proprietary Profile Builder software and the NIST database, suggest that the emission lines of the elements C, Al, Fe, Si, Ti, Ca, Na, Li and K are able to explain the variability and influence the prediction models of clay, silt, sand and SOC (Fig. 8).

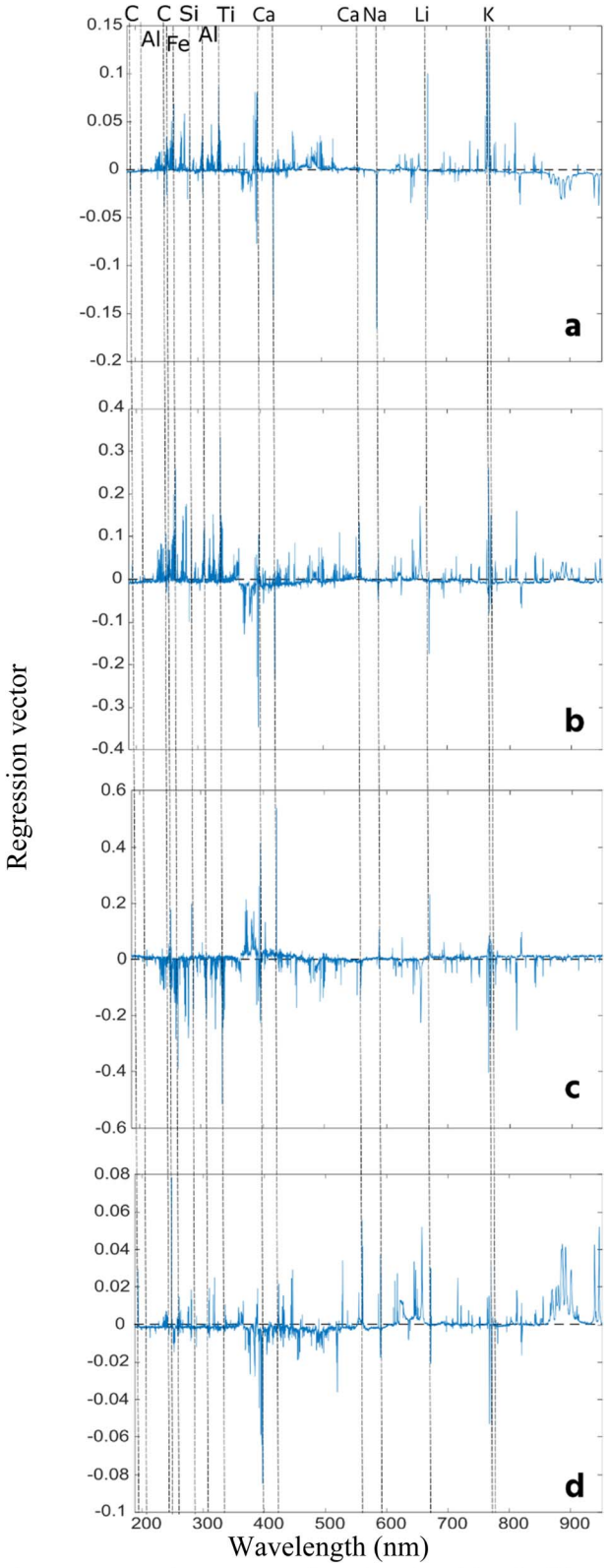

Fig. 8 Regression vector plots for clay (a), silt (b), sand (c), and SOC (d).

3.5. Regression vector analysis

The prediction of clay was mainly influenced by the highly negative regression vector score of Na 588.96 nm, and the highly positive scores of Ca 422.67, Li 670.99 nm, and K 766.81 nm, and 770.12 nm and the less-prominent negative scores of C 193.09 and 247.82 nm. The slight influence of C emission lines might be due to the clay complexation with SOC, while the effect

of Ca and K emission lines might be associated with clay mineralogy, and that of Na and Li emissions with the parent rock material.

The silt variability was characterized by the highly positive regression vector scores of Si 288, Al 309.29 nm, Ti 334.9 nm and K 766.81 nm and 770.12, the highly negative scores of Ca 396.19 nm and 422.67 nm, and Li 670.99 nm, and the less-prominent positive regression vector score of C 193.09 nm.

The variability of sand was influenced by the highly negative regression vector scores of Ti 334.9 nm and K 766.81 nm and 770.12 nm, and the prominent positive scores of Ca 396.19 nm and 422.67 nm, and Li 670.99 nm, the negative scores of Si 288 nm and Ca 558.82 nm, the positive scores of Fe 251.58 nm and Na 588.82 nm and the less-prominent scores of C 193.09 nm and Si 212 nm. Apparently, the sand regression vector plot included the most identifiable emission lines that influence the variability of sand, *i.e.*, the model was more efficient in relating the elements influencing sand variability, thus resulting in a more accurate sand prediction.

Finally, SOC variability was influenced by the highly positive regression vector scores of C 193.09 nm and 247.82, Ca 558.82 nm, Na 588.82 nm and Fe 900.37 nm, the highly negative scores associated with Ca 396.19 nm and K 766.81 nm and K 770.12 nm, and the negative scores of Fe, Mg, and Li and the less-prominent positive scores associated with Al, Ca, and Ti. Apparently, Ca explained most of SOC variability, which might be ascribed to the influence of carbonates. In a previous study, SOC was reported to have a minor correlation with rock-forming elements such as Al, Fe, Ca, Ti and Si, and a major correlation with Mg.[64] Differently, the results of this study showed that most emission lines in the SOC regression vector plot were associated with rock-forming elements as compared to Mg emission lines, which suggested that the variability of SOC was largely influenced by inorganic C.

4. Conclusion

The performance of handheld LIBS for predicting soil texture components, *i.e.*, clay, sit and sand and SOC in a wide range of soils covering a continental scale was assessed based on the repeatability and reproducibility of the LIBS signal from selected elemental emission lines. Furthermore, the prediction accuracy of the models used was compared for the soil properties investigated.

Overall, the repeatability achieved by handheld LIBS for predicting soil texture and SOC, as assessed by the median RSD of the signal ratios ranged from 4 to 9%, while the reproducibility ranged from 7 to 10%, whereas the repeatability and reproducibility for predicting soil texture and SOC were <25%. Furthermore, handheld LIBS exhibited a relatively stable performance even under changing environmental conditions such as temperature and humidity, as shown by the minimal (between 0.1 and 0.5) signal ratio variations of the selected C, Si, Ca, and K emission lines for measurements conducted over five days. The higher prediction error and signal ratio measured during a high-humidity day suggested that moisture is an important environmental factor affecting LIBS analysis. Of the investigated soil properties, the sand prediction model exhibited the lowest error followed by clay and silt models, which were comparable, and SOC. The higher prediction error for SOC might be attributed to the high variability of SOC content and the different land use practices.

To improve the prediction accuracy of handheld LIBS, future studies should focus on the classification of samples by integrating information about soil mineralogy before modeling. Incorporating soil maps could potentially enable rapid soil classifications, or prescreening of samples before detailed analysis, based on the relative abundance of elements related to the soil properties of interest. In conclusion, the results of this study confirm the promising potential of handheld LIBS as a tool for large-scale applications to determine agronomically related soil properties that would enable timely farm interventions.

Data availability

The data supporting this article have been included as part of the ESI.†

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This research received funding from the Innovation Fund Denmark (grant no. 0153-00146B). Additional funding for this work was covered by the European Union. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the Research Executive Agency (REA) as granting authority. Neither the European Union nor the granting authority can be held responsible for them. We thank the European Commission, Joint Research Centre for providing the samples used in this study.

References

1 N. H. Batjes, *Eur. J. Soil Sci.*, 1996, **47**(2), 151–163.
2 H. H. Janzen, *Agric. Ecosyst. Environ.*, 2004, **104**(3), 399–417.
3 A. Bauer and A. L. Black, *Soil Sci. Soc. Am. J.*, 1994, **58**(1), 185–193.
4 T. Maruthaiah, S. K. Vajravelu, V. Kaliyaperumal and D. Kalaivanan, *Optik*, 2023, **278**, 170691.
5 W. M. Post and K. C. Kwon, *Global Change Biol.*, 2000, **6**(3), 317–327.
6 I. Barra, S. M. Haefele, R. Sakrabani and F. Kebede, *TrAC, Trends Anal. Chem.*, 2021, **135**, 116166.
7 J. A. M. Demattê, A. C. Dotto, L. G. Bedin, V. M. Sayão and A. B. e. Souza, *Geoderma*, 2019, **337**, 111–121.
8 A. B. McBratney, M. L. Mendonça Santos and B. Minasny, *Geoderma*, 2003, **117**(1), 3–52.
9 G. Nicolodelli, J. Cabral, C. R. Menegatti, B. Marangoni and G. S. Senesi, *TrAC, Trends Anal. Chem.*, 2019, **115**, 70–82.

10 G. S. Senesi, J. Cabral, C. R. Menegatti, B. Marangoni and G. Nicolodelli, *TrAC, Trends Anal. Chem.*, 2019, **118**, 453–469.
11 G. S. Senesi, R. S. Harmon and R. R. Hark, *Spectrochim. Acta, Part B*, 2021, **175**, 106013.
12 A. Erler, D. Riebe, T. Beitz, H. G. Lohmannsroben, M. Leenen, S. Patzold, M. Ostermann and M. Wojcik, *Sensors*, 2023, **23**(16), 7178.
13 F. J. Fortes and J. J. Laserna, *Spectrochim. Acta, Part B*, 2010, **65**(12), 975–990.
14 R. T. Wainner, R. S. Harmon, A. W. Miziolek, K. L. McNesby and P. D. French, *Spectrochim. Acta, Part B*, 2001, **56**(6), 777–793.
15 R. M. da Silva, D. M. B. P. Milori, E. C. Ferreira, E. J. Ferreira, F. J. Krug and L. Martin-Neto, *Spectrochim. Acta, Part B*, 2008, **63**(10), 1221–1224.
16 N. G. Glumac, W. K. Dong and W. M. Jarrell, *Soil Sci. Soc. Am. J.*, 2010, **74**(6), 1922–1928.
17 R. C. Izaurralde, C. W. Rice, L. Wielopolski, M. H. Ebinger, J. B. Reeves 3rd, A. M. Thomson, R. Harris, B. Francis, S. Mitra, A. G. Rappaport, J. D. Etchevers, K. D. Sayre, B. Govaerts and G. W. McCarty, *PLoS One*, 2013, **8**(1), e55560.
18 G. S. Senesi and N. Senesi, *Anal. Chim. Acta*, 2016, **938**, 7–17.
19 B. Connors, A. Somers and D. Day, *Appl. Spectrosc.*, 2016, **70**(5), 810–815.
20 J. Cunat, F. J. Fortes and J. J. Laserna, *Anal. Chim. Acta*, 2009, **633**(1), 38–42.
21 A. Erler, D. Riebe, T. Beitz, H. G. Lohmannsroben and R. Gebbers, *Sensors*, 2020, **20**(2), 18.
22 D. Meng, N. Zhao, M. Ma, L. Fang, Y. Gu, Y. Jia, J. Liu and W. Liu, *Appl. Opt.*, 2017, **56**(18), 5204–5210.
23 Z. Hao, L. Liu, M. Shen, R. Zhou, J. Li, L. Guo, X. Li, Y. Lu and X. Zeng, *J. Anal. At. Spectrom.*, 2018, **33**(9), 1564–1570.
24 Y. Tian, H. C. Cheung, R. Zheng, Q. Ma, Y. Chen, N. Delepine-Gilon and J. Yu, *Spectrochim. Acta, Part B*, 2016, **124**, 16–24.
25 D. W. Hahn and N. Omenetto, *Appl. Spectrosc.*, 2012, **66**(4), 347–419.
26 M. H. Ebinger, M. L. Norfleet, D. D. Breshears, D. A. Cremers, M. J. Ferris, P. J. Unkefer, M. S. Lamb, K. L. Goddard and C. W. Meyer, *Soil Sci. Soc. Am. J.*, 2003, **67**(5), 1616–1619.
27 Y. He, X. Liu, Y. Lv, F. Liu, J. Peng, T. Shen, Y. Zhao, Y. Tang and S. Luo, *Sensors*, 2018, **18**(5), 1526.
28 X. Xu, C. Du, F. Ma, Y. Shen, Y. Zhang and J. Zhou, *Soil Sci. Soc. Am. J.*, 2020, **84**(4), 1307–1326.
29 S. Wold, M. Sjöström and L. Eriksson, *Chemom. Intell. Lab. Syst.*, 2001, **58**(2), 109–130.
30 V. Costa, D. Babos, J. Castro, D. Andrade, R. Gamela, R. Machado, M. Sperança, A. Araújo, J. Garcia and E. Pereira-Filho, *J. Braz. Chem. Soc.*, 2020, **31**, 2439–2451.
31 D. Riebe, A. Erler, P. Brinkmann, T. Beitz, H. G. Lohmannsroben and R. Gebbers, *Sensors*, 2019, **19**(23), 5244.
32 T.-L. Zhang, S. Wu, H.-S. Tang, K. Wang, Y.-X. Duan and H. Li, *Chin. J. Anal. Chem.*, 2015, **43**(6), 939–948.
33 E. C. Ferreira, D. M. B. P. Milori, E. J. Ferreira, R. M. Da Silva and L. Martin-Neto, *Spectrochim. Acta, Part B*, 2008, **63**(10), 1216–1220.
34 K. H. Lepore, C. I. Fassett, E. A. Breves, S. Byrne, S. Giguere, T. Boucher, J. M. Rhodes, M. Vollinger, C. H. Anderson, R. W. Murray and M. D. Dyar, *Appl. Spectrosc.*, 2017, **71**(4), 600–626.
35 S. Sanchez-Esteva, M. Knadel, R. Labouriau, G. H. Rubaek and G. Heckrath, *Appl. Spectrosc.*, 2021, **75**(1), 22–33.
36 A. Wangeci, D. Adén, T. Nikolajsen, M. H. Greve and M. Knadel, *Geoderma*, 2024, **444**, 116865.
37 B. Minasny and A. B. McBratney, *Comput. Geosci.*, 2006, **32**(9), 1378–1388.
38 A. Orgiazzi, C. Ballabio, P. Panagos, A. Jones and O. Fernández-Ugalde, *Eur. J. Soil Sci.*, 2017, **69**(1), 140–153.
39 A. Jones, O. F. Ugalde and S. Scarpa, *LUCAS 2015 Topsoil Survey, EUR 30332 EN*, European Commission JRC, Luxembourg, 2020.
40 ISO, *ISO 13320:2009 Particle Size Analysis Laser Diffraction Methods*, International Organization for Standardization, Geneva, 2009.
41 ISO, *ISO 10694:1995 Soil Quality Determination of Organic and Total Carbon after Dry Combustion (Elementary Analysis)*, International Organization for Standardization, Geneva, 1995.
42 ISO, *ISO 10693:1995 Soil Quality Determination of Carbonate Content Volumetric Method*, International Organization for Standardization, Geneva, 1995.
43 R. J. Barnes, M. S. Dhanoa and S. J. Lister, *Appl. Spectrosc.*, 1989, **43**(5), 772–777.
44 P. H. Eilers, *Anal. Chem.*, 2003, **75**(14), 3631–3636.
45 H. Martens and T. Næs, *Multivariate Calibration*, Wiley, 1992.
46 V. Bellon-Maurel, E. Fernandez-Ahumada, B. Palagos, J.-M. Roger and A. McBratney, *TrAC, Trends Anal. Chem.*, 2010, **29**(9), 1073–1081.
47 D. Brunet, B. G. Barthès, J.-L. Chotte and C. Feller, *Geoderma*, 2007, **139**, 106–117.
48 A. R. Dexter, G. Richard, D. Arrouays, E. A. Czyż, C. Jolivet and O. Duval, *Geoderma*, 2008, **144**(3), 620–627.
49 K. Adhikari, A. E. Hartemink, B. Minasny, R. Bou Kheir, M. B. Greve and M. H. Greve, *PLoS One*, 2014, **9**(8), e105519.
50 K. M. Donaldson and X. T. Yan, *Geoderma*, 2019, **337**, 701–709.
51 R. S. Harmon, F. C. DeLucia Jr, A. LaPointe, R. J. Winkel Jr and A. W. Miziolek, *Anal. Bioanal. Chem.*, 2006, **385**(6), 1140–1148.
52 A. Wangeci, D. Adén, M. H. Greve and M. Knadel, *Spectrochim. Acta, Part B*, 2023, **206**, 106712.
53 V. C. Costa, S. dos Santos Ferreira, L. N. Santos, M. A. Sperança, C. S. da Silva, G. A. Sodré and E. R. Pereira-Filho, *J. Appl. Spectrosc.*, 2020, **87**(2), 378–386.
54 C. Sun, Y. Tian, L. Gao, Y. Niu, T. Zhang, H. Li, Y. Zhang, Z. Yue, N. Delepine-Gilon and J. Yu, *Sci. Rep.*, 2019, **9**(1), 11363.
55 X. Fu, S. Ma, G. Li, L. Guo and D. Dong, *Spectrochim. Acta, Part B*, 2020, **167**, 105817.
56 A. M. Popov, S. M. Zaytsev, I. V. Seliverstova, A. S. Zakuskin and T. A. Labutin, *Spectrochim. Acta, Part B*, 2018, **148**, 205–210.

57 J. El Haddad, L. Canioni and B. Bousquet, *Spectrochim. Acta, Part B*, 2014, **101**, 171–182.
58 D. Fernandes Andrade, E. R. Pereira-Filho and D. Amarasiriwardena, *Appl. Spectrosc. Rev.*, 2020, **56**(2), 98–114.
59 B. Bousquet, G. Travaillé, A. Ismaël, L. Canioni, K. Michel-Le Pierrès, E. Brasseur, S. Roy, I. le Hecho, M. Larregieu, S. Tellier, M. Potin-Gautier, T. Boriachon, P. Wazen, A. Diard and S. Belbèze, Development of a mobile system based on laser-induced breakdown spectroscopy and dedicated to in situ analysis of polluted soils, *Spectrochim. Acta, Part B*, 2008, **63**(10), 1085–1090.
60 M. Knadel, R. Gislum, C. Hermansen, Y. Peng, P. Moldrup, L. W. de Jonge and M. H. Greve, *Biosyst. Eng.*, 2017, **156**, 157–172.
61 P. R. Villas-Boas, R. A. Romano, M. A. de Menezes Franco, E. C. Ferreira, E. J. Ferreira, S. Crestana and D. M. B. P. Milori, *Geoderma*, 2016, **263**, 195–202.
62 M. C. Pelagio, D. A. Navarro, L. J. Janik and R. B. Lamorena, *ChemistrySelect*, 2020, **5**(13), 3798–3804.
63 A. Kramida, Y. Ralchenko, J. Reader and NIST ASD Team, *NIST Atomic Spectra Database (Version 5.10)*, National Institute of Standards and Technology, 2023.
64 M. Z. Martin, M. A. Mayes, K. R. Heal, D. J. Brice and S. D. Wullschleger, *Spectrochim. Acta, Part B*, 2013, **87**, 100–107.