

ORIGINAL ARTICLE

Fundamental Soil Science

Optimization of sample preparation and data evaluation techniques for X-ray fluorescence prediction of soil texture, pH, and cation exchange capacity of loess soils

Isabel Greenberg¹  | Anja Sawallisch¹ | Jan Stelling² | Michael Vohland^{3,4,5}  | Bernard Ludwig¹ 

¹Department of Environmental Chemistry, University of Kassel, Witzenhausen, Germany

²Bruker AXS GmbH, Karlsruhe, Germany

³Geoinformatics and Remote Sensing, Institute for Geography, Leipzig University, Leipzig, Germany

⁴Remote Sensing Centre for Earth System Research, Leipzig University, Leipzig, Germany

⁵German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany

Correspondence

Isabel Greenberg, Department of Environmental Chemistry, University of Kassel, 37213 Witzenhausen, Germany. Email: isabel.greenberg@uni-kassel.de

Assigned to Associate Editor Tetsuhiro Watanabe.

Funding information

Deutsche Forschungsgemeinschaft, Grant/Award Numbers: LU 583/24-1, VO 1509/10-1

Abstract

Use of X-ray fluorescence (XRF) spectrometry for estimation of soil texture, pH, and cation exchange capacity (CEC) is desirable given the time-consuming nature of traditional methods. Recent studies have shown promising results; however, further investigation is required to determine the effects of sample preparation and data evaluation techniques on accuracy. Our objective was to compare (I) a simple but well-founded approach, combining measurement of powder samples and modeling with elemental contents as predictors in stepwise multiple linear regressions (MLR), with alternative approaches including (II) use of partial least squares regression (PLSR), (III) sample preparation as a pressed pellet, and (IV) spectral intensities as predictors (20 kV, 40 kV, and concatenated 20 + 40 kV). A total of 395 loess soils from three arable fields were used with a fivefold random training-testing approach and a hold-one-site-out training-testing approach. With random partitioning, clay, silt, and sand accuracy with approach I was excellent (ratio of performance to interquartile distance in validation (RPIQ_v) = 8.5–12.9), while pH and CEC estimations were satisfactory to excellent (RPIQ_v = 2.0–2.5 and 2.2–3.3, respectively). Differences between MLR and PLSR were negligible. Increases in accuracy with pellet samples were 1%–13% of RPIQ_v for 20 kV intensities, but effects were inconsistent for other predictors. The optimal predictor varied by property, and differences ranged from 3% to 13% of RPIQ_v. Improvements to accuracy from Approach I to the best alternative were largest for texture (10%–15%) but may be superfluous given the excellent accuracy across all approaches. Although the leave-one-site-out training resulted in variable performance, inclusion of soils from the target site in training assured reliable accuracy.

Abbreviations: CEC, cation exchange capacity; ED-XRF, energy dispersive X-ray fluorescence; MLR, multiple linear regression; PLSR, partial least squares regression; RMSE, root mean squared error; RPIQ, ratio of performance to interquartile distance; XRF, X-ray fluorescence.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Soil Science Society of America Journal* published by Wiley Periodicals LLC on behalf of Soil Science Society of America.

1 | INTRODUCTION

The use of sensors for prediction of soil properties has gained attention due to the time-consuming nature of traditional laboratory analyses. In this content, X-ray fluorescence (XRF) spectrometry—which detects fluorescence at discrete energies resulting from the interaction of X-rays and atoms to quantify the total elemental contents of a sample—has proven to be very useful. XRF is already a well-established approach for quantifying the contents of heavy metal contamination in soils (Nawar et al., 2020). However, in recent years, more wide-ranging applications for XRF have been explored as devices have become increasingly accurate and versatile. Modern-day energy dispersive X-ray fluorescence (ED-XRF) devices are capable of accurately detecting the contents of elements from sodium (Na) to uranium (U) (Haschke et al., 2021)—a range that includes the elements that make up the majority of the soil matrix as well as a number of plant macro- and micronutrients. Thus, soil properties that co-vary with total elemental contents from Na to U can be predicted by XRF following model calibration with values determined by traditional laboratory methods. While soil organic fractions are therefore dominated by elements too light to be quantified by XRF, prediction of soil inorganic fractions and the properties they strongly influence is of interest.

Recent studies have therefore explored application of XRF for the estimation of soil texture (Silva et al., 2020), pH (Javadi & Mouazen, 2021), and cation exchange capacity (CEC) (Tavares et al., 2020) with promising results. The mechanisms by which XRF can predict contents of the soil particle size classes are based on the unique assemblages of elements present in soil minerals associated with coarse versus fine particles. For example, Zhu et al. (2011), O'Rourke, Stockmann et al. (2016), and Tavares et al. (2020) found iron (Fe) content was useful for the prediction of clay content across a diverse range of soils. This is explained by the presence of Fe in minerals in the clay fraction, for example, in the Fe-oxides goethite and hematite and as an interlayer cation in montmorillonite and vermiculite (Sparks, 2003). Prediction mechanisms for pH and CEC are more complex due to their dependence on not only inorganic but also organic soil components, which are dominated by elements too light to be quantified by XRF (Haschke et al., 2021). Nevertheless, Tavares et al. (2020) and Silva et al. (2017) both found calcium (Ca) content to be a useful predictor for pH and CEC, which could be due to the presence of Ca-containing carbonates (e.g., calcite and dolomite) and other Ca-containing primary minerals (e.g., amphiboles and pyroxenes) that contribute to base saturation as they weather, buffering soil acidity, and retained in soil by CEC (Sparks, 2003).

Since relationships between elemental contents and soil properties may be site or region specific, it is essential that studies demonstrate model performance both with and with-

Core Ideas

- With random partitioning into training and test sets, XRF estimation of soil texture, pH, and cation exchange capacity was successful.
- Effects of preparation (powder vs. pellet), predictor (elements vs. spectra), and algorithm (multiple linear regression vs. partial least squares regression) were minor.
- Accuracy improvements with optimized approaches were largest for texture prediction but may be superfluous.
- Inclusion of soils from the target site in training improved the reliability of models.

out soils from the target site in model calibration to inform best practices. For example, infrared spectroscopy studies have demonstrated high model robustness in independent validation for spectrally active soil properties (e.g., clay), but that local calibration is often essential for spectrally inactive properties (e.g., pH and CEC) (Soriano-Disla et al., 2014). While a large number of XRF studies predicting soil properties have been conducted, typically only one model calibration/validation strategy has been applied in each study (e.g., Javadi & Mouazen, 2021; Kandpal et al., 2022; Li et al., 2022; O'Rourke, Minasny et al., 2016; O'Rourke, Stockmann et al., 2016; Sharma et al., 2015; Tavares et al., 2020; Zhu et al., 2011). Thus, a mechanistic investigation of the robustness of XRF predictions for various soil properties with and without independent validation is required.

Further optimizations of XRF application are also worth exploring, including the benefits of various approaches to sample preparation and data evaluation. The degree of sample preparation affects the trade-off between the accuracy and efficiency of soil characterization with XRF (Li et al., 2022). While field measurement of intact soil samples saves time and expense (Goff et al., 2020), enabling a much higher spatiotemporal resolution of data to be collected, soil moisture content absorbs X-rays and scatters primary radiation, resulting in poor detection of light elements (Haschke et al., 2021) and thus lower model accuracy. Sample representativeness is also a challenge with intact soil samples (Stockmann, Cattle et al., 2016). To overcome these issues, use of dried and sieved (Xu et al., 2019) or ground (O'Rourke, Minasny et al., 2016; Towett et al., 2015) soil samples are common approaches to sample preparation for recent XRF studies indirectly predicting soil texture or fertility attributes. Finally, studies implementing XRF to achieve highly accurate determination of elemental contents have often measured pressed pellets formed from ground soil material (Byers et al., 2016;

Goff et al., 2020; Zhao et al., 2013). This approach, applied with or without the use of additives (e.g., cellulose and wax) for binding, enables higher accuracy due to the smoother sample surface and more uniform sample density (Haschke et al., 2021). However, it remains to be explored whether the indirect estimation of soil texture and fertility parameters by XRF can be meaningfully improved by measurement of pressed pellets versus powder samples.

Regarding the predictors used to model soil properties with XRF, it is possible to use either the fluorescence spectral intensities directly, or instead use total elemental contents quantified by spectral deconvolution from the spectra. Spectral deconvolution involves calculation of the area of element-specific fluorescence peaks, which may be overlapping with peaks of other elements and scattering peaks (Brouwer, 2010). Thus, this approach is based on knowledge of physicochemical relationships, and therefore is capable of handling spectral artifacts arising from the X-ray tube source and the detection process (Beckhoff et al., 2006; Haschke et al., 2021). In studies utilizing XRF to indirectly estimate soil properties, a common approach is to use these elemental contents as predictors in multiple linear regression (MLR), often with stepwise model simplification to eliminate elements whose inclusion does not substantially lower the error of prediction (O'Rourke, Stockmann et al., 2016; Sharma et al., 2015; Wang et al., 2013; Zhu et al., 2011).

Rather than calculating elemental contents first to use as model predictors, an alternative approach is to directly use intensities across the XRF spectra (Javadi & Mouazen, 2021; Kandpal et al., 2022; Xu et al., 2019). Direct use of the spectral intensities can be considered an empirical modeling approach with a high number of correlated predictors, containing both useful information as well as spectral artifacts and noise. In this context, partial least squares regression (PLSR) is a useful linear modeling approach which has often been applied in the field of infrared soil spectroscopy (Soriano-Disla et al., 2014). PLSR calculates orthogonal latent variables to maximize the covariance between independent variables (spectral intensities) and dependent variables (measured laboratory data) and eliminate irrelevant information (Wehrens, 2020). Although not a standard approach, modeling with elemental data achieved by spectral deconvolution could also be implemented with the PLSR algorithm, as the elemental contents are also often highly correlated and contain both useful information as well as noise. Thus, a comparison of models created using spectral predictors with PLSR versus elemental predictors with both PLSR and MLR could be useful to separate the effect of using different model predictors as well as different algorithms.

The use of elemental contents versus spectral intensities also raises considerations about how ED-XRF measurements at different voltages should be handled. Measurements at two or more voltages are often carried out due to the complementary nature of the information collected: measurement at

lower voltages (e.g., 15–20 kV) is optimal for detection of lighter elements due to higher background-to-peak ratios in the lower energy portion of the spectra, whereas measurement at higher voltages (e.g., 35–50 kV) provides sufficient excitation for optimal detection of heavier elements (Haschke et al., 2021). While elemental contents are typically determined using a single spectrum with optimal voltage for a given atomic number, the optimal approach using spectral intensities as predictors is unclear. While O'Rourke, Minasny et al. (2016) and O'Rourke, Stockmann et al. (2016) analyzed both single and concatenated spectral intensities measured at three voltages (15, 40, and 50 kV), some studies do not specify the voltage(s) of the measurements used in modeling with ED-XRF (Xu et al., 2019) or do not specify whether spectra measured at both 15 and 40 kV were utilized in modeling (Javadi & Mouazen, 2021; Kandpal et al., 2022). Due to the time-consuming nature of consecutive analyses of multiple single spectrum or concatenated spectra, more information on the improvements from including measurements with various voltages may be beneficial.

Our objective was therefore to optimize the use of ED-XRF for the prediction of soil properties by starting with Approach I—measurement of powder samples, with use of XRF elemental contents as predictors, and modeling with stepwise MLR—and then systematically testing changes to model accuracy resulting from alternative approaches. Approach I may be regarded as the simplest and most well-founded of the tested approaches because it is the least time-consuming in terms of both sample preparation and modeling due to the smaller number of predictors, and due to its reliance on physicochemical relationships to extract elemental information from the XRF spectra rather than empirical modeling with the spectral intensities. The alternative approaches included (II) use of another algorithm (PLSR), (III) sample preparation as a press pellet; and (IV) use of spectral intensities as predictors (20 kV, 40 kV, and concatenated 20 + 40 kV spectra). To test these approaches, we utilized a set of 395 loess soils from three arable fields for which wet chemistry analyses were already conducted (Greenberg et al., 2022) and subsequently carried out XRF measurements on powder and pressed sample material to address the aforementioned research objective. Finally, we compared model performance following random dataset partitioning into training and testing sets with a leave-one-site-out approach to determine the robustness of the models at independent sites.

2 | MATERIALS AND METHODS

2.1 | Field sampling and laboratory analysis

The soils under investigation came from three loess-derived arable fields in eastern Germany. The soil types according to the World Reference Base for Soil (IUSS Working Group

WRB, 2022) included a silt Haplic Luvisol in Lüttewitz, a silt loam Gleyic Luvisol in Zschortau, and a silty clay loam Haplic Phaeozem in Friemar. Luvisols are categorized as Alfisols, and Phaeozems are categorized as Mollisols according to the US Soil Taxonomy (Jelinski et al., 2023; Soil Survey Staff, 2022). Long-term tillage experiments have been carried out at each of these sites by the Südzucker AG in cooperation with the Institute of Sugar Beet Research at the University of Göttingen since 1992 (Lüttewitz and Freimar) and 1997 (Zschortau) including conventional tillage (i.e., annual moldboard plowing to a depth of 0.3 m), reduced tillage (i.e., mixing tillage with a cultivator and/or disc harrow to a depth of 0.1–0.15 m), and no till (i.e. direct seeding) (Koch et al., 2009). For the present study, the tillage treatments are only relevant because they increased the field-scale heterogeneity of certain soil properties. At each site, a crop rotation consisting of sugar beet (*Beta vulgaris* L.)—winter wheat (*Triticum aestivum* L.)—winter wheat was cultivated, with white mustard (*Sinapis alba* L.) grown as a cover crop following the second winter wheat harvest. Crop management was in accordance with recommended agricultural practices for the region, and mineral fertilizers (nitrogen [N], phosphorus [P], potassium [K], and magnesium [Mg]) were applied equally across all treatments in line with results of electro-ultrafiltration soil analysis. Thus, soils from these three sites have both similar parent material (loess) and long-term agricultural management.

In September and October 2018, sample material was collected at $n = 50$ points per tillage treatment at each site, resulting in a total of $n = 150$ per site and a grand total of $n = 450$ across all three sites, as described in Greenberg et al. (2022). Soils from each combination of site and tillage treatment were taken from a 2-ha area with regular grid sampling (see Greenberg et al. (2022) for a map of the fields and sampling design). At each sampling point, a 15 cm \times 15 cm area was cleared of crop residues, and soil was collected to a depth of 2 cm. In the present study, only $n = 395$ sample units were analyzed ($n = 123$ from Friemar and Zschortau each and $n = 149$ from Lüttewitz) due to lack of sufficient material following the previous experiment for the rest of the sample units.

Soils were sieved to <2 mm and dried at 40°C for 48 h. Total carbon (C) contents were measured on ball-milled samples with a CN elemental analyzer (Elementar Vario El, Heraeus). Contents of inorganic carbon (IC) were determined with the Scheibler Method. Briefly, 10% HCl was added to the dried, ball-milled soil and the evolved CO₂ was measured volumetrically. Organic C (OC) content was calculated by subtracting the IC from total C. Soil texture was determined with the pipette method (DIN ISO 11277, 2002). The pH values were determined with 2.5 g field-moist soil in 6.25 mL 0.01 M CaCl₂ (DIN ISO 10390, 2005). For determination of CEC, the soil was first slowly leached with 0.1 M BaCl₂, with a soil to solution ratio of 1:10. Next, exchangeable K⁺, Na⁺, Ca²⁺,

and Mg²⁺ were measured in the filtered extracts with ion chromatography (850 Professional IC, 237 Metrohm) and CEC was calculated as the sum of exchangeable cations (Koenig & Fortmann, 1996). A total of 226 of the 395 soils in the current experiment contained carbonate, which can be dissolved during cation extraction with BaCl₂, leading to overestimation of CEC in highly calcareous soils (Jaremko & Kalembasa, 2014). To determine the relevance of this source of error for the carbonate-containing soils, we compared the amount of Ca potentially bound to carbonates in this soil—quantified by the Scheibler Method—with the amount of Ca extracted with BaCl₂. The amount extracted with BaCl₂ was two to four orders of magnitude higher than the amount potentially contained within and thus dissolved from carbonates. This source of error was therefore considered to be negligible in the current experiment.

2.2 | Spectral measurements

Prior to XRF measurements, 12 g of soil for each sample unit were ball-milled using a Retsch MM 400 with 10 ZrO₂ balls at 30 Hz for 5 min. A test of this milling method using other soils found an average particle diameter of ca. 13 μ m, with 90% of the particles with a diameter <33 μ m. From this, 5 g was taken for XRF measurement as powder and subsequently recombined with the rest of the milled soil prior to formation of the press pellets. For this, 11 g of powder sample were mixed with 1.87 g Hoechst Wax C Micropowder (C₃₈H₇₆N₂O₄) from MERCK KGaA, resulting in pellets with 17% wax on a mass basis (DIN EN 15309, 2007). Soils and wax were thoroughly mixed in the ball mill with two polyamide balls for 10 min at 12 Hz. The mixtures were then pressed in a HTP 40 half-automated press (Herzog Maschinenfabrik GmbH & Co. KG) for 39 s in total, building up for 10 s to a maximum pressure of 200 kN, which was held for 10 s, and then released over 9 s, resulting in pellets with a 40-mm diameter and 5-mm height, which matches the recommended pressure to diameter ratio of Haschke et al. (2021).

XRF measurements were made with the bench-top Bruker S2 PUMA EDXRF spectrometer. This device has a silver X-ray tube with a maximum power of 50 W, a Peltier-cooled HighSense Silicon Drift Detector which can detect approximately 300,000 counts per second (cps) output count rate. Measurements were conducted in helium (He) atmosphere with reduced pressure of 0.2 bar. Intensities were recorded across 4096 channels across energies from 0 to 44 keV, using two measurement ranges as follows: (1) 20 kV and 150 s spectral acquisition time with 25% dead time on average and (2) 40 kV and 200 s spectral acquisition time with 39% dead time on average. For 20 kV, the current was automatically adjusted to 0.59 and 0.65 mA on average for pressed and powder samples, respectively, and 1.00 and 1.20 mA on

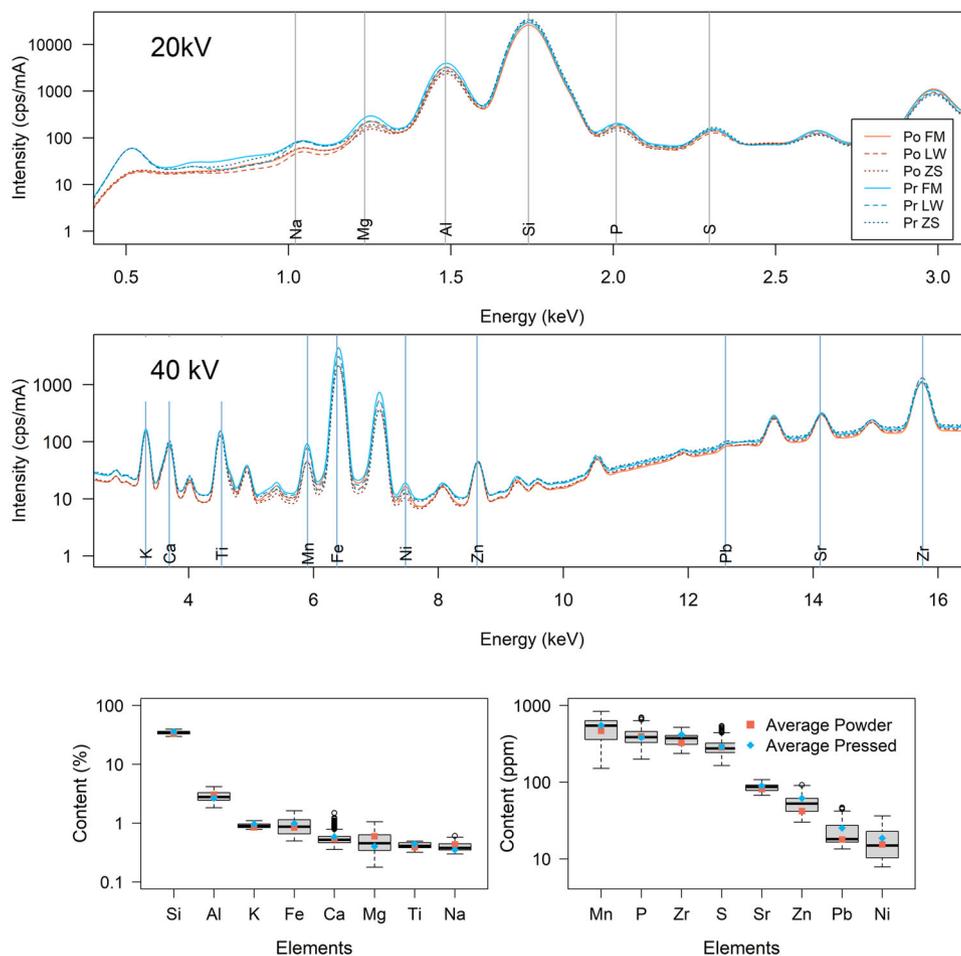


FIGURE 1 Average 20 and 40 kV X-ray fluorescence (XRF) spectra for Friemar (FM), Lüttewitz (LW), and Zschortau (ZS) for powder (Po) or pressed (Pr) soils. $K\alpha_1$ lines for each element are shown, besides for Pb, which utilized the $L\beta_1$ line. Boxplots show elemental contents across the three sites determined by XRF measurement of Po or Pr soil samples.

average for pressed and powder samples, respectively, at 40 kV. For both sample preparations, sample cups with a diameter of 3.6 cm and a mask size of 3.4 cm were used. Cups were sealed with a SpectroMembrane Prolene Thin-Film with 4.0- μm thickness (Chemplex Industries, Inc.) and rotated during measurement. Two replicate measurements per sample unit were taken with each method (powder and press pellet) and values (estimates of elemental contents or spectral intensities) were averaged.

For the analysis approach utilizing the 20 and/or 40 kV spectra, raw spectra were first standardized by converting total count data over the duration of the measurement period [measurement time = total time – (total time \times percentage of dead time)] to counts per second per mA (Figure 1). Due to periodic drift corrections that were required during the XRF measurements, the measured energy channels shifted very slightly. Measured intensities were thus resampled using the resample function of the prospectr package (Stevens et al., 2022) with the interpolation method set to “spline” in order to have consistent wavelength predictors throughout the dataset.

For the XRF approach utilizing elemental contents, an internal calibration was performed using acid digestion followed by inductively coupled plasma optical emission spectroscopy (ICP-OES) as a reference method and the Bruker AXS SPECTRA.ELEMENTS software to correct XRF elemental contents accordingly (version 2.4, Bruker AXS GmbH). A total of 43 sample units used for internal calibration were selected to cover the variability in CEC existing across the three fields. The reference values for total elemental contents were obtained by acid digestion with HNO_3 , HCl , HClO_4 , and HF followed by ICP-OES, as described in Greenberg et al. (2023). Silicon (Si) contents, which cannot be measured using ICP-OES due to loss during acid evaporation, were obtained via wavelength dispersive XRF on using a Malvern Panalytical Axios advanced spectrometer (Rh X-ray tube) and the software SuperQ (version 4) at the University of Göttingen (Greenberg et al., 2023). For this, measurements were made on glass disks containing 0.42 g soil mixed with 4.2 g of A12 flux (66% di-lithium tetraborate and 34% lithium metaborate). Si was measured using a PE 002C crystal, a

550- μm collimator and a flow detector. Measurements (60 s counting time) were performed at a peak angle of 109.0258 ($2\theta^\circ$) and with a voltage of 25 kV and a current of 160 mA. Three soil standards covering the SiO_2 range of the investigated samples as well as JR-1, a Geological Survey of Japan international rock standard with similar SiO_2 , were processed and measured with the sample set. For all standards, SiO_2 was within 3.9% of the accepted values (Govindaraju, 1994, see Table S1). With these reference methods, 18 elements had concentrations above the detection limit. The Bruker AXS SPECTRA.ELEMENTS software was then utilized to create calibrations between the elemental contents and the spectral intensities, with the 20 kV spectra used for contents of lighter elements (Na, Mg, aluminum [Al], Si, P, and sulfur [S]) and the 40 kV spectra for heavier elements (K, Ca, titanium [Ti], chromium [Cr], manganese [Mn], Fe, nickel [Ni], copper [Cu], zinc [Zn], strontium [Sr], zirconium [Zr], and lead [Pb]). A total of 16 elements were ultimately used in the subsequent prediction models. Cr was eliminated due to low R^2 between measured and estimated contents within the internal calibration ($R^2 = 0.30$ for powder samples and 0.25 for pressed samples), and Cu was eliminated due to negative contents predicted for some sample units in the set of $n = 395$ soils, which is implausible. For other elements, R^2 values for the internal calibration were comparable for pressed versus powder samples for most elements, with $R^2 \geq 0.9$ for Mg, Al, Si, P, Mn, Fe, Zn, and Sr, $R^2 \geq 0.8$ for S, K, Ti, Ni, and Pb, $R^2 \geq 0.7$ for Na, and ≥ 0.5 for Zr (see Table S2 for R^2 values by element and preparation method). Deviations between the reference values and XRF elemental contents could result from minor differences in the 50 mg subsample used for acid digestion compared to the samples measured by XRF, despite fine milling. This internal calibration was then used to determine element concentrations for all $n = 395$ soils. In contrast to our approach, it is typical to apply correction factors to XRF elemental contents using standard reference materials from a national bureau of standards (Soil Survey Staff, 2014). This may be less accurate than our chosen approach since the ranges of elemental contents and sample preparations would be less well-suited to the target dataset, but certainly more time and cost-effective than performing a site-specific internal calibration.

Figure 1 shows boxplots of the total elemental contents estimated from the XRF spectra.

2.3 | Modeling

The goal of modeling was to predict soil clay, silt, and sand content, as well as pH and CEC from the XRF elemental contents or spectral intensities. For all modeling approaches, a fivefold model training and testing procedure was carried out using $n = 300$ randomly selected sample units for train-

ing and $n = 95$ sample units for model testing. Additionally, to test the robustness of Approach I (MLR models for powder samples with elemental predictors) as well as the single most accurate approach for application at unknown sites, a leave-one-site-out approach to training and testing was also conducted, whereby two sites were used for model training, and the third site was used for testing. Modeling was performed with the statistical software R version 4.2.2 (R Core Team, 2022).

2.3.1 | Stepwise multiple linear regressions

Modeling with XRF elements was first performed with MLR with stepwise simplification. For this, XRF elemental data were first z-transformed to ease comparison of the importance of the element predictors in the models despite vastly different magnitudes. MLR began with all 16 elements using the `lm()` command, and models were subsequently simplified using the `step()` command of the “stats” package based on the Akaike Information Criterion (AIC) calculated as follows:

$$\text{AIC}_{\text{MLR}} = n \times \log_e \left(\frac{\text{SSE}}{n} \right) + 2k,$$

where n is the sample size, SSE is the sum of squares of the error, and k is the number of parameters. This resulted in optimal regressions for each of the five training sets, which were then applied to the test sets and evaluated also according to the performance measures described above. We also calculated the variance inflation factor (VIF) to determine the degree of multicollinearity between our model predictors using the `vif` function of the “car” package (Fox & Weisberg, 2023).

2.3.2 | Partial least squares regressions

PLSR was carried out for both elemental data and spectral data using the “pls” package (Mevik et al., 2019). Variants for the spectral data included separate PLSR with the 20 kV spectral intensities, the 40 kV spectral intensities, and a concatenated matrix containing both. Although methods of spectral pretreatment are well-developed for soils measured by infrared spectroscopy, standardized approaches have not yet been established for XRF (Xu et al., 2019). Transformations with the Savitzky–Golay algorithm have, however, shown some usefulness for XRF (Javadi & Mouazen, 2021; O’Rourke, Stockmann et al., 2016; Xu et al., 2019). Thus, 13 spectral pretreatments were carried out with the “prospectr” package (Stevens et al., 2022), including use of the full spectra without manipulation, calculation of moving averages (calculated over 5, 11, 17, or 23 data points), and application of the Savitzky–Golay algorithm for the reduction of noise applied with the polynomial degree (PD) set to 2, the order

TABLE 1 Pretreatments applied to X-ray fluorescence (XRF) spectral intensities using the “prospectr” package (Stevens et al., 2022) and the Savitzky–Golay algorithm for the reduction of noise.

Variant	Smoothing window	Polynomial degree	Derivative
1	0	0	0
2	5	0	0
3	11	0	0
4	17	0	0
5	23	0	0
6	5	2	1
7	11	2	1
8	17	2	1
9	23	2	1
10	5	2	2
11	11	2	2
12	17	2	2
13	23	2	2

of the derivative (DER) ranging from 1 to 2 (with PD–DER: 2–1 or 2–2), and a window smoothing size of 5, 11, 17, or 23 (Table 1). Using the best spectral pretreatment determined in model training compared to no spectral pretreatment decreased error of prediction in the validation sets by 4.9% on average across all tested models. Thus, only results from models utilizing spectral pretreatments will be presented.

To determine the optimal number of latent variables for PLSR using either elemental or spectral predictors, training included leave-one-out cross-validation for each of the five training folds. The maximum number of latent variables was set to 15. To create more robust, parsimonious models, the optimal number of latent variables was determined in cross-validation by considering minimization of Akaike information criterion (AIC) (Viscarra Rossel & Behrens, 2010) which is calculated as follows:

$$AIC_{PLSR} = n \times \log_e (RMSE) + 2v,$$

where n is the sample size, v is the number of latent variables, and root mean squared error (RMSE) is calculated as follows:

$$RMSE = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n}},$$

where y_i is the measured soil property, \hat{y}_i is the modeled soil property, and n is the sample size. The optimal spectral pretreatment was the model with the highest ratio of performance to interquartile distance (RPIQ) in cross-validation, calculated as follows:

$$RPIQ = \frac{IQR}{RMSE},$$

where IQR is the interquartile range of the measured soil property. RPIQ was calculated rather than ratio of prediction to deviation (RPD) due to non-normality of the distributions of the measured soil properties for this set of $n = 395$ soils according to the Shapiro–Wilk test. Results were evaluated according to the classification system of Chang et al. (2001) by converting the RPD classification system to RPIQ values. For a normally distributed variable and large sample size, $RPIQ = 1.89$ corresponds to $RPD = 1.4$ and $R^2 = 0.5$. Thus, a model with $RPIQ < 1.89$ is considered poor, $RPIQ = 1.89–2.70$ is satisfactory, and $RPIQ > 2.70$ is very good. Nevertheless, the usefulness of the models must also be determined based on the context in which they are applied. The optimal PLSR models generated in training for the five dataset partitions were then tested with the validation sets. The performance measures RMSE and RPIQ were calculated for the validation sets of both PLSR and MLR models for comparison.

3 | RESULTS AND DISCUSSION

In presenting and discussing the results, we will begin with results from random dataset partitioning into training and testing sets applied with Approach I, a simple and well-founded approach using MLR models for powder samples with elemental predictors. This approach will then be compared with alternative approaches sequentially, including variation of the algorithm applied, sample preparation, and predictor type. Finally, Approach I and the most accurate alternative approach for each soil property will be applied with a hold-one-site-out training-testing procedure to determine the robustness of these models at independent sites.

3.1 | Performance of MLR models for powder samples with elemental predictors

For Approach I applied with random dataset partitioning into training and testing sets, texture predictions were excellent for all model partitions, with average and range of RMSE in validation (RMSE_v) for clay of 1.1% (0.9%–1.3%), silt of 2.0% (1.8%–2.3%), and sand of 1.8% (1.7%–2.1%), and average and range of RPIQ in validation (RPIQ_v) for clay of 11.3 (9.1–12.9), silt of 9.6 (8.5–10.7), and sand of 13.2 (12.7–14.8) (Table 2, Figure 2). The plots of measured versus estimated contents (Figure 3) show the accurate and unbiased predictions produced by this approach.

To aid in interpretation of the prediction mechanisms, Figure 4a shows the interrelation between all measured soil properties and Figure 4b shows the interrelation between measured soil properties and the elemental contents for powder samples, as well as the average slopes

TABLE 2 Average (Avg.) and standard deviation (SD) of performance in model testing for five random dataset partitions with eight approaches to predicting clay, silt, and sand content as well as pH and cation exchange capacity (CEC).

Property	Predictor	Preparation	Algorithm	RMSE _v		RPIQ _v	
				Avg.	SD	Avg.	SD
Clay (%)	Elements	Powder	MLR	1.1	0.2	11.3	1.7
	Elements	Powder	PLSR	1.1	0.1	11.2	1.5
	Elements	Pressed	PLSR	1.1	0.2	11.0	1.7
	20 kV	Powder	PLSR	1.0	0.2	11.8	2.1
	20 kV	Pressed	PLSR	1.0	0.2	12.3	2.4
	40 kV	Powder	PLSR	1.0	0.1	11.7	1.4
	40 kV	Pressed	PLSR	1.0	0.1	11.5	1.6
	20 + 40 kV	Powder	PLSR	1.0	0.1	12.6	2.2
Silt (%)	Elements	Powder	MLR	2.0	0.2	9.6	0.9
	Elements	Powder	PLSR	2.1	0.2	9.4	0.9
	Elements	Pressed	PLSR	2.1	0.3	9.4	1.4
	20 kV	Powder	PLSR	2.0	0.2	9.5	0.7
	20 kV	Pressed	PLSR	1.8	0.2	10.5	0.8
	40 kV	Powder	PLSR	2.0	0.1	9.8	0.5
	40 kV	Pressed	PLSR	2.0	0.2	9.7	1.1
	20+40 kV	Powder	PLSR	1.9	0.2	10.3	0.8
Sand (%)	Elements	Powder	MLR	1.8	0.2	13.2	1.7
	Elements	Powder	PLSR	1.8	0.2	13.3	1.8
	Elements	Pressed	PLSR	1.7	0.2	13.8	1.8
	20 kV	Powder	PLSR	1.7	0.1	13.4	1.5
	20 kV	Pressed	PLSR	1.5	0.2	15.2	2.3
	40 kV	Powder	PLSR	1.6	0.2	14.3	1.5
	40 kV	Pressed	PLSR	1.6	0.1	14.5	1.2
	20+40 kV	Powder	PLSR	1.7	0.1	14.0	1.1
pH	Elements	Powder	MLR	0.34	0.05	2.17	0.17
	Elements	Powder	PLSR	0.33	0.05	2.18	0.15
	Elements	Pressed	PLSR	0.35	0.05	2.11	0.20
	20 kV	Powder	PLSR	0.33	0.06	2.22	0.19
	20 kV	Pressed	PLSR	0.33	0.06	2.24	0.14
	40 kV	Powder	PLSR	0.34	0.05	2.14	0.12
	40 kV	Pressed	PLSR	0.32	0.03	2.24	0.12
	20 + 40 kV	Powder	PLSR	0.31	0.05	2.32	0.14
CEC (mmol_c kg⁻¹)	Elements	Powder	MLR	22	3.7	2.61	0.39
	Elements	Powder	PLSR	23	4.3	2.59	0.43
	Elements	Pressed	PLSR	22	4.1	2.68	0.49
	20 kV	Powder	PLSR	23	3.7	2.53	0.37
	20 kV	Pressed	PLSR	22	3.6	2.63	0.38
	40 kV	Powder	PLSR	23	3.9	2.56	0.41
	40 kV	Pressed	PLSR	21	3.1	2.76	0.35
	20 + 40 kV	Powder	PLSR	23	4.0	2.52	0.41

Note: Results of Approach I (use of elemental predictors for powder samples with multiple linear regression [MLR]) and the best performing approach are bold. Abbreviations: PLSR, partial least squares regression; RMSE_v, root mean squared error in validation; RPIQ_v, ratio of prediction to deviation in validation.

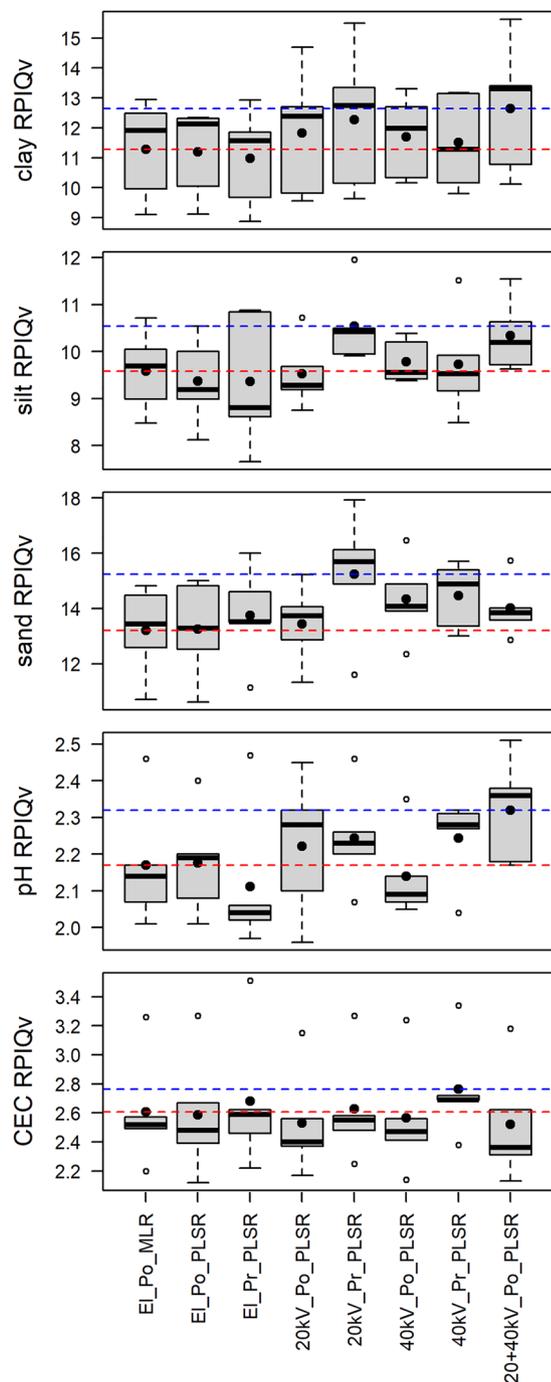


FIGURE 2 Boxplots of ratio of performance to interquartile distance in validation (RPIQv) for five random dataset partitions depending on the predictor type (elemental contents [EI], 20 kV, 40 kV, or concatenated [20 + 40 kV] spectral intensities), sample preparation (powder [Po] or pressed [Pr] samples), and the algorithm applied (multiple linear regression [MLR] or partial least squares regression [PLSR]). Means of each variant are given with the black dots. Outliers, shown with open circles, are defined as points lying 1.5 times the interquartile range above or below the interquartile range. The red dashed line shows the mean of approach I (EI_Po_MLR) and the blue dashed line shows the mean of the optimal approach.

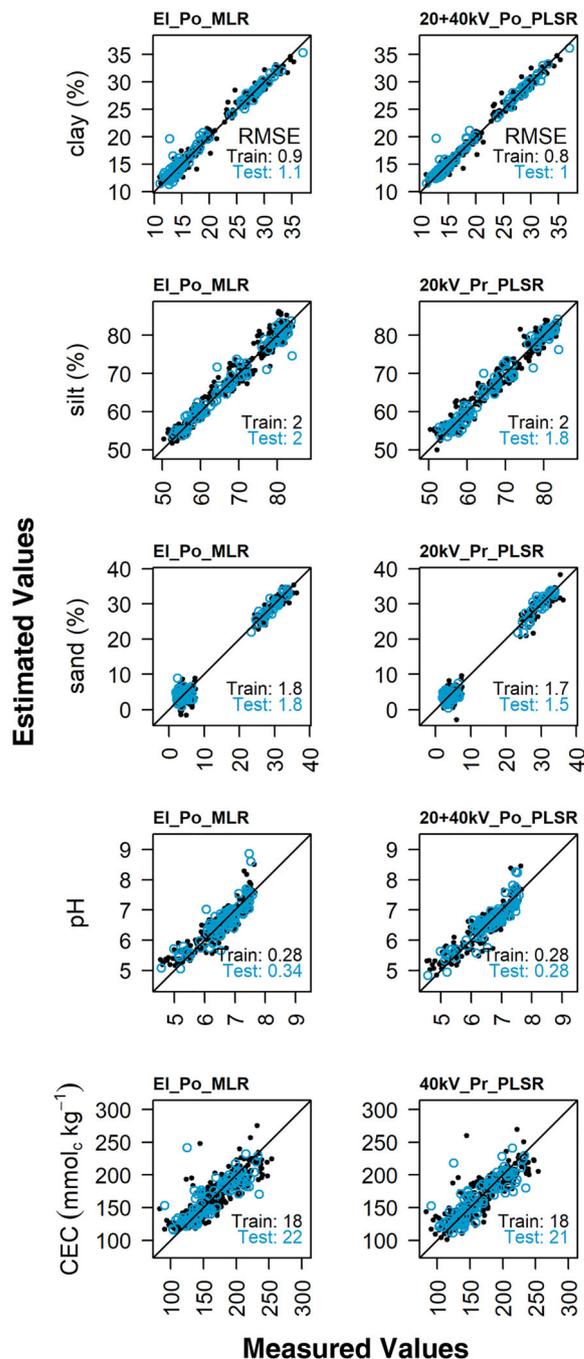


FIGURE 3 Measured versus estimated soil properties in training (black filled circles) and testing (blue open circles) after random partitioning. Data from the partition with median validation performance is shown, while an average root mean squared error (RMSE) for training (train) and testing (test) across all five dataset partitions is given. The plots on the left and right show performance for Approach I and the optimal approach, respectively, where elements (EI), 20 kV, 40 kV, or 20+40 kV (concatenated spectra) refers to the model predictor, powder (Po) or pressed (Pr) is the sample preparation, and multiple linear regression (MLR) or partial least squares regression (PLSR) is the algorithm. CEC, cation exchange capacity.

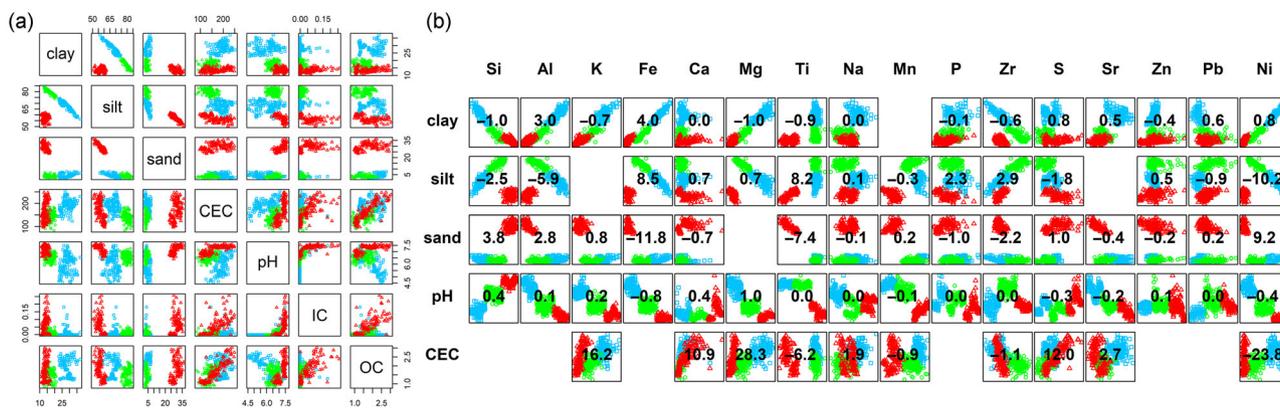


FIGURE 4 Interrelation of (a) clay, silt, and sand contents (%), cation exchange capacity (CEC, $\text{mmol}_c \text{kg}^{-1}$), pH, inorganic carbon content (IC, %), and organic carbon content (OC, %) at Friemar (blue squares), Lüttewitz (green circles), and Zschortau (red triangles) and (b) the subset of measured soil properties to be predicted by X-ray fluorescence (XRF) with elemental contents determined by XRF after measurement on powder samples. Numbers give the average slope of the element predictor in the stepwise multiple linear regressions across the five model partitions. Elemental contents (%) were z-transformed prior to modeling to allow comparison of slopes despite varying magnitudes of contents. No plot is shown for elements that were not utilized in any of the regressions for a given property.

of the element predictors in the MLR after stepwise simplification. For clay, 15 of the 16 element predictors remained after stepwise simplification in at least one of the five dataset partitions used for model training (Figure 4b). Since the matrix of element predictors was z-transformed prior to modeling, the importance of the element predictors can be compared using their slopes. Here, elements with an absolute value of the slope larger than the average of the absolute values of the slopes are mentioned. The most important element predictors for clay were, in descending order, Fe (slope = 4.0), Al (3.0), Si (-1.0), Mg (-1.0), and Ti (-0.9). To interpret the mechanisms behind these results, it must be considered that the overall direction of the relationship between elemental contents and soil properties (sign of the coefficients in the MLR) is determined by the elemental contents in one size fraction relative to the others. For example, although Si is a key component of aluminosilicates in the clay fraction, Si makes up a larger proportion of quartz minerals in the sand fraction (Sparks, 2003), and was therefore negatively related to clay content for this dataset. The positive relationships between Fe and Al with clay could be due to their presence as interlayer metal cations in montmorillonite and vermiculite (Sparks, 2003), which are high-activity clays commonly found in Luvisols (IUSS Working Group WRB, 2022). The positive relationships could also be explained by the presence of Fe-oxides in the clay fraction, which may also undergo isomorphic substitution to include Al (Schwertmann, 2008). Silva et al. (2020) likewise found for $n = 502$ surface soils collected throughout five Brazilian states that Fe, Al, and Si were the most important element predictors for clay content, with coefficients in their stepwise generalized linear models matching the signs of those in the present study. Finally, Zhu et al. (2011) and O'Rourke, Stockmann et al.

(2016) likewise found a useful positive relationship between Fe and clay contents.

For silt, 14 elements were used in at least one of the dataset partitions. Important element predictors included Ni (-10), Fe (8.5), Ti (8.2), Al (-5.9), and Zr (2.9). Relationships between silt contents and elemental contents were consistent at Lüttewitz and Friemar for several elements, whereas relationships differed substantially at Zschortau. This indicates some variability in the mineralogy or extent of weathering between the sites, which poses a problem for XRF prediction models (Zhu et al., 2011), and makes interpretation of prediction mechanisms more difficult. Interpretation is further hampered by the mineralogical diversity of the silt-sized fraction, which can contain a range of primary and secondary minerals, whereas the sand fraction is dominated by primary minerals and the clay fraction by secondary minerals (Sparks, 2003). A mineralogical analysis of particle size fraction by X-ray diffraction may give more insights into the prediction mechanisms. However, despite differing relationships between elemental and silt contents between the sites, silt estimation models were nonetheless successful, with an average RMSEv of 2.0%.

For sand, 15 elements were used in at least one of the dataset partitions. Important elements overlapped substantially with those for clay and silt prediction, and included Fe (-12), Ni (9.2), Ti (-7.4), Si (3.8), and Al (2.8). The negative coefficient of Fe could be due to the inverse relationship between sand content and Fe-containing minerals in the clay fraction, and the positive coefficient for Si could be due to the positive correlations of sand with quartz (Drewnik et al., 2014). Some results, such as the positive coefficient for Al despite a seemingly similar relationship as Fe to sand, are harder to explain and seem to arise from multicollinearity

between element predictors for sand. This issue is discussed further in Section 3.2, comparing use of MLR versus PLSR. With random partitioning into training and test sets, sand prediction was successful in validation despite substantially higher sand contents at Zschortau (29% on average) compared to Lüttewitz (3%) and Friemar (4%).

pH predictions were satisfactory for all model partitions with random partitioning into training and test sets, with an average and range of RMSE_v of 0.34 (0.28–0.42) and RPIQ_v of 2.17 (2.01–2.46) (Table 2, Figure 2). Measured versus estimated values show some tendency for overestimation of pH at low and high values (Figure 3). All 16 element predictors were used in at least one model partition (Figure 4b). Important element predictors included Mg (slope = 1.0), Fe (−0.8), Ni (−0.4), Ca (0.4), Si (0.4), and S (−0.3). Prediction mechanisms are difficult to interpret, first, given multicollinearity between the element predictors and, second, due to the influence of agricultural management as well as organic and inorganic soil fractions on pH (Curtin & Trolove, 2013). Based on the positive nonlinear relationship between pH and IC content in these loess soils (Figure 4a), it is feasible that Mg and Ca bound to carbonates in the soil, which buffer soil acidity, were useful for predicting pH (Stevens & Blanchar, 1992), whereas relationships between pH and OC as well as soil texture fractions were much less consistent across the sites. Tavares et al. (2020) and Silva et al. (2017) likewise found that Ca content was a useful predictor for pH.

Finally, CEC predictions ranged from satisfactory to very good for all model partitions with random partitioning into training and test sets, with an average and range of RMSE_v of 22 mmol_c kg^{−1} (18–28 mmol_c kg^{−1}) and RPIQ_v of 2.61 (2.20–3.26) (Table 2, Figure 2). The plot of measured versus estimated values shows unbiased predictions (Figure 3). Ten element predictors were used in at least one model partition (Figure 4b). Important element predictors included Mg (slope = 28), Ni (−24), K (16), S (12), and Ca (11). In general, CEC is affected by both the amount and pH-dependent exchange capacity of soil organic matter and clay colloids (Curtin & Rostad, 1997). For these loess soils, a relationship between CEC and clay was absent, but there was a positive correlation between CEC and OC content, particularly at Zschortau (Figure 4a). Several of the key elemental predictors (Mg, K, S, and Ca) for CEC are important components of both plant and microbial biomass (Khan et al., 2009; Shrivastav et al., 2020), from which soil organic matter is derived. Furthermore, Mg, K, and Ca are additionally present in soil as cations held against leaching by CEC. Silva et al. (2017) likewise found positive coefficients for Ca in regressions to predict CEC, while Tavares et al. (2020) found positive coefficients for both Ca and K.

3.2 | Effect of the algorithm on model performance

In addition to the use of stepwise MLR with elemental contents measured on powder soils, we additionally tested the use of PLSR with this data. As spectral intensities are evaluated with more sophisticated algorithms such as PLSR out of necessity due to the large number of predictors, comparison of PLSR and MLR for elemental data was used to isolate the effect of using different algorithms from the effect of using different predictors. In our case, the difference in performance between models calculated with MLR versus PLSR was very minor and did not show a consistent advantage of one algorithm over another. The difference in error ranged from 2.3% lower with MLR for silt prediction to 0.4% lower with PLSR for pH (Table 2, Figure 2). Thus, in terms of accuracy, neither algorithm was consistently superior.

However, calculation of VIF for the model predictors for the MLR equations showed severe multicollinearity between element predictors for all soil properties, with an average of 55 for clay, 59 for silt, 61 for sand, 64 for pH, and 20 for CEC considering all predictors retained after stepwise model simplification and across all dataset partitions. According to Myers (1990), a VIF of 10 or higher is a cause for concern. Multicollinearity among explanatory variables is very common in observational studies such as this one (Crawley, 2015). Multicollinearity results in higher standard error of slope predictions between samples, and thus lower trustworthiness of slope estimates (Field et al., 2013). Multicollinearity also makes it more difficult to interpret prediction mechanisms because predictors are somewhat interchangeable (Field et al., 2013). Therefore, despite frequent and successful application of MLR in studies using XRF elemental contents to predict soil properties with high accuracy, our results show that MLR prediction mechanisms must be interpreted with caution, and that PLSR may be appropriate due to their adequate handling of multicollinearity between predictors. Despite these considerations, both the PLSR and MLR models presented above were useful, as demonstrated by the successful prediction of all soil properties in model testing with a closed population of soils. The robustness of such models for unknown soils (leave-one-site-out training-testing strategy) is explored subsequently (Section 3.6).

3.3 | Effect of sample preparation on model performance

As sample homogeneity reduces the influence of matrix effects and surface roughness (Haschke et al., 2021), we additionally tested the effect of forming soil powders into press pellets (17% wax on a mass basis, 200 kN) in combination

with use of elemental contents, 20 kV spectral intensities, or 40 kV spectral intensities as model predictors. When elements were used as predictors, sample preparation as a press pellet did not make a large or consistent difference on model performance, with the scale of the effect of pressing ranging from a 3% decrease in RPIQ_v for pH to a positive 4% increase in RPIQ_v for sand (Table 2, Figure 2). Pressing was the most consistently advantageous for the 20 kV spectra, with the effect ranging from a marginal positive effect for pH (1% increase in RPIQ_v), to a small positive effect for clay and CEC (4% increase in RPIQ_v), to more substantial improvements for silt and sand (11% and 13% increases in RPIQ_v, respectively). Finally, for 40 kV spectra, the effect of pressing ranged from inconsistent for texture estimations (−2% to 1% change in RPIQ_v), to a small positive effect for pH and CEC (5% and 8% increase in RPIQ_v, respectively). The explanation for the inconsistent to minor improvements from measurement of press pellets could be that sufficient steps were already taken to ensure the homogeneity of powder samples. These steps included drying and grinding (Goff et al., 2020; Li et al., 2022), rotation of the sample during measurement to compensate for minor sample inhomogeneities and surface roughness, and measurement in He atmosphere with reduced air pressure to limit fluorescence attenuation, which could vary for uncompacted powder samples (Haschke et al., 2021).

3.4 | Effect of the XRF predictor type on model performance

Considering only powder soils analyzed by PLSR, a comparison was made between models using elemental contents, 20 kV spectral intensities, 40 kV spectra intensities, or concatenated 20 + 40 kV spectral intensities as model predictors. For clay, the performance rankings in descending order were 20 + 40 kV (RPIQ_v = 12.6), 20 kV (11.8), 40 kV (11.7), and elements (11.2), with a 13% increase in RPIQ_v from the worst to best model (Table 2, Figure 2). For silt, the performance rankings in descending order were 20 + 40 kV (RPIQ_v = 10.3), 40 kV (9.8), 20 kV (9.5), and elements (9.4), with a 10% increase in RPIQ_v from the worst to best model. For sand, the performance rankings in descending order were 40 kV (RPIQ_v = 14.3), 20 + 40 kV (14.0), 20 kV (13.4), and elements (13.3), with an 8% increase in RPIQ_v from the worst to best model. For pH, the performance rankings in descending order were 20 + 40 kV (RPIQ_v = 2.32), 20 kV (2.22), elements (2.18), and 40 kV (2.14) with an 8% increase in RPIQ_v from the worst to best model. Finally, for CEC, the performance rankings in descending order were elements (RPIQ_v = 2.59), 40 kV (2.56), 20 kV (2.53), and 20 + 40 kV

(2.52), with a 3% increase in RPIQ_v from the worst to best model. Thus, the optimal predictors varied by soil property and comparative benefits ranged from marginal (3% for CEC) to more substantial (13% for clay).

Use of elemental contents determined by spectral deconvolution is considered an optimized data evaluation technique for XRF due to the complexity of physicochemical relationships, which create a risk of poor model outcomes due to spectral artifacts arising from the X-ray tube source, but also artifact peaks arising from the detection process (Arkadiev et al., 2006; Haschke et al., 2021). Thus, extraction of element predictors from the spectra is a mechanistic approach based on knowledge of physicochemical relationships, whereas use of spectral predictors is an empirical approach using automated selection of spectral pretreatments and multivariate statistical modeling to extract useful information from noise. Despite these differences, neither approach was consistently more accurate. Though the spectral approach lacks knowledge-based extraction of information from noise, the scattering present in the XRF spectra could actually contain useful information about the presence of light elements (Morona et al., 2017) and matrix properties (Haschke et al., 2021) that assist in predicting complex soil properties.

3.5 | Summary of performance improvement through alternative approaches

The measured versus validation estimated values in model training and testing following random dataset partitioning for Approach I versus the single best approach for each soil property are shown in Figure 3. Of the eight tested approaches for the five soil properties of interest, Approach I (MLR modeling of powder samples with elemental contents as predictors) was never optimal. For clay and pH, the optimal approach was use of the 20 + 40 kV spectral intensities for powder samples, increasing RPIQ_v by 12% (from 11.3 to 12.6) and 7% (from 2.17 to 2.32), respectively. For silt and sand, the optimal approach was use of 20 kV spectral intensities for pressed samples, increasing RPIQ_v by 10% (from 9.6 to 10.5) and 15% (from 13.2 to 15.2), respectively. Finally for CEC, the optimal approach was use of the 40 kV spectral intensities for pressed samples, increasing RPIQ_v by 6% (from 2.61 to 2.76). Thus, the optimal approach depended on the soil property of interest. While benefits of alternative approaches were more moderate for pH and CEC (7% and 6% increase in RPIQ_v, respectively) than those for soil texture (10%–15% increase), optimizations may be superfluous for texture estimations given the excellent performance across all tested approaches. Further testing is required to determine the generalizability of these results.

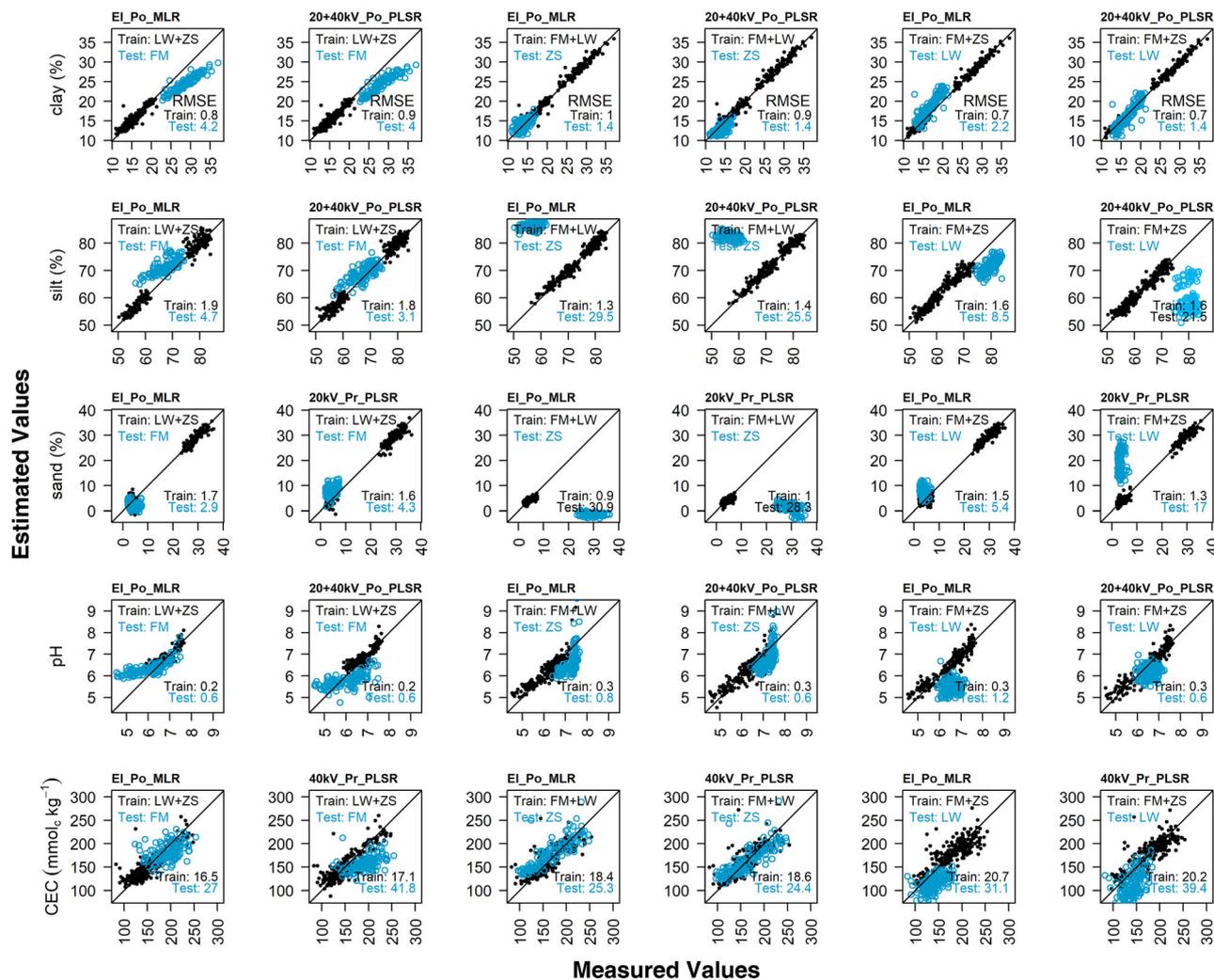


FIGURE 5 Measured versus estimated soil properties in training with data from two sites (black filled circles) and testing with data from the third site (blue open circles) for the hold-one-site-out training-testing procedure. Data and root mean squared error (RMSE) are given for all three dataset partitions. For each partition, models using Approach I as well as the optimal approach previously determined with random dataset partitioning are shown. Elements (EI), 20 kV, 40 kV or 20 + 40 kV (concatenated spectra) refers to the model predictor, powder (Po), or pressed (Pr) refers to the sample preparation, and multiple linear regression (MLR) or partial least squares regression (PLSR) refers to the algorithm applied. CEC, cation exchange capacity; FM, Friemar; LW, Lüttewitz; ZS, Zschortau.

3.6 | Effect of dataset partitioning on model performance

Finally, to test the robustness of Approach I as well as the single most accurate approach for application at new sites, a leave-one-site-out approach to training and testing was conducted, whereby two sites were used for model training and the third site was used for testing. Although the RMSE in model training versus testing was very similar following random dataset partitioning (Figure 3), RMSE increased in model testing in all cases for models trained with a hold-one-site-out approach, and often severely (Figure 5). While the RMSE of the test sets for clay may be within an acceptable range for certain applications (1.4%–4.2%), silt and sand predictions for Zschortau and Lüttewitz test sets were unreli-

able. This is to be expected based on the relationships between element predictors and soil properties in Figure 4b, which showed relatively consistent relationships between elements and clay across the three sites, whereas relationships between elements and silt were different at Zschortau compared to the other two sites, and sand contents were of a much higher magnitude at Zschortau compared to the other two sites. For pH, element predictors had partially consistent relationships across the three sites, resulting in the magnitude of RMSE roughly doubling to quadrupling from model training to testing. Finally, the range of CEC overlapped across the three sites to a greater extent than other properties, and RMSE at most roughly doubled from model training to testing with the hold-one-site-out approach. Thus, for datasets where the relationships between the property of interest and the elemental

contents are not consistent and/or the magnitude of properties do not overlap across sites, inclusion of the samples from the target site in model training is strongly advised. This is consistent with findings in the field of visible/near- and mid-infrared spectroscopy, where the benefits of creating local soil calibration models or spiking national spectral libraries with local soils are well established (Briedis et al., 2020; Seidel et al., 2019). Also of note is that the best approach following random dataset partitioning was not always superior to Approach I with the hold-one-site-out training-testing procedure. Thus, the optimal sample preparation, predictor type, and algorithm were not only property specific, but also dependent on the training-testing strategy.

4 | CONCLUSION

This study demonstrated the usefulness of XRF spectrometry for prediction of soil texture, CEC, and pH at three loess sites. When model training included soils from the target site, validation estimates were reliable, and texture estimations in particular were highly accurate. The generally minor differences in performance regardless of sample preparation (powder vs. pellet), predictor type (elements vs. spectral intensities), and algorithm (MLR vs. PLSR) demonstrate the robustness of this approach and indicate that certain steps to increase the efficiency of the method can be taken without loss of accuracy.

These positive findings encourage exploration of further efficiency gains to the modeling approaches demonstrated here through the use of standard reference materials rather than site-specific corrections to XRF elemental contents, as well as through measurement of intact, field-moist soils, as in the recent study by Li et al. (2022). For this, empirical corrections for soil moisture have proven helpful (Stockmann, Jang et al., 2016), but effects of poor detection of light elements (Na, Mg, Al, Si, P, and S)—which were highly important predictors in the current study—must be considered. The combination of laboratory XRF with visible/near- and/or mid-infrared handheld spectrometers (Greenberg et al., 2023; Javadi & Mouazen, 2021) is a promising means to improve model accuracy by providing complimentary information about spectrally active organic molecules and soil minerals. Therefore, combining handheld infrared and portable XRF spectrometers could be a solution to achieving sufficient accuracy with field measurement.

AUTHOR CONTRIBUTIONS

Isabel Greenberg: Conceptualization; data curation; formal analysis; investigation; methodology; validation; visualization; writing—original draft. **Anja Sawallisch:** Data curation; formal analysis; investigation; methodology; validation. **Jan Stelling:** Methodology; validation; writing—review and

editing. **Michael Vohland:** Conceptualization; funding acquisition; project administration; supervision; writing—review and editing. **Bernard Ludwig:** Conceptualization; formal analysis; funding acquisition; investigation; methodology; project administration; resources; supervision; validation; writing—review and editing.

ACKNOWLEDGMENTS

The authors would like to thank Theresa Mohr and the Staatliche Betriebsgesellschaft für Umwelt und Landwirtschaft, Sachsen for consultation and assistance in the creation and measurement of the press pellets. The authors would like to thank the Department of Sedimentology and Environmental Geology, University of Göttingen, for carrying out the acid digestion and ICP-OES measurements. The authors would like to thank the Geochemistry and Isotope Geology Department at the University of Göttingen for carrying out the WD-XRF measurements. Finally, the authors would like to thank Anja Sawallisch and her team at the Department of Environmental Chemistry, University of Kassel, for technical assistance with all other measurements. This research was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft), Grant Numbers LU 583/24-1 and VO 1509/10-1.

Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

Jan Stelling is an employee of Bruker AXS GmbH. The other authors declare no conflicts of interest.

ORCID

Isabel Greenberg  <https://orcid.org/0000-0002-4762-8474>

Michael Vohland  <https://orcid.org/0000-0002-6048-1163>

Bernard Ludwig  <https://orcid.org/0000-0001-8900-6190>

REFERENCES

- Arkadiev, V., Knüpfner, W., & Langhoff, N. (2006). X-ray tubes. In B. Beckhoff, B. Kanngießer, N. Langhoff, R. Wedell, & H. Wolff (Eds.), *Handbook of practical X-ray fluorescence analysis* (pp. 36–52). Springer.
- Beckhoff, B., Kanngießer, B., Langhoff, N., Wedell, R., & Wolff, H. (Eds.). (2006). *Handbook of practical X-ray fluorescence analysis*. Springer.
- Briedis, C., Baldock, J., de Moraes Sá, J. C., dos Santos, J. B., & Milori, D. M. B. P. (2020). Strategies to improve the prediction of bulk soil and fraction organic carbon in Brazilian samples by using an Australian national mid-infrared spectral library. *Geoderma*, 373, 114401. <https://doi.org/10.1016/j.geoderma.2020.114401>
- Brouwer, P. (2010). *Theory of XRF* (3rd ed.). PANalytical B.V.
- Byers, H. L., McHenry, L. J., & Grundl, T. J. (2016). Forty-nine major and trace element concentrations measured in soil reference materials NIST SRM 2586, 2587, 2709a, 2710a and 2711a using ICP-MS and wavelength dispersive-XRF. *Geostandards and*

- Geoanalytical Research*, 40(3), 433–445. <https://doi.org/10.1111/j.1751-908X.2016.00376.x>
- Chang, C.-W., Laird, D. A., Mausbach, M. J., & Hurburgh, C. R., Jr. (2001). Near-infrared reflectance spectroscopy—Principal components regression analyses of soil properties. *Soil Science Society of America Journal*, 65(2), 480–490. <https://doi.org/10.2136/sssaj2001.652480x>
- Crawley, M. J. (2015). *Statistics: An introduction using R* (2nd ed.). Wiley.
- Curtin, D., & Rostad, H. P. (1997). Cation exchange and buffer potential of Saskatchewan soils estimated from texture, organic matter and pH. *Canadian Journal of Soil Science*, 77(4), 621–626.
- Curtin, D., & Trolove, S. (2013). Predicting pH buffering capacity of New Zealand soils from organic matter content and mineral characteristics. *Soil Research*, 51(6), 494–502. <https://doi.org/10.1071/SR13137>
- DIN EN 15309. (2007). *Charakterisierung von Abfällen und Böden—Bestimmung der elementaren Zusammensetzung durch Röntgenfluoreszenz-Analyse; Deutsche Fassung EN_15309:2007*. Beuth Verlag GmbH.
- DIN ISO 10390. (2005). *Soil quality—Determination of pH (ISO 10390:2005)*. Beuth Verlag GmbH.
- DIN ISO 11277. (2002). *Soil quality—Determination of particle size distribution in mineral soil material—Method by sieving and sedimentation (ISO 11277:1998 + ISO 11277:1998 Corrigendum 1:2002)*. Beuth Verlag GmbH.
- Drewnik, M., Skiba, M., Szymański, W., & Żyła, M. (2014). Mineral composition vs. soil forming processes in loess soils—A case study from Kraków (Southern Poland). *Catena*, 119, 166–173. <https://doi.org/10.1016/j.catena.2014.02.012>
- Field, A., Miles, J., & Field, Z. (2013). *Discovering statistics using R*. Sage Publications Ltd.
- Fox, J., & Weisberg, S. (2023). *An R companion to applied regression*. <https://CRAN.R-project.org/package=car>
- Goff, K., Schaeztl, R. J., Chakraborty, S., Weindorf, D. C., Kasmerchak, C., & Bettis, E. A. (2020). Impact of sample preparation methods for characterizing the geochemistry of soils and sediments by portable X-ray fluorescence. *Soil Science Society of America Journal*, 84(1), 131–143. <https://doi.org/10.1002/saj2.20004>
- Govindaraju, K. (1994). 1994 Compilation of working values and sample description for 383 geostandards. *Geostandards and Geoanalytical Research*, 18, 1–158. <https://doi.org/10.1046/j.1365-2494.1998.53202081.x-i1>
- Greenberg, I., Seidel, M., Vohland, M., Koch, H.-J., & Ludwig, B. (2022). Performance of in situ vs laboratory mid-infrared soil spectroscopy using local and regional calibration strategies. *Geoderma*, 409, 115614. <https://doi.org/10.1016/j.geoderma.2021.115614>
- Greenberg, I., Vohland, M., Seidel, M., Hutengs, C., Bezar, R., & Ludwig, B. (2023). Evaluation of mid-infrared and X-ray fluorescence data fusion approaches for prediction of soil properties at the field scale. *Sensors*, 23(2), 662. <https://doi.org/10.3390/s23020662>
- Haschke, M., Flock, J., & Haller, M. (2021). *X-ray fluorescence spectroscopy for laboratory applications*. Wiley-VHC GmbH.
- IUSS Working Group WRB. (2022). *World reference base for soil resources: International soil classification system for naming soils and creating legends for soil maps* (4th ed.). International Union of Soil Sciences (IUSS). https://wrb.isric.org/files/WRB_fourth_edition_2022-12-18.pdf
- Jaremko, D., & Kalembsa, D. (2014). A comparison of methods for the determination of cation exchange capacity of soils/porównanie metod oznaczania pojemności wymiany kationów i sumy kationów wymiennych w glebach. *Ecological Chemistry and Engineering S*, 21(3), 487–498. <https://doi.org/10.2478/eces-2014-0036>
- Javadi, S. H., & Mouazen, A. M. (2021). Data fusion of XRF and vis-NIR using outer product analysis, Granger–Ramanathan, and least squares for prediction of key soil attributes. *Remote Sensing*, 13(11), 2023. <https://doi.org/10.3390/rs13112023>
- Jelinski, N. A., Richardson, J. B., & Nater, E. A. (2023). Soils of humid cool temperate regions. In M. J. Goss, & M. Oliver (Eds.), *Encyclopedia of soils in the environment* (2nd ed., pp. 289–298). Academic Press.
- Kandpal, L. M., Munaf, M. A., Cruz, C., & Mouazen, A. M. (2022). Spectra fusion of mid-infrared (MIR) and X-ray fluorescence (XRF) spectroscopy for estimation of selected soil fertility attributes. *Sensors*, 22(9), 3459. <https://doi.org/10.3390/s22093459>
- Khan, K. S., Heinze, S., & Joergensen, R. G. (2009). Simultaneous measurement of S, macronutrients, and heavy metals in the soil microbial biomass with CHCl₃ fumigation and NH₄NO₃ extraction. *Soil Biology and Biochemistry*, 41(2), 309–314. <https://doi.org/10.1016/j.soilbio.2008.11.001>
- Koch, H.-J., Dieckmann, J., Büchse, A., & Märländer, B. (2009). Yield decrease in sugar beet caused by reduced tillage and direct drilling. *European Journal of Agronomy*, 30(2), 101–109. <https://doi.org/10.1016/j.eja.2008.08.001>
- Koenig, N., & Fortmann, H. (1996). *Probenvorbereitungs-, Untersuchungs- und Elementbestimmungs-Methoden des Umweltanalytik-Labors der Niedersaechsischen Forstlichen Versuchsanstalt und des Zentrallabor 2 des Forschungszentrums Waldoekosysteme*. Universitaet Goettingen.
- Li, S., Shen, J., Bishop, T. F. A., & Viscarra Rossel, R. A. (2022). Assessment of the effect of soil sample preparation, water content and excitation time on proximal X-ray fluorescence sensing. *Sensors*, 22(12), 4572. <https://doi.org/10.3390/s22124572>
- Mevik, B.-H., Wehrens, R., & Liland, K. H. (2019). *pls: Partial least squares and principal component regression*. <https://CRAN.R-project.org/package=pls>
- Morona, F., Dos Santos, F. R., Brinatti, A. M., & Melquiades, F. L. (2017). Quick analysis of organic matter in soil by energy-dispersive X-ray fluorescence and multivariate analysis. *Applied Radiation and Isotopes: Including Data, Instrumentation and Methods for Use in Agriculture, Industry and Medicine*, 130, 13–20. <https://doi.org/10.1016/j.apradiso.2017.09.008>
- Myers, R. H. (1990). *Classical and modern regression with applications* (2nd ed.). *The Duxbury advanced series in statistics and decision sciences*. PWS-KENT.
- Nawar, S., Cipullo, S., Douglas, R. K., Coulon, F., & Mouazen, A. M. (2020). The applicability of spectroscopy methods for estimating potentially toxic elements in soils: State-of-the-art and future trends. *Applied Spectroscopy Reviews*, 55(7), 525–557. <https://doi.org/10.1080/05704928.2019.1608110>
- O'Rourke, S. M., Minasny, B., Holden, N. M., & McBratney, A. B. (2016). Synergistic use of Vis-NIR, MIR, and XRF spectroscopy for the determination of soil geochemistry. *Soil Science Society of America Journal*, 80(4), 888–899. <https://doi.org/10.2136/sssaj2015.10.0361>
- O'Rourke, S. M., Stockmann, U., Holden, N. M., McBratney, A. B., & Minasny, B. (2016). An assessment of model averaging to improve

- predictive power of portable vis-NIR and XRF for the determination of agronomic soil properties. *Geoderma*, 279, 31–44. <https://doi.org/10.1016/j.geoderma.2016.05.005>
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Schwertmann, U. (2008). Iron oxides. In W. Chesworth (Ed.), *Encyclopedia of soil science* (pp. 363–369). Springer.
- Seidel, M., Hutengs, C., Ludwig, B., Thiele-Bruhn, S., & Vohland, M. (2019). Strategies for the efficient estimation of soil organic carbon at the field scale with vis-NIR spectroscopy: Spectral libraries and spiking vs. local calibrations. *Geoderma*, 354, 113856. <https://doi.org/10.1016/j.geoderma.2019.07.014>
- Sharma, A., Weindorf, D. C., Wang, D., & Chakraborty, S. (2015). Characterizing soils via portable X-ray fluorescence spectrometer: 4. Cation exchange capacity (CEC). *Geoderma*, 239–240, 130–134. <https://doi.org/10.1016/j.geoderma.2014.10.001>
- Shrivastav, P., Prasad, M., Singh, T. B., Yadav, A., Goyal, D., Ali, A., & Dantu, P. K. (2020). Role of nutrients in plant growth and development. In M. Naeem, A. A. Ansari, & S. S. Gill (Eds.), *Contaminants in agriculture: Sources, impacts and management* (pp. 43–60). Springer International Publishing.
- Silva, S. H. G., dos Santos Teixeira, A. F., de Menezes, M. D., Guilherme, L. R. G., de Souza Moreira, F. M., & Curi, N. (2017). Multiple linear regression and random forest to predict and map soil properties using data from portable X-ray fluorescence spectrometer (pXRF). *Ciência e Agrotecnologia*, 41(6), 648–664. <https://doi.org/10.1590/1413-70542017416010317>
- Silva, S. H. G., Weindorf, D. C., Pinto, L. C., Faria, W. M., Acerbi, F. W. Jr., Gomide, L. R., de Mello, J. M., de Pádua, A. L. Jr., de Souza, I. A., dos Santos Teixeira, A. F., Guilherme, L. R. G., & Curi, N. (2020). Soil texture prediction in tropical soils: A portable X-ray fluorescence spectrometry approach. *Geoderma*, 362, 114136. <https://doi.org/10.1016/j.geoderma.2019.114136>
- Soil Survey Staff. (2014). *Soil survey field and laboratory methods manual. Soil survey investigations* (Version 2.0, Report No. 51), Version 2.0. U.S. Department of Agriculture, Natural Resources Conservation Service.
- Soil Survey Staff. (2022). *Keys to Soil Taxonomy* (13th ed.). USDA Natural Resources Conservation Service.
- Soriano-Disla, J. M., Janik, L. J., Viscarra Rossel, R. A., Macdonald, L. M., & McLaughlin, M. J. (2014). The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. *Applied Spectroscopy Reviews*, 49(2), 139–186. <https://doi.org/10.1080/05704928.2013.811081>
- Sparks, D. L. (2003). *Environmental soil chemistry* (2nd ed.). Elsevier.
- Stevens, A., Ramirez-Lopez, L., & Hans, G. (2022). *prospectr: Miscellaneous functions for processing and sample selection of spectroscopic data*. <https://cran.r-project.org/web/packages/prospectr/index.html>
- Stevens, J. J., & Blanchar, R. W. (1992). Soil pH gradients near calcite and dolomite particles. *Soil Science Society of America Journal*, 56(3), 967–972. <https://doi.org/10.2136/sssaj1992.03615995005600030047x>
- Stockmann, U., Cattle, S. R., Minasny, B., & McBratney, A. B. (2016). Utilizing portable X-ray fluorescence spectrometry for in-field investigation of pedogenesis. *Catena*, 139, 220–231. <https://doi.org/10.1016/j.catena.2016.01.007>
- Stockmann, U., Jang, H. J., Minasny, B., & McBratney, A. B. (2016). The effect of soil moisture and texture on Fe concentration using portable X-ray fluorescence spectrometers. In A. Hartemink & B. Minasny (Eds.), *Digital soil morphometrics* (pp. 63–71). Progress in soil science. Springer. https://doi.org/10.1007/978-3-319-28295-4_5
- Tavares, T. R., Molin, J. P., Javadi, S. H., de Carvalho, H. W. P., & Mouazen, A. M. (2020). Combined use of vis-NIR and XRF sensors for tropical soil fertility analysis: Assessing different data fusion approaches. *Sensors*, 21(1), 148. <https://doi.org/10.3390/s21010148>
- Towett, E. K., Shepherd, K. D., Sila, A., Aynekulu, E., & Cadisch, G. (2015). Mid-infrared and total X-ray fluorescence spectroscopy complementarity for assessment of soil properties. *Soil Science Society of America Journal*, 79(5), 1375–1385. <https://doi.org/10.2136/sssaj2014.11.0458>
- Viscarra, R., & Behrens, T. (2010). Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma*, 158(1–2), 46–54. <https://doi.org/10.1016/j.geoderma.2009.12.025>
- Wang, S., Li, W., Li, J., & Liu, X. (2013). Prediction of soil texture using FT-NIR spectroscopy and PXRF spectrometry with data fusion. *Soil Science*, 178(11), 626–638. <https://doi.org/10.1097/SS.000000000000026>
- Wehrens, R. (2020). *Chemometrics with R: Multivariate data analysis in the natural and life sciences* (2nd ed.). Springer-Verlag Berlin Heidelberg.
- Xu, D., Zhao, R., Li, S., Chen, S., Jiang, Q., Zhou, L., & Shi, Z. (2019). Multi-sensor fusion for the determination of several soil properties in the Yangtze River Delta, China. *European Journal of Soil Science*, 70(1), 162–173. <https://doi.org/10.1111/ejss.12729>
- Zhao, H., Guo, B., Wei, Y., & Zhang, B. (2013). Multi-element composition of wheat grain and provenance soil and their potentialities as fingerprints of geographical origin. *Journal of Cereal Science*, 57(3), 391–397. <https://doi.org/10.1016/j.jcs.2013.01.008>
- Zhu, Y., Weindorf, D. C., & Zhang, W. (2011). Characterizing soils using a portable X-ray fluorescence spectrometer: 1. Soil texture. *Geoderma*, 167–168, 167–177. <https://doi.org/10.1016/j.geoderma.2011.08.010>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Greenberg, I., Sawallisch, A., Stelling, J., Vohland, M., & Ludwig, B. (2024). Optimization of sample preparation and data evaluation techniques for X-ray fluorescence prediction of soil texture, pH, and cation exchange capacity of loess soils. *Soil Science Society of America Journal*, 88, 27–42. <https://doi.org/10.1002/saj2.20598>