

Using pXRF and vis-NIR spectra for predicting properties of soils developed in loess

Gafur GOZUKARA^{1,2,*}, Yakun ZHANG¹ and Alfred E. HARTEMINK¹

¹University of Wisconsin-Madison, Department of Soil Science, FD Hole Soils Lab, 1525 Observatory Drive, Madison, WI 53706 (USA)

²Eskisehir Osmangazi University, Department of Soil Science and Plant Nutrition, Eskisehir 26160 (Turkey)

(Received November 16, 2020; revised January 15, 2021; accepted March 4, 2021)

ABSTRACT

Visible near-infrared (vis-NIR) and portable X-ray fluorescence (pXRF) spectrometers have been increasingly utilized for predicting soil properties worldwide. However, only a few studies have focused on splitting the predictive models by horizons to evaluate prediction performance and systematically compare prediction performance for A, B, and combined A+B horizons. Therefore, we investigated the performance of pXRF and vis-NIR spectra, as individual or combined, for predicting the clay, silt, sand, total carbon (TC), and pH of soils developed in loess, and compared their prediction performance for A, B, and A+B horizons. Soil samples (176 in A horizon and 172 in B horizon) were taken from Mollisols and Alfisols in 136 pedons in Wisconsin, USA and analyzed for clay, silt, sand, pH, and TC. The pXRF and vis-NIR spectrometers were used to measure the pXRF and vis-NIR soil spectra. Data were separated into calibration ($n = 244$, 70%) and validation ($n = 104$, 30%) datasets. The Savitzky-Golay filter was applied to preprocess the pXRF and vis-NIR spectra, and the first 10 principal components (PCs) were selected through principal component analysis (PCA). Five types of predictor, *i.e.*, PCs from vis-NIR spectra, pXRF of beams at 0–40 and 0–10 keV (XRF40 and XRF10, respectively) spectra, combined XRF40 and XRF10 (XRF40+XRF10) spectra, and combined XRF40, XRF10, and vis-NIR (XRF40+XRF10+vis-NIR) spectra, were compared to predict soil properties using a machine learning algorithm (Cubist model). A multiple linear regression (MLR) model was applied to predict clay, silt, sand, pH, and TC using pXRF elements. The results suggested that pXRF spectra had better prediction performance for clay, silt, and sand, whereas vis-NIR spectra produced better TC and pH predictions. The best prediction performances for sand ($R^2 = 0.97$), silt ($R^2 = 0.95$), and clay ($R^2 = 0.84$) were achieved using vis-NIR+XRF40+XRF10 spectra in B horizon, whereas the best prediction performance for TC ($R^2 = 0.93$) and pH ($R^2 = 0.79$) were achieved using vis-NIR+XRF40+XRF10 spectra in A+B horizon. For all soil properties, the best MLR model had a lower prediction accuracy than the Cubist model. It was concluded that pXRF and vis-NIR spectra can be successfully applied for predicting clay, silt, sand, pH, and TC with high accuracy for soils developed in loess, and that spectral models should be developed for different horizons to achieve high prediction accuracy.

Key Words: Cubist model, machine learning algorithm, portable X-ray fluorescence spectra, soil elements, visible near-infrared spectra

Citation: Gozukara G, Zhang Y, Hartemink E A. 2022. Using pXRF and vis-NIR spectra for predicting properties of soils developed in loess. *Pedosphere*. 32(4): 601–614.

INTRODUCTION

Proximal soil sensing techniques have been widely used to provide information on soils. Visible near-infrared (vis-NIR) and portable X-ray fluorescence (pXRF) spectra have been used as rapid, environment-friendly, and low-cost techniques to measure or predict soil properties with high accuracy, and in a non-destructive way compared to traditional chemical methods (Taylor *et al.*, 2004; Waiser *et al.*, 2007; Rossel *et al.*, 2010, 2011; Rouillon and Taylor, 2016). Vis-NIR and pXRF spectrometers have been increasingly utilized for the prediction or estimation of soil types and properties and for environmental monitoring (Rossel *et al.*, 2009; Weindorf *et al.*, 2012; Hartemink and Minasny, 2014; Sharma *et al.*, 2014; Stockmann *et al.*, 2016; Duda *et al.*, 2017; Tümsava *et al.*, 2019; Benedet *et al.*, 2020).

To date, vis-NIR spectra of soils have been increasingly used to predict soil texture (clay, silt, and sand) (Waiser *et al.*,

2007; Rossel *et al.*, 2009; Tümsava *et al.*, 2019; Zhang and Hartemink, 2020), soil organic carbon (SOC) (Gomez *et al.*, 2008), total carbon (TC) (Wang *et al.*, 2015; Cardelli *et al.*, 2017; Zhang and Hartemink, 2020), carbonate, soil moisture (Yost and Hartemink, 2019), cation exchange capability (CEC), pH (Ben-Dor and Banin, 1995; Stockmann *et al.*, 2016), soil aggregate stability, and aggregate size distribution (Shi *et al.*, 2020) by means of the relationships between reflectance spectra in the range of 350 to 2 500 nm and soil properties (Duda *et al.*, 2017; Rawal *et al.*, 2019; Zhang and Hartemink, 2019). The pXRF spectrometer identifies and measures multiple elements using different energy spectra of X-ray (Weindorf *et al.*, 2014). Additionally, pXRF has been used to predict TC (Zhang and Hartemink, 2020), SOC (Gomez *et al.*, 2008), CEC (Sharma *et al.*, 2015), and soil texture (Zhu *et al.*, 2011; Silva, 2020; Zhang and Hartemink *et al.*, 2020). Recently, vis-NIR and pXRF spectral data

*Corresponding author. E-mail: ggozukara@ogu.edu.tr.

from different sensors have been increasingly combined with machine learning algorithms to predict soil properties, such as TC and total nitrogen (TN) (Wang *et al.*, 2015; Zhang and Hartemink, 2019), soil texture (Zhang and Hartemink, 2019), pH (O'Rourke *et al.*, 2016b; Zhang and Hartemink, 2019), CEC (O'Rourke *et al.*, 2016b; Wan *et al.*, 2020), elements (O'Rourke *et al.*, 2016a), and heavy metal pollution (Hu *et al.*, 2017; Xu *et al.*, 2019). The vis-NIR spectra provide more information on soil organic materials, whereas the pXRF spectrometer provides more information on soil mineral components (O'Rourke *et al.*, 2016a, b; Zhang and Hartemink, 2019). According to the results of these researchers, using combined vis-NIR and pXRF spectral data leads to increased soil property prediction accuracy.

Prediction performance for soil properties is influenced by land use and soil management practices (Mancini *et al.*, 2019; Andrade *et al.*, 2020a), different ranges of the dataset (Santana *et al.*, 2018; Zhang and Hartemink, 2019), soil chemical properties, soil types (Benedet *et al.*, 2020), soil depths (Mancini *et al.*, 2019; Silva *et al.*, 2020), soil horizons (Mancini *et al.*, 2019; Andrade *et al.*, 2020b; Benedet *et al.*, 2020) or soil diagnostic superficial horizons (Dos Santos *et al.*, 2014), and mineralogical properties (Andrade *et al.*, 2020a). To date, very few studies have investigated the splitting of predictive models by horizon (combined and separately) to achieve high prediction performance for predicting soil properties using individual pXRF (Mancini *et al.*, 2019; Andrade *et al.*, 2020b; Silva *et al.*, 2020b) and vis-NIR or combined pXRF and vis-NIR data (Benedet *et al.*, 2020) with machine learning algorithms. Additionally, only a few studies have investigated splitting the predictive models by horizon to achieve high prediction performance and systematically compare the prediction performance of A, B, and A+B horizons (Mancini *et al.*, 2019; Andrade *et al.*, 2020b; Benedet *et al.*, 2020).

Therefore, the objectives of this study were: 1) to investigate the performance of pXRF and vis-NIR spectra, individually or combined, in predicting the clay, silt, sand, TC, and pH of soils developed in loess, and 2) to compare their prediction performance for A, B, and combined A+B horizons. We hypothesize that combining pXRF and vis-NIR spectra will achieve better prediction performance than either alone, and that splitting the predictive models by horizon (A or B) will provide better prediction performance for some soil properties than when horizons combined (A+B).

MATERIALS AND METHODS

Soil sampling and analysis

The study was conducted in the Driftless Area, Dane County, Wisconsin, USA. The study area was approximately

7 000 ha and was mainly agricultural land, with maize, soybean, and alfalfa as the most prevalent field crops. The climate in this region is characterized by warm, moist summer and cold, dry winter. The mean annual precipitation and mean annual temperature are 857 mm and 7.3 °C, respectively. Loess materials cover much of the landscape in the study area. The soils were described and sampled in 136 pedons along 15 catenas in the study area. The catenas were selected depending on the aspect and slope of the study area. The detailed distribution of the selected catenas and pedons can be found in Evans and Hartemink (2014a, b). In this study, 22 pedons were classified as Alfisols and 114 pedons as Molisols (Soil Survey Staff, 2014). A 7-cm Edelman auger was used to sample the soil profiles. A total of 348 soil samples were taken from the 136 pedons, including 176 samples from A horizon and 172 samples from B horizon. The A horizon consisted of Ap, Ap1, Ap2, A1, A2, and A3 horizons, while B horizon consisted of B1, Bt1, Bt2, Bt3, Bt4, Bt5, Btg, and Bw horizons. The thickness of A horizon varied from 6 to 70 cm, while the thickness of B horizon varied from 5 to 90 cm. Soil samples were collected from the center of each horizon. The soils in the A and B horizons were developed from the loess material.

All the soil samples collected were air-dried, sieved using a 2-mm sieve, and analyzed for soil physical and chemical properties. Soil texture was analyzed using the hydrometer method (Gee and Bauder, 1979). Soil pH was determined using a pH meter (510 Series, OAKTON Instruments, Vernon Hills, USA) with a glass electrode in 1:1 (weight:volume) soil water solutions (Soil Survey Staff, 2014). TC was measured using the dry combustion method (Flash EA 1112 series NC, Thermo Electron Corporation, USA) (Soil Survey Staff, 2014).

Vis-NIR and pXRF analyses

Approximately 50 g ground (< 2 mm) soil was placed in a plastic weighing boat and slightly pressed for a uniform and flat surface before vis-NIR and pXRF measurements. The spectra of the 348 soil samples were measured using a vis-NIR spectrometer (Spectral Evolution, Lawrence, USA) with a spectral range of 350 to 2 500 nm. The vis-NIR spectrometer was fitted with a contact probe and a fiber-optic cable, along with an integrated 5 W halogen light source to emit and capture light from the sample surface. The spectral resolutions of the vis-NIR spectrometer were 3 nm between 350 and 1 000 nm, 8 nm between 1 000 and 1 900 nm, and 6 nm between 1 900 and 2 500 nm. At the end of spectral measurement, each spectral reflectance was exported at 1 nm intervals, which resulted in 2 151 spectral points. After scanning every ten soil samples, a white plate made from polytetrafluoroethylene was used to calibrate the vis-NIR spectrometer. The mean spectral reflectance was calculated from three replicates for each soil sample. The mean vis-NIR spectra of soil samples from A, B, and A+B horizons are shown in Fig. 1.

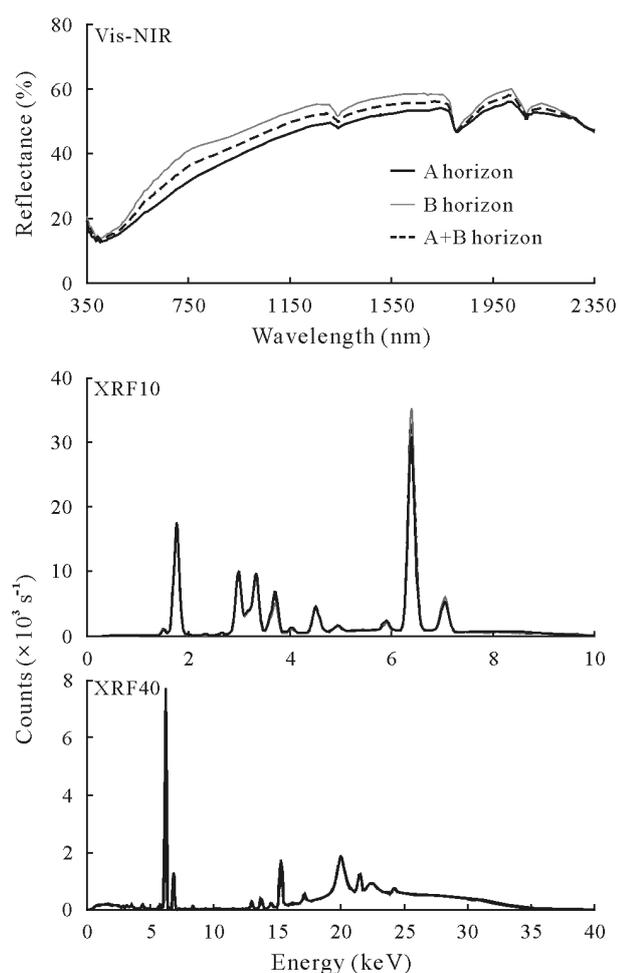


Fig. 1 Mean spectra of visible near-infrared (vis-NIR) and portable X-ray fluorescence (pXRF) of beams at 0–40 (XRF40) and 0–10 keV (XRF10) of soils, developed from loess material, in A, B, and A+B horizons.

The same setup was used for measurement with a pXRF spectrometer (Delta Premium, Olympus Scientific Solutions Americas Inc., Waltham, USA) in Geochem Mode, which runs for a duration of 60 s in a two-beam configuration: beam 1 at 0–40 keV and beam 2 at 0–10 keV. Before measurements, 316 stainless steel was used to calibrate the pXRF spectrometer. The pXRF spectra were used to obtain elemental concentrations according to an internal Compton normalization method. Some elements had low concentrations in the soil samples and were not detected using the pXRF spectrometer. The concentrations of elements (Zr, Sr, Rb, Zn, Mn, Ti, Ca, Si, Fe, Al, and Mg) were measured for all soil samples. The spectrum of beam 1 (0–40 keV) had 2048 spectral points and the spectrum of beam 2 (0–10 keV) had 512 spectral points. Fig. 1 shows the mean pXRF spectra of beam 1 (XRF40) and beam 2 (XRF10) of soil samples from A, B, and A+B horizons.

Vis-NIR and pXRF spectra preprocessing

The vis-NIR spectra were preprocessed as follows: i) all

spectral reflectances were converted to absorbance spectra; ii) Savitzky-Golay filter with a window size of 11 nm and a polynomial order of 2 was applied to smooth the spectra. The Savitzky-Golay filter matches a specific polynomial regression on the signal with defined polynomial order and window size (Savitzky and Golay, 1964); iii) the spectral ranges from 350 to 500 nm and 2 450 to 2 500 nm were deleted to remove noise, and every 10 th spectrum ranging from 500 to 2 450 nm was chosen to minimize high dimensionality and collinearity (Zhang and Hartemink, 2019; Benedet *et al.*, 2020; Wan *et al.*, 2020); and iv) the standard normal variate transformation was applied to normalize the spectra (Barnes *et al.*, 1989). R version 3.4.3 software was used for these processes (R Core Team, 2016).

The Savitzky-Golay filter was used to smooth both XRF40 and XRF10 spectra, with window sizes of 11 and 7 nm, respectively, and a polynomial order of 2 in R version 3.4.3 software, using the Savitzky-Golay function in the prospectr package (Stevens *et al.*, 2013). The Savitzky-Golay filter was suggested for preprocessing the pXRF spectra by other studies (O'Rourke *et al.*, 2016b; Xu *et al.*, 2019; Zhang and Hartemink, 2019).

The processed vis-NIR and pXRF (XRF40 and XRF10) spectra were individually scaled to one standard deviation and zero mean, and then a principal component analysis (PCA) was applied to select the first 10 principal components (PCs) for each type of spectrum. The PCA analysis was undertaken using the prcomp function in R version 3.4.3. Five types of predictor were compared for predicting soil properties: i) 10 PCs from vis-NIR spectra, ii) 10 PCs from XRF40 spectra, iii) 10 PCs from XRF10 spectra, iv) concatenating 20 PCs from combined XRF40 and XRF10 (XRF40+XRF10) spectra, and v) concatenating 30 PCs from combined XRF40, XRF10, and vis-NIR (XRF40+XRF10+vis-NIR) spectra.

Predictive models

The 348 soil samples collected from A and B horizons were randomly separated into a calibration ($n = 244$, 70%) and a validation ($n = 104$, 30%) dataset using the sample function in R version 3.4.3. The calibration dataset contained 123 samples from A horizon and 121 samples from B horizon, and the validation dataset contained 53 samples from A horizon and 51 samples from B horizon. The Cubist model was applied to predict clay, silt, sand, TC, and pH from different combinations of PCs of soil spectra. The Cubist model is a rule-based method derived from Quinlan's M5 model tree (Quinlan, 1992), in which each tree branch is built on a linear regression model instead of a discrete value (Minasny and McBratney, 2008). Other models (including multiple linear regression, random forest, and partial least square regression) were briefly compared with the Cubist model, and the results did not show any superiority over

the Cubist model (data not shown). The Cubist model was developed using the train function of caret package (Kuhn, 2008) in R version 3.4.3. A 10-fold cross validation was applied to the calibration dataset to prevent overfitting in the train function. In the cross validation, parameters including committees (1, 10, and 20) and neighbors (0, 5, and 9) were tested for the Cubist model. The final model with specific committee and neighbor was chosen based on the lowest root mean squared error (RMSE) value. The validation dataset was individually checked for the Cubist model. The coefficient of determination (R^2), RMSE, and residual prediction deviation (RPD) were determined for the training process and independent validation (Chang *et al.*, 2001).

A multiple linear regression (MLR) model was applied to predict clay, silt, sand, TC, and pH directly from important geochemical elemental concentrations (Zr, Sr, Rb, Zn, Mn, Ti, Ca, Si, Fe, Al, and Mg). Different combinations of the elemental variables were used in the MLR models to select the optimal MLR model with the highest R^2 value with certain elements. An exhaustive search of the optimal MLR model was conducted on the calibration dataset *via* the RegBest function of FactoMineR package (Husson *et al.*, 2018) in R version 3.4.3. The optimal MLR model was then evaluated using the independent validation dataset. The R^2 , RMSE, and RPD were determined for both the calibration and validation processes.

For each soil property (clay, sand, silt, TC, and pH) and different combinations of pXRF and vis-NIR spectral data, prediction models (Cubist and MLR) were evaluated for various datasets: i) A horizon, ii) B horizon, and iii) combined A+B horizon.

RESULTS

Characterization of soil horizons

The distributions of soil particle size fractions (clay, silt, and sand), TC, pH, and elements of the 348 soil samples are shown in Fig. 2 for A, B, and A+B horizons. The A horizon had an average of 12% sand, 66% silt, and 22% clay. The Bt horizon had an average of 15% sand, 58% silt, and 27% clay. In the soil profiles, clay, Mg, Al, Fe, and Sr increased from A to Bt horizon, whereas TC, TN, pH, Ca, Mn, and Zn noticeably decreased from A to Bt horizon. The soil samples had higher concentrations of Al, Mg, Fe, Ca, and Si and lower concentrations of Mn, Ti, Rb, Zn, Zr, and Sr. The B horizon had noticeably higher coefficients of variation (CV) than A horizon for pH, TC, silt, Mg, Al, Ca, Ti, Zn, Mn, Rb, Zr, and Sr. In particular, TC, Ca, and Mn had very high CV values in B horizon (Table I). The distributions of clay, silt, sand, TC, and pH values in A+B horizon of the calibration and validation datasets are shown as histograms in Fig. 3.

Clay, silt, sand, pH, and TC had similar distributions in the calibration and validation datasets.

The Pearson correlation coefficients (r) between soil properties and elements in A, B, and A+B horizons are shown in Fig. 4. As expected, sand content was strongly negatively correlated with Al, Ti, Rb, Sr, and Zr in A, B, and A+B horizons, whereas silt and clay contents were strongly positively correlated with Fe, Al, Ti, Rb, Sr, and Zr in A, B, and A+B horizons. In particular, sand and silt contents showed strong correlation with soil elements such as Fe, Al, Ti, Rb, Sr, and Zr in each horizon. Generally, strong correlations were observed between soil properties (clay, sand, silt, TC, and pH) and elements (Fe, Al, Ca, Ti, Rb, Sr, and Zr) in A and B horizons. The correlations were weaker between soil properties (pH and TC) and elements (Si, Fe, Ca, Ti, Zr, and Zn) when the data for A and B horizons were combined (Fig. 4).

Principal component analysis and spectral variance

The 10 PCs explained approximately 100% of the cumulative variance in the vis-NIR spectra, and approximately 90% and 80% of the cumulative variance in the XRF10 and XRF40 spectra, respectively (Table I). The loadings of the first three principal components (PC1, PC2, and PC3) of the PCA on vis-NIR, XRF10, and XRF40 spectra in the validation dataset for A, B, and A+B horizons are shown in Fig. 5. Valleys and peaks of the loadings showed important features of vis-NIR, XRF10, and XRF40 spectra on the first three PCs (Fig. 5). PC1 for the vis-NIR spectra explained 52% of the variance in A horizon and had strong positive weightings (> 0.60) at 1 000–1 400 nm and strong negative weightings (< -0.6) above 1 900 nm. PC2 for the vis-NIR spectra explained 28% of the variance in A horizon and had strong positive weightings at 550–600 and 1 450–1 880 nm and strong negative weightings at 800–1 060 nm. PC3 for the vis-NIR spectra explained 15% of the variance in A horizon, and had strong positive weightings at 640 nm and strong negative weightings at 1 400 nm, corresponding to the O–H bonds in water molecules. The loading patterns of vis-NIR spectra were very similar in A, B, and A+B horizons, except that PC1 in B and A+B horizons had strong negative weightings at 1 400 nm. The loading patterns of XRF10 and XRF40 spectra showed more peaks and vibrations compared to that of vis-NIR spectra (Fig. 5), because the raw XRF spectra contained more abrupt peaks (Fig. 1). The XRF10 and XRF40 spectra had stronger positive weightings than negative weightings. The loading patterns of XRF40 spectra were very similar in A, B, and A+B horizons, in which PC1 had broad positive weightings above 20 keV.

Soil property prediction performance with pXRF and vis-NIR spectra

The calibration and validation results of different in-

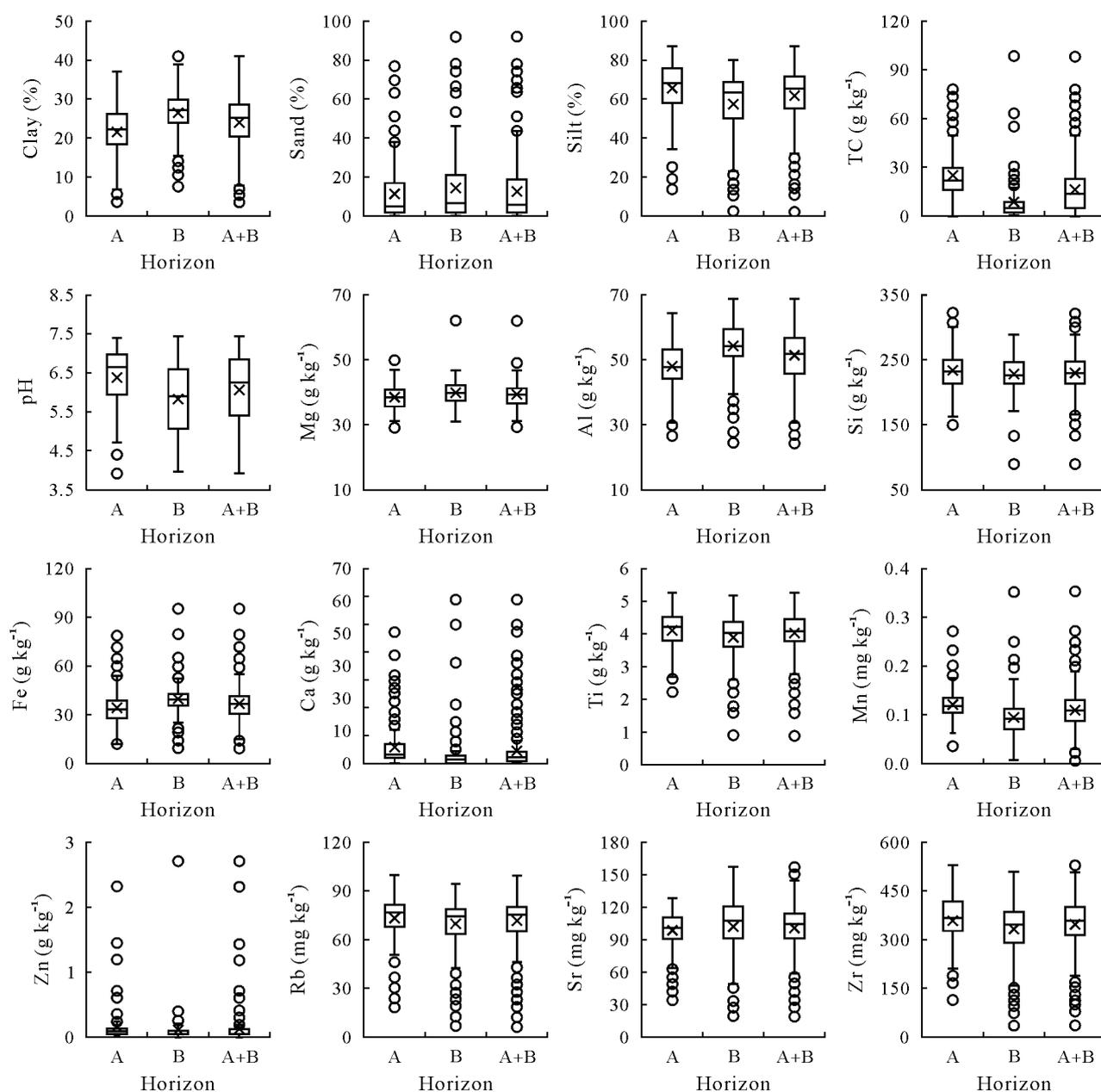


Fig. 2 Distributions of clay, sand, silt, total carbon (TC), pH, and elements (Mg, Al, Si, Fe, Ca, Ti, Mn, Zn, Rb, Sr, and Zr) of soils, developed from loess material, in A, B, and A+B horizons. Each box shows the minimum (lower line end), 1 st quartile (Q1, lower box section), median (line within box), 3rd quartile (Q3, upper box section), maximum (upper line end), and outliers (open circles). Interquartile range (IQR) = $Q3 - Q1$; minimum = $Q1 - 1.5IQR$; maximum = $Q1 + 1.5IQR$.

put variables and at different horizons for predicting soil properties with Cubist model are shown in Table SI (see supplementary material for Table SI). The predictions of the Cubist model using different input variables and spectral data from A, B, and A+B horizons had high overall accuracy. The prediction accuracy with the validation dataset was similar to that with the calibration dataset, which indicated that there was no overfitting issue in the Cubist model.

In general, soil clay, silt, and sand had higher prediction accuracy in B horizon than in A or A+B horizon. The Cubist model with XRF40 spectra had higher validation R^2 values

for predicting soil particle size fractions (validation $R^2 = 0.79, 0.93,$ and 0.96 for clay, silt, and sand, respectively) than that with XRF10 spectra (validation $R^2 = 0.79, 0.84,$ and 0.87 for clay, silt, and sand) in B horizon (Fig. 6). Combining the two spectra (XRF40+XRF10) only slightly improved the prediction performance for sand (validation $R^2 = 0.97$), silt (validation $R^2 = 0.94$), and clay (validation $R^2 = 0.79$) over using XRF40 spectra in B horizon. The vis-NIR spectra provided the better prediction performance for clay in A horizon (validation $R^2 = 0.84$) than in B (validation $R^2 = 0.42$) and A+B horizons (validation $R^2 = 0.62$). Combining

TABLE I

Cumulative variance explained by the first 10 principal components (PCs) for visible-near infrared (vis-NIR) and portable X-ray fluorescence (pXRF) of beams at 0–10 (XRF10) and 0–40 keV (XRF40) spectra of soils, developed from loess material, in A, B, and A+B horizons in the validation dataset

PC	Vis-NIR			XRF10			XRF40		
	A	B	A+B	A	B	A+B	A	B	A+B
1	52	68	66	43	36	43	54	57	53
2	80	85	87	69	69	64	63	66	63
3	95	93	96	77	77	74	69	70	68
4	98	97	98	83	81	80	72	73	71
5	99	98	99	85	84	83	73	74	72
6	99	99	99	87	85	85	75	76	74
7	100	100	100	89	87	87	76	77	74
8	100	100	100	90	88	88	77	78	75
9	100	100	100	91	89	89	78	79	76
10	100	100	100	92	90	89	79	80	77

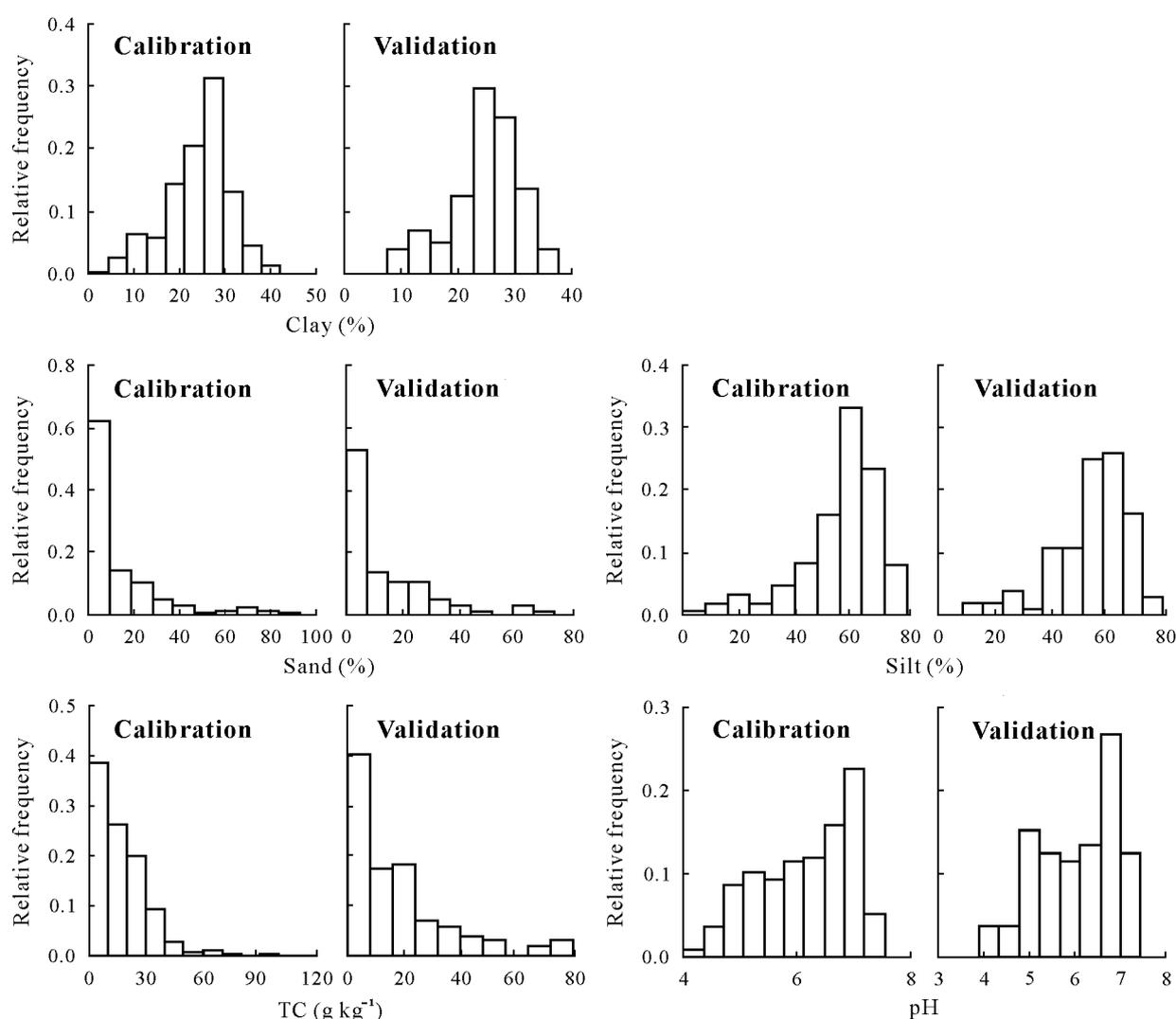


Fig. 3 Histograms of clay, sand, silt, total carbon (TC), and pH of soils, developed from loess material, in A+B horizon in the calibration and validation datasets.

XRF40+XRF10 and vis-NIR spectra slightly improved the prediction performance for clay, silt, and sand in A, B, or A+B horizon (Fig. 6).

The Cubist model for the prediction of TC in A+B horizon generally had higher prediction accuracy than in the individual A or B horizon. Using XRF40 spectra produced a

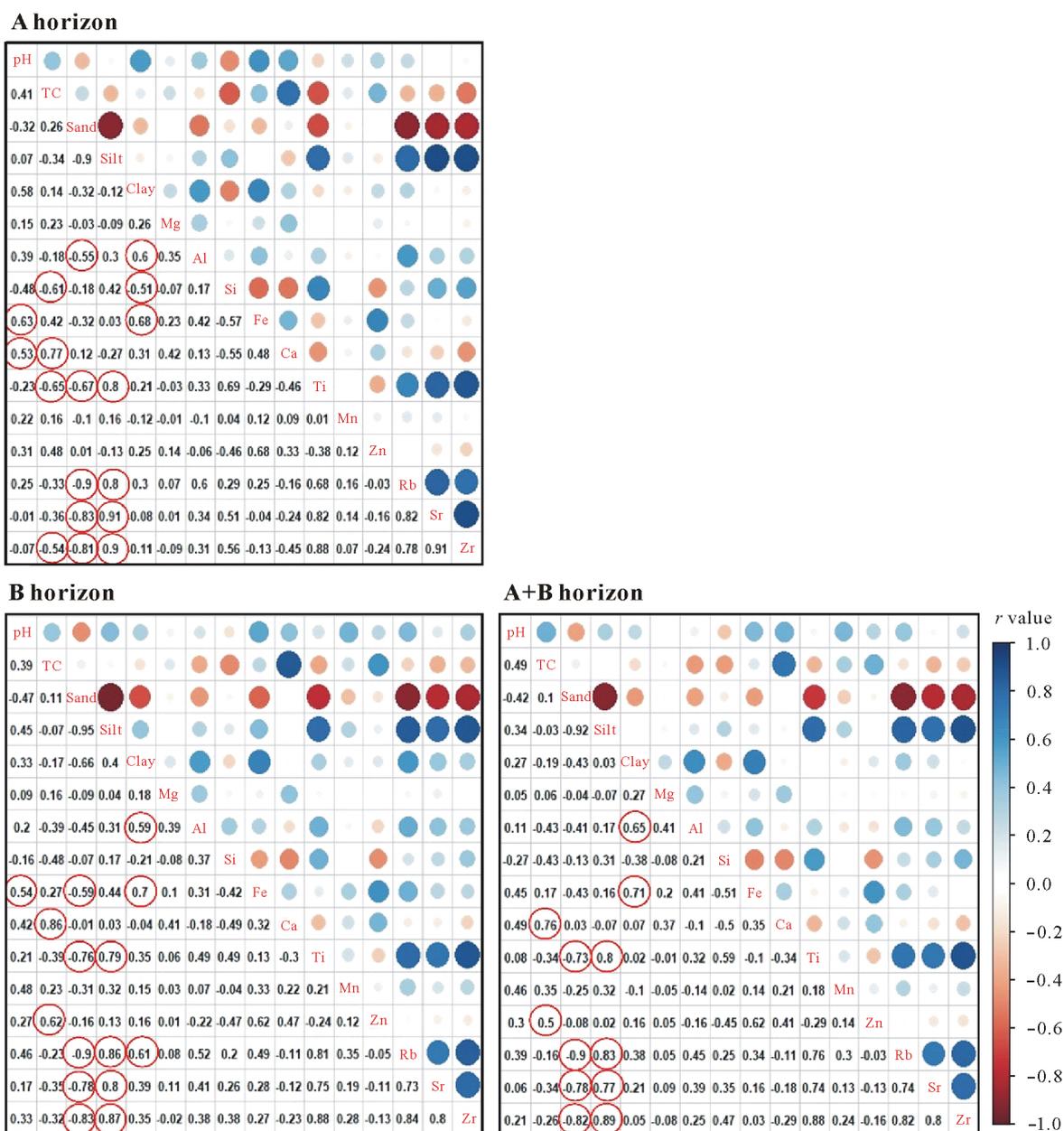


Fig. 4 Pearson correlations between properties and elements of soils, developed from loess material, in A, B, and A+B horizons. Blue circles indicate greater positive correlations, while filled red circles indicate greater negative correlations. The numbers in red open circles show correlation coefficient $r \geq 0.5$ or $r \leq -0.5$.

higher validation R^2 (0.84) than using XRF10 spectra (0.81) in A+B horizon. Combining the two pXRF (XRF40+XRF10) spectra produced a better prediction for TC (validation $R^2 = 0.87$) than using individual pXRF (XRF40 or XRF10) spectra in A+B horizon. In particular, the use of vis-NIR spectra generated higher prediction accuracy for TC (validation $R^2 = 0.91$) in A+B horizon than the use of pXRF spectra. Combining vis-NIR, XRF40, and XRF10 (vis-NIR+XRF40+XRF10) spectra made the best prediction for TC (validation $R^2 = 0.93$) compared to using individual pXRF or vis-NIR spectra.

The prediction performances for pH in A, B, and A+B horizons differed from those for other soil properties (clay,

silt, sand, and TC); the prediction accuracy was slightly lower for pH. The XRF40 and vis-NIR spectra had higher prediction accuracy (validation $R^2 = 0.63$ and 0.79 , respectively) in A horizon, whereas XRF10 and XRF40+XRF10 had higher prediction accuracy (validation $R^2 = 0.69$ and 0.66 , respectively) in B horizon. The vis-NIR+XRF40+XRF10 spectra explained more variation than using individual XRF40 or XRF10 spectra in relation to pH. In contrast to those for the other soil properties, the highest prediction accuracy (validation $R^2 = 0.79$) for pH was obtained when using only vis-NIR spectra in A horizon, while using XRF40+XRF10+vis-NIR spectra had the same prediction performance (validation

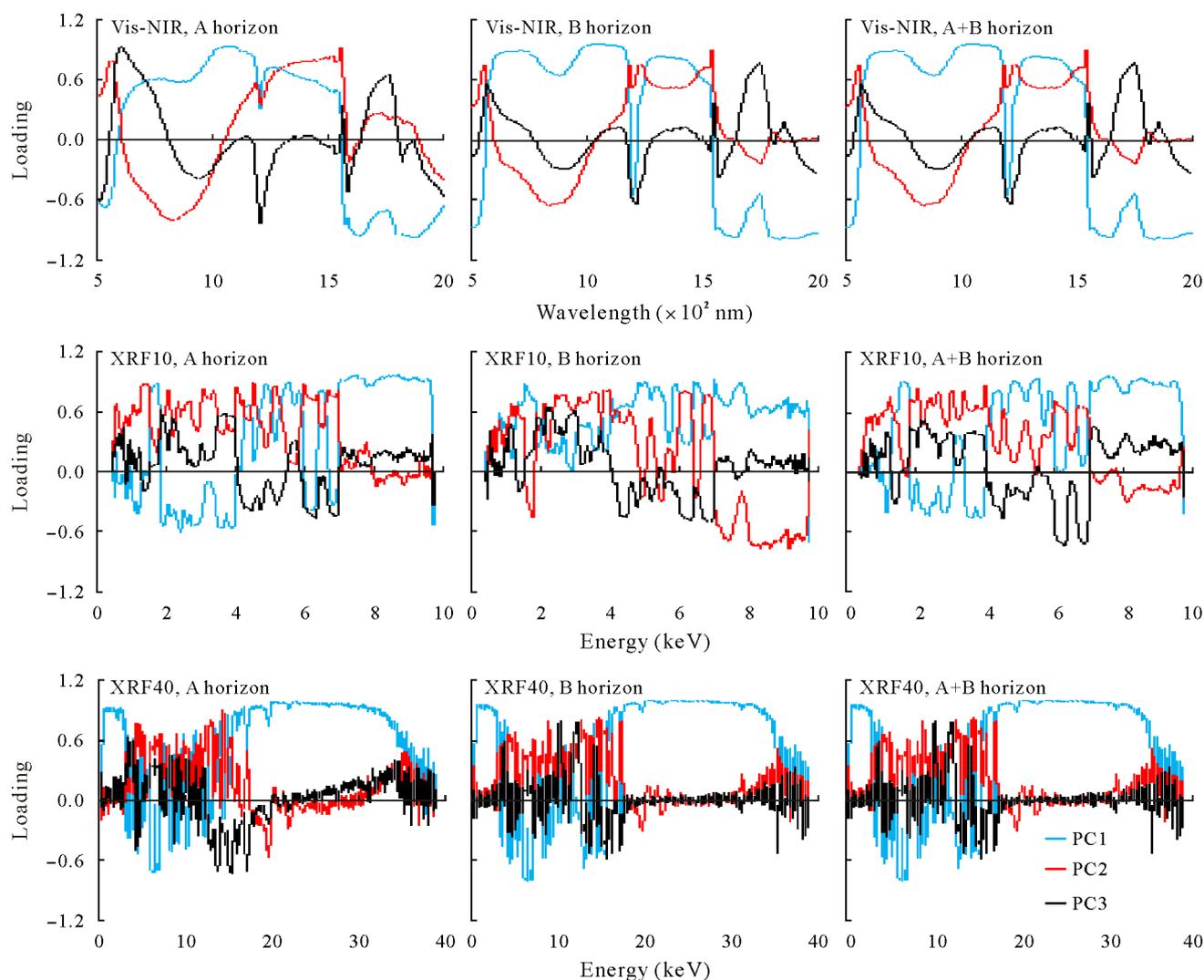


Fig. 5 Loadings of the first three principal components (PC1, PC2, and PC3) of the principal component analysis on visible-near infrared (vis-NIR) and portable X-ray fluorescence (pXRF) of beams at 0–10 (XRF10) and 0–40 keV (XRF40) spectra of soils, developed from loess material, in A, B, and A+B horizons in the validation dataset.

$R^2 = 0.79$) in A+B horizon (Fig. 6).

Considering the obtained results, the predictions of clay, silt, and sand were better in B horizon, whereas the predictions of TC and pH were better in A+B horizon when combining vis-NIR and pXRF spectra in the model. The pXRF spectra performed better in predicting clay, sand, and silt, whereas the vis-NIR spectra performed better in predicting TC and pH. Combining all the spectra improved the prediction accuracy of most soil properties. The scatterplots of the predicted and measured clay, silt, sand, TC, and pH in these horizons are shown in Fig. 7.

Effect of chemical element selection on prediction performance of soil properties

Table II shows the calibration and validation results of the MLR models for predicting soil properties with pXRF elements. Different pXRF elements with the highest overall

accuracy were used for A, B, and A+B horizons. It showed that the elements were used successfully to predict the clay, silt, sand, TC, and pH in A horizon. In particular, the MLR models for clay, silt, sand, and pH had the highest prediction accuracy (validation $R^2 = 0.81, 0.87, 0.94$, and 0.53 , respectively) in A horizon, whereas TC was predicted most accurately (validation $R^2 = 0.87$) in A+B horizon (Table II). The Al, Ca, and Rb were used to predict almost all clay, silt, sand, TC, and pH in individual A or B and combined A+B horizons (Table II). The Al, Ca, Fe, Si, Zn, Ti, Rb, Zr, and Sr were most important for predicting soil particle size (clay, silt, and sand) in almost every horizon. Sand was better predicted (validation $R^2 = 0.94$) by MLR with elements than the other soil properties (silt, validation $R^2 = 0.87$; clay, validation $R^2 = 0.81$; TC, validation $R^2 = 0.81$; and pH, validation $R^2 = 0.53$). The Al, Ca, Zn, and Sr

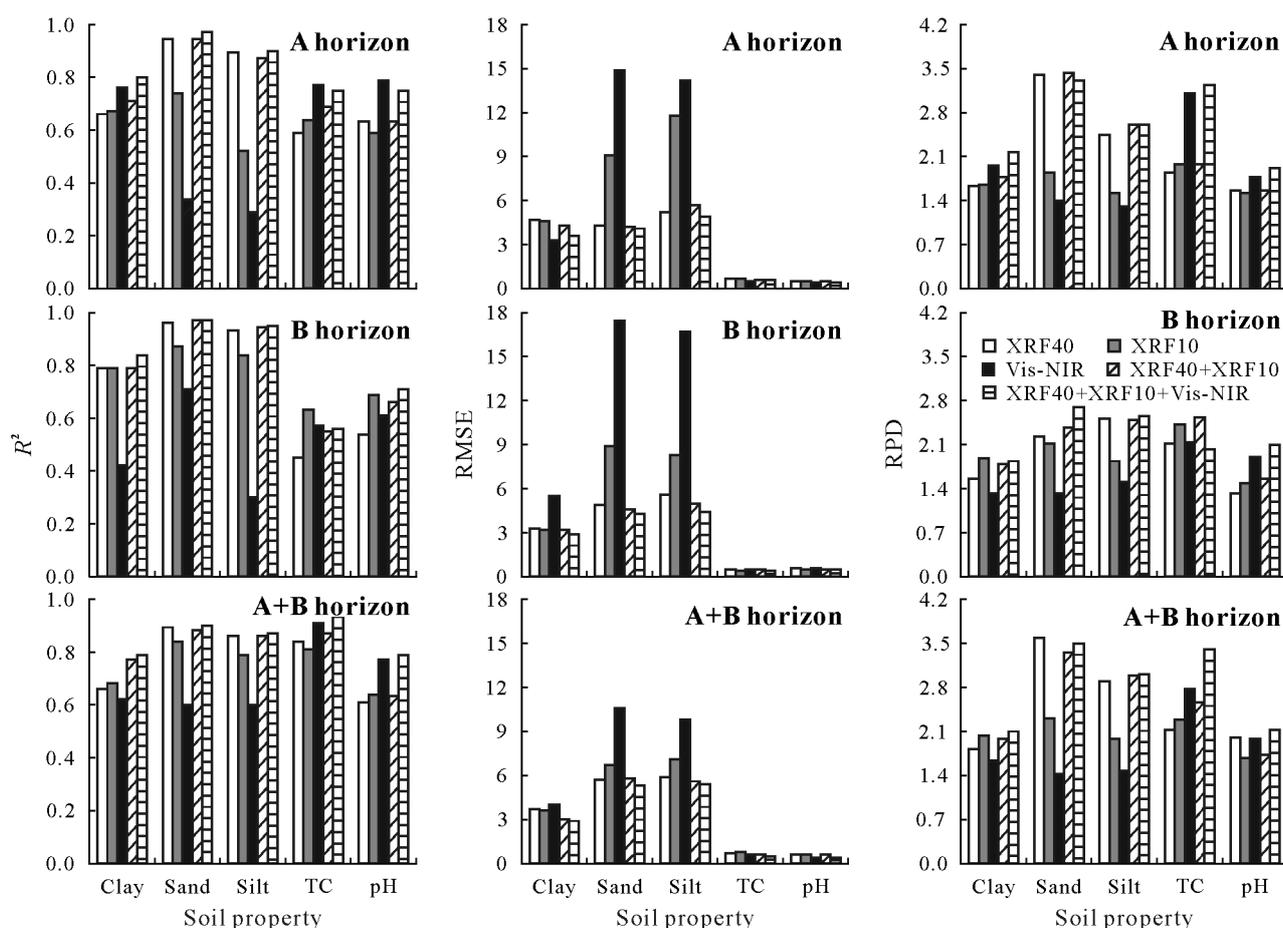


Fig. 6 Performance statistics of Cubist model validations set for five properties of soils, developed from loess material, in A, B, and A+B horizons using spectral data of visible-near infrared (vis-NIR) and portable X-ray fluorescence (pXRF) of beams at 0–10 (XRF10) and 0–40 keV (XRF40). TC = total carbon; R^2 = coefficient of determination; RMSE = root mean squared error; RPD = residual prediction deviation.

were used to predict TC, whereas Mg, Al, Ca, Mn, and Zr were used to predict pH in all horizons (Table II).

DISCUSSION

Model performance using pXRF and vis-NIR spectra for prediction of soil properties

In our study, good results were obtained for predicting clay, silt, sand, pH, and TC from vis-NIR and pXRF spectra. The PCA+Cubist method predicted clay, silt, and sand from vis-NIR spectra with good accuracy (validation $R^2 = 0.84$, 0.60, and 0.71, respectively). Clay and sand were more accurately predicted than silt, which may be due to the special characteristics of clay (O–H in water, FeOH–, AlOH–, and MgOH– in the structure of mineral crystal lattice) and corresponding spectral features in the vis-NIR spectra (Bendor and Banin, 1995). Other studies have also reported moderate to high prediction accuracy (R^2 ranging from 0.42 to 0.83) for clay content using vis-NIR spectra (e.g., Ng *et al.*, 2019; Zhang and Hartemink, 2020) (Table III). Duda *et al.* (2017) and Zhang and Hartemink (2020) observed

lower prediction accuracy for sand (validation $R^2 = 0.18$ and 0.39, respectively) and silt contents (validation $R^2 = 0.41$ and 0.53), whereas Tümsava *et al.* (2019) observed high prediction accuracy (validation $R^2 = 0.82$ and 0.91). Rossel *et al.* (2009) related the low prediction accuracy for clay and sand contents using the vis-NIR spectrometer to an overlap of different effects that hide the absorption features at 1 400 nm, due to the presence of FeOH– and/or AlOH–. In addition, the prediction accuracy for clay, sand, and silt differed for different horizons.

The prediction models for clay, silt, and sand using pXRF spectra showed better performance than those using vis-NIR spectra in our study. O'Rourke *et al.* (2016a) and Zhang and Hartemink (2020) reported similar results. According to Tóth *et al.* (2019) and Zhang and Hartemink (2020), pXRF elements had higher correlations with clay, silt, and sand than vis-NIR spectra. Additionally, when pXRF (XRF40+XRF10) and vis-NIR spectra were applied together to predict clay, silt, and sand, the prediction performance for clay was similar to that of pXRF (XRF40+XRF10) alone. The combination of pXRF and vis-NIR spectra was better than using individual

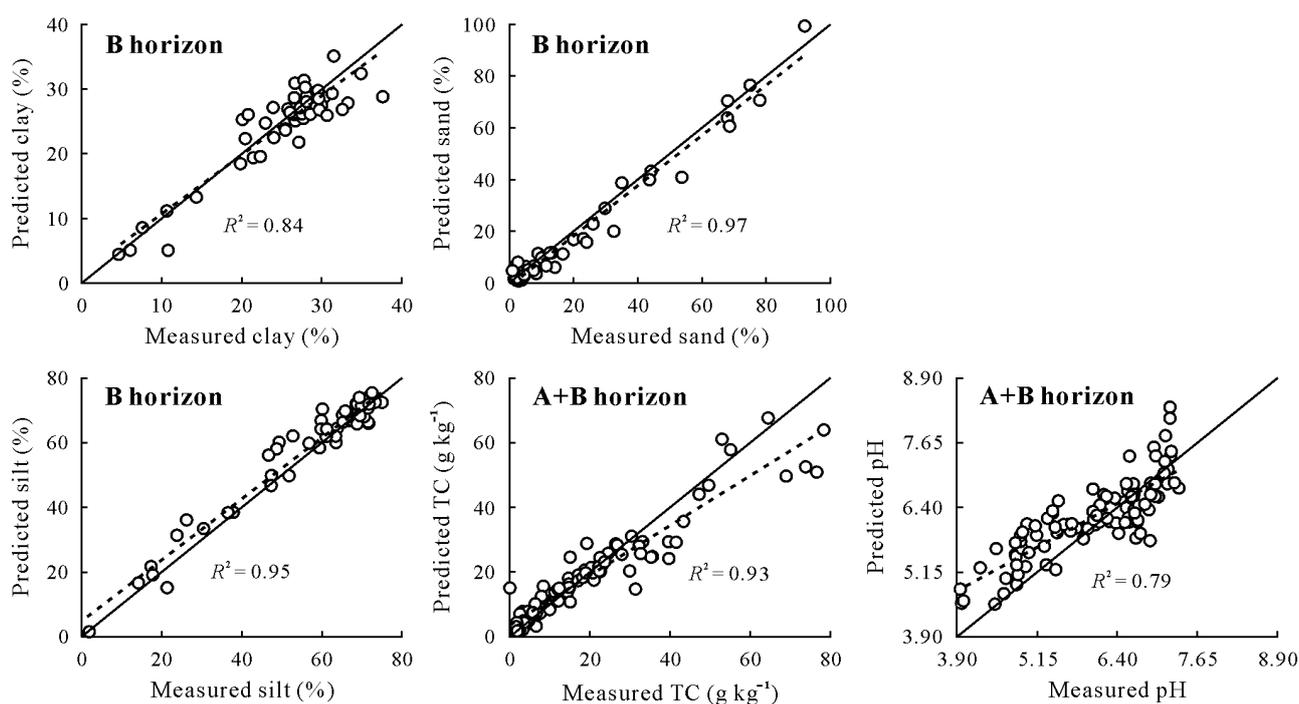


Fig. 7 Scatterplots of measured and predicted properties of soils, developed from loess material, in the validation dataset. The soils were predicted using a Cubist model with the spectra of the combined vis-NIR+XRF10+XRF40 as the input data. The solid line is the 1:1 line and the dotted line is the trend line. The best predictions were obtained for clay, sand, and silt contents in B horizon and for total carbon (TC) and pH in A+B horizon. vis-NIR = visible-near infrared; XRF10 = portable X-ray fluorescence (pXRF) of beams at 0–10 keV; XRF40 = pXRF of beams at 0–40 keV.

TABLE II

Calibration and validation results^{a)} of multiple linear regression for predicting properties of soils, developed from loess material, in A, B, A+B horizons with elemental concentrations

Soil property	Horizon	Elements	Calibration			Validation		
			R^2	RMSE	RPD	R^2	RMSE	RPD
Clay	A	Al, Si, Ca, Rb, Zr	0.74	2.94	1.92	0.81	3.58	2.26
	B	Al, Si, Fe, Ca, Rb	0.65	3.28	1.66	0.74	2.65	1.96
	A+B	Al, Si, Fe, Ca, Mn, Zn, Rb, Zr	0.78	3.22	2.10	0.69	3.32	1.76
Sand	A	Al, Si, Fe, Zn, Rb, Sr, Zr	0.85	4.53	2.53	0.94	4.12	4.03
	B	Si, Fe, Ti, Zn, Rb, Sr, Zr	0.85	5.42	2.57	0.89	5.52	2.88
	A+B	Si, Fe, Ca, Ti, Mn, Rb, Sr, Zr	0.87	5.07	2.76	0.78	6.48	2.10
Silt	A	Al, Fe, Zn, Sr, Zr	0.84	5.25	2.45	0.87	5.41	2.76
	B	Al, Ca, Ti, Zn, Rb, Sr, Zr	0.85	5.04	2.54	0.86	4.76	2.64
	A+B	Mg, Al, Ca, Ti, Zn, Rb, Sr, Zr	0.84	5.48	2.50	0.81	5.89	2.22
TC	A	Al, Ca, Ti, Zn, Sr	0.82	0.64	2.29	0.56	0.68	1.47
	B	Al, Si, Ca, Ti, Mn, Zn, Sr	0.95	0.31	4.36	0.66	0.43	1.57
	A+B	Al, Fe, Ca, Mn, Zn, Rb, Sr	0.79	0.65	2.15	0.81	0.79	2.30
pH	A	Mg, Al, Si, Ca, Ti, Mn, Zr	0.57	0.42	1.49	0.53	0.48	1.47
	B	Mg, Al, Ca, Mn, Rb, Sr, Zr	0.55	0.58	1.45	0.44	0.63	1.33
	A+B	Mg, Al, Si, Ca, Mn, Rb, Sr, Zr	0.58	0.52	1.51	0.52	0.56	1.43

^{a)} R^2 = coefficient of determination; RMSE = root mean squared error; RPD = residual prediction deviation; TC = total carbon.

^{b)} Data in bold show the best validation model.

vis-NIR spectra for prediction of clay content in B horizon, while the prediction performance of combined pXRF and vis-NIR spectra was worse than that of individual vis-NIR spectra for clay content in A horizon. However, the prediction performance of combined pXRF and vis-NIR spectra was better than that of individual vis-NIR spectra for sand and silt contents in every horizon. Several studies reported that predictions of soil particle size (clay, silt, and sand) improved

when vis-NIR and pXRF spectra were combined (Table III) (Duda *et al.*, 2017; Zhang and Hartemink, 2020). In particular, Zhang and Hartemink (2020) predicted clay, sand, and silt contents fairly well with a wide range of datasets. However, they reported that prediction performance of clay, sand, and silt contents was affected by textural classes. For example, sandy loam, sandy clay loam, and clay loam soils had better prediction than loam sandy soil.

TABLE III

Summary of studies predicting soil properties using visible-near infrared (vis-NIR) spectra, portable X-ray fluorescence (pXRF) spectra, and pXRF elements

Model ^{a)}	Input	Validation coefficient of determination					Reference
		Clay	Sand	Silt	Total C	pH	
PLSR	Vis-NIR spectra	–	–	–	–	0.72	Zornoza <i>et al.</i> (2008)
		0.67	0.44	–	0.54	0.66	Ng <i>et al.</i> (2019)
MLR RF	pXRF elements	0.88, 0.98	0.85, 0.89	0.68, 0.88	–	–	Zhu <i>et al.</i> (2011)
	pXRF spectra	–	–	–	0.77	–	Wang <i>et al.</i> (2015)
	pXRF elements	–	–	–	0.58	–	Cardelli <i>et al.</i> (2017)
	Vis-NIR spectra	–	–	–	0.81	–	Wang <i>et al.</i> (2015)
	–	–	–	–	0.60	–	Cardelli <i>et al.</i> (2017)
PSR	Vis-NIR+pXRF spectra	–	–	–	0.83	–	Wang <i>et al.</i> (2015)
	Vis-NIR spectra	–	–	–	0.90	–	Wang <i>et al.</i> (2015)
PSR	Vis-NIR+pXRF spectra	–	–	–	0.93	–	Wang <i>et al.</i> (2015)
Cubist	pXRF spectra	–	–	–	–	0.42	O'Rourke <i>et al.</i> (2016a)
		0.77	0.79	–	0.76	0.60	O'Rourke <i>et al.</i> (2016b)
	Vis-NIR spectra	–	–	–	–	0.56	O'Rourke <i>et al.</i> (2016a)
	–	0.86	0.35	–	0.38	0.75	O'Rourke <i>et al.</i> (2016b)
	–	0.76	0.60	–	0.70	0.76	Ng <i>et al.</i> (2019)
ENET	pXRF elements	–	–	–	0.62	–	Cardelli <i>et al.</i> (2017)
	Vis-NIR spectra	–	–	–	0.73	–	Cardelli <i>et al.</i> (2017)
SVR	Vis-NIR spectra	0.53	0.18	0.39	0.70	–	Duda <i>et al.</i> (2017)
	pXRF spectra	0.15	0.24	0.13	0.68	–	Duda <i>et al.</i> (2017)
	Vis-NIR+pXRF spectra	0.49	0.25	0.45	0.80	–	Duda <i>et al.</i> (2017)
MLR	pXRF elements	0.85	0.89	0.91	0.65	0.15	Zhang and Hartemink (2020)
	PCA+Cubist	pXRF spectra	0.86	0.90	0.92	0.61	0.12
–		0.77	0.88	0.86	0.87	0.63	This study ^{b)}
Vis-NIR spectra		0.62	0.60	0.60	0.91	0.77	This study ^{b)}
–	Vis-NIR+pXRF spectra	0.79	0.90	0.87	0.93	0.79	This study ^{b)}

^{a)}PLSR = partial least square regression; MLR = multiple linear regression; RF = random forest; PSR = penalized spline regression; ENET = elastic net regression; SVR = support vector regression; PCA = principal component analysis.

^{b)}Data in this study are from the combined A+B horizon.

The PCA+Cubist method achieved good prediction of TC using the vis-NIR spectrometer. In particular, TC was predicted more accurately in A+B horizon (validation $R^2 = 0.91$) than in A (validation $R^2 = 0.77$) or B (validation $R^2 = 0.57$) horizon. In general, organic matter, TC, and organic carbon of soils have been quite well predicted using vis-NIR spectra. For instance, some studies achieved fairly high prediction accuracy ($R^2 = 0.61$ – 0.90) for TC with vis-NIR spectra (Duda *et al.*, 2017; Zhang and Hartemink, 2020). However, TC prediction had low accuracy ($R^2 = 0.38$) using vis-NIR spectra in O'Rourke *et al.* (2016a) (Table III). Soil organic matter has high absorbance in some regions of the vis-NIR spectra owing to the N–H, C–H, and C–O bonds present, and can be better predicted using vis-NIR spectra (Rossel and Webster, 2011). In our study, the XRF40 and XRF10 spectra were less accurate in predicting TC than vis-NIR spectra. Cardelli *et al.* (2017) ($R^2 = 0.58$), Duda *et al.* (2017) ($R^2 = 0.68$), and Zhang and Hartemink (2020) ($R^2 = 0.61$) reported that the prediction accuracy for TC was similarly low using pXRF data (Table III). However, in our study, when vis-NIR and pXRF (XRF40+XRF10) spectra were applied together, the prediction accuracy of TC slightly improved, with the validation R^2 increasing from 0.91 (vis-NIR) to 0.93 (vis-NIR+pXRF) in A+B horizon. Similar improvement in prediction accuracy has been reported when

both vis-NIR and pXRF spectra were used (Duda *et al.*, 2017; Zhang and Hartemink, 2020).

In this study, pH prediction was less accurate than those for other soil properties. The pH prediction was better in A horizon when only vis-NIR spectra were used, while combining pXRF (XRF40+XRF10) and vis-NIR spectra produced similar prediction performance for pH in A+B horizon. Soil pH was indirectly predicted by vis-NIR spectra because of its strong correlations with other soil properties. O'Rourke *et al.* (2016b) and Zhang and Hartemink (2020) reported that pXRF and vis-NIR spectra were not enough to accurately predict pH, whereas Sharma *et al.* (2014) and Teixeira *et al.* (2018) reported that pH was predicted well by pXRF spectra, with R^2 of 0.77 and 0.85 (Table III), because of its strong correlations with elemental data from pXRF (*e.g.*, Mn and Al). Zhang and Hartemink (2020) obtained very low prediction performance for pH due to poor correlations between pXRF elements/soil properties and pH. However, in our study, moderate correlations between soil properties (such as clay, sand, silt, Fe, and Ca) and pH were observed. In addition, some elements from pXRF such as Mg, Al, Ca, Mn, and Zr used in MLR were effectively used to predict soil pH in every horizon (A, B, and A+B). These results indicated that prediction performance of pH

was affected by correlations between pXRF elements, soil particle size, and pH.

According to Mancini *et al.* (2019), when elemental concentrations in A, B, and C horizons were predicted separately from pXRF data, the prediction performances were different. They observed higher prediction accuracy in C horizon than in A and B horizons due to several effects, including land use (agricultural activities), soil management practices (less human influence), and variation of soil properties in the deep soil compared to surface and subsurface horizons. Andrade *et al.* (2020a) reported that prediction models for soil fertility performed better in A horizon than in B horizon when pXRF and vis-NIR spectra were used, while Benedet *et al.* (2020) reported that prediction models for some soil properties of soil subgroup performed better in B horizon than in A horizon when using pXRF spectra. However, Andrade *et al.* (2020a) reported that the best prediction result was obtained using pXRF spectra from combined A and B horizon. Similarly, Santana *et al.* (2018) and Zhang and Hartemink (2019) successfully used pXRF spectra, without considering soil horizons, to improve model accuracy for predicting soil properties. The parent material, weathering state, and leaching rate of soil samples increased the variations in the physical and chemical composition during the formation and pedogenetic processes (Zhu *et al.*, 2011).

From a practical perspective, combining vis-NIR and pXRF spectra produced better prediction performance for clay, silt, sand, TC, and pH. In particular, higher prediction accuracy for clay, sand, and silt was obtained in B horizon, whereas better prediction performance for TC and pH was obtained without considering soil horizons (A+B).

Assessment of model applicability

In this study, the model used was a simple method for predicting clay, silt, sand, TC, and pH using combined pXRF and vis-NIR spectral data with a machine learning algorithm (Cubist model). The observations clearly showed that the models using pXRF spectra had better prediction performance than those using vis-NIR spectra for silt and sand, and predicted soil particle size fractions (clay, silt, and sand) very well (validation $R^2 = 0.79, 0.97, \text{ and } 0.95$, respectively) in B horizon. When pXRF and vis-NIR spectra were combined, prediction performance showed little improvement for soil particle size fractions. The models using vis-NIR spectra alone had better prediction performance for TC and pH than those using pXRF spectra, and predicted TC and pH very well (validation $R^2 = 0.91 \text{ and } 0.79$, respectively) in A+B and A horizons. Combining pXRF and vis-NIR spectra produced a small improvement in prediction for TC.

Prediction performance is affected not only by soil physical and chemical properties, but also by different features such as environmental conditions, land use (Mancini *et al.*,

2019; Andrade *et al.*, 2020a), soil types (Benedet *et al.*, 2020), accumulation-leaching (Andrade *et al.*, 2020b), soil depths (Mancini *et al.*, 2019; Silva *et al.*, 2020), soil diagnostic horizons (Dos Santos *et al.*, 2014), mineralogical properties (Andrade *et al.*, 2020a), soil color, and different ranges of the dataset (Santana *et al.*, 2018; Zhang and Hartemink, 2019) from A, B, and A+B horizons. Therefore, from a practical perspective, the performance of this methodology should be applied and tested further in order to develop models for soils that developed from different parent materials, environmental conditions, and land uses.

CONCLUSIONS

The Cubist model was used with individual or combined pXRF and vis-NIR spectra through a PCA to investigate its performance in predicting clay, silt, sand, TC, and pH in soils developed in loess. Its prediction performance was evaluated for data from A, B, and A+B horizons. We found that the models using individual pXRF spectra produced better predictions for clay, silt, and sand, whereas those using individual vis-NIR spectra performed better in TC and pH predictions. Combining vis-NIR and pXRF spectra achieved higher prediction accuracy than using individual vis-NIR and pXRF spectra for most soil properties. The MLR model had the highest prediction accuracy for clay, silt, sand, and pH in A horizon and for TC in A+B horizon. We conclude that proximal sensors, namely vis-NIR and pXRF spectrometers, are good tools for predicting soil particle size fractions (sand, silt, and clay), TC, and pH in soils developed from loess. We suggest splitting the predictive models by horizon to achieve high prediction accuracy for clay, sand, and silt when using pXRF and vis-NIR spectra combined with the Cubist model. However, the best prediction performance for TC and pH was obtained without considering soil horizons (A+B) when using combined pXRF and vis-NIR spectra together with the Cubist model. The performance of this methodology should be applied and tested further in order to develop models for soils developed from different parent materials, environmental conditions, and land uses, with different machine learning algorithms.

ACKNOWLEDGEMENTS

The authors are grateful to Mr. Dave Evans, who sampled these pedons for his master project. The first author is supported by the Scientific Research Projects (BAP) (No. 2019-2757) of Eskisehir Osmangazi University for Postdoc research at the Department of Soil Science, University of Wisconsin-Madison.

SUPPLEMENTARY MATERIAL

Supplementary material for this article can be found in the online version.

REFERENCES

- Andrade R, Faria W M, Silva S H G, Chakraborty S, Weindorf D C, Mesquita L F, Guilherme L R G, Curi N. 2020a. Prediction of soil fertility via portable X-Ray fluorescence (pXRF) spectrometry and soil texture in the Brazilian Coastal Plains. *Geoderma*. **357**: 113960.
- Andrade R, Silva S H G, Weindorf D C, Chakraborty S, Faria W M, Mesquita L F, Guilherme L R G, Curi N. 2020b. Assessing models for prediction of some soil chemical properties from portable X-ray fluorescence (pXRF) spectrometry data in Brazilian Coastal Plains. *Geoderma*. **357**: 113957.
- Barnes R J, Dhanoa M S, Lister S J. 1989. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl Spectrosc*. **43**: 772–777.
- Ben-Dor E, Banin A. 1995. Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. *Soil Sci Soc Am J*. **59**: 364–372.
- Benedet L, Faria W M, Silva S H G, Mancini M, Guilherme L R G, Dematté J A M, Curi N. 2020. Soil subgroup prediction via portable X-Ray fluorescence and visible near-infrared spectroscopy. *Geoderma*. **365**: 114212.
- Cardelli V, Weindorf D C, Chakraborty S, Li B, De Feudis M, Cocco S, Agnelli A, Choudhury A, Ray D P, Corti G. 2017. Non-saturated soil organic horizon characterization via advanced proximal sensors. *Geoderma*. **288**: 130–142.
- Chang C W, Laird D A, Mausbach M J, Hurburgh C R. 2001. Near-infrared reflectance spectroscopy—principal components regression analyses of soil properties. *Soil Sci Soc Am J*. **65**: 480–490.
- Dos Santos W J R, Curi N, Silva S H G, Da Fonseca S, Da Silva E, Marques J J. 2014. Detailed soil survey of an experimental watershed representative of the Brazilian Coastal Plains and its practical application. *Ciência Agrotecnol*. **38**: 50–60.
- Duda B M, Weindorf D C, Chakraborty S, Li B, Man T, Paulette L, Deb S. 2017. Soil characterization across catenas via advanced proximal sensors. *Geoderma*. **298**: 78–91.
- Evans D M, Hartemink A E. 2014a. Digital soil mapping of a red clay subsoil covered by loess. *Geoderma*. **230-231**: 296–304.
- Evans D M, Hartemink A E. 2014b. Terra rossa catenas in Wisconsin, USA. *Catena*. **123**: 148–152.
- Gee G W, Bauder J W. 1979. Particle size analysis by hydrometer: A simplified method for routine textural analysis and a sensitivity test of measurement parameters. *Soil Sci Soc Am J*. **43**: 1004–1007.
- Gomez C, Viscarra R R A, McBratney A B. 2008. Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: An Australian case study. *Geoderma*. **146**: 403–411.
- Hartemink A E, Minasny B. 2014. Towards digital soil morphometrics. *Geoderma*. **230-231**: 305–317.
- Hu B F, Chen S C, Hu J, Xia F, Xu J F, Li Y, Shi Z. 2017. Application of portable XRF and VNIR sensors for rapid assessment of soil heavy metal pollution. *PLOS ONE*. **12**: e0172438.
- Husson F, Josse J, Le S, Mazet J. 2018. FactoMineR: Multivariate exploratory data analysis and data mining. R package version 1.41. Available online at <https://cloud.r-project.org/web/packages/facto/index.html> (verified on February 7, 2022).
- Kuhn M. 2008. Building predictive models in R using the caret package. *J Stat Softw*. **28**: 1–26.
- Mancini M, Weindorf D C, Chakraborty S, Silva S H G, Dos Santos Teixeira A F, Guilherme L R G, Curi N. 2019. Tracing tropical soil parent material analysis via portable X-ray fluorescence (pXRF) spectrometry in Brazilian Cerrado. *Geoderma*. **337**: 718–728.
- Minasny B, McBratney A B. 2008. Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. *Chemometr Intell Lab Syst*. **94**: 72–79.
- Ng W, Minasny B, Montazerolghaem M, Padarian J, Ferguson R, Bailey S, McBratney A B. 2019. Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra. *Geoderma*. **352**: 251–267.
- O'Rourke S M, Minasny B, Holden N M, McBratney A B. 2016a. Synergistic use of Vis-NIR, MIR, and XRF spectroscopy for the determination of soil geochemistry. *Soil Sci Soc Am J*. **80**: 888–899.
- O'Rourke S M, Stockmann U, Holden N M, McBratney A B, Minasny B. 2016b. 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.
- Quinlan J R. 1992. Learning with continuous classes. In Proceedings of 5th Australian Joint Conference on Artificial Intelligence. World Scientific, Hobart. pp. 343–348.
- R Core Team. 2016. R: A Language and Environment for Statistical Computing. R Core Team, Vienna.
- Rawal A, Chakraborty S, Li B, Lewis K, Godoy M, Paulette L, Weindorf D C. 2019. Determination of base saturation percentage in agricultural soils via portable X-ray fluorescence spectrometer. *Geoderma*. **338**: 375–382.
- Rossel R A V, Adamchuk V I, Sudduth K A, McKenzie N J, Lobsey C. 2011. Proximal soil sensing: An effective approach for soil measurements in space and time. *Adv Agron*. **113**: 243–291.
- Rossel R A V, Cattle S R, Ortega A, Fouad Y. 2009. In situ measurements of soil colour, mineral composition and clay content by vis-NIR spectroscopy. *Geoderma*. **150**: 253–266.
- Rossel R A V, McBratney A B, Minasny B. 2010. Proximal Soil Sensing. Springer, Dordrecht.
- Rossel R A V, Webster R. 2011. Discrimination of Australian soil horizons and classes from their visible-near infrared spectra. *Eur J Soil Sci*. **62**: 637–647.
- Rouillon M, Taylor M P. 2016. Can field portable X-ray fluorescence (pXRF) produce high quality data for application in environmental contamination research? *Environ Pollut*. **214**: 255–264.
- Santana M L T, Ribeiro B T, Silva S H G, Poggere G C, Guilherme L R G, Curi N. 2018. Conditions affecting oxide quantification in unknown tropical soils via handheld X-ray fluorescence spectrometer. *Soil Res*. **56**: 648–655.
- Savitzky A, Golay M J E. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem*. **36**: 1627–1639.
- Sharma A, Weindorf D C, Man T, Aldabaa A A A, Chakraborty S. 2014. Characterizing soils via portable X-ray fluorescence spectrometer: 3. Soil reaction (pH). *Geoderma*. **232-234**: 141–147.
- 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.
- Shi P, Castaldi F, Van Wesemael B, Van Oost K. 2020. Vis-NIR spectroscopic assessment of soil aggregate stability and aggregate size distribution in the Belgian loam belt. *Geoderma*. **357**: 113958.
- Silva S H G, Weindorf D C, Pinto L C, Faria W M, Junior F W A, Gomide L R, De Mello J M, De Pádua Junior A L, 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.
- Soil Survey Staff. 2014. Kellogg soil survey laboratory methods manual. In Burt R, Soil Survey Staff (eds.) Soil Survey Investigations Report No. 42. Version 5.0. U.S. Department of Agriculture, Natural Resources Conservation Service, Lincoln.
- Stevens A, Ramirez-Lopez L, Hans G. 2013. Prospectr: Miscellaneous functions for processing and sample selection of vis-NIR diffuse reflectance data. R package version 0.2.2. Available online at <https://cran.r-project.org/web/packages/prospectr/index.html> (verified on January 5, 2022).
- 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.
- Taylor P D, Ramsey M H, Potts P J. 2004. Balancing measurement uncertainty against financial benefits: Comparison of in situ and ex situ analysis of contaminated land. *Environ Sci Technol*. **38**: 6824–6831.
- Teixeira A F D S, Weindorf D C, Silva S H G, Guilherme L R G, Curi N. 2018. Portable X-ray fluorescence (pXRF) spectrometry applied to

- the prediction of chemical attributes in Inceptisols under different land uses. *Ciência Agrotecnol.* **42**: 501–512.
- Tóth T, Kovács Z A, Rékási M. 2019. XRF-measured rubidium concentration is the best predictor variable for estimating the soil clay content and salinity of semi-humid soils in two catenas. *Geoderma.* **342**: 106–108.
- Tümsava Z, Tekin Y, Ulusoy Y, Mouazen A M. 2019. Prediction and mapping of soil clay and sand contents using visible and near-infrared spectroscopy. *Biosyst Eng.* **177**: 90–100.
- Waiser T H, Morgan C L S, Brown D J, Hallmark C T. 2007. *In situ* characterization of soil clay content with visible near-infrared diffuse reflectance spectroscopy. *Soil Sci Soc Am J.* **71**: 389–396.
- Wan M X, Hu W Y, Qu M K, Li W D, Zhang C R, Kang J F, Hong Y S, Chen Y, Huang B. 2020. Rapid estimation of soil cation exchange capacity through sensor data fusion of portable XRF spectrometry and Vis-NIR spectroscopy. *Geoderma.* **363**: 114163.
- Wang D D, Chakraborty S, Weindorf D C, Li B, Sharma A, Paul S, Ali M N. 2015. Synthesized use of VisNIR DRS and PXRF for soil characterization: Total carbon and total nitrogen. *Geoderma.* **243-244**: 157–167.
- Weindorf D C, Bakr N, Zhu Y D. 2014. Advances in portable X-ray fluorescence (PXRF) for environmental, pedological, and agronomic applications. *Adv Agron.* **128**: 1–45.
- Weindorf D C, Zhu Y D, Haggard B, Lofton J, Chakraborty S, Bakr N, Zhang W, Weindorf W C, Legoria M. 2012. Enhanced pedon horizonation using portable X-ray fluorescence spectrometry. *Soil Sci Soc Am J.* **76**: 522–531.
- Xu D Y, Chen S C, Rossel R A V, Biswas A, Li S, Zhou Y, Shi Z. 2019. X-ray fluorescence and visible near infrared sensor fusion for predicting soil chromium content. *Geoderma.* **352**: 61–69.
- Yost J L, Hartemink A E. 2019. Effects of carbon on moisture storage in soils of the Wisconsin Central Sands, USA. *Eur J Soil Sci.* **70**: 565–577.
- Zhang Y K, Hartemink A E. 2019. Soil horizon delineation using vis-NIR and pXRF data. *Catena.* **180**: 298–308.
- Zhang Y K, Hartemink A E. 2020. Data fusion of Vis-NIR and pXRF spectra to predict soil physical and chemical properties. *Eur J Soil Sci.* **71**: 316–333.
- Zhu Y D, Weindorf D C, Zhang W T. 2011. Characterizing soils using a portable X-ray fluorescence spectrometer: 1. Soil texture. *Geoderma.* **167-168**: 167–177.
- Zornoza R, Guerrero C, Mataix-Solera J, Scow K M, Arcenegui V, Mataix-Beneyto J. 2008. Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. *Soil Biol Biochem.* **40**: 1923–1930.