Geoderma 243–244 (2015) 157–167

Contents lists available at ScienceDirect

Geoderma

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

# Synthesized use of VisNIR DRS and PXRF for soil characterization: Total carbon and total nitrogen ☆

Dandan Wang [a], Somsubhra Chakraborty [b], David C. Weindorf [c,*], Bin Li [d], Aakriti Sharma [c], Sathi Paul [b], Md. Nasim Ali [b]

[a] College of Geography and Remote Sensing, Nanjing University of Information Science and Technology, Nanjing, Jiangsu, China
[b] Ramakrishna Mission Vivekananda University, Kolkata, India
[c] Department of Plant and Soil Sciences, Texas Tech University, Lubbock, TX, USA
[d] Louisiana State University, Baton Rouge, LA, USA

ARTICLE INFO

*Article history:*
Received 9 October 2014
Received in revised form 8 December 2014
Accepted 14 December 2014
Available online 9 January 2015

*Keywords:*
Proximal sensing
Random forest
Penalized spline regression

ABSTRACT

Soil total carbon (TC) and total nitrogen (TN) both play critical roles in soil health and ecosystem dynamics. The former is involved in soil structural formation, atmospheric carbon sequestration, and improved soil tilth while the latter is a plant essential element which is often deficient in agronomic production systems. Traditionally, both TC and TN were limited to laboratory determination using techniques such as wet or dry combustion, ion sensing electrodes, loss on ignition, or via chemical assays. These techniques, while generally accurate, require extensive soil sampling, laboratory analysis, and are inherently destructive to the sample analyzed. An approach which could quantify both TC and TN in situ would result in considerable time and cost savings and provide the analyst with the ability to capture more data for a given field of interest. Portable x-ray fluorescence (PXRF) and visible near infrared (VisNIR) diffuse reflectance spectrometry were used to scan 675 soil samples in a laboratory with diverse physicochemical properties from three states of the USA, then compared via random forest (RF) regression and penalized spline regression (PSR) to TC and TN data obtained through traditional laboratory analysis (Dumas method high temperature combustion). Results clearly demonstrated that merging the PXRF and VisNIR datasets improved the power of predictive models by improving the residual prediction deviation (RPD) and $R^2$ statistics. Using synthesized (PXRF + VisNIR) models, independent validation data produced quality predictive statistics for soil TC (RPD = 2.90; $R^2$ = 0.88 via PSR) and TN (RPD = 2.99; $R^2$ = 0.89 via RF). Both proximal sensing techniques were also used to independently predict TC and TN, with results less robust than the synthesized approach. The general order of optimal prediction can be summarized as PXRF + VisNIR > VisNIR > PXRF. In conclusion, the use of synthesized proximal data from PXRF and VisNIR was shown to be a solid, stable predictor of soil TC and TN with widespread agronomic and environmental science applications.



## 1. Introduction

Both carbon and nitrogen are critical elements in soils. Soil total carbon (TC) can improve soil fertility, quality, and water retention, and ultimately maintain and increase crop production (Muñoz and Kravchenko, 2011). In addition, the soil carbon pool, as the largest reservoir in the terrestrial ecosystem, is 3.3 times the size of the atmospheric pool and 4.5 times the size of the biotic pool (Lal, 2004). Small changes in the soil carbon pool may influence global climate change. Soil TC loss due to cultivation degrades soil fertility and quality, reduces biomass productivity, and adversely impacts water quality; depletions exacerbated by projected global warming (Lal, 2004; McDowell et al., 2012). Soil total nitrogen (TN), a critical macronutrient for plant growth, is a major determinant and indicator of soil fertility and quality, and also the most commonly deficient soil nutrient (Reeves, 1997; Miller and Gardiner, 1998). However, excessive nitrogen contents in soil not only lead to non-point source pollution, such as eutrophication and associated water-quality problems (Carpenter et al., 1998; Zhao et al., 2011), but also can be released to the atmosphere as greenhouse gases (e.g., nitrous oxide, $N_2O$) (Ledley et al., 1999). Moreover, the C:N ratio is a good indicator of the degree of decomposition and quality of the organic matter held in the soil (Batjes, 1996). Soil TC is the driving force of biological activity, serving as the primary source of energy and nutrients for many soil organisms (Craswell and Lefroy, 2001), and an important factor affecting nitrogen mineralization and immobilization in soils (Compton and Boone, 2002; Hoylea et al., 2006; Lang et al., 2010). Soil

*Abbreviations:* PSR, Penalized spline regression; PXRF, Portable x-ray fluorescence spectrometry; RPD, Residual prediction deviation; RF, Random forest; TN, Total nitrogen; TC, Total carbon; VisNIR DRS, Visible near-infrared diffuse reflectance spectroscopy.

☆ Type of contribution: Conceived and designed the experiments: DW, SC, and DCW. Performed the experiments: DW, SC, and AS. Wrote the paper: DW, SC, DCW, BL, and AS.

* Corresponding author at: Texas Tech University, Department of Plant and Soil Sciences, Box 42122, Lubbock, TX 79409, USA.
*E-mail address:* david.weindorf@ttu.edu (D.C. Weindorf).

http://dx.doi.org/10.1016/j.geoderma.2014.12.011
0016-7061/

nitrogen, as a key nutrient, can directly influence carbon sequestration in terrestrial ecosystems (Oren et al., 2001). Therefore, spatial predictions of soil TC and TN contents are needed for a wide range of agricultural and environmental applications (Muñoz and Kravchenko, 2011; Wang et al., 2013a).

For decades, classical laboratory-based methods have been utilized for quantifying soil TC and TN content. Two basic approaches are used to quantify TC in soils, namely, dry combustion and wet combustion (Nelson and Sommers, 1996). Dry combustion requires separate determinations for inorganic- and organic-C, is time consuming, relatively expensive, and not adaptable to in situ determinations (Reeves et al., 2002). Wet combustion is a semi-quantitative estimate of soil carbon due to the lack of a universal conversion factor for each soil analyzed; it is time-consuming, tedious, and generates toxic waste that must be disposed of properly (Nelson and Sommers, 1996; Craswell and Lefroy, 2001). Another relatively inexpensive and rapid technique, loss-on-ignition (LOI) has been shown to be inaccurate in some instances due to the decomposition of certain mineral fractions at high temperatures (Nelson and Sommers, 1996; Lal et al., 2001). The Dumas (1831)(dry combustion) and Kjeldahl (1983)(wet oxidation) methods have gained general acceptance for determination of TN in the laboratory. However, both methods are time consuming, destructive to the sample being analyzed, and fail to recover some forms of N, particularly N in certain heterocyclic compounds and compounds containing N–N and N–O linkages (Bermner, 1996). The Dumas method is also expensive and has lower precision than Kjeldahl approaches (Bermner, 1996). Another method of quantification involves ion sensing electrodes, as a quick and reliable alternative to chemical-based laboratory methods for nitrate measurements. However, interference from other similar and undesired ions can be problematic; sometimes causing instability in attaining equilibrium (Kuang et al., 2012). Also, cell membranes, reference electrodes, and amplifier distortions may cause anomalous readings (Kuang et al., 2012). The disadvantages of all these traditional laboratory analysis methods are compounded by the large number of samples required for accurate assessment (McCarty and Reeves, 2006; McDowell et al., 2012). Although these traditional methods are relatively accurate and widely accepted, they require extensive lab work and destroy the sample during processing. Therefore, there is a growing demand for rapid, cost effective, and nondestructive approaches for predicting C and N in situ. Proximal soil sensing techniques have the potential to eliminate the aforementioned constraints.

One popular proximal soil sensing technique, visible near infrared (VisNIR) spectroscopy, is quick, cost-effective, non-destructive, requires little sample preparation with no hazardous chemicals used, and is highly adaptable to automated and in situ measurements (McCarty et al., 2002; Viscarra Rossel et al., 2006). Such approaches have attracted widespread interest in soil science since the 1980s (Stenberg et al., 2010). The same spectra from scanning a soil with VisNIR spectroscopy can be used for the prediction of a variety of soil properties simultaneously (Sarkhot et al., 2011), especially soil carbon. Many recent studies have been conducted on quantifying TC (Vasques et al., 2008; McDowell et al., 2012), soil organic carbon, inorganic carbon (Sankey et al., 2008; Morgan et al., 2009; Gogé et al., 2014) and other soil carbon fractions (Vasques et al., 2009; Sarkhot et al., 2011) using VisNIR spectroscopy in the laboratory, in situ, or using airborne imaging spectroscopy (Mouazen et al., 2007; Morgan et al., 2009; Stevens et al., 2010; Ge et al., 2014). Comparatively fewer studies have focused on estimating soil nitrogen (Selige et al., 2006; Brunet et al., 2008; An et al., 2014) through these approaches, let alone simultaneously with soil carbon. Although these studies obtained excellent results, which showed that it is a viable alternative for the routine quantitative analysis of soil carbon, lab-based VisNIR can only provide semi-quantitative estimation with residual prediction deviation (RPD) = 1.5–2.0 (Bricklemyer and Brown, 2010). Many factors, including moisture, particle size, mineral composition, and the presence of Fe, influence the reflectance of soils (Hunt, 1977; Lobell and Asner, 2002; Reeves et al., 2002) and soil VisNIR spectra are largely non-specific, quite weak, and broad due to overlapping absorptions of soil constituents (Rossel and Behrens, 2010). Morgan et al. (2009) found VisNIR spectra alone do not provide sufficient accuracy for stand-alone C sequestration measurement, monitoring and verification. VisNIR spectroscopy alone will never provide complete soil characterization, so application in parallel with other sensing technologies should be a focus of future research (Brown et al., 2006).

Another proximal soil sensing technique, x-ray fluorescence spectrometry has been used since the 1930s (Jones, 1982). Given technological advances in recent years, portable x-ray fluorescence (PXRF) spectrometry has been developed and improved greatly with a number of significant advantages including minimal sample preparation, high sample throughputs, and the rapid, nondestructive, accurate, low cost, and in situ identification of many elements (Ulmanu et al., 2011; Zhu et al., 2011). Therefore, PXRF has become increasingly popular and been adopted by environmental consultancies, research institutions, and governmental agencies such as the US Environmental Protection Agency via Method 6200 (USEPA, 2007), the International Organization for Standardization (ISO) (International Organisation for Standardisation, 2013), and the National Institute for Occupational Safety and Health (NIOSH) Method 7702 (NIOSH, 1998) for the analysis of soil and sediments. Portable x-ray fluorescence spectrometry can quantify elements from z = 15 (P) through 94 (Pu) and is useful for environmental monitoring of many elements in soils and other geological materials (Weindorf et al., 2013a; Wiedenbeck, 2013). Applied to soil science, many studies have focused on metal contamination assessment using PXRF (Weindorf et al., 2012a, 2013b; Parsons et al., 2013; Hu et al., 2014), and PXRF elemental data has been used as a proxy for a wide number of soil parameters such as pH (Sharma et al., 2014), cation exchange capacity (Sharma et al., 2015), soil calcium and gypsum (Weindorf et al., 2009, 2013a; Zhu and Weindorf, 2009), soil texture (Zhu et al., 2011), soil salinity (Swanhart et al., in press), and soil horizon differentiation (Weindorf et al., 2012b, 2012c). However, PXRF cannot presently be used to quantify lighter elements (Olympus, 2013) (e.g., Na, N, C, H, Li) given their stable electron configurations and low fluorescent energies. To our knowledge, few studies have attempted to predict soil carbon and nitrogen by using PXRF, given that indirect approaches to these determinations are required as direct measurements are not possible. However, Weindorf et al. (2012c) tried to link organic carbon content with PXRF elemental concentration while differentiating spodic horizons, and found some associations of soil carbon and nitrogen with PXRF elemental data.

Many studies have reported successful prediction of soil properties using a single instrument (e.g., VisNIR, PXRF, and so on). However, single sensors provide no robust capability to measure soil properties successfully at different sampling sites because of the complex nature of soils (Wang et al., 2013b). Therefore, the integrated use of proximal soil sensor technologies such as PXRF, multispectral satellite imagery, and VisNIR may enhance predictive models. For example, Wang et al. (2013b) predicted soil texture using Fourier transform near-infrared (NIR) spectroscopy and PXRF spectrometry with data fusion and concluded soil textural fractions predicted with sample data and sensor data fusion methods (e.g., clay, validation $R^2$ = 0.83–0.86, RPD = 1.94–2.39) were more accurate than those with individual sensors and individual data sets (e.g., clay, validation $R^2$ = 0.61–0.74, RPD = 1.56–2.36). Also, Aldabaa et al. (2015) evaluated surface soil salinity using data from multispectral satellite imagery, PXRF, and VisNIR in playas of Texas, and showed that the synthesis of all three datasets produced the best predictive results; better than any one technique taken independently. The synthesis of PXRF and VisNIR datasets have even been filed for provisional patent in the USA with numerous agronomic, environmental, and pedological applications. Given the success of these previous investigations, the application of synthesized spectroscopy models for TC and TN quantification seems timely. Therefore, the objectives of this paper were to: 1) evaluate soil TC and TN predictability

using proximal soil sensing methods (VisNIR and PXRF) independently and, 2) investigate soil TC and TN predictability by combining VisNIR and PXRF approaches. In this study, we hypothesize that soil TC and TN can be predicted from the direct elemental readings of PXRF and VisNIR spectra. Furthermore, predictive models from synthesized proximal soil sensing techniques will provide the best predictive ability; considerably better than each approach individually.

## 2. Materials and methods

### 2.1. Site selection and field sampling

A total of 675 soil samples were collected from Seward County, Nebraska, Kern County, California, and Lubbock County, Texas, USA during 2014. Specifically, in each state, 75 sampling points were randomly selected within a single agricultural field and collected at three depths (0–15, 15–30, and 30–45 cm). The fields in Nebraska, California, and Texas supported corn, wheat, and cotton, respectively. At each site, soils were sampled with a sharpshooter or auger, placed in sealed plastic bags, and shipped to the Texas Tech University Pedology Laboratory for analysis. Generally, the Nebraska soils were from the Hastings (Fine, smectitic, mesic Udic Argiustolls), Fillmore (Fine, smectitic, mesic Vertic Argialbolls), and Butler (Fine, smectitic, mesic Vertic Argiaquolls) soil series in major land resource area (MLRA) 75-Central Loess Plains (Soil Survey Staff, 2014a). In California, soils were from the Wasco (Coarse-loamy, mixed, superactive, nonacid, thermic Typic Torriorthents) series in MLRA 17-Sacramento and San Joaquin Valleys (Soil Survey Staff, 2014a). In Texas, soils were from the Amarillo (Fine-loamy, mixed, superactive, thermic Aridic Paleustalfs), Acuff (Fine-loamy, mixed, superactive, thermic Aridic Paleustolls), and Estacado (Fine-loamy, mixed, superactive, thermic Aridic Paleustolls) soil series in MLRA 77C-Southern High Plains, Southern Part (Soil Survey Staff, 2014a). In selecting the soil sampling sites for this project, we purposefully chose soils from disparate locations such that considerable variation would be represented in soil physicochemical properties. Importantly, samples at each depth at each site were composites of two subsamples, which were not necessarily vertically associated at the same collection point. Furthermore, all fields sampled were regularly plowed by farm machinery, causing wholesale mixing of the upper 45 cm of soil.

### 2.2. Physicochemical lab analysis

Upon arrival in the laboratory, all samples were air dried, and ground to pass a 2-mm sieve prior to general laboratory analysis. Soil samples for high temperature combustion analysis were further sieved to pass a 0.25 mm sieve prior to analysis. TC and TN values were determined for soil samples via Dumas method combustion using a Truspec® CHN analyzer (LECO Corp., St. Joseph, MI, USA) per Soil Survey Staff (2014b) and Fultz et al. (2013).

### 2.3. VisNIR scanning and spectral pretreatments

The spectra of the 675 soil samples were obtained using a PSR-3500® portable VisNIR spectroradiometer (Spectral Evolution, Lawrence, MA, USA) with a spectral range of 350 to 2500 nm. The spectroradiometer had a 2-nm sampling interval and a spectral resolution of 3.5, 10, and 7-nm from 350 to 1000 nm, 1500 nm and 2100 nm, respectively. Scanning was facilitated with a contact probe featuring a 5 W built-in light source. Full contact with the sample was ensured to avoid outside interference. Each sample was scanned four times with a 90° rotation between scans to obtain an average spectral curve. Each individual scan was an average of 10 internal scans over a time of 1.5 s. The detector was white referenced (after four samples) using a 12.7 cm × 12.7 cm NIST traceable radiance calibration panel, ensuring that fluctuating downwelling irradiance could not saturate the detector.

Raw reflectance spectra were processed via a statistical analysis software package, R version 2.11.0 (R Development Core Team, 2008) using custom "R" routines (Chakraborty et al., 2013, 2014b). These routines involved (i) a parabolic splice to correct for "gaps" between detectors, (ii) averaging replicate spectra, (iii) fitting a weighted (inverse measurement variance) smoothing spline to each spectra with direct extraction of smoothed reflectance at 10 nm intervals. This study used one spectral pretreatment to prepare the smoothed soil spectra for analysis, and two multivariate algorithms to develop the TC and TN predictive models. Spectral pretreatments helped in reducing the influence of the side information contained in the spectra. The pretreatment transformation applied was Savitzky–Golay (SG) first derivative using a first-order polynomial across a ten band window. The transformation was implemented in the Unscrambler®X 10.3 software (CAMO Software Inc., Woodbridge, NJ).

### 2.4. PXRF scanning

All samples were also scanned using a DP-6000 Delta Premium PXRF (Olympus, Waltham, MA, USA). The instrument features a Rh x-ray tube operated at 15–40 keV with quantification via ultra-high resolution (<165 eV) silicon drift detector. Prior to soils analysis, the instrument was calibrated using a 316 alloy clip, containing 16.130% Cr, 1.780% Mn, 68.760% Fe, 10.420% Ni, 0.200% Cu, and 2.100% Mo, tightly fitted over the aperture. The instrument was operated in "Soil Mode" capable of detecting the following suite of elements: Sr, Zr, Mo, Ag, Cd, Sn, Sb, Ti, Ba, Cr, Mn, Fe, Co, Ni, Cu, Zn, Hg, As, Se, Pb, Rb, P, S, Cl, K, Ca, and V. Soil Mode consists of three beams which operate sequentially. Each beam was set to scan for 30 s such that one whole scan was completed in 90 s. As the PXRF analysis window is small (approximately 2 $cm^2$), it is important that the average composition of the sample is well represented within this area. Soil samples were carefully mixed to ensure sample homogeneity before scanning. Also, the aperture of the instrument was kept clean by air blowing to prevent soil or dust from contaminating the aperture window after each scan. Elemental data was stored in the on-board computer, and then downloaded into MS Excel for analysis.

### 2.5. Machine learning

The two multivariate methods tested included random forest (RF) regression and penalized spline regression (PSR). Both VisNIR-SG spectra and PXRF elements were combined in a single table and the whole

**Table 1**
Descriptive statistics of PXRF elements, total N and total C used in the multivariate models.

| Statistic | Total C | Total N | K | Ca | Ti | V | Mn | Fe | Zn | Rb | Sr | Zr | Pb |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | % | | mg $kg^{-1}$ | | | | | | | | | | |
| Minimum | 0.067 | 0.006 | 6738.0 | 1729.0 | 871.0 | 18.0 | 56.0 | 4993.0 | 16.1 | 26.1 | 36.7 | 168.0 | 4.8 |
| Maximum | 6.160 | 0.330 | 21,001.0 | 350,673.0 | 4379.0 | 108.0 | 880.0 | 29,694.0 | 120.0 | 111.7 | 339.0 | 683.0 | 111.0 |
| 1st Quartile | 0.321 | 0.038 | 12,529.0 | 4623.5 | 2112.0 | 41.0 | 201.0 | 13,324.5 | 36.6 | 63.0 | 100.0 | 313.5 | 11.3 |
| Median | 0.523 | 0.060 | 17,277.0 | 5721.0 | 3085.0 | 64.0 | 384.0 | 18,576.0 | 51.0 | 77.5 | 166.0 | 429.0 | 13.5 |
| 3rd Quartile | 1.355 | 0.135 | 18,466.0 | 15,521.0 | 3295.0 | 78.0 | 438.0 | 23,292.0 | 63.2 | 94.1 | 286.0 | 526.0 | 20.9 |
| Mean | 0.832 | 0.084 | 15,801.2 | 11,018.7 | 2764.9 | 60.6 | 338.6 | 18,145.4 | 49.7 | 76.9 | 183.5 | 420.6 | 15.3 |
| Standard deviation (n − 1) | 0.712 | 0.062 | 3469.8 | 20,032.0 | 714.4 | 21.3 | 135.2 | 6041.2 | 16.8 | 19.9 | 87.0 | 122.6 | 6.2 |

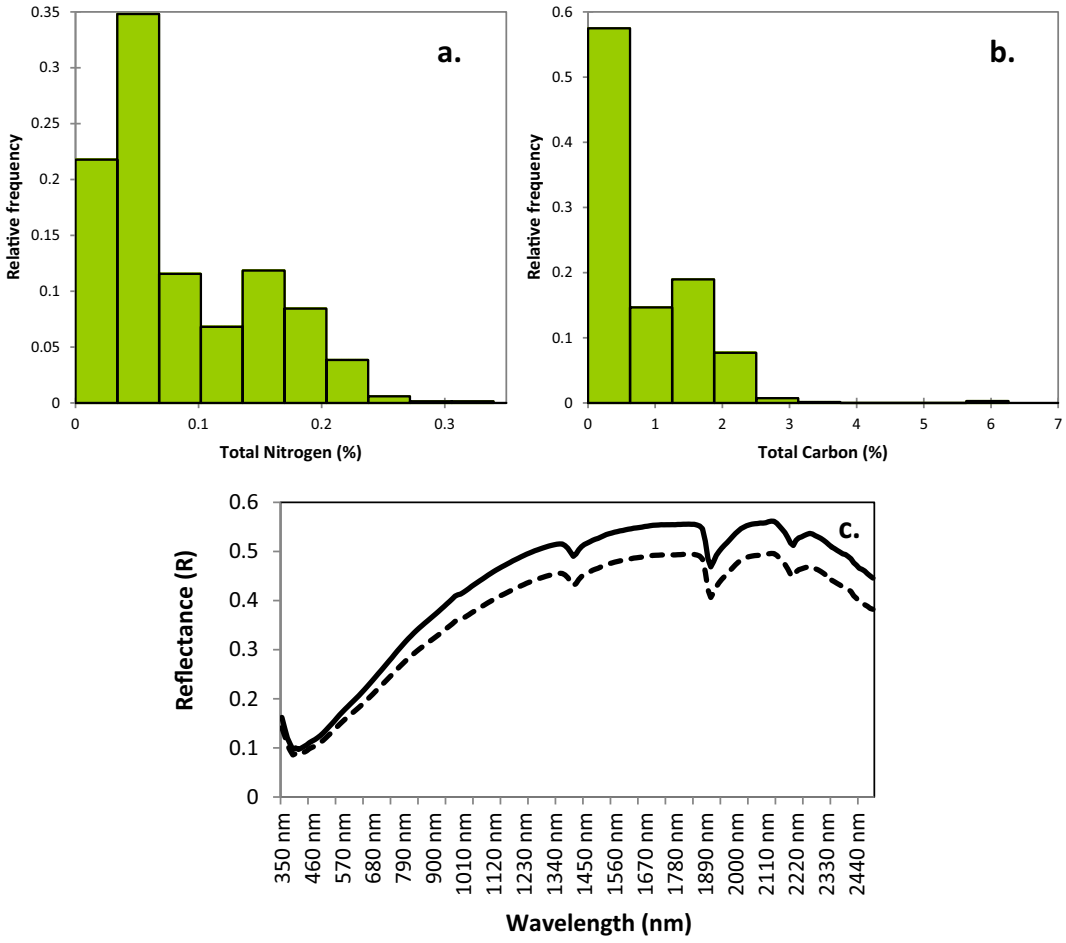


**Fig. 1.** Plots showing histograms of a) soil total N, b) total C, and c) average reflectance spectra of two randomly selected soil samples. Both soils exhibited two distinct absorption peaks around 1900 (water absorption bands) and 2200 nm (metal–hydroxyl stretching).

dataset was randomly divided twice into i) a ~70% training set (n = 472) for calibration and a ~30% independent test set (n = 203) and ii) for a more honest evaluation of the prediction performance, a ~70% training set (n = 472) for calibration and a ~30% independent test set (n = 203) for validation keeping the three depths of a given profile together. Note that, no pre- or post-processing was applied on PXRF data.

Classical least squares modeling approaches usually fail on high-dimensional multivariate calibration problems because the size of the regressors is larger than the sample size. Unlike most other approaches, a PSR model makes use of the ordered structure among the regressors (Eilers and Marx, 1996). The objective function on which the penalized spline minimizes is a tradeoff between the goodness-of-fit on the data through sum squared error (SSE) and the penalty function on the smoothness of the estimated regression coefficients (i.e. constraining the difference between the neighboring regression coefficients) (Marx and Eilers, 1999). Because of the additional smoothness penalty, penalized spline is well-suited for ill-posed problems (the dimensionality is much larger than the sample size) such as signal regression problems. Since PSR is a straightforward extension of the linear regression model, it inherits all the beneficial properties from linear regression,

**Table 2**
Correlation matrix (p-values) of PXRF elements, total N and total C used in the multivariate models.

| Variables | Total N | K | Ca | Ti | V | Mn | Fe | Zn | Rb | Sr | Zr | Pb | Total C |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Total N | **0** | | | | | | | | | | | | |
| K | <**0.0001**[a] | **0** | | | | | | | | | | | |
| Ca | <**0.0001** | 0.774 | **0** | | | | | | | | | | |
| Ti | <**0.0001** | <**0.0001** | **0.032** | **0** | | | | | | | | | |
| V | <**0.0001** | <**0.0001** | **0.007** | <**0.0001** | **0** | | | | | | | | |
| Mn | <**0.0001** | <**0.0001** | 0.132 | <**0.0001** | <**0.0001** | **0** | | | | | | | |
| Fe | 0.193 | <**0.0001** | 0.656 | <**0.0001** | <**0.0001** | <**0.0001** | **0** | | | | | | |
| Zn | <**0.0001** | <**0.0001** | **0.022** | <**0.0001** | <**0.0001** | <**0.0001** | <**0.0001** | **0** | | | | | |
| Rb | <**0.0001** | <**0.0001** | <**0.0001** | <**0.0001** | <**0.0001** | <**0.0001** | <**0.0001** | <**0.0001** | **0** | | | | |
| Sr | <**0.0001** | <**0.0001** | <**0.0001** | <**0.0001** | <**0.0001** | <**0.0001** | <**0.0001** | <**0.0001** | <**0.0001** | **0** | | | |
| Zr | <**0.0001** | **0.029** | <**0.0001** | 0.505 | <**0.0001** | 0.066 | <**0.0001** | <**0.0001** | <**0.0001** | <**0.0001** | **0** | | |
| Pb | <**0.0001** | <**0.0001** | <**0.0001** | <**0.0001** | <**0.0001** | <**0.0001** | <**0.0001** | <**0.0001** | <**0.0001** | 0.240 | <**0.0001** | **0** | |
| Total C | <**0.0001** | **0.003** | <**0.0001** | **0.000** | <**0.0001** | 0.059 | **0.001** | <**0.0001** | <**0.0001** | <**0.0001** | <**0.0001** | <**0.0001** | **0** |

[a] Values in bold are different from 0 with a significance level alpha = 0.05.

**Table 3**
Validation statistics of multivariate models using 30% validation set (n = 203).

| Property (%) | Approach | Model | $R^2$ | RMSE (%) | RPD[a] | Bias (%) |
|---|---|---|---|---|---|---|
| *Random splitting scheme* | | | | | | |
| Total C | VisNIR + PXRF | RF[b] | 0.83 | 0.319 | 2.42 | −0.039 |
| | | PSR[c] | 0.93 | 0.209 | 3.69 | −0.026 |
| Total N | VisNIR + PXRF | RF | 0.91 | 0.019 | 3.39 | −0.001 |
| | | PSR | 0.91 | 0.019 | 3.35 | −0.002 |
| Total C | VisNIR | RF | 0.81 | 0.331 | 2.33 | −0.043 |
| | | PSR | 0.90 | 0.233 | 3.31 | −0.025 |
| Total N | VisNIR | RF | 0.90 | 0.019 | 3.23 | −0.0009 |
| | | PSR | 0.89 | 0.020 | 3.14 | −0.0015 |
| Total C | PXRF | RF | 0.77 | 0.366 | 2.11 | −0.051 |
| Total N | PXRF | RF | 0.90 | 0.020 | 3.20 | −0.002 |
| *Without separating soil profile* | | | | | | |
| Total C | VisNIR + PXRF | RF | 0.86 | 0.216 | 2.69 | 0.044 |
| | | PSR | 0.88 | 0.201 | 2.90 | 0.039 |
| Total N | VisNIR + PXRF | RF | 0.89 | 0.018 | 2.99 | 0.005 |
| | | PSR | 0.88 | 0.018 | 2.93 | 0.003 |
| Total C | VisNIR | RF | 0.84 | 0.232 | 2.50 | 0.040 |
| | | PSR | 0.87 | 0.202 | 2.87 | 0.041 |
| Total N | VisNIR | RF | 0.85 | 0.020 | 2.65 | 0.005 |
| | | PSR | 0.88 | 0.019 | 2.89 | 0.003 |
| Total C | PXRF | RF | 0.85 | 0.222 | 2.61 | 0.110 |
| Total N | PXRF | RF | 0.85 | 0.020 | 2.66 | 0.010 |

[a] RPD, Residual prediction deviation.
[b] RF, Random forest regression.
[c] PSR, Penalized spline regression.

such as the confidence interval of estimated regression coefficients. The computation of the PSR model, including cross-validation, is relatively less penalizing. For leave-one-out cross-validation (LOOCV), a penalized spline model can directly output the validation error without recomputation of the model omitting each sample. In this study, the cubic B-spline was used via R version 2.14.1 (R Development Core Team, 2008) as the base function with 100 equally spaced knots. The order of the penalty was set to the default value of three. The optimal value for the penalty-tuning parameter was selected by minimizing the LOOCV error on the training set.

Random forest (Breiman, 2001) is an ensemble learning method that combines hundreds of individual trees. To build each tree, first a bootstrap sample is drawn from the training samples; then a tree is built using the bootstrap sample of the data, and at each node split the candidate set of the regressor is a random subset (the size of the subset is denoted as *mtry*) of all the regressors. The final prediction of a new observation is the average of the predicted values (for the new observation) from all the trees in the forest. Studies have shown that the prediction accuracy of the random forest ensemble is usually better than the one from an individual tree (Breiman, 2001). In this study, the 'random forest' package was used in R to build the random forest model. The number of trees in random forest was set to the default value of 500. The coefficient of determination ($R^2$), RMSE, residual prediction deviation (RPD) (Williams, 1987), and bias were used as rubrics for judging model generalizing capability. Since RPD is the ratio of standard deviation and RMSE, model predictability is enhanced when the validation set standard deviation (SD) is comparatively larger than the estimation error (RMSE). Chang et al. (2001) categorized the accuracy and stability of their spectroscopy models based on the RPD values of the validation set. RPDs > 2.0 were considered stable and accurate predictive models; RPD values between 1.4 and 2.0 indicated fair models that could be improved by more accurate predictive techniques; RPD values <1.4 indicated poor predictive capacity.

In this study, five modeling approaches were employed as follows:

1. At first, both TC and TN were predicted using simply concatenating PXRF elements with VisNIR spectra via PSR and RF.
2. Next, both TC and TN were predicted using VisNIR spectra alone via PSR and RF.
3. A RF model was used to predict TC and TN using only PXRF elements.
4. As a conservative estimate of model performance, entire profiles were selected for calibration and validation to maintain independence between calibration and validation data (Brown et al., 2005). Then, RF was applied to PXRF data only, PSR and RF were applied to VisNIR data only, and PSR and RF were applied to fused PXRF + VisNIR datasets. Scans for individual soil profiles were not split between calibration and validation datasets.

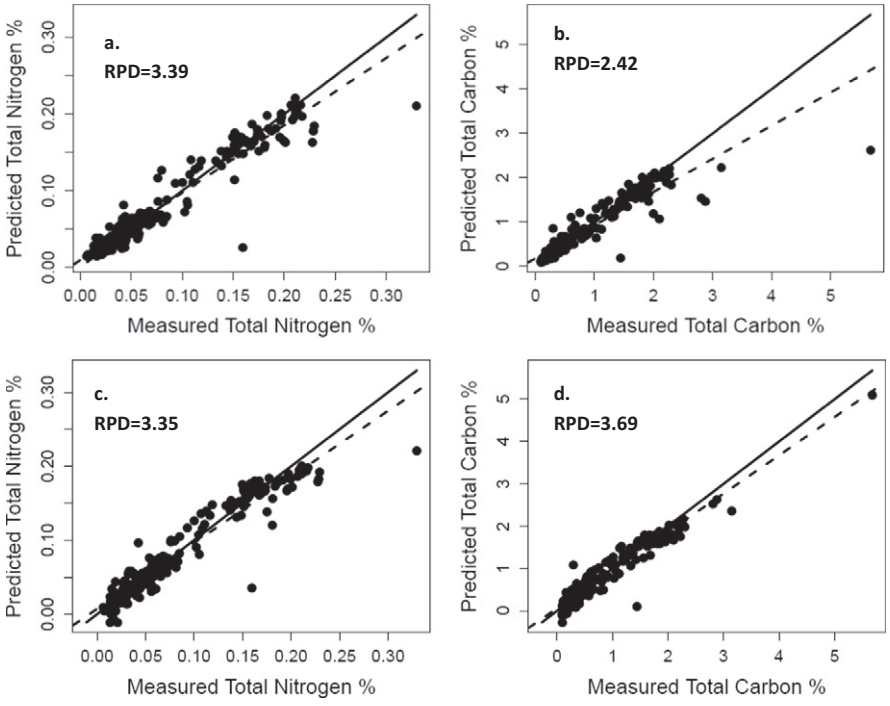


**Fig. 2.** Plots showing a) random forest model for total N, b) random forest model for total C, c) penalized spline model for total N, and d) penalized spline model for total C using randomly selected VisNIR spectra and PXRF elements.

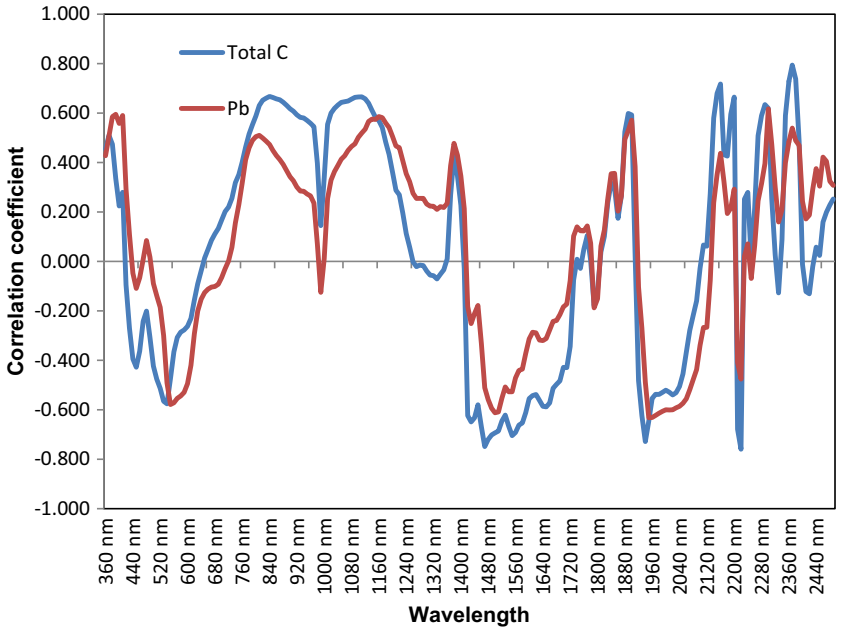


**Fig. 3.** Correlation patterns between reflectance of VisNIR reflectance spectra and Pb contents of soil and total C.

5. Finally, whole-field holdout validation was done for both TC and TN to determine how field to field heterogeneity affected prediction accuracies (Brown et al., 2005). Whole-field holdouts were achieved by calibrating a model using RF with two of the three fields. The third field was held out as the validation sample. A total of 12 models were created to represent all three fields as a validation set using a) PXRF elements + VisNIR spectra, and b) PXRF elements only.

## 3. Results and discussion

### 3.1. Soil properties and reflectance spectra

Importantly, not all elements were quantified with discreet values via PXRF. Thus, only 10 elements (K, Ca, Ti, V, Mn, Fe, Zn, Rb, Sr, Zr, and Pb) were viable for regression, with discreet values across all soil samples scanned. Soil samples varied widely in their estimated

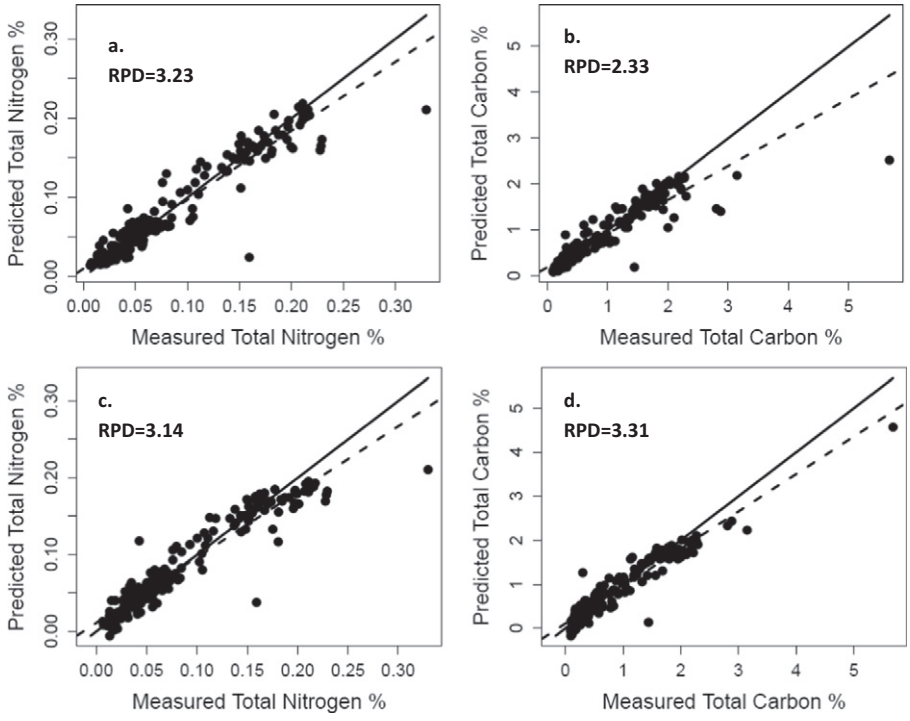


**Fig. 4.** Plots showing a) random forest model for total N, b) random forest model for total C, c) penalized spline model for total N, and d) penalized spline model for total C using randomly selected VisNIR spectra only.

properties, likely a result of differences in land use, sampling depth, vegetation cover, and geologic origin. Statistical moments related with soil TN, TC, and PXRF sensed elements are explained as follows: compared to other properties, PXRF-sensed K (~3-fold), Rb (~4-fold), Zr (~4-fold), Fe (5-fold), Ti (5-fold), V (6-fold), Zn (7-fold) and Sr (9-fold) exhibited relatively lower ranges of variation (Table 1). The ranges of variation of other properties were markedly larger [15-(Mn) to 91-fold (TC)]. PXRF-sensed total Ca produced the highest range of variation (1729-fold). Both TN and TC were non-normally distributed (Shapiro–Wilk p-values < 0.0001) from 0.006 to 0.33% and 0.06 to 6.16%, respectively (Fig. 1a and b). Most of the variables were significantly correlated between themselves with some exceptions (Table 2). Notably, TC was significantly correlated with TN ($\rho = 0.87$). The C:N ratios for California (max 16.1, min 4.0, $\bar{x}$ 8.5), Texas (max 97.9, min 5.9, $\bar{x}$ 10.5), and Nebraska (max 15.2, min 4.8, $\bar{x}$ 9.4) were all similar. Neither N nor C is measured by PXRF; both elements are too light. But previous research (Sharma et al., 2014) has shown that other elements can and do have significant associations with light elements. Thus, while the PXRF does not directly measure C or N, the inclusion of other elements quantified by PXRF strengthens the predictive algorithms as auxiliary input data.

Average reflectance spectra of two randomly selected soil samples are shown in Fig. 1c. In general, reflectance spectra for both soils were similar regarding high optical density [log (1/R)] in the visible light region (350–750 nm) and two distinct absorption peaks around 1900 (water absorption bands) and 2200 nm (metal–hydroxyl stretching), as reported in the literature (Clark, 1999).

### 3.2. Multivariate modeling

Initially, among the two multivariate algorithms tested (RF and PSR) by concatenating both VisNIR spectral data and PXRF elements irrespective of keeping the same profile intact, TC was estimated with greater accuracy by PSR while TN was estimated slightly better by RF (Table 3). Although both RF and PSR produced an identical $R^2$ (0.91) for TN, the former exhibited a slightly higher RPD value of 3.39. Conversely, the PSR model for TC increased both validation $R^2$ (0.93) and RPD (3.69) relative to its RF counterpart. In general, both RF and PSR models for TN exhibited slight underestimation at higher values (Fig. 2). Some of these underestimations could be due to the relative scarcity of observations at the higher ends of the property scales, as previously outlined by Brown et al. (2006). On the contrary, while measured vs. PSR predicted TC closely approximated the 1:1 line with slight underestimation at higher values (Fig. 2d), exhibiting a greater deviation from the 1:1 line the prediction worsened further for the RF model (Fig. 2b). Given that the consistency of a NIR spectroscopic model is normally restricted to the range of parameter values, the wide ranges of soil TC (0.067–6.16%) and TN (0.006–0.33%) perhaps contributed to the overall satisfactory results (Table 1). Since regression coefficients of TN were rather similar to those of TC, it can be concluded that VisNIR detected a combination of soil constituents containing organic functional groups, which contain organic forms of N. Although intense bands in the VisNIR spectra are not directly associated with the presence of metals or other constituents, it is already established that metals can interact with the main spectrally active components of soil like soil C (Song et al., 2012). To further validate this hypothesis, we randomly selected PXRF sensed Pb which exhibited both positive and negative correlations with VisNIR reflectance spectra of soil (Fig. 3). Further, Pb showed strong negative correlations at ~520 nm, ~1480 nm, ~1890 nm, and ~2200 nm; where there was also a strong negative relationship with the TC content of the soils (Fig. 3). Thus, the above correlations gave an indication of the good relationship of the Pb (and perhaps for other PXRF elements) with TC. This explanation was supported by the significant correlations of the PXRF sensed elements (except Mn) with TC which represents soil organic matter

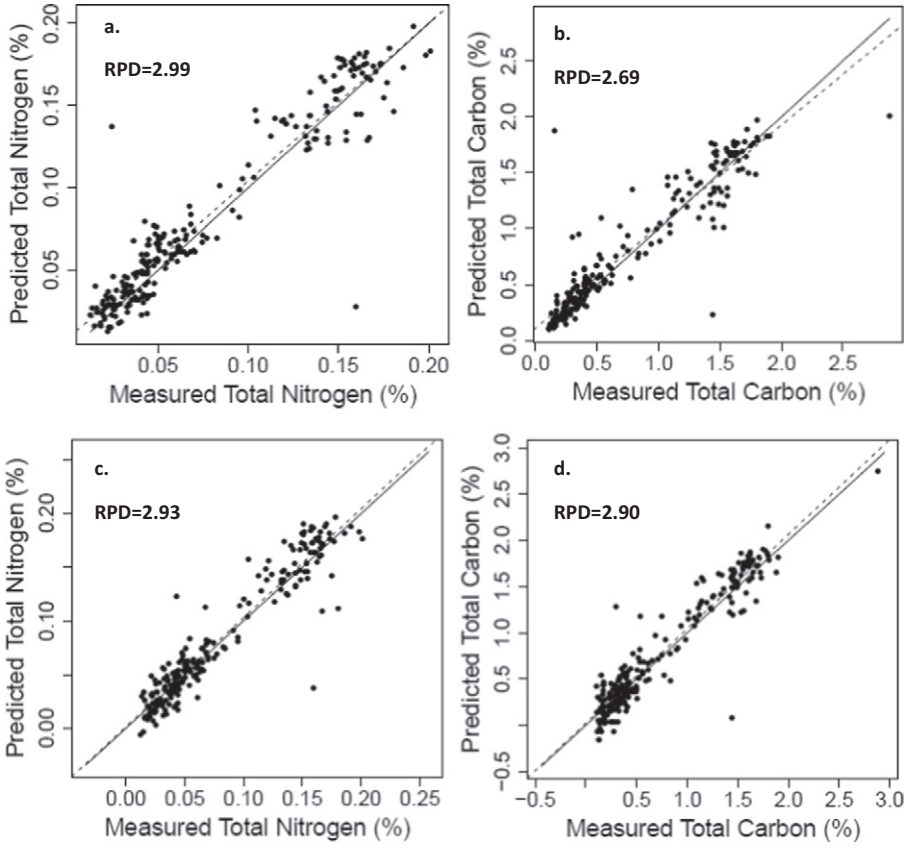


**Fig. 5.** Plots showing a) random forest model for total N, b) random forest model for total C, c) penalized spline model for total N, and d) penalized spline model for total C using whole soil profile VisNIR spectra and PXRF elements.

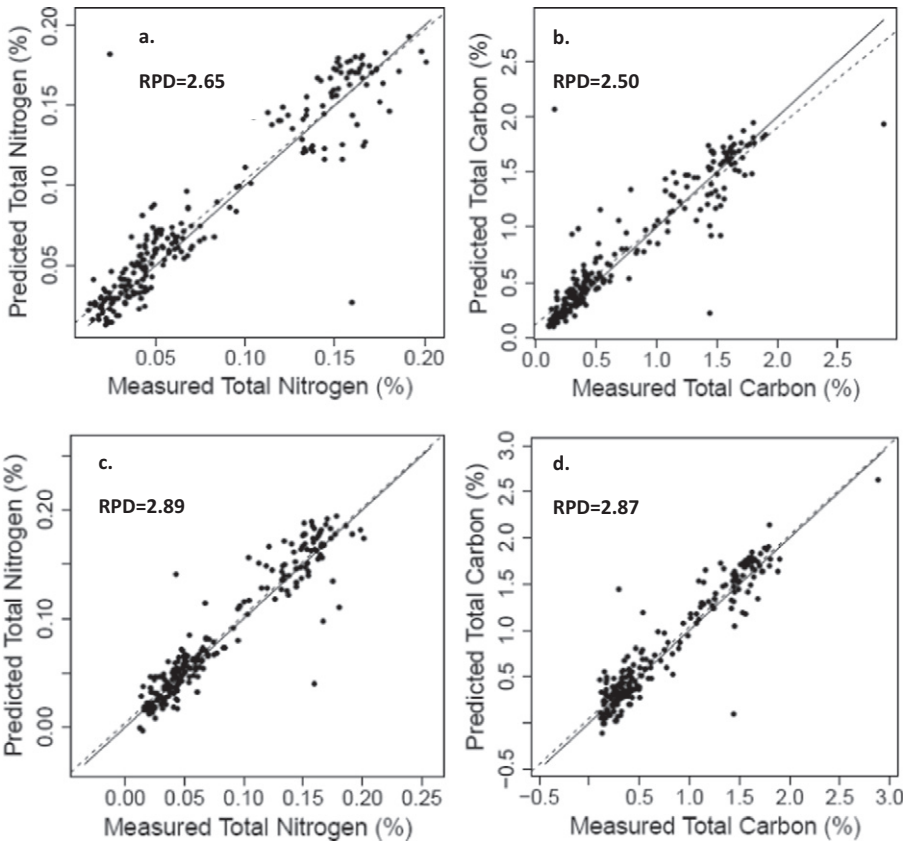


**Fig. 6.** Plots showing a) random forest model for total N, b) random forest model for total C, c) penalized spline model for total N, and d) penalized spline model for total C using whole soil profile VisNIR spectra only.

(Table 2). Association with organic matter may be a main binding form of PXRF sensed elements in soils. These findings further rationalized the use of PXRF elements in predicting TC and TN of soil samples. These results also provide support for the synergistic use of diffuse reflectance spectra and PXRF elements in predicting TC contents of soil samples.

When both TC and TN were modeled with PSR and RF using only VisNIR data irrespective of keeping the same profile intact, a similar trend was observed. Indeed, while both PSR and RF models for TN exhibited almost the same generalization capacity, the PSR-TC model again showed improved predictability ($R^2 = 0.90$, RPD = 3.31) as compared to the RF-TC ($R^2 = 0.81$, RPD = 2.33) model (Table 3). Following the same trends of VisNIR + PXRF models for TN, both PSR-TN and RF-TN models using VisNIR spectra showed close approximation to the 1:1 line with subtle underestimation at higher values (Fig. 4). Furthermore, the prediction for TC deviated more from the 1:1 line in the RF model (Fig. 4b) than in the PSR model (Fig. 4d), corroborating the lesser predictability of the former. One important observation was that concatenation of VisNIR and PXRF data produced better predictability than using VisNIR spectra alone for modeling both TC and TN.

Additionally, to test the hypothesis that predictive models from synthesized proximal soil sensing techniques provide the best predictive ability irrespective of keeping the same profile intact, noticeably better than each approach individually, we modeled TC and TN with PXRF elements only. While predicting with PXRF elements via RF algorithm, a substantial reduction in model generalization was obtained in the case of TC ($R^2 = 0.77$, RPD = 2.11) (Table 3). Moreover, while considering RF-TN with PXRF elements only, a steady decrease of RPD values was obtained from its VisNIR + PXRF (RPD = 3.39) and VisNIR only (RPD = 3.23) counterparts.

While keeping the same profile intact for calibration and validation, similar trends were observed, although with slightly lowered model accuracy than those obtained when samples from the same profile were mixed between calibration and validation sets. Models obtained by combining both VisNIR and PXRF outperformed their VisNIR only and PXRF only counterparts (Table 3). While TC was estimated with noticeably greater accuracy by PSR (RPD = 2.90) than RF, TN was estimated slightly better by RF (RPD = 2.99) than its PSR counterpart (RPD = 2.93). In general, while RF models for TC exhibited slight underestimation at higher values, PSR closely approximated the 1:1 line (Figs. 5 and 6). There were only a few predictions falling beyond ±10% of the reference data. Summarily, the results reported a remarkable accuracy suggesting that combination of VisNIR and PXRF could be convenient for estimating several soil properties with advanced empirical calibrations.

Nevertheless, holding out whole fields reduced prediction accuracies in all cases where the soils of the field held out were not represented by another field, creating extrapolation (Table 4). Addition of PXRF data did not produce any visible improvement. Validation with an individual

**Table 4**
Prediction accuracies of TC and TN using whole field holdout validation. Models were created using random forest regression with two fields and then validated using the third field. Twelve separate models were calibrated and validated.

| Validation field | Parameter | RMSE-test (%) | | RPD[a] | | Bias (%) | |
|---|---|---|---|---|---|---|---|
| | | VisNIR | VisNIR + PXRF | VisNIR | VisNIR + PXRF | VisNIR | VisNIR + PXRF |
| Nebraska | TC | 0.82 | 0.80 | 0.53 | 0.55 | −0.64 | −0.64 |
| | TN | 0.07 | 0.06 | 0.54 | 0.62 | −0.06 | −0.05 |
| Lubbock | TC | 0.82 | 0.80 | 0.76 | 0.78 | 0.48 | 0.44 |
| | TN | 0.06 | 0.06 | 0.32 | 0.32 | 0.05 | 0.05 |
| California | TC | 1.96 | 2.00 | 0.07 | 0.07 | 1.90 | 1.96 |
| | TN | 0.12 | 0.11 | 0.13 | 0.13 | 0.11 | 0.10 |

[a] RPD, Residual prediction deviation.

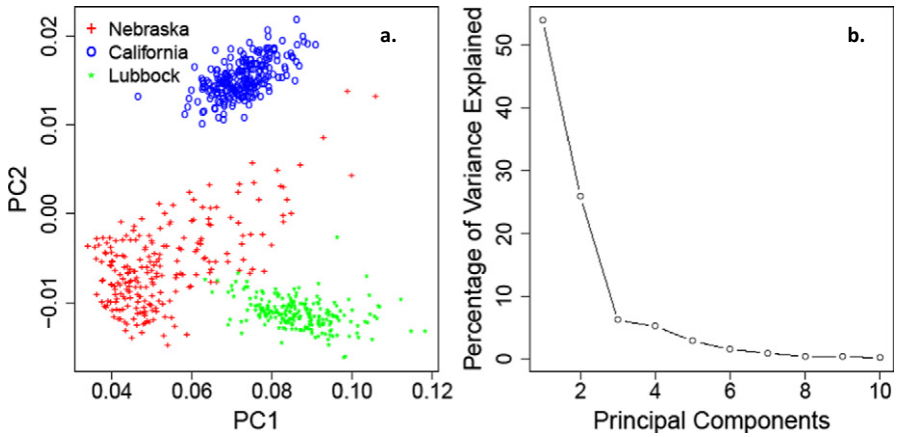


**Fig. 7.** Plots showing a) principal component plot for PC1 vs. PC2 of the first-derivative of VisNIR reflectance spectra. The red crosses, green dots, and blue circles represent soils from Nebraska, Lubbock, and California, respectively. Plot b) represents the "Screeplot" of the first ten principal components of the first derivative spectra.

field exhibited significant increases in RMSEs and a reduction in RPDs relative to the 30% random validation set. As expected, when principal component analysis was implemented on the VisNIR spectrum only, it clearly demonstrated different spectral behaviors of the three locations (Fig. 7). For example, the soil textures of Nebraska (silty clay loam, silty clay), California (sandy loam), and Texas (sandy loam, sandy clay loam) were largely different (Fig. 7).

### 3.3. Prospects of VisNIR + PXRF for rapidly quantifying soil parameters

Summarily, both PXRF and VisNIR DRS showed considerable promise in providing rapid TC and TN prediction in soils with reasonable accuracy, but the most pronounced one was VisNIR DRS. Adding PXRF elemental data as predictors along with soil VisNIR spectra was beneficial and should be included to model soil TC and TN. One possible reason for the improved estimation of TC and TN by the PXRF + DRS approach can be the covariation of non-chromophores or spectrally inactive components (PXRF elements except Fe and Mn) with relevant chromophores or spectrally active components (especially organic carbon).

The better performance of PSR can be attributed to its stability and flexibility; more so than other parametric approaches like partial least squares regression and principal component regression since the shape of the functional relationship between covariates and the dependent variable (soil TC and TN, in this study) was managed by the data (Eilers and Marx, 1996; Marx and Eilers, 1999). Non-parametric PSR which has become a very powerful and applicable smoothing technique over the last 10–15 years can be considered as a simplification of smoothing splines with a more flexible choice of bases and penalties (Krivobokova, 2006). Given that, P-spline smoothing tied with ridge regression, mixed models, and Bayesian statistics, allows more sound handling of signal regression problems such as VisNIR DRS. The improvement of model generalization capability by implementing PSR followed the same trend reported elsewhere for other soil and compost properties (Chakraborty et al., 2012, 2014a). However, in this study PSR could not be applied on only PXRF elements since PSR is only applicable for signal regression. In other words, since PSR assumes smoothness of the coefficient for each channel, the X should be a signal (reflectance).

Acquisition of PXRF data is rapid, easy, and cost-effective. Both evaluated techniques share the advantage of determining TC and TN especially for unusual circumstances where non-destructive sampling is required. Evaluated independently, the techniques were generally ranked as VisNIR DRS > PXRF. Conversely, the synthesis of both datasets produced the best predictive results; better than any one technique taken independently. Our results strongly support a previous study where synthesizing PXRF data with VisNIR spectra boosted the prediction of soil EC (Aldabaa et al., 2015).

In this study, one obvious question was: is it possible to directly analyze soil TC and TN in fresh samples by VisNIR DRS + PXRF? This question was critical as we attempted to develop models which could ultimately lead to in situ measurements in future. Given that VisNIR DRS can sense the changes in the matrix material scanned, particularly moisture (as it relates to O–H bonding and color) (Bishop et al., 1994; Zhu et al., 2010), bringing the soil samples to standard water content prior to scanning is critical for obtaining consistent results. In the present study, soil samples were air-dried. Despite that, it cannot be excluded that even in low soil moisture conditions, there was still remaining water adsorbed on the surface areas of clay minerals (e.g., hygroscopic water) and organic matter in equilibrium with atmospheric water vapor. Interestingly, even pre-treatment methods like quick-freezing and freeze-drying were unable to remove the water completely from the layer minerals of soil (Terhoeven-Urselmans et al., 2008). However, these minor variations perhaps did not cause much difference in the NIR spectra, as previously identified by Minasny et al. (2011). Maintaining homogeneous water content in the field during in situ scanning is not easy. Establishing the best sample pretreatment (field-moist or air-dried) was beyond the scope of this study and requires further investigations. Hence, the question of "how best to develop calibrations when moisture is present" is one of the big issues that remains to be properly answered about calibrations for soil TC and TN with VisNIR DRS and PXRF.

## 4. Conclusions

In conclusion, 675 soil samples representing disparate physicochemical conditions from three different states in the USA were evaluated for TC and TN using synthesized PXRF and VisNIR spectroscopy datasets, and traditional laboratory analysis (Dumas method high temperature combustion). Specifically, random forest (RF) regression and penalized spline regression (PSR) were used as multivariate tools for relating the two datasets in terms of $R^2$, RMSE, and RPD. Results were compelling, with validation sub-datasets producing optimized total carbon predictions via PSR (RPD = 3.69; $R^2$ = 0.93) and total nitrogen predictions via RF (RPD = 3.39; $R^2$ = 0.91). Even while keeping the same profile intact for calibration and validation, independent validation data for synthesized (PXRF + VisNIR) models produced quality predictive statistics for soil TC (RPD = 2.90; $R^2$ = 0.88 via PSR) and TN (RPD = 2.99; $R^2$ = 0.89 via RF) and outperformed other models tested. Predictions were also made with both PXRF and VisNIR data independently; both were less robust than the synthesized dataset approach utilizing both in tandem for predictions. The general predictive ability can be summarized as follows: PXRF + VisNIR > VisNIR > PXRF. The synthesized dataset

approach is a revolutionary step forward in soil chemical analysis as proximally captured data can effectively predict both TC and TN contents in soils quickly, non-destructively, and with little need for traditional laboratory analysis. Furthermore, as has been shown in previous studies (e.g., Aldabaa et al., 2015), the data from such proximal sensing (both spectral and elemental) can be used for the simultaneous prediction of any number of different soil properties, by simply applying the appropriate modeling equation. Such an approach has widespread applicability in both agronomic (e.g., plant essential element) and environmental (e.g., carbon stock assessment, carbon sequestration) applications. Future studies should evaluate the synthesized approach under even more diverse soil conditions, and in situ, where soil moisture, organic matter, and other ancillary factors can be considered while developing an optimized predictive equation for TC and TN.

## Acknowledgments

The authors are grateful for financial support from the BL Allen Endowment in Pedology at Texas Tech University.

## References

Aldabaa, A.A.A., Weindorf, D.C., Chakraborty, S., Sharma, A., 2015. Combination of proximal and remote sensing methods for rapid soil salinity quantification. Geoderma 239–240, 34–46. http://dx.doi.org/10.1016/j.geoderma.2014.09.011.

An, X., Li, M., Zheng, L., Liu, Y., Sun, H., 2014. A portable soil nitrogen detector based on NIRS. Precis. Agric. 15, 3–16.

Batjes, N.H., 1996. Total carbon and nitrogen in the soils of the world. Eur. J. Soil Sci. 47, 151–163.

Bermner, J.M., 1996. Nitrogen-Total. In: Sparks, D.L. (Ed.), Methods of Soil Analysis — Part 3: Chemical Methods. SSSA, Madison, WI, pp. 1085–1122.

Bishop, J.L., Pieters, C.M., Edwards, J.O., 1994. Infrared spectroscopic analyses on the nature of water in montmorillonite. Clay Clay Miner. 42, 702–716.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Bricklemyer, R.S., Brown, D.J., 2010. On-the-go VisNIR: potential and limitations for mapping soil clay and organic carbon. Comput. Electron. Agric. 70, 209–216.

Brown, D.J., Bricklemeyer, R.S., Miller, P.R., 2005. Validation requirements for diffuse reflectance soil characterization models with a case study of VNIR soil C prediction in Montana. Geoderma 129, 251–267.

Brown, D.J., Shepherd, K.D., Walsh, M.G., Dewayne Mays, M., Reinsch, T.G., 2006. Global soil characterization with VNIR diffuse reflectance spectroscopy. Geoderma 132, 273–290.

Brunet, D., Bernoux, M., Barthès, B.G., 2008. Comparison between predictions of C and N contents in tropical soils using a Vis–NIR spectrometer including a fibre-optic probe versus a NIR spectrometer including a sample transport module. Biosyst. Eng. 100, 448–452.

Carpenter, S.R., Caraco, N.F., Corell, D.L., Howarth, R.W., Sharpley, A.N., Smith, V.H., 1998. Nonpoint pollution of surface waters with phosphorus and nitrogen. Ecol. Appl. 8, 559–568.

Chakraborty, S., Weindorf, D.C., Zhu, Y., Li, B., Morgan, C.L.S., Ge, Y., Galbraith, J., 2012. Spectral reflectance variability from soil physicochemical properties in oil contaminated soils. Geoderma 177–178, 80–89.

Chakraborty, S., Weindorf, D.C., Ali, N., Li, B., Ge, Y., Darilek, J.L., 2013. Spectral data mining for rapid measurement of organic matter in unsieved moist compost. Appl. Opt. 52, B82–B92.

Chakraborty, S., Das, B.S., Ali, N., Li, B., Sarathjith, M.C., Majumdar, K., Ray, D.P., 2014a. Rapid estimation of compost enzymatic activity by spectral analysis method combined with machine learning. Waste Manag. 34, 623–631.

Chakraborty, S., Weindorf, D.C., Li, B., Ali, N., Majumdar, K., Ray, D.P., 2014b. Analysis of petroleum contaminated soils by spectral modeling and pure response profile recovery of *n*-hexane. Environ. Pollut. 190, 10–18.

Chang, C., Laird, D.A., Mausbach, M.J., Hurburgh, C.R., 2001. Near infrared reflectance spectroscopy: principal components regression analysis of soil properties. Soil Sci. Soc. Am. J. 65, 480–490.

Clark, R.N., 1999. Spectroscopy of rocks and minerals, and principles of spectroscopy. In: Rencz, N. (Ed.), Remote Sensing for the Earth Sciences: Manual of Remote Sensing. John Wiley & Sons, New York, pp. 3–52.

Compton, J.E., Boone, R.D., 2002. Soil nitrogen transformations and the role of light fraction organic matter in forest soils. Soil Biol. Biochem. 34, 933–943.

Craswell, E.T., Lefroy, R.D.B., 2001. The role and function of organic matter in tropical soils. Nutr. Cycl. Agroecosyst. 61, 7–18.

Dumas, J.B.A., 1831. Procedes De l'analyse organique. Ann. Chim. Phys. 247, 198–213.

Eilers, P.H.C., Marx, B.D., 1996. Flexible smoothing with B-spline and penalties (with comments and rejoinder). Stat. Sci. 11, 89–121.

Fultz, L.M., Moore-Kucera, J., Zobeck, T.M., Acosta-Martínez, V., Wester, D.B., Allen, V.G., 2013. Organic carbon dynamics and soil stability in five semiarid agroecosystems. Agric. Ecosyst. Environ. 181, 231–240.

Ge, Y., Morgan, C.L.S., Ackerson, J.P., 2014. VisNIR spectra of dried ground soils predict properties of soils scanned moist and intact. Geoderma 221–222, 61–69.

Gogé, F., Gomez, C., Jolivet, C., Joffre, R., 2014. Which strategy is best to predict soil properties of a local site from a national Vis–NIR database? Geoderma 213, 1–9.

Hoylea, F.C., Murphya, D.V., Fillery, I.R.P., 2006. Temperature and stubble management influence microbial $CO_2$–C evolution and gross N transformation rates. Soil Biol. Biochem. 38, 71–80.

Hu, W., Huang, B., Weindorf, D.C., Chen, Y., 2014. Metals analysis of agricultural soils via portable X-ray fluorescence spectrometry. Bull. Environ. Contam. Toxicol. 92, 420–426.

Hunt, G.R., 1977. Spectral signatures of particulate minerals in the visible and near infrared. Geophysics 42, 501–513.

ISO (International Organisation for Standardisation), 2013. ISO 13196: 2013(E). Soil Quality — Screening Soils for Selected Elements by Energy-Dispersive X-Ray Fluorescence Spectrometry Using a Handheld or Portable Instrument. BSI Standards Publication, London, UK.

Jones, A.A., 1982. X-ray fluorescence spectrometry. In: Page, A.L., Miller, R.H., Keeney, D.R. (Eds.), Methods of soil analysis — Part 2: Chemical and microbiological properties, 2nd Ed. ASA-SSSA, Madison, WI, pp. 85–118.

Kjeldahl, J., 1983. Neue methode zur Bestmmung des Stickstoffs in organischen Korpern. Z. Anal. Chem. 22, 366–382.

Krivobokova, T., 2006. Theoretical and Practical Aspects of Penalized Spline Smoothing. (Ph.D. dissertation). Bielefeld University, Bielefeld, Germany.

Kuang, B., Mahmood, H.S., Quraishi, M.Z., Hoogmoed, W.B., Mouazen, A.M., van Henten, E.J., 2012. Chapter four — sensing soil properties in the laboratory, in situ, and on-line: a review. In: Donald, L.S. (Ed.), Advances in Agronomy. Academic Press, pp. 155–223.

Lal, R., 2004. Soil carbon sequestration impacts on global climate change and food security. Science 304, 1623–1627.

Lal, R., Kimble, J.M., Follett, R.F., Stewart, B.S., 2001. Methods of Assessment of Soil Carbon. CRC Press, Boca Raton, FL.

Lang, M., Cai, Z., Mary, B., Hao, X., Chang, S.X., 2010. Land-use type and temperature affect gross nitrogen transformation rates in Chinese and Canadian soils. Plant Soil 334, 377–389.

Ledley, T.S., Sundquist, E.T., Schwartz, S.E., Hall, D.K., Fellows, J.D., Killeen, T.L., 1999. Climate change and greenhouse gases. EOS Trans. Am. Geophys. Union 80, 453–474.

Lobell, D.B., Asner, G.P., 2002. Moisture effects on soil reflectance. Soil Sci. Soc. Am. J. 66, 722–727.

Marx, B.D., Eilers, P.H.C., 1999. Generalized linear regression on sampled signals and curves: a P-spline approach. Technometrics 41, 1–13.

McCarty, G.W., Reeves, J.B., 2006. Comparison of IR and MIR diffuse reflectance spectroscopy for field-scale measurement of soil fertility parameters. Soil Sci. 171, 94–102.

McCarty, G.W., Reeves III, J.B., Reeves, V.B., Follett, R.F., Kimble, J.M., 2002. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. Soil Sci. Soc. Am. J. 66, 640–646.

McDowell, M.L., Bruland, G.L., Deenik, J.L., Grunwald, S., Knox, N.M., 2012. Soil total carbon analysis in Hawaiian soils with visible, near-infrared and mid-infrared diffuse reflectance spectroscopy. Geoderma 189–190, 312–320.

Miller, R.W., Gardiner, D.T., 1998. Soils in Our Environment. Prentice Hall, Upper Saddle River, NJ.

Minasny, B., McBratney, A., Bellon-maurel, V., Roger, J.M., Gobrecht, A., Ferrand, L., Joalland, S., 2011. Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon. Geoderma 167–168, 118–124.

Morgan, C.L.S., Waiser, T.H., Brown, D.J., Hallmark, C.T., 2009. Simulated in situ characterization of soil organic and inorganic carbon with visible near-infrared diffuse reflectance spectroscopy. Geoderma 151, 249–256.

Mouazen, A.M., Maleki, M.R., De Baerdemaeker, J., Ramon, H., 2007. On-line measurement of some selected soil properties using a VIS–NIR sensor. Soil Tillage Res. 93, 13–27.

Muñoz, J.D., Kravchenko, A., 2011. Soil carbon mapping using on-the-go near infrared spectroscopy, topography and aerial photographs. Geoderma 166, 102–110.

Nelson, D.W., Sommers, L.E., 1996. Total carbon, organic carbon, and organic matter. In: Sparks, D.L. (Ed.), Methods of soil analysis — Part 3: Chemical methods. SSSA, Madison, WI, pp. 961–1010.

NIOSH, 1998. NIOSH method 7702: Lead by field portable XRF. Available online at, http://www.caslab.com/Test-Method-7702/ (verified 7 Oct. 2014).

Olympus, 2013. Periodic table of detectable elements. Available online at, http://www.olympus-ims.com/en/innovx-xrf-xrd/pdf-brochures/ (verified 7 Oct. 2014).

Oren, R., Ellsworth, D.S., Johnsen, K.H., Phillips, N., Ewers, B.E., Maier, C., Schäfer, K.V.R., McCarthy, H., Hendrey, G., McNulty, S.G., Katul, G.G., 2001. Soil fertility limits carbon sequestration by forest ecosystems in a $CO_2$ enriched atmosphere. Nature 411, 469–472.

Parsons, C., Margui Grabulosa, E., Pili, E., Floor, G.H., Roman-Ross, G., Charlet, L., 2013. Quantification of trace arsenic in soils by field-portable X-ray fluorescence spectrometry: considerations for sample preparation and measurement conditions. J. Hazard Mater. 262, 1213–1222.

R Development Core Team, 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (Available online at http://www.cran.r-project.org. (verified 18 Jul. 2014)).

Reeves, D.W., 1997. The role of soil organic matter in maintaining soil quality in continuous cropping systems. Soil Tillage Res. 43, 131–167.

Reeves III, J., McCarty, G., Mimmo, T., 2002. The potential of diffuse reflectance spectroscopy for the determination of carbon inventories in soils. Environ. Pollut. 116 (Supplement 1), S277–S284.

Rossel, R.A.V., Behrens, T., 2010. Using data mining to model and interpret soil diffuse reflectance spectra. Geoderma 158, 46–54.

Sankey, J.B., Brown, D.J., Bernard, M.L., Lawrence, R.L., 2008. Comparing local vs. global visible and near-infrared (VisNIR) diffuse reflectance spectroscopy (DRS) calibrations for the prediction of soil clay, organic C and inorganic C. Geoderma 148, 149–158.

Sarkhot, D.V., Grunwald, S., Ge, Y., Morgan, C.L.S., 2011. Comparison and detection of total and available soil carbon fractions using visible/near infrared diffuse reflectance spectroscopy. Geoderma 164, 22–32.

Selige, T., Böhner, J., Schmidhalter, U., 2006. High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures. Geoderma 136, 235–244.
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., 2015. Characterization of soils via portable x-ray fluorescence spectrometer: 4. Cation exchange capacity (CEC). Geoderma http://dx.doi.org/10.1016/j. geoderma .2014.10.001.
Soil Survey Staff, 2014a. Official soil series descriptions. Available online at, http://www.nrcs.usda.gov/wps/portal/nrcs/detailfull/soils/home/?cid=nrcs142p2_053587 (verified 19 Sept. 2014).
Soil Survey Staff, 2014b. Kellogg soil survey laboratory methods manual. Soil survey investigations report No. 42, Version 5.0. USDA-NRCS, Lincoln, NE.
Song, Y., Li, F., Yang, Z., Ayoko, G.A., Frost, R.L., Ji, J., 2012. Diffuse reflectance spectroscopy for monitoring potentially toxic elements in the agricultural soils of Changjiang River Delta, China. Appl. Clay Sci. 64, 75–83.
Stenberg, B., Viscarra Rossel, R.A., Mouazen, A.M., Wetterlind, J., 2010. Chapter five — visible and near infrared spectroscopy in soil science. In: Sparks, D.L. (Ed.), Advances in Agronomy. Academic Press, pp. 163–215.
Stevens, A., Udelhoven, T., Denis, A., Tychon, B., Lioy, R., Hoffmann, L., van Wesemael, B., 2010. Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy. Geoderma 158, 32–45.
Swanhart, S., Weindorf, D.C., Chakraborty, S., Bakr, N., Zhu, Y., Nelson, C., Shook, K., Acree, A., 2014. Measuring soil salinity via portable x-ray fluorescence spectrometry. Soil Sci. (in press).
Terhoeven-Urselmans, T., Schmidt, H., Georg Joergensen, R., Ludwig, B., 2008. Usefulness of near-infrared spectroscopy to determine biological and chemical soil properties: importance of sample pre-treatment. Soil Biol. Biochem. 40 (5), 1178–1188.
Ulmanu, M., Anger, I., Gament, E., Mihalache, M., Plopeanu, G., Ilie, L., 2011. Rapid determination of some heavy metals in soil using an x-ray fluorescence portable instrument. Res. J. Agric. Sci. 43, 235–241.
USEPA, 2007. Method 6200: Field Portable X-ray fluorescence Spectrometry for the Determination of Elemental Concentrations in Soil and Sediment. Available online at, http://www.epa.gov/osw/hazard/testmethods/sw846/pdfs/6200.pdf (verified 7 Oct. 2014).
Vasques, G.M., Grunwald, S., Sickman, J.O., 2008. Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. Geoderma 146, 14–25.
Vasques, G.M., Grunwald, S., Sickman, J.O., 2009. Modeling of soil organic carbon fractions using visible-near-infrared spectroscopy. Soil Sci. Soc. Am. J. 73, 176–184.
Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131, 59–75.
Wang, K., Zhang, C., Li, W., 2013a. Predictive mapping of soil total nitrogen at a regional scale: a comparison between geographically weighted regression and cokriging. Appl. Geogr. 42, 73–85.
Wang, S., Li, W., Li, J., Liu, X., 2013b. Prediction of soil texture using FT-NIR spectroscopy and PXRF spectrometry with data fusion. Soil Sci. 178, 626–638.
Weindorf, D.C., Zhu, Y., Ferrell, R., Rolong, N., Barnett, T., Allen, B.L., Herrero, J., Hudnall, W., 2009. Evaluation of portable X-ray fluorescence for gypsum quantification in soils. Soil Sci. 174, 556–562.
Weindorf, D.C., Zhu, Y., Chakraborty, S., Bakr, N., Huang, B., 2012a. Use of portable X-ray fluorescence spectrometry for environmental quality assessment of peri-urban agriculture. Environ. Monit. Assess. 184, 217–227.
Weindorf, D.C., Zhu, Y., Haggard, B., Lofton, J., Chakraborty, S., Bakr, N., Zhang, W., Weindorf, W.C., Legoria, M., 2012b. Enhanced pedon horizonation using portable X-ray fluorescence spectrometry. Soil Sci. Soc. Am. J. 76, 522–531.
Weindorf, D.C., Zhu, Y., McDaniel, P., Valerio, M., Lynn, L., Michaelson, G., Clark, M., Ping, C.L., 2012c. Characterizing soils via portable x-ray fluorescence spectrometer: 2. Spodic and albic horizons. Geoderma 189–190, 268–277.
Weindorf, D.C., Herrero, J., Castañeda, C., Bakr, N., Swanhart, S., 2013a. Direct soil gypsum quantification via portable X-ray fluorescence spectrometry. Soil Sci. Soc. Am. J. 77, 2071–2077.
Weindorf, D.C., Paulette, L., Man, T., 2013b. In-situ assessment of metal contamination via portable X-ray fluorescence spectroscopy: Zlatna, Romania. Environ. Pollut. 182, 92–100.
Wiedenbeck, M., 2013. Field-portable XRF: a geochemist's dream? Elements 9, 7–8.
Williams, P.C., 1987. Variables affecting near-infrared reflectance spectroscopic analysis. In: Williams, P., Norris, K. (Eds.), Near-infrared technology in the agricultural and food industries. Am. Assoc. of Cereal Chemists, St. Paul, MN, pp. 143–167.
Zhao, Y., Xia, X.H., Yang, Z.F., Xia, N., 2011. Temporal and spatial variations of nutrients in Baiyangdian Lake, North China. J. Environ. Inf. 17, 102–108.
Zhu, Y., Weindorf, D.C., 2009. Determination of soil calcium using field portable x-ray fluorescence. Soil Sci. 174, 151–155.
Zhu, Y., Weindorf, D.C., Chakraborty, S., Haggard, B., Johnson, S., Bakr, N., 2010. Characterizing surface soil water with field portable diffuse reflectance spectroscopy. J. Hydrol. 391, 133–140.
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.