Science of the Total Environment 514 (2015) 399–408

Contents lists available at ScienceDirect

Science of the Total Environment

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

Development of a hybrid proximal sensing method for rapid identification of petroleum contaminated soils ☆

Somsubhra Chakraborty [a], David C. Weindorf [b,*], Bin Li [c], Abdalsamad Abdalsatar Ali Aldabaa [d], Rakesh Kumar Ghosh [e], Sathi Paul [a], Md. Nasim Ali [a]

[a] Ramakrishna Mission Vivekananda University, Kolkata, India
[b] Department of Plant and Soil Science, Texas Tech University, Lubbock, TX, USA
[c] Department of Experimental Statistics, Louisiana State University, Baton Rouge, LA, USA
[d] Desert Research Center, Cairo, Egypt
[e] National Institute of Research on Jute and Allied Fibre Technology, Kolkata, India

HIGHLIGHTS

- PXRF elemental data and VisNIR DRS spectra were correlated to soil TPH.
- VisNIR + PSR was used to predict TPH followed by PXRF + RF to predict the residuals.
- The fused model produced better results than other multivariate models tested.

ARTICLE INFO

Article history:
Received 20 November 2014
Received in revised form 20 January 2015
Accepted 26 January 2015
Available online xxxx

Editor: Kevin V. Thomas

Keywords:
Penalized spline model
Portable X-ray fluorescence spectrometry
Soil petroleum contamination
Random forest
Visible near-infrared diffuse reflectance spectroscopy

ABSTRACT

Using 108 petroleum contaminated soil samples, this pilot study proposed a new analytical approach of combining visible near-infrared diffuse reflectance spectroscopy (VisNIR DRS) and portable X-ray fluorescence spectrometry (PXRF) for rapid and improved quantification of soil petroleum contamination. Results indicated that an advanced fused model where VisNIR DRS spectra-based penalized spline regression (PSR) was used to predict total petroleum hydrocarbon followed by PXRF elemental data-based random forest regression was used to model the PSR residuals, it outperformed (R^2 = 0.78, residual prediction deviation (RPD) = 2.19) all other models tested, even producing better generalization than using VisNIR DRS alone (RPD's of 1.64, 1.86, and 1.96 for random forest, penalized spline regression, and partial least squares regression, respectively). Additionally, unsupervised principal component analysis using the PXRF + VisNIR DRS system qualitatively separated contaminated soils from control samples.

Capsule: Fusion of PXRF elemental data and VisNIR derivative spectra produced an optimized model for total petroleum hydrocarbon quantification in soils.

1. Introduction

Soil petroleum contamination is a serious environmental concern because of its neurotoxic effects on humans and animals (Brevik and Burgess, 2013). The growth of the petroleum industry worldwide and marketing of petroleum products have resulted in countless chances for spillage. Most commonly, when an underground storage tank is removed, soil petroleum contamination is discovered which may pose an even more worrisome problem, groundwater contamination. Moreover, environmental pollution resulting from crude oil drilling has put numerous food crops under considerable risk (Oyedeji et al., 2012). On April 20, 2010, the largest accidental marine oil spill in the history of the petroleum industry occurred following a sea-floor oil spill gusher from the Deepwater Horizon drilling rig explosion in the Gulf of Mexico south of Louisiana, USA. To date, the total costs associated with lost jobs, contaminated food and water, cleanup, restoration, and environmental damage have not been fully determined (Camilli et al., 2010) and may not be for many years. Initial estimates placed the cost of damages to the Oil Company, environment, and US Gulf Coast economy at $36.9 billion (Smith et al., 2011), but later estimates by the Oil Company were

Abbreviations: PLS, partial least squares regression; PSR, penalized spline regression; PXRF, portable X-ray fluorescence; RF, random forest regression; RPD, residual prediction deviation; TPH, total petroleum hydrocarbon; VisNIR DRS, visible near infrared diffuse reflectance spectroscopy.

☆ Type of contribution: Conceived and designed the experiments: SC and DCW. Performed the experiments: SC, DCW, BL, and AAAA. Wrote the paper: SC, DCW, BL, RKG, SP, and MNA.

* Corresponding author at: 2011 Fulbright Scholar, Texas Tech University, Department of Plant and Soil Science, Box 42122, Lubbock, TX 79409, USA.
E-mail address: david.weindorf@ttu.edu (D.C. Weindorf).

http://dx.doi.org/10.1016/j.scitotenv.2015.01.087
0048-9697/

closer to $41.0 billion. Undoubtedly, rapid and cost-effective means of identifying total petroleum hydrocarbon (TPH) content in contaminated soils could substantially reduce the cost involved in their restoration.

Rapid and wide scale characterization of soil petroleum contamination is not feasible with traditional gas chromatography based methods since these are prohibitively expensive, extremely laborious, time consuming, sometimes show high variability (an order of magnitude) in TPH results across commercial laboratories, lack field-portability, and warrant rigorous field sampling (Dent and Young, 1981; Malley et al., 1999).

Remote sensing tools appear to be a viable technology to provide a comprehensive solution to this problem (Vasques et al., 2010). Specifically, a mounting body of literature underlines the usefulness of visible and near-infrared (VisNIR) diffuse reflectance spectroscopy (DRS) (350–2500 nm) as a rapid and noninvasive technique for the estimation of several soil properties simultaneously and in-situ with minimum or no sample pretreatments (Chang et al., 2001; Brown et al., 2006; Viscarra Rossel et al., 2006; Vasques et al., 2009). Chakraborty et al. (2010, 2012a, 2012b, 2014) demonstrated the capability of VisNIR DRS to estimate soil petroleum contamination from a single reflectance spectrum of the contaminated soil by means of multivariate regression models. Other researchers also independently showed the robustness of VISNIR DRS models for rapidly estimating TPH and polycyclic aromatic hydrocarbons (Graham, 1998; Forrester et al., 2010; Schwartz et al., 2012; Okparanma et al., 2014). The underlying principle is based on the diagnostic absorption bands (primarily overtones and combinations) in the VisNIR region arising from the C–H bond in hydrocarbons, helping in both qualitative and quantitative analysis of contaminated soils.

Another soil sensing technique, X-ray fluorescence spectrometry, has been used since 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; Weindorf et al., 2014; Zhu et al., 2011). Therefore, PXRF has become increasingly popular for soil/sediment analysis with references such as Method 6200 (USEPA, 2007) among others. PXRF 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; Paulette et al., 2015; 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., 2014a), cation exchange capacity (Sharma et al., 2014b), 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 horizonation (Weindorf et al., 2012b,c).

While the value, utility, and application of these two techniques (VisNIR DRS and PXRF) have been proven repeatedly in independent studies, these technologies have not yet been combined into a singular approach for rapid, on-site environmental quality assessment. Aldabaa et al. (2015) recently proved the utility of the PXRF + VisNIR DRS system in rapid assessment of soil salinity. Similarly, Wang et al. (2015) used a combined PXRF + VisNIR DRS approach for quantifying soil carbon and nitrogen simultaneously with excellent results. Combining these two approaches will allow for comprehensive assessment of both organic and inorganic soil contaminants and allow for instant quantitative results on-site. Because these approaches require no consumables and provide results within seconds, they offer tremendous advantages over traditional sampling, allowing for pervasive, high density spatial and temporal assessment of soil contaminants. Previous studies have elucidated the relationship of various elements with crude oil (hydrocarbons) (Groudeva et al., 2001; Gondal et al., 2006). Furthermore, the traditional analytical test procedures used to evaluate soil contamination by petroleum products are petroleum hydrocarbons and heavy metals quantification (Onianwa, 1995; Massoud et al., 1996; Kelly and Tate, 1998; Adeniyi and Afolabi, 2002). We hypothesize that the inclusion of metal/elemental data into VisNIR DRS hydrocarbon predictive models will enhance TPH predictive ability.

Thus, the objective of this study was to examine whether predictive models from synthesized proximal soil sensing techniques can provide better predictive ability than VisNIR DRS alone. This study is the first attempt to utilize VisNIR DRS spectra in combination with PXRF elemental data for the determination of petroleum hydrocarbons in contaminated soils. The proposed methodology could be applied in real time to provide rapid soil petroleum contamination assessment.

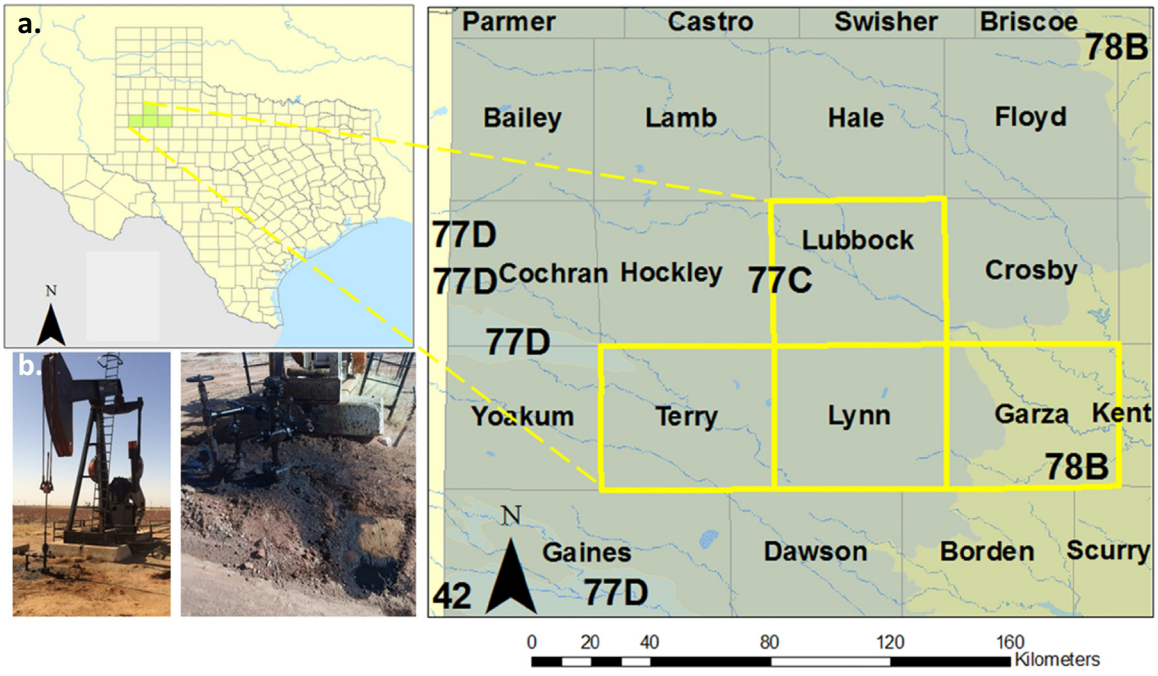

Fig. 1. a) Map showing the counties in West Texas, USA where the soil samples were collected and, b) pump jack with contaminated soil in the foreground and active cotton production in the background; pump jack with pervasive contamination at a rangeland site.

Table 1
Soil series and taxonomic classification of samples (n = 108) utilized for a PXRF + VisNIR DRS study of hydrocarbon contamination in West Texas, USA.

n	Soil Series	Classification[a]
2	Portales	Fine-loamy, mixed, superactive, thermic Aridic Calciustolls
40	Olton	Fine, mixed, superactive, thermic Aridic Paleustolls
2	Layton/Mobeetie	Mixed, mesic Psammentic Haploxerolls/coarse-loamy, mixed, superactive, thermic Aridic Haplustepts
5	Vernon	Fine, mixed, active, thermic Typic Haplustepts
3	Miles	Fine-loamy, mixed, superactive, thermic Typic Paleustalfs
27	Mobeetie/Potter	Coarse-loamy, mixed, superactive, thermic Aridic Haplustepts/loamy-skeletal, carbonatic, thermic Petronodic Ustic Haplocalcids
7	Patricia/Amarillo	Fine-loamy, mixed, superactive, thermic Aridic Paleustalfs/fine-loamy, mixed, superactive, thermic Aridic Paleustalfs
15	Amarillo	Fine-loamy, mixed, superactive, thermic Aridic Paleustalfs
6	Acuff	Fine-loamy, mixed, superactive, thermic Aridic Paleustolls
1	Abilene	Fine, mixed, superactive, thermic Pachic Argiustolls

[a] Soil Survey Staff (2014).

2. Materials and methods

2.1. Soil sample collection

Soil samples were collected at active oil production sites in Lubbock, Garza, Lynn, and Terry Counties, Texas, USA in January, 2014. For each site, efforts were made to collect surface (0–2 cm) soil samples from sites clearly impacted by hydrocarbon contamination (e.g., odor, visual evidence of oil) as well as from adjacent soils nearby with no evidence of contamination (control samples) (Fig. 1). Some of the sampling sites were in active crop fields, while other sites were in native rangeland for beef cattle production. In total, 108 samples were collected from the soils presented in Table 1. Sampling sites were within major land resource areas (MLRAs) 77C (Southern High Plains — Southern Part) and 78B (Central Rolling Red Plains — Western Part) (Soil Survey Staff., 2006). These areas are characterized by a semi-arid climate with average annual precipitation totaling 405–560 mm and average annual temperatures of 13 to 17 °C. Prevailing winds are from the south with an average freeze free period of 225 days y^{-1}. Geologically, the area is covered largely with aeolian deposits of the Blackwater Draw formation (Pleistocene age). In Garza County, Permian shale, sandstone, gypsum and dolomite in the Whitehorse, and Blain Formations are common. Aeolian deposits are on top of a large escarpment (caprock), extend westward some 100 km, and generally are used for irrigated agricultural production of cotton (*Gossypium hirsutum* L.). Rangeland to the east of the escarpment is commonly comprised of mesquite (*Prosopis glandulosa* Torr.), mixed grasses, and cacti.

At each sampling site, samples were placed in glass jars and sealed for transport back to the Texas Tech University Pedology Laboratory for spectral analysis. Upon arrival, samples were refrigerated at 5 °C. At each site, a brief site description was made and mapped soil recorded with the use of the SoilWeb application for iPhone; an application which uses geolocation to reference current position with soil survey geographic data for a given location.

2.2. VisNIR scanning and spectral pretreatments

The spectra of the soil samples were obtained using a PSR-3500® portable VisNIR spectroradiometer (Spectral Evolutions, 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. VisNIR experimental parameters and scanning procedures followed as part of this study are given by Aldabaa et al. (2015).

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 per Brown et al. (2006). 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 1 nm intervals. Furthermore, in order to reduce the influence of the side information contained in the original reflectance spectra, Savitzky–Golay first derivative transformation was applied 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). Only first derivative spectra were used in subsequent multivariate models.

2.2.1. 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 soil analysis, the instrument was calibrated using a 316 alloy clip, containing 16.13% Cr, 1.78% Mn, 68.76% Fe, 10.42% Ni, 0.20% Cu, and 2.10% 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

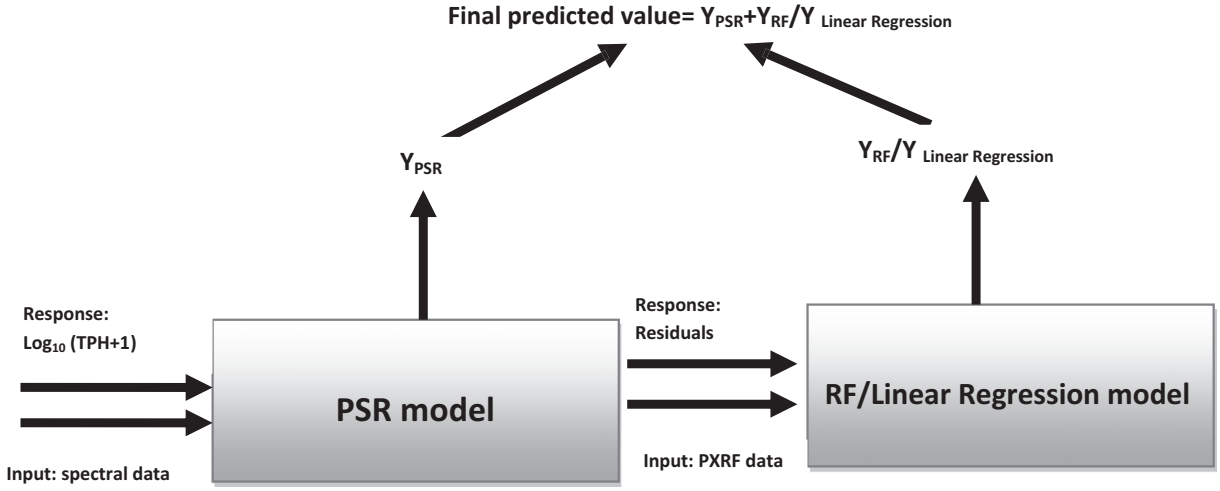

Fig. 2. Schematic diagram of fused PSR + RF and PSR + linear regression prediction models used in the study.

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.

2.3. Total petroleum hydrocarbon analysis

In this study, AR grade (Sigma) chemicals were used without further purification. All solutions were prepared with MilliQ™ (18.2 M Ω) water and sterilized by filtration (0.44 μm pore) or by autoclave at 120 °C.

A general TPH method was used to analyze oil content in soil samples (EPA method 8015 M); soils were first extracted using methylene chloride, and then extracts were analyzed by an HP6890 gas chromatograph (Hewlett-Packard, CA, USA) with flame ionization detection (GC-FID). Initial oven temperature (100 °C) was held for 1 min, followed by a 10 °C min^{-1} ramp to 275 °C and final hold for 5 min. The FID temperature was 330 °C.

A 7-point calibration curve was constructed using weathered (30 days) Macondo collected off the coast of Louisiana following the 2010 Deepwater Horizon oil spill. Quantification of TPHs was accomplished by summing the areas for 2–3 integrations of all gas chromatographic peaks beginning with n-hexane (n-C6) and ending with n-pentatriacontane (n-C35) similar to TNRCC Method 1005.

2.4. Machine learning

All statistical modeling was performed using R version 2.11.0 (R Development Core Team, 2008) software. A normality check was performed using the Shapiro–Wilk test statistic at a 5% significance level. In the present study, Box–Cox transformation (Box and Cox, 1964) was applied to the original positively skewed TPH data (Pearson skewness coefficient: 1.10) using $\lambda = 0$ ($\log_{10}$ transformed). Moreover, due to the presence of several control samples with 0 mg kg^{-1} TPH values, $\log_{10}$ (TPH + 1) was predicted as a target instead of $\log_{10}$ TPH.

Initially, we targeted $\log_{10}$ (TPH + 1) with VisNIR data via partial least squares regression (PLS), random forest regression (RF), and penalized spline regression (PSR) (Breiman, 2001; Guyon et al., 2002). The whole dataset was randomly divided only into a ~75% training set (n = 81) for calibration and a ~25% independent validation set (n = 27) to prevent overfitting. Each time, the three methods were applied on the training set and validated by test samples. The optimum number of PLS latent factors (rotations of principal components for a slightly different optimization criterion) was selected on the basis of the number of factors with the smallest total residual validation Y-variance or highest total explained validation Y-variance. We carefully examined outliers by marking them one-by-one and plotting the X–Y relation outliers for several model factors to monitor their influence on the shape of the X–Y relationship. For PSR, 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 leave-one-out-cross-validation (LOOCV) error on the training set. Moreover, the 'randomForest' 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) (Eq. (1)), ratio of performance to inter-quartile distance (RPIQ), and bias (Eq. (2)) were used as rubrics for judging model generalizing capability (Chang et al., 2001; Gauch et al., 2003; Bellon-Maurel et al., 2010)

$$RPD = \left[\frac{1/(n-1)\sum_{i=1}^{n}(Yobs-Ymean)^2}{1/n\sum_{i=1}^{n}(Yobs-Ypred)^2}\right]^{0.5}_{Validation} \quad (1)$$

Table 2
Summary statistics of soil (n = 108) samples utilized for a PXRF + VisNIR DRS study of hydrocarbon contamination in West Texas, USA.

	TPH	$\log_{10}$ (TPH + 1)	Al	Si	K	Ca	Ti	Mn	Fe	Cu	Zn	Rb	Sr	Y	Zr	Pb
	mg kg^{-1}															
Minimum	0.00	0.000	8891.28	46,772.44	3327.63	275.27	1010.35	65.59	7072.50	7.45	14.27	17.08	25.18	4.93	90.24	6.05
Maximum	326,294.48	5.514	34,595.56	220,289.36	14,740.18	149,770.88	6067.06	504.76	44,104.87	485.16	1828.91	107.82	1985.51	26.11	382.28	594.11
1st quartile	14,057.06	4.148	18,806.46	103,589.78	7848.56	14,858.24	1950.74	205.79	14,775.25	15.96	53.40	39.20	83.07	10.80	173.00	14.00
Median	54,678.88	4.738	23,811.48	134,800.24	8933.81	32,825.52	2282.81	255.09	18,046.71	21.69	77.65	46.32	109.39	12.93	201.37	22.62
3rd quartile	117,384.43	5.069	26,522.36	152,328.08	9980.20	53,295.69	2525.84	302.09	21,311.67	33.32	135.47	54.19	138.37	15.98	262.33	43.49
Mean	80,421.66	4.265	22,760.72	130,253.44	8951.10	35,561.34	2351.78	261.30	19,106.42	41.36	147.85	47.61	185.42	13.41	213.70	42.41
Variance (n − 1)	6,388,692,219.66	1.834	35,559,837.21	1,533,794,474.09	4,167,447.98	669,698,764.13	601,855.61	8589.18	46,883,696.62	4055.53	46,145.58	233.83	83,880.78	19.07	4149.32	4447.61
Standard deviation	79,929.30	1.354	5963.21	39,163.69	2041.43	25,878.54	775.79	92.68	6847.17	63.68	214.82	15.29	289.62	4.37	64.42	66.69

$$Bias = \sum_{i=1}^{n}(Ypred - Ymean)/n \quad (2)$$

where, *Yobs* and *Ypred* are the observed and predicted response variables, respectively, *Ymean* is the mean of the *Yobs* values, and *n* is the number of soils in the validation data set.

Subsequently, to test whether simple concatenation of PXRF elemental data with VisNIR DRS spectra can improve TPH accuracy, the aforementioned algorithms were used further. Besides, an advanced fused modeling approach (PSR + RF) was employed where PSR was used to fit the training set (containing spectra only) using LOOCV to choose the tuning parameter. Next, RF was used to fit the residual on the PXRF elemental data (Fig. 2). Further, we tried another fused model (PSR + linear regression) to compare with PSR + RF, since linear regression is a popular classic model.

Additionally, principal component analysis (PCA) was applied for qualitative PXRF + VisNIR DRS-based discrimination of contaminated and control soil samples. The cumulative proportion of variance explained by the leading principal components (PC) was used to extract optimum PCs. Furthermore, pairwise scatterplots of the first two PCs were produced to provide visual assessment of how contaminated and control samples were separated in the PC space. PCA was performed using R version 2.11.0 (function: prcomp).

3. Results and discussion

3.1. Soil TPH and elemental concentrations

Descriptive statistics of the different soil properties analyzed in this study are given in Table 2. Notably, not all elements were quantified with discreet values via PXRF. Thus, only Al, Si, K, Ca, Ti, Mn, Fe, Cu, Zn, Rb, Sr, Y, Zr, and Pb were viable for regression, with discreet values across all soil samples scanned. Half of the collected samples exhibited the incidence of V, Ni, and As which are normal constituents of crude oil. As reported earlier, some of the abovementioned metals perhaps originated as organo-metallic compounds in crude oil from which researchers have identified the geoporphyrins of V, Ni, Cu, and Zn (Gondal et al., 2006).

Our results converged with Grujic et al. (2004) who reported contamination of Pb, Cu, and Zn while investigating heavy metals in petroleum-contaminated surface soils of Serbia. Our results also corroborated the salient findings of Gondal et al. (2006), where laser induced breakdown spectroscopy revealed the presence of several trace elements in the residue of crude oil samples. While original TPH values were non-normally (Shapiro–Wilk stat = 0.873, p-value < 0.0001) distributed from 0 to ~3×10^6 mg kg^{-1}, $\log_{10}$ (TPH + 1) contents ranged from 0 to 5.51 mg kg^{-1} and were used as the dependent variable for subsequent predictive modeling (Table 2). For control samples, the mean TPH content of 203.2 mg kg^{-1} was expectedly lower than mean TPH of contaminated samples (89,518.6 mg kg^{-1}) and also exhibited less variability. A similar trend was observed in earlier studies (Albers, 1995; Onianwa, 1995; Chukwuma, 1996; Adeniyi and Afolabi, 2002), signifying petroleum hydrocarbon pollution. Strikingly, the magnitude of mean TPH obtained for the control samples (203.2 mg kg^{-1}) was several folds higher than the range of 1.00–26.63 mg kg^{-1} earlier reported for unpolluted soils (Adekambi, 1989; Hewari et al., 1995; Onianwa, 1995; Adeniyi and Afolabi, 2002), underlining that several seemingly control samples were actually somewhat contaminated, yet undetected by simple visual inspection.

3.2. Qualitative spectral analysis and PCA

Average reflectance spectra for two soil samples from the Olton soil series with high TPH content (243,792 mg kg^{-1}) and no TPH (control) are shown in Fig. 3. In general, reflectance spectra for both contaminated and control 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 et al., 1990). In the NIR region (700–2500 nm), mean spectral reflectance decreased as contamination increased causing higher absorbance and reflecting less light than control samples (Hoerig et al., 2001). The specific absorption minima of petroleum at ~1645 nm (C–H stretching modes of ArCH linked to polycyclic aromatic hydrocarbons), 1752 nm (C–H stretching mode of saturated CH_2 group in the first overtone region), and ~2240 nm (stretch + bend) of the NIR band were obvious (Mullins et al., 1992;

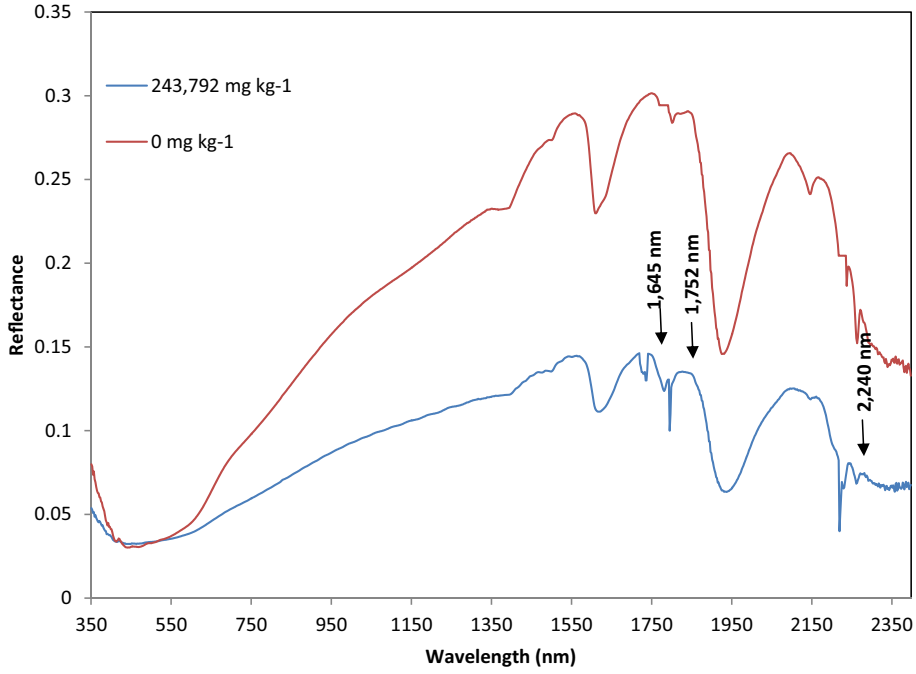

Fig. 3. Average reflectance spectra are shown for two soil samples of the Olton soil series with high total petroleum hydrocarbon content (243,792 mg kg^{-1}, blue) and zero total petroleum hydrocarbon (control, red). Spectral absorption maximums of petroleum at ~1645 nm, 1752 nm, and ~2240 nm are apparent in mean spectral reflectance curve of contaminated sample.

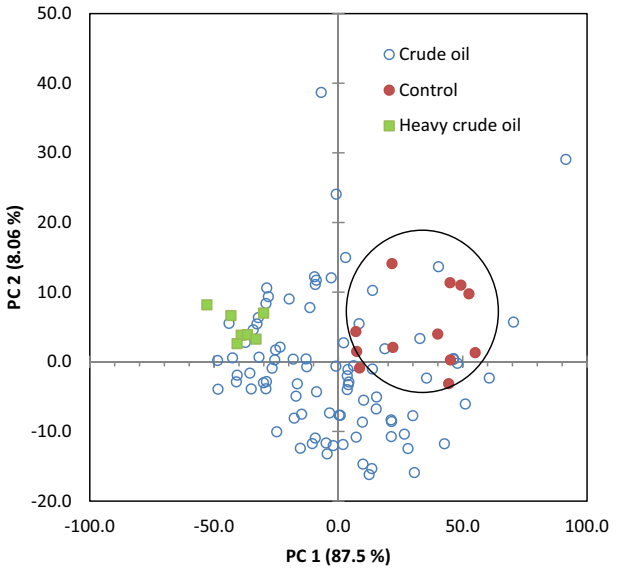

Fig. 4. Pairwise principal component plot for PC1 vs. PC2 for qualitative PXRF + VisNIR DRS based discrimination of contaminated and control soil samples. Red circles, hollow circles, and green squares represent control samples, crude oil contaminated samples, and heavy crude oil contaminated samples, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Osborne et al., 1993; Workman and Weyer, 2008). While the locations of aforementioned first and last signature were a bit shifted from the exact anticipated positions (1647 nm and 2298 nm, respectively), it was natural in the sense that real molecules do not behave totally harmonically. Other researchers identified that the first overtone of the C–H band makes the most important contribution for analysis of oil systems (Balabin and Safieva, 2007). However, these features were practically absent in the control reflectance curve.

Despite the high dimensionality of the spectral data (2151 channels from 350 to 2500 nm at 1-nm intervals) and PXRF data (14 elements), almost 100% of the variation was primarily explained by the first six PCs (99.22%). Fig. 4 illustrates how the pairwise PC score plot discriminated control soils from soils contaminated with crude oil and heavy crude oil (six samples containing >90,000 mg kg^{-1} TPH) along PC 1. Conversely, heavy crude oil samples were mixed with crude oil samples due to their compositional resemblance. Since the samples in the heavy crude oil set belonged to a single soil series (Mobeetie/Potter), they showed smaller spectral variation and their data cloud was compact. Conversely, the higher spectral diversity exhibited by the crude oil contaminated set indicated that it was compositionally diverse. Our results justified the ability of the PXRF + VisNIR DRS system to capture the intrinsic data structure in a two-dimensional representation.

3.3. Multivariate modeling

Three multivariate regression techniques were used to relate the PXRF elements and VisNIR DRS reflectance spectra to TPH contents with independent validation. Accuracy and stability of different multivariate models were evaluated according to the RPD-based rules of Chang et al. (2001). 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). Good, fair, and unreliable prediction models are characterized by RPDs of >2.0, 1.4–2.0, and <1.40, respectively. In this study, model RMSE and bias of (TPH + 1) were reported as $\log_{10}$ mg kg^{-1} and these values were only used for comparing multivariate models as they did not represent the expected error in concentration.

While using VisNIR spectra only, the PLS model ($R^2 = 0.73$, RMSE = 0.59 $\log_{10}$ mg kg^{-1}, RPD = 1.96, RPIQ = 0.63) outperformed both PSR and RF models (Table 3). In order to investigate whether using PXRF elemental data as predictors in addition to VisNIR spectra can improve TPH predictability, a PSR + RF model was constructed and produced the best accuracy relative to other models tested. Moreover, while incorporating PXRF data, the PLS model was somewhat able to prevent overfitting with fewer PLS latent factors (1). Apart from producing the highest RPD of 2.19, the PSR + RF model also exhibited the highest validation coefficient of determination (0.78) and RPIQ (0.75), suggesting satisfactory and effective calibration of (TPH + 1). Conversely, PSR + linear regression performed similarly as PSR, producing a fair RPD of 1.93. While comparing the PSR + RF and PSR + linear regression, it was evident that using RF on the residuals achieved better performance than linear regression. Fig. 5 comprehensively explains the rationale behind the better performance of PSR + RF, where Fig. 5a exhibits the RF relative variable importance for each of the PXRF elements. It was apparent that Sr and Zr were the two most influential ones, corroborating earlier studies (Collins, 1975). Moreover, RF partial dependence plots (Fig. 5b and c) exhibit clear non-linear effects of Sr and Zr on the residuals of the PSR model, respectively. Note that the partial dependence function shows the effect of a variable on the response in deciles (e.g., 10 percentile, 20 percentiles, etc.) after accounting for the average effects of all other variables in the model. Partial dependence plots provide a useful tool for interpreting the effects of a variable on the response particularly when there are no or weak interactions in the data or predictors. We refer the readers who are interested in the details of the partial dependence function to Friedman (2001). Another possible explanation for better performance of RF could be its ability to handle the interactions among these elemental variables implicitly while in a linear regression model the researcher has to select wisely and explicitly the interaction terms in the model.

From lab-measured versus model predicted (TPH + 1) plots, it was clear that the PSR model showed overestimation at lower values [i.e.,

Table 3
Summary model statistics obtained for soil (TPH + 1) by two different approaches using derivative spectra and three different multivariate algorithms. Also ratios of results of PSR+RF model with other model results are shown. For example, a value of 1.26 for PLS model only means the R^2 of PSR+RF using derivative spectra is ~26% higher than the former.

Approach	Model[a]	PLS LF[b]	R^2	RMSE ($\log_{10}$ mg kg^{-1})	RPD[c]	RPIQ[d]	Bias ($\log_{10}$ mg kg^{-1})	R^2	RMSE ($\log_{10}$ mg kg^{-1})	RPD
			Independent validation					Ratio with PSR + RF results for PXRF + VisNIR		
VisNIR only	PLS	3	0.73	0.59	1.96	0.63	−0.167	1.07	0.89	1.12
	PSR	–	0.70	0.75	1.86	0.60	−0.140	1.11	0.71	1.18
	RF	–	0.61	0.70	1.64	0.57	−0.241	1.28	0.75	1.34
VisNIR + PXRF	PLS	1	0.62	0.70	1.64	0.55	−0.244	1.26	0.75	1.34
	PSR	–	0.73	0.59	1.96	0.65	−0.167	1.07	0.89	1.18
	PSR + RF	–	0.78	0.53	2.19	0.75	−0.166	1.00	1.00	1.00
	PSR + linear regression	–	0.72	0.60	1.93	0.65	−0.060	1.08	0.88	1.14

[a] PLS, partial least squares regression; PSR, penalized spline regression; RF, random forest.
[b] PLS LF, partial least squares regression latent factor.
[c] RPD, residual prediction deviation.
[d] RPIQ, ratio of performance to inter-quartile distance.

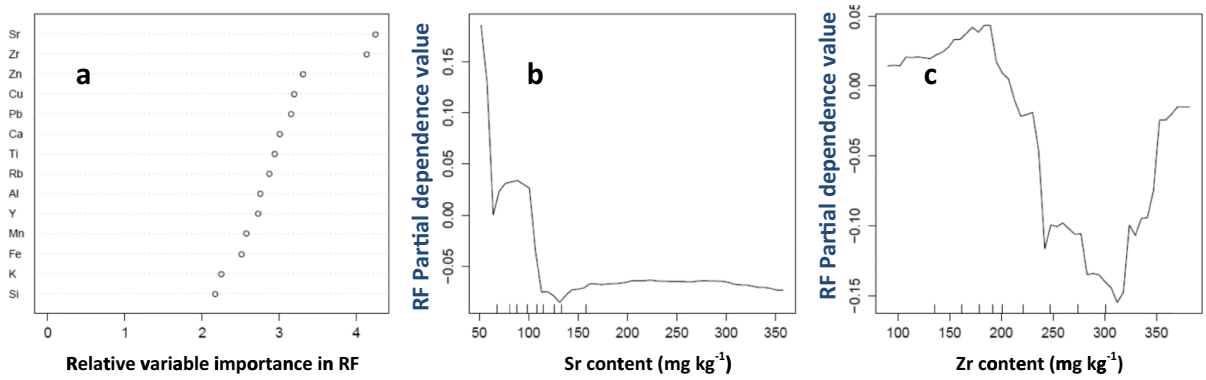

Fig. 5. Plots showing a) RF relative variable importance for each of the PXRF elements used in PSR + RF model and RF partial dependence function of b) Sr and c) Zr on the residuals of PSR model.

$\log_{10}$ (TPH + 1) is <2 mg kg^{-1}] and underestimation at higher values [i.e., $\log_{10}$ (TPH + 1) between 2 and 4.5 mg kg^{-1}] (Fig. 6). Exhibiting a similar trend and scattering from the 1:1 line, the prediction decreased further for the PLS model. There were only a few observations with zero $\log_{10}$ (TPH + 1) mg kg^{-1} which contributed to the overprediction of low (TPH + 1) values. Conversely, PSR + RF models closely approximated the 1:1 line, improving prediction accuracy. However, bias made a significant contribution to the overall lack of validation fit (>20% of MSE) for all models tested. The bias was expected due to the fact that the heterogeneity of sample origin was so large and that the validation set was perhaps quite different from the calibration set. A key requirement for model validation is that the validation samples should be independent and preferably scanned with different spectroradiometers (Brown et al., 2006). While we were unable logistically to achieve the latter requirement, the samples used in this study were otherwise independent.

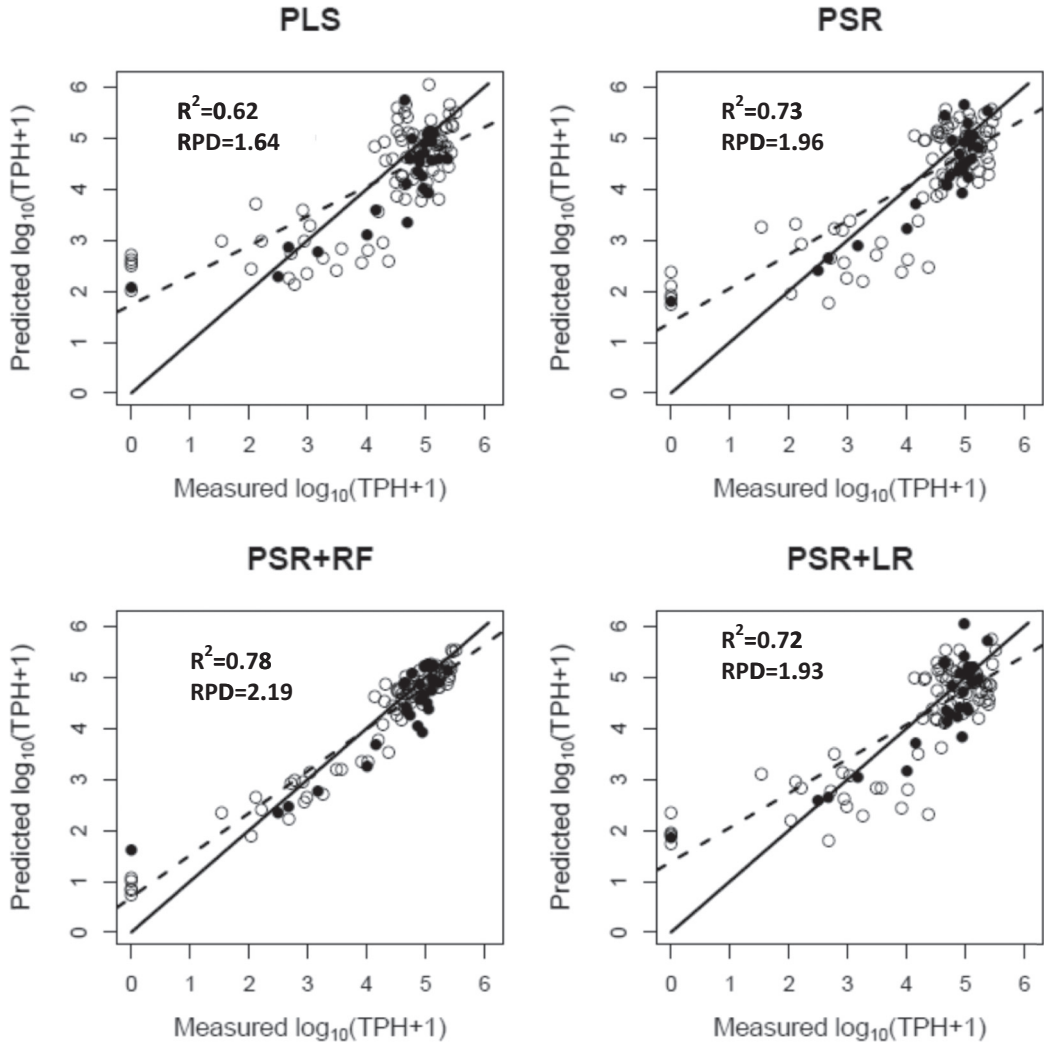

Fig. 6. Laboratory measured versus PXRF + VisNIR DRS predicted (TPH + 1) concentration ($\log_{10}$ mg kg^{-1}) using four multivariate models. LR represents linear regression. The dashed line is the regression line, and the solid line is a 1:1 line. White and black circles represent calibration and validation samples, respectively.

To clearly visualize the prediction improvement by using the fused PSR + RF model, Fig. 7 represents a scatterplot matrix produced in R using the spm function in the car library. The diagonal elements are the density plots for the three competitors [observed value (TPH + 1), PSR, and PSR + RF]. For example, the upper left one is the density plot of TPH [i.e., $\log_{10}$ (TPH + 1)]. While the tickmarks at the bottom axis show the observed values in the data, the curve is the smoothed density function of TPH and skewed to the left. The off-diagonal elements are the pairwise scatter plots of three competitors, together with the best linear and nonlinear smoothers. For example, the upper middle one shows the scatter plot of observed TPH (on the vertical axis) and predicted PSR + RF values (on the horizontal axis) using spectral data. The black and red solid lines are the fitted linear regression line and the loess smoother (a popular nonlinear smoother using local linear regression) fit, respectively. The red dash lines represent one standard error above and below the estimated function. It can be observed that the density functions of TPH and PSR + RF were more similar to each other than PSR to TPH. PSR estimates had a heavier tail than the ones from the observed TPH and PSR + RF estimates. This implied that PSR tended to underestimate the samples with small values of TPH between 2 and 4.5 [i.e., $\log_{10}$ (TPH + 1)]. This was further confirmed by observing the lower left scatter plot of TPH (on the horizontal axis) and PSR estimates (on the vertical axis). It can be seen that many points were below the linear regression line when TPH was between 2 and 4.5, implying that those underestimated samples from the PSR model have been lifted up by applying the RF on the PSR residuals. This was not surprising given the non-linear and contingent relationships between VisNIR DRS reflectance and soil composition (Clark, 1999).

In addition, the ratio of PSR + RF to other model results revealed some interesting trends, where the PSR + RF fused model clearly improved the coefficient of determination and RPD and lowered RMSE as compared to other models tested. A 14% increase in RPD was observed while using PSR + RF as compared to PSR+ linear regression. Notably, the fused model alone produced ~20% higher RPD than average RPDs of VisNIR DRS-based models, clearly highlighting the improvement by including PXRF elements with reflectance spectra.

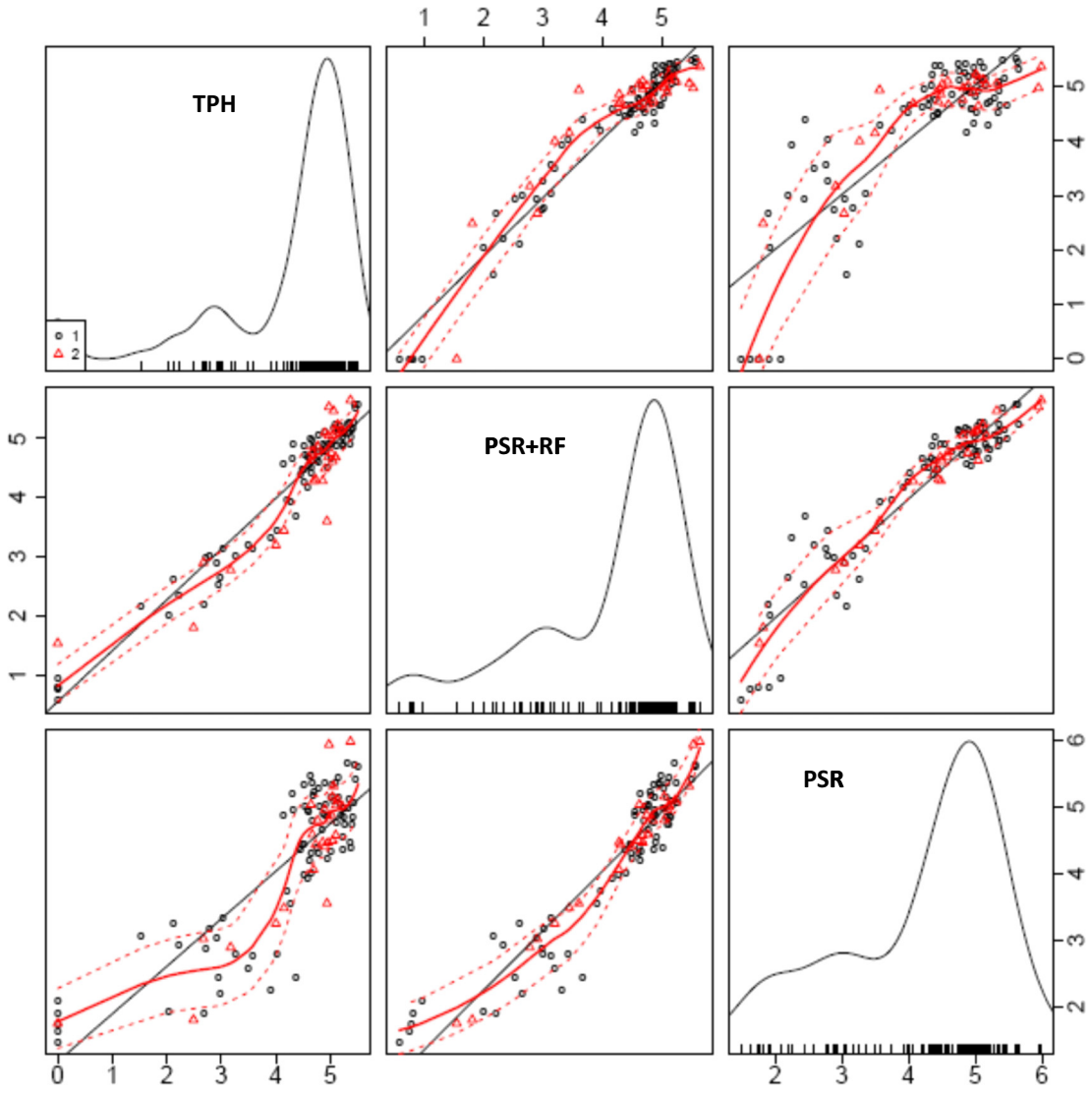

Fig. 7. Diagnostic scatter plot matrix showing density plots for three competitors: observed value (TPH + 1), penalized spline regression (PSR), and fused penalized spline random forest regression (PSR + RF). Black circles and red triangles are training and validation samples, respectively. The black and red solid lines are the fitted linear regression line and the loess smoother fit, respectively. The red dash lines represent one standard error above and below the estimated function. For diagonal plots, the vertical axis shows the density function for its corresponding values. For example, the top left one is the density plot of $\log_{10}$ (TPH + 1). In the off-diagonal plots, their axes are all in the $\log_{10}$ mg kg^{-1} unit. For example, for the bottom left one, the horizontal axis is the $\log_{10}$ (TPH + 1) mg kg^{-1} and the vertical axis title is the predicted $\log_{10}$ (TPH + 1) mg kg^{-1} values from PSR. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.4. Prospect of PXRF/VisNIR system for rapid soil petroleum contamination monitoring

In this study, one obvious question was whether it is possible to directly analyze soil petroleum contamination in fresh samples by PXRF + VisNIR DRS. This question was critical as we attempted to develop models which could ultimately lead to in-situ measurements. 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 (field capacity) prior to scanning is critical for obtaining consistent results. However, although under laboratory controlled conditions soil can be scanned under uniform moisture content, maintaining unique water content in the field is not easy. Even in air-dried conditions, water remains adsorbed on the surface areas of clay minerals (e.g., hygroscopic water) and organic matter in equilibrium with atmospheric water vapor. Interestingly, even pretreatment 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). Thus, in this study we relied on our earlier finding that the presence of moisture does not substantially reduce the predictability of TPH (Chakraborty et al., 2014). Researchers have also reported that NIR spectra calibrated using field soil samples can be used for field prediction of soil properties (Christy, 2008; Kusumo et al., 2008; Bricklemyer and Brown, 2010). Despite that, establishing the best sample pretreatment (field-moist or air-dried) was beyond the scope of this study and requires further investigations. Hence, influence of variable moisture remains one of the big issues to be addressed about calibrations for soil petroleum contamination. Even so, our approaches may be minimally impacted by soil moisture in many arid regions of the world where oil production is commonplace (e.g., Middle East, West Texas).

Summarily, although VisNIR DRS earlier showed considerable promise in providing rapid soil petroleum contamination prediction with reasonable accuracy, adding PXRF elemental data as auxiliary predictors was beneficial and should be included to model soil TPH. Brown et al. (2006) postulated that VisNIR DRS alone will never give complete soil characterization, thus application in parallel with other sensing technologies should be a focus of future research. Acquisition of PXRF data is rapid, easy, and cost-effective. While we concede that the initial purchase/investment of the equipment might seem high (perhaps ~ $100,000), the savings moving forward quickly recover the capital investment. As stated earlier, the proposed hybrid PXRF + VisNIR DRS system requires no consumables, minimum to no laboratory preparation time, and no laboratory facilities. In a traditional laboratory, analysis of TPH and heavy metals can easily cost $100 per sample or more and require days or weeks of laboratory processing. At that price, the cost of the hybrid PXRF + VisNIR DRS system would be recovered after 1000 or fewer samples; an attractive benefit for regulators and remediation specialists alike. Besides, as these instruments are field portable, direct assessment in-situ is also possible. Both evaluated techniques share the advantage of determining soil TPH especially for unusual circumstances where non-destructive sampling is required. Finally, our approaches provide more than just soil petroleum contamination, they provide specific elemental analyses; thus, more comprehensive information than simple TPH readings. Additional research should be continued to include larger geographical ranges along with other soil properties, but the future of PXRF + VisNIR DRS-based soil TPH characterization appears promising.

4. Conclusions

Summarily, we have demonstrated the potential of using a PSR + RF model using PXRF + VisNIR DRS data as a viable method for rapid and low-cost quantification of soil petroleum contamination as an addition to the standard methods for soil TPH analysis. We intended to test the capability of PXRF + VisNIR DRS viability instead of making a lab-grade predictive model. This study used a total of 108 petroleum contaminated and control soil samples from West Texas for scanning via VisNIR spectrometer and PXRF, followed by correlating diffuse reflectance data and elemental data with lab-determined soil TPH values. The advanced fused PSR + RF model using first derivative spectra outperformed all models tested, producing 78% variability of the independent validation set and a quality RPD (2.19). Unsupervised PCA also qualitatively separated contaminated soils from control samples. Hence, PXRF + VisNIR DRS showed the potential to be used as a post-spill rapid soil contamination monitoring tool. In this present feasibility study, we have presented a preliminary contribution to this problem and further intensive research is recommended to confirm the results here obtained. We anticipate that the future development of synthesized PXRF + VisNIR DRS methods and the expansion of soil-spectral libraries will support the assessment of soil petroleum contamination variability at a scale and resolution not previously possible.

Acknowledgments

The authors are grateful for the financial support from the BL Allen Endowment in Pedology at Texas Tech University. The synthesized use of PXRF + VisNIR DRS for soil physicochemical analysis is patent pending.

References

Adekambi, E.O., 1989. Petroleum Hydrocarbon Pollution in Nigerian Waters and Sediments Around Lagos and Niger Delta Areas of Nigeria. (PhD thesis). University of Ibadan, Ibadan, Nigeria.

Adeniyi, A.A., Afolabi, J.A., 2002. Determination of total petroleum hydrocarbons and heavy metals in soils within the vicinity of facilities handling refined petroleum products in Lagos metropolis. Environ. Int. 28, 79–82.

Albers, P.H., 1995. Petroleum and individual polycyclic aromatic hydrocarbons. In: Haffman, D.T., Rattner, B.A., Burton, G.A., Cairns, J. (Eds.), Handbook of Ecotoxicology. Lewis, London, pp. 330–355.

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.

Balabin, R.M., Safieva, R.Z., 2007. Capabilities of near infrared spectroscopy for the determination of petroleum macromolecule content in aromatic solutions. J. Near Infrared Spectrosc. 15, 343–349.

Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.M., McBratney, A., 2010. Critical reviews of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. Trends Anal. Chem. 29, 1073–1081.

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.

Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. J. R. Stat. Soc. Ser. B Methodol. 26, 211–252.

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

Brevik, E.C., Burgess, L.C. (Eds.), 2013. Soils and Human Health. Taylor Francis Press, Boca Raton, FL.

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

Camilli, R., Reddy, C.M., Yoerger, D.R., Van Mooy, B.A.S., Jakuba, M.V., Kinsey, J.C., McIntyre, C.P., Silva, S.P., Maloney, J.P., 2010. Tracking hydrocarbon plume transport and biodegradation at Deepwater Horizon. Science 330, 201–204.

Chakraborty, S., Weindorf, D.C., Morgan, C.L.S., Ge, Y., Galbraith, J., Li, B., Kahlon, C.S., 2010. Rapid identification of oil contaminated soils using visible near-infrared diffuse reflectance spectroscopy. J. Environ. Qual. 39, 1378–1387.

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

Chakraborty, S., Weindorf, D.C., Zhu, Y., Li, B., Morgan, C.L.S., Ge, Y., Galbraith, J., 2012b. Assessing spatial variability of soil petroleum contamination using visible near-infrared diffuse reflectance spectroscopy. J. Environ. Monit. 14, 2886–2892.

Chakraborty, S., Weindorf, D.C., Li, B., Ali, N., Majumdar, K., Ray, D.P., 2014. 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.

Christy, C.D., 2008. Real-time measurement of soil attributes using on-the-go near infrared reflectance spectroscopy. Comput. Electron. Agric. 61, 10–19.

Chukwuma, C.S., 1996. Evaluating baseline data for trace elements pH, organic matter content and bulk density in agricultural soils in Nigeria. Water Air Soil Pollut. 86, 13–34.
Clark, R.N., 1999. Spectroscopy of rocks and minerals, and principles of spectroscopy. In: Rencz, A.N. (Ed.), Remote Sensing for the Earth Sciences: Manual of Remote Sensing. Am Soc Photogramm Remote Sensing, Bethesda, MD, pp. 3–58.
Clark, R.N., King, T.V.V., Klejwa, M., Swayze, G.A., Vergo, N., 1990. High spectral resolution reflectance spectroscopy of minerals. J. Geophys. Res. 95 (B8), 12653–12680.
Collins, A.G., 1975. Geochemistry of Oilfield Waters. Elsevier, The Netherlands (508 pp.).
Dent, A., Young, A., 1981. Soil Survey and Land Evaluation. George Allen& Unwin Publ, Boston, MA.
Forrester, S., Janik, L., McLaughlin, M., 2010. An infrared spectroscopic test for total petroleum hydrocarbon (TPH) contamination in soils. Proceedings of the 19th World Congress of Soil Science, Soil Solutions for a Changing World, Brisbane, Australia, August 1–6, pp. 13–16.
Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29, 1189–1232.
Gauch, H.G., Hwang, J.T.G., Fick, G.W., 2003. Model evaluation by comparison of model-based predictions and measured values. Agron. J. 95, 1442–1446.
Gondal, M.A., Hussain, T., Yamani, Z.H., Baig, M.A., 2006. Detection of heavy metals in Arabian crude oil residue using laser induced breakdown spectroscopy. Talanta 69 (5), 1072–1078.
Graham, K.N., 1998. Evaluation of analytical methodologies for diesel fuel contaminants in soil. MS thesis.The University of Manitoba, Canada (Unpublished results).
Groudeva, V.I., Groudev, S.N., Doycheva, A.S., 2001. Bioremediation of waters contaminated with crude oil and toxic heavy metals. Int. J. Miner. Process. 62, 293–299.
Grujic, S., Ristic, M., Lausevic, M., 2004. Heavy metals in petroleum-contaminated surface soils in Serbia. Ann. Chim. 94 (12), 961–970.
Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene selection for cancer classification using SVM. Mach. Learn. 46, 389–422.
Hewari, J., Beaulie, C., Oullete, D., Pontbriand, Y., Halsaz, A.M., Vanata, H., 1995. Determination of petroleum hydrocarbons in soil: SFE versus Soxhlet and water effect on recovery. Int. J. Environ. Anal. Chem. 60, 123–137.
Hoerig, B., Kuehn, F., Oschuetz, F., Lehmann, F., 2001. HyMap hyperspectral remote sensing to detect hydrocarbons. Int. J. Remote Sens. 8, 1413–1422.
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.
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.
Kelly, J.J., Tate, R.L., 1998. Effects of heavy metals contamination and remediation on soil microbial communities in the vicinity of a Zn smelter. J. Environ. Qual. 27, 609–617.
Kusumo, B.H., Hedley, C.B., Hedley, M.J., Hueni, A., Tuohy, M.P., Arnold, G.C., 2008. The use of diffuse reflectance spectroscopy for in situ carbon and nitrogen analysis of pastoral soils. Aust. J. Soil Res. 46, 623–635.
Malley, D.F., Hunter, K.N., Barrie Webster, G.R., 1999. Analysis of diesel fuel contamination in soils by near-infrared reflectance spectrometry and solid phase microextraction–gas chromatography. J. Soil Contam. 8 (4), 481–489.
Massoud, M.S., Al-Abdali, F., Al-Ghadban, A.N., Al-Sarawi, M., 1996. Bottom sediments of the Arabian Gulf: II. TPH and TOC contents as indicators of oil pollution and implications for the effect and fate of the Kuwait oil slick. Environ. Pollut. 93, 27–284.
Mullins, O.C., Mitra-Kirtley, S., Zhu, Y., 1992. The electronic absorption edge of petroleum. Appl. Spectrosc. 46, 1405–1411.
Okparanma, R.N., Coulon, F., Mouazen, A.M., 2014. Analysis of petroleum-contaminated soils by diffuse reflectance spectroscopy and sequential ultrasonic solvent extraction–gas chromatography. Environ. Pollut. 184, 298–305.
Onianwa, P.C., 1995. Petroleum hydrocarbon pollution of urban top soil in Ibadan City, Nigeria. Environ. Int. 21, 341–343.
Osborne, B.G., Fearn, T., Hindle, P.H., 1993. Practical NIR Spectroscopy with Applications in Food and Beverage Analysis. 2nd ed. Longman Group UK Limited, England.
Oyedeji, A.A., Adebiyi, A.O., Omotoyinbo, M.A., Ogunkunle, C.O., 2012. Effect of crude oil-contaminated soil on germination and growth performance of *Abelmoschus esculentus* L. Moench—a widely cultivated vegetable crop in Nigeria. Am. J. Plant Sci. 3, 1451–1454.
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.
Paulette, L., Man, T., Weindorf, D.C., Person, T., 2015. Rapid assessment of soil and contaminant variability via portable X-ray fluorescence spectroscopy: Copşa Mică, Romania. Geoderma 243–244, 130–140.
R Development Core Team, 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (Available online with updates at http://www.cran.r-project.org, verified 6 January 2014).
Schwartz, G., Ben-Dor, E., Eshel, G., 2012. Quantitative analysis of total petroleum hydrocarbons in soils: comparison between reflectance spectroscopy and solvent extraction by 3 certified laboratories. Appl. Environ. Soil Sci. 2012, 1–11.
Sharma, A., Weindorf, D.C., Man, T., Aldabaa, A.A.A., Chakraborty, S., 2014a. Characterizing soils via portable X-ray fluorescence spectrometer: 3. Soil reaction (pH). Geoderma 232–234, 141–147.
Sharma, A., Weindorf, D.C., Wang, D., 2014b. Characterization of soils via portable X-ray fluorescence spectrometer: 4. Cation exchange capacity (CEC). Geoderma 239–240, 130–134.
Smith, L.C., Smith, L.M., Ashcroft, P.A., 2011. Analysis of environmental and economic damages from British Petroleum's Deepwater Horizon oil spill. Albany Law Rev. 74 (1), 563–585.
Soil Survey Staff, 2014. Official Soil Series Descriptions. USDA-NRCS. Available at, https://soilseries.sc.egov.usda.gov/osdname.asp (verified 29 Sept. 2014).
Soil Survey Staff., 2006. Land resource regions and major land resource areas of the United States, the Caribbean, and the Pacific Basin. USDA-NRCS. US Gov. Print. Office, Washington, DC.
Swanhart, S., Weindorf, D.C., Chakraborty, S., Bakr, N., Zhu, Y., Nelson, C., Shook, K., Acree, A., 2015. Measuring soil salinity via portable X-ray fluorescence spectrometry. Soil Sci. (Accepted; in press 12.5.14).
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., 2009. Modeling of soil organic carbon fractions using visible–near-infrared spectroscopy. Soil Sci. Soc. Am. J. 73, 176–184.
Vasques, G.M., Grunwald, S., Harris, W.G., 2010. Spectroscopic models of soil organic carbon in Florida, USA. J. Environ. Qual. 39, 923–934.
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, 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., 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.
Weindorf, D.C., Bakr, N., Zhu, Y., 2014. Advances in portable X-ray fluorescence (PXRF) for environmental, pedological, and agronomic applications. Adv. Agron. 128, 1–45.
Wiedenbeck, M., 2013. Field-portable XRF: a geochemist's dream? Elements 9, 7–8.
Workman Jr., J., Weyer, L., 2008. Practical Guide to Interpretive Near-infrared Spectroscopy. CRC Press, Taylor and Francis Group, Boca Raton, FL, USA.
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.