Analysis of petroleum-contaminated soils by diffuse reflectance spectroscopy and sequential ultrasonic solvent extraction–gas chromatography

Reuben N. Okparanma

Frederic Coulon

Abdul M. Mouazen*

a.mouazen@cranfield.ac.uk

mouazen2002@yahoo.com

Department of Environmental Science and Technology, Cranfield University, Cranfield, MK43 0AL Bedfordshire, United Kingdom

*Corresponding author.

Abstract

In this study, we demonstrate that partial least-squares regression analysis with full cross-validation of spectral reflectance data estimates the amount of polycyclic aromatic hydrocarbons in petroleum-contaminated tropical rainforest soils. We applied the approach to 137 field-moist intact soil samples collected from three oil spill sites in Ogoniland in the Niger Delta province (5.317°N, 6.467°E), Nigeria. We used sequential ultrasonic solvent extraction–gas chromatography as the reference chemical method. We took soil diffuse reflectance spectra with a mobile fibre-optic visible and near-infrared spectrophotometer (350–2500 nm). Independent validation of combined data from studied sites showed reasonable prediction precision (root-mean-square error of prediction = 1.16–1.95 mg kg^{-1}, ratio of prediction deviation = 1.86–3.12, and validation r^2 = 0.77–0.89). This suggests that the methodology may be useful for rapid assessment of the spatial variability of polycyclic aromatic hydrocarbons in petroleum-contaminated soils in the Niger Delta to inform risk assessment and remediation.

This approach may be used to collect large spatial data at reduced cost and time to assess the variability of polycyclic aromatic hydrocarbons in petroleum release sites.

Keywords: Chemometrics; Near-infrared spectroscopy; Hydrocarbon; Soil

1 Introduction

Rapid measurement of polycyclic aromatic hydrocarbons (PAHs) in soils can reduce the costs associated with their management, and the likely impacts of future pollution incidents through early identification. For effective management of PAHs in the environment, knowledge of their concentration and compositional distribution in the environment is vital. The information is essential to be able to identify the origin or source of the PAH (Hites and Gschwend, 1982; Okoro and Ikolo, 2007; Yunker and MacDonald, 1995) for risk-based assessment and remediation purposes (Askari and Pollard, 2005). In the analysis of petroleum-contaminated soils for PAH, gas chromatography–mass spectrometry (GC–MS) is mostly preferred because of their relative selectivity and sensitivity (Brassington et al., 2010; Wang and Fingas, 1995). Nevertheless, GC–MS method is relatively expensive, involves time-consuming sample preparation protocols, and relies on the use of noxious extraction solvents (Okparanma and Mouazen, 2013a). This has prompted increased research activity recently by scientists to evolve simpler, faster, and cost-effective methods of measuring PAH in contaminated soils to complement the conventional methods.

Among these innovative methods are techniques based on vibrational spectroscopy, particularly those employing diffuse reflectance sample interfaces. These techniques include Fourier transform infrared (FTIR) spectroscopy and visible and near-infrared (VIS–NIR) spectroscopy. Diffuse reflectance in the IR spectrum came to limelight in 1976, and became known as diffuse reflectance infrared Fourier transform (DRIFT) spectroscopy (Willey, 1976). Before its advent in the IR spectrum, diffuse reflectance has been commonly available in the visible and NIR spectrum (Willey, 1976). Early applications of DRIFT have been widely reported (e.g., Fuller and Griffiths, 1978). Both DRIFT and VIS–NIR are very much similar in many ramifications, but maybe more importantly are the differences between the two of them. Both techniques are now available in portable devices that can be deployed for in-field measurements without sample preparation. In terms of capital equipment cost, the prices of their latest portable models are comparable. The commercially available DRIFT (portable) system developed by Forrester et al. (2011), the 4100 ExoScan FTIR (Agilent Technologies, CA, USA), currently costs $61 301 while the latest portable model of VIS–NIR analyser, the LabSpec 5000 (Analytical Spectral Devices, Inc., CO, USA), costs $61 078 (Okparanma and Mouazen, 2013a). The former system predicted TPH with

root-mean-square error of 903 mg kg^{-1} in cross-validation for TPH ranging from 0 to 11 000 mg kg^{-1} (Forrester, 2010). On the other hand, VIS–NIR spectroscopy has been successfully used for on-line (tractor-mounted) measurement of soil properties (e.g., Mouazen et al., 2007a; Kuang and Mouazen, 2013), which opens the possibility for on-line measurement of PAHs using the technology. So far, no reports about the use of DRIFT for on-line measurement can be found in the literature.

In VIS–NIR spectroscopy (350–2500 nm), energy absorption by hydrocarbon derivatives is due to overtones and combinations of fundamental vibrational C–H stretching modes of saturated CH$_2$ and terminal CH$_3$, or aromatic C–H (ArCH) functional groups (Aske et al., 2001). The difficulty in interpreting VIS–NIR spectra because of broad and overlapping bands (Stenberg, 2010), to a large extent has been overcome by the use of advanced chemometrics and data-processing techniques (Pasquini, 2003). In the analysis of spectroscopic data, multivariate calibration generally solves the problem of interference from compounds closely related to the analyte thereby eliminating the need for selectivity (Naes et al., 2002). VIS–NIR spectroscopy has been used in conjunction with various multivariate analytical techniques to predict TPH in contaminated soils (Chakraborty et al., 2010, 2012; Forrester et al., 2010; Graham, 1998; Malley et al., 1999; Schwartz et al., 2012). However, studies on the application of the methodology to predict PAH in contaminated soils are few in literature (e.g., Bray et al., 2010; Okparanma and Mouazen, 2013b). High false positive rates were reported for PAH prediction by Bray et al. (2010), which underscore the need for further research on the application of the approach. On the other hand, models reported by Okparanma and Mouazen (2013b) for PAH prediction appear to be local models. This limits extrapolation of the models to studied soil types. Moreover, none of these two studies considered contaminated tropical rainforest soils in their applications. Therefore, to set up a VIS–NIR-based methodology to model VIS–NIR spectra for broader environmental application, it is essential to expand the approach to cover a wider range of soil types and environmental conditions.

The objective of this study was to evaluate the performance of VIS–NIR-based methodology in the prediction of PAH in contaminated tropical rainforest soils. To do this we used sequential ultrasonic solvent extraction–gas chromatography (SUSE–GC) as the benchmark method. Sites investigated in this study are in the tropical rainforests of the oil-rich Niger Delta province in Nigeria. To the best of our knowledge this study is the first attempt to adopt VIS–NIR spectroscopy for the determination of hydrocarbons in contaminated arable lands in Nigeria.

2 Materials and methods

2.1 Brief description of the study area

Available geological data for the study area show that Ogoniland is located within the Niger Delta basin with soils that are broadly classified as tropical rainforest soils, which occur in the southern part of Nigeria (SPDC, 2006). Niger Delta province covers a total land area of 70 000 km^2 (Niger Delta Environmental Survey, 1995). According to the United States Department of Agriculture (USDA) soil taxonomic order, soils in the Niger Delta belong to the Oxisols. In this study, soils were collected from within the shallow geology (top-soils) of Ogoniland, which consists of sandy clay (UNEP, 2011). Typical ranges of soil nutrients concentrations at all soil depths reported for similar ecosystems in the Niger Delta show that nitrate-nitrogen range from 0.01 to 1.96 mg kg^{-1}, phosphorous range from 0.21 to 6.92 mg kg^{-1}, sulphate range from 0.20 to 10.91 mg kg^{-1} and soil pH range from 5.2 to 6.4 (SPDC, 2006). Total organic carbon range from 3.63 to 4.11 % (Tanee and Albert, 2011).

2.2 Sample collection

In this study, we collected soil samples from three oil spill sites located at Baraboo (4.652°N, 7.249°E), Bomu 1 (4.662°N, 7.277°E), and Bomu 2 (4.662°N, 7.249°E) all in K-Dere, Ogoniland in the Niger Delta province of Nigeria (Fig. 1). By directed sampling method, we collected soil samples from the top 15-cm depth in plastic containers and preserved them in a cooler containing ice blocks until shipment. We adopted directed sampling method because of the need to cover as much of the visible hot-spots in the contaminated sites as possible. Sample management was done according to the standards of the Nigerian Government's Department of Petroleum Resources (2002). We carried out both reference SUSE–GC analysis and VIS–NIR optical scanning of soil samples in the Environmental Analytical Facility of Cranfield University, United Kingdom.

Online view   Print view
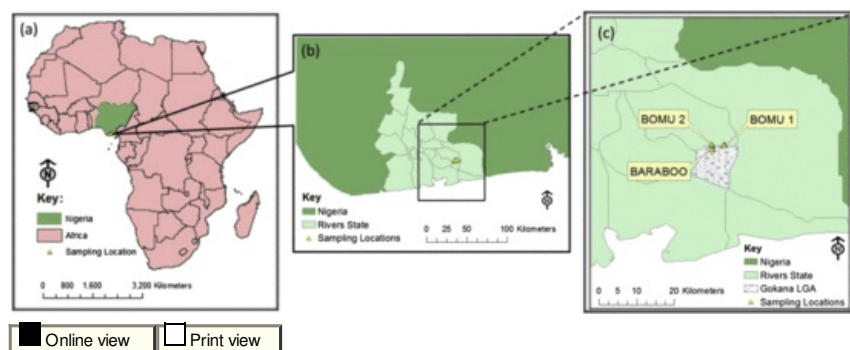
Fig. 1 Sampling locations: (a) Nigeria, (b) Rivers State in the Niger Delta province, and (c) Gokana Local Government Authority in Ogoniland (Datum and Projection: GCS WGS 1984. Shape files source: ESRI®, CA, USA).

2.3 Reference chemical analysis of PAHs

We used SUSE–GC method to determine the concentrations of PAH compounds in the soil samples as described by Risdon et al. (2008). But, in place of acetone we used dichloromethane (Rathburn Chemicals Ltd., Walderburn, UK). Analysis of PAHs was carried with a 6890N Network GC system coupled to a 5973 Network mass selective detector (Agilent Technologies Inc., USA) operated at 70 eV in positive ion mode. Each PAH compound was quantified by the internal standard method. In this study, we adopted the same limit of quantitation (LOQ) of 0.02 mg kg^{-1} routinely used by laboratories in Nigeria for PAH quantitation in Nigerian soils because our soil samples were collected from Nigeria. LOQ is the lowest concentration at which an analyte can be reliably detected (Mitra, 2003). Consequently, concentrations less than the LOQ were removed from the total PAH computation as they were considered unreliable. A priori 5-level calibration was carried out with a calibration solution mix. The calibration solution mix was made up with EPA 525 PAH Mix-A standard solution (Sigma–Aldrich Co. Ltd., Dorset, UK), surrogate standard solution mix, and solution mix of deuterated PAHs as internal standards. Deuterated PAHs used were Naphthalene-d8, Anthracene-d10, Chrysene-d12, and Perylene-d12 while surrogates used were 2-Fluorobiphenyl and o-Terphenyl (Sigma–Aldrich Co. Ltd., Dorset, UK). Matrix spikes, duplicates, solvent and method blanks were also analysed as quality control samples.

2.4 Accuracy, precision, and experimental uncertainty of reference PAH analytical method

Accuracy of the reference SUSE–GC chemical method was evaluated from the percent recovery of the surrogate spiked into the sample prior to extraction according to equation (1).

$$\% \text{ Recovery} = 100 \times \frac{\text{Measured concentration}}{\text{Theoretical concentration}} \tag{1}$$

Where, theoretical concentration is equal to the concentration of the surrogate standard spiked into the sample. PAH final concentrations were bias-corrected since percent spike recoveries were between 0 and 100 as recommended in literature (Patnaik, 1997). Biases in measured concentrations of 2- and 3-ring PAHs were corrected with the percent recovery of 2-Fluorobiphenyl while the percent recovery of o-Terphenyl was used to correct for biases in measured concentrations of 4-, 5-, and 6-ring PAHs. Individual PAH concentrations were then summed up to get total PAH. Since several replicate analyses of samples were not possible in this study, reproducibility (precision) of the analytical method was determined by estimating the relative percent difference of duplicate analyses of one sample in each batch of samples as recommended in literature (Patnaik, 1997). Relative percent difference was deduced by means of equation (2).

$$\text{Relative percent difference } (\%) = \frac{(a_1 - a_2)}{\frac{(a_1+a_2)}{2}} \times 100 \tag{2}$$

Where a_1 and a_2 are the PAH concentrations in the first and second duplicate sample, respectively. For each batch of samples, experimental uncertainty was estimated using confidence interval at 95 % level of confidence for duplicate analyses of one randomly chosen sample. The confidence interval was deduced by the well-known expression in equation (3).

$$\text{Confidence interval} = \overline{x} \pm \frac{ts}{\sqrt{n}} \tag{3}$$

Where, $\overline{x}$ = the sample mean, t = Student's t for a desired level of confidence, s = the sample standard deviation, and n = the number of measurements. Then, percent error (absolute) in the current measurement was checked against PAH results obtained from a commercial laboratory in Nigeria. To do this, we ran duplicate analyses of the same sample from Baraboo site that was previously analysed in the commercial laboratory. Percent error was then computed by means of equation (4).

$$\% \text{ Error} = \left| \frac{T - E}{T} \right| \times 100 \tag{4}$$

Where, T = "known" PAH value from commercial laboratory, and E = measured mean PAH value from current study.

2.5 Optical measurement of soil samples

We took diffuse reflectance spectra from the soil samples with a mobile fibre-optic LabSpec2500® VIS–NIR spectrophotometer (350–2500 nm) coupled to a high-intensity probe (Analytical Spectral Devices Inc., CO, USA). The spectrophotometer has one Si array (350–1000 nm) and two Peltier-cooled InGaAs detectors (1000–1800 nm and 1800–2500 nm). Spectral sampling interval of the instrument was 1 nm across the entire spectral range. However, the spectral resolution was 3 nm at 700 nm and 10 nm at 1400 and 2100 nm. The high-intensity probe has a built-in light source made of a quartz-halogen bulb of 2727 °C. The light source and detection fibres are assembled in the high-intensity probe enclosing a 35-degree angle. Before use, and after every 30 min, the instrument was optimised by white-referencing with a white Spectralon disc of almost 100 % reflectance. Reflectance spectra were taken from each soil sample, tightly packed and levelled out in a cuvette, at three different positions,

120° apart. Each sample was scanned three times at each position and averaged before spectral pre-processing and multivariate analysis.

2.6 Spectral pre-processing

The pre-processing of soil spectra was carried out with the Unscrambler® X (CAMO Software AS, Oslo, Norway). Noisy portions at the extremes of the spectrum (i.e., from 350 to 449 nm and 2451–2500 nm) were removed due to low instrument sensitivity at these wavelengths. Spectral truncation was followed by smoothing (by averaging successive 5-nm wavelengths), leaving a total of 401 wavelengths in the working range of 452–2450 nm. Then spectral transformation was carried out by successive combination of maximum normalization and Savitzky–Golay first derivative of polynomial order of two and two smoothing points. Normalization helps to bring all data to approximately the same scale or to get a more even distribution of the variances and the average values. First derivative removes additive baseline shifts in the data and smoothing reduces the impact of noise. These measures were aimed at reducing spurious peaks that do not hold physical or chemical information (Aske et al., 2001; Naes et al., 2002).

2.7 Partial least-squares (PLS) regression analysis

Before calibration, spectral reflectance (R) was transformed to the logarithm of the relative intensity (1/R), or absorption (Naes et al., 2002). The PLS regression analysis combines both the independent variables (reference values of PAH) and the dependent variables (wavelengths) using them as regression generators for the independent variables (Maleki et al., 2007). Detailed information about the PLS can be found in Martens and Naes (1989). We used PLS regression analysis with full cross-validation to relate the variation in a single-component variable (e.g., PAH) to the variation in a multi-component variable (e.g., wavelength) by means of Unscrambler® X. The optimal number of latent factors for future predictions was determined on the basis of the number of factors with the smallest total residual validation Y-variance or highest total explained validation Y-variance (CAMO Software, 2012). Site-specific calibration models were developed with VIS–NIR spectral data and reference SUSE–GC chemical data for each site. Also, a generalised model was developed for all three sites with the entire dataset. To develop the generalised model, 78 % of the samples were used for cross-validation (calibration) while the remaining 22 % were used for validation (prediction). The ratio of calibration/validation samples was chosen to ensure an equal representation of samples in the validation set by randomly choosing ten samples from each site.

2.8 Statistical evaluation of model performance

Criteria used to evaluate the quality of the models were based on the root-mean-square error (RMSE) of cross-validation and prediction (equations (5) and (6), respectively), ratio of prediction deviation (RPD) (equation (8)), and corresponding coefficient of determination (r^2) (Williams and Sobering, 1986). Model performance classification system adopted was based on the following criteria: very poor model predictions if $RPD < 1.0$; poor if $1.0 \le RPD < 1.4$; fair if $1.4 \le RPD < 1.8$; good if $1.8 \le RPD < 2.0$; very good if $2.0 \le RPD < 2.5$; and excellent if $RPD > 2.5$ (Viscarra Rossel et al., 2006a).

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{N} \left(\widehat{y}_{CV,i} - y_i\right)^2}{N}} \tag{5}$$

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{N_P} \left(\widehat{y} - y_i\right)^2}{N_P}} \tag{6}$$

$$RMSE = \sqrt{MSE} = \sqrt{E(\widehat{y} - y)^2} \tag{7}$$

$$RPD = \frac{SD}{RMSE} \tag{8}$$

Where $\widehat{y}_{CV,i} =$ estimate for y_i based on the calibration equation with i deleted; $\widehat{y}$ and $y_i =$ predicted and measured reference values respectively; $E(.) =$ statistical expectation (average) over the population of future samples; $N =$ number of samples in the set; and SD = standard deviation of the measured reference values.

2.9 Outlier detection techniques

During model calibration, a priori outlier detection techniques were adopted to remove influential X- and Y-outliers. The influence of sample outliers was ascertained using the influence plot (Fig. 2). This was done by plotting the residual X- and Y-variances against leverages. Samples with a high leverage and high residual X- or Y-variance were considered as potential outliers (CAMO Software, 2012). Then the projected influence plots were used to confirm those samples with high residuals. Nonetheless, before such samples were treated as outliers, they were studied in more details by marking them one-by-one and plotting the X–Y relation outliers for several model factors to observe their influence on the shape of the X–Y relationship, as
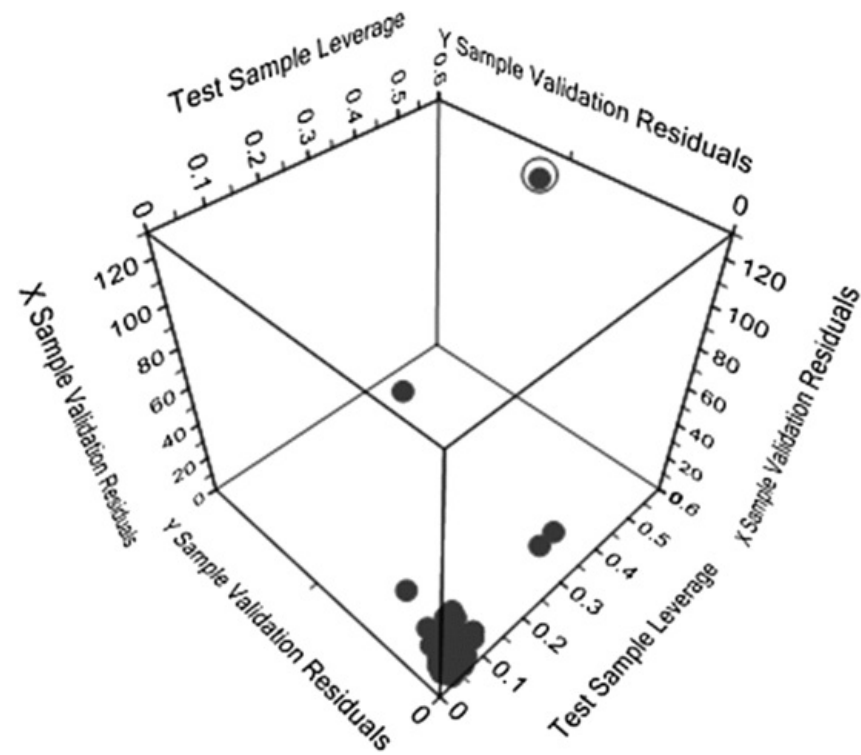
recommended in literature (CAMO Software, 2012).

Fig. 2 Detection of outliers after partial least-squares regression analysis. A potential sample outlier, marked in circle, detected among samples from Baraboo site in Ogoniland, Niger Delta province of Nigeria is shown as an example.

3 Results and discussion

3.1 Accuracy, precision and level of uncertainty in reference SUSE–GC analysis of PAH

Table 1 shows the statistical results of the chemical SUSE–GC analysis showing the sum of individual PAHs quantified for each site. The high spatial variability of the compounds within and among sites is clearly demonstrated in the table. It is important that the sum of a PAH fraction in the entire sample set (Table 1) is differentiated from the sum of the entire PAH fraction in a sample (Data not shown due to the number of samples involved). The latter is the total PAH in a sample used for model development. In this study, total PAH concentration is the sum of the 13 individual PAH concentrations in a sample. The 13 PAHs in Table 1 were the relevant compounds contained in the PAH standard solution used for prior instrument calibration. The concentrations of the lower boiling 2- to 4-ring PAHs (i.e., acenaphthylene to chrysene) were higher than the higher boiling 5- to 6-ring PAHs. The difficulty in quantitation increased with boiling point. The lower boiling PAHs eluted the column before the higher boing PAHs because of their shorter retention times. Overall, Bomu 2 appeared to be the most contaminated site with the highest mean total PAH (5.39 mg kg^{-1}) compared to either Baraboo site (3.38 mg kg^{-1}) or Bomu 1 (4.47 mg kg^{-1}). These values can also be confirmed from Table 1 by adding the respective mean values.

Table 1 Statistics of the chemical analysis result showing the sum of individual polycyclic aromatic hydrocarbons (PAHs) quantified for each site by reference sequential ultrasonic solvent extraction–gas chromatography (SUSE–GC).

PAH	LOQ (mg kg^{-1})	Baraboo					Bomu 1					Bomu 2				
		N	Min. (mg kg^{-1})	Max. (mg kg^{-1})	Mean (mg kg^{-1})	Range (mg kg^{-1})	N.	Min. (mg kg^{-1})	Max. (mg kg^{-1})	Mean (mg kg^{-1})	Range (mg kg^{-1})	N	Min. (mg kg^{-1})	Max. (mg kg^{-1})	Mean (mg kg^{-1})	Range (mg kg^{-1})
Acenaphthylene	0.02	43	0.07	3.06	0.83	2.98	58	0.02	2.40	0.46	2.38	36	<0.02	8.53	1.41	8.53

PAH	LOQ	N				N					N				
Fluorene	0.02	43	<0.02	4.80	0.73	4.80	58	<0.02	2.70	0.45	2.70	36	<0.02	6.25	0.88	6.25
Phenanthrene	0.02	43	<0.02	3.64	0.47	3.64	58	0.02	3.15	0.81	3.13	36	<0.02	6.86	0.76	6.86
Anthracene	0.02	43	<0.02	4.22	0.41	4.22	58	<0.02	2.57	0.36	2.57	36	<0.02	6.86	0.75	6.86
Pyrene	0.02	43	<0.02	1.03	0.30	1.03	58	<0.02	2.83	0.56	2.83	36	<0.02	1.73	0.52	1.73
Benz[a]anthracene	0.02	43	<0.02	0.88	0.18	0.88	58	<0.02	2.61	0.39	2.61	36	<0.02	1.01	0.15	1.01
Chrysene	0.02	43	<0.02	1.17	0.18	1.17	58	<0.02	1.07	0.22	1.07	36	<0.02	0.55	0.09	0.55
Benz[b]fluoranthene	0.02	43	<0.02	0.15	0.03	0.15	58	<0.02	1.33	0.23	1.33	36	<0.02	0.18	0.04	0.18
Benzo[k]fluoranthene	0.02	43	<0.02	0.15	0.03	0.15	58	<0.02	1.23	0.19	1.23	36	<0.02	0.18	0.04	0.18
Benzo[a]pyrene	0.02	43	<0.02	0.73	0.17	0.73	58	<0.02	1.80	0.14	1.80	36	<0.02	9.36	0.68	9.36
Indeno[1,2,3-cd]pyrene	0.02	43	<0.02	0.22	0.02	0.22	58	<0.02	1.48	0.23	1.48	36	<0.02	0.82	0.06	0.82
Dibenzo[a,h]anthracene	0.02	43	<0.02	0.15	0.03	0.15	58	<0.02	1.82	0.22	1.82	36	<0.02	<0.02	–	–
Benzo[g,h,i]perylene	0.02	43	<0.02	0.22	0.02	0.22	58	<0.02	1.48	0.22	1.48	36	<0.02	0.33	0.02	0.33

LOQ, Limit of quantitation. It is the lowest concentration at which an analyte can be reliably detected (Mitra, 2003).

N, Number of samples.

PAH, Polycyclic aromatic hydrocarbons.

Percent recoveries of spiked surrogates for the three sites, as shown in Table 2, are within the acceptable range of 40–120 % (USEPA, 1999), which is typical of a reasonably accurate PAH extraction procedure.

Table 2 Accuracy of the reference sequential ultrasonic solvent extraction–gas chromatography (SUSE–GC) method used in the chemical analysis of polycyclic aromatic hydrocarbons (PAH) in topsoils from petroleum-contaminated sites in Ogoniland, Niger Delta province of Nigeria.

Sampling site	No. of samples	Spiked (mg kg⁻¹)	Measured (mg kg⁻¹)[a]		% Recovery of spiked surrogates
			2-Fluorobiphenyl	*o*-Terphenyl	
Baraboo	43	2.50	1.36	1.33	53–54
Bomu 1	58	2.50	1.94	1.63	65–78
Bomu 2	36	2.50	1.19	1.40	48–56

[a] Mean concentrations of the spiked surrogates.

Table 3 shows the reproducibility and experimental uncertainty for 95 % confidence interval ($n = 2$) of the reference PAH measurement procedure. The relative percent difference for duplicate samples for all sites is less than 20 %, suggesting that the reference SUSE–GC method was within precision standards (Mayer, A. S. [2008] – Lecture Note, GE3850: Geohydrology, Department of Civil and Environmental Engineering, Michigan Technology University, MI.). The rather large confidence intervals in Table 3 are attributed to the fewer number of measurements ($n = 2$). It is known that the confidence interval reduces with number of measurements as the sample mean approaches the true population mean (Harris, 2010). Nevertheless, the estimated confidence interval suggests with 95 % confidence that the true mean PAH value from the duplicate measurements would lie within the estimated range.

Table 3 Experimental uncertainty and precision of reference sequential ultrasonic solvent extraction–gas chromatography (SUSE–GC) method used for analysis of polycyclic aromatic hydrocarbon (PAH) in topsoils from petroleum-contaminated sites in Ogoniland, Niger Delta of Nigeria. Test sample was randomly chosen for each site.

Sampling site	First duplicate (mg kg⁻¹)	Second duplicate (mg kg⁻¹)	Relative percent difference (%)	Confidence interval
				Mean ± uncertainty[a] (mg kg⁻¹)
Baraboo	3.67	3.22	13	3.45 ± 2.86
Bomu 1	4.49	3.88	14	4.18 ± 3.84
Bomu 2	2.89	2.66	8	2.78 ± 1.42

Inter-laboratory differences in reported PAH values are shown in Table 4. For the test soil sample, Table 4 shows that mean PAH value obtained in this study is comparable to PAH result from the commercial laboratory in Nigeria. This implies that the current PAH measurement by Cranfield University's operating procedures described above is in agreement with those of the commercial laboratory. The margin of error may be attributed to differences in operating procedures used in both laboratories (Table 4). Moreover, inter-laboratory differences in reported values of an analyte in chemical analysis are a common occurrence (Risdon et al., 2008).

Table 4 Inter-laboratory differences in reported PAH values and selected rubrics in operating procedures for test soil sample from petroleum-contaminated site in Baraboo in Ogoniland, Niger Delta of Nigeria.

Analyte and selected rubrics	This study	Commercial laboratory
PAH (mg kg^{-1})	3.45 ± 2.86	2.73[a]
Percent error (%)[b]	26	–
Quantitation method	Internal standard	External standard
Extraction method	Sequential ultrasonic solvent extraction	Sonication water bath (5-hour sonication)
Extracting solvent(s)	Dichloromethane and hexane (1:1)	n-Pentane
Surrogate standard(s)	2-Fluorobiphenyl and o-Terphenyl	1-Chlorooctadecane

a Uncertainty was not reported.

b Commercial laboratory PAH result was taken as the "known" value.
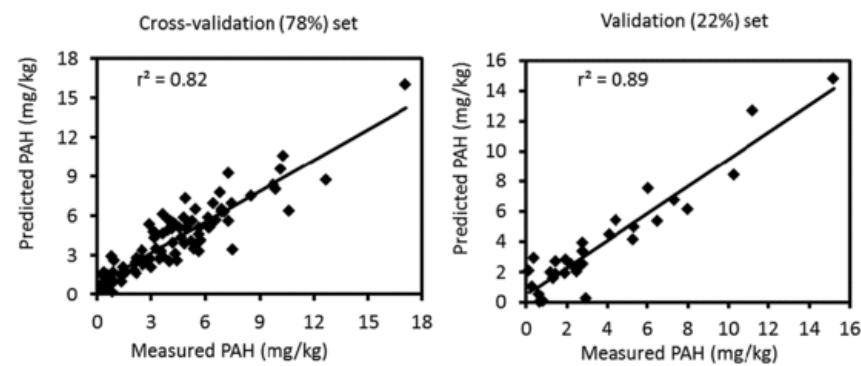
3.2 The PAH partial least-squares regression models

Table 5 summarises the statistical results of the PLS models. The cross-validation models for pre-processed spectra used smaller number of latent variables than raw spectra. Similarly, cross-validation models developed by pre-processed spectra had better quality than raw reflectance models. Overall, the quality of the cross-validation models is classified as ranging from very good to excellent on the basis of RPD values. Fig. 3 shows the linear relationship between reference SUSE–GC-measured and VIS–NIR-predicted PAH data for pre-processed spectra of entire dataset. Only three sample outliers were removed from the cross-validation datasets while none was removed from the independent validation set. On the basis of RPD values, the model quality is classified as excellent prediction. The quality of the model demonstrates the possibility of using VIS–NIR spectroscopy for quantitative determinations (Viscarra Rossel et al., 2006a). Moreover, as shown in Table 5 the similarity between cross-validation and validation standard deviations indicates that the prediction models are not skewed. These results are comparable with RPD values of 1.7 and 2.5 reported by previous researchers for TPH prediction in field-moist soils by PLS regression analysis (Chakraborty et al., 2010, 2012).

Table 5 Statistical results of partial least-squares (PLS) site-specific calibration and general prediction models for polycyclic aromatic hydrocarbons (PAH) in petroleum-contaminated tropical rainforest soils from Ogoniland in the Niger Delta province of Nigeria developed by visible and near-infrared (VIS–NIR) spectroscopy.

Sampling site	No. of samples	Reflectance spectra					Combined pre-processing spectra[a]				
		r^2	RMSE (mg kg^{-1})	SD	RPD	LV	r^2	RMSE (mg kg^{-1})	SD	RPD	LV
Cross-validation set											
Baraboo	43	0.78	0.82	1.81	2.20	8	0.84	0.64	1.81	2.81	6
Bomu 1	58	0.76	1.41	2.95	2.09	11	0.83	1.22	2.95	2.41	3
Bomu 2	36	0.77	1.98	4.08	2.07	5	0.83	1.42	4.08	2.87	7
General	107	0.69	1.61	3.05	1.81	11	0.82	1.30	3.05	2.34	8
Validation set											
General	30	0.77	1.95	3.63	1.86	11	0.89	1.16	3.63	3.12	8

LV, Latent variable.

RMSE, Root-mean-square error.

RPD, Ratio of prediction deviation.

SD, Standard deviation.

[a] Combination of maximum normalisation, and first derivative and smoothing by Savitzky–Golay.

Fig. 3 Scatter plots of chemically measured vs. predicted values of polycyclic aromatic hydrocarbons (PAH). These were developed by partial least-squares (PLS) regression analysis with pre-processed spectra of 137 soil samples from petroleum-contaminated tropical rainforest soils collected from Ogoniland, Niger Delta, Nigeria.

3.3 Spectral reflectance of petroleum-contaminated tropical rainforest soils

Fig. 4 shows mean VIS–NIR spectral reflectance curves of field-moist tropical rainforest soils consisting of sandy clay from three oil spill sites in Ogoniland, Niger Delta, Nigeria. Generally, soil spectral reflectance decreased with increasing mean PAH concentration particularly in the NIR region (700–2500 nm). There was however, a slight shift in this trend for Baraboo and Bomu 1 reflectance curves in the visible range (452–700 nm). This is attributed to soil colour-associated changes in the visible range brought about when water and/or oil (e.g., diesel) is added to soil (Mouazen et al., 2005; Okparanma and Mouazen, 2013b). In the reflectance curves for the contaminated sites, spectral absorption minima of hydrocarbon-based oil were observed around 1647, 1712, and 1752 nm in the first overtone region of the NIR band (Fig. 4). The absorption around 1647 nm is attributed to C–H stretching modes of ArCH linked to PAH. Absorptions around 1712 and 1752 nm are attributed to C–H stretching modes of terminal CH_3 and saturated CH_2 groups linked to TPH, both present in the contaminating hydrocarbon-based oil (Forrester et al., 2010; Osborne et al., 1993; Workman and Weyer, 2008). These features are practically absent in the uncontaminated reflectance curve as shown by the number of reflectance shoulders that appeared in the reflectance curves between 1452 and 1952 nm (Fig. 4).
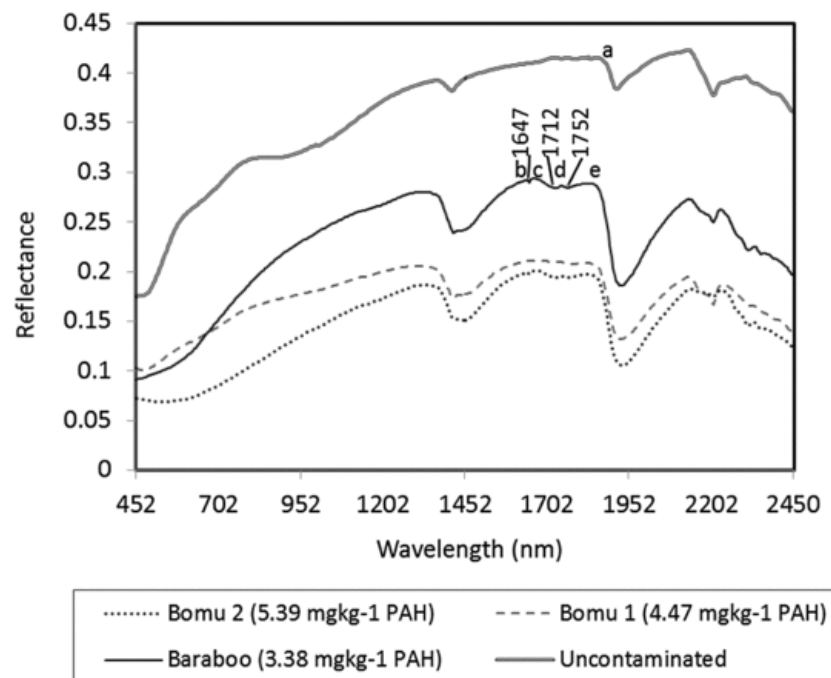
Fig. 4 Mean VIS–NIR spectral reflectance curves of petroleum-contaminated tropical rainforest soils from three oil spill sites in Ogoniland in the Niger Delta province of Nigeria. Values in the legend are average PAH concentrations. Point a, b, c, d, and e are reflectance shoulders.

One reflectance shoulder appeared around 1830 nm (point a) in the uncontaminated reflectance curve while four appeared around 1630, 1675, 1737, and 1830 nm (point b, c, d, and e, respectively) in the contaminated reflectance curves (Fig. 4). A reflectance shoulder appears between two absorption wavelengths of colour, water and/or hydrocarbon (Mouazen et al., 2007b; Okparanma and Mouazen, 2013b). The reflectance shoulder at 1830 nm is attributed to water absorption in the first overtone band around 1950 nm (Stenberg, 2010) and hydrocarbon absorption in the first overtone band around 1712 nm (Osborne et al., 1993; Workman and Weyer, 2008). The shoulder at 1630 nm is attributed to water absorption in the second overtone region around 1450 nm (Mouazen et al., 2007b) and hydrocarbon absorption in the first overtone band around 1647 nm (Workman and Weyer, 2008). The shoulder at 1675 nm is attributed to hydrocarbon absorptions in the first overtone band around 1647 and 1712 nm (Workman and Weyer, 2008). The shoulder at 1737 nm is attributed to hydrocarbon absorptions in the first overtone band around 1712 and 1752 nm (Workman and Weyer, 2008). Therefore, the hydrocarbon absorption bands around 1647 and 1712 nm differentiate the uncontaminated from contaminated reflectance curves (Fig. 4).

3.4 Regression coefficients

Regression coefficients in PLS regression analysis provide a summary of all predictors and a given response. The plot of regression coefficients is used to identify important wavelengths for the prediction of relevant soil properties. Fig. 5 shows bar plots of regression coefficients versus wavelength derived after PLS regression analysis using raw VIS–NIR spectral data and reference SUSE–GC chemical data for each site. In the bar plots, the absolute value of the regression coefficients indicate the relative importance of the wavelength on the basis of explained X-variance in the model. In Fig. 5, negative coefficients around 1712 and 1752 nm show a positive link to absorptions due to vibrational C–H stretching modes of terminal CH_3 and saturated CH_2 functional chemical groups linked to TPH. But positive coefficients around 1647 nm are consistent with absorptions due to vibrational C–H stretching modes of ArCH functional groups linked to PAH. These results agree with absorption bands observed in the reflectance curves (see Fig. 4). Over the modelling wavelength range of 452–2450 nm, the bar plots show that intensities of regression coefficients vary in magnitude for each site. Nonetheless, the positions of important wavelengths are largely similar (Fig. 5) even though the samples are from different sites. This shows a similarity in soil property among the three sites. We observed larger regression coefficients in Fig. 5b and c than in Fig. 5a. The intensity of regression coefficients for Baraboo site was the least (Fig. 5a), and samples from Baraboo site showed the least level of contamination among the three sites as well (see Fig. 4). This may be explained by the fact that the mechanism of PAH prediction is attributed to co-variation of PAH with other soil properties that have direct spectral responses in the NIR spectral range, particularly water, clay minerals and organic carbon (Okparanma and Mouazen, 2013c). This observation was also typified in Fig. 5 by the sizes of coefficients around 950, 1450, 1950, 2200, 2212, and 2300 nm. Absorptions around 950, 1450 and 1950 nm in the NIR band are due to O–H stretching modes of water in the O–H second and first overtones, and combinations band respectively (Whalley and Stafford, 1992). Absorption features linked to metal-OH bend plus O–H stretch combinations around 2200 and 2300 nm are characteristic of clay minerals (Clark et al., 1990; Viscarra Rossel et al., 2006b). Absorption features attributed to long-chain C–H + C–H and C–H + C–C stretch combinations

around 2150 nm and 2212 nm are unique to soil organic matter (Forrester et al., 2010).
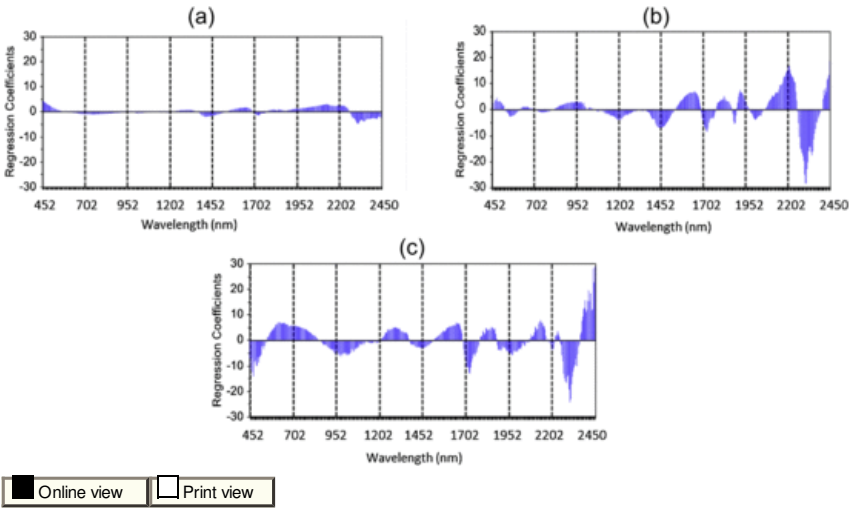
Fig. 5 Plots of regression coefficients vs. wavelength after partial least squares (PLS) regression analysis by visible and near infrared (VIS–NIR) diffuse reflectance spectroscopy for petroleum-contaminated tropical rainforest soils collected from (a) Baraboo, (b) Bomu 1, and (c) Bomu 2 in Ogoniland, Niger Delta, Nigeria.

4 Conclusions

In this study, results show that soil diffuse reflectance decreased with increasing PAH concentration. Positive regression coefficients around 1647 nm show a link to PAH. Additionally, the quality of generalized PLS models for PAH predictions ranged from good to excellent (validation r^2 = 0.77–0.89, RMSEP = 1.16–1.95 mg kg^{-1}, and RPD = 1.86–3.12). This demonstrates the possibility of using VIS–NIR spectroscopy and PLS regression analysis for rapid quantitative determination of PAH in petroleum-contaminated tropical rainforest soils in the Niger Delta province of Nigeria. However, it should be pointed out that the extrapolation of the model is limited to the three sites investigated in this study namely Baraboo, Bomu 1, and Bomu 2 in Ogoniland, Niger Delta region of Nigeria. To set up a model based on VIS–NIR spectroscopy for general application in the Niger Delta province of Nigeria, we recommend the use of larger dataset covering both the concentration range and all the other sources of variability in oil spill sites in the Niger Delta region.

Acknowledgements

References

Askari K. and Pollard S.J.T., The UK Approach for Evaluating Human Health Risks from Petroleum Hydrocarbons in Soils, Science Report P5-080/TR32005, Environment Agency; Rio House, Waterside Drive, Aztec West, Almondsbury, Bristol, UK, 1–22.

Aske N., Kallevik H. and Sjoblom J., Determination of saturate, aromatic, resin, and asphaltenic (SARA) components in crude oils by means of infrared and near-infrared spectroscopy, Energy Fuels **15**, 2001, 1304–1312.

Brassington K.J., Pollard S.J.T. and Coulon F., Weathered hydrocarbon wastes: a risk management primer, In: Timmis K.N., (Ed), Handbook of Hydrocarbon and Lipid Microbiology, 2010, Springer-Verlag; Berlin Heidelberg, 2488–2499.

Bray J.G., Viscarra Rossel R.A. and McBratney A.B., Diagnostic screening of urban soil contaminants using diffuse reflectance spectroscopy, In: Viscarra Rossel R.A., McBratney A.B. and Minasny B., (Eds.), Proximal Soil Sensing, 2010, Springer-Verlag; Berlin Heidelberg, 191–199.

CAMO Software, The Unscrambler® X Version 10.2 User's Guide, 2012, CAMO Software AS; Nedre Vollgate, Oslo, Norway.

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

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

Clark R.N., King T.V.V., Klejwa M., Swayze G. and Vergo N., High spectral resolution reflectance spectroscopy of minerals, *J. Geophys. Res.* **95**, 1990, 12653–12680.

Department of Petroleum Resources, Environmental Guidelines and Standards for the Petroleum Industry in Nigeria (EGASPIN), 2002, Ministry of Petroleum and Natural Resources; Abuja, Nigeria, 314.

Forrester S., Janik L. and McLaughlin M., An infrared spectroscopic test for total petroleum hydrocarbon (TPH) contamination in soils, In: *Proceedings of the 19th World Congress of Soil Science, Soil Solutions for a Changing World, Brisbane, Australia, August 1–6,* 2010, 13–16.

Forrester S., In-situ determination of total petroleum hydrocarbon (TPH) contamination: a quick infrared spectroscopic test for TPH at contaminated sites, In: *Poster Presentation at the 19th World Congress of Soil Science, Soil Solutions for a Changing World, Brisbane, Australia, August 1–6,* 2010.

Forrester, S.F., Janik, L.J., McLaughlin, M.J., 2011. Method of Contaminant Prediction. PCT Patent WO/2011/035391. Date issued: 31 March.

Fuller M.P. and Griffiths P.R., Diffuse reflectance measurements by infrared Fourier transform spectrometry, *Anal. Chem.* **50** (13), 1978, 1906–1910.

Graham K.N., Evaluation of Analytical Methodologies for Diesel Fuel Contaminants in Soil, (M.Sc. thesis)1998, The University of Manitoba; Canada, (Unpublished results).

Harris D.C., Quantitative Chemical Analysis, eighth ed., 2010, W.H. Freeman and Company; New York.

Hites R.A. and Gschwend P.M., The ultimate fate of polycyclic aromatic hydrocarbons in marine and lacustrine sediments, In: Cooke M., Dennis A.J. and Fisher F.L., (Eds.), *Poly-nuclear Aromatic Hydrocarbons – Physical and Biological Chemistry,* 1982, Bettelle Press; Columbus, Ohio, USA, 357–365.

Kuang B. and Mouazen A.M., Effect of spiking strategy and ratio on calibration of on-line visible and near infrared soil sensor for measurement in European farms, *Soil Till. Res.* **128**, 2013, 125–136.

Maleki M.R., Mouazen A.M., Ramon H. and De Baerdemaeker J., Optimization of soil VIS–NIR sensor-based variable rate application system of soil phosphorus, *Soil Till. Res.* **94**, 2007, 239–250.

Malley D.F., Graham K.N. and Webster G.R.B., Analysis of diesel contaminated soils by near-infrared reflectance spectroscopy and solid phase micro extraction-gas chromatography, *J. Soil Contam.* **8**, 1999, 481–489.

Mitra S., Sample Preparation Techniques in Analytical Chemistry, 2003, Wiley and Sons, Inc., Publication; Hoboken, NJ, USA.

Mouazen A.M., De Baerdemaeker J. and Ramon H., Towards development of on-line soil moisture content sensor using a fibre-type NIR spectrophotometer, *Soil Till. Res.* **80**, 2005, 171–183.

Mouazen A.M., Maleki M.R., De Baerdemaeker J. and Ramon H., On-line measurement of some selected soil properties using a VIS–NIR sensor, *Soil Till. Res.* **93** (1), 2007a, 13–27.

Mouazen A.M., Karoui R., Deckers J., De Baerdemaeker J. and Ramon H., Potential of visible and near-infrared spectroscopy to derive colour groups utilizing the Munsell soil colour charts, *Biosyst. Eng.* **97**, 2007b, 131–143.

Naes T., Isaksson T., Fearn T. and Davies T., A User Friendly Guide to Multivariate Calibration and Classification, 2002, NIR Publications; Chichester, UK.

Niger Delta Environmental Survey, Background and Mission: Briefing Note 1, 1995, Publication of the Steering Committee, NDES; Falomo, Lagos, Nigeria, 1–7.

Okoro D. and Ikolo A.O., Sources and compositional distribution of polycyclic aromatic hydrocarbons in soils of Western Niger Delta, *J. Appl. Sci. Technol.* **12** (1–2), 2007, 35–40.

Okparanma R.N. and Mouazen A.M., Determination of total petroleum hydrocarbon (TPH) and polycyclic aromatic hydrocarbon (PAH) in soils: a review of spectroscopic and non-spectroscopic techniques, *Appl. Spectrosc. Rev.* **48** (6), 2013a, 458–486.

Okparanma R.N. and Mouazen A.M., Combined effects of oil concentration, clay and moisture contents on diffuse reflectance spectra of diesel-contaminated soils, *Water Air Soil Pollut.* **224** (5), 2013b, 1539–1556.

Okparanma R.N. and Mouazen A.M., Visible and near-infrared spectroscopy analysis of a polycyclic aromatic hydrocarbon in soils, *ScientificWorldJournal* 2013c, (Article in review ~~submitted for publication~~).

Osborne B.G., Fearn T. and Hindle P.H., Practical NIR Spectroscopy – with Applications in Food and Beverage Analysis, second ed., 1993, Longman Group UK Limited; England.

Patnaik P., Handbook of Environmental Analysis: Chemical Pollutants in Air, Water, Soil and Solid Wastes, 1997, CRC Press Inc.; Boca Raton, FL.

Risdon G.C., Pollard S.J.T., Brassington K.J., McEwan J.N., Paton G.I., Semple K.T. and Coulon F., Development of an analytical procedure for weathered hydrocarbon contaminated soils within a UK risk-based framework, *Anal. Chem.* **80** (18), 2008, 7090–7096.

Schwartz G., Ben-Dor E. and Eshel G., Quantitative analysis of total petroleum hydrocarbons in soils: comparison between reflectance spectroscopy and solvent extraction by 3 certified laboratories, *Appl. Environ. Soil Sci.* **2012**, 2012, 1–11.

SPDC, Environmental Impact Assessment (EIA) of Rumuekpe (OML 22) and Etelebou (OML 28) Area 3 Dimensional Seismic Survey, 2006, Publication of the Shell Petroleum Development Company (SPDC) of Nigeria Limited; Abuja, Nigeria, Available: http://www.shell.com.ng/environment-society/environment-impact-assessments.html, (accessed 25.02.13.).

Stenberg B., Effects of soil sample pretreatments and standardized rewetting as interacted with sand classes on VIS–NIR predictions of clay and soil organic carbon, *Geoderma* **158**, 2010, 15–22.

Tanee F.B.G. and Albert E., Post-remediation assessment of crude oil polluted site at Kegbara-Dere Community, Gokana L.G.A. of Rivers State, Nigeria, *Bioremediation Biodegradation* **2**, 2011, DOI: 10.4172/2155-6199.1000122, (accessed 26.02.13.).

UNEP, Environmental Assessment of Ogoniland, 2011, United Nations Environment Programme (UNEP); Nairobi, Kenya, Available: http://www.unep.org, (accessed 29.10.12.).

USEPA, Method 8270: Semi Volatile Organic Compounds (SVOCS) by Gas Chromatography–Mass Spectrometry (GC–MS), SW-846 Manual1999, United States Environmental Protection Agency (USEPA), Government Printing Office; Washington, DC, USA.

Viscarra Rossel R.A., Walvoort D.J.J., McBratney A.B., Janik L.J. and Skjemstad J.O., Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties, *Geoderma* **131**, 2006a, 59–75.

Viscarra Rossel R.A., McGlynn R.N. and McBratney A.B., Determining the composition of mineral-organic mixes using UV–vis–NIR diffuse reflectance spectroscopy, *Geoderma* **137**, 2006b, 70–82.

Wang Z. and Fingas M., Differentiation of the sources of spilled oil and monitoring of the oil weathering process using gas chromatography–mass spectroscopy, *J. Chromatogr. A* **712**, 1995, 321–343.

Whalley W.R. and Stafford J.V., Real-time sensing of soil water content from mobile machinery: options for sensor design, *Comput. Electron. Agr.* **7**, 1992, 269–358.

Willey R.R., Fourier transform infrared spectrophotometer for transmittance and diffuse reflectance measurements, *Appl. Spectrosc.* **30** (6), 1976, 593–601.

Williams P.C. and Sobering D.C., Attempts at standardization of hardness testing of wheat. II. The near-infrared reflectance method, *Cereal Foods World* **31**, 1986, 417–420.

Workman J., Jr. and Weyer L., , Practical Guide to Interpretive Near-infrared Spectroscopy, 2008, CRC Press, Taylor and Francis Group; Boca Raton, FL, USA.

Yunker M.B. and MacDonald R.W., Composition and origins of polycyclic aromatic hydrocarbons in the Mackenzie River and on the Beaufort Sea Shelf, *J. Arct. Inst. N. Am.* **48** (2), 1995, 118–129.

Highlights

- We model NIR diffuse reflectance spectra for PAH prediction in contaminated soils.
- Soil diffuse reflectance decreases with increasing PAH concentration.
- Mechanism of prediction relies on co-variation of PAH with other soil properties.
- Positions of important wavelengths are largely similar for studied sites.
- Positive regression coefficients around 1647 nm show a link to PAH.

Queries and Answers

Query: Refs. "Pasquini, 2003, Martens and Naes (1989), Mayer, A. S. (2008)" are cited in the text but not provided in the reference list. Please provide them in the reference list or delete these citations from the text.

Answer:

Pasquini, C., Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. *J. Braz. Chem. Soc.*, 14(2), 2003, 198–219.

Martens, H. and Naes, T., Multivariate Calibration, second ed., 1989, John Wiley and Sons; Chichester, UK.

Mayer, A.S., Lecture Note, GE3850: Geohydrology, Department of Civil and Environmental Engineering, Michigan Technology University, MI, 2008, Unpublished data.

Query: Please update Okparanma and Mouazen, submitted for publication.

Answer: Article is in review.

Query: Please confirm that given names and surnames have been identified correctly.

Answer: Confirmed ok

Analysis of petroleum-contaminated soils by diffuse reflectance spectroscopy and sequential ultrasonic solvent extraction-gas chromatography

Okparanma, R. N.