Science of the Total Environment, Vol. 626, 1 June 2018, pp. 1108-1120
DOI:10.1016/j.scitotenv.2018.01.122

Evaluation of vis-NIR reflectance spectroscopy sensitivity to weathering for enhanced assessment of oil contaminated soils

Abstract

This study investigated the sensitivity of visible near-infrared spectroscopy (vis-NIR) to discriminate between fresh and weathered oil contaminated soils. The performance of random forest (RF) and partial least squares regression (PLSR) for the estimation of total petroleum hydrocarbon (TPH) throughout the time was also explored. Soil samples (n=13) with 5 different textures of sandy loam, sandy clay loam, clay loam, sandy clay and clay were collected from 10 different locations across the Cranfield University's Research Farm (UK). A series of soil mesocosms was then set up where each soil sample was spiked with 10 ml of Alaskan crude oil (equivalent to 8450 mg/kg), allowed to equilibrate for 48 h (T2d) and further kept at room temperature (21^{o}C). Soils scanning was carried out before spiking (control TC) and then after 2 days (T2d) and months 4 (T4m), 8 (T8m), 12 (T12m), 16 (T16m), 20 (T20m), 24 (T24m), whereas gas chromatography mass spectroscopy (GC-MS) analysis was performed on T2d, T4m, T12m, T16m, T20m, and T24m. Soil scanning was done simultaneously using an AgroSpec spectrometer (305 to 2200 nm) (tec5 Technology for Spectroscopy, Germany) and Analytical Spectral Device (ASD) spectrometer (350 to 2500 nm) (ASDI, USA) to assess and compare their sensitivity and response against GC-MS data. Principle component analysis (PCA) showed that ASD performed better than tec5 for discriminating weathered versus fresh oil contaminated soil samples. The prediction results proved that RF models outperformed PLSR and resulted in coefficient of determination (R^2) of 0.92, ratio of prediction deviation (RPD) of 3.79, and root mean square error of prediction (RMSEP) of 108.56 mg/kg. Overall, the results demonstrate that vis-NIR is a promising tool for rapid site investigation of weathered oil contamination in soils

and for TPH monitoring without the need of collecting soil samples and lengthy hydrocarbon extraction for further quantification analysis.

Keywords: visible near-infrared diffuse reflectance spectroscopy; weathering; hydrocarbon; land management; chemometrics.

1. Introduction

Globally petroleum hydrocarbons are used widely but their uses have caused contamination of soil, water and air mainly during oil production activities, storage and distribution of petroleum products and spillage accidents (ATSDR, 1999). Petroleum hydrocarbons are a complex mixture of aliphatic and aromatic hydrocarbon compounds, among which certain compounds can pose a significant risk to human health and or the environment (Wartini et al., 2017; Cipullo et al., 2018). While there have been a great deal of studies that have been carried out on developing and validating analytical framework for characterizing and quantifying petroleum hydrocarbons in soil matrices, they often require soil sampling and then rely on lengthy extraction procedure that needs to be carried out in the laboratory (Paiga et al., 2012; Douglas et al., 2017). There is a need for the rapid measurement of petroleum hydrocarbons in soil to allow better and swifter site characterization and increased confidence in prioritizing remediation actions. Most importantly, the concept of taking 'the lab to the field' for measuring hydrocarbon contamination in soil without compromising data quality and information needs to be demonstrated (Horta et al., 2015; Douglas et al., 2017). To this end, field-based techniques offer rapid, non-destructive and cost-effective means of defining levels and distribution of petroleum hydrocarbons on-site. They also provide real-time monitoring data useful for initial site assessment and inform future sampling campaign for detailed risk assessment of the contaminated sites. However, one drawback of

field-based techniques is that they often fail to determine and quantify the entire range of petroleum hydrocarbons, the aliphatic and aromatic hydrocarbon fractions, in soil (Douglas et al., 2017).

Once petroleum hydrocarbon are discharged to the environment, they undergo physical, chemical and biological processes that further alter their composition, toxicity, availability, and distribution in the environment. Such weathering (degradation) processes include adsorption, volatilization, dissolution, biotransformation, photolysis, oxidation, and hydrolysis (Brassington et al., 2007; Jiang et al., 2016). These processes shift the chemical composition of the hydrocarbons towards recalcitrant, asphaltenic products of increased hydrophobicity (Coulon et al., 2010). Weathered hydrocarbons are highly complex mixture and are known soil contaminants, which in the face of 40 years of petroleum research, are still not adequately understood or appropriately characterize for informing contaminated land risk categorization (Coulon et al., 2010). Recently, research has been intensified in developing robust analytical technique for the identification of weathered hydrocarbons, which are the main sources of the organic carcinogens or suspected carcinogens that drive quantitative risk assessment (e.g., Benz[a]anthracene, benzo[a]pyrene, chrysene) at oil-contaminated sites (Environment Agency, 2005). Analytical methods including gas chromatography mass spectroscopy (GC-MS), gas chromatography coupled with flame ionization detector (GC-FID), gravimetric analysis, and infrared spectroscopy are available for analyzing weathered hydrocarbons; however, the choice of technique may be influenced by the risk assessment being used during the remediation of contaminated land (API, 2001).

Table 1: Previous results of visible near-infrared (vis-NIR) technology performance for the analysis of petroleum-contaminated soils at field-scale.

Targeted analyte	N	Spectral range (nm)	Modeling method	Statistical parameters	References
TPH	85	350-2500	RF	R^2=0.68, RMSEP=69.64 mg/kg, RPD=1.85	Douglas et al., 2018a
			PLSR	R^2=0.54, RMSEP=75.86 mg/kg, RPD=1.51	
PAH	85	350-2500	RF	R^2=0.71, RMSEP=0.99 mg/kg, RPD=1.99	Douglas et al., 2018b
			PLSR	R^2=0.56, RMSEP=1.12 mg/kg, RPD=1.55	
TPH	108	350-2500	PSR	R^2=0.70, RMSEP=0.75 mg/kg, RPD=1.86	Chakraborty et al., 2015
			RF	R^2=0.61, RMSEP=0.70 mg/kg, RPD=1.64	
			PLSR	R^2=0.73, RMSEP=0.59 mg/kg, RPD=1.96	
TPH	164	350-2500	FD (PSR)	R^2=0.87, RMSEP=0.528 mg/kg, RPD=2.78	Chakraborty et al., 2014
			SNV-DT (PSR)	R^2=0.80, RMSEP=0.66 mg/kg, RPD=2.21	
			FD (RF)	R^2=0.58, RMSEP=0.95 mg/kg, RPD=1.56	
			SNV-DT (RF)	R^2=0.58, RMSEP=0.94 mg/kg, RPD=1.57	
PAH	137	350-2500	PLSR	R^2=0.89, RMSEP=1.16 mg/kg, RPD=3.12	Okparanma et al., 2014
PAH	150	350-2500	PLSR	R^2=0.89, RMSEP=0.20 mg/kg, RPD=2.75	Okparanma et al., 2013b
TPH	205	2000-2500	PLSR	R^2=0.63, RMSEP=5224 mg/kg, RPD=1.5	Forrester et al., 2013
TPH	45	1560-1800	PLSR	R^2=0.94, RMSECV=1590 mg/kg, Bias=0.003	Hauser et al., 2013
TPH	46	350-2500	PLSR	R^2=0.64, RMSEP=0.34 mg/kg, RPD=1.70	Chakraborty et al., 2010
TPH	26	1100-2498	SMLR	R^2=0.71, SEP=770 mg/kg, RPD=1.80	Malley et al., 1999

N=number of samples, TPH=total petroleum hydrocarbon, PAH=polycyclic aromatic hydrocarbon, R^2 = coefficient of determination, RMSEP = root mean square error of prediction, SEP= standard error of prediction, RPD = residual prediction deviation, RF=random forest, SMLR, = stepwise multiple linear regression, PLSR=partial least square regression, PSR=penalized spline regression, FD = first derivative preprocessing, SNV-DT= standard normal variate preprocessing followed by detrending

Reflectance spectroscopy, including visible and near-infrared (vis-NIR) or mid-infrared (MIR) spectroscopy, has been shown to be a suitable rapid acquisition method for the measurement of hydrocarbon concentration in soil without the need of any sample preparation (Chakraborty et al., 2010; Okparanma and Mouazen, 2013a; Horta et al., 2015; Douglas et al., 2018a). More details on previous works on the use of vis-NIR spectroscopy for quantifying hydrocarbons in soils can be found in Table 1. However, to the best of our knowledge, the application of vis-NIR-based techniques to differentiate between freshly contaminated *versus* weathered crude oil contaminated soils has not been investigated. Furthermore, no attempts to implement the vis-NIR

spectroscopy to quantify the total petroleum hydrocarbon (TPH) in soil, across different stages of weathering can be found in the literature.

The objectives of this study were (i) to investigate the sensitivity of two portable vis-NIR spectrophotometers (ASD and tec5) for the discrimination between weathered and fresh oil spill in soils using principal component analysis (PCA), and (ii) to quantify TPH in these soils during weathering, using partial least squares regression (PLSR) and random forest (RF) modeling methods.

2. Materials and methods

2.1 Study area and soil sampling

A total of thirteen (n=13) surface soil samples (0-15 cm) with approximately 5 kg per sample were collected using a shovel from 10 sites located in Bedfordshire, namely, Avenue, Downings, Orchard, Mound, Wood, Copse, Ivy ground, Near Warden, Showground, and Sandpit; all from the Cranfield University's Research Farm, Bedfordshire, UK (Fig. 1). Samples were taken with Ziploc bags to the laboratory and stored in the freezer at 4 ^{o}C prior to utilization. Two and three samples were collected for Avenue and Ivy ground fields, respectively, while one samples was collected from each of the remaining five fields. The collected soil samples were subjected to soil physical and chemical analyses. The soil moisture content (MC) was measured by oven-drying soil samples at 105 ± 5^{o}C for 24 h. Soil pH was measured following the Standard Operating Procedure (SOP) of the British Standard BS ISO 10390:2005; the total organic carbon (TOC) was determined using a Vario III Elemental Analyser using SOP based on British Standard BS 7755 Section 3.8: 1995 and the particle size was determined using SOP based on British Standard BS 7755 Section 5.4:1995.

Fig. 1. Location of the study area and sampling points collected from 10 sites in Bedfordshire, UK.

2.2 Mesocosms setup

Using 1 kg soil, 13 soil mesocosms (representing all the 13 samples) were set up. Each soil sample was spiked with 10 ml of Alaskan crude oil (equivalent to 8450 mg/kg) and allowed to equilibrate at room temperature (21 ^{o}C) for 48 h. Vis-NIR scanning was performed on pristine soil (control (TC) - pristine samples dried at room temperature to reduce moisture effect) and then after 2 days (T2d) and months 4 (T4m), 8 (T8m), 12 (T12m), 16 (T16m), 20 (T20m), 24 (T24m); whereas gas chromatography mass spectroscopy (GC-MS) analysis was performed on

T2d, T4m, T12m, T16m, T20m, and T24m. Therefore, data of T8m was excluded from the quantitative analysis of TPH.

2.3 Optical measurement and spectra preprocessing

Soil spectral measurements were done in the laboratory using two vis-NIR spectrophotometers, namely, an AgroSpec vis-NIR spectrometer with a spectral range of 305-2200 nm (tec5 Technology for Spectroscopy, Germany) and an ASD LabSpec2500® (Analytical Spectral Devices, Inc., USA), which covers a spectral range of 350–2500 nm. Both spectrometers are portable, but use different detectors; ASD uses monochromatic detector while tec5 is equipped with a diode array detector.

Spectral measurement by ASD LabSpec2500® spectrometer in this study followed the protocols described by Douglas et al. (2018a). Before scanning, samples were air-dried in order to eliminate the effect of moisture content on soil spectral analysis (Mouazen et al., 2006). After removal of all plants and pebble materials, three subsamples were prepared from each soil sample; these were placed into 3 different Petri dishes (1 cm height x 5.6 cm in diameter), and the surface was smoothened gently with a spatula before scanning (Mouazen et al., 2005). This was done to achieve optimal diffuse reflection and, thus, a good signal-to-noise ratio. A high-intensity probe was used for scanning of soil samples, which has a built-in light source made of a quartz-halogen bulb of 2727 °K. The light source and detection fibres are assembled in the high-intensity probe enclosing a 35° angle. The device was calibrated using a 100 % white Spectralon disc before use, and after every 30 min. The spectral measurements were made in the dark in order to both, control the illumination conditions and reduce the effects of stray light. The three replicates of each sample were scanned at three different spots, and an average spectrum was

obtained for further analysis. A total of 10 scans were acquired from each replicate, and the average spectrum of the three replicates was considered as the sample spectrum.

Prior to multivariate analysis, three standardized spectral pre-treating approaches (including maximum normalization, first derivative, and smoothing) were carried out using R software (R Core Team, 2013). Maximum normalization divides each row (spectrum) by its maximum absolute value to achieve an even distribution of the variances; the first derivative removes the baseline shift to improve the accuracy of quantification (Okparanma et al. 2014; Demetriades-Shah et al., 1990); and smoothing reduces the impact of noise (Okparanma and Mouazen, 2013b). These routines were aimed at keeping all useful chemical and physical information in the spectra for analysis.

2.4 Gas chromatography and peak integration

Chemical analysis for TPH concentration was carried out using sequential ultrasonic solvent extraction-gas chromatography (SUSE-GC) as described by Risdon et al. (2008) with some modifications. Briefly, 5 g of soil sample was mixed with 20 ml of dichloromethane (DCM): hexane (Hex) solution (1:1, v/v) and shaken for 16 h at 150 oscillations per min over 16 h; and finally sonicated for 30 min at 20 °C. After centrifugation, extracts were cleaned on Florisil® columns by elution with hexane. Deuterated alkanes and polycyclic aromatic hydrocarbons (PAHs) internal standards were added to extracts at appropriate concentrations. The final extract was diluted (1:10) for GC-MS analysis. Deuterated alkanes ($C10^{d22}$, $C19^{d40}$ and $C30^{d62}$) and PAH (naphthalene ^{d8}, anthracene ^{d10}, chrysene ^{d12} and perylene ^{d12}) internal standards were added to extracts at 0.5 µg ml^{-1} and 0.4 µg ml^{-1}, respectively. Aliphatic hydrocarbons and PAHs were identified and quantified using an Agilent 5973N GC-MS operated at 70 eV in positive ion mode. The column used was a Zebron fused silica capillary column (30 x 0.25 mm internal

diameter, Phenomenex) coated with 5MS (0.25 µm film thickness). Splitless injection with a sample volume of 1 µL was applied. The oven temperature was increased from 60 °C to 220 °C at 20 °C min^{-1} then to 310 °C at 6 °C min^{-1} and held at this temperature for 15 min. The mass spectrometry was operated using the full scan mode (range *m/z* 50-500) for quantitative analysis of target alkanes and PAHs. For each compound, quantification was performed by integrating the peak at specific *m/z* using auto-integration method with Mass Selective Detector (MSD) ChemStation software. External multilevel calibrations were carried out for both alkanes and PAH quantification ranging from 0.5 to 2500 µg ml^{-1} and from 1 to 5 µg ml^{-1}, respectively. For quality control, a 500 µg ml^{-1} diesel standard solution (ASTM C_{12}-C_{60} quantitative, Supelco) and mineral oil mixture Type A and B (Supelco) were analyzed every 20 samples. The variation of the reproducibility of extraction and quantification of soil samples were determined by successive injections (n=7) of the same sample and estimated to ± 8%. In addition, duplicate reagent control and reference material were systematically used. The reagent control was treated following the same procedure as the samples without adding soil sample. The reference material was an uncontaminated soil of known characteristics, and was spiked with a diesel and mineral oil standard at a concentration equivalent to 16,000 mg kg^{-1}. Relative standard deviation (RSD) values for all the soils was <10%. From the results obtained for alkanes and PAHs, TPH was obtained for each sample, and further used for modelling purposes.

2.5 Multivariate analyses

2.5.1 Principal component analysis (PCA)

PCA was used for qualitative vis-NIR discrimination of soil samples based on the spectral properties of the different contaminated weathering groups. PCA is a multivariate technique that

reduces the dimensionality of large multivariate datasets. PCA helps to transform the wavelengths (independent variables) into principle components (PCs). Plotting the PCs enables one to examine interrelationships among different variables, and detect and interpret sample patterns, groupings, similarities, or differences (Martens and Naes, 1989; Mouazen et al., 2006). The preprocessed spectra have been used in the PCA; the results showed a similarity map of principal PCs, as well as the loadings that can be used to investigate the significant wavebands for hydrocarbons. The PCA was performed using FactorMine R-package (R Core Team, 2013).

2.5.2 Quantitative assessment of TPH using PLSR and RF methods

The preprocessed vis-NIR soil spectra for both ASD and tec5 spectrophotometers coupled with the reference laboratory TPH measured by SUSE-GC were used to develop calibration models for quantifying TPH through 2 years weathering period. The total number of samples used for both PLSR and RF modelling were 78, obtained from 13 soil samples scanned at six occasions through 24 months. Sixty (n=60) samples were selected for calibration while eighteen (n=18) for prediction (validation). The same calibration and validation datasets used in PLSR were utilized for RF analysis. The selection of the samples in the calibration and prediction set was done based on the Kennard-Stone algorithm (Kennard and Stone, 1969). Two groups of calibration models for TPH were developed, one for tec5 and the second one for ASD spectral data. The intension was to evaluate the effect of the spectral range of the prediction accuracy of TPH in the soil during 2 years weathering period.

PLSR is a commonly used multivariate regression technique available in standard statistical and chemometric software. It is a combination of both the independent variables (TPH values) and the dependent variables (wavelengths), which are used as regression generators for the

independent variables. In this study, we use PLSR with leave-one-out cross validation (LOOCV) to develop TPH prediction model, using pls package (R Core Team, 2013). It is documented that LOOCV annul the possible effect of model under- or over-fittings (Efron and Tibshirani, 1993).

Random forest is a nonparametric and nonlinear classification and regression algorithm using assembly learning strategy that integrates hundreds of individual trees (Breiman, 2001). A bootstrap sample is first drawn from the training dataset to build each tree. At each node split, the candidate set of the regressor is a random subset of all the regressors. The final prediction of a new observation is the average of the predicted values from all the trees in the forest. The tuning parameters of RF have been defined based on function implemented in the R software package and were set to 500, 2, and 2 for the number of trees (*ntree),* the number of predictor variables used to split the nodes at each partitioning (*mtry*), and the minimum size of the leaf (*nodesize*), respectively. Models were developed with R program using the software package randomForest Version 4.6-12 (Liaw and Wiener, 2015), based on Breiman and Cutler's Fortran code (Breiman, 2001).

2.6 Evaluation of model performance

The performance of TPH prediction models was assessed by means of three parameters: (i) the coefficient of determination in prediction R^2, (ii) root mean square error of prediction (RMSEP), and (iii) residual prediction deviation (RPD), which is a ratio of standard deviation (SD) to RMSEP. In this study, we adopted the model classification criterion of Viscarra Rossel et al. (2006): RPD < 1.0 indicates very poor model predictions, 1.0 ≤ RPD < 1.4 indicates poor, 1.4 ≤ RPD < 1.8 indicates fair, 1.8 ≤ RPD < 2.0 indicates good, 2.0 ≤ RPD < 2.5 indicates very good,

and excellent if RPD > 2.5. In general, a best model performance would have the highest values of R^2 and RPD, and smallest value of RMSEP.

3. Results and discussion

3.1 Soil physiochemical properties

Soil physio-chemical properties (*viz*. partial size distribution, TOC, and MC) of the different soil samples are presented in Table 2. Clay content ranged between 14% and 57%, silt between 15% and 27%, and sand between 16% and 63%. However, examining the soil texture type according to the United State Department of Agriculture (USDA) classification system, indicates the majority of soils in the study fields are on the heavy side of the texture triangle. TOC was high with minimum and maximum of 1.62 and 4.48%, respectively. Results indicated a high variation in soil texture and TOC among the soil samples. Apart from soil MC, soil texture is the other main factor to affect accuracy of vis-NIR spectroscopy. However, since soil samples were scanned after air drying, the effect of MC was excluded from spectral analysis. It has been reported that small particle size (high clay content) can result in a better model performance (Fontán et al., 2010) of soil organic carbon, whereas prediction was reported to be to be less accurate in coarse soil textures (Stenberg, 2010). Since the majority of soil textures of the samples analyzed in this work were on the heavy side of the texture triangle, the similarity in texture is assumed to have minor effect on prediction accuracy of TPH.

Table 2: Soil physio-chemical properties of 13 surface soil samples (0-15 cm) collected from ten different locations across the Cranfield University's Research Farm, Bedfordshire, UK.

Location name	Sample No.	Clay %	Silt %	Sand %	TOC %	Texture
Avenue	1	17	20	63	2.02	Sandy loam
	2	30	19	51	1.67	
Downings	3	28	19	53	2.3	Sandy clay loam
Orchard	4	33	26	41	2.32	Clay loam
Mound	5	16	21	63	1.96	Sandy loam
Wood	6	42	25	33	2.28	Clay
Copse	7	38	26	36	2.7	Clay loam
	8	57	27	16	4.48	Clay
Ivy ground	9	57	27	16	4.48	Clay
	11	57	27	16	4.48	Clay
Near warden	10	57	25	18	3.1	Clay
Showground	12	24	17	59	1.87	Sandy clay loam
Sand pit	13	14	15	71	1.62	Sandy loam

TOC=total organic carbon.

3.2. Spectral data analysis

Illustrative raw air dry soil spectra and pre-processed soil spectra changes overtime are presented in Fig. 2 (note that only T2d, T12m and T20m are shown for clarity). In both Fig. 2a and c, the control soil (TC) reflects higher than the contaminated soils or, in other words, absorb less light energy due to the lighter color of samples without oil added.

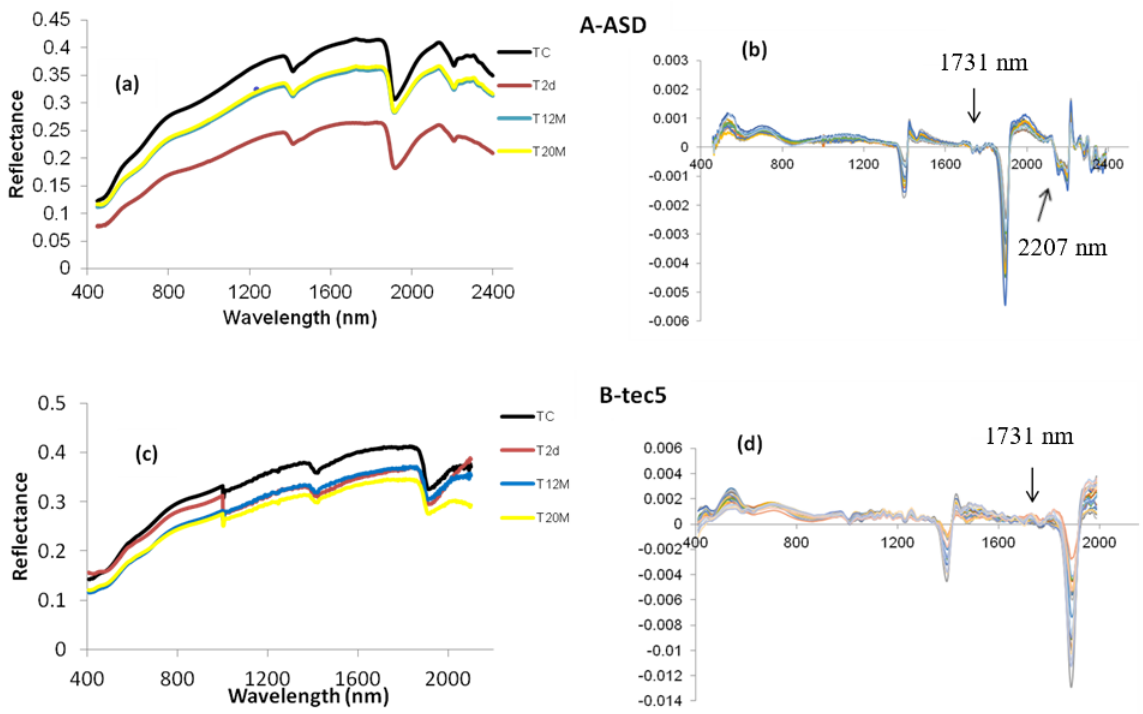

Fig 2. Illustrative example of visible and near infrared (vis-NIR) soil spectra overtime: Control pristine soil (TC), and contaminated soil after 48 hours (T2d), 12 months (T12m) and 20 months (T20m); Panels a & b showed raw spectra and preprocessed spectra obtained with ASD spectrometer; Panels c & d showed raw spectra and preprocessed spectra obtained with the tec5 spectrophotometer.

It is clearly demonstrated that reflectance decreased or absorption increased when adding crude oil, due to the darker color. Among the contaminated soils, the spectral reflectance increased (i.e., less absorbance) as weathering of hydrocarbons in soils progresses. Thus, T2d samples had the highest absorbance, and this decreased with weathering time. In terms of equipment performance, a better discrimination between groups' average spectra was achieved with the ASD spectrometer compared to tec5 spectrometer (Fig. 2).

The behavior of control and contaminated spectra observed herein is in line with the conclusions drawn by Hoerig et al. (2001). Both ASD and tec5 spectrophotometers captured hydrocarbon features around 1731 nm in the first overtone region (Fig. 2b and 2d), which is linked with TPH. Our result is not far from those identified by other scientists e.g., 1732, 1758 nm (Douglas et al. 2018a), 1752 nm (Chakraborty et al., 2015), 1712, 1752 nm (Okparanma et al., 2014). An absorption band of hydrocarbons around 2207 nm in the combination region (Fig. 2) was also observed in the ASD spectra, a wavelength that is close to those reported by other researchers e.g., 2240 nm by Chakraborty et al. (2015), and 2460 nm by Forrester et al. (2013). The other absorption bands are associated with other soil properties, e.g., water, clay mineralogy, and organic carbon. More details about the hydrocarbon signatures in soils are presented in section 3.2.

3.2. Qualitative discrimination of weathering groups by PCA

In order to examine the variability between spectra of the contaminated soils overtime, spectra were subjected to PCA, with the aim to extract distinctive spectral features that can assemble similar weathered contaminated soils together in distinguished groups. If this can be achieved, we can claim that the vis-NIR spectrometers used in this study can differentiate weathered *vs.* fresh oil spill in soils. An scatter diagram of component score for the first and second principal components (PC-1, PC-2) is shown in Fig 3a for the ASD spectrometer and Fig 3b for the tec5 spectrometer. With the ASD spectrometer, PC1 accounted for 94.50% while PC 2 accounted for 5.10% of variance, with a total of 99.6%. However, a slightly less variance was accounted for by the PCA performed on the tec5 spectra (Fig. 3b), with PC1 accounting for 93.30% and PC2 accounting for 5.12%, which sums up to 98.42% of the total variance. It is noteworthy that the

separation patterns of the various weathering group soils achieved with the two portable vis-NIR instruments are different; with ASD (Fig. 3a) providing the best visual separation in the principal component space. The separation was particularly clear between the non-contaminated (TC) and freshly contaminated samples at T2d, obtained with the ASD spectrometer. Different weathering groups were formed along the PC1 of the ASD-PCA plot, showing different degree of overlap between soil groups of different weathering time, where overlap becomes more evident after month 12 and up to month 24 in Fig. 3a. Soil samples at T2d and T4m are better separated from the remaining weathering groups (Fig. 3a). Few samples from T4m overlapped with those of T2d, whereas one T4m and few T8m samples were in the neighborhood of the T12m and T24m samples. In the case of the T2d and T4m samples, there is less compositional resemblance reflected on different spectral signature, whereas more compositional resemblance exists within the T12m to T24m samples, resulting in smaller spectral differences of the same sample throughout weathering time, and hence the increase of sample overlap. The tec5-PCA plot shows less clear separation between different weathering groups (Fig. 3b) compared to the ASD-PCA plots. Separation here occurs along the diagonal access between PC1 and PC2 (Fig. 3b). It is obvious that TC samples are clearly separated from the other groups, and that more clear overlap exists between the remaining groups compared to the ASD-PCA plots. For example, it is odd to observe that T4m samples are closer to TC samples, in comparison with T2d samples, which were further away from TC samples. Furthermore, samples of T24m and T20m are closer to TC samples than the remaining groups with smaller weathering time (e.g., T4m, T8m, T12m and T16m).

Overall, we can conclude that, the ASD spectrometer provided logical and clearer separation of the different weathering groups and that instrument's sensitivity to weathering reduces overtime

due to the reduction of the TPH concentration (see discussion below). On the other hand, the clear separation observed between the contaminated and TC samples indicate that the two groups are compositionally dissimilar. This is in agreement with the results reported by Chakraborty et al. (2010), who assessed the ability of vis-NIR spectroscopy to distinguish contaminated and non-contaminated soils qualitatively using PCA.

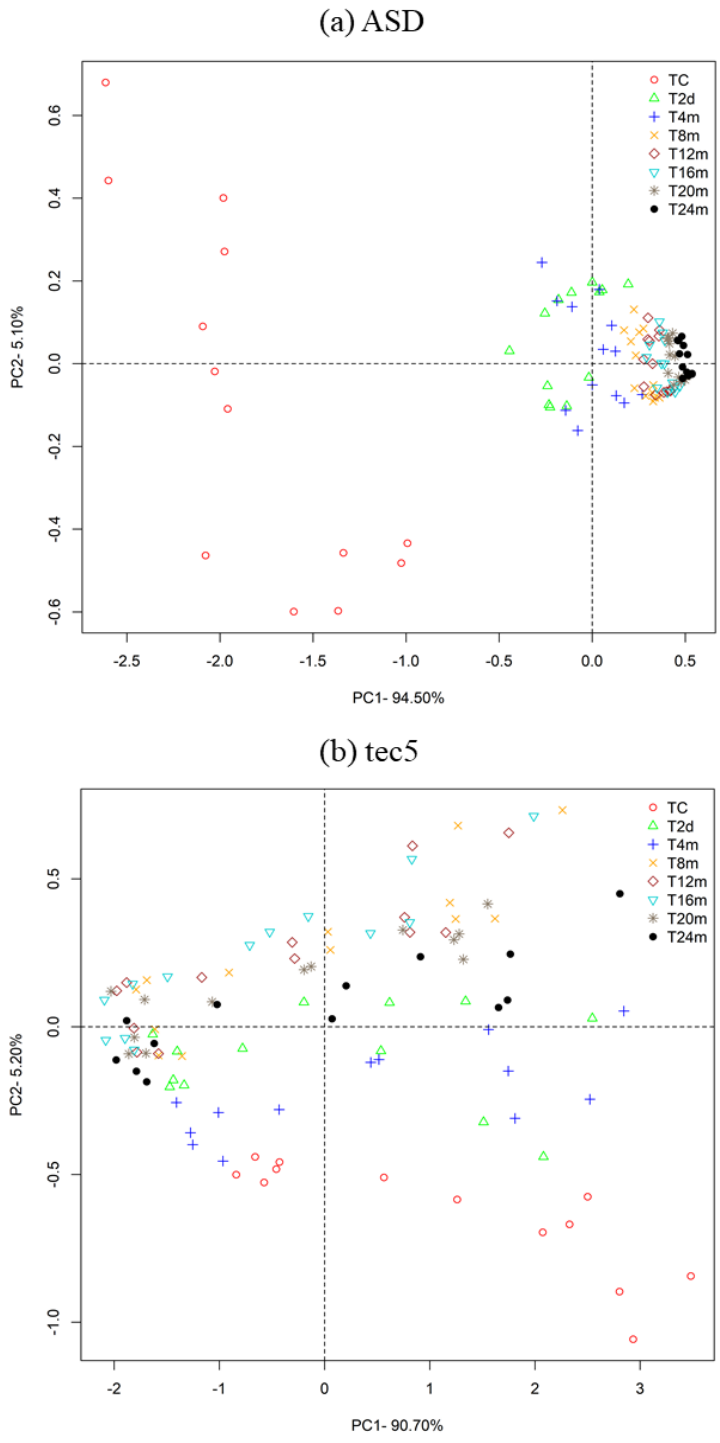

Fig. 3 Principal component analysis of the soil scanning profile overtime obtained using (**a**) ASD and (**b**) tec5 spectrophotometers (TC: control samples (pristine); and contaminated soil samples after 48 h (T2d), and months 4 (T4m), 8 (T8m), 12 (T12m), 16 (T16m), 20 (T20m) and 24 (T24m).

Furthermore, PCA loadings were produced to investigate potential wavelengths associated with diesel originated hydrocarbon contamination (Fig. 4). In the PCA loadings, an absorption minimum was observed at 1730 nm in both ASD and tec5 spectrometers, which is attributed to C-H stretching modes of terminal CH_3 and saturated CH_2 groups linked to TPH in the first overtone region. This result is in line with observations from others researchers (Okparanma et al., 2014; Workman and Weyer, 2008). Furthermore, typical spectral signatures around 1452 nm and 1950 nm were clearly observed in both ASD and tec5 spectrometers. These are associated with the second and first overtones of water absorption around 1450 nm and 1950 nm, previously reported (Mouazen et al., 2005; Mouazen et al., 2006). Absorption features around 2279 and 2340 nm were also observed in ASD spectrometer alone. These are associated with metal-OH bend and O-H stretch combination and characteristic of clay minerals. The results obtained here are similar to those at 2200 and 2300 nm, reported in the literature (Clark et al., 1990; Viscarra Rossel et al., 2006b). The absorption band at 2207 nm can be attributed to either amides (C=O) absorption (Viscarra Rossel and Behrens, 2010) or crude oil spectral signatures (stretch+bend) (Mullins et al., 1992). Furthermore, this band can be linked to the hydrocarbon concentration that can be effective to discriminate between weathering groups (Fig 4a). Therefore, the ASD showed a high capability to discriminate between the weathering group, and this is because its full vis-NIR range spectrum including all the effective waveband associated with hydrocarbons.

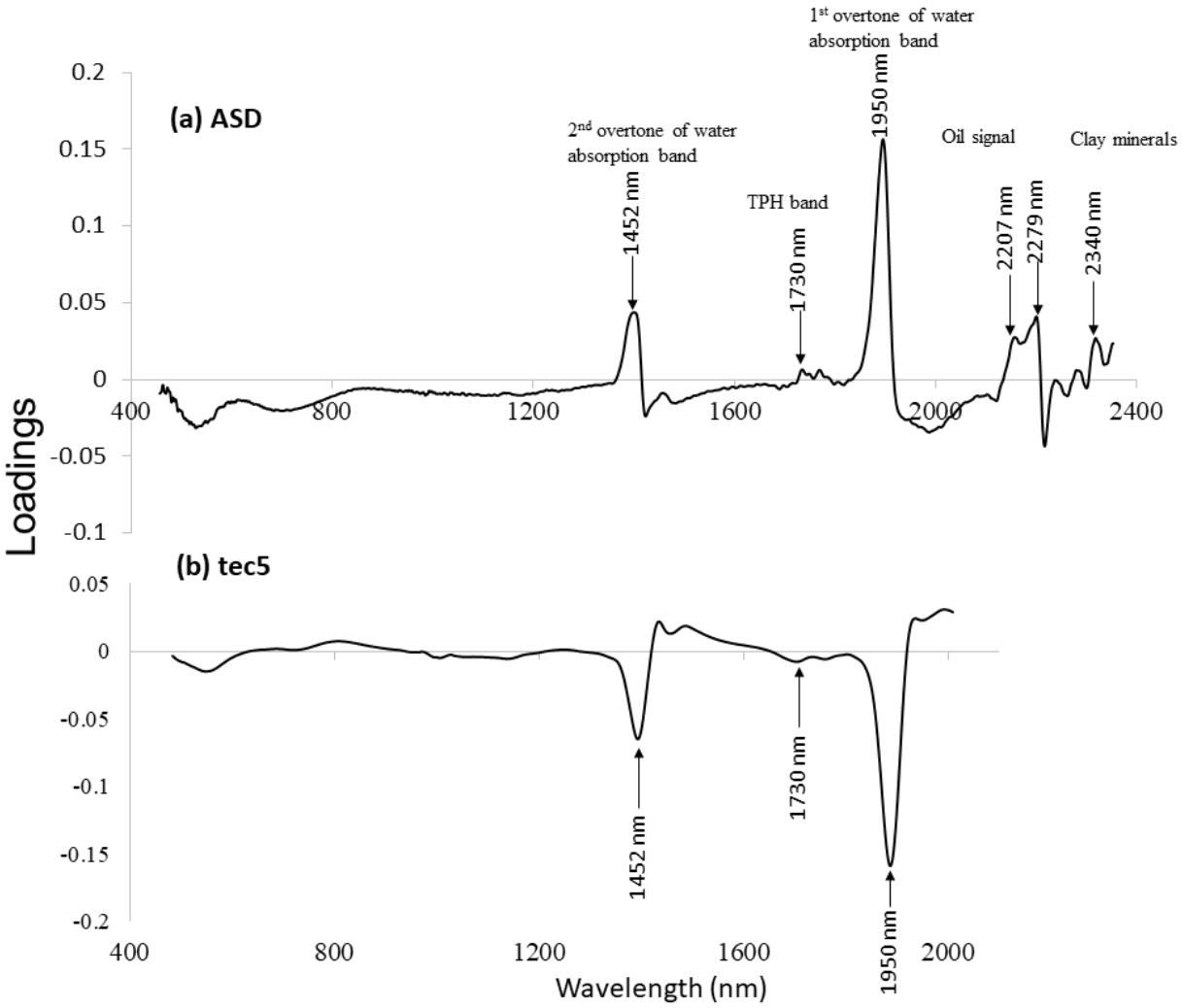

Fig. 4 Principal Component analysis loadings of the spectral patterns showing the wavelengths associated with hydrocarbon fractions, water and mineralogy.

3.3 Soil TPH analysis

The petroleum hydrocarbon profiles and change overtime are illustrated in Fig. 5. Chromatogram showed a well-developed series of n-alkanes distribution with carbon band range C10 – C36, but with about 85 % of the mixture existing within the range C12 - C28 (Fig 5; T2d). The distribution confirms that the hydrocarbon source is weathered (degraded) over time. After month 16 and 24, the most prominent residual hydrocarbon fractions were the aliphatic fractions C_{16}-C_{35} and C_{35}-C_{40}, and the aromatic fractions C_{12}-C_{16} and C_{16}-C_{21}, respectively.

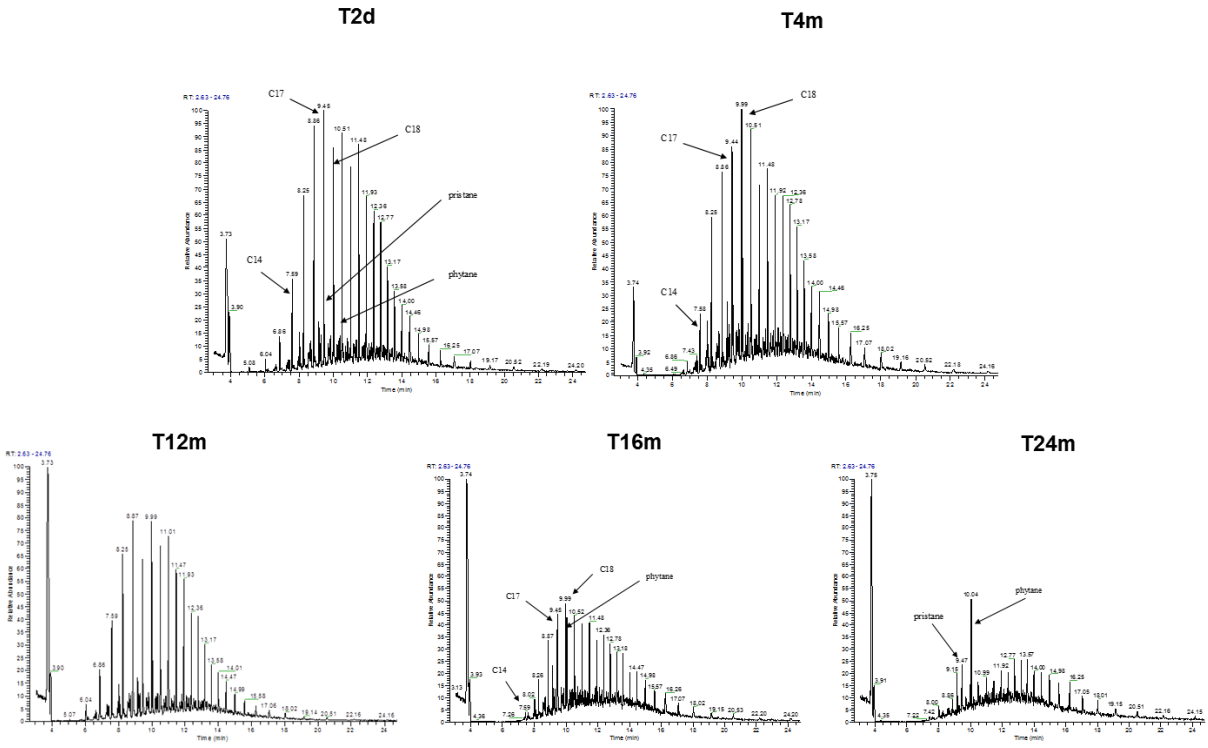

Fig. 5 Illustrative gas chromatography-mass spectrometry (GC-MS) chromatogram showing petroleum hydrocarbons fingerprint change overtime. Results are shown for contaminated soil samples after 48 h (T2d), after months 4 (T4m), 12 (T12m), 16 (T16m), and 24 (T24m)

Summary statistics of the aliphatic and aromatic fractions as well as the TPH concentrations which equal to sum of aliphatic and aromatic fractions are provided in Table 3. These TPH values were used for the vis-NIR spectra modeling. Samples were divided into calibration and

prediction sets. In the calibration set, the minimum and maximum concentrations of TPH were 187.5 and 1761.5 mg kg^{-1}, respectively. The minimum and maximum concentrations of TPH in the prediction set were 186.7 and 1362.4 mg kg^{-1}, respectively (Tables 3 and 4). The largest reduction in both the aliphatic and aromatic fractions were obtained after month 16 where 50% and 38% of the aliphatic and aromatic fractions, respectively, were degraded. Further to this, TPH reduction reached 72% by month 20 and 85% by month 24.

Table 3: Descriptive statistics of aliphatic and aromatic fraction concentrations (mg/kg) in 13 soil samples overtime (n = 78). Results are shown for diesel contaminated soil samples after 48 h (T2d), and months 4 (T4m), 12 (T12m), 16 (T16m), 20 (T20m) and 24 (T24m).

Hydrocarbon fractions		T2d			T4m			T12m		
		Med	Min	Max	Med	Min	Max	Med	Min	Max
Aliphatic	nC10-nC12	48.24	1.55	121.02	40.90	0.87	119.20	32.51	0.44	82.67
	nC12-nC16	63.20	32.97	99.04	50.54	4.19	105.84	37.56	4.54	81.10
	nC16-nC35	34.09	0.14	161.69	18.64	0.15	107.60	25.95	0.21	75.43
	nC35-nC40	0.56	0.02	4.18	0.78	0.04	4.68	0.73	0.01	13.57
	Total	1259	1113	1642	887.11	813.70	1214.75	880	721	1055
Aromatic	nC10-nC12	nd	nd	Nd	nd	nd	nd	nd	nd	nd
	nC12-nC16	3.32	3.25	3.96	3.33	3.26	4.29	3.36	3.10	3.53
	nC16-nC21	3.97	3.26	30.11	4.06	3.21	12.20	3.64	3.10	14.03
	nC21-nC35	6.49	3.50	15.24	5.72	3.36	16.51	3.92	3.10	16.63
	Total	82.68	71.59	134.37	81.71	64.77	100.47	56.09	43.10	94.93
	TPH	1343.28	1190.78	1716.49	963.83	884.15	1315.23	959.90	802.45	1101.30

		T16m			T20m			T24m		
		Med	Min	Max	Med	Min	Max	Med	Min	Max
Aliphatic	nC10-nC12	21.65	0.64	56.75	6.38	0.14	23.50	2.43	1.01	6.99
	nC12-nC16	29.05	13.82	50.42	13.83	1.80	31.91	7.48	1.05	18.81
	nC16-nC35	19.96	0.16	75.91	10.80	0.03	44.85	3.28	0.01	20.88
	nC35-nC40	0.25	0.01	2.19	0.24	0.01	2.35	0.02	0.01	0.76
	Total	678	628	774	326	233	421	162	133	185
Aromatic	nC10-nC12	nd	nd	Nd	nd	nd	nd	nd	nd	nd
	nC12-nC16	3.24	2.92	3.55	3.30	2.22	3.42	2.37	2.29	3.41
	nC16-nC21	4.29	2.05	7.64	3.77	3.31	6.29	3.32	3.10	4.40
	nC21-nC35	4.47	2.94	7.46	3.54	3.10	5.27	3.34	3.10	6.90
	Total	59.12	53.07	62.83	47.35	43.41	62.50	46.20	43.38	50.76
	TPH	733.87	687.62	833.98	380.53	279.68	465.40	207.83	178.47	232.64

nd= not detected, **med**=median, **min**=minimum, **max**=maximum

Table 4: Statistical summary of total petroleum hydrocarbons (TPH) concentrations of the collected soil samples measured with gas chromatography-mass spectrometry (GC-MS) for the different weathering stages in cross-validation and independent validation.

	N	Minimum	Mean	Median	1st Qu.	3rd Qu.	Maximum	St. dev
TPH (mg/kg)								
Cross-validation	60	187.5	773.70	789.20	383.60	990.10	1761.50	133.13
Independent validation	18	186.7	800.40	838.20	372.50	1121.4	1362.40	40.20

N = number of samples, 1st Qu. = first quartile; 3rd Qu. = third quartile; St. dev = standard deviation.

3.4 Models performance for estimating TPH

Table 5 and Figures 6 and 7 summaries the cross-validation and prediction results of TPH based on PLSR and RF analyses obtained with both the ASD and tec5 spectrophotometers. Generally, the RF models outperformed the PLSR in cross-validation and prediction for both ASD and tec5 measurements. The results of prediction based on ASD spectra indicated that RF model resulted in R^2 of 0.92, RMSEP of 108.56 mg/kg, RPD of 3.79, and RPIQ of 6.90, which outperformed PLSR model (R^2 = 0.83, RMSEP = 164.87 mg/kg, RPD = 2.49, RPIQ = 4.54). This was also the case for tec5 spectra as the RF model (R^2 = 0.22, RMSEP = 352.71 mg/kg, RPD = 1.16, and RPIQ = 2.13) outperformed PLSR (R^2 = 0.11, RMSEP = 422.50 mg/kg, RPD = 0.97, and RPIQ = 1.77). The current results for both PLSR and RF prediction are better than those reported by Douglas et al. (2018a, 2018b) using 85 naturally contaminated soil samples collected from the Niger Delta region of Nigeria. Furthermore, our results for RF prediction are better than those reported by Chakraborty et al. (2015) using 108 contaminated soil samples (West Texas, USA) with i) RF modeling method only (R^2 = 0.61, RMSE = 0.70 mg kg^{-1}, RPD = 1.64 and RPIQ = 0.57), and ii) RF combined with penalized spline regression (PSR) RF+PSR (R^2 = 0.78, RMSE = 0.53 $mgkg^{-1}$, RPD = 2.19 and RPIQ = 0.75). Also, the PLSR prediction in the current study are better than the results reported by Chakraborty et al. (2010 and 2015), who achieved RPD values of 1.7 and 1.96, respectively, for field-moist soils (Table 1). A possible reason for the observed

difference in the present study may be attributed to the combination of spectral pre-processing (maximum normalization, 1st derivative and smoothing) that represents a vital step in multivariate calibration and improves the model performance (Mouazen et al., 2010; Buddenbaum and Steffens, 2012; Nawar et al. 2016). According to Viscarra Rossel et al. (2006) model classification for RPD, excellent and very good predictions for TPH were achieved with RF-ASD (RPD = 3.79) and PLSR-ASD (2.49), respectively, whereas using tec5, poor and very poor results were obtained with RF-tec5 (RPD = 1.16) and PLSR-tec5 (RPD = 0.97), respectively.

Table 5: Summary results of partial least squares regression (PLSR) and random forest (RF) models in calibration (cross-validation) and prediction (independent validation) for total petroleum hydrocarbons (TPH) prediction in oil-contaminated soil samples using ASD and tec5 spectrophotometers.

		PLSR					RF				
Instrument		R^2	RMSEP (mg/kg)	RPD	RPIQ	LV	R^2	RMSEP (mg/kg)	RPD	RPIQ	*ntrees*
ASD	Calibration (n=60)	0.92	113.42	3.60	5.34	6	0.98	44.07	9.28	13.76	500
	Prediction (n=18)	0.83	164.87	2.49	4.54	4	0.92	108.56	3.79	6.90	200
tec5	Calibration (n=60)	0.83	164.26	2.47	3.70	8	0.92	111.65	3.63	5.45	500
	Prediction (n=18)	0.11	422.50	0.97	1.77	8	0.22	352.71	1.16	2.13	200

R^2 = coefficient of determination, RMSEP = root mean square error of prediction, RPD = residual prediction deviation, LV = number of latent variables, *ntrees* = number of trees, and RPIQ = ratio of performance to interquartile range.

The scatter plots of GC-MS measured *versus* ASD and tec5 predicted TPH concentrations (based on PLSR and RF models) are shown in Fig. 6 and Fig.7, respectively. Both the ASD and tec5 instruments quantitatively discriminated the soils at their various stages of weathering; however, a better discrimination was achieved with the ASD instrument. The results herein support the qualitative separation of the various soil groups by PC score plots illustrated in Fig. 3.

The TPH wavelength regression coefficients plots shown in Fig. 8 illustrate important wavebands around 1452, 1730, and 1950 nm for both ASD and tec5 spectrometers. The 1730 nm wavelength is attributed TPH absorption in the first overtone, which is close to the previous findings (Douglas et al., 2018a; Okparanma et al., 2014; Workman and Weyer, 2008; Osborne et al., 2007). The significant spectral signals around 1452 and 1950 nm are associated with water absorption bands in the second and first overtones, respectively, which accord findings reported in previous studies (Douglas et al., 2018a; Mouazen et al., 2007). In the ASD spectra, the spectral signature at 2207 nm may be due to the effect of hydrocarbon in the combination region around 2220 nm (Chakraborty et al., 2015 Forrester et al., 2013). Interestingly, the absorption feature around 2279 nm and 2340 nm is the same with the one observed in the PCA loadings (Fig. 4a). This is characteristic of clay minerals around 2300 nm (Clark et al., 1990). The low performance of tec5 in separating the different weathering groups (Fig. 4b) and quantitative assessment of TPH may be attributed to the smaller spectral range (losing important spectral features to TPH), compared to that of ASD.

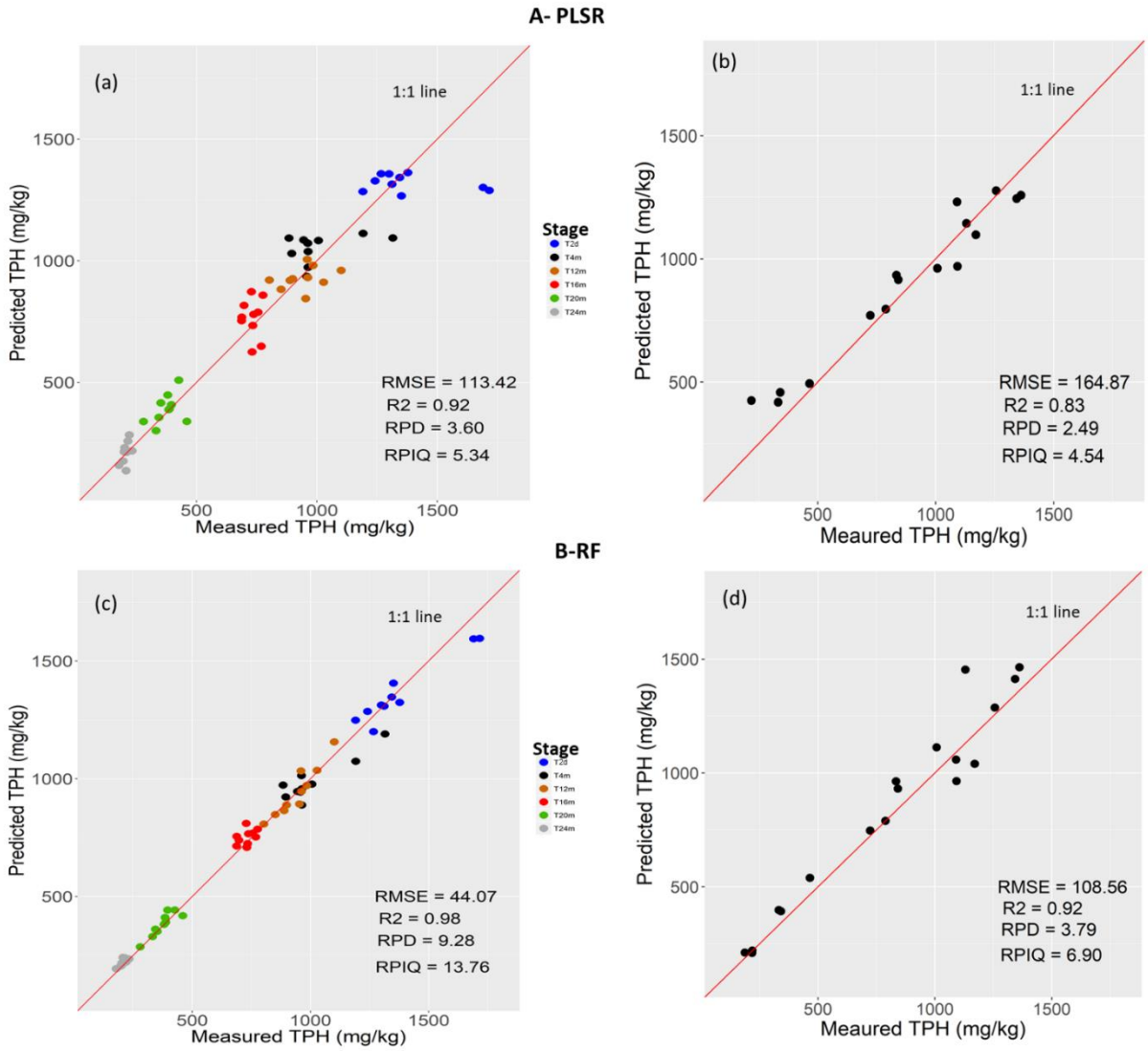

Fig. 6 Scatter plots of measured total petroleum hydrocarbons (TPH) using gas chromatography-mass-spectrometry (GC-MS) versus visible and near infrared (vis-NIR) ASD spectrometer predicted concentrations based on (A) partial least squares regression (PLSR) in (a) cross-validation and (b) prediction, and (B) random forest (RF) in (c) cross-validation and (d) prediction. Results show clear separation of diesel contaminated groups of different weathering stages of 48 h (T2d), and months 4 (T4m), 12 (T12m), 16 (T16m), 20 (T20m) and 24 (T24m).

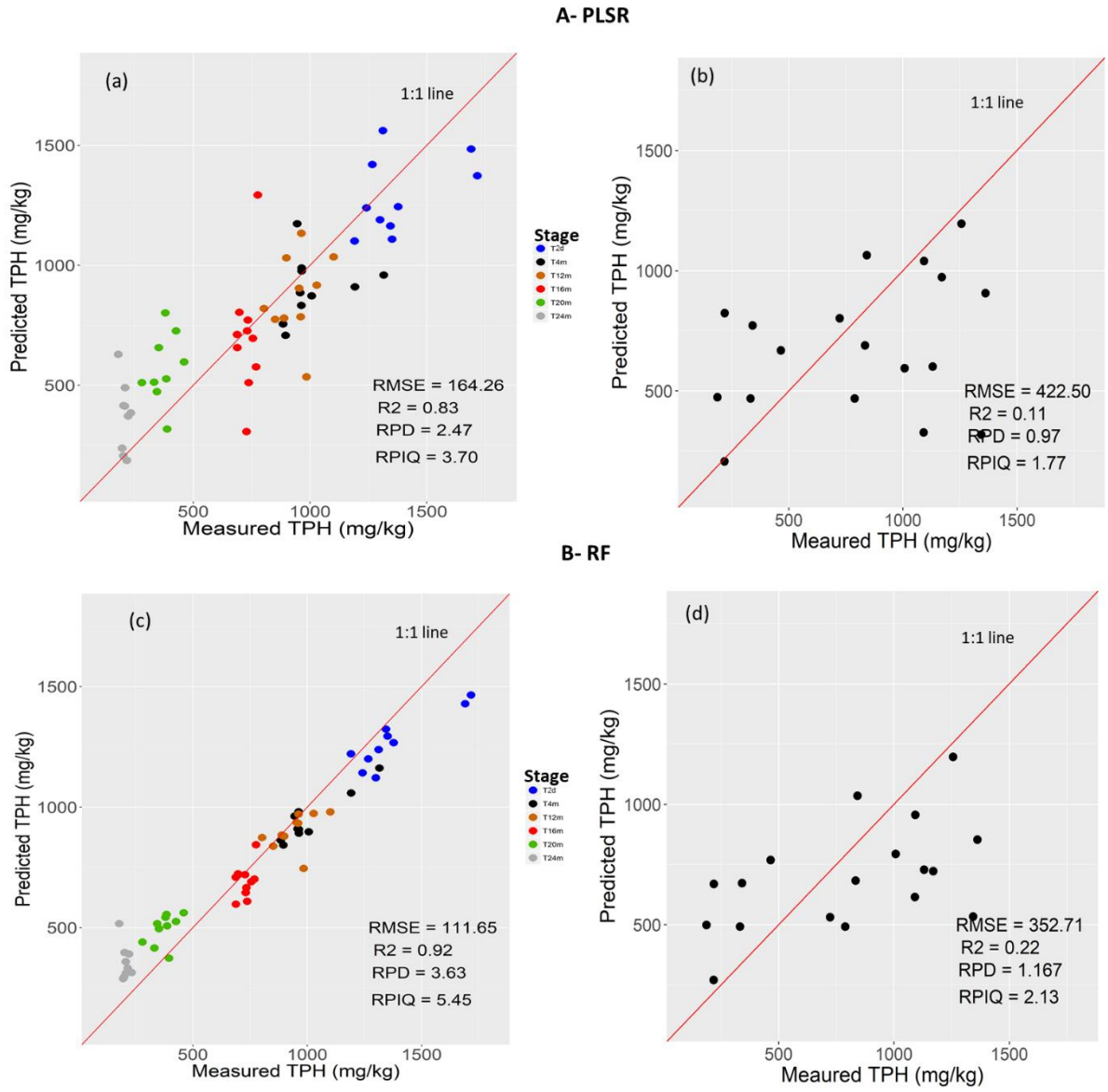

Fig. 7 Scatter plots of measured total petroleum hydrocarbons (TPH) using gas chromatography-mass-spectrometry (GC-MS) versus visible and near infrared (vis-NIR) tec5 spectrometer predicted concentrations based on (A) partial least squares regression (PLSR) in (a) cross-validation and (b) prediction, and (B) random forest (RF) in (c) cross-validation and (d) prediction. Results show clear separation of diesel contaminated groups of different weathering stages of 48 h (T2d), and months 4 (T4m), 12 (T12m), 16 (T16m), 20 (T20m) and 24 (T24m).

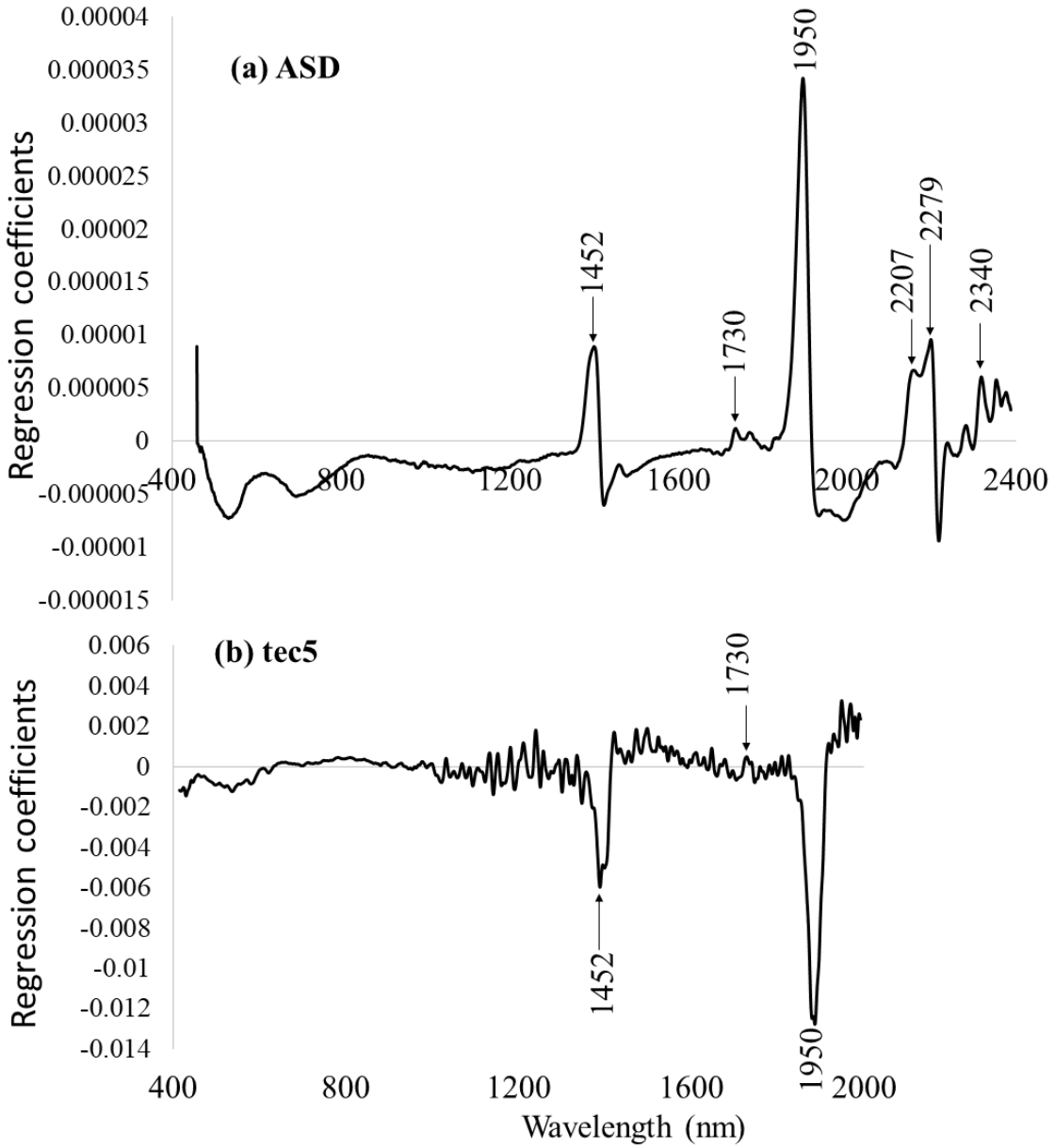

Fig. 8 Regression coefficients plots resulted from partial least squares regression (PLSR) analysis for total petroleum hydrocarbons (TPH) based on visible and near infrared (vis-NIR) spectra of oil-contaminated soil samples using (a) ASD and (b) tec5 spectrophotometers. Wavelengths highlighted on the plot are the potential features for TPH.

4. Conclusions

This pilot study evaluated visible and near infrared (vis-NIR) diffuse reflectance spectroscopy sensitivity to hydrocarbon concentration differences attributed to weathering for enhanced

assessment of crude oil contamination in soils. It compared the performance between a full vis-NIR range of 350-2500 nm spectrometer (e.g., ASD) with a short range of 305-2200 nm spectrometer (e.g., tec5), using two calibration methods of random forest (RF) and partial least squares regression (PLSR). From the results reported the following conclusions can be drawn:

- Principal component analysis (PCA) showed reasonable separation between the different weathered soil groups over time. This was true for the ASD spectrometer only, which was attributed to the large wavelength range of 350-2500 nm, compared to the short wavelength range (305-2200 nm) of the tec5 spectrometer. However, since total petroleum hydrocarbon (TPH) content is soil samples decreases with time due to weathering, the sensitivity of the ASD spectrometer for detecting changes due to weathering in soils decreases, particularly after 8 months of contamination.
- Both RF and PLSR analyses supported the PCA results for the ASD spectrometer in separation between different weathering groups, which was again much better that the separation obtained with the tec5 spectrometer. However, the RF model provided clearer separation than PLSR.
- Both RF and PLSR demonstrated that TPH can be estimated throughout time up to two years weathering. However, better estimation of TPH was obtained with RF-ASD model (R^2 = 0.92, RPD = 3.79, RMSE = 108.56 mg/kg), compared to PLSR-ASD model (R^2 = 0.83, RPD = 2.49, RMSE = 164.87 mg/kg).

Overall, the results demonstrated the potential of vis-NIR spectroscopy with a spectral range of 350-2500 nm for the successful estimation and discrimination of different weathering groups in oil-impacted soils. It is a rapid measurement tool for quick on-site investigation

and monitoring through weathering (up to 2 years), without the need for collecting soil samples and lengthy hydrocarbon extraction associated to traditional laboratory analysis.

Acknowledgements: This work was supported financially by the Petroleum Technology Development Fund (PTDF) of Nigeria (PTDF/OSS/PHD/DRK/711/14), REMEDIATE (Improved decision-making in contaminated land site investigation and risk assessment) Marie-Curie Innovation Training Network from the European Union's Horizon 2020 Programme (grant agreement No. 643087) and Flemish Scientific Research (FWO) funded SiTeMan Odysseus I Project (Nr. G0F9216N).

References

API, 2001. Risk-based Methodologies for Evaluating Petroleum Hydrocarbon Impacts at Oil and Natural Gas E&P Sites, API Publication 4709, API Publishing Services, Washington DC. Available at http://api-ep.api.org/industry/index.cfm?bitmask=002007001005009000.

Brassington, K.J., Pollard, S.T.J., Coulon, F., 2010. Weathered hydrocarbon wastes: a risk assessment primer, in Handbook of hydrocarbon and Lipid Microbioloy In: Timmis, K.N., McGenity, T., Van Der Meer, J.R., De Lorenzo, V. (Eds.), Handbook of Hydrocarbon and Lipid Microbiology. Springer Berlin, 2488–2499.

Brassington, K.J., Hough, R.L., Paton, G.I., Semple, K.T., Risdon, G.C., Crossley, J., Hay, I., Askari, K., Pollard, S.J.T, 2007. Weathered hydrocarbon wastes: a risk assessment management primer. Crit. Rev Environ Sci Technol 37,199–232.

British Standard BS 7755 Section 5.4, 1998. Determination of particle size distribution in mineral soil material-Method by sieving and sedimentation which is identical to ISO 11277:1998.

British Standard BS 7755 Section 3.8, 1995. Determination of organic and total organic after dry

combustion (elementary analysis) which is identical to ISO 10694:1995.

British Standard BS ISO 10390, 2005. Determination of pH.

Chang, C-W., Laired, D.A., Mausbach, M.J., and Hurburgh, C.R., 2001. Near-Infrared Reflectance Spectroscopy-Principal Component Regression Analyses of Soil Properties. Soil Sci. Soc. Am. J. 65, 480–490.

Chakraborty, S., Weindorf, D.C., Li, B., Aldabaa, A.A.A., Gosh, R.K., Paul, S., Ali, M.N., 2015. Development of a hybrid proximal sensing method for rapid identification of petroleum contaminated soils. Sci. Total Environ. 514, 399–408.

Chakraborty, S., Weindorf, D.C., Li, B., Ali, M.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.

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

Chakraborty, S., Weindorf, D. C., Morgan, C. L. S., Ge, Y., Galbraith, J. M., 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.

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

Cipullo S., Prpich G., Campo P., Coulon F. 2018. Assessing bioavailability of complex mixtures in contaminated soils: progress made and research needs. Sci. Total Environ. 615, 708–723.

Coulon, F., Whelan, M.J., Paton, G.I., Semple, K.T., Villa, R., and Pollard, S.J.T., 2010. Multimedia fate of petroleum hydrocarbons in the soil: oil matrix of constructed biopiles.

Chem., 81, 1454–62.

Demetriades-Shah, T.H., Steven, M.D., Clark, J.A., 1990. High Resolution Derivative Spectra in Remote Sensing. Remote Sens. Environ. 33:55–64.

Douglas, R.K., Nawar, S., Alamar, M.C., Coulon, F., Mouazen, A.M., 2017. Almost 25 years of chromatographic and spectroscopic analytical method development for petroleum hydrocarbons analysis in soil and sediment: state-of-the-art, progress and trends. Crit. Rev Environ Sci Technol., 47(16), 1497–1527.

Douglas, R.K., Nawar, S., Alamar, M.C., Mouazen, A.M., Coulon, F., 2018a. Rapid prediction of total petroleum hydrocarbons concentration in contaminated soil using vis-NIR spectroscopy and regression techniques. Sci. Total Environ., 616-617, 147–155.

Douglas, R.K., Nawar, S., Alamar, M.C., Coulon, F., Mouazen, A.M., 2018b. Rapid detection of alkanes and polycyclic aromatic hydrocarbons in oil-contaminated soils using visible near-infrared spectroscopy. Eur. J. Soil Sci. (in revision).

Drozdova, S., Ritter, W., Lendl, B., Rosenberg, E., 2013. Challenges in the determination of petroleum hydrocarbons in water by gas chromatography (hydrocarbon index). Fuels. 113, 527–536.

Fontán, J. M., Calvache, S., López-Bellido, R. J. & López-Bellido, L. 2010. Soil carbon measurement in clods and sieved samples in a Mediterranean Vertisol by Visible and Near-Infrared Reflectance Spectroscopy. *Geoderma*, 156, 93–98.

Forrester, S.T., Janik, L.J., McLaughlin, M.J., Soriano-Disla, J.M., Stewart, R., Dearman, B., 2013. Total Petroleum Hydrocarbon Concentration Prediction in Soils Using Diffuse Reflectance Infrared Spectroscopy. Soil Sci. Soc. Am. J. 77, 450–460.

182 Environment Agency, 2005. The UK approach for evaluating human health risks from petroleum
183 hydrocarbons in soils, Science Report P5-080/TR3, Environment Agency, Almondsbury,
184 Bristol.
185 Geladi, P., Kowalski, B.P., 1986. Partial least-squares regression: A tutorial. Analytica Chimica
186 Acta 185(1), 1–17.
187 Hauser, A., Ali, F., Al-Dosari, B., Al-Sammar, H., 2013. Solvent-free determination of TPH in
188 soil by near-infrared reflectance spectroscopy. Int. J. Sustain. Dev. Plan. 8, 413–421.
189 Hoerig, B., Kuehn, F., Oschuetz, F., Lehmann, F., 2001. HyMap hyperspectral remote sensing
190 to detect hydrocarbons. Int. J. Remote Sens. 8, 1413–1422
191 Horta, A., Malone, B., Stockmann, U., Minasny, B., Bishop, T.F.A., McBratney, A.B., Pallasser,
192 R. & Pozza, L. 2015. Potential of integrated field spectroscopy and spatial analysis for
193 enhanced assessment of soil contamination: A prospective review. Geoderma, 241-242, 180–
194 209.
195 Jiang Y., Brassington K.J., Prpich G., Paton G.I., Semple K.T., Pollard S.J.T., Coulon F. 2016.
196 Insights into the biodegradation of weathered hydrocarbons in contaminated soils by
197 bioaugmentation and nutrient stimulation. Chemosphere. 161: 300–307.
198 Kennard, R.W. & Stone, L.A. 1969. Computer aided design of experiments. Technometrics, **11**,
199 137–148.
200 Malley, D.F., Hunter, K.N., Webster, G.R.B., 1999. Analysis of Diesel Fuel Contamination in
201 Soils by Near-Infrared Reflectance Spectrometry and Solid Phase Microextraction-Gas
202 Chromatography. J. Soil Contam. 8, 481–489.
203 Martens, H., and T. Naes. 1989. Multivariate calibration. 2nd ed. John Wiley & Sons,
204 Chichester, UK.

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

207 Mouazen, A.M., Karoui, R., De Baerdemaeker, J., Ramon, H., 2006. Characterization of soil
208 water content using measured visible and near infrared spectra. Soil Science Society of
209 America Journal, 70, 1295–1302.

210 Mouazen, A.M., De Baerdemaeker, J., Ramon, H., 2005. Towards development of on-line soil
211 moisture content sensor using a fibre-type NIR spectrophotometer. Soil Tillage Res. 80, 171–
212 183.

213 Mullins, O.C., Mitra-Kirtley, S., Zhu, Y., 1992. The electronic absorption edge of petroleum.
214 Appl. Spectrosc. 46, 1405–1411.

215 Okparanma, R.N., Coulon, F., Mouazen, A.M., 2014. Analysis of petroleum-contaminated soils
216 by diffuse reflectance spectroscopy and sequential ultra sonic solvent extraction-gas
217 chromatography. Environ. Pollut. 184, 298–305.

218 Okparanma, R. N., Mouazen, A. M., 2013a. Determination of Total Petroleum Hydrocarbon
219 (TPH) and Polycyclic Aromatic Hydrocarbon (PAH) in soils. A Review, Appl. Spectrosco.
220 Rev, 46 (6), 458–486.

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

224 Paíga, P., Mendes, L., Albergaria, J.T., Delerue-Matos, C.M., 2012. Determination of total
225 petroleum hydrocarbons in soil from different locations using infrared spectrophotometry and
226 gas chromatography. Chem. 66, 711–721.

R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (URL http://www.R-project.org/).

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

Schwartz, G., Ben-Dor, E., and 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.

Stenberg, B., 2010. Effects of soil sample pre-treatments and standardised rewetting as interacted with sand classes on Vis-NIR predictions of clay and soil organic carbon. Geoderma, 158(1-2), 15–22.

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

Viscarra Rossel, R.A., Walvoort, D.JJ., 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.

Wartini, Ng., Brendan, P.M. & Budiman, M., 2017. Rapid assessment of petroleum-contaminated soils with infrared spectroscopy. Geoderma, 289, 150–160.

Wold, S., 2010. Personal memories of the early PLS development. Chemometrics and Intelligent Laboratory Systems 58, 83–84.

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