



Rapid prediction of total petroleum hydrocarbons concentration in contaminated soil using vis-NIR spectroscopy and regression techniques



R.K. Douglas^a, S. Nawar^a, M.C. Alamar^a, A.M. Mouazen^{a,b}, F. Coulon^{a,*}

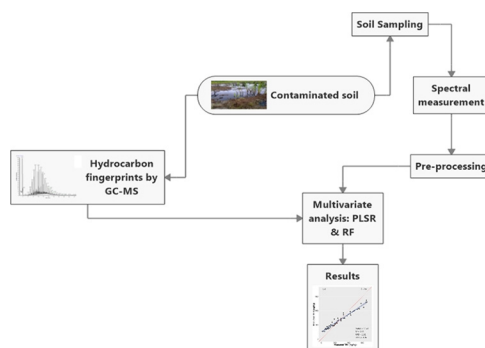
^a School of Water, Energy and Environment, Cranfield University, Cranfield MK43 0AL, UK

^b Department of Soil Management, Ghent University, Coupure 653, 9000 Gent, Belgium

HIGHLIGHTS

- vis-NIR spectroscopy and chemometrics as rapid measurement tool for hydrocarbons
- Better TPH prediction with random forest modelling
- Random forest handles well the non-linear behaviour of the spectral response.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 10 September 2017

Received in revised form 30 October 2017

Accepted 30 October 2017

Available online xxx

Editor: D. Barcelo

Keywords:

Total petroleum hydrocarbons

vis-NIR spectroscopy

Chemometric methods

Partial least squares regression

Random forest regression

ABSTRACT

Visible and near infrared spectrometry (vis-NIRS) coupled with data mining techniques can offer fast and cost-effective quantitative measurement of total petroleum hydrocarbons (TPH) in contaminated soils. Literature showed however significant differences in the performance on the vis-NIRS between linear and non-linear calibration methods. This study compared the performance of linear partial least squares regression (PLSR) with a nonlinear random forest (RF) regression for the calibration of vis-NIRS when analysing TPH in soils. 88 soil samples (3 uncontaminated and 85 contaminated) collected from three sites located in the Niger Delta were scanned using an analytical spectral device (ASD) spectrophotometer (350–2500 nm) in diffuse reflectance mode. Sequential ultrasonic solvent extraction–gas chromatography (SUSE-GC) was used as reference quantification method for TPH which equal to the sum of aliphatic and aromatic fractions ranging between C₁₀ and C₃₅. Prior to model development, spectra were subjected to pre-processing including noise cut, maximum normalization, first derivative and smoothing. Then 65 samples were selected as calibration set and the remaining 20 samples as validation set. Both vis-NIR spectrometry and gas chromatography profiles of the 85 soil samples were subjected to RF and PLSR with leave-one-out cross-validation (LOOCV) for the calibration models. Results showed that RF calibration model with a coefficient of determination (R^2) of 0.85, a root means square error of prediction (RMSEP) 68.43 mg kg⁻¹, and a residual prediction deviation (RPD) of 2.61 outperformed PLSR ($R^2 = 0.63$, RMSEP = 107.54 mg kg⁻¹ and RDP = 2.55) in cross-validation. These results indicate that RF modelling approach is accounting for the nonlinearity of the soil spectral responses hence, providing significantly higher prediction accuracy compared to the linear PLSR. It is recommended to adopt the vis-NIRS coupled with RF modelling approach as a portable and cost effective method for the rapid quantification of TPH in soils.

© 2017 Elsevier B.V. All rights reserved.

* Corresponding author.

E-mail addresses: Abdul.Mouazen@UGent.be (A.M. Mouazen), f.coulon@cranfield.ac.uk (F. Coulon).

1. Introduction

Petroleum hydrocarbons contamination in soil is a worldwide significant environmental issue which has raised serious concerns for the environment and human health (Brevik and Burgess, 2013). Petroleum hydrocarbons encompass a mixture of short and long-chain hydrocarbon compounds. However the difference between the term petroleum hydrocarbons (PHC) as such and the term total petroleum hydrocarbons (TPH) should be noted. PHC typically refer to the hydrogen and carbon containing compounds that originate from crude oil, while TPH refer to the measurable amount of petroleum-based hydrocarbons in an environmental matrix and thus to the actual results obtained by sampling and chemical analysis (Coulon and Wu, 2017). TPH is thus a method-defined term. Among a range of techniques, gas chromatography is preferred for the measurement of hydrocarbon contamination in environmental samples, since it allows to detect a broad range of hydrocarbons and can provide both sensitivity and selectivity depending on the detector and hyphenated configuration used (Brassington et al., 2010; Drozdova et al., 2013). However, GC-based techniques can be time consuming and expensive and do not allowed rapid and broad scale analysis of petroleum contamination on-site (Okparanma and Mouazen, 2013; Okparanma et al., 2014).

Among potential rapid measurement techniques that can be carried out on-site, reflectance spectroscopy, including the visible and near infrared (vis-NIRS) and mid infrared ranges, is one of the most promising techniques for detecting and quantifying TPH (Okparanma and Mouazen, 2013). Reflectance spectroscopy measures the diffuse reflected electromagnetic energy from samples (i.e. soil or sediment) subjected to a light source; by modelling the sample spectral data against samples with known chemical composition and concentration levels, calibration models for quantifying key attributes can be established. However, to date very limited studies on the use of reflectance spectroscopy for the analyses of TPH in soil can be found in the literature.

There are also several factors affecting the measurement accuracy of reflectance spectroscopy, including among others the quality of the laboratory reference data and spectra, and adopted pre-processing and modelling techniques (Viscarra Rossel and Behrens, 2010; Nawar et al., 2016). Partial least-squares regression (PLSR) is the most common multivariate analysis method, as it is capable to model several response variables simultaneously while effectively addressing strongly collinear and noisy predictor variables (Wold, 2010). It is important to mention that PLSR is a linear approach that may not perform well when solving nonlinear behaviour, e.g., like those of soil. Random forest (RF) is typically known as a hierarchical nonparametric method that estimates complex nonlinear relationships among independent and dependent variables. RF method was reported to be outperformed by PLSR, adaptive regression splines (MARS), artificial neural network (ANN) and support vector machine (SVM) for the analysis of soil organic carbon, clay content and pH (Viscarra and Behrens, 2010; Breiman, 2001) whereas Knox et al. (2015) reported that RF outperformed PLSR for the analysis of soil total carbon (TC) with residual prediction deviation (RPD) of 2.7 and 2.6 for RF and PLSR, respectively. For TPH analysis using vis-NIRS, a recent study by Chakraborty et al. (2015) showed PLSR outperformed both penalised spline regression (PSR) and RF modelling approaches; the authors reported residual prediction deviation (RPD) of 1.64, 1.86, and 1.96 for RF, PSR, and PLSR, respectively. This single study comparing the performance of RF with PLSR for the analysis of TPH may not confirm this trend to be correct, as previous work reported RF to outperform PLSR for modelling of other soil properties (Knox et al., 2015). Therefore, it is essential to evaluate the capability of the RF as a nonlinear modelling approach for modelling TPH content in the soil and to confirm whether or not TPH can be predicted with RF with higher accuracy than with PLSR. To the best of our knowledge, there is to date no study where RF modelling has been applied to estimate TPH in soils based on vis-NIR spectroscopy with a limited soil

data set. Thus, the aim of this study is to compare the performance of PLSR linear modelling technique with RF nonlinear technique to predict TPH in oil-contaminated soils from Niger Delta, Southern Nigeria, using vis-NIR spectroscopy.

2. Materials and methods

2.1. Study area and sample collection

The study area located in Bayelsa and Rivers State, Niger Delta, Southern Nigeria has a tropical rain forest climate characterised by two seasons: the rainy season lasts for about 7 months between April and October with an overriding dry period in August (known as August break); and the dry season lasts for about 5 months, between November and March. The temperature varies between 25 °C and 35 °C. The regional geology of the Niger Delta is relatively simple, consisting of Benin, Agbada (the kitchen of kerogen) and Akata Formations, overlain by various types of Quaternary deposits (Kogbe, 1989; Wright et al., 1985). A total of 85 representative spot sample points were collected randomly from three oil contaminated sites (Ikarama: 31 samples; Kalabar: 21 samples; and Joinkrama: 33 samples) in August 2015. The soil samples (approx. 5 kg) were collected in the top 15-cm soil layer using a shovel. In addition, three uncontaminated samples were collected (2 samples from Joinkrama, 1 sample from Kalabar) for control purpose. Fig. 1 shows the sampling location map. Soil samples were kept in air-tight centrifuge tubes and stored at 4 °C using ice block to avoid hydrocarbon volatilisation and preserve field-moist status until shipment to Cranfield University. The samples were then stored at –20 °C prior to GC–MS analysis.

2.2. Soil physicochemical properties

Prior to soil physicochemical properties analysis, soil samples were grouped based on the variation of the soil texture using the “Feel Method” (Thien, 1979). Then two representative samples were selected from each texture class with a total of 10 samples per site. Therefore soil physicochemical properties were determined on 30 soil samples selected to represent soil spatial variation in the study. This approach was used due to limited amount of soil that could be transported back to the UK for analysis. 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.

2.3. Soil scanning and spectral analysis

The diffuse reflectance spectra of the soil samples were measured using an ASD LabSpec2500® Vis–NIR spectrophotometer which covers a spectral range of 350–2500 nm (Analytical Spectral Devices, Inc., USA). With a spectral interval resolution varying of 3 nm at a wavelength of 700 nm and of 6 nm between 1400 and 1200 nm, the ASD LabSpec2500® spectrometer recorded a total 2151 spectral bands. The spectral measurements were made in the dark in order to both, control the illumination conditions and reduce the effects of stray light. 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° angle. Before use, and after every 30 min, the instrument was calibrated by white-referencing with a white Spectralon disc of ca. 100% reflectance. Three subsamples (field-moist) from each soil sample were packed into plastic Petri dishes (1 cm height, 5.6 cm diameter) for vis-NIR DRS spectra measurement. To obtain optimal diffuse reflection, and hence, a good signal-to-noise ratio, all plant and pebble particles were removed and

Download English Version:

<https://daneshyari.com/en/article/8862287>

Download Persian Version:

<https://daneshyari.com/article/8862287>

[Daneshyari.com](https://daneshyari.com)