# Total Petroleum Hydrocarbon Concentration Prediction in Soils Using Diffuse Reflectance Infrared Spectroscopy

**Article** *in* Soil Science Society of America Journal · March 2013

DOI: 10.2136/sssaj2012.0201

**6 authors**, including:

Sean T Forrester
ForrEnviro
**73** PUBLICATIONS **3,050** CITATIONS
SEE PROFILE

Mike John McLaughlin
The University of Adelaide
**605** PUBLICATIONS **36,094** CITATIONS
SEE PROFILE

L. Janik
The Commonwealth Scientific and Industrial Research Organisation
**69** PUBLICATIONS **7,579** CITATIONS
SEE PROFILE

JOSE M SORIANO DISLA
**54** PUBLICATIONS **2,735** CITATIONS
SEE PROFILE

# Total Petroleum Hydrocarbon Concentration Prediction in Soils Using Diffuse Reflectance Infrared Spectroscopy

**Sean T. Forrester**
**Les J. Janik***
CSIRO Water for a Healthy Country
Flagship Program
CSIRO Land and Water
Waite Campus
Waite Road
Urrbrae, SA 5064, Australia

**Michael J. McLaughlin**
CSIRO Water for a Healthy Country
Flagship Program
CSIRO Land and Water
Waite Campus
Waite Road
Urrbrae, SA 5064 Australia
*and*
Univ. of Adelaide
PMB 2
Glen Osmond, SA 5064, Australia

**José M. Soriano-Disla**
CSIRO Water for a Healthy Country
Flagship Program
CSIRO Land and Water
Waite Campus
Waite Road
Urrbrae, SA 5064 Australia
*and*
Dep. of Agrochemistry and Environment
Univ. Miguel Hernández of Elche
Avenida de la Universidad S/N
03202 Elche, Spain

**Richard Stewart**
Ziltek Pty Ltd
PO Box 373
Kent Town, SA 5071, Australia

**Ben Dearman**
18 Birks Street
Parkside, SA 5063, Australia

**Partial least squares (PLS) calibration models, from diffuse reflectance infrared Fourier-transform (DRIFT) spectra, were developed for the prediction of total petroleum hydrocarbon (TPH) concentrations in contaminated soils. Soils were collected from sites in southeastern Australia known to be contaminated with TPH. Regression models were derived for concentration ranges of 0 to 32,600, 0 to 5000, 0 to 15,000, and 15,000 to 32,600 mg kg$^{-1}$. For each range, PLS models using selected near-infrared (NIR) and mid-infrared (MIR) frequencies were tested. The aliphatic alkyl stretching vibration regions were the most sensitive to TPH: NIR frequencies at 4500 to 4100 cm$^{-1}$ and MIR at 3000 to 4600 cm$^{-1}$. The MIR range included two –CH$_3$ peaks, one at 2950 cm$^{-1}$ and the other near 2730 cm$^{-1}$, having strong correlation with TPH at low and high TPH concentrations, respectively. These peaks were considered to be either weak or absent in natural soil organic matter relative to the usual –CH$_2$ region at 2930 to 2850 cm$^{-1}$. The PLS regression analysis using the combined 2980 to 2950 and 2777 to 2650 cm$^{-1}$ MIR frequency regions for the 0 to 15,000 mg kg$^{-1}$ set resulted in a ratio of prediction deviation of 3.7 (“analytical quality”), coefficient of determination ($R^2$) = 0.93, and root mean square error of cross-validation of 564 mg kg$^{-1}$. Using these MIR frequencies, the DRIFT infrared technique showed the potential to be a rapid and accurate nondestructive method to determine TPH concentrations in contaminated soils, potentially adaptable to in-field use.**

**Abbreviations:** DRIFT, diffuse reflectance infrared Fourier transform; GC-FID, gas chromatography with flame ionization detection; IR, infrared; MIR, mid-infrared; NIR, near infrared; PLS, partial least squares; $R^2$, coefficient of determination; RPD, ratio of prediction deviation; SOM, soil organic matter; TPH, total petroleum hydrocarbon.

Total petroleum hydrocarbons typically comprise a mixture of short- and medium-length hydrocarbon compounds that are derived from crude oil. The less volatile components of petroleum hydrocarbons can become environmental contaminants in the event of slow leakage from storage tanks or as the result of a rapid and catastrophic oil leak due to an accident. Such pollution may remain in the environment for extended periods, often being toxic to wildlife, flora, and humans. With urban growth, city fringes are gradually encroaching on areas that were formerly disused or predominantly industrial in nature. While large-scale contamination from petroleum production related accidents usually receives considerable press coverage and requires urgent or immediate attention for remediation, many sites have been contaminated by petroleum hydrocarbon leakage from previous industrial uses and are much more common. For example, fuel from underground storage tanks at current and previous service station sites often leaks into soils and the site requires remediation as a prerequisite for urban development or to avoid groundwater contamination.

Soil Sci. Soc. Am. J. 77:450–460
doi:10.2136/sssaj2012.0201
Received 22 June 2012.
*Corresponding author (Les.Janik@csiro.au).

Government jurisdictions have set soil quality standards for TPH, above which further investigation, risk assessment, or remediation is required. In Australia, health investigation levels for aliphatic TPH concentrations in soils, set to protect human health by the National Environmental Protection Council, vary from 5600 to 28,000 mg $kg^{-1}$ for $C_{16}$ to $C_{35}$ and up to 280,000 mg $kg^{-1}$ for $>C_{35}$ depending on land use (National Environment Protection Council, 1999a). Thus, sites contaminated by TPH require analysis of the soils for these compounds before the commencement of risk assessment or remediation.

Currently, most soil analyses for TPH are performed via supercritical fluid extraction of the TPH components from the soil, followed by gas chromatography with flame ionization detection (GC-FID). Although this method is the industry standard, as required by regulatory agencies (National Environment Protection Council, 1999b; USEPA, 2003; ISO, 2004), it is time consuming and requires a nationally accredited laboratory and thus is not suited to rapid analysis or on-site applications. A simpler and faster alternative method for screening contaminated sites for total TPH concentrations would be desirable to speed up risk assessment processes and to assist remediation.

Infrared (IR) spectroscopy has potential for the simple and rapid determination of TPH concentrations. One standard method, using Freon 113 as a soil extractant and subsequent analysis of the extracted TPH by MIR attenuated total reflection, has been proposed (USEPA, 1978). Because of restrictions on the use of ozone-depleting substances, however, this method was modified to use non-ozone-depleting substances for the extraction step (Hazel et al., 1997; Freeman and Krishnan, 2007; ASTM, 2011). The extraction step is, however, somewhat tedious to carry out, is not rapid, and does not lend itself to adaptation for in-field application as a direct measurement method.

The development of an alternative, faster IR technique for the prediction of TPH would have considerable benefits in terms of saving money and time for site risk assessment and remediation. A rapid, accurate, and simple-to-use method, based on diffuse reflectance infrared Fourier-transform (DRIFT) spectroscopy of undiluted whole, neat (without the use of any dispersing matrix such as KBr powder) soils combined with multivariate regression such as partial least squares (PLS) regression, offers a means to achieve this requirement. Further advantages can be made by adapting the DRIFT-PLS technique from laboratory-based spectrometers to in-field measurements using new hand-held IR technologies.

An important advantage of the DRIFT technique is that unprocessed, neat, whole soil samples can be studied directly, where the samples are simply placed under the incoming IR beam and the reflected signal is analyzed; thus, often no sample preparation is required. Immediacy in TPH analysis (on-site, rapid, or real-time techniques) is not only important where a response action is needed quickly after an accident to minimize major environmental damage but also for urban development where delays to construction can be costly. For this reason, samples need to be presented as received (neat, unground samples) in an attempt to develop a rapid technique ultimately adaptable for in-field application.

In practice, both the NIR (12,500–4000 $cm^{-1}$) and MIR (4000–400 $cm^{-1}$) spectral ranges can be used for TPH prediction when combined with PLS regression. The combined DRIFT spectroscopy and PLS regression technique proposed here for quantification of TPH draws on procedures similar to those used to determine other soil properties on neat soils using MIR or NIR spectroscopy (Janik and Skjemstad, 1995; Janik et al., 1998; Reeves et al., 1999, 2006; Viscarra-Rossel et al., 2006; Kuang et al., 2012). A literature search on the use of IR for TPH prediction in soils, and in particular using unprocessed, neat, whole soils, resulted in very few studies. The first reported IR study on the prediction of TPH in contaminated soils by Malley et al. (1999) demonstrated the potential of NIR for routine TPH determination, an application later confirmed by others using PLS and other multivariate methods (Chakraborty et al., 2010, 2012). Near-infrared spectroscopy together with PLS was also used by Sorak et al. (2012) for the prediction of diesel and oil concentrations in contaminated soil samples. While soil moisture was seen as a major issue in analyzing soils for TPH using NIR (Malley et al., 1999; Chakraborty et al., 2010) and MIR (Hazel et al., 1997), these studies did not attempt to examine the sensitivity of various IR frequency regions with regard to TPH concentration ranges. The MIR studies by Hazel et al. (1997) did, however, demonstrate the adverse effect of moisture on relatively small sample sets and showed that it was possible to counter the negative effects of soil moisture by the use of PLS regression methods. Furthermore, Chakraborty et al. (2012) tested the effects of different clay contents and levels of organic C but also did not investigate the impacts of specific spectral frequencies. Our study aimed to examine the impact of various specific frequencies in the usual alkyl frequency range and to test each of these frequencies on the regression precision and accuracy for low to very high ranges of TPH concentrations.

Central to the use of the IR–DRIFT method is that both the NIR and MIR spectra are sensitive to alkyl functional chemical groups in organic materials, including TPH compounds (Hazel et al., 1997; Malley et al., 1999; Forrester et al., 2009; Chakraborty et al., 2010). The sensitivity of IR spectra to organic aliphatic components therefore provides the possibility of accurately screening soils for TPH contamination.

Although rapid, there are inherent problems with the application of the DRIFT technique to whole soil analysis for TPH. The first problem is the overlap of TPH-sensitive IR peaks with those of naturally occurring soil organic matter (SOM), particularly those associated with the alkyl $-CH_2$ groups. Identification of spectral peaks unique to TPH can thus be difficult (Forrester et al., 2009). Apart from the interference by SOM peaks, there is also overlap of TPH peaks in the MIR by spectra of soil carbonates. Other known problems are due to the physical effects inherent when examining soils, for example, shielding of the internal structure of soil microaggregates to IR radiation and heterogeneity of the samples. Variable soil water content can also be a problem

**Table 1. Reference standard soils ACU-1 and Cungena showing sampling depth, total organic C (TOC), pH ($CaCl_2$), clay, silt, sand, and carbonate concentrations, and major mineral composition according to infrared spectra.**

| Soil | pH | TOC | Clay | Silt | Sand | Carbonate | Dominant mineralogy |
|---|---|---|---|---|---|---|---|
| | | g kg$^{-1}$ | % | | | | |
| ACU-1 | 4.7 | 25.6 | 21 | 69 | 10 | 0 | illite, smectite, kaolinite, minor quartz |
| Cungena | 7.6 | 1.2 | 14 | 4 | 82 | 36 | illite, smectite, kaolinite, calcite |

in field situations where it can confound predictions, particularly where the IR beam is reflected from high levels of surface liquid (Hazel et al., 1997; Gallagher et al., 2008), or cause a mismatch of the spectra with the predictive model(s) when models have been derived from dry samples (Gallagher et al., 2008). Drying the samples would add considerable sample preparation time (Hazel et al., 1997). Fortunately, some of these problems can be minimized by additional principal components to the PLS models or by careful frequency range selection. This study thus aimed to determine the potential accuracy and detection limits of predicted TPH concentration using IR spectroscopy.

# MATERIAL AND METHODS

## Samples

### Reference Materials for Spiking with Total Petroleum Hydrocarbons

Two standard reference soils were used: an organic silty loam ACU-1 (Agrixeroll) from Gumeracha (South Australia) and a calcareous sandy soil (Calcixerollic Xerochrept) from Cungena, Eyre Peninsula (South Australia). The basic soil properties were determined according to standard methods (Rayment and Higginson, 1994) and are presented in Table 1. The soils were dried at 40°C and sieved to <2 mm. Soil organic matter extracted from a reference soil (CSIRO standard soil SS6) using HF was available for use as a SOM "signature" (Skjemstad et al., 1998). Reference hydrocarbons (TPH) were crude oil (light crude blend, BP oil refinery), diesel (Navy diesel, BP oil refinery), and *n*-hexane (Sigma-Aldrich, 95% spectrophotometric grade) were used as spectral representatives of TPH. The crude oil was also used for the spiking experiments. Diluted aliquots of the crude oil stock solution were prepared with 0, 0.025, 0.05, 0.10, 0.25, 0.50, 0.70, 1.00, 1.20, 1.50, 2.00 and 2.50 mL of stock crude oil diluted with cyclohexane to give 10 mL total volume of each aliquot, equivalent to concentrations of 0 to 25,000 mg kg$^{-1}$. Aliquots of the crude oil stock solutions were mixed with fixed weights of each sample, 10 g for the Cungena soil and 2.5 g for the ACU-1 soil, in a tumbler for 12 h to give spiked soils with concentrations ranging from 257 to 25,000 mg kg$^{-1}$. The samples were allowed to dry for 18 h at 40°C to remove all traces of the cyclohexane solvent, and the concentrations were confirmed using solvent extraction and GC-FID. Spectra of reference minerals commonly found in soil were also determined, i.e., calcite, kaolinite, smectite and quartz, as well as those of the diesel, crude oil, and *n*-hexane.

### Field-Contaminated Soils

A set of 205 samples was collected for calibration modeling from four sites in southern Australia: Victoria (45 from Melbourne and 89 from Ballarat, noted as M and B, respectively), New South Wales (41 from near Sydney, noted as A), and 30 from South Australia (near Adelaide, noted as S). These are described in Table 2. The soils at each site had generally been deposited fill material from other contaminated soil locations, as is a common practice at industrial sites, and were therefore extremely heterogeneous, as is typical for urban contaminated sites. At each site, samples were taken from predetermined sampling points chosen from site maps or pits. Sampling was conducted at various depths on the pit faces and placed into glass jars, lightly mixed by hand, and then split, one for laboratory analysis and the other placed in cold storage at 4°C for spectral scanning.

Laboratory TPH analyses were performed using the GC-FID method (Sadler and Connell, 2003) on one of the two soil subsamples by an accredited analytical reference laboratory (ALS Laboratories, Campbell Bros. Ltd). The contaminated soil data and descriptions are presented in Table 2.

## Infrared Spectra

The MIR and NIR DRIFT spectra were scanned with a PerkinElmer Spectrum-One FTIR spectrometer. The spectrometer was equipped with an extended-range KBr beam splitter, a high-intensity ceramic source, a deuterium triglycine sulfate Peltier-cooled detector, and a 60-sample autosampler (Auto-Diff, Pike Technologies). Soil subsamples (approximately 100 mg) were placed, as

**Table 2. Contaminated soils from Victoria (Melbourne and Ballarat), Sydney, and Adelaide (M, B, A and S, respectively). Data show sampling depth, number of samples (*n*) available, maximum total petroleum hydrocarbon (TPH) content, SD of TPH, and site description.**

| Soil | Depth | *n*† | Max.TPH | SD | Description |
|---|---|---|---|---|---|
| | m | | — mg kg$^{-1}$ — | | |
| M | 0–2.5 | 45 | 30,000 | 6848 | "fill material" from unknown sources and quaternary-aged Melbourne sand; <1 m thick; hard, orange/brown, high plasticity sandy clay; some rich organic topsoil; silty sand to clayey sand at depth with low plasticity |
| B | 1–2.5 | 89 | 32,600 | 4600 | saw-mill site with heavy TPH contamination at various point sources; dark brown surface soil with reddish-yellow clay subsoil; USDA classification Alfisol |
| A | 0–1 | 41 | 12,200 | 2560 | old gas works and workshops, heavy contamination of TPH; brown sandy loam subsoil under gravel overburden |
| S | – | 30 | 6,200 | 1981 | two piles of contaminated waste: Pile 1, yellow to red sand with carbonate nodules throughout; Pile 2, darker soil with organic material, clay, ink waste |

† Number of samples actually used was less than this number due to missing data and outliers.

received, into 10-mm diameter stainless steel sample cups, leveled, and loaded into the autosampler. Spectra were initially recorded on as-received samples in the frequency (wavenumber) range 7800 to 450 $cm^{-1}$ at a resolution of 8 $cm^{-1}$, using a silicon carbide reference disc (PerkinElmer Life and Analytical Sciences Pty Ltd) as a background (assumed to have a reflectance = 1 or 100%). The scans were repeated after allowing the subsamples to air dry overnight in their sample cups. Spectra were expressed in pseudo absorbance (A) units (where $A = \log_{10}$ reflectance$^{-1}$)

Spectra were converted from the PE instrument (.SP) format into Grams (.SPC) files using the Grams-AI software (Thermo Fisher Scientific) before importing into Unscrambler V9.8 (Camo AS) spreadsheets for regression analysis. Spectral assignments were based on MIR soil studies by Nguyen et al. (1991) and Janik et al. (1998)

## Partial Least Squares Regression Modeling

Partial least squares regression models were derived from the DRIFT spectra ($x$, predictor variables) and analytical TPH data ($y$, dependent variables). Model training was validated by cross-validation (Wold et al., 1983; Geladi and Kowalski, 1986) and by prediction validation using a randomly selected "test" set of 50 samples in the TPH range 0 to 15,000 mg $kg^{-1}$. Prediction models for TPH concentration were developed for the reference soil–TPH mixtures and the field-contaminated soils.

Spectra were preprocessed with mean centering and the *Unscrambler* linear Detrend correction across the full spectral range. Different PLS models were performed across the following spectral ranges to assess the impact of the various frequencies on the regression accuracy: (i) the NIR first-overtone 4540 to 4060 $cm^{-1}$ frequency region; (ii) the MIR 3000 to 2600 $cm^{-1}$ range corresponding to the alkyl $-CH_2$ and $-CH_3$ fundamental stretching frequencies, including the weak peak at 2730 $cm^{-1}$; (iii) the MIR 2980 to 2950 $cm^{-1}$ shoulder corresponding to the $-CH_3$ asymmetric stretching frequency; (iv) the MIR 2880 to 2800 $cm^{-1}$ range corresponding to the alkyl $-CH_2$ fundamental stretching frequency; (v) the MIR 2730 $cm^{-1}$ peak; and (vi) the MIR 2980 to 2950 $cm^{-1}$ shoulder plus the 2730 $cm^{-1}$ peak (2770–2650 $cm^{-1}$).

Partial least squares models were cross-validated against the measured TPH values during model training until a minimum regression residual and maximum coefficient of determination ($R^2$) were obtained with the optimum number of PLS factors. The PLS regression statistics are reported in terms of $R^2$, root mean square error (RMSE), and the ratio of prediction deviation (RPD = SD/RMSE), where SD is standard deviation. The value of RPD was used to indicate the quality of the predictions, where values of RPD < 1.5 are considered "poor", 1.5 to 2.0 considered "indicator," 2.0 to 3.0 considered "good", and >3.0 as "analytical" quality (Sudduth and Hummel, 1996; Chang et al., 2001; Fearn, 2002; Janik et al., 2009). Further validation was performed by random selection of a separate 50- sample field "test" set from the set of 199 (0–15,000 mg $kg^{-1}$) samples and predicting TPH concentration from a calibration built from the remaining samples.

# RESULTS AND DISCUSSION

## Spectral Interpretation

Figure 1 depicts portions of the NIR (5000–4000 $cm^{-1}$) and MIR (4000–1250 $cm^{-1}$) DRIFT spectra of liquid diesel, crude oil, and *n*-hexane deposited as liquid films onto a powdered KBr substrate. These first two hydrocarbon compounds often form a significant portion of TPH contaminants in soils. Peaks due to the fundamental alkyl stretching vibrations were observed in the frequency region near 2964 $cm^{-1}$ ($-CH_3$), 2930 $cm^{-1}$ ($-CH_2$), and 2855 $cm^{-1}$ ($-CH_2$). They are close to, but slightly higher than, those reported by Kalme et al. (2008) for diesel (2955, 2924, and 2855 $cm^{-1}$), with the increased frequency observed possibly due to the effects of specular reflectance distortion on the liquid films. Corresponding combination or overtone vibrations can be seen in the NIR for diesel at 4390, 4329, and 4257 $cm^{-1}$.

Interestingly, a small peak was observed for the liquid hydrocarbons near 2730 $cm^{-1}$. This peak has been previously interpreted as being due to the $-CH$ stretching mode of the $-CH=O$

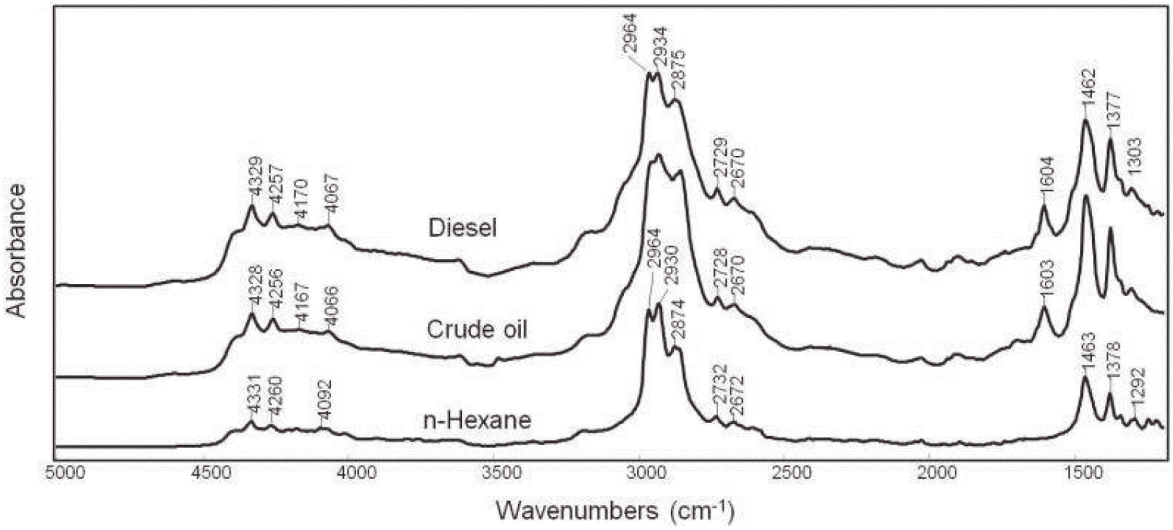

**Fig. 1. Portions of the Fourier-transform near-infrared (5000–4000 $cm^{-1}$) and mid-infrared (4000–2000 $cm^{-1}$) spectra of diesel, crude oil, and *n*-hexane in powdered KBr.**

aldehyde group in diesel after enzymatic degradation (Kalme et al., 2008). Aldehyde compounds, however, must also contain the –C=O group, which absorbs near 1740 $cm^{-1}$. Because such a carbonyl peak was not observed for crude oil or diesel spectra in this study, and importantly was not observed in *n*-hexane, which also shows the 2730 $cm^{-1}$ peak, an alternative explanation is needed. The explanation of the 2730 $cm^{-1}$ peak proposed here is that it is probably due to an overtone of the $-CH_3$ deformation vibration near 1377 $cm^{-1}$. As such, this peak has potential use for the detection and quantification of TPH because it is very weak in natural SOM compared with the fundamental alkyl intensities (Forrester et al., 2011).

The peaks and general spectral patterns of the liquid film hydrocarbons on the surfaces of a nonabsorbing matrix such as sand or, in this case, KBr depicted in Fig. 1 are, however, different from those of hydrocarbons. Liquid films are characterized by shifts, or distortions, of the usual spectral features due to the effects of a refractive index in the specular reflectance component of diffuse reflectance (Griffiths and Fuller, 1982). The "real" spectrum of a hydrocarbon, sorbed onto a matrix such as soil, should in fact resemble that of an alkyl material sorbed onto clay or SOM in soil rather than the spectra of liquid films as illustrated in Fig. 1. Figure 2 depicts the Fourier-transform MIR spectra of diesel sorbed onto a clay (a smectite) and onto a sand in the alkyl stretching region compared with that of natural SOM extracted from a soil (standard soil SS6).

The fundamental $-CH_3$ stretching vibration at 2964 $cm^{-1}$ in the diesel film was shifted down to a weaker peak at 2954 $cm^{-1}$, the $-CH_2$ peak at 2938 $cm^{-1}$ shifted to 2925 $cm^{-1}$ (now far stronger), the 2872 to 2861 $cm^{-1}$ peak envelope shifted to 2854 $cm^{-1}$, and the peaks near 2730 and 2670 $cm^{-1}$ were significantly reduced in intensity. These changes were attributed to a change in the reflection mechanism from specular reflectance to diffuse reflectance, and the spectra now closely resembled those of SOM and agreed with those of Kalme et al. (2008). This observation has implications in real soil systems contaminated with TPH where the amount of clay or other absorbing matrix in a sample can, for some soil types, have a profound effect on the intensity of the measured TPH spectrum.

Soil organic matter alkyl material is mostly long-chain $-CH_2$ groups in cell wall lipid material, with relatively few $-CH_3$ groups (Janik et al., 1998). Because medium-length alkanes, such as $C_{15}$ to $C_{28}$ diesel type of compounds, have a higher proportion of methyl $-CH_3$ to $-CH_2$ than SOM, the relatively stronger peaks attributed to methyl groups in TPH may be able to be used to predict TPH even in the presence of SOM. The $-CH_3$ 2954 $cm^{-1}$ peak in the diesel–clay spectrum in Fig. 2 is clearly stronger than the weak shoulder in this region for SOM. This is confirmed by the strong peak remaining at 2956 $cm^{-1}$ resulting from the subtraction of the SOM spectrum from the diesel–clay spectrum in Fig. 2. The contribution from $-CH_3$ groups in the shorter C chains of diesel also appeared relatively much stronger than in the longer chain SOM molecules.

In addition to the possible effects of surface reflection and interference by SOM peaks with those of TPH, there is also the question of overlap or masking by soil mineral components. Representative MIR spectra of some reference minerals are illustrated in Fig. 3. Most soils have significant amounts of one or more of these components, represented in the DRIFT spectra by peaks due to quartz (sand), 2:1 layer aluminosilicate clays (for example smectite and illite, often with interlayer water), kaolinite (a nonhydrated 1:1 layer aluminosilicate clay), and calcite ($CaCO_3$) (Nguyen et al., 1991).

In general, most soil minerals are free of MIR absorption peaks in the TPH alkyl stretching region. There is, however, considerable overlap due to carbonate minerals near 2980 to 2870, 2600 to 2500, and 1810 $cm^{-1}$ and in the alkyl deformation region near 1470 to 1360 $cm^{-1}$ with the fundamental $-CO_3$ stretching vibration near 1450 $cm^{-1}$ (shifted by specular reflectance distortion to 1375 $cm^{-1}$).

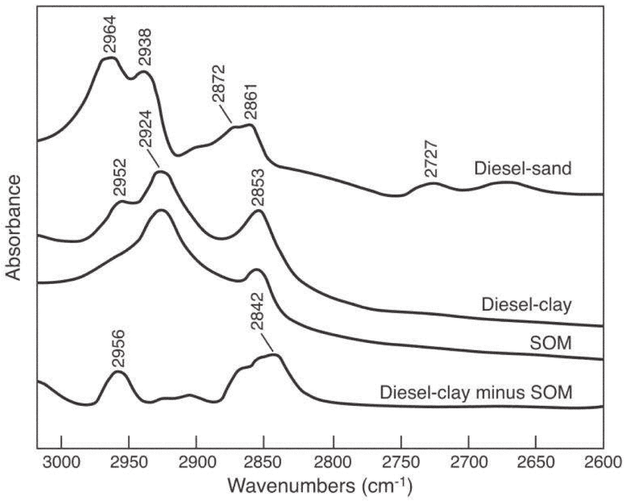

**Fig. 2. Fourier-transform mid-infrared (3000–2600 $cm^{-1}$) spectra of diesel on sand, diesel on clay, soil organic matter (SOM) from HF extraction of Soil SS6, and the subtracted spectrum of diesel on clay minus SOM.**

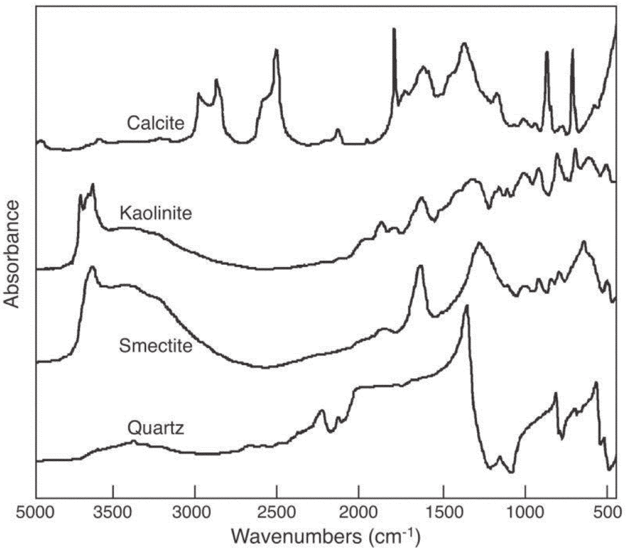

**Fig. 3. Diffuse-reflectance infrared Fourier-transform spectra of the reference minerals calcite, kaolinite, smectite, and quartz.**

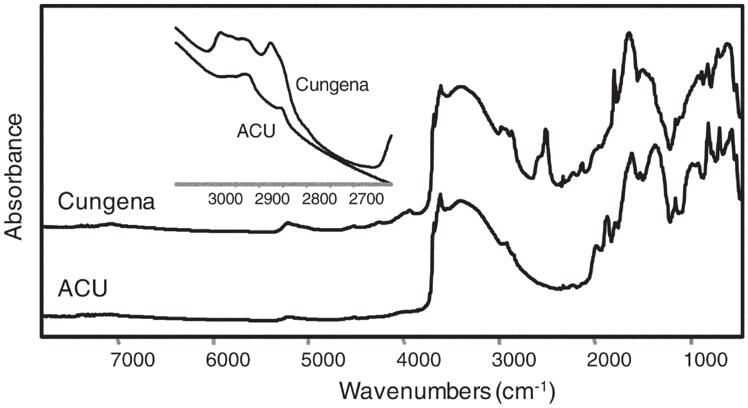

**Fig. 4. Fourier-transform infrared spectra of the Cungena and ACU-1 reference standard soils. The inset shows details of the overlap of carbonate peaks in the Cungena soil with soil organic matter alkyl peaks in the 3100–2700 cm$^{-1}$ region evident in the ACU soil.**

The spectra for the standard reference soils from Cungena and ACU-1 are depicted in Fig. 4. These two soils were used because they represented extremes of some major components found in soils that may impact on the spectral prediction of TPH. For example, the Cungena soil was high in carbonate and sand-sized particles. In contrast, the ACU-1 soil was high in SOM and silt. Other soils could have been chosen but these two were thought sufficient to demonstrate the possible interference of these soil characteristics for TPH predictions. The spectrum of the Cungena soil shows peaks due to smectite, illite, kaolinite clays (Al–OH at 3695 and 3630 cm$^{-1}$ and interlayer water at 3400 cm$^{-1}$), and carbonate (calcite near 2980–2870, 2518, and 1800 cm$^{-1}$) (see inset in Fig. 4). There is also evidence for some minor SOM alkyl peaks near 2930 and 2850 cm$^{-1}$. The ACU-1 soil was also characterized by smectite, illite, and kaolinite and SOM but no carbonate. Water peaks, possibly associated with SOM, smectite, and disordered illite, were observed near 3600 to 3000 cm$^{-1}$. Peaks in the rest of the spectrum, from about 2000 to 450 cm$^{-1}$, were due mostly to quartz.

The spectrum for the TPH field-contaminated soils across the full NIR plus MIR frequency range is characterized by peaks due to smectite and illite clays, organic components (SOM plus TPH alkyl peaks), and carbonate (Fig. 5). The peaks below 2000 cm$^{-1}$ are due mostly to quartz. The NIR spectra showed the presence of water (5220 cm$^{-1}$), kaolinite (7060 and 4530 cm$^{-1}$), and alkyl peaks (4320, 4250, and 4190 cm$^{-1}$). The water peaks were thought to be associated with the SOM and clays. Peaks due to alkyl components in SOM and TPH are shown in more detail in the inset.

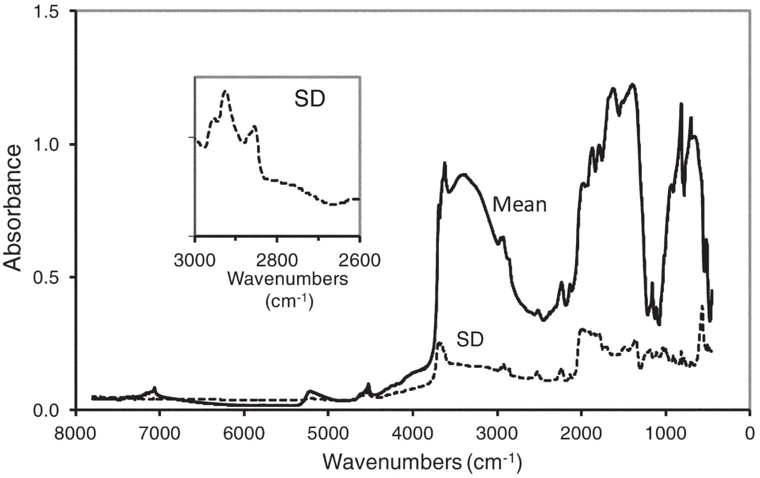

**Fig. 5. Fourier-transform infrared spectra (mean and standared deviation, SD) of the combined set of total petroleum hydrocarbon polluted soils from four sites. The inset shows SD plot in the 3100 to 2600 cm$^{-1}$ region.**

## Regression Analysis

### Standard Spiked Soils

According to Grabiec-Raczak et al. (2005), IR-based methods have typically measured the absorbance at a single frequency (e.g., close to 2930 cm$^{-1}$), which corresponds to the stretching of aliphatic $-CH_2$ groups. Others have used multiple frequencies including absorption bands at about 2960 cm$^{-1}$ for $-CH_3$ groups. Hazel et al. (1997) showed that there was a strong reduction in the TPH NIR intensity with respect to soil water content, but it was still possible to predict diesel concentrations in wet soils from MIR spectral frequencies in the 3500 to 2500 cm$^{-1}$ range using PLS regression models. In contrast, their results using univariate regression models based solely on the band at 2850 cm$^{-1}$ were meaningless.

The PLS regression coefficients for the models developed for the two standard soils spiked with crude oil are depicted in Fig. 6. The plots show peaks in the NIR (4388, 4330, and 4254 cm$^{-1}$) and MIR (2950, 2920, and 2850 cm$^{-1}$) spectral regions due to the high correlation between spectra and TPH concentration.

The results of cross-validation for the ACU-1 and Cungena soils are presented in Table 3. The coefficient of determinations ($R^2$) for the full MIR 3000 to 2650 cm$^{-1}$ spectral region were 0.96 and 0.90, respectively, for the two soils. Similar regressions were obtained for the ACU-1 soil by using only the narrow regions near 2980 to 2950 cm$^{-1}$ and at 2750 to 2650 cm$^{-1}$. The optimum model for Cungena ($R^2$ = 0.94) was obtained with a model using only the 2980 to 2950 cm$^{-1}$ frequencies, but significantly worse accuracy resulted from using only the 2750 to 2650 cm$^{-1}$ region ($R^2$ = 0.62 for five principal components). It thus appeared that PLS regression for the calcareous Cungena soil using the 2750 to 2650 cm$^{-1}$ region may have been adversely affected by the presence of carbonate peaks in the 2600 to 2520 cm$^{-1}$ region of the spectra. Clearly, therefore, the 2730 cm$^{-1}$ peak alone was not able to fully model crude oil concentrations spiked into the calcareous Cungena soil. With regard to the NIR frequency range (4780–4000 cm$^{-1}$), the precision of THP regression for the ACU-1 soil was significantly lower than for the MIR range ($R^2$ = 0.82). Results for the Cungena soil using the NIR were improved compared with those of the ACU-1 soil, resulting in an $R^2$ = 0.93. These results suggested that MIR PLS may be capable of accurate prediction of TPH in both calcareous and noncalcareous soils using the appropriate frequency regions.

## Field-Contaminated Soils

The soils were combined to form a composite set for PLS modeling. Because of the very large concentration range and skewed distribution, however, it was considered unrealistic to fit a single regression model to all the samples. Four sets of samples were therefore selected for PLS modeling: the full range (205 samples, 0–32,600 mg kg$^{-1}$), a low-range set (181 samples, 0–5000 mg kg$^{-1}$), an intermediate-range set (199 samples, 0–15,000 mg kg$^{-1}$), and a high-range set (21 samples, 5000–32,600 mg kg$^{-1}$).

There appeared to be a significant difference in the first PLS loadings for the low and high TPH ranges (0–5000 and 5000–32,600 mg kg$^{-1}$) in the 3000 to 2600 cm$^{-1}$ region (see Fig. 7). Peaks characteristic of TPH (peaks at 4954, 2925, and 2854 cm$^{-1}$) were observed in Loading 1 for the low-range TPH, while the strongest peaks for the high TPH range in Loading 1 were at 2954 to 2914, 2848, and 2730 to 2670 cm$^{-1}$. The cross-validation statistics and regression plots (measured vs. PLS-predicted) derived from the air-dried soil models are presented in Table 4.

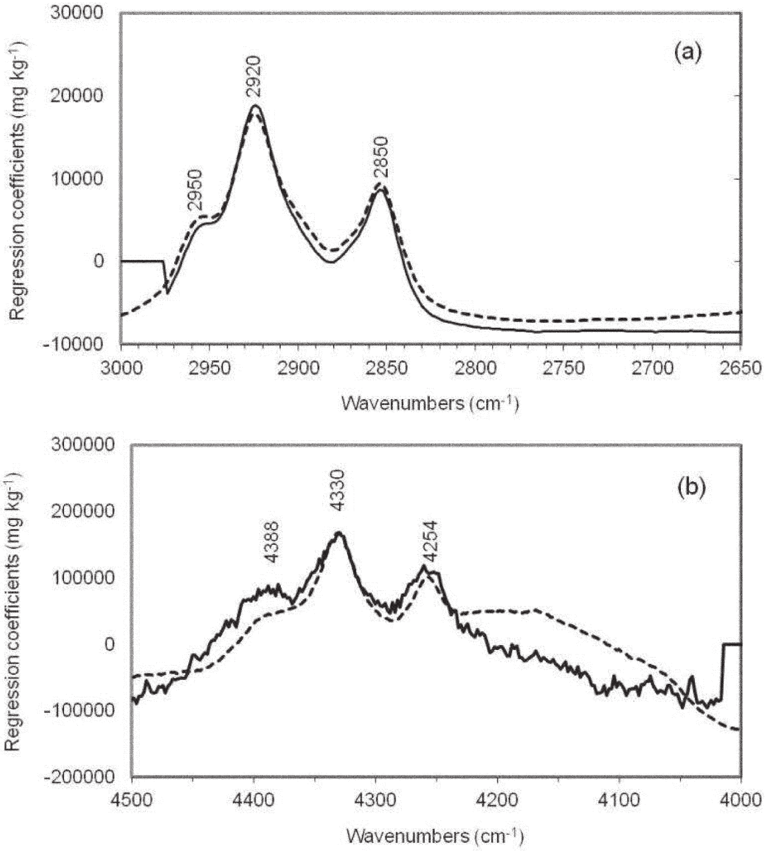

**Fig. 6. Cross-validation regression coefficients for total petroleum hydrocarbon contents of crude oil sorbed into the ACU-1 soil (solid line) and Cungena soil (dashed line) using the (a) near-infrared spectral range and (b) the mid-infrared spectral range.**

Partial least squares regression for the complete set of combined samples, using the NIR in the 4540 to 4120 cm$^{-1}$ region and the four MIR spectral ranges in the 3000 to 2600 cm$^{-1}$ region, resulted in varying calibration precisions ranging from $R^2$ = 0.67 to 0.44. Six samples with high leverage were thought to be the cause of poor regression and were subsequently omitted from further regression analyses as outliers. Two of these (26B in duplicate) had very much higher TPH concentrations than the other samples (32,600 mg kg$^{-1}$ laboratory value vs. 15,240 mg kg$^{-1}$ predicted with the 3000–2600 cm$^{-1}$ high-range sample set) and were thus possibly outside the range of the calibration models based on the remaining majority of samples.

Three samples (12B at 12,100 mg kg$^{-1}$ laboratory vs. 3170 mg kg$^{-1}$ predicted, 30C at 7350 mg kg$^{-1}$ laboratory vs. 4531 mg kg$^{-1}$ predicted, and 30D at 25,800 mg kg$^{-1}$ laboratory vs. 9523 mg kg$^{-1}$ predicted) had spectral characteristics of adsorbed water in the 3500 to 3200 cm$^{-1}$ range due to very high levels of either smectite or natural SOM, which may cause sorption of most of the TPH within the soil matrix. The sixth sample (ALN-1-1 at 520 mg kg$^{-1}$ laboratory vs. 6873 mg kg$^{-1}$ predicted) was characterized by quartz sand and had an unusual alkyl signature with very intense $-CH_2$ peaks but almost no observed $-CH_3$ peak.

The majority of samples (a total of 181) were in the 0 to 5000 mg kg$^{-1}$ range and resulted in $R^2$ values of 0.51 (RMSE = 702 mg kg$^{-1}$, RPD = 1.1) for the NIR to 0.75 (RMSE = 497 mg kg$^{-1}$, RPD = 2.0) for the 2980 to 2950 plus 2770 to

**Table 3. Partial least squares cross-validation statistics for crude oil containing total petroleum hydrocarbon (TCH) spiked into the standard soils ACU-1 and Cungena.**

| Soil | *n* | TPH range | Spectrum range | Spectrum type† | Principal components | $R^2$ | RMSE |
|---|---|---|---|---|---|---|---|
| | | mg kg$^{-1}$ | cm$^{-1}$ | | | | mg kg$^{-1}$ |
| ACU-1 | 11 | 257–25,000 | 4780–4000 | FT-NIR | 4 | 0.82 | 3097 |
| | 11 | 257–25,000 | 3000–2650 | FT-MIR | 3 | 0.96 | 1433 |
| | 11 | 257–25,000 | 2980–2950 | FT-MIR | 2 | 0.95 | 1681 |
| | 11 | 257–25,000 | 2750–2650 | FT-MIR | 4 | 0.93 | 1888 |
| Cungena | 11 | 257–25,000 | 4780–4000 | FT-NIR | 3 | 0.93 | 1975 |
| | 11 | 257–25,000 | 3000–2650 | FT-MIR | 2 | 0.90 | 2494 |
| | 11 | 257–25,000 | 2980–2950 | FT-MIR | 3 | 0.94 | 1838 |
| | 11 | 257–25,000 | 2750–2650 | FT-MIR | 5 | 0.62 | 4758 |

† FT-NIR, Fourier-transform near infrared; FT-MIR, Fourier-transform mid-infrared; $R^2$, coefficient of dermination; RMSE, root mean square error.

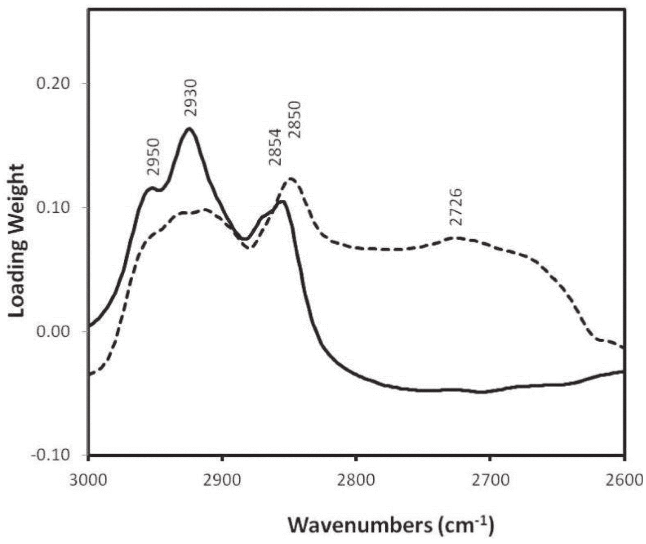

**Fig. 7. The partial least squares first loading weights for the combined contaminated soil set in the mid-infrared spectral range (3000–2600 cm$^{-1}$) for concentration ranges 0 to 5000 mg kg$^{-1}$ (solid line) and 5000 to 32,600 mg kg$^{-1}$ (dashed line).**

2650 cm$^{-1}$ range calibration. An RPD = 2 suggested an "indicator" quality prediction model. Extending this range of TPH to 15,000 mg kg$^{-1}$ markedly improved the calibration, probably resulting from the inclusion of some highly leveraging samples in the 5000 to 15,000 mg kg$^{-1}$ range of TPH. Cross-validation statistics now showed a range of $R^2$ from 0.84 (RMSE = 853 mg kg$^{-1}$, RPD = 2.4) for the NIR to 0.93 (RMSE = 564 mg kg$^{-1}$, RPD = 3.7) for the 2980 to 2950 plus 2770 to 2650 cm$^{-1}$ range calibration. The calibration using the full 3000 to 2600 cm$^{-1}$ spectral range gave very similar results, with an $R^2$ = 0.92, RMSE = 601 mg kg$^{-1}$, and RPD = 3.4, and thus both of these MIR calibrations could be considered to be of "analytical" quality.

As shown in Table 4, the inclusion of samples in the 5000 to 15,000 mg kg$^{-1}$ range of TPH not only improved the $R^2$ of the models obtained within the 0 to 5000 mg kg$^{-1}$ range but also increased the RPD, resulting in similar low RMSE values in spite of having doubled the TPH range and standard deviation. The regression for the air-dried samples in the 0 to 15,000 mg kg$^{-1}$ range is illustrated in Fig. 8. It is clear from Fig. 8 that most of the samples are distributed at low concentrations and only a few at high concentrations of TPH.

**Table 4. Partial least squares cross-validation statistics for total petroleum hydrocarbon (TPH) concentration ranges (0–5000, 5000–32,600, 0–32,600, and 0–15,000 mg kg$^{-1}$) in the air-dried (AD) and as-received (AR) contaminated soils using models derived for the six frequency ranges.**

| Frequency range | Statistic† | AD | | | | AR |
|---|---|---|---|---|---|---|
| | | 0–5000 | 0–15,000 | 5000–32,600 | 0–32,600 | 0–15,000 |
| cm$^{-1}$ | | | | | | |
| All | *n* | 181 | 199 | 21 | 205 | 199 |
| | SD | 1002 | 2102 | 8601 | 4185 | 2102 |
| 4540–4120 | PCs, no. | 9 | 9 | 7 | 9 | 14 |
| | $R^2$ | 0.51 | 0.84 | 0.63 | 0.44 | 0.46 |
| | RMSE, mg kg$^{-1}$ | 702 | 853 | 5224 | 3122 | 7461 |
| | RPD | 1.1 | 2.4 | 1.5 | 1.0 | 1.4 |
| 2777–2650 | PCs, no. | 6 | 6 | 8 | 8 | 6 |
| | $R^2$ | 0.64 | 0.85 | 0.78 | 0.52 | 0.80 |
| | RMSE, mg kg$^{-1}$ | 562 | 806 | 4008 | 2917 | 1011 |
| | RPD | 1.7 | 2.6 | 2.1 | 1.4 | 2.1 |
| 2980–2950 | PCs, no. | 4 | 3 | 3 | 9 | 2 |
| | $R^2$ | 0.70 | 0.91 | 0.14 | 0.61 | 0.83 |
| | RMSE, mg kg$^{-1}$ | 549 | 621 | 8272 | 2631 | 928 |
| | RPD | 1.8 | 3.4 | 1.0 | 1.6 | 2.2 |
| 2880–2800 | PCs, no. | 4 | 5 | 10 | 9 | 6 |
| | $R^2$ | 0.70 | 0.90 | 0.40 | 0.61 | 0.86 |
| | RMSE, mg kg$^{-1}$ | 549 | 649 | 7864 | 2631 | 852 |
| | RPD | 1.8 | 3.2 | 1.1 | 1.6 | 2.5 |
| 2980–2950, 2777–2650 | PCs, no. | 8 | 6 | 3 | 11 | 3 |
| | $R^2$ | 0.75 | 0.93 | 0.14 | 0.67 | 0.82 |
| | RMSE, mg kg$^{-1}$ | 497 | 564 | 8272 | 2426 | 971 |
| | RPD | 2.0 | 3.7 | 1.0 | 1.7 | 2.1 |
| 3000–2600 | PCs, no. | 8 | 7 | 14 | 11 | 5 |
| | $R^2$ | 0.73 | 0.92 | 0.46 | 0.55 | 0.85 |
| | RMSE, mg kg$^{-1}$ | 495 | 601 | 7461 | 2860 | 876 |
| | RPD | 1.7 | 3.4 | 1.4 | 1.3 | 2.4 |

† *n*, number of samples; PCs, partial least squares principal components; $R^2$, coefficient of determination; RMSE, root mean square error; RPD, residual predictive deviation; SD, standard deviation.

Modeling the high-range TPH samples, from 5000 to 32,600 mg kg$^{-1}$, was clearly a problem and presented difficulties. Table 4 shows that, apart from the NIR ($R^2 = 0.63$, RMSE = 5224 mg kg$^{-1}$, RPD = 1.5) and the 2770 to 2650 cm$^{-1}$ ($R^2 = 0.78$, RMSE = 4008 mg kg$^{-1}$, RPD = 2.1) models, none of the other calibration frequency ranges used in this study were able to derive a viable PLS model for the high concentrations. Values of $R^2$ ranged from near 0 to 0.14 with RPD values of about 1.0, reflective of the fact that the usual alkyl spectral profile for the low TPH concentration samples was different from that of the high-concentration samples (illustrated in Fig. 7). Also important, particularly at low concentrations, is the relatively low signal intensity of the 2770 to 2650 cm$^{-1}$ peak compared with the noise level of the IR spectrometer. Part of this noise is due to detector noise, but there are often the very weak background peaks caused by uncompensated water vapor vibrations superimposed on the spectra of the soils.

At low TPH concentrations, spectral noise can significantly downgrade the calibration lower detection limits of the 2770 to 2650 cm$^{-1}$ model. Because the spectral signatures of these high-concentration samples were in some cases markedly different from those of the lower concentration samples calibration, a separate PLS model for high concentrations was therefore required. The model from the MIR 2770 to 2650 cm$^{-1}$ range suggested that it could be considered to be of "indicator" quality, while the NIR 4540 to 4120 cm$^{-1}$ range was only marginally regarded as "poor-indicator" quality. Furthermore, it appears that the 2770 to 2650 cm$^{-1}$ range may have the most use in predicting high TPH concentrations rather than the low to medium concentrations.

With all MIR frequency ranges capable of producing RMSE errors <1000 mg kg$^{-1}$ for the 0 to 15,000 mg kg$^{-1}$ concentration range, the results suggest that the accuracy of such calibrations could be regarded as sufficient for screening purposes. As illustrated in this study, the IR spectra are highly sensitive to TPH compounds and therefore should, in principle, result in accurate predictions. The robustness of the mid-range calibration model was further tested by using a randomly selected "test" set as a more independent validation for prediction of TPH concentrations in the MIR range (3000–2600 cm$^{-1}$). Prediction of the 50 "test" set samples resulted in an $R^2 = 0.89$, RMSE = 767 mg kg$^{-1}$, and an RPD = 2.7. As expected, this was slightly less precise than that of the cross-validation of the complete set of 199 samples ($R^2 = 0.92$, RMSE = 601 mg kg$^{-1}$, and RPD = 3.4) but nevertheless confirmed from the high RPD that the prediction model still achieved a good quality.

Our results showed that, in spite of the high $R^2$ values for many of the models developed, there are still sources of error that affect the model accuracy. It should be noted that, when assessing the prediction errors, not all the error is due to the IR calibrations ; a significant component of the error can be attributed to the laboratory reference data, possibly due either to attributing part of the measured TPH to SOM, to sample heterogeneity (Malley et al., 1999), or, in some cases as shown in this study, to soil matrix composition.

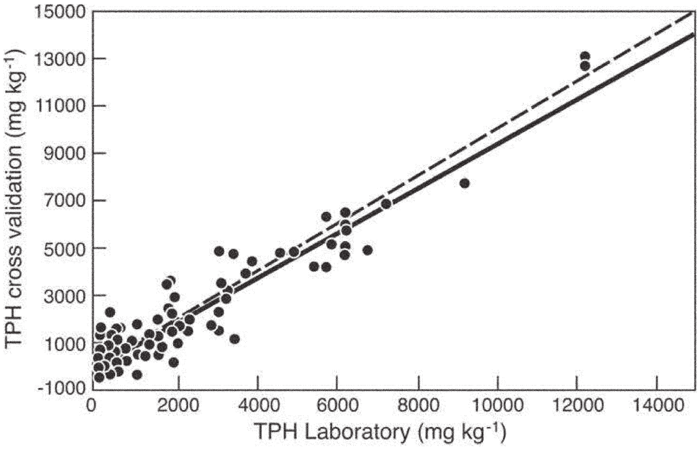

**Fig. 8. Cross-validation predicted vs. laboratory-determined total petroleum hydrocarbon (TPH) content (mg kg$^{-1}$) for 199 air-dried soil samples using the mid-infrared region 3000 to 2600 cm$^{-1}$ at a concentration range of 0 to 15,000 mg kg$^{-1}$. The solid line indicates the line of best fit (slope = 0.93, intercept = 93, and $R^2 = 0.92$) and the dashed line the 1:1 line.**

The scanning of neat, unground or unsieved samples can result in major sources of error, often due to inter- and intraparticle sample heterogeneity. While interparticle heterogeneity can be partly addressed by repetitive sample scanning, the effect of intraparticle heterogeneity is more difficult to overcome because it is caused by a difference in the composition of soil micro- and macroaggregates from the outer surface to the depth penetrated by the IR beam.

Further complications can arise from the use of as-received field samples. The presence of water in as-is (field-moist) samples has been suggested as one source of error in the models developed for the prediction of TPH in contaminated soils using NIR, with $R^2$ values approximately 0.71 (Malley et al., 1999). Hazel et al. (1997), however, demonstrated that MIR spectroscopy could be used for identifying and quantifying a fixed concentration of marine diesel fuel in soils across a wide range of moisture content by using multivariate calibration techniques including principal component analysis and PLS regressions. Similar conclusions could be taken from the study of Chakraborty et al. (2010), who compared the performance of NIR-PLS models for the prediction of TPH in field-moist, air-dried, and air-dried ground soil samples. In their investigation, TPH estimated by the field-moist intact first-derivative PLS model had the greatest accuracy ($R^2 = 0.64$), with the worst corresponding to the model using air-dried ground samples. These factors may all have implications for the use of IR-based reflection methods in direct field application.

As a comparison with the dried samples, and to illustrate the effect of using as-received wet field samples on prediction accuracy and precision, Table 4 presents data for the 0 to 15,000 mg kg$^{-1}$ TPH range samples modeled in this present study using as-received (and often wet) samples. Cross-validation precision of the as-received calibration was reduced in all spectral ranges from that of the air-dry sample set (see Table 4), possibly due partly to the effects of reflection from the surface films of water and also on the basis of TPH analytical data being presented on an air-dry basis. In spite of these reservations, the MIR-PLS method for TPH determinations in as-received samples appears to be viable. Further research on the effects of soil water contents, the treatment of sample heterogeneity and spatial variability, wider ranges of TPH concentration, and adaptation to hand-held technology still require further study.

## CONCLUSIONS

This study confirmed that the use of DRIFT spectroscopy with PLS regression was able to provide viable and accurate models for the prediction of TPH in the 0 to 15,000 mg $kg^{-1}$ concentration range, with the MIR range outperforming the NIR region in this study. Initial experiments with two reference soils representing contrasting soil properties such as organic matter and carbonate, demonstrated the ability of PLS regression to deal with potential interferences with these compounds for the prediction of TPH concentration.

Further experiments with a range of soils from contaminated sites confirmed the high accuracy achievable with the IR–PLS method. The optimum frequencies for regression modeling were in the combined 2980 to 2950 and 2777 to 2650 $cm^{-1}$ spectral range, resulting in "analytical" quality calibrations (RPD = 3.7, $R^2 = 0.93$, and RMSE = 564). Similar results could be obtained for a calibration using the full 3000 to 2600 $cm^{-1}$ spectral range. The use of a separate 2770 to 2650 $cm^{-1}$ range was required for high TPH concentrations, giving an "indicator" quality model. A small decrease in precision for wet, as-received soils was shown for the 3000 to 2600 $cm^{-1}$ frequencies.

The peak near 2730 $cm^{-1}$ was identified as potentially specific to TPH, with very little overlap with natural SOM and soil carbonate. While exclusive use of this peak was successful in the prediction of very high TPH concentrations, its effectiveness was compromised by the relatively low intensity of this peak at low TPH concentrations. It appears that the determination of the relative proportion of methyl $-CH_3$ to alkyl $-CH_2$ and the intensity of the methyl peak near 2950 $cm^{-1}$ was sufficient in most cases to overcome any overlap with natural SOM or carbonate.

The technique offers the potential to provide a very rapid and inexpensive method to characterize TPH concentrations at contaminated sites. Given the specificity of the spectral signatures identified, the method also offers the potential to overcome errors in laboratory analysis by GC-FID due to SOM. With a range of smaller hand-held IR devices becoming available, the method also provides an opportunity to adapt the technique for in-field analysis.

## ACKNOWLEDGMENT

José M. Soriano gratefully acknowledges the Department of Education (Government of Valencia) for a post-doctoral fellowship (APOSTD/2011/034). The authors also acknowledge the support and funding for this work by Ziltek Pty Ltd. (Thebarton, South Australia).

## REFERENCES

ASTM. 2011. D7678-11: Standard test method for total petroleum hydrocarbons (TPH) in water and wastewater with solvent extraction using mid-IR laser spectroscopy. ASTM Int., West Conshohocken, PA.

Chakraborty, S., D.C. Weindorf, C.L.S. Morgan, Y.F. Ge, J.M. Galbraith, B. Li, and C.S. Kahlon. 2010. Rapid identification of oil-contaminated soils using visible near-infrared diffuse reflectance spectroscopy. J. Environ. Qual. 39:1378–1387. doi:10.2134/jeq2010.0183

Chakraborty, S., D.C. Weindorf, Y. Zhu, B. Li, C.L.S. Morgan, Y. Ge, and J. Galbraith. 2012. Spectral reflectance variability from soil physicochemical properties in oil contaminated soils. Geoderma 177–178:80–89. doi:10.1016/j.geoderma.2012.01.018

Chang, C.W., D.A. Laird, M.J. Mausbach, and C.R. Hurburgh, Jr. 2001. Near-infrared reflectance spectroscopy: Principal components regression analyses of soil properties. Soil Sci. Soc. Am. J. 65:480–490. doi:10.2136/sssaj2001.652480x

Fearn, T. 2002. Assessing calibrations: SEP, RPD, RER and $R^2$. NIR News 13:12–14

Forrester, S.T., L.J. Janik, and M.J. McLaughlin. 2009. Application of MIR spectroscopy and multivariate statistical analysis to the analysis of hydrocarbon-contaminated soils and solid wastes. Sci. Rep. Ser. CSIRO Land and Water, Adelaide, SA, Australia.

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

Freeman, B.S., and P.K. Krishnan. 2007. Replacing the method 418.1 test for total petroleum hydrocarbons using non-ozone depleting substances. Integrated Environ. Solutions, Safat, Kuwait. http://www.iestech.net/download/kwlab_replace418.pdf.

Gallagher, N.B., P. Gassman, and T. Blake. 2008. Strategies for detecting organic liquids on soils using mid-infrared reflection spectroscopy. Environ. Sci. Technol. 42:5700–5705. doi:10.1021/es8005404

Geladi, P., and B.R. Kowalski. 1986. Partial least-squares regression: A tutorial. Anal. Chim. Acta 185:1–17. doi:10.1016/0003-2670(86)80028-9

Grabiec-Raczak, E., M. Kocon, B. Kaminska, A. Sztuczynska, and D. Dabrowska. 2005. Determination of oil in environmental samples using FTIR spectrometry technique. In: M. Waclawek and W. Waclawek, editors, ECOpole '05, Conference Proceedings, Jamrozowa Polana, Poland. 20–22 Oct. 2005. Towarzystwo Chemii i Inzynierii Ekologicznej, Jamrozowa, Poland. p. 85–89.

Griffiths, P.R., and M.P. Fuller. 1982. Mid-infrared spectrometry of powdered samples. In: R.J.H. Clark and R.E. Hester, editors, Advances in infrared and Raman spectroscopy. Vol. 9. Heyden & Sons, London. p. 63–129.

Hazel, G., F. Bucholtz, I.D. Aggarwal, G. Nau, and K.J. Ewing. 1997. Multivariate analysis of mid-IR FT-IR spectra of hydrocarbon-contaminated wet soils. Appl. Spectrosc. 51:984–989. doi:10.1366/0003702971941359

ISO. 2004. Method 16703:2004. Soil quality: Determination of content of hydrocarbons in the range $C_{10}$ to $C_{40}$ by gas chromatography. ISO, Geneva, Switzerland.

Janik, L.J., S.T. Forrester, and A. Rawson. 2009. The prediction of soil chemical and physical properties from mid-infrared spectroscopy and combined partial least-squares regression and neural networks (PLS-NN) analysis. Chemom. Intell. Lab. Syst. 97:179–188. doi:10.1016/j.chemolab.2009.04.005

Janik, L.J., R.H. Merry, and J.O. Skjemstad. 1998. Can mid infrared diffuse reflectance analysis replace soil extractions? Aust. J. Exp. Agric. 38:681–696. doi:10.1071/EA97144

Janik, L.J., and J.O. Skjemstad. 1995. Characterization and analysis of soils using mid-infrared partial least squares: II. Correlations with some laboratory data. Aust. J. Soil Res. 33:637–650. doi:10.1071/SR9950637

Kalme, S., G. Parshetti, S. Gomare, and S. Govindwar. 2008. Diesel and kerosene degradation by *Pseudomonas desmolyticm* NCIM 2112 and *Nocardia hydrocarbonoxydans* NCIM 2386. Curr. Microbiol. 56:581–586. doi:10.1007/s00284-008-9128-6

Kuang, B., H.S. Mahmood, M.Z. Quraishi, W.B. Hoogmoed, A.M. Mouazen, and E.J. van Henten. 2012. Sensing soil properties in the laboratory, in situ, and on-line: A review. Adv. Agron. 114:155–223. doi:10.1016/B978-0-12-394275-3.00003-1

Malley, D.F., K.N. Hunter, and G.R.B. Webster. 1999. Analysis of diesel fuel contamination in soils by near-infrared reflectance spectrometry and solid phase microextraction–gas chromatography. J. Soil Contam. 8:481–489. doi:10.1080/10588339991339423

National Environment Protection Council. 1999a. Schedule B(1): Guideline on the Investigation Levels for Soil and Groundwater. NEP (Assessment of Site Contamination) Measure 1999. NEPC, Parkes, ACT. http://www.ephc.gov.au/sites/default/files/ASC_NEPMsch__01_Investigation_Levels_199912.pdf.

National Environment Protection Council. 1999b. Schedule B(3): Laboratory Analysis of Potentially Contaminated Soils, NEP (Assessment of Site Contamination) Measure 1999. NEPC, Parkes, ACT. http://www.ephc.gov.au/sites/default/files/ASC_NEPMsch__03_Lab_Analysis_199912.pdf.

Nguyen, T.T., L.J. Janik, and M. Raupach. 1991. Diffuse reflectance infrared Fourier transform (DRIFT) spectroscopy in soil studies. Aust. J. Soil Res. 29:49–67. doi:10.1071/SR9910049

Rayment, G.E., and F.R. Higginson. 1994. Australian laboratory handbook of soil and water chemical methods. Inkata Press, Melbourne, VIC, Australia. p. 29.

Reeves, J.B., III, R.F. Follett, G.W. McCarty, and J.M. Kimble. 2006. Can near or mid-infrared diffuse reflectance spectroscopy be used to determine soil carbon pools? Commun. Soil Sci. Plant Anal. 37:2307–2325. doi:10.1080/00103620600819461

Reeves, J.B., III, G.W. McCarty, and J.J. Meisinger. 1999. Near infrared reflectance spectroscopy for the analysis of agricultural soils. J. Near Infrared Spectrosc. 7:179–193. doi:10.1255/jnirs.248

Sadler, R., and D. Connell. 2003. Analytical methods for the determination of total petroleum hydrocarbons in soil. In: A. Langley et al., editors, Proceedings of the 5th National Workshop on the Assessment of Site Contamination, Adelaide, SA. May 2003. NEPC Serv. Corp., Adelaide, SA, Australia. p. 133–150.

Skjemstad, J.O., L.J. Janik, and J.A. Taylor. 1998. Non-living soil organic matter: What do we know about it? Aust. J. Exp. Agric. 38:667–680. doi:10.1071/EA97143

Sorak, D., L. Herberholz, S. Iwascek, S. Altinpinar, F. Pfeifer, and H.W. Siesler. 2012. New developments and applications of handheld Raman, mid-infrared, and near-infrared spectrometers. Appl. Spectrosc. Rev. 47:83–115. doi:10.1080/05704928.2011.625748

Sudduth, K.A., and J.W. Hummel. 1996. Geographic operating range evaluation of a NIR soil sensor. Trans. ASAE 39:1599–1604.

U.S. Environmental Protection Agency. 1978. Method 418.1: Test method for evaluating total recoverable petroleum hydrocarbon (spectrophotometric, infrared). USEPA, Washington, DC. http://www.oilbarriers.co.nz/tech/totalhydrocarbontestmethods418_1.pdf.

U.S. Environmental Protection Agency. 2003. Method 8015: Nonhalogenated organics using GC/FID. USEPA, Washington, DC. http://www.epa.gov/osw/hazard/testmethods/pdfs/8015d_r4.pdf.

Viscarra-Rossel, R.A., D.J.J. Walvoort, A.B. McBratney, L.J. Janik, and J.O. Skjemstad. 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131:59–75. doi:10.1016/j.geoderma.2005.03.007

Wold, S., H. Martens, and H. Wold. 1983. The multivariate calibration method in chemistry solved by the PLS method. In: A. Ruhe and B. Kagstrom, editors, Proceedings of the Conference on Matrix Pencils, Pite Havsbad, Sweden. 22–24 Mar. 1982. Lect. Notes Math. Springer-Verlag, Heidelberg. p. 286–293.