Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 207 (2019) 183–188

Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

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

Application of infrared spectrum for rapid classification of dominant petroleum hydrocarbon fractions for contaminated site assessment

Liang Wang [a,b,*], Ying Cheng [a,b], Dane Lamb [a,b], Raja Dharmarajan [a,b], Sreenivasulu Chadalavada [b], Ravi Naidu [a,b]

[a] Global Centre for Environmental Remediation, Research and Innovation Division, University of Newcastle, Callaghan, NSW 2308, Australia
[b] CRC for Contamination Assessment and Remediation of Environment, University of Newcastle, Callaghan, NSW 2308, Australia

ARTICLE INFO

Article history:
Received 20 July 2018
Received in revised form 13 September 2018
Accepted 14 September 2018
Available online 15 September 2018

Keywords:
Handheld FTIR
Carbon chain length (C_n)
Petroleum hydrocarbon
Alkanes
Rapid site assessment

ABSTRACT

In this study, the infrared spectrum (4000–400 cm^{-1}) was applied to identify and classify the different alkanes based on carbon chain length (C_n). It was found there were two bands coherent to the doublet at location 2954 and 2872 cm^{-1}, respectively can be applied to identify the fraction of carbon chains. From C_{20} to C_{37}, by the increase of the C_n, the intensities of the two bands were reduced as demonstrated. There were another two doublets existed at the region from 1480 to 1450 cm^{-1} and the region at 750 and 730 cm^{-1}. It was observed the intensity of one coherent band at each of these regions was increased following the increase on the C_n. The bands center at 1462 and 730 cm^{-1} were increased from C_{20} to C_{37}. The intensity ratio of the coherent bands can be applied to identify the C_n. Successfully identify four different petroleum products with different fractions of carbon chains in soil samples, is evidence the theory can be applied to investigate the fraction of carbon chains in soil. Coupling with handheld FTIR, it is possible to rapidly estimate the dominant fraction of C_n in soil in field.

1. Introduction

Petroleum is comprised of highly complex and varied mixtures of hydrocarbons [1]. For example, crude oil can consist of hundreds of individual petroleum hydrocarbon (PHC) compounds that represent from 50 to 98% of the total weight of crude oil [2]. The PHC fractions are defined by chemical structure, carbon number, and structure-activity relationships [3]. Equivalent carbon numbers (C_n) are used to describe the PHC fractions, which are based on a range of physical-chemical properties and simple partitioning models [4]. The chains under C_7, called naphthas, which are normally used as dry cleaning solvents, are very light and easily vaporised. The chains from C_7 to C_{11} are normally blended and used for petrol, which is more volatile than water. The sequence that follows is kerosene, from C_{12} to C_{15}, and then diesel and other heavier fuel oils, from C_{12} to C_{30}. Lubricating oils, such as engine and motor oils, have carbon chains above C_{30}. Hydrocarbons with higher carbon fractions, or longer the carbon chains, are considered as heavier fractions of PHCs. The heavier fractions of PHCs have high viscosity, density, boiling points and long residence times in soil [3].

When an oil spill or leakage occurs on land, the sequestration and diffusion of PHCs are subject to physical, chemical, and biological processes that further change their composition, toxicity, and distribution (partitioning) within soil, and are referred to as weathering processes [5]. The highly volatile fractions of PHCs will evaporate into the gas phase of the porous medium and may be lost to the atmosphere, adsorbed on to soil solids, or dissolved into soil water or groundwater [6]. Less volatile PHCs will diffuse as oil or non-aqueous liquid (NAPLs) forms, through the porous medium and be trapped in pores or adsorbed at mineral and organic matter surfaces. Linear compounds up to C30, for example, may dissipate within a month by volatilisation, decomposition and degradation. By contrast, those with up to C_{37} need at least 200 days. Indeed, the tar fraction consisting of long and complex chains can persist for years [7]. Investigating the carbon fractions in a PHC contaminated site can provide valuable site assessment information for remediation.

To measure the fractions of C_n of PHC contaminants in soil, they need to be extracted by standardised soil extraction procedures before instrumental analysis. For volatile compounds, such as BTEX and petrol, methanol extraction is suggested under EPA method 5035, or direct headspace analysis can be performed. For semi-volatile compounds Soxhlet, sonication and supercritical fluid extractions (SFE) are commonly employed [1]. High-performance liquid chromatography (HPLC) and gas chromatography (GC) are the commonly used instrumental techniques for PHC analyses. GC with mass spectrometry (GC/MS) or flame ionisation detection (GC/FID) are preferably used for individual PHC component determinations [8]. However, these approaches suffer from disadvantages compared with online *in situ* monitoring, including the associated laboratory costs and lengthy processing time, sample degradation and cross contamination. In addition, these conventional methodologies provide only a ‘snapshot’ in time. Consequently,

---

* Corresponding author at: Global Centre for Environmental Remediation, Research and Innovation Division, University of Newcastle, Callaghan, NSW 2308, Australia.
E-mail address: Liang.Wang@newcastle.edu.au (L. Wang).

https://doi.org/10.1016/j.saa.2018.09.024
1386-1425/

these methods require frequent sampling intervals to provide representative temporal variations in PHC levels. An alternative method for measuring PHCs is to use an infrared (IR) technique, such as EPA method 8440, whereby the PHCs are identified by their spectral bands. The analysis time and cost of IR methods are typically far less than the GC and HPLC based methodologies.

Nevertheless, the major disadvantage of Fourier Transformed IR (FTIR) is that the spectrum line shape (band) typically consists of a series of more or less overlapping bands representing absorbed or scattered individual molecules, which can be considered as interferences [9–12] [9–12]. Extracting the information and identifying the components from an overlapping IR spectrum is a key challenge to analysing a sample with an unknown mixture of PHCs. Czarnecki and Ozaki noted that the separation of very overlapping bands requires both an experienced researcher and some knowledge about the system being studied [12]. Therefore, infrared analyses in the region of 4000 to 400 cm^{-1} are commonly used for such chemical structure studies.

Irrespective of the length of the carbon chains, all the PHC fractions have similar band patterns in infrared spectral modes. As demonstrated, a similar doublet IR band between 3000 and 2800 cm^{-1}, and a smaller single band between 1500 and 1450 cm^{-1} presented for all the long-chain alkanes. These IR bands indicate the vibration of carbon hydrogen bindings. Therefore, the slight difference of PHC fractions in the mid-IR 'fingerprint' region (1500–700 cm^{-1}) become crucial for PHC determination. However, the signals at the mid-IR 'fingerprint' region are significantly weaker than higher wavelength regions. Additionally, they possess low signal to noise ratios with intensive bands overlapping. Hence, there have only been limited studies on applying these spectral techniques for determination of PHC fractions.

Coupling with a handheld FTIR, this study demonstrated a rapid PHC fraction classification method for *in situ* PHC weathering assessment. Long chain alkanes from Icosane (C_{20}) to heptatriacontane (C_{37}) have been investigated using the infrared spectral signals. A novel infrared spectral data processing method was applied to automatically baseline correct and band decompose the IR spectra without visual inspection. Quantitative analysis for each alkane was investigated using the decomposed sub-bands. Orthogonal experimental design (OED) was applied to generate the alkane mixtures with designed heterogeneous concentrations [14].

2. Methodology

2.1. Apparatus

Pure chemicals of the alkanes: icosane (C_{20}), hexacosane (C_{26}), octacosane (C_{28}), dotriacontane (C_{32}) and heptatriacontane (C_{37}) were obtained from Sigma Aldrich. The pure alkane chemicals were added into hexane (95%, Sigma Aldrich) to create 20 mL of 10 g/L standard solutions for C_{20} to C_{32}, and 5 g/L for C_{37}, respectively. It was observed that the alkanes with higher carbon chain numbers have lower a dissolution rate in hexane. 5 g/L C_{37} solutions needed to be prepared with the assistance of ultrasonic vibrations. The individual alkanes were spiked into potassium bromide (KBr) for characteristic sub-band identification.

All the samples were measured triplicated using a handheld FTIR (Agilent 4300), with an 8 cm^{-1} resolution, 32 sample scans, 64 background scans were co-added in the infrared 4000 to 600 cm^{-1} region, at a scanning velocity of 2.5 kHz, and 255 beam energy. All measurements for the spiked soil samples were made in diffuse reflectance infrared spectroscopy (DRIFTS) mode, sample non-destructively. It should be mentioned that there are several levels of spectral resolution: from 8 to 2 cm^{-1}. The lower the number present, the higher resolutions, and the detection limits and spectral features for a given compound can be improved through higher resolutions which need high-resolution interferometers. However, the relative cost to incorporate such a system for field monitoring would be exorbitant. Further, it would take a considerable amount of time for a scan and obtain point data for computational analysis. The sensitivity level chosen for this experiment is appropriate for reduced downtime and rapid screening for field related applications. With setting the resolution at 8 cm^{-1}, one measurement can be completed within thirty seconds with 32 scans. The infrared spectrum data can be obtained using Agilent Microlab PC software (Agilent). The spectrum data was transferred to the Resolutions Pro FTIR software for observation. The spectrum data was saved in comma separated value format (CSV), which is editable using Excel 2013. MATLAB R2016b was applied for method implementation, data processing, and analysis.

2.2. Baseline Correction

In this study, a new baseline correction method, computational recursion, was developed and applied to automated baseline correction for MIR (4000 to 400 cm^{-1}) spectra. The algorithm developed will run through all the spectral details and locate the lowest valley; then the baseline can be simply drawn by connecting the valleys to each side of the spectrum using straight lines. If any part of the spectrum is intersected by the straight lines after baseline correction, the spectrum will create negative absorbance values. In this case, the algorithm will be recursively run for the spectral regions where the intersecting straight lines connected. This simple and rapid baseline correction method can be applied to any IR spectral region, without any intervention. Furthermore, since all the PHCs have similar spectral features, the adopted baseline correction method can give unified baseline corrected IR data for all of the analytes of interest.

2.3. Band Decomposition

Automatic band decomposition method was applied in this study. The sub-bands were established using a second derivation curve (SDC). The band number can be controlled by eliminating the number of SDC valleys [13]. The dominant bands from the original spectrum were presented as the lower valleys in the SDC. The small valleys, representing secondary bands, could be eliminated using a Gaussian low pass filter, with a standard deviation 1.5. The spectrum band can be properly decomposed using Gaussian curves, which the amplitude, width, and location were optimized using Monte Carlo algorithm (MCA), which is a heuristic algorithm based on randomness and statistics to get an optimisation result.

3. Results and Discussion

3.1. Alkane Determination

The measurements for alkane identification was conducted using four different alkanes, C_{20}, C_{26}, C_{32} and C_{37} with the concentration of 5000 ppm for each alkane, respectively. To prepare these standards, the pure alkane chemicals were diluted with potassium bromide (KBr). The four alkanes' spectra after background subtraction and baseline correction are shown in Fig. 1.a. According to the similarity of the IR bands in the entire region indicates the alkanes have identical molecular structures. The doublet observed between 3000 and 2800 cm^{-1}, which is the evidence of carbon hydrogen bonding for long-chain alkanes due to —C—H stretching vibrations. From the Fig. 1.b, it is found there were two bands coherent to the doublet at location 2954 and 2872 cm^{-1}, respectively. From C_{20} to C_{37}, by the increase of the C_n, the intensities of the two bands were reduced as demonstrated. This phenomenon matched the similar results from Snyder and co-workers, when they are comparing the Raman spectrum data of C_8 and C_{20} [14,15]. As demonstrated in one of their figures, both the spectra of C_8 and C_{20} contained all the similar bands, including the bands at location 2954 and 2872 cm^{-1}. It is observed the band intensities of the C_{20} were less than C_8 at these two locations [14]. On the opposite, there were another two doublets existed in the region from 1480 to 1450 cm^{-1} and the

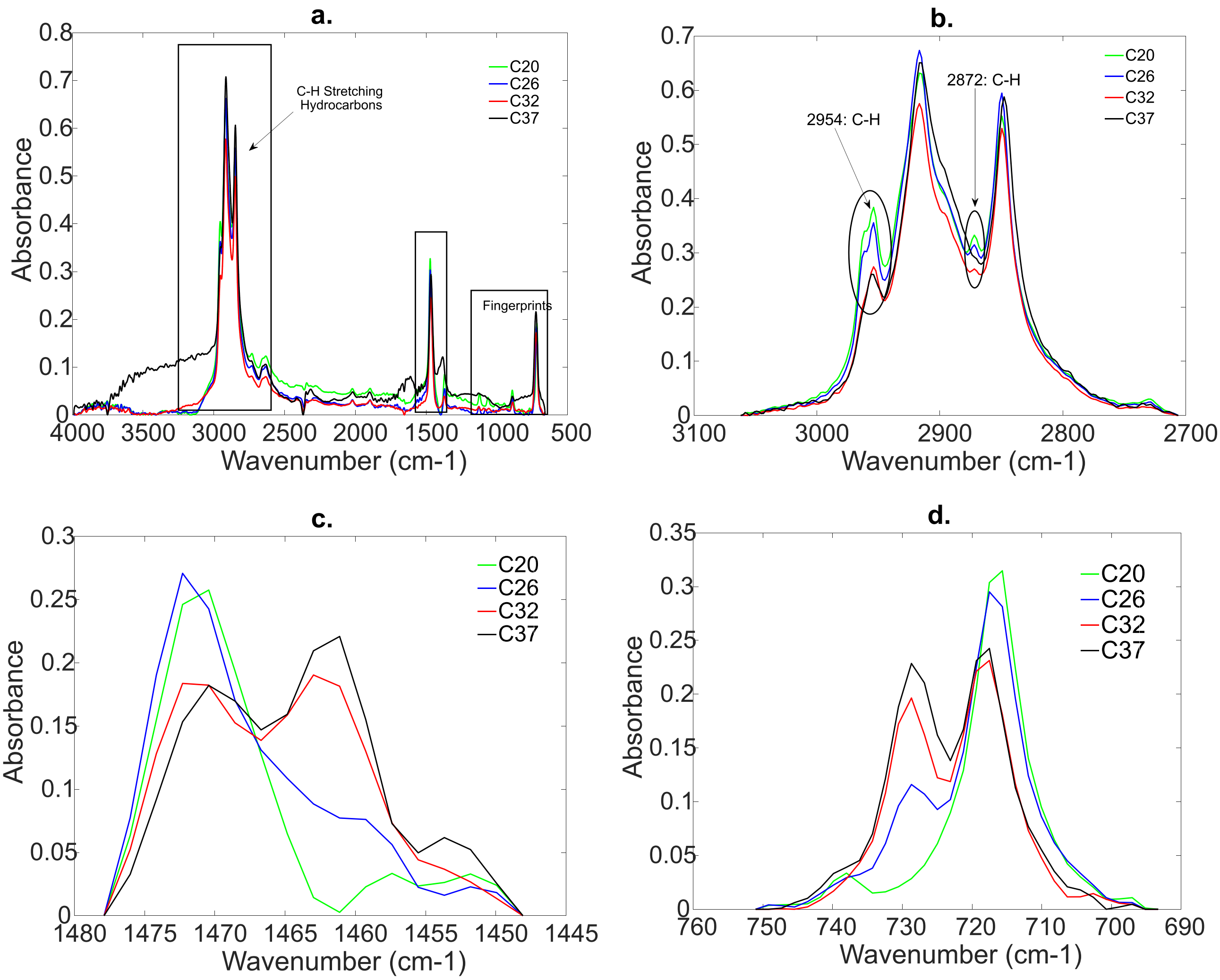

Fig. 1. The IR spectra of the five alkanes after baseline correction.

region at 750 to 730 cm^{-1}. It was observed the intensity of one coherent band at each of these regions were increased following the increase on the C_n. As shown in Fig. 1.c and d, the band center at 1462 and 730 was increased from C_{20} to C_{37}. As mentioned in [16], the band near 730 cm^{-1} can be seen only in long chain alkanes. The phenomena of the coherent bands in the three regions can be applied for C_n identification.

For the quantification studies, three different concentration levels of each selected alkane were mixed with KBr, to generate the calibration standards of 500, 5000 ppm and 10,000 ppm for C_{20}, C_{26}, C_{32} and 500, 1000, 2500 and 5000 ppm for alkane C_{37}. All the calibration spectra of each selected alkane, including the baseline corrected and band decomposed, are demonstrated in Fig. 2. It is noted that the increasing concentration did significantly effect on the absorbance magnitudes. Accordingly, based on the measurements it was difficult to identify the characteristic bands when the alkanes present were less than 500 ppm for each alkane. In an ideal scenario when only a single alkane presented, the calibration for each alkane can be applied using the bands at this IR region after baseline correction. However, this ideal scenario does not fit into the petroleum products. For example, crude oil can consist of hundreds of individual petroleum hydrocarbon compounds. For the mixtures, the calibration should vary according to the characteristic bands of the subbands, after the decomposition, as shown in Fig. 2 for each alkane. The calibration results of the four alkanes are illustrated in Fig. 3. Comparatively, all alkanes provided reasonable predictions when linking the absorbance area values to the spiked concentrations. Fig. 3 provides evidence that the concentrations were able to be predicted using the calibration curves generated with the single alkane standards.

In order to study the mixture scenario, orthogonal experimental design (OED) was applied to generate the alkane mixtures with designed heterogeneous concentrations [15]. In this study, an orthogonal design table (ODT) was generated and contained nine synthetic mixtures of the selected alkanes, with three concentration levels mentioned before. The mixtures were also prepared using an artificial soil containing 80% of quartz, 5% of humic acid and 15% of kaolinite. The details are listed in Table 1 and the spectral data for these nine mixtures are shown in Fig. 4. From Fig. 4.b, it is observed the minor band at 2873 cm^{-1} which indicates the amount of C_{20}. For example, the band exists in the spectrum of OED 8 rather than OED 9, demonstrated with black continuous line and dash line, respectively. After band decomposition, the length carbon chains, ECn can be indicated with the ratio of band 2850 cm^{-1} *versus* band 2956 cm^{-1}. Additionally, ECn can also be indicated by the ratio of the other two doublets at the regions 1489 to 1450 cm^{-1} (Fig. 4.c) and Region 750 to 700 cm^{-1} (Fig. 4.d). For example, the mixture OED 2 and 8, the proportion of shorter carbon chain (C_{20} and C_{26}) were higher than the longer carbon chain (C_{32} and C_{37}). It is observed the bands of 1470, and 715 cm^{-1} were suppressed of 1460 and 730 cm^{-1}, respectively, as shown in Fig. 4.c and d. These suppress phenomena are not shown in other spectral data from more homogeneous mixtures. The total concentration of the alkane mixture can be determined using absorbance area of a region below 3000 to 2800 cm^{-1}, as the EPA method 8440. In a mixture scenario, it is difficult

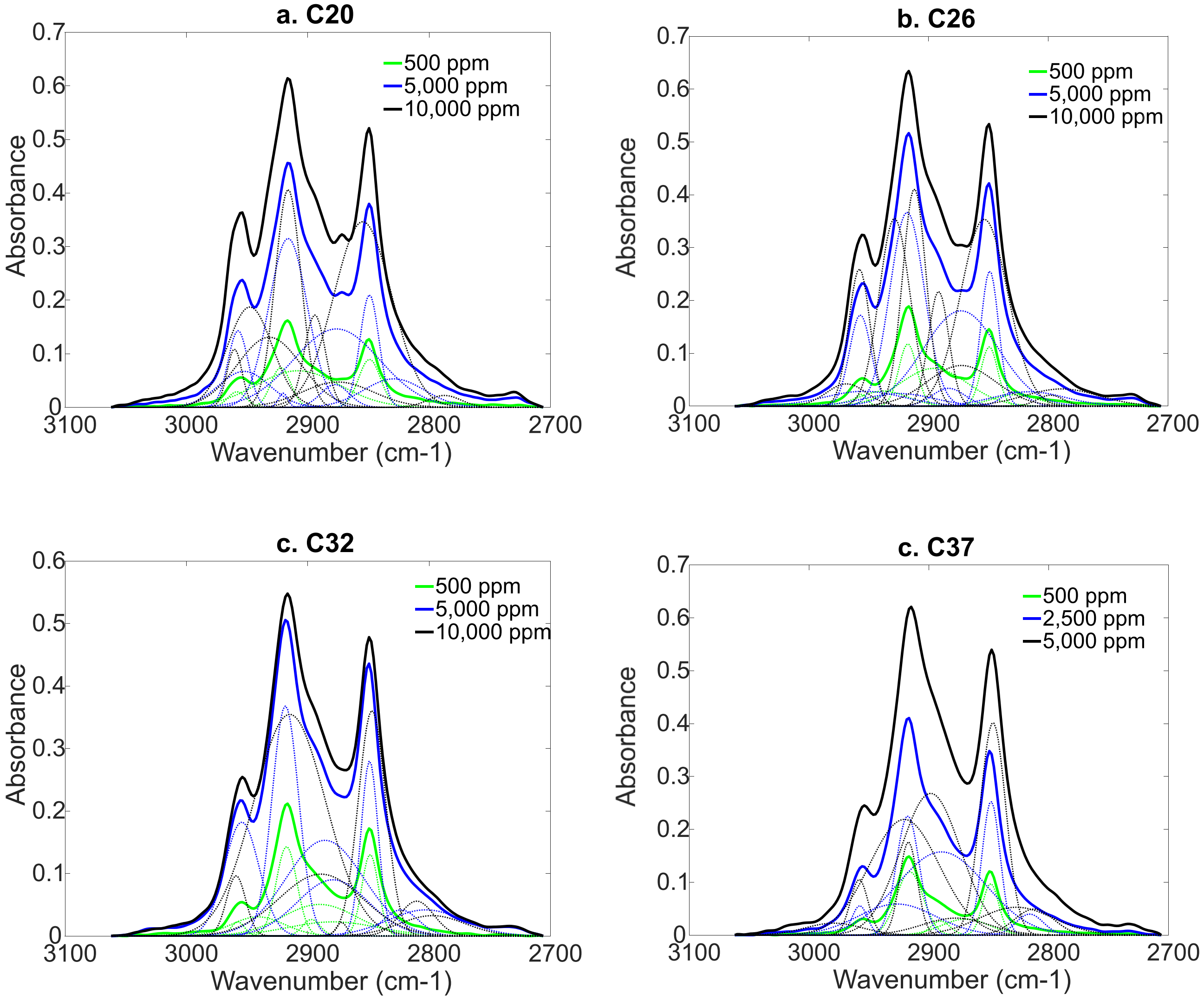

Fig. 2. Calibration spectral data for the selected alkanes (Spectra after baseline correction and Sub-bands after decomposition).

to identify the specified C_ns from C_{20} to C_{37}. Fortunately, the proportion of the shorter carbon chain (C_{20} and C_{26}) and the longer carbon chain (C_{32} and C_{37}) can be identified using the ratios as mentioned above.

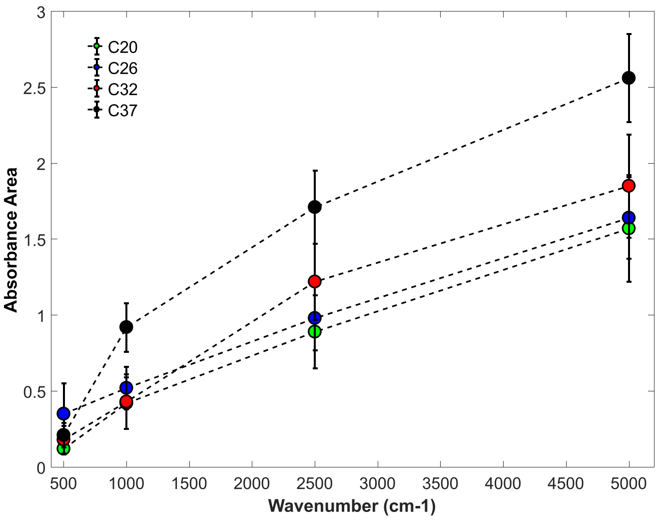

Fig. 3. The absorbance area value related to the concentrations for each alkane.

3.2. Petroleum Product Separation

To further validate the prediction concept, 100 mg of four different petroleum products, were spiked separately into 1 g of the artificial soil. The product details are listed in Table 2. The products include petrol, kerosene, diesel, motor/lubricating oils, and grease wax. The density of each product was measured by weighing samples in 50 mL volume containers. From Table 2, as expected, the densities increased with the carbon chain numbers. All samples were prepared fresh daily, and measurements were carried out at room temperature (22 °C) in triplicate, and the average values are presented. As mentioned the artificial soil

Table 1
Orthogonal experimental design (unit: ppm or mg/kg).

OED	C20	C26	C32	C37	Total
1	10,000	5000	10,000	500	25,500
2	10,000	10,000	500	2500	23,000
3	5000	500	10,000	2500	18,000
4	5000	10,000	5000	500	20,500
5	5000	5000	500	5000	15,500
6	500	10,000	10,000	5000	25,500
7	500	500	500	500	2000
8	10,000	500	5000	5000	20,500
9	500	5000	5000	2500	13,000

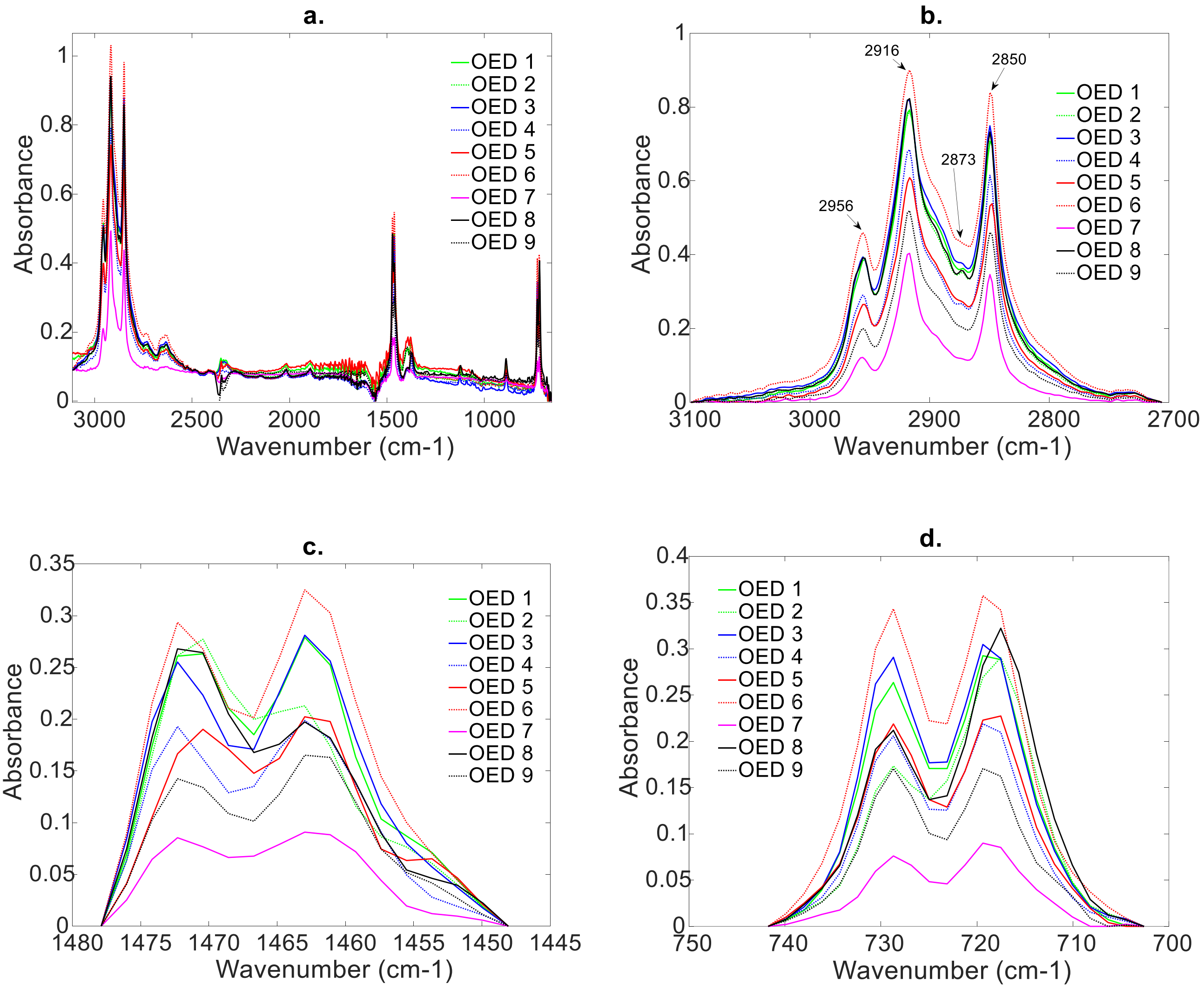

Fig. 4. Spectral data for these nine mixtures. a. Entire spectra; b. Region 3100 to 2700 cm^{-1}; c. Region 1489 to 1450 cm^{-1}; d. Region 750 to 700 cm^{-1}.

containing 80% of quartz, 5% of humic acid and 15% of kaolinite. It is observed the characteristic band at 804.5 cm^{-1}, indicating the Si—O bending vibrations for quartz identification. Kaolinite can be identified with the strong O—H stretching vibration presenting as the doublets at 3690–3620 cm^{-1}. Humic acid can be identified as the small bands like noise covered around 3500 cm^{-1} and water content band central at around 3400 cm^{-1}. According to the spectrum in Fig. 5, the signal in below 2000 cm^{-1} was dominated by quartz, and two of the hydrocarbon identical band regions at 740 to 710 cm^{-1} and 1480 to 1450 cm^{-1} were submerged in the dominated bands. Therefore, to avoid the interference from soil chemical components, the bands at 3000 to 2800 cm^{-1} can be applied to the alkane studies. Fig. 6 demonstrated the bands of the four products in the region. As mentioned the band at 2950 cm^{-1} increases adversely with the C_n from C_{20} to C_{37}. By comparison, it is clear to observe that a higher intensive band at 2950 cm^{-1} existed in IR spectrum of diesel and kerosene, rather than in the lubricating oil and grease wax.

Table 2
Selected petroleum products.

Product type	Product name	Carbon chains	Density (g/mL)
Kerosene	Diggers	C_{10} to C_{18}	0.76
Diesel	Caltex	C_{19} to EC_{30}	0.81
Lubricating oil	Gear oil, 85W-140, Penrite	$>EC_{30}$	0.90
Grease Wax	Grease Valplex, Valvoline	$>EC_{30}$	0.97

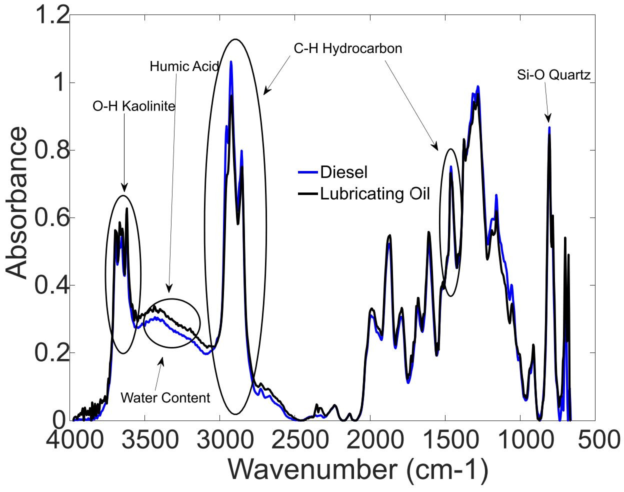

Fig. 5. The baseline corrected spectral data for spiked samples in the artificial soil.

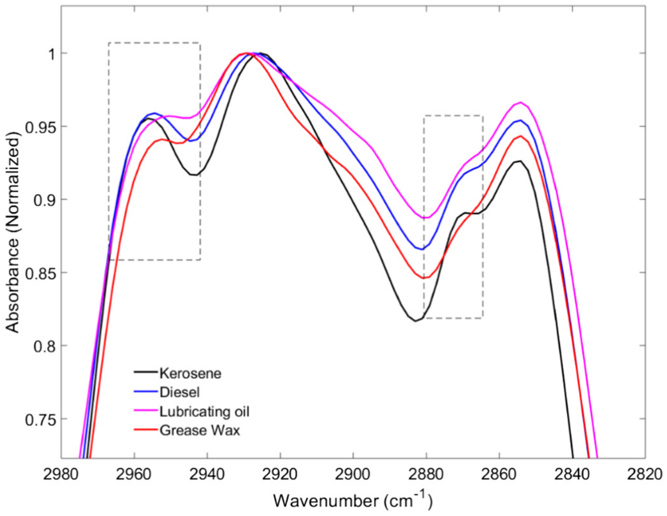

Fig. 6. IR spectral data for the four petroleum products at region 3000 to 2800 cm^{-1}.

4. Conclusion

The study has shown that it is possible to identify the alkanes with different C_n using their most presented characteristic bands. The phenomena of the three regions, 2954 to 2872 cm^{-1}, 1480 to 1450 cm^{-1} and 750 to 730 cm^{-1} can be applied for C_n identification. However, unlike applying pure alkane chemicals to a KBr background. When a petroleum product spat to soil, the soil IR spectrum will submerge the characteristic bands below 2000 cm^{-1}, due to the quartz and other soil components. Hence, realistically, the first IR region mentioned above, should be applied majorly to the alkane studies. By using this method, it is possible to estimate the dominant fraction of C_n in soil. However, the method is not able to determinate each alkane in heterogeneous mixtures, especially in a complex background, such as a neutral soil.

Acknowledgments

The authors would like to thank the Cooperative Research Centre for Contamination Assessment and Remediation of the Environment (CRC CARE), BHP Billiton Ltd. and the University of Newcastle for making this research possible. This research was funded by CRC CARE Pty Ltd. The research was conducted in the laboratories of the Global Centre for Environmental Remediation (GCER) at the University of Newcastle.

References

[1] W. Weisman, Analysis of petroleum hydrocarbons in environmental media, Total Petroleum Hydrocarbon Criteria Working Group Series 1 (1998) 98.
[2] R.J. Irwin, M. V, Mark Van Mouwerik, L. Stevens, M.D. Seese, W. Basham, Environmental Contaminants Encyclopaedia, National Government Publication, 1998.
[3] Massachusetts Department of Environmental Protection, Executive Office of Environmental Affairs Boston, MA, U. S., Interim Final Petroleum Report: Development of Health-based Alternative to the Total Petroleum Hydrocarbon (TPH) Parameter, 1994.
[4] TNRCC: Texas, U.S., Total Petroleum Hydrocarbons, TNRCC Method 1005, Revision 03, 2001.
[5] K.J. Brassington, R.L. Hough, G.I. Paton, K.T. Semple, G.C. Risdon, J. Crossley, I. Hay, K. Askari, S.J.T. Pollard, Weathered hydrocarbon wastes: a risk management primer, Crit. Rev. Environ. Sci. Technol. 37 (2007) 199–232.
[6] P. Fine, E.R. Graber, B. Yaron, Soil interactions with petroleum hydrocarbons: abiotic processes, Soil Technol. 10 (1997) 133–153.
[7] C. Gudin, Interaction Between Oil-Vegetation and Soil. International Symposium on Ground Water Pollution by Oil Hydrocarbons, International Press Praha, 1978 411–417 Prague.
[8] G. Zadora, R. Borusiewicz, J. Zieba-Palus, Differentiation between weathered kerosene and diesel fuel using automatic thermal desorption-GC–MS analysis and the likelihood ratio approach, J. Sep. Sci. 28 (2005) 1467–1475.
[9] B.B. Kebbekus, S. Mitra, Environmental Chemical Analysis, Thomson Science, U. K., 1998
[10] P.R. Griffiths, J.A. De Haseth, Fourier Transform Infrared Spectrometry, John Wiley & Sons, U. S., 2007
[11] M. Bradley, Curve Fitting in Raman and IR Spectroscopy: Basic Theory of Line Shapes and Applications, Thermo Fisher Scientific, U. S., 2007
[12] M.A. Czarnecki, Y. Ozaki, Determination of integrated intensities of overlapped IR bands by curve-fitting, Fourier self-deconvolution and a combination of both methods, Spectrochim. Acta A 52 (1996) 1593–1601.
[13] L. Wang, D. Yang, C. Fang, Z. Chen, P.J. Lesniewski, M. Mallavarapu, R. Naidu, Novel methodologies for anutomatically and simultaneously determining BTEX components using FTIR spectra, Talanta 144 (2015) 1104–1110.
[14] R.G. Snyder, S.L. Hsu, S. Krimm, Vibrational spectra in the C—H stretching region and the structure of the polymethylene chain, Spectrochim. Acta 34A (1977) 395–406.
[15] R.G. Snyder, H.L. Strauss, C.A. Elliger, C—H stretching modes and the structure of n-alkyl chains. 1. Long, disordered chains, J. Phys. Chem. 86 (26) (1982) 5145–5150.
[16] B.C. Smith, Infrared Spectral Interpretation: A Systematic Approach, CRC Press, 1998.