Detection of hydrocarbons in clay soils: A laboratory experiment using spectroscopy in the mid- and thermal infrared

M. van der Meijde *, N.M. Knox 1, S.L. Cundill, M.F. Noomen, H.M.A van der Werff, C. Hecker

University of Twente, Faculty for Geo-information Science and Earth Observation (ITC), P.O. Box 6, 7500 AA Enschede, The Netherlands

ARTICLE INFO
Article history:
Received 12 July 2012
Accepted 9 November 2012

Keywords:
Hydrocarbon
Soils
Hyperspectral
Pollution
Thermal infrared

ABSTRACT
Remote sensing has been used for direct and indirect detection of hydrocarbons. Most studies so far focused on indirect detection in vegetated areas. We investigated in this research the possibility of detecting hydrocarbons in bare soil through spectral analysis of laboratory samples in the short wave and thermal infrared regions. Soil/oil mixtures were spectrally measured in the laboratory. Analysis of spectra showed development of hydrocarbon absorption features as soils became progressively more contaminated. The future application of these results airborne seems to be a challenge as present and future sensors only cover the diagnostic regions to a limited extent.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction
Remote sensing in the near- and shortwave infrared can be used for direct and indirect detection of hydrocarbons. Previous research largely focused on direct spectral detection of hydrocarbons (e.g. Lammoglia and Filho, 2011; Bihong et al., 2007; van der Werff et al., 2006; Winkelmann, 2005; Kühn et al., 2004; Hörgå et al., 2001; Malley et al., 1999). The presence of hydrocarbons in a soil can also lead to chemical and mineralogical alterations. Bacterial oxidation of hydrocarbons can establish anomalous redox zones that favour the development of a diverse array of chemical changes (Schumacher and Abrams, 1996). An overview of hydrocarbon-induced alterations, and consequences for remote sensing is given by van der Meer et al. (2002), Schumacher and Abrams (1996), and Cloutis (1989).

The aforementioned studies focused on natural hydrocarbon seepage and/or controlled experiments. Recent work, related to pipeline leakage, in the laboratory (Noomen et al., 2006; Smith et al., 2004), field (van der Meijde et al., 2009) and airborne (van der Werff et al., 2008), showed that indirect detection of hydrocarbons through analysis of vegetation reflectance is possible. These studies, however, focused only on areas with relatively dense vegetation cover, leaving areas with little or no vegetation unaccounted for.

In this research, we investigate the possibility of detecting hydrocarbons in bare soil through spectral analysis of the short wave and thermal infrared wavelength regions. Based on the discriminative wavelengths found and available thermal infrared sensors, a feasibility study shows that the direct applicability of such techniques for airborne based detection and monitoring is currently limited.

2. Methods
2.1. Laboratory experiment
The laboratory experiment consisted of two stages. At first, 10.00 g soil samples were created from a mixture of air-dried clay soil and a hydrocarbon (Shell mineral engine oil). The hydrocarbons were added to the soil in increments of 10% weight, from 10% till 100% hydrocarbon content. In the second stage, the air-dried clay soil was moistened (8.50 g soil to 1.50 g water) and well mixed to make it comparable to field conditions. The sample was prepared with hydrocarbon concentrations that ranged between 0 and 10% of the moistened soil weight, in increments of 1%. Directly following sample preparation spectral samples were measured thereby minimising the possibility of evaporation.

2.2. Instrument and spectral measurement set-up
Thermal and shortwave infra-red spectra were recorded with a Bruker Vertex 70 Fourier Transform Infrared (FTIR) spectrometer (Hecker et al., 2011). Diffuse Reflectance Infrared Fourier Transform (DRIFT) spectral measurements were recorded in the wavenumber

* Corresponding author. Tel.: +31 53 4874322.
E-mail address: vandermeijde@itc.nl (M. van der Meijde).
1 Presently at: Research and Development Group, South African National Space Agency (SANSA) – Earth Observation Division, P.O. Box 484, Silverton, Pretoria 0127, South Africa.
0303-2434/5 – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jag.2012.11.001
range of 5000–500 cm\(^{-1}\). For laboratory samples a spectral resolution of 2 cm\(^{-1}\) and 64 scans per sample was used. For field samples the number of scans was increased to 500 to enhance the signal-to-noise ratio. The sample was stirred in its canister and a random sub-sample was drawn to fill the sample holder. The sub-sample was levelled to ensure a flat measurement surface and placed in the DRIFT compartment (following a background reference measure) to take a sample measurement. The sub-sample was then rotated 180\(^\circ\) and measured again, thereby capturing brightness differences resulting from scattering due to soil structure. Thus, for each sub-sample two spectra were measured. A total of five sub-samples (10 spectra in total) were measured per sample.

2.3. Statistical analysis

Spectral data shows strong co-linearity between wavelengths, therefore analysis should consider how to account for this. The large sample size of the laboratory data set allowed for the application of a partial least squares regression (PLSR, Wold et al., 2001) to initially reduce the dimensionality of the data by identifying significantly contributing wavelengths and to reduce the impact of co-linearity on the modelling.

The FTIR dataset was split randomly into 75\% for training and 25\% for model validation. Applying PLSR (see Hecker et al. (2012)) for an application of PLSR to thermal infrared spectra, latent variables were selected which reduced the prediction residual error sum of squares values observed in the training data. The summed absolute loadings of the selected latent variables were then used to identify which wavelengths contributed most to the derived model. Once the number of latent variables had been selected, the (absolute) loading values of these variables were summed, to determine which wavenumbers were contributing most. A major limitation is that the results of PLSR analysis are dependent on the hydrocarbon type it was trained for and the uncertainty estimates are only valid for a dataset that is compositionally comparable (in average as well as extreme values) to the training dataset.

To account for this limitation of the PLSR and in an attempt to create a model that can generalise to new data, we selected the 10 highest loading peaks identified in the PLSR modelling and applied a stepwise multiple linear regression (SMLR, Crawley, 2006; Grossman et al., 1996) to these wavelengths. Since the data has been strongly reduced and co-linearity already addressed in the PLSR modelling this can now deliver a model that is more independent from the hydrocarbon type used. Prior to performing a forward/backward SMLR procedure it was verified that the data followed a non-normal distribution and these were therefore arc-sine transformed (Crawley, 2006). The SMLR models were then validated using the independent validation set.

3. Results

3.1. Trends in spectra

Analysis of all dry soil to oil mixtures (0–100% hydrocarbon contamination) as well as the moist soil to oil mixtures (0–10% hydrocarbon contamination) shows that there was development of hydrocarbon features and reduction in the soil features as soils became progressively more contaminated (see Fig. 1 for an example of the moist mixture). The most dominant and specific hydrocarbon features (Stuart, 2004) between 2850 and 2960 cm\(^{-1}\) were already visible from a hydrocarbon concentration of approximately 1\%. Furthermore, an absorption feature is present in the C–H fingerprint region of 1000–1275 cm\(^{-1}\) which is becoming increasingly visible with increasing hydrocarbon content. Other spectral features are related to water content and/or soil composition and are getting smaller with increasing hydrocarbon content.

![Fig. 1. Average spectrum for HC (oil) contaminated soil ranging between 0 and 10% contamination levels. For visualisation purposes only the even percentages pollution levels are shown and offset in steps of 0.02 with respect to the lowest concentration spectrum. The wavenumbers highlighted (vertical lines) are those selected in the PLSR analysis for inclusion in the SMLR modelling. Typical regions for C–H spectral features are indicated in grey (after Stuart (2004)).](image-url)
wavenumbers selected through SMLR to form a model for predicting hydrocarbon contamination levels. The model had a goodness of fit of 0.95 (Adjusted $R^2$), and RMSE of 0.019 and displays a close fit with the validation data (Fig. 2).

4. Discussion

4.1. Application for different hydrocarbons and soils

In the laboratory an oil hydrocarbon was used. In real field situations often other hydrocarbons will be present. It was not possible, unfortunately, to measure high concentrations of highly toxic substances, such as e.g. benzene condensates, without a fume hood. We believe, however, that the impact of different hydrocarbons is not necessarily large. An overview of hydrocarbon absorption features (Stuart, 2004) shows that hydrocarbon absorption features in the so-called C–H fingerprint region of 1000–1275 cm$^{-1}$ are associated with methyl and methylene C–H stretching and bending (Stuart, 2004), and are generic for many types of hydrocarbons. Variations in absorption features between the different hydrocarbons used are most likely to occur within this wavelength region.

Thermal signatures can be strongly influenced by the type of soil that has been used. Clay does not have very strong features in the TIR, contrary to sandy soils. The quartz content of sands is strongly dominating features in the range of 1050–1250 cm$^{-1}$ and 714–800 cm$^{-1}$ (Hecker et al., 2012). This is close to the diagnostic 1045 cm$^{-1}$ feature found in the modelling. The produced model is therefore valid for clay soils with little to no quartz content.

4.2. Airborne applicability

The lack of applications of airborne TIR remote sensing in hydrocarbon detection can be attributed to two main causes: (a) a relatively restrictive terrestrial atmosphere that limits thermal remote sensing to roughly 2000–3333 cm$^{-1}$ and 714–1250 cm$^{-1}$ and (b) the scarcity of high spectral resolution sensors in the thermal infrared (see Table 1).

In terms of sensor availability, large improvements are to be expected within the next couple of years. For many years the SEBASS instrument (Hackwell et al., 1996) was the only high spectral resolution thermal instrument available, but data availability for civil applications was very limited. Currently, SEBASS data are flown and available commercially (pers. communication D. Riley, The Aerospace Corporation). Several other airborne thermal imaging systems have recently been deployed or are close to completion (compare Table 1). All of them measure emission in the window starting at 800 cm$^{-1}$ and ranging up to 1235–1333 cm$^{-1}$ with a maximum of 1429 cm$^{-1}$. Luckily, the C–H fingerprint region falls largely within the terrestrial atmospheric windows but many of the other diagnostic wavenumbers found in this study cannot be used since they are not covered by the spectral range of the sensors. Practically, this results in the unfortunate situation that for all sensors, except MAGI, only the 1045 cm$^{-1}$ diagnostic feature can be used. The MAGI sensor extends far enough to cover the 1416 cm$^{-1}$ diagnostic feature to be included in future analysis. The other 5 diagnostic wavenumbers on which the prediction model is based are not covered by the present range of active and

<table>
<thead>
<tr>
<th>Sensor name</th>
<th>Coordinating organisation</th>
<th>Spectral range</th>
<th>No. of bands</th>
<th>Spectral resolution</th>
<th>NEΔT</th>
<th>Availability</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEBASS</td>
<td>Aerospace Corporation, USA</td>
<td>7.5–13.5 μm 741–1333 cm$^{-1}$</td>
<td>128</td>
<td>10 nm</td>
<td>0.05 K</td>
<td>1995–</td>
<td>Kirkland et al. (2002) and Hackwell et al. (1996)</td>
</tr>
<tr>
<td>AHI</td>
<td>University of Hawaii, USA</td>
<td>7.5–11.5 μm 870–1333 cm$^{-1}$</td>
<td>256</td>
<td>125 nm</td>
<td>0.19 K</td>
<td>1998–</td>
<td>Vagni (2007) and Lucey et al. (2001)</td>
</tr>
<tr>
<td>TASI-600</td>
<td>ITRES, Canada</td>
<td>8.0–11.5 μm 870–1250 cm$^{-1}$</td>
<td>64</td>
<td>125 nm</td>
<td>?</td>
<td>2006–</td>
<td>IT Research Limited (2012b) and Shimoni (2009)</td>
</tr>
<tr>
<td>AISA Owl</td>
<td>SPECIM, Finland</td>
<td>8.0–12.0 μm 833–1250 cm$^{-1}$</td>
<td>84</td>
<td>100 nm</td>
<td>0.12 K</td>
<td>Early 2010</td>
<td>Spectral Imaging Ltd. (2012), Holma et al. (2009), and Shimoni (2009)</td>
</tr>
<tr>
<td>MAKO</td>
<td>Aerospace Corporation, USA</td>
<td>7.8–13.4 μm 746–1282 cm$^{-1}$</td>
<td>128</td>
<td>40 nm</td>
<td>0.1 K</td>
<td>2011–</td>
<td>Hall et al. (2011a) and Tratt and Hall (2011)</td>
</tr>
<tr>
<td>MAGI</td>
<td>Aerospace Corporation, USA</td>
<td>7.0–12.7 μm 787–1429 cm$^{-1}$</td>
<td>32</td>
<td>≈190 nm</td>
<td>0.1 K</td>
<td>2012–</td>
<td>Hall et al. (2011b) and Hall et al. (2008)</td>
</tr>
<tr>
<td>HyTES</td>
<td>JPL, USA</td>
<td>7.5–12.0 μm 833–1333 cm$^{-1}$</td>
<td>256</td>
<td>35 nm</td>
<td>0.1 K</td>
<td>Late 2012</td>
<td>Hook et al. (2009)</td>
</tr>
<tr>
<td>SIELETERS</td>
<td>ONERA, France</td>
<td>8.0–11.5 μm 870–1250 cm$^{-1}$</td>
<td>n/a</td>
<td>&lt;230 nm</td>
<td>&lt;0.2 K</td>
<td>?</td>
<td>Shimoni (2009) and Thetas et al. (2009)</td>
</tr>
</tbody>
</table>

Modified after Hecker et al. (2010).
upcoming TIR sensors. This is in line with work by Lammoglia and Filho (2011) who also noted that hydrocarbon related TIR features have little to no response in the range covered by the SEBASS sensor. An overview of multi- and hyperspectral sensors and their applications by van der Meer et al. (2012) further confirms that the airborne sensors have limited contribution to the direct detection of hydrocarbons.

The noise level in thermal imagery data varies a lot between the different systems. For geological purposes, Richter et al. (2005) determined that noise equivalent difference in temperature (NEΔT) of around 0.1 K still gave good retrieval of mineral spectra for the proposed ARES sensor. Table 1 shows that SEBASS and the proposed HyTES combine high spectral resolution (number of bands) with NEΔT levels at or below 0.1 K. The MAGI sensor is at 0.1 K and is therefore the most promising sensor with the highest potential on the market for hydrocarbon detection.

There are a few sensors specifically developed for the mid-infrared (MIR) region. SEBASS has a MIR extension which has 128 bands in the range of 1887–4000 cm⁻¹ offering an alternative for TIR spectroscopy and covering three of the diagnostic features from the model. TASI has also a MIR equivalent, called MASI (IT Research Limited, 2012a). MASI is covering a slightly smaller range, 2000–3333 cm⁻¹, in 64 bands. Two much lower spectral resolution alternatives are two NASA sensors, MAS (Modis airborne simulator, citetMAS) with 16 bands in the MIR, and MASTER (Modis/Aster airborne simulator, NASA, 2012) with 15 bands.

5. Conclusion

We have investigated the possibility of detecting hydrocarbons in bare soil through spectral analysis of the short wave and thermal infrared regions. For this purpose a laboratory experiment was set up. Hydrocarbon polluted soil samples were analysed for diagnostic wavenumbers that can predict the amount of hydrocarbons in the soil. A two-step approach modelling using PLSR and SMLR resulted in a model that had a goodness of fit around 0.95 (Adjusted R²) and is based on spectral features that all relate to hydrocarbon absorption features indicating that at least these features are stable for the prediction of hydrocarbon oil contamination in soil. An assessment of the applicability of airborne TIR sensors for detecting hydrocarbons shows that the present and future sensors do cover the C–H fingerprint region but do not cover the spectral range of 1500 cm⁻¹ and beyond in which most of the diagnostic features were found in this study.

Acknowledgements

The authors wish to thank two reviewers for their constructive comments. We thank Boudewijn de Smeth, head of the geochemical laboratory at UT-ITC, for his support during the measurements. This research was financially supported by the Nederlandse Aardolie Maatschappij BV.

References


