Journal Pre-proof

Diagnostic features emerging in near-infrared reflectance spectroscopy for low petroleum hydrocarbon pollution after spectral subtraction

Qijia Lou, Mei Lei, Yu Wang, Shaobin Wang, Guanghui Guo, Wencheng Xiong, Ying Jiang, Tienan Ju, Xiaofeng Zhao, Frederic Coulon

PII: S0048-9697(24)02407-0

DOI: https://doi.org/10.1016/j.scitotenv.2024.172264

Reference: STOTEN 172264

To appear in: *Science of the Total Environment*

Received date: 5 December 2023

Revised date: 4 April 2024

Accepted date: 4 April 2024

Please cite this article as: Q. Lou, M. Lei, Y. Wang, et al., Diagnostic features emerging in near-infrared reflectance spectroscopy for low petroleum hydrocarbon pollution after spectral subtraction, *Science of the Total Environment* (2023), https://doi.org/10.1016/j.scitotenv.2024.172264

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2024 Published by Elsevier B.V.

Journal Pre-proof

Science of The Total Environment, Volume 928, June 2024, Article number 172264
DOI:10.1016/j.scitotenv.2024.172264

Diagnostic Features Emerging in Near-infrared Reflectance Spectroscopy for Low Petroleum Hydrocarbon Pollution after Spectral Subtraction

Qijia Lou [a,b*], Mei Lei [a,b*], Yu Wang [c], Shaobin Wang [a], Guanghui Guo [a], Wencheng Xiong [d], Ying Jiang [e], Tienan Ju [a], Xiaofeng Zhao [f], Frederic Coulon [e]

a. Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

b. University of Chinese Academy of Sciences, Beijing 100049, China

c. Technical Center for Soil, Agriculture and Rural Ecology and Environment, Ministry of Ecology and Environment, Beijing, 100012, China

d. Satellite Application Center for Ecology and Environment, Ministry of Ecology and Environment, Beijing, 100094, China

e. School of Water, Energy and Environment, Cranfield University, Cranfield MK43 0AL, UK

f. Command Center for Natural Resources Comprehensive Survey, China Geological Survey, Beijing 100055, China

Journal Pre-proof

ABSTRACT

Diagnostic features in near-infrared reflectance spectroscopy (NIRS) are the foundation of knowledge-based approach of petroleum hydrocarbon determination. However, a significant challenge arises when analyzing samples with low levels of petroleum hydrocarbon pollution, as they often lack distinctive diagnostic features in their sample NIRS spectra, limiting the effectiveness of this approach. To address this issue, we have developed a

Published by Elsevier. This is the Author Accepted Manuscript issued ith: Creative Commons Attribution Non-Commercial No Derivatives License (CC:BY:NC:ND 4.0).
The final published version (version of record) is available online at DOI:10.1016/j.scitotenv.2024.172264. Please refer to any applicable publisher terms of use.

technical workflow for diagnostic spectrum construction and parameterization based on spectral subtraction. This method was applied on a set of NIRS spectra from soil samples that were contaminated with petroleum hydrocarbons (ranged between 178 to 1,716 mg/kg of total petroleum hydrocarbon). Then two diagnostic features for low-level petroleum hydrocarbon pollution were found: (1) An overall downward concave emerged on diagnostic spectrum within both 2,290–2,370 nm and 1,700–1,780 nm for all low pollution levels even below 200 mg/kg; (2) An indicative pattern of asymmetric "W-shaped" double absorption valley occurred for those exceeding 1,000 mg/kg, and its valleys located near 2,310 nm, 2,348 nm or 1,727 nm, 1,762 nm stably. These two features on diagnostic spectrum could be parameterized to detect, and the detection limit was at least about 10-50 times lower than that based on sample spectrum. These findings update our understanding on the detectability of spectral response from low petroleum hydrocarbon pollution, and widely extend the application of knowledge-based NIRS approach in either field detection or remote sensing identification for environmental management.

Keywords: absorption feature; fast Fourier transform filtering; knowledge-based approach; soil matrix

1. Introduction

Petroleum hydrocarbon pollution represents a global concern, as highlighted in previous studies(Albers, 2002; Ellis Jr & Adams Jr, 1961; Khan et al., 2018; Kuppusamy et al., 2020; Tang et al., 2011). Petroleum hydrocarbons are ubiquitous in environment, primarily originating from activities such as petroleum extraction, refinement, transportation, storage,

and utilization, as well as natural seepage from oil reservoirs. Their presence poses significant threats to ecosystems and human health. To address these issues, it is imperative to accurately identify the sources of contamination and map the affected areas, enabling timely remediation and control measures.

Near-infrared reflectance spectroscopy (NIRS) uses near-infrared light with wavelengths in the range of 780–2,500 nm. It is a non-destructive analytical technique used to analyze the composition and properties of various substances, especially organic materials. In both field detection and remote sensing applications, it has the potential to be an indispensable tool for large-scale petroleum hydrocarbon determination. Several studies have employed field NIRS spectroradiometers and airborne hyperspectral imagers to assess petroleum hydrocarbon-polluted soil and oil seepages, utilizing various technical approaches including direct or indirect and knowledge-based or data-driven (Asadzadeh et al., 2022; Asadzadeh & de Souza Filho, 2016, 2017; Brekke & Solberg, 2005; Dubucq et al., 2021; Ng et al., 2017; Pasquini, 2018; Seo et al., 2023). However, the majority of these efforts remain in the exploratory phase, with limited practical applications. Among them, knowledge-based direct detection is highly interpretable, robust and universally applicable in real-world scenario, therefore it is considered as one of the most promising approach for regional, national and even global identification of petroleum hydrocarbon sources and their contaminated areas.

Diagnostic features are the foundation of knowledge-based NIRS approach for direct detection of petroleum hydrocarbon. Petroleum hydrocarbon and its contaminated materials exhibit distinct absorption bands at approximately 1,730 and 2,300 nm, attributed to primary

combinations and overtones of the C-H stretching modes of saturated CH_2 and terminal CH_3 (Aske et al., 2001; Cloutis, 1989; Scafutto & de Souza Filho, 2016; Yin et al., 2017). These absorption features are commonly referred to as the diagnostic features of petroleum hydrocarbon. These features have been instrumental in developing various detection indices, including the hydrocarbon index, normalized difference hydrocarbon index, hydrocarbon detection index, and hydrocarbon index area (Achard et al., 2021; Allen & Satterwhite, 2006; Kühn et al., 2004; Short, 2011). In addition, they have been employed to assess hydrocarbon contamination levels, determine oil density, and identify the speciation of hydrocarbons present in soil (Achard et al., 2021; Pabón et al., 2019; Scafutto & de Souza Filho, 2016; Scafutto et al., 2016).

Previous studies have revealed a notable challenge in acquiring diagnostic features in NIRS when on low-levels of petroleum hydrocarbon polluted soil samples, therefore significantly limiting the practical applications of NIRS, particularly in environmental management. From an analytical perspective, it has been established that in order to clearly observe absorption features in hydrocarbon contaminated soil sample spectra, the concentration of hydrocarbon need to be at least above 10,000 mg/kg and sometimes even much higher depending on variables including oil properties, soil types, and detection conditions (Pabón & de Souza Filho, 2016; Pelta & Ben-Dor, 2019; Winkelmann, 2005). Considering in many countries the regulatory threshold values of soil hydrocarbon contamination typically range from hundreds to thousands of mg/kg, the conventional diagnostic NIRS due to its detection limit, is not sufficient to meet the analytical requirements for determination of petroleum hydrocarbon contamination (MEE, 2018; VROM, 2012).

In the context of remote sensing application of NIRS, the detection limit defined as the smallest areal extent of oil seep within a pixel detectable spectrally, was reported to vary from 2.5 to 25 % of a pixel (Asadzadeh & de Souza Filho, 2017; Ellis et al., 2001; Hörig et al., 2001; Kokaly et al., 2013; Kühn et al., 2004). Whereas, areal extent of the common onshore seepage occurred on earth surface is often small to negligible by comparing to the spatial resolution of spaceborne or airborne hyperspectral imagery. From this perspective, it is evident that knowledge-based NIRS approach may not be effective for routine investigations and monitoring of petroleum hydrocarbon pollution as required for environmental management.

To overcome the challenge faced by the knowledge-based NIRS approach, several critical questions must be addressed: Can low-polluted materials exhibit spectral responses in the form of non-absorption features, and can these responses be transformed into diagnostic features for detection? The absence of absorption features in the sample spectra of lightly contaminated materials can be attributed to the relatively weak absorption characteristics of petroleum hydrocarbons in NIRS, as well as the strong spectral interferences from the background or soil matrix (Pabón & de Souza Filho, 2016). Spectral subtraction, as an established and powerful signal processing technique, is efficient at removing the contributions of non-target components from a mixture spectrum, enhancing the absorption features of the target component and enabling them to be detectable (Smith, 2021). It suggests that spectral subtraction may play an important role to find answer for this question.

This study aimed to explore diagnostic features of low petroleum hydrocarbon pollution.

We proposed a technical workflow of diagnostic spectrum construction and parameterization based on spectral subtraction, and executed it on a set of experimental data about before and after slightly polluted soil samples. Then, we recorded the morphology of diagnostic spectra visually and quantified them by series of spectral parameters. This was the first study intending to extract and reveal diagnostic features of low petroleum hydrocarbon polluted soils for detection, which might play a profound impact to the development of knowledge-based NIRS determination methods for both field detection and remote sensing identification.

2. Materials and Methodologies

2.1. Data source

A dataset including physicochemical properties, petroleum hydrocarbon concentrations, and visible NIRS of soil samples (https://doi.org/10.17862/cranfield.rd.5794791.v1) was used in this study. This dataset was the research data of an influential paper on the visible NIRS sensitivity to weathering of oil contaminated soils, and was obtained through well-designed and strictly controlled experiments. It was desirable for our study, because the visible NIRS about soil samples at pristine statuses and their polluted statuses in different weathering stages could be paired to construct ideal subtraction spectra with little impact from the heterogeneity of soil matrix. For understanding easily and utilizing readily, a brief introduction and simple reorganization of the data were done as follows. More detailed information can be acquired from the reference article (Douglas et al., 2018).

Thirteen pristine soil samples were collected from the research farm of Cranfield University in Bedfordshire, United Kingdom. Each sample was oven-dried, extracted at 1 kg soil, spiked with 10 ml Alaskan crude oil, and dried at room temperature (21 °C). Visible NIRS scanning was performed for each soil sample 8 times: pristine status (TC) and their polluted statues in weathering stages of 2 days (T2d) and months 4 (T4m), 8 (T8m), 12 (T12m), 16 (T16m), 20 (T20m), 24 (T24m). All of these were obtained in laboratory using ASD LabSpec2500 ® (Analytical Spectral Devices, Inc., USA), which covers a spectral range of 350–2,500 nm with a spectral resolution of 1 nm. Total petroleum hydrocarbon (TPH) concentrations were measured at T2d, T4m, T12m, T16m, T20m, and T24m, using gas chromatography-mass spectrometry (GC-MS) analysis. The THP concentrations in these samples ranged from 178 to 1,716 mg/kg. The mean values of THP concentrations in these samples at each weathering stage were 1,366, 1,037, 952, 742, 376, 207 mg/kg respectively, exhibiting a linearly downward trend with weathering time (Fig.1).

Based on the similarities of spectra and petroleum hydrocarbons concentrations, seven weathering stages were divided into two phases. Phase I included T2d, T4m, and T8m; while Phase II included T12m, T16m, T20m, and T24m. The spectral features of the soil samples would be presented and discussed thoroughly in the following sections.

2.2. Data analysis

2.2.1. Technical workflow

A diagnostic spectrum construction and parameterization workflow was developed for low

petroleum hydrocarbon pollution detection, by integrating spectral subtraction with fast Fourier transform filtering, diagnostic spectrum extraction and parameterization together (Fig.2). This workflow was designed to reduce spectral interference from untargeted components, eliminate signal noise of the measuring instrument, enhance and stabilize diagnostic features, and then enable the detection of low petroleum hydrocarbon pollution with parameters.

2.2.2. Spectral Subtraction

Spectral subtraction is an established and powerful technique for removing the contributions of non-target components from a mixture spectrum (Smith, 2021). Although many researchers have successfully applied spectral subtraction method to analyze soil organic matter through infrared spectroscopy (Margenot et al., 2016; Margenot et al., 2019), to the best of the authors' knowledge, few studies have explored the feasibility of using this method for measuring petroleum hydrocarbon content in soil using NIRS. The formula for spectral subtraction is as follows:

$$S_D = S_O - (S_R \times F) \quad (1)$$

where S_D is the result of spectral subtraction, named the subtraction spectrum; S_O is the original spectrum of contaminated soil, named the sample spectrum; S_R is the original spectrum of pristine soil without a target component, named the reference spectrum; F is an adjustment coefficient depending on the content level of the target component, also called the subtraction factor.

In this study, TC spectrum was selected as the reference spectrum for each soil sample, whereas the spectra of T2d, T4m, T8m, T12m, T16m, T20m, and T24m were used as separate sample spectra. A uniform subtraction factor of 1 was applied to each reference spectrum due to the low concentration of petroleum hydrocarbon in these samples.

2.2.3. Fast Fourier Transform Filtrating

A smoothing algorithm for the fast Fourier transform filter was applied to eliminate high-frequency noise from mainly from the measuring instrument and help to fit absorption valleys in subsequent processing. Previous studies have suggested setting a smoothing window equal to half the Full Width at Half Maximum (FWHM) of the absorption valley (Brown, 2006; Huguenin & Jones, 1986). Exercising caution is important in this step. If the smoothing window is set smaller, a large amount of noise will not be cleared; however, if it is set larger, a narrow absorption valley will be removed from the spectrum. We performed an initial analysis of the FWHM attributed to petroleum hydrocarbons. The results revealed that a relatively narrow FWHM of 10 mm was present. Therefore, the smoothing window was set to 5 nm in this study.

2.2.4. Diagnostic Spectrum Extraction

Diagnostic spectrum refers to a specific spectral segment encompassing the regions where diagnostic features of petroleum hydrocarbons are typically located. Petroleum hydrocarbons exhibit five distinctive absorption characteristics within the NIR range (780–2,500 nm), primarily attributed to C-H vibrations: combinations around 2,310 nm, the first

overtone around 1,730 nm, first overtone combinations around 1,400 nm, second overtone around 1,200 nm, and third overtone around 900 nm (Workman & Weyer, 2012). For the sample spectra of soil or other materials polluted by petroleum hydrocarbons, absorption valleys around 2,310 nm and 1,730 nm are generally more visible and easier to interpret, whereas those at other positions are often imperceptible or inconclusive (Achard et al., 2021; Pabón et al., 2019). However, precise boundaries of these characteristics absorption regions are not universally defined, because number of absorption valleys and their positions can vary depending on the diversity and complexity of the composition within petroleum hydrocarbons. Relatively, 2,290–2,370 nm and 1,700–1,780 nm are optimal for detection of hydrocarbon in soil (Pabón & de Souza Filho, 2016), for these two regions on almost all sample spectra of heavily petroleum hydrocarbon polluted soil have most visible concave features. Thus, we intercepted spectral segments within these two regions and took them after baseline correction as diagnostic spectra. In addition, other than absorption valleys, overall downward concave of the diagnostic spectra was also regarded as the diagnostic feature, which is more tolerant for the variable composition of petroleum hydrocarbon.

Journal Pre-proof

2.2.5. Parameterization

Both downward concave and absorption valley were quantified with common characteristic parameters, which could be developed into mathematics criterion for machine identification. According to the physical meaning of parameters and their operability within common software tools, integral area (S) of diagnostic spectrum was selected to depict overall downward concave feature, and wavelength position (P), number of absorption valleys

(N), FWHM (W), depth (D), and fitting area (A) were used to portray absorption valley. Either S or P is the essential indicator for petroleum hydrocarbon detection. All parameters were obtained with OriginPro, and P, N, W, D, A were calculated after Gaussian curve fitting. Table 1 lists the descriptions of each parameter.

3. Results

3.1. Morphological Characteristics

3.1.1. Global feature in full spectral range

In order to examine the spectral responses of soil samples with low petroleum hydrocarbon contamination, spectral curves spanning the full range (400–2,400 nm) were generated for soil samples at different weathering stages (Fig.3). In Fig.3 (a), the average reflectance of sample spectra is lower than that of the reference spectrum (TC), with Phase I showing lower reflectance than Phase II. However, it's worth noting that the average reflectance within either Phase I or Phase II does not exhibit a linear correlation with weathering time (pollution level). Additionally, the morphological features of the sample spectra in the full range generally resemble those of the reference spectrum. Yet, the spectral segments at 2,290–2,370 nm and 1,700–1,780 nm in Phase I are much flatter than those of in Phase II and TC, though all sample spectra show no observable absorption features at these two regions. These phenomena indicate that spectral responses of low petroleum hydrocarbons contaminated soil are existing in both spectral reflectance in full range and morphological feature at petroleum hydrocarbons related regions, but they are not exclusive

and are observable only through comparison.

Fig.3 (b) illustrates mean relative reflectance of subtraction spectra obtained by subtracting the sample spectrum from the reference spectrum. Similarly, the mean relative reflectance demonstrates no linear correlation with pollution levels within each phase, with Phase I exhibiting lower values than Phase II. Morphological features across the full range remain inconclusive. However, a noteworthy change becomes evident on the subtraction spectra in Phase I. Two distinct absorption valleys and a weak concave feature emerge at 2,290–2,300 nm and 1,700–1,780 nm. These absorption features are consistent with the diagnostic features on sample spectrum of high petroleum hydrocarbon polluted soil. This suggests that the spectral response of low petroleum hydrocarbon-polluted soil, particularly within Phase I, becomes detectable following the transformation of sample spectra into subtraction spectra.

3.1.2. Local shape of diagnostic spectra

To emphasize the distinct characteristics in the local shapes of spectra related to petroleum hydrocarbon, a fast Fourier transform filtering method was applied to smooth both sample and subtraction spectra. Subsequently, two segments at both 2,290–2,370 nm and 1,700–1,780 nm were extracted and corrected by subtracting the baseline. These spectral segments are presented in Fig. 4. Fig. 4 (a) and (b) illustrate the local shapes of sample spectra (including reference spectra) and subtraction spectra at seven weathering stages for each soil sample. Local segments of sample spectra in Phase II exhibit consistent and mutually similar pattern, closely resembling those of reference spectra. However, none of

these segments reveal absorption features at the diagnostic bands of petroleum hydrocarbon. In contrast, local segments in Phase I appear irregular and inconsistent, with some of them exhibiting at least one absorption feature near 2,310, 2,348, 1,727, and 1762 nm. This consistency observed between the sample spectra in Phase II and reference spectra suggests that the spectral responses in soil with low petroleum hydrocarbon and long weathering times (not less than 12 months) are nearly negligible. Whereas, those discrepancies in Phase I, especially the appearance of absorption features at the diagnostic bands of petroleum hydrocarbons, reaffirm that slightly polluted soil with shorter weathering times exhibits observable spectral responses.

Fig. 4 (c) and (d) provide detailed insights into local shapes of subtraction spectra at seven weathering stages for each soil sample. Notably, the local segments in the regions associated with petroleum hydrocarbons on the subtraction spectra show an overall downward concave feature. Especially, those in Phase I display a distinct homogeneously asymmetric "W-shaped" double absorption feature. There are obvious absorption valleys near 2,310, 2,348 nm in Fig. 4 (c) and 1,727, 1,762 nm in Fig. 4 (d), and the absorption intensity of the left valley is higher than that of the right for both regions. Conversely, local segments in Phase II display an irregular and inconsistent pattern, with most of them featuring more than two absorption valleys. Both rough concave and specific valley on these spectral segments can be regarded as detectable spectral response of petroleum hydrocarbon. It indicates that almost all of the polluted soils in this study are detectable based on concave features, whose lowest concentration of petroleum hydrocarbons is 178 mg/kg. Meanwhile, the polluted soils in Phase I also can be detected using absorption valleys due to their regular and consistent

pattern.

3.2. Parameters of diagnostic spectra

3.2.1. Integral area of spectral segment

To quantify the concave feature of diagnostic spectra, the integral areas (S) of spectral segments within either 2,290–2,370 nm or 1,700–1,780 nm were calculated. The results are summarized in Table 2. Based on the proportion of negative S values, it is clear that most subtraction spectra exhibit a pronounced downward concave feature. From the perspective of weathering phase, 100% subtraction spectra in Phase I and more than 80% in Phase II display negative S values. Notably, Phase I exhibits S values with an absolute mean approximately had about 3.5 times larger than Phase II along with a lower coefficient of variation. These results validate the detectability of slight spectral responses of low petroleum hydrocarbons polluted soil using parameter S. Moreover, shorter weathered or more polluted soil is easier to detect due to stronger absorption intensity and higher stability.

Considering the spectral region, local segments within both 2,290–2,370 nm and 1,700–1,780 nm in Phase I display an equal proportion of negative S values. The absorption intensity (absolute value of mean S) at former region is about 2 times stronger than the later with similar coefficient of variation. Yet, the former region possesses higher proportion of negative S and stronger absorption intensity with lower coefficient of variation in Phase II. It suggests that the concave feature within 2,290–2,370 nm on subtraction spectrum is more pronounced and stable, especially for those longer weathered or lower contaminated soil, in this study. In

other words, this spectral region is more competitive for petroleum hydrocarbons determination.

3.2.2. Parameters of absorption valley

To check the presence of absorption valley and its validity to indicate hydrocarbon, valley positions (P) were analyzed and sifted, according to the spectra-structure correlations for near infrared spectroscopy. Then Number (N), FWHM (W), Depth (D) and fitting Area (A) were chose and calculated to parameterize the main features of absorption valley. Table 3 summarizes the all parameters for each absorption valley in each phase. In the subtraction spectra of Phase I, there were two main absorption valleys within either 2,290–2,370 nm or 1,700–1,780 nm. Their positions locate near 2,310, 2,348 nm or 1,727, 1762 nm, whose occurrence proportions are 100, 100, 100 and 90% respectively. W of the absorption valleys near 2,310, 2,348, and 1,762 nm are close, while the one near 1,727 nm is approximately 1.5 times wider. Both of H and A present decreasing trend when the absorption position shift left. These parameters delineate the common asymmetric "W-shaped" double absorption feature in Phase I quantitatively. It not only confirms that 2,310 nm and 1,727 nm are the most typical positions within 2,290–2,370 nm and 1,700–1,780 nm for petroleum hydrocarbons detection, but also reveals that 2,348 nm or 1,762 nm can be utilized as auxiliary.

In Phase II, more absorption valleys appear within both the 2,290–2,370 nm and 1,700–1,780 nm ranges. Most of these valleys cluster around 2,306, 2,321, 2,339, 2,351, 2,360, 1,709, 1,724, 1,743, and 1,762 nm, but their occurrence proportions are relatively low, ranging from 32.69 % to 57.69 %. Their W, D, and A values decrease consequentially. These

phenomena represent that absorption valleys will diversify and their intensity will weaken after long-term weathering (not less than 12 months). These changes bring challenges to petroleum hydrocarbons detection based on absorption valley.

4. Discussion

4.1. Diagnostic spectral features

This study revealed that sample spectrum of low petroleum hydrocarbon contaminated soil possessed slightly non-conclusive and even negligible spectral response, but diagnostic features — downward concave or absorption valley occurred on their diagnostic spectrum after transformation at either 2,290-2,370 nm or 1,700-1,780 nm. Specifically, a shared pattern of asymmetric "W-shaped" double absorption valley was found consistently in Phase I, while more than two weak absorption valleys with changeable numbers and positions were observed in Phase II. Overall, a roughly downward concave feature could be viewed in either Phase I or Phase II. These features appearing on diagnostic spectrum of low petroleum hydrocarbon contaminated soil are consistent with the diagnostic features on sample spectra of heavily contaminated soil reported in previous studies (Lammoglia & de Souza Filho, 2011; Pabón & de Souza Filho, 2016; Pabón et al., 2019; Scafutto & de Souza Filho, 2016; Scafutto et al., 2017).

These diagnostic features and their change with weathering time can be interpreted by the near-infrared absorbing mechanism of C-H bond vibration (Cloutis, 1989; Workman & Weyer, 2012). Relatively fixed absorption valleys locating near 2,310 nm, 2,348nm, 1,727 nm

and 1,762 nm in Phase I, can attributed to several main aliphatic C-H bonds which dominating in petroleum hydrocarbon within a short-term weathering. Spectral superposition by dominant type of C-H bond weakens with the volatilization and degradation in Phase II, more but weak absorption valleys, caused by other C-H bonds like those on aromatic base, appear subsequently.

4.2. Detection capability

This study demonstrated that both downward concave and absorption valley of diagnostic spectrum could be parameterized for the detection of low petroleum hydrocarbon contamination. In relative terms, the integral area (S) of the diagnostic spectrum is a more accessible and sensitive indicator for detecting low petroleum hydrocarbons pollution. This parameter, even when extracted from sample spectrum is less susceptible to noise and the shifting effect caused by changes in petroleum composition and measurement conditions (Achard et al., 2021; Winkelmann, 2005). However, it has not gained widespread usage. One possible explanation is that the specific materials exhibiting a concave feature in the petroleum hydrocarbon-related regions of their sample spectra remain largely unexplored. Spectral region within 2,290-2,370 nm on diagnostic spectrum is better for the detection of low petroleum hydrocarbons pollution than that within 1,700-1,780 nm, from the perspective of absorption intensity. However, previous studies have shown that there are wider range of other materials, including dolomitic soil, clays, latosol, gibbsite, alunite, gypsum, exhibit absorption features within the 2,290-2,370 nm region (Asadzadeh & de Souza Filho, 2017; Scafutto et al., 2016). Therefore, 2,290-2,370 nm range is preferable to 1,700-1,780 nm for

petroleum hydrocarbons determination only when interferences from other materials can be effectively excluded. Similarly, though the absorption intensities near 2,310 nm, 2,348nm, 1,727 nm and 1,762 nm are decreasing in sequence measured by Depth (D) or fitting Area (A) of absorption valley, their priorities for petroleum hydrocarbons determination should consider interferences. Besides, this study revealed the FWHM of the absorption valley near 1,727 nm was much wider. It implies that it had the least impact by valley shift, which might be one of the reasons why previous studies favored this position for petroleum hydrocarbons detection (Kühn et al., 2004).

The detection limit based on absorption valley is between 952 mg/kg (T12m) and 1,037 mg/kg (T4m), and that based on downward concave feature is at least as low as 207 mg/kg (T24m). Comparing to the detection limit of knowledge-based NIRS method based on sample spectrum (exceeding 10,000 mg/kg), it is about 10-50 times lower (Pabón & de Souza Filho, 2016; Pelta & Ben-Dor, 2019; Winkelmann, 2005; Yin et al., 2017).

4.3. Implications

This study verified the advantages of the diagnostic spectrum construction and parameterization workflow for low petroleum hydrocarbon detection. The success of this workflow originated from (1) reducing the non-target spectrum attributed to soil matrix with spectral subtraction method, (2) eliminating the noise through fast Fourier transform filtering, (3) enhancing the spectral responses of petroleum hydrocarbons by spectral segment interception and baseline correction, and then (4) parameterizing to detect diagnostic features (downward concave and absorption valley). The findings based on this workflow update our

understanding about the spectral response of low petroleum hydrocarbon pollution. It extended the application potential of knowledge-based NIRS detection approach, especially for environmental management, because its detection limit is much lower than the most common control value for soil quality management in many countries. Besides, this workflow is generic and feasible for both field detection and remote sensing identification.

4.4. Limitations

A theoretical hypothesis of knowledge-based NIRS detection approach is that absorption features (overall downward concave and specific absorption valley) within either 2,290-2,370 nm or 1,700-1,780 nm can be regarded as the indicator of petroleum hydrocarbon detection. In real scenario, plastics, paints, dry vegetation and other minerals possess similar or even same diagnostic features, which may interfere the detection of petroleum hydrocarbon. In this study, we excluded these interferences mainly by pairing the spectra of the same soil sample before (reference spectrum) and after pollution (sample spectrum) to construct subtraction spectrum. Obviously, it is difficult to select ideal reference spectrum with the same background as sample spectrum. Roughly, maybe two types of spectra could be selected as the alternative reference spectra in practice: (1) historical spectra of the target before hydrocarbon pollution, (2) spectra of material with the same or similar compositions as the target without hydrocarbon pollution in the same region. However, the impact of the difference between ideal and alternative reference spectrum remains unknown, and the selection process of alternative reference spectrum need the experiences and judgements of the technical personnel. Thus, more comprehensive and thorough researches should be explore

in the future. Especially, the impact of the spatialtemporal heterogeneity between sample and reference should be assessed, and specific workflow and criterion of alternative reference spectrum selection should be constructed.

5. Conclusions

This study proposed a technical workflow of diagnostic spectrum construction and parameterization for the detection of low petroleum hydrocarbon pollution. The results revealed that nearly all diagnostic spectra possessed overall downward concave features within either 2,290-2,370 nm or 1,700-1,780 nm, and specially most of those with shorter weathered or higher polluted soil (weathered time more than 8 months or concentration exceeding 1,000 mg/kg) appeared an asymmetric "W-shaped" double absorption features. The detection limit based on diagnostic features could be about 200 mg/kg, which is much lower than control values of total petroleum hydrocarbons in many countries. These findings not only update our understanding on the spectral response of low petroleum hydrocarbon pollution, but also widely extend the application of knowledge-based NIRS approach in either field detection or remote sensing identification for environmental management.

Funding

This research was supported by the National Key R&D Program of China (No. 2018YFC1800100).

Acknowledgements

We thank Editage (www.editage.cn) for English language editing. We also thank the

relevant editor and the anonymous reviewers for their valuable comments and suggestions.

References

Achard, V., Foucher, P.-Y., & Dubucq, D. (2021). Hydrocarbon pollution detection and mapping based on the combination of various hyperspectral imaging processing tools. *Remote Sensing*, *13*(5), 1020. https://doi.org/10.3390/rs13051020.

Albers, P. H. (2002). Petroleum and individual polycyclic aromatic hydrocarbons. In *Handbook of ecotoxicology* (pp. 365-396). CRC Press. https://doi.org/10.1201/9781420032505.ch14.

Allen, C. S., & Satterwhite, M. B. (2006). Reflectance spectra of three liquid hydrocarbons on a common sand type. *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XII* (Vol. 6233, pp. 575-586). SPIE. https://doi.org/10.1117/12.665586.

Asadzadeh, S., de Oliveira, W. J., & de Souza Filho, C. R. (2022). UAV-based remote sensing for the petroleum industry and environmental monitoring: State-of-the-art and perspectives. *Journal of Petroleum Science and Engineering*, *208*, 109633. https://doi.org/10.1016/j.petrol.2021.109633.

Asadzadeh, S., & de Souza Filho, C. R. (2016). A review on spectral processing methods for geological remote sensing. *International journal of applied earth observation and geoinformation*, *47*, 69-90. https://doi.org/10.1016/j.jag.2015.12.004.

Asadzadeh, S., & de Souza Filho, C. R. (2017). Spectral remote sensing for onshore seepage characterization: A critical overview. *Earth-Science Reviews*, *168*, 48-72.

https://doi.org/10.1016/j.earscirev.2017.03.004.

Aske, N., Kallevik, H., & Sjöblom, J. (2001). Determination of saturate, aromatic, resin, and asphaltenic (SARA) components in crude oils by means of infrared and near-infrared spectroscopy. *Energy & Fuels*, *15*(5), 1304-1312. https://doi.org/10.1021/ef010088h.

Brekke, C., & Solberg, A. H. (2005). Oil spill detection by satellite remote sensing. *Remote sensing of environment*, *95*(1), 1-13. https://doi.org/10.1016/j.rse.2004.11.015.

Brown, A. J. (2006). Spectral curve fitting for automatic hyperspectral data analysis. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(6), 1601-1608. https://doi.org/10.1109/TGRS.2006.870435.

Cloutis, E. A. (1989). Spectral reflectance properties of hydrocarbons: remote-sensing implications. *Science*, *245*(4914), 165-168. https://doi.org/10.1126/science.245.4914.165.

Douglas, R. K., Nawar, S., Alamar, M. C., Mouazen, A., & Coulon, F. (2018). Rapid prediction of total petroleum hydrocarbons concentration in contaminated soil using vis-NIR spectroscopy and regression techniques. *Science of the Total Environment*, *616*, 147-155. https://doi.org/10.1016/j.scitotenv.2017.10.323.

Dubucq, D., Turon, L., Blanco, B., & Bideaud, H. (2021). Earth observation remote sensing for oil and gas: A new era. *The Leading Edge*, *40*(1), 26-34. https://doi.org/10.1190/tle40010026.1.

Ellis, J. M., Davis, H., & Zamudio, J. A. (2001). Exploring for onshore oil seeps with hyperspectral imaging. *Oil and Gas Journal*, *99*(37), 49-58.

Ellis Jr, R., & Adams Jr, R. S. (1961). Contamination of soils by petroleum hydrocarbons.

Advances in agronomy, *13*, 197-216. https://doi.org/10.1016/S0065-2113(08)60959-1.

Hörig, B., Kühn, F., Oschütz, F., & Lehmann, F. (2001). HyMap hyperspectral remote sensing to detect hydrocarbons. *International Journal of Remote Sensing*, *22*(8), 1413-1422. https://doi.org/10.1080/01431160120909.

Huguenin, R., & Jones, J. (1986). Intelligent information extraction from reflectance spectra: Absorption band positions. *Journal of Geophysical Research: Solid Earth*, *91*(B9), 9585-9598. https://doi.org/10.1029/JB091iB09p09585.

Khan, M. A. I., Biswas, B., Smith, E., Naidu, R., & Megharaj, M. (2018). Toxicity assessment of fresh and weathered petroleum hydrocarbons in contaminated soil-a review. *Chemosphere*, *212*, 755-767. https://doi.org/10.1016/j.chemosphere.2018.08.094.

Kokaly, R. F., Couvillion, B. R., Holloway, J. M., Roberts, D. A., Ustin, S. L., Peterson, S. H., Khanna, S., & Piazza, S. C. (2013). Spectroscopic remote sensing of the distribution and persistence of oil from the Deepwater Horizon spill in Barataria Bay marshes. *Remote Sensing of Environment*, *129*, 210-230. https://doi.org/10.1016/j.rse.2012.10.028.

Kühn, F., Oppermann, K., & Hörig, B. (2004). Hydrocarbon Index–an algorithm for hyperspectral detection of hydrocarbons. *International Journal of Remote Sensing*, *25*(12), 2467-2473. https://doi.org/10.1080/01431160310001642287.

Kuppusamy, S., Maddela, N. R., Megharaj, M., & Venkateswarlu, K. (2020). Total petroleum hydrocarbons. Environmental Fate, Toxicity, and Remediation. https://doi.org/10.1007/978-3-030-24035-6.

Lammoglia, T., & Filho, C. R. D. S. . (2011). Spectroscopic characterization of oils yielded

from brazilian offshore basins: potential applications of remote sensing. *Remote Sensing of Environment*, 115(10), 2525-2535. https://doi.org/10.1016/j.rse.2011.04.038.

Margenot, A. J., Calderón, F. J., & Parikh, S. J. (2016). Limitations and potential of spectral subtractions in Fourier-transform infrared spectroscopy of soil samples. *Soil Science Society of America Journal*, *80*(1), 10-26. https://doi.org/10.2136/sssaj2015.06.0228.

Margenot, A. J., Parikh, S. J., & Calderón, F. J. (2019). Improving infrared spectroscopy characterization of soil organic matter with spectral subtractions. *JoVE (Journal of Visualized Experiments)*(143), e57464. https://doi.org/10.3791/57464.

MEE. (2018). Soil environmental quality Risk control standard for soil contamination of development land. In (Vol. GB 36600-2018). Beijing: China Environmental Science Press.

Ng, W., Malone, B. P., & Minasny, B. (2017). Rapid assessment of petroleum-contaminated soils with infrared spectroscopy. *Geoderma*, *289*, 150-160. https://doi.org/10.1016/j.geoderma.2016.11.030.

Pabón, R. E. C., & de Souza Filho, C. R. (2016). Spectroscopic characterization of red latosols contaminated by petroleum-hydrocarbon and empirical model to estimate pollutant content and type. *Remote Sensing of Environment*, *175*, 323-336. https://doi.org/10.1016/j.rse.2016.01.005.

Pabón, R. E. C., de Souza Filho, C. R., & de Oliveira, W. J. (2019). Reflectance and imaging spectroscopy applied to detection of petroleum hydrocarbon pollution in bare soils. *Science of the Total Environment*, *649*, 1224-1236.

https://doi.org/10.1016/j.scitotenv.2018.08.231.

Pasquini, C. (2018). Near infrared spectroscopy: A mature analytical technique with new perspectives–A review. *Analytica chimica acta*, *1026*, 8-36. https://doi.org/10.1016/j.aca.2018.04.004.

Pelta, R., & Ben-Dor, E. (2019). Assessing the detection limit of petroleum hydrocarbon in soils using hyperspectral remote-sensing. *Remote Sensing of Environment*, *224*, 145-153. https://doi.org/10.1016/j.rse.2019.01.026.

Scafutto, R. D. P. M., & de Souza Filho, C. R. (2016). Quantitative characterization of crude oils and fuels in mineral substrates using reflectance spectroscopy: Implications for remote sensing. *International journal of applied earth observation and geoinformation*, *50*, 221-242. https://doi.org/10.1016/j.jag.2016.03.017.

Scafutto, R. D. P. M., de Souza Filho, C. R., & de Oliveira, W. J. (2017). Hyperspectral remote sensing detection of petroleum hydrocarbons in mixtures with mineral substrates: Implications for onshore exploration and monitoring. *ISPRS Journal of Photogrammetry and Remote Sensing*, *128*, 146-157. https://doi.org/10.1016/j.isprsjprs.2017.03.009.

Scafutto, R. D. P. M., de Souza Filho, C. R., & Rivard, B. (2016). Characterization of mineral substrates impregnated with crude oils using proximal infrared hyperspectral imaging. *Remote Sensing of Environment*, *179*, 116-130. https://doi.org/10.1016/j.rse.2016.03.033.

Seo, J., Yu, J., & Wang, L. (2023). Indicator Spectral Bands and Logistic Models for Detecting Diesel and Gasoline Polluted Soils based on Close-Range Hyperspectral Image Data.

IEEE Transactions on Geoscience and Remote Sensing. https://doi.org/10.1109/TGRS.2023.3264967.

Short, N. (2011). NASA remote sensing tutorial oil and gas. In *technical report*. NASA.

Smith, B. (2021). Spectral subtraction. *Spectroscopy*, *36*(5), 14–19-14–19.

Tang, J., Wang, M., Wang, F., Sun, Q., & Zhou, Q. (2011). Eco-toxicity of petroleum hydrocarbon contaminated soil. *Journal of Environmental Sciences*, *23*(5), 845-851. https://doi.org/10.1016/S1001-0742(10)60517-7.

VROM. (2012). Soil Remediation Circular 2009. In (Vol. N 6563). The Hague, the Netherlands: Ministry of Housing, Spatial Planning and the Environment.

Winkelmann, K. H. (2005). On the applicability of imaging spectrometry for the detection and investigation of contaminated sites with particular consideration given to the detection of fuel hydrocarbon contaminants in soil (Doctoral dissertation, BTU Cottbus-Senftenberg).

Workman, J., & Weyer, L. (2012). Practical guide and spectral atlas for interpretive near-infrared spectroscopy *(Second Edition)*. CRC. https://doi.org/10.1201/b11894.

Yin, W., Chen, Z., Jiao, Y., Liu, H., & Liu, Q. (2017). Research on Visible-Near Infrared Spectral Characterization of Purplish Soil Contaminated with Petroleum Hydrocarbon and Estimation of Pollutant Content. *Spectroscopy and Spectral Analysis*, 37(12), 3924-3931. https://doi.org/10.3964/j.issn.1000-0593(2017)12-3924-08.

Zhao, H., Zhang, L., Zhang, X., Liu, J., Wu, T., & Wang, S. (2015). Hyperspectral feature extraction based on the reference spectral background removal method. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6):

2832-2844. https://doi.org/10.1109/JSTARS.2015.2401052.

Author Contributions Statement

Qijia Lou: Conceptualization, Data Curation, Methodology, Software, Visualization, Writing – original draft. **Mei Lei:** Conceptualization, Writing – review & editing, Supervision, Funding acquisition. **Yu Wang:** Data Curation, Software, Writing – original draft. **Shaobin Wang:** Conceptualization, Writing – review & editing. **Guanghui Guo:** Methodology, Writing – review & editing. **Wencheng Xiong:** Conceptualization, Methodology. **Ying Jiang:** Writing – review & editing. **Tienan Ju:** Data Curation, Methodology, Software, Project administration. **Xiaofeng Zhao:** Data Curation, Formal Analysis. **Frederic Coulon:** Data curation, Supervision.

Journal Pre-proof

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof

Figures

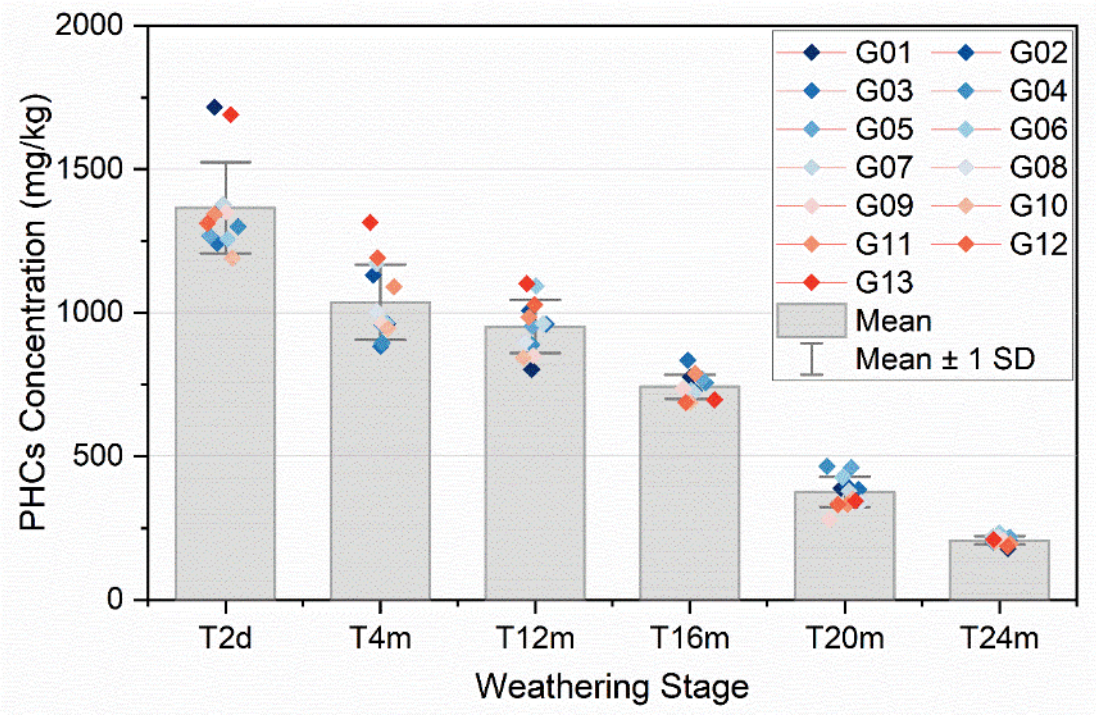


Fig. 1. Petroleum hydrocarbon concentrations of each soil sample at six weathering stages. G01~13 represent 13 soil samples respectively.

Journal Pre-proof

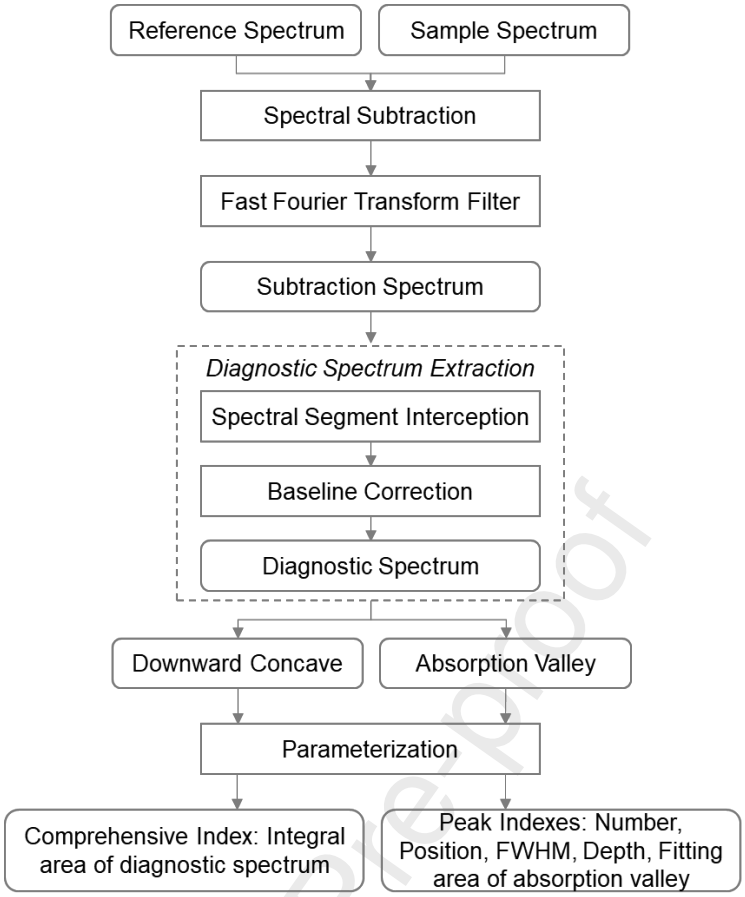
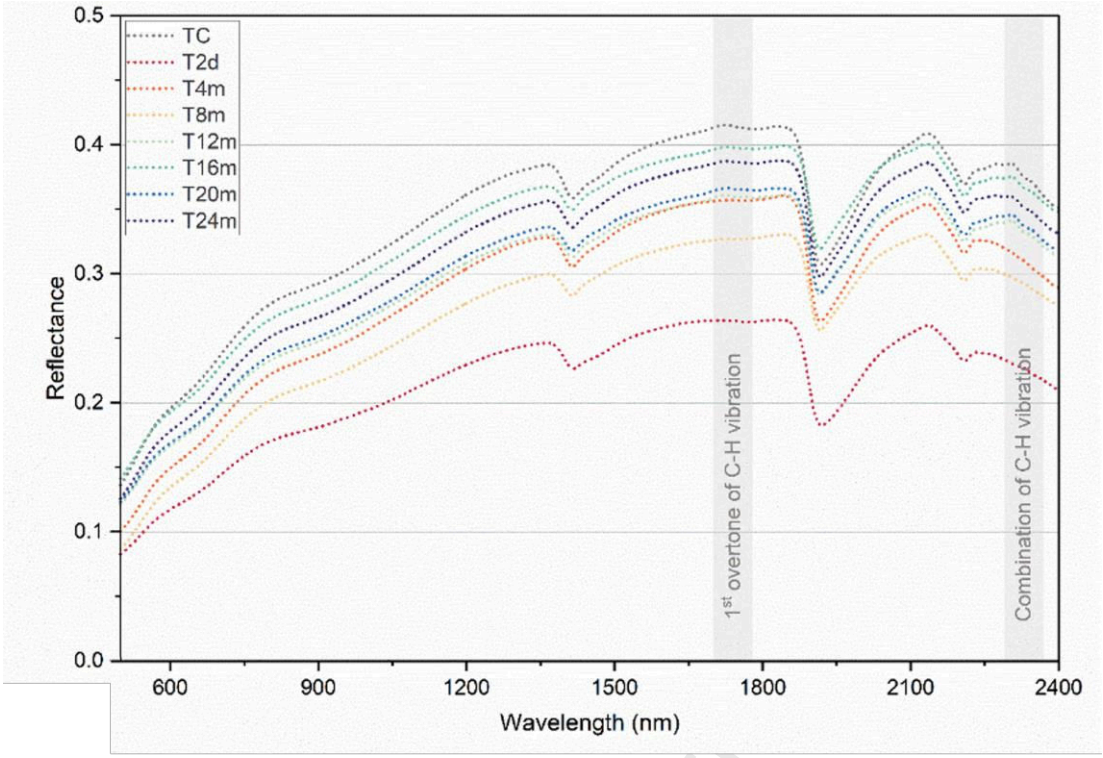
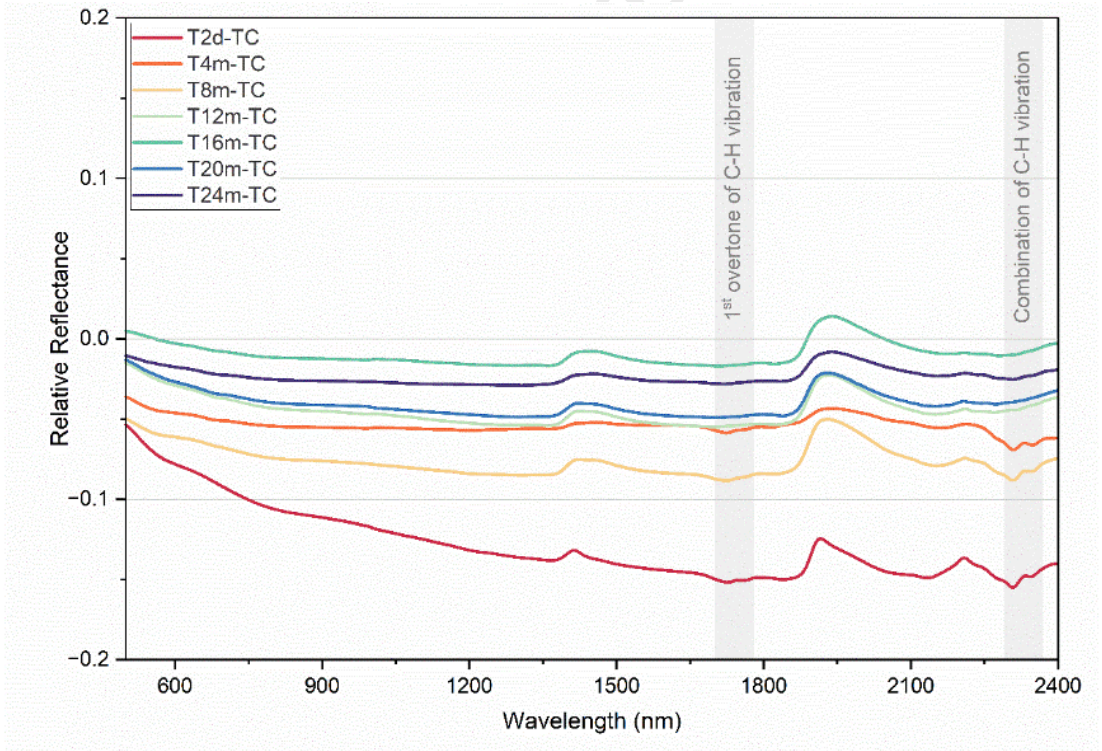


Fig. 2. Technical workflow of diagnostic spectrum construction and parameterization

Journal Pre-proof



(a) Sample spectra



(b) Subtraction spectra

Fig. 3. Spectral curve in full range (400–2,400 nm) at different weathering stages: (a) Mean reflectance of sample spectra; (b) Mean relative reflectance of subtraction spectra. Two gray

regions at 2,290-2,370 nm and 1,700-1,780 nm refer to petroleum hydrocarbon related regions attributed to the combination and 1st overtone of C-H vibration.

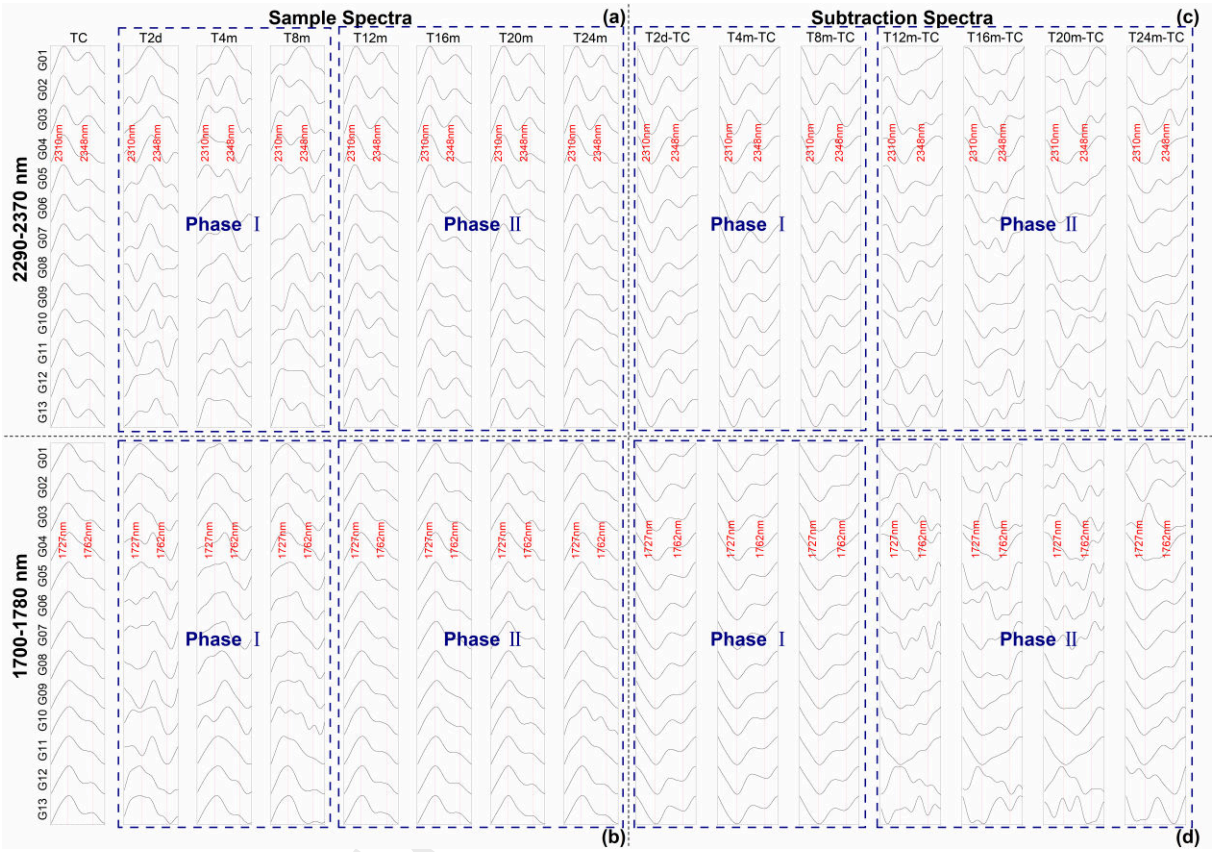


Fig. 4. Local shapes at diagnostic regions of both sample spectra and subtraction spectra: (a) Sample spectra at 2,290-2,370 nm; (b) Sample spectra at 1,700-1,780 nm; (c) Subtraction spectra at 2,290-2,370 nm; (d) Subtraction spectra at 1,700-1,780 nm. Positions of absorption valley near 2,310 nm, 2,348 nm, 1,727 nm, 1,762 nm are the common indicator of petroleum hydrocarbon. Phase I (T2d, T4m, T8m) represents relatively higher polluted, while Phase II (T12m, T16m, T20m, T24m) represents lower polluted.

Tables

Table 1. Parameters of diagnostic spectral features

Diagnostic feature	Parameter	Description
Downward concave	Integral area (S)	A comprehensive index, negative S indicates overall downward concave of diagnostic spectrum. Used to detect petroleum hydrocarbon roughly. The larger the absolute S, the more severe the pollution. It is tolerant for the changeable and shiftable diagnostic features.
Absorption valley	Position (P)	Refers to wavelength positions of absorption valley attributed to C-H vibrations. Used to detect petroleum hydrocarbon and distinguish its dominant composition. The corresponding relationship between P and the composition type can be queried from previous study (Workman & Weyer, 2012).
	Number (N)	Numbers of P within diagnostic spectrum. N correlates simultaneously with the component types and their proportions. A large N value implies the presence of multiple component types and each of them have fair proportions,

Journal Pre-proof

	while a small N value suggests a few dominant component types.
FWHM (W)	Refers to Full Width at Half Maximum. It is relatively fixed for a specific absorption valley and changes with P.
Depth (D)	Indicates the absorption strength. It is positively correlated with petroleum hydrocarbons content. The larger the D, the more severe the pollution.
Fitting area (A)	A synthetic indicator related with W and D. Its physical meaning is similar with D.

Journal Pre-proof

Table 2. Statistics of integral areas of spectral segments for the subtraction spectra.

Weathering Phase	Spectral Segment (nm)	Proportion of Negative S (%)	Mean	Standard Deviation	Coefficient of Variation
Phase I	2290-2370	100.00	-0.2068	0.0413	-0.1997
	1700-1780	100.00	-0.0957	0.0187	-0.1954
Phase II	2290-2370	94.23	-0.0656	0.0316	-0.4817
	1700-1780	80.77	-0.0255	0.0161	-0.6314

Note: S-Integral area of the absorption region.

Journal Pre-proof

Table 3. Spectral parameters of the absorption valleys of the subtraction spectra.

Weathering Phase	Spectral Segment (nm)	N	P (nm)	Proportion of existing P (%)	W (nm)	D	A
I	2290-2370	2	2310±1	100.00	23±2	-0.0068±0.0012	-0.1635±0.0370
			2348±2	100.00	24±6	-0.0041±0.0008	-0.1079±0.0386
	1700-1780	2	1727±2	100.00	33±7	-0.0026±0.0006	-0.0947±0.0371
			1762±2	89.74	22±4	-0.0014±0.0003	-0.0325±0.0099
II	2290-2370	5	2306±2	40.38	17±8	-0.0018±0.0013	-0.0426±0.0655
			2321±2	46.15	15±6	-0.0013±0.0006	-0.0230±0.0143
			2339±2	51.92	18±16	-0.0013±0.0007	-0.0342±0.0456
			2351±2	36.53	13±5	-0.0012±0.0005	-0.0173±0.0114
			2360±1	57.69	12±7	-0.0012±0.0010	-0.0208±0.0462
	1700-1780	4	1709±2	32.69	10±2	-0.0003±0.0002	-0.0038±0.0020

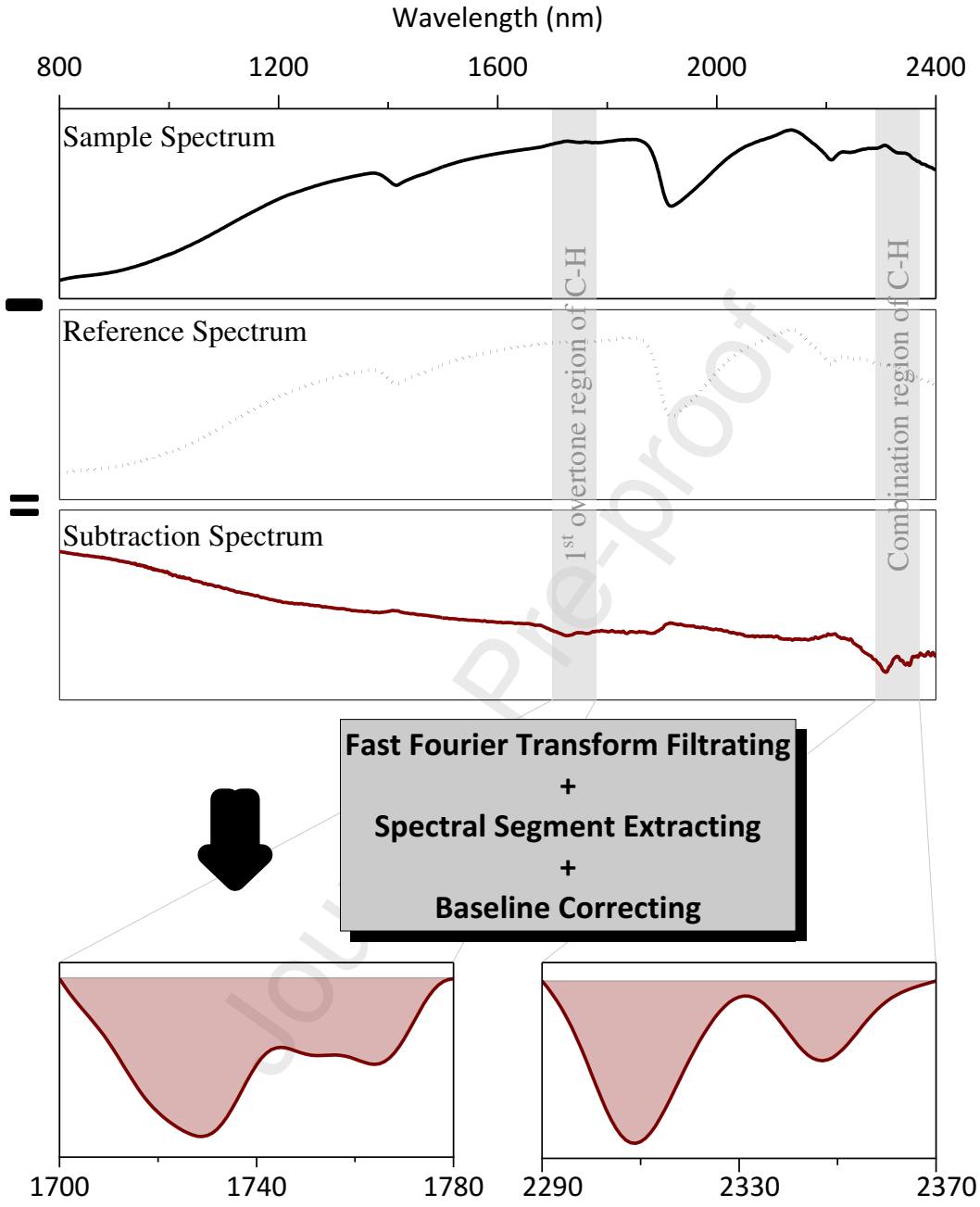
Journal Pre-proof

	1724±2	42.31	23±13	-0.0008±0.0005	-0.0249±0.0294
	1743±1	38.46	15±9	-0.0004±0.0002	-0.0082±0.0098
	1762±2	50.00	20±25	-0.0005±0.0004	-0.0172±0.0405

Notes: N-Numbers of absorption valley; P-Positions of absorption valley; W-Full width at half maxima; D-Absorption depth; A-Fitting area of absorption valley.

Journal Pre-proof

Graphical abstract



HIGHLIGHTS

- A new workflow was developed to construct and parameterize diagnostic spectrum
- NIRS diagnostic features for low-level petroleum hydrocarbon pollution were found
- Detection limit significantly improved which enables detection even below 200 mg/kg

Journal Pre-proof

https://dspace.org/

Diagnostic features emerging in near-infrared reflectance spectroscopy for low petroleum hydrocarbon pollution after spectral subtraction

Lou, Qijia; Lei, Mei; Wang, Yu; Wang, Shaobin; Guo, Guanghui ...

2024-06-10
Elsevier
https://doi.org/10.1016/j.scitotenv.2024.172264
Lou Q, Lei M, Wang Y, et al., (2024) Diagnostic features emerging in near-infrared reflectance spectroscopy for low petroleum hydrocarbon pollution after spectral subtraction. Science of The Total Environment, Volume 928, June 2024, Article number 172264

Downloaded from a DSpace repository. https://dspace.org/