Estimating persistent oil contamination in tropical region using vegetation indices and random forest regression

Guillaume Lassalle, Anthony Crédoz, Rémy Hédacq, Georges Bertoni, Dominique Dubucq, Sophie Fabre, Arnaud Elger

Estimating persistent oil contamination in tropical region using vegetation indices and Random Forest regression

Guillaume Lassalle[a, b], *Anthony Credoz*[b], *Rémy Hédacq*[b], *Georges Bertoni*[c], *Dominique Dubucq*[d], *Sophie Fabre*[a], *Arnaud Elger*[e]

AUTHOR ADDRESS

[a] Office National d'Études et de Recherches Aérospatiales (ONERA), Toulouse, France

[b] TOTAL S.A., Pôle d'Études et de Recherches de Lacq, Lacq, France

[c] DynaFor, Université de Toulouse, INRA, Castanet-Tolosan, France

[d] TOTAL S.A., Centre Scientifique et Technique Jean-Féger, Pau, France

[e] EcoLab, Université de Toulouse, CNRS, INPT, UPS, Toulouse, France

*Corresponding author: Guillaume Lassalle, Office National d'Études et de Recherches Aérospatiales, 2 Avenue Edouard Belin, 31055 Toulouse, France; E-mail: guillaume.lassalle@onera.fr

Keywords: reflectance spectroscopy, soil contamination, Total Petroleum Hydrocarbons, vegetation indices, Random Forest

ABSTRACT

The persistence of soil contamination after cessation of oil activities remains a major environmental issue in tropical regions. The assessment of the contamination is particularly difficult on vegetated sites, but promising advances in reflectance spectroscopy have recently emerged for this purpose. This study aimed to exploit vegetation reflectance for estimating low concentrations of Total Petroleum Hydrocarbons (TPH) in soils. A greenhouse experiment was carried out for 42 days on *Cenchrus alopecuroides* (L.) under realistic tropical conditions. The species was grown on oil-contaminated mud pit soils from industrial sites, with various concentrations of TPH. After 42 days, a significant decrease in plant growth and leaf chlorophyll and carotenoid contents was observed for plants exposed to 5 to 19 $g.kg^{-1}$ TPH in comparison to the controls ($p < 0.05$). Conversely, pigment contents were higher for plants exposed to 1 $g.kg^{-1}$ TPH (hormesis phenomenon). These modifications proportionally affected the reflectance of C. alopecuroides at leaf and plant scales, especially in the visible region around 550 and 700 nm. 33 vegetation indices were used for linking the biochemical and spectral responses of the species to oil using elastic net regressions. The established models indicated that chlorophylls a and b and β-carotene were the main pigments involved in the modifications of reflectance ($R^2 > 0.7$). The same indices also succeeded in estimating the concentrations of TPH using random forest regression, at leaf and plant scales (RMSE = 1.46 and 1.63 $g.kg^{-1}$ and RPD = 5.09 and 4.44, respectively). Four out of the 33 indices contributed the most to the models (>75%). This study opens up encouraging perspectives for monitoring the cessation of oil activities in tropical regions. Further researches will focus on the application of our approach at larger scale, on airborne and satellite imagery.

1. Introduction

Along with its development during the last century, oil and gas industry has become a major source of contamination in the environment (Barraza et al., 2018; Durango-Cordero et al., 2018; Ogri, 2001; Romero et al., 2017). Crude oil and by-products (e.g. petroleum products, wastewater, oil sludge) are frequently released in soils, following accidental facility failures, bad practices and, more rarely, storm events (Ahmadun et al., 2009; Chang and Lin, 2006; Correa Pabón et al., 2019; Hu et al., 2013). They cause important ecological disturbances, because oil

contaminants – especially Total Petroleum Hydrocarbons (TPH) – are highly toxic toward organisms (Bejarano and Michel, 2016; Finer et al., 2008; Merkl et al., 2004). The contamination of soil may persist in brownfields and mud pits after cessation of the oil production activity and affect ecosystems on the long term (Lassalle et al., 2019b, 2019a). The monitoring of oil activities has therefore become a critical environmental issue, and gave rise to an increasing need in reliable and cost-effective methods for assessing soil contamination. For this purpose, reflectance spectroscopy provided promising results when applied to bare soils (Correa Pabón and de Souza Filho, 2016; Scafutto and de Souza Filho, 2016). Its application to multi- and hyperspectral remote sensing imagery makes possible to detect and quantify soil TPH content at large scale using airborne sensors (Correa Pabón et al., 2019; Scafutto et al., 2017). More recently, a new approach has been proposed to extend its use to vegetated areas, where oil cannot be detected directly at the soil surface (Lassalle et al., 2018; Rosso et al., 2005; Sanches et al., 2013a). This approach shows great interest in tropical regions, where vegetation is particularly dense (Achard et al., 2018; Adamu et al., 2018; Arellano et al., 2015).

TPH affect vegetation health – especially pigment and water contents – and optical properties in the reflective domain (400 – 2500 nm) (Balliana et al., 2017; Emengini et al., 2013a; Rosso et al., 2005; Tran et al., 2018). By exploiting the reflectance of leaves and canopies at particular wavelengths, it is possible to detect the changes induced by TPH in leaf biochemistry (Gürtler et al., 2018; Lassalle et al., 2018; Sanches et al., 2013a). This approach is however still under development and needs further improvements in order to be used operationally. Methods based on vegetation indices (VI) (i.e. reflectance ratios) have succeeded in distinguishing healthy and oil-exposed vegetation and in discriminating among various types of oil contamination, under controlled and natural conditions (Emengini et al., 2013a; Lassalle et al., 2019b). In recent studies, TPH were accurately detected from satellite images (Arellano et al., 2015), but their quantification remains rarely considered, although it is of great interest for assessing the environmental risks deriving from the contamination.

The quantification of TPH remains a major challenge, as it implies tracking little variations in leaf biochemistry from its reflectance. This might be particularly difficult on contaminated brownfields and mud pits, because the established species are particularly tolerant to oil exposure (Credoz et al., 2016; Lassalle et al., 2019b). This difficulty is moreover enhanced when dealing with low TPH concentrations, for two main reasons. In most cases, the lower the concentration,

the lower the effects on vegetation health and reflectance (Emengini et al., 2013a; Sanches et al., 2013a). These effects become more difficult to detect below a certain concentration (Lassalle et al., 2019b). Moreover, some species undergo a stimulation of growth under exposure to low oil contamination (Kirk et al., 2002; Lin et al., 2002). It is particularly frequent for tropical species. These phenomena make the quantification of TPH very challenging using vegetation reflectance.

Previous studies focusing on heavy metals – which is a very different context of soil contamination – achieved accurate quantification of these compounds by exploiting reflectance data in empirical univariate regression models and in Partial Least Square Regression (PLSR) (Shi et al., 2014b, 2014a). Similar approaches proved efficient for quantifying high levels of TPH contamination (up to 77 $g.kg^{-1}$) in our previous study (Lassalle et al., 2019a), but became ineffective below a certain concentration (~ 20 $g.kg^{-1}$). So, to date, there is still no evidence that low levels of TPH can be quantified using vegetation reflectance, especially in tropical region. To achieve this, alternative methods are necessary. Emerging machine learning methods, which proved efficient for solving complex regression problems, could help quantifying low TPH concentrations (Hastie et al., 2017; Breiman, 2001).

This study aims to assess the potential and limits of vegetation reflectance for quantifying low TPH concentrations in soils, under controlled conditions. The proposed approach relies on the combination of VI and a machine learning method – namely Random Forest (RF) regression. This method was tested on reflectance measurements performed at various acquisition scales, under realistic tropical conditions.

2. Materials and Methods

2.1. Study site and greenhouse experiment

An experiment was carried out for 42 days, for which a realistic case of oil contamination was reproduced under controlled tropical conditions. An industrial vegetated mud pit located in tropical region was identified for this purpose. Residues of oil production have been accumulated on this site resulted in persistent oil contamination in soils. After cessation of the activity, the site has been colonized by herbaceous vegetation, mainly *Cenchrus alopecuroides* (L.). The experiment focused on this species, which proved good indicator of oil contamination in previous studies (Emengini et al., 2013c; Lassalle et al., 2018). To reproduce as faithfully as

possible the conditions of the mud pit, the soil was sampled on two locations of the site and used in the experiment. Soil analyses revealed low concentrations of BTEX and Polycyclic Aromatic Hydrocarbons (PAHs) in both samples, and C_5-C_{40} TPH concentrations of 5 and 21 $g.kg^{-1}$, respectively. The lower (C_5-C_{10}) and intermediate (C_{10}-C_{21}) hydrocarbon fractions have been partially degraded, so the remaining contamination mostly came from the dense ones (C_{21}-C_{40} hydrocarbons), which are poorly degradable and whose uptake by roots is very limited (Lassalle et al., 2019b). Detailed soil analyses can be found in the Supporting Information section. No substantial change in TPH concentrations was observed for the different treatments at the end of the experiment (data not shown). In addition, an uncontaminated soil with similar texture was sampled on another site and used in the experiment. All the soils were homogenized manually and sieved to 10 mm to remove coarse root residuals.

Five treatments were applied to *C. alopecuroides* for the greenhouse experiment, with TPH concentrations ranging from 0 to 19 $g.kg^{-1}$. For this purpose, the soils sampled on the mud pit and the control site were mixed in varying proportions to obtain five levels of contamination, respectively 0 (control), 1, 5, 13 and 19 $g.kg^{-1}$ TPH (see Supporting Information). Cultivated young plants were acclimated for 15 days in greenhouse and reached 15-cm height before being transplanted in individual pots filled with a 3-cm layer of clay balls and 3 L of the corresponding treatment soil. For each treatment, 11 replicates were grown for 42 days in May and June at 27° C and 70% hygrometry, with a 12:12 light:dark photoperiod provided by natural and artificial light. N-P-K (6-6-6) fertilization was applied weekly to the plants, which were irrigated to field capacity every day. At the end of the experiment, plant shoots were harvested, oven-dried for 48 h at 60° C and then weighed.

2.2. Biochemical analyses and reflectance acquisitions

The biochemical and spectral responses of *C. alopecuroides* to the treatments were measured during the study. For this purpose, we followed the procedure described in Lassalle et al. (2019b). Young leaves were sampled on five different replicates per treatment and served for determining leaf water (LWC) and pigment contents (Arellano et al., 2015; Diepens et al., 2017). Leaf pigment content was quantified by High Pressure Liquid Chromatography (HPLC), and included chlorophylls a and b and various carotenoids, such as β-carotene, lutein, antheraxanthin, violaxanthin and zeaxanthin (Barlow et al., 1996). Differences among the treatments were

analyzed through ANOVA and Tukey post-hoc tests. In addition, the spectral reflectance of the same sampled leaves was measured following the protocol described below, and linked to biochemical analyses (see section 2.3). This procedure was carried out after 21 and 42 day of experiment. No leaf was sampled before in order to avoid influencing plant growth during early stages.

Reflectance measurements were also performed directly on the plants (i.e. without picking leaves) at day 0, 21 and 42, using an ASD FieldSpec 4 Hi-Res spectroradiometer (Malvern Panalytical, Malvern, UK). Data were acquired in radiance in the reflective domain (400 – 2400 nm) and converted to reflectance using a white reference calibration panel (Spectralon, Labsphere Inc., North Sutton, USA) (Milton, 1987). For each treatment, three leaves per replicate were measured at random (n = 165 spectra per date) using a leaf-clip with an internal light source (10 measurements averaged per leaf). Additional measurements were conducted at plant scale, by placing a 10°-FOV fore-optic above the pots to obtain a 5-cm wide acquisition footprint (10 measurements averaged per plant). These measurements were performed on each replicate (n = 55 spectra per date) under natural light between 11.30 am and 1.30 pm, under clear sky. The reflectance data from the 1350 – 1450 and 1800 – 1950 nm ranges were not conserved, because of low atmospheric transmission at plant scale. A Savitzky-Golay smoothing filter was then applied at the remaining wavelengths, improving the signal-to-noise ratio (Savitzky and Golay, 1964).

2.3. Vegetation indices

In this study, 33 VI were computed and used in two ways. They are listed in the Supporting Information section. These indices have been specifically exploited for detecting oil-induced changes in leaf biochemistry in our previous study, so here they were tested for quantifying TPH (Lassalle et al. 2019b). Most of them exploited reflectance in the visible (VIS) region (450 – 700 nm), which is the most important spectrum region for assessing oil contamination. It has been shown that under exposure to oil, changes in the values of a single vegetation index result from the alteration of several inter-correlated pigments that share common light absorption features (e.g. chlorophylls and carotenoids). Consequently, the pigments to which an index has been initially linked in the literature may differ from those involved in the response to oil exposure. So, it is necessary to identify which pigments contribute the most to index changes, in order to

understand how vegetation reflectance is affected by oil. For this purpose, the 33 indices were first computed from the reflectance data of leaves sampled for biochemical analyses, and each of them was linked to leaf pigment and water contents using Elastic net (ENET) multiple regression (Zou and Hastie, 2005), as described in Lassalle et al. (2019b). ENET is a penalized least squared regression method that allows selecting predictors under multicollinearity. Multicollinearity appears when predictors – in our case, leaf pigment contents – are linear combination of each other, and results in confusions when selecting those contributing the most to the target variable (i.e. VI) (Belsley et al., 1980; Dormann et al., 2013). ENET regression prevents from such consequences and has already proved efficient for identifying biochemical parameters involved in reflectance changes under exposure to contaminants (Lassalle et al., 2019b). For each VI, the best predictors were retained and the R^2 of the model was calculated.

In a second time, the same 33 indices were computed from the reflectance data acquired at leaf and plant scales and used for predicting TPH concentrations using RF regression (Breiman, 2001). RF is an ensemble method that builds and average a lot of independent decision trees to model the relationship between predictors (i.e. the 33 VI) and a response variable – here, TPH concentration (Hastie et al., 2017; Hutengs and Vohland, 2016; Mutanga et al., 2012). Each tree is constructed from a set of decision rules that results from the recursive fragmentation of the original space into successively smaller sub-regions. To define these sub-regions, binary splits are applied independently to each vegetation index. A simple model is then adjusted between VI and TPH concentrations with respect to the splits, and those minimizing the mean squared error between the measured and predicted concentrations are retained. The succession of the selected index splits represents the branches of the decision tree, in such a way that each set of decision rules leads to an estimation of TPH concentration (see Hutengs & Vohland (2016) for a complete description). RF accounts for non-linear relationships between the predictors and the response variable – which can occur for low concentrations of contaminants, so it was of great interest in our case. Moreover, RF informs on the relative contribution of VI to the quantification of TPH (Breiman, 2001; Grömping, 2009), which is essential for operational applications of the method. The RF regression was fitted to half of the data (50% training set) and tested on the remaining part (50% test set), at leaf and plant scales separately (Lassalle et al., 2019b, 2019a). For this purpose, we used the data from day 42, because no difference among the treatments was observed on the previous dates, so no quantification of TPH was possible (see section 3.1). The

predictions of TPH made on the test set were compared to those from initial soil analyses based on the R^2, the Root Mean Squared Error (RMSE) and the Residual Predictive Deviation (RPD) (Shi et al., 2014). The RPD denotes the ratio of the standard deviation of the measured TPH concentrations to the RMSE calculated between the measured and predicted concentrations.

3. Results and discussion

3.1. Biochemical and spectral responses to oil contamination

As expected, *C. alopecuroides* was highly tolerant to oil, as none of the plants from the contaminated treatments died during the experiment. This observation is consistent with previous studies carried out on the same species with various types of oil contamination (engine oil and oil and gas waste mud) (Emengini et al., 2013c; Lassalle et al., 2018). No visible stress symptom was observed on any of the replicates before the last stages of the experiment (35 – 42 days), so no difference in leaf pigment contents and in leaf and plant reflectance was observed among the treatments on the previous dates (0 and 21 days, data not shown). On day 42, leaf discoloration was observed for the individuals exposed to 5 $g.kg^{-1}$ TPH or more, expressing a significant decrease in leaf chlorophyll and carotenoid contents when compared to the control ($p < 0.05$) (**Fig. 1**). Plant growth was also significantly affected on this date. The alterations in pigment contents induced an increase of reflectance by 5 to 10% in the VIS region, especially around 550 and 700 nm (**Fig. 2**). Other regions of the spectrum were also affected in the same way by contamination. This response was more pronounced at leaf scale, because plant reflectance is also influenced by plant architecture, background soil and illuminating and viewing geometry (Asner, 1997; Sanches et al., 2013a). As suggested in previous works (Emengini et al., 2013c; Lassalle et al., 2018), the late response of *C. alopecuroides* to oil highlights its interest for monitoring long-term contaminated sites, especially brownfields and mud pits remaining after cessation of oil and gas activities.

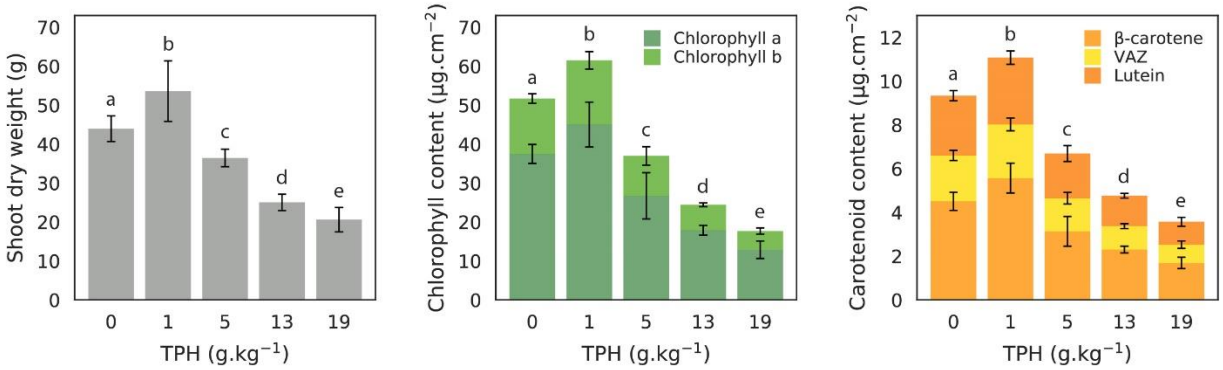

Figure 1. Shoot dry weight and leaf chlorophyll and carotenoid contents (mean ± SD; n = 11 and 5 samples per treatment, respectively) of *C. alopecuroides* after 42 days of exposure to various levels of Total Petroleum Hydrocarbons (TPH). Significant differences among the treatments are denoted by different lowercase letters (ANOVA and Tukey post-hoc tests, $p < 0.05$). (VAZ: Violaxanthin + Antheraxanthin + Zeaxanthin.)

Alterations in leaf biochemistry and optical properties are very common for vegetation exposed to contaminants (Balliana et al., 2017; Lassalle et al., 2019b; Rosso et al., 2005; Sanches et al., 2013a). They depend on multiple factors, especially the species, the contamination type and concentration and the time of exposure, and have been previously observed on *C. alopecuroides* and other tropical species (Arellano et al., 2017b; Emengini et al., 2013c; Lassalle et al., 2017). In the case of oil, these alterations result from the reduction of water and nutrient availability in soils and uptake capacities of roots (Athar et al., 2016; Balasubramaniyam and Harvey, 2014; Nie et al., 2011). Additional effects may come from the accumulation of hydrocarbons in tissues (Baruah et al., 2014), but they mostly involve low-carbon TPH, so they are unlikely to explain the responses observed in our experiment. The amplitude of these biochemical and spectral responses was determined by the level of contamination, which confirmed previous results obtained for higher levels of oil (Emengini et al., 2013c; Sanches et al., 2013a). The more the TPH, the lower the leaf pigment contents and the higher the reflectance (**Fig. 2**). Consequently, leaf chlorophyll content felt to 17.7 (± 3.0) µg.cm^{-2} under exposure to 19 g.kg^{-1} TPH, while it reached 51.7 (± 3.5) µg.cm^{-2} for control at the end of the experiment (**Fig. 1**). A similar trend was observed on carotenoids, and suggested being able to discriminate among the different levels of contamination using reflectance data.

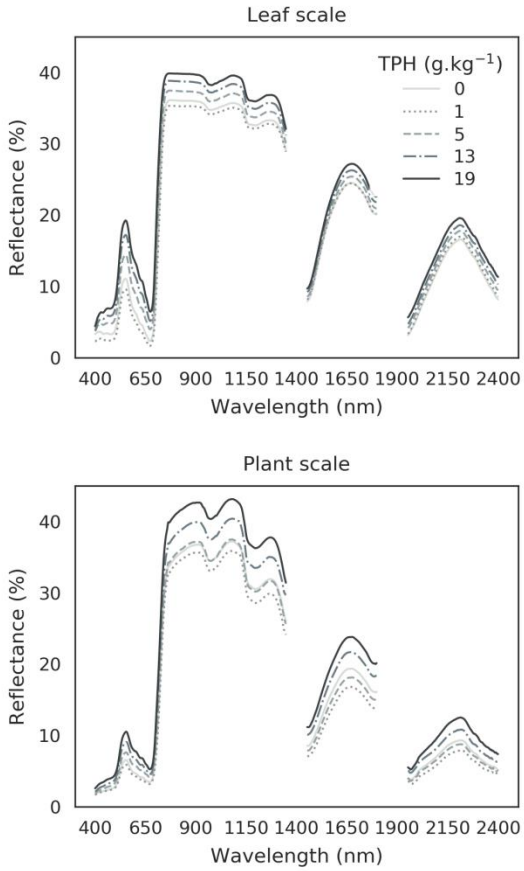

Figure 2. Mean reflectance of *C. alopecuroides* in the reflective domain after 42 days of exposure to various levels of Total Petroleum Hydrocarbons (TPH), at leaf and plant scales. Reflectance data from the 1350–1450 and 1800–1950 intervals were removed, because of low atmospheric transmission at plant scale. (The legend is common to both figures.)

In contrast to the other treatments, a stimulating effect was observed for plants exposed to 1 $g.kg^{-1}$ TPH. Shoot dry weight and leaf chlorophyll and carotenoid contents were significantly higher for this treatment than those of the control ($p < 0.05$), and reflectance was slightly lower at leaf and plant scales. This particular response – called hormesis – has already been observed on the growth of other oil-exposed species in previous studies (Kirk et al., 2002; Lin et al., 2002; Salanitro et al., 1997). However, its causes remain poorly documented. To our knowledge, this study is the first to bring evidence of this phenomenon using reflectance measurements. Previous experiments aimed to simulate accidental oil spills (e.g. pipeline leakage), by applying high

doses of oil – mainly diesel or engine oil – to crop species, which are particularly sensitive to such contamination (Emengini et al., 2013a; Gürtler et al., 2018; Sanches et al., 2013a). Consequently, strong alterations in leaf biochemistry and reflectance have been observed regardless of the contamination level, so no hormesis has been noticed. The responses of *C. alopecuroides* observed in our study substantially differed, because we focused on a totally different context. When oil contamination persists in industrial brownfield and mud pit soils – such as that reproduced in our experiment, the most phytotoxic fractions of petroleum hydrocarbons (e.g. BTEX, PAH) are almost absent from the mixture. The effects on oil-tolerant established vegetation are therefore much less pronounced than those observed for other types of oil contamination with high proportion of these compounds (e.g. crude oil, diesel, gasoline), especially at low concentration (Lassalle et al., 2019b). Here, PAHs and BTEX were found below the detection limits (0.02 and 0.05 $mg.kg^{-1}$, respectively) at 1 $g.kg^{-1}$ of TPH and in the control soil, so they did not affect the plants negatively. Conversely, these compounds are likely to induce positive effects at very low concentration (Maliszewska-Kordybach and Smreczak, 2000), but they cannot be the only contaminants responsible for the differences observed between the two treatments. Some other hydrocarbon fractions – which were absent from the control soil – enhance the availability of nutrients and organic matter for plants by stimulating microorganisms (Li et al., 1990). Therefore, the very low concentrations of PAHs and BTEX combined to those of other hydrocarbon fractions might have induced the hormetic effect observed in this study.

3.2. Vegetation indices

3.2.1. Elastic net regressions

The VI succeeded in linking the biochemical and spectral responses of *C. alopecuroides* to oil contamination. 20 out of the 33 indices tested were strongly correlated to leaf pigment or water content ($R^2 \geq 0.7$) (**Tab. 1**). For most of them, at least three contributing pigments were identified, the others having not being retained by the ENET regression. Chlorophylls and β-carotene were often the most contributing ones, and provided some of the best models ($R^2 \geq 0.8$) when used together (i.e. without addition of other carotenoids). Therefore, these pigments explained most of the modifications in reflectance observed among the treatments in the VIS

region. Other carotenoid pigments – especially lutein and violaxanthin – were frequently retained in the models (e.g. GM2 and PSSRb indices), thus confirming previous results obtained on another oil-tolerant species (Lassalle et al., 2019b). These pigment are usually of less contribution to leaf optical properties, because of the masking effect of chlorophylls, which are present at higher concentrations in leaves and share common light absorption features (Feret et al., 2008; Zhang et al., 2017). Our study shows that they also contribute to the spectral response of vegetation to oil contamination.

Table 1. Results of the elastic net regressions performed on the 33 vegetation indices. For each index, the R^2 of the model is presented, along with the contributing set of biochemical parameters (in order of importance). (Chl a: Chlorophyll a, Chl b: Chlorophyll b, B-car: β-carotene, Lut: Lutein, Ant: Antheraxanthin, Vio: Violaxanthin, Zea: Zeaxanthin, LWC: Leaf Water Content.)

Index	R^2	Pigments
CARI	0.76	B-car, Chl b
CCI	0.73	B-car, Chl b, Chl a
CTR1	0.24	Vio, B-car, Chl b, Lut, Ant, Zea, Chl a
CTR2	0.81	Chl a, Chl b
GM1	0.84	B-car, Chl b, Chl a
GM2	0.87	Chl a, Chl b, B-car, Lut
LI1	0.65	Chl a, Chl b
LI2	0.18	Chl a
LI3	0.42	Vio, Lut, Ant, Zea, B-car, Chl b, Chl a
mCARI1	0.43	Ant, B-car, Lut, Chl b, Zea, Chl a
mCARI2	0.12	Chl a
mSR705	0.72	Chl a
MTCI	0.84	Chl a, Chl b
ND705	0.82	B-car, Chl b, Chl a
PRI1	0.54	B-car, Vio, Chl b, Zea, Ant, Chl a, Lut
PRI2	0.67	Chl b, Chl a
PRI3	0.63	Chl b, Chl a, Lut
PSRI	0.43	B-car, Chl b, Lut, Ant, Zea, Chl a
PSSRa	0.72	Chl a
PSSRb	0.88	Chl a, Chl b, Lut, B-car, Vio
PSSRc	0.83	Chl a
SIPI1	0.71	Chl a, Chl b
SIPI2	0.81	Chl a, Chl b
SIPI3	0.73	Chl a, Chl b
SR705	0.83	B-car, Chl b, Chl a
TCARI	0.68	B-car, Chl b, Chl a
TCARI_OSAVI	0.75	B-car, Chl b, Chl a
VOG1	0.75	B-car, Chl b, Chl a
VOG2	0.7	B-car, Chl b, Chl a
VOG3	0.71	B-car, Chl b, Vio, Chl a
DWSI	0.82	LWC

The close relationship between some indices and contributing pigments are illustrated in Figure 3. As observed, these indices exhibited strong link with single pigment content. ENET regression allowed identifying those contributing the most to index changes, in order of

importance. These results confirmed the interest of combining pigment contents into multiple models for better understanding the effects induced by oil contamination on vegetation reflectance (Lassalle et al., 2019b). Most of the indices exploited reflectance around 550 and 700 nm, which was particularly affected by oil (**Fig. 2**), so they were particularly adapted to our context of study. The same wavelengths also proved efficient for assessing soil contaminated by oil and by-products in other situations (oil extraction, pipeline leakages, etc.). Zhu et al. (2014) exploited reflectance at 700 nm for distinguishing among various levels of alteration in pigment contents induced by phenanthrene contamination on *Suaeda salsa*. Likewise, Sanches et al. (Sanches et al., 2013b) used the same spectrum region for assessing the effects of oil leakages on crops species. One of the main advantages of VI relies on their robustness. Some of them remain only little affected by bare soil and plant architecture, so they can be applied for tracking subtle alterations in leaf pigment contents induced by oil at different acquisition scales (e.g. leaf, plant and canopy) (Blackburn, 1998; Dash and Curran, 2007). For example, the same 33 indices succeeded in discriminating among various types of oil contamination from leaf to canopy scales in previous study (Lassalle et al., 2019b). Here, they were suitable for estimating TPH concentrations in soils.

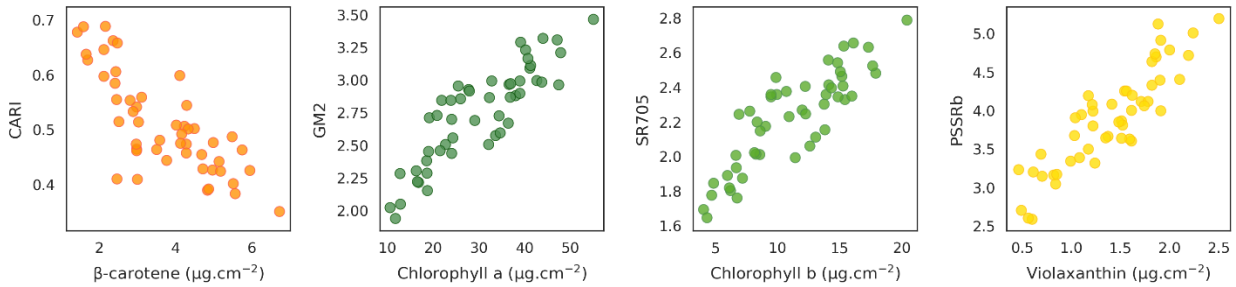

Figure 3. Relationship between vegetation indices and leaf pigment contents. Figures include data from leaves sampled on all the treatments, after 21 and 42 days of experiment (n = 50).

3.2.2. Random Forest regressions

RF regressions performed using the 33 VI provided accurate predictions of TPH concentrations on the test set, at leaf and plant scales. These results are presented in Figure 4. High R^2 values – respectively 0.96 and 0.95 at leaf and plant scales – indicated strong correlation between the measured and predicted concentrations on the test sets. Predictions of TPH were

very close to the true concentrations, as indicated by the low RMSE values obtained at both scales. This accuracy was confirmed by RPD values greater than 4.4. RPD provides a better interpretation of predictions, because it compares the RMSE to the variability in the true TPH concentrations. RPD values above 2 or 3 are usually needed for considering the models reliable for use. The 33-VI based RF models were therefore particularly adapted for estimating TPH in our study. The analysis of residuals revealed that the highest level of contamination (19 $g.kg^{-1}$) was the most difficult to predict, as it was almost systematically underestimated (**Fig. 4**). This may be caused by saturation in the spectral response of *C. alopecuroides* to this type of oil contamination, as described for other species in previous study. Confusions also arose at the lowest concentrations (0 – 1 $g.kg^{-1}$). They highlighted the difficulty to detect the hormesis phenomenon using reflectance data, for which only little differences were observed among the two treatments (**Fig. 2**). These results thus helped identifying the detection limit of our approach.

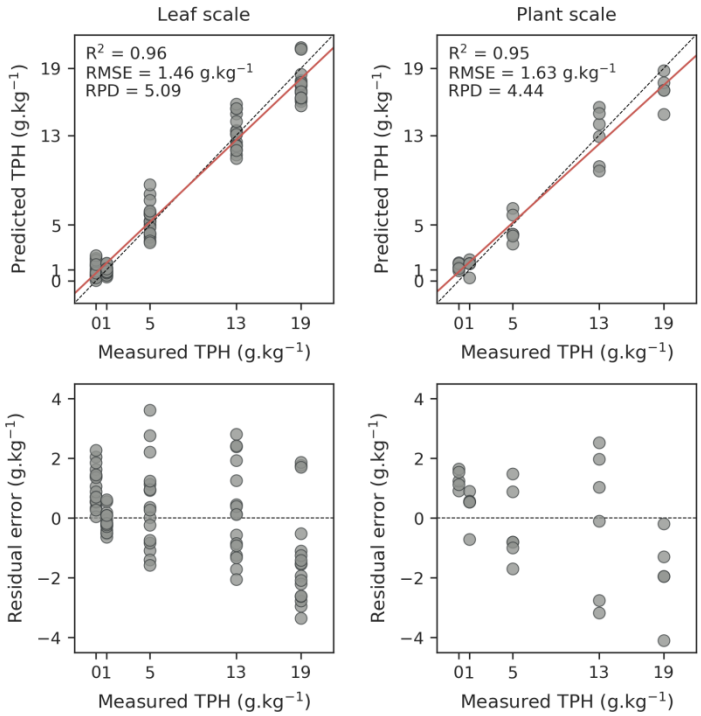

Figure 4. Comparison between the measured and predicted concentrations of Total Petroleum Hydrocarbons (TPH) using the 33 vegetation indices and the random forest regression on the test set, at leaf and plant scales (top figures), and residuals of the predictions (bottom figures).

The relative contribution of VI to the RF models was consistent from leaf to plant scales. Four out of the 33 indices contributed the most to TPH predictions (> 75%, **Fig. 5**). These indices were among those closely linked to leaf pigment contents ($R^2 > 0.7$, **Tab. 1**), and more precisely chlorophylls and β-carotene. They exploit reflectance around 550 and 700 nm, but also 670 – 680 and 750 nm, which helped enhancing the differences among the treatments. Only few wavelengths are therefore necessary for estimating TPH accurately using reflectance spectroscopy. Conversely, the other indices remained of little contribution to the models (< 5% each), especially those linked to additional carotenoids (e.g. lutein, violaxanthin), because they brought redundant information.

The consistence of the results between acquisition scales opens up promising perspectives for operational applications of our approach under natural conditions. Some of the main indices – especially the MTCI and TCARI / OSAVI – have been originally designed for imagery application purposes (Dash and Curran, 2007; Haboudane et al., 2002), and proved efficient for detecting – but not for quantifying – oil contamination in tropical regions. For example, Arellano et al. (2015) used the MTCI index to discriminate among control and oil-contaminated sites in the Amazon forest using hyperspectral satellite images. However, the application of these indices for estimating the level of contamination remained unexplored until now. This study was a first attempt in that direction, but further researches are needed to assess the reliability of our approach under natural conditions, using field measurements, and in the long term using multi- and hyperspectral imagery. High to very high spatial resolution might help achieving this, because the contamination can occur on a few square meters and be therefore difficult to detect using medium to low spatial resolution imagery (Adamu et al., 2018; Arellano et al., 2015). As discussed in section 3.1, the case of persistent low contamination in brownfields and mud pits makes the estimation of TPH very challenging, because of the composition of the contamination and the tolerance of the species. Under natural conditions, additional difficulties should be considered before application of our approach. Vegetation established on brownfields and mud pits is exposed to a multitude of environmental factors that affect its health and reflectance (e.g. drought, waterlogging), especially in the tropical regions with marked seasonality (Adamu et al., 2016; Smith et al., 2004). These factors might overlap with the effects induced by TPH, making their estimation more difficult. Thanks to recent advances, it seems however possible to discriminate among oil and other stressors using reflectance spectroscopy (Emengini et al.,

2013b; Lassalle et al., 2019b, 2018). Because of the selective growing conditions imposed, only few species are generally established on soils with oil (Noomen et al., 2012), as observed on our study site. This makes the estimation of TPH feasible, provided that these species are oil-sensitive. Species richness is however very high in some tropical regions subject to oil contamination (Arellano et al., 2017a, 2017b). Such diversity in plant species means important differences in sensitivity to oil among them. Some species are totally tolerant to oil, whereas others are affected even at very low concentration. This makes the estimation of TPH more difficult, so it is necessary to identify the most suitable species before, as we did under controlled conditions.

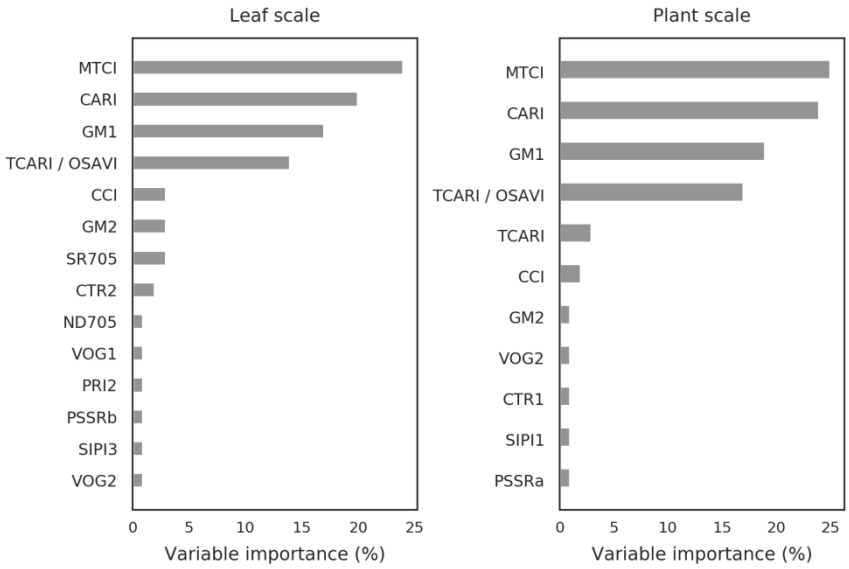

Figure 5. Relative contribution of the vegetation indices to the estimation of Total Petroleum Hydrocarbons (TPH) using random forest regression, at leaf and plant scales. Only indices with non-zero contribution are displayed.

4. Conclusion

This study aimed to quantify low TPH concentrations in soils using *C. alopecuroides* reflectance under controlled tropical conditions. Modifications in leaf biochemistry appeared after a long-term exposure, depending on the level of contamination. Pigment contents were reduced for TPH concentrations of 5 $g.kg^{-1}$ and above, whereas they increased at lower

concentration when compared to control. These modifications were linked to the reflectance data through VI, which brought evidence of the implication of chlorophylls and various carotenoids in the spectral response of the species to oil. The same indices succeeded in predicting TPH concentrations with good accuracy using RF regressions, at leaf and plant scales. Four out of the 33 indices tested were almost sufficient to achieve these predictions. This study emphasizes the potential of reflectance spectroscopy for quantifying oil contamination in tropical regions with dense vegetation. More specifically, the long-term response of *C. alopecuroides* to oil highlights its interest for assessing persistent contamination, for example after cessation of the oil production activity. Although our approach is at early stage and needs further improvements, we are convinced that it could soon result in imagery applications. The emergence of new satellite- and UAV-embedded hyperspectral sensors is sparking a growing interest by oil and gas companies, because they could help assessing oil contamination locally or at large scale. Our study under controlled conditions was the first necessary step prior to such applications. In its continuity, our upcoming research will focus on the adaptation of the approach to hyperspectral imagery with high spatial resolution, and its assessment in tropical region with heavy past oil and gas production activities.

Acknowledgements

This interdisciplinary work was achieved in the frame of the NAOMI R&D project between TOTAL and the ONERA, with the support of Ecolab and Dynafor research units of Toulouse. The authors gratefully acknowledge E. Buffan-Dubau, D. Lambrigot and O. Berseille for their assistance in pigment analysis.

Declarations of interest: None.

Funding: Financial support of this work was provided by TOTAL.

433 References

434

435 Achard, V., Fabre, S., Alakian, A., Dubucq, D., Déliot, P., 2018. Direct or indirect on-shore
436 hydrocarbon detection methods applied to hyperspectral data in tropical area, in: Michel, U.,
437 Schulz, K. (Eds.), Earth Resources and Environmental Remote Sensing/GIS Applications
438 IX. SPIE, p. 22. https://doi.org/10.1117/12.2325097

439 Adamu, B., Tansey, K., Ogutu, B., 2018. Remote sensing for detection and monitoring of
440 vegetation affected by oil spills. Int. J. Remote Sens. 39, 3628–3645.
441 https://doi.org/10.1080/01431161.2018.1448483

442 Adamu, B., Tansey, K., Ogutu, B., 2016. An investigation into the factors influencing the
443 detectability of oil spills using spectral indices in an oil-polluted environment. Int. J.
444 Remote Sens. 37, 2338–2357. https://doi.org/10.1080/01431161.2016.1176271

445 Ahmadun, F.-R., Pendashteh, A., Abdullah, L.C., Awang Biak, D.R., Madaeni, S.S., Abidin,
446 Z.Z., 2009. Review of technologies for oil and gas produced water treatment. J. Hazard.
447 Mater. 170, 530–551. https://doi.org/10.1016/j.jhazmat.2009.05.044

448 Arellano, P., Tansey, K., Balzter, H., Boyd, D.S., 2017a. Field spectroscopy and radiative
449 transfer modelling to assess impacts of petroleum pollution on biophysical and biochemical
450 parameters of the Amazon rainforest. Environ. Earth Sci. 76, 1–14.
451 https://doi.org/10.1007/s12665-017-6536-6

452 Arellano, P., Tansey, K., Balzter, H., Boyd, D.S., 2015. Detecting the effects of hydrocarbon
453 pollution in the Amazon forest using hyperspectral satellite images. Environ. Pollut. 205,
454 225–239. https://doi.org/10.1016/j.envpol.2015.05.041

455 Arellano, P., Tansey, K., Balzter, H., Tellkamp, M., 2017b. Plant family-specific impacts of
456 petroleum pollution on biodiversity and leaf chlorophyll content in the Amazon rainforest of
457 Ecuador. PLoS One 12. https://doi.org/10.1371/journal.pone.0169867

458 Asner, G.P., 1997. Biophysical and Biochemical Sources of Variability in Canopy Reflectance.
459 Remote Sens. Environ. 64, 234–253.

460 Athar, H. ur R., Ambreen, S., Javed, M., Hina, M., Rasul, S., Zafar, Z.U., Manzoor, H., Ogbaga,
461 C.C., Afzal, M., Al-Qurainy, F., Ashraf, M., 2016. Influence of sub-lethal crude oil
462 concentration on growth, water relations and photosynthetic capacity of maize (Zea mays
463 L.) plants. Environ. Sci. Pollut. Res. 23, 18320–18331. https://doi.org/10.1007/s11356-016-
464 6976-7

465 Balasubramaniyam, A., Harvey, P.J., 2014. Scanning electron microscopic investigations of root
466 structural modifications arising from growth in crude oil-contaminated sand. Environ. Sci.
467 Pollut. Res. 21, 12651–12661. https://doi.org/10.1007/s11356-014-3138-7

468 Balliana, A.G., Moura, B.B., Inckot, R.C., Bona, C., 2017. Development of Canavalia ensiformis
469 in soil contaminated with diesel oil. Environ. Sci. Pollut. Res. 24, 979–986.
470 https://doi.org/10.1007/s11356-016-7674-1

471 Barlow, R.G., Cummings, D.G., Gibb, S.W., 1996. Improved resolution of mono- and divinyl
472 chlorophylls a and b and zeaxanthin and lutein in phytoplankton extracts using reverse
473 phase C-8 HPLC. Mar. Ecol. Prog. Ser. 161, 303–307.

474 Barraza, F., Maurice, L., Uzu, G., Becerra, S., López, F., Ochoa-Herrera, V., Ruales, J., Schreck,
475 E., 2018. Distribution, contents and health risk assessment of metal(loid)s in small-scale
476 farms in the Ecuadorian Amazon: An insight into impacts of oil activities. Sci. Total
477 Environ. 622–623, 106–120. https://doi.org/10.1016/j.scitotenv.2017.11.246

478 Baruah, P., Saikia, R.R., Baruah, P.P., Deka, S., 2014. Effect of crude oil contamination on the
479 chlorophyll content and morpho-anatomy of Cyperus brevifolius (Rottb.) Hassk. Environ.
480 Sci. Pollut. Res. 21, 12530–12538. https://doi.org/10.1007/s11356-014-3195-y

481 Bejarano, A.C., Michel, J., 2016. Oil spills and their impacts on sand beach invertebrate
482 communities: A literature review. Environ. Pollut. 218, 709–722.
483 https://doi.org/10.1016/j.envpol.2016.07.065

484 Belsley, D.A., Kuh, E., Wemsch, R.E., 1980. Detecting and Assessing Collinearity, in: John
485 Wiley & Sons, I. (Ed.), Regression Diagnostics: Idenhfiing Injuential Data and Sources of
486 Collineariq.

487 Blackburn, G.A., 1998. Quantifying chlorophylls and carotenoids at leaf and canopy scales: An
488 evaluation of some hyperspectral approaches. Remote Sens. Environ. 66, 273–285.
489 https://doi.org/10.1016/S0034-4257(98)00059-5

490 Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
491 https://doi.org/https://doi.org/10.1023/A:1010933404324

492 Chang, J.I., Lin, C.C., 2006. A study of storage tank accidents. J. Loss Prev. Process Ind. 19, 51–
493 59. https://doi.org/10.1016/j.jlp.2005.05.015

494 Correa Pabón, R.E., de Souza Filho, C.R., 2016. Spectroscopic characterization of red latosols
495 contaminated by petroleum-hydrocarbon and empirical model to estimate pollutant content
496 and type. Remote Sens. Environ. 175, 323–336. https://doi.org/10.1016/j.rse.2016.01.005

497 Correa Pabón, R.E., de Souza Filho, C.R., Oliveira, W.J. de, 2019. Reflectance and imaging
498 spectroscopy applied to detection of petroleum hydrocarbon pollution in bare soils. Sci.
499 Total Environ. 649, 1224–1236. https://doi.org/10.1016/j.scitotenv.2018.08.231

500 Credoz, A., Hédacq, R., Barreau, C., Dubucq, D., 2016. Experimental study of hyperspectral
501 responses of plants grown on mud pit soil, in: Earth Resources and Environmental Remote
502 Sensing/GIS Applications VII. p. 100051E. https://doi.org/10.1117/12.2239606

503 Dash, J., Curran, P.J., 2007. Evaluation of the MERIS terrestrial chlorophyll index (MTCI). Adv.
504 Sp. Res. 39, 100–104. https://doi.org/10.1016/j.asr.2006.02.034

505 Diepens, N.J., Buffan-Dubau, E., Budzinski, H., Kallerhoff, J., Merlina, G., Silvestre, J., Auby,
506 I., Nathalie Tapie, Elger, A., 2017. Toxicity effects of an environmental realistic herbicide
507 mixture on the seagrass Zostera noltei. Environ. Pollut. 222, 393–403.
508 https://doi.org/10.1016/j.envpol.2016.12.021

509 Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G.,
510 Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E.,
511 Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013. Collinearity:
512 a review of methods to deal with it and a simulation study evaluating their performance.
513 Ecography (Cop.). 36, 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

514 Durango-Cordero, J., Saqalli, M., Laplanche, C., Locquet, M., Elger, A., 2018. Spatial Analysis
515 of Accidental Oil Spills Using Heterogeneous Data: A Case Study from the North-Eastern
516 Ecuadorian Amazon. Sustainability 10, 4719. https://doi.org/10.3390/su10124719

517 Emengini, E.J., Blackburn, G.A., Theobald, J.C., 2013a. Early detection of oil-induced stress in
518 crops using spectral and thermal responses. J. Appl. Remote Sens. 7.
519 https://doi.org/10.1117/1.jrs.7.073596

520 Emengini, E.J., Blackburn, G.A., Theobald, J.C., 2013b. Detection and discrimination of oil and
521 water deficit-induced stress in maize (Zea mays L.) using spectral and thermal responses.
522 IOSR J. Environ. Sci. Toxicol. Food Technol. 3, 53–57.

523 Emengini, E.J., Ezeh, F.C., Chigbu, N., 2013c. Comparative Analysis of Spectral Responses of
524 Varied Plant Species to Oil Stress. Int. J. Sci. Eng. Res. 4, 1421–1427.

525 Feret, J.B., François, C., Asner, G.P., Gitelson, A.A., Martin, R.E., Bidel, L.P.R., Ustin, S.L., le
526 Maire, G., Jacquemoud, S., 2008. PROSPECT-4 and 5: Advances in the leaf optical
527 properties model separating photosynthetic pigments. Remote Sens. Environ. 112, 3030–
528 3043. https://doi.org/10.1016/j.rse.2008.02.012

529 Finer, M., Jenkins, C.N., Pimm, S.L., Keane, B., Ross, C., 2008. Oil and gas projects in the
530 Western Amazon: Threats to wilderness, biodiversity, and indigenous peoples. PLoS One 3.
531 https://doi.org/10.1371/journal.pone.0002932

532 Grömping, U., 2009. Variable importance assessment in regression: Linear regression versus
533 random forest. Am. Stat. 63, 308–319. https://doi.org/10.1198/tast.2009.08199

534 Gürtler, S., de Souza Filho, C.R., Sanches, I.D., Alves, M.N., Oliveira, W.J., 2018.
535 Determination of changes in leaf and canopy spectra of plants grown in soils contaminated
536 with petroleum hydrocarbons. ISPRS J. Photogramm. Remote Sens. 146, 272–288.
537 https://doi.org/10.1016/j.isprsjprs.2018.09.011

538 Haboudane, D., Miller, J.R., Tremblay, N., Zarco-Tejada, P.J., Dextraze, L., 2002. Integrated
539 narrow-band vegetation indices for prediction of crop chlorophyll content for application to
540 precision agriculture. Remote Sens. Environ. 81, 416–426. https://doi.org/10.1016/S0034-
541 4257(02)00018-4

542 Hastie, T., Tibshirani, R., Friedman, J., 2017. The Elements of Statistical Learning The Elements
543 of Statistical Learning.

544 Hu, G., Li, J., Zeng, G., 2013. Recent development in the treatment of oily sludge from
545 petroleum industry: A review. J. Hazard. Mater. 261, 470–490.
546 https://doi.org/10.1016/j.jhazmat.2013.07.069

547 Hutengs, C., Vohland, M., 2016. Downscaling land surface temperatures at regional scales with
548 random forest regression. Remote Sens. Environ. 178, 127–141.
549 https://doi.org/10.1016/j.rse.2016.03.006

550 Kirk, J.L., Klironomos, J.N., Lee, H., Trevors, J.T., 2002. Phytotoxicity assay to assess plant
551 species for phytoremediation of petroleum-contaminated soil. Bioremediat. J. 6, 57–63.
552 https://doi.org/10.1080/10889860290777477

553 Lassalle, G., Credoz, A., Fabre, S., Elger, A., Hédacq, R., Dubucq, D., 2017. Hyperspectral
554 signature analysis of three plant species to long-term hydrocarbon and heavy metal
555 exposure, in: Michel, U., Schulz, K. (Eds.), Earth Resources and Environmental Remote
556 Sensing/GIS Applications VIII. SPIE, p. 33. https://doi.org/10.1117/12.2277709

557 Lassalle, G., Credoz, A., Hédacq, R., Fabre, S., Dubucq, D., Elger, A., 2018. Assessing Soil
558 Contamination Due to Oil and Gas Production Using Vegetation Hyperspectral Reflectance.
559 Environ. Sci. Technol. 52, 1756–1764. https://doi.org/10.1021/acs.est.7b04618

560 Lassalle, G., Fabre, S., Credoz, A., Hédacq, R., Bertoni, G., Dubucq, D., Elger, A., 2019a.
561 Application of PROSPECT for estimating total petroleum hydrocarbons in contaminated
562 soils from leaf optical properties. J. Hazard. Mater. 377, 409–417.
563 https://doi.org/10.1016/j.jhazmat.2019.05.093

564 Lassalle, G., Fabre, S., Credoz, A., Hédacq, R., Borderies, P., Bertoni, G., Erudel, T., Buffan-
565 Dubau, E., Dubucq, D., Elger, A., 2019b. Detection and discrimination of various oil-
566 contaminated soils using vegetation reflectance. Sci. Total Environ. 655, 1113–1124.
567 https://doi.org/10.1016/j.scitotenv.2018.11.314

568 Li, Y., Morris, J.T., Yoch, D.C., 1990. Chronic Low Level Hydrocarbon Amendments Stimulate
569 Plant Growth and Microbial Activity in Salt-Marsh Microcosms. J. Appl. Ecol. 27, 159–
570 171. https://doi.org/10.2307/2403575

571 Lin, Q., Mendelssohn, I.A., Suidan, M.T., Lee, K., Venosa, A.D., 2002. The dose-response
572 relationship between No. 2 fuel oil and the growth of the salt marsh grass, Spartina
573 alterniflora. Mar. Pollut. Bull. 44, 897–902. https://doi.org/10.1016/S0025-326X(02)00118-
574 2

575 Maliszewska-Kordybach, B., Smreczak, B., 2000. Ecotoxicological Activity of Soils Polluted
576 with Polycyclic Aromatic Hydrocarbons (PAHs) - Effect on Plants. Environ. Technol. 21,
577 1099–1110. https://doi.org/10.1080/09593330.2000.9618996

578 Merkl, N., Schultze-Kraft, R., Infante, C., 2004. Phytoremediation in the tropics-the effect of
579 crude oil on the growth of tropical plants. Bioremediat. J. 8, 177–184.
580 https://doi.org/10.1080/10889860490887527

581 Milton, E.J., 1987. Principles of field spectroscopy. Int. J. Remote Sens. 8, 1807–1827.
582 https://doi.org/https://doi.org/10.1080/01431168708954818

583 Mutanga, O., Adam, E., Cho, M.A., 2012. High density biomass estimation for wetland
584 vegetation using worldview-2 imagery and random forest regression algorithm. Int. J. Appl.

585 Earth Obs. Geoinf. 18, 399–406. https://doi.org/10.1016/j.jag.2012.03.012

586 Nie, M., Lu, M., Yang, Q., Zhang, X., Xiao, M., Jiang, L., Yang, J., Fang, C., Chen, J., Li, B.,
587 2011. Plants ' use of different nitrogen forms in response to crude oil contamination 159,
588 157–163. https://doi.org/10.1016/j.envpol.2010.09.013

589 Noomen, M.F., van der Werff, H.M.A., van der Meer, F.D., 2012. Spectral and spatial indicators
590 of botanical changes caused by long-term hydrocarbon seepage. Ecol. Inform. 8, 55–64.
591 https://doi.org/10.1016/j.ecoinf.2012.01.001

592 Ogri, O.R., 2001. A review of the Nigerian petroleum industry and the associated environmental
593 problems. Environmentalist 21, 11–21.
594 https://doi.org/https://doi.org/10.1023/A:1010633903226

595 Romero, I.C., Toro-Farmer, G., Diercks, A.R., Schwing, P., Muller-Karger, F., Murawski, S.,
596 Hollander, D.J., 2017. Large-scale deposition of weathered oil in the Gulf of Mexico
597 following a deep-water oil spill. Environ. Pollut. 228, 179–189.
598 https://doi.org/10.1016/j.envpol.2017.05.019

599 Rosso, P.H., Pushnik, J.C., Lay, M., Ustin, S.L., 2005. Reflectance properties and physiological
600 responses of Salicornia virginica to heavy metal and petroleum contamination. Environ.
601 Pollut. 137, 241–252. https://doi.org/10.1016/j.envpol.2005.02.025

602 Salanitro, J.P., Dorn, P.B., Huesemann, M.H., Moore, K.O., Rhodes, I.A., Rice Jackson, L.M.,
603 Vipond, T.E., Western, M.M., Wisniewski, H.L., 1997. Crude oil hydrocarbon
604 bioromediation and soil ecotoxicity assessment. Environ. Sci. Technol. 31, 1769–1776.
605 https://doi.org/10.1021/es960793i

606 Sanches, I.D., de Souza Filho, C.R., Magalhães, L.A., Quitério, G.C.M., Alves, M.N., Oliveira,
607 W.J., 2013a. Assessing the impact of hydrocarbon leakages on vegetation using reflectance
608 spectroscopy. ISPRS J. Photogramm. Remote Sens. 78, 85–101.
609 https://doi.org/10.1016/j.isprsjprs.2013.01.007

610 Sanches, I.D., Souza Filho, C.R., Magalhães, L.A., Quitério, G.C.M., Alves, M.N., Oliveira,
611 W.J., 2013b. Unravelling remote sensing signatures of plants contaminated with gasoline
612 and diesel: An approach using the red edge spectral feature. Environ. Pollut. 174, 16–27.
613 https://doi.org/10.1016/j.envpol.2012.10.029

614 Savitzky, A., Golay, M.J.E., 1964. Smoothing and Differentiation of Data by Simplified Least
615 Squares Procedures. Anal. Chem. 36, 1627–1639. https://doi.org/10.1021/ac60214a047

616 Scafutto, R.D.M., de Souza Filho, C.R., 2016. Quantitative characterization of crude oils and
617 fuels in mineral substrates using reflectance spectroscopy: Implications for remote sensing.
618 Int. J. Appl. Earth Obs. Geoinf. 50, 221–242. https://doi.org/10.1016/j.jag.2016.03.017

619 Scafutto, R.D.P.M., de Souza Filho, C.R., de Oliveira, W.J., 2017. Hyperspectral remote sensing
620 detection of petroleum hydrocarbons in mixtures with mineral substrates: Implications for
621 onshore exploration and monitoring. ISPRS J. Photogramm. Remote Sens. 128, 146–157.
622 https://doi.org/10.1016/j.isprsjprs.2017.03.009

623 Shi, T., Chen, Y., Liu, Y., Wu, G., 2014a. Visible and near-infrared reflectance spectroscopy —
624 An alternative for monitoring soil contamination by heavy metals. J. Hazard. Mater. 265,
625 166–176. https://doi.org/10.1016/j.jhazmat.2013.11.059

626 Shi, T., Liu, H., Wang, J., Chen, Y., Fei, T., Wu, G., 2014b. Monitoring arsenic contamination in
627 agricultural soils with reflectance spectroscopy of rice plants. Environ. Sci. Technol. 48,
628 6264–6272. https://doi.org/10.1021/es405361n

629 Smith, K.L.L., Steven, M.D.D., Colls, J.J.J., 2004. Use of hyperspectral derivative ratios in the
630 red-edge region to identify plant stress responses to gas leaks. Remote Sens. Environ. 92,
631 207–217. https://doi.org/10.1016/j.rse.2004.06.002

632 Tran, T.H., Gati, E.M., Eshel, A., Winter, G., 2018. Germination , physiological and biochemical
633 responses of acacia seedlings (Acacia raddiana and Acacia tortilis) to petroleum
634 contaminated soils *. Environ. Pollut. 234, 642–655.
635 https://doi.org/10.1016/j.envpol.2017.11.067

636 Zhang, Y., Huang, J., Wang, F., Blackburn, G.A., Zhang, H.K., Wang, X., Wei, Chuanwen,
637 Zhang, K., Wei, Chen, 2017. An extended PROSPECT : Advance in the leaf optical
638 properties model separating total chlorophylls into chlorophyll a and b. Sci. Rep. 7.
639 https://doi.org/10.1038/s41598-017-06694-y

640 Zou, H., Hastie, T., 2005. Regression and variable selection via the elastic net. J. R. Stat. Soc.
641 Ser. B (Statistical Methodol. 67, 301–320. https://doi.org/10.1111/j.1467-
642 9868.2005.00503.x

643