

Soil TPH Concentration Estimation Using Vegetation Indices in an Oil Polluted Area of Eastern China

Linhai Zhu^{1,2}, Xuechun Zhao^{1,2}, Liming Lai¹, Jianjian Wang¹, Lianhe Jiang¹, Jinzhi Ding¹, Nanxi Liu¹, Yunjiang Yu³, Junsheng Li³, Nengwen Xiao³, Yuanrun Zheng^{1*}, Glyn M. Rimmington⁴

1 Key Laboratory of Resource Plants, Beijing Botanical Garden, West China Subalpine Botanical Garden, Institute of Botany, Chinese Academy of Sciences, Xiangshan, Beijing, China, **2** University of Chinese Academy of Sciences, Beijing, China, **3** Chinese Research Academy of Environmental Sciences, Beijing, China, **4** Global Learning College of Engineering, Wichita State University, Wichita, Kansas, United States of America

Abstract

Assessing oil pollution using traditional field-based methods over large areas is difficult and expensive. Remote sensing technologies with good spatial and temporal coverage might provide an alternative for monitoring oil pollution by recording the spectral signals of plants growing in polluted soils. Total petroleum hydrocarbon concentrations of soils and the hyperspectral canopy reflectance were measured in wetlands dominated by reeds (*Phragmites australis*) around oil wells that have been producing oil for approximately 10 years in the Yellow River Delta, eastern China to evaluate the potential of vegetation indices and red edge parameters to estimate soil oil pollution. The detrimental effect of oil pollution on reed communities was confirmed by the evidence that the aboveground biomass decreased from 1076.5 g m⁻² to 5.3 g m⁻² with increasing total petroleum hydrocarbon concentrations ranging from 9.45 mg kg⁻¹ to 652 mg kg⁻¹. The modified chlorophyll absorption ratio index (MCARI) best estimated soil TPH concentration among 20 vegetation indices. The linear model involving MCARI had the highest coefficient of determination ($R^2 = 0.73$) and accuracy of prediction ($RMSE = 104.2$ mg kg⁻¹). For other vegetation indices and red edge parameters, the R^2 and $RMSE$ values ranged from 0.64 to 0.71 and from 120.2 mg kg⁻¹ to 106.8 mg kg⁻¹ respectively. The traditional broadband normalized difference vegetation index (NDVI), one of the broadband multispectral vegetation indices (BMVIs), produced a prediction ($R^2 = 0.70$ and $RMSE = 110.1$ mg kg⁻¹) similar to that of MCARI. These results corroborated the potential of remote sensing for assessing soil oil pollution in large areas. Traditional BMVIs are still of great value in monitoring soil oil pollution when hyperspectral data are unavailable.

Citation: Zhu L, Zhao X, Lai L, Wang J, Jiang L, et al. (2013) Soil TPH Concentration Estimation Using Vegetation Indices in an Oil Polluted Area of Eastern China. PLoS ONE 8(1): e54028. doi:10.1371/journal.pone.0054028

Editor: Ben Bond-Lamberty, DOE Pacific Northwest National Laboratory, United States of America

Received: October 20, 2012; **Accepted:** December 7, 2012; **Published:** January 16, 2013

Copyright: © 2013 Zhu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the National Key Technology R&D Program of China (2008BAC43B01). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: zhengyr@ibcas.ac.cn

Introduction

Oil is a crucial energy resource and vital industrial raw material. Most oil is produced on land [1]. With increasing industrial production, oil pollution has become a serious worldwide environmental problem, especially at the oil mining stage in the field. It has many detrimental effects on the composition, structure and functioning of terrestrial ecosystems [2,3,4,5], including loss of biodiversity [6]. Oil pollutants also transfer via food chains and eventually harm human health [7]. In addition, residual oil hydrocarbons can persist in the soil for decades [3,8] and have a chronic effect on ecosystems and human beings. Therefore, it is crucial to monitor oil pollution accurately and quickly in areas where oil is produced.

Oil pollution at the mining stage is usually monitored by field sampling of soil, water, atmosphere and vegetation, and laboratory analysis of oil pollutants. However, these methods, especially when applied in large areas, are difficult, time-consuming and expensive. Remote sensing can cover large areas simultaneously and periodically in a non-destructive fashion, and offers a valid alternative to traditional ground-based methods [9,10,11]. Among remote sensing data, hyperspectral data consisting of hundreds of

contiguous spectral bands narrower than 10 nm throughout the visible-infrared spectrum could prove to be invaluable in developing environmental monitoring capabilities [12].

The reflectance of a leaf is governed by its pigment concentrations, morphological and anatomical properties, water content and other biochemical properties [13,14]. Canopy reflectance of plants is determined by the optical properties of leaves, and is modified by leaf area index (LAI), the amount of green biomass, canopy architecture and leaf angle distribution [15,16]. Alterations in these variables caused by oil pollution and other environmental stresses induce changes in canopy reflectance that can be detected by remote sensors and might be used as indicators for oil pollution of soils [10,17,18]. However, the spectral signal captured by sensors includes not only spectral characteristics of plants, but also optical properties of soils and air within the field of view [19,20]. Remote sensing by sensors is also influenced by viewing angle and illumination geometry [16]. Therefore, vegetation indices (VIs) [19,21] are proposed to enhance the spectral contribution from green plants and minimize those from soil, air and other external factors.

Vegetation indices combine two or more spectral bands to quantitatively characterize the physiological status of plants [22].

According to spectral resolution of data, vegetation indices can be classified into two classes: multispectral vegetation indices (MVIS) and hyperspectral vegetation indices (HVIs) [21]. Multispectral vegetation indices are first formulated with multispectral bands (such as those of Landsat ETM+ and NOAA AVHRR), but they can also be calculated from hyperspectral data using the reflectance at specific wavelengths [17,21]. Multispectral vegetation indices calculated in these two fashions can be defined as broadband multispectral vegetation indices (BMVIS) and narrow-band multispectral vegetation indices (NMVIS). The normalized difference vegetation index (NDVI) is one of the oldest and most widely used multispectral vegetation indices [23]. Hyperspectral vegetation indices are proposed based on the absorption and reflectance properties in specific regions of the high resolution spectrum [24,25].

Hyperspectral remote sensing not only provides new data for the development of vegetation indices, but also can detect some characteristics that traditional multispectral remote sensing cannot. Red edge is a unique and most notable feature of green plants, and cannot be measured using spectral bands broader than 100 nm. It is a sharp change in reflectance between low red reflectance caused by chlorophyll absorption near 680 nm and high infrared reflectance governed by internal leaf scattering near 750 nm [26]. Three parameters, red edge position (REP), slope (RES) and area (REA), can be calculated using the first derivative of reflectance. Red edge position is the maximum of the first derivative of reflectance between 680 nm and 750 nm. Red edge slope and red edge area are the maximum and integral of the first derivative in this region. In addition, the linear four-point interpolation technique [27], the Gaussian function fitting technique [28] and the polynomial fitting technique [21,29] have been proposed to parameterize the red edge.

Remote sensing has previously been used to detect stresses in plants [30,31,32]. Visible reflectance increased consistently in stressed leaves for eight stress agents and among six vascular plant species [30], and ratios of leaf reflectance were further evaluated as indicators of these stresses for these plants [31]. Remote sensing has been used to estimate metal pollutions in soils [17,33]. Rosso et al reported that lightweight petroleum induced significant changes in reflectance between 500 nm and 1550 nm [18]. Current and older foliage from spruce/fir forest affected by air pollution showed an approximately 5 nm shift away from the normal value of red edge position toward shorter wavelengths [34]. A hyperspectral derivative ratio in the red edge region has been used to identify plant stress due to gas leaks before visible symptoms were observed [32]. Although vegetation indices and red edge have been employed to estimate soil metal contamination [10,17,33], there are very few studies concerning the estimation of soil oil pollution by remote sensing. Therefore, it is necessary to evaluate current remote sensing methods and to develop a suitable method for oil pollution monitoring.

As the youngest land and one of the largest wetlands in China, the Yellow River Delta provides many ecological functions, such as habitats for numerous species [35]. However, this region is widely subjected to serious crude oil contamination due to oil mining. Shengli oilfield, as the second largest in China, lies mainly in this region. Since oil production began in 1964, intensive exploitation of oil for decades has produced detrimental effects on wetland ecosystems. Shengli oilfield discharges more than 100 000 tons of oily sludge into soils each year [36], and oil pollution has become the main anthropogenic source of ecological risks [37]. Therefore, estimating oil pollution accurately and quickly is urgently needed to enable control of its adverse effects and protect the remaining wetlands.

In this study, the feasibility of remote sensing to estimate soil oil pollution in the Yellow River Delta, Shandong Province, eastern China was investigated, and an effective remote sensing method was developed to monitor oil pollution accurately and quickly. This will be useful for establishing an early warning system and mitigating the harmful effects of oil pollution.

Materials and Methods

Ethics Statements

The field study was conducted with the permission of the Hekou Oil Production Plant, Shengli Oilfield Company, SINOPEC and the owners of the lands around the oil wells. No protected plant species were present in quadrats of the plant communities.

Site and Sample Design

This study was carried out in the Chengdong Oilfield belonging to the Shengli Oilfield Company (118°34'38.35" E-118°38'08.53" E, 37°57'34.07" N-38°01'28.50" N) (Figure 1). The sites are adjacent to the ancient Yellow River, the National Natural Reserve of the Yellow River Delta and the National Forest Park of the Yellow River Estuary. The region is sparsely populated and oil exploitation is the major production activity. Reeds dominate this region and several plant species grow under the reeds (Figure 2). The soil is a saline soil.

A field study was conducted in August 2009. Sample plots around each of five oil wells and three sample plots (bare soil) away from oil wells were established. Another bare soil plot was established for testing spectroradiometer. Soil color was uniform in all study sites. The mining history of the oil wells ranged from 7 to 13 years, and had an average of 10 years, with a standard deviation of 2.5 years. All wells produced heavy oils. Three sample lines were laid out in directions away from the wells. One line was set up along the prevailing wind direction, while the other two were established at a 120° angle from the first line to represent different oil pollution conditions in different directions. Along each sample line, three 1m×1m quadrats were established around each point at different distances from each oil well, 0 m, 5m, 10m, 20m, 30m, 50m, and 100m. This sampling procedure represents possible differences in oil pollution with different distances from the oil wells. At same distance points to oil well center in three lines of each oil well, e.g. at 5 m, data of three quadrats were averaged. Therefore, there are totally seven averaged data for one well.

Reflectance Measurement

The reflectance spectra of each quadrat and the bare soils were recorded using a field spectroradiometer ASD FieldSpec FR (Analytical Spectral Devices, Inc., Boulder, Colorado, USA). This spectroradiometer was fitted with a fiber optic probe having a 25° field of view. It covers the spectrum between 350 nm and 2500 nm and its spectral sampling interval is 1.4 nm for the region 350–1000 nm and 2 nm for the region 1000–2500 nm. Its spectral resolution (full-width-half-maximum) is 3 nm for the region 350–1000 nm and 10 nm for the region 1000–2500 nm. The spectra were then interpolated by the ASD software to produce readings at every 1 nm. Measurements were taken between 10:00 and 12:00 am on clear days without clouds and winds. Scans were taken from a height of 50 cm above the canopies of the plant communities looking towards the nadir position so that the field of view was a circular area having a diameter of about 22 cm. Ten replicates were measured for each quadrat. Before measurement of each quadrat, the radiance of a white standard panel coated with BaSO₄ and the dark current (Systematic noise from the instrument electronics and detectors) were recorded for reference and

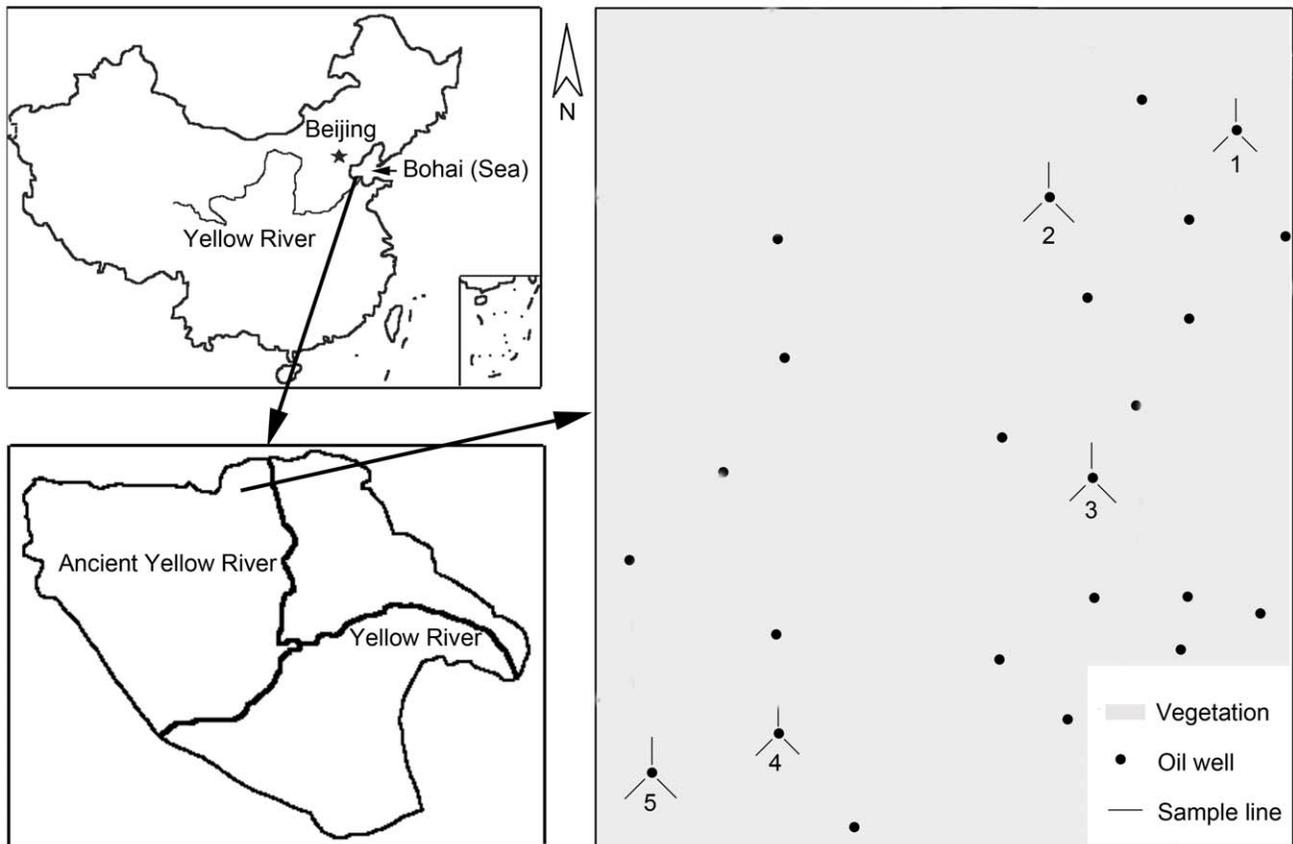


Figure 1. The location of the study site in eastern China. The spatial extent of the third panel is approximately 7 km by 8 km.
doi:10.1371/journal.pone.0054028.g001

optimization of measurement [18,32]. Because the energy incident on plant or soil is approximately equivalent to that reflected by the white panel, the reflectance (ρ , %) can be calculated as follows [38].

$$\rho = \frac{\text{target-darkcurrent}}{\text{reference-darkcurrent}} \times 100$$

where, 'target' is the energy reflected off plant canopies or bare soils, and 'reference' is that reflected off BaSO₄ white panel.

After the individual spectra of reflectance were inspected for bad data and outliers, they were averaged by quadrat [38]. The

averaged reflectance values were used to calculate vegetation indices. The reflectances of broadbands similar to Landsat ETM+ were calculated as the average reflectance of all wavelengths in the corresponding bands of spectroradiometer ASD FieldSpec FR.

Aboveground Biomass and Species of the Plant Communities

After measurement of canopy reflectance, the aboveground portion of plants in each quadrat was harvested. After the fresh aboveground material was weighed, it was dried at 80°C for 72 h and then reweighed. Species in every quadrat were recorded (Table 1).



Figure 2. Reed communities with different oil pollution.
doi:10.1371/journal.pone.0054028.g002

Table 1. Species in every quadrat in study site.

Oil wells	Quadrats	Species
1	1	<i>Phragmites australis</i>
	2	<i>P. australis</i> , <i>Tamarix austromongolica</i> (Rare)
	3	<i>P. australis</i> , <i>Suaeda. salsa</i>
	4	<i>P. australis</i> , <i>S. salsa</i> , <i>Cynanchum chinense</i>
	5	<i>P. australis</i> , <i>S. salsa</i> , <i>C. chinense</i>
	6	<i>P. australis</i> , <i>S. salsa</i> , <i>Aeluropus sinensis</i>
	7	<i>P. australis</i> , <i>S. salsa</i> , <i>T. austromongolica</i> (Rare), <i>C. chinense</i> (Very Rare), <i>Setaria viridis</i> (Very Rare)
2	1	<i>P. australis</i> , <i>A. sinensis</i> (Rare)
	2	<i>P. australis</i> , <i>Scorzonera mongolica</i> (Rare)
	3	<i>P. australis</i> , <i>S. salsa</i>
	4	<i>P. australis</i> , <i>S. salsa</i> , <i>Limonium bicolor</i> (Very Rare)
	5	<i>P. australis</i> , <i>S. salsa</i> , <i>C. chinense</i> (Rare)
	6	<i>P. australis</i> , <i>S. salsa</i> , <i>Apocynum venetum</i> (Rare)
	7	<i>P. australis</i> , <i>S. salsa</i> , <i>C. chinense</i> (Rare), <i>S. viridis</i> (Very Rare)
3	1	<i>P. australis</i>
	2	<i>P. australis</i> , <i>A. sinensis</i> (Rare)
	3	<i>P. australis</i> , <i>S. salsa</i> , <i>S. viridis</i> (Very Rare)
	4	<i>P. australis</i> , <i>S. salsa</i> , <i>C. chinense</i> (Very Rare)
	5	<i>P. australis</i> , <i>S. salsa</i>
	6	<i>P. australis</i> , <i>C. chinense</i> (Rare), <i>S. viridis</i> (Very Rare)
	7	<i>P. australis</i> , <i>S. salsa</i> , <i>C. chinense</i> (Rare), <i>S. viridis</i> (Very Rare)
4	1	<i>P. australis</i>
	2	<i>P. australis</i>
	3	<i>P. australis</i> , <i>A. sinensis</i> (Rare)
	4	<i>P. australis</i> , <i>S. viridis</i> (Rare)
	5	<i>P. australis</i> , <i>S. salsa</i> , <i>C. chinense</i> (Very Rare)
	6	<i>P. australis</i> , <i>S. salsa</i> , <i>C. chinense</i> (Rare)
	7	<i>P. australis</i> , <i>S. salsa</i> , <i>C. chinense</i> (Rare), <i>S. viridis</i> (Rare)
5	1	<i>P. australis</i>
	2	<i>P. australis</i>
	3	<i>P. australis</i> , <i>S. salsa</i> , <i>S. viridis</i> (Very Rare)
	4	<i>P. australis</i> , <i>S. salsa</i>
	5	<i>P. australis</i> , <i>S. salsa</i> , <i>S. viridis</i> (Rare)
	6	<i>P. australis</i> , <i>S. salsa</i> , <i>S. viridis</i> (Rare), <i>C. chinense</i> (Very Rare)
	7	<i>P. australis</i> , <i>S. salsa</i> , <i>S. viridis</i> (Rare), <i>C. chinense</i> (Very Rare)

doi:10.1371/journal.pone.0054028.t001

Soil Total Petroleum Hydrocarbon Measurement

Three samples from the top 30 cm of soils were gathered from each quadrat. Soil total petroleum hydrocarbon (TPH) concentrations in the samples were measured using the infrared spectrophotometry method [39,40]. In this method, dried (60°C/24 hour) soil samples (10 g) were extracted with 20 mL of carbon tetrachloride for 30 min by ultrasonication (40 kHz) at 40°C. The infrared absorption of the extract was measured at the band of 2930 cm⁻¹. The amount of TPH was calculated as the difference between the amount of total TPH in the polluted soil minus the biogenic TPH content in the control [39].

Multispectral Vegetation Indices (MVs)

a. Normalized Difference Vegetation Index (NDVI). This index has been proposed to eliminate seasonal sun angle

differences and to minimize the effect of atmospheric attenuation [41].

$$NDVI = \frac{(\rho_{NIR} - \rho_R)}{(\rho_{NIR} + \rho_R)}$$

Where *NIR* and *R* denote the near-infrared band and red band, respectively.

b. Transformed Soil Adjusted Vegetation Index (TSAVI). Many vegetation indices (such as TSAVI and SAVI2) rely on the existence of the soil line. TSAVI is a measure of the angle between the soil line and the line that joins the vegetation point and the intercept of the soil line [42].

identified empirically as 470 nm [24].

$$TSAVI = a \frac{\rho_{NIR} - a\rho_R - b}{a\rho_{NIR} + \rho_R - ab}$$

$$PSSR_c = \frac{\rho_{800}}{\rho_{470}}$$

$$PSND_c = \frac{\rho_{800} - \rho_{470}}{\rho_{800} + \rho_{470}}$$

Where a and b represent the slope and intercept of the soil line, respectively. For $a=1$ and $b=0$, TSAVI is equivalent to NDVI. TSAVI equals 0 for bare soil and is close to 1 for high LAI.

The soil line is a linear relationship between bare soil reflectance observed in two different wavebands, and bare soil reflectance lies on a single line in the space generated by the wavelength bands. The soil line usually is derived from bare soil reflectance in red and near-infrared wavebands [42]. The soil lines were obtained using broadband and narrowband data [42] of bare soils from 16 quadrats that we investigated in field.

The broadband soil line is:

$$\rho_{NIR} = 0.01581 + 1.1141\rho_R$$

The narrowband soil line is:

$$\rho_{800} = 0.01624 + 1.08661\rho_{680}$$

Where the numbers in the subscript represent the wavelengths used for calculation.

c. The second version of Soil Adjusted Vegetation Index (SAVI2). SAVI2 uses the ratio of the intercept of the soil line b to its slope a as the soil adjustment factor [43].

$$SAVI2 = \frac{\rho_{NIR}}{\rho_R + b/a}$$

d. Atmospherically Resistant Vegetation Index (ARVI). ARVI uses the difference in the radiance between the blue and the red bands to eliminate the atmospheric effect on the red band [20]. This correction enhances the resistance of ARVI to atmospheric effects (in comparison to NDVI).

$$ARVI = \frac{\rho_{NIR} - [\rho_R - \gamma(\rho_B - \rho_R)]}{\rho_{NIR} + [\rho_R - \gamma(\rho_B - \rho_R)]} = \frac{\rho_{NIR} - [\rho_R - (\rho_B - \rho_R)]}{\rho_{NIR} + [\rho_R - (\rho_B - \rho_R)]}$$

Where γ equals 1, this value is suitable for vegetated areas with small to moderate aerosol particle size and for arid regions with large particle size, and the subscript B denotes the blue band.

e. Optimization of Soil-Adjusted Vegetation Index (OSAVI). OSAVI uses a different soil adjustment factor ($L=0.16$) [44].

$$OSAVI = (1 + 0.16) \frac{(\rho_{NIR} - \rho_R)}{(\rho_{NIR} + \rho_R + 0.16)}$$

Multispectral vegetation indices were calculated with broadband and narrowband data in this study. Broadbands were blue (450–515 nm), red (630–690 nm) and near-infrared (750–900 nm), and narrowband were blue (470 nm), red (680 nm) and near-infrared (800 nm) (Figure 3).

Hyperspectral Vegetation Indices (HVIs)

a. Pigment Specific Simple Ratio (PSSR) and Pigment Specific Normalized Difference (PSND). The optimal individual wavelengths for the estimation of carotenoids are

These indices provide a measure of the depth of the pigment absorption relative to the highly reflective near-infrared plateau. The near-infrared wavelength of 800 nm can be considered to minimize the effects of radiation interactions at the leaf surface and internal structures in mesophyll.

b. Modified Chlorophyll Absorption Ratio Index (MCARI). MCARI is a modification of the Chlorophyll Absorption Ratio Index, which can minimize the effects of non-photosynthetic materials on the estimation of absorbed photosynthetically active radiation. The MCARI is the depth of the chlorophyll absorption at 670 nm relative to the reflectance at 550 nm and 700 nm [25].

$$MCARI = \frac{\rho_{700}}{\rho_{670}} \times [(\rho_{700} - \rho_{670}) - 0.2 \times (\rho_{700} - \rho_{550})]$$

The ratio (ρ_{700}/ρ_{670}) can minimize the combined effects of the underlying soil reflectance and the canopy non-photosynthetic materials [45].

c. Transformed Chlorophyll Absorption in Reflectance Index (TCARI). The formula of TCARI is similar to that of MCARI with a different position of the ratio (ρ_{700}/ρ_{670}) [45]. TCARI uses this ratio to counteract the background influence only on the difference ($\rho_{700} - \rho_{550}$).

$$TCARI = 3 \times [(\rho_{700} - \rho_{670}) - 0.2 \times (\rho_{700} - \rho_{550}) \times \frac{\rho_{700}}{\rho_{670}}]$$

d. Maximum first derivative spectrum (deRES). deRES was calculated as the maximum of the first derivative of the reflectance between 680 nm and 750 nm [26,46].

e. Fifth-order polynomial fitting technique (fpnRES). A fifth-order polynomial function is fitted to the reflectance spectrum between 680 and 750 nm [29].

$$\rho = \alpha + \beta_1 \lambda + \beta_2 \lambda^2 + \beta_3 \lambda^3 + \beta_4 \lambda^4 + \beta_5 \lambda^5$$

Where λ is the wavelength, α , β are parameters obtained by fitting. fpnRES was determined as the maximum of the first derivative of the fitted reflectance.

f. c2RES. Because there are two or more peaks in the first derivative reflectance of the red edge region, the peaks near 720 nm are identified as the main peak or dominant peak [21,26]. c2RES was calculated as the maximum of the first derivative near 720 nm.

g. Sum of the first derivative (sumREA). The sumREA was calculated as the sum of the first derivative of reflectances from 680 nm to 750 nm using data with a 1 nm interval from ASD software.

h. Gaussian function fitting technique (GauREA). The first derivative of reflectance was fitted with the Gaussian function.

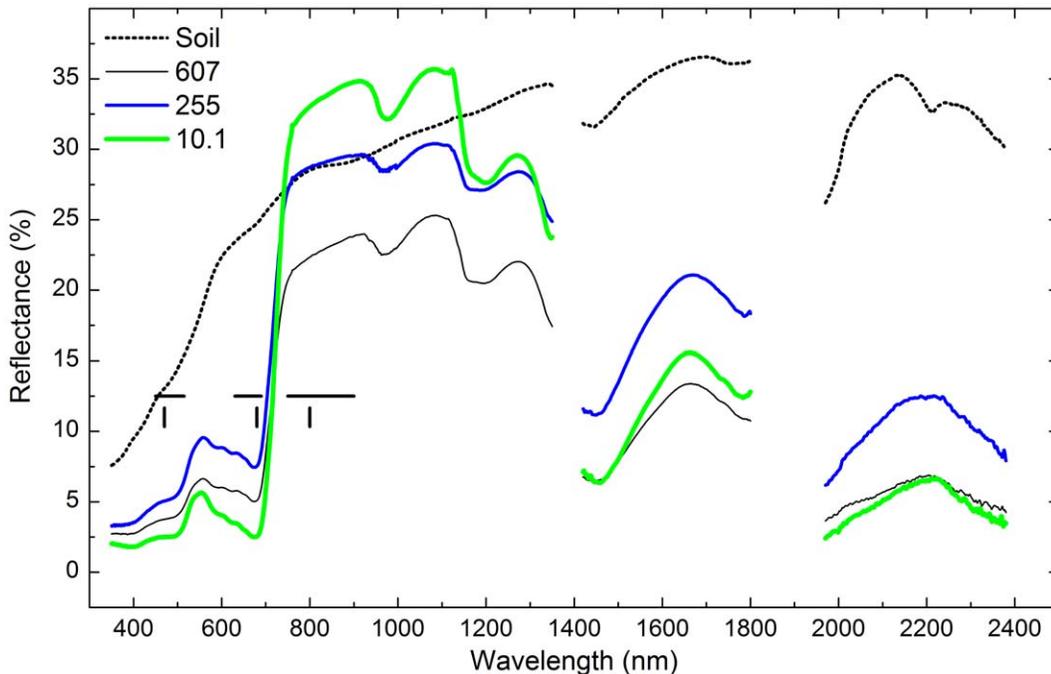


Figure 3. Reflectance of bare soil and reed communities with different soil TPH concentrations. Horizontal lines denoted broad wavebands used to calculate BMVIs. From left to right, they were blue (450–515 nm), red (630–690 nm) and near-infrared (750–900 nm). Vertical lines indicated wavelengths used to calculate NMVIs. From left to right, they were blue (470 nm), red (680 nm) and near-infrared (800 nm). The reflectance of bare soil was the average of 16 quadrats of bare soil. Each reflectance curve of vegetation was the average of 3 quadrats of reed communities at the same distance to the oil well in the same plot. The numbers in the legend were soil TPH concentrations (mg kg⁻¹). doi:10.1371/journal.pone.0054028.g003

$$y = y_0 + A \exp\left(-\frac{(\lambda - \lambda_c)^2}{2w^2}\right)$$

Where y is the first derivative of reflectance, and y_0 , A , λ_c and w are four parameters that determine the shape of the curve of the first derivative and are calculated by iteration.

Then the fitted curve was integrated as GauREA. This method is a modification of the inverted-Gaussian reflectance model that was used to fit the reflectance in the spectral region of 670–800 nm [28].

i. Sixth-order polynomial fitting technique (spnREA). A sixth-order polynomial function is fitted to the reflectance spectrum between 680 and 750 nm [21].

$$\rho = \alpha + \beta_1\lambda + \beta_2\lambda^2 + \beta_3\lambda^3 + \beta_4\lambda^4 + \beta_5\lambda^5 + \beta_6\lambda^6$$

spnREA was calculated as the integral of the first derivative of the fitted reflectance.

The calculation of RES and REA were accomplished using OriginPro 8.5.1 (OriginLab Corporation, Northampton, MA, USA).

Regression Analysis and Validation

Soil TPH concentrations, aboveground biomass, vegetation indices at the same distance from the oil wells in the same plot were averaged. The curve estimation module in IBM SPSS Statistics 19 was used to fit the relationship of aboveground biomass to soil TPH concentrations, and that of the reflectance at specific wavelengths, vegetation indices to soil TPH concentrations. Among 35 data of five wells, 30 data were used to fit above

regression equations, other five data were used to validate the regression equations. The best equations were chosen from the linear, logarithmic, S, exponential, inverse and power models. The equations with a high R^2 , small root mean square error (RMSE) and good validation were regarded as good (Table 2).

To evaluate the regression equations performance, a measure used by Qiu *et al.* (1998) [47] was adopted to examine the correlation between the estimated and observed data. Employing this method, the observed and estimated data were used to derive a linear regression equation ($y = a+bx$), which was then compared with the 1:1 observed data line ($y = x$). Residual plots were used to evaluate the regression equations' performance.

Results

Effect of Soil Oil Pollution on Aboveground Biomass of the Reed Community

Soil TPH concentrations ranged from 9.45 mg kg⁻¹ to 652 mg kg⁻¹ with an average of 190 mg kg⁻¹ and a standard deviation of 200 mg kg⁻¹. Residual hydrocarbons were significantly and negatively correlated with the aboveground biomass of the reed community, which decreased from 1076.5 g m⁻² to 5.3 g m⁻² (Figure 4).

Effect of Soil Oil Pollution on Reflectance of the Reed Community

Reflectance of reed community under three concentrations of soil TPH (Figure 3), which representing low, intermediate and high levels of oil pollution, shows that oil pollution decreased reflectance in the near-infrared between 800 nm and 1300 nm. However, reflectance in the other spectral bands did not follow the variations in TPH concentrations.

Table 2. Regression equations of different spectral indices and soil TPH concentrations (y , mg kg^{-1}) in the reed communities. $p < 0.001$.

Category	Spectral indices (x)	Regression equations	R^2	Root mean square error (RMSE, mg kg^{-1})
BMVIs	Broadband TSAVI	$y = 540.3 - 677.7x$	0.70	109.3
	Broadband SAVI2	$y = 527.0 - 283.9\ln x$	0.67	114.3
	Broadband ARVI	$y = 461.5 - 591.0x$	0.68	112.3
	Broadband OSAVI	$y = 597.8 - 928.7x$	0.68	112.7
NMVIs	Narrowband SAVI2	$y = 519.9 - 271.3\ln x$	0.69	112.0
	Narrowband ARVI	$y = 453.7 - 557.4x$	0.70	109.4
	Narrowband OSAVI	$y = 589.7 - 899.4x$	0.69	110.6
HVIIs	PSSRc	$y = 776.2 - 327.1\ln x$	0.70	108.7
	PSNDc	$y = 1017 - 1205x$	0.71	107.5
	TCARI	$y = 587.3 - 5136x$	0.70	110.1
RES	f _p nRES	$y = -1207 - 242.2\ln x$	0.64	120.2
	c2RES	$y = -1352 - 269.8\ln x$	0.67	112.9
REA	GauREA	$y = -311.8 - 264.2\ln x$	0.64	119.1
	spnREA	$y = -316.6 - 266.6\ln x$	0.65	118.7

doi:10.1371/journal.pone.0054028.t002

There were significant differences between reflectance of the reed community and that of bare soils (Figure 3). In the visible band and near-infrared band of 1400–2400 nm, the reflectance curve of the reed community was lower than that of bare soils. For the near-infrared band between 800 nm and 1300 nm, the difference in reflectance between the reed community and bare soils was smaller.

Effect of Soil Oil Pollution on the First Derivative of Reflectance of the Reed Community

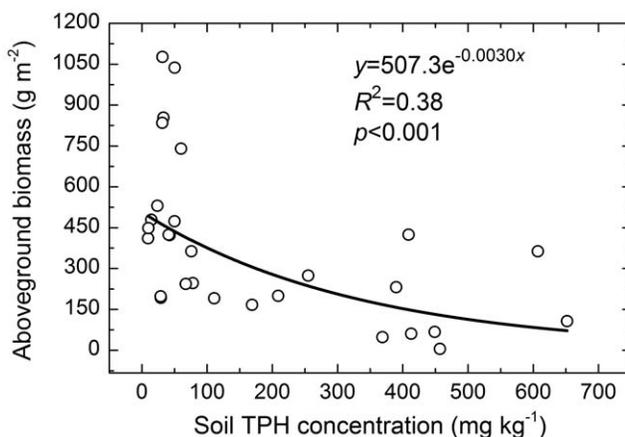
The first derivative of reflectance in the region of red edge distinguished bare soil and reed communities with different soil TPH concentrations (Figure 5). The first derivative of reflectance of bare soil was approximately a constant and close to zero, whereas that of the reed community was larger and had three or four asymmetric peaks. The first derivative of reflectance from

three soil TPH concentrations shows that red edge slope and red edge area decreased with the increase in soil TPH concentrations.

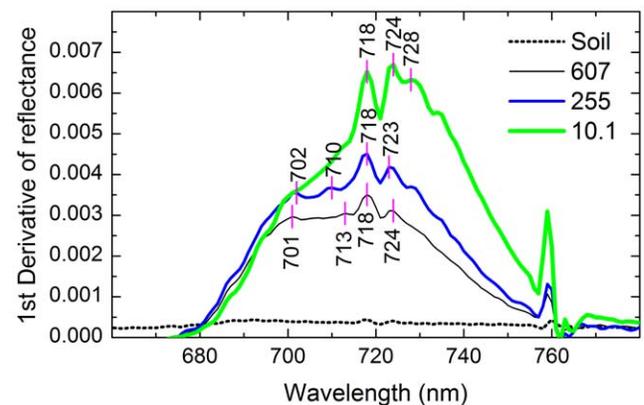
Performance of Spectral Indices for Estimating Soil Oil Pollution

Reflectances at specific wavelengths performed poorly for indicating soil oil pollution except for that at 680 nm, and reflectances at visible wavelengths performed better than that at near-infrared wavelengths (Figure 6).

Compared with reflectances at single wavelengths, vegetation indices estimated soil TPH concentrations better, and most coefficients of determination of them exceeded 0.65 (Figure 7, Table 2). Compared with vegetations indices induced from red edge, other vegetation indices achieved comparable results and performed better. The results from red edge slope and red edge area were also comparable.

**Figure 4.** Relationship between soil TPH concentrations and aboveground biomass of the reed communities.

doi:10.1371/journal.pone.0054028.g004

**Figure 5.** The first derivative of reflectance of bare soil and reed communities with different soil TPH concentrations. Numbers in the legend were soil TPH concentrations (mg kg^{-1}). Vertical lines and numbers indicated the wavelengths of peaks (nm). doi:10.1371/journal.pone.0054028.g005

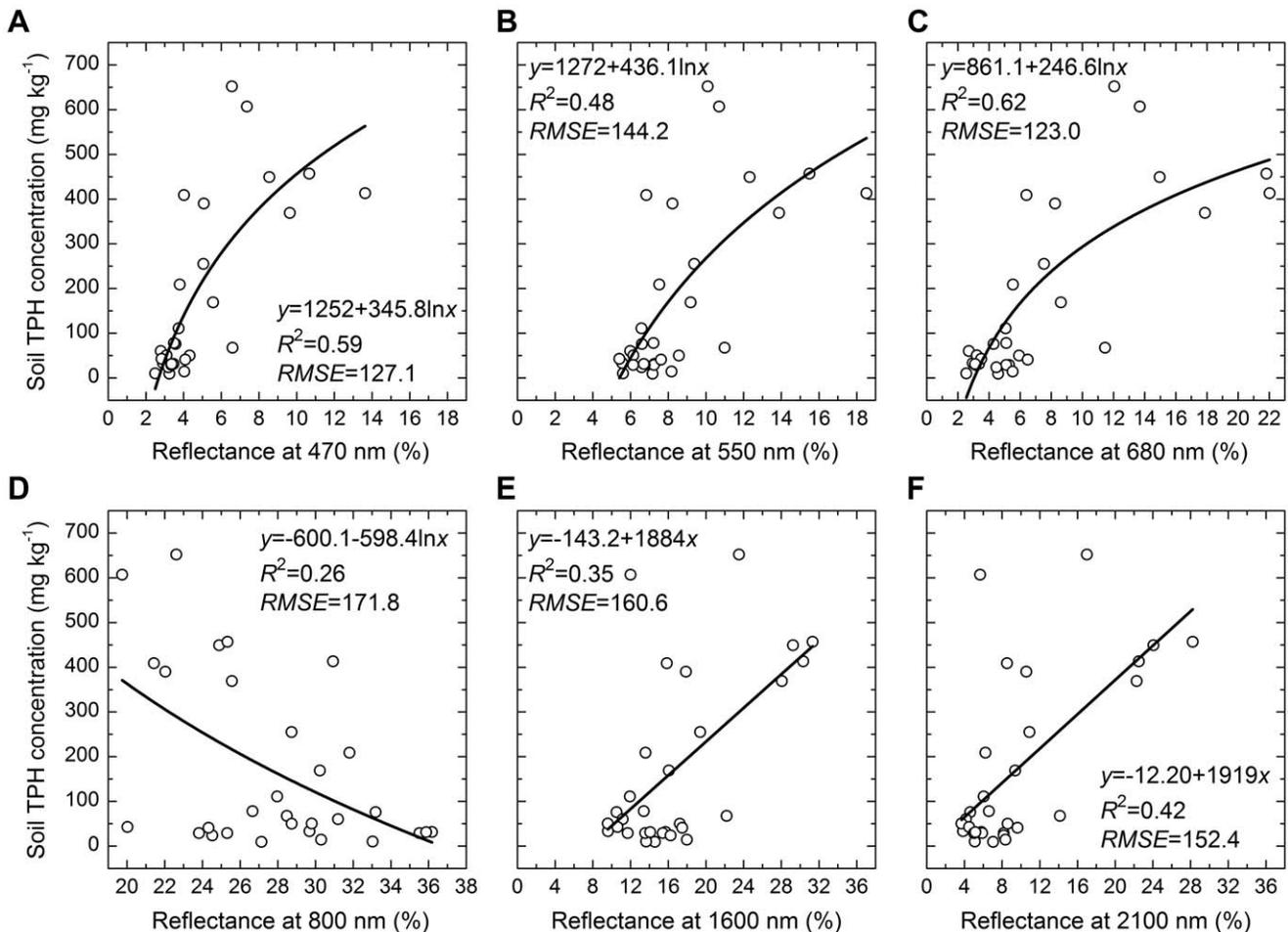


Figure 6. Performance of reflectance at specific wavelengths for estimating soil TPH concentrations. RMSE was the root mean square error. $p < 0.001$.

doi:10.1371/journal.pone.0054028.g006

MCARI was the best among the 20 vegetation indices (Figure 7, 8, 9, Table 2). Among multispectral vegetation indices, the best estimations were given by NDVI and TSAVI, followed by narrowband ARVI. Narrowband NDVI and TSAVI were slightly better than their broadband versions. For hyperspectral vegetation indices, MCARI yielded the highest accuracy of prediction. PSNDc, PSSRc and TCARI could also estimate oil pollution well, and their coefficients of determination were larger than or equivalent to 0.70. Among red edge slope parameters calculated using different methods, deRES gave the best estimation of oil pollution. Red edge area values calculated using different methods achieved similar results (Figure 7, 8, 9, Table 2). It was clear from Figure 8 and 9 that the validation was in agreement with above results and residual error was acceptable.

Discussion

Persistence Times of Oil Pollutants in Soil

Hydrocarbons may persist in soils for 39 years and aboveground biomass was continuously decreased in hydrocarbon-contaminated areas [3]. Although it is recognized that reed dominated wetlands have potential for mitigating hydrocarbon pollution, reed biomass significantly decreased in the early stages of serious oil pollution [40], while very light short-term oil pollution led to an increase in reed biomass because of the added increment of carbon

[48]. Further, ecological effects caused by long-term oil pollution were different from those resulting from short-term pollution [2]. In this study, there were residues of petroleum hydrocarbons around oil wells with a history of approximately 10 years, and aboveground biomass of the reed community decreased with increasing oil pollution (Figure 4). This result indicates that oil pollution resulted in significant change for the plant communities.

Effects of Oil Pollution on Spectral Characteristics

Changes in the physiological status of plants caused by pollution and other environmental stresses can alter their spectral characteristics [33]. Also, reflectance at different wavelengths might show different sensitivities to these stresses [30]. Visible reflectance increased consistently in leaves subjected to eight stresses including pollution for six vascular plant species, while infrared reflectance was comparatively unresponsive to these stresses [30]. In this study, oil pollution altered the spectral characteristics of the reed community. Oil pollution increased the reflectance of light in the visible and near-infrared bands greater than that of 1400 nm, and decreased that in the near-infrared between 800 and 1300 nm (Figure 6). These effects were more significant for the reflectance of visible light than for that in the near-infrared (Figure 6). Thus, visible reflectance rather than that

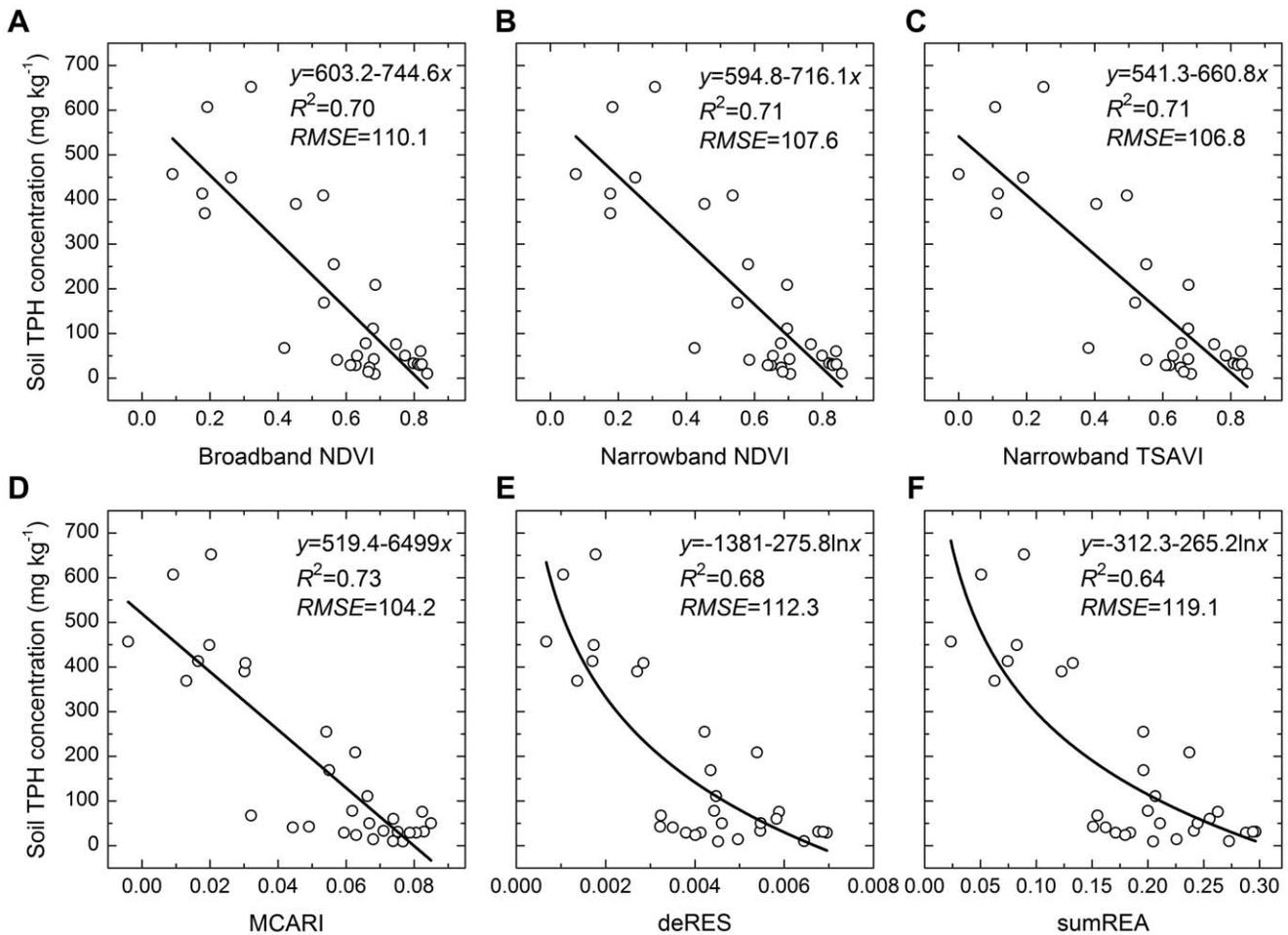


Figure 7. Performance of the leading vegetation indices for estimating soil TPH concentrations. NDVI, TSAVI, MCARI, deRES and sumREA denoted Normalized Difference Vegetation Index, Transformed Soil Adjusted Vegetation Index, Modified Chlorophyll Absorption Ratio Index, red edge slope calculated using maximum first derivative spectrum and red edge area calculated using the sum of the first derivative, respectively. RMSE was the root mean square error. $p < 0.001$. doi:10.1371/journal.pone.0054028.g007

in the near-infrared was the more sensitive and reliable indicator of stresses.

Generally, changes of reflectance of plant canopy are mainly governed by leaf pigment concentrations, morphological and anatomical properties, water content and other biochemical properties [13,14], and also can be modified by leaf area index, the amount of green biomass, canopy architecture and leaf angle distribution [15,16]. Although most of these parameters were not measured in this study, aboveground biomass decrement (Figure 4) and reed communities coverage decrement (Figure 2) resulted from oil pollution might imply changes of leaf properties such as leaf pigment concentrations, water content and leaf angle distribution, and therefore leaf reflectance were changed.

Performance of Various Indices for Indicating Oil Pollution

Effects of soil on canopy reflectance caused a lower accuracy of prediction for estimating soil oil pollution using reflectance of single wavelengths compared with those using vegetation indices [22]. Our results also agree well with this conclusion (Figure 6, 7, Table 2). Therefore, it is necessary to mitigate the effect of soil on reflectance.

Most of the selected vegetation indices utilize red reflectance dominated by chlorophyll absorption and near-infrared reflectance governed by leaf scattering. These combinations effectively enhance the vegetation signal and suppress soil background effects [21]. Therefore, these indices more accurately characterize the spectral changes in canopy reflectance of vegetation under oil pollution. The MCARI and TCARI indices introduced a green band into their formulae and quantified the variations in the triangular area comprising the green peak, red reflectance minimum and near-infrared reflectance shoulder [21,25,49]. Therefore, these two indices also could estimate oil pollution well, and MCARI even produced the most accurate prediction of soil TPH concentrations in this study (Figure 7, 8, 9).

Multispectral vegetation indices and hyperspectral vegetation indices have also been used to estimate soil Zn concentrations. Compared to BMVIs, their narrowband variants showed a small improvement [17]. Similar results were found for the estimation of LAI and canopy chlorophyll density, with BMVIs yielding better predictions than their narrowband variants and hyperspectral vegetation indices [21]. In this study, all categories of vegetation indices achieved comparable accuracy of prediction for soil TPH concentrations (Figure 7, 8, 9, Table 2). These results indicate that

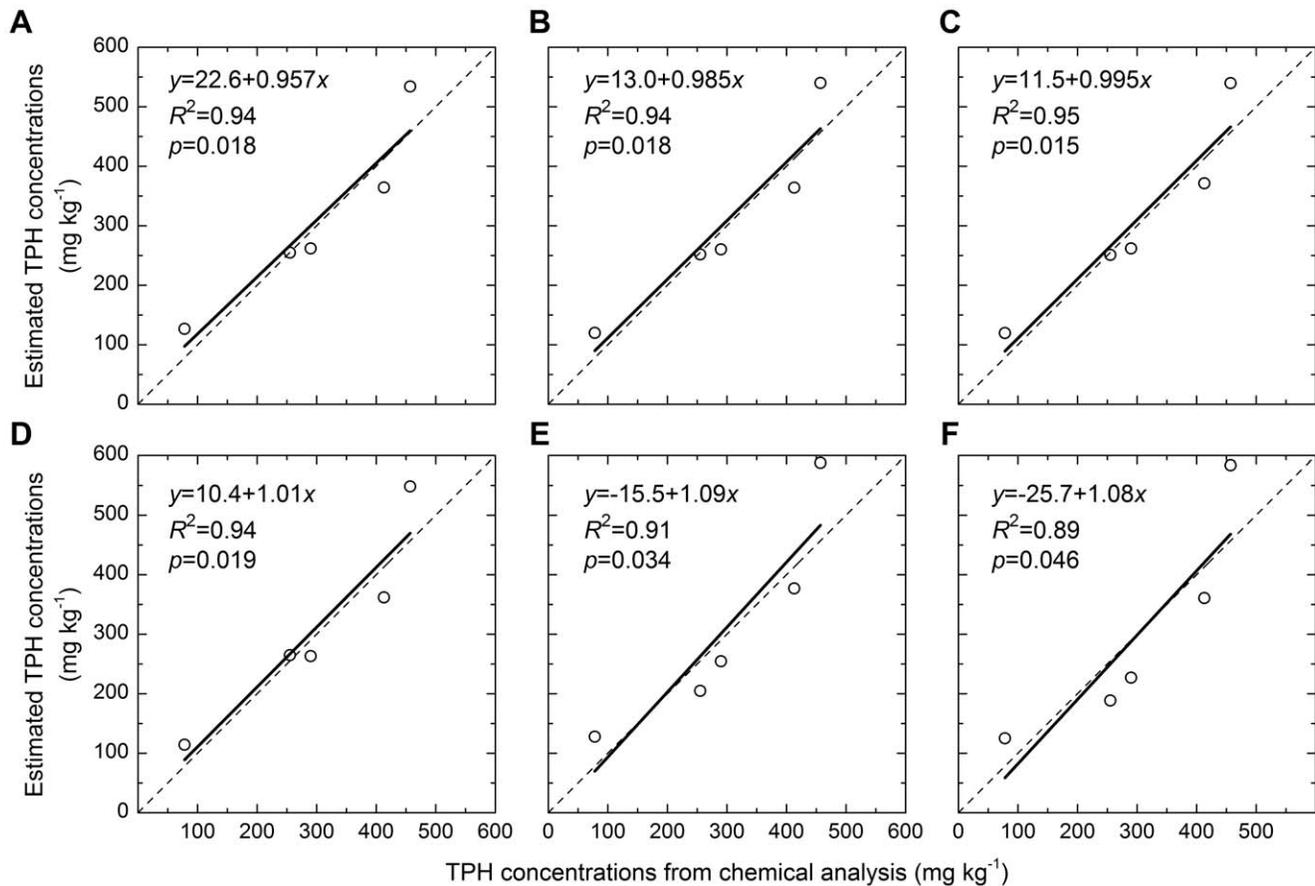


Figure 8. Comparison of observed and simulated TPH concentrations (mg kg^{-1}). The dashed line showed the 1:1 relationship, the solid line, the fitted regression equations. A, B, C, D, E, F denoted the validation for the regression equations derived from broadband NDVI, narrowband NDVI, narrowband TSAVI, MCARI, deRES and sumREA, and abbreviations were the same as Figure 7. doi:10.1371/journal.pone.0054028.g008

NMVI and hyperspectral vegetation indices had no absolute advantage over BMVIs.

Red edge parameters have the same theoretical foundation as the other vegetation indices consisting of red reflectance and near-infrared reflectance, because the red edge effect results from chlorophyll absorption of red light and leaf scattering of near-infrared light [21]. Because the first derivative of reflectance of bare soil was approximately a constant and close to zero, the derivative calculation of red edge parameters has the advantages of enhancing vegetation signals and suppressing the effects of bare soils [50]. Therefore, red edge slope and red edge area could estimate TPH concentrations with a higher accuracy than did the reflectance of single wavelengths (Figure 6, 7, 8, 9, Table 2) [17].

Compared to other vegetation indices, red edge parameters had a poorer performance in this study (Figure 7, 8, 9, Table 2). Firstly, this was dependent on their capacity to parameterize red edge. Polynomial and Gaussian function fitting have usually been used to calculate red edge slope and red edge area values. These methods often have the effect of smoothing out small spectral features that may contain important pollution-related information [32,33]. In addition, because the first derivative curve in the region of red edge is asymmetric (Figure 5) and the Gaussian function is symmetric, Gaussian function fitting will average out such asymmetry [21]. Also, the existence of several peaks makes it more difficult to parameterize the red edge. For example, several plants from the greenhouse and the field also showed two peaks at

around 700 and 725 nm [26]; A minor peak at 700 nm and a double peak around 720 nm were identified for annual bluegrass and perennial ryegrass [33]; One grass and two crops had four peaks at 702 nm, 718 nm, 725 nm and 735 nm in the red edge region [32]. Similarly, three or four peaks were present in the region of 700–730 nm in this study (Figure 5). However, methods of fitting assume that there is only a single peak in the red edge region. This might be another reason for the poor predictive capability of red edge in this study.

Therefore, traditional BMVIs such as NDVI still have a great potential for application in environmental monitoring because of their low cost, and more available data resources, such as Landsat TM, SPOT etc. compared with hyperspectral data [21], although Broadband NDVI has a lower R^2 (0.70) than that of MCARI (0.73). NMVIs and hyperspectral vegetation indices could be used to guide the design of field portable radiometer for vegetation monitoring [51], and only reflectance values at two or three wavelengths are required to quantify the growth status of vegetation and to estimate soil oil pollution.

Conclusions

Residual petroleum hydrocarbon resulting from oil exploitation of approximately 10 years' duration was associated with lowered aboveground biomass of plant communities dominated by reeds and altered their spectral characteristics. Spectral changes of the reed community under oil pollution provide an essential pre-

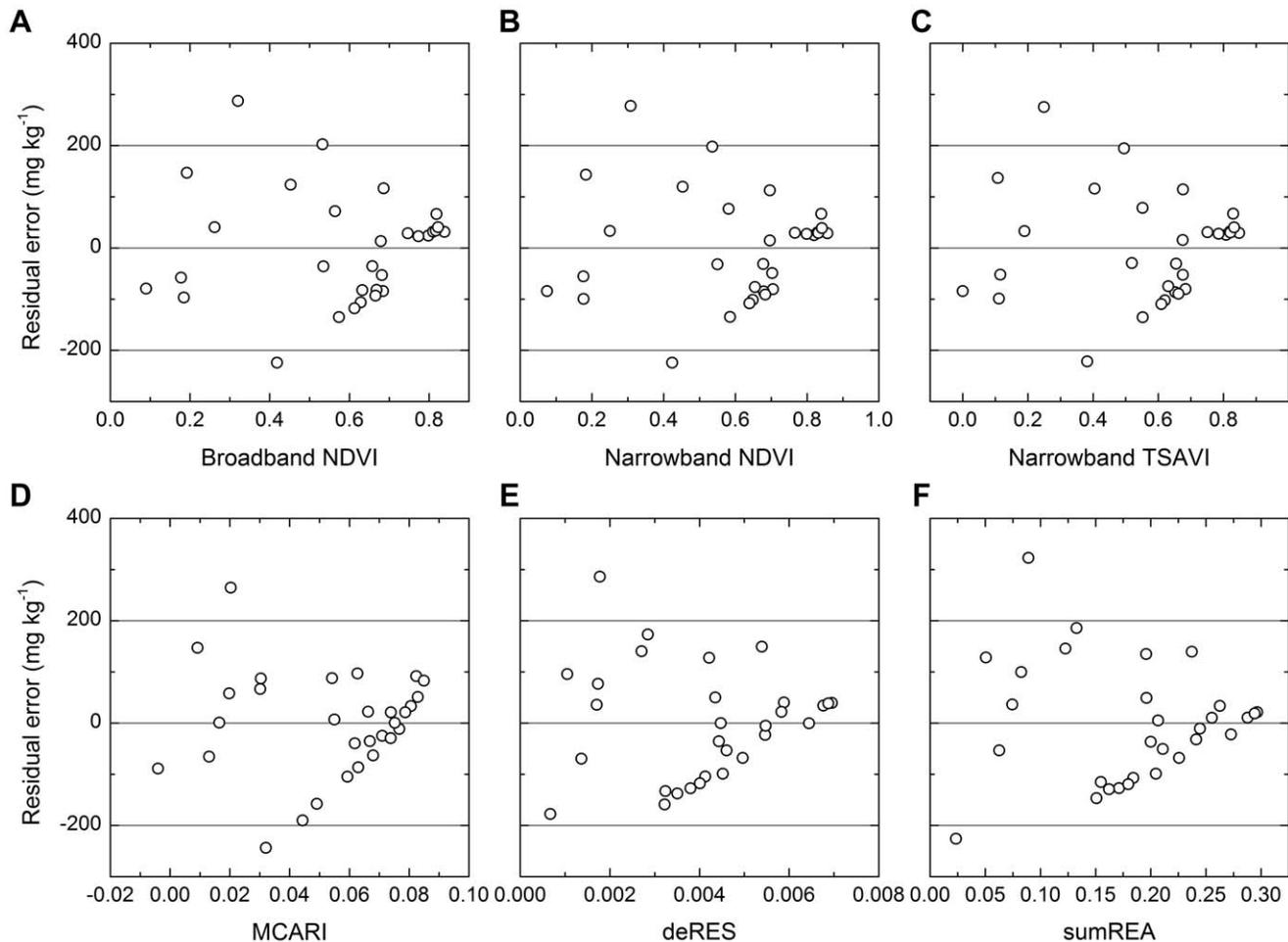


Figure 9. Residual plots for comparing prediction of TPH concentrations by different equations. Abbreviations were the same as Figure 7.

doi:10.1371/journal.pone.0054028.g009

requisite for the estimation of soil oil pollution. Compared to reflectance at specific wavelengths, vegetation indices better characterized these spectral alternations. Therefore, they could be used to estimate soil TPH concentration with a higher accuracy. Among 20 vegetation indices, MCARI produced the highest accuracy of prediction and traditional broadband NDVI, one of the BMVIs, yielded a prediction similar to that of MCARI.

These results confirm that remote sensing has great potential for estimating oil pollution in soils over large areas under appropriate conditions. Traditional BMVIs still have great value for monitoring soil oil pollution when hyperspectral data are unavailable.

References

- Ren L, Huang TL (2000) Contamination of soils by petroleum. *Agro-environ Prot* 19: 360–363.
- Pezeshki SR, Hester MW, Lin Q, Nyman JA (2000) The effects of oil spill and clean-up on dominant US Gulf Coast marsh macrophytes: a review. *Environ Pollut* 108: 129–139.
- Culbertson JB, Valiela I, Pickart M, Peacock EE, Reddy CM (2008) Long-term consequences of residual petroleum on salt marsh grass. *J Appl Ecol* 45: 1284–1292.
- Nie M, Wang Y, Yu J, Xiao M, Jiang L (2011) Understanding plant-microbe interactions for phytoremediation of petroleum-polluted soil. *PLoS ONE* 6: e17961. doi:10.1371/journal.pone.0017961.
- Fodrie FJ, Heck KL (2011) Response of coastal fishes to the Gulf of Mexico oil disaster. *PLoS ONE* 6: e21609. doi:10.1371/journal.pone.0021609.
- Osuji LC, Adesiyun SO, Obute GC (2004) Post-impact assessment of oil pollution in Agbada West Plain of Niger Delta, Nigeria: field reconnaissance and total extractable hydrocarbon content. *Chem Biodivers* 1: 1569–1578.
- Shields PG, Bowman ED, Harrington AM, Doan VT, Weston A (1993) Polycyclic aromatic hydrocarbon-DNA adducts in human lung and cancer susceptibility genes. *Cancer Res* 53: 3486–3492.
- Peterson CH, Rice SD, Short JW, Esler D, Bodkin JL (2003) Long-term ecosystem response to the Exxon Valdez oil spill. *Science* 302: 2082–2086.
- Austin JM, Mackey BG, Van Niel KP (2003) Estimating forest biomass using satellite radar: an exploratory study in a temperate Australian *Eucalyptus* forest. *Forest Ecol Manag* 176: 575–583.
- Kooistra L, Salas EAL, Clevers JGPW, Wehrens R, Leuven RSEW (2004) Exploring field vegetation reflectance as an indicator of soil contamination in river floodplains. *Environ Pollut* 127: 281–290.

Acknowledgments

We thank Zhongxin Chen and Zongnan Li for their help with the fieldwork and laboratory analysis. We also greatly appreciate the academic editor Prof. Bond-Lamberty and three anonymous reviewers for their insightful comments.

Author Contributions

Conceived and designed the experiments: LZ YZ LJ JD JL NX YY. Performed the experiments: LZ XZ LL JW JD NL. Analyzed the data: LZ YZ XZ LL JW. Wrote the paper: LZ YZ LJ GR XZ LL JW.

11. Mason DC, Davenport IJ, Neal JC, Schumann GJP, Bates PD (2012) Near real-time flood detection in urban and rural areas using high-resolution synthetic aperture radar images. *IEEE Trans Geosci Remote Sens* 50: 3041–3052.
12. Goetz AFH, Vane G, Solomon JE, Rock BN (1985) Imaging spectrometry for earth remote sensing. *Science* 228: 1147–1153.
13. Gates DM, Keegan HJ, Schleter JC, Weidner VR (1965) Spectral properties of plants. *Appl Opt* 4: 11–20.
14. Woolley JT (1971) Reflectance and transmittance of light by leaves. *Plant Physiol* 47: 656–662.
15. Knipling EB (1970) Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. *Remote Sens Environ* 1: 155–159.
16. Colwell JE (1974) Vegetation canopy reflectance. *Remote Sens Environ* 3: 175–183.
17. Kooistra L, Leuven RSEW, Wehrens R, Nienhuis PH, Buydens LMC (2003) A comparison of methods to relate grass reflectance to soil metal contamination. *Int J Remote Sens* 24: 4995–5010.
18. Rosso PH, Pushnik JC, Lay M, Ustin SL (2005) Reflectance properties and physiological responses of *Salicornia virginica* to heavy metal and petroleum contamination. *Environ Pollut* 137: 241–252.
19. Huete AR (1988) A soil-adjusted vegetation index (SAVI). *Remote Sens Environ* 25: 295–309.
20. Kaufman YJ, Tanré D (1992) Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE Trans Geosci Remote Sens* 30: 261–270.
21. Broge NH, Leblanc E (2000) Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sens Environ* 76: 156–172.
22. Bannari A, Morin D, Bonn F, Huete AR (1995) A review of vegetation indices. *Remote Sens Rev* 13: 95–120.
23. Pettorelli N, Vik JO, Mysterud A, Gaillard JM, Tucker CJ (2005) Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends Ecol Evol* 20: 503–509.
24. Blackburn GA (1998) Spectral indices for estimating photosynthetic pigment concentrations: a test using senescent tree leaves. *Int J Remote Sens* 19: 657–675.
25. Daughtry CST, Walthall CL, Kim MS, de Colstoun EB, McMurtrey JE (2000) Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sens Environ* 74: 229–239.
26. Horler DNH, Dockray M, Barber J (1983) The red edge of plant leaf reflectance. *Int J Remote Sens* 4: 273–288.
27. Guyot G, Baret F (1988) Utilisation de la haute résolution spectrale pour suivre l'état des couverts végétaux. Proceedings of the 4th International colloquium on spectral signatures of objects in remote sensing. ESA SP-287, Aussois, France (pp. 279–286).
28. Miller JR, Hare EW, Wu J (1990) Quantitative characterization of the vegetation red edge reflectance 1. an inverted-Gaussian reflectance model. *Int J Remote Sens* 11: 1755–1773.
29. Pu RL, Gong P, Biging GS, Larrieu MR (2003) Extraction of red edge optical parameters from Hyperion data for estimation of forest leaf area index. *IEEE Trans Geosci Remote Sens* 41: 916–921.
30. Carter GA (1993) Responses of leaf spectral reflectance to plant stress. *Am J Bot* 80: 239–243.
31. Carter GA (1994) Ratios of leaf reflectances in narrow wavebands as indicators of plant stress. *Int J Remote Sens* 15: 697–703.
32. Smith KL, Steven MD, Colls JJ (2004) Use of hyperspectral derivative ratios in the red-edge region to identify plant stress responses to gas leaks. *Remote Sens Environ* 92: 207–217.
33. Clevers JGPW, Kooistra L, Salas EAL (2004) Study of heavy metal contamination in river floodplains using the red-edge position in spectroscopic data. *Int J Remote Sens* 25: 1–13.
34. Rock BN, Hoshizaki T, Miller JR (1988) Comparison of in situ and airborne spectral measurements of the blue shift associated with forest decline. *Remote Sens Environ* 24: 109–127.
35. Chen J, Wang SY, Mao ZP (2011) Monitoring wetland changes in Yellow River Delta by remote sensing during 1976–2008. *Prog Geogr* 30: 587–592.
36. Liu WX, Luo YM, Teng Y, Li ZG, Wu LH (2007) A survey of petroleum contamination in several Chinese oilfield soils. *Soils* 39: 247–251.
37. Xu XG, Lin HP, Fu ZY, Bu RC (2001) Regional ecological risk assessment of wetland in the Huanghe River Delta. *Beijing Da Xue Xue Bao Zi Ran Ke Xue Bao* 37: 111–120.
38. Thenkabail PS, Smith RB, De Pauw E (2000) Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sens Environ* 71: 158–182.
39. Chaineau CH, Yepremian C, Vidalie JF, Ducreux J, Ballerini D (2003) Bioremediation of a crude oil-polluted soil: biodegradation, leaching and toxicity assessments. *Water Air Soil Pollut* 144: 419–440.
40. Ji GD, Yang YS, Zhou Q, Sun T, Ni JR (2004) Phytodegradation of extra heavy oil-based drill cuttings using mature reed wetland: an in situ pilot study. *Environ Int* 30: 509–517.
41. Rouse JW, Haas RH, Schell JA, Deering DW (1974) Monitoring vegetation systems in the Great Plains with ERTS. NASA. Goddard Space Flight Center 3d ERTS-1 Symp., Vol. 1, Sect. A: 309–317.
42. Baret F, Guyot G, Major DJ (1989) TSAVI: a vegetation index which minimizes soil brightness effects on LAI and APAR estimation, in 12th Canadian Symp. on Remote Sensing and IGARSS'90, Vancouver, Canada, 10–14 July 1989, 4 pp.
43. Major DJ, Baret F, Guyot G (1990) A ratio vegetation index adjusted for soil brightness. *Int J Remote Sens* 11: 727–740.
44. Rondeaux G, Steven M, Baret F (1996) Optimization of soil-adjusted vegetation indices. *Remote Sens Environ* 55: 95–107.
45. Haboudane D, Miller JR, Tremblay N, Zarco-Tejada PJ, Dextraze L (2002) Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sens Environ* 81: 416–426.
46. Cho MA, Skidmore AK (2006) A new technique for extracting the red edge position from hyperspectral data: the linear extrapolation method. *Remote Sens Environ* 101: 181–193.
47. Qiu GY, Yano T, Momii K (1998) An improved methodology to measure evaporation from bare soil based on comparison of surface temperature with a dry soil surface. *J Hydrol* 210: 93–105.
48. Ji GD, Sun TH, Ni JR (2007) Impact of heavy oil-polluted soils on reed wetlands. *Ecol Eng* 29: 272–279.
49. Haboudane D, Miller JR, Pattey E, Zarco-Tejada PJ, Strachan IB (2004) Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: modeling and validation in the context of precision agriculture. *Remote Sens Environ* 90: 337–352.
50. Demetriades-Shah TH, Steven MD, Clark JA (1990) High resolution derivative spectra in remote sensing. *Remote Sens Environ* 33: 55–64.
51. Peñuelas J, Filella I (1998) Visible and near-infrared reflectance techniques for diagnosing plant physiological status. *Trends Plant Sci* 3: 151–156.