Normalized Difference Vegetation Vigour Index: A New Remote Sensing Approach to Biodiversity Monitoring in Oil Polluted Regions

Nkeiruka Nneti Onyia 1,*, Heiko Balzter 1,2 and Juan-Carlos Berrio 1

1 School of Geography, Geology and Environment, University of Leicester, University Road, Leicester LE1 7RH, UK; hb91@le.ac.uk (H.B.); jcb34@leicester.ac.uk (J.-C.B.)
2 National Centre for Earth Observation, University of Leicester, University Road, Leicester LE1 7RH, UK
* Correspondence: nno4@le.ac.uk; Tel.: +44-79-563-610-73

Received: 1 May 2018; Accepted: 5 June 2018; Published: 7 June 2018

Abstract: Biodiversity loss remains a global challenge despite international commitment to the United Nations Convention on Biodiversity. Biodiversity monitoring methods are often limited in their geographical coverage or thematic content. Furthermore, remote sensing-based integrated monitoring methods mostly attempt to determine species diversity from habitat heterogeneity somewhat reflected in the spectral diversity of the image used. Up to date, there has been no standardized method for monitoring biodiversity against the backdrop of ecosystem or environmental pressures. This study presents a new method for monitoring the impact of oil pollution an environmental pressure on biodiversity at regional scale and presents a case study in the Niger delta region of Nigeria. It integrates satellite remote sensing and field data to develop a set of spectral metrics for biodiversity monitoring. Using vascular plants of various lifeforms observed on polluted and unpolluted (control) locations, as surrogates for biodiversity, the normalized difference vegetation vigour index (NDVVI) variants were estimated from Hyperion wavelengths sensitive to petroleum hydrocarbons and evaluated for potential use in biodiversity monitoring schemes. The NDVVI ranges from 0 to 1 and stems from the presupposition that increasing chlorophyll absorption in the green vegetation can be used as a predictor to model vascular plant species diversity. The performances of NDVVI variants were compared to traditional narrowband vegetation indices (NBVIs). The results show strong links between vascular plant species diversity and primary productivity of vegetation quantified by the chlorophyll content, vegetation vigour and abundance. An NDVVI-based model gave much more accurate predictions of species diversity than traditional NBVIs (R-squared and prediction square error (PSE) respectively for Shannon’s diversity = 0.54 and 0.69 for NDVVI and 0.14 and 0.9 for NBVI). We conclude that NDVVI is a superior remote sensing index for monitoring biodiversity indicators in oil-polluted areas than traditional NBVIs.

Keywords: biodiversity monitoring; species diversity; hyper-spectral imaging; Hyperion; spectral metrics; vegetation vigour; oil pollution

1. Introduction

A major concern over the global loss of biodiversity is that its interference with nature’s balance may cause irreversible damage by altering key processes essential for the productivity and sustainability of life on earth [1,2]. Natural water, carbon and other cycles are dependent on the interactions and interdependence of various species that inhabit the earth, and as noted by [3], the diversity and identity of organisms control the functioning of ecosystems. In their review paper evaluating the impact of biodiversity loss on humanity, Cardinale et al. [4] showed evidence of the
Several studies have established linear relationships between species diversity and ecosystem productivity with the implication that biodiversity loss may lead to loss or decrease in the supply of ecosystem goods and services [5–9]. Norris [7,9] agree that the loss of biodiversity inadvertently results in the loss of ecosystem functions and services. The unprecedented loss of biodiversity over the years led to the signing of convention on biological diversity (CBD) in 1992 by the United Nations member countries. The overarching goal was the conservation and sustainable use of biological diversity in the present and future. To achieve this, targets were set to halt or reduce biodiversity loss with the most recent being the Aichi 2020 biodiversity targets [10,11]. As Nigeria is a signatory to the CBD, the need to monitor the rate and extent of biodiversity loss becomes paramount, particularly in threatened biodiversity hotspots like the Niger delta where oil spills continue to degrade the ecosystem.

The application of remote sensing technology in biodiversity monitoring has been an area of considerable research in the recent past, for instance [12–14]. Several studies have combined remote sensing and field data to determine the spatial and temporal distributions of biodiversity [15,16]. Motivated by the need to devise a standardized methodology for monitoring biodiversity at regional and global scales and to overcome some limitations of conventional monitoring methods such as in situ field surveys, researchers have developed integrated methods using remote sensing and field data [17–21]. The Group on Earth Observations Biodiversity Observation Network (GEOBON) have spearheaded these efforts leading to the development of the essential biodiversity variables (EBVs). An EBV is defined as a measurement required for the study, reporting and management of biodiversity change [22]. They provide standardized measurements and observations necessary to calculate indicator transformations hence, are useful for monitoring biodiversity [23–25]. For instance, changes in measurable aspects of biodiversity such as species population, abundance and distribution can provide early warning signs of significant alteration in biodiversity of the region [22–27].

Although there are inherent challenges associated with such integrated schemes, studies [28–30], have confirmed that they provide very useful information on the response of biodiversity to natural and anthropogenic changes in the environment. While some studies have investigated the link between spectral diversity and species diversity through the biochemical diversity of vegetation [31], others used land cover classifications derived from multispectral sensors such as Landsat to estimate the species diversity of the area of interest [32]. The land cover approach has been criticized as inadequate for collecting fine-scale detail of vegetation structure and chemistry due to the coarse spectral and spatial resolution [14,32] of multispectral sensors. Generally, mapping of species diversity estimates are empirically supported by defining the relationship between variation in spectral signal and variation in species or habitat diversity [15,33–35] and in some cases, variations in pigment concentrations [36–38]. However, this procedure may not be enough if the Aichi 2020 targets are achievable. There is need to develop methodologies that estimate species diversity against the backdrop of environmental pressures such as oil pollution. Such methods will not only reveal the state of the ecosystem following impact (for instance, biodiversity change), but also reveal ecosystem resilience to particular external pressure.

Hyperspectral data not only measure vegetation biochemical and biophysical properties including water content, leaf pigments, nitrogen, cellulose and lignin concentrations [30,39–41], but also how these parameters vary across the ecosystem [14,38]. The Hyperion sensor on board NASA’s Earth Observation-1 satellite is an example of a hyperspectral satellite mission [42]. It was the first satellite hyperspectral sensor launched on-board the Earth Orbiter-1 platform by the United States National Aeronautics and Space Administration (NASA) New Millennium Program (NMP). The Hyperion acquire 16 bits, 30 m spatially resolved data in 220 discrete narrow-bands between the spectral range of 400 and 2500 nm. The sensor capture about 75 times more data than multispectral sensors from a similar area, hence providing a large volume of data that need enhanced analyzing techniques [43].

Vegetation vigour defined as active, healthy, well-balanced and robust growth of plants [44] is recognized as an essential environmental quality index. Vigorous vegetation is characterized by
enhanced growth and extent as well as increased productivity [45]. Several studies on vegetation have linked the index to climate change [46]; soil erosion [47]; and biological conservation [48]. Some organizations such as [49,50] recommend the use of vegetation vigour index in tests to evaluate the effect of chemical substances such as pesticides on the growth of various plant species.

In the present study, vegetation vigour is associated with vegetation productivity and/or biomass production. This is in line with previous publications such as [2–9,51–57], which have also revealed strong relationship between these parameters and vascular plant species diversity. Tillman et al. [55] as well as [56–58], following their various experiments found a positive relationship between plant species diversity and plant productivity (measured as above ground biomass) which they attributed to several factors including the complementarity of resource use.

More recently, determination of vegetation vigour has been from remotely sensed images using proxy vegetation indices such as the normalized difference vegetation index (NDVI) and net primary productivity (NPP) [51,59–61]. Other studies have used vegetation cover as an indicator of vegetation vigour and established a positive linear relationship between both variables [62–64]. With advancement in earth observation technology in the way of increasing spectral and spatial resolutions, researchers have exploited these relationships with concerted effort to identify individual species and estimate species diversity of a given area using spectral metrics derived from satellite imagery [13,65–70].

In this study, spectral metrics were created from Hyperion bands that were sensitive and those that were insensitive to total petroleum hydrocarbon (TPH) concentration in soil. Eigemeier et al. [71] stated that the performance of a vegetation index is better when both most sensitive and insensitive bands to a parameter of interest are included in the algorithm. Hence, the new index were used in this study to investigate the capability of spectral metrics derived from Hyperion image to estimate species diversity of vascular plants on oil polluted transects in the Niger Delta region of Nigeria. The following research questions have been posed.

1. Are vascular plants susceptible to oil pollution and does this affect their spectral signatures?
2. Is there any relationship between plant spectral signatures and vascular plant species diversity?
3. Can this relationship be modelled to estimate the diversity of vascular plants on oil-polluted transects?

2. Materials and Methods

2.1. Study Area

The study area is around the Shell Petroleum Development Corporation (SPDC) 28” Nkporo-Bomu Pipeline at Kporghor located in Tai Local Government Area (LGA) of Rivers State in Nigeria (Figure 1). Tai LGA is one of the four administrative units that make up Ogoniland. Two spill locations were selected because they are among the few oil spill locations within Ogoniland for which a hyperspectral image is available. The United States Geological Survey Earth Observation 1 (USGS EO-1) Hyperion sensor acquired the image following our data acquisition request (DAR) submitted in 2015.

The study area, Kporghor, is located in the Niger Delta region of Nigeria which is one of the largest wetlands in the world that has substantially degraded due to oil pollution necessitating in 2011, the assessment of area by a special team of the United Nations Environment Program [72,73]. It receives some of the highest rainfall in the world of up to 3000 mm annually [74]. The temperature ranges from 20° to 30 °C during the day. The Intertropical Front wind originating where the humid air masses from the Gulf of Guinea and dry air masses from the north meet is constantly blowing over the delta resulting in high levels of humidity with an average of 75% [75]. The area is also characterized by high cloud cover which invariably affects the quality of satellite data obtained at certain times of the year [74]. Due to the heavy rainfall experienced in this region, there is continuous vegetation growth all year round; however, most agricultural activities occur between March and October each year.
Ecologically, Kporghor has been classified as a coastal plain and freshwater rainforest zone populated by forest tree species, palms, shrubs, ferns, lianas and so on [76,77]. Dominant tree species include the indigenous oil palm \( (Elaeis guineensis) \), raphia palm \( (Raphia vinifera) \) and timber species \( (Symphonia globulifera) \) [76]. Others are \( Eichornia crassipes \); \( Calamus \) and \( Alchornea \) species, \( Alstonia boonei, Berlinia \) species [79]. Osuji and Ezeibu [80] also reported the occurrence of \( Antidesma \) species, \( Paulina pinnata, Oureta \) species, \( Chassalia \) species, \( Cuvaria \) species, \( Dryopteris \) species, \( Memsecylon \) species, \( Blackoides \) species, \( Agelea oblique \) and \( Psychotria manii \) in the study area.

The first spill epicentre (P1S0) is located at 4.71842°N latitude and 7.22523°E longitude (WGS 84 Datum). The landscape of the area is flat with elevation of less than 14 m above sea level. The Nigerian Oil Spill Monitor website [81] reported that the spill incident, which occurred on the 12th of September 2015, was attributed to crude oil theft (illegal bunkering). About 864 barrels of crude oil (137,365 litres) spilled onto a land area of 4190 m\(^2\) while only about 557 barrels (88,556 litres) was recovered through the containment measures implemented by the company. This implied that about 307 barrels (48,809 litres) remained in the impacted environment damaging the ecosystem. The second spill epicentre (P2S0) is located at 4.719906°N latitude and 7.22375°E longitude. Data about this site was obtained from the United Nations report on the Environmental Assessment of Ogoniland carried out between 2009 and 2011. According to the report, the oil spill incident occurred on the 1 June 2007 following sabotage and oil theft activities [72,81].

2.2. Establishment of Study Transects

The study was conducted on polluted and unpolluted (control) locations. Polluted locations were identified from the Nigerian oil spill records website [81] while control locations were randomly selected from unpolluted areas determined as such from past and present oil spill records. All the investigated transects share similar environmental and ecological conditions as they are located within...
the same ecological zone (coastal plain and freshwater rainforest zone) \cite{76,77} with oil pollution being the main distinguishing factor. Four transects of 100 m × 0.6 m were established on polluted and control locations. Polluted transects A, B, C and D proceeded northwards, southwards, eastwards and westwards, respectively, starting from the spill epicentre (P1S0). Each transect was subdivided into 5 segments of 20 m each labelled SS1 to SS5. At the second spill location, field survey was done at four 20 m by 20 m segments overlaying the UNEP sampling point. These were labelled P2S0, P2S1, P2S2, and P2S3. Control transects were randomly established in areas that were unpolluted at the time of the survey. Locations of control transects provided in supplementary materials (Supplementary A: Table S1). Randomization of control transects was done to remove bias. Each transect was subdivided into segments of 20 m length. Transects were established using fibre tapes, hammer, and iron pegs coated in white paint and a hand held GPS device. In total 33 segments were investigated, 17 on polluted transects and 16 on control transects. Vegetation and soil TPH data were obtained from these segments. Similarly, the spectral data analyzed in this study were extracted from pixels overlaying these segments.

2.3. Field Survey of Vascular Plant Species

Vegetation survey was implemented through the line-intercept sampling method. The survey was conducted in two campaigns during the months of February to March 2016 and April to May 2017. The first vegetation survey of polluted transects A, B, C, D and control transects C1 and C2 was performed in 2016, while the survey of the second polluted location (P2) and control transects C3 and C4 was performed in 2017. The choice of using the line-intercept method was because it offered the most potential at capturing adequate field data to determine species composition and abundance within the unique constraints of the study area. Buckland et al. \cite{82} affirmed that vegetation records obtained from transects provide sufficient information with which to calculate the numerical abundance, frequency, coverage and other relevant vegetation characteristics. It is also less time consuming and gives relatively accurate information.

Additionally, \cite{83} noted that this sampling method is useful for highlighting transition zones (ecotones) in vegetation and it is well adapted for investigating the relationships between changes in floristic compositions and environmental variables. Another advantage of using the line-intercept method is that the ease with which species are recorded as they occur along the line gives little room for errors \cite{84}.

Occurrence and number of individuals for each species of vascular plants was recorded per segment of transects. Plants that occurred within 30 to 100 cm on both sides of the transect lines were included in the count. To reduce the effect of spatial autocorrelation, only three alternating segments were used in the analysis, namely SS1, SS3 and SS5. In situ species identification was done with the aid of species inventory sheets compiled from previous literature \cite{76–78,85,86}. Photographs of unidentified species were obtained and taken to the herbarium in the local University (College of Natural Resources and Environmental Management, Michael Okpara University of Agriculture, Umudike Umuahia, Nigeria) for identification.

The Sorensen’s similarity index of transects, Menhinick’s species richness index (M), Shannon’s (H) and Simpson’s (D) diversity indices were computed in the Palaeontological Statistics Software (PAST), \cite{87} for selected segments.

The similarity index of investigated transects was computed to measure the degree of agreement between the species composition of polluted and control transects in order to validate their comparison. The index is computed using the presence-absence vegetation data and range from 0 to 1 as similarity of the units increase. Several researchers have used the Sorensen’s index as an appropriate test to determine the similarity between the units being studied. The units may be quadrants \cite{88} or localities \cite{89}. In this study, segments of polluted and control transects were assessed for similarity in species composition.
The alpha diversity ($\alpha$-diversity) of vascular plant species was also computed to measure the diversity of vascular plant species in the study area. Alpha diversity describes the number of species present (species richness) and the distribution pattern of the individual members of these species (species evenness or species equitability) within an ecological unit (habitat level) [90–93]. Commonly used diversity indices include those that measure only species richness (Menhinick’s and Chao-1 index) as well as those that measure species richness and evenness (Simpson’s and Shannon’s diversity indices).

2.4. Soil Sampling and Analysis

To determine the concentration of crude oil in the soil, soil samples were collected from each segment at 30 cm depth with the aid of hand augurs. At each segment, three samples were taken from random points and mixed together to form a composite sample. A portion of the composite sample was stored in a sterilized and labelled air-tight glass bottle and taken to the laboratory in a plastic box filled with ice packs. This was done to control further chemical reactions in the soil. Determination of TPH in samples followed standard procedures stipulated in the United States Environmental Protection Agency (U. S. EPA) method 8015 document involving gas chromatography (GC) technique and flame ionization detector (FID) electrodes [94]. All analyses were carried out in International Energy Services Limited laboratory located in Port Harcourt Nigeria, according to national and international quality standards. Ancillary data comprising location and TPH concentration of soil samples was obtained from a UNEP report following the environmental assessment of Ogoniland in 2011 [72].

2.5. Hyperspectral Image Description and Preparation

Researchers have established strong connections between species diversity and spectral variation [15,34,35,95]. These studies have mostly relied on hyperspectral data with hundreds of bands capable of detecting very subtle changes in plant attributes. Hyperspectral data not only measure vegetation biochemical and biophysical properties including water content, leaf pigments, nitrogen, cellulose and lignin concentrations [30,39–41], but also how these parameters vary across the ecosystem [14,38]. Therefore, in order to achieve the aim of this study, a hyperspectral image acquired by the Hyperion sensor on-board the Earth Orbiter 1 (EO-1) satellite launched in November 2000 by NASA was used. The sensor provides radiometrically calibrated spectral data acquired by a push-broom system in single frames measuring 7.65 km (cross-track) by 185 km long (along-track), acquired in NADIR position from an altitude of 705 km. Each pixel in a Hyperion image is approximately 30 m$^2$ on the ground and surface reflectance is measured in 242 spectral channels ranging from 400 nm to 2500 nm. There are, however, only 196 effective and calibrated bands [96,97] while the others suffer from the effect of bad detectors that result in ‘bad pixels’ which are set to zero during level 1 processing. Calibrated bands include the VNIR bands 8–57 (427.55 nm to 925.85 nm) and SWIR bands 79 to 224 (932.72 nm–2395.53 nm). Due to the strong absorption of water vapour and oxygen at wavelengths ranging from 1356 nm–1417 nm, 1820 nm–1932 nm and >2395 nm, as well as the overlap of wavelengths in the VNIR (band 56 and 57) and SWIR (77 and 78) regions, the list of useful bands was limited to 176 [97].

The Hyperion image was downloaded as a Level 1T geotiff file from the United States Geological Survey (USGS) website using the earth explorer data download tool. The image was acquired on the 23 November 2015. Due to the narrow swath width of the image of 7.65 km, only two polluted sites in Kporghor were captured.

Generally, image pre-processing is performed to eliminate noise and other artefacts arising from atmospheric interferences and internal sensor defects [98–100]. In this study, the Hyperion image was pre-processed as shown in Table 1. During the atmospheric correction using the FLAASH module in ENVI 5.3, reflectance value was scaled by 10,000 to remove decimal points and reduce computational time. The resulting image had 68 columns and 66 lines, 4488 pixels and 164 spectral bands.
Table 1. Pre-processing steps performed with the Hyperion image.

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Purpose</th>
<th>Module/Tool</th>
<th>Software</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image subset</td>
<td>Focus on study area and reduce processing time</td>
<td>Region of Interest (ROI)</td>
<td>ENVI 5.3</td>
</tr>
<tr>
<td>Atmospheric correction</td>
<td>Removal of atmospheric interference to retrieve surface reflectance.</td>
<td>Fast Line-of-sight Atmospheric Analysis of Hypercubes (FLAASH)</td>
<td>ENVI 5.3 [100]</td>
</tr>
<tr>
<td>Removal of smile effect</td>
<td>Enhance retrieval of surface reflectance</td>
<td>Cross-Track Illumination Correction (CTIC)</td>
<td>ENVI 5.3 [101,102]</td>
</tr>
<tr>
<td>Noise reduction and Destripping</td>
<td>Maximize signal to noise ratio (SNR) and minimise data dimensionality</td>
<td>Minimum Noise Fraction Transformation (MNFT)</td>
<td>ENVI 5.3 [103,104]</td>
</tr>
</tbody>
</table>

2.6. Processing Landsat 8 and Sentinel 2A Images

Two other satellite datasets were used in this study during implementation of the best performing model on randomly selected pixels to evaluate its performance. NDVI of the study area was computed from Landsat 8 Surface Reflectance and Sentinel 2A images acquired on the 4 January 2016 and the 22 December 2015 respectively. These acquisition dates were close enough to that of the Hyperion image. The specifications of both images are given in Table 2 below.

Table 2. Specifications of Landsat 8-OLI and Sentinel 2A Images used to compute NDVI.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Landsat 8-OLI</th>
<th>Sentinel 2A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product ID</td>
<td>LC08_L1TP_188057_20160104_20170404_01_T1, LC81880572016004LGN02</td>
<td>S2A_OPER_MSI_L1C_TL_SGS_20151222T100120_20151222T151903_A002607_T32NKL_N02_01_01</td>
</tr>
<tr>
<td>Bands</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td>Spectral</td>
<td>0.435–1.384</td>
<td>0.44–2.22</td>
</tr>
<tr>
<td>Spatial</td>
<td>30 m</td>
<td>10 m, 20 m, 60 m</td>
</tr>
<tr>
<td>Radiometric</td>
<td>16 bits</td>
<td>12 Bits</td>
</tr>
<tr>
<td>Temporal</td>
<td>16 days</td>
<td>10 days</td>
</tr>
<tr>
<td>Sensor</td>
<td>Operational Land Imager (OLI)</td>
<td>Multispectral Instrument (MSI)</td>
</tr>
<tr>
<td>Type</td>
<td>Multispectral</td>
<td>Multispectral</td>
</tr>
<tr>
<td>Satellite</td>
<td>Landsat 8</td>
<td>Sentinel-2A</td>
</tr>
<tr>
<td>Mission</td>
<td>Landsat Program</td>
<td>Sentinel</td>
</tr>
<tr>
<td>Operator</td>
<td>U.S. Geological Survey (USGS)</td>
<td>European Space Agency (ESA)</td>
</tr>
</tbody>
</table>

The sentinel 2A dataset was preprocessed using the ESA Sentinel Application Platform (SNAP). This toolbox provides functions for clipping the dataset to the study area, performing atmospheric correction using the Sen2Cor plug-in and computing the NDVI using relevant bands. Details for these procedures are provided in the Sentinel 2 documentation available at the ESA website [105–107].

The Landsat Surface Reflectance Image, on the other hand is a Level 2 processed product, hence needed only spatial subset to the study area before the NDVI was computed in ENVI 5.3. The Level 2 product measured the fraction of incoming solar radiation reflected back from the earth’s surface to the Landsat sensor [108] hence; atmospheric interferences are removed at level 2 pre-processing.

2.7. Vegetation Indices

Vegetation indices are derived from the spectral reflectance of plants. Plant species produce unique spectral signatures dependent on their physiological state at the time of acquisition. Photosynthetic and protective processes involving pigments such as chlorophylls, carotenoids, and anthocyanins are the major determinants of plant physiological status and productivity. These pigments essentially convert light energy to the chemical energy needed for plant productivity. The proportion of the different pigments in the leaf determine the amount of solar radiation absorbed and reflected hence, they serve
as valuable channels for information transfer from plants to space-based sensors. Many researchers have used satellite derived vegetation indices to investigate, explain [109] and estimate environmental phenomena [110]. Additionally, they have been incorporated in models developed for estimating, monitoring, mapping and analyzing ecosystem structures such as vegetation cover and species composition as well as functions such as productivity and biomass (for instance [111–114]). Here, narrowband vegetation indices (NBVIs) were computed from the Hyperion bands. NBVIs are derived from narrowband reflectance obtained from hyperspectral sensors. Several studies have applied NBVIs to determine the structure, biochemical, biophysical and physiological or stress status of vegetation in various habitats [111,114–117]. The main parameters measured by NBVIs include

a. Chlorophyll Content: used to monitor changes in green biomass, chlorophyll content and leaf structure. High values indicate increased chlorophyll content, green biomass and vegetation vigour and

b. Primary Productivity: measure changes in the photosynthetic light use efficiency of plants. High values indicate reduced light use efficiency, hence reduced productivity.

NBVIs have been shown to overcome the saturation problem easily associated with broadband vegetation indices such as NDVI [118]. The NBVIs and NDVI evaluated in this paper are listed in Table 3. The indices were computed in ENVI 5.3 and index values were extracted for each segment in polluted and control transects.

<table>
<thead>
<tr>
<th>Index</th>
<th>Formula</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normalized Difference Vegetation Index (NDVI)</td>
<td>( \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}} )</td>
<td>[119]</td>
</tr>
<tr>
<td>Red-Edge NDVI (RENDVI)</td>
<td>( \frac{\text{R}<em>{750} - \text{R}</em>{705}}{\text{R}<em>{750} + \text{R}</em>{705}} )</td>
<td>[120]</td>
</tr>
<tr>
<td>Modified Red-Edge NDVI (MRENDVI)</td>
<td>( \frac{\text{R}<em>{750} - \text{R}</em>{705}}{\text{R}<em>{750} + \text{R}</em>{705} - 2 \times \text{R}_{445}} )</td>
<td>[121]</td>
</tr>
<tr>
<td>Modified Red-Edge Simple Ratio Index (MRESRI)</td>
<td>( \frac{\text{R}<em>{750} - \text{R}</em>{445}}{\text{R}<em>{705} + \text{R}</em>{445}} )</td>
<td>[122]</td>
</tr>
<tr>
<td>Vogelmann Red Edge Index 1 (VREI1)</td>
<td>( \frac{\text{R}<em>{740}}{\text{R}</em>{720}} )</td>
<td>[123]</td>
</tr>
<tr>
<td>Photochemical Reflectance Index (PRI)</td>
<td>( \frac{\text{R}<em>{531} - \text{R}</em>{570}}{\text{R}<em>{531} + \text{R}</em>{570}} )</td>
<td>[124]</td>
</tr>
<tr>
<td>Structure Insensitive Pigment Index (SIPI)</td>
<td>( \frac{\text{R}<em>{800} - \text{R}</em>{445}}{\text{R}<em>{800} + \text{R}</em>{445}} )</td>
<td>[125]</td>
</tr>
</tbody>
</table>

2.8. Derivation of Stress-Sensitive Wavelengths by Sensitivity Analysis

Sensitivity analysis is a mathematical procedure which determines how changes in levels of an independent variable affect changes in levels of a response variable [126]. The response level (wavelengths of vegetation reflectance, in this study) showing the largest relative change in response to changes in levels of the independent variable (soil TPH) is considered to be the most sensitive response level [127]. According to [126], this procedure is necessary to evaluate the influence of variables and rank their significance based on their influence. Sensitivity analysis was performed for both polluted and control vegetation on the assumption that the environmental and edaphic conditions are homogenous and that polluted vegetation would be more stressed than control vegetation due to the influence of TPH in the soil.

To investigate the effect of oil pollution (independent variable) on vegetation reflectance (response variable) using Hyperion data, two ideas were explored. Firstly, the wavelengths that were sensitive to TPH concentration in the soil were identified and the reflectance at these wavelengths were compared to determine if there was significant difference between polluted and control transects. The criterion for identifying sensitive wavelengths was vegetation response to TPH concentration in the soil which was expected to influence spectral reflectance [128]. Sensitivity analysis of vegetation reflectance spectrum to soil TPH was also necessary to differentiate between the plant stress caused by TPH concentration in the soil and other edaphic factors while also correcting for irradiance, leaf orientation, irradiance angle and shading [129].
Since vegetation reflectance at the VNIR generally increases with plant stress [130,131], sensitivity to TPH-induced stress in vegetation for VNIR Hyperion wavelengths was computed by subtracting the reflectance of control vegetation (non-stressed) from that of stressed vegetation (polluted). The resulting difference was normalized by further dividing by the reflectance of the non-stressed vegetation to establish the sensitivity of each wavelength to soil TPH [129]. The formulae for computing the reflectance difference and sensitivity are as follows:

\[ R_\Delta = R_u - R_n, \]  
\[ R_s = \frac{R_u - R_n}{R_n}, \]

where

- \( R_n \) is reflectance of non-stressed vegetation
- \( R_u \) is reflectance of stressed vegetation
- \( R_\Delta \) is reflectance difference
- \( R_s \) is reflectance sensitivity

Following sensitivity analysis and ranking of the VNIR Hyperion wavelengths, the most sensitive wavelengths with the highest sensitivity values and correspondingly high reflectance difference values positioned at the top of the rank order, while the least sensitive bands with sensitivity values closer to zero were positioned at the bottom of the rank order. The five most sensitive and five least sensitive wavelengths were selected and used in creating the normalized difference vegetation vigour index (NDVVI). These wavelengths were selected and combined to create an index with maximum sensitivity to vegetation response to soil TPH because index performance is improved when both sensitive and insensitive wavelengths are used in its creation [71]. Additionally, vegetation sensitivity to soil TPH appeared to be limited to specific wavelengths in the blue, red and NIR channels. To investigate the full range of oil pollution impact on vegetation, the NDVVI variants were created from these channels.

### 2.9. Calculation of the Normalized Difference Vegetation Vigour Index (NDVVI)

Vegetation vigour has long been recognized as an essential environmental quality index used in studies investigating various aspects of the environment [46,47,132]. The index has also been recommended for use in tests to evaluate the effect of chemical substances on the growth of various plant species [49,50]. Several studies use the vegetation vigour index derived from remote sensing data to measure vegetation productivity and biomass production.

In the present study, a new vegetation vigour index was created to measure the response of vegetation to the presence of TPH in the soil, and compare how the response differs on polluted and control transects. The index referred to as the normalized difference vegetation vigour index (NDVVI) was computed by normalizing reflectance difference at the least and most sensitive wavelengths. The NDVVI was computed for each segment using the formula

\[ \text{NDVVI} = \frac{R_i - R_j}{R_i + R_j}, \]

where

- \( R_i \) = reflectance at least sensitive wavelength
- \( R_j \) = reflectance at most sensitive wavelength

The most sensitive wavelengths were those which exhibited large difference in reflectance between polluted and control transects while the least sensitive wavelengths were those whose reflectance values hardly changed in the presence of TPH. Previous studies showed that NIR reflectance did not vary much between healthy leaves and stressed leaves [133,134]. Since reflectance at these NIR wavelengths is almost constant, the index thus relies on reflectance at the most sensitive wavelengths.
which occurred in the blue and red channels. Thus, NDVI value is zero when reflectance at either the red or blue channel is high and 1 when reflectance at these channels is low. The inclusion of the least sensitive wavelengths in creating the NDVVI automatically corrects for reflectance from non-vegetated areas (such as bare soils and buildings) which show high reflectance at the NIR region. Six NDVVI variants were created combining the five most sensitive and least sensitive Hyperion wavelengths. These were NDVVI814,437, NDVVI824,427, NDVVI844,447, NDVVI752,630, NDVVI773,641, and NDVVI844,630. The range of index values are shown in Figure 2.

![Figure 2. Raster images of the NDVVI variants used in the model. The low index value of polluted transects is clearly seen in the images. Additionally, roads, buildings and waterbody (areas with low vegetation density) are clearly seen to have very low index values, which is a reflection of the new index properties.](image)

### 2.10. Statistical Analyses

The potential of the new NDVVI variants to estimate the species richness and diversity of the study area was assessed using non-parametric statistics. Vegetation indices such as the normalized difference vegetation index (NDVI) are commonly used to predict species richness and diversity [32,95,135–137], therefore, we compared the capability of the new NDVVI variants and traditional narrow-band vegetation indices (NBVIs) listed in Table 3 for predicting vascular plant species diversity. As the NDVVI is derived from Hyperion wavelengths with different levels of sensitivity to soil TPH, it is expected that the NDVVI values will be higher for control transects than for polluted transects and that consequently, estimated species richness and diversity indices will follow similar patterns. Thus, underscoring the need for the application of the new index in biodiversity monitoring in oil producing regions.

Each diversity index was modelled as a function of the spectral metrics derived from the Hyperion image and the traditional NBVIs. We tested for differences between the median NDVVI and median NBVIs of the polluted vs. control transects with the Mann-Whitney U test (M-W test). The Null hypothesis is that the median index of the polluted transects is the same as the median of the control transect. For all tests, the type I error was controlled at $\alpha = 0.05$. The spectral metrics and vegetation indices were then regressed with field measurements of biodiversity indices to determine the strength and significance of the relationships. As the NDVVI is a good indicator of vegetation...
vigour, chlorophyll content and productivity, it is expected to have a strong positive relationship with field biodiversity indicators in accordance with [5–9].

The spectral metrics and vegetation indices were used as predictors in models designed to estimate various species diversity indices of investigated transects. Partial least square regression (PLS) and non-parametric multivariate regression (NPM) procedures were employed to model the relationship between derived spectral metrics and field-measured diversity indices (Shannon’s, Simpson’s, Menhinick’s and Chao-1). These regression models were selected because they are not limited by assumptions of data distribution common with parametric regression procedures.

PLS regression was applied to evaluate the predictive capability of the new spectral index (NDVVI variants) for comparison with traditional NBVIs. The PLS technique performs multivariate regression without the restrictions associated with the standard regression methods. It is particularly useful when predictor variables outnumber response variables and when there is high multicollinearity between the predictor variables. The procedure transforms the predictor data (in this case, the NDVVI variants derived from the Hyperion data and the traditional NBVIs) into a smaller set of uncorrelated components and performs least square regression on these components instead of the original data. PLS has been shown to be effective for the analysis of hyperspectral data [38,138] due to high multicollinearity of the bands. Selection of the optimum number of component is ascertained from coefficient of determination ($R^2$) values which refer to how much of the variance in the predictors and between the predictors and response is explained by each component. For highly correlated predictors, it is normal for fewer components to be used in the model.

NPM regression analysis was performed to account for any violations of the assumptions about the distribution of the data. Non-parametric methods allow the modelling of densities and local polynomial regression on both continuous and categorical data which do not necessary follow any pre-defined distribution [139]. The $np$ package in R developed by [140] was used for this analysis. The procedure commences with selection of the optimum bandwidths estimated from second-order Gaussian kernel densities. The bandwidth objects are then assigned to an appropriate regression function, which determines the fitting of the curve and calculate the fitted, predicted and error values. The $np$ package has a multi-start function, which helps to avoid errors that occur in the presence of local minima. Since the NPM is based on kernel density estimation, choosing the smoothing parameter (bandwidth) is very crucial. In this study, optimum bandwidths were selected using the Akaike information criterion (AIC), a classical method of unbiased estimation that minimizes the expected Kullback-Leibler divergence [141].

For model validation purposes, the original dataset was randomly sub-divided into training data and test data in the ratio of 6:4. The training data used to calibrate the models was made up of 20 observations (10 segments each from polluted and control transects) while the test data used for model validation comprised 13 observations (7 polluted and 6 control segments). Model performance was evaluated in terms of model type (PLS or NPM) and predictors (NDVVI variants or NBVIs). Each model type was tested with both sets of spectral metrics (NDVVI variants and NBVIs) to determine their performance in vascular plant species diversity estimation. In total, four different models were developed and tested. The characteristics of each model is defined in Table 4. The coefficient of determination ($R^2$), residuals, bias and random error of all four models were compared to identify the subset of spectral metrics best suited for vascular plant species diversity estimation.

<table>
<thead>
<tr>
<th>Model ID</th>
<th>Regression Method</th>
<th>Predictors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1A</td>
<td>Partial Least Squares (PLS)</td>
<td>NDVVIs</td>
</tr>
<tr>
<td>1B</td>
<td>&quot;</td>
<td>NBVIs</td>
</tr>
<tr>
<td>2A</td>
<td>Non-Parametric Multivariate Regression (NPM)</td>
<td>NDVVIs</td>
</tr>
<tr>
<td>2B</td>
<td>&quot;</td>
<td>NBVIs</td>
</tr>
</tbody>
</table>
Based on the performance of the models, the NDVVI-based NPM model was selected for implementation. Vascular plant species diversity indices of randomly selected pixels (predsites) were estimated from their NDVVI values. The spectral values for the predsites were extracted from the NDVVI variants (NDVVI_{752,630}, NDVVI_{814,437}, NDVVI_{824,427}, NDVVI_{773,641}, NDVVI_{844,447}, NDVVI_{844,630}, as illustrated in Figure 3, using the Raster and GISTools packages in R. In total, 30 pixels were randomly selected in such a manner that all the visible land cover types (water body, swamp, farmland, mixed vegetation and forested) within the study area were captured in the dataset. The decision for randomly selecting pixels from within the study area is due to the fact that it has been classified as polluted following reports of oil spills at various sites across the area in the past. However, soil TPH concentrations are expected to be lower in the predsites than is observed at polluted transects and consequently, higher species diversity values are anticipated from implementing the selected model.

Figure 3. Map of NDVVI_{814,437} for Kporghor displaying the locations of the randomly selected pixels (predsites) used for evaluating the regression model. Shannon’s, Simpson’s, Menhinick’s and Chao-1 diversity indices were estimated for the predsites using the NDVVI variants.

The selected model was then applied to the new dataset of derived spectral metrics (preddata). For this procedure, only the satellite derived spectral metrics were utilized. Each diversity index (Shannon’s, Simpson’s, Menhinick’s and Chao1 index) was estimated separately for the predsites. Due to the absence of field data for the predsites, estimation accuracy was determined by correlating predicted values with corresponding NDVI values computed from a Landsat 8 and Sentinel-2A images of the study area. NDVI was selected because it is a well-known vegetation index, which measures vegetation health and density and is commonly used as a surrogate for measuring species diversity in some studies. Since vegetation productivity increases with species diversity [2–9,51–57], we presume that the NDVI values will strongly correlate with the diversity estimates of the ‘predsites’. However, as NDVI is known to saturate at high vegetation densities, researchers have advocated for more robust indices that can handle to range of vegetation densities frequently observed in the tropical rain forests. The choice of a different sensor to calculate the NDVI was to minimize bias from using an NDVI
image calculated from Hyperion data. Both the Landsat 8 and Sentinel 2A images were downloaded from the USGS earth explorer tool and the vegetation index computed using ENVI 5.3. Additionally, accuracy of predictions was visually evaluated using very high-resolution imagery from digital globe freely available in Google Earth (GE, hereafter). Several researchers have utilized the GE images as a visualization tool for land use and land cover maps [142–144].

3. Results

3.1. Sorensen’s Similarity and Diversity Indices of Transects

The similarity of polluted and control transects investigated in this study was determined using the Sorensen’s similarity index. Results in Table 5 show that vegetation composition of polluted and control transect were adequately similar for comparison. The Sorensen’s index for all pairs of transects was greater than 0.5 except for observations on the second polluted location P2 which showed less similarity with the other transects (similarity index values <0.5). The higher index values portray greater similarity among transects; however, it is apparent that intra-transect similarity (similarity among polluted transects or among control transects) was greater than inter-transect similarity (similarity between polluted and control transects). The observed pattern suggest that differences in species composition and vegetation reflectance may be attributed to soil TPH concentration.

Table 5. Sorensen’s similarity index of paired transects showing strong similarity in species composition of polluted and control transects.

<table>
<thead>
<tr>
<th>Transect A</th>
<th>Transect B</th>
<th>Transect C</th>
<th>Transect D</th>
<th>P2</th>
<th>Transect C1</th>
<th>Transect C2</th>
<th>Transect C3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transect B</td>
<td>0.62</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transect C</td>
<td>0.53</td>
<td>0.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transect D</td>
<td>0.55</td>
<td>0.64</td>
<td>0.72</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P2</td>
<td>0.18</td>
<td>0.31</td>
<td>0.25</td>
<td>0.37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transect C1</td>
<td>0.57</td>
<td>0.55</td>
<td>0.58</td>
<td>0.69</td>
<td>0.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transect C2</td>
<td>0.57</td>
<td>0.51</td>
<td>0.58</td>
<td>0.62</td>
<td>0.22</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>Transect C3</td>
<td>0.50</td>
<td>0.51</td>
<td>0.53</td>
<td>0.62</td>
<td>0.32</td>
<td>0.86</td>
<td>0.78</td>
</tr>
<tr>
<td>Transect C4</td>
<td>0.51</td>
<td>0.51</td>
<td>0.54</td>
<td>0.63</td>
<td>0.32</td>
<td>0.87</td>
<td>0.76</td>
</tr>
</tbody>
</table>

In total 60 vascular plant species belonging to 31 families were recorded on polluted and control (non-polluted) transects in the study area. The average number of vascular plant species per segment on polluted transects was 16 while on control transects it was 34. The full list of vascular plant species recorded on polluted and control transects in the study area is shown in Table S1 (Supplementary A).

3.2. Vegetation Data Analysis

Vegetation data was analyzed to determine the differences in characteristics of polluted and control vegetation using the vegan and labdsv packages in R. Median values of vascular plant species richness and diversity were greater for control vegetation than for polluted vegetation. The Mann-Whitney test showed that this difference was significant ($p < 0.05$) which suggest that the presence of TPH in soil adversely affected vegetation characteristics. The summary of the results are shown in Table 6.

Table 6. Result of Mann-Whitney test of differences between polluted and control vegetation.

<table>
<thead>
<tr>
<th>Index</th>
<th>Polluted (n1)</th>
<th>Control (n2)</th>
<th>Confidence Interval (95%)</th>
<th>M-W U</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lower</td>
<td>Upper</td>
</tr>
<tr>
<td>Taxa</td>
<td>14</td>
<td>37</td>
<td>17</td>
<td>26</td>
</tr>
<tr>
<td>Shannon’s</td>
<td>2.64</td>
<td>3.61</td>
<td>0.7</td>
<td>1.47</td>
</tr>
<tr>
<td>Simpson’s</td>
<td>0.93</td>
<td>0.97</td>
<td>0.03</td>
<td>0.09</td>
</tr>
<tr>
<td>Menhinick’s</td>
<td>3.74</td>
<td>6.08</td>
<td>1.75</td>
<td>3.1</td>
</tr>
<tr>
<td>Log_{10}Chao-1</td>
<td>2.02</td>
<td>2.85</td>
<td>0.59</td>
<td>1.22</td>
</tr>
<tr>
<td>Abundance</td>
<td>1.25</td>
<td>1.45</td>
<td>0.01</td>
<td>0.48</td>
</tr>
</tbody>
</table>
Furthermore, Figure 4 illustrates the magnitude of this difference. The species accumulation curve shows that vascular plant species richness in polluted transects was lower than in control transects and that species accumulated more rapidly in control transects.

![Species accumulation curves](image)

**Figure 4.** Species accumulation curves comparing species number on polluted and control transects in the study area. Curves show that vascular plant species richness and rate of accumulation (rate at which new species are observed in segments) was greater on control transects than on polluted transects.

### 3.3. Sensitivity Analysis and Comparison of Hyperion Wavelengths

The maximum, mean and minimum reflectance of the polluted and control transects is shown in Figure 5A. The reflectance in the visible region was higher and the NIR reflectance was lower for the polluted transects than for the control transects. The greatest reflectance difference between the control and polluted transects were observed at the wavelength range 420 nm–470 nm (blue channels) and 620 nm–670 nm (red channels, Table 7). Reflectance at these wavelengths increased significantly (p < 0.05) on the polluted transects which can be attributed to the presence of TPH in the soil. As chlorophyll absorption is highest at the wavelengths of 430 nm, 460 nm, 640 nm, and 660 nm [145], the spectral absorption from chlorophyll in plants was adversely affected by oil pollution. Median reflectance in the visible wavelengths is shown in the boxplots in Figure 5B. They differ significantly (p < 0.05) between the polluted and control transects according to the M-W test.

The results of the sensitivity analysis indicate that reflectances at 440 ± 10 nm (blue channels) and 640 ± 10 nm (red channels) substantially increase in the presence of soil TPH. Conversely, at the wavelength range of 670 nm–900 nm (NIR), the reflectance of the polluted transects decreased slightly but was not significantly different from the NIR reflectance of the control transects (p > 0.05).

Minimum reflectance difference (near zero difference) occurred at 730 nm–830 nm (Figure 5C). The highest reflectance sensitivity to soil TPH (Figure 5D) was observed at wavelengths 440 ± 10 nm (blue channels) and 640 ± 10 nm (red channels).

The M-W results in Table 8 reveal that blue and red reflectance from vegetation on control transects is significantly lower than from polluted transects (p < 0.05). Vegetation reflectance at 426.8 nm (chlorophyll absorption feature in the blue range) is significantly lower (p < 0.05) for the control than for the polluted transects. Reflectance in the NIR wavelengths does not differ significantly between polluted and control transects. This may be due to the presence of TPH in polluted transects.

Earlier studies have reported that oil-contaminated substrates exhibit increased NIR reflectance, which have been attributed to thickness of the crude oil [146,147]. Although hydrocarbon absorption features occur in the 1730–2310 nm wavelengths in the SWIR region [148], in the NIR region the absorption from oil is decreased substantially leading to increased reflectance [146]. With the increased NIR reflectance from both polluted and control vegetation, the characteristics of reflectance in the visible range is differentiating between polluted and non-polluted vegetation.
Figure 5. (A) Reflectance of control (C) transects, \( n = 16 \) and polluted (P) transects \( n = 17 \) in Kporghor spill site measured in November 2015 by the Hyperion EO-1 sensor. The plots displayed are the maximum, mean and minimum reflectance of vegetation on transects at the VNIR region. Inset: reflectance at 427 to 500 nm zoomed in to highlight differences between polluted and control transects. (B) Comparison of median reflectance of specific wavelengths that were sensitive to soil TPH concentration. Boxplots are for polluted and control transects. In (A,B), reflectance values are scaled by 10,000 during atmospheric correction in ENVI 5.3 to remove decimals and reduce computational time. (C) Reflectance difference of vegetation growing on polluted and control transects computed by subtracting the mean reflectance of vegetation on control transects \( n = 16 \) from that of polluted vegetation \( n = 17 \); (D) Reflectance sensitivity to stress or relative change in reflectance computed by dividing the reflectance difference (Figure 5C) by the mean reflectance of the control transects. M-W test results show that the reflectance in the most sensitive wavelengths differ significantly between the polluted and control transects.

Table 7. Hyperion bands and wavelengths with maximum and minimum differences in reflectance and those with least and most sensitivity to TPH-induced stress.

<table>
<thead>
<tr>
<th>Maximum-Difference Bands</th>
<th>9</th>
<th>10</th>
<th>28</th>
<th>8</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength</td>
<td>436.99</td>
<td>447.17</td>
<td>630.32</td>
<td>426.8</td>
<td>650</td>
</tr>
<tr>
<td>Difference</td>
<td>195.96</td>
<td>170.87</td>
<td>123.1</td>
<td>122.48</td>
<td>102.25</td>
</tr>
<tr>
<td>Minimum-Difference Bands</td>
<td>49</td>
<td>46</td>
<td>40</td>
<td>47</td>
<td>38</td>
</tr>
<tr>
<td>Wavelength</td>
<td>844</td>
<td>813.48</td>
<td>752.43</td>
<td>823.65</td>
<td>732</td>
</tr>
<tr>
<td>Difference</td>
<td>1.79</td>
<td>3.49</td>
<td>3.67</td>
<td>4.11</td>
<td>6.22</td>
</tr>
<tr>
<td>Stress-Insensitive Bands</td>
<td>49</td>
<td>46</td>
<td>47</td>
<td>40</td>
<td>42</td>
</tr>
<tr>
<td>Wavelength (nm)</td>
<td>844</td>
<td>813.48</td>
<td>823.65</td>
<td>752.43</td>
<td>772.78</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.00087</td>
<td>0.0018</td>
<td>0.002</td>
<td>0.002</td>
<td>0.0034</td>
</tr>
<tr>
<td>Stress-Sensitive Bands</td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>28</td>
<td>29</td>
</tr>
<tr>
<td>Wavelength (nm)</td>
<td>447.17</td>
<td>436.99</td>
<td>426.8</td>
<td>630.32</td>
<td>640.5</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.77</td>
<td>0.68</td>
<td>0.48</td>
<td>0.19</td>
<td>0.16</td>
</tr>
</tbody>
</table>
Table 8. Median reflectance of polluted and control transects at selected Hyperion wavelengths. P-values were less than 0.05.

<table>
<thead>
<tr>
<th>Wavelength</th>
<th>Polluted (n1)</th>
<th>Control (n2)</th>
<th>Difference (n1–n2)</th>
<th>Confidence Interval (95%) M-W U</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N = 17</td>
<td>N = 16</td>
<td></td>
<td>Lower Limit</td>
</tr>
<tr>
<td>426.8 nm</td>
<td>372.43</td>
<td>238.83</td>
<td>118.56</td>
<td>75.47</td>
</tr>
<tr>
<td>436.99 nm</td>
<td>453.35</td>
<td>270.36</td>
<td>175.62</td>
<td>134.52</td>
</tr>
<tr>
<td>447.17 nm</td>
<td>392.74</td>
<td>221.87</td>
<td>164.2</td>
<td>110.1</td>
</tr>
<tr>
<td>630.32 nm</td>
<td>763.24</td>
<td>640.14</td>
<td>124.82</td>
<td>91.36</td>
</tr>
<tr>
<td>640.5 nm</td>
<td>725.66</td>
<td>624.38</td>
<td>115.71</td>
<td>83.26</td>
</tr>
<tr>
<td>650 nm</td>
<td>740.05</td>
<td>637.8</td>
<td>114.28</td>
<td>59.6</td>
</tr>
</tbody>
</table>

3.4. Modelling the Vascular Plant Species Diversity of the Investigated Transects

3.4.1. Model Calibration Using Training Data

NDVVI variants and NBVIs (see Table 3) were computed and extracted for selected segments in polluted and control transects. The M-W test was applied to test for differences between the median NDVVI of polluted and control transects. The results show significant differences between the median NDVVI for the polluted and control transects (p < 0.05). We infer that TPH concentration affects the vegetation vigour, composition and abundance on the polluted transects.

PLS regression commenced with an initial transformation of the predictor datasets (6 NDVVI variants and 6 NBVIs listed in Table 3) into a smaller set of uncorrelated components with the optimum number selected from R² value associated with each component. A maximum of 5 components was chosen to run the procedure; however, optimum number of components varied for different response variables as shown Table 9. For the NDVVI dataset, only 1–2 components, which best explained the variation in the dataset were selected for the regression analysis. For the NBVIs, 1–4 components were used in the models. Cross-validation was performed by a leave-two-out procedure on the components before selecting the optimal number. The NDVVI-based PLS model had larger R-squared (R²) values than the NBVI-based PLS model. Additionally, prediction error sum of squares (PRESS) is smaller for the NDVVI predictors than for the NBVIs. This confirms that the PLS model of NDVVI variants have greater predictive ability than that of traditional NBVIs. The results from model calibration are summarized in Table 9.

Table 9. Calibration parameters of NDVVI and NBVI-based models used in the PLS and NPM regression methods.

<table>
<thead>
<tr>
<th>Response</th>
<th>Components Selected</th>
<th>R²</th>
<th>PRESS</th>
<th>F</th>
<th>p</th>
<th>R²</th>
<th>RSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDVVIs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shannon’s</td>
<td>2</td>
<td>0.67</td>
<td>12.3</td>
<td>17.56</td>
<td>&lt;0.05</td>
<td>0.71</td>
<td>0.61</td>
</tr>
<tr>
<td>Simpson’s</td>
<td>2</td>
<td>0.66</td>
<td>1.11</td>
<td>16.25</td>
<td>&lt;0.05</td>
<td>0.69</td>
<td>0.17</td>
</tr>
<tr>
<td>Menhinick’s</td>
<td>1</td>
<td>0.54</td>
<td>44</td>
<td>20.82</td>
<td>&lt;0.05</td>
<td>0.61</td>
<td>1.23</td>
</tr>
<tr>
<td>Log(Chao-1)</td>
<td>2</td>
<td>0.6</td>
<td>8.69</td>
<td>12.75</td>
<td>&lt;0.05</td>
<td>0.69</td>
<td>0.49</td>
</tr>
<tr>
<td>Canopy Chlorophyll</td>
<td>2</td>
<td>0.56</td>
<td>2181</td>
<td>10.92</td>
<td>&lt;0.05</td>
<td>0.58</td>
<td>9.08</td>
</tr>
<tr>
<td>NBVIs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shannon’s</td>
<td>3</td>
<td>0.39</td>
<td>25.28</td>
<td>3.38</td>
<td>&lt;0.05</td>
<td>0.49</td>
<td>0.82</td>
</tr>
<tr>
<td>Simpson’s</td>
<td>1</td>
<td>0.3</td>
<td>1.71</td>
<td>8.23</td>
<td>&lt;0.05</td>
<td>0.46</td>
<td>0.23</td>
</tr>
<tr>
<td>Menhinick’s</td>
<td>4</td>
<td>0.48</td>
<td>67.78</td>
<td>3.54</td>
<td>&lt;0.05</td>
<td>0.58</td>
<td>1.31</td>
</tr>
<tr>
<td>Log(Chao-1)</td>
<td>3</td>
<td>0.50</td>
<td>11.74</td>
<td>5.35</td>
<td>&lt;0.05</td>
<td>0.55</td>
<td>0.58</td>
</tr>
<tr>
<td>Canopy Chlorophyll</td>
<td>1</td>
<td>0.11</td>
<td>4355</td>
<td>2.12</td>
<td>ns</td>
<td>0.59</td>
<td>8.89</td>
</tr>
</tbody>
</table>

ns = not significant.
The significance of the relationship between the predictors (NDVVI variants and NBVIs) and the response (diversity indices) was analyzed using the F-statistic. The results show that each diversity index was statistically related with the selected NDVVI components \((R^2 > 0.5, p < 0.05)\). Similarly, diversity indices also significantly regressed with the NBVI components; however as stated earlier, the \(R^2\) values were much lower \((\leq 0.5, p < 0.05)\) except for the Chao-1 index (Table 9). The significant relationship observed between satellite-based indices (NDVVI and NBVI) and field measured diversity indices is in line with previous results such as [149] who reported \(R^2\) as high as 0.87 between NDVI and plant richness; [137] who reported \(R^2\) values between 0.32 and 0.72 for NDVI and Shannon’s diversity; and [95] who reported \(R^2\) values of 0.51 to 0.83 for first order hyperspectral indices and diversity indices including Shannon-Weiner, Pielou, Simpson, Margalef and Gleason. The scatterplots of observed versus predicted diversity indices are shown in Figure 6.

![Figure 6. Observed versus predicted diversity indices using PLS NDVVI-based regression model.](image)

There appears to be a linear relationship between both sets of data leading to the high \(R^2\) values. This result is consistent with results from previous studies predicting species diversity from vegetation indices.

From the scatterplots in Figure 6, it is apparent that the NDVVI variants performance in estimating species diversity is comparable to results reported in other studies. The mechanism explaining the relationship between satellite derived indices and field measured diversity indices is not yet well understood; however, judging from the results of this study, we infer that vegetation biochemical parameters, particularly those strongly influenced by variations in pigment absorption at wavelengths sensitive to soil TPH are important drivers of this relationship. The scatterplot of residuals versus predicted diversity index from model calibration using the training data are shown in Figure 7. The plots suggest that the PLS model provide a good fit for the data. The residuals generally satisfy the goodness of fit requirements with randomness, homoscedasticity and linearity.
Similarly, the NDVVI-based NPM model has much smaller error values than the NBVI-based NPM model. The NDVVI-based model perform better during calibration with higher $R^2$ values (0.61–0.71 at calibration stage) compared to NBVI-based models with $R^2 < 0.59$. Residual standard error (RSE) values from model calibration are smaller for the NDVVI NPM model and larger for the NBVI model.

3.4.2. Model Validation Using Test Data

Validation of the trained models was performed using the test data ($n = 13$, polluted = 7, control = 6). The predictive capability of the spectral metrics is inferred from predicted $R^2$, RSE, root mean square error (RMSE), bias and residual analysis of the different models. The performance of the NDVVI-based models was uniform across both PLS and NPM model types. Analysis of residuals following model validation also affirms the superiority of NDVVI for estimating vascular plant species diversity over NBVI. The F-statistics, $p$, $R^2$, RMSE and Bias are summarized in Table 9 for all models. The NDVVI-based models (Models 1A and 2A) had the highest $R^2$ as well as lowest RSE values. Although non-parametric models are generally not as powerful as parametric ones, the spectral NDVVI metrics derived from TPH-sensitive Hyperion wavelengths consistently outperformed the traditional NBVIs as estimators of species diversity in all the models. Poor estimates for Simpson’s diversity index are obtained from NDVVI and NBVIs-based models, particularly using the NPM regression method, although the error values were very low. From the results, the best index for estimating the Menhinick’s Richness index is the NDVVI variants. The $R^2$ and RMSE values for NDVVI-based PLS model are 0.57 and 1.13 respectively, while for NBVIs-based PLS model, they are 0.37 and 1.58 respectively. Results of the model validation using the test data are summarized in Table 10.

Generally, all the models underestimated the response variables (Shannon’s, Simpson’s, Menhinick’s, Chao-1, and Canopy Chlorophyll) as evident in the negative bias scores, although the biases were greater for the NBVI-based models. With respect to monitoring biodiversity, this effect may be an advantage and reduces the risk of overestimating the vascular plant species diversity of an oil affected location or a protected area.

The NDVVI-based model predictions were over 50% accurate for Shannon’s and Menhinick’s diversity indices, and less than 50% for Simpson’s and Chao-1’s indices. The best predictions were for Menhinick’s index as illustrated in the closeness of the fitted lines to the 1:1 lines in all four models shown in Figure 8 for PLS models and Figure 9 for NPM models. In contrast, Simpson’s index was
the least accurate as the plots showed little or no relationship between the predicted and observed field measurements.

**Table 10.** Results of the species diversity and canopy chlorophyll estimation of investigated transects using two different models for each set of predictors. Models 1 and 2 are the partial least square (PLS) and non-parametric (NPM) regression models respectively. Letters A and B indicate the set of predictors (spectral metrics) used in each model, A = NDVVI and B = NBVI; \( n = 13, \text{df} = 12 \); ns = not significant.

<table>
<thead>
<tr>
<th>Response Variable</th>
<th>Model</th>
<th>( F )</th>
<th>( p )</th>
<th>( R^2 )</th>
<th>RSE</th>
<th>RMSE</th>
<th>Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shannon’s Diversity Index</td>
<td>1A</td>
<td>12.82</td>
<td>&lt;0.05</td>
<td>0.54</td>
<td>0.51</td>
<td>0.69</td>
<td>-11.4</td>
</tr>
<tr>
<td></td>
<td>1B</td>
<td>1.77</td>
<td>ns</td>
<td>0.14</td>
<td>0.69</td>
<td>0.9</td>
<td>-16.2</td>
</tr>
<tr>
<td></td>
<td>2A</td>
<td>13.08</td>
<td>&lt;0.05</td>
<td>0.54</td>
<td>0.5</td>
<td>0.5</td>
<td>-6.2</td>
</tr>
<tr>
<td></td>
<td>2B</td>
<td>2.67</td>
<td>ns</td>
<td>0.2</td>
<td>0.67</td>
<td>0.94</td>
<td>-17.2</td>
</tr>
<tr>
<td>Simpson’s Diversity Index</td>
<td>1A</td>
<td>6.66</td>
<td>&lt;0.05</td>
<td>0.38</td>
<td>0.05</td>
<td>0.24</td>
<td>-15.9</td>
</tr>
<tr>
<td></td>
<td>1B</td>
<td>0.11</td>
<td>ns</td>
<td>0.01</td>
<td>0.07</td>
<td>0.22</td>
<td>-18.1</td>
</tr>
<tr>
<td></td>
<td>2A</td>
<td>1.163</td>
<td>0.3</td>
<td>0.1</td>
<td>0.07</td>
<td>0.14</td>
<td>-9.2</td>
</tr>
<tr>
<td></td>
<td>2B</td>
<td>0.09</td>
<td>ns</td>
<td>0.01</td>
<td>0.07</td>
<td>0.21</td>
<td>-14.9</td>
</tr>
<tr>
<td>Menhinick’s Richness Index</td>
<td>1A</td>
<td>14.32</td>
<td>&lt;0.05</td>
<td>0.57</td>
<td>1.15</td>
<td>1.13</td>
<td>-7.5</td>
</tr>
<tr>
<td></td>
<td>1B</td>
<td>6.37</td>
<td>&lt;0.05</td>
<td>0.37</td>
<td>1.38</td>
<td>1.58</td>
<td>-21.6</td>
</tr>
<tr>
<td></td>
<td>2A</td>
<td>5.35</td>
<td>&lt;0.05</td>
<td>0.33</td>
<td>1.42</td>
<td>1.32</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2B</td>
<td>7.4</td>
<td>&lt;0.05</td>
<td>0.4</td>
<td>1.34</td>
<td>1.31</td>
<td>-10.2</td>
</tr>
<tr>
<td>Log (Chao-1)</td>
<td>1A</td>
<td>8.55</td>
<td>&lt;0.05</td>
<td>0.44</td>
<td>0.24</td>
<td>0.57</td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td>1B</td>
<td>2.12</td>
<td>ns</td>
<td>0.16</td>
<td>0.3</td>
<td>0.58</td>
<td>-1.1</td>
</tr>
<tr>
<td></td>
<td>2A</td>
<td>10.16</td>
<td>&lt;0.05</td>
<td>0.48</td>
<td>0.23</td>
<td>0.56</td>
<td>2.1</td>
</tr>
<tr>
<td></td>
<td>2B</td>
<td>1.93</td>
<td>ns</td>
<td>0.15</td>
<td>0.3</td>
<td>0.51</td>
<td>3.2</td>
</tr>
<tr>
<td>Canopy Chlorophyll Content</td>
<td>1A</td>
<td>7.89</td>
<td>&lt;0.05</td>
<td>0.42</td>
<td>7.85</td>
<td>7.87</td>
<td>6.7</td>
</tr>
<tr>
<td></td>
<td>1B</td>
<td>1.14</td>
<td>ns</td>
<td>0.09</td>
<td>9.79</td>
<td>9.29</td>
<td>4.7</td>
</tr>
<tr>
<td></td>
<td>2A</td>
<td>10.49</td>
<td>&lt;0.05</td>
<td>0.49</td>
<td>7.36</td>
<td>7.59</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td>2B</td>
<td>1.64</td>
<td>ns</td>
<td>0.13</td>
<td>9.6</td>
<td>13.49</td>
<td>9.9</td>
</tr>
</tbody>
</table>

**Figure 8.** Observed versus predicted plots for the various PLS models. For each species diversity index, scatterplots of observed values versus the NDVVI variants (blue) and NBVI (red) predicted values are shown (\( n = 13 \)). The regression equations are also shown with the \( R^2 \) values, \( y_1 \) = response to NDVVI variants, \( y_2 \) = response to NBVI. The line of best fit for each model is plotted to compare with the 1:1 line (in black).
Figure 9. Observed versus predicted plots for the NPM models. For each species diversity index, scatterplots of observed values versus the NDVVI variants (blue) and NBVIs (red) predicted values are shown ($n = 13$). The regression equations are also shown with the $R^2$ values, $y_1 =$ response to NDVVI variants, $y_2 =$ response to NBVIs. The line of best fit for each model is plotted to compare with the 1:1 line (in black).

All models clearly distinguished between polluted and control transects with the diversity estimates; however, the NDVVI-based models performed better. The residual versus predicted scatterplot in Figure 10 show that the NDVVI-based model is a good fit for Shannon’s index and the SPAD chlorophyll estimates. However, this was absent for the other indices.

Figure 10. Scatterplots of residual versus predicted values of NDVVI-based model. Predicted values are from the NPM regression using test data ($n = 13$). The charts clearly show that the model was a good fit for Shannon’s diversity index and SPAD chlorophyll estimates.
Using the model equations from the NDVVI PLS model, spatial maps of vascular plants species diversity indices were created for the investigated area. Figure 11 shows the maps for the Shannon’s, Simpson’s, Menhinick’s and log transformed Chao-1’s indices as well as the canopy chlorophyll content. A quick look at the images for Shannon’s or Simpson’s diversity and the canopy chlorophyll content shows that pixels with high diversity index were also high in canopy chlorophyll.

![Figure 11](image_url)

**Figure 11.** Spatial maps of vascular plant species diversity estimated from NDVVI PLS model. Location of control and polluted transects on the maps correspond with the estimated diversity index and chlorophyll content. From the images, polluted transects are seen to have low diversity and canopy chlorophyll values while control transects have high diversity and canopy chlorophyll values. This further highlights the relationship between vegetation productivity indicated by canopy chlorophyll content and vascular plant species diversity.

### 3.5. Model Implementation and Evaluation Using Random Pixels

The new dataset of derived spectral metrics (preddata) was used as predictors in the NPM model in order to estimate the Shannon’s, Simpson’s, Menhinick’s and Chao1 index values for the predsites. The estimations were done separately for each variable. Average species diversity index estimated for each land cover type visible from high-resolution image available in google earth is shown in Table 11. Expectedly higher diversity index values were predicted for forests and mixed vegetation, while swamps and waterbodies had lower diversity prediction. Moderate diversity indices were predicted for farmlands.

NDVI values computed from both Landsat and Sentinel 2A images (Figure 12) were extracted for the predsites. The NDVI values were generally low for the different land cover types compared to the NDVVI values. NDVVI values for forested pixels ranged from 0.53 to 0.94 while NDVI values were 0.2 and 0.12 respectively for L8-NDVI and S2A-NDVI. Similarly, higher NDVVI values than NDVI values were extracted from pixels categorized as farmland and mixed. Despite the large margin between NDVVI and NDVI values, the pattern of vascular plant species diversity estimation was similar. As evident in Table 11, the higher the index value, the higher the estimated species diversity value and vice versa.
Table 11. Average diversity values predicted for randomly selected pixels according to observed land cover type. N = number of 30m pixels in each class, L8-NDVI = NDVI derived from Landsat 8 image and S2A-NDVI = NDVI derived from Sentinel 2A image. Due to its higher spatial resolution, average NDVI values was calculated using a $2 	imes 2$ pixel window from the S2A-NDVI.

<table>
<thead>
<tr>
<th>Land Cover Type</th>
<th>Farmland</th>
<th>Forested</th>
<th>Mixed</th>
<th>Swamp</th>
<th>Waterbody</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>7</td>
<td>5</td>
<td>10</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>L8-NDVI</td>
<td>0.17</td>
<td>0.2</td>
<td>0.18</td>
<td>0.13</td>
<td>0.11</td>
</tr>
<tr>
<td>S2A-NDVI</td>
<td>0.11</td>
<td>0.12</td>
<td>0.11</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>NDVI$_{844,447}$</td>
<td>0.48</td>
<td>0.57</td>
<td>0.49</td>
<td>0.29</td>
<td>0.27</td>
</tr>
<tr>
<td>NDVI$_{814,437}$</td>
<td>0.73</td>
<td>0.83</td>
<td>0.76</td>
<td>0.61</td>
<td>0.57</td>
</tr>
<tr>
<td>NDVI$_{824,427}$</td>
<td>0.5</td>
<td>0.58</td>
<td>0.51</td>
<td>0.3</td>
<td>0.28</td>
</tr>
<tr>
<td>NDVI$_{752,630}$</td>
<td>0.44</td>
<td>0.53</td>
<td>0.46</td>
<td>0.25</td>
<td>0.23</td>
</tr>
<tr>
<td>NDVI$_{772,641}$</td>
<td>0.71</td>
<td>0.80</td>
<td>0.73</td>
<td>0.56</td>
<td>0.53</td>
</tr>
<tr>
<td>NDVI$_{844,630}$</td>
<td>0.85</td>
<td>0.94</td>
<td>0.88</td>
<td>0.73</td>
<td>0.7</td>
</tr>
<tr>
<td>Shannon’s</td>
<td>2.59</td>
<td>3.42</td>
<td>2.77</td>
<td>0.88</td>
<td>0.94</td>
</tr>
<tr>
<td>Simpson’s</td>
<td>0.82</td>
<td>0.95</td>
<td>0.87</td>
<td>0.25</td>
<td>0.3</td>
</tr>
<tr>
<td>Menhinick’s</td>
<td>3.64</td>
<td>5.34</td>
<td>4.13</td>
<td>1.41</td>
<td>1.16</td>
</tr>
<tr>
<td>LogChao-1</td>
<td>2.16</td>
<td>2.79</td>
<td>2.4</td>
<td>1</td>
<td>1.06</td>
</tr>
<tr>
<td>Canopy Chlorophyll</td>
<td>56.12</td>
<td>64.8</td>
<td>56.46</td>
<td>39.57</td>
<td>34.09</td>
</tr>
</tbody>
</table>

As an additional step in evaluating the performance of the NDVVI-based model, NDVI was used as surrogates for the vascular plant species diversity index. Since NDVI has been shown to correlate strongly with species diversity, it is expected to exhibit similar behaviour with the predicted vascular species diversity indices if the predictions were correct. Due to the higher spatial resolution of the Sentinel 2A image, average NDVI values was computed for each segment using a $2 	imes 2$ pixel window. The result of the correlation analysis in Table 12 suggests that the estimated values have a strong linear relationship with NDVI values from both images. The correlation coefficients ranged from 0.73 to 0.85 for the diversity indices.

Figure 12. NDVI computed from (A) Landsat 8 OLI and (B) Sentinel 2A images acquired by the sensors on the 4 January 2016 and 22 December 2015 respectively from the study area.
Figure 12. NDVI computed from (A) Landsat 8 OLI and (B) Sentinel 2A images acquired by the sensors on the 4 January 2016 and 22 December 2015 respectively from the study area.

Figure 13. A high-resolution Digital Globe 2006 true color image of the study area extracted from Google Earth showing the location of predsites. This image was selected because it depicted the land cover types in the study area better than more recent high-resolution images. From the estimated Shannon’s diversity index shown next to the predsites, it is obvious that most of the predictions correspond with the visible land cover type.

Visual evaluation of high resolution Google Earth imagery (Figure 13) shows that most predicted values correspond with the land cover type on the ground surface. For instance, the predsites that were located on swamps and water bodies had low estimated values for vascular plant species diversity. However, the location of the predsite P2 with predicted Shannon diversity index of 2.68 appears to be bare soil in this image (acquired by Digital Globe in December 2006), the most current image acquired in January 2016 (not used due to cloud obstruction) shows the presence of vegetation regrowth at the location. This may explain the predicted high diversity values for the pixel.

Table 12. Pearson’s correlation coefficients of NDVI and estimated species diversity indices for predsites. All the results are significant ($p < 0.05$).

<table>
<thead>
<tr>
<th>Diversity Index</th>
<th>NDVI\textsubscript{Landsat-8}</th>
<th>NDVI\textsubscript{Sentinel-2A}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shannon’s</td>
<td>0.77</td>
<td>0.78</td>
</tr>
<tr>
<td>Simpson’s</td>
<td>0.73</td>
<td>0.75</td>
</tr>
<tr>
<td>Menhinick’s</td>
<td>0.78</td>
<td>0.79</td>
</tr>
<tr>
<td>Chao-1</td>
<td>0.84</td>
<td>0.84</td>
</tr>
</tbody>
</table>
4. Discussion

TPH pollution in soil amplifies the difference between the reflectance of polluted and control transects in wavelengths associated with chlorophyll absorption in the blue (440 ± 10 nm) and red (640 ± 10 nm) spectral channels. These wavelengths are most sensitive to TPH concentration in the soil. Since chlorophyll absorption occurs in the wavelength range of 430 nm–460 nm and 650 nm–700 nm [133,150], the presence of TPH in the soil affects the absorption of chlorophyll in vegetation, which is also supported by occurrence of the most sensitive wavelength at 447.17 nm (sensitivity = 0.77). Previously, [133] found that the reflectance at 420 ± 5 nm varies little with stress in plants, and they reported increased sensitivity to plant stress for reflectances at 600 nm and 695 nm.

Although there was a significant difference in the NIR reflectance (700 nm–900 nm) of polluted and control vegetation, the sensitivity analysis using mean reflectance showed that this region was least sensitive to TPH-induced stress. Other researchers reported similar patterns in the NIR reflectance of stressed vegetation. For instance, [133] reported that at 730 nm, reflectance in stressed plants did not significantly change, while [151] found that NIR reflectance did not vary between healthy leaves and stressed leaves. They attributed this phenomenon to the increase in the size and length of the assemblages in the spongy parenchyma in stressed leaves. Moreover, [147,148,152] analyzed polluted substrates and attributed the increased NIR reflectance in polluted vegetation to the presence of hydrocarbons.

Other factors may have contributed to this response. Firstly, as suggested by [38,130,153] there may have been increased presence of invasive species, which are tolerant to hydrocarbons. Secondly, it may also be that the plant assemblages (cell walls, mesophyll cells and intercellular spaces) responsible for NIR reflectance in vegetation were yet to succumb to the stress caused by TPH in the soil. This is most likely the case along polluted transects as soil TPH concentration decreased, thereby delaying the onset of physiological damage in plants tissues.

Results of the sensitivity analysis differentiated the response of chlorophyll pigments a (Chl-a) and b (Chl-b) to TPH concentration in the soil. The most sensitive wavelength in the blue range occurred at the Chl-a absorption maximum (447.17 nm) and in the red range at the Chl-b absorption maximum (630.32 nm). Although these wavelengths showed sensitivity to soil TPH concentrations, Chl-a absorption in the blue range was most affected as the reflectance difference between polluted and control vegetation at that wavelength was up to 300%. In contrast, [98] found that the wavelength around 650 nm was more sensitive to chlorophyll content in vegetation than the chlorophyll absorption features in the blue range.

Since Chl-a is the principal pigment for photosynthesis, this may explain the severe effect associated with oil pollution in plants. Arellano et al. [114,154,155] reported that increasing crude oil contamination caused a significant decrease in the chlorophyll content, which sometimes led to plant mortality.

Several researchers have propounded theories on how TPH influences chlorophyll content in affected plants. Investigations of crude oil effect on plant anatomy such as [155–157] discovered structural deformations in the form of thickening of the epicuticular region, compression of the palisade and spongy parenchyma, compression of the vascular bundles, reduction of intercellular air spaces, distortion and reduction of the stomata. These changes generally inhibit chlorophyll synthesis thereby affecting plant growth and productivity [155]. Considering the response of the Chl-a absorption features to soil TPH concentration, these physiological effects have been linked to oil pollution in various environments, causing a decrease in Chl-a production and consequently vegetation growth, health and productivity.

The modelling results of vascular plants species diversity indices provide strong evidence of a relationship with narrowband chlorophyll-related vegetation indices. This relationship is stronger when hyperspectral wavelengths sensitive to soil TPH are used in calculating the vegetation index as is the case with the NDVVI. This further emphasizes the need for incorporating the new index in biodiversity monitoring and conservation schemes.
NDVVI is indicative of chlorophyll content and is hence an important plant biochemical parameter for vegetation productivity and health [131,155,158]. Not only did NDVVI significantly differ between polluted and control transects, it also strongly correlated with the vascular plants species diversity. This result is consistent with [159] who reported changes in vegetation pattern in polluted fields and [160] who found lower diversity indices for contaminated sites than for uncontaminated sites. Additionally, reports from [14,161,162] were in agreement with our results. Hence, the low NDVVI values found over polluted transects can be attributed to reduced species composition, reduced abundance and deteriorating health of the vegetation. Since vegetation vigour is characterized by vegetation productivity and health [63], the presence of TPH in soils adversely affected both traits.

As stated earlier, the need to clarify the mechanism defining relationships between vegetation reflectance and species diversity remains primal, several researchers have linked it to variations in vegetation biochemical parameters. For instance, [38] following their study on airborne spectranomics reported that plant species have unique chemical fingerprints which correspond with spectral and species diversity. The chemical fingerprints are exhibited via differences in photosynthetic and photo-protective pigments, water and leaf structure and remotely measurable. Similarly, [34] observed that interspecific variability in pigment (chlorophyll, anthocyanins, and carotenoids) levels in plants contributed in species differentiation using spectral metrics. Additionally, [163] successfully classified seven tree species using hyperspectral metrics derived from wavelengths sensitive to vegetation chemistry and structure.

In view of these, we infer that the NDVVI variants superior performance in estimating species diversity is attributed to the selection of wavelengths sensitive to soil TPH which is known to affect vegetation chemistry. Thus, making it the ideal index for use in known crude oil polluted regions. The superior performance of the NDVVI variants in estimating vascular plants species diversity may be attributed to the selection of particular wavelengths that strongly responded to changes in vegetation pigments due to oil pollution. This procedure not only extracted relevant wavelengths from hundreds of hyperspectral wavelengths that are potentially redundant, but also reduced the presence of noise from the data. Jacquemond et al. [38,41] stated that plant spectra may contain additional information unrelated to pigment concentration.

Previous studies have shown that changes in vegetation productivity and species diversity are common symptoms of ecosystem stress. Rapport, Regier and Hutchinson [164] reported that environmental stress including oil pollution induces changes “in the size of dominant species, species diversity and a shift in species dominance to opportunistic shorter-lived forms”. Accordingly, NDVVI-based models identify low species diversity indices for polluted transects and higher indices for unpolluted transects.

High NDVVI values extracted from predsites following implementation of the best performing model contrasted with the very low NDVI values and suggest that the new index is more capable of detecting vegetation presence than the NDVI in oil polluted regions. Due to the adverse effect of oil pollution on vegetation (reduced growth [165–167] and increased mortality [168–170], the NDVVI designed to have maximum sensitivity to soil TPH, appears to be a more suitable index to measure vegetation characteristics because of its ability to detect even sparse areas of vegetative growth/presence. Furthermore, the NDVVI variants successfully predicted the diversity indices for the randomly selected sites from the satellite image of the study area. The low index values predicted for the swamps and water bodies are consistent with expectations. According to [171], the waterways of the Niger delta harbour invasive species, particularly the water hyacinth (Eichornia crassipes (Mart.) Solms). In their work, [172] asserted that invasive species adversely affect species richness, diversity and composition of invaded habitats. Hence, it is not surprising that the diversity indices are low for those pixels even though there is abundance of green vegetation.

As vascular plants are the common biodiversity indicators in an ecosystem, any condition that leads to drastic changes in vegetation characteristics (such as oil pollution) is bound to interfere with the composition, structure and functions of the entire ecosystem. Due to its ability to detect oil-induced
stress in vegetation, the new NDVVI has potential as a spectral metric for measuring changes in ecosystem functions, an essential biodiversity variable as well as providing information about the condition and vulnerability of ecosystems, a biodiversity indicator.

When incorporated in a temporal analysis, the NDVVI can reveal the extent of habitat degradation resulting from oil pollution. Since the variants are derived from remote sensing data, their application is standardized, scalable and repeatable making it a very useful tool to achieve some of the Aichi 2020 targets set by the United Nations Convention on Biological Diversity (CBD) [173].

At local or regional scales, routine application of the NDVVI over areas with oil installations will facilitate detection of oil seepages, unreported spills and illegal bunkering activities. In essence, the index will facilitate effective biodiversity monitoring and conservation by providing decision makers with relevant information on areas of high or low biodiversity. This information will ensure the efficient management of meagre resources by reducing the frequency and scale of cost intensive field surveys.

5. Conclusions

A new index, known as the NDVVI, is introduced. The index and its variants are better at discriminating between oil-polluted and natural vegetation and are more strongly related to vascular plant species diversity indices than traditional NBVIs. NDVVIs have potential as an essential biodiversity variable (EBV) for monitoring biodiversity and offer better solutions than NBVIs for assessing oil-polluted vegetation.

The performance of the NDVVI in this study provides evidence of the deleterious effect of oil pollution on the chlorophyll systems in vegetation. Given that vegetation productivity is intricately linked to plant species richness and diversity, these effects potentially extend to the biodiversity of the area.

The adverse effect of oil pollution on ecosystem function, structure and composition is evident in the NDVVI values over the polluted and control transects. The differences between these transects were significant. Changes in vegetation characteristics observed in the field data were manifest in spectral reflectance signals and were detected by the Hyperion sensor.

Supplementary Materials: The following are available online at http://www.mdpi.com/2072-4292/10/6/897/s1. Supplementary A: Kporghor Vegetation Survey Data; Supplementary B: Kporghor Soil Properties Data; Supplementary C: SPAD Chlorophyll Data.

Author Contributions: Conceptualization, N.N.O.; Formal analysis, N.N.O.; Funding acquisition, N.N.O.; H.B.; Methodology, N.N.O., H.B. and J.-C.B.; Project administration, H.B.; Resources, H.B.; Supervision, H.B. and J.-C.B.; Validation, H.B., J.-C.B.; Writing—original draft, N.N.O.; Writing—review & editing, N.N.O., H.B. and J.-C.B.

Acknowledgments: This study was financed partly by the Tertiary Education Trust Fund (TETFUND), Michael Okpara University of Agriculture, Umudike, Nigeria, the National Center for Earth Observation (NCEO) and the Center for Landscape and Climate Research (CLCR), University of Leicester, United Kingdom. The authors are grateful to the traditional ruler of Kporghor community and the members of the Youth leadership for their support during the field campaign.

Conflicts of Interest: The authors declare no conflict of interest.

References


42. Vrieling, A. Satellite remote sensing for water erosion assessment: A review. CATENA 2006, 65, 2–18. [CrossRef]


59. Hector, A. The Effect of Diversity on Productivity: Detecting the Role of Species Complementarity. Oikos 1998, 82, 597–599. [CrossRef]


69. Galidaki, G.; Gitas, I. Mediterranean Forest Species Mapping Using Classification of Hyperion Imagery. Geocarto Int. 2015, 30, 48–61. [CrossRef]

73. Agbagwa, I.O.; Ndukwu, B.C. Oil and Gas Pipeline Construction-Induced Forest Fragmentation and Biodiversity Loss in the Niger Delta, Nigeria. Nat. Resour. 2014, 5, 698–718. [CrossRef]
88. Sattout, E.; Caligari, P.D.S. Forest Biodiversity Assessment in Relic Ecosystem: Monitoring and Management Practice Implications. Diversity 2011, 3, 531–546. [CrossRef]


101. Dadon, A.; Ben-Dor, E.; Karnieli, A. Use of Derivative Calculations and Minimum Noise Fraction Transform for Detecting and Correcting the Spectral Curvature Effect (Smile) in Hyperion Images. TGRS 2010, 48, 2603–2612. [CrossRef]


114. Arelanno, P.; Tansley, K.; Balzter, H.; Boyd, D.S. Detecting the effects of hydrocarbon pollution in the Amazon forest using hyperspectral satellite images. Environ. Pollut. 2015, 205, 225–239. [CrossRef] [PubMed]


144. Kaimaris, D.; Georgoula, O.; Patias, P.; Stylianidis, E. Comparative analysis on the archaeological content of imagery from Google Earth. *J. Cult. Heritage* 2011, 12, 263–269. [CrossRef]


157. Lopes, A.; da Rosa-Osman, S.M.; Piedade, M.T.F. Effects of crude oil on survival, morphology, and anatomy of two aquatic macrophytes from the Amazon floodplains. *Hydrobiologia* 2009, 636, 295–305. [CrossRef]


159. Noomen, M.F.; van der Werff, H.M.A.; van der Meer, F.D. Spectral and spatial indicators of botanical changes caused by long-term hydrocarbon seepage. *Ecol. Inform.* 2012, 8, 55–64. [CrossRef]


