UNIVERSITY OF LEICESTER

Detection and Mapping of Terrestrial Oil Spill Impact Using Remote Sensing Data in Combination with Machine Learning Methods. A Case Site within the Niger Delta Region of Nigeria.

Thesis Submitted for the Degree of

Doctor of Philosophy

at the University of Leicester

By

Mohammed Shuaibu Jamiu Ozigis (M.Sc.)

2019

Abstract

In this novel research, the use of optical and SAR images for separate and joint investigations were explored, with the aid of machine learning classifiers for detecting hydrocarbon spill impact on cropland, grassland and dense forested vegetation types. Optical image spectral bands across the Visible, Near-Infrared and Shortwave-Infrared spectrum, various vegetation health indices (including NDVI, NDWI, LAI and SAVI) and SAR derived variables (including backscatter, coherence and textural variables) were used to detect and map oil spill sites within cropland, grassland and TCA vegetation types. Results generally showed that the integration of multi-frequency L, C and X band SAR in the wet (summer) season yielded the best overall classification accuracy in discrimination of polluted and oil-free vegetation types. An overall accuracy (OA) of 82.3%, 66.67% and 70.93% were obtained for Cropland, Grassland and Tree Cover Areas (TCA) vegetation types, respectively. The accuracies recorded were significantly better ($P>0.05$) than when the spill was classified using optical imagery only and when integrated optical and SAR image variables are classified. These results were further corroborated by the multi-temporal Sentinel – 1 backscatter analysis, which showed that mean backscatter difference between polluted and oil-free cropland and grassland vegetation are significantly different ($P>0.05$) in the wet season than in the dry season. Furthermore, the new fuzzy forest method used for multi-frequency (C and X band) SAR and Optical variable reduction was able to achieve a good result in addressing high dimensionality in cropland and grassland vegetation. This research demonstrates that SAR based monitoring of petroleum hydrocarbon impacts on vegetation is feasible and has high potential for oil pipeline monitoring and facility management. The research also presented a new paradigm into terrestrial oil spill detection, which largely can replace the use of optical data.

Dedication

I dedicate this thesis to Allah the almighty without whom all effort end in futility. To my wife Rahmatullah Suleiman, Daughter Aneesa, Son's Anwar and Abdullah (Imam), to my parents (Mr S B Ozigis and Mrs H A Ozigis), to my sister (Mrs Rabiat Ozigis Ibrahim), brothers (Muftau Ozigis, Abdulmumeen O, Faruq O, Bashir O and Habib O), to my extended family members and all well-wishers.

Acknowledgement

This research is funded by the Petroleum Technology Development Fund under scholarship number PTDF/ED/TR/767. Data for Sentinel – 1 SAR and Sentinel – 2 optical © ESA (2015), Data for TanDEM – X © DLR (2017), Data for Cosmo Skymed © ALR, ALOS PALSAR 2 © JAXA (2018) and Landsat – 8 OLI © USGS. To process some of the data, the ALICE High Performance Computing Facility at the University of Leicester was used.

I am most grateful to Almighty Allah for the gift of life, giving me the patience and making it possible for me to complete this study. My first and most sincere gratitude goes to my supervisors Dr Jorg, D. Kaduk and Dr Claire, H. Jarvis for their invaluable guidance, comments, observations and inputs, which have greatly shaped the direction of this research. I also owe a great thanks to my reviewers Professor Kevin Tansey and Dr Kirsten Barret for their critical evaluation, comments and suggestions during the three annual reviews. I would also like to say a big thank you to Professor Heiko Balzter of the Centre for Landscape and Climate Research (CLCR) for the useful feedback, comments, support and observations given in the course of this research. I would like to specially thank Professor Kevin Tansey, Dr Polyanna Bispo and Dr James Wheeler for the several trainings in the area of SAR processing, using R software and using the ALICE HPC computing facility, which all proved very vital in delivering the objectives of my research.

I am also grateful to the management of the National Space Research and Development Agency for the financial support offered through a study leave with pay arrangement. My sincere gratitude also goes out to Dr Ogochukwu Amukali of the Niger Delta University, Mr Goodluck Nakaima and Dr Bolaji Babatunde of the Department of Animal and Environmental Biology, University of Port Harcourt for their support during the course of the various fieldworks conducted in the Niger Delta. I would also like to appreciate Bikhtiyar Ameen, Mohammed Alhedyan, Akram Abdulla, Ahmed Al-Budeiri, Pengyuan Liu, Genna Tyrell, Chris Martin and others too numerous to mention for their friendship and support throughout my period of study. I greatly appreciate my family for their patience and support through my study period, especially my wife – Rahmatullah and kids Anwar, Aneesa and Abdullah.

Table of Content

List of Figures

List of Tables

List of Acronyms

ADW	Water Absorption
ALOS PALSAR	Advance Land Observation Satellite Phased Array type L-band Synthetic Aperture Radar
ALOS PALSAR	Advanced Land Observation Phased Array type L-band Synthetic Aperture Radar
ANIR	Angle at NIR
ARED	Angle at Red
ARI	Anthocyanin Reflectance Index
ARVI	Adjusted Resistant Vegetation Index
ASTM	American Society for Testing and Materials
AVIRIS	Advanced Visible Infrared Imaging Spectrometer
BMVI	Broadband Multispectral Vegetation Indices
BP DWH	British Petroleum Deep Water Horizon
BVI	Broadband Vegetation Indices
CARET	classification and regression trees
CART	Classification and Regression Tree
CI	Chlorophyll Index
CP	Cloude Pottier
CTR	Carter Index
Der	Ratio of derivatives
deRES	Maximum first derivative spectrum
DNA	Deoxyribo-Nucleic Acid
ENDVI	Enhanced Normalized Differences Vegetation Index
EO	Earth Observation
ESA	European Space Agency
ESA ECCI	European Space Agency - Climate Change Initiative
EVI	Enhanced Vegetation Index
FD	Freeman Durden
FF	Fuzzy forest
fpnRES	Fifth-order polynomial fitting technique
G/NIR	Green Near Infrared
G/RED	Green Red
G/SWIR	Green Short Wave Infrared
GauREA	Gaussian function fitting technique
GBNDVI	Green-blue NDVI
GEC	Geocoded Ellipsoid Corrected
GLCM	Gray-level co-occurrence matrix
GLI	Green leaf index
GNDVI	Green Normalised Difference Vegetation Index
GRD	Ground Range Detected
GRNDVI	Green-red NDVI
HH	Horizontal Horizontal
HI	Hydrocarbon Index
HPC	High Performance Computing
HV	Horizontal Vertical
JIV	Joint Investigative Report
LAD	Leaf Angle Distribution

LAI	Leaf Area Index
LIC	Lichtenthaler Index
LIDAR	Light Detection and Ranging
MCARI	Modified Chlorophyll Absorption Ratio Index
MDA	Mean Decrease in Accuracy (MDA)
MDG	Mean Decrease in Gini index (MDG)
MICA	Material Identification and Characterization Algorithm
MIVIS	Multispectral Infrared Visible Imaging Spectrometer
MODIS	Moderate Resolution Imaging Spectroradiometer
MRENDVI	Modified Red-Edge NDVI
MRESRI	Modified Red-Edge Simple Ratio Index
MRF	Markov Random Fields
MSAVI	Modified soil adjusted vegetation index
MSI	Moisture soil index
MSI	Multi Spectral Imager
MSIC	Multiseasonal Image Classification
MSIF	multi-frequency SAR Image Fusion
MSOIF	multi-frequency SAR Optical Image Fusion
MSR	Modified simple ratio
MTCI	MERIS Terrestrial Chlorophyll Index
NASA	National Aeronautics and Space Administration
NBR	Normalized burn ratio
NBVI	Narrow Band Vegetation Indices
NDDC	Niger Delta Development Commission
NDII	Normalised Difference Infrared Index
NDVI	Normalized Difference Vegetation Index
NDVVI	Normalized Difference Vegetation Vigour Index
NDWI	Normalised Difference Water Index
NGO	Non-governmental Organisation
NHI	Normalised Heading Index
NIR	Near Infrared
NIR/R	Near Infrared Red
NNPC	Nigerian National Petroleum Corporation
NOSDRA	National Oil Spill Detection and Response Agency
NPC	National Population Commission
NPV	Non-Polluted Vegetation
OBIA	Object Based Image Analysis
OSAVI	Optimization of Soil-Adjusted Vegetation Index
OSCAR	Oil Spill Contamination mapping in Russia
PCA	Principal Component Analysis
PolSAR	Polarimetric SAR
PPR	plant pigment ratio
PRI	Photochemical Reflectance Index
PSND	Pigment Specific Normalized Difference
PSSR	Pigment Specific Simple Ratio
RENDVI	Red-Edge NDVI
REP	Red-Edge Position

RF	Random Forest
RF	Random forest
RFE-RF	Recursive Feature Elimination Random Forest
RG	Red Green Ratio
SAR	Synthetic Aperture Radar
SAVI	Soil Adjusted Vegetation Index
SIPI	Structure Insensitive Pigment Index
SLAR	side-looking airborne radar
SLC	Single Look Complex
SLR	side-looking radar
SMAP	Soil Moisture Active Passive
SNAP	Sentinel Application Platform
SPDC	Shell Development Corporation
spnREA	Sixth-order polynomial fitting technique
SR	Simple ration
SRI	Simple ratio index
SRTM	Shuttle Radar Topography Mission
sumREA	Sum of the first derivative
SVM	Support Vector Machines
SVM	Support Vector Machine
SWIR	Shortwave Infrared
TCA	Tree cover areas
TCARI	Transformed Chlorophyll Absorption in Reflectance
THC	Total Hydrocarbon content
TNDVI	Transformed normalized vegetation index
TPH	Total Petroleum Hydrocarbon
TSAVI	Transformed Soil Adjusted Vegetation Index
UAV	Unmanned Air Vehicle
UAVSAR	Unmanned Air Vehicle Synthetic Aperture Radar
UNEP	United Nation Environmental Programme
USGS	United States Geologic Survey
VHI	Vegetation Health Indices
VI	Vegetation Indices
VIS	Visible
VOC	volatile organic compounds
VOGI	Vogelmann Red-Edge Index
VRE	Vogelmann Red Edge Index
WBI	Water Band Index
WGCNA	Weighted Gene Correlation Network Analysis
WGS UTM	World Geodetic Survey Universal Transverse Mercator

Chapter 1 : Introduction

1.1 Oil Exploration and Spill in the Niger Delta

Rapid human capital development and the expansion of socio – economic activities worldwide have led to evolvement and demand for various forms of energy. This is required to cope with the urgent needs of the rapid growing global population in the areas of electricity, agriculture, transportation, housing and infrastructure development, which are fundamentally driven by multiple sources of energy.

Since the early emergence of the first oil producing area in Pennsylvania in 1861 there has been an increase in oil demand and at an annual rate higher than 20% since 2000 (Glover and Economides, 2010). This has led to the steady development of allied industries such as the aviation sector and automobile industry (Rodrigue, 2009). Activities in these main stream industries are largely dependent on by-products of crude oil, which also accounts for more than 75% of the global utilization of the product (Aleklett *et al.*, 2010). By virtue of hydrocarbon's portability, dense energy source and as the base of many industrial chemicals, petroleum is one of the world's most important commodities (Sarma, 2012). Today about 90% of fuel needs are met by oil. Other by-products such as carbon blacks, greases, transformer and cable oil, waxes, lubricating oil, white oils, rust preventatives and miscellaneous, solvents are among other derivatives of the petroleum industry. These products are however in high demand as they are necessary to fulfill basic human needs to foster and sustain livelihood (Younes, 2012).

Crude oil is an actively traded commodity globally and its demand has grown steadily over the decades, from 60 million barrels per day in the 1980s to 84 million barrels per day in the 1990's (Hasan, Ghannam and Esmail, 2010). Figure 1-1 shows the increasing trend in global oil demand and refining activities of petroleum. The current global daily demand for oil is roughly around 90 million barrels of oil to satisfy the world's daily needs, 'compared with 31.2 barrels in 1965' (Rodrigue, 2009) and the International Energy Agency forecasts that demand will be 121 million barrels a day by 2030 (Judkoff and Neymark, 1995).

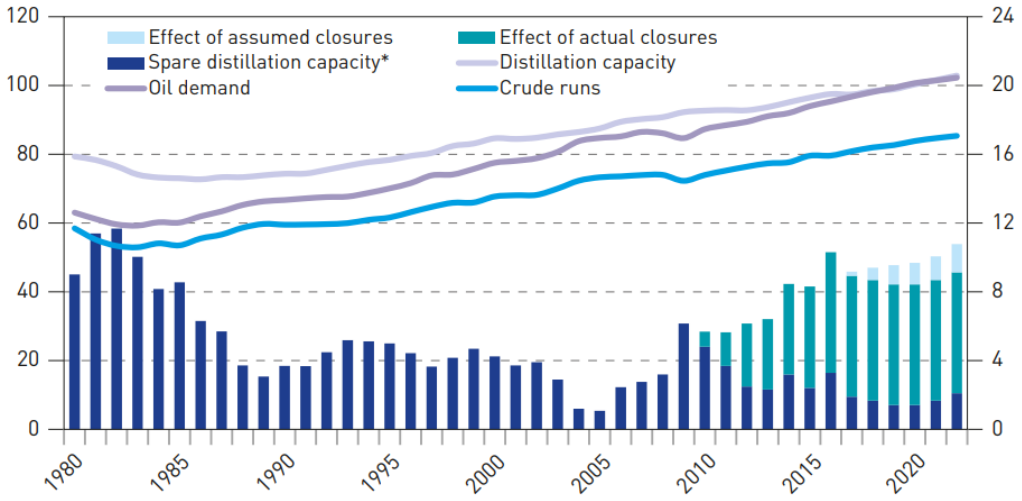


Figure 1-1: Global Oil Demand, refining capacity and crude runs, 1980 – 2022
Source: (Countries, 2017)

Generally, the petroleum industry is a large scale industry characterized by large scale mechanization process of drilling for extraction, transportation, refining and distribution of the final end products (Hasan, Ghannam and Esmail, 2010). In most cases, when the product is extracted at the initial stage, large scale well-coordinated piping systems are required for effective transportation to refining stations, where a fractional distillation process is used to separate the petroleum products into various primary by-products (Ohsol *et al.*, 1999). Similarly, upon successful extraction of the various by-products, another highly mechanized pipeline system is required to distribute the by-products for third party consumption or for further processing to obtain other essential commodities (Ohsol *et al.*, 1999).

Thus, due to the large-scale production, distribution and transmission, pipeline networks are required to move crude oil in a large-scale production process. This is usually at very high pressure of more than 1000 pounds per square inch (psi). The slightest malfunction or accident can result to catastrophic spill incidents. These spills are characterized by discharge of huge volume of raw unrefined hydrocarbon crude into the marine or terrestrial environment (depending on where it occurs). The results are usually very devastating and detrimental to the environment. Also, complexities involved in the operational control of pipeline facilities can often result to hydrocarbon crude oil gush for several days and weeks before normalcy is restored (O'Rourke and Connolly, 2003), further complicating these incidents.

Crude oil exploration in Nigeria began around 1903, while the first discovery was made in 1956. Nigeria joined the ranks of oil producers in 1958 producing around 5100 Barrels Per Day when its first oil field came on stream and further expanded in its exploration, as concessionaire rights were extended to other companies to explore the onshore and offshore areas of the Niger Delta after 1960. Nigeria is currently the largest oil producer in Africa (Taiwo, Otolorin and Afolabi, 2012) and has the 8th largest crude oil reserves in the world (Biersteker, 2014). ChevronTexaco, ExxonMobil, Total, Agip, and ConocoPhillips are the major multinationals involved in the Nigeria oil sector (Taiwo, Otolorin and Afolabi, 2012).

Crude oil, which is the main stay of the Nigerian economy, accounts for more than 80% of the country's revenue and derived foreign exchange, and has seen production increase from as little above 1000 barrels per day in 1970 to over 3000 barrels per day in 2010 (Figure 1-2) (Taiwo, Otolorin and Afolabi, 2012). This has largely accounted for the huge economic prosperity the country currently enjoys having substituted for agriculture which used to be the main stay of the economy (Biersteker, 2014). In spite of the tremendous gains, activities of oil exploration in the Niger Delta region has led to concomitant increase in oil spill incidents. Reasons for these spill incidents have been largely linked to decay in oil pipeline infrastructure, operational failure and sabotage (UNEP, 2011).

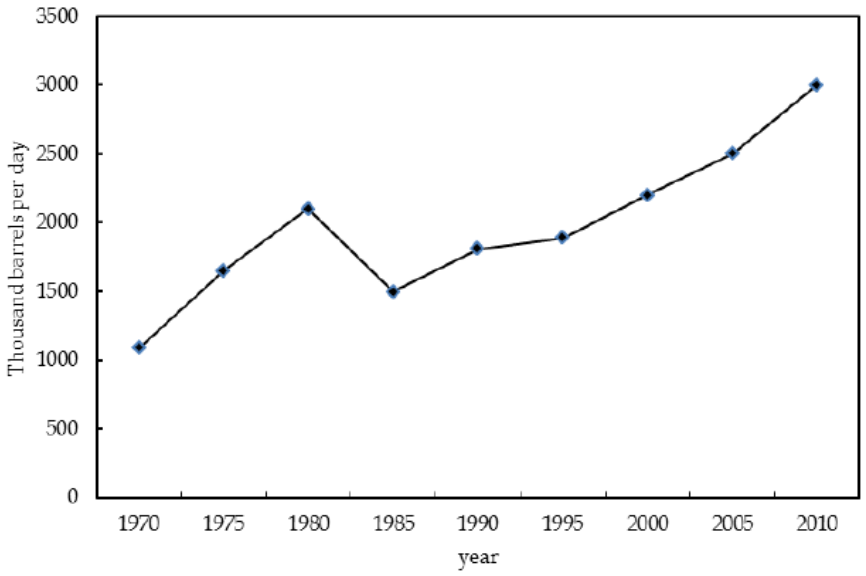


Figure 1-2: Nigeria Crude Oil Production Trend since 1970 up until 2010. Source: (Ayanlade, 2016)

Crude oil spill incidents have been observed to be on the rise since 1975 to 2015. Data provided by the Nigerian National Petroleum Corporation (NNPC) in Figure 1-3 shows that spill incidents has been on the rise since 1976 when oil production export began to 2009. The increased oil production activities have substantially led to the expansion of oil pipeline facilities, which is not unconnected to the rise in oil spill incidents (Taiwo, Otolorin and Afolabi, 2012). Other factors such as rising population trends and neglect of the Niger Delta host community have also been noted as prime cause of sabotage to oil facilities (UNEP, 2011) leading to some of the very large devastating spill incidents.

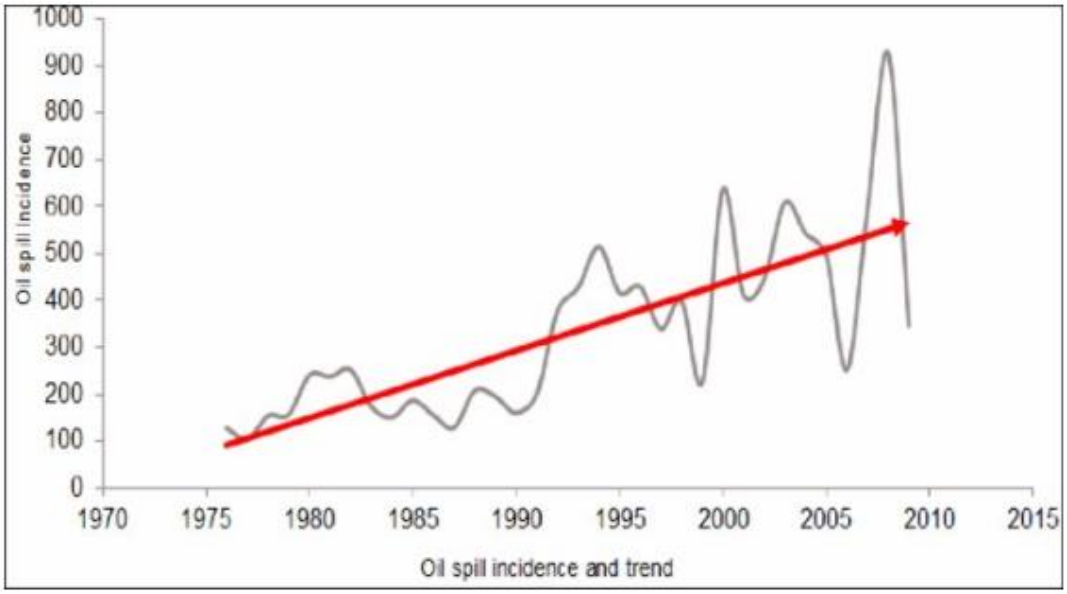


Figure 1-3: Oil Spill Incident Trend between 1970 and 2010. Source: (Ayanlade, 2016).

Major oil spills in the Niger delta include the Forcado tank 6 Terminal incident in Delta state where 570,000 barrels of crude was spilt into the Forcados estuary in July 1979. This affected the aquatic environment and surrounding swamp forest. The Funiwa No.5 Well in Funiwa field is also another major spill where an estimated 421,000 barrels of oil gushed into the ocean from January 17 to January 30 1980. The incident was reported to have affected marine lives and destroyed over 836 acres of mangrove forest (Gabriel, 2004; Tolulope, 2004; Ukoli, 2005). The Oyakama oil spill occurred on the 10th May 1980, with a spill volume of approximately 30,000 barrels of unrefined raw crude into the environment (Ukoli, 2005).

Furthermore, in August 1983 the Oshika village in Rivers state witnessed a spill incident where 5,000 barrels of oil from the Ebocha-Brass (Ogada-Brass 24) pipeline flooded the lake and swamp forest leading to the mortality of various aquatic life forms such as crabs, fish and shrimp. The impact of the spill was reported to have persisted eight months after its occurrence and there was high mortality in embryonic shrimp and reduced reproduction due to retained oil in the lake sediments (Gabriel, 2004). The Ogada-Brass pipeline oil spillage in February 1995 was also reported to have spilled around 24,000 barrels of crude oil over freshwater swamp forest and into the brackish water mangrove swamp (Kadafa, 2012; Ukoli, 2005).

Other devastating spill incidents have also occurred in recent time in the Niger delta. The Shell Bonga oil spill is one of such incidents, where over 40,000 barrels of crude was spilt into the adjoining environment. The incident which occurred in December 20th 2011 affected over 20 riverine communities across Akwa Ibom, Bayelsa and Delta State in Nigeria (Ladan, 2012; Ogbonna and Ebimobowei, 2012). The worse yet, is the Bodo oil spill in Bodo Village where a 55-year-old pipeline owned by Shell ruptured twice spilling an estimated 600,000 barrels of crude oil into the surrounding creeks in 2008. Residents of the community were compensated in 2015 by Shell petroleum in a payout of $83.2m in compensation for the various impacts and hardship caused by the spills (Aljazeera report, 2017).

The Nigerian Conservation Foundation in a study in 2006 put the figure for oil spilt, onshore and offshore, at 9 to 13 million barrels of oil over the past 50 years, creating a black-spotted landscape that can be seen from space (Jernelöv, 2018). This has led to continuous contamination of all types of land and water, dams, canals, delta arms, agricultural lands, forests and mangroves alike with oil at different stages of weathering.

Jernelöv (2018) observed that oil fractions like asphalt and tar dominate polluted areas given the many decades of oil spills and as new oil is intermittently spilt, the recovery processes are repeatedly interrupted, resulting in widespread oil damage to waterbodies, land and ground water.

1.2 Environment and Social Impact of Oil Pollution in the Niger Delta

1.2.1 Land Degradation

Various harmful and organic compounds introduced into the natural environment in an event of oil spill can alter the geochemical and physical properties of soil, river, and other components of the environment. The potential for fire hazard is also very high due to the volatility of crude oil and also because spilt crude oil is occasionally burnt rendering the soil unfit for cultivation. This in turn can affect agriculture and lead to drastic decline in output from farming activities (Nriagu, 2011). Water pollution of aquatic ecosystems also has the ability to damage fisheries and contaminate water in shallow wells, which residents depend on for drinking and other domestic purposes. Between 50% and 70% of the Niger Delta inhabitants depend on the natural environment for agriculture, fishing, and the collection of forest products as their principal source of livelihood. For most of the people of the Niger Delta, unsoiled and sustainable environment is fundamental to their overall well-being and development (Nriagu, 2011; UNEP, 2011).

In the long-term, oil spill can lead to land degradation and emigration of local inhabitants to other rural and urban areas, exerting pressure on the often inadequate and dilapidated infrastructure, which leads to increase in poverty (Nriagu, 2011; UNEP, 2011). Investigations carried out by Lindén and Pålsson (2013) as part of the UNEP-lead study to assess the environmental conditions in the Niger Delta region of Nigeria showed that oil induced impact on vegetation led to extensive mortality of mangrove vegetation. Laboratory tests of drinking water and soil sediment samples also indicated that inhabitants were exposed to contaminated drinking water 14000 times higher than the Nigerian standard of drinking (Jernelöv, 2018; Nriagu, 2011; UNEP, 2011).

1.2.2 Physical Health Impact

Oil spills also have an associated impact on the physical health of people within affected communities. Direct or indirect contact with crude oil through for example, inhalation of vapors, coating on skin and consumption of tainted seafood can cause deleterious health effects ranging from dizziness, skin rashes, nausea, cancer (in extreme cases) and issues with the central nervous system (Aguilera *et al.*, 2010; Chang *et al.*, 2014; Herrington *et al.*, 2006; Jenssen, 1996; Major and Wang, 2012).

Although long-term hydrocarbon toxicity effects on humans are less understood (Aguilera *et al.*, 2010; Jenssen, 1996), they have been linked with volatile organic compounds (VOCs) that are capable of causing severe DNA degradation, cancers, birth and reproductive defects, irreversible neurological and endocrine damage, and impaired cellular immunity (Aguilera *et al.*, 2010; Chang *et al.*, 2014; Jenssen, 1996).

Similarly, Orisakwe (2009), also observed that the Funiwa oil well blow-out, which occurred in January 1980 (involving about 200,000 barrels of crude oil) deprived the inhabitants of the affected area of access to potable water. Inhabitants were also deprived of their fishing occupation, resulting in a massive displacement of thousands of farmers and fishermen. Just like in the case of Valdez oil spill where medical doctors and dispensers in the area reported a prompt increase in the incidence of vaginitis, gastroenteritis, conjunctivitis, dermatitis and vulvitis among the people (Aguilera *et al.*, 2010).

1.2.3 Psychological Health Impact

Similarly, Chang *et al.* (2014) have particularly noted that technological disasters such as oil spill in the United States, were more psychologically stressful than natural disasters. This threatened the stability of livelihoods, coupled with new flows of recovery money into the community, leading to various forms of stress and social breakdown (Chang *et al.*, 2014). It was observed that following the Valdez oil spill, high rates of alcohol and drug use were associated with recovery jobs, especially in indigenous communities, and there were higher rates of domestic violence and crime more broadly (Chang *et al.*, 2014). This in turn led to increased demands on clinic, mental health, and rehabilitation programs and further impacted not only on direct costs through hospital bills and related expenses, but also on broader social cost through, work stoppages and shorter life expectancy (Chang *et al.*, 2014).

1.2.4 Social Structure to Mitigate Impact

The use of social and welfare initiatives is a common practice to mitigate the impacts of oil pollution on inhabitants. Social impacts have however been alleviated in some spill events. An example is in the case of the Hebei-Spirit and Prestige spill (Jernelöv, 2018), where interim aid payments, NGO support, volunteer mobilization and support from unaffected populations were provided to assist the affected populations (Chang *et al.*, 2014). A study of the Galician coast affected by the Prestige spill found

that, in contrast to the Valdez oil spill, there was no significant impact on mental health, social relationships, or confidence in institutions, because of the strong support of social groups and appropriate levels of interim financial aid (Chang *et al.*, 2014).

1.3 Rationale for this Research

Spills on terrestrial landscapes, which are the primary focus of this research are far more frequent compared to the marine oil spill. Large scale spills on the terrestrial landscape can cause immeasurable damage to vegetation and inland aquatic life as toxins can totally render an area completely in inhabitable. Studies in this regard (Aislabie *et al.*, 2001; Ribeiro *et al.*, 2013; Zabbey and Uyi, 2014) have shown that the microbial impact caused on soil organism can prevent optimal vegetation growth and crop yield for up to a decade if proper post spill remediation measures are not implemented. Similarly, human health and socio – economic impact of oil is not uncommon and can prove very detrimental, as livelihood of rural agrarian and peasant farmers can be affected.

It becomes therefore imperative to devise means through which the precise area affected by oil pollution can be readily established. This would foster all post spill recovery, remediation and rehabilitation work. In the past, approaches such as field reconnaissance survey and helicopter flybys have been relied on to visually reconcile areas affected by oil pollution (Murvay and Silea, 2012). However, field based methods are heavily constrained by inaccessible terrains such as marsh, mangrove and densely forested areas. Similarly, there is also the challenge of cost efficiency, loss of manpower and valuable time associated with these approaches.

Recently, remote sensing satellite image-based detection is becoming a more popular method for oil spill detection. This has proven useful in formulating mitigation strategies to facilitate post disaster recovery and damage assessment. Remote sensing satellites platforms are particularly equipped with wide synoptic view and the capability of capturing extensive areas at a single pass, potentially saving the huge human and financial cost associated with other techniques. In addition, satellite images from remote sensing platforms provide the opportunity of providing more detailed and unbiased spatial information of an area, most of which human efforts cannot effectively provide or reconcile. Also, satellite remote sensing platforms

offers repetitive and revisit capabilities, this can be weekly, monthly or yearly, providing an avenue for studying changes caused by oil spill impact on both flora and fauna communities.

As the orbital presence of satellite sensors increase, so does their potential, hence the opportunity to better understand the environmental trends due to oil pollution. Over the years, there has been increasing capability of orbital satellites to capture electromagnetic radiation at different part of the spectrum, from optical, near infrared, shortwave infrared and to microwave. These images come in various temporal, spatial and spectral resolutions, providing a huge potential for obtaining more detailed information to study oil spill impacts on the environment. The emphasis here is on both short, medium and long term responses of the terrestrial ecosystem. This research focuses on assessing the spatial and temporal dimensions of terrestrial oil spills, with emphasis on detection and mapping, whilst bridging existing gaps in the literatures. It also aims at utilizing some of the most innovative tools, satellite image processing methodologies and a range of complex data for detecting and mapping terrestrial oil spill impact.

1.4 Thesis Structure

This thesis comprises of seven chapters. Chapters 1, 2 and 3 are the introduction, literature review and methodology, respectively. Chapters 4, 5 and 6 presents the result of the data analysis, while chapter 7 provides a summary discussion, conclusion, limitations and areas of possible further research. Figure 1-4 is a graphical illustration showing the various chapters of this thesis and the connection between them.

Chapter 1: Provides a general overview of the research. It provides a general background of oil exploration in Nigeria and some notable oil spill incidents over the past 4 decades. The environmental and social impact caused by oil spill incidents in the Niger Delta are also highlighted, while the rationale for detecting and mapping terrestrial oil spills with satellite remote sensing is also discussed.

Chapter 2: Presents some of the previous studies on terrestrial oil spill detection and the findings from various approaches used for this purpose. It also critically evaluates the reflectance characteristics from field and image-based spectroscopy of various oiled and oil-free surface types. The general characteristics of vegetation bio-

chemical and biophysical responses to crude oil and the potential of SAR image for monitoring changes as a result of oil spill impact was also discussed. This chapter goes further to assess some notable machine learning methods used for image classification and how they have improved for better classification accuracy. Furthermore, previous studies on the assessment of oil spill impact and land cover change in the Niger Delta using satellite remote sensing methods are discussed. While finally, the observed gaps in the literature, the rationale for the current research, the research questions, aims and objectives are also highlighted.

Chapter 3: Provides a general methodology for this research. It further provides information on the various remote sensing satellite data, spill incident data, field data and existing land cover data used to address the various research objectives. It also provides an overview of the study area in the Nigeria's Niger Delta region, with emphasis on the location, size, population, ecology, soil geology and climate of the study area.

Chapters 4: Presents the result of the first research objective to answer the research question of 'to what extent can optical image spectral band and vegetation health indices be used to discriminate oil-free and oil-polluted cropland, grassland and tree cover area'. Here, machine learning RF was used to classify Landsat – 8 spectral bands and vegetation health indices in two separate classification scenarios namely; full study area image and land cover type specific image classification to discriminate oil-free and polluted vegetation.

Chapter 5: Presents the result of the second research objective to answer the research question of 'to what extent can the integration of optical derived variables, multi-frequency (C and X band) SAR and geo-environmental variables be used to detect and discriminate between oil-free and oil-impacted cropland, grassland and tree cover areas. The emphasis here is on the detection of oil pollution impact using variables derived from Sentinel – 1, Sentinel – 2, Cosmo Skymed and TanDEM X images. The utility and performance of a novel method 'fuzzy forest' in selecting unbiased variables for the classification process is tested and compared to the result of the traditional random forest classifier.

Chapter 6: Presents the result of the third research objective to answer the research question of 'to what extent can multi-temporal, multi-frequency and multi-seasonal

SAR images be used for detecting and discriminating oil-free and oil-polluted cropland, grassland and tree cover area. The chapter specifically examines sentinel – 1 time series backscatter characteristics of a typical oil-free and oil-polluted cropland, grassland and tree cover areas to infer vegetation temporal characteristics. It further analyzes wet and dry season stack of multi-frequency (L, C and X band) SAR classification with random forest and support vector machine classifiers to improve classification and discrimination accuracy.

Chapter 7: Provides a summary discussion of the results obtained from the three analysis chapters (Chapters 4, 5 & 6) with a view of harmonizing all findings and evaluating how these compares with existing literature in addressing the research gaps observed. It also provides conclusion to the thesis, highlighting implication of the findings and contributions to wider knowledge, the several limitations encountered in the course of the research and finally areas of possible further research.

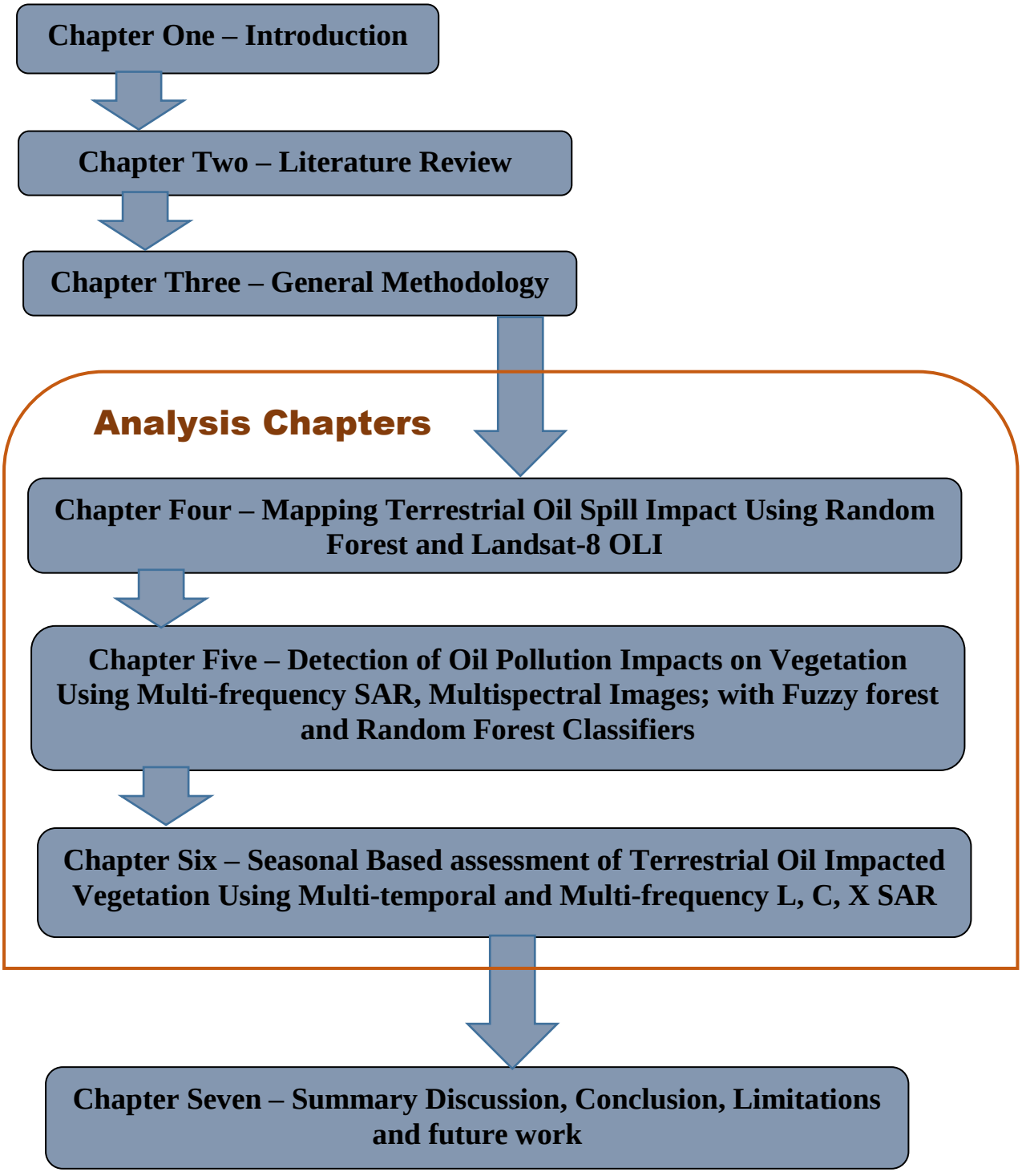


Figure 1-4: A graphical illustration of the various chapters in the thesis and the connection between them

Chapter 2 : Literature Review

2.1 Terrestrial Oil Spill Remote Sensing: An Overview

Terrestrial oil spills have been identified as one of the main causes of environmental degradation (Orisakwe, 2009). These occurs through the uncontrolled discharge of oil or its by-products including chemicals and wastes, as a result of equipment failure, operational errors, or willful damage. Other major causes of hydrocarbon pollution on land include: near shore drilling, off shore oil pollution and hydrocarbon micro and macroseepage. Pollution from oil seepage occurs through a natural process of vertical or near-vertical movement of oil from a sub-surface reservoir to the surface (Yang *et al.*, 2000). Tedesco (2012) has emphasized that Macroseeps and Microseeps do not always indicate the presence of economic recoverable hydrocarbons at depth, but rather may represent leakage from a temporarily stationary sources of petroleum.

Methods and approach used for detecting hydrocarbon microseeps, macroseeps and oil spills on terrestrial landscape are similar. The hydrocarbon compositions from the two sources are largely same, as they both constitute raw and unrefined hydrocarbon crude content. Hence, it can be assumed that microseeps, macro seeps and oil spills pose the same kind of effects on both vegetation and bare surfaces. Schumacher (1996), in describing the major characteristics of hydrocarbon on land, noted that hydrocarbon beneath soil subsurface creates a chemically reducing zone in the soil column at depths shallower than would be expected in the absence of oil. This stimulates the activity of hydrocarbon-oxidizing bacteria, which decreases soil oxygen concentration while increasing the concentration of CO_2 and organic acids. These changes affect pH and Eh in soils, which in turn affect the solubility of trace elements altering the root structure of vegetation, ultimately influencing plant vigor and hence, spectral change.

Several advances have been made in the detection and characterization of terrestrial oil spills since the early periods of 1950 before the launch of the first satellite. Traditional methods often relied on airborne camera images and field reconnaissance surveys to establish precise oiled surface extent (Murvay and Silea, 2012). Many satellite platforms now provide synoptic views of the landscape at various temporal, spatial and spectral resolutions; and within the Visible (VIS), Near Infrared (NIR), Shortwave Infrared (SWIR) to the microwave part of the electromagnetic spectrum. Field based methods, such as ground spectroscopy have also received considerable attention in exploring the characteristics of polluted landscapes to compliment

measured reflectance spectra of oil polluted land and vegetation from the satellite sensor.

Brooks (1979) noted that the applications of remote sensing for oil monitoring on land should ideally cover two aspects for effectiveness: the first is to research into the spectral characteristics of the anomalous or stress vegetation species caused by hydrocarbon, while the second should focus on mapping the anomalous or stress vegetation species extent from imagery. Furthermore, Siegel (1974) has stated that in order to identify vegetation anomalies caused by oil pollution using remote sensing, "normal" plant variability has to be assessed by judging the plant distribution, presence of indicator plants, or morphological changes in plants induced by excesses or deficiencies in available soil nutrients (Brooks, 1979; Yang *et al.*, 2000).

The use of remote sensing techniques to evaluate and map the extent of impact of oil spill on land has not received a lot of attention. Fewer efforts have been devoted for monitoring onshore oil spills and reasons for this are partly due to the complexities associated with clearly distinguishing oil from other features with similar spectral characteristics (Mahdianpari *et al.*, 2018). Similarly, oil under a dense vegetation canopy layer is even more difficult to detect, especially with optical sensors, as they rely on earth reflected energy from the sun. As such, the sensor only captures top of vegetation canopy reflectance. Figure 2-1 demonstrates this problem.

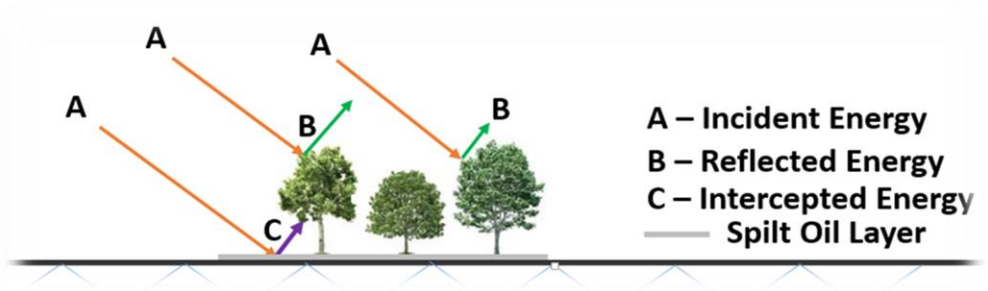


Figure 2-1: Earth Energy Interaction in an Oil Spill Polluted Site

In addition, the difficulty of acquiring cloud free high spatial and spectral resolution optical images has also contributed to the use of optical data for inland oil spill detection (Figure 2-2). This is unfortunate given the large numbers of oil spills that occur annually in terrestrial areas globally and their devastating economic and environmental impacts (Mahdianpari *et al.*, 2018).

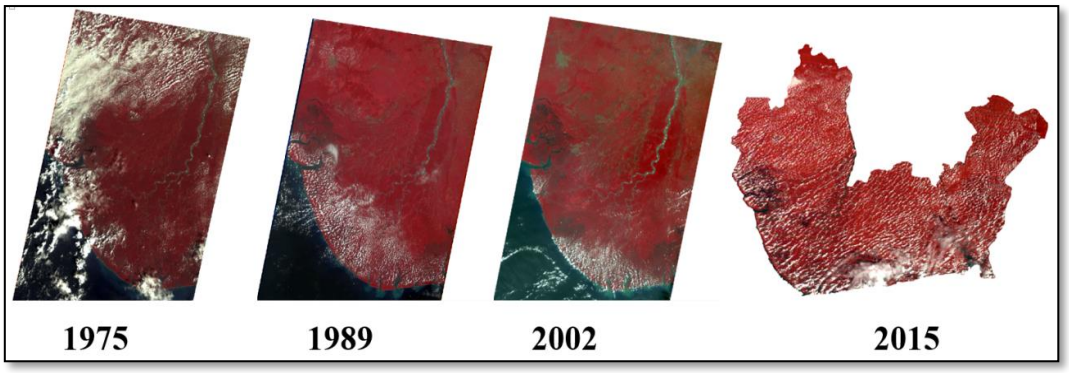


Figure 2-2: Images of the Niger Delta Showing Persistent Cloud Cover

As a result of these myriad of factors, some research (Li, Ustin and Lay, 2005; Noomen *et al.,* 2015) has focused on the use of vegetation greenness and health status as a proxy to detect the presence (through impact) or absence of Petroleum crude oil products from in-land spillages. However, relatively few studies (Ramsey Iii *et al.,* 2011; Ramsey, Rangoonwala and Jones, 2015) have focused on SAR images to detect oil contamination on terrestrial landscape. SAR has the capability of penetrating and imaging through cloud as it is an active sensor, which does not rely on the earth reflected sun energy. Evidence from research has generally shown that advance sensors such as Microwave sensors, Laser fluoro sensors and LIDAR has not been extensively explored in the detection and mapping of hydrocarbon crude oil impact on the terrestrial landscape.

In this chapter, the literature is surveyed with a specific focus on understanding spectral responses of various landscapes (both vegetated and non-vegetated) to oil spill impact, within the VIS-NIR-SWIR and Microwave part of the spectrum. The potential of microwave-based vegetation responses to physical changes caused by oil and the potential of ground spectroscopy methodology is also assessed and discussed. Similarly, the capabilities of optical and SAR images to detect and map oil impact of affected areas using machine learning classifiers are highlighted.

2.2 Detection of Hydrocarbon Pollution Using Hyperspectral and Multispectral Images

2.2.1 Characteristics of Oil Affected Surface

Terrestrial oil spill remote sensing in the visible (Red, Green and Blue) spectrum is complex because there are limited mechanism for positive and accurate oil detection

(Fingas and Brown, 2014). Research has shown that oil on shorelines is even more difficult to identify positively because other features such as non-photosynthetic vegetation, dense vegetation, shallow waterbodies as well as rich soil may exhibit similar spectral characteristics as oil. This would however necessitate adequate reconciliation of visible optical sensors with precise field spectral reflectance measurements of oiled surface to improve detection and classification accuracy. This approach was partly implemented in Khanna *et al.* (2013), where Advanced Visible Infrared Imaging Spectrometer (AVIRIS) data and field based observations such as species composition, canopy condition, presence of oil and penetration of oil into the marsh of Barateria bay were used to map polluted marsh following the deep water horizon oil spill. This significantly helped in the final characterization of oil-impacted and oil-free marshes.

Oil-contaminated soils generally exhibit a significant decrease in reflectance and this is due to strong absorptions induced by electronic transitions in molecular orbits (Gaffey *et al.,* 1993). Cloutis (1989) observed that the decrease in VIS-NIR reflectance is due to charge transfer between carbon atoms. Orlov *et al.* (1993) in an experimental study also noted a decrease in reflectance of oil-contaminated soils in the VIS-NIR region (400–800 nm), following progressive contamination with oil. This produced a rather flattened spectral curve for strongly contaminated soils compared to the oil free soil (Li, Ustin and Lay, 2005).

Cloutis (1989) further found that the wavelength positions of hydrocarbon overtones and band combinations is usually at 1700nm, but bands between 2200nm and 2600nm were the most promising regions to search for hydrocarbons using spectroscopy and analysis of satellite data. It was further emphasized that carbon–hydrogen stretch overtones and combinations dominate the 1700nm region, and the region from 2200 to 2600nm is affected by overlapping combination. This position is also supported by the work of Hörig *et al.*, (2001). Hörig *et al.* (2001), presented similar observations in the spectral region of 1400–2500nm with GER Mark V Infrared Intelligent Spectroradiometer and HyMap data. They observed that the decrease in the NIR-SWIR reflectance is primarily due to the most intense overtones and combinations of major fundamental absorptions of hydrocarbons (Li, Ustin and Lay, 2005).

Recent studies carried out by Douglas *et al.*, (2018) specifically showed that contaminated soils were characterized by high absorbance particularly in the NIR range of 700 – 2500nm. They further observed two absorption peaks at 1415nm and 1914nm, which they attributed to water absorption overtones, and a third peak at 2200nm, which they also attributed to metal-hydroxyl stretching. The results obtained in their study was further validated using Total Petroleum Hydrocarbon (TPH) measurement values from ultrasonic solvent extraction-gas chromatography in a linear partial least square regression (PLSR) and a non-linear random forest regression.

Oil spill remote sensing for vegetation stress detection is based upon vegetation response to solar radiation (Li, Ustin and Lay, 2005). The spectral response of healthy plants to solar radiation is in general similar, though difference exists between plants due to morphology and physiology, background soil types, and the climate (Li, Ustin and Lay, 2005). Noomen *et al.* (2015) observed that stressed vegetation typically exhibits a shift of the red edge position towards shorter wavelengths when exposed to natural hydrocarbon seepage or to simulated gas leakage. In this regard, reflectance properties of vegetation in the visible part of the spectrum are dominated by the absorption properties of photosynthetic pigments including chlorophyll, which have absorption at 660nm and 680nm (Smith *et al.*, 1990). And changes in the chlorophyll concentration produce spectral shifts of the "red edge" absorption near 700nm.

Hutchinson and Hellebust (1974) particularly noted that vegetation response to oil presents stages of effects in which, leaves progressively loose photosynthetic pigments leading to color changes from green to pale-green, yellowish-green, and then finally yellow. In advance stages the leaves, stems and trunks would become ash-brown, dark-brown or darkened after long exposure.

In the biochemical realm of things, plants begin to exhibit chlorosis due to the loss of both chlorophyll a and b, defoliation, enhanced litter fall, dwarfed seedlings and in some cases mortality. In addition, plants affected by oil pollution experience desiccation of the mesophyll tissue and collapse of cell walls resulting to substantial reduction in intercellular surface area and air space. This causes more reflection of red light and less reflectance in the near-infrared region compared to green healthy plants. Greater reflection of red light (by polluted vegetation) is due to the loss of

photosynthetic pigments, resulting in less absorption (Hutchinson and Hellebust, 1974; Li, Ustin and Lay, 2005; Rosso *et al.*, 2005). Healthy plants however have diagnostic high reflectance in the near-infrared region of solar radiation because of strong internal scattering of incident light from cell walls and intercellular spaces (Hutchinson and Hellebust, 1974).

Crude oil on vegetated land can lead to disruption of plant and seed water balance, metabolism, and toxicity. This can consequently lead to a reduction in plant germination and growth, inhibition of seedling and development process, which can characteristically change plant spectral reflectance (Li, Ustin and Lay, 2005; McCown and Deneke, 1972).

2.2.2 Field and Image Spectroscopy for Terrestrial Oil Spill Detection and Mapping

The detection and mapping of oil-affected areas on land is an important process that helps to reconcile and discriminate between actual contaminated vegetation from oil-free vegetation. It further forms the basis of establishing the total area affected by oil pollution. Maps established from this process are a useful medium to facilitate post oil spill recovery, remediation and rehabilitation exercise. Effective discrimination of oil-polluted and oil-free land cover can provide information on the location of oil pipeline leakages and the extent of land area affected by oil in regions with limited accessibility. In the long term however, this approach can be used to formulate robust and transferable image processing models, which can be used to track future terrestrial oil spills leveraging on the pool of spectral library generated.

The detection (Arellano *et al.*, 2015; Khanna *et al.*, 2013; Kokaly *et al.*, 2013) and mapping (Bianchi *et al.*, 1995b; Van der Werff *et al.*, 2007) of inland (terrestrial) based oil spills using high end spectral data have proved valuable in several remote sensing applications. This is because the oil signal is characterized by absorption features in the near-infrared (NIR) and shortwave-infrared (SWIR) regions of the spectrum, which are rather effectively captured by hyperspectral sensors in radiance of more than 300 distinctive narrow bands. Khanna *et al*. (2013) have emphasized that the NIR 1730nm and SWIR 2300nm bands are best suited for detecting oil on bare soil. The result of their assessment is presented in figure 2-3 and figure 2-4, and this shows that oil does not overlap with absorption features of soil background

materials and non-photosynthetic vegetation (Cloutis, 1989; Kühn *et al*., 2004) in this spectral region.

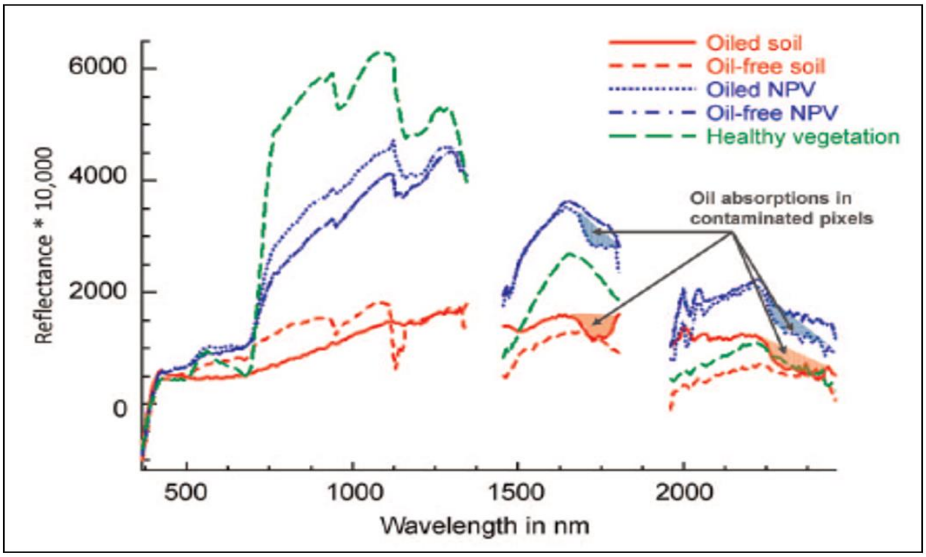


Figure 2-3: Reflectance Curve for Oiled soil, Oil-free soil, Oiled Non-photosynthetic Vegetation (NPV), Oil-free NPV and Healthy vegetation from AVIRIS data (Khanna *et al.*, 2013).

Kokaly *et al.* (2013) also argued that hyperspectral sensors for oil spill detection have the potential for detailed materials identification and precise estimation of their abundance. It was further noted that with more than one hundred wavelengths provided by a hyperspectral sensor, the spectral signature of oil can be harnessed to distinguish between different oil types (crude or light oil) and among features of interest, which can also eliminate the false alarm rate of features that have the same appearance as oil (Brekke and Solberg, 2005).

Past laboratory and field studies have indicated the potential for using imaging and field spectrometer data for detecting oil contamination on land (Kühn *et al*., 2004). Kühn *et al*. (2004) developed a hydrocarbon index (HI), which focused on the detection of oil based on a single absorption feature in the radiance data measured by imaging spectrometers. Background materials with overlapping absorption features, such as dry vegetation, were observed to have had high Hydrocarbon Index values similar to hydrocarbons (Kokaly *et al.*, 2013).

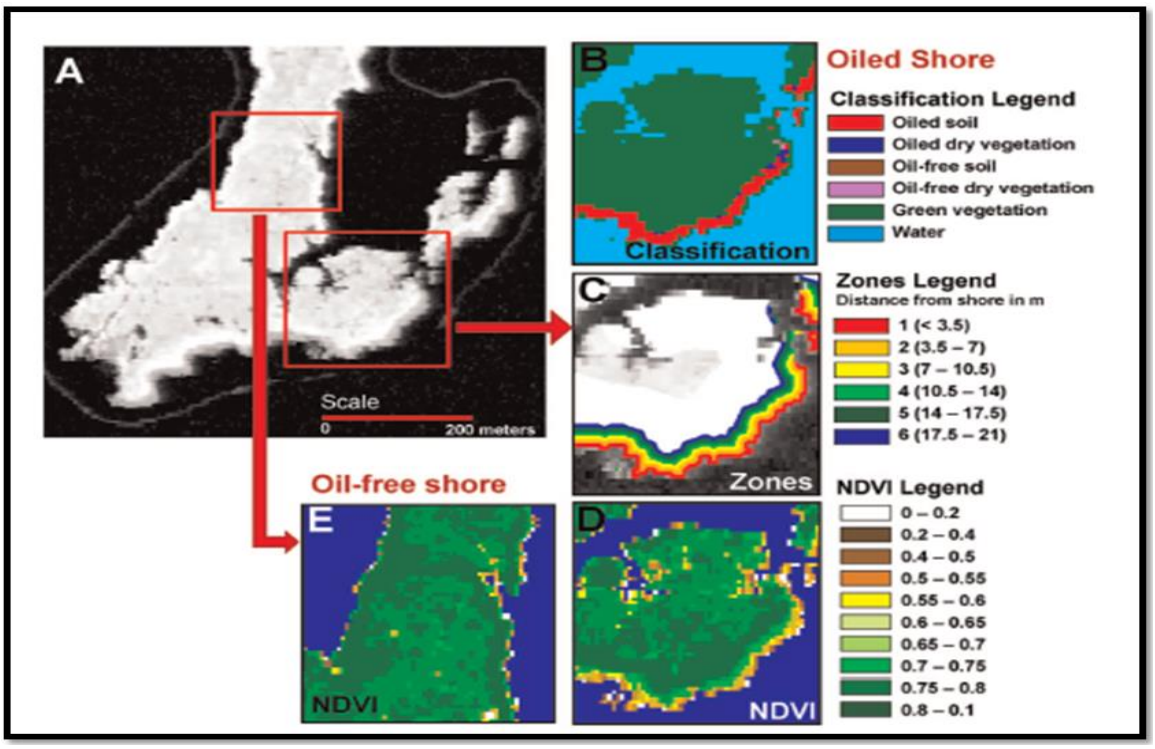


Figure 2-4: NDVI and Classified Image used to analyze oil Spill Impact (a) Gray scale NDVI image (b) The classified image (c) The zones from the shoreline (d) The low-NDVI bands of pixels right next to the oiled shoreline. For comparison, (e) shows the NDVI profile for an oil-free section of the shoreline (Khanna *et al.*, 2013)

Goodman (1994) has particularly stressed that reliable detection typically requires highly detailed spectral information provided by imaging spectrometers on terrestrial environments. Khanna *et al.* (2013) have also noted that high spectral resolution data are necessary to establish oil presence and attribute whether observed stress is due to oil or other stress factors. These have formed the basis of very recent studies where field and image spectroscopy have been applied to detect hydrocarbon crude oil impact on vegetation.

In addition, Mishra *et al.* (2012) used hyperspectral sensor to quantify the short-term impacts of oil spill on the photosynthetic activity and physiological status of coastal salt marshes. This study was able to demonstrate the impact of oil spill on salt marsh vegetation drawing from the phenological indicators from polluted and non-polluted sites. This further helped in the successful delineation of the critical hotspots of marsh stress, for the prioritization of immediate restoration. Arellano *et al.* (2015) also investigated the suitability of satellite image spectrometry for the detection of oil contamination in the Ecuadorian Amazon forest. EO-1 (Earth-Observation 1) Hyperion images were used for analysis in conjunction with supporting field data on

soils and foliar properties. The results indicated that tropical dense forests exposed to petroleum pollution show reduced levels of chlorophyll content, accompanied by higher levels of foliar water content.

Khanna *et al.* (2013) also demonstrated the viability and efficiency of high spectral resolution imaging spectroscopy in the detection and mapping of oil spill impact in the near shore marshes of Barataria Bay, Louisiana in September 2010 and August 2011. The study used AVIRIS data to assess the impact of oil spill on the salt marsh plant community. Oil contamination was mapped using oil absorption features in pixel spectra to examine the impact of oil along the oiled shorelines. Results from the study showed that stress was restricted to the tidal zone extending 14m inland from the shoreline in September 2010.

Kokaly *et al.* (2013) also successfully mapped oil contamination along the shoreline after the British Petroleum – Deep Water Horizon (BP – DWH) spill using AVIRIS data acquired on July 31, September 14 and October 2 of 2010 over Barataria Bay. They used hydrocarbon absorption features centered at 1720nm and 2300nm as inputs to the United States Geologic Survey (USGS) Material Identification and Characterization Algorithm (MICA) to map oil presence. Huang *et al.* (2019) also used airborne hyper-spectral images for the detection of plant stress as a result of hydrocarbon micro-seepage based on the PROSPECT model in a laboratory setting. Similarly, other researchers (Lassalle *et al.*, 2018; Reséndez-Hernández, Prudencio-Csapek and Lozano-García, 2018) have also used hyperspectral images to assess vegetation affected by oil pollution, all as indirect methods of detecting hydrocarbon crude oil presence and impact.

2.2.3 Classifying oil-polluted area

In recent time however, newer studies have sought to explore other image derived mediums and techniques for classifying oil polluted areas. Susantoro *et al.* (2018) explored image-derived NDVI, LAI and ENDVI variables to detect and map vegetation affected by oil and gas micro seepages. Results obtained showed that NDVI and ENDVI allowed good identification of oil polluted areas. While Mahdianpari *et al.* (2018) also used high spatial resolution UAV and electromagnetic (EM) induction data for terrestrial oil spill detection and mapping in a small area in Dixonville, Alberta, Canada. Their results showed that soil salinity from EM data

coupled with UAV derived variables have much potential for the classification and discrimination of oil-polluted land from adjoining oil-free vegetation and other land cover components. In addition Achard *et al.* (2018) also investigated airborne hyperspectral derived variables and SVM methodology for terrestrial oil spill mapping.

Image spectroscopy provides a very unique and reliable opportunity of observing vegetation stress posed by crude oil pollution. Many studies rely on these vegetation spectra to detect and map hydrocarbon impacted areas. For example Bianchi *et al.* (1995a) used airborne hyperspectral images from the Multispectral Infrared Visible Imaging Spectrometer (MIVIS) to map the spatial distribution of oil spilled during the February 1994 Trecate blow-out in Northern Italy. The study used hyperspectral unmixing methods, oil fractional areas, as well as other classes (such as water, woods, cultivated fields, smooth surface soil, grooved surface soil and rice fields) as end-members in the image analysis process. Results obtained and validated in the study using ground reference points showed reasonable agreement between the unconstrained unmixing technique and the ground truth relating to the volume of oil information.

Van der Werff *et al.* (2007) also experimented two pixel based classifications methods (minimum distance to class means and Spectral Angle Mapper) to classify an hyperspectral Probe – 1 (HyMap) image to discriminate crude oil resulting from macroseepage and an anomalous halo bare soil resulting from microseepage. The result however showed that at best only 48% and 29% of the pixels that respectively contain crude oil and seepage-affected soil were detected. They further observed that confusion mainly resulted from the physical characteristics of the anomalies, as these are not unique to seepages. It was recommended that remote sensing of natural hydrocarbon seepages or oiled surface can be improved by image processing algorithms that makes use of spatial information.

Similarly, Hese and Schmullius (2009) in the Oil Spill Contamination Mapping in Russia (OSCAR) project used Landsat images to discriminate between oil-free vegetation, oil-contaminated vegetation, soil and industrial land use. Results obtained showed that NDVI and shortwave infrared bands significantly improved

discrimination of polluted vegetation from other land cover types. This underlines the intrinsic importance of vegetation indices for detecting oil impacts.

However, several factors can hinder the effective utilization of Spaceborne hyperspectral imaging in mapping oil polluted areas, especially within developing economies and areas within the mangrove ecosystem, such as in the Niger Delta region of Nigeria. Constraints such as the cloudy weather conditions experienced almost all year round (i.e. between February and November) limits the applicability of field and image spectroscopy for the detection and mapping of terrestrial oil spills. In addition, limited freely accessible hyperspectral images exist for the Niger Delta region, especially after the decommissioning of the NASA EO – 1 satellite. In addition, the cost prohibitive nature of Airborne hyperspectral imaging also limits applicability of image spectroscopy for the assessment of terrestrial oil spills.

These necessitates the use of alternative datasets, image processing methods and techniques to harness the utility of available NIR and SWIR of readily accessible multispectral images, such as the freely accessible Sentinel – 2A and 2B, which has high spatial and temporal resolution of 10m and 5 days respectively. It also captures reflectance in three unique Red-Edge bands and multiple shortwave infrared bands. This can potentially help in the discrimination of polluted and oil-free land cover types (i.e. both on land and on vegetation). In addition, the use of Spaceborne Synthetic Aperture Radar (SAR) images within the microwave region of the spectrum has seen little application for terrestrial oil spill mapping and can provide reliable temporal monitoring of affected areas. SAR Images are increasingly been used to assess and monitor the state of earth surface dynamics, especially within forest and cropland vegetation, landslide detection, deforestation mapping and mapping soil moisture. As such their application for mapping terrestrial oil spill would greatly profit the oil spill remote sensing research community.

2.2.4 Application of Vegetation Health Indices

Vegetation Indices (VIs) are mathematical combinations of various satellite image bands. They have been used by scientists since the 1960s (Jensen, 2009). Satellite vegetation indices products are commonly used in a wide variety of terrestrial science applications aimed at monitoring and characterizing Earth's vegetation cover from space (Myneni *et al.*, 2002). Indices retrieved from remotely sensed satellite data can

provide information on vegetation growth, vigor and their dynamics, and also provide useful insights for various application areas such as: in environmental monitoring, agriculture, biodiversity conservation, urban green infrastructures, forestry and other related fields (Xue and Su, 2017).

VIs have been extensively explored using spectral bands from airborne and satellite sensors to examine vegetation spectral signature characteristics, in the visible and near-infrared part of the electromagnetic spectrum. This is because vegetation stress associated with various forms of disturbance are likely to present symptoms, such as stunted growth and defoliation, most of which can only be significantly detected when sensitive and non-sensitive spectral bands of the image are logically manipulated (White *et al.*, 2008).

Stagakis *et al.* (2010) have observed that canopy reflectance in the visible and near infrared is strongly dependent on both structural (i.e., amount of leaves per area, leaf orientation, canopy structure) and biochemical properties (i.e. chlorophyll, carotenoids) of the canopy. Hence, it is difficult to develop and utilize a "unique" Vegetation Index that is exclusively sensitive to one plant variable, or a specific kind of disturbance alone (Agapiou *et al.*, 2012a).

Several studies have developed and utilized a number of indices to robustly capture the structural and bio-chemical stress caused by oil pollution on vegetation. Table 2-1 shows some of the VIs developed and used in several studies for detecting vegetation impact of hydrocarbon oil pollution. VIs often used can be either retrieved from multispectral or hyperspectral satellite images. Most studies in this regard often explore several indices in an attempt to find the best Vegetation Index that is most suitable to detect hydrocarbon presence or impact on plant health. This is however necessitated due to natural factors such as the influence of soil background, solar illumination, atmospheric conditions, soil moisture, surface water and the sensor observation geometry, which vary from ecosystem to ecosystem and can alter spectral reflectance characteristics (Hadjimitsis *et al.*, 2010).

Table 2-1: Vegetation Health Indices derived from Multi-spectral Broad-band Vegetation Health Indices (BVI's) and Hyper-spectral Narrow-Band Vegetation Indices (NBVI's) Investigated by several authors on the detection of stress caused by oil pollution.

Author	**Indices Explored**	**Category**	**Main result**
(Adamu *et al.*, 2015)	ARVI2, ClGreen, EVI, EVI2, GBNDVI, GLI, G\NIR, G\RED, G\SWIR, GRNDVI, MSAVI2, MSI, MSR705, NBR, NDVI, NIR\R, PPR, SAVI, SRI, TNDVI	BVIs	NDVI, SAVI, ARVI2, G/NIR and G/SWIR; were found to be sensitive to the effects of oil pollution
(Adamu *et al.*, 2016)	NDVI and NDWI	BVIs	NDVI showed good sensitivity to oil induced stressed over time than NDWI
(Adamu *et al.*, 2018)	NDVI, ARVI2, G-NIR	BVIs	All indices explored showed significant changes in a multi-temporal detection of stress caused by hydrocarbon crude
(Zhu *et al.*, 2013b)	TSAVI, SAVI2, ARVI, OSAVI, SAVI2, ARVI, OSAVI, PSSRc, PSNDc, TCARI, fpnRES, c2RES, GauREA, spnREA, sumREA	BVIs and NBVIs	MCARI and NDVI was observed to best estimate Total Petroleum Hydrocarbon (TPH) concentration than other NBVI's and BVI's respectively

(Arellano *et al.*, 2015)	SR, NDVI, GNDVI, EVI, ARVI, SG, PSSRa, NDVI705, mSR705, mNDVI705, CTR 2, LIC 1, PSNDa, OSAVI, MCARI, Der725-702, REP, VOGI, CI590, MTCI, SIPI, RG, ARI1, ARI2, WBI, NDWI, MSI, NDII, NHI	BVIs and NBVIs	SR, NDVI and $NDVI_{705}$ were found to be more appropriate to detect the effects of petroleum pollution
(Onyia *et al.*, 2018)	NDVVI, NDVI, RENDVI, MRENDVI, MRESRI, VREI1, PRI, SIPI	NBVIs	NDVVI showed strong linkage with vascular plant species diversity and associated stress caused hydrocarbon pollution compared to other indices
(Khanna *et al.*, 2013)	NDVI, mNDVI, NDII, ANIR, ARED, ADW1, ADW2	NBVIs	All indices were found suitable in detecting oil induced stress with the exception of the ANIR
(Khanna *et al.*, 2018)	NDVI, NDVII, ANIR, ARed	NBVI's and BVIs	It was observed that ARed and NDII exhibited better performance at differentiating vegetation stress due to oil contamination
(Shapiro *et al.*, 2016)	NDVI, mNDVI, NDII, ANIR, ARed, ADW1, ADW2	BVIs and NBVIs	Non of the indices were best at measuring oil impact across the sites investigated but the indices consistently differed between oiled and unoiled pixels

(Susantoro *et al.*, 2018)	NDVI, ENDVI, LAI	BVIs	Observed that ENDVI is better than NDVI in detecting vegetation stress as result of oil pollution
(Ramsey *et al.*, 2015)	LAI, LAD	BVIs	The two indices explored showed substantial changes and relationship with Polarimetric SAR images in detecting stressed marsh
(Hese and Schmullius, 2009)	NDVI	BVIs	NDVI was able to differentiate between impacted and oil free vegetation

In addition, Clerici *et al.* (2017) have particularly noted that the use of multiple vegetation health indices for detection and classification purpose can also systematically help to counteract the effect of diversity in vegetation structural stand and species (especially in image classification operation). This is because the use of a single index may not capture the variability of the diverse species present in a particular area.

2.3 Application of SAR data for On-shore Oil Spill Detection

Synthetic Aperture Radar (SAR) is a form of Radar that is used to create two-dimensional images of objects. It uses the motion of its antenna over a target area to provide finer spatial resolution than the conventional Real Aperture Radar, such as the Side-Looking Airborne Radar (SLAR) or Side-Looking Radar (SLR) (figure 2-5).

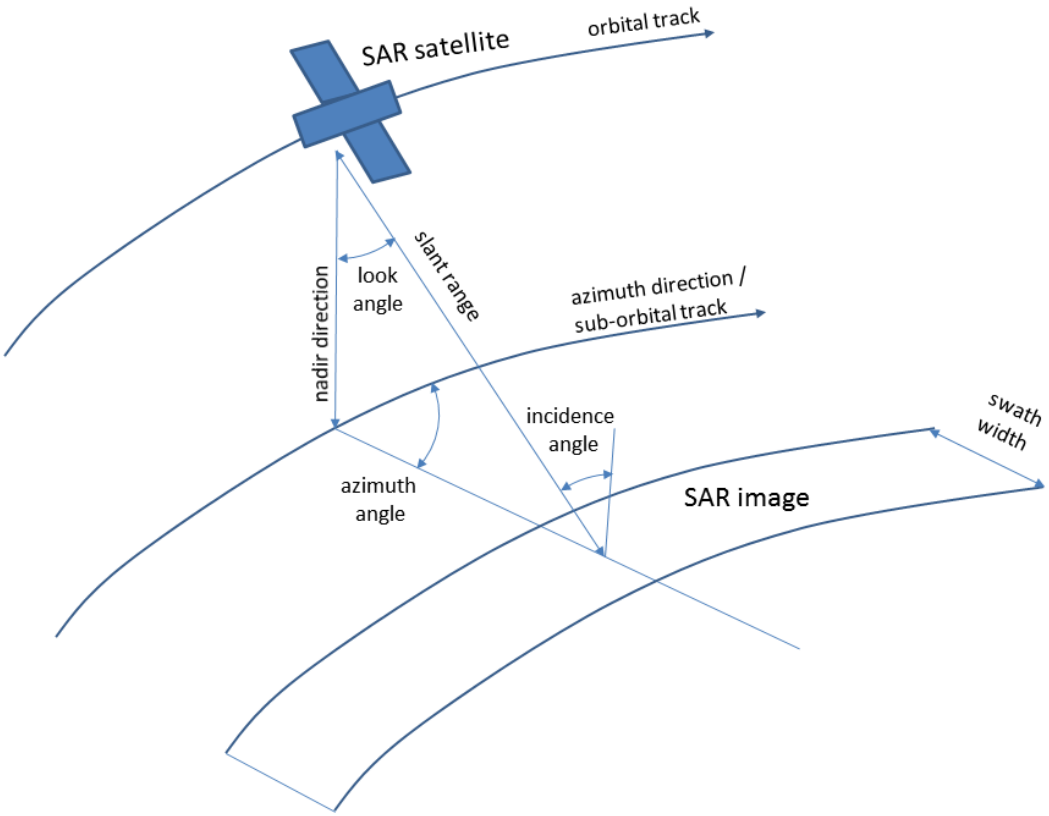


Figure 2-5: Acquisition geometry of a Satellite SAR system. Source: (Spies *et al.*, 2013)

Synthetic Aperture Radar (SAR) sensors are active sensors, as they do not rely on the sun energy. Rather they are equipped to beam electromagnetic pulses over a target. SAR generally operates within the microwave part of the spectrum and capable of observing the Earth's surface in all-weather conditions, day and night. The operational principle behind the capture of a SAR image relies on emitting of pulsed beam on the side of the antenna, unto a target (earth surface) and measuring the intensity and time taken for the signal to return to the platform. Signals that take a longer time to return are reflected much farther from the sensor in the range direction and as the platform moves, each pulse gives information about the target in the azimuth (Lillesand *et al.*, 2014).

The intensity of a SAR image is a measure of the strength of the returned signal, and is affected by geometric and dielectric properties (the moisture content) of the target. The strength of the returned signal for a particular pixel depends on a complex addition of numerous signal interactions with surface scatterers contained within the target area (Ban *et al.*, 2015).

Emitted signals from a SAR sensor is polarized in either Horizontal (H) or Vertical (V) planes, and the returned signal is also received in either Horizontal or Vertical planes. Co-polarized SAR data (VV - vertical send, vertical receive, and HH – horizontal send, horizontal receive) has less utility than cross-polarized (HV and VH) SAR data for vegetation monitoring as the configuration of cross-polarized sensor is more sensitive to the changes in polarization produced by volume scattering elements within a tree canopy (Mitchard *et al.*, 2011).

Full Polarimetric SAR data provides a scattering matrix (consisting of intensity and phase) of a terrain's surface at the four polarizations channels of HH, HV, VH, and VV (Boerner *et al.*, 1997; Henderson and Lewis, 2008). Recorded Polarimetric data allows for a better characterization of features of interest based on backscatter returns or feedback from the backscatter decomposition (Cloude and Pottier, 1996).

In principle, the backscattering properties of a Radar signal is a function of the electromagnetic wave frequency, its incidence angle, its polarization, the target geometry and dielectric properties of the observed medium (Ouarzeddine *et al.*, 2007; Ulaby *et al.*, 2015). Figure 2-6 shows the position of the Radar portion within the microwave part of the electromagnetic spectrum.

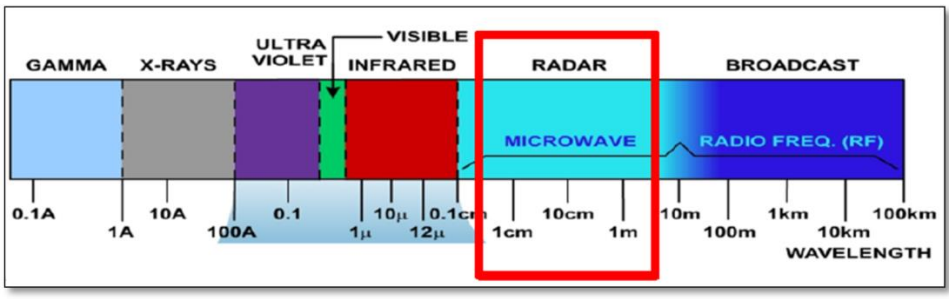


Figure 2-6: The Electromagnetic Spectrum Highlighting the Microwave/Radar Region of the Spectrum (Fingas and Brown, 2014)

Synthetic aperture radar signals in longer wavelengths have higher penetration of vegetation canopies. Figure 2-7 is an illustration of the penetration capabilities of the various Bands of SAR Sensors. L-band and P-band sensors have the highest penetration capabilities, as opposed to C-band and X-band SAR. The longer wavelength SAR has the utility of depicting detailed structural changes in vegetation canopies at the understory and sub canopy level compared to shorter wavelengths (Ghulam *et al.*, 2014; Lee and Pottier, 2009).

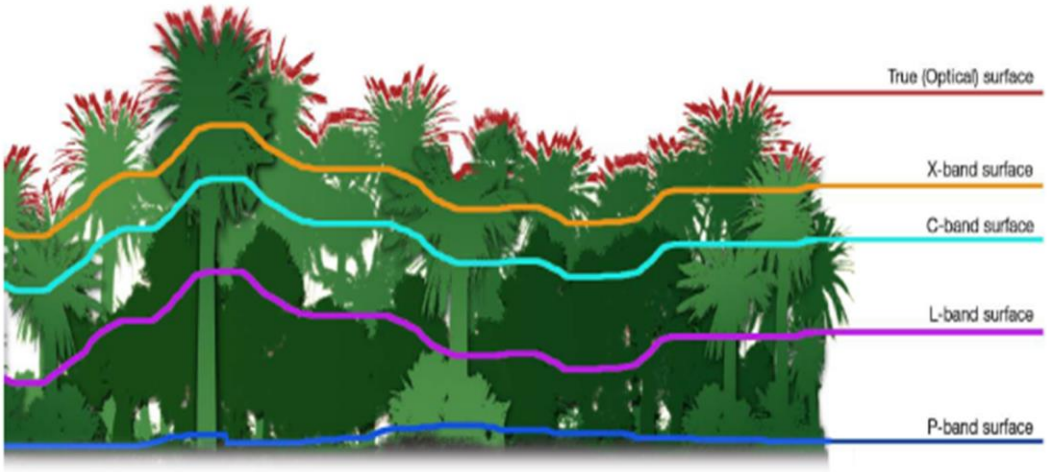


Figure 2-7: The Penetration Capability of the Different Bands of SAR Sensors (Woodhouse, 2005)

Literature survey on existing application areas of Polarimetric SAR data have shown that 3 common approaches are normally used in the utilization and analysis of SAR images for land cover assessment, change detection, anomaly detection, biomass estimation and modelling, and image classification. These are backscatter decomposition, Multi-frequency SAR Image Integration and Multi-temporal SAR Image Analysis.

2.3.1 Backscatter Decomposition

The concept behind a Polarimetric decomposition is that backscattered SAR signals can be deconstructed to extract meaningful information about the structure of the ground target, the scattering mechanism of the return signal as well as the apparent shift in the phase of the signal from the target (Henderson and Lewis, 2008; van Beijma *et al.*, 2014). Several methods have been used to obtain earth surface characteristics from full Polarimetric data. Eigenvector decompositions of a coherency or covariance matrix were proposed by Cloude and Pottier (1996). An approach to the three-component scattering mechanism model was presented by Freeman and Durden (1998), and a four-component approach was presented by Yamaguchi *et al.* (2005). Figure 2-8 shows the various components of the Freeman and Durden; and Cloudier Pottier decomposition algorithms.

These methods have formed the basis for several Polarimetric SAR (PolSAR) applications in the areas of vegetation mapping and biomass estimation (Kattenborn *et al.*, 2015) as well as invasive species mapping (Ghulam *et al.*, 2014).

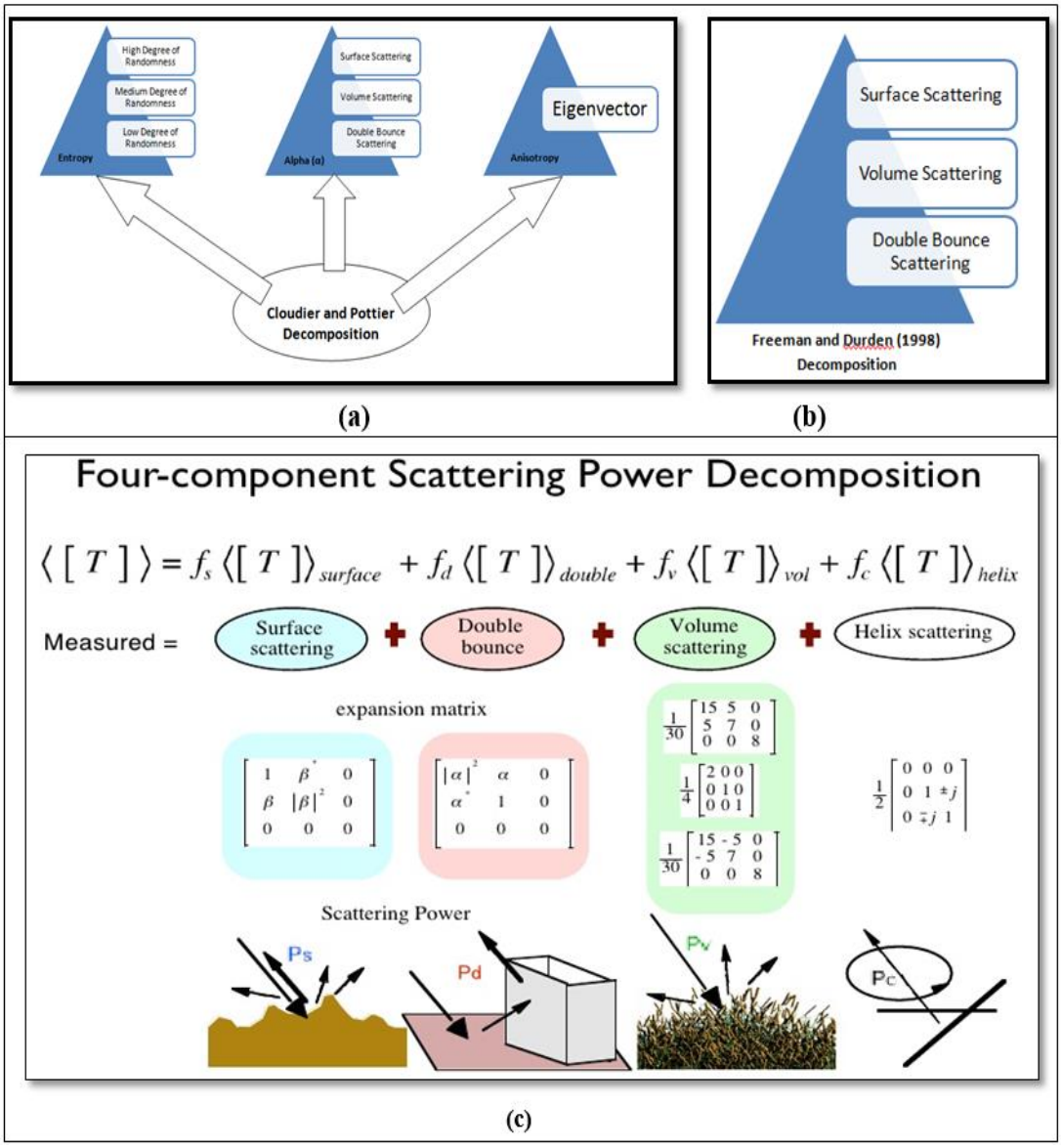


Figure 2-8: Polarimetry SAR Decomposition Techniques. (a) Cloudier and Pottier Decomposition (b) Freeman and Durden Decomposition. (c) Four-component scattering Power Decomposition. Adapted from (Cloude and Pottier, 1996; Freeman and Durden, 1998; Yajima *et al.*, 2008; Yamaguchi *et al.*, 2005)

Polarimetric decomposition relies on the scattering matrix, which describes mainly how radar targets or the elements on the ground surface scatter the electromagnetic energy. The matrix is of dimension (2x2) and built for each Pixel. (Touzi and Chabonneau, 2002) define it as:

$$S = \begin{pmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{pmatrix} \quad (1)$$

Where

S= Scattering Matrix

S_{HH}= Horizontal Send and Receive Polarization

S_{HV}= Horizontal Send and Vertical Receive Polarization

S_{VH}= Vertical Send and Horizontal Receive Polarization

S_{VV}= Vertical Send and Vertical Receive Polarization

The pauli vector, similar to other vectors such as the Stokes Parameter, is obtained from the complex pauli spin matrix basis set. This splits the scattering matrix into the sum of elementary scattering matrices, each one defining a deterministic scattering mechanism (Cloude and Pottier, 1996). The Pauli vector is given by:

$$\acute{K}_p = \frac{1}{\sqrt{2}} \begin{bmatrix} S_{hh} + & S_{vv} \\ S_{hh} - & S_{vv} \\ 2S_{hv} & \end{bmatrix} \quad (2)$$

The first element of the vector (S_{hh} + S_{vv}) expresses odd bounce scatter such as the sphere, the plane surface or reflectors of trihedral type. The second vector (S_{hh} - S_{vv}) is related to a dihedral scatter or double isotropic bounce, while the third element ($2S_{hv}$) is related to horizontal and cross polarization associated to the diffuse scattering or volume scattering. The polarimetry decomposition is based on the coherency matrix given by Cloude and Pottier (1996).

However, there is limited application of this SAR processing approach for general land cover assessment owing to several difficulties relating to limited capability of separating PolSAR signals of different surface types (Cable *et al.*, 2014). Recently, the characterization and classification of marshland and oil spill features using Airborne polarimetric SAR data was investigated by Ramsey Iii *et al.* (2011) and Ramsey *et al.*, (2015). The study utilized Freeman-Durden (Freeman and Durden, 1998) and Cloude-Pottier (Cloude and Pottier, 1996) decomposition to deconstruct SAR backscatter signals before implementing a Wishart classification (Cloude and Pottier, 1996). It was found to be particularly useful given the constraints of weather primarily affecting optical images in the mapping of inland oil pollution. Their result

sufficiently proved the utility of UAVSAR data in depicting changes in interior marshes, possibly as a result of the impact of the deep water horizon spill. They further stressed that high variability in canopy structure even within the same environmental composition can be profound. As such, this suggests that a multi-polarization SAR sensor can be better suited than one with single or double polarizations, because other additional information can be used to better differentiate changes in backscatter that are caused by changes in vegetation structure (Ghulam *et al.*, 2014).

2.3.2 Multi-frequency SAR Image Integration

This approach to the use of SAR images for detecting and mapping land cover has gained a lot of attention recently, owing to its capability in capturing vegetation variability and further providing precise land cover extent. The approach simply entails the integration of multiple SAR images acquired from different SAR sensors across different wavelengths and polarizations for classification. The approach is complementary, as dominant backscatter return and phase, from the different frequencies (multi-frequency), polarizations and temporality of SAR images adequately captures the heterogeneity and characterize the variability that often exists in vegetation canopy stands in a typical terrestrial landscape.

Improved overall classification result using multi-frequency SAR integration obtained in Mohammadimanesh *et al.*, (2018) sufficiently justifies this position. Multi-frequency SAR features ranging from intensity, interferometric coherence and Polarimetric scattering were derived from RADARSAT – 2 and TanDEM-X images to map the wetland extent in the north eastern part of Newfoundland. Similarly results obtained from van Beijma *et al.,* (2014) have also shown that the integration of multi-frequency SAR image variables has the potential to increase overall classification accuracy, as the multi-polarization and multi-frequency of integrated SAR images captures the variability of vegetation structural change across space. Several studies (Ghulam *et al.*, 2014; Rajah *et al.*, 2018) have also used multi-frequency SAR variables for mapping and distinguishing native from invasive species. Their results show moreover that the integration of optical in addition to multi-frequency SAR image derivatives can improve intra-vegetation class discrimination further. This is especially the case when more advanced techniques as machine learning, data fusion and object based image classification methods are used.

To the best of my knowledge, no current published study has used the multi-frequency SAR analysis approach to study the variability and changes in vegetation composition as a result of the impact of oil spills on the terrestrial habitat. This approach could have a huge potential in characterizing vegetation affected by oil pollution, as the basic backscatter theorem stipulates that variability in the structural composition of the vegetation canopy would significantly alter backscatter attributes (Cable *et al.*, 2014). As such, vegetation exposed to consistent stress conditions may lose foliage and canopy cover might drop, which has the potential of altering the scattering characteristics of the canopy. Most of these changes would be detected if various frequencies of SAR images across different wavelengths and temporal domain are used harmoniously for the combined process of detecting and mapping polluted vegetation.

2.3.3 Multi-temporal SAR Image Analysis

The increasing availability and improvements in both spatial and temporal resolution of SAR images have seen the emergence of various multi-temporal application of SAR images to monitor vegetation changes and capture the phenological characteristics. Several studies have employed SAR as a meaningful alternative to optical data in multi-temporal time series analysis. Gao *et al.* (2018) used Sentinel – 1 SAR in a time series assessment for irrigation mapping. Vreugdenhil *et al.* (2018) also relied on multi-temporal sentinel – 1 SAR backscatter data to establish vegetation dynamics and map various cropland features in Australia. Similarly, Rüetschi *et al.* (2017) also used multi-temporal sentinel – 1 SAR backscatter data to monitor phenology and map deciduous and coniferous forests in Northern Switzerland.

Multi-temporal SAR image analysis has also seen little to no application in terrestrial oil spill monitoring. As previously mentioned, SAR is an all – weather, day and night sensor, capable of capturing the structural change of vegetation through temporal backscatter return. As such it can offer the potential for the extensive assessment of an oil-affected area over a long period of time. This could give the opportunity of fostering a deeper understanding of how factors such as seasonality and environmental changes can influence the impact recovery cycle of oil-affected vegetation.

Research by Ramsey Iii *et al.* (2011) and Ramsey *et al.* (2015) as previously mentioned remains the foremost application of SAR data (in single frequency mode and Polarimetric decomposition) for characterizing inland vegetation as a result of oil pollution. There is certainly need for more research in the use of microwave datasets, especially within the tropical and densely forested ecosystems like the Niger Delta region of Nigeria, where persistent spills, most of which are located inland, occur frequently. The varying physical and geographic conditions within the Niger Delta region, coupled with the extensive crude oil pipeline facilities and oil incidents provides a rare opportunity for the assessment of multi-frequency and multi-temporal SAR in detecting the effects of crude oil spills on vegetation.

In addition, the potentials of Spaceborne SAR as opposed to UAVSAR (Ramsey *et al.*, 2015; Ramsey Iii *et al.*, 2011); and single frequency SAR (Ramsey *et al.*, 2015; Ramsey Iii *et al.*, 2011) compared to multi-frequency SAR data has not been fully implemented to assess and map the impact posed by oil spills on the terrestrial landscape. This is particularly important given the cost prohibitive nature of acquiring airborne/UAV SAR data, the penetration capabilities of various multi-frequency Spaceborne SAR and many of which are becoming freely accessible. As such, the potential of freely accessible Space borne SAR data needs to be investigated and effectively harnessed to establish the impact posed by oil pollution on the various vegetation types on the terrestrial landscape.

2.4 Oil Spill Induced Land Cover Assessment in the Niger Delta Using Remote Sensing techniques

While terrestrial oil spills in the Niger Delta region of Nigeria remain recurring phenomena, the methodological approaches employed to assess the detrimental consequences of this situation on the environment still lack cutting edge precision. The use of popular remote sensing satellite data and techniques has not produced exceptional results, and the reason for this is mostly associated with the immense cloud cover experienced within the region all year round, which inadvertently constrains the use of optical satellite images for near real-time monitoring, characterization and mapping changes associated with terrestrial oil pollution.

Attempts have been made to map land cover changes in the Niger Delta as a result of the direct influence of anthropogenic activities, and secondary influences such as oil

pollution and climate change. Kuenzer *et al.* (2014) used Landsat time series images to assess Niger Delta land surface dynamics over a period of 3 decade spanning from 1986 to 2013. This depicted the extensive changes in the broad land cover classes, assessment of coastline dynamics and manifestation of oil exploitation activities. Results showed a significant amount of mangrove lost, expansion of the urban cities, increase in gas flaring and expansion of oil exploitation activities have occurred over the period.

Similarly, Ayanlade and Howard (2016) also used Landsat images with decadal intervals spanning between 1984 and 2011 to assess the spatiotemporal change in land cover in Tsekelew area of the Niger Delta. The study integrated both social media data and remote sensing image analysis to establish the environmental impact of oil production within the study area. Result of the NDVI analysis showed that several mangrove forests in the study area were degraded as a result of oil exploration activities, while they also observed that assessing the impact of oil spills empirically was not possible in an image classification operation, as a result of confusion among several land cover features. Contrary to this position, Ochege *et al.* (2017) employed Landsat image derived NDVI and maximum likelihood image classification to assess dominant vegetation change between 1987 and 2013 to establish the impact of human induced oil exploration activities and oil pollution. Result from their study suggested a decline in healthy forests and vegetation cover between 1987 and 2002, conversely in 2013 a vegetation gain was observed based on NDVI differential, which they linked to the effect of oil and gas exploration within the case study.

In the area of microwave remote sensing, research carried out by Balogun (2015) on the utility of microwave remote sensing in oil spill detection on mangrove did not yield any significant results. This possibly is due to the fewer variables and algorithms explored in the study. However, they came to the conclusion that further research on the use of SAR data for terrestrial oil spill monitoring is needed due to the limited ability to discriminate between backscatter of oil on land and that of soil with high water content. As such it is recommended that a synergy and integration of optical and SAR image could provide a strong medium for mapping oil spill on mangrove.

In general, there is limited application of contemporary SAR imagery in the detection and mapping of oil-impacted vegetation within the tropical-mangrove ecosystem. Studies as (Ramsey *et al.*, 2015; Ramsey Iii *et al.*, 2011) still remains the only substantial application of single frequency SAR imagery for terrestrial oil spill on marsh land cover assessment.

2.5 Machine Learning Algorithms for Satellite Image Processing

The last decade has seen a rise in the development and utilization of more sophisticated algorithms such as artificial intelligence and big data analytic approaches for remotely sensed satellite images. Of particular interest are the use of machine learning algorithms. Machine learning itself has been defined by Arthur Samuel in 1959 as a "Field of study that gives computers the ability to learn without being explicitly programmed", through pattern recognition and computational learning theory in artificial intelligence (Puissant *et al.*, 2014). While some machine learning algorithms do not require and assume specific data distribution to separate multi-dimensional feature space into classes, others such as linear regression methods do require and assume specific data distribution characteristics to make inferences. They rely extensively on learning data characteristic values from which tentative inferences are deduced. Examples of commonly used non-parametric machine learning classifiers include Decision Trees, Support Vector Machines, Artificial Neural Networks and Random Forest.

A wide range of methods have been employed to analyze multi-temporal remote sensing satellite images for detecting oil induced stress on vegetation change. While some image processing methods are based on the post classification comparison of direct image classification for multi-temporal change (Li and Yeh, 1998; Mas, 1999; Singh, 1989), others are based on image regression, band ratio, change-vector analysis and principal component analysis (Li *et al.*, 2005; Singh, 1989).

Several terrestrial oil spill studies have used a number of machine learning algorithms for the assessment of the impact of oil contamination on mangrove and marshland. Giri *et al.* (2011), used a decision tree classifier based on a univariate decision tree (C45.5) algorithm to classify Landsat and Airborne Photography of the Louisiana Mangrove, to depict the spatio-temporal characteristics of ecosystem shift, in terms of expansion, retraction and disappearance. Results of visual comparison with

Google Earth and discussion with wetland experts indicate high confidence of classification results. Khanna *et al.* (2013) used a binary decision tree based on vegetation index, angle index, and depth of oil absorption continuum to produce a classification map for six classes (i.e. oiled soil, oiled dry vegetation, oil-free soil, oil-free dry vegetation, green vegetation and water) to asses oil impact on marshland vegetation on the Louisiana coast. Results also demonstrated the utility of machine learning algorithms in depicting the impact of oil, using vegetation indices localized within pixels of contaminated and non-contaminated pixel. Recently, Mahdianpari *et al.* (2018) also employed the random forest machine learning method to map and discriminate land-based oil spills from other land cover features using high spatial resolution unmanned aerial vehicle imagery and electromagnetic induction survey data.

Random forest is one of the most popular and effective machine learning methods commonly used for satellite image processing (Jhonnerie *et al.*, 2015; Juel *et al.*, 2015; Mutanga *et al.*, 2012; Rodriguez-Galiano *et al.*, 2015; van Beijma *et al.*, 2014). Proposed by Breiman (2001), the Random Forest (RF) method itself is an ensemble of classification trees, where each tree contributes with a single vote for the assignment of the most frequent class to the data (Breiman, 2001). RF typically uses a random subset of the input features or predictive variables in the subdivision of every node, as opposed to using the best variables, thereby reducing error from generalization (Breiman, 2001; Guo *et al.*, 2011; Rodriguez-Galiano *et al.*, 2012). It uses bootstrapping to form an ensemble of classification trees and it generally works well with the identification of strong predictors of a specified outcome without making assumptions about an underlying model.

However, a common problem of high-dimensionality in the presence of correlated predictors impacts RF ability to identify the strongest predictors by decreasing the estimated importance scores of correlated variables (Darst *et al.*, 2018; Nicodemus and Malley, 2009; Strobl *et al.*, 2008). To this end, Millard and Richardson(2015) have recommended that high dimensionality owing to use of high number of input feature variables should be eliminated as much as possible from RF operation, using only uncorrelated important variables for classification for a better prediction to be achieved. Studies such as Mohammadimanesh *et al.* (2018) resolved the issue of high dimensionality with Spearman's rank order correlation amongst and between input

variables. This was used to select only important non-correlated features for a random forest image classification. This allowed for a high overall classification accuracy of 85.40% in a wetland classification in the Newfoundland and Labrador area of Canada.

In this research, the potential of Fuzzy Forest (FF) machine learning method is explored for the prime purpose of effective selection and use of only high performing non-correlated variables for image classification process. FF is a relatively new concept introduced in 2015 by Conn *et al.* (2015) and it is an extension of the generic random forests methodology.

FF is mainly formulated to yield less biased variable importance rankings when there is high correlation among variables. The method is still based on the traditional bagging approach, which infers that each base classifier in the tree is constructed with different subset samples of the input variables (Conn *et al.*, 2015). The theoretical underpinning of the FF method is specifically aimed at using a piecewise screening process to eliminate features from initially assigned variable clusters through Weighted Gene Correlation Network Analysis (WGCNA) for detecting correlation networks. Then a selection phase is implemented through the Recursive Feature Elimination Random Forest RFE-RF process to allow for the interaction between different variable clusters for the selection of unique/important variables from each cluster.

Weighted Gene Correlation Network Analysis (WGCNA) is a biological statistical network tool that is primarily used to analyze genes through a network correlation assessment across micro array of samples (Conn *et al.*, 2015). The method is robust for finding the linkages among gene clusters, which are necessary steps for developing sound clinical gene and cell therapies. The application of fuzzy forest and WGCNA in this research thesis is, to the best of my knowledge, the first in the area of satellite image processing for image derived variable selection and classification, especially for terrestrial oil spill mapping. The method is employed to address the challenge associated with high dimensionality often associated with the use of multiple variables in a single classification operation. This should improve the variable selection process and reduce the redundancy presented by less important variables in the classification. The latter can often translate to less accurate maps

(Darst *et al.*, 2018; Schmidt *et al.*, 2017; Strobl *et al.*, 2008) due to the effect of overfitting or underfitting in the final model from the *n* variables to be used for the classification process.

For the purpose of this research, the random forest, fuzzy forest and support vector machine methods will be interactively explored for the classification of Optical only, Optical and SAR; and SAR only images for effective detection and mapping of oil-impacted vegetation in a case site within the Niger Delta region of Nigeria.

2.6 Research Gap

The following Gaps were identified in the course of **literature review**:

- The integration of optical derived image variables and spectral bands in a machine learning classifier has not been tested for oil spill detection with specific focus on different land cover types.
- The application of multi-frequency and multi-temporal Spaceborne SAR data has not been fully assessed for the specific purpose of terrestrial oil spill impact mapping.
- The effect of seasonality on detectability of oil impact on vegetation has not been investigated and less understood most especially with SAR data.
- No existing study at the moment has leveraged on the variable importance information of random forest to discriminate between oil polluted and oil-free vegetation.
- The fuzzy forest methodology as adapted by Conn *et al.*, (2015) has seen no application in the area of remote sensing satellite image classification to date.
- The potential of the fuzzy forest methodology as an effective variable selection method has not been tested in mapping terrestrial oil spill impact.

2.7 Research Questions

In order for the identified gaps to be critically examined, the following research questions are therefore asked:

I. To what extent can **multispectral optical image only, vegetation health indices and land cover specific classification** be used to detect, map and discriminate between oil-free and oil impacted vegetation types **with machine learning classifier**?

II. To what extent can the **integration of optical derived variables (i.e. multispectral bands and vegetation indices), multi-frequency SAR and geo-physical variables** improve the classification accuracy of polluted and oil-free vegetation types?

III. Can the use of **fuzzy forest resolve the high dimensionality problem associated with the integration of Optical, SAR and Geo-physical variables** (in RQ II above) to improve classification accuracy?

IV. To what extent can **SAR only (i.e. multi-temporal and multi-frequency SAR) variables improve discrimination and classification accuracy (between oil-free and oil-impacted vegetation)** using a machine learning classifier?

V. Can the result in Research Question (RQ) IV be improved by **seasonal stack classification, using multi-temporal and multi-frequency SAR** variables in the study area.

VI. Can **multitemporal SAR analysis depict dominant characteristics** of the polluted and oil-free vegetation

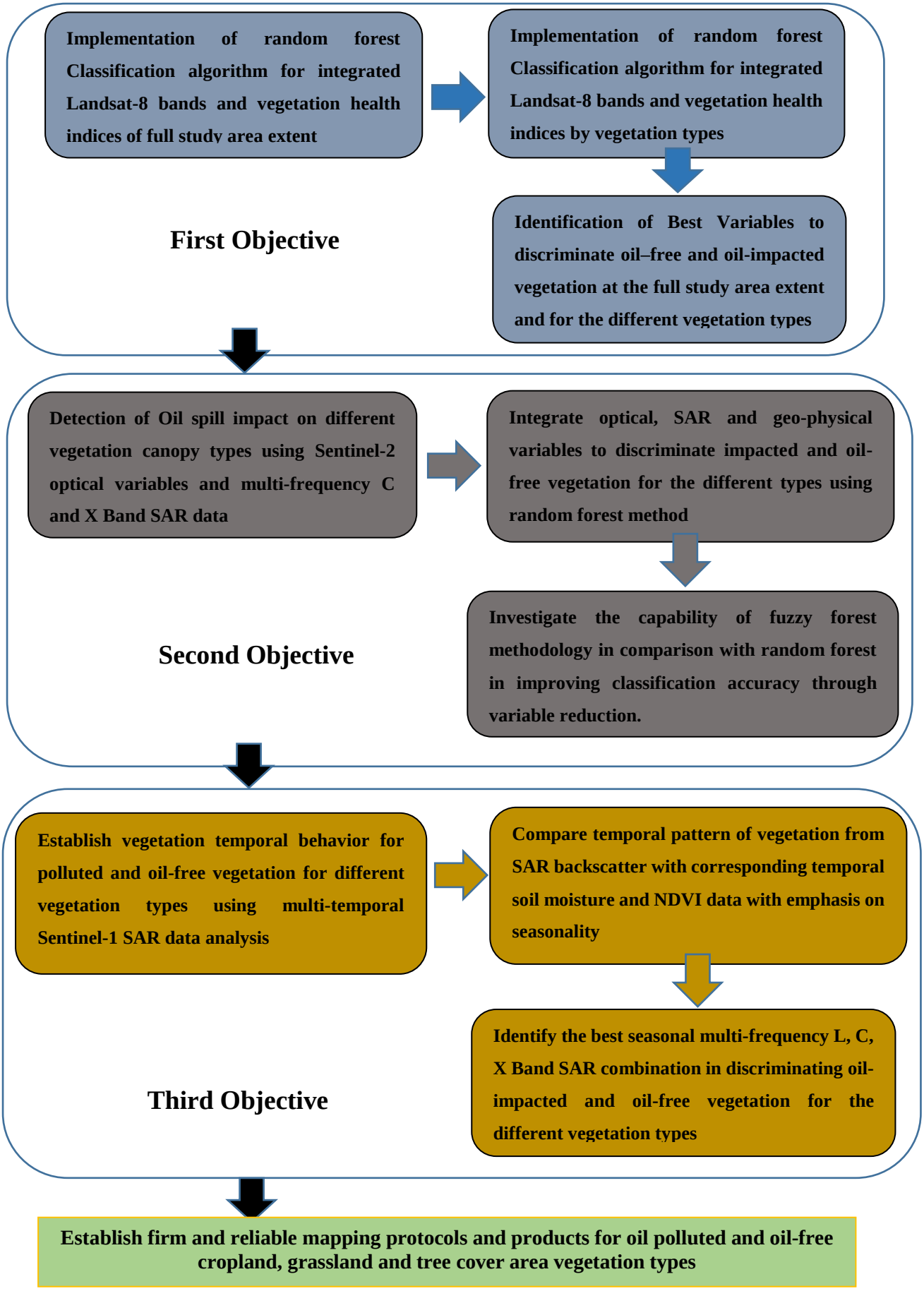
2.8 Rationale of Implementation

Following the establishment of the research questions, the first objective was to integrate optical image bands and the vegetation indices generated thereof to discriminate between healthy and oil-impacted vegetation using a random forest classifier. Emphasis here was to investigate the utility of micro level classification (i.e. for the individual land cover vegetation type) compared to the macro level classification (i.e. at the entire study area image comprising of all land cover types). Particular emphasis was also given to the variable importance as a means of identifying the best discriminatory indices. These approaches has never been tested in the literature for oil pollution mapping, as previous studies achieve this through the use of study area wide image and parametric methods.

However, in order to improve classification accuracy further in the discrimination of polluted and oil-free vegetation, multi-frequency C and X Band SAR variables, as well as other geologic variables (including Digital Elevation Model, Soil - Geology) were introduced into the random forest classifier. Furthermore, the operational paradigm of random forest method was re-worked to further improve classification accuracy. In this regard, the fuzzy forest method was introduced to resolve the problem of high dimensionality associated with the use of multiple SAR derived image variables and Optical variables in a classification process.

In the final analysis, the role of seasonality in the discriminability of polluted and oil-free vegetation was investigated. Here the potential of multi-temporal Sentinel – 1 SAR was assessed to show vegetation temporal characteristics of polluted and oil-free vegetation types. Furthermore, the dry and wet season image stack of the multi-temporal Sentinel – 1 C band SAR, Cosmo Skymed and TanDEM-X X band SAR, and ALOS PALSAR 2 L – Band SAR, were investigated and classified to discriminate between healthy/oil-free vegetation and oil spill impacted vegetation.

2.9 Research Conceptualization into achievable objectives



2.10 Aim and Objectives

2.10.1 Aim

The aim of this PhD is to apply state of the earth remote sensing data and methods to facilitate the detection and mapping of oil-polluted areas on the terrestrial landscape.

2.10.2 Objectives

1. To evaluate the capabilities of Optical Derived Indices and Spectral bands of Landsat – 8 OLI and Machine Learning *RF* to discriminate between oil-polluted and oil-free vegetation (i.e. cropland, grassland and tree cover area in dry and wet seasons).
    a. To investigate the potential of optical derived vegetation indices and spectral bands in a machine learning random forest classifier to discriminate between oil-free and oil (spill) impacted vegetation (land cover).
    b. To investigate the capabilities of a land cover specific (vegetation specific) image classification operations to discriminate oil-impacted and oil-free vegetation types.
    c. To identify the best vegetation indices that best characterize and discriminate between the oil-impacted and the oil-free vegetation.
2. To analyze the potential of Integrating Optical (Sentinel 2) derived variables and Multi-frequency C and X band – SAR backscatter using a fuzzy forest and random forest classifiers to discriminate between the oil-polluted and oil-free vegetation types (i.e. cropland, grassland and tree cover area in dry and wet seasons).
    a. To investigate the potential of various frequencies of SAR (C and X band) backscatter to detect hydrocarbon crude oil impact on vegetation.
    b. To assess the relationship between various optical derived indices and multi-frequency SAR (C and X band) in various vegetation types
    c. To implement random forest classification using the multi-frequency C and X band SAR to discriminate between oil-free and oil-impacted vegetation.

d. To investigate the potentials of the fuzzy forest classifier in comparison with the results of the random forest classifier (above) in addressing high dimensionality associated with SAR (i.e. multi-frequency C and X band SAR) and Optical image integration to discriminate oil-free and oil-impacted vegetation types.

3. To evaluate the potential of Multi-temporal, Multi-seasonal and Multi-frequency L, C and X band SAR; with machine learning support vector machine and random forest classifier in discriminating between oil – polluted and oil-free vegetation types (i.e. cropland, grassland and tree cover area in dry and wet seasons).
    a. To investigate the temporal backscatter characteristics of a typical oil-free and oil-impacted vegetation types.
    b. To assess and compare the temporal SAR backscatter response to optical image derived health response from MODIS NDVI and the SMAP soil moisture.
    c. To investigate the role of seasonality in the detection and discrimination of polluted and oil-free vegetation types in a number of classification scenarios using multi-temporal, multi-frequency and multi-seasonal SAR images, together with support vector machine and random forest classifiers, in an attempt to evolve the best data combination and methodology for discriminating oil-polluted and oil-free vegetation types.

Chapter 3 : General Methodology

3.1 Introduction

The general methodology chapter describes all the datasets used in addressing all of the research objectives. It also highlights the image classification methods and data pre-processing used during the course of data analysis. It goes further to describe the study area with respect to location and size, population, climatic conditions, ecology and geology.

3.2 Data

The dataset used in this research comprises of both optical multispectral image dataset (from Landsat – 8 and Sentinel – 2 A), Synthetic Aperture Radar (from Sentinel – 1 A, Cosmo Skymed, TanDEM – X and ALOS PALSAR), Oil spill Incident data (from SPDC and NOSDRA), Existing Land cover dataset (from ECCI), Soil Moisture data (from SMAP), MODIS NDVI data (from NASA), Digital Elevation Model - SRTM (from NASA), Soil and Geology data (NGSA) and field acquired (Spectroradiometric) primary data.

3.2.1 Optical Reflectance Data

Three optical reflectance data were used in the course of this research to investigate the potential of optical spectral bands and their derived vegetation health indices in discriminating between oil-impacted and oil-free vegetation types. These include Landsat – 8 Optical Land Imager (OLI), Sentinel – 2 A Multispectral Imager (MSI) and the MODIS 1Km Vegetation Indices.

3.2.1.1 Landsat – 8 Imagery

The Landsat 8 – OLI data for the year 2016 was the first multispectral optical image data used in the course of this research. The data was downloaded from the USGS website (earthexplorer.usgs.gov/). The acquired image is a Landsat Surface reflectance Higher – Level Data Product, processed using the Landsat Surface Reflectance Code (LaSRC). The LaSRC makes use of the coastal aerosol band to perform aerosol inversion tests using auxiliary climate data from MODIS and a unique radiative transfer model (Roy *et al.*, 2014). Additionally, LaSRC hardcodes the view zenith angle to "0" and solar zenith, which are used for calculations as part of the atmospheric correction process.

The Landsat – 8 OLI image was used in the 1st analysis chapter (i.e. chapter 4) of this thesis to investigate the capabilities of multispectral optical image spectral bands and derived vegetation indices in discriminating between oil-impacted and oil-free vegetation types in a machine learning random forest classifier. Results obtained from the normalized variable importance chart also offered insight on the best optical variable for discriminating between oil-impacted and oil-free cropland, grassland and tree cover area vegetation types.

The study used a post-spill dry season multispectral image acquired on the 6th of December, 2016. This ensured that there was little cloud, aerosol and haze cover, as images between the months of February to November have significant cloud cover due to the wet season.

3.2.1.2 Sentinel – 2 Imagery

Sentinel 2 is an European Space Agency (ESA) and European Commission satellite sensor launched under the Copernicus joint initiative programme. The satellite was launched on the 23rd of June 2015 with a Multi Spectral Imager (MSI) sensor for monitoring the variability in coastal and terrestrial environment. The sensor acquires multispectral image across 13 spectral channels through Visible (VIS), Near Infrared (NIR) and Shortwave Infrared (SWIR) at a spatial resolution of 10m, 20m and 60m. The Sentinel – 2 data is open access, which means the data is available for the public to download and use free of charge from any of the download platforms. However, the Sentinel – 2 data for this research was obtained from the Sentinel Copernicus Hub (https://scihub.copernicus.eu/dhus/#/home).

In this research, the Sentinel – 2A image was used in the 2nd analysis chapter (i.e. chapter 5) of this thesis to evaluate the capabilities of integrating optical and multi-frequency SAR derived image variables to discriminate oil-impacted and oil-free vegetation types in a machine learning fuzzy forest and random forest classifiers. For this purpose, the post-spill Orthorectified image with Universal Transverse Mercator (UTM) projection and the World Geodetic Survey (WGS) 1984 Datum, Zone 32N acquired in December 2016 was used for analysis. The choice of image date (i.e. December 2016) is premised on the fact that most images of the study area had extensive cloud and haze cover all year round (especially between February and November). The research used all the image bands with 20m spatial resolution, as

this provided extensive coverage of the near infrared and shortwave infrared (NIR-SWIR) part of the spectrum. This range of the spectrum is particularly efficient for oil spill detection on terrestrial landscape as emphasized by Cloutis (1989) and Hörig *et al.* (2001).

3.2.2 Synthetic Aperture Radar (SAR) Images

SAR images from four sensors with various polarizations were obtained from their respective vendors and used interactively in the course of this research. They include; Sentinel – 1 A (C-Band SAR), TanDEM X (X-Band SAR), Cosmo Skymed (X-Band SAR) and ALOS PALSAR (L-Band SAR) images.

3.2.2.1 Sentinel – 1

- SLC Product

In this study, single Sentinel 1 A interferometric wide swath (IWS) single look complex (SLC) image in the ascending mode was acquired for individual months from April 2015 to December 2016 (i.e. one image per month). Thus, 22 sentinel – 1 A-image scenes for the period of April 2015 to December 2016 were acquired, pre-processed and analyzed for multitemporal assessment. The data was obtained from the Sentinel Copernicus Sci-hub (https://scihub.copernicus.eu/dhus/#/home). Sentinel – 1 A is part of the European Space Agency and European Commission Copernicus launched satellite sensors. It was launched on the 3rd of April 2014 to monitor both marine and terrestrial ecosystem for the effective management and monitoring of earth's resources. The sensor is a C – band SAR that acquires amplitude and phase information from VH and VV polarization, with a 250-km swath width and at 5 x 20m spatial resolution. The images used in the course of this research were the VH and VV polarization, as not all locations have the two polarizations in a single capture. Table 3-1 shows the date of the Sentinel – 1 A image acquisition.

Table 3-1: Sentinel – 1 Image List Acquisition Dates

Date	Date	Date	Date
2015 – April 05	2015 – Oct 26	2016 – March 06	2016 - Sept 26
2015 – May 23	2015 – Nov 19	2016 – April 11	2016 – Oct 20
2015 – June 16	2018 – Dec 01	2016 – May 17	2016 – Nov 25
2015 – July 22	2015 – Dec 25	2016 – June 10	2016 – Dec 07
2015 – August 27	2016 – Jan 30	2016 – July 28	-
2015 – Sept 20	2016 – Feb 22	2016 – August 21	-

The Sentinel – 1 A Single look Complex (SLC) image datasets were used in the 3rd analysis chapter (i.e. chapter 6) of this thesis for two major analysis. First, to evaluate the capabilities of Multi-temporal Sentinel – 1 C band SAR backscatter to infer vegetation temporal characteristics to investigate the role of seasonality in the detectability of the impact of oil spill on the predominant vegetation types. Secondly, the Sentinel 1 multitemporal images were integrated with L – Band Advance Land Observation Satellite Phased Array type L-Band Synthetic Aperture Radar 2 (ALOS PALSAR-2) and X – Band TanDEM – X in a seasonal image classifications to further discriminate between oil-impacted and oil-free vegetation types.

- GRD Product

The Sentinel 1 C – band SAR, Level-1 Ground Range Detected (GRD) format, was downloaded and used in the course of this research, primarily because of its utility in providing a radar cross-section of both distributed and point targets. This image was used in the 2nd analysis chapter (i.e. chapter 5) of this thesis to evaluate the capabilities of integrating multispectral optical and SAR derived image variables in a machine learning fuzzy forest and random forest classification, to discriminate oil-impacted from oil-free vegetation types. For the purpose of this study, a single post spill image of January 27 2017 with VH and VV polarizations was downloaded, pre-processed and used in fulfilling this research objective.

3.2.2.2 TanDEM – X

Tandem – X data consisting of multiple scenes covering the study area, were acquired from the German Space Agency (DLR) and used in the 2nd analysis chapter (i.e. chapter 5) of this thesis. The satellite, which is jointly operated by the German Space Agency (DLR) and EADS Astrium, is a bi-static X-band sensor, which consist of two twin satellites, namely: TerraSAR-X (launched June 15, 2007) and the TanDEM-X (launched June 21, 2010). It was the first X-band single-pass SAR interferometer in space. The images used in this study are both post spill wet and dry season images captured in April 2016 and February 2017, respectively. The two image sets (i.e. both wet and dry season) were used in chapter 6 for the multi-temporal, multi-frequency and multi-seasonal based image classification. Only the dry season post-spill image was used in chapter 5 for the multi-frequency SAR and Optical image classification, using the fuzzy forest and random forest methods. The images are Level 1b Geocoded Ellipsoid Corrected (GEC) images captured in Strip map mode.

3.2.2.3 Cosmo Skymed

Cosmo Skymed is a high resolution SAR launched and operated by the Italian space agency (ASI) acquiring backscatter return in dual polarization mode of HH and HV. Several image scenes covering the study area were requested and acquired through the COSAIR programme. The image used in this study is a level 1A image, which needed to be corrected for the Range Spreading loss effect using antenna pattern gain compensation and incidence angle effect (Sportouche *et al.*, 2012). The images used in the course of this research were post-spill images with acquisition dates in December 2016 in the dry season. The Cosmo Skymed images was used in both the 2nd analysis chapter (i.e. chapter 5) and the 3rd analysis chapter (i.e. chapter 6) of this thesis. These were used to evaluate the capabilities of integrating optical and SAR derived image variables in both machine learning fuzzy forest and random forest classifications, as well as multifrequency SAR seasonal image-based classification respectively, to discriminate between oil-impacted and oil-free vegetation types.

3.2.2.4 ALOS PALSAR

The Advance Land Observation Satellite Phased Array type L-band Synthetic Aperture Radar (ALOS PALSAR) data as Single Look Complex (SLC) at processing level 1.1, in slant range and HBQ mode of acquisition (high sensitivity) was acquired in CEOS format for this research. The images where provided by the japan space

agency (JAXA) for this research. The images acquired were ascending orbit in a left looking off nadir angle mode and has a nominal resolution of 2.86m and 3.13m in range and azimuth respectively. Two image pairs covering the study area for both dry and wet season acquisitions on 28th December 2016 and 23rd May 2016 in HH, HV, VV and VH polarizations were used in the course of this research. The images were used mainly in the 3rd analysis chapter (i.e. chapter 6) of this thesis for the multi-temporal, multi-frequency and multi-season based image classification.

3.2.3 Oil Spill Incident Data

Nigeria's oil-rich Niger Delta is characterized by a prevalence of both old and new oil spills totaling about 9,343 incidents in the last 10 years, according to official records by National Oil Spill Detection and Response Agency (NOSDRA). This translates to an average of nearly one thousand spills yearly, the highest rate of spills globally. NOSDRA reported that, within the period from 2006 to 2015, there were over 5000 spill sites from the over 9000 spills. The average volume of oil spill annually is about 115,000 barrels, totaling around 1,150,000 in the past 10 years (Ndimele *et al.*, 2018).

The spill incident data is an important dataset for this research, as it provides actual locations of oil spill sites and other relevant information for past and present spills. The oil spill datasets used for this research were obtained from two published sources. These are the Shell Petroleum Development Corporation (SPDC) (https://www.shell.com.ng/sustainability/environment/oil-spills.html) oil spill incident data and the National Oil Spill Detection and Response Agency (NOSDRA) (https://oilspillmonitor.ng/) oil spill data. Shell Petroleum Development Corporation is the largest private crude oil extractive company, both on marine and terrestrial ecosystems within the Niger Delta region of Nigeria, while the National Oil Spill Detection and Response Agency is the government outfit tasked with responsibility of capturing and monitoring all oil spill incidents in the country. The agency has published oil spill incidents on its website since 2006 to date.

The incident data are obtained through a joint investigative visits (JIV) that normally follows an oil spill incident. Appendix I and II shows a sample of the SPDC Joint Investigative Visit report and picture photograph of an oil-polluted site visited in Bodo West community. The investigative team is usually comprised of necessary

stakeholders in the oil and gas industry. This includes the government agency, civil society, security agencies, representative of the host community and the principal oil company. SPDC started this mode of spill data capture from 2011 and it remains the only mode through which spills are recorded and published across the country. Information published includes the location (in actual coordinates), cause of spill, estimated volume of oil spilled, estimated area coverage of oil spilled, observed impacts on the environment, the spill boundary coordinates as well as necessary remediation activities recommended or ongoing as at the time of the joint investigative visit.

Figure 3-1 shows the oil spill incident points obtained from both NOSDRA and SPDC from 2011 to 2017 combined, as well as the 2015 and 2016 oil spill incidents within the proposed study site specifically used in this research.

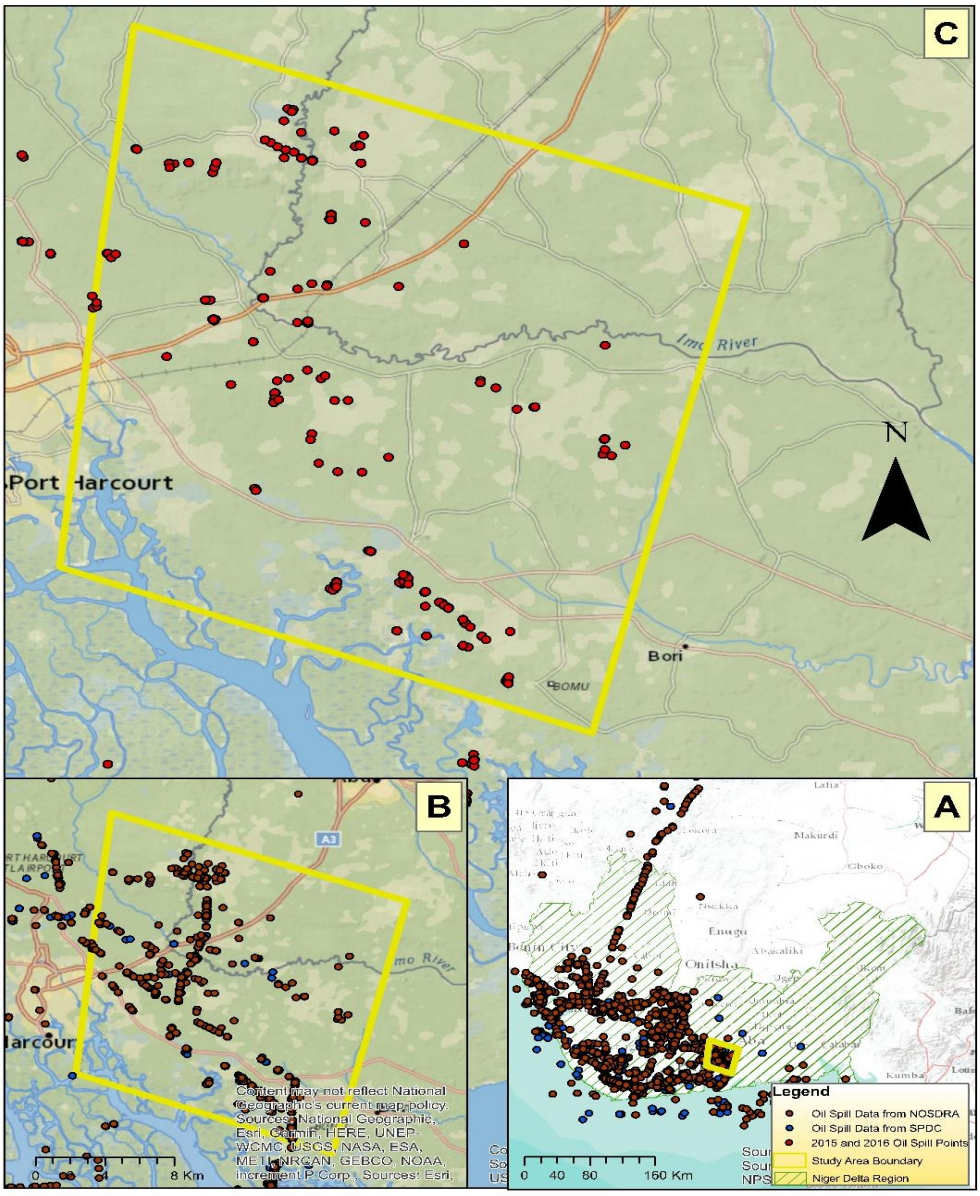


Figure 3-1: Shows the oil spill incident points obtained from both NOSDRA and SPDC from 2011 to 2017. 'inset A and B' shows the combined oil spill datasets from the two published sources, while 'inset C' shows the 2015 and 2016 SPDC/NOSDRA data for the study site investigated.

The two sources (NOSDRA and SPDC) are similar in most respects, that is, they both capture, record and publish oil spill incidents following a joint investigative visit of an oil spill site. Both datasets were used in this research to ensure that a more comprehensive oil spill incident database is established, upon which further deductions and processing would be carried out. For the purpose of this research, the spill data for the year 2015 and 2016 were used for data analysis (Figure 3-1).

The rationale for using the 2015/2016 spill incident data is premised on the fact that it provides a spatial concentration and number of incident points sufficient for image analysis, classification and validation process. In addition, the 2015 and 2016 oil spill dataset also coincided more with the most available optical and SAR image datasets available for analysis. Figure 3-1 shows the oil spill point spatial distribution, while Table 3-2 shows the requisite information concerning the date of spill, location of spill and the total reference points selected within the spill area.

The spill data as shown in Table 3-2 were used in the data analysis in chapter 4 and 5, while the number of reference points was expanded for the analysis in chapter 6. The rationale for the expansion of the number of reference points for the multi-temporal, multi-frequency SAR and multi-season assessment (i.e. in chapter 6), is to ensure that there is spatio – temporal stability between the number of image datasets and the number of samples used for machine learning calibration and classification (Maxwell *et al.*, 2018). Similarly, Khatami and Mountrakis (2018) also noted in this regard that the use of a higher number of training sample can substantially help in improving learning and classification accuracy in a multi-temporal image classification.

3.2.4 Existing Land Cover data

The existing land cover map for the African Continent produced by the ESA Climate Change Initiative (ECCI) 2016 is a freely accessible dataset (accessible via: https://www.esa-landcover-cci.org/?q=node/187) used in the course of this research. This was used to establish the baseline land cover types and their spatial extent for subsequent image analysis. The product contains 10 classes for different land cover classes including built – up areas, waterbody and various vegetation types produced from a 10m high spatial resolution Sentinel – 2A image over Africa. The major land cover classes used in this research are cropland, grassland and tree covered areas. These were primarily used to establish the three vegetation types (cropland, grassland and TCA) investigated for separate spill detection and image classification process. Essentially, the land cover extent was used to subset the various images to extract the various vegetation types in the three analysis chapters of this thesis. Figure 3-2 shows the predominant land cover types used in this research from the ECCI land cover dataset.

Table 3-2: Oil spill incident locations and the total number of ground reference selected

Location	Year/Month	State	LGA	Ref. Points
Atali	01/10/2015	Rivers	Obio/Akpor	3
Ayama	01/06/2016	Rivers	Oyigbo	4
Ayama - Noke	01/12/2015	Rivers	Oyigbo	3
Ayama - Noke	14/01/2016	Rivers	Oyigbo	2
B-Dere	01/06/2015	Rivers	Gokana	3
B-Dere	01/06/2015	Rivers	Gokana	1
B-Dere	14/01/2016	Rivers	Gokana	1
B-Dere	28/05/2015	Rivers	Gokana	3
B-Dere	16/02/2016	Rivers	Gokana	1
B-Dere	01/06/2015	Rivers	Gokana	2
B-Dere	01/06/2015	Rivers	Gokana	2
Egberu	01/08/2015	Rivers	Obigbo	3
Egberu	16/02/2016	Rivers	Obigbo	5
Egberu	01/08/2015	Rivers	Obigbo	1
Ejamah Ebubu	01/06/2015	Rivers	Eleme	1
Ejamah Ebubu	21/02/2015	Rivers	Eleme	3
Ejamah Ebubu	01/06/2015	Rivers	Eleme	1
Gio	01/08/2015	Rivers	Gokana	7
Gio	01/09/2016	Rivers	Tai	7
Gio	01/08/2015	Rivers	Gokana	1
Imeh	01/09/2015	Rivers	Etche	1
Komkom	01/05/2015	Rivers	Oyigbo	4
Komkom	12/05/2015	Rivers	Oyigbo	4
Komkom	01/09/2016	Rivers	Oyigbo	1
Komkom	01/05/2015	Rivers	Oyigbo	1
Kpoghor and Gio	01/05/2015	Rivers	Tai	3
Kpoghor and Gio	16/02/2016	Rivers	Tai	6
Kpoghor and Gio	01/09/2015	Rivers	Tai	5
Norkpo	01/09/2015	Rivers	Tai	4
Odagwa	01/09/2016	Rivers	Etche	1
Ogale	01/08/2016	Rivers	Eleme	5
Ogale	16/02/2016	Rivers	Eleme	1
Owaza	01/05/2015	Abia	Ukwa West	1
Owaza	28/05/2015	Abia	Ukwa West	5

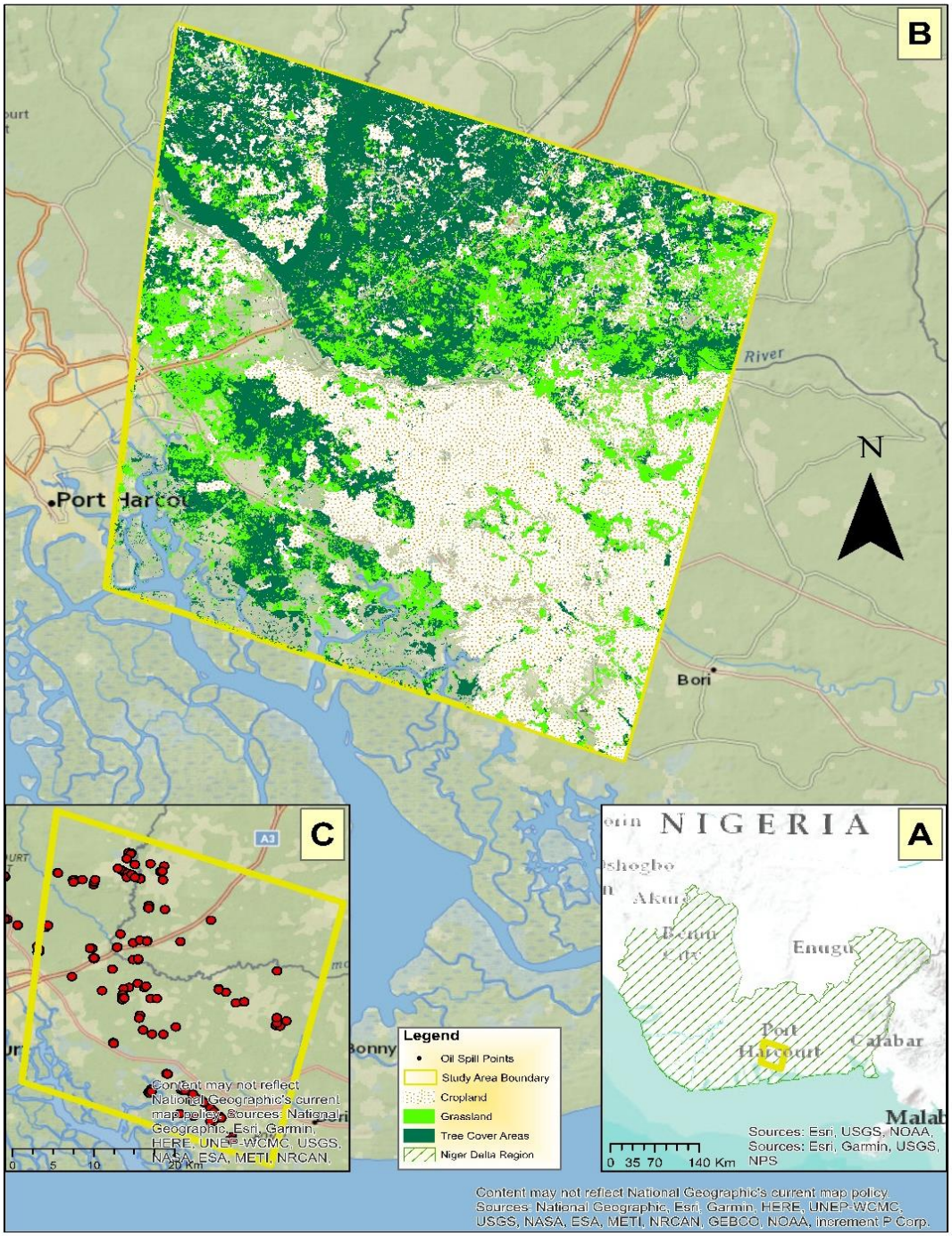


Figure 3-2: The Study Area (B) is shown within the Niger Delta Region of Nigeria (A), indicated by the yellow outline. The red points in 'C' are the oil spill sites as retrieved from SPDC and NOSDRA database. 'B' shows the predominant land cover types (Cropland, Grassland and Tree Cover Areas) as provided by the European Climate Change Initiative Derived from Sentinel 2A dataset. The northern part of the study area has a higher concentration of riparian forest and dense vegetation, while patches of grassland are concentrated in the central part of the study area connecting the much cultivated part of the study area to the south eastern corner.

Pictures of the cropland, grassland and tree cover areas are shown in Figure 3-3a, 3-3b and 3-3c respectively. Features such as built – up areas, waterbody and bare surfaces were excluded from this study as most oil pipelines and the corresponding spill incidents occur on vegetated landscape. Thus, their exclusion should help in reducing artifacts and mis-classification errors as this was major issue observed in previous studies (e.g. Hese and Schmullius, 2009; Van der Werff *et al.* 2007 and Bianchi *et al.* 1995a).

For the purpose of this research, the study area shape file was used to subset and extract existing land cover extent upon which further analysis was carried out. Principal among them is the implementation of the class wise or vegetation specific detection of oil spill impact and the subsequent mapping of the corresponding extents to discriminate between spill-impacted vegetation and healthy vegetation for the different land cover (i.e. vegetation) types in this study.

(a) (b)

(c)

Figure 3-3: Pictures of the predominant land cover types within the study area investigated (a) Cropland (b) Grassland (c) Tree Cover Areas. (Source: Authors fieldwork, January 2017)

3.2.5 Soil Moisture Data – Soil Moisture Active Passive (SMAP)

Soil moisture was used in this study to establish the dominant state of the top soil activity as it contributes to vegetation growth across multi-temporal scale. This has been observed by Kramer and Boyer (1995) as being more efficient and influences plant growth than precipitation, as the influence of rainfall runoff especially in high elevated areas can impede on soil retention capacity. This study used the Soil Moisture Active Passive (SMAP) Soil Moisture (Radiometer) Level 2 data, which nominally represents volumetric soil moisture in the top soil layer of $5m^3$. The data is derived from an L – Band radiometer measurements with a 9km spatial resolution and with a revisit period of 3 days.

In this study, a single monthly acquisition was obtained for the study area for the year 2015 and 2016 (amounting to a total of 24 images for the two years), to effectively align with the dates for the Sentinel 1 A temporal backscatter and MODIS temporal NDVI retrievals for effective comparison. Data acquired for this study are for the descending overpass (6 a.m. local time) as the data effectively demarcates the daily periods as observed by Draper *et al.* (2018). The data were mainly used in the analysis of biochemical and biophysical factors in the 3rd analysis chapter (i.e. chapter 6) of this thesis. This essentially provided an avenue to evaluate and compare the observed multi-temporal backscatter trends and soil moisture.

3.2.6 Moderate Resolution Imaging Spectroradiometer (MODIS) NDVI

The 1 – Kilometer spatial resolution monthly Aqua Moderate Resolution Imaging Spectroradiometer (MODIS) Vegetation Indices (MYD13A3) Version 6 data was used in the course of this research to investigate the bio-physical response of vegetation to oil spill. The monthly product is generated from the integration of all the 16 – day 1 kilometer product using a weighted temporal average. The NDVI image scene covering the study area was obtained free of charge (from https://modis.gsfc.nasa.gov/data/) for the period April 2015 through December 2016 to coincide with the monthly Sentinel – 1 A (SAR) temporal backscatter. The vegetation index was used to establish the temporal variability of vegetation status (as affected by oil pollution) within the study area to provide a basis for evaluating the results of the multi-temporal SAR backscatter analysis. The data were mainly used in the analysis of biochemical and biophysical factors in the 3rd analysis chapter

(i.e. chapter 6) of the thesis essentially to provide an avenue to compare the result of the seasonal and temporal SAR backscatter of polluted and oil-free vegetation types.

3.2.7 Elevation Data (Digital Elevation Model)

Digital elevation model obtained from the Shuttle Radar Topography Mission (SRTM) was used in the course of this research. The SRTM 1 Arc-Second Global dataset acquired 23rd September 2014 was obtained from the USGS website (https://earthexplorer.usgs.gov/). The void filled product resampled from the initial 3Arc-Seconds to 1 Arc-Second was used, which corresponds to a 30m spatial resolution elevation model dataset.

The SRTM mission was one of the earliest missions to produce a global terrain dataset using a C-band Spaceborne imaging radar and X-Band SAR on board the space shuttle endeavor in 1994. These data were used in the chapter 5 of this thesis to investigate the role of elevation and topography in the detectability and discrimination of oil-polluted and oil-free vegetation in both a fuzzy forest and a random forest classification. Previous studies have suggested that elevation plays a significant role when discriminating oil-free and oil-impacted vegetation (Hester and Mendelssohn, 2000; Hester *et al.*, 2016).

3.2.8 Soil and Geology Data

Soil type and geology as compiled by the Nigerian Geologic Survey Agency (NGSA) http://ngsa.gov.ng/GeoMaps were obtained, pre-processed and incorporated into the image classification process in the 2nd analysis chapter (i.e. chapter five) of this thesis. The data were incorporated to investigate the role of soil characteristics and the underlying geologic landforms in the detectability and discrimination of oil-polluted and oil-free vegetation. Previous studies (Abdel-Moghny *et al.*, 2012; Klamerus-Iwan *et al.*, 2015; Wang *et al.*, 2013) have suggested that soil type and geological composition of an area can play a significant role in the detection of hydrocarbon and oil spill impact, as this to a large extent determines penetrability of oil through the soil column to the vegetation root layer.

3.2.9 Field Validation Exercise

Two fieldwork exercises were undertaken in the course of this research. The exercise were mainly conducted to evaluate the results obtained in the course of this research. This are discussed under fieldwork 1 referring to the oldest (January 2017) and fieldwork 2 referring to the most recent (November 2018)

3.2.9.1 Fieldwork – 1

The objective of the first fieldwork exercise was to implement field spectroscopy with the aid of a Handheld 2 Spectroradiometric device to establish vegetation spectral properties. This fieldwork exercise was conducted between 24th and 28th January 2017 (see Table 3-3) in Bayelsa state of Nigeria, specifically in Ikarama community located in Kaiama Local Government Area.

One recent spill site, one old spill site and a non – spill (oil-free) site were visited as part of the investigation conducted. The Old spill site is a spill location where the oil spill incident occurred as far back as January 2010, while the recent spill site is a spill incident site which occurred in November 2016.

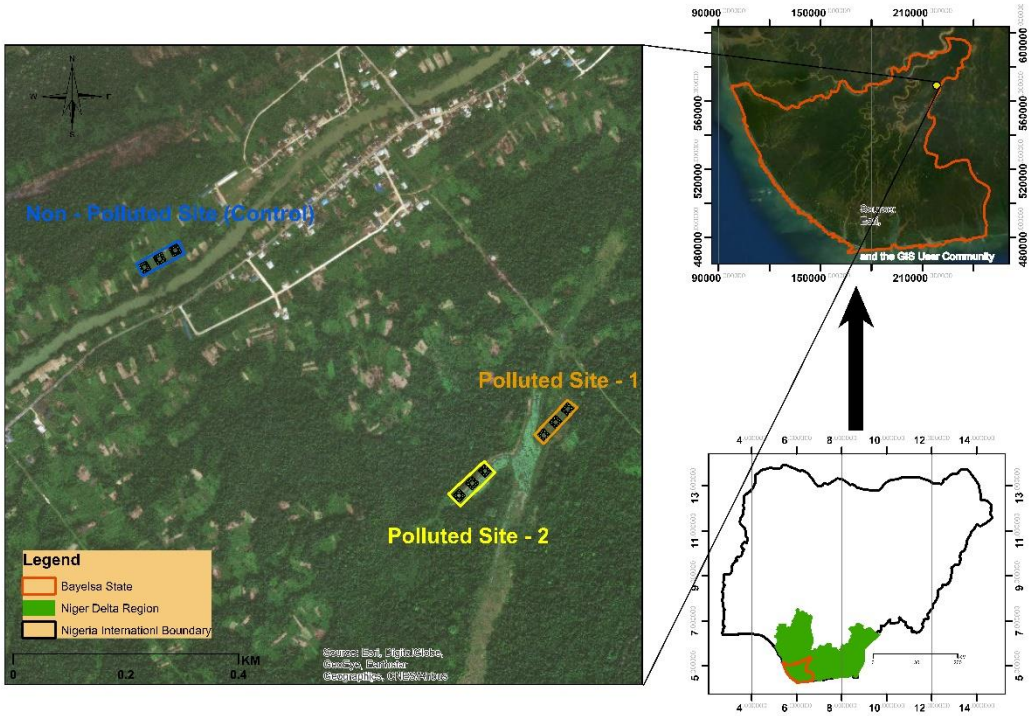


Figure 3-4: Google Earth Image Showing the field established Plots within the Niger Delta Region of Nigeria. Spectroradiometric measurement of dominant vegetation reflectance were acquired from the various micro-plots.

The location of the non – spill site, which served as the control site is located about 1km from the current site (as shown in Figure 3-4). The various sites were chosen as part of recommendation made by Siegel (1974) to allow for comparison between a site of known impact to a normal site (i.e. healthy plant void of any stress) before a conclusion can be reached on the dominant biochemical state of the vegetation samples investigated as a result of oil spill impact.

The field visit was carried out through the technical support of an NGO called SACA (Shareholder Alliance for Corporate Accountability). The organization played a significant role in providing logistics and access to the spill site.

Table 3-3: Showing the Date of Field Experimental Activities were carried out

S/No	Date/Day	Activity Carried out
1.	24th Tuesday January 2017	Implementation of Plots for the New/Recent Spill Site
2.	26th Thursday January 2017	Implementation of Plots for the Old Oil Spill Site
3.	28th Saturday January 2017	Implementation of Plots for the Control Site

Evidence of oil slick coating on vegetation stem layer were still observed in the most recent spill site during the field exercise as shown in Figure 3-5. In order for the current research to effectively assess the influence of oil pollution on vegetation characteristics, 3 plots and 16 micro plots measuring 20 by 20 meters and 5 by 5 meters respectively, were set up along a transect of 120m in each of the three sites.

The coordinates of each of the corner plots and the center points were acquired with the aid of a GPS device. This was later mapped into a GIS environment for image spectral information retrieval. The result of this field experiment is presented in Chapter four and six of this thesis.

(a)

(b)

(c)

Figure 3-5: Showing Oil Coating on Vegetation Sub Canopy

3.2.9.2 Fieldwork – 2

The second fieldwork was carried out between 18th October and 22nd October 2018 (Table 3-4 shows the date the various locations were visited). The main objectives and activities carried out during the course of the fieldwork exercise is summarized into four broad parts.

Table 3-4: Showing the Date of Field Experimental Activities were carried out

S/No	Date/Day	Location	Activity Carried out
1.	19th October 2018	Etche	Soil Sample Collection in Control Site
2.	20th October 2018	TAI	Soil Sample Collection in Recent Spill Site
3.	21st October 2018	GOKANA	Soil Sample Collection in Recent Spill Site
4.	22nd October 2018	Eleme	Soil Sample Collection in Recent Spill Site

- Toxicology analysis

Three recent spill sites located in TAI, GOKANA and ELEME local government areas of Rivers State were visited. Soil samples were taken from each of the spill sites for toxicology analysis to measure and establish the Total Hydrocarbon Carbon Content (THC) volume in each of the site. Figure 3-6 is a map showing the location of the three spill and one non – spill (control) site visited during the second fieldwork exercise.

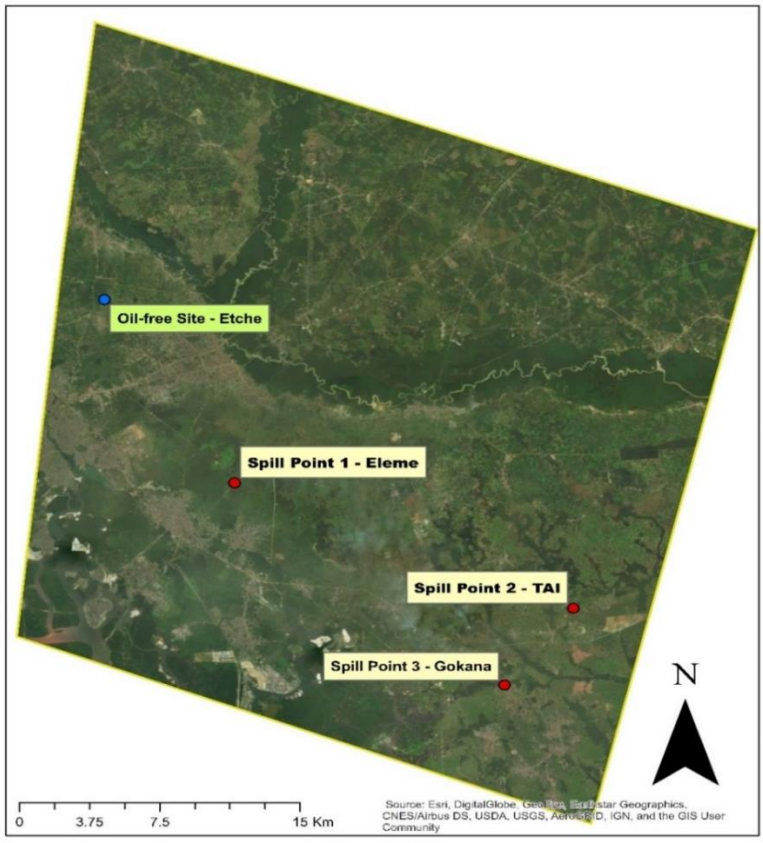
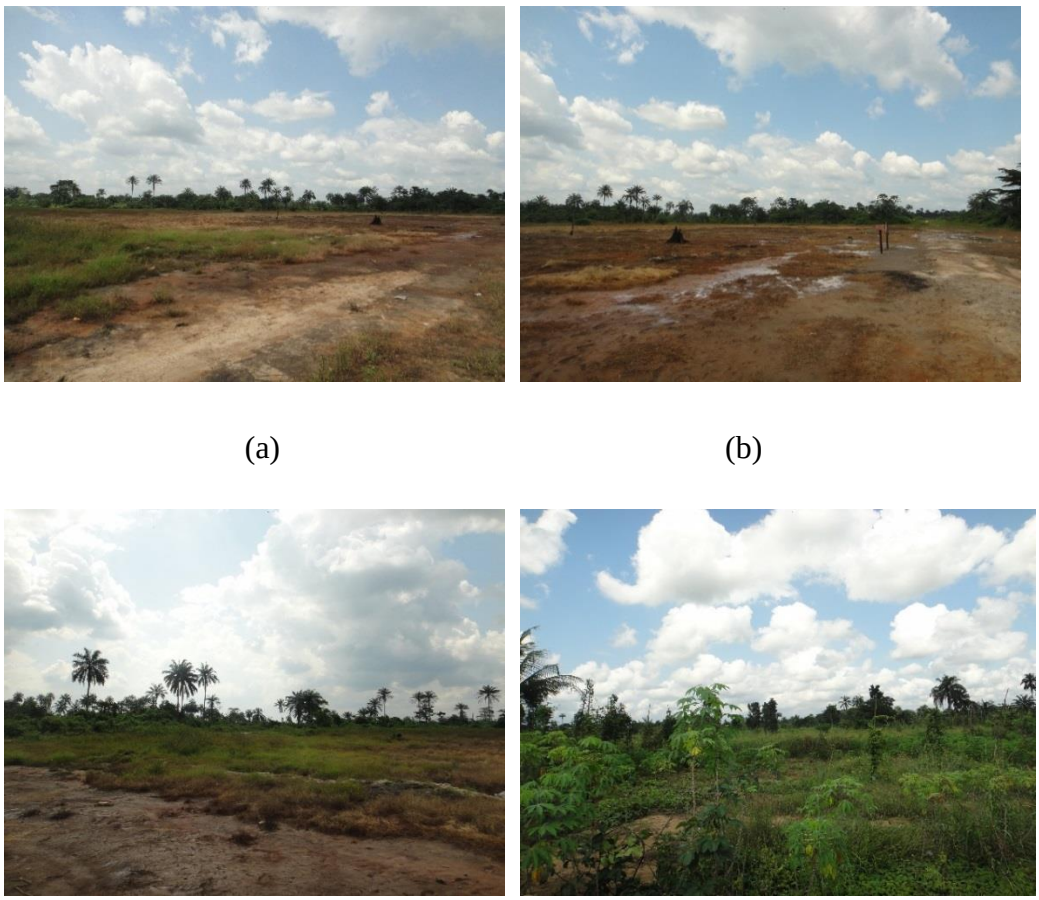


Figure 3-6: Showing Oil Spill Site and the Oil free Site Visited during the fieldwork

- Soil Sampling and Preparation

Three different samples of contaminated soil were collected around the spill epicenter location. A hand held auger was used to bore a hole and excavate the soil to a depth of around 40cm before soil samples were taken. This was repeated across the different locations within the affected spill area. The collected samples were then combined to form one sample for each spill site visited, making a total of three soil samples, one each for TAI, GOKANA and ELEME spill site. Figure 3-7 a, b and c shows the recent spill site visited in Eleme, while figure 3-7d is a picture of the oil free non – polluted site visited in Etche.



(a) (b)

(c) (d)

Figure 3-7: Showing Picture of Oil Spill Site and Oil Coating on Vegetation Sub Canopy (a) Eleme Site (b) TAI (c) Gokana and Oil free Site (d) Etche

The same procedure was repeated for the non – spill/control site location in Etche. Collected samples were stored in a zip lock and kept in a cool and iced environment

to prevent vaporization. It was ensured that Samples collected got to the laboratory before the end of the day in order for proper storage to be carried out.

- Hydrocarbon Test Procedure

The methods adopted for the laboratory analysis were in line with the American Society for Testing and Materials (ASTM) Analytical Protocol. For the purpose of this study, toxicology analysis was made to reveal the total hydrocarbon content (THC) present in each of the soil samples. Method used in the laboratory for the test procedure was the ASTMD3921, which is a standard test method for oil and grease and petroleum hydrocarbon concentration in sediment. The method uses CFCs for infrared spectroscopy to separate the different liquid content before the actual volume of the THC is ascertained. The result of the hydrocarbon test analysis was compared to the result of the image classification in chapter 5 of this thesis in the four locations to establish if classes are assigned to the corresponding oil-free and oil-impacted areas.

- Ground Truth of False Positives sites.

The second part of the fieldwork was a qualitative ground truth exercise. This generally involved going round the study area to directly validate and compare the results obtained from the image classification process to actual realities on ground. Attention was focused mainly on only false positives in the classified image, as initially visited sites (for toxicology analysis) were true positive sites. However, due to logistics and political volatility of the study area, only four (4) large expanse of fields were visited. The locations visited were Obio/Akpor, Oyigbo, Etche and Emohua.

3.3 Image Classification Methods

The image classification methods highlights the methods adopted in the assignment of pixels to specific classes based on the training sites of polluted and oil-free vegetation established in section 3.2.3. Three machine learning classifiers namely: Random forest (RF), Fuzzy forest (FF) and Support Vector Machine (SVM) methods were used interactively to address the various research questions (section 2.7) and research objectives (section 2.10.2) posed in the course of the research. While the RF method was used in chapter 4 to investigate the utility of the vegetation specific

classification against the general image classification, both RF and FF methods were used in chapter 5 of this thesis to assess the integrated use of optical – SAR image variables, as well as high dimensionality effect in discrimination of polluted and oil-free vegetation. Lastly, the RF and SVM methods were both used in chapter 6 of the thesis in the multifrequency SAR, Multitemporal and Multi-seasonal based image classification.

3.3.1 Random forest

The Random Forest (RF) is an ensemble classification method proposed by Breiman (2001). It is an ensemble method for supervised classification and regression, based on classification and regression trees (Caret). It works on the assumption that an aggregation of correctly predicted classes from the decision trees that make-up the RF can improve the general classification accuracy. Figure 3-8 shows a graphical representation of a typical random forest ensemble with three classification trees.

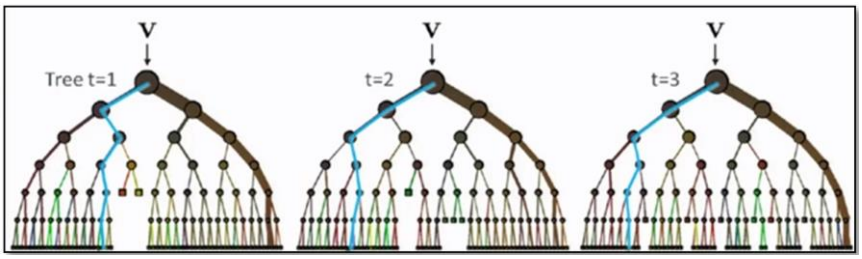


Figure 3-8: Random forest with '3' trees. Image source: (Gislason *et al.*, 2006)

Each ensemble in the classifier is trained on a subset of the various input variables using two thirds of these samples. The remaining one third is used to generate the out-of-bag error, which is an internal validation process that evaluates the final model for classification or regression. Random forest is one of the most efficient machine learning based classifiers used for various land cover and land use classifications (Adam *et al.*, 2014; Barrett *et al.*, 2014; Du *et al.*, 2015; Gislason *et al.*, 2006; Guo *et al.*, 2011; Juel *et al.*, 2015; Samat *et al.*, 2018; Son *et al.*, 2018; van Beijma *et al.*, 2014; Ye *et al.*, 2013).

As a machine learning base classifier, RF offers the opportunity to incorporate as many spatial variables in a single classification operation, which in turn should contribute positively to the prediction of the final class. The importance of the individual variables is estimated through random permuting of the value of the out-

of-bag samples, which produces a measure of how important a specific variable or data layer is in the classification operation (Genuer *et al.*, 2010). Random forest also offers huge flexibility in the tuning of the two major parameters, the number of trees in the forest (*ntree*) and the number of variables to use at each node (*mtry*). Similarly, the method also seeks to generate a measure of importance for each of the subsampled variables used in the classification process on the account of the Gini index and mean decrease in Gini.

However, it is believed that *RF* variable importance measures could be biased when highly correlated features are incorporated in a single classification operation (Nicodemus and Malley, 2009; Strobl *et al.*, 2008), thereby generally influencing the overall classification accuracy. In order to arrive at an optimal *ntree* value, several iterative runs were carried out in this research, and it was discovered that an optimal ntree value of 500 produced the best accuracy. Similarly, the square root of the number of input variables was used as *mtry* in the random forest image classification both in the EnMap box (Waske *et al.*, 2012) used in chapter 4 and the R software (TeamR, 2017) used in chapter 5 and 6 of this thesis.

3.3.2 Fuzzy forest

Fuzzy Forest (FF) is an extension of the RF approach which seeks to obtain less biased variable importance rankings in the presence of high correlated features. This is accomplished in two fundamental steps. First is a screening process aimed at eliminating unimportant variables by assigning features to separate variable clusters called 'modules'. Here, the target is for fuzzy forest to produce a partition of the features with high correlation using Weighted Gene Correlation Network Analysis (WGCNA).

This feature can be denoted by the set$Q = \{Q_1, \dots\dots, Q_m\}$. Let $q_l = |Q_l|$ so that $\sum_{l=1}^{m} q_l = q$ (Conn *et al.*, 2015). (3)

The screening step operates independently on each partition and each element of the partition Q_l is relieved of unimportant variables using the Recursive Feature Elimination Random Forest (RFE-RF). Starting with all features in partition Q_l, a random forest is fit and the least important features are eliminated. The resultant features after the first round of elimination can be called $Q_l^{(1)}$. Consequently, a

second random forest is then fit using features in the $Q_l^{(1)}$ and the least important features are also eliminated leading to a further reduced set of features $Q_l^{(2)} \subset Q_l^{(1)} \subset Q_l$. The subset obtained after iteration *t* can be denoted as $Q_l^{(t)}$ which is the number of features in $Q_l^{(t)}$. The process of feature elimination continues until a user defined threshold is reached, which could be for instance until 25% of the most important variables in Q_l are remaining (Conn *et al.*, 2015). For the full potential of the most important variables to be selected, the user must specify how many features are to be dropped after each iteration by specifying and tuning various screening parameters and specifying a stopping criterion (Conn *et al.*, 2015). This process was implemented in the R software environment (TeamR, 2017) in comparison with the RF method.

However, given the variable reduction potential of fuzzy forest and its ability to select only high performing non-correlated variables for a single classification operation. It is expected that its utilization in this research would help cope with the diversity and complexity associated with the integration of multifrequency and multitemporal SAR, optical and geophysical variables for effective discrimination of polluted and oil-free vegetation.

3.3.3 Support Vector Machine

Propounded by Cortes and Vapnik (1995), the SVM algorithm in a classification operation seeks to find a linear separating hyperplane with the maximum margin in a higher dimensional space to distinguish input data into distinct number of classes. Given a training set of an instance – label pair ($x_{t,}\, y_t$), i = 1, ….., l where ($x_t \in R^n$) and $y \in \{1, -1\}^l$, the support vector machines (SVM) require the solution of the following optimization problem (Hsu *et al.*, 2003):

$$\underset{w,b,\varepsilon}{min} \quad \frac{1}{2} w^T w + C \sum_{i=1}^{l} \varepsilon_i$$

Subject to $\quad y_i(w^T \phi(x_i) + b) \geq 1 - \varepsilon_{i,} \quad \varepsilon_i \geq 0.$ (4)

Here x_i and *w* are training vector mapped into a higher (maybe infinite) dimensional space by the function $\emptyset$ and *b*, which are bias. SVM finds a linear separating hyperplane with the maximal margin in this higher dimensional space. To accomplish this, the penalty parameter of the error term $C > 0$ and the kernel function $K(x_{i,}\, y_j) \equiv \phi(x_i)^T \emptyset(x_j)$ are used to efficiently distinguish the training data into predefined classes for a better prediction of the test data. Figure 3-9 is an illustration of how a SVM seeks to find the best separating hyperplane for a test data, usually through a grid search operation.

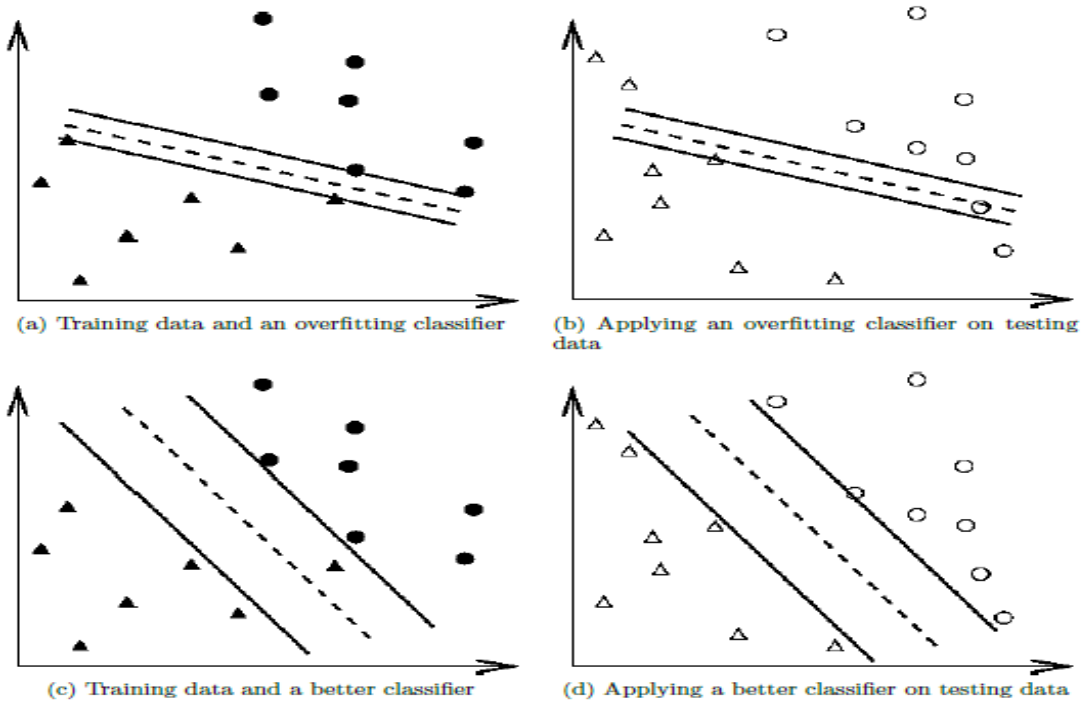


Figure 3-9: Demonstrates how *SVM* seeks to project test data in a hyperplane to obtain the best class separation. Image Source: (Chang and Lin, 2011)

The Support Vector Machine image classification was carried out using the 'rpart' (Therneau *et al.*, 2015) and 'e1071' (Meyer *et al.*, 2017) packages in the R Software environment (TeamR, 2017).

3.4 Study Area

3.4.1 Location and Size

The Niger Delta region is one of the 8 geopolitical zones in Nigeria. It is located on latitude 4º10' to 6º20' north and longitude 2º35' east of the equator with a total landmass of ~72,000km^2, making it the world's third largest and Africa's largest delta (Akegbejo-Samsons and Omoniyi, 2009). The region is comprised of 8 states; Abia, Akwa Ibom, Bayelsa, Cross River, Delta, Edo, Imo, Ondo and Rivers states (Figure 3-10).

The region is situated right about the Gulf of Guinea on the coast of Atlantic Ocean (Hooper *et al.*, 2002; Imoobe and Iroro, 2009; Onuoha, 2007). The Niger Delta is a low lying area, approximately 3.5m above mean sea level and an area of about 112,110 square kilometers, making up 12% of Nigeria's total land mass (NDDC, 2004). The region cuts across the coasts of Ondo, Delta, Bayelsa, Rivers, Akwa Ibom to Cross Rivers' states (Imoobe and Iroro, 2009). The Delta is a large flood plain area built up as a result of accumulation of sediments washed down for over 100 million years from the Benue and Niger Rivers (Lindén and Pålsson, 2013).

The region has an extensive hydrology system connecting rivers, creeks, and estuaries flowing towards the Atlantic Ocean (Abam, 2001; Akpokodje, 1987). The River Niger axis of the Delta has its headwaters from Guinea, passing through Mali to Republic of Benin, Niger republic and Nigeria. Similarly, the river Benue axis of the Delta has its own headwaters in from the Cameroon highlands and traverses both the Cameroon and Adamawa highlands in Nigeria (Abam, 2001).

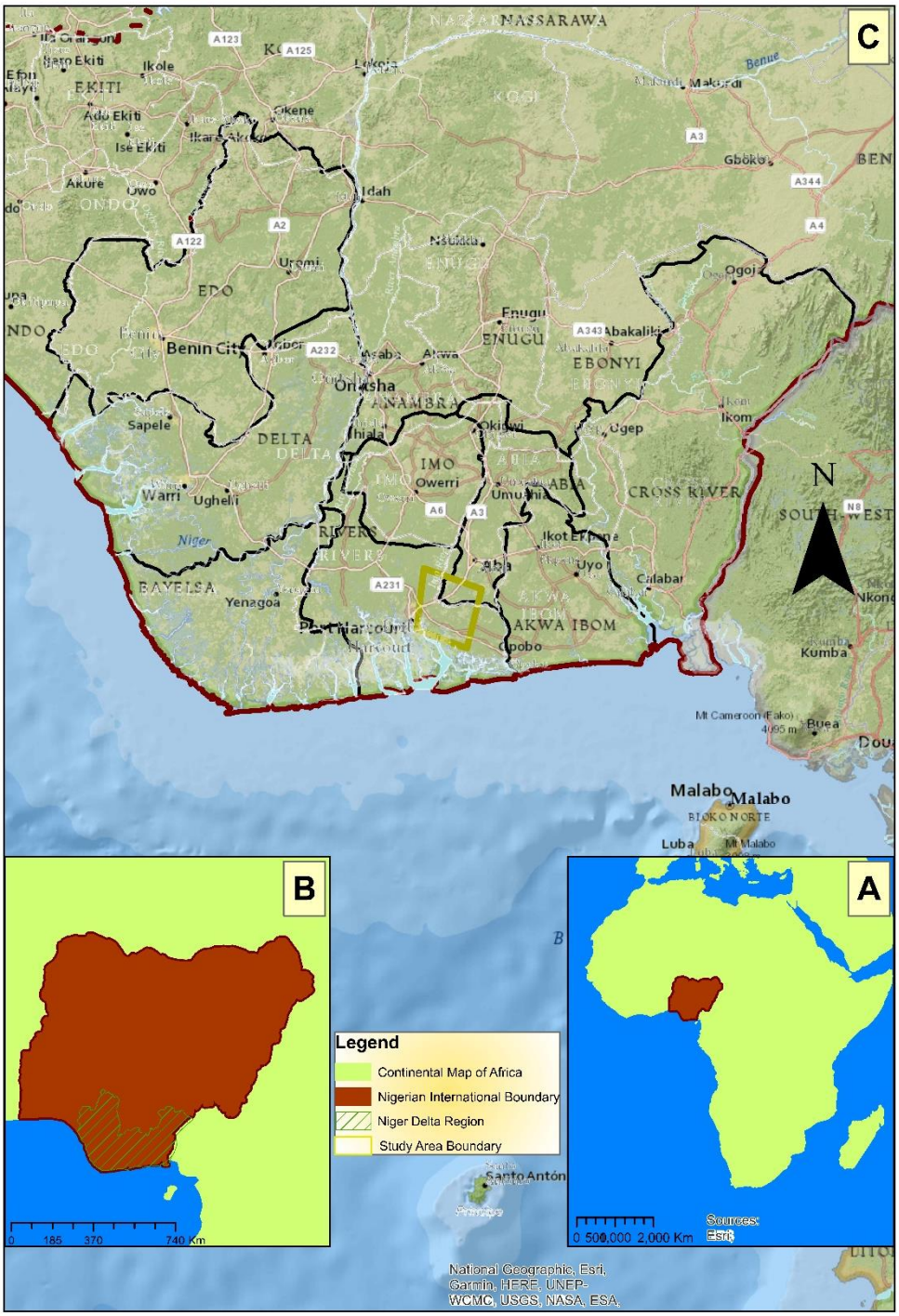


Figure 3-10: Showing the Location of the Study Area used for this research (A) Map of Africa Showing Nigeria (B) Map of Nigeria Showing the Niger Delta Region (C) Study Area

The actual study site for this research is bounded by four corner coordinates of (6.957^0E, 5.025^0N), (7.247^0E, 5.025^0N), (6.96^0E, 4.795^0N) and (7.254^0E, 4.804^0N), and covers an area of 1320 km^2 (Figure 3-11), which represents a small part of the region. This site was chosen because it represents a rare oil spill hotspot area, with a

high concentration of oil spill incidents (see figure 3-1). This particularly made the area suitable for image analysis to carry out the important process of calibration and validation of the machine learning random forest, fuzzy forest and SVM models.

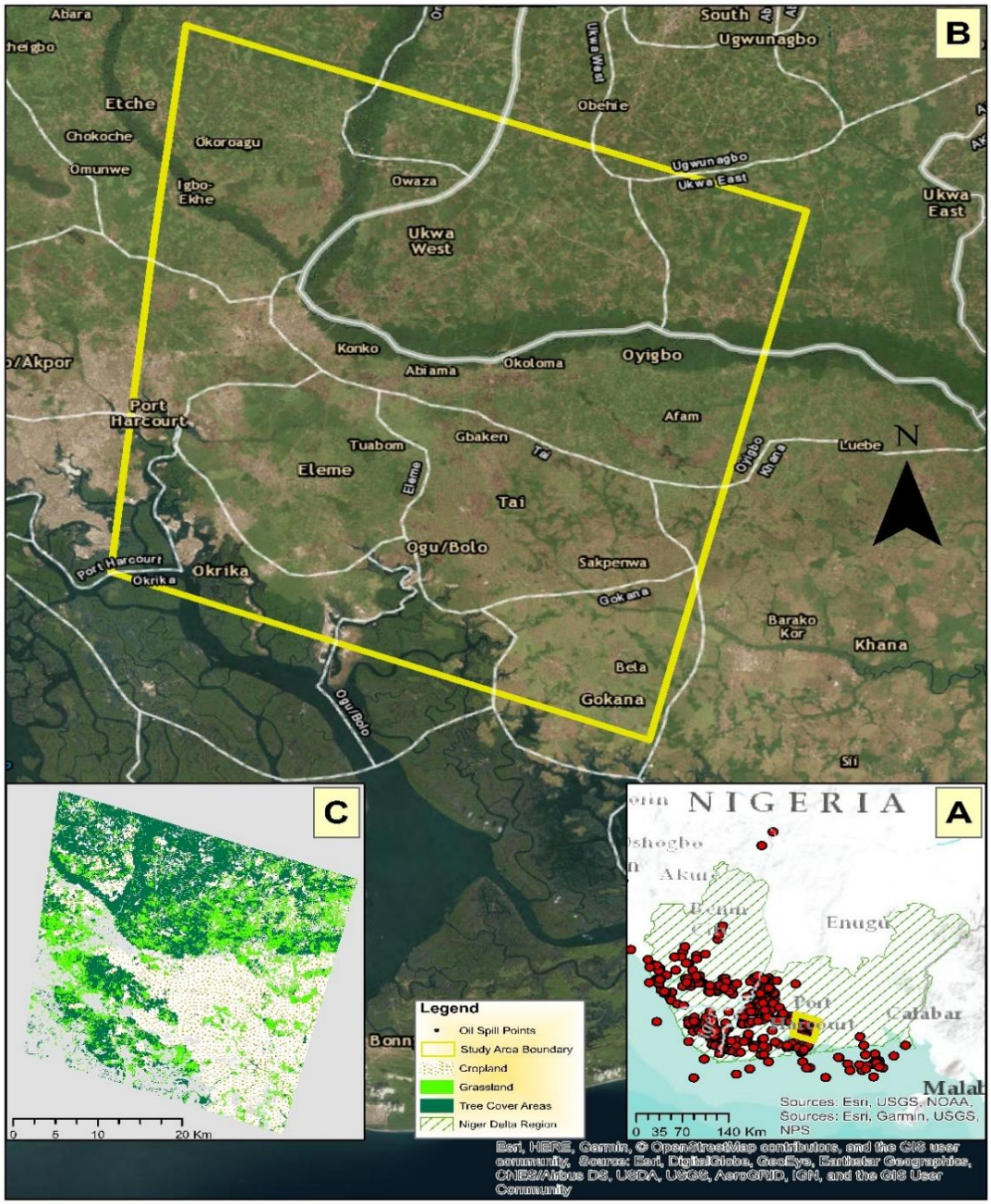


Figure 3-11: Showing the Location of the Study Area used for this research with emphasis on the Landcover Map and Oil Spill Incident Data (A) Map of the Niger Delta Region Showing the Oil Spill Incident Points (B) Map of the Study Area showing names of communities within the study site (C) Predominant Land cover within the Study Area

3.4.2 Population

The population of the Niger Delta region stood at about 20 million at the time of the 1991 census (Figure 3-12), accounting for 23% of the Nigeria's population. Current estimates from government sources put the total population of the region at 27 million in 2005 (NDDC, 2005). The National Population Commission (NPC) has estimated that the population of the Niger Delta region would be nearly 33 million people by the end of 2013 (NPC, 2006). The 2006 census also showed that the Niger Delta region has the highest population density in the country. A report by NDDC, (2004) estimated that the population density is about ~450 persons per km^2, making it the most densely populated region in Africa (Ayanlade, 2015).

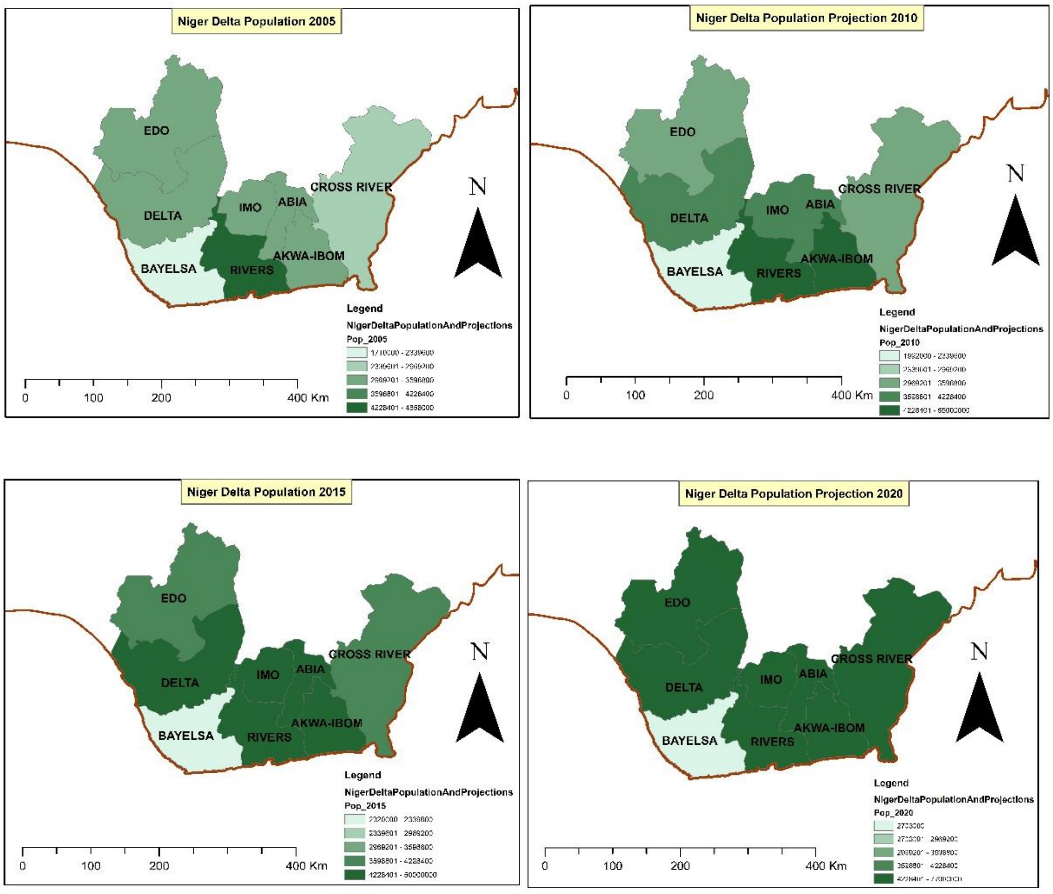


Figure 3-12: Showing the Population Map and Projection of the various States of the Niger Delta Region for 2005, 2010, 2016 and 2020

Major factors responsible for the sharp population increase in the Niger Delta region, on the account of previous studies (Ayanlade, 2015; Ebeku, 2006; Eregha and Irughe, 2009; International, 2009), are due to high rate of migration into the Delta region by people from other parts of Nigeria seeking opportunities and employment into the oil

industry. The majority of the population in the region (about 90%) depend mainly on agriculture and forest material for their fuel and means of livelihood (Ayanlade, 2015; Ibaba, 2010; Mmom and Arokoyu, 2010).

3.4.3 Climate

The Niger Delta area lies within the wet equatorial climate with high cloud cover and it is influenced by the localized convection of the West African monsoon, with less contribution from the mesoscale and synoptic system of the Sahel (Adejuwon, 2012; Okoro *et al.*, 2014). In between the months of March and October the climate of the region is characterized by the tropical maritime air mass, while in the months of November to February the region is under the influence of dry tropical continental air mass (Dada *et al.*, 2015). The average annual rainfall of the region ranges from 2000 to 4021mm (Adejuwon, 2012), with precipitation increasing from North of the Delta towards the coast (Figure 3-13) (Ayanlade, 2015).

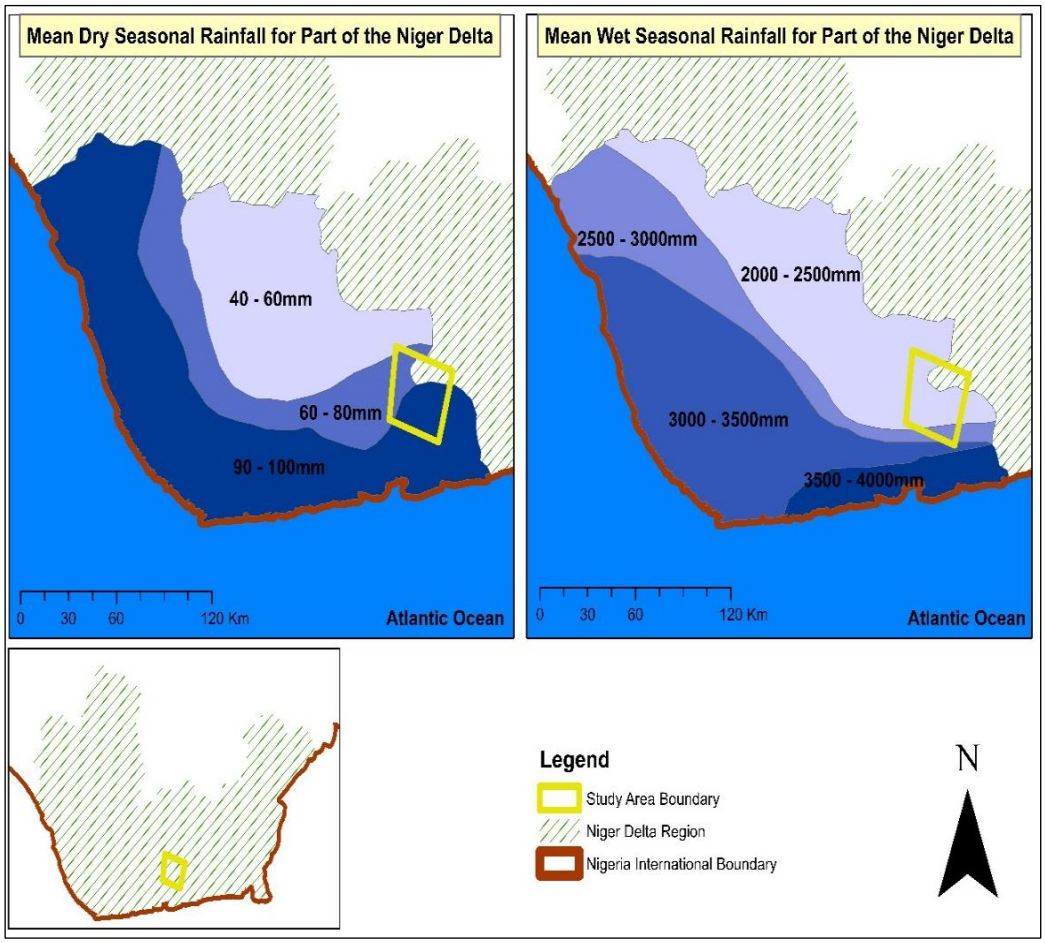


Figure 3-13: Mean Seasonal Wet and Dry Season rainfall map of part of the Niger Delta Region of Nigeria. Source: (Adejuwon, 2012).

The rainy season generally lasts for about nine months in the south, and less than four months in the north (Nigerian Meteorological Agency, 2010). Precipitation however increases from the Northern axis of the Delta region down towards the coast (Ayanlade, 2015). This region is also characterized with high surface humidity that rarely dips below 75% and the coastal areas also experience much higher humidity than the inland areas (Sorgwe, 1997).

The temperature variability across the region is relatively constant throughout the year over with average annual temperature of about 27^{o}C (Figure 3-14) and with little seasonal variation (Adejuwon, 2012). In addition, much of this high humidity (80% to 90%) is usually recorded within the months of June through September, and lower humidity occurs from December to March (Adejuwon, 2012).

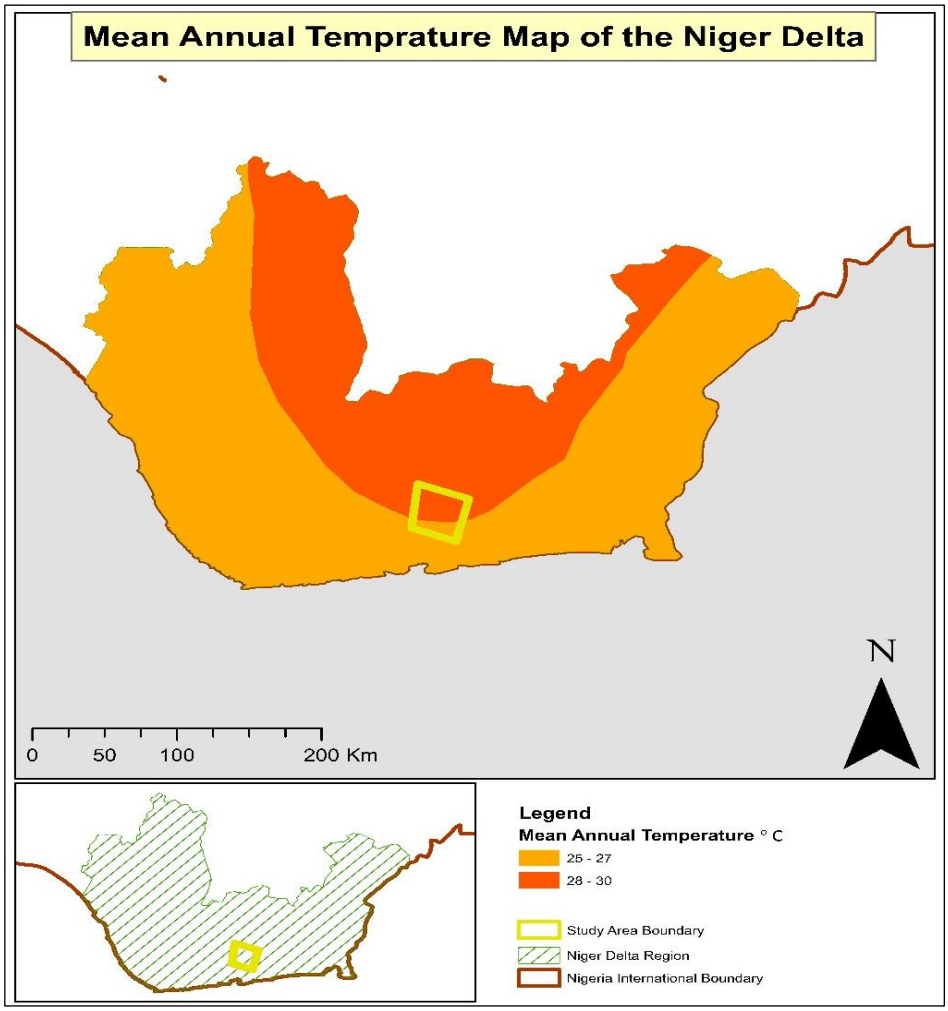


Figure 3-14: Mean Annual Temperature Map of the Niger Delta Region. Source: Federal Ministry of Environment

3.4.4 Ecology of the Niger Delta

The Niger Delta region has diverse mosaics of ecological types (as shown in Figure 3-15) made up of five distinct ecological zones ranging from barrier island forest and coastal vegetation areas to montane habitats (Twumasi and Merem, 2006). Twumasi and Merem (2006) conceptualized the Niger Delta ecology into several zones: mangrove and coastal vegetation, fresh water swamp and rainforest zones. The first eco-zone has the features of a mangrove forest and coastal vegetation zone with a chain of low sandy barrier islands that protects the coast of the Niger Delta, between the Benin and Imo estuaries. Omo-Irabor *et al.* (2011) also noted that the mangrove forest zone extends from Lagos to Sapele (Delta state) connecting with the freshwater swamp some few kilometers inland, which in turn gives way to the rainforest inland.

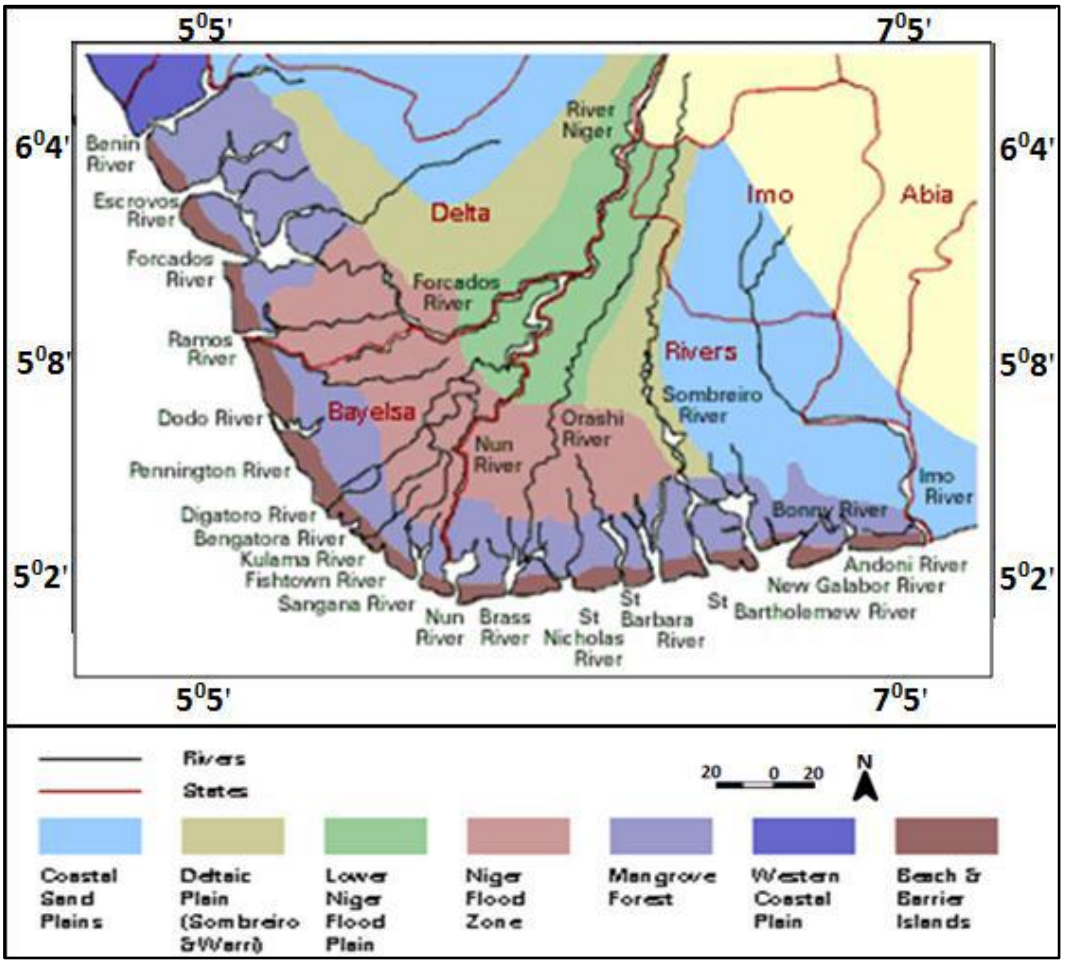


Figure 3-15: Map of the Niger Delta Showing the Geomorphological units (Source: Urhobo Historical Society, 2009)

The second eco-zone is characterized by the freshwater swamp forest zone, covering approximately 17,000 square kilometers or about half of the region. The third eco-

zone is predominated with mostly lowland rain forest zone, which stretches over non-riverine areas flanking the Delta, while the fourth eco-zone is made up of the derived savannah and this is found in the northern part of the Niger Delta region.

However, the fifth and last eco-zone constitutes mostly of the montane zone and this is also concentrated in the north eastern part of Cross River state of the region (NDDC, 2005; Twumasi and Merem, 2006). Osuji and Opiah (2007) have particularly noted that generally, mangrove forest, freshwater swamp and rainforest, which are the most predominant ecological zone of the Delta are dominated by the tree species *Elaeis guineensis* (Shittu, 2014).

Similarly, predominant land cover types as provided by the European Climate Change Initiative for the Niger Delta region include: Cropland, Grassland, Marsh, Mangrove, Bareland, Waterbody and Tree Cover Areas. Three of the 8 land cover types were given greater attention in this research, due to their extensive coverage in the study area. From the data provided (Figure 3-2), the northern axis of the study area has higher concentration of riparian forest and dense vegetation types, while patches of grassland are concentrated in the central part of the study area connecting the much cultivated part of the study area to the south eastern corner.

3.4.5 Geology

The landforms of the Niger Delta region were created from the accumulation of marine and deltaic sediment of over 50 million years ago in the upper Cretaceous period (UNEP, 2011). The sediments deposited by fluvial processes centuries ago led to the formation of a relatively flat alluvium basin with natural levees and ox-bow lakes (Abam, 2001). The sediments in the Coastal sand plains vary from medium-coarse grained sands, fine clayey sands in natural levees to clayey and peaty deposits in swamps and lagoons as shown in Figure 3-16 (Tuttle *et al.*, 1999).

The deltaic plain lies flat at about 40m above sea level towards the interior, and less than 8m above sea level on approaching the coast (Akpokodje, 1987). Deltaic Plain and Western coastal plain are non-tidal zones, which are characterized by seasonally flooded small lakes. The soils obtainable in this region were formed through the deposition of alluvium sediment materials during the late Pleistocene to early Holocene time (Ugochukwu and Ertel, 2008).

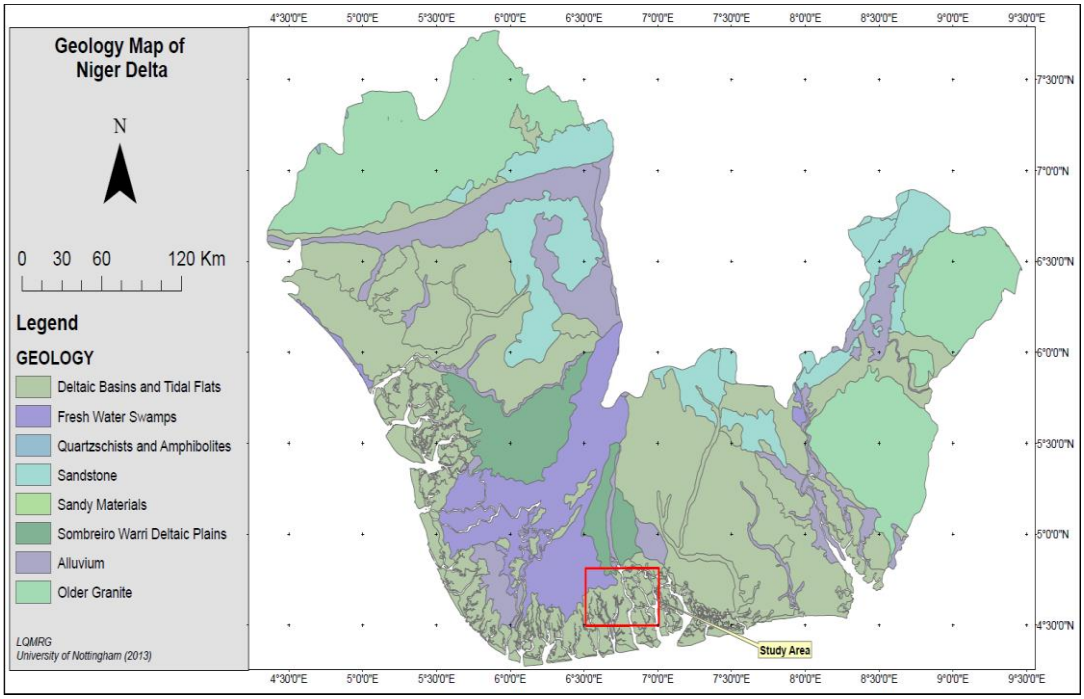


Figure 3-16: Geologic Map of the Niger Delta. Source: Ayanlade (2015

The alluvial soils are usually found in the flood plains of the deltas and along coastal flats. They extend from the inland coastal areas to the valleys of the Niger and the Benue rivers, cutting across the various vegetation zones (Akpokodje, 1987). Owing to the high rainfall regime, shallow aquifer, and the predominant flat topography, there is usually perennial river inundation and overflow in the coastal axis of the region occasionally causing flooding (Akpokodje, 1987).

Abam (2001) noted that soils in the flood plains are characteristic of freshwater soil of grey to white sand, grey clay and sandy clay with humus topsoil. Another group consists of brownish to black saline mangrove soils, with a mat of rootlets. Additionally, the soils of this region are all of fluvial origin, except for the coastal areas that consist of marine sand overlain with an organic surface layer.

Chapter 4 : Mapping Terrestrial Oil Spill Impact Using Machine Learning Random Forest and Landsat 8 OLI Imagery

The work presented in this chapter has been published as:

M. S Ozigis; J. D Kaduk and C. H Jarvis (2019). Mapping Terrestrial Oil Spill Impact Using Machine Learning Random Forest and Landsat 8 OLI Imagery: A Case Site within the Niger Delta Region of Nigeria. *Environmental Science and Pollution Research*

4.1 Introduction

This chapter explores the potentials of the non-parametric – Random Forest Machine Learning Classifier to discriminate between pixels of oil-polluted from oil-free vegetation types within the Niger Delta region of Nigeria, using Landsat 8 Visible (VIS), Near Infrared (NIR), Short wave Infrared (SWIR) spectral bands and derived vegetation health indices. It goes further to identify the variables that provide most information for this discrimination/classification process, as previous studies (Adamu *et al.*, 2015; Arellano *et al.*, 2015; Khanna *et al.*, 2013; Zhou *et al.*, 2014) have tested the sensitivity of some of these variables to detect oil spill using parametric methods. The role of subset classification for the separate vegetation types (i.e. cropland, grassland and tree cover area) in the possible reduction of confusion between oil-impacted and oil-free classes was explored. Results of the field spectroscopy carried out in the course of the research is also presented and discussed in this chapter.

4.2 Methodology and Data Pre-processing

4.2.1 Oil Spill Incident Data Pre-processing and Sampling Regime

4.2.1.1 Spill Incident Harmonization

The oil spill data harmonization sought to integrate and expand the oil spill database for this research. The harmonization operation was carried out by overlaying both datasets (NOSDRA and SHELL) in a GIS environment. Points with repeated information as a result of duplicate capture and multiple spill incidents over the years were identified and marked. Duplicates (in most cases the SPDC data) were deleted since the dataset provided by NOSDRA is all encompassing as the Government's regulatory agency with the responsibility of documenting all spill incidents. The spill information relating to volume, size and date of spill was checked, as this provided the basis for tracking the spill intensity on the different vegetation types. The minimum area covered by the spill data used for this exercise is 1000sqm, which is greater than a single Landsat – 8 image pixel of 900sqm. This is to ensure that pixels used for training, testing and validation of the final model as well as the image classification have the dominant spectral reflectance of a typical oil polluted site.

4.2.1.2 Assignment of Spill Incidents to Land cover

The assignment of oil spill incidents to the corresponding vegetation type is an important step in this study, as the RF method relies on the spectral signatures provided by these training sites to build a robust model. For each vegetation class (cropland, grassland and tree cover areas) spill incidents located within them were identified and extracted using requisite GIS selection tools. This provided the various training and validation sites for the identification of oil-impacted (polluted vegetation) classes (See Table 3-2 and Table 4-1).

4.2.1.3 Selecting Non – Polluted Sites for the Different Landover

Non – polluted sample sites are necessary in this study for two main reasons. First for the identification of oil-free (Non – Polluted) vegetation within the study area and secondly for an effective discrimination between pixels of oil-free and oil-impacted vegetation as stressed by Siegel (1974). Proximity analysis as suggested by a number of researchers (Obida *et al.*, 2018; Park *et al.*, 2016; Whanda *et al.*, 2016) provided the basis for the selection of the polluted and oil-free vegetation pixels. The minimum rule was set that all non – polluted sites must be located at least 600m away from all

polluted sites based on the maximum area of spill recorded. This resulted in an 800m buffer ring around all existing spill points, which avoided any overlap with any likely spill-impacted area. The procedure ensured that sample sites selected for the respective oil-free vegetation are reasonably well spaced from the respective oil-polluted sites/vegetation. Thereafter the training sites for the non – polluted vegetation were selected at random outside the buffer ring established. Furthermore, specifically only healthy vegetation as inferred from high-resolution google earth image were chosen.

4.2.1.4 Pixel selection using buffer analysis

Following the reconciliation and extraction of the oil spill points and the non – polluted sites respectively according to their respective vegetation class (cropland, grassland and TCA), the points were then sub-divided into two categories for training and validation purpose. A total of 60% of the points for individual vegetation type were randomly selected for training, while the other 40% were set aside for validation in a post classification confusion error matrix. Table 4-1 shows the distribution of the Polluted spill sites and oil-free sites according to their respective vegetation type. To this end, 30m buffer ring polygons were established around all the training sites to ensure that only adjacent pixels within the High Consequence Area close to the point of impact were selected for the polluted sites (Adamu *et al.*, 2015; Alexakis *et al.*, 2016).

Table 4-1: Total number of Sites used for Calibrating and Validating the Random Forest Classification operation.

Class Label	Number of spill sites
Non – Polluted Cropland	41
Non – Polluted Grassland	27
Non – Polluted Tree Cover Areas	25
Polluted Cropland	44
Polluted Grassland	26
Polluted Tree Cover Areas	26

4.2.2 Image Preprocessing

As the Landsat surface reflectance higher – level data product was obtained, there was no need to carry out any atmospheric correction operations.

4.2.2.1 Geometric Correction

In order to ensure that the Landsat 8 – OLI satellite image co-registered properly with other datasets (such as the oil spill sites and boundary dataset), the satellite image was re-projected to the Universal Transverse Mercator Projection and the World Geodetic Survey 1984 Datum of Zone 32 North (UTM WGS84 Zone 32N).

4.2.2.2 Retrieval of Important Vegetation Health Indices

Eight vegetation health indices were generated using the formulae presented in Table 4-2. The indices were generated from the pre-processed Landsat 8 – OLI image of the study area using the Red, Green, Blue, Near Infra-Red, Short-Wave Infra-Red 1 and Short Wave Infra-Red 2 bands.

4.2.2.3 Land cover Image Masking

Following the geometric correction of the study area image, the three dominant existing vegetation class extent extracted from the ESA CCI data (section 3.2.4) were used to subset the final layer stacked image comprising of the spectral bands and vegetation health indices for cropland, grassland and TCA. This provided the basis of implementing a general study area wide classification operation (at macro level) and individual vegetation subset classification (at micro level). The image extents generated were for cropland, grassland and tree cover areas (i.e. dense woody vegetation), in which the harmonized oil spill and oil-free training sites were used to implement a macro and micro level classification. This produced six different class schemes, that is, polluted (oil-impacted) cropland, polluted grassland, polluted TCA, non – polluted (oil-free) cropland, non – polluted grassland and non – polluted TCA.

Table 4-2: Vegetation Health Indices Generated Using the RED, GREEN, BLUE, NIR and SWIR Bands

Vegetation Indices	Formula	Author
Difference Vegetation Index	$R_{NIR} - R_{RED}$	(Tucker, 1980)
Modified Soil Adjusted Vegetation Index	$1/2[2R_{NIR} + 1/\sqrt{(2R_{NIR}+1) - 8(R_{NIR} - R_{RED})}]$	(Qi *et al.*, 1994)
Moisture Stress Index	R_{MidIR}/R_{NIR}	(Doraiswamy and Thompson, 1982)
Normalized Difference Vegetation Index	$(R_{NIR} - R_{RED})/(R_{NIR} + R_{RED})$	(Rouse Jr *et al.*, 1974)
Normalized Differential Water Index	$(R_{NIR} - R_{SWIR})/(R_{NIR} + R_{SWIR})$	(Hardisky, Klemas and Smart, 1983)
Renormalized Difference Vegetation Index	$R_{NIR} - R_{RED}/\sqrt{R_{NIR} + R_{RED}}$	(Roujean and Breon, 1995)
Ratio Vegetation Index	R_{RED}/R_{NIR}	(Jordan, 1969)
Soil and Atmospherically Resistant Vegetation Index	$(1 + 0.5)\ (R_{NIR} - R_{RB})/(R_{NIR} + R_{RB} + 0.5)$	(Qi *et al.*, 1994)
Soil Adjusted Vegetation Index	$(1 + L)(R_{NIR} - R_{RED})/(R_{NIR} + R_{RED} + L)$	(Huete, 1988)
Transformed Difference Vegetation Index	$\sqrt{R_{NIR} - R_{NIR}}/((R_{NIR} + R_{RED}) + 0.5)$	(Bannari, Asalhi and Teillet, 2002)

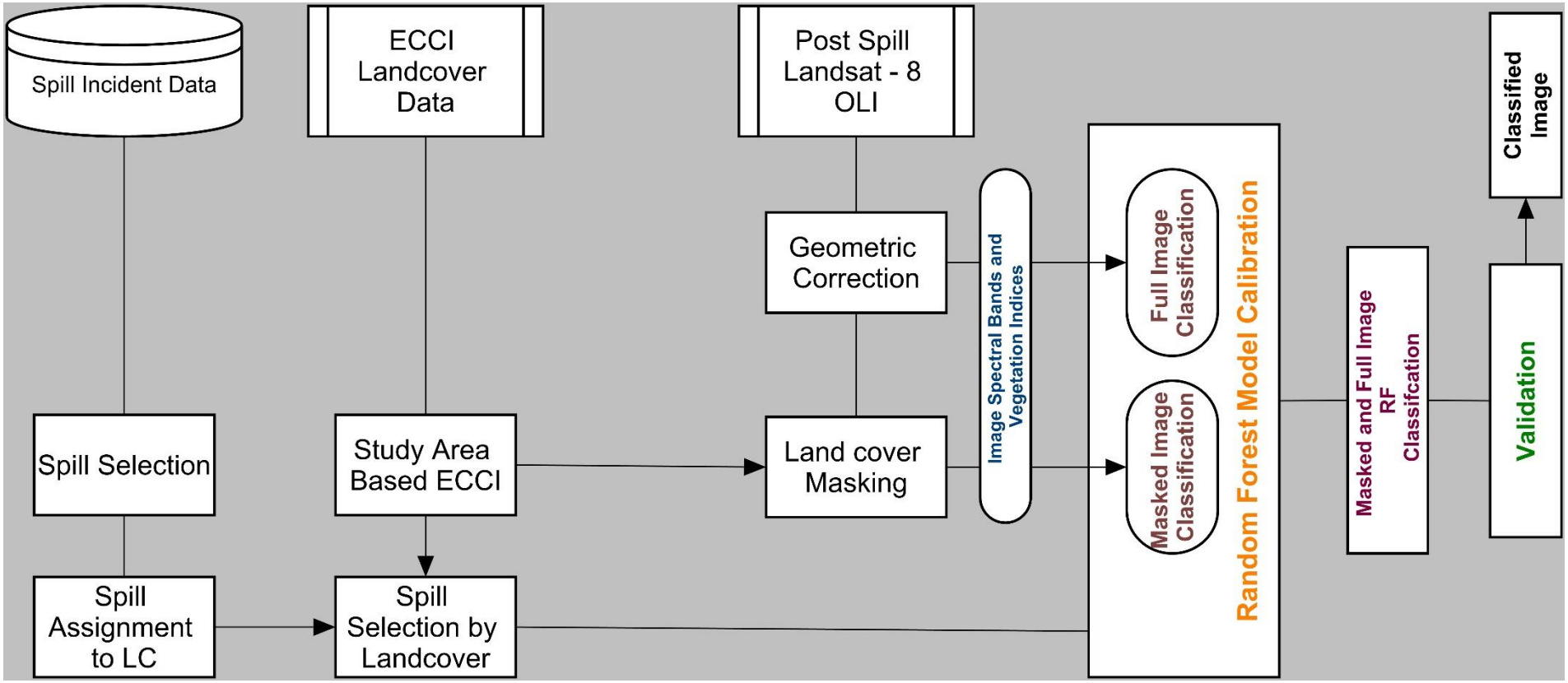


Figure 4-1: Methodological Flowchart of the Data Analysis for **Chapter 4**

4.2.3 Random Forest Classifier

The RF classification was used to distinguish and effectively characterize vegetation impacted by oil pollution from oil-free healthier vegetation. The analysis was carried out using the ImageRF component of the EnMap Box (Waske *et al.*, 2012) application. To achieve this, various vegetation health indices (Table 4-2) together with the seven Landsat – 8 OLI bands (across Visible, NIR and SWIR) were fused for the classification process. The tree size (*ntree*) used for classification was determined through repetitive runs before an optimal value of 500 (*ntree*) was arrived at and used for parametrization in all classification scenarios implemented. Table 4-3 outlines the list of variables used for the RF classification.

Table 4-3: List of Variables used for the *RF* Classification.

S/No	Spectral Variables
1	Band 2 - Blue
2	Band 3 - Green
3	Band 4 - Red
4	Band 5 - Near Infrared (NIR)
5	Band 6 - Shortwave Infrared (SWIR) 1
6	Band 7 - Shortwave Infrared (SWIR) 2
7	Difference Vegetation Index (DVI)
8	Modified Soil Adjusted Vegetation Index (MSAVI)
9	Moisture Stress Index (MSI)
10	Normalized Differential Vegetation Index (NDVI)
11	Normalized Differential Water Difference (NDWI)
12	Renormalized Difference Vegetation Index (RDVI)
13	Ratio Vegetation Index (RVI)
14	Soil and Atmospherically Resistant Vegetation Index (SARVI)
15	Soil Adjusted Vegetation Index (SAVI)
16	Transformed Normalized Difference Vegetation Index (TNDVI)
17	Band 1 – Ultra-Blue Band

4.2.4 Accuracy Assessment

Two performance indicators were employed to assess the RF Calibration Model and the resulting classified image obtained. First is the F1 Accuracy, which is the harmonic mean of Precision and Sensitivity (Recall) accuracy statistics. This is used in the ImageRF to assess the out of bag error of the RF calibration. The precision is the ratio of correctly predicted positive pixels to the total positive observations (incorporating true positives and false positives), while the recall is the ratio of correctly predicted positive observations to the sum of true positives and false negative observations. This however can be further interpreted as the measure of truly assigned pixels to a particular class (recall) and the measure of truly assigned pixels in the image space. The F1 score is a robust accuracy measure for model performance measurement, because it seeks to balance the influence of recall and precision through the use of harmonic mean of both measures.

This is denoted by the formulae below (Kuhn, 2012):

$$F1\ Accuracy = 2 \times \frac{Precision \times Recall}{Recall + Precision} \quad (5)$$

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

Where

TP = True Positives

FP = False Positives

FN = False Negatives

The error matrix as described by Congalton, (1991) was also used to assess the classified image output from the RF classification using the 40% validation points (section 3.2.3). This enabled an effective comparison of the classified image outputs to the original reference sites. Specific attention was given to the users, producers and the overall accuracies in the assessing classified image.

4.2.5 Field Spectral Measurement

4.2.5.1 Field Spectroscopy Experimental Procedure

Field spectral reflectance was acquired using a Handheld 2 Field spectroradiometer device as implemented in Arellano *et al.*, (2015) and Zhu *et al.*, (2013b). The device was used to measure the absorption and reflectance characteristics across the visible and NIR spectrum for the different micro – plots within the sites investigated (as explained in section 3.9.2.1). The spectral device covers between 325nm and 1075nm of the spectrum, with a spectral interval of 1nm. Vegetation spectra were acquired using the handheld device positioned at an angle of 90^0 pointing vertically to the target (vegetation canopy) and all experimental readings were acquired when the sun was at nadir between 11:30am and 12:30pm under bright skies without cloud cover. This significantly affected and accounted for variability in the dates in which the experiments were carried out since cloud conditions are an issue in the Niger Delta. Before readings were obtained the device was calibrated using the White Background Spectralon and dark current as prescribed by Zhu *et al.*, (2013b). The device was set at obtaining 20 readings for a single acquisition at one point before re-calibration was carried out. A minimum of two readings were obtained from individual micro plots and in total of 40 readings were obtained from the main plots (see section 3.2.9.1). The Handheld device was position at an approximate height of 45cm above the vegetation canopy.

In order to retrieve the spectral reflectance from the Digital Number (DN) measurements made in the field, the formula below was used as prescribed by Thenkabail *et al.*, (2000):

$$P = \frac{\text{Target} - \text{Darkcurrent}}{\text{Reference} - \text{Darkcurrent}} \times 100 \qquad (8)$$

4.2.5.2 Analysis

The extracted spectral reflectance values were further averaged for individual macro plots as implemented by a number of other researchers, e.g. (Arellano *et al.*, 2017a; Davidson *et al.*, 2016; Domínguez-Beisiegel *et al.*, 2016; Zhu *et al.*, 2013a) to obtain a mean representative spectral signature of a particular plot. The mean spectral reflectance values were used to establish the dominant and general reflectance characteristics from the plant communities within the three sites. This result was obtained by summing up all the 120 spectral reflectance values from the three main plots in each site (i.e. the recent, old and non-polluted site). Reflectance values recorded from 325nm to 1061nm of the spectrum

were summed up and divided by the total number of readings (in this case 120) to obtain a unified reading for the recent polluted, old polluted and non – polluted sites. This created the basis for comparing the vegetation canopy health among sample plots, with emphasis on the absorption and reflectance characteristics. To further ensure that the results from the spectral reflectance values are free from external noise interference, the reflectance values were smoothed with a Savitzky – Golay Filter (Chen *et al.*, 2004) to remove noise and improve the signal to noise ratio. The final result from this experiment compared the spectral reflectance of the recent, old and non-spill site.

4.3 Results

4.3.1 Random Forest Calibration

Figure 4-2 shows the result of the RF out of bag error. In general, the results indicated that the calibration carried out on the stacked image subset for the different vegetation types had lower out of bag errors and consequently higher calibration accuracy, compared to the result obtained from the full image calibration. This further showed that of the six classes, the Non – Polluted (NP) and Polluted (P) TCA and grassland vegetation had better calibration results ranging from 45% to 70% F1 Accuracy. On the contrary, both the P (oil-impacted) and NP (oil-free) cropland had lower calibration accuracies when the full study area image was calibrated. It also showed that of the six different schemes investigated the NP – grassland and NP TCA had the best prediction to error ratio of 86% and 84% as indicated in the F1 Accuracy when the respective vegetation types were trained separately. In contrast, the Polluted (P) and NP – cropland had the least calibration accuracy.

In terms of the implication for inter – class separability and model fit, it is observed that calibration accuracy increased gradually from zero and mostly attained saturation when the tree size (*ntree*) in the RF reached 50, although for some cases the F1 accuracy increased up to 100 *ntree* before maximum saturation was reached. This implied that an even lower *ntree* value could yield sufficient calibration result.

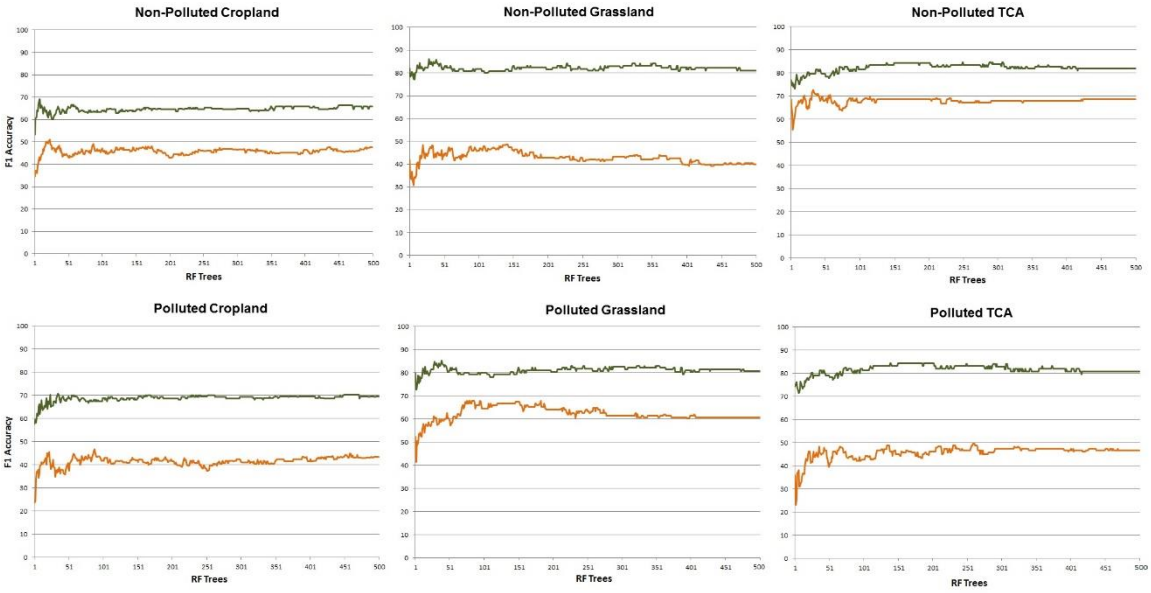


Figure 4-2: RF parameterization result for the full study area image and individual vegetation types subset images, using training samples of oil-free and oil-impacted vegetation. The green line represents parameterization F1 accuracy for the individual vegetation landcover subset images, while the brown line represents F1 accuracy for the full study area image parameterization result.

4.3.2 Landcover Subset Vs Full Image Classification

Figure 4-3 and 4-4 shows the image classification result from the two scenarios implemented. It was observed that the image classification at the micro-level of different vegetation types had better representation of polluted and impacted classes with more generalized spatial extents compared to the full image classification, which had a crisper and noisy representation. This supports various assertions in several studies where subpixel classification has been implemented (Aplin and Atkinson, 2001; Arif *et al.*, 2015; MacLachlan *et al.*, 2017). A major reason for the observed disparity could be as a result of the presence of multiple signatures from conflicting land cover features causing high spectral mixing for the RF classifier. Fröhlich *et al.* (2013) also observed that textural characteristics of neighboring adjacent features can inadvertently cause false representation of image features. Similarly, the spectral diversity of the features investigated had a smaller separability index as observed from the out of bag error for the full study area image. This can indirectly inhibit the RF classifier in adequately producing generalizable extents.

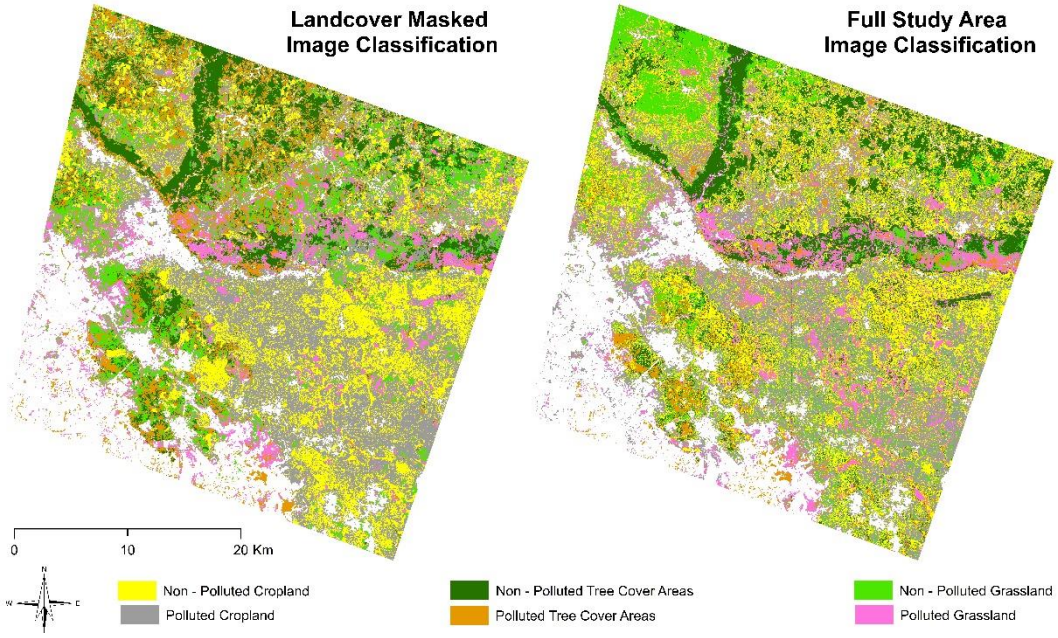


Figure 4-3: *RF* Image Classification Result for the Full Study Area Image and Individual vegetation Subsets. It is observed that the former produced a more generalized representation of vegetation extents compared to the crisp output from the full study area image

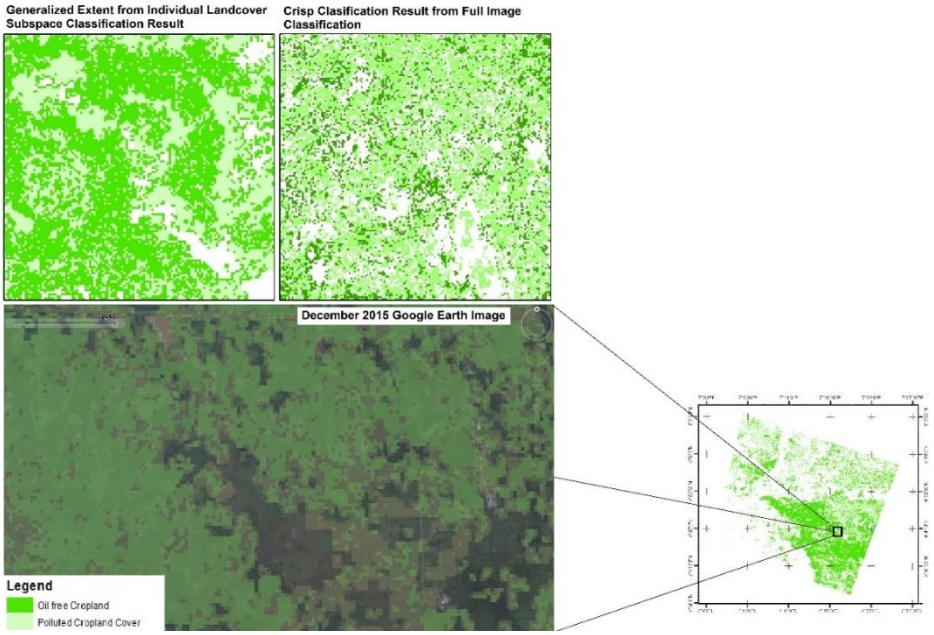


Figure 4-4: Subset of the study area showing the *RF* classified image for cropland vegetation extent into polluted and oil-free cropland. Inset is high-resolution image from google earth for the same area. This showed that spill impacted and oil-free cropland were better captured by the image subset classification (left), compared to the more crisp extent from the full study area image classification (right).

4.3.3 Variable Importance

The variable importance chart is a standard output from a typical RF classification operation (Gislason *et al.*, 2006). This essentially provides an insight on the performance of the sub-sampled variables that gave most of the best splits in the classification process. Figure 4-5 shows the variable importance chart for the full study area and vegetation type (subset) Image Classification results.

The Near Infra Band had the highest contribution to the assignment of endmember classes for the six classes (i.e. oil-free cropland, oil-free grassland, oil-free TCA, polluted cropland, polluted grassland and polluted TCA) when the full study area image with all land cover types was classified at once (Figure 4-5). Other variables however such as moisture stress index, normalized difference water index, shortwave infrared 1 (Mid infrared region) and the green band also contributed substantially in the classification process.

At the subset level, the result showed that the Normalized Difference Water Index and Moisture Stress Index were very influential in providing the best splits between polluted and oil-free cropland vegetation class. This conforms with results obtained in Kalubarme and Sharma (2014) where NDWI index is observed to be sensitive to water stress conditions in wheat cultivated farm plantations. Similarly, results obtained by Benabdelouahab *et al.*, (2015) also showed that MSI and NDWI are sensitive indicators of stress also in a wheat cultivated farm field.

However, the Near Infrared and Shortwave Infrared bands were also observed to have the highest contribution in splitting oil-contaminated and oil-free grassland vegetation class. The Difference Vegetation Index (DVI) and Normalized Differential Water Index clearly had strong contribution in splitting oil-polluted from oil-free TCA.

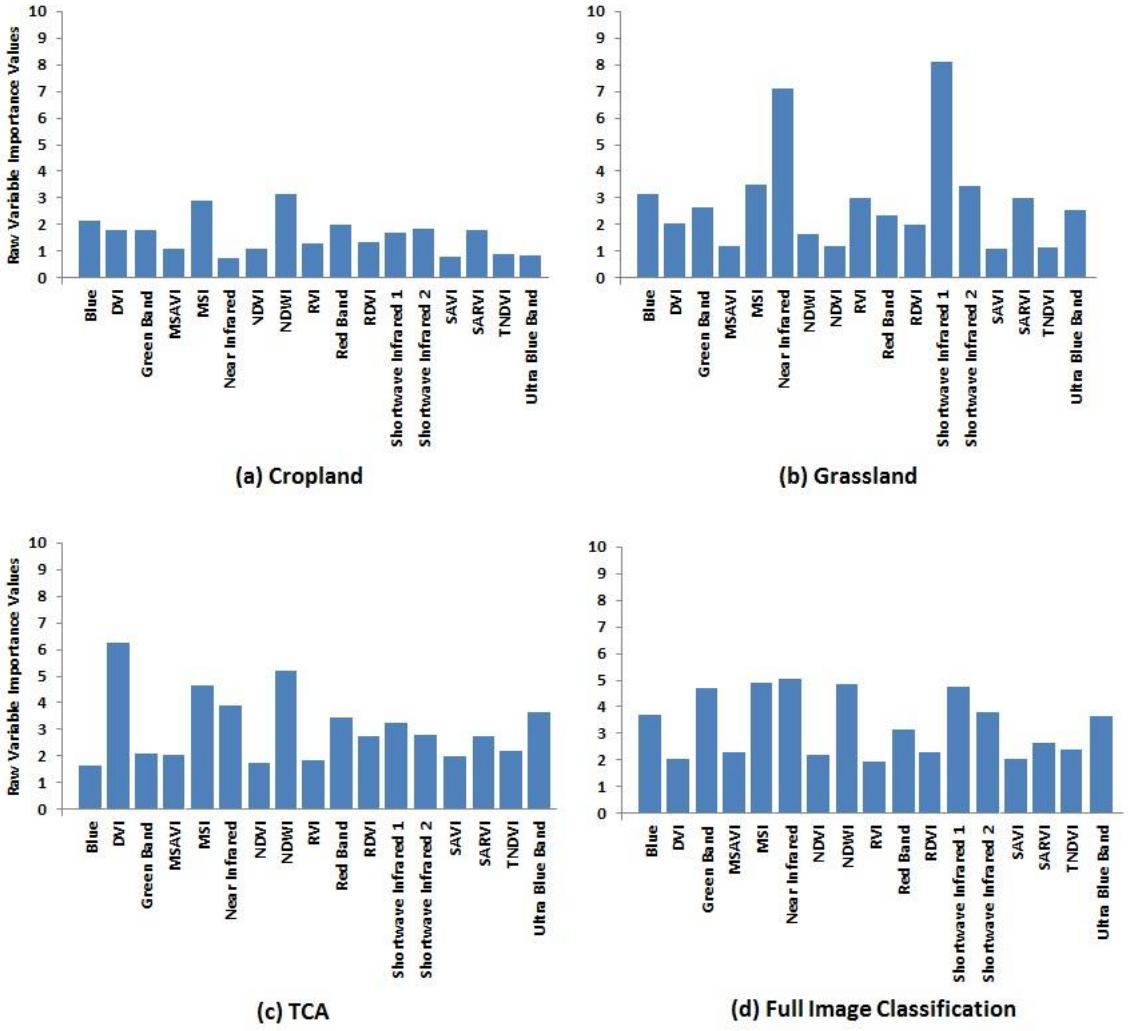


Figure 4-5: Variable Importance chart of the RF classification operation for the full study area image and vegetation masked image classification

In general, the moisture related indices and sensitive bands (shortwave infrared 1) were observed to have a more significant contribution in distinguishing oil-polluted from oil-free vegetation types both at the macro level of the entire study area and at the micro level of the individual land cover subsets. This is expected as the fundamental characteristics of stressed vegetation are their inability to carry out basic life supporting functions such as respiration, transpiration and photosynthesis (Arellano *et al.*, 2015), which the classifier can evidently rely on from the distinctions provided by the indices for class assignment. Figure 4-6 shows the most important variable (i.e. NDWI, SWIR and DVI) in the classification process for cropland, grassland and TCA vegetation subset respectively and their respective oil-free and oil-polluted extents.

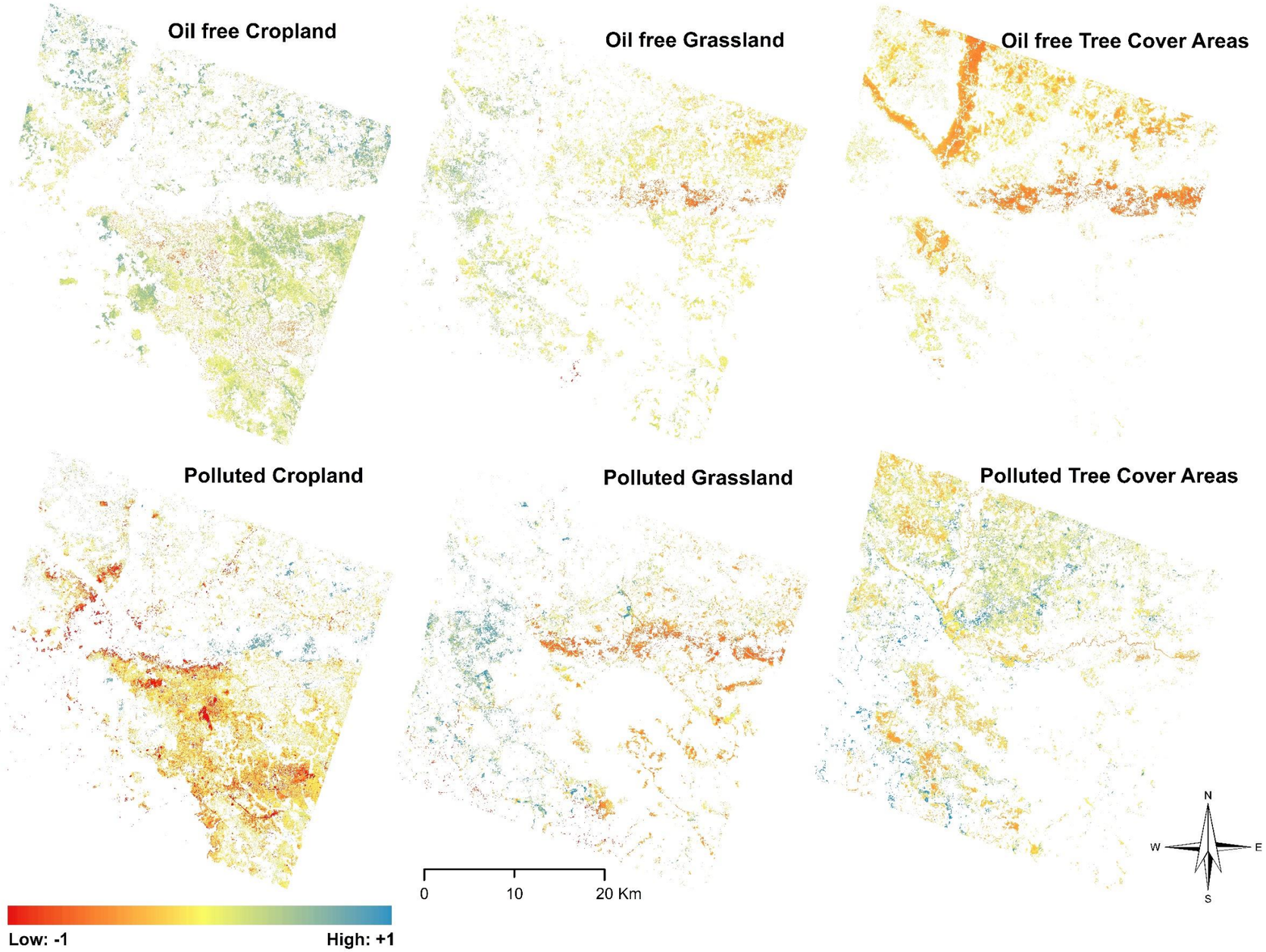


Figure 4-6: The most important variable for each vegetation type classification. NDWI, SWIR and DVI for cropland, grassland and TCA. Result shows that the most important variable for cropland and grassland classification had the best split into oil-impacted and oil-free vegetation, as opposed to the TCA subset where the most important variable did not give favorable split into oil-impacted and oil-free TCA.

Figure 4-6 indicates that areas with high vegetation health indices and greenness are predominantly associated with oil-free vegetation especially for the oil-free cropland and grassland vegetation classes. Areas with low vegetation health and greenness are mostly associated with polluted vegetation classes, in this case the polluted cropland and grassland.

However, TCA had a poor split as indicative of the most important variable from the RF classification (Figure 4-6). This could be associated to the fact that most areas within the Niger Delta are characterized by dense and mangrove forest vegetation (James *et al.*, 2007), in which case the impact of crude oil would pose minimal discernible effect within a typical oil free vegetation.

4.3.4 Vegetation Greenness Distribution

Figure 4-7 is a box plot showing vegetation greenness retrieved from NDWI for the various polluted and oil-free vegetation training sites. This was the most influential index when the full study area image was classified together with the moisture stress index (MSI). Their performance in the classification operation further reinforces the strength of moisture-based indices in depicting stress on vegetation. The plot shows the degree of variation in the health status of the oil-impacted and oil-free vegetation classes. Non-polluted tree cover areas were observed to have the highest NDWI index compared to the non – polluted cropland and grassland.

Generally, polluted grassland and cropland had the least NDWI greenness compared to their respective non – polluted classes. This is an indication that their health status could have been affected by the oil spill in the respective locations, thereby accounting for lower health indices compared to the respective oil-free vegetation. Similarly, the distribution of the indices for the six classes shows little to no overlap between oil-polluted and oil-free vegetation, a trend which could have accounted for the high performance of the NDWI in the classification process.

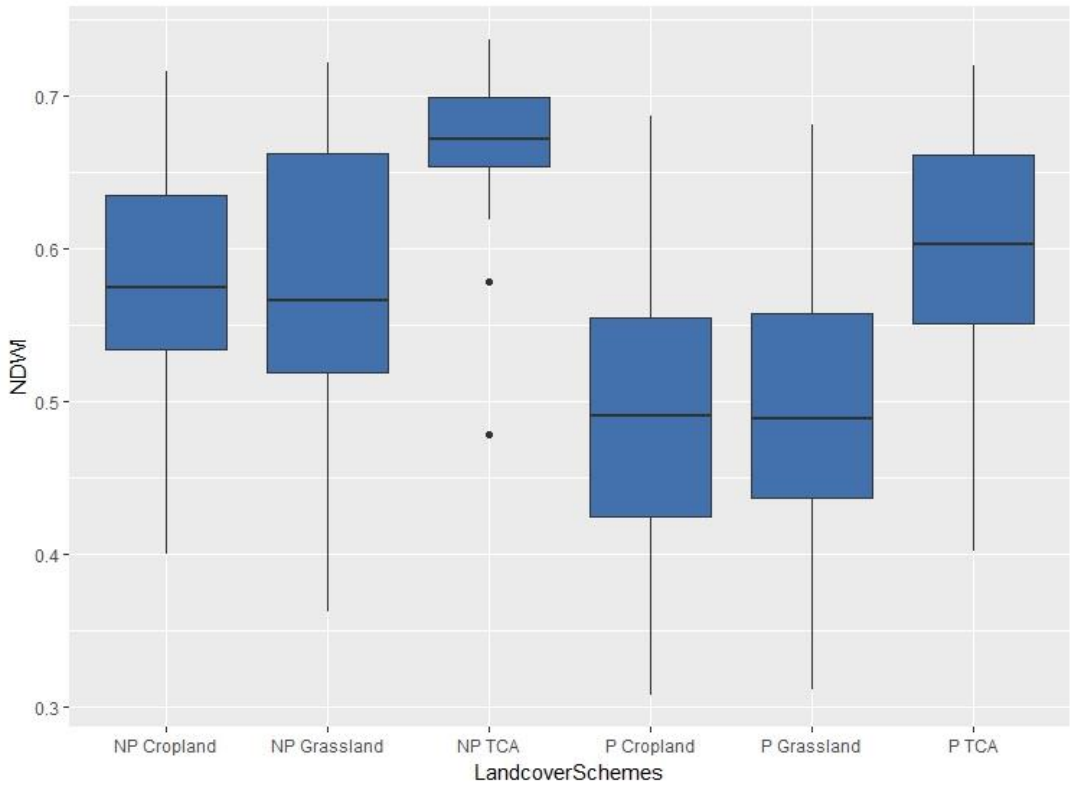


Figure 4-7: Box plot of vegetation greenness retrieved from NDWI for the polluted and oil-free vegetation samples.

4.3.5 Accuracy Assessment

The confusion matrix was computed with the validation data (section 3.2.3) (Table 4-4) and used to evaluate the result of the RF classification for the two scenarios implemented. The overall accuracy from the full image classification gave much lower accuracy of 45.45%, 33.33% and 28.5% overall accuracies for cropland, grassland and tree cover areas, respectively compared to the result obtained from the various land cover vegetation subset. Results from the subset classification of tree cover areas gave the highest accuracy of 70%, while the grassland and cropland classification gave accuracies of 65% and 60.61% overall accuracy respectively. In terms of inter class accuracy, the result from the confusion matrices showed that the highest user accuracies were obtained from the Non – Polluted grassland and Polluted TCA with 80% from the subset classification.

Similarly, the vegetation class with the highest accuracy when the full study area image was classified is the Polluted and oil free TCA classes with Producer and User Accuracy's of 50% and 40% respectively. This is not surprising as the result from the

parameterization operation in Figure 4-2 showed that the training sites used for classification had better characterization between polluted and oil-free TCA.

Furthermore, the result also showed that most of the classes that had better calibration also recorded higher accuracy. An example is in the case of TCA and grassland vegetation which recorded high accuracies of above 80% out of bag error, and also 70% and 65% overall accuracies at the micro-level classification.

Table 4-4: Accuracy Assessment Result for the Full Study Area and individual vegetation subset Classification

Map Class	Full-Image Classification			Masked-Image Classification		
	User's Accuracy [%]	Producer's Accuracy [%]	Overall Accuracy [%]	User's Accuracy [%]	Producer's Accuracy [%]	Overall Accuracy [%]
Non-Polluted Cropland	25	18.75	45.45	58.82	62.5	60.61
Polluted Cropland	29.17	41.18		62.5	58.82	
Non-Polluted Grassland	18.18	20	33.33	61.54	80	65
Polluted Grassland	30	30		71.43	50	
Non-Polluted Tree Cover Areas	50	40	28.5	75	60	70
Polluted Tree Cover Areas	37.5	30		66.67	80	

4.3.6 Spill Impacted Vs Non Spill Spatial Extent

Figure 4-8a and 4-8b present a stacked bar plot comparing the total estimated area covered by oil-impacted and oil-free vegetation classes from the full study area and individual vegetation subset classification, respectively. The results were compared to the total area coverage for the different land cover as provided by the ECCI. Generally, the results showed that aggregated areas of polluted and oil-free classes were closer to the areas from the ECCI when the image subsets are classified than when the full image is classified.

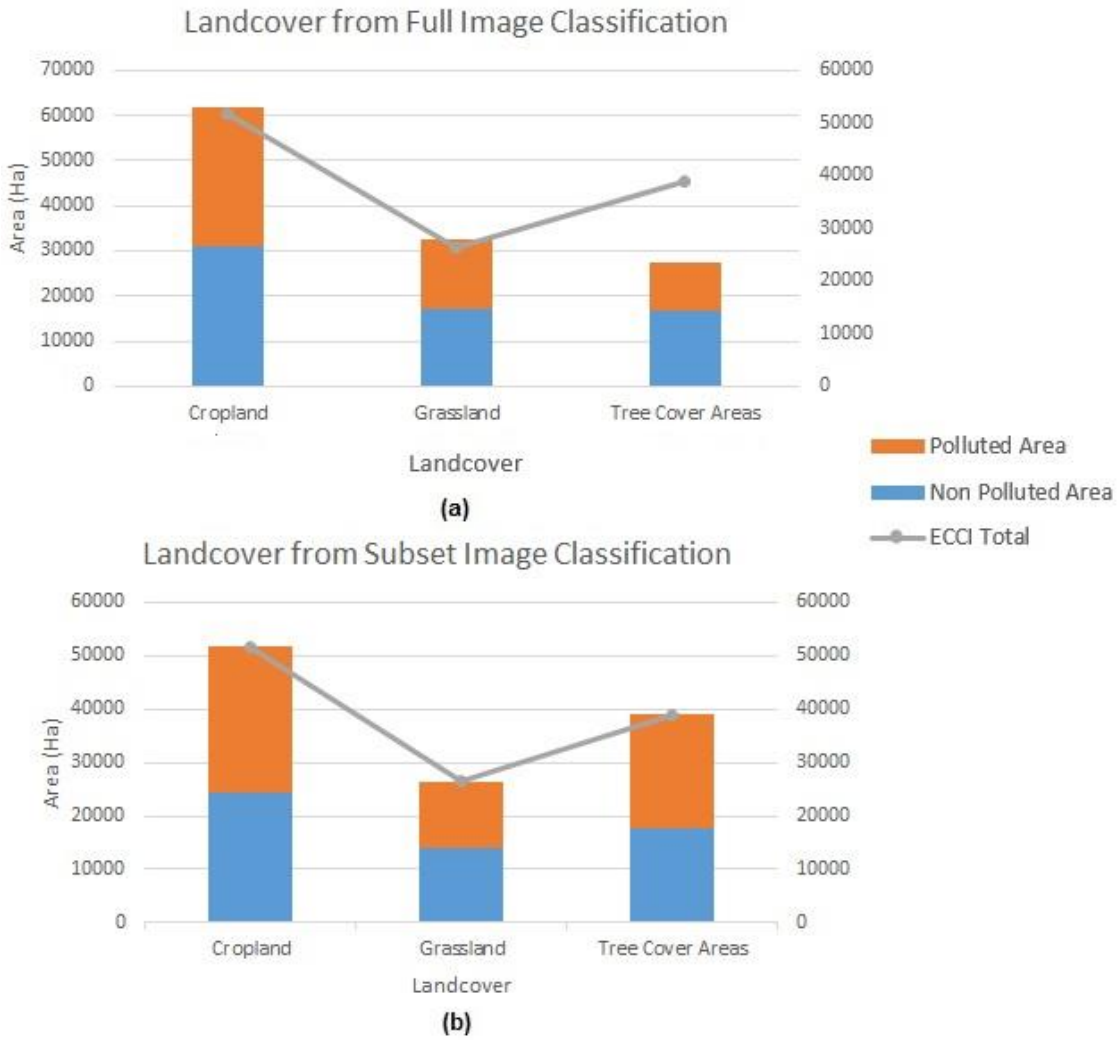


Figure 4-8: Spatial extent of oil-impacted and oil-free classes retrieved from **(a)** full study area image and **(b)** vegetation subset classification. The orange and blue stacks represents polluted and oil-free vegetation classes, while the ash coloured line represents the aggregate obtained from the ECCI land cover dataset.

Similarly, the extent of spill impacted grassland and TCA were larger than their respective oil-free vegetation, except in cropland vegetation where the area covered by oil-free cropland had a larger coverage than the oil-impacted cropland. In addition, of the six vegetation classes used in the analysis, the spatial extent of oil-impacted cropland in the full study area image and cropland vegetation in the subset image classification remained similar. This suggests that the spectral characteristics of the polluted cropland remained largely unchanged in the two experimental classifications. This is an indication that this class could have been more heavily impacted from the 2015 and 2016 spill incidents in the area.

4.3.7 Field Validation Result Using Field Spectroscopy

Figure 4-9 shows the mean spectral reflectance curves for vegetation assessed in the recent, old and non-polluted sites. The result of this exercise indicates that plants within the non – polluted site are predominantly healthier and highly photosynthetic. This is because of the high absorption of visible energy between the 325nm and 650nm regions of the wavelength spectrum and high reflectance of NIR energy starting from the Red edge region of 750nm to 1075nm of the spectrum. The absorption – reflectance characteristic of vegetation within the two polluted sites shows that they have much lower absorption and reflectance of visible and NIR energy respectively.

The vegetation spectral profile of plants within the recent polluted site showed that much of the vegetation within this site is partly photosynthetic, which could be as a result of the impact of hydrocarbon crude oil pollution. Similarly, the result obtained from the old polluted site also showed that the plant community were partially photosynthetic, as the mean spectral reflectance within the Near Infrared region fell between the healthy non – polluted plant community and the stressed plant community within the recent polluted site. This however suggests that the impact of crude oil pollution might still be present and pose some impact on plant community within the old polluted site.

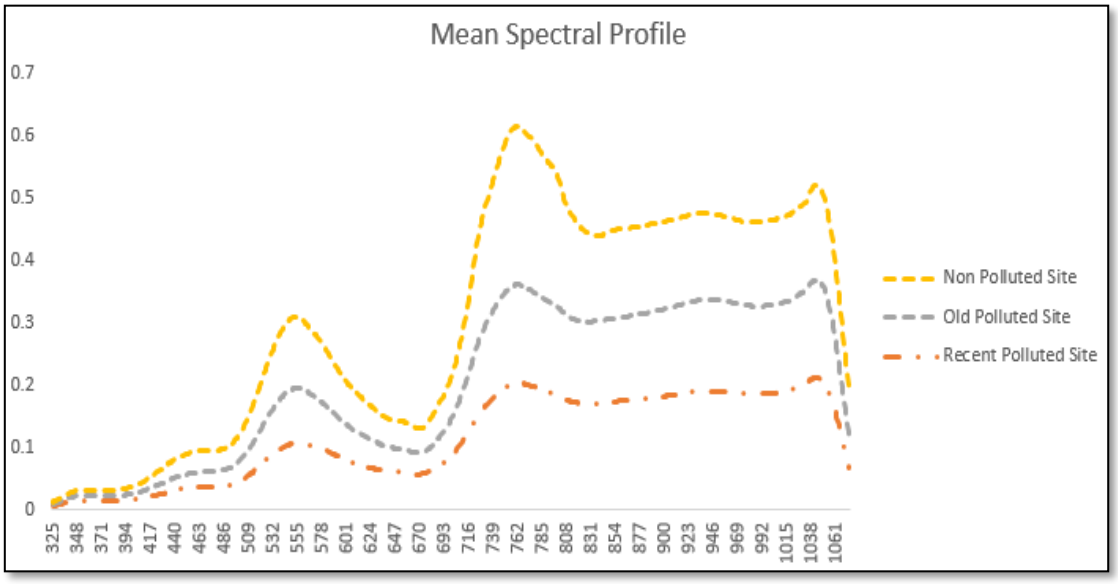


Figure 4-9: Spectral Reflectance Curves for the three sites investigated. Separate lines denote the mean reflectance spectra for the recent, old and non – polluted sites.

Results obtained from the field spectroscopy using the Handheld spectroradiometric device depicts that the potential elevated levels of hydrocarbon crude oil observed during the field visit (Figure 3.5) within the recent polluted site had significant impact on the vegetation within the experimental plots. This is because the result clearly indicates that there is low absorption of green and red energy within the visible portion of the spectrum and low NIR reflectance. However, when compared to the old and non – polluted site it is observed that there is high absorption of green and red energy, as well as high reflectance of NIR energy from vegetation within these sites.

This is in line with several studies (Arellano *et al.*, 2015; Kokaly *et al.*, 2013; Mishra *et al.*, 2012; Zhu *et al.*, 2013b) where similar technique have been employed to investigate the adverse effect of hydrocarbon crude oil on vegetation. It was observed that hydrocarbon crude oil significantly alters the biochemical components of plants, leading to low chlorophyll generation, which is essential for the absorption of green and blue energy for carrying out photosynthesis in the oil polluted site (Arellano *et al.*, 2015; Arellano *et al.*, 2017b). This can also lead to partial or complete defoliation of plant leaves in oil affected vegetation (Mishra *et al.*, 2012).

In contrast, the vegetation within the non – polluted site did not show as much of physical stress also compared to the old – polluted site, as the reflectance values retrieved were quite high within the near – infrared region of the spectrum.

4.4 Discussion

Oil pollution and contamination of vegetation canopies within the Niger Delta region is a common and almost a consistent phenomenon. Few studies have focused on leveraging on the potentials of machine learning approaches (such as RF) to map the exact oil spill extent for different vegetation types. This study attempted to bridge this gap by using RF classification to first establish the precise extent of oil-impacted and oil-free vegetation types and secondly to identify the most useful optical discriminators for distinguishing oil-impacted vegetation from the corresponding oil-free vegetation. The result obtained from these experiments after calibration of sample sites and implementation of the classification operations showed that RF algorithm has the potential of providing reliable maps of oil-free and oil-impacted vegetation, especially at the micro-subset level of classification. The RF Classifier produced better results with the different vegetation subsets (especially for grassland and TCA) as opposed to

when the full study area image is classified at once, reinforcing the findings of Arroyo, Johansen and Phinn, (2010) were image space delineation for automatic classification of land cover features proved very successful.

The high calibration accuracy obtained from the out of bag errors during the parameterization exercise of the RF at the micro level clearly accounts for the high accuracies of 70% and 65% obtained for the tree cover areas and grassland vegetation types respectively. These values were much higher than accuracies obtained in Bianchi *et al.* (1995a) and Van der Werff *et al.,* (2007) where minimum distance to means, spectral angle mapper and hyperspectral unmixing methods were used to process hyperspectral images. Overall accuracy assessment result obtained in Van der Werff *et al.,* (2007) ranged from 48% and 29% in the mapping of microseepage induced anomalies on baresurface. While spectral angle mapper determines end member classes by matching reference pixels to target class based on proximity of spectral angle to established threshold in the *n*-Dimensional space. Minimum-distance to mean method assigns classes to target pixels based on pixel values arithmetically close to mean spectral values of training sample spectra. Random forest method on the hand uses boosting and bagging technique where multiple arrays of input features are selected at random without replacement to train the classier. This to a large extent ensures that the effect of over-fitting is reasonably addressed; as such classes are not assigned based on predefined threshold which could give rise to over-generalization.

The result of the confusion matrix for cropland when the masked image classification operation was implemented shows that more than half of the sample sites were correctly classified into their respective class of polluted and oil-free cropland. However, when compared to the confusion matrix (Appendix – XII) results of the full image classification, it is observed that only 3 and 7 of the total reference classes were classified correctly for the non-polluted and polluted cropland classes respectively. It was also evident from Appendix - XII that reference sites for polluted cropland were mostly mis-classified as non-polluted cropland or non-polluted grassland. This implied that the RF classifier was able to identify more polluted and non-polluted sites correctly when the masked image classification process was implemented than when the full study area image classification was implemented. Non-polluted grassland cover had the best prediction to error ratio as 80% of the total reference sites were correctly classified accordingly, while polluted grassland had an average of 50%

reference site prediction accuracy. In tree cover areas, there were better reference site prediction accuracy for the polluted TCA, as 80% of the total reference sites were correctly classified as polluted TCA at the subset level of classification. In addition, more than half (i.e. 60%) of the total reference sites for Non-polluted TCA were also correctly classified at the subset level representing nearly 200% improvement in classification accuracy. This is because the classification accuracy at the entire image level only yielded 28.5% overall classification accuracy, as only 40% and 30% of the original reference sites of Non-polluted TCA and polluted TCA were correctly classified.

However, the result of the most important variable in the classification process (Figure 4-5) does not mirror an excellent split for TCA and grassland subsets compared to cropland, which had a good split between polluted and oil-free classes. A major reason for this trend can be attributed to the fact that most cropland vegetation are distinctly sparse in nature and a huge volume of the oil spilt in these areas experience significant seepage into the soil sub surface and immediately causing detectable impact on crops. This invariably account for a better split of oil-polluted and oil-free cropland, as indicative of the NDWI.

Similarly, the exposed soil in cropland fields also means that much of the oiled sand surface is reflective, accounting for the significant influence of shortwave infrared band (Ben-Dor *et al.*, 1997; Cloutis, 1989; Kühn *et al.*, 2004) and its derived indices in distinguishing oil-impacted from oil-free croplands (Adamu *et al.*, 2015; Ben-Dor *et al.*, 1997; Brekke and Solberg, 2005; Khanna *et al.*, 2013; Kühn *et al.*, 2004).

The variable importance plot obtained from the RF image analysis also showed that the Near Infrared, Shortwave infrared bands, Normalized Difference Water Index, Difference Vegetation Index and Moisture Stress Index are particularly influential in pixel class assignment. Some of these variables (shortwave infrared, moisture stress index and NDWI) are mostly sensitive to vegetation moisture content (Gao, 1996). Several studies (Agapiou *et al.*, 2012b; Arellano *et al.*, 2015; Benabdelouahab *et al.*, 2015; Dotzler *et al.*, 2015; Kalubarme and Sharma, 2014) have also demonstrated that SWIR, MSI and NDWI variables are a useful indicators of stress in vegetation canopy as a result of their sensitivity to water net loss or gain.

Similarly, the NIR band is also well known for its ability to distinguish between stressed and stress free vegetation. This is because a major characteristic of a stress free vegetation will be the absorption of visible light for photosynthesis necessary to propagate the high reflectance of near infrared energy (Ben-Dor *et al.*, 1997; Knipling, 1970). It is without doubt that these variables have the most ideal spectral information to characterize oil-free from oil-polluted vegetation. The complex interaction of these variables is a major reason for their investigation in this research, suggesting that vegetation stress as a result of oil pollution can be better characterized and mapped.

In addition, the result obtained from the spatial extent of the classified maps for polluted and oil-free vegetation further suggests that cropland cover had the most significant impact, as the area coverage and extent obtained from the full study area image and cropland vegetation subset classification remained similar. This contrast to the results obtained from the TCA and grassland vegetation classification, where the spatial extent of the polluted class had a much higher area than their non-polluted/oil-free class. A possible reason for this trend could be as a result of over generalization of the extent of spill-impacted vegetation overlapping with other areas where vegetation stress by other types of stressors exist.

A post-classification ground truth exercise carried out in October 2018 showed that features such as water logged areas, dried vegetation, burned vegetation and cleared/exposed vegetation surface often exhibited similar spectral signatures as polluted sites and were classified as such. This is in line with observations made by Khanna *et al.*, (2013) and Kokaly *et al.*, (2013). Most of the aforementioned misclassification anomalies are also vegetation stress related, accounting for the superior performance of the NDWI, NDVI, SWIR and NIR in the classifications processes. Figure 4-10 shows some the areas that exhibited similar spectral response.

(a) (b) (c)

Figure 4-10: Potential influences to pixel mis-classification of oil polluted site. (a) Water Log areas (b) Cleared and exposed surfaces (c) Dried Vegetated areas which often leads to burning/burn scars.

In addition, the results from the field spectroscopy analysis of polluted and oil-free vegetation undertaken in the course of this research also demonstrated clearly that impacts of oil pollution are strong in both the recent and old spill site, due to reduced NIR reflectance compared to the high NIR in the oil-free site. Generally, the polluted sites and other aforementioned features such as water, exposed surface (bare land) and burnt surface share similar characteristic low reflectance within the NIR region as indicated in figure 4-11. This is an indication that net primary productivity of vegetation within oil-affected areas can remain characteristically low long after the incident, thereby influencing detectability. As such, for effective detection and characterization to take place between polluted and oil-free vegetation, healthy vegetation must be properly identified. However, this may not completely eliminate pixel mis-match in the classification process.

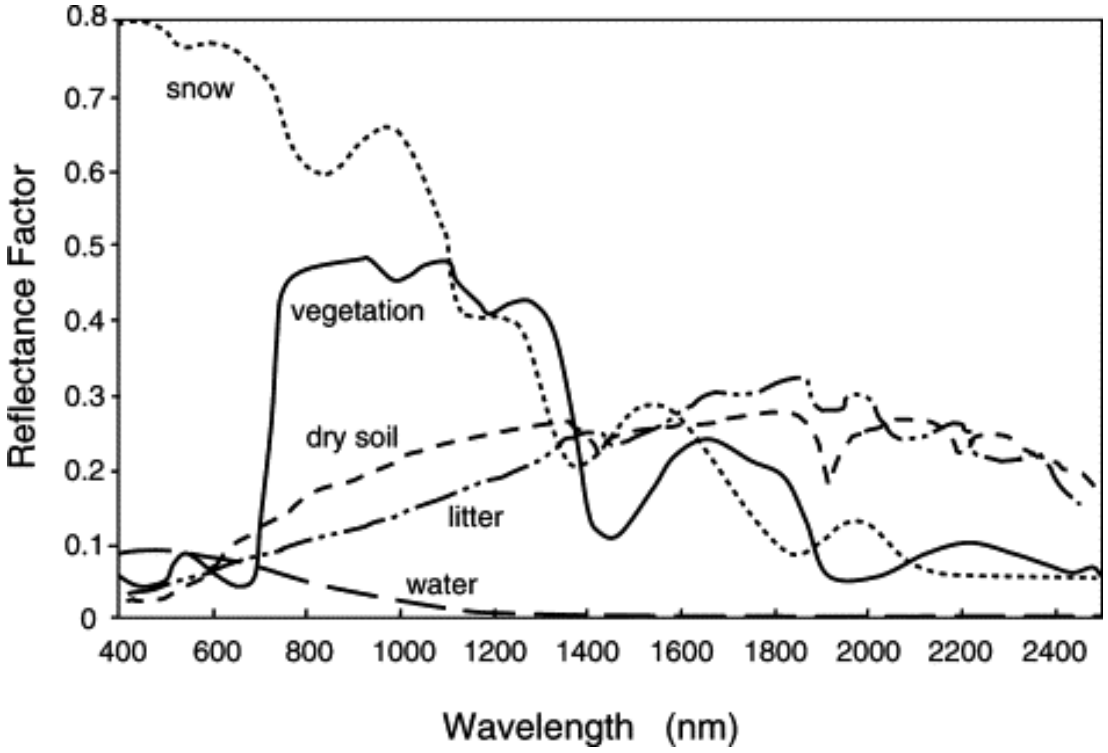


Figure 4-11: Spectral reflectance curve for various land cover features with emphasis on the low NIR reflectance of dry soil, litter and water body, compared to the high NIR reflectance from healthy vegetation. Source: (Huete, 2004)

The problem of pixel mis-classification in image classification is a general problem as also observed by other researchers e.g. (Ishida *et al.*, 2018; Xiao and McPherson, 2005; Zlinszky *et al.*, 2012) where the characterization of a single vegetation type into a narrower group by species or health status delineation has been implemented. The occurrence of pixel mis-match and over generalization of land cover spatial extent is very much apparent in this study. One way of addressing this problem is the incorporation of other relevant variables (such as Radar datasets, Digital Elevation Model, Soil - Geology map, as well as soil moisture) which generally do not specifically rely only on the bio-chemical components of vegetation, rather the structural characteristics of vegetation and environmental factors are relied on for class assignment.

Similarly, the result of the RF image classification at the macro (study area) level showed that the NIR band was characteristically useful in discriminating between polluted and oil-free vegetation. This could be connected with the presence of other features such as bare surface, built-up areas and waterbody in the classification process, reinforcing the utility of the NIR band in the discrimination of these features in the classification process. In contrast however, the influence of the NIR band is

minimal in the land cover specific classification (i.e. at the micro level), as the influence of bare soil and waterbody signature is substantially reduced. This saw the influence of other wave bands such as the SWIR and moisture stress indices.

However, it must be noted that the concentration and size of spill also plays a significant role in the detection and mapping of affected areas using satellite image. A number of other studies (Adamu *et al.*, 2016; Khanna *et al.*, 2013; Kokaly *et al.*, 2013; Mendelssohn *et al.*, 2012; Mishra *et al.*, 2012; Noomen *et al.*, 2015; Onwurah *et al.*, 2007; Ramsey *et al.*, 2015; Ramsey III *et al.*, 2011; Shi *et al.*, 2007; Sun *et al.*, 2016; Zabbey and Uyi, 2014) have shown that the size of oil spill with respect to volume and age of oil are a major determinant of detectability of spill effect. This is largely predicated on the fact that not all spill incidents come in large sizes or quantities that can be meaningfully captured by the satellite sensors, or pose detectable stress on vegetation communities. In this study, this challenge was addressed by using only spills with 1000sqm or above in size to ensure that the characteristics of a typical spill site are reasonably captured within the spill epicenter and adjacent pixel used for classification. It was however observed that other stress factors and features with same spectral characteristics can be potentially misclassified as oil-polluted vegetation with this threshold. These certainly call for further research, especially using fuzzy techniques in establishing precise spill threshold values for adequate detection and classification purpose, or the incorporation of other variables to improve the machine learning models used for the discrimination exercise.

4.5 Summary and Main Findings

This chapter of the thesis aimed at implementing a novel method of mapping terrestrial oil spill through the integration of optical spectral reflectance bands and vegetation health indices in a machine learning random forest classifier to discriminate polluted and oil-free vegetation. Another novel approach was also introduced by specifically focusing the detection and discrimination process for different vegetation types, which has not been tested in the literatures. Initial results from field spectroscopy analysis of polluted and oil-free vegetation before image analysis showed that the impact of oil pollution was substantial in recent and old oil spill sites compared to the oil-free (control) site. This is owing to reduced NIR reflectance in the former compared to the high NIR in the latter, further reinforcing the rationale for the use of both old and recent spill sites in the calibration and classification process in this research.

Results obtained from the implementation of the image classification process in this chapter showed that RF algorithm provides better results of oil-free and oil-impacted vegetation when the image classification is carried out at the micro-subset level of different vegetation types, as opposed to when the entire full study area image containing all classes are classified at ones. However, the problem of pixel mis-match and over generalization of land cover spatial extent was imminent in the results obtained. Ground truth exercise carried out in October 2018 to ascertain the major reason for these confusion showed that water logged areas, dried vegetation, burned vegetation and cleared/exposed surface (which are also stressed vegetation) were very much present in the study area. They also had similar spectral signatures as polluted sites and were mis-classified as such.

This invariably necessitate further exploration of other relevant variables which generally do not specifically rely on the bio-chemical components of vegetation (i.e. using optical reflectance data), rather, the structural characteristics of vegetation and environmental factors are used and incorporated in the machine learning method as dependencies to unravel class membership. In the next chapter, these variables are incorporated into the machine learning random forest model to explore their capabilities in improving classification accuracy. Specific variables to be explored are Radar Images, Digital Elevation Model, Soil geology and Soil type map into the machine learning classification to attempt to improve classification accuracy. In addition, the potential of variable reduction using the fuzzy forest method would as be explored as the integration of the optical and SAR derived variables could present issues of high dimensionality.

Chapter 5 : Detection of Oil Pollution Impacts on Vegetation Using Multi-frequency SAR, Multispectral Images with Fuzzy forest and Random forest methods

5.1 Introduction

In this Chapter, the potentials of C and X band SAR derived textures, interferometric coherence, backscatter, topography and soil-geology variables, as well as multispectral optical image derived variables for identifying oil pollution using machine learning fuzzy forest (FF) (Conn *et al.*, 2015) and random forest (RF) (Breiman, 2001) classifiers were investigated. The rationale for the choice of classifier is that RF classifier has high flexibility in the use of input variables for describing different situations. This should allow for optimal use of diverse range of input variables in the discrimination of polluted and oil-free vegetation. In addition, the potential of FF was also explored. The FF method seeks to use only high performing non-correlated variables for classification. The potentials of multi-frequency SAR image fusion (MSIF) only classification; and multi-frequency SAR and Optical image fusion (MSOIF) classification was specifically investigated in the discrimination of oil-impacted and oil-free cropland, grassland and tree cover areas.

Furthermore, effort was made to model the relationship between bio-chemical indicators of vegetation stress as captured in the derived vegetation indices, with vegetation structural component as captured in the various multi-frequency SAR backscatter, to understand the relationship between optical and SAR data integration for monitoring vegetation affected by oil pollution. It is also worthy of note to mention that the image analysis implemented in this chapter was carried out on the different vegetation subset, as opposed to the entire study area image (as previously implemented in chapter 4) . This is because results from previous analysis chapter (i.e. chapter 4) of this thesis suggests that detection and characterization of vegetation affected by oil pollution yield substantially better results at the micro-level of vegetation type than at the macro-level consisting of several land cover features in the image space.

5.2 Data Preprocessing

Datasets used in this chapter ranged from remote sensing multispectral optical images and Synthetic Aperture Radar (SAR) images (Table 5-1), as well as other secondary derived ancillary dataset such as digital elevation model, soil and geologic map. Figure 5-1 shows the overall methodological and data processing flow chart.

Table 5-1: List of all Remote sensing data

Platform	Sensor	Swath	Spatial res.	Image dates	Season
Sentinel 1	C SAR	250km	5x20m	Jan. 2017	Dry Season
Tandem X	X SAR	30km	6m	Feb. 2016	Dry Season
Cosmo Skymed	X SAR	40km	3m x 3m	Dec. 2016	Dry Season
Sentinel 2	Optical	290km	20m	Dec. 2016	Dry Season
Endeavour Shuttle	SRTM	225km	30m	Sep. 2014	-

5.2.1 Optical reflectance data and vegetation indices from Sentinel 2

The major pre-processing operation implemented was the radiometric correction using the Sen2Cor tool in the Sentinel Application Platform (SNAP) software environment (Zuhlke *et al.*, 2015). This converted the data from radiance to top of atmosphere reflectance. The image pixels as contained in the six spectral bands with 20m spatial resolution were then resampled to 10m spatial resolution to ensure that information across the different spectral bands were uniform for analysis. The radiometrically corrected images were converted to native coordinate system of UTM Zone 32 North and WGS1984 Datum. This was done to ensure that spatial information contained in the spill data, the SAR image derived variables (i.e. interferometric coherence and backscatter), digital elevation model and geo-physical variables (Soil – Geology) effectively coincides and geometrically aligned with the S2 image.

The final pre-processed bands were used as direct input in the experimental image classification and generation of Vegetation Health Indices. Three vegetation health indices were generated and incorporated into the machine learning classification, as they were particularly efficient in identifying stressed vegetation in previous studies (Adamu *et al.*, 2015; Arellano *et al.*, 2015). This include the Normalized Differential Vegetation Index (NDVI), Leaf Area Index (LAI) and the Normalized Differential Water Index (NDWI). The NDVI is widely used for remote sensing of vegetation health because of its ability to depict stress in vegetation. It derived using radiance from the red channel, which is the strong chlorophyll absorption region, and Near Infrared, which represents a high reflectance plateau of vegetation canopies (Eq. 8).

$$NDVI = \frac{NIR-RED}{NIR+RED} \quad (8)$$

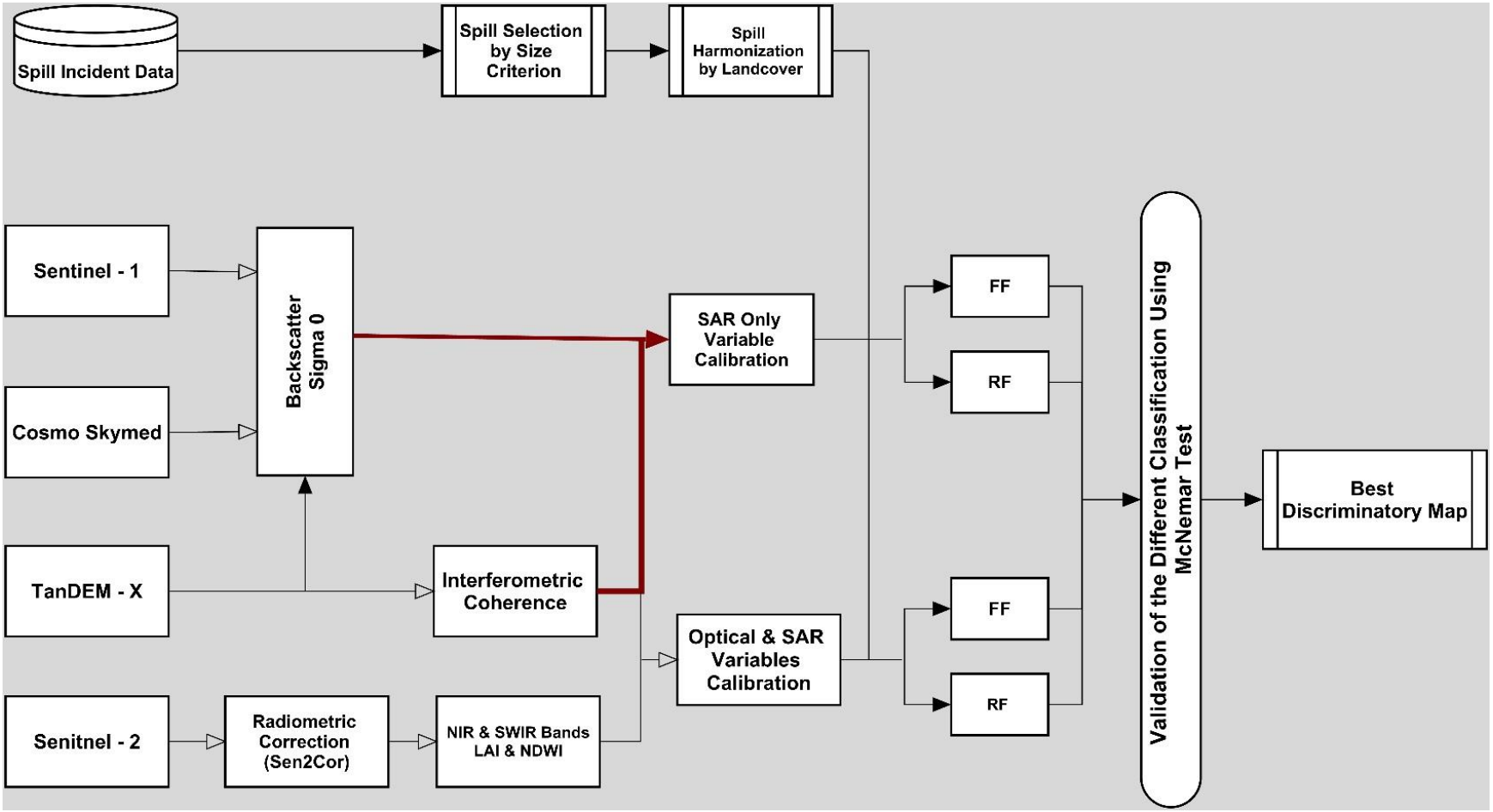


Figure 5-1: Methodological Flowchart of the Data Analysis for **Chapter 5**

The NDWI is another index proposed by Gao (1996). The NDWI basically uses the Mid Infrared and near infrared bands, which are located in the high reflectance plateau of vegetation (Eq. 9). However, owing to the weak liquid absorption in the mid infrared, the index is sensitive to change in liquid water content of vegetation, and vegetation with near or absolute water loss is effectively detected with the index than with NDVI.

$$NDWI = \frac{NIR-SWIR}{NIR+SWIR} \quad (9)$$

The leaf area index is commonly defined as the ratio of green leaf area to horizontal ground surface. Various methods ranging from field based measurement and remote sensing data are used to compute LAI. LAI is an important variable for establishing gross photosynthesis, net primary productivity, evapotranspiration and bi–directional reflectance as it depicts the structural properties of vegetation. The index can reveal a lot about the health and structural state of vegetation. In this chapter, LAI from Sentinel 2 optical imagery using the bio-physical processor in the SNAP software (Zuhlke *et al.*, 2015) was generated. The processor primarily relies on radiative light transfer process from 8 Sentinel 2 bands, sun zenith and viewing zenith and viewing geometry to estimate LAI within a three layered neural network.

5.2.2 Normalized Radar Cross Section Backscatter

5.2.2.1 Sentinel 1 – GRD Product

In order to generate the normalized radar cross section backscatter for the Sentinel – 1 A GRD image. The scene was first radiometrically calibrated before multilooking (one look in range and four in azimuth), then geocoded based on Shuttle Radar Topography Mission (SRTM) data, and radiometrically calibrated with a final pixel spacing of 10m × 10m. Pixel values were then converted to backscattering (or normalized radar cross section), measured in decibels (Db) using the formula in Equation 10 within the SNAP software.

$$\sigma^0\text{dB} = 10 \log 10\, \sigma^0 \quad (10)$$

Where

dB Normalized radar cross section

σ^0 is the backscatter for a specific polarization

log is the decimal logarithm.

5.2.2.2 TanDEM X

The TanDEM-X is the second SAR image used in this study. However, only the return signal in the HH polarized channel was used in this analysis, as the cross polarization HV channel data was not readily available for the desired period and for the study area. The acquired image was first radiometrically calibrated before Terrain correction was implemented using SRTM data. The final backscatter values were further converted from DN to linearized cross sectional values in decibels using the formula in equation 11 below (Sportouche *et al.*, 2012):

$$\sigma_{0_{db}}^{TSX}(i,j) = 10 * \log\left(\left|x_{(i,j)}\right|^2 * \sin\left(\theta_{inc}^{loc}\right) * k_s\right) \quad (11)$$

Where

θ_{inc}^{loc} is the local incidence angle

k_s is the calibration and processor scaling factor

$x_{(i,j)}$ is the image value at pixel (i,j)

log is the decimal logarithm.

5.2.2.3 Cosmo Skymed

The cosmo skymed is the third SAR image used in this chapter. Major pre-processing on the cosmo skymed image include correction for the range spreading loss effect using antenna pattern gain compensation and incidence angle effect. The corrected image was further multilooked (one look in range and two in azimuth), geocoded based on the SRTM, and radiometrically calibrated with a final pixel spacing of 10m × 10m. Pixels values were converted to backscattering (or normalized radar cross section), measured in decibels using formula in equation 12 (Sportouche *et al.*, 2012).

$$\sigma_{0_{db}}^{CSK}(i,j) = 10 * \log\left(\frac{\left|x_{(i,j)}\right|^2 * \sin\left(\theta_{inc}^{ref}\right) * RSR^{2\alpha}}{CC * RF^2}\right) \quad (12)$$

Where

$x_{(i,j)}$ is the image value at pixel (i,j)

θ_{inc}^{ref} is the reference incidence angle

RSR is the reference slant range

α is the reference slant range exponent

CC is the calibration constant

RF is the rescaling factor.

Terrain correction was also implemented using the range Doplar terrain correction module in the Sentinel Application Platform (SNAP) (Zuhlke *et al.*, 2015).

5.2.3 Coherence

The interferometric coherence is a function of decorrelation of two SAR images, which measures the degree of similarity between two co-registered SAR image pairs. Coherence was only generated for the TanDEM-X data, as the image typically consist of two pairs in a simultaneous acquisition mode from different side-track of the TerraSAR and TanDEM-X. Coherence decreases with increasing volume scattering and temporal changes due to abrupt movement of targets such as vegetation and flooded areas (Bamler and Hartl, 1998). Coherence values range between 0 and 1, where 0 represents a low and incoherent target and 1 represents high and absolute coherence. Interferometric coherence was generated for the post spill TanDEM-X image in the SNAP software (Zuhlke *et al.*, 2015). Topographic phase was removed with the aid of the SRTM 3 data and the final product was multi-looked with ratio 2:2 to obtain the same spatial resolution as the Sentinel 1, Sentinel 2 and Cosmo Skymed images. The Refined Lee filter was used for noise suppression before range Doplar terrain correction was applied to geometrically correct the final coherence image.

5.2.4 Texture features

Textural features and extraction in the spatial domain were first introduced and applied using the Gray-level co-occurrence matrix (GLCM) by Haralick and Shanmugam, (1973). They provide meaningful spatial statistical descriptors of image features, which generally help to understand how the grey level scale of an image affects feature presentation (Eckert, 2012). Textural features were used in this study because they have the ability to depict important rapid change in vegetation structural composition, which in turn can influence grey level tonation of SAR images (Hlatshwayo *et al.*, 2019; Jin *et al.*, 2014). Eight GLCM were generated as prescribed by Haralick and Shanmugam, (1973) from the

high resolution Cosmo Skymed and TanDEM-X images. These include Contrast, Correlation, Dissimilarity, Homogeneity, Mean, Second Moment, Variance and Entropy.

5.2.5 Digital elevation model

The major pre-processing operation implemented on the SRTM data is re-sampling of 60m to fit the pixel resolution of 10m baseline used for all other images in the data analysis process.

5.2.6 Oil Spill Point and Incident data

The full pre-processing operations implemented on the oil spill incident data can be found in the overall methodology in Chapter 3 (section 3.2.3) of this thesis. The incident data from the two published sources were combined and large spill incidents covering areas of not less than 1000m^2 identified to ensure that the area covered is greater than a single pixel of 100m^2 of the Sentinel – 2 pixel of 10m spatial resolution. As such, the greater the spill size, the higher the number of sample points selected within the spill area (Table 5-2) and conversely, only a single sample point was selected for relatively small spill sites.

Furthermore, as the focus of the study is to distinguish between oil spill impacted and oil-free vegetation, it was necessary to select sample location of non-spill/oil-free sites sufficiently separated from the known spill sites (Siegel, 1974). To achieve this, a buffer ring of 600m was implemented around the spill areas to create the High Consequence Area (HCA) isolating all existing spill points. Outside the HCA, non – polluted sites were selected at random (Whanda *et al.*, 2015). The selected spill site and non – spill points were thereafter assigned to different vegetation types (cropland, grassland and tree cover areas) as provided by the ECCI dataset (http://2016africalandcover20m.esrin.esa.int/) and as given in Table 5-2. The processed points served two major purposes in this analysis.

First, they were used for the class-wise assignment of pixels in the fuzzy forest and random forest image classification operation. Here points were divided into two sets by a ratio of 60:40 for training and validation respectively. Secondly, the processed points were used in extracting requisite spectral and backscatter information from the images for further statistical analysis.

Table 5-2: Total number of sample points used for training and validation

Class Label	Total Number of Sites	Training	Validation
Non – Polluted Cropland	41	26	15
Non – Polluted Grassland	27	17	10
Non – Polluted Tree Cover Areas	25	15	10
Polluted Cropland	44	27	17
Polluted Grassland	26	16	10
Polluted Tree Cover Areas	26	16	10

5.2.7 Soil Map and Geologic data

Several studies have suggested that the dominant soil type and geology are capable of influencing oil seepage, which in part can rapidly influence fragile plant mortality as a result of crude oil pollution. The soil and geology data (see section 3.4.5) were first georeferenced using the map coordinates provided by Longitudes and Latitudes in the World Geodetic Survey (WGS) Datum and were later re-projected to the UTM projection. The study area extent was extracted from the entire map before intersecting layers were digitized. The vector map was further rasterized using the kriging interpolation technique.

The soil geology data comprised of two predominant soil types, Ferrosols and Fluvisols, while the geology comprised two predominant geologic layers of Coastal Plains Sands and Alluvium. The layers were all incorporated into the classification process as several studies (Abdel-Moghny *et al.*, 2012; Klamerus-Iwan *et al.*, 2015; Wang *et al.*, 2013) have suggested that soil type and geologic characteristics can to a large extent influence hydrocarbon crude seepage and runoff effect. This to a large extent determines the extent of the impact posed on vegetation.

5.3 Methodology

5.3.1 Evaluation of Discriminatory Potential of the Developed Variables

Backscatter from the various SAR sensors and the optical-derived vegetation health indices were tested for their potential to discriminate between oil-polluted and oil-free areas within each of the three vegetation types (cropland, grassland and tree cover areas). Box plots, corresponding median values and interquartile range for each variable for the polluted and oil-free reference sites were analyzed and tested for significant differences.

The paired sample t-test was used to compare pairwise differences in means between oil-free and oil-impacted vegetation as implemented in Khanna *et al.*, (2013). A critical value = 0.05 was used at 95% confidence to infer significant difference in the backscatter means between polluted and oil-free vegetation. The information content of the SAR backscatter characteristics independent of the optical variables was also assessed using linear regression of the SAR variables on the three vegetation indices for polluted and oil-free sites.

5.3.2 Image Classification and Experimental Scenarios

In order to investigate the functionality of the fuzzy forest methodology for image classification considering the variable reduction potential of the algorithm, the random forest image classification was implemented for comparison. A bilinear interpolation technique was used to resize and re-sample all 37 derived image variables (consisting of textural measures, interferometric coherence, normalized backscatter values, optical reflectance bands, vegetation indices, digital elevation model, aspect, slope and geophysical variables) into a uniform pixel size of 10m. The WGS 1984 UTM Zone 32N projection system was used for all input features to ensure that sufficient and reliable geometric alignment among all variables.

Two experimental image classification scenarios were implemented in this analysis to assess the potentials of multi-frequency SAR Image Fusion (MSIF) and a combination with optical variables i.e. ‘multi-frequency SAR Optical Image Fusion’ (MSOIF). The total number of variables for the MSIF and MSOIF were 29 and 37 respectively. The Fuzzy forest and Random forest classifiers were trained separately on each of the three vegetation types with the reference spill and oil-free sites before classification was implemented. Table 5-3 is list of variables used in the image classification process.

Table 5-3: List of SAR, Optical, Elevation and Geophysical Variables used for image classification. S1 (Sentinel 1), CSM (Cosmo Skymed), TDX (TanDEM X) and S2 (Sentinel 2)

S/No	Variables	S1	CSM	TDX	S2	S/No	Variables	S1	CSM	TDX	S2
1	Sigma Nought	v	v	v		15	SLOPE			v	
2	VV/VH	v				16	ASPECT			v	
3	VV+VH	v				17	Soil Map				
4	VV-VH	v				18	Soil Geology				
5	Texture-Variance		v	v		19	NDWI				v
6	Texture-Second Moment		v	v		20	LAI				v
7	Texture-Mean		v	v		21	NDVI				v
8	Texture-Homogeneity		v	v		22	WDVI				v
9	Texture-Entropy		v	v		23	B11				v
10	Texture-Dissimilarity		v	v		24	B12				v
11	Texture-Correlation		v	v		25	B5				v
12	Texture-Contrast		v	v		26	B6				v
13	Coherence			v		27	B7				v
14	DEM			v		28	B8A				v

5.3.3 Random Forest

In this chapter, RF image classification was implemented in the R software (TeamR, 2017) using the Caret package (Kuhn, 2012). Several calibration/parameterization runs were carried out to determine the optimal *ntree* and *mtry* values for training the random forest models in each of the 3 vegetation cover type, using all the image input variables. Results as experimented and as shown in Figure 5-2 indicates that the *ntree* = 500 and *mtry*=6 yielded the best calibration accuracy with the lowest out-of-bag error. It also showed that the calibration result for the tree cover area vegetation had the lowest out-of-bag-error of around 0.06% while grassland had the highest out-of-bag-error of around 0.2%.

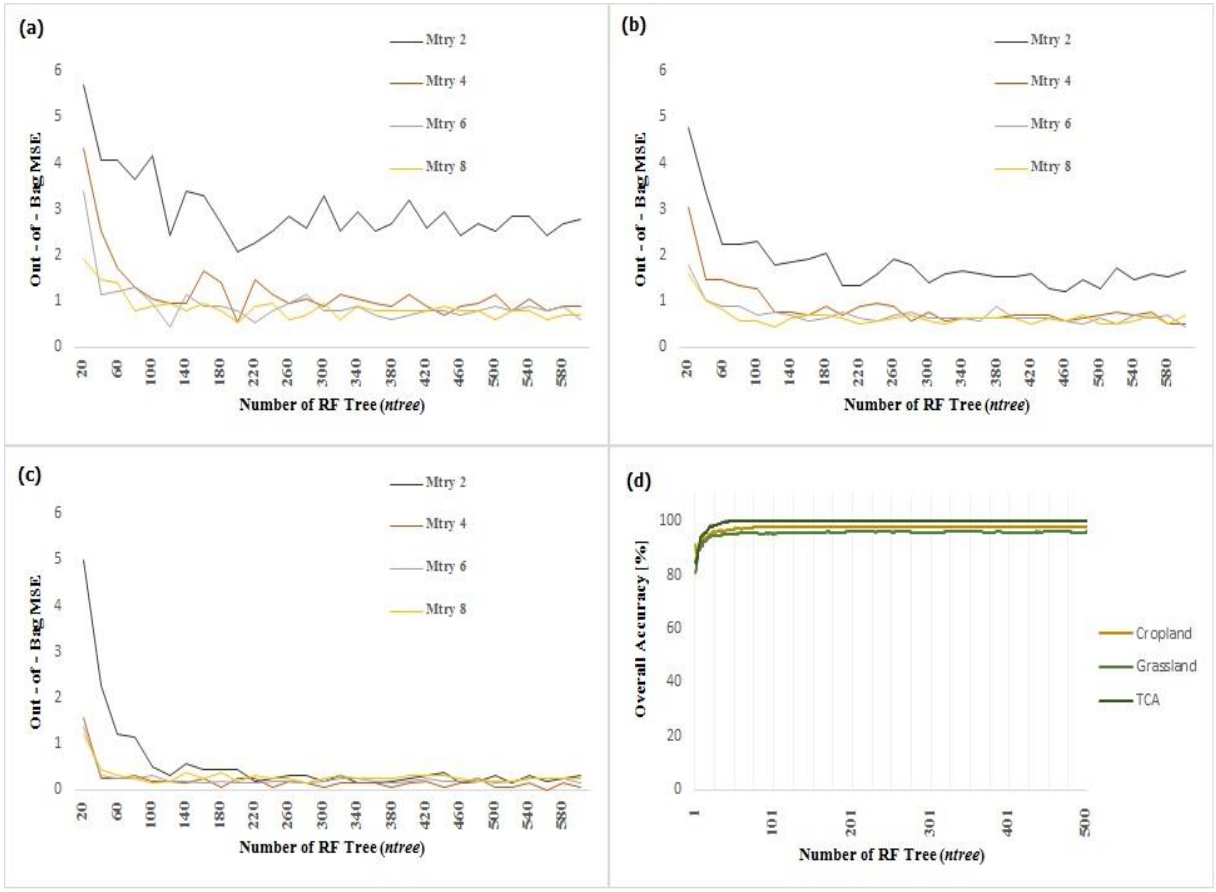


Figure 5-2: Out-of-bag accuracy (1 - OOB error) as a function of number of decision trees for the three vegetation class subsets. (a) Cropland (b) Grassland (c) TCA and; (d) Overall Calibration Accuracy

5.3.4 Fuzzy Forest

In this chapter, FF image classification was also implemented in the R Studio package version 1.0.143 using the fuzzy forest and WGCNA packages. Several calibration/parameterization runs were also carried out to determine the optimal *model parameters*. Similar results with the RF method was obtained, in that, *ntree* value of 500 and square root of the number of input variables for *mtry* produced the best calibration accuracy.

However, several other parameters such as the drop fraction, keep fraction and the number selected must be specified prior to the implementation of fuzzy forest classification. The drop fraction refers to the number of features (mostly weak and correlated input variables) to be dropped after each iteration of the recursive feature elimination – random forest. While the keep fraction refers to the percentage of the original input features to be retained at the end of the process, which must also logically corresponds to the fraction of the 'number selected' for the final classification. All input parameters must be specified by the user at the beginning of the model calibration phase of implementation.

For the purpose of this study, screening parameters were set as: *ntree* = 500; *drop_fraction* = 0.5; *keep_fraction* = 0.5; *number_selected* = 5. These were used in the final classification to ensure that only 10% of the original variables, which corresponds to the most important, uncorrelated and unbiased 5 variables are kept for the final fuzzy forest classification.

Five variables were chosen as the most preferred choice of the final number of variables to be used as it offered an optimal and balanced representation of the input variables used for effective comparison with the top 5 random forest variables and variable importance information. An experimentation of 75%, 50% and 25% of the original input variables was carried out, however, the final variable importance chart didn't give better avenue for robust comparison. Hence 10% yielded a much simplistic value for baseline comparison with the random forest method.

5.3.5 Confusion Matrix

The results produced from the RF and FF classification were assessed using the error matrix (Congalton, 1991) produced with the remaining 40 percent oil-free and oil-polluted ground reference data, resulting in 180 pixels per vegetation type. Hence, selected pixels representing actual classes from classification result were compared to the ground truth reference class as determined in (3.2.3). The validation process evaluated whether the True Positive Sites known as oil spill sites were correctly classified as oil polluted (oil-impacted) vegetation and if the

known non-polluted (oil-free) sites were also positively classified as oil-free vegetation. The performances of the two methods were further assessed using the overall accuracy (OA), producer's accuracy (PA) and user's accuracy (UA).

In addition, whether the FF or RF classification provided better results was evaluated using McNemars's test (de Leeuw *et al.*, 2006). This has been applied in several studies (Onojeghuo *et al.*, 2018a; Son *et al.*, 2018; Whyte *et al.*, 2018). McNemars's test is a nonparametric test based on χ^2 using a 2 *X* 2 contingency matrix to assess the level of performance of multiple classifier outputs based on the number of correctly predicted samples. The accuracies were considered as statistically significant at a confidence level of 95% if the calculated χ^2 (from equation 13) score is larger than the critical value of 1.5. The samples are labelled as f_{12} and f_{21} which represents the correctly classified samples for FF that were mis-classified by RF, and the number of correctly classified samples for RF that were mis-classified by FF, respectively (Whyte *et al.*, 2018).

$$X^2 = \frac{(f_{12}-f_{21})^2}{f_{12}+f_{21}} \qquad (13)$$

5.3.6 Field and Qualitative Validation

A qualitative validation to assess prediction performance was also undertaken using high-resolution google earth image. The spatial extent of classified image as determined by the different classification processes to the known oil spill extent was visually evaluated and compared between the different methods. In addition, field validation data were collected during a post-spill fieldwork assessment carried out in October 2018 (3.2.8) in some of the oil spill sites. It formed the basis of a toxicology analysis carried out during the fieldwork to establish the volume of hydrocarbon content present in the soil. The toxicology analysis of soil sediment sample from three spill site locations and one oil-free site location are displayed on the high-resolution image for comparison purpose. The toxicology analysis tested for the Total Hydrocarbon Content (THC) levels within the respective location and this was compared visually to the result of the classified map from the two classifiers.

5.4 Results

5.4.1 Vegetation Health Response from Different Derived Sentinel – 2 Indices

In general, result of the retrieved vegetation indices for polluted and oil-free vegetation types showed that the values are normally distributed with the exception of the NDWI values for the polluted TCA which were slightly skewed. Pairwise comparison using t-test was used to compare the means of polluted and oil-free health indices retrieved from the various vegetation types. For cropland, the vegetation indices LAI, NDWI and NDVI tended to be higher for unpolluted cropland than for polluted cropland, with significantly different means for LAI, NDVI and NDWI (P<0.05) as obtained from the paired sample t-test (Figure 5-3). The range of the indices was smaller (median < 0.3 and smaller interquartile range) for oil-free cropland compared to the large median and interquartile range for the polluted cropland (median > 0.3). For grassland areas, the results indicate significant differences between the means for oil-impacted and oil-free grassland (P<0.05 for all three indices), as the retrieved LAI and NDVI for oil-free grassland often had a higher median and interquartile range. This indicates a larger heterogeneity of the unpolluted sites. With respect to TCA, the results shows that all three indices retrieved and explored were more dissimilar and heterogeneous, as P<0.005 and the median for oil-free TCA were higher than those retrieved from polluted TCA.

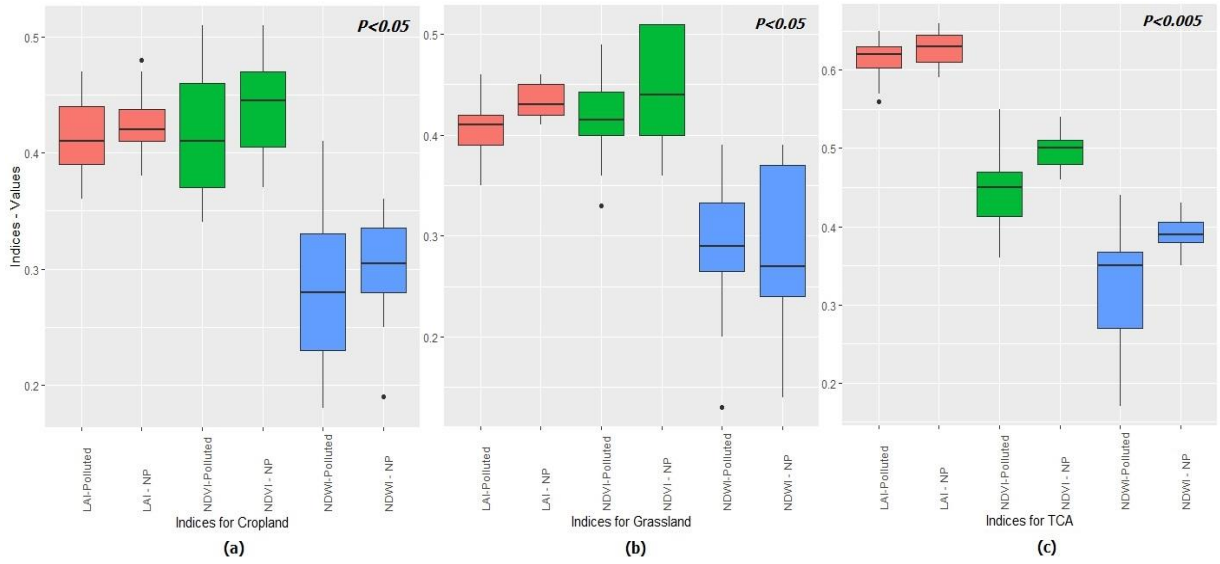
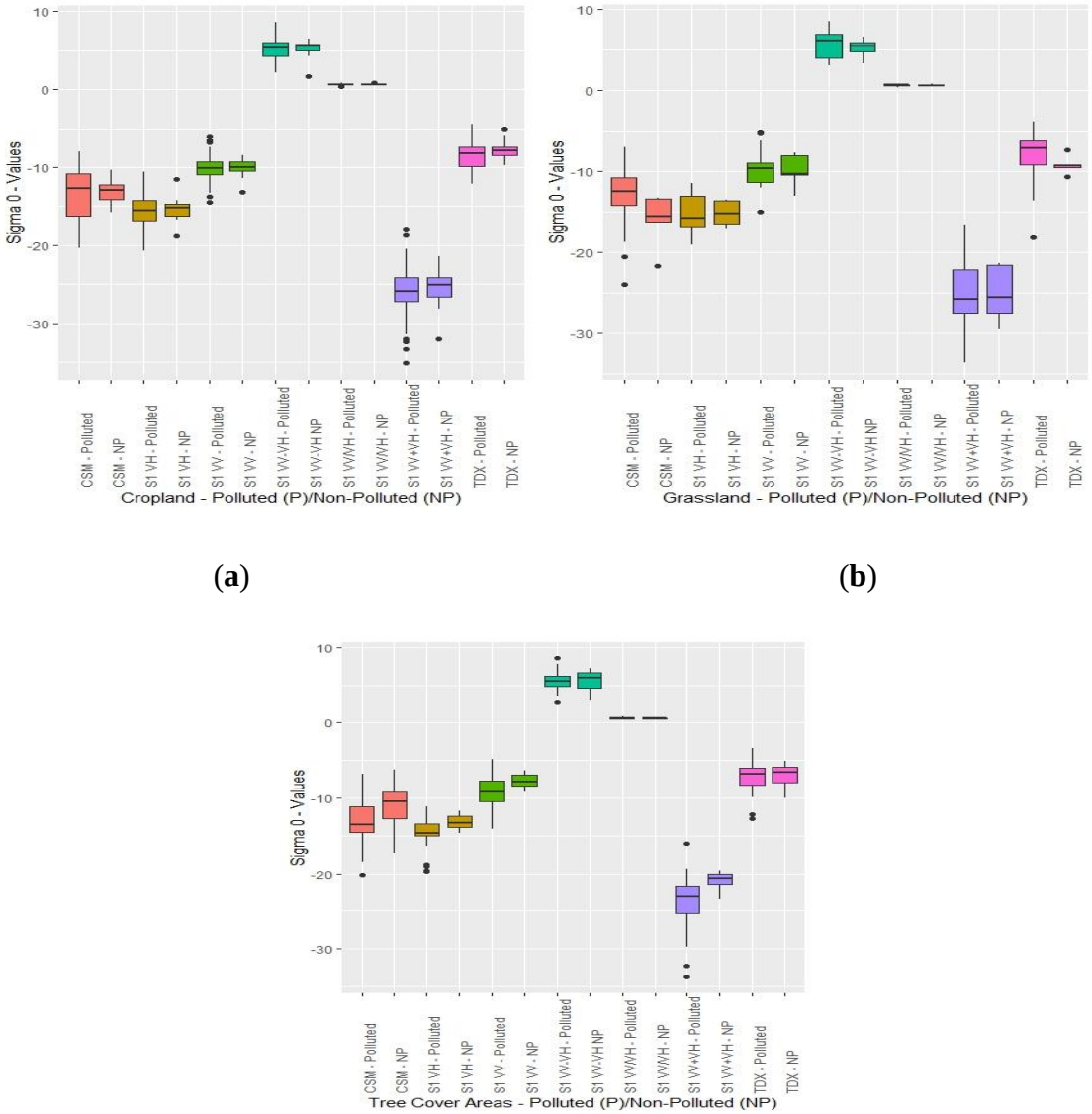


Figure 5-3: Distribution of retrieved NDWI (blue), LAI (pink) and NDVI (green) for both Polluted and oil-free land cover types in the study area (a) cropland (b) grassland (c) forest. Results show that median values of indices for oil-free land cover types are mostly significantly higher than the polluted land cover.

5.4.2 SAR C- and X-band Backscatter for Detecting Hydrocarbon Pollution

Figure 5-4 shows boxplots of retrieved backscatter values for different classes. Backscatter values for both polluted and oil-free vegetation types had similar distribution characteristics. It was observed that backscatter values from the cross polarization ratio of sentinel-1 VV/VH had low variability across the three vegetation types investigated. To this end, pairwise comparison of polluted and oil-free backscatter values was implemented for the various vegetation types.



(**a**) (**b**)

(**c**)

Figure 5-4: The distribution of TDX Backscatter, TDX coherence, Cosmo Skymed and Sentinel 1 VV – VH backscatter for polluted and oil-free land cover types in the study area. (a) Cropland (b) Grassland (c) TCA. Result shows that median backscatter values and interquartile range in oil-free TCA are significantly higher than the polluted TCA.

The results indicate that backscatter values from unpolluted cropland often had lower interquartile ranges with median values > -35dB, compared to the polluted cropland, which often had larger interquartile spread and median values > -35dB. The significant variations were observed more with the Sentinel-1 derived polarizations and cross-polarization ratios. A similar trend was also observed for grassland areas as retrieved backscatter values for unpolluted grassland had much lower variability and lower median backscatter. Variations were also associated more with the Sentinel-1 cross-polarization ratios and the TanDEM-X data. However, these results were not statistically significant (P>0.05) for cropland and grassland.

In TCA however, the results generally showed that backscatter values retrieved from the unpolluted sites had higher medians of -10dB, -17dB, -8dB, 7dB and -19dB from the CSM, S1 VH, S1 VV, S1 VV–VH and S1 VV+VH respectively. In contrast, the polluted TCA had median values of -13dB, -14dB, -9dB, 6dB and -17dB from the CSM, S1 VH, S1 VV, S1 VV – VH and S1 VV/VH respectively. Interquartile range of backscatter between the oil-free (unpolluted TCA) and polluted TCA were uniform across the different sensors. The paired sample t-test result showed that the difference between means was statistically significant (P < 0.05). These were mostly obtained with S1 derived products of S1 VV (P=0.0006), S1 VV + VH (P=0.0008) and S1 VH (P=0.0229).

5.4.3 Relationship Between and Among Various Biophysical Variables

For croplands, there is generally a poor association between the SAR variables and the optical derived LAI indices as indicated by the results of least square regressions (Table 5-4). The NDWI was observed to have a strong association of R >0.4 (R^2 >0.17) or above with the Sentinel 1 VV, VH, and VV + VH derived backscatter (P<0.05). TanDEM-X (TDX) backscatter had R = 0.3 (R^2 = 0.13) with the NDWI and this was statistically significant. LAI did not show strong association with backscatter, as R and R^2 values were much lower and the corresponding P values were much higher than the critical value.

However, the result for NDVI showed a rather strong linear association with CSM, S1 VV, VH + VH and VV-VH. R and R^2 values were above 0.3 and 0.17 respectively, and the result was statistically significant (as P<0.05).

For grassland, there is generally a strong relationship between backscatter and plant health indices (Table 5-5). The NDWI and LAI were observed to have the strongest association with the S1 Variables. R^2 values obtained for NDWI with S1 VV, S1 VH, S1 VV+VH, S1 VV/VH, were 0.38, 0.30, 0.39 and 0.22 respectively. R^2 values between LAI and S1 VV, S1 VH, VV + VH, VV/VH, TDX Coherence, were 0.12, 0.22, 0.21, 0.25, 0.10 respectively. The results obtained were statistically significant ($P < 0.05$).

The result of linear regressions for TCA vegetation is shown in Table 5-6. This indicates that the NDWI and LAI had the strongest association with the various backscatter for this vegetation type. High R^2 values obtained for NDWI were with TDX coherence, S1 VV, S1 VH, S1 VV+VH, S1 VV-VH and S1 VV/VH, and this gave R^2 values of 0.23, 0.53, 0.21, 0.39, 0.19 and 0.49 respectively. Similarly, high R^2 values recorded for LAI were with S1 TDX Coherence, S1 VV, S1 VV + VH, S1 VV – VH and VV/VH. Values obtained are 0.25, 0.32, 0.21, 0.20 and 0.33 respectively. The results obtained were also statistically significant, as P values were in most cases below 0.005.

Table 5-4: Linear regressions result of C and X band SAR backscatter retrieved from the polluted cropland sites against the optical derived health indices. Results with statistical significance (P <0.05) regression coefficient (R) are indicated in green colour.

	NDWI				LAI				NDVI			
Variable	R	R^2	P Value	Std. Error	R	R^2	P Value	Std. Error	R	R^2	P Value	Std. Error
TDX	0.36	0.13	0.004	0.063	0.26	0.07	0.037	0.112	0.24	0.01	0.059	0.031
Coherence	0.28	0.08	0.027	0.065	0.1	0.01	0.441	0.116	0.06	0.00	0.595	0.032
CSM	0.28	0.08	0.026	0.065	0.10	0.01	0.433	0.116	0.32	0.10	0.011	0.03
S1 VV	0.40	0.17	0.001	0.062	0.04	0.00	0.714	0.117	0.20	0.04	0.112	0.031
S1 VH	0.44	0.20	0.000	0.06	0.07	0.01	0.575	0.116	0.43	0.20	0.001	0.029
S1 VV+VH	0.45	0.21	0.000	0.06	0.06	0.00	0.616	0.116	0.34	0.01	0.006	0.03
S1 VV-VH	0.16	0.03	0.214	0.067	0.05	0.00	0.700	0.117	0.39	0.16	0.002	0.029
S1 VV/VH	0.08	0.01	0.501	0.067	0.03	0.00	0.786	0.117	0.21	0.05	0.094	0.031

Table 5-5: Linear regressions result of C and X band SAR backscatter retrieved from the polluted grassland sites against the optical derived health indices. Results with statistical significance (P <0.05) regression coefficient (R) are indicated in green colour.

	NDWI				LAI				NDVI			
Model	R	R^2	P Value	Std. Error	R	R^2	P Value	Std. Error	R	R^2	P Value	Std. Error
TDX	0.13	0.02	0.536	0.062	0.00	0.00	0.569	0.039	0.122	0.15	0.569	0.026
Coherence	0.23	0.05	0.286	0.061	0.35	0.12	0.087	0.037	0.206	0.04	0.334	0.025
CSM	0.17	0.03	0.439	0.062	0.16	0.03	0.433	0.039	0.043	0.00	0.840	0.026
S1 VV	0.62	0.38	0.001	0.049	0.46	0.22	0.021	0.035	0.107	0.01	0.619	0.026
S1 VH	0.55	0.30	0.006	0.053	0.46	0.21	0.023	0.035	0.055	0.00	0.799	0.026
S1 VV+VH	0.62	0.39	0.001	0.049	0.49	0.25	0.13	0.034	0.027	0.00	0.899	0.026
S1 VV-VH	0.09	0.01	0.666	0.063	0.00	0.00	0.998	0.039	0.226	0.05	0.289	0.025
S1 VV/VH	0.44	0.20	0.032	0.056	0.31	0.10	0.134	0.037	0.181	0.03	0.398	0.025

Table 5-6: Linear regressions result of C and X band SAR backscatter retrieved from the polluted TCA sites against the optical derived health indices investigated in the study. Results with statistical significance (P <0.05) regression coefficient (R) are indicated in green

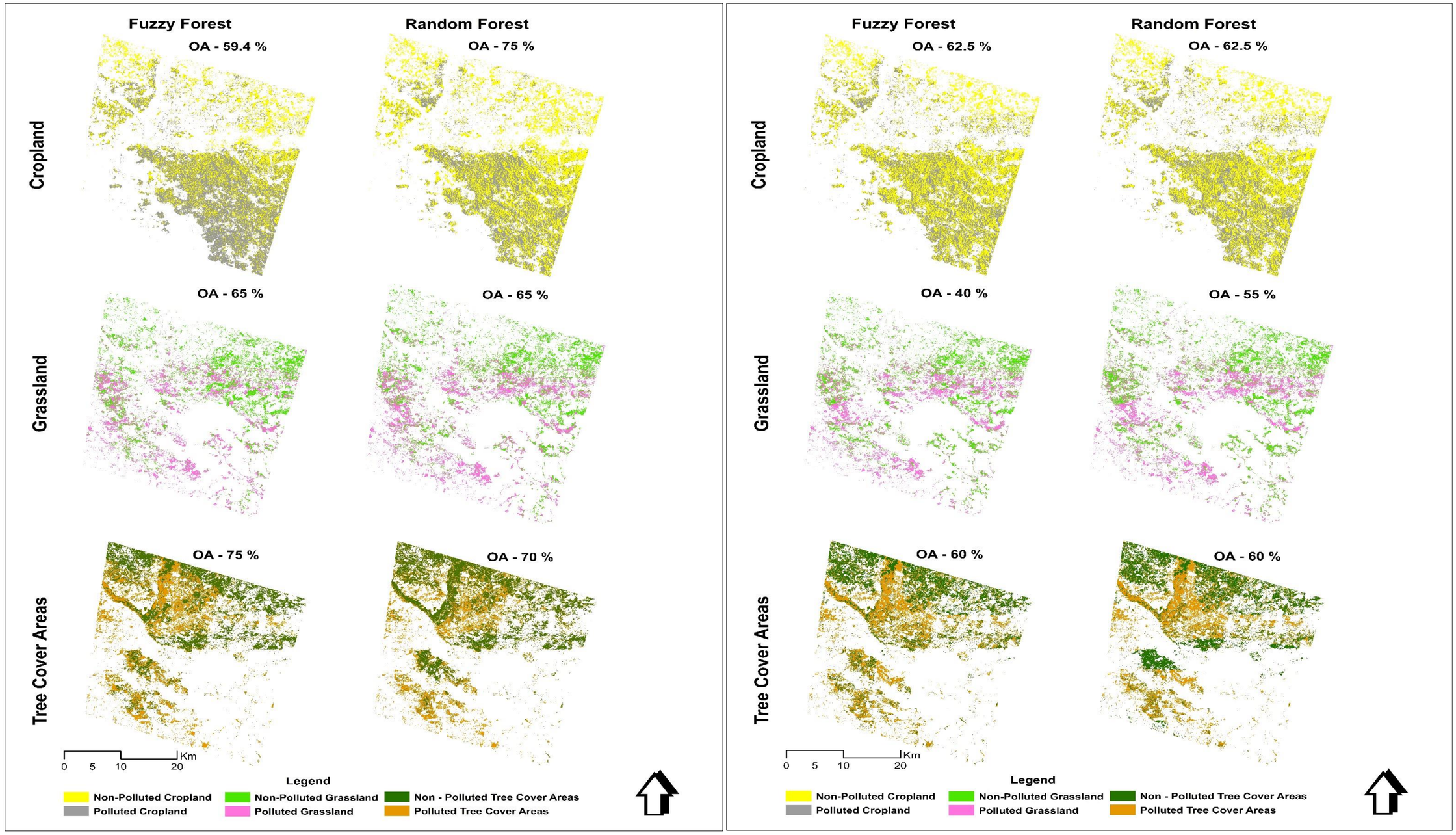
	NDWI				LAI				NDVI			
Model	R	R^2	P Value	Std. Error	R	R^2	P Value	Std. Error	R	R^2	P Value	Std. Error
TDX	0.03	0.00	0.881	0.073	0.01	0.00	0.932	0.046	0.02	0.00	0.917	0.024
Coherence	0.48	0.23	0.007	0.064	0.50	0.25	0.005	0.04	0.102	0.01	0.590	0.024
CSM	0.22	0.04	0.247	0.071	0.24	0.10	0.185	0.045	0.037	0.00	0.844	0.024
S1 VV	0.73	0.53	0.000	0.05	0.56	0.32	0.001	0.038	0.143	0.02	0.451	0.024
S1 VH	0.45	0.21	0.12	0.065	0.30	0.01	0.099	0.041	0.115	0.01	0.546	0.024
S1 VV+VH	0.62	0.39	0.000	0.057	0.46	0.21	0.010	0.041	0.136	0.02	0.473	0.024
S1 VV-VH	0.44	0.19	0.015	0.066	0.415	0.20	0.023	0.042	0.047	0.00	0.806	0.024
S1 VV/VH	0.7	0.49	0.000	0.052	0.575	0.33	0.001	0.038	0.142	0.02	0.455	0.024

5.4.4 Classifying and Mapping Oil Polluted Sites

Figure 5-5 shows the classification results. The best overall accuracies were obtained when the FF and RF methods were used to classify the multi-frequency SAR, optical image, DEM, Soil and Geology fusion (MSOIF) for tree cover areas and cropland areas (Figure 5-5). Generally, overall accuracies presented slight differences in the output from the FF and RF; however, MSOIF yielded about 10% higher overall accuracy than the multi-frequency SAR image fusion (MSIF) only. This implied that the exclusion of the optical variables in the second classification scenario increased inter-class errors, thereby reducing the overall accuracy.

A visual assessment of the outputs showed that the spatial extent of oil-polluted cropland within the cropland vegetation subset from the MSOIF using FF was larger than the extent from RF, which could have led to lower classification accuracy. The results for the TCA using the RF with the MSOIF dataset also had large oil-impacted class especially within the central parts of the study area. This also must have accounted for the lower classification accuracy, compared to the results obtained from the FF, which had smaller segments of polluted areas. However, the results obtained for the grassland areas did not show as much contrast and dissimilarity as in the cropland and TCA classes.

A similar overall accuracy was obtained when FF and RF were used to classify the MSIF variables for cropland and TCA areas (60% for *FF* and 60% for *RF*). The RF outperformed the FF when the same data were used to classify grassland. Overall accuracies of 55% and 40% were obtained for RF and FF respectively.



(a) (b)

Figure 5-5: Classified maps of polluted and oil-free areas for the three vegetation types. (a) Using multi-frequency SAR Image Variables, Sentinel 2 Spectral Bands and Vegetation Health Indices (MSOIF). (b) Using only multi-frequency SAR Image Variables (MSIF). Results show that FF had highest performance (OA 75%) in TCA using the MSOIF variables, while RF had highest performance (75%) in Cropland Using MSOIF variables. Note that crop cover is more dominant in the south west of the study area and tree cover more in the north (Figure 3-9).

5.4.5 Multi-frequency SAR – Optical Image Fusion (MSOIF) Variable Importance

The result of the variable importance from the RF and FF using the MSOIF dataset is presented in Figure 5-6. Elevation-derived variables including the DEM were the most important variables. For cropland vegetation classification, the Red Edge (band 7), Aspect and NDVI were the most important variables in the discrimination of oil-free and oil-impacted cropland areas when FF method was used for classification. In contrast, for RF, slope, aspect and the DEM were the three most important variables. The result implied that more diverse features were selected for the FF than the RF in the classification of cropland vegetation.

For grassland areas, it was observed that the DEM, SWIR and Red bands were the most important explanatory variables in the fuzzy forest classification. However, the result from the RF classification showed that the three elevation variables (Aspect, DEM and Slope) and SWIR bands had higher importance. The result obtained from the classification of tree cover vegetation showed that NDVI, DEM, SWIR and the Red Edge (band 7) were the most important variables in the FF classification. Only the DEM and Red Edge bands featured in the Top 5 variables when the RF method was used for classification.

This exploration of the variable importance highlights that the use of the reduced variables, which were free of high dimensionality, in the FF classification yielded comparable or improved performance in the discrimination of oil-polluted and oil-free grassland and TCA compared to RF (Figure 5-6). This further indicates that the FF was able to optimize the *n* input variables to select the most important uncorrelated variables for the classification of the MSOIF data. However, the reduced variables for cropland did not necessarily translate into better classification accuracy.

5.4.6 Multi-frequency SAR Image Fusion (MSIF) Variable Importance

The results of the variable importance obtained from the RF and FF classifications from the MSIF variables are shown in Figure 5-7. Elevation variables had greater importance than any other variables used in the classifications. In cropland areas, the results of the RF showed that the 5 most important variables were mostly derived from the elevation model, including slope, aspect and DEM, and textural variables.

However, the variable importance chart from the FF showed that both elevation-derived variables and Sentinel-1 backscatter data were the variables selected by the FF classifier in the classification process. This is partly different for grassland areas, as the 5 most important variables obtained from the RF classification were mostly the S1 and elevation-derived variables. On the contrary, important variables obtained from the FF classifier were more diverse, as all three elevation-derived variables, the soil map and Sentinel-1 data were used for the classification of grassland areas.

Results for TCA areas also followed a similar trend as the cropland areas, as the most important variables in the FF classification were Sentinel-1 and elevation variables. In contrast, S1, elevation and soil type variables were the top 5 variables in the RF classification. It must be mentioned that the result of the classification using only the MSIF variables in the FF performed comparatively with RF, especially in cropland and TCA, as the best overall classification accuracies of 62.5% and 60% respectively were obtained. However, RF outperformed FF in the grassland areas, overall accuracy assessment results were 55% and 40% respectively.

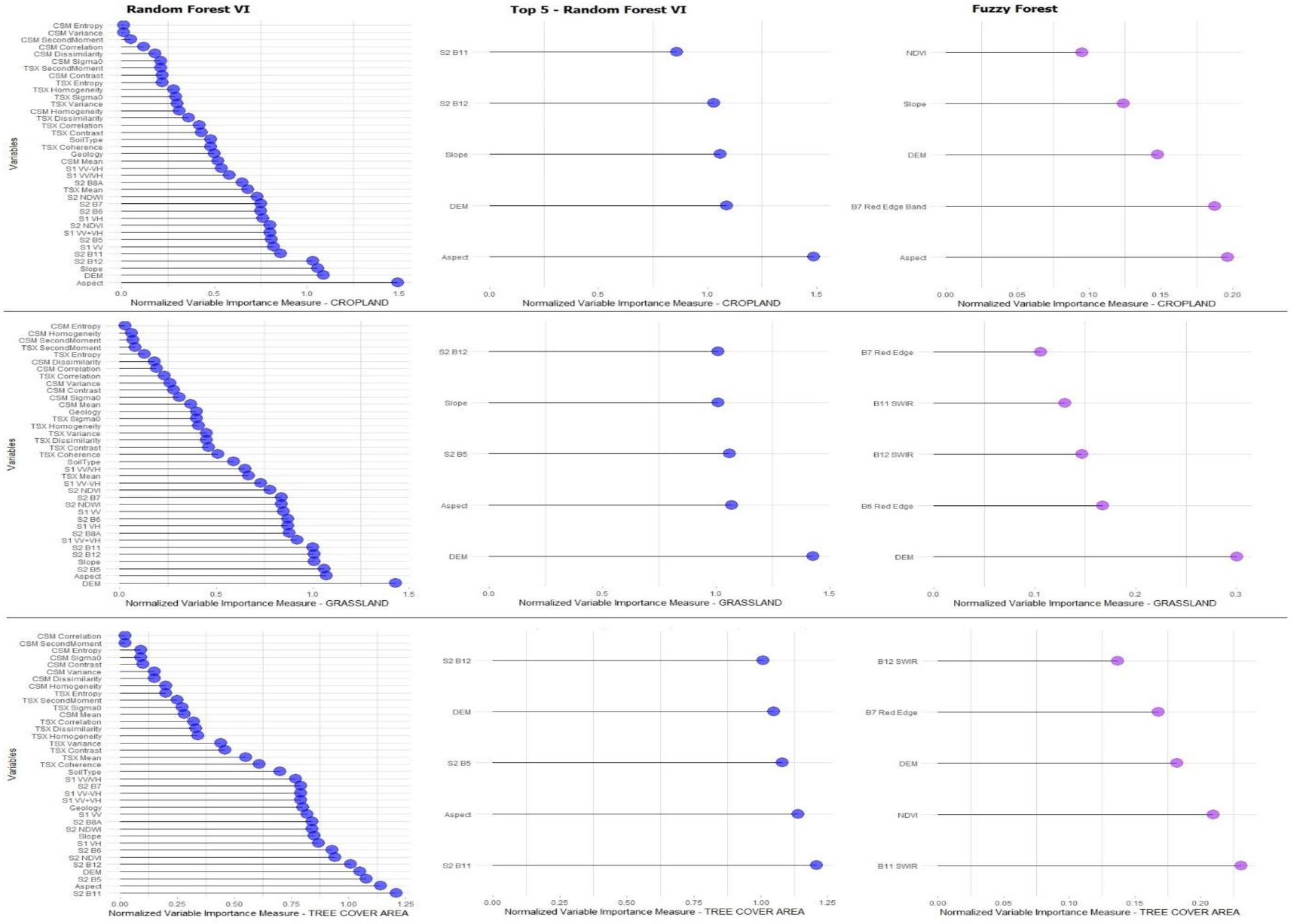


Figure 5-6: Variable Importance (VI) plot from the Fuzzy forest and Random forest classification of the multi-frequency SAR and Optical Image Fusion (MSOIF) derived variables. This shows that Aspect, DEM and SWIR have been consistently identified by the *FF* and *RF* classifiers as the most important variables for discriminating polluted Cropland, Grassland and TCA from their respective oil-free cover types.

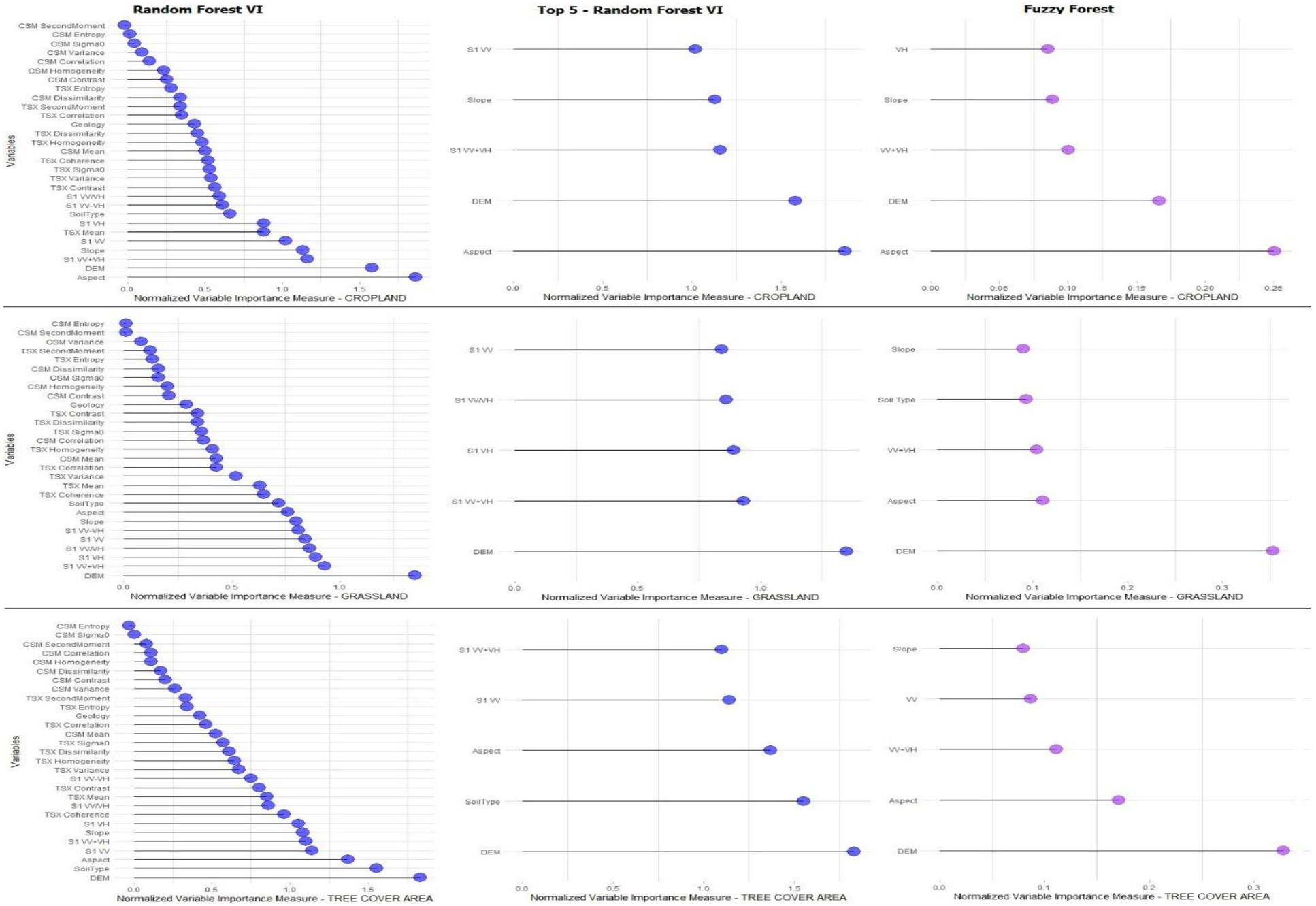


Figure 5-7: Variable Importance (VI) plot from the Fuzzy forest and Random forest classification of the multi-frequency SAR Image Fused (MSIF) derived variables. This shows that Aspect, DEM and DEM have been consistently identified by the *FF* and *RF* classifiers as the most important variables for discriminating polluted Cropland, Grassland and TCA from their respective oil-free cover types.

5.4.7 Accuracy Assessment

Table 5-7 and 5-8 presents the overall accuracy (OA), user accuracy (UA) and producer's accuracy (PA) (Congalton, 1991) for cropland, grassland and TCA classification using the RF and FF methodologies. The OA often measures the accuracy of a map with respect to its ability to correctly identify the ground reference points for all classes utilized, while PA and UA measures individual map class accuracy in terms of its ability to correctly identify class specific ground reference points.

This concomitantly accounts for the error of omission i.e. as a result of the classified map not correctly identifying a ground reference (Producer Accuracy) and error of commission i.e. as a result of wrong assignment of a map class in a classified image (User Accuracy). The results in this study showed that the cropland and TCA classified maps with RF and FF respectively, using the MSOIF dataset, gave the highest overall classification accuracy of 75% each. It is clear from the results (in Table 5-7 and Table 5-8) that RF had better prediction of the ground reference site for polluted cropland as user accuracy was above 75%, while FF also had superior performance in polluted TCA prediction as user accuracy also reached 100%.

Table 5-7: OA, UA and PA of the MSOIF data with Fuzzy forest and Random forest

		Cropland		
		Producer's Accuracy [%]	User's Accuracy [%]	Overall Accuracy [%]
FF	Oil-free Cropland	56.25	60	59.4
	Polluted Cropland	62.5	41.2	
RF	Oil-free Cropland	73.33	73.33	75
	Polluted Cropland	76.47	76.47	
		Grassland		
		Producer's Accuracy [%]	User's Accuracy [%]	Overall Accuracy [%]
FF	Oil-free Cropland	63.63	70	65
	Polluted Cropland	66.66	60	
RF	Oil-free Cropland	70	63.6	65
	Polluted Cropland	60	66.7	
		Tree Cover Areas		
		Producer's Accuracy [%]	User's Accuracy [%]	Overall Accuracy [%]
FF	Oil-free TCA	100	50	75
	Polluted TCA	66.67	100	
RF	Oil-free TCA	70	70	70
	Polluted TCA	70	70	

Table 5-8: OA, UA and PA of the MSIF data with Fuzzy forest and Random forest

		Cropland		
		Producer's Accuracy [%]	User's Accuracy [%]	Overall Accuracy [%]
FF	Oil-free Cropland	58.82	66.66	62.5
	Polluted Cropland	66.66	58.82	
RF	Oil-free Cropland	66.67	58.82	62.5
	Polluted Cropland	58.82	66.67	
		Grassland		
		Producer's Accuracy [%]	User's Accuracy [%]	Overall Accuracy [%]
FF	Oil-free Grassland	40	40	40
	Polluted Grassland	40	40	
RF	Oil-free Grassland	50	55.56	55
	Polluted Grassland	60	54.55	
		Tree Cover Areas		
		Producers Accuracy	Users Accuracy	Overall Accuracy [%]
FF	Oil-free TCA	75	30	60
	Polluted TCA	56.25	90	
RF	Oil-free TCA	40	66.67	60
	Polluted TCA	80	57.14	

The result of McNemar's test (de Leeuw *et al.*, 2006) presented in Table 5-9 shows that RF outperformed FF in cropland areas, as the difference between the errors was significant (χ^2-test $P<0.05$), compared to the low 0.2 χ^2 – value ($P>0.05$) obtained when the MSIF data were classified. The result of McNemar's test for grassland show no significant difference between the errors in the FF and RF classifications (overall classification accuracies of 65%). However, of the two classes investigated, the oil-free grassland had higher UA=70% when FF was used to classify the MSOIF. McNemar's test also showed that there is no significant difference between the oil-free and oil-impacted grassland areas for the MSOIF and MSIF ($P>0.05$).

Results for tree cover areas showed that FF outperformed *RF* in the discrimination of oil-free and oil-polluted TCA vegetation. OA was 75% and 70% for FF and RF respectively when the MSOIF data were classified. McNemar's test showed that a significant difference between the OA values ($P<0.05$). In addition, a low χ^2 of 0.3 was also obtained when the MSIF data for TCA was classified. An overall accuracy of 60% was recorded by the two classifiers, explaining why there is no significant difference between them ($P>0.05$).

Table 5-9: McNemar's test used to test the performance of Fuzzy forest and Random forest. Number in bold represents classification scenario with statistical significant difference in overall accuracy at 95% confidence

Cropland		RF	
		MSOIF	MSIF
FF	MOIF	**4.17**	
	MSIF		0.2

Grassland		RF	
		MSOIF	MSIF
FF	MSOIF	0.2	
	MSIF		0.57

TCA		RF	
		MOIF	MSIF
FF	MOIF	**1.62**	
	MSIF		0.3

5.4.8 Field and Qualitative Validation

The results of the field and qualitative validation are presented in Figure 5-8. The results of the lab test showed that spill sites visited in Eleme, TAI and Gokana had THC values of 641 mg/kg, 620 mg/kg and 605 mg/kg respectively. The areas around the oil spill locations were correctly classified as true positive sites of oil-impacted cropland by both the FF and RF Classifiers. The non-spill site visited in Etche had a THC volume of 548 mg/kg, which is slightly lesser than the THC for the spill sites. Very-high-resolution google earth imagery in a true color composite over the polluted sites was used to evaluate the extent classified as oil-polluted by the two classifiers. The spatial extent of oil-polluted sites obtained from FF were broader and captured the extensive vegetated areas impacted by the oil spill extent. While the extent of the RF on the other side was also able to capture the extent of the spill site but with less coverage as captured by the high spatial resolution imagery.

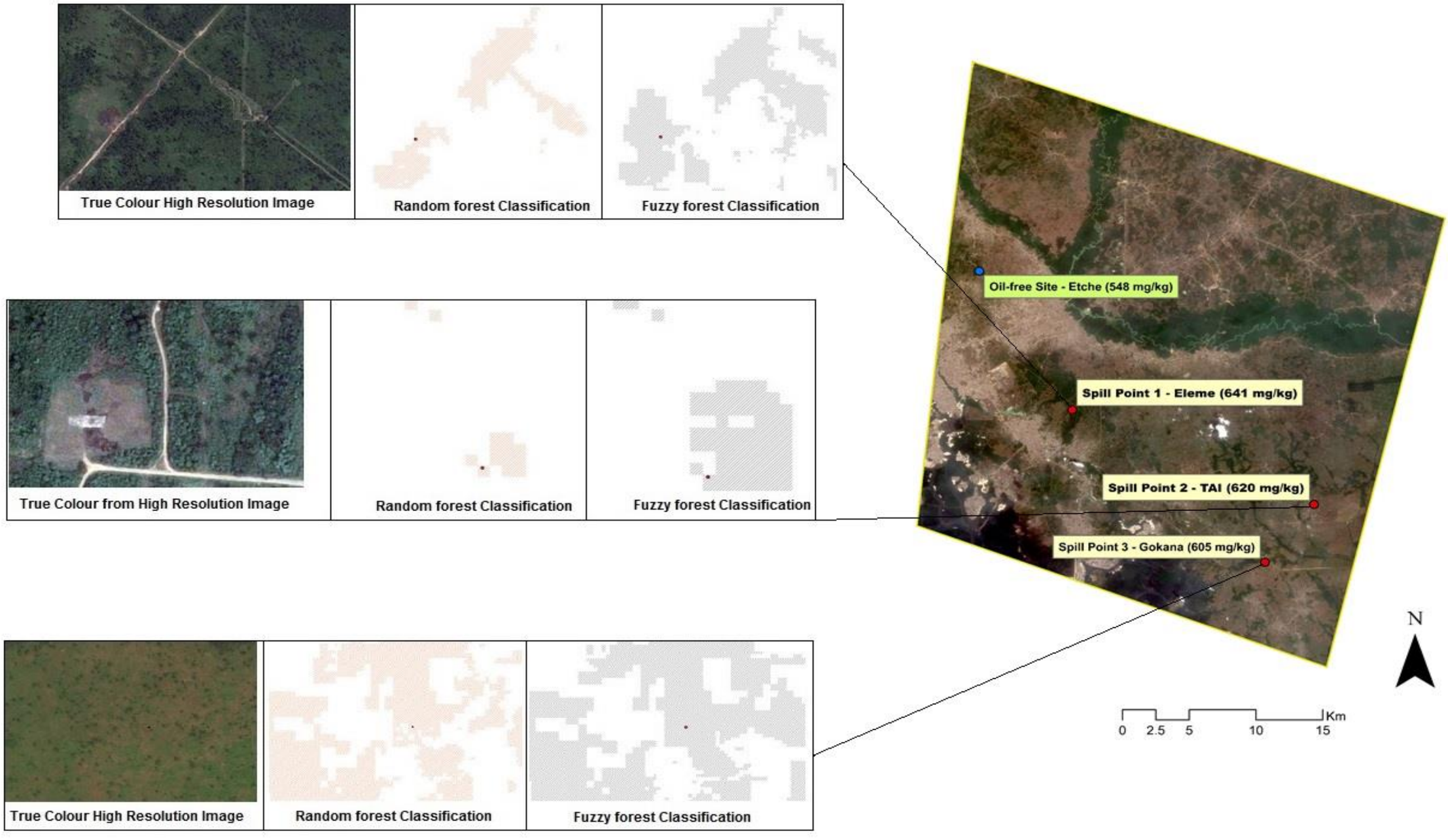


Figure 5-8: Spill points and Total Hydrocarbon Content (THC) on a true colour composite Sentinel 2A image of December 2016 (Source: https://scihub.copernicus.eu/dhus/#/home) of the study area. Results show that THC around the spill points is much higher than at the unpolluted site. FF represented the polluted sites better than RF.

5.5 Discussion

This chapter aimed at mapping and categorizing cropland, grassland and TCA vegetation into distinct class of oil-polluted and oil-free vegetation cover using training samples of oil spill incident locations and oil-free sites. Results obtained from Sentinel-2 vegetation indices (NDWI, NDVI and LAI) for detecting the effects of hydrocarbon crude oil largely showed that oil-free vegetation types were healthier due to the higher vegetation health indices values (with respect to median and interquartile range) obtained, compared to the low values retrieved from the oil spill sites. This is an indication that vegetation in the oil-polluted sites could be undergoing stress due to the presence of oil. Fundamental characteristics of typical oil-polluted vegetation are immediate alteration of the bio-physical and bio-chemical properties leading to reduction of plant growth and active function causing collapse of mesophyll cells, cuticle fragmentation and the disintegration of chloroplasts and nuclei (Samat *et al.*, 2018). This often results in changes to plant functional traits and characteristics such as biomass, leaf area index (LAI), water content and chlorophyll concentration (Douaoui *et al.*, 2006; Kozlowski, 1997; Leone *et al.*, 2007; Samat *et al.*, 2018).

The result obtained in this study have shown that NDVI and NDWI are effective indicators of plant stress. However, in this chapter the potentials of LAI was further investigated to depict the relationship between plant stress as a result of hydrocarbon crude oil pollution and the corresponding SAR backscatter signals. The results obtained here are also in keeping with previous studies in this regard, as LAI also showed significant levels of heterogeneity between the oil-impacted and oil-free vegetation types, especially in grassland and TCA.

Investigation of the potential of backscatter from the various SAR sensors for detecting and discriminating oil-free and oil-impacted vegetation yielded comparable result with the Sentinel-2 indices especially with the C-SAR variables. The distribution of backscatter values for the oil-polluted vegetation class usually had higher variation in backscatter, compared to oil-free vegetation for both X band (Cosmo – Skymed and TanDEM-X) SAR and the C band (Sentinel – 1) SAR. Initial assessment of the cropland and grassland vegetation showed that both C band (Sentinel – 1) and X band (Cosmo – Skymed and TanDEM-X) SAR backscatter had

higher variance and range for the oil-polluted cropland and grassland areas, while the corresponding oil-free vegetation had smaller variability. A possible reason for this is that the leaves and branches within healthy cultivated farmland and grassland fields are wider and broader in nature (Baudena *et al.*, 2015) than those of stressed fields. This can ultimately lead to stable scattering in the former, accounting for the narrow interquartile range of backscatter values when compared to the unstable backscatter trend and concomitant large interquartile range within the polluted fields (Pampaloni and Paloscia, 1985).

With respect to TCA, the result showed that Sentinel-1 single-polarization (i.e. VV and VH), co-polarization (VV) and cross polarization (VH) bands gave a more distinct difference between higher and lower backscatter distribution for oil-free and oil-polluted areas respectively. Cosmo Skymed (CSM) backscatter also showed a similar trend, reinforcing the notion that oil-free woody vegetation stands often presents higher backscatter, represented by the volume scattering compared to the lower backscatter from oil-impacted TCA. This position is supported by Costa, (2004) where backscatter from densely matted vegetation stands (which are typical characteristics of oil-free vegetation) are in most cases higher than returns in loose vegetation stands (which also characterizes stressed vegetation). This is because the oil-free woody canopy generally has a higher volume scattering as a result of the presence of leaves, foliage and branches presenting a higher density of scattering elements in the canopy volume. In contrast, the polluted sites exhibit lower backscatter because of the ground surface scattering as a result of loss of leaves and foliage.

However, microwave radiation interacts differently with different plant types and communities. The characteristics of the plant in terms of density, distribution, orientation, dielectric constant, height and branches and the condition of the ground in terms of dryness, moisture, as well as the sensor characteristics in terms of polarization, incidence angle and wavelength are important determining factors of the strength of backscatter radiation towards the antenna (Costa, 2004; Dobson *et al.*, 1996).

The result obtained from the least square regression is an indication that moisture and biomass content are very important determinants of vegetation impacted by oil pollution in the terrestrial ecosystem. The utility of NDWI and LAI has been demonstrated in Arellano *et al.*, (2015) and Ramsey *et al.*, (2015) respectively for monitoring terrestrial oil spills. Owing to the dielectric property of the targets imaged by SAR sensors, which make the imagery sensitive to surface roughness and moisture, they can be very useful in studies where moisture and biomass content of vegetation are important. Variables such as SWIR and Red Edge bands have been analyzed by a number of researchers (Adamu *et al.*, 2015; Cloutis, 1989; Hörig *et al.*, 2001; Khanna *et al.*, 2013; Kokaly *et al.*, 2013; Kühn *et al.*, 2004) to be within the fundamental spectral region for terrestrial oil spill detection. They also form the basic component of the NDVI, NDWI and LAI, utilized in this study for the primary purpose of detecting the impact of oil pollution.

NDWI is more associated with the backscatter from the three vegetation types investigated. This is in accordance with Vreugdenhil *et al.*, (2018) where a moisture derived measure (Vegetation Water Content) had the best association with Sentinel 1 derived cross polarization ratios. Similarly, in this chapter LAI showed significant association with backscatter in tree cover areas, as R^2 often reached 0.5 with S1 VV single-polarization, VV/VH cross polarization ratios and interferometric coherence. This is to be expected, as TCA in the study area are mostly composed of ever green broad-leaf woody vegetation, whose attributes are better captured with the LAI measures compared to any other indices. In addition, studies carried out by Stankevich *et al.* (2017) have also demonstrated that Sentinel 1 C-Band SAR has huge potentials for predicting LAI in forested areas, while Rüetschi *et al.*, (2017) have shown that Sentinel-1 C SAR was able to generate forest maps with classification accuracy of as high as 86% OA. This implies that polarization is an important factor in the use of SAR data for monitoring vegetation dynamics within the terrestrial landscape, primarily because stronger levels of associations were observed for the C-band backscatter compared to the X-band backscatter. This is largely a result of the high capability of C-band SAR to reasonably penetrate through the vegetation canopy compared to the surface penetration in the X-band.

The study by Spies *et al.*, (2013) has substantiated this further. Results obtained from the experimentation of L, C and X band SAR have shown that L-band had better prediction and mapping capabilities for vegetation and land cover in parts of Africa, followed by C and X band data. This is due to the wavelength and frequency of the SAR products used, although, derivatives such as DSM (Chen *et al.*, 2007), Interferometric Digital Elevation Model (Solberg *et al.*, 2018), Canopy Height Models (Schlund *et al.*, 2016) from X-band SAR such as TanDEM-X and COSMO SkyMed have performed well in mapping forest cover.

Results from the image classification have shown that the combination of optical and SAR-derived variables offer better discrimination between oil-free and oil-polluted vegetation. This showed better improvement compared to the results obtained from the previous analysis chapter (i.e. in Chapter 4) where only optical data and random forest method were used for classification. This has also been the conclusion of various other contemporary studies (Clerici *et al.*, 2017; Gao *et al.*, 2017; Mansaray *et al.*, 2017; Onojeghuo *et al.*, 2018a; Schmidt *et al.*, 2017; Whyte *et al.*, 2018) where the integration of optical and SAR derived variables improved the classification accuracy. This reinforces the position of Joshi *et al.*, (2016) that the combined use of SAR and optical data has great potential for land cover mapping based on their biophysical and biochemical properties as they are affected by external factors such as oil spills. The type of methodology is often experimental and its adoption can sometimes affect the outcome of these studies. In this study, fuzzy forest proved to be a good match with contemporary classification methods, such as random forest, especially in the grassland and TCA areas where the classification accuracy assessment result was either equal to or higher than the result obtained from random forest.

The fuzzy forest model can be used for the optimization and selection of unbiased input variables to achieve a better classification result. Results obtained from the variable importance chart have confirmed this, as only few, most diverse, uncorrelated variables were used for classification and an improved result in terms of overall classification accuracy was obtained. Furthermore, the spatial extent of the predicted oil polluted site had less generalization error, better prediction and was more in tune with the high-resolution image when the fuzzy forest was used for

classification. Results obtained here show that the FF methodology can be a remedial alternative to Random forest, especially when high volume of input variables are used for classification, most of which often present problems of high dimensionality. In this regard, elevation-derived variables (DEM, Aspect and Slope) consistently featured as the highest contributors in the variable importance charts from the two classifiers. This highlights the importance of topography in the characterization of oil-impacted areas in terrestrial landscape. Areas with low-lying elevation can be more susceptible to the impacts of oil due to the influence of gradient and slope in oil flow and runoff mechanism. However, some researchers (Hester and Mendelssohn, 2000; Hester *et al.*, 2016) have observed that the influence of elevation on plant stress can be due to soil erosion and flooding as well, and not necessarily the impact of oil. Further investigation into the reliability and influence of elevation models and variables for terrestrial oil spill impact are required, as it is possible that oil residues have greater concentrations in low-lying areas.

Observed limitations with the fuzzy forest method in this study is its inability to further improve classification accuracy on the multifrequency SAR image fusion (MSIF) classification for the three vegetation types investigated. More also, classification accuracy was also observed to be lower in cropland vegetation when the multifrequency SAR optical image fusion (MSOIF) classification was implemented. Reasons for the poor performance of fuzzy forest in this case could be attributed to the in ability of the fewer variables used for the classification to significantly improve classification accuracy. In this regard, Maxwell *et al.* (2018) have argued that the diversity often provided by multiple variables in a machine learning classification yields much more substantially improved results than when only fewer variables are used. The random forest particularly uses random samples across all input variables for training purpose, hence creates a unique generalization balance of false positives and true positives in a classification operation thereby translating to increase accuracy. Conversely, when fewer or reduced variables are used, only strong variables capable of instigating true positive characterization are used and hence true negative representation are seldomly captured in the classification result. As such reduced number of input variables may not necessarily translate to improve classification accuracy.

5.6 Summary and Main Findings

This chapter aimed at implementing a novel approach (of integrating optical image variables, multifrequency SAR image derived variables, Digital elevation model variables and geo-physical variables) into a machine learning random forest classifier to discriminate oil-free from oil-polluted vegetation types. In addition, a novel fuzzy forest (FF) image classification method was also introduced to reduce dimensionality and cope with multi-collinearity associated with the variables used for classification. Results obtained from NDWI, NDVI and LAI for detecting the effects of hydrocarbon crude oil on vegetation largely showed that oil-free vegetation were healthier due to the higher indices values obtained (with respect to median and interquartile range), compared to the low values retrieved for the oil spill sites.

Multifrequency backscatter from the various SAR sensors were also retrieved for the various site to detect and map oil-free and oil-impacted vegetation. Results obtained showed comparative result with the health indices, as the distribution of backscatter values for the oil-polluted vegetation had higher variation in backscatter, compared to oil-free vegetation. Significant difference in means were mostly associated with X band (Cosmo – Skymed and TanDEM-X) SAR and the C band (Sentinel – 1) SAR, in which LAI and NDWI for grassland and TCA vegetation had higher correlation values with SAR backscatter, reinforcing a stronger linkage between LAI of vegetation and longer wavelength C-band SAR.

Furthermore, image classification result have also shown that the combination of optical, SAR and geo-physical derived variables offer better discrimination potentials. This showed better improvement compared to the results obtained from the previous analysis chapter (i.e. in Chapter 4) where only optical data and random forest method were used for classification. Similarly, the fuzzy forest method also proved as effective variable reduction method as the optimization and selection of unbiased input variables achieved better classification results for TCA. Results obtained from the variable importance chart have also shown that elevation-derived variables (DEM, Aspect and Slope) had higher contribution in the discrimination of oil-free from oil-polluted vegetation. This highlights the importance of topography in the characterization of oil-impacted areas in terrestrial landscape, which is also

quite different from the result obtained from the first analysis chapter where health indices and spectral bands were the most important variables.

In the next chapter, the utility of longer wavelength L-band SAR would be tested in the discrimination and classification process. In addition, the seasonal dimensions and implication to mapping terrestrial oil spill would be evaluated using multitemporal Sentinel-1 SAR and multifrequency L-C-X band SAR in both dry and wet season classifications, in attempt to identify the best possible dataset, method and season to detect and map oil pollution in vegetated area.

Chapter 6 : Wet and Dry Season Assessment of Oil Polluted Vegetation Using Multi-temporal and Multi-frequency L, C and X Band SAR

6.1 Introduction

The analysis of data implemented in Chapter's 4 and 5 has shown that the integration of optical and SAR data improves the classification and discrimination of impacted and oil-free vegetation compared to using optical only data. Most importantly, the field spectroscopy experiment showed that for effective detection of oil-impact on terrestrial vegetation, there is need for healthy vegetation component to exhibit exceptional plant functional traits with respect to photosynthesis and structural development.

Figure 6-1 shows the spectral reflectance curves of vegetation in three different locations with varying degree of oil spill impact based on the age of oil. Similarly, Figure 6-2 shows comparison of spectral reflectance of oil-impacted and oil-free vegetation in other studies as well as pictures during the authors fieldwork exercise. Figure 6-1 shows that the vegetation within the non-polluted site had strong NIR reflectance, compared to the much lower NIR reflectance from the old and the most recent spill sites. This implies that proper differentiation between healthy and stressed vegetation stands within a particular area of interest would greatly influence detectability and discriminability of impacted and oil-free vegetation. The healthier the vegetation stand is, the better the detection, classification and discrimination accuracy.

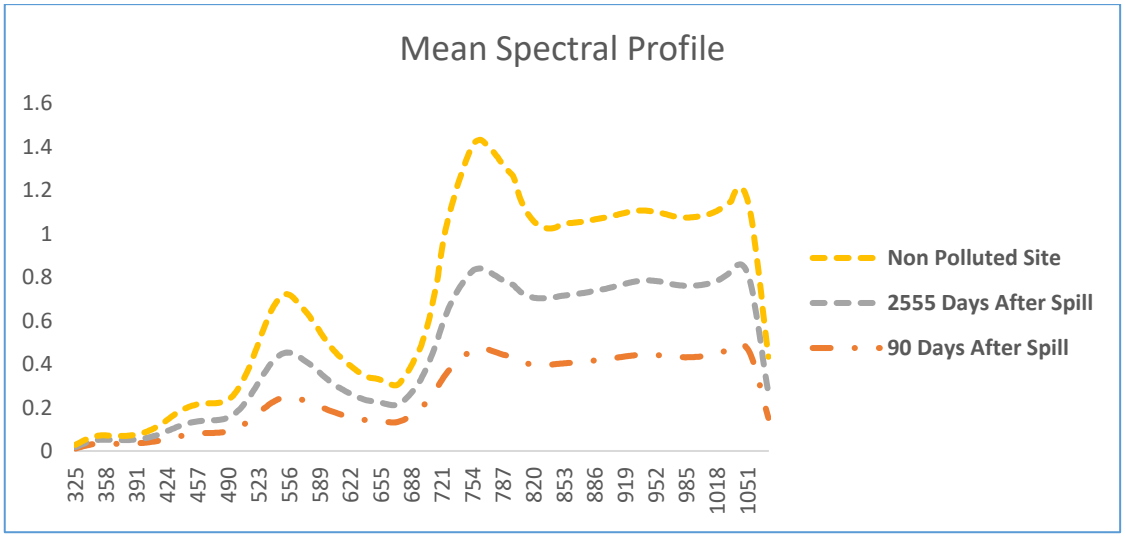


Figure 6-1: Mean vegetation Reflectance spectra for the recent, old and non – polluted sites. (Source: Authors fieldwork)

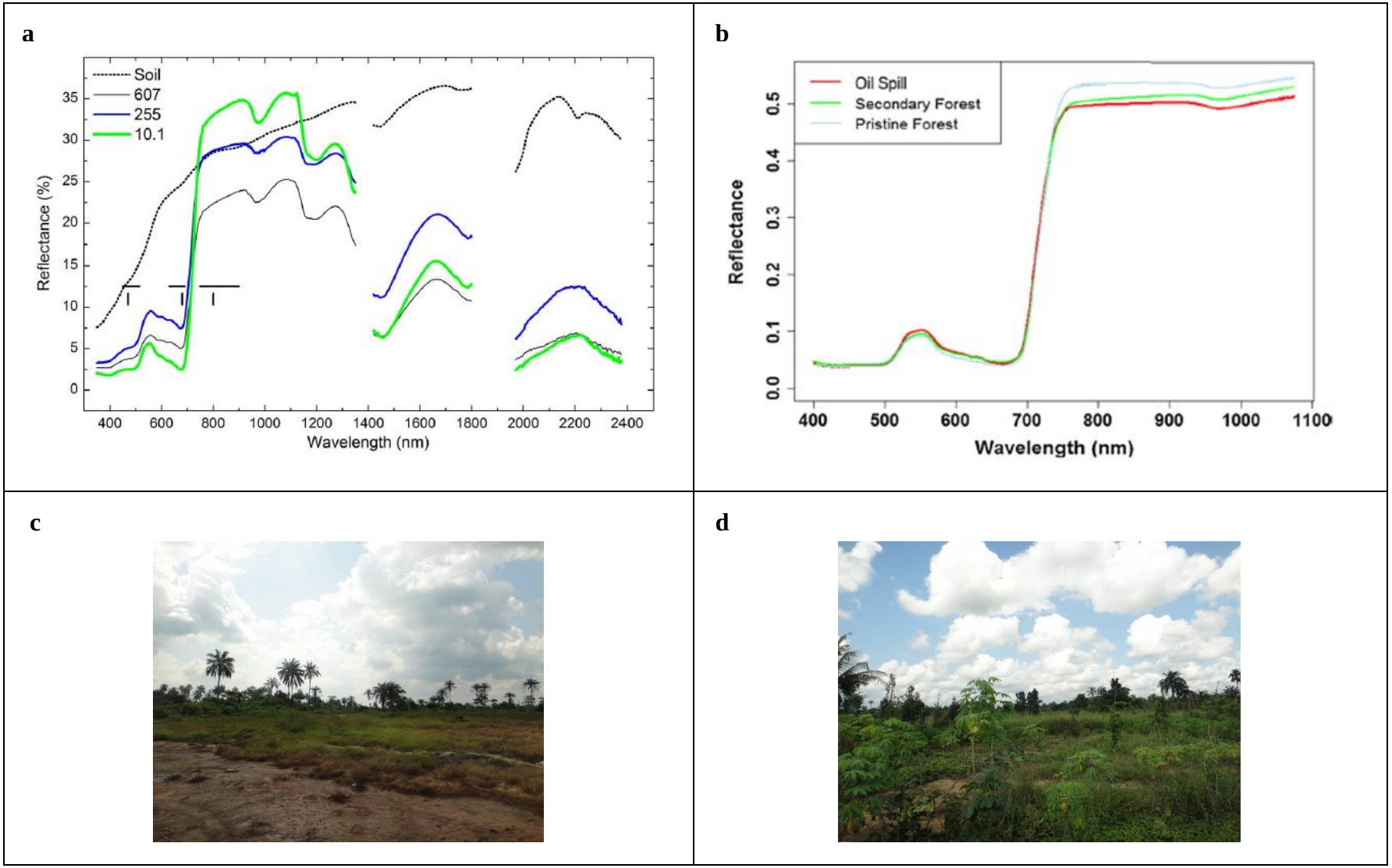


Figure 6-2 (**a**) Reflectance average of Reed vegetation communities and bare soil at same distance to source of contamination. The numbers in the legend are soil TPH concentration for each plot mg kg $^{-1}$ (Source: Zhu *et al.*, 2013b). (**b**) Averaged reflectance values of an oil polluted vegetation, oil-free Secondary forest and oil-free Pristine Forest (Source: Arellano *et al.*, 2015). (**c**) Oil-polluted cropland vegetation along an oil pipeline facility (Source: Authors fieldwork). (**d**) Oil-free cropland vegetation (Source: Authors fieldwork).

A major factor that militates against proper discrimination in this context could be the effect of seasonality. Dry season/wintery conditions in most part of the world leads to increased moisture stress and vegetation senescence (Cissé *et al.*, 2016). This tends to lead to loss of green pigment, defoliation, and increased litter fall.

In this chapter, the potential of deriving vegetation phenological descriptors and temporal characteristics from multi-temporal Sentinel 1 VH VV backscatter in detecting oil spill impact on vegetation was investigated. Machine learning RF and SVM methods were used to classify and discriminate between oil-impacted and oil-free vegetation, on separate dry and wet season multi-frequency L, C and X band SAR (for ALOS PALSAR-2, Sentinel – 1 VV and VH backscatter, TanDEM-X and Cosmo Skymed images). This provided a perspective on the potential influence and effects of seasonality on discrimination and classification accuracy.

The role of L – Band SAR in further improving classification accuracy is particularly highlighted in a multi-frequency SAR analysis, since a previous study (Ramsey *et al.*, 2015; Ramsey III *et al.*, 2011) has used only single frequency L-Band SAR in the characterization of oil-impacted marsh vegetation of Barateria bay.

6.2 Data Preprocessing

Datasets used in this chapter consist of remote sensing optical and SAR images for different dry and wet season acquisitions (Table 6-1). Figure 6-3 shows the methodology and data processing flow chart; the list of multi-temporal Sentinel – 1 data and dates acquired are detailed in section 3.2.2.

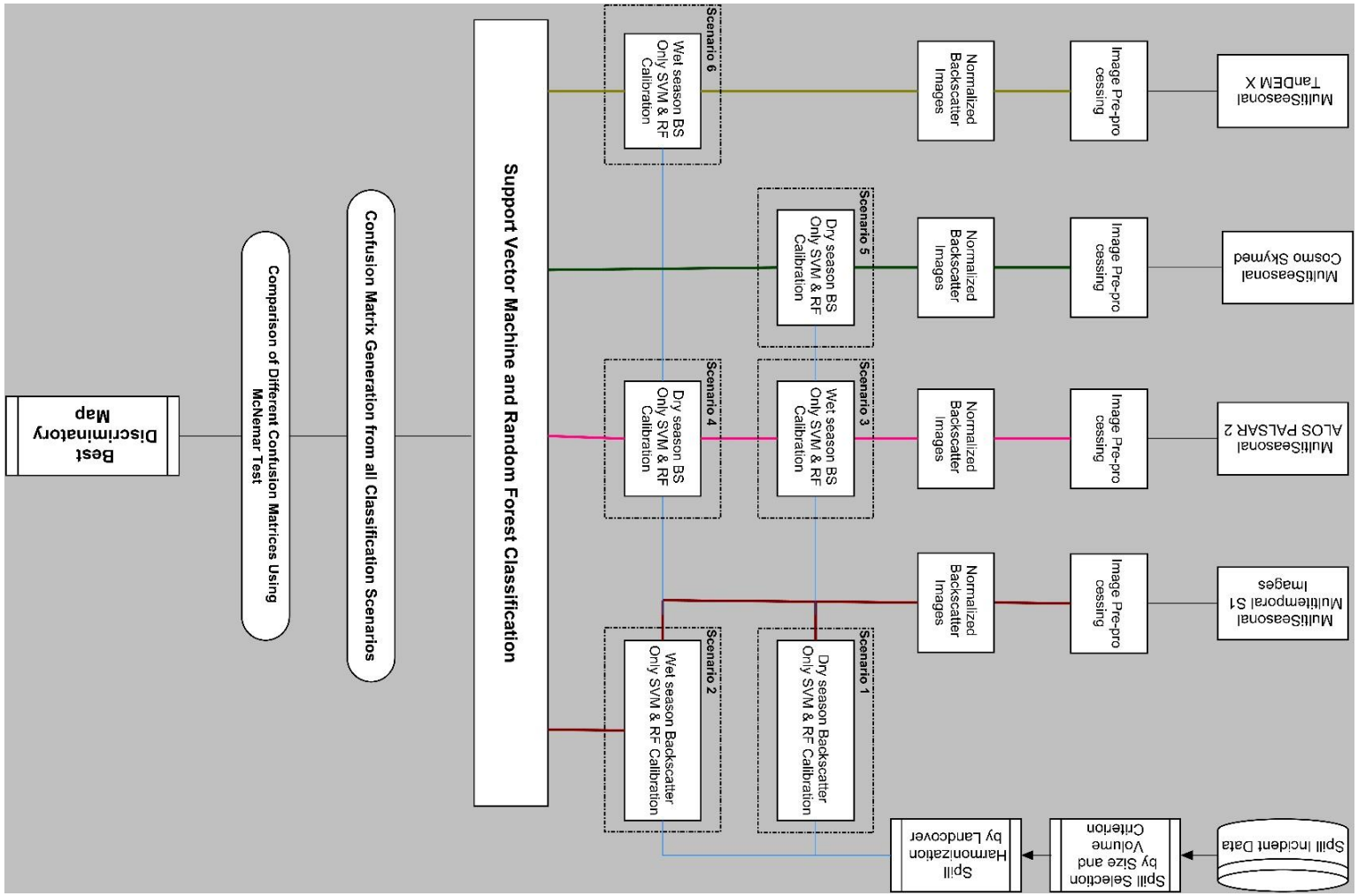


Figure 6-3: Methodological Flowchart of the Data Analysis for **Chapter 6.** Refer to further sections on the Data pre-processing (6.2)

Table 6-1**:** List of all Remote sensing data

Platform	Sensor	Swath	Spatial res.	Number of Scenes
Sentinel 1	C SAR	250km	5x20m	12 Dry and 30 Wet season images (VV VH)
TanDEM X	X SAR	30km	6m	2 Wet Season (HH)
Cosmo Skymed	X SAR	40km	3m	2 Dry Season (HH)
Endeavour Shuttle	SRTM	225km	30m	1
ALOS PALSAR 2	L SAR	70km	10m	2 Dry and 4 Wet season images (HH HV VV VH)
SMAP Data	L SAR	36km	9km	24 Scenes
MODIS NDVI	Optical	2330km	1km	24 Scenes

6.2.1 Sentinel 1 – SLC

The sentinel – 1 Single Look Complex (SLC) product as explained in section 3.2.2 was used in this Chapter. The first pre-processing step was the cropping operation; here, individual scenes were splitted to extract study area specific swaths to reduce processing time. Subsequently, scenes were radiometrically corrected to obtain initial backscatter values as DN before the multilooking operation was implemented. One look in range and four in azimuth (1:4) was used as the multilooking specification. The scenes were then geometrically rectified using the SRTM 1 Arc Second. Furthermore, the product was radiometrically calibrated from digital number values to a normalized cross section backscattering values in decibels using equation 14 (Laurin *et al.*, 2018).

$$\sigma 0 \text{ dB} = 10 \log 10 \sigma 0 \qquad (14)$$

Finally, the images were subjected to terrain correction using the range Doplar terrain correction module within the Sentinel Application Platform (SNAP) software environment (Zuhlke *et al.*, 2015). The final pre-processed backscatter images were used for two separate operations. First, they served as direct input to the experimental seasonal image stack classification. Secondly, backscatter values were retrieved for the pre-processed polluted and oil-free sample site/points for the three different vegetation types (cropland, grassland and TCA) investigated in this research. Extracted values form the basis of the multi-temporal backscatter analysis used to explore temporal characteristics of polluted and oil-free vegetation.

6.2.2 COSMO SKYMED

The Cosmo Skymed image used in this analysis is a level 1 A image which needed to be corrected for the Range Spreading loss effect using antenna pattern gain compensation and incidence angle effect following (Sportouche *et al.*, 2012). The corrected image was further multilooked (one look in range and two in azimuth), geocoded based on the SRTM, and radiometrically calibrated with a final pixel spacing of 10m × 10m. Pixels values were converted to backscatter (or normalized radar cross section), measured in decibel using the formula in equation 15 (Sportouche *et al.*, 2012).

$$\sigma_{0_{db}}^{CSK}(i,j) = 10 * \log\left(\frac{|x_{(i,j)}|^2 * \sin\left(\theta_{inc}^{ref}\right) * RSR^{2\alpha}}{CC * RF^2}\right) \qquad (15)$$

where

$x_{(i,j)}$ is the image value at pixel (i,j),

θ_{inc}^{ref} is the reference incidence angle,

RSR is the reference slant range,

α is the reference slant range exponent,

CC is the calibration constant,

RF is the rescaling factor,

Terrain correction was also implemented as for the Sentinel – 1 data (6.2.1).

6.2.3 TanDEM X

The TanDEM-X co polarized channel return signal in the HH channel was used in this study as the cross polarization HV channel data was not readily available for the desired period and study area. The acquired image was first radiometrically calibrated using the formula in equation (16) (Sportouche *et al.*, 2012). Terrain correction was subsequently carried out using SRTM data, Backscatter values were then converted from DN to linearized cross sectional backscatter values in Decibel.

$$\sigma_{0_{db}}^{TSX}(i,j) = 10 * \log\left(\left|x_{(i,j)}\right|^2 * \sin\left(\theta_{inc}^{loc}\right) * k_s\right) \quad (16)$$

where

θ_{inc}^{loc} is the local incidence angle

k_s is the calibration and processor scaling factor

$x_{(i,j)}$ is the image value at pixel (i,j)

log is the decimal logarithm.

6.2.4 ALOS PALSAR

The multiple ALOS PALSAR scenes with HH, HV, VV and VH polarizations were pre – processed using a standard pre-processing chain operation. First, the scenes were radiometrically calibrated and thereafter they were multilooked (one look in range and two in azimuths) before being geocoded based on Shuttle Radar Topography Mission (SRTM) data, and finally radiometrically calibrated with a final pixel spacing of 10m × 10m. Pixels values were converted to backscattering measured in decibel using the formula below (Eqn 17) (Sportouche *et al.*, 2012)in SNAP.

$$\sigma_{slc}^{0} = 10 \cdot log_{10}\langle I^2 + Q^2\rangle + CF_1 - A \quad (17)$$

where

DN Digital number (or raw pixel value)

σ^0 is the backscattering coefficient (Sigma naught) for a specific polarization

CF_1, A: is Calibration factor (Unit: dB).

6.2.5 Soil Moisture Data

Monthly SMAP raster data were imported into the ArcGIS software package and layer stacked without resampling (as the SMAP data were not used in the classification). The preprocessed points were then used to retrieve the soil moisture values across the multitemporal SMAP stack. The data was subjected to statistical analysis to infer secondary relationship between dominant soil moisture characteristics, health of vegetation as indicated by NDVI and temporal backscatter characteristics.

6.2.6 MODIS NDVI

A major preprocessing operation on the MODIS NDVI data is similar to the SMAP data, as the dataset is a ready to use product.

6.2.7 Oil Spill Point and Incident Data

The processed points (3.2.3) were used for class-wise assignment of pixels into oil-free and polluted vegetation class using the random forest and support vector machine image classification operation. Informational classes include oil-free cropland, oil-free grassland, oil-free TCA, polluted cropland, polluted grassland and polluted TCA. The selected points (Table 6-2) were divided into two sets by a ratio of 60:40 for training and validation respectively. The processed spill and oil-free sample points were also used to retrieve backscatter values from the pre-processed SAR images for the multi-temporal/multi-seasonal analysis. Table 6-2 shows the locations of the various oil spill incidents and number of points selected from the spill sites, while Table 6-3 shows the total number of training and validation sites used for model calibration and classification.

Table 6-2: Oil spill incident locations and the total number of ground reference selected

Location	Year/Month	LGA	Ref. Points
Atali	01/10/2015	Obio/Akpor	10
Ayama	01/06/2016	Oyigbo	7
Ayama - Noke	01/12/2015	Oyigbo	4
Ayama - Noke	14/01/2016	Oyigbo	4
B-Dere	01/06/2015	Gokana	4
B-Dere	01/06/2015	Gokana	2
B-Dere	14/01/2016	Gokana	3
B-Dere	28/05/2015	Gokana	5
B-Dere	16/02/2016	Gokana	2
B-Dere	01/06/2015	Gokana	2
B-Dere	01/06/2015	Gokana	1
Egberu	01/08/2015	Obigbo	2
Egberu	16/02/2016	Obigbo	8
Egberu	01/08/2015	Obigbo	1
Ejamah Ebubu	01/06/2015	Eleme	1
Ejamah Ebubu	21/02/2015	Eleme	2
Ejamah Ebubu	01/06/2015	Eleme	1

Location	Year/Month	LGA	Ref. Points
Gio	01/08/2015	Gokana	23
Gio	01/09/2016	Tai	22
Gio	01/08/2015	Gokana	5
Imeh	01/09/2015	Etche	4
Komkom	01/05/2015	Oyigbo	9
Komkom	12/05/2015	Oyigbo	10
Komkom	01/09/2016	Oyigbo	4
Komkom	01/05/2015	Oyigbo	4
Kpoghor and Gio	01/05/2015	Tai	7
Kpoghor and Gio	16/02/2016	Tai	10
Kpoghor and Gio	01/09/2015	Tai	7
Norkpo	01/09/2015	Tai	7
Odagwa	01/09/2016	Etche	3
Ogale	01/08/2016	Eleme	10
Ogale	16/02/2016	Eleme	3
Owaza	01/05/2015	Ukwa West	3
Owaza	28/05/2015	Ukwa West	11

Table 6-3: Specific Number of Polluted and Oil free (Non – Polluted) Sites for the Different Landcover

Class Label	Total Number of Sites
Non – Polluted Cropland	131
Non – Polluted Grassland	84
Non – Polluted Tree Cover Areas	124
Polluted Cropland	148
Polluted Grassland	79
Polluted Tree Cover Areas	114

6.3 Methodology

6.3.1 Establishing Temporal Variability Using Sentinel 1 Backscatter

To investigate potential variations of bio-physical properties of vegetation due to oil pollution, temporal backscatter profiles were produced for different polluted and oil-free vegetation types investigated in both VV and VH channel. In addition, mean backscatter values were computed for the different vegetation types and finally seasonal variability was tested using statistical measures to ascertain if there is statistically significant difference between the observed mean seasonal trends.

6.3.2 Temporal Backscatter Profile

Furthermore, temporal backscatter descriptors were retrieved to capture the predominant vegetation dynamics because of the presence or absence of the impact of oil and cross section backscatter values from the representative samples of the six vegetation classes evaluated. This had the advantage of showing the upper and lower interquartile ranges, as well as highlighting the lowest, highest and median backscatter values.

6.3.3 Mean Backscatter Profile

Results from the temporal backscatter profile were further condensed to give a generalized figure for the respective polluted and oil-free vegetation types. The mean backscatter values were computed for the respective polluted and oil-free cropland, grassland and TCA sites from the VV and VH channel of the Sentinel – 1. This gave a better overview of how vegetation canopy structure changed intermittently and

varied over the 22 months investigated. Mean backscatter values were considered appropriate for this purpose because they represent global averages which can effectively capture the characteristics of all the different samples and reduce the effect of variance from the upper and lowest return signal. Similarly, Rüetschi *et al.*, (2017) have also recommended mean backscatter value to describe and infer dominant temporal backscatter trends. However, the rationale behind the use of the mean values is that the initial investigation carried out to assess backscatter difference before and immediately after spill incident in Ramsey *et al.*, (2015) did not show any significant and conspicuous change in backscatter.

6.3.4 Comparison of Backscatter for Various Targets

The result of the mean temporal backscatter profile from the polluted and oil-free vegetation types were further compared to other land cover types within the study area. This included the mean temporal backscatter from built – up areas, bareland and waterbody within the study site. This helped in conceptualizing the result of the temporal returns from the primary targets (oil-free and polluted-vegetation) in this research with results from other land cover components in the study area.

6.3.5 Assessing Seasonal Variability

Seasonality plays a significant role in the determination of the health status of vegetation (Cissé *et al.*, 2016), as environmental conditions (with respect to soil moisture, precipitation and humidity) particular to a specific season can positively influence plant growth or negatively influence it through senescence and moisture stress. In order to establish if the difference observed in the mean backscatter values across different seasons for the polluted and oil-free vegetation are substantial enough to infer dependable change in vegetation attributes. The paired samples t – test was used to compare the result of the retrieved mean backscatter values for oil-polluted with oil-free vegetation in the dry season and in the wet season. A critical value = 0.05 was used at 95% confidence to infer significant difference in the mean backscatter values for the polluted and oil-free vegetation in a particular season assessed.

6.3.6 Relationship between Dominant Bio – Physical Characteristics and Temporal Backscatter

To further assess and validate trends of the multi-temporal backscatter characteristics, the retrieved mean temporal soil moisture index from SMAP data (https://smap.jpl.nasa.gov/data/) and the mean temporal NDVI index from the Terra MODIS sensor (https://search.earthdata.nasa.gov/search) were visually compared using line plots. This was used to check for temporal variability and analyze for seasonal dynamics of vegetation health owing to the impact of oil.

For an independent assessment of vegetation biophysical characteristics within the study area, mean values from separate independent oil-free vegetation sample points were used to retrieved temporal NDVI, soil moisture and backscatter (VV – VH) values. The temporal NDVI and soil moisture values were then used to establish a relationship with the corresponding VV and VH backscatter using least square regression. This was done to assess and establish trends, dependencies and agreement between the various biophysical factors (soil moisture and NDVI) within a typical vegetated area in the study site and their specific relationship to dominant temporal backscatter.

The R^2 value from the modelled relationship was used to characterize the dominant backscatter against, the soil moisture and NDVI relationships. This provided an avenue to compare the result of the retrieved temporal backscatter for the polluted and oil-free vegetation types to show if observed temporal backscatter trends are in alignment with available optical derived indices, and examine the role of soil moisture condition.

Emphasis was not given to establishing a similar modelled relationship for both the polluted and oil-free sites separately used in this study. This is because of the coarse resolution of the data sets, that is, 9km for SMAP soil moisture and 250km for the MODIS NDVI. Compared to the 10m spatial resolution used for the derived temporal backscatter, as observation may not reflect the actual realities on ground.

6.3.7 Multi Seasonal Image Classification (MSIC)

The multiseasonal image classification (MSIC) followed the image pre-processing of the Sentinel – 1 dual polarization (VV and VH), ALOS PALSAR quad polarization (HH, HV, VH and VV), the Cosmo Skymed single Polarization (HH) and TanDEM – X single polarization (HH) images for the various months of the dry and wet season. Both dry and wet season image stack for the aforementioned images were generated in the order of Table 6-4 for experimental image classification scenarios.

The classification operation was on the image subset for the three vegetation types of cropland, grassland and TCA to discriminate between polluted and oil-free vegetation components. Machine learning Random Forest classifier (3.3.1) and Support Vector Machine (3.3.3) were used for the image classification.

Table 6-4: Showing input data for the different image classification scenario's implemented

Classification Scenarios	Season	Datasets Used	Total No. of Images
Scenario – 1	Dry Season	Multi-temporal S1	12
Scenario – 2	Wet Season	Multi-temporal S1	30
Scenario – 3	Wet Season	Multi-temporal S1 + ALOS 2	34
Scenario – 4	Dry Season	Multi-temporal S1 + ALOS 2	14
Scenario – 5	Dry Season	Multi-temporal S1 + ALOS 2 + CSM	16
Scenario – 6	Wet Season	Multi-temporal S1 + ALOS 2 + TDX	36

6.3.8 Assessing Classification Accuracy

The process of classification accuracy was carried out to assess the performance of the classification methodologies (RF and SVM) in discriminating between polluted and oil-free vegetation types using wet and dry season multifrequency L, C and X band SAR images. The data were combined classified separately for different vegetation types. Error matrix (Congalton, 1991), was first used to determine the correctness of the map generated, by comparing with the 40% reference points for polluted and oil-free sites (oil-free cropland, oil-free grassland, oil-free TCA, polluted cropland, polluted grassland and polluted TCA). Here, emphasis was on the Producers, Users and Overall Accuracy.

Secondly, the McNemar's tests (de Leeuw *et al.*, 2006) was also employed to measure and assess class predictions errors among the two classifiers and between the various seasonal classifications to correctly predict the reference oil-impacted and oil-free sites.

The McNemar's test is a nonparametric test based on χ^2 using a 2 *x* 2 contingency matrix to assess the performance of multiple classifier outputs based on the number of correctly predicted samples. The accuracies were considered as statistically significant at a confidence level of 95% if the calculated χ^2 (from equation 18) is larger than the critical value of 2.5. The samples are labelled as f_{12} and f_{21} which represents the correctly classified samples by a particular scenario (e.g. sc1) that were misclassified by sc2, and the number of correctly classified samples for sc2 that were misclassified by sc1, respectively (Whyte *et al.*, 2018).

$$X^2 = \frac{(f_{12} - f_{21})^2}{f_{12} + f_{21}} \qquad (18)$$

6.4 Results

6.4.1 Temporal Sentinel – 1 VV – VH Backscatter

The temporal VV – VH backscatter for the polluted and oil-free vegetation types is a representation of the dominant characteristics of the reflective surface illuminated by the Radar sensor. This was obtained from the single monthly Sentinel – 1 single look complex (SLC) image from April 2015 to December 2016 (i.e. 22 months), allowing for reasonable deductions such as: the maximum, minimum and mean temporal backscatter values for individual vegetation types to be assessed.

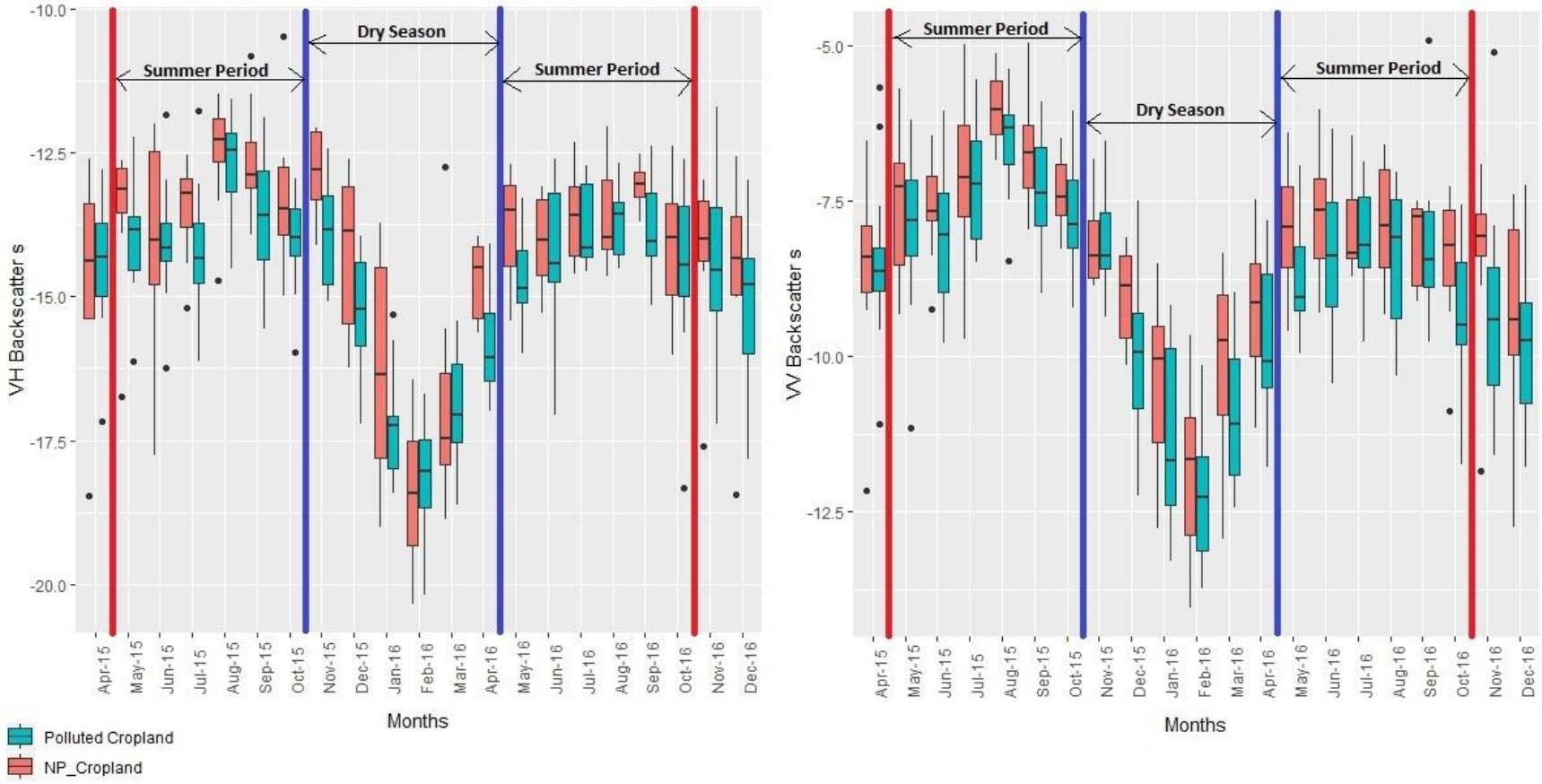
- Cropland

Retrieved VH backscatter values from the polluted and oil-free cropland vegetation have similar scattering tendencies (Figure 6-4a and 6-4b). Backscatter from the polluted and oil-free vegetation often ranged between -17 and -13dB in the late spring month of March and the summer month of August. The signals in the VH channels from the two cropland vegetation were observed to attenuate from the early winter and generally reached their minimum in the late winter. Backscatter ranges in this case was between -20 and -14. In addition, the median backscatter values for oil-free cropland in both the VV and VH channels were observed to be much higher than those of the polluted cropland vegetation.

The retrieved backscatter values from the VV channel for polluted and oil-free cropland vegetation (Figure 6-4b), show that backscatter strength from this channel is higher than the VH channel for both polluted and oil-free cropland. Backscatter values were slightly different between the two targets, owing to the wider range and intra backscatter variation in polluted cropland vegetation compared to the oil-free cropland.

Backscatter for polluted cropland were observed to range between -11 and -6 during the late spring month of March. These generally reached their peak in the summer month of August and thereafter a decline in the backscatter strength followed. Generally, low backscatter returns were recorded in the winter months of November, December through to February and often ranged between -14 and -7. For oil-free cropland vegetation, results show that backscatter in the late spring month of March

ranged between -12 and -6; and this gradually rose to -7 and -5 during the peak summer period of August, before a steady decline was observed in September through into the winter months of October and November. Low backscatter values were also recorded during the winter period for the oil-free cropland vegetation, which was slightly different to the polluted cropland; the backscatter range in this case was between -14 and -8dB.



(a) (b)

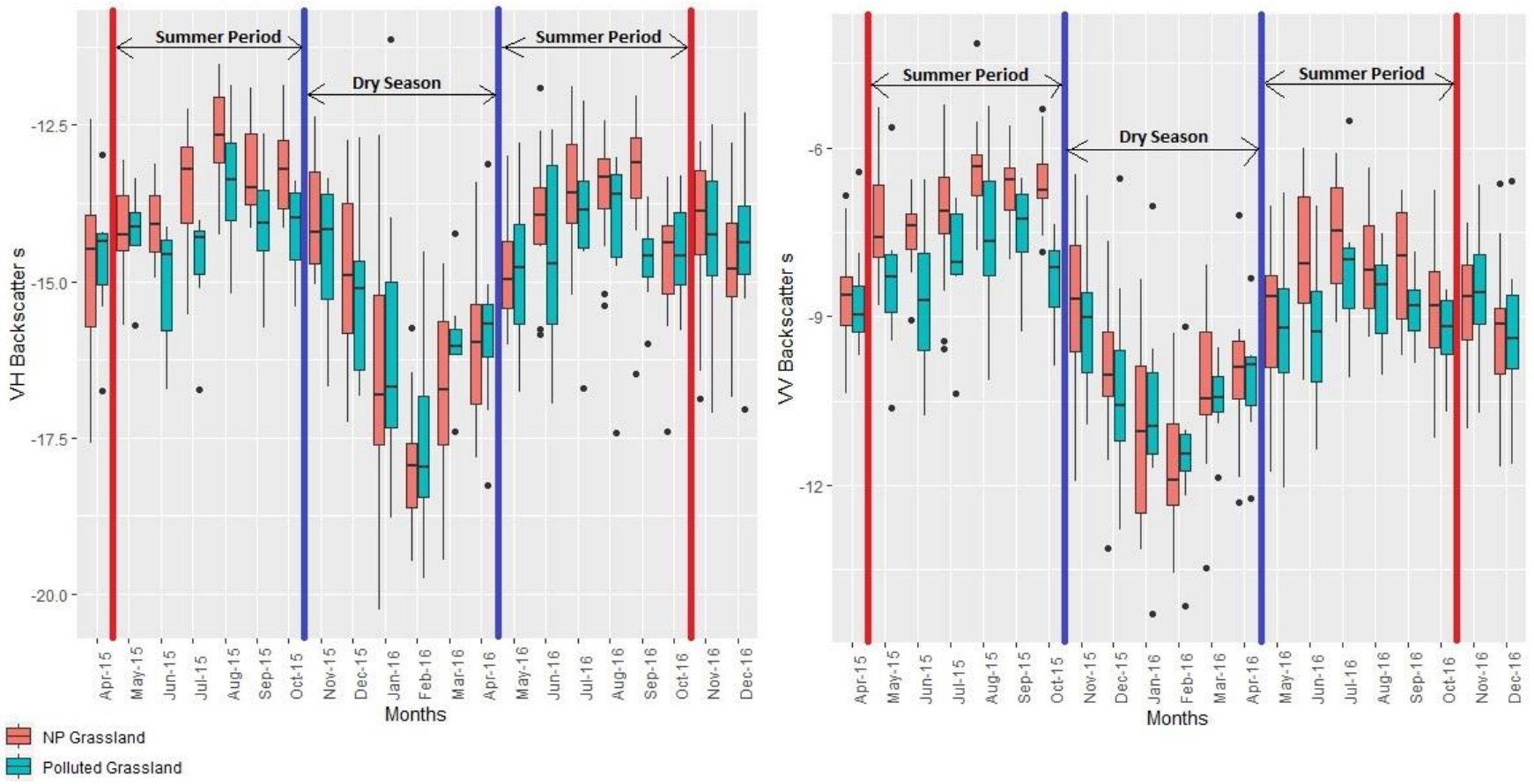
Figure 6-4: Box plot showing the multi-temporal sentinel – 1 backscatter distribution of Polluted and Oil free (Non – Polluted) Cropland. (a) VH Channel (b) VV Channel

- Grassland

Figure 6-5a presents the result of the multi-temporal VH backscatter for polluted and oil-free grassland vegetation. This shows a slight difference from the cropland VH backscatter. The backscatter in this case showed little multi-temporal variability across the months assessed. Backscatter for polluted grassland often ranged between -13 and -17dB at the beginning of the year (during the spring month of March), which largely remained same throughout most parts of the spring and summer months. It was also observed that the VH backscatter strength gradually declined from early winter month of November and generally reached its minimum in the month of February. Minimum backscatter in the winter months for this vegetation type ranged between -19 and -17 at the peak of winter.

However, for oil-free grassland, the results showed meaningful variability in the strength of retuned backscatter across the 22 months assessed, mostly accounted for as a result of the close values across multiple sites. Backscatter values often ranged between -16 and -13 in the early spring month of April and generally reached its peak scatter intensity in the late summer month of October. Minimum backscatter values were between -20 and -14 in the winter months of January and February.

Figure 6-5b presents the result of the multi-temporal backscatter from the VV channel for the Polluted and Oil-free grassland vegetation. This shows that there is a huge contrast between retrieved backscatter values for Polluted and Oil-free grassland vegetation. While oil-free grassland showed meaningful multi-temporal seasonal variance in VV backscatter, the polluted grassland showed little variability. In addition, backscatter intensity values were substantially higher in the VV channel compared to the VH channel for grassland, and median backscatter value for oil-free grassland vegetation is much higher than the corresponding polluted grassland vegetation. Backscatter often ranged between -10 and -7 at the beginning of the year during the late spring period of March and mostly reached saturation and maximum return in the summer month of September.



(a) (b)

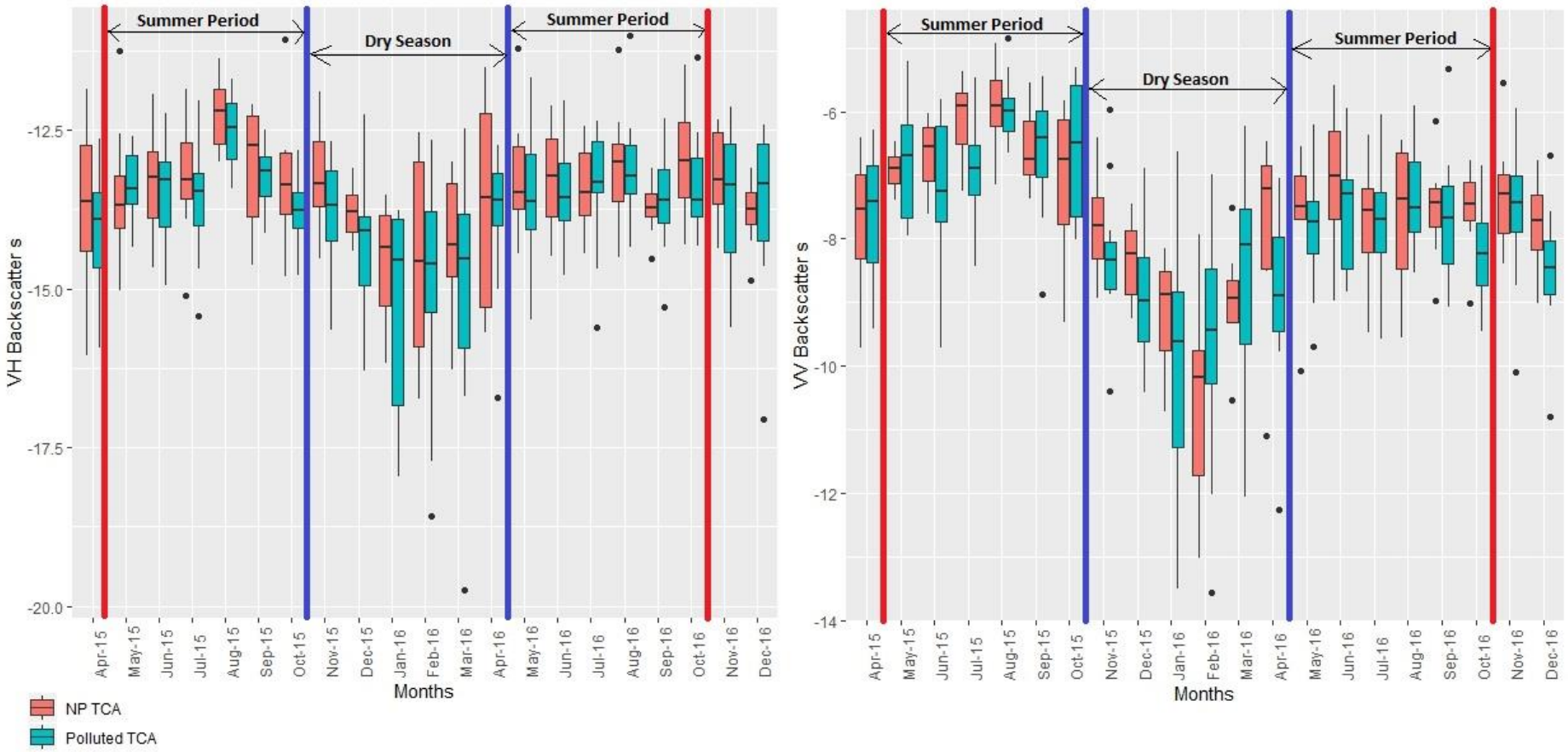
Figure 6-5: Box plot showing the multi-temporal sentinel – 1 backscatter distribution of Polluted and Oil free (Non – Polluted) Grassland. (a) VH Channel (b) VV Channel

Backscatter in this case ranged between -7 and -5 during the peak summer period, before a gradual decline in backscatter strength began in October and generally reached its minimum in February. Minimum backscatter also ranged between -14 and -10. In contrast, polluted-grassland was observed to have backscatter range of between -10 and -6 in the early spring period of April and generally reached its peak intensity in the summer month of August with a maximum backscatter range of -9 and -7. Minimum backscatter values were between -15 and -11dB in the winter months of January and February.

- Tree Cover Areas

The result for the retrieved multi-temporal VH backscatter for tree cover area vegetation is presented in Figure 6-6. Multi-temporal trends observed in this case were somewhat similar to the result obtained for grassland vegetation in the VH channel. Figure 6-6 shows that the polluted TCA had little multi-temporal variance across the months assessed. Backscatter often ranged between -14 and -8dB in the early months of March and remained largely same throughout the subsequent months. Variations as a result of seasonality is not as pronounced as was observed with the grassland and cropland vegetation.

In contrast, however, the result obtained for oil-free TCA in the VH channel showed high temporal variability. Backscatter ranges in the early spring month of March were between -16 and -12. Backscatter strength was highest in the summer month of September, before a decline was observed as the dry winter period set in. Minimum retrievals were between -17 and -14 in the peak of the winter month of February.



(a) (b)

Figure 6-6: Box plot showing the multi-temporal sentinel – 1 backscatter distribution of Polluted and Oil free (Non – Polluted) Tree Cover Areas. (a) VH Channel (b) VV Channel

Figure 6-6b presents the result of the temporal VV backscatter for the polluted and oil-free TCA vegetation. The VV returns for the polluted-TCA shows considerable dissimilarity when compared with the result obtained in the VH. The result show that there is more temporal variance, suggesting the effect and influence of seasonality in the backscatter return for the polluted-TCA. In addition, backscatter strength was much higher in the VV channel as was the case with cropland and grassland vegetation, compared to the low VH backscatter values. Backscatter often ranged between -9 and -6 decibel in the early spring period and this gradually rose to reach peak backscatter return in the summer month of August. Similarly, as the autumn and winter months set in, the backscatter values are observed to decline accordingly, until a minimum backscatter return range of -14 and -7 is reached. This result tends to be similar to the oil-free TCA vegetation, where backscatter range is between -10 and -5dB in the early spring months, and also gradually increases to attain peak backscatter intensity values in the summer months.

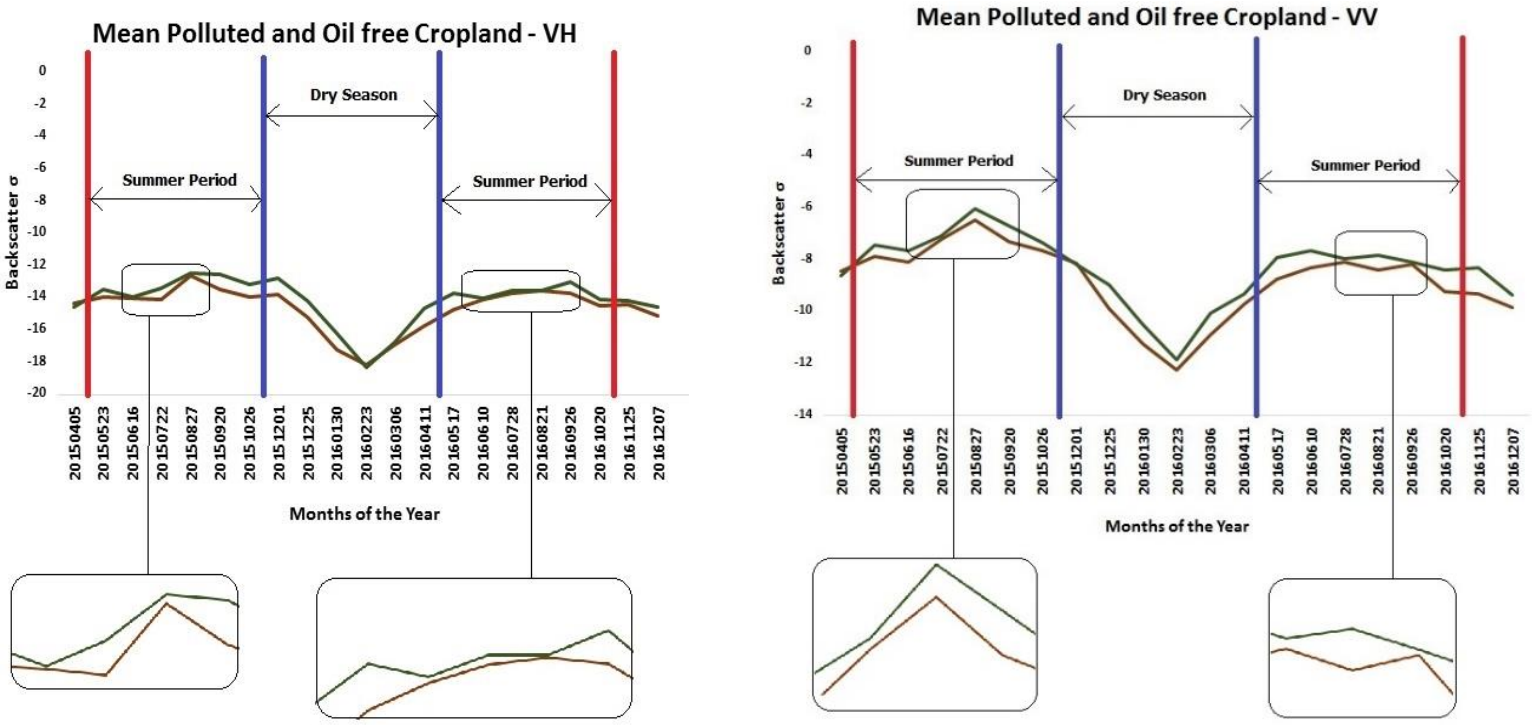
6.4.2 Mean Temporal Backscatter Profile

The mean temporal backscatter profile shows the average backscatter from the sampled sites of polluted and oil-free cropland, grassland and tree cover areas for individual months from April 2015 through to December 2016. This allowed for general comparison to be made of the dominant vegetation characteristics within the respective polluted and oil-free vegetation types. In addition, mean values of polluted and oil-free cropland, grassland and tree cover area that fell within the corresponding wet season (March to October) and dry season (November to February) were analyzed for seasonal differences. Here, pairwise comparison of backscatter values for polluted and oil-free vegetation were tested using paired sample t-test to test for difference in means for specific season (i.e. wet or dry) at a 95% confidence interval. This allowed for conclusive deductions to be made regading the characteristics of oil-free and oil-polluted vegetation within a particular season.

6.4.2.1 Cropland

The results of the mean temporal VV and VH backscatter for polluted and oil-free cropland is presented in figure 6-7a. This figure shows that during the spring month of April (in the early rainy season), mean temporal backscatter values for polluted and oil-free cropland are marginally separable in the VH channel. Separability between

oil-free and polluted cropland however gradually increased as the rainfall season progressed through the months of June, July and August, to the peak rainfall summer month of September. Disparity between the means of polluted and oil-free cropland trend continued into the early dry season months of October, November and December. However, in the peak of the winter (dry) season months of February and March, mean temporal backscatter from oil-free and oil-polluted cropland became inseparable in the VH channel until April of the preceding year, before mean backscatter became separable. Result of the paired samples t – test showed that the difference in the mean VH backscatter values for polluted and oil-free cropland during the dry season is not significant ($P>0.05$). While the t – test showed that the difference in the mean temporal VH backscatter values in the summer period for polluted and oil-free cropland is significant ($P<0.001$).



(a) (b)

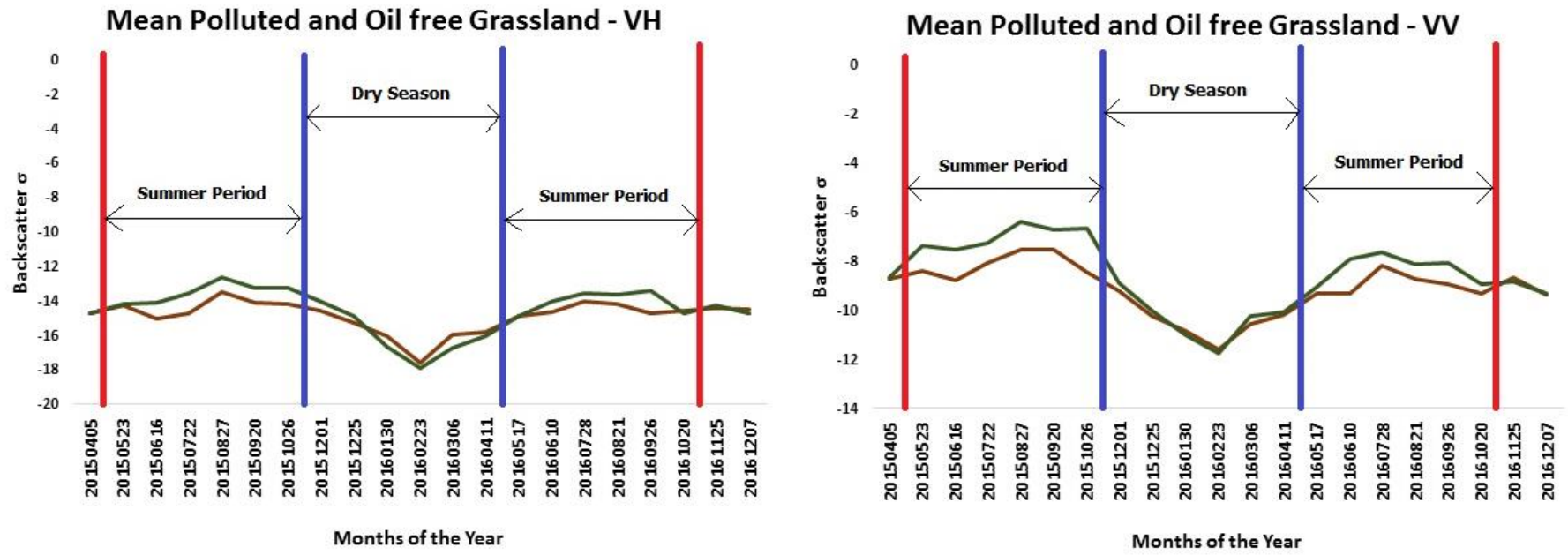
Figure 6-7: Mean temporal backscatter values for Polluted (maroon coloured line) and Oil free (Non – Polluted) Cropland vegetation (green coloured line). (a) VH Channel (b) VV Channel. The figure indicates that mean backscatter values for polluted cropland is significantly lower than the mean backscatter of oil-free cropland vegetation during the rainy/summer season in the VH channel. While in the dry/winter season, the difference is not statistically significant. In the VV channel the difference between the polluted and oil free cropland are significantly different throughout both wet and dry season.

The result obtained in the VV channel shows a huge contrast to the result obtained in the VH channel. Figure 6-7b shows that polluted and oil-free cropland vegetation remained highly separated for most part of the wet and dry season. Result of the paired sample t – test showed that the difference in mean temporal backscatter between polluted and oil-free cropland vegetation in the two seasons investigated were significantly different (P>0.001).

6.4.2.2 Grassland

The results for the comparison of the mean temporal VV and VH backscatter values for the polluted and oil-free grassland vegetation are presented in Figure 6-8. This figure shows that there is high mean temporal backscatter separability between polluted and oil-free grassland in VH channel. This is also reflected in the paired sample t – test result, which showed that the difference between polluted and oil-free grassland vegetation in the summer (wet) season is statistically significant (as P value = 0.003). In contrast, result obtained for the winter (dry) season showed very little to no separability as the t – test result showed that the difference in mean VH backscatter between polluted and oil-free grassland is not significant (as P value = 0.608).

Trends observed in the VV channel are not different from the observations in the VH channel for polluted and oil-free grassland. At the start of the spring of 2015, mean VV backscatter of polluted and oil-free grassland were observed to exhibit conspicuously high degree of separability and this trend was largely sustained through the summer – wet season. However, in the onset of the autumn and further into the dry (winter) season, backscatter strength gradually reduced until the signal from both polluted and oil-free grassland became inseparable. Results of the paired sample t – test also affirmed that the difference in mean VV temporal backscatter (between polluted and oil-free grassland) in the dry (winter) season is slightly significant (with P value = 0.034). In addition, the result of the paired sample t – test showed that the difference in mean temporal VV backscatter retrieval between polluted and oil-free grassland in the wet (summer) season is statistically significant.



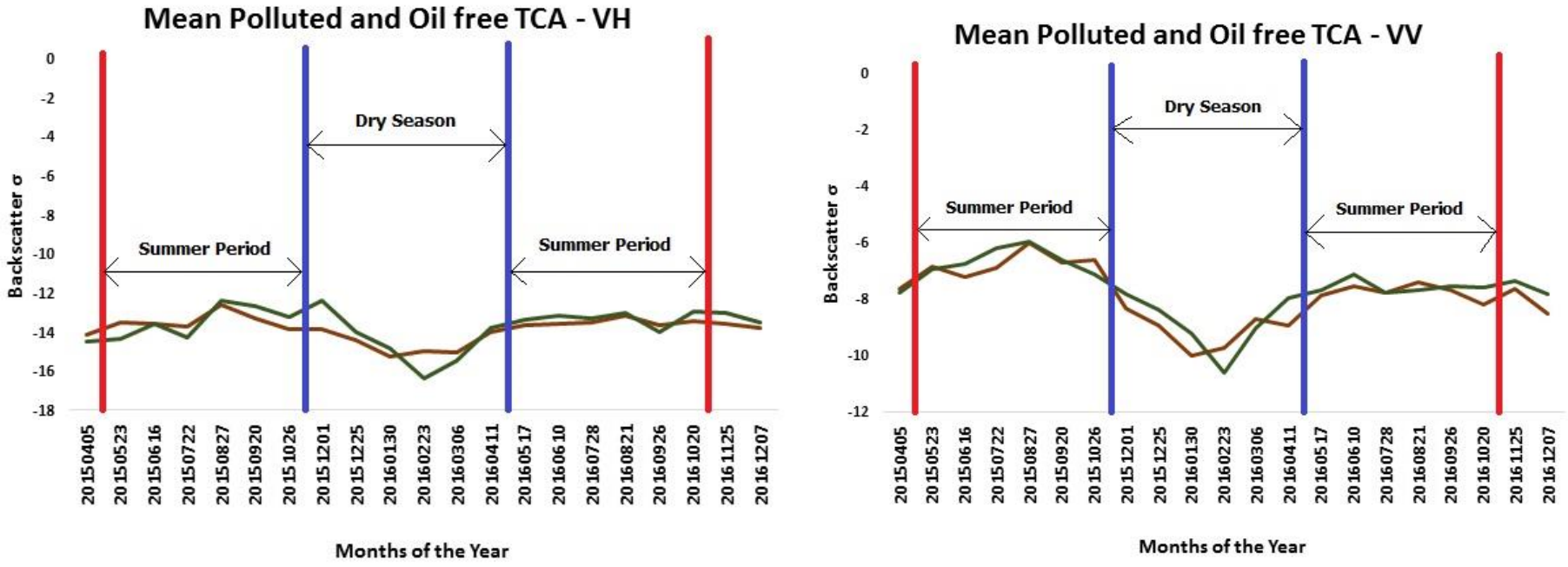
(a) (b)

Figure 6-8: Mean temporal backscatter profiles for Polluted (maroon coloured line) and Oil free (Non – Polluted) Grassland vegetation (green coloured line). (a) VH Channel (b) VV Channel. The figure also indicates that mean backscatter values for polluted grassland and oil-free grassland is significantly different during both the rainy/summer season in both VH and VV channel. While in the dry/winter season, the difference between the polluted and oil free grassland vegetation is not significantly different throughout.

6.4.2.3 Tree Cover Areas

The result of the mean temporal VV and VH backscatter for polluted and oil-free TCA is presented in Figure 6-9. This showed that the mean VH temporal backscatter between polluted and oil-free TCA exhibited little to no separability both in the dry (winter) and wet (summer) seasons. Results from the paired sample t – test also confirmed this position as *P* values (*P*>0.804 and *P*>0.720 were obtained for the dry winter and wet summer seasons respectively) did not record any significant difference. Similar characteristics were also observed for the mean temporal VV backscatter, as a rather unstable mean temporal scatter trend was observed for both polluted and oil-free TCA, although, the t – test was significant for the VV channel.

However, it is observed that the mean scattering trend was largely influenced by seasonal variations, as maximum scattering values for both polluted and oil-free TCA reached their respective peak values during the summer (wet) season, while minimum scattering values were observed during the dry (winter) season. The paired sample t – test result showed that the difference between the mean VV temporal backscatter of polluted and oil-free TCA vegetation is not significant in the dry (winter) season (*P* value = 0.296 was recorded). However, the results of the t – test showed that the difference in mean VV temporal backscatter between the two TCA vegetation was slightly significant (as *P* value = 0.044 was obtained) in the dry – summer season.



(a) (b)

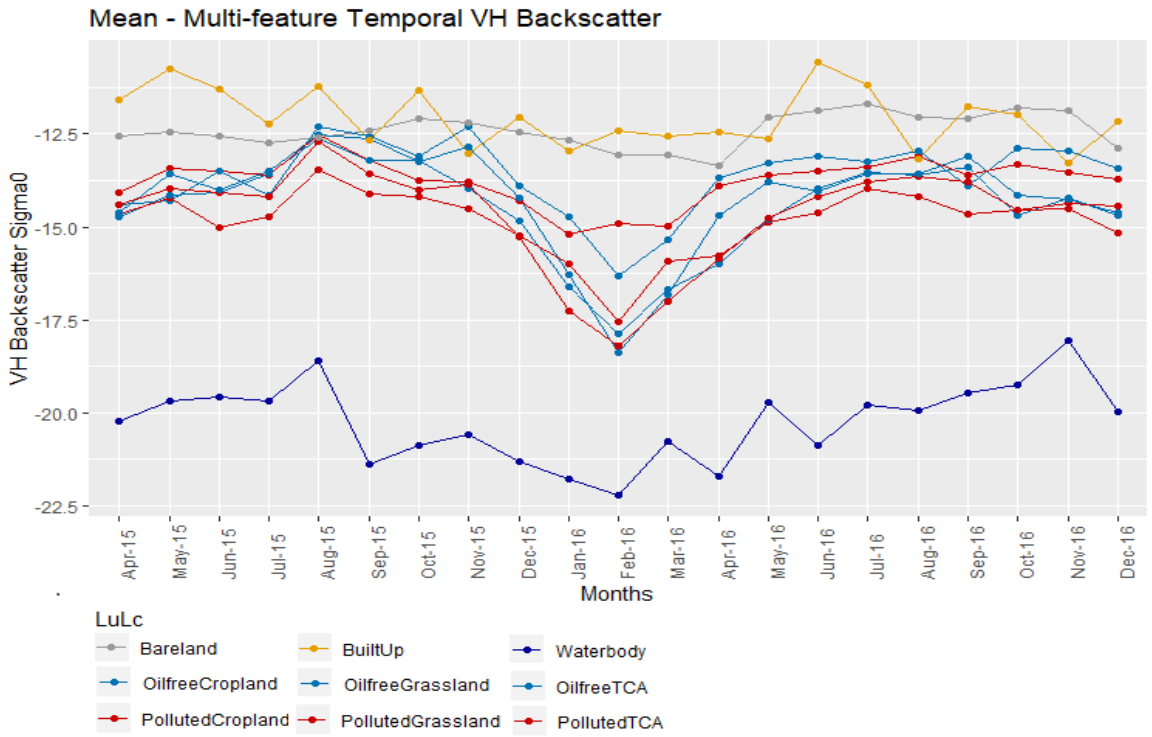
Figure 6-9: Mean temporal backscatter values for Polluted (maroon coloured line) and Oil free (Non – Polluted) Tree Cover Area vegetation (green coloured line). (a) VH Channel (b) VV Channel. Results did show any significant difference between polluted and oil free TCA in both dry and wet season in both the VH – VV channel

6.4.3 Comparison with other Land cover types

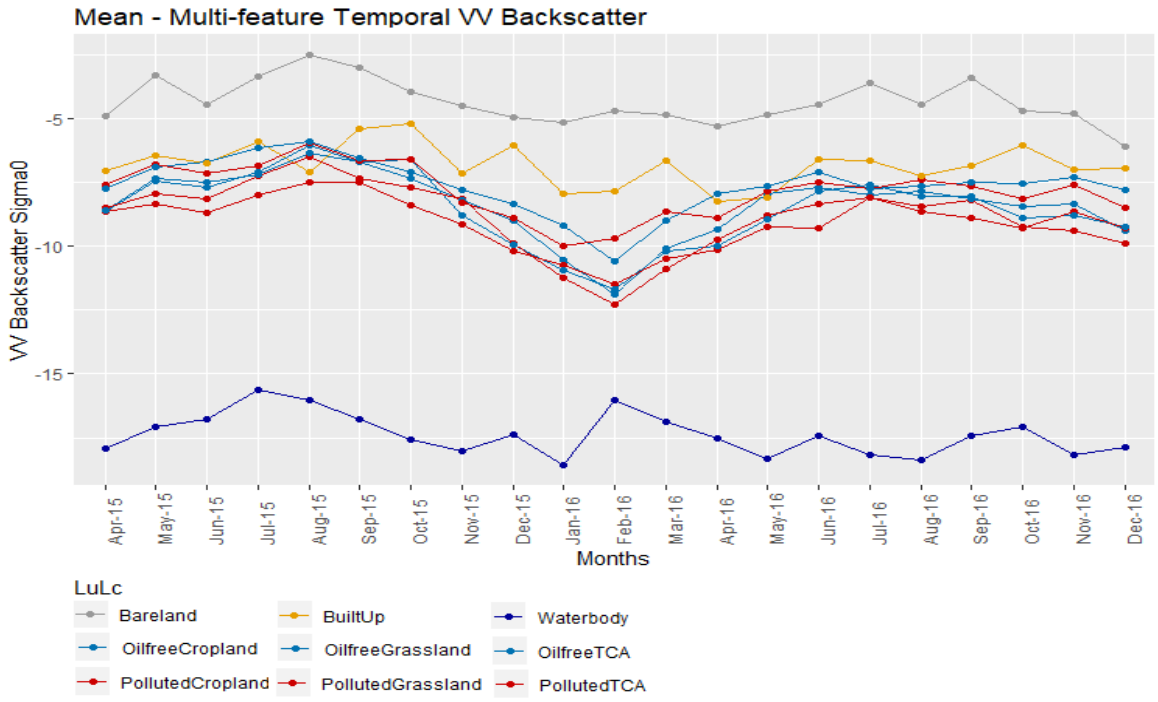
Figure 6-10 shows line plots of other land cover features (built-up areas, bare land and waterbody) including the oil-free and oil-polluted vegetation in the VV and VH channel. This showed that the vegetation generally has a more distinct temporal backscattering pattern, suggesting the influence of seasonality in the backscatter strengths in both VV and VH channel. Most of the vegetation showed that backscatter is much higher during the summer – wet season than in the winter – dry season. The backscatter range of vegetation is often between -18 and -13dB in the VH channel. In addition, the backscatter range for bare land ranged between -11 and -12 for most part of the year, while waterbodies had the lowest and an uncommon backscatter range of -18 and -22 for most part of the year in the VH channel. Built-Up areas had the highest backscatter values in the VH channel, which ranged from -13 and -10dB for the entire year. These observation agree with previous research (Cable *et al.*, 2014).

However, in the VV channel (Figure 6-10b) backscatter values are much higher in strength than the VH channel. The reason for this is not farfetched, as several studies and investigations have posited that returns from co-polarized SAR backscatter feedback are usually higher than those of cross polarized backscatter (Mitchard *et al.*, 2011). Result obtained in this study show that bare land had much higher backscatter of between -6 and -3dB than most of the other features in the study area. It is also observed that the temporal VV backscatter for bare land is slightly influenced by seasonality, as backscatter strength slightly reduced during the winter months.

Waterbodies also had the lowest backscatter range of -18 and -16dB for most parts of the year, while built-up areas had optimal temporal VV backscatter ranging between -8 and -6dB. With respect to the various vegetation types, results from the temporal profile show that there is a distinct temporal characterization of most oil-free and oil-polluted vegetation, as temporal backscatter for the former tends to be meaningfully lower than the initial's values. However, oil-polluted TCA had a much higher temporal backscatter in comparison with the other oil-polluted vegetation. This certainly suggests a huge potential for the use of multi-temporal backscatter for terrestrial oil spill monitoring.



(a)



(b)

Figure 6-10: Mean temporal backscatter profile for the respective Polluted and Oil-free vegetation cover types, compared to other land cover types (i.e. built – up areas, bare land and waterbody). (a) VH Channel (b) VV Channel.

6.4.4 Relationship between Biophysical factors and Backscatter

6.4.4.1 MODIS NDVI

Results of the temporal MODIS NDVI for polluted and oil-free vegetation as presented in Figure 6-11 show that all of the six vegetation types (i.e. oil-free and polluted, cropland grassland and TCA) have similar temporal trends. In general, the mean retrieval of vegetation greenness is largely characterized by low NDVI values in the months of August and February across the different vegetation types, while high greenness values are recorded in the months of May and October.

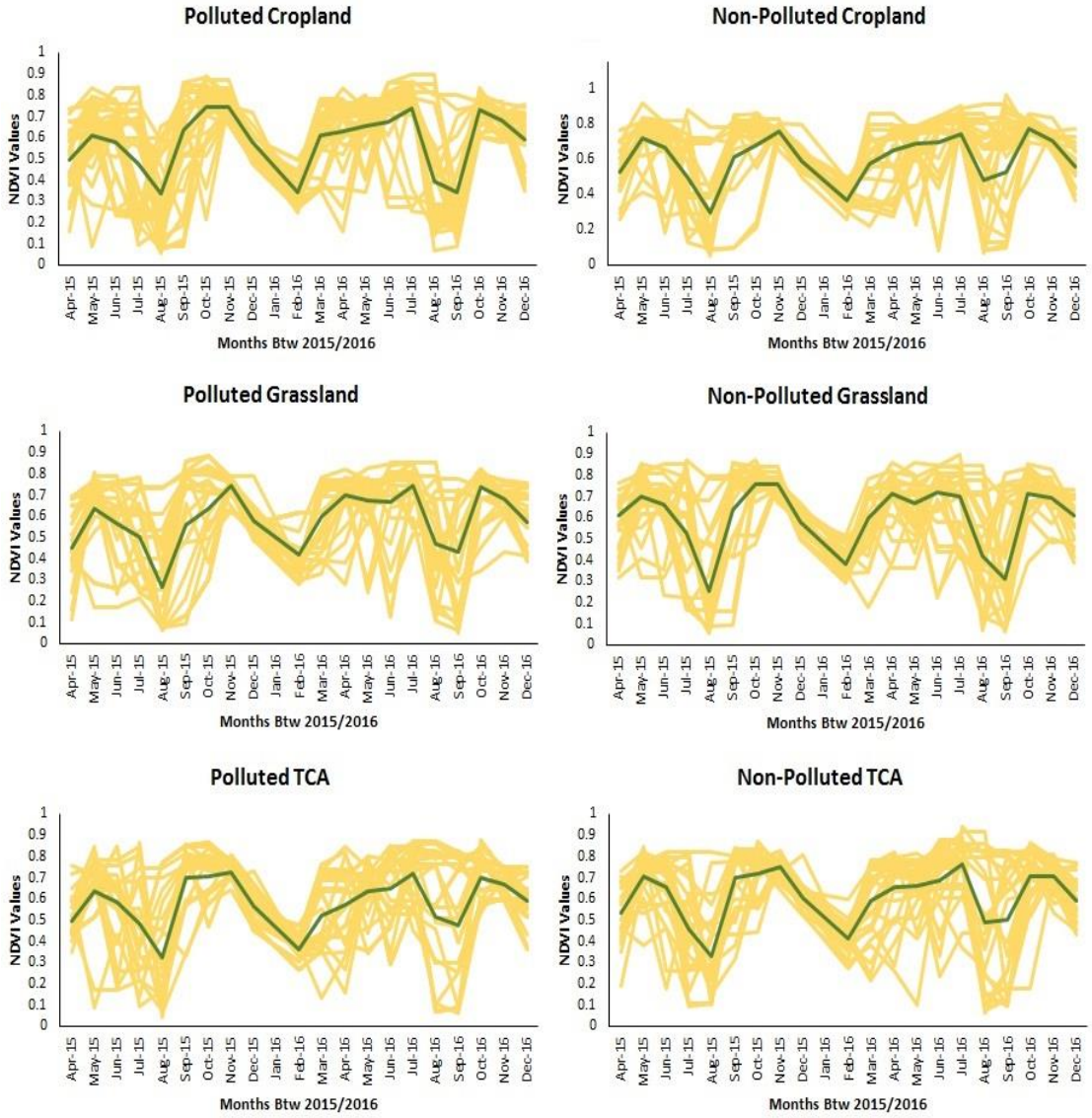


Figure 6-11: Line plots of retrieved MODIS NDVI for both polluted and oil free range of vegetation; the green coloured line represents the mean temporal value.

High mean NDVI values of up to 0.8 were often recorded in the peak vegetation greening month of October, while high values of up to 0.7 are recorded during the early growing summer month of May. Conversely, the lowest mean NDVI greenness values of 0.3 were often recorded in the peak of the rainy season months of August, compared to the much improved low values of 0.5 recorded at the peak of the winter (dry) season of February.

6.4.4.2 SMAP Soil Moisture

Results of the temporal soil moisture index as retrieved from the SMAP data are presented in Figure 6-12. The results show considerable differences from the temporal NDVI values. However, certain uniquely similar seasonal trends are peculiar between observed soil moisture and vegetation growth. In general, soil moisture values ranged from as low as $0m^3$ to $0.6m^3$ across the six vegetation types. However, the computed mean moisture index varied between $0.2m^3$ and $0.4m^3$ within the peak of the rainy (summer) season of July, August and September, while the lowest moisture index was often recorded in the peak of dry (winter) season of December through to February. Temporal trends showed that of the six vegetation types investigated, oil-free (non – polluted) cropland, grassland and TCA had better alignment with the seasonal rainfall trends of the Niger delta as observed by Adejuwon, (2012), where high rainfall are recorded in the months of July through to September and low rainfall dominates between the months of November through to February. This particularly explains why most of the observations (believed to be intense rainfall, flooding or runoff) are well situated within the rainy season of July through September, while the months of November through March are characteristically smoother.

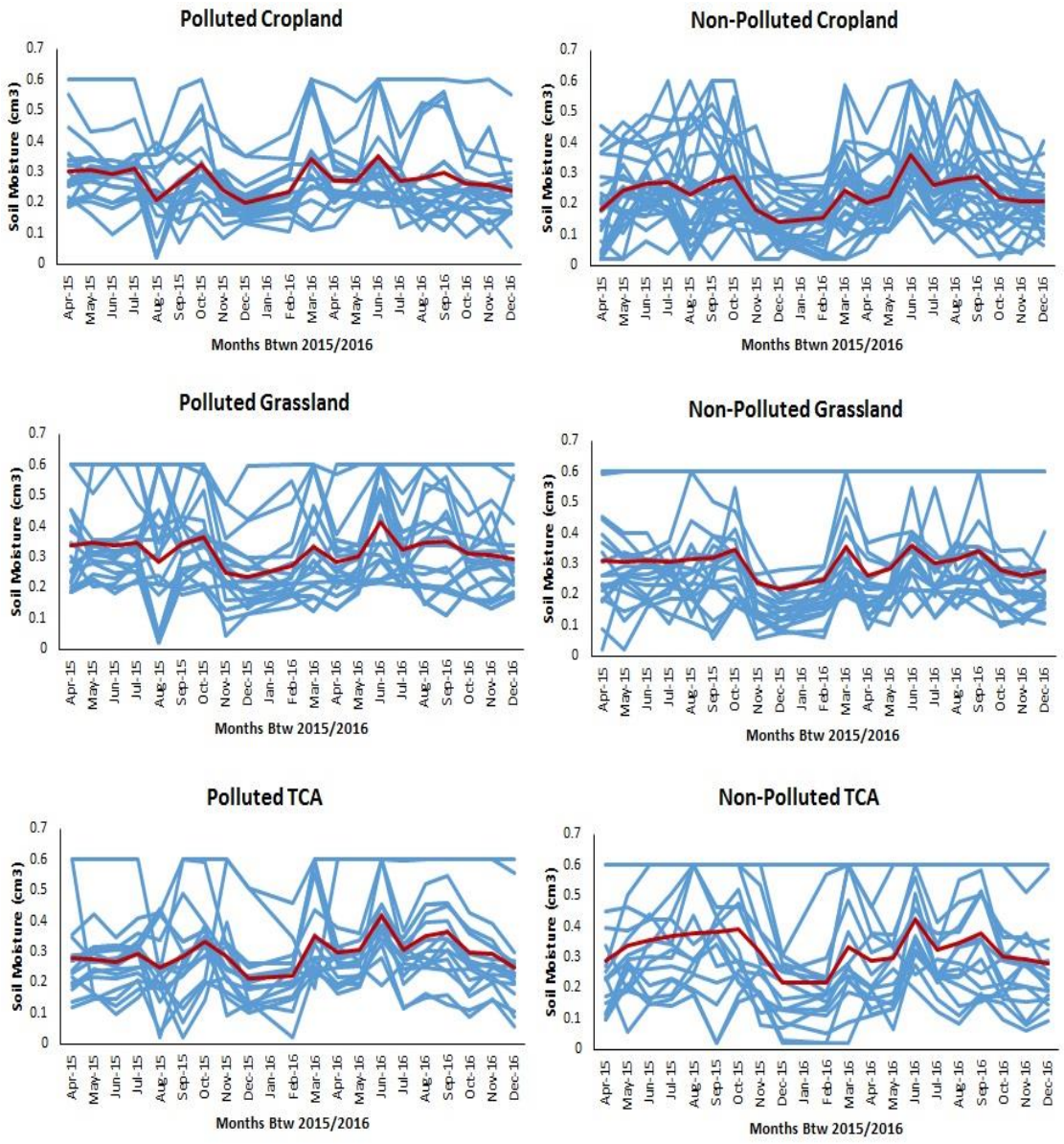


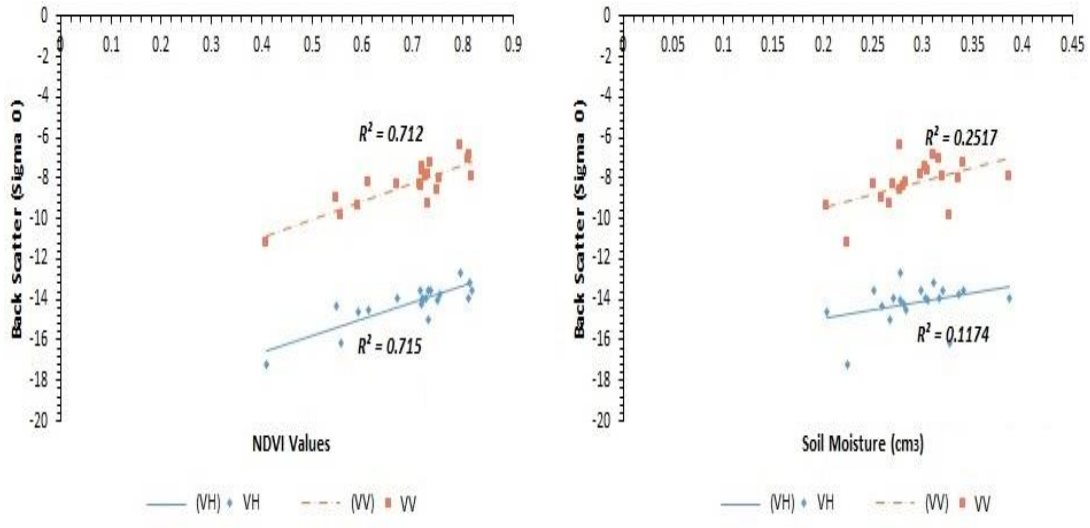
Figure 6-12: Line plots of retrieved SMAP soil moisture for both polluted and oil free range of vegetation; the brown coloured line represents the mean temporal value.

6.4.5 Relationship between Temporal NDVI, Soil Moisture and Backscatter

Figure 6-13 presents the result of the least square regression comparing the temporal VV and VH backscatter with corresponding NDVI and Soil moisture. This shows that temporal VV and VH backscatter from vegetation across the study area are in agreement with the temporal NDVI retrievals. This is as indicated by the high R^2 values of 0.715 and 0.712 obtained from the VH and VV backscatter respectively. The results also showed that a rather weak agreement exist between the backscatter and the SMAP soil moisture. In this case, the R^2 value from the VV channel (0.252) is observed to be much higher than those of the VH channel (0.117). In general, it is particularly

observed that the VH backscatter have a slightly and negligible higher R^2 value with the NDVI. This is however expected as most feedback from the cross polarization channel has a more shallow depth penetration and greater interaction with vegetation canopy, branches and trunks causing a double bounce scatter and consequently capturing much of the variability in the vegetation canopy.

Similarly, it is also observed that the R^2 value for the VV channel Vs soil moisture is slightly higher than the R^2 value (0.117) for the VH channel Vs soil moisture. The reason for this is also premised on the fact that most SAR co-polarization (i.e. VV and HH) returns have greater penetration into vegetation canopy with double bounce scatter, capturing much of the variability below the canopy layer and the consequent soil moisture condition.



(a) (b)

Figure 6-13: Mean Temporal Soil Moisture and MODIS NDVI Retrieved from study area vegetation modelled against Backscatter in VV and VH Channel (a) NDVI (b) Soil Moisture

6.4.6 RF and SVM Classification of Multi-temporal, Multi-frequency and Multi-seasonal SAR images

Table 6-5 shows the various multi-frequency L, C and X band SAR data combinations according to the season they were acquired. The dataset were integrated and formulated into a number of experimental image classification scenarios. Six different scenarios were implemented in this analysis: 3 wet seasons scenarios and 3 dry season scenarios. Scenario 1 and 2 are single (C-Band) frequency dry and wet season image

classification, while scenario 3 and 4 are dual frequency (L and C) band wet and dry season classification. The last two scenarios of 5 and 6 represents multiple frequency (L, C and X) band SAR dry and wet season classification. In addition, the wet season image classification scenario's had higher number of images as they represent the longest season (between March to October). While conversely, the dry season image scenarios had fewer images due to the short duration (i.e. between November to February) of the season. The overall classification accuracy from the six experimental image classification scenarios for cropland, grassland and TCA vegetation are shown in Table 6-6 below.

Table 6-5: Input data for the different image classification scenario's implemented

Classification Scenarios	Season	Datasets Used	Total No. of Images
Scenario – 1	Dry Season	Multi-temporal S1	12
Scenario – 2	Wet Season	Multi-temporal S1	30
Scenario – 3	Wet Season	Multi-temporal S1 + ALOS 2	34
Scenario – 4	Dry Season	Multi-temporal S1 + ALOS 2	14
Scenario – 5	Dry Season	Multi-temporal S1 + ALOS 2 + CSM	16
Scenario – 6	Wet Season	Multi-temporal S1 + ALOS 2 + TDX	36

Table 6-6: Shows the Overall Classification Accuracy (OA). The overall accuracy measures the correctness of the map class to the total ground truth used for validation.

	Cropland		Grassland		TCA	
Classification Operations	RF OA (%)	SVM OA (%)	RF OA (%)	SVM OA (%)	RF OA (%)	SVM OA (%)
Dry/Sc - 1	63.7	58.4	60.3	62.8	64.8	52.8
Wet/Sc - 2	71.7	67.3	60.3	62.8	67	61.5
Wet/Sc - 3	82.3	80.5	59	66.7	71	62.8
Dry/Sc - 4	63.7	64.6	59	65.4	60.4	52.8
Dry/Sc - 5	68.1	65.5	61.5	65	62.6	52.8
Wet/Sc - 6	71.7	74.3	59	61.5	74.7	67

Results indicates a high overall accuracy (OA) for cropland vegetation when the wet season Sentinel 1 (VV and VH) image stack are classified in scenario 2, compared to the low overall accuracy obtained when the dry season Sentinel 1 (VV and VH) image stack was classified in scenario 1. The best output for multi-temporal Sentinel – 1 (single frequency) image to discriminate between polluted and oil-free cropland vegetation was obtained with the random forest classifier in scenario 2. An overall accuracy of 71.7% was obtained, while an OA of 67.3% was obtained when the Support Vector Machine classifier was used for the same classification.

In grassland areas, results obtained from the multi-temporal Sentinel – 1 (single frequency) image classification (i.e. for both dry and wet season) were same across the different classifiers used. Overall accuracies of 62.8% OA was obtained for both the SVM and RF classifier outputs. In the TCA, the result obtained from the confusion matrix shows that the best OA result of 67% was obtained when the random forest classifier was used to classify the multi-temporal Sentinel – 1 (single frequency) image stack of scenario 2. An OA of 64.8% was also obtained when the dry season multi-temporal Sentinel – 1 (single frequency) image stack was classified for TCA. The best SVM result was obtained in the wet season stack classification with OA of 61.5%.

The integration of the L-Band ALOS PALSAR 2 with the single frequency multi-temporal Sentinel – 1 VV VH in the third (Sc3) and fourth (Sc4) scenarios led to substantial improvement in classification accuracy. In the wet season (scenario – 3), classification accuracy for cropland increased to 82.3% OA when the RF classifier was used for the classification. Similar improvement was recorded with the SVM classifier as OA increased from 67.3% to 80.5%. Result obtained for grassland vegetation classification also showed an improvement with the incorporation of the L – Band ALOS PALSAR 2 wet season images. The best OA result of 66.7% and 33.1% was obtained when the SVM classifier was used for the grassland classification, while a low OA of 59% was obtained with the RF classifier. In addition, the classification accuracy of TCA was also observed to have improved, as an OA result of 71% was obtained when the RF method was used to classify the multitemporal Sentinel – 1 and ALOS PALSAR (L and C band) wet season image stack.

In contrast, however, the integration of the L-Band ALOS PALSAR 2 dry season images did not improve classification accuracy as was observed with the wet season.

The best OA result of 64.6% was obtained for cropland when the SVM classifier was used to classify the multitemporal Sentinel – 1 and ALOS PALSAR 2 (L and C band) stack. For grassland, the best OA result of 65.4% was obtained with the SVM classifier; while for TCA 60.4% was obtained with the RF classifier. This represented a reduction of nearly 10% in overall accuracy when compared to the result of scenario – 3 where the wet season L – Band images were integrated with the wet season C – Band SAR.

Furthermore, attempt was made to integrate X - band SAR image from Cosmo Skymed and TanDEM – X into the Dry and Wet Season Sentinel – 1 and ALOS PALSAR combinations (previous scenarios). This only successfully improved classification accuracy for TCA in the wet summer season in scenario - 6. An OA result of 74.7% was obtained using the RF classifier, representing nearly 5% improvement from the initial 71% OA obtained in scenario 3. Additionally, cropland dry season image classification also saw an improvement by almost 5% following the incorporation of the Cosmo Skymed (X-Band SAR) images. An overall accuracy assessment result of 68.1% was obtained with the RF classifier, compared to the best result obtained in scenario 1 and 4 dry season (i.e. OA 63.7%). Figure 6-14, 6-15 and 6-16 shows the result of the various experimental image classifications.

In addition, the results from the McNemar test are presented in Table 6-7, 6-8 and 6-9, for cropland, grassland and TCA respectively. These tables were used to compare the performance of individual image combinations based on SAR frequency, the season undertaken and the classification methodology employed and against the number of samples correctly or incorrectly assigned to the respective polluted and oil-free vegetation ground reference samples.

Figure 6-14: Image Classification Result Using Random forest and Support Vector Machine for Cropland Vegetation

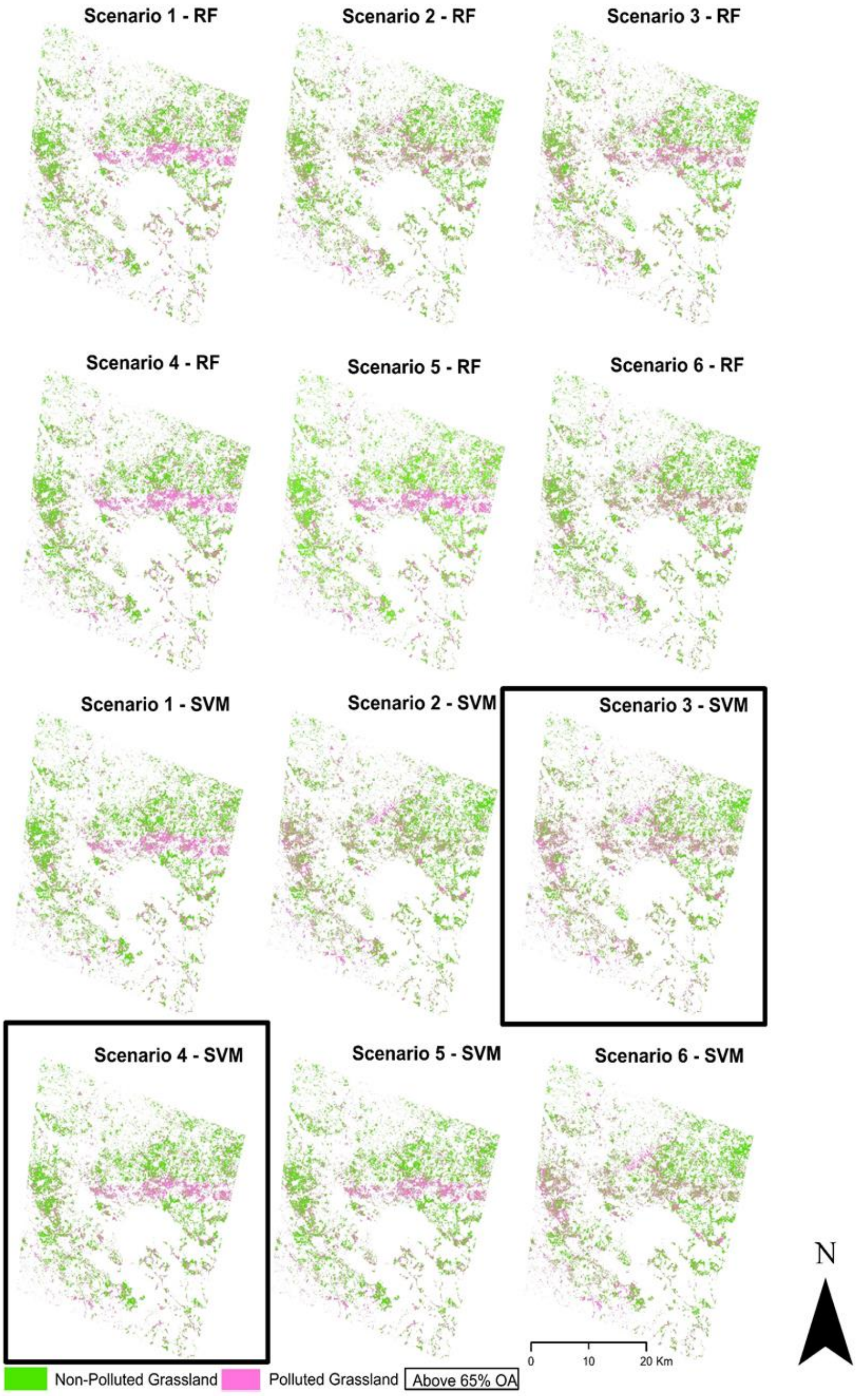


Figure 6-15: Image Classification Result Using Random forest and Support Vector Machine for Grassland Vegetation

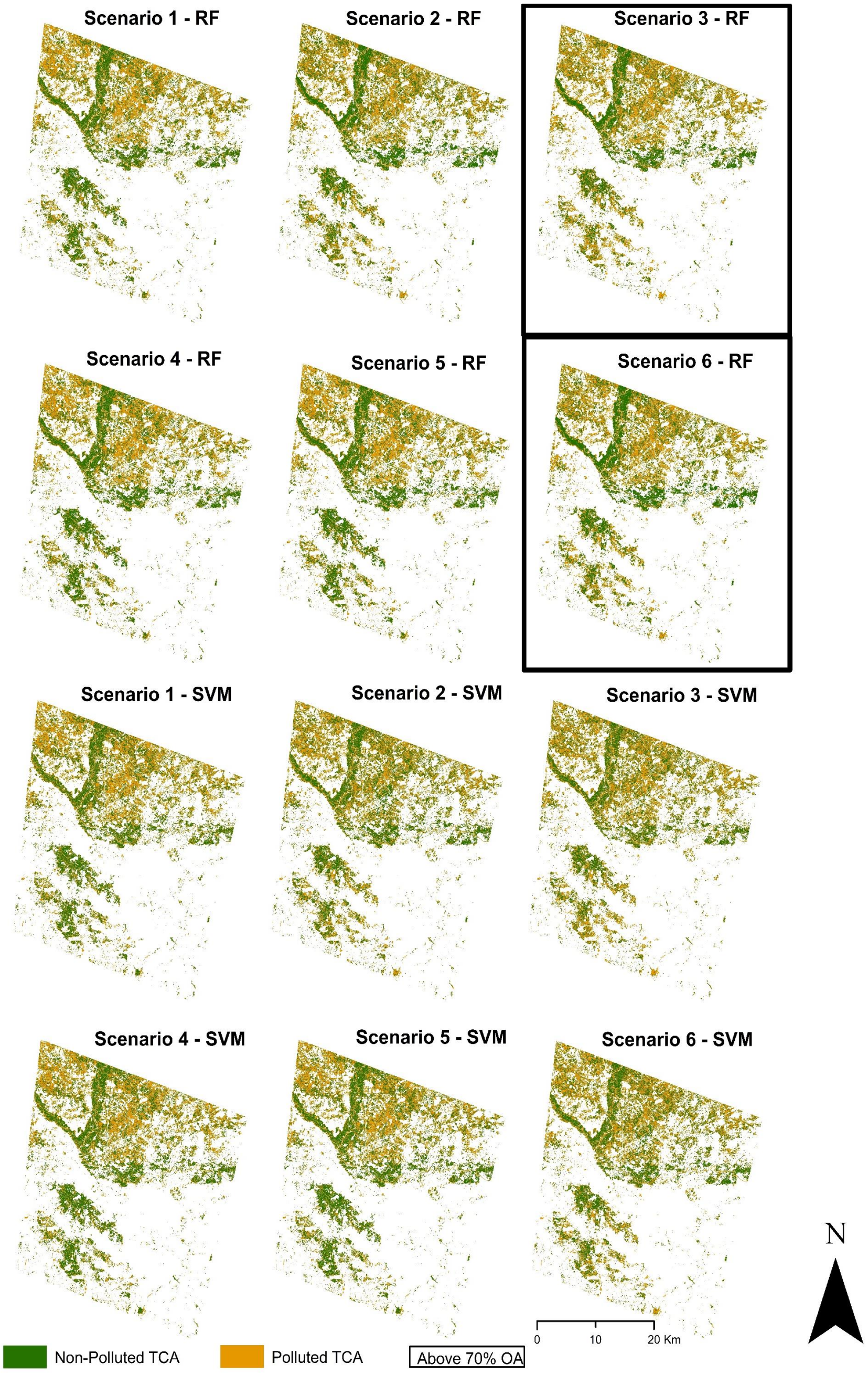


Figure 6-16: Image Classification Result Using Random forest and Support Vector Machine for Tree Cover Area vegetation

The statistical comparison of the corrected predict reference points against the wrongly assigned class reference in the various classification scenarios of the McNemar Test is presented in Table 6-7, 6-8 and 6-9.

This shows that the use of the wet season images with the Random forest classifier gave a better classification accuracy when compared with the result obtained from the dry season image classification using both the random forest and support vector machine.

In cropland areas, the result of the McNemar Test showed that the use of wet season multitemporal Sentinel – 1 VV VH and ALOS PALSAR 2 images (Sc2) improved overall classification accuracy and significantly outperformed the result obtained from classifying the Sentinel 1 VV VH (single frequency) dry season images. This gave the highest Z – value of 14.0 and 19.3 respectively when SVM and RF were used for the classification. Similarly, the use of the wet season multitemporal Sentinel – 1 and ALOS PALSAR 2 product significantly outperformed the dry season multi-frequency (L, C and X band) SAR stacked image classification for polluted and oil-free cropland discrimination. In addition, the use of the wet season multi-frequency (L, C, and X band) SAR stacked product using both RF and SVM significantly outperformed most of the dry season experimental scenarios in Scenario 1, 4 and 5 for cropland.

In grassland areas, the result of the McNemar test showed that the incorporation of the wet season L – Band ALOS PALSAR 2 and multitemporal Sentinel 1 images significantly outperformed most of the other classification scenarios. The most significant difference was the comparison between the aforementioned and the Dry season Sentinel 1 VV VH multi-temporal image pair (scenario 1) with Z – Value of 3.27. However, in general the result for grassland areas did not show as much variability as the cropland classification. This can be attributed to the general low performance of the classification scenarios as depicted by the overall classification accuracy.

In TCA, the result showed that the integrated use of wet season multi-frequency (L, C and X band) SAR data using the random forest classifier outperformed most of the other scenarios. It was observed that this scenario (Scenario – 6) outperformed most of the other scenarios by a greater margin, largely due to the 15% improvement recorded in the overall classification accuracy between the last scenario (scenario 6) tested against all other classification scenarios implemented.

Furthermore, the results have also shown that random forest classifier had the highest number of best results as evident from the overall classification accuracy and the McNemar test Z – score.

Table 6-7: McNemar Test of statistical significance between the various results comparing the Multi-temporal/Multi-frequency SAR Classification Using Random forest and Support Vector Machine in Cropland Areas

Code	SVM_Sc1	RF_Sc1	SVM_Sc2	RF_Sc2	SVM_Sc3	RF_Sc3	SVM_Sc4	RF_Sc4	SVM_Sc5	RF_Sc5	SVM_Sc6	RF_Sc6
SVM_Sc1	0	1.25	2.13	5.93	14.045	19.3	4	2.72	6.13	5.3	9	6.32
RF_Sc1		0	0.26	2.06	8.31	13.8	0	0.17	0.06	2.3	4.03	2.21
SVM_Sc2			0	0.76	7.84	8.3	0.12	0.03	0.03	0	3.1	1.7
RF_Sc2				0	1.94	0.03	1.75	1.88	1.33	0.38	0.31	0.17
SVM_Sc3					0	0.06	8.5	6.6	7.3	4.5	1.44	2.533
RF_Sc3						0	13	11.17	11.17	8.7	2.37	5.04
SVM_Sc4							0	0	0	0.267	3.23	1.88
RF_Sc4								0	0.07	1.33	3.12	1.57
SVM_Sc5									0	0.27	2.9	1.44
RF_Sc5										0	0.64	0.41
SVM_Sc6											0	0.31
RF_Sc6												0

Table 6-8: McNemar Test of statistical significance between the various results comparing the Multi-temporal/Multi-frequency SAR Classification Using Random forest and Support Vector Machine in Grassland Areas

Code	SVM_Sc1	RF_Sc1	SVM_Sc2	RF_Sc2	SVM_Sc3	RF_Sc3	SVM_Sc4	RF_Sc4	SVM_Sc5	RF_Sc5	SVM_Sc6	RF_Sc6
SVM_Sc1	0	0.05	0.3	0.32	2.8	0.16	1.53	0.16	1.16	0.64	3.2	0.24
RF_Sc1		0	0.32	0.03	2.12	0	0.83	0.04	0.55	0.16	0.31	0
SVM_Sc2			0	0.07	0.13	0.15	0.07	0.267	0	0	0	0.15
RF_Sc2				0	0.6	0	0.75	0	0.36	0	0	0
SVM_Sc3					0	3.27	0	1.04	0.35	0.35	0.45	3.06
RF_Sc3						0	0.55	0.45	0.32	0	0.05	0.83
SVM_Sc4							0	1.23	0	0.36	0.14	0.6
RF_Sc4								0	1.46	0.25	0.19	0
SVM_Sc5									0	0.1	0.036	0.35
RF_Sc5										0	0	0.05
SVM_Sc6											0	0.07
RF_Sc6												0

Table 6-9: McNemar Test of statistical significance between the various results comparing the Multi-temporal/Multi-frequency SAR Classification Using Random forest and Support Vector Machine in Tree Cover Areas

Code	SVM_Sc1	RF_Sc1	SVM_Sc2	RF_Sc2	SVM_Sc3	RF_Sc3	SVM_Sc4	RF_Sc4	SVM_Sc5	RF_Sc5	SVM_Sc6	RF_Sc6
SVM_Sc1	0	5.3	1.17	2.75	0.02	0.02	0.25	2.4	0.125	3.76	3.51	7.9
RF_Sc1		0	0.11	0.03	1.02	1.31	5.26	0.9	5.88	0.13	0.03	2.06
SVM_Sc2			0	1.23	0.88	0.66	1.225	0	2.75	0.103	1.45	6.72
RF_Sc2				0	2.63	3.51	3.51	0.88	3.51	0.63	0	0
SVM_Sc3					0	2.12	0	0.57	0	0.55	3.18	7.54
RF_Sc3						0	0.085	0.76	0.02	0.74	2.7	8.31
SVM_Sc4							0	2.12	0.17	3.77	3.51	9.5
RF_Sc4								0	2.12	0.5	0.66	3.89
SVM_Sc5									0	4.27	3.51	4.45
RF_Sc5										0	0.43	2.86
SVM_Sc6											0	2.12
RF_Sc6												0

6.5 Discussion

This Chapter set out to investigate the potential of multi-temporal backscatter analysis to infer the primary response of vegetation (based on SAR backscatter sample mean difference) affected by oil pollution. Secondly, the chapter inferred secondary attributes to determine the spatial extent of the oil-free and polluted vegetation through a random forest and support vector machine classification, with particular emphasis on seasonal difference. The results show that retrieved temporal backscatter profiles generated from the 2015 – 2016 images using oil-free and polluted sample points for cropland, grassland and TCA vegetation indicate that vegetation exhibits similar temporal behaviors in terms of response to seasonal effects.

Similarly, the inter-quartile range and median values for oil-free vegetation were much higher across temporal scale and evolution when compared to the polluted vegetation. However, mean descriptors for cropland and grassland vegetation were largely similar, as they both depicted higher separability in the wet (summer) season than the dry (winter) season. In addition, results of the paired sample t-test showed that the difference was statistically significant ($P<0.05$) for cropland and grassland.

In contrast, results obtained for TCA showed little separability between oil-free and polluted TCA vegetation in both wet and dry season in the VH channel. However, during the wet season the VV channel separability was observed to be marginal as *P* value of 0.04 was obtained. A possible reason for this trend might be the fact that both cropland and grassland vegetation are mostly shrub vegetation types, where water stress and senescence during the dry (winter) season can profoundly affect their structural condition (both polluted and the oil-free components) making them exhibit similar structural and biochemical deficiencies to a typical oil-polluted site. This increases the likelihood of lower backscatter return in the dry season. In contrast, in the dense woody TCA vegetation, they are predominantly deciduous broad-leafed vegetation types and hence little separability is observed especially in the VH channel.

The general seasonal traits observed in this study is similar and agrees with results obtained in a number of research (Betbeder *et al.*, 2014; Gao *et al.*, 2018; Laurin *et al.*, 2018; Rüetschi *et al.*, 2017; Vreugdenhil *et al.*, 2018) using Sentinel 1 data. Additionally, studies carried out by Osunmadewa *et al.* (2018) in the spatio-temporal analysis of vegetation phenology using AVHRR NDVI in the southern part of Nigeria also showed

that annual rainfall trend coincided effectively with vegetation growth pattern. As such, this is most likely to be responsible for the observed temporal backscatter variations.

Furthermore, results of the retrieved mean temporal values for the respective oil-free and polluted vegetation were compared to the mean temporal values of other features, which include waterbody, built-up areas and bare/exposed surfaces. The results obtained demonstrated contemporary similar trends as observed in Cable *et al.*, (2014), where backscatter return from water are usually lower owing to higher dispersion of incident energy as a result of the smooth surface, while built – up areas and exposed surfaces have similar characteristic higher backscatter returns in both the co (VV) and cross (VH) polarized channels. This is mostly owing to the diffuse scattering surfaces of edges of buildings and terrain leading to a double bounce and smooth surface specular reflectance respectively.

In terms of assessment of bio-physical assessment of vegetation attributes as deduced from MODIS NDVI and SMAP soil moisture index, the results obtained reinforce the notion that the temporal response of vegetation as obtained from the NDVI and soil moisture are very much sensitive to seasonal influence. Major low and high vegetation health and growth peaks coincided strongly with peak rainfall months (in the month of August) and senescence in the winter (dry season) month of February. Further assessment showed that lowest mean NDVI was observed in the month of August (rainfall peak period) compared to the mean observed in the winter period. This certainly suggests that factors such as intense rainfall and concomitant flooding events in most vegetated areas, especially within the sampled location could be responsible.

Results obtained from the temporal soil moisture retrieval also reaffirms this position, as more unstable soil moisture variation and feedback are predominant in the rainy season, compared to the much smoother feedback in the dry season. In addition, Ologunorisa and Adeyemo (2005) have observed that most flooding incidents experienced in the country are mostly between the months of July through September, most of which occur around states bordering the Niger – Benue River (Ologunorisa, 2004) and the Niger delta region where the current study has been implemented.

Comparison of SAR based temporal trends with an optical dataset is one of the most effective ways of ascertaining the validity of results obtained. This has been demonstrated in several studies (e.g. Gao *et al.*, 2018; Gao *et al.*, 2017; Laurin *et al.*, 2018; Lussem *et al.*, 2016; Mansaray *et al.*, 2017). The R^2 result of the modelled relationship between the temporal NDVI and VV – VH backscatter is an indication that observed temporal backscatter trends and analysis carried out in this study for polluted and oil-free vegetation are in agreement.

With respect to the multi-temporal, multi-seasonal and multi-frequency image classification, the research has demonstrated that the use of wet season multi-frequency SAR images for classification and discrimination of oil-free and polluted vegetation gave the best result. Temporal seasonal assessment showed that there is statistically significant difference between retrieved backscatter for oil-free and polluted vegetation in the wet season compared to the insignificant difference observed in the dry season. Mean temporal backscatter for oil-free cropland and grassland were always higher than the backscatter for polluted cropland and grassland vegetation. Similarly, the wet-summer season is more suitable for plant growth and greenness especially in the mangrove ecosystem. This is largely owing to increased precipitation events and concomitant increase in soil moisture, which inadvertently leads to increased vegetation net primary productivity.

It is thus, expected that healthy uncontaminated vegetation would record peak productivity, more structural viability and greenness during the wet season than a typical contaminated vegetation, allowing for better discrimination accuracy. In the dry season, in contrast, productivity is expected to be greatly diminished in both healthy and contaminated vegetation samples owing to the effect of water stress and unfavorable temperature, which may lead to increased senescence, reduced greenness and defoliation in both healthy and unhealthy vegetation. The effect of this is that both healthy (oil-free vegetation) and contaminated vegetation (as a result of oil pollution) exhibits similar structural and spectral characteristics leading to poor separability in the image classification.

Differentiation and classification accuracy for cropland vegetation improved significantly with the incorporation of L Band SAR together with the multi-temporal Sentinel – 1 (C Band SAR) for the wet season, while in grassland and TCA vegetation the incorporation

of multifrequency (L and X Band) SAR led to significant improvement in classification accuracy. The results also supports findings from Numbisi *et al.* (2019) where the use of multi – date seasonal SAR and textural measures offered the highest classification accuracy.

In terms of the performance of the classifier used in this chapter, results have shown that RF outperformed SVM in most of the scenarios implemented. This is evident from the McNemar test implemented, further reinforcing the position of the mean seasonal backscatter difference, classification operation (multi-seasonal/multi-frequency classification) and more specifically the classifier performance implemented in this study. High Z – values were mostly recorded when the combined multi-frequency wet season, all three SAR band types and random forest classification was used in the course of the classification process.

The discrimination between polluted and oil-free cropland vegetation recorded the most significant highest OA (82.3%) when the multi-temporal/multi-frequency wet season images (Scenario – 3) were classified using the random forest method. In contrast, the lowest OA (58.4%) was obtained when the SVM classifier was used to classify the dry season multi-temporal Sentinel – 1 images (Scenario – 1). This was followed by TCA, which also recorded a significant increase in OA (74.7%) when the combined multi-temporal, multi-frequency wet season (Scenario – 6) images were classified using the RF method. Compared to the low OA (60.4%) when the dry season multi-temporal Sentinel – 1 and ALOS PALSAR 2 (Scenario – 4) was classified. Overall accuracy assessment results obtained for cropland area classification in the wet season were higher than results obtained in the 2nd and 3rd analysis chapters (chapter four and chapter five) and certainly better than results obtained in other similar studies (e.g. Bianchi *et al.*, 1995a; Mahdianpari *et al.*, 2018; Hese and Schmullius, 2009; Van der Werff *et al.*, 2007).

However, in grassland areas, the most significant highest classification OA (66.7%) was obtained when the wet season multi-temporal Sentinel – 1 and ALOS PALSAR 2 images (Scenario – 3) was classified using the SVM. The lowest OA (59%) was obtained when the same dataset was classified with the Random forest classifier, suggesting high influence of methodological performance in improving the result over the influence of the various input data used.

6.6 Summary and Main Findings

This chapter implemented a novel study of assessing multi-temporal backscatter from Sentinel-1 SAR data analysis to infer the primary response of vegetation affected by oil pollution. Secondly, the chapter also saw the implementation of multitemporal, multifrequency and multiseasonal SAR image classification to infer and determine the spatial extent of the oil-free and polluted vegetation using both random forest and support vector machine (SVM) methods. In general, the chapter assessed the effect and influence of seasonality in the detection and mapping of oil-polluted vegetation, results of which provided new perspectives and fundamental basis and reasons for low discrimination and classification accuracy.

Results of the temporal backscatter profiles show that oil-free and oil-polluted vegetation exhibits similar temporal behaviors in terms of response to seasonality. Cropland and grassland vegetation had higher separability in the wet (summer) season than in the dry (winter) season, while TCA showed little separability in both wet and dry season. Furthermore, results from the comparison of mean temporal backscatter values for the respective oil-free and polluted vegetation with other features (i.e. waterbody, built-up areas and bare/exposed surfaces) showed good agreement with previous literatures, an indication that result obtained in this study were valid. Effort to further validate the results obtained in this study through the comparison of bio-physical factors (as deduced from MODIS NDVI and SMAP soil moisture index) showed that temporal response of vegetation as obtained from the NDVI and soil moisture are very much sensitive to influence of seasonality as with the results obtained from the multitemporal backscatter. Major lows and high vegetation health index values as well as vegetation growth peaks coincided strongly with backscatter trends and mildly with soil moisture.

In terms of the multi-temporal, multi-seasonal and multi-frequency image classification, results obtained showed that the implementation of the image classification process in the wet season using multi-frequency L and C band SAR images provides a better result tha when this is implemented in the dry season. This reinforced the position of earlier findings in this research where discriminability is observed to be better in the wet season than the dry season, owing to significant backscatter difference in the wet season than in the dry season especially for cropland and grassland vegetation. Discrimination and classification accuracy for cropland vegetation improved significantly with the integration of L-band SAR together with C Band SAR for the wet season, while in grassland and TCA

vegetation the incorporation of multifrequency L, C and X Band SAR led to significant improvement in classification accuracy. This generally gave the best overall classification accuracy of 82.3% for cropland 66.7% for grassland and 75% for TCA vegetation.

Chapter 7 : Summary, Conclusion, Limitations and Possibilities for Future Research

7.1 Introduction

This section discusses and summarizes the results obtained in the analysis Chapters (i.e. Chapters 4, 5 and 6), which focused on evaluating the potential of optical variables, integrated optical – SAR variables and, multi-temporal – multi-frequency SAR respectively, with machine learning (Random forest, Fuzzy forest and Support Vector Machines) methods. This were used to discriminate between oil-polluted and oil-free cropland, grassland and tree cover areas. The chapter highlights the important findings that emanated in the course of this research, stressing the contribution to the existing body of knowledge within the specific aspect of terrestrial oil spill remote sensing. The chapter also highlights the summary findings related to the various research questions, the limitations encountered in the course of this research and finally areas of possible future work.

7.2 Summary Discussion

This research explored the capability of a range of remote sensing data and new methodologies for detecting terrestrial oil pollution. It was evident from Chapter 4 that the discrimination of oil-polluted and oil-free vegetation types yielded better results when vegetation types are assessed separately as opposed to when entire images with all land cover features are assessed at once. This formed the rationale for the subsequent use of this image processing technique (i.e. at the micro – scale level of the different vegetation types) for the optical – SAR variable integration (using FF and RF) and the multi-temporal – multi-frequency SAR (using RF and SVM) in Chapters 5 and 6 respectively. Similarly, results obtained from the field validation work suggested that the differences in spectral reflectance of polluted and oil-free vegetation were larger in periods when vegetation is in the best condition, rather than in periods when vegetation is under stress. This specifically suggested the implementation of the multi-seasonal image analysis, to investigate the potential role of seasonality in the discrimination of polluted and oil-free vegetation. Specific outcomes from the research are discussed as follows:

- **Overcoming the limitation with Multispectral Optical Sensors**

Using multispectral optical data for the detection and mapping of oil-impacted vegetation canopy has been implemented in several studies. While most of these studies represent small-scale application of oil spill detection and mapping, there are also challenge of spectral mis-match among features. This occurred as a result of the characteristic nature of absorption features within the shortwave infrared spectral wavelength as mentioned by Cloutis (1989) and Hörig *et al.* (2001). This issue has characterized most oil spill detection studies that have had relatively little success, especially in cases where multispectral images are used. Bianchi *et al.* (1995a) used airborne hyperspectral images from the Multispectral Infrared Visible Imaging Spectrometer (MIVIS), hyperspectral unmixing methods and oil fractional areas for water, woods, cultivated field, smooth surface soil, grooved surface soil and rice fields classes were used as end-members in the image analysis process. Similarly, Achard *et al.* (2018) explored several un-supervised classifications including an original method of classification combining an unmixing approach and SVM on digitally acquired hyperspectral images. They noted the difficulty associated with comparing unsupervised classification results and assigning class members to potential features; which links to the problem of handling multiple image features. They also reported that the water feature (class) in the maximum abundance classification had several displaced pixels owing to the threshold limit.

In addition, Mahdianpari *et al.* (2018) also reported confusion among successional classes of low and moderate contaminated sites and between moderate and high contamination site from their level – 2 classification to characterize polluted land into different classes using random forest and electromagnetic induction data. The reason for confusion in the level – 2 classification to discern the intensity levels of polluted land was however not stated in the study, but evidence from the base map suggests that features such as trees, herbaceous vegetation, man-made path and water were present within the polluted land, thus possibly accounting for the minuet scale confusion. Hese and Schmullius (2009) particularly noted that the spectral characteristics of oil spills complicated the analysis in their study because specific spectral signatures of oil-polluted features are difficult to create. This according to the authors is because crude oil has the spectral properties of a complete absorber just like features such as waterbodies and bare surface in the VIS/NIR range of the spectrum. Another source of spectral confusion according to the authors has to do with the complexity created by additional oil spill events on areas that have already

been contaminated earlier. This study, which relied mostly on Landsat images and object-based classification sought to discriminate between oil free vegetation, oil contaminated vegetation, soil and industrial land use areas. Results showed low discrimination accuracy primarily because of limited capabilities in separating oil-polluted areas from other features with similar spectral characteristics (such as soil reflections and regenerating vegetation signatures) using the Landsat TM 5 imagery.

This research employed novel image processing techniques and explored the potential of wide array of datasets (from visible through to the microwave) part of the spectrum to overcome the mix-pixel problem. In the use of optical reflectance data (i.e. in the Chapter four) in the course of this research, machine learning random forest and assessment of specific vegetation types were introduced which have not been previously experimented prior to this research. The results showed that the new approach can reasonably resolve the discrimination issues between oil-impacted vegetation and other feature types, since image analysis is concentrated on specific feature (vegetation) types within the image space. This effectively reduces the effect of spectral confusion among features. In addition, the use of machine learning methods benefits from an increased learning rate if different thematic feature layers are introduced compared to when only restricted few input variables are explored in the discrimination process.

This thesis implemented this approach for monitoring the impact of oil on vegetation, through the exploration of optical – spectral reflectance data within the VIS-NIR-SWIR and various vegetation health indices using the random forest method. The use of these variables together with random forest methods for different vegetation types yielded significant improvement in the discrimination accuracy of end member classes compared to results obtained in previous studies (e.g. Hese and Schmullius, 2009; Van der Werff *et al.*, 2007 and Bianchi *et al.*, 1995a).

- **Overcoming the limitation with Hyperspectral Sensors**

The use of hyperspectral images should reasonably resolve and overcome the problem of spectral confusion among features in the detection and mapping of oil polluted areas. However, one particular challenge in the use of both multispectral and hyperspectral sensor for terrestrial oil spill mapping is the lack of sufficient cloud free optical images to limited capability in validating results obtained. Adamu *et al.*, (2018) and Ayanlade, (2015) had reported that persistent cloud cover in the Niger Delta especially in the wet

season had restricted their image processing to dry season analysis. Similarly, Arellano *et al.*, (2015) also reported that analysis of hyperspectral Hyperion image used for detecting oil pollution impact in the Amazonian rain forest was also restricted to dry season image of February 2005 owing to cloudy weather condition preventing new acquisition.

This research sought to overcome this challenge by first exploring the potential of integrating optical, SAR and geo-physical variables in detecting oil pollution within different vegetation canopy types and further mapping the corresponding extent of the vegetation types. Secondly, SAR only image analysis was also implemented to counteract the challenges associated with cloud cover affecting effective use of spaceborne hyperspectral and multispectral images especially in data scarce region. Results provided new perspectives, insights into terrestrial oil spill monitoring than were previously known as the implementation of multi-temporal, and multi-frequency SAR image analysis for both dry and wet seasons revealed astonishing trends in a typical oil-polluted sites. It was specifically discovered that the wet season is the best time to implement detection protocols of oil pollution. Result obtained for the wet season image analysis yielded the highest classification accuracy in this research. Results also obtained from the exploration of multitemporal C-SAR images to establish vegetation temporal characteristic within a typical polluted and oil-free site further corroborated the classification result, temporal results demonstrated that characterization, and discrimination of impacted and oil-free vegetation in the wet season offers better results, hence the best period for carrying out activities as fieldwork and data analysis.

- **Implication of the findings and Contribution to Wider Knowledge**

The results obtained in this research have shown that class specific vegetation assessment can help resolve the problem of mis-classification among features and has proved very useful in this study to improve classification accuracy. The use of SAR images for terrestrial oil spill detection in itself have seen very limited application previously. Challenges hindering the effective deployment and use of SAR images are the restrictions of accessibility to the imagery, huge memory requirement and the time required for processing multiple SAR image scenes. In addition, signal responses from SAR images are usually measured in terms of backscatter (Decibel) return strength, which does not

necessarily depict the changes in bio-chemical processes of plant ecology. Rather, it relies on the physiognomic and structural changes in vegetation canopy composition.

Results obtained from the operational integration of multi-temporal, multi-frequency and multi-seasonal SAR from various sensor as demonstrated in this research (Chapter six) yielded the highest overall accuracy than any other data combination (Optical Only: Chapter Four and Optical – SAR Integration: Chapter five) implemented in the course of this research. Results obtained were far better than the results obtained in other similar studies (e.g. Bianchi *et al.*, 1995a; Mahdianpari *et al.*, 2018; Van der Werff *et al.*, 2007). The singular inclusion of wet season L – Band SAR together with the wet season C and X band SAR led to a significant improvement in the classification accuracies. Thus, it is likely that L – Band SAR has a better performance in detecting hydrocarbon-induced stress as a result of the high accuracy obtained for cropland vegetation (OA of 82%) in this research.

Another major important finding and novelty from this research is the applicability of multi-temporal SAR backscatter data to explore vegetation temporal characteristics in oil-affected areas. Evidence from the results obtained through statistical test showed that better detectability is more assured during the wet (summer) season than the dry (winter) season. Possible reasons are that in the winter (dry) season, there is generally low precipitation and soil moisture in the Niger Delta mangrove region. This tends to induce negative vegetation feedback such as stress, senescence and defoliation to the entire vegetation within the study area (i.e. both oiled and oil-free), thereby reducing detectability and discrimination accuracy. In contrast, in the summer (wet) season, precipitation and soil moisture are usually high, which ultimately presents good conditions for oil-free vegetation to thrive (in both biochemical processes and structural development), especially in sparser vegetation, while the oil-impacted vegetation typically exhibits limited growth as a result of the impact of hydrocarbon in the wet season.

The methods and approach used in this research with specific reference to land cover specific image processing and the machine learning methodologies (random forest, fuzzy forest and support vector machines) is another novel contribution to existing knowledge space. This data and methodological integration approach holds huge potential for monitoring leakages in oil and gas transportation facilities, hydrocarbon micro and macro

seepage detection and estimating the general area impacted by oil pollution. The approach provides a basis for further developing an early warning detection system for pipeline monitoring, especially when a staggered processing chain is followed to reduce the influence of spectral mixing as machine learning methods in this case offers flexibility in using wide array of useful thematic layers for both detection, training (calibration phase) and classification. In addition, the results of this research present a more robust, cost efficient and environmentally friendly approach for pipeline monitoring when compared to field inspection and other intrusive methods, owing both to the revisit capabilities of remote sensing satellite platforms to an area of interest and recent improved spectral, spatial and temporal resolutions. This provides a unique opportunity for monitoring changes along pipeline right of ways, oil production facilities and oil well locations, especially in areas with limited accessibility and intermittent cloud cover.

Furthermore, another important novelty of this research is the introduction and use of fuzzy forest method. Certainly, the remote sensing community would profit immensely from the application and use of this methodology in other relevant studies where the influence of high dimensionality is critical owing to use of high number of input variables. This research represents the first application of this methodology in the processing and analysis of remote sensing earth observation satellite images, and also certainly the first application in the discrimination and mapping of terrestrial oil pollution. Results obtained in this research demonstrated quite clearly that the method can effectively reduce variable bias through the arbitrary elimination of non-performing variables in the classification process, thereby improving classification accuracy (as was obtained for grassland and TCA vegetation types). However, as mentioned in the concluding parts of section 5.5, reduced variables may not always translate into improve classification accuracy, hence the reason for the poor performance of the FF method in cropland vegetation and the entire MSIF image classification.

On the other hand, Fuzzy forest offers a unique ability of resizing, reshaping and reducing the number of input variables where there is high number of input features. Thus, can be operationally prudent when computational efficiency is lacking especially in data scarce regions. The Fuzzy forest approach can replace operationally intensive procedures associated with such data preprocessing as it provide one stop solution for variable elimination based on machine learning random forest, thereby improving reliability in variable importance.

7.3 Conclusion

In conclusion, this research was able to answer and fulfill all of the research questions and objectives it set out to achieve. Based on the result presented in the analysis chapters of this thesis and the follow on discussion, the following concluding points are further noted based on the research question:

I. To what extent can **multispectral optical image only, vegetation health indices and land cover specific classification** be used to detect, map and discriminate between oil-free and oil impacted vegetation types **with machine learning classifier**?

Results obtained from the first research objective (**Chapter Four**) on the utilization of Landsat – 8 spectral bands and derived vegetation health indices in a machine learning random forest classifier, showed that this procedure can be very cumbersome especially when there are ambiguities among features in the image space. This was characterized by low classification accuracy caused by spectral confusion amongst the diverse features present in the image space. Similar observation were also made by Ayanlade, (2015) and Hese and Schmullius, (2008; 2009) and other previously mentioned studies where attempts were made to map oil-polluted areas on terrestrial landscape without specific recourse to land cover specific classification. However, following the implementation of the image classification on vegetation specific extents, there was generally improved accuracy in the result obtained (as shown in table 4.4). Classification accuracy significantly improved from 45.45%, 33.33% and 28.5% OA when the entire study area image was classified to 60.61%, 65% and 70% for cropland, grassland and tree cover areas respectively, representing significant improvement in accuracy. The improved accuracies recorded in this chapter were also better than those reported in previous studies (Bianchi *et al.* 1995b; Van der Werff *et al.*, 2007).

II. To what extent can the **integration of optical derived variables (i.e. multispectral bands and vegetation indices), multi-frequency SAR and geo-physical variables** improve the classification accuracy of polluted and oil-free vegetation types?

This research question was answered in **Chapter Five** of this thesis. The focus here was to improve the overall classification accuracy and further subdue the effect of spectral confusion between oil-impacted and oil-free vegetation stands. The incorporation of SAR derived backscatter images (from C and X band SAR), interferometric coherence and other geo-physical variables was necessary for this purpose. In addition, machine learning

fuzzy forest classifier was introduced to further resolve high dimensionality in multiple variable inputs. This was achieved by specific selection of only high performing biophysical, bio-chemical and geophysical variables to improve the discrimination further between oil-free and oil-impacted vegetation. This is also as a result of the fact that the random forest has a limitation of inability to effectively handle the effect of high dimensionality and multi-collinearity as observed by Nicodemus and Malley, (2009) and Strobl *et al.*, (2008).

In this chapter, 37 different variables consisting of SAR derived backscatter from TanDEM X, Sentinel 1 and Cosmo Skymed images, interferometric coherence and textural variables, Sentinel 2 spectral bands, NDWI and LAI, soil type and geology were integrated and classified to map oil-polluted vegetation. Results showed that the exploration of other datasets and integration of optical and SAR variables gave an improved classification accuracy compared to when only optical image derived variables (chapter 4) are used for classification. The reason for this improvement is because the integration of vegetation biochemical response (optical variables) and vegetation bio-physical responses (SAR) (Joshi *et al.*, 2016; van Beijma *et al.*, 2014) have greater potential in reducing the effects of over generalization, spectral confusion and thus, an improved classification accuracy, than when only either bio-chemical and bio-physical indicators are assessed separately.

In addition, the fuzzy forest method achieved a good selection of the most important non – correlated variables for classification, especially for the TCA and grassland vegetated areas. The result of the confusion matrix, high-resolution image validation and field toxicology analysis showed that fuzzy forest variable reduction is able to improve classification accuracy for TCA. This represented almost a 10% increment in the overall accuracy's recorded in Chapter 4. The Random forest approach also had a superior performance, which was significantly different from result obtained with Fuzzy forest for cropland vegetation when the entire 37 image variables were used.

III. To what extent can **SAR only (i.e. multi-temporal and multi-frequency SAR) variables improve discrimination and classification accuracy (between oil-free and oil-impacted vegetation)** using a machine learning.

This research question was answered in **Chapter Six** of this thesis, where the potential of multi-temporal Sentinel – 1 backscatter in depicting temporal characteristics for impacted and oil-free vegetation was investigated. The potential influence and effect of seasonality in detecting and discriminating oil-impacted vegetation from oil-free healthy vegetation was also explored using multi-temporal – multi-frequency L, C and X band SAR classification. This became necessary as the field spectral reflectance of polluted and oil-free vegetation showed huge contrast between sites of varied oil impact, suggesting that externalities such seasonal influence can hinder effective discrimination. Results from this Chapter indicated that retrieved backscatter from polluted and oil-free vegetation types depicted reliable vegetation temporal characteristics, compared to coarse resolution MODIS NDVI data. This particularly demonstrated the suitability of multi-temporal SAR to unravel seasonal dimensions associated with terrestrial oil spills. This is as a result of significant differences in mean backscatter descriptors between polluted and oil-free vegetation during the summer (wet) season relative to the winter (dry) season.

The effect of seasonality as observed in this chapter displayed similar temporal backscatter trends to those observed in Onojeghuo *et al.*, (2018b) and Rüetschi *et al.*, (2017), where high backscatter values are recorded during the summer (wet season) growing season, while low backscatter values dominates in the winter (dry) harvest season. Classification accuracies also showed that the integration of L, C and X band SAR data in the summer season gave the best result of over 80% OA in the discrimination of polluted and oil-free cropland vegetation, representing an over 10% improvement over the previous best classification in Chapter five (using dry season images). Grassland vegetation also had a marginal improvement of more than 1.6% from the previous implemented classification. A major reason for the improved accuracies can be attributed to the use of multi-frequency and multi-polarization stacked images. This have been previously recommended by Ramsey *et al.,* (2015) to effectively capture the structural variability of sparse vegetation across space and depth in comparison with denser vegetation. This is also the position of Shang *et al.* (2009) where multi-frequency SAR proved to be a reliable solution for crop vegetation classification due to high accuracies

observed from the integration of L-C-X band SAR as opposed to when single or two frequencies of SAR images were integrated.

Similarly, Hagensieker and Waske (2018) also came to the conclusion that the integration of multifrequency – multiSAR sensor is worthwhile as it significantly improves tropical land cover classification. Martinis and Rieke (2015) also concluded that the use of multifrequency SAR in the mapping of flood areas yielded remarkable improvement in accuracy due to the suitability of all three (L-C-X) frequencies of SAR in mapping flooded vegetation even in sparse vegetation.

The inclusion of L – Band SAR in the classification led to a significant improvement in the general classification accuracy. The reason for this is that L – Band SAR has higher penetration into the vegetation sub canopy (15m depth), representing radiation interaction with vegetation trunks, crown, ground and vegetation branches. In contrast, the C and X Band SAR have a much shallow penetration (of 7m and 3m, respectively), which represents backscattering radiation interaction with vegetation canopy, leaves, and secondary branches. In addition, results obtained in this study are also in line with findings from Ramsey *et al.*, (2015), where L – Band SAR gave a positive result in the detection of structural defect caused by oil pollution on marsh vegetation owing to the better penetration of incident beam from the L – Band sensor.

To conclude, this research discovered that using multispectral optical image derivatives for the prime purpose of detecting and mapping terrestrial oil spill gives stronger results when specific land cover types are assessed separately than when the entire image consisting of different features are assessed at once. Image spectral bands, especially those within the Shortwave and Near – Infrared parts of the spectrum, are better suited for discriminating between polluted and oil-free vegetation than vegetation health indices in the machine learning approach. Thirdly, the incorporation and integration of SAR with multispectral optical images presents greater potential for the mapping of polluted vegetation with higher accuracy than when optical data alone is used. The use of SAR images for the purpose of detecting and mapping oil-affected vegetation are more assured in areas with moderate vegetation canopy, in terms of biomass density. Better results were also gained when prior knowledge of the predominant land cover and vegetation type are known and used individually, than when the entire image scene with all other features are assessed at once. The reliability and assurance of an improved detection and mapping of

oil-impacted vegetation using SAR images is much higher during the summer/wet season than during the dry/winter season due to homogenous representation of both stressed and healthy vegetation in the dry season. It is also the case that L – Band SAR are more suited for detecting oil-impacted vegetation of all types than C and X band SAR.

7.4 Limitations

Several limitations encountered in the course of this research are discussed below:

Oil Spill Incident data: The oil spill data used in this research were obtained from the website of the national oil spill detection and response agency (NOSDRA), which is the government agency responsible for documenting all spill incidents and the shell petroleum development corporation (SPDC). The research relied mostly on the information provided in the joint investigative reports (JIV) on size, area and date of spill. Hence, independent verification of most of the information provided by the stakeholder was not achievable. Therefore, critical assumption and judgements had to be applied in the selection of spill sites used in the course of this research. An example is in the case of reference coordinate system conversion from the old 'Minna Mid-Belt' coordinate system to contemporary Universal Transverse Mercator (UTM) reference system. This was to ensure proper alignment of all datasets used in the course of this research. Another example in this regard is the fact that not all the spill incidents reported had GPS coordinates establishing the spatial extent and coverage of the spill. As such, certain assumptions and generalizations had to be made considering the large volume and spatial coverage of spill reported. These were considered with care in the number of training sites selected.

Existing Land cover: This research greatly relied on the existing land cover map produced by the European Space Agency - Climate Change Initiative (ECCI) in 2015, using Sentinel – 2 optical reflectance data and random forest method. The land cover classes served as the primary basis for establishing the different vegetation types used throughout this research. In this regard, it was impossible to independently verify the authenticity of different land cover extent on ground, as this would have required extensive fieldwork. However, visual comparison with high resolution Google Earth image (HRGEI) was carried out to verify that the vegetation extent as provided by ECCI were in agreement with the realities on ground.

Generalization of the extent of polluted/impacted vegetation extent: An important exercise undertaken in the course of this research was to map the spatial extent of oil-impacted vegetation through machine learning classification, following the detection of impact posed on vegetation using a number of statistical tools. The spatial extent outputs representing oil-impacted or polluted vegetation does not exclusively represent polluted areas with active pools of oil, but provide reliable information on vegetation undergoing stress as a result of the impact posed by oil pollution. In addition, evidence from the field exercise also indicated certain areas affected by other stressors may also be wrongly assigned as polluted vegetation.

Limited field validation: Field validation is an essential requirement in satellite remote sensing based studies. This usually provides basis for establishing the performance of methodologies employed to adequately and accurately predict land cover classes. It also provides a means through which the authenticity of map features can be reconciled with actual ground features, as such areas of false positives can then be noted and improved on subsequently. In the course of this research, only six spill sites were accessible for visit and used to conduct ground truth. Challenges such as political insecurity, host community conflict, general bureaucratic procedures and clearance to visit major spill sites severely hampered efforts to validate the results of this research further. This greatly impeded effort to ascertaining the status of some of the polluted areas predicted from the image classification.

7.5 Future Work

- **Exploring the potential of Hyperspectral data in machine learning**

Hyperspectral images and Hyperspectral-derived variables as mentioned in section 2.2 provide a reliable means for detecting oil-polluted areas. Evidence from previous studies shows that the use of this data can sufficiently overcome the spectral confusion challenge associated with oil spill detection and mapping on terrestrial landscape. However, their incorporation into machine learning classifiers for mapping oil-polluted areas has not been experimented. This is especially necessary given that a number of hyperspectral satellite missions have been proposed as EnMAP Hyperspectral Imager, Spaceborne Hyperspectral Applicative Land and Ocean Mission (SHALOM) and Hyperspectral Infrared Imager (HyspIRI). This certainly would form a substantive area for further

research as the integration of this data can improve the general process of oil spill detection, mapping and oil pipeline monitoring on the terrestrial landscape.

- **Exploration of Newer Image Processing Methods**

The last decade has seen the emergence of very sophisticated image processing methodology drawn from the field of computer science. Notable among them is the Deep Learning methods, which are further improvements on the existing neural network methods. The potential and capability of deep learning for the specific processing and discrimination of oil-impacted from oil-free vegetation is an area that needs to be explored further owing to achievements already recorded by the methodology in the area of signal processing.

- **What role does elevation data plays in terrestrial oil spill detection?**

There is certainly need to further investigate the role of elevation models in the detection and mapping of hydrocarbon-induced stress, since results obtained in chapter five of this thesis consistently showed that digital elevation model derived variables (i.e. slope, aspect and DEM) were the most important variable in the classification process. The utility of DEM and associated variables such as terrain and topography are yet to be explored in detail when detecting terrestrial oil spills. An in-depth analysis of the dynamics of oil spill movement on the terrestrial landscape and its influences in producing accurate maps of polluted areas, especially with machine learning would greatly enrich the terrestrial oil spill remote sensing research community.

- **Further SAR Investigation Using Scatterometer**

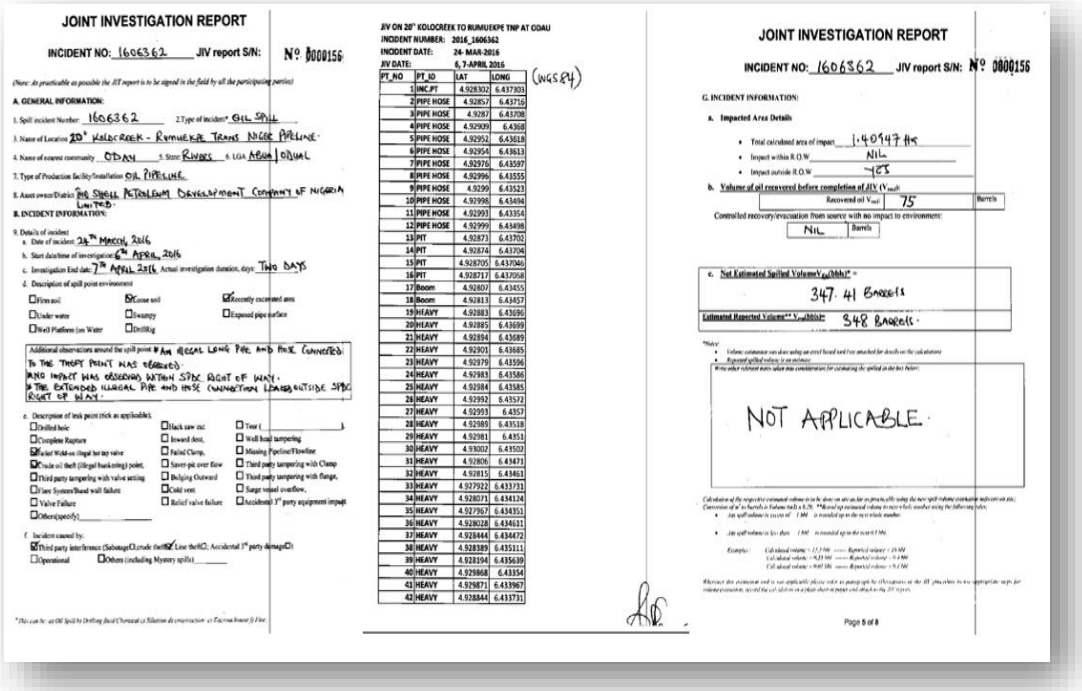
A scatterometer device works on the same operational principle of a SAR sensor. It operates by transmitting a pulse of microwave energy towards a target on the earth's surface and measuring the reflected energy. This device can also be manually operated to obtain and characterize the dominant scattering tendencies of a specific target. The operational use of this device for the specific monitoring and measurement of a typical oil polluted site is an area that is worth investigating. This would help in further reconciling and complimenting the progress made in this research in terms of the image derived scattering characteristics of oil polluted land cover types as compared to the field observed scattering tendencies. Effort such as this would further improve the understanding of physical characteristics of various oil-polluted surfaces.

Appendix

Appendix I – Oil polluted sites within the study area captured by the National Oil Spill Detection and Response Agency.

Overview: Extent of the spill incident impact on 28'' Bomu-Bonny Pipeline at Bodo West . Pictures taken during Joint Investigation of 26th February, 2011

Appendix II – Sample of Joint Investigative Visit (JIV) form used to capture oil spill incidents



JOINT INVESTIGATION REPORT

INCIDENT NO: 1606362 JIV report S/N: Nº 0000156

A. GENERAL INFORMATION:

1. Spill incident Number: 1606362 2.Type of incident*: OIL SPILL

3. Name of Location: 28" KOLOCREEK - RUMUEKPE TRANS NIGER PIPELINE.

4. Name of nearest community: ODIAM 5. State: RIVERS 6. LGA: ABUA / ODUAL

7. Type of Production facility/Installation: OIL PIPELINE.

8. Asset owner/Distric: THE SHELL PETROLEUM DEVELOPMENT COMPANY OF NIGERIA LIMITED.

B. INCIDENT INFORMATION:

9. Details of incident:

a. Date of incident: 26TH MARCH, 2016

b. Start date/time of investigation: 6TH APRIL, 2016

c. Investigation End date: 7TH APRIL 2016 Actual investigation duration, days: TWO DAYS

d. Description of spill point environment

☐Firm soil ☒Loose soil ☑Recently excavated area

☐Under water ☐Swampy ☐Exposed pipe surface

☐Well Platform /on Water ☐Drill Rig

Additional observations around the spill point: AN ILLEGAL LONG PIPE AND HOSE CONNECTED TO THE TROPY POINT WAS OBSERVED.
NO IMPACT WAS OBSERVED WITHIN SPDC RIGHT OF WAY.
THE EXTENDED ILLEGAL PIPE AND HOSE CONNECTION LOCATED OUTSIDE SPDC RIGHT OF WAY.

e. Description of leak point (tick as applicable):

☐Drilled hole ☐Hack saw cut ☐Tear (_____)

☐Complete Rupture ☐Inward dent, ☐Well head tampering

☑Relief Weld-on illegal hot tap valve ☐Failed Clamp, ☐Missing Pipeline/Flowline

☑Crude oil theft (illegal hacksawing) point, ☐Sleeve-pin over flow ☐Third party tampering with Clamp

☐Third party tampering with valve setting ☐Bulging Outward ☐Third party tampering with flange,

☐Tank System/Bund wall failure ☐Cold vent ☐Surge vessel overflow,

☐Valve Failure ☐Relief valve failure ☐Accidental 3rd party equipment impact

☐Other(specify)_____

f. Incident caused by:

☑Third party interference (Sabotage☐,crude theft☑ Line theft☐; Accidental 3rd party damage☐)

☐Operational ☐Others (including Mystery spills)_____

*This can be: an Oil Spill by Drilling fluid Chemical or Siltation de construction et Excavation based fl Flow

JIV ON 28" KOLOCREEK TO RUMUEKPE TNP AT ODAU

INCIDENT NUMBER: 2016_1606362

INCIDENT DATE: 26- MAR-2016

JIV DATE: 6, 7-APRIL 2016

(WGS84)

PT_NO	PT_ID	LAT	LONG
1	INC_PT	4.928302	6.437303
2	PIPE HOSE	4.92857	6.43716
3	PIPE HOSE	4.9287	6.43708
4	PIPE HOSE	4.92909	6.4368
5	PIPE HOSE	4.92952	6.43618
6	PIPE HOSE	4.92954	6.43613
7	PIPE HOSE	4.92976	6.43597
8	PIPE HOSE	4.92996	6.43555
9	PIPE HOSE	4.9299	6.43523
10	PIPE HOSE	4.92998	6.43494
11	PIPE HOSE	4.92993	6.43354
12	PIPE HOSE	4.92999	6.43498
13	PIT	4.92873	6.43702
14	PIT	4.92874	6.43704
15	PIT	4.928705	6.437040
16	PIT	4.928717	6.437058
17	Boom	4.92807	6.43455
18	Boom	4.92813	6.43457
19	HEAVY	4.92883	6.43696
20	HEAVY	4.92885	6.43699
21	HEAVY	4.92894	6.43689
22	HEAVY	4.92903	6.43685
23	HEAVY	4.92979	6.43596
24	HEAVY	4.92983	6.43586
25	HEAVY	4.92984	6.43585
26	HEAVY	4.92992	6.43572
27	HEAVY	4.92993	6.4357
28	HEAVY	4.92989	6.43518
29	HEAVY	4.92981	6.4351
30	HEAVY	4.93002	6.43502
31	HEAVY	4.92806	6.43471
32	HEAVY	4.92815	6.43463
33	HEAVY	4.927922	6.433731
34	HEAVY	4.928071	6.434124
35	HEAVY	4.927967	6.434351
36	HEAVY	4.928028	6.434631
37	HEAVY	4.928444	6.434472
38	HEAVY	4.928389	6.435111
39	HEAVY	4.928194	6.435639
40	HEAVY	4.929868	6.43354
41	HEAVY	4.929871	6.433967
42	HEAVY	4.928844	6.433731

JOINT INVESTIGATION REPORT

INCIDENT NO: 1606362 JIV report S/N: Nº 0000156

G. INCIDENT INFORMATION:

a. Impacted Area Details

- Total calculated area of impact: 1.40947 Ha
- Impact within R.O.W: NIL
- Impact outside R.O.W: YES

b. Volume of oil recovered before completion of JIV (V_{recd})

Recovered oil V_{recd}: 75 Barrels

Controlled recovery/evacuation from source with no impact to environment: NIL Barrels

c. Net Estimated Spilled Volume V_{est}(bbls)* =

347. 41 Barrels

Estimated Reported Volume** V_{rep}(bbls)= 348 Barrels

*Note:

- Volume estimation was done using an excel based tool (see attached for details on the calculations
- Reported spilled volume is an estimate
- Any other relevant notes taken into consideration for estimating the spilled in the box below:

NOT APPLICABLE.

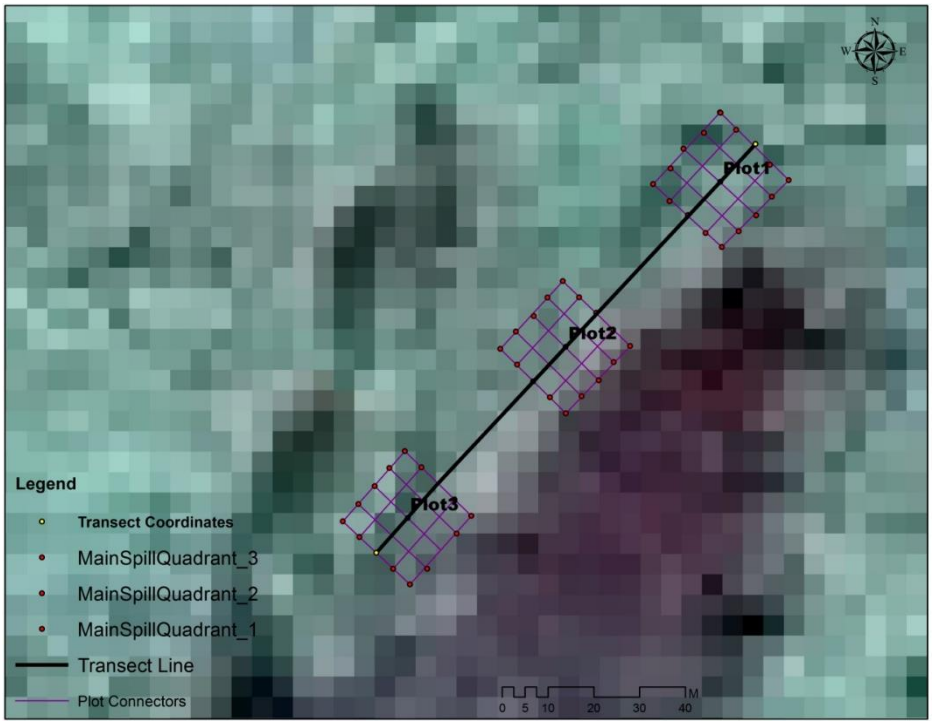
Page 5 of 8

Appendix III – Google Earth Image Showing the Established Plot within the Recent Oil Spill Site

Appendix IV – Google Earth Image Showing the Established Plot within the Recent Oil Spill Site

Appendix V – Google Earth Image Showing the Established Plot within the Non-Polluted Site

Appendix VI – Showing the Coordinates of the Plots



Appendix VII – Several Pictures Taken during the course of the Field Spectroradiometric Data acquisition

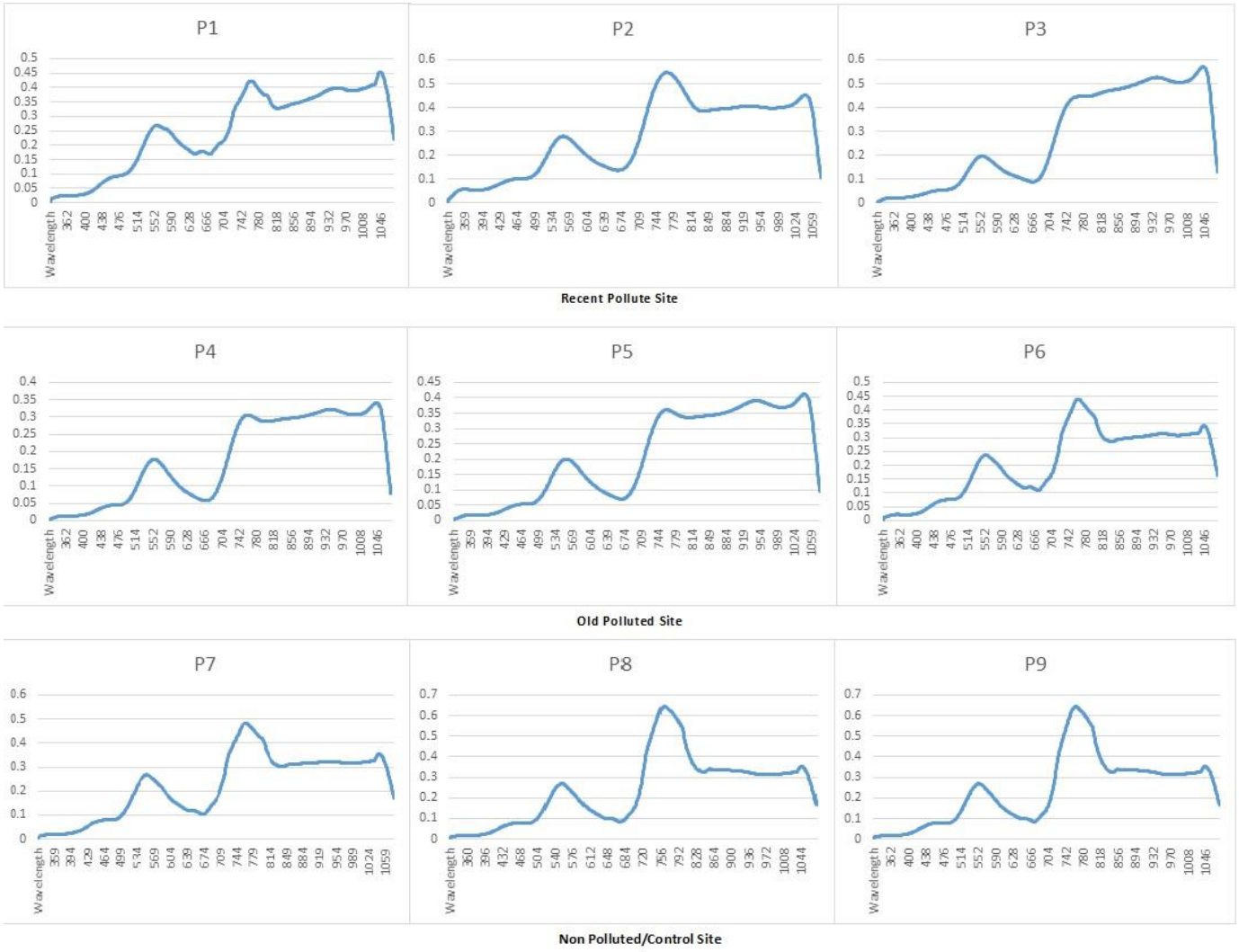
(a)

(b)

(c)

(d)

Appendix VIII – Mean Spectral Reflectance from individual plots of the recent, old and non-polluted site



Appendix IX – **Detailed Confusion Matrix – Masked Image Classification – Cropland (Samples)**

Map class	Non-Polluted Cropland	Polluted Cropland	Total
Non-Polluted Cropland	10	7	17
Polluted Cropland	6	10	16
Total	16	17	33

Appendix X - Detailed Confusion Matrix – Masked Image Classification – Grassland (Samples)

Map class	Non-Polluted Grassland	Polluted Grassland	Total
Non-Polluted Grassland	8	5	13
Polluted Grassland	2	5	7
Total	10	10	20

Appendix XI - Detailed Confusion Matrix – Masked Image Classification – TCA (Samples)

Map class	Non-Polluted TCA	Polluted TCA	Total
Non-Polluted TCA	6	2	8
Polluted TCA	4	8	12
Total	10	10	20

Appendix XII - Detailed Confusion Matrix – Entire Image Classification (Samples)

Map class	Non-Polluted Cropland	Non-Polluted Grassland	Non-Polluted TCA	Polluted Cropland	Polluted Grassland	Polluted TCA	Total
Non-Polluted Cropland	**3**	1	1	4	1	2	12
Non-Polluted Grassland	4	**2**	2	1	1	1	11
Non-Polluted TCA	1	0	**4**	0	1	2	8
Polluted Cropland	8	4	0	**7**	3	2	24
Polluted Grassland	0	1	1	5	**3**	0	10
Polluted TCA	0	2	2	0	1	**3**	8
Total	16	10	10	17	10	10	**73**

Bibliography

Abam, T. (2001). Regional hydrological research perspectives in the Niger Delta. *Hydrological sciences journal, 46*(1), 13-25.

Abdel-Moghny, T., Mohamed, R. S., El-Sayed, E., Mohammed Aly, S., & Snousy, M. G. (2012). Effect of soil texture on remediation of hydrocarbons-contaminated soil at El-Minia district, Upper Egypt. *ISRN Chemical Engineering, 2012*.

Achard, V., Fabre, S., Alakian, A., Dubucq, D., & Déliot, P. (2018). *Direct or indirect on-shore hydrocarbon detection methods applied to hyperspectral data in tropical area*. Paper presented at the Earth Resources and Environmental Remote Sensing/GIS Applications IX.

Adam, E., Mutanga, O., Odindi, J., & Abdel-Rahman, E. M. (2014). Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the performance of random forest and support vector machines classifiers. *International Journal of Remote Sensing, 35*(10), 3440-3458. doi:10.1080/01431161.2014.903435

Adamu, B., Tansey, K., & Ogutu, B. (2015). Using vegetation spectral indices to detect oil pollution in the Niger Delta. *Remote Sensing Letters, 6*(2), 145-154. doi:10.1080/2150704x.2015.1015656

Adamu, B., Tansey, K., & Ogutu, B. (2016). An investigation into the factors influencing the detectability of oil spills using spectral indices in an oil-polluted environment. *International Journal of Remote Sensing, 37*(10), 2338-2357. doi:10.1080/01431161.2016.1176271

Adamu, B., Tansey, K., & Ogutu, B. (2018). Remote sensing for detection and monitoring of vegetation affected by oil spills. *International Journal of Remote Sensing, 39*(11), 3628-3645.

Adejuwon, J. O. (2012). Rainfall seasonality in the Niger delta belt, Nigeria. *Journal of Geography and Regional Planning, 5*(2), 51-60.

Agapiou, A., Hadjimitsis, D., & Alexakis, D. (2012). Evaluation of broadband and narrowband vegetation indices for the identification of archaeological crop marks. *Remote Sensing, 4*(12), 3892-3919.

Agapiou, A., Hadjimitsis, D. G., & Alexakis, D. D. (2012). Evaluation of broadband and narrowband vegetation indices for the identification of archaeological crop marks. *Remote Sensing, 4*(12), 3892-3919.

Aguilera, F., Méndez, J., Pásaro, E., & Laffon, B. (2010). Review on the effects of exposure to spilled oils on human health. *Journal of Applied Toxicology, 30*(4), 291-301.

Aislabie, J., Fraser, R., Duncan, S., & Farrell, R. L. (2001). Effects of oil spills on microbial heterotrophs in Antarctic soils. *Polar Biology, 24*(5), 308-313.

Akegbejo-Samsons, Y., & Omoniyi, I. (2009). Management challenges of mangrove forests in Africa: a critical appraisal of the coastal mangrove ecosystem of Nigeria. *Nature and Faune, 24*, 50-55.

Akpokodje, E. G. (1987). The engineering-geological characteristics and classification of the major superficial soils of the Niger Delta. *Engineering Geology, 23*(3-4), 193-211.

Aleklett, K., Höök, M., Jakobsson, K., Lardelli, M., Snowden, S., & Söderbergh, B. (2010). The peak of the oil age–analyzing the world oil production reference scenario in world energy outlook 2008. *Energy Policy, 38*(3), 1398-1414.

Alexakis, D. D., Sarris, A., Kalaitzidis, C., Papadopoulos, N., & Soupios, P. (2016). Integrated use of satellite remote sensing, GIS, and ground spectroscopy

techniques for monitoring olive oil mill waste disposal areas on the island of Crete, Greece. *International Journal of Remote Sensing, 37*(3), 669-693.

Aljazeera report, A. (2017). *https://www.aljazeera.com/indepth/inpictures/2017/07/long-term-effects-oil-spills-bodo-nigeria-170717090542648.html, [Accessed 21/01/2019*.

Aplin, P., & Atkinson, P. M. (2001). Sub-pixel land cover mapping for per-field classification. *International Journal of Remote Sensing, 22*(14), 2853-2858.

Arellano, P., Tansey, K., Balzter, H., & Boyd, D. S. (2015). Detecting the effects of hydrocarbon pollution in the Amazon forest using hyperspectral satellite images. *Environ Pollut, 205*, 225-239. doi:10.1016/j.envpol.2015.05.041

Arellano, P., Tansey, K., Balzter, H., & Boyd, D. S. (2017). Field spectroscopy and radiative transfer modelling to assess impacts of petroleum pollution on biophysical and biochemical parameters of the Amazon rainforest. *Environmental Earth Sciences, 76*(5), 217.

Arellano, P., Tansey, K., Balzter, H., & Tellkamp, M. (2017). Plant Family-Specific Impacts of Petroleum Pollution on Biodiversity and Leaf Chlorophyll Content in the Amazon Rainforest of Ecuador. *PLoS One, 12*(1), e0169867. doi:10.1371/journal.pone.0169867

Arif, M., Suresh, M., Jain, K., & Dundhigal, S. (2015). Sub Pixel Classification of High Resolution Satellite Imagery. *International Journal of Computer Applications, 129*(1).

Arroyo, L. A., Johansen, K., & Phinn, S. (2010). Mapping Land Cover Types from Very High Spatial Resolution Imagery: Automatic Application of an Object Based Classification Scheme. *Proceedings of the GEOBIA*.

Ayanlade, A. (2015). *Remote sensing of environmental change in the Niger Delta, Nigeria*. King's College London,

Ayanlade, A. (2016). Evaluating environmental change impacts on ecological services in the Niger Delta of Nigeria. *Ife Research Publications in Geography, 11*(1), 111-125.

Ayanlade, A., & Howard, M. T. (2016). Environmental impacts of oil production in the Niger Delta: Remote sensing and social survey examination. *African Geographical Review, 35*(3), 272-293.

Balogun, T. F. (2015). Utility of Microwave and Optical Remote Sensing in Oil Spill Detection in the Mangrove Region of Nigeria. *Journal of Geoscience and Environment Protection, 3*(01), 16.

Bamler, R., & Hartl, P. (1998). Synthetic aperture radar interferometry. *Inverse problems, 14*(4), R1.

Ban, Y., Jacob, A., & Gamba, P. (2015). Spaceborne SAR data for global urban mapping at 30m resolution using a robust urban extractor. *ISPRS Journal of Photogrammetry and Remote Sensing, 103*, 28-37. doi:10.1016/j.isprsjprs.2014.08.004

Bannari, A., Asalhi, H., & Teillet, P. (2002). *Transformed difference vegetation index (TDVI) for vegetation cover mapping*. Paper presented at the Geoscience and Remote Sensing Symposium, 2002. IGARSS'02. 2002 IEEE International.

Barrett, B., Nitze, I., Green, S., & Cawkwell, F. (2014). Assessment of multi-temporal, multi-sensor radar and ancillary spatial data for grasslands monitoring in Ireland using machine learning approaches. *Remote Sensing of Environment, 152*, 109-124. doi:10.1016/j.rse.2014.05.018

Baudena, M., Dekker, S. C., van Bodegom, P. M., Cuesta, B., Higgins, S. I., Lehsten, V., . . . Yin, Z. (2015). Forests, savannas and grasslands: bridging the knowledge gap

between ecology and Dynamic Global Vegetation Models. *Biogeosciences, 12*, 1833-1848.

Ben-Dor, E., Inbar, Y., & Chen, Y. (1997). The reflectance spectra of organic matter in the visible near-infrared and short wave infrared region (400–2500 nm) during a controlled decomposition process. *Remote Sensing of Environment, 61*(1), 1-15.

Benabdelouahab, T., Balaghi, R., Hadria, R., Lionboui, H., Minet, J., & Tychon, B. (2015). Monitoring surface water content using visible and short-wave infrared SPOT-5 data of wheat plots in irrigated semi-arid regions. *International Journal of Remote Sensing, 36*(15), 4018-4036.

Betbeder, J., Rapinel, S., Corpetti, T., Pottier, E., Corgne, S., & Hubert-Moy, L. (2014). Multitemporal classification of TerraSAR-X data for wetland vegetation mapping. *Journal of Applied Remote Sensing, 8*(1), 083648.

Bianchi, R., Cavalli, R. M., Marino, C. M., Pignatti, S., & Poscolieri, M. (1995a). *Use of airborne hyperspectral images to assess the spatial distribution of oil spilled during the Trecate blow-out (Northern Italy).* Paper presented at the Remote Sensing for Agriculture, Forestry, and Natural Resources.

Bianchi, R., Cavalli, R. M., Marino, C. M., Pignatti, S., & Poscolieri, M. (1995b). *Use of airborne hyperspectral images to assess the spatial distribution of oil spilled during the Trecate blow-out (Northern Italy).* Paper presented at the Satellite Remote Sensing II.

Biersteker, T. J. (2014). *Multinationals, the state and control of the Nigerian economy* (Vol. 498): Princeton University Press.

Boerner, W.-M., Mott, H., & Luneburg, E. (1997). *Polarimetry in remote sensing: Basic and applied concepts.* Paper presented at the Geoscience and Remote Sensing, 1997. IGARSS'97. Remote Sensing-A Scientific Vision for Sustainable Development., 1997 IEEE International.

Breiman, L. (2001). Random forests. *Machine learning, 45*(1), 5-32.

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

Brooks, R. (1979). Indicator plants for mineral prospecting—a critique. *Journal of Geochemical Exploration, 12*, 67-78.

Cable, J. W., Kovacs, J. M., Shang, J., & Jiao, X. (2014). Multi-temporal polarimetric RADARSAT-2 for land cover monitoring in Northeastern Ontario, Canada. *Remote Sensing, 6*(3), 2372-2392.

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST), 2*(3), 27.

Chang, S., Stone, J., Demes, K., & Piscitelli, M. (2014). Consequences of oil spills: a review and framework for informing planning. *Ecology and Society, 19*(2).

Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., & Eklundh, L. (2004). A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. *Remote Sensing of Environment, 91*(3-4), 332-344.

Chen, Y., Shi, P., Deng, L., & Li, J. (2007). Generation of a top-of-canopy Digital Elevation Model (DEM) in tropical rain forest regions using radargrammetry. *International Journal of Remote Sensing, 28*(19), 4345-4349.

Cissé, S., Eymard, L., Ottlé, C., Ndione, J. A., Gaye, A. T., & Pinsard, F. (2016). Rainfall intra-seasonal variability and vegetation growth in the Ferlo Basin (Senegal). *Remote Sensing, 8*(1), 66.

Clerici, N., Valbuena Calderón, C. A., & Posada, J. M. (2017). Fusion of Sentinel-1A and Sentinel-2A data for land cover mapping: a case study in the lower Magdalena region, Colombia. *Journal of Maps, 13*(2), 718-726.

Cloude, S. R., & Pottier, E. (1996). A review of target decomposition theorems in radar polarimetry. *IEEE Transactions on Geoscience and Remote Sensing, 34*(2), 498-518.

Cloutis, E. (1989). Spectral reflectance properties of hydrocarbons: remote—sensing implications. *Science, 245*(4914), 1657168.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment, 37*(1), 35-46.

Conn, D., Ngun, T., Li, G., & Ramirez, C. (2015). Fuzzy Forests: Extending Random Forests for Correlated, High-Dimensional Data.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning, 20*(3), 273-297.

Costa, M. (2004). Use of SAR satellites for mapping zonation of vegetation communities in the Amazon floodplain. *International Journal of Remote Sensing, 25*(10), 1817-1835.

Countries, O. o. t. P. E. (2017). 2017 World Oil Outlook 2040.

Dada, O. A., Qiao, L., Ding, D., Li, G., Ma, Y., & Wang, L. (2015). Evolutionary trends of the Niger Delta shoreline during the last 100 years: Responses to rainfall and river discharge. *Marine Geology, 367*, 202-211.

Darst, B. F., Malecki, K. C., & Engelman, C. D. (2018). Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. *BMC genetics, 19*(1), 65.

Davidson, S., Santos, M., Sloan, V., Watts, J., Phoenix, G., Oechel, W., & Zona, D. (2016). Mapping Arctic Tundra Vegetation Communities Using Field Spectroscopy and Multispectral Satellite Data in North Alaska, USA. *Remote Sensing, 8*(12), 978. doi:10.3390/rs8120978

de Leeuw, J., Jia, H., Yang, L., Liu, X., Schmidt, K., & Skidmore, A. (2006). Comparing accuracy assessments to infer superiority of image classification methods. *International Journal of Remote Sensing, 27*(1), 223-232.

Dobson, M. C., Pierce, L. E., & Ulaby, F. T. (1996). Knowledge-based land-cover classification using ERS-1/JERS-1 SAR composites. *IEEE Transactions on Geoscience and Remote Sensing, 34*(1), 83-99.

Domínguez-Beisiegel, M., Castañeda, C., Mougenot, B., & Herrero, J. (2016). Analysis and mapping of the spectral characteristics of fractional green cover in saline wetlands (NE Spain) using field and remote sensing data. *Remote Sensing, 8*(7), 590.

Doraiswamy, P. C., & Thompson, D. (1982). A crop moisture stress index for large areas and its application in the prediction of spring wheat phenology. *Agricultural Meteorology, 27*(1-2), 1-15.

Dotzler, S., Hill, J., Buddenbaum, H., & Stoffels, J. (2015). The potential of EnMAP and Sentinel-2 data for detecting drought stress phenomena in deciduous forest communities. *Remote Sensing, 7*(10), 14227-14258.

Douaoui, A. E. K., Nicolas, H., & Walter, C. (2006). Detecting salinity hazards within a semiarid context by means of combining soil and remote-sensing data. *Geoderma, 134*(1-2), 217-230.

Douglas, R., Nawar, S., Alamar, M., Mouazen, A., & Coulon, F. (2018). Rapid prediction of total petroleum hydrocarbons concentration in contaminated soil using vis-NIR spectroscopy and regression techniques. *Science of The Total Environment, 616*, 147-155.

Draper, C. S., Reichle, R. H., & Koster, R. D. (2018). Assessment of MERRA-2 land surface energy flux estimates. *Journal of Climate, 31*(2), 671-691.

Du, S., Zhang, F., & Zhang, X. (2015). Semantic classification of urban buildings combining VHR image and GIS data: An improved random forest approach. *ISPRS Journal of Photogrammetry and Remote Sensing, 105*, 107-119.

Ebeku, K. S. (2006). *Oil and the Niger Delta people in international law: resource rights, environmental and equity issues* (Vol. 2): Köppe Köln.

Eckert, S. (2012). Improved forest biomass and carbon estimations using texture measures from WorldView-2 satellite data. *Remote Sensing, 4*(4), 810-829.

Eregha, P. B., & Irughe, I. (2009). Oil induced environmental degradation in the Nigeria's Niger Delta: The multiplier effects. *Journal of Sustainable Development in Africa, 11*(4), 160-175.

Fingas, M., & Brown, C. (2014). Review of oil spill remote sensing. *Mar Pollut Bull, 83*(1), 9-23. doi:10.1016/j.marpolbul.2014.03.059

Freeman, A., & Durden, S. L. (1998). A three-component scattering model for polarimetric SAR data. *IEEE Transactions on Geoscience and Remote Sensing, 36*(3), 963-973.

Fröhlich, B., Bach, E., Walde, I., Hese, S., Schmullius, C., & Denzler, J. (2013). Land cover classification of satellite images using contextual information. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 3*, W1.

Gabriel, A. (2004). Women in the Niger Delta: environmental issues and challenges in the third millennium. *The Journal of Sustainable Development in Africa, 6*(2).

Gaffey, S., McFadden, L., Nash, D., & Pieters, C. (1993). Ultraviolet, visible, and near-infrared reflectance spectroscopy: Laboratory spectra of geologic materials. *Remote geochemical analysis: Elemental and mineralogical composition, 43*, 77.

Gao, B.-C. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment, 58*(3), 257-266.

Gao, Q., Zribi, M., Escorihuela, M., Baghdadi, N., & Segui, P. (2018). Irrigation mapping using Sentinel-1 time series at field scale. *Remote Sensing, 10*(9), 1495.

Gao, Q., Zribi, M., Escorihuela, M. J., & Baghdadi, N. (2017). Synergetic use of Sentinel-1 and Sentinel-2 data for soil moisture mapping at 100 m resolution. *Sensors, 17*(9), 1966.

Genuer, R., Poggi, J.-M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters, 31*(14), 2225-2236.

Ghulam, A., Porton, I., & Freeman, K. (2014). Detecting subcanopy invasive plant species in tropical rainforest by integrating optical and microwave (InSAR/PolInSAR) remote sensing data, and a decision tree algorithm. *ISPRS Journal of Photogrammetry and Remote Sensing, 88*, 174-192. doi:10.1016/j.isprsjprs.2013.12.007

Giri, C., Long, J., & Tieszen, L. (2011). Mapping and Monitoring Louisiana's Mangroves in the Aftermath of the 2010 Gulf of Mexico Oil Spill. *Journal of Coastal Research, 277*, 1059-1064. doi:10.2112/jcoastres-d-11-00028.1

Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern Recognition Letters, 27*(4), 294-300.

Glover, P. C., & Economides, M. J. (2010). *Energy and Climate Wars: How naive politicians, green ideologues, and media elites are undermining the truth about energy and climate*: Bloomsbury Publishing.

Goodman, R. (1994). Overview and future trends in oil spill remote sensing. *Spill Science & Technology Bulletin, 1*(1), 11-21.

Guo, L., Chehata, N., Mallet, C., & Boukir, S. (2011). Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests. *ISPRS Journal of Photogrammetry and Remote Sensing, 66*(1), 56-66.

Hadjimitsis, D. G., Papadavid, G., Agapiou, A., Themistocleous, K., Hadjimitsis, M., Retalis, A., . . . Clayton, C. (2010). Atmospheric correction for satellite remotely sensed data intended for agricultural applications: impact on vegetation indices. *Natural Hazards and Earth System Sciences, 10*(1), 89-95.

Hagensieker, R., & Waske, B. (2018). Evaluation of multi-frequency SAR images for tropical land cover mapping. *Remote Sensing, 10*(2), 257.

Haralick, R. M., & Shanmugam, K. (1973). Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*(6), 610-621.

Hardisky, M., Klemas, V., & Smart, M. (1983). The influence of soil salinity, growth form, and leaf moisture on the spectral radiance of. *Spartina alterniflora*, 77-83.

Hasan, S. W., Ghannam, M. T., & Esmail, N. (2010). Heavy crude oil viscosity reduction and rheology for pipeline transportation. *Fuel, 89*(5), 1095-1100.

Henderson, F. M., & Lewis, A. J. (2008). Radar detection of wetland ecosystems: a review. *International Journal of Remote Sensing, 29*(20), 5809-5835.

Herrington, P., Ball, G., O'Halloran, K., & Tremblay, L. (2006). *Aquatic ecotoxicity of cutback bitumen*: Land Transport New Zealand.

Hese, S., & Schmullius, C. STRUCTURAL IMAGE CLASSIFICATION FOR OIL SPILL CONTAMINATION MAPPING IN WEST SIBERIA USING QUICKBIRD DATA.

Hese, S., & Schmullius, C. (2008). Object oriented oil spill contamination mapping in West Siberia with Quickbird data. In *Object-Based Image Analysis* (pp. 367-382): Springer.

Hese, S., & Schmullius, C. (2009). High spatial resolution image object classification for terrestrial oil spill contamination mapping in West Siberia. *International Journal of Applied Earth Observation and Geoinformation, 11*(2), 130-141.

Hester, M., & Mendelssohn, I. (2000). Long-term recovery of a Louisiana brackish marsh plant community from oil-spill impact: vegetation response and mitigating effects of marsh surface elevation. *Marine Environmental Research, 49*(3), 233-254.

Hester, M. W., Willis, J. M., Rouhani, S., Steinhoff, M. A., & Baker, M. C. (2016). Impacts of the Deepwater Horizon oil spill on the salt marsh vegetation of Louisiana. *Environmental pollution, 216*, 361-370.

Hlatshwayo, S. T., Mutanga, O., Lottering, R. T., Kiala, Z., & Ismail, R. (2019). Mapping forest aboveground biomass in the reforested Buffelsdraai landfill site using texture combinations computed from SPOT-6 pan-sharpened imagery. *International Journal of Applied Earth Observation and Geoinformation, 74*, 65-77.

Hooper, R. J., Fitzsimmons, R. J., Grant, N., & Vendeville, B. C. (2002). The role of deformation in controlling depositional patterns in the south-central Niger Delta, West Africa. *Journal of Structural Geology, 24*(4), 847-859.

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

Hsu, C.-W., Chang, C.-C. and Lin, C.-J. (2003) 'A practical guide to support vector classification'

Huang, S., Chen, S., Wang, D., Zhou, C., van der Meer, F., & Zhang, Y. (2019). Hydrocarbon micro-seepage detection from airborne hyper-spectral images by

plant stress spectra based on the PROSPECT model. *International Journal of Applied Earth Observation and Geoinformation, 74*, 180-190.
Huete, A. (2004). Remote sensing for environmental monitoring. In *Environmental Monitoring and Characterization* (pp. 183-206): Elsevier.
Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment, 25*(3), 295-309.
Hutchinson, T. C., & Hellebust, J. A. (1974). *Oil spills and vegetation at Norman Wells, NWT* (Vol. 73): Information Canada.
Ibaba, I. S. (2010). Environmental protection laws and sustainable development in the Niger Delta. *Africana, 4*(1).
Imoobe, T. O., & Iroro, T. (2009). Ecological restoration of oil spill sites in the Niger Delta, Nigeria. *Journal of Sustainable Development in Africa, 11*(2), 54-65.
International, A. (2009). *Fatal Flaws*: Amnesty International.
Ishida, T., Kurihara, J., Viray, F. A., Namuco, S. B., Paringit, E. C., Perez, G. J., . . . Marciano, J. J. (2018). A novel approach for vegetation classification using UAV-based hyperspectral imaging. *Computers and Electronics in Agriculture, 144*, 80-85.
James, G. K., Adegoke, J. O., Saba, E., Nwilo, P., & Akinyede, J. (2007). Satellite-based assessment of the extent and changes in the mangrove ecosystem of the Niger Delta. *Marine Geodesy, 30*(3), 249-267.
Jensen, J. R. (2009). *Remote sensing of the environment: An earth resource perspective 2/e*: Pearson Education India.
Jenssen, B. M. (1996). An overview of exposure to, and effects of, petroleum oil and organochlorine pollution in grey seals (Halichoerus grypus). *Science of The Total Environment, 186*(1-2), 109-118.
Jernelöv, A. (2018). Environmental Effects of Terrestrial Oil Spills.
Jhonnerie, R., Siregar, V. P., Nababan, B., Prasetyo, L. B., & Wouthuyzen, S. (2015). Random forest classification for mangrove land cover mapping using Landsat 5 TM and ALOS PALSAR imageries. *Procedia Environmental Sciences, 24*, 215-221.
Jin, H., Mountrakis, G., & Stehman, S. V. (2014). Assessing integration of intensity, polarimetric scattering, interferometric coherence and spatial texture metrics in PALSAR-derived land cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing, 98*, 70-84.
Jordan, C. F. (1969). Derivation of leaf-area index from quality of light on the forest floor. *Ecology, 50*(4), 663-666.
Joshi, N., Baumann, M., Ehammer, A., Fensholt, R., Grogan, K., Hostert, P., . . . Mitchard, E. T. (2016). A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring. *Remote Sensing, 8*(1), 70.
Judkoff, R., & Neymark, J. (1995). *International Energy Agency building energy simulation test (BESTEST) and diagnostic method*. Retrieved from
Juel, A., Groom, G. B., Svenning, J.-C., & Ejrnæs, R. (2015). Spatial application of Random Forest models for fine-scale coastal vegetation classification using object based analysis of aerial orthophoto and DEM data. *International Journal of Applied Earth Observation and Geoinformation, 42*, 106-114.
Kadafa, A. A. (2012). Environmental impacts of oil exploration and exploitation in the Niger Delta of Nigeria. *Global Journal of Science Frontier Research Environment & Earth Sciences, 12*(3), 19-28.
Kalubarme, M., & Sharma, A. VEGETATION WATER STRESS ASSESSMENT USING SHORT WAVE INFRARED (SWIR) INDICES IN WHEAT.

Kattenborn, T., Maack, J., Faßnacht, F., Enßle, F., Ermert, J., & Koch, B. (2015). Mapping forest biomass from space – Fusion of hyperspectral EO1-hyperion data and Tandem-X and WorldView-2 canopy height models. *International Journal of Applied Earth Observation and Geoinformation, 35*, 359-367. doi:10.1016/j.jag.2014.10.008

Khanna, S., Santos, M. J., Ustin, S. L., Koltunov, A., Kokaly, R. F., & Roberts, D. A. (2013). Detection of salt marsh vegetation stress and recovery after the Deepwater Horizon Oil Spill in Barataria Bay, Gulf of Mexico using AVIRIS data. *PLoS One, 8*(11), e78989. doi:10.1371/journal.pone.0078989

Khanna, S., Santos, M. J., Ustin, S. L., Shapiro, K., Haverkamp, P. J., & Lay, M. (2018). Comparing the potential of multispectral and hyperspectral data for monitoring oil spill impact. *Sensors, 18*(2), 558.

Khatami, R., & Mountrakis, G. (2018). The interacting effects of image acquisition date, number of images, classifier, and number of training samples on accuracy of binary classification of impervious cover. *Remote Sensing Letters, 9*(2), 189-198.

Klamerus-Iwan, A., Błońska, E., Lasota, J., Kalandyk, A., & Waligórski, P. (2015). Influence of oil contamination on physical and biological properties of forest soil after chainsaw use. *Water, Air, & Soil Pollution, 226*(11), 389.

Knipling, E. B. (1970). Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. *Remote Sensing of Environment, 1*(3), 155-159.

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

Kozlowski, T. (1997). Responses of woody plants to flooding and salinity. *Tree physiology, 17*(7), 490.

Kramer, P. J., & Boyer, J. S. (1995). *Water relations of plants and soils*: Academic press.

Kuenzer, C., van Beijma, S., Gessner, U., & Dech, S. (2014). Land surface dynamics and environmental challenges of the Niger Delta, Africa: Remote sensing-based analyses spanning three decades (1986–2013). *Applied Geography, 53*, 354-368.

Kühn, F., Oppermann, K., & Hörig, B. (2004). Hydrocarbon index–an algorithm for hyperspectral detection of hydrocarbons. *International Journal of Remote Sensing, 25*(12), 2467-2473.

Kuhn, M. (2012). The caret package. *R Foundation for Statistical Computing, Vienna, Austria. URL https://cran. r-project. org/package= caret*.

Ladan, M. T. (2012). Review of NESREA act 2007 and regulations 2009-2011: a new Dawn in environmental compliance and enforcement in Nigeria. *Law Env't & Dev. J., 8*, 116.

Lassalle, G., Credoz, A., Hédacq, R., Fabre, S., Dubucq, D., & Elger, A. (2018). Assessing Soil Contamination Due to Oil and Gas Production Using Vegetation Hyperspectral Reflectance. *Environmental science & technology, 52*(4), 1756-1764.

Laurin, G. V., Balling, J., Corona, P., Mattioli, W., Papale, D., Puletti, N., . . . Urban, M. (2018). Above-ground biomass prediction by Sentinel-1 multitemporal data in central Italy with integration of ALOS2 and Sentinel-2 data. *Journal of Applied Remote Sensing, 12*(1), 016008.

Lee, J.-S., & Pottier, E. (2009). *Polarimetric radar imaging: from basics to applications*: CRC press.

Leone, A., Menenti, M., Buondonno, A., Letizia, A., Maffei, C., & Sorrentino, G. (2007). A field experiment on spectrometry of crop response to soil salinity. *Agricultural Water Management, 89*(1-2), 39-48.

Li, L., Ustin, S. L., & Lay, M. (2005). Application of AVIRIS data in detection of oil-induced vegetation stress and cover change at Jornada, New Mexico. *Remote Sensing of Environment, 94*(1), 1-16.

Li, X., & Yeh, A. (1998). Principal component analysis of stacked multi-temporal images for the monitoring of rapid urban expansion in the Pearl River Delta. *International Journal of Remote Sensing, 19*(8), 1501-1518.

Lillesand, T., Kiefer, R. W., & Chipman, J. (2014). *Remote sensing and image interpretation*: John Wiley & Sons.

Lindén, O., & Pålsson, J. (2013). Oil contamination in Ogoniland, Niger delta. *Ambio, 42*(6), 685-701.

Lussem, U., Hütt, C., & Waldhoff, G. (2016). COMBINED ANALYSIS OF SENTINEL-1 AND RAPIDEYE DATA FOR IMPROVED CROP TYPE CLASSIFICATION: AN EARLY SEASON APPROACH FOR RAPESEED AND CEREALS. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, 41*.

MacLachlan, A., Roberts, G., Biggs, E., & Boruff, B. (2017). Subpixel land-cover classification for improved urban area estimates using Landsat. *International Journal of Remote Sensing, 38*(20), 5763-5792.

Mahdianpari, M., Salehi, B., Mohammadimanesh, F., Larsen, G., & Peddle, D. R. (2018). Mapping land-based oil spills using high spatial resolution unmanned aerial vehicle imagery and electromagnetic induction survey data. *Journal of Applied Remote Sensing, 12*(3), 036015.

Major, D. N., & Wang, H. (2012). How public health impact is addressed: a retrospective view on three different oil spills. *Toxicological & Environmental Chemistry, 94*(3), 442-467.

Mansaray, L. R., Huang, W., Zhang, D., Huang, J., & Li, J. (2017). Mapping rice fields in urban Shanghai, southeast China, using Sentinel-1A and Landsat 8 datasets. *Remote Sensing, 9*(3), 257.

Martinis, S., & Rieke, C. (2015). Backscatter analysis using multi-temporal and multi-frequency SAR data in the context of flood mapping at River Saale, Germany. *Remote Sensing, 7*(6), 7732-7752.

Mas, J.-F. (1999). Monitoring land-cover changes: a comparison of change detection techniques. *International Journal of Remote Sensing, 20*(1), 139-152.

Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: an applied review. *International Journal of Remote Sensing, 39*(9), 2784-2817. doi:10.1080/01431161.2018.1433343

McCown, D. D., & Deneke, F. J. (1972). *Plant germination and seedling growth as affected by the presence of crude petroleum.* Paper presented at the Symp. Proc. Impact of Oil Resource Development on Northern Plant Communities, Published by University of Alaska.

Mendelssohn, I. A., Andersen, G. L., Baltz, D. M., Caffey, R. H., Carman, K. R., Fleeger, J. W., . . . Overton, E. B. (2012). Oil impacts on coastal wetlands: implications for the Mississippi River Delta ecosystem after the Deepwater Horizon oil spill. *BioScience, 62*(6), 562-574.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2017). e1071: misc functions of the department of statistics, probability theory group (Formerly: E1071), TU Wien. R package version 1.6-8. In.

Millard, K., & Richardson, M. (2015). On the importance of training data sample selection in random forest image classification: A case study in peatland ecosystem mapping. *Remote Sensing, 7*(7), 8489-8515.

Mishra, D. R., Cho, H. J., Ghosh, S., Fox, A., Downs, C., Merani, P. B. T., . . . Mishra, S. (2012). Post-spill state of the marsh: Remote estimation of the ecological impact of the Gulf of Mexico oil spill on Louisiana Salt Marshes. *Remote Sensing of Environment, 118*, 176-185. doi:10.1016/j.rse.2011.11.007

Mitchard, E. T., Saatchi, S. S., Lewis, S., Feldpausch, T., Woodhouse, I. H., Sonké, B., . . . Meir, P. (2011). Measuring biomass changes due to woody encroachment and deforestation/degradation in a forest–savanna boundary region of central Africa using multi-temporal L-band radar backscatter. *Remote Sensing of Environment, 115*(11), 2861-2873.

Mmom, P. C., & Arokoyu, S. B. (2010). Mangrove forest depletion, biodiversity loss and traditional resources management practices in the Niger Delta, Nigeria. *Research Journal of Applied Sciences, Engineering and Technology, 2*(1), 28-34.

Mohammadimanesh, F., Salehi, B., Mahdianpari, M., Motagh, M., & Brisco, B. (2018). An efficient feature optimization for wetland mapping by synergistic use of SAR intensity, interferometry, and polarimetry data. *International Journal of Applied Earth Observation and Geoinformation, 73*, 450-462.

Murvay, P.-S., & Silea, I. (2012). A survey on gas leak detection and localization techniques. *Journal of Loss Prevention in the Process Industries, 25*(6), 966-973.

Mutanga, O., Adam, E., & Cho, M. A. (2012). High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. *International Journal of Applied Earth Observation and Geoinformation, 18*, 399-406.

Myneni, R. B., Hoffman, S., Knyazikhin, Y., Privette, J., Glassy, J., Tian, Y., . . . Smith, G. (2002). Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sensing of Environment, 83*(1-2), 214-231.

NDDC. (2004). Niger Delta Development Commission. *Biodiversity and Natural Resources Sector Study*.

NDDC. (2005). Niger Delta Development Commission [NDDC]. *Niger Delta Regional Development Master Plan, Port Harcourt, Nigeria*.

Ndimele, P. E., Saba, A. O., Ojo, D. O., Ndimele, C. C., Anetekhai, M. A., & Erondu, E. S. (2018). Remediation of Crude Oil Spillage. In *The Political Ecology of Oil and Gas Activities in the Nigerian Aquatic Ecosystem* (pp. 369-384): Elsevier.

Nicodemus, K. K., & Malley, J. D. (2009). Predictor correlation impacts machine learning algorithms: implications for genomic studies. *bioinformatics, 25*(15), 1884-1890.

Nigerian Meteorological Agency, N. (2010). Nigerian Climate Review.

Noomen, M., Hakkarainen, A., van der Meijde, M., & van der Werff, H. (2015). Evaluating the feasibility of multitemporal hyperspectral remote sensing for monitoring bioremediation. *International Journal of Applied Earth Observation and Geoinformation, 34*, 217-225. doi:10.1016/j.jag.2014.08.016

NPC. (2006). Population census and household data survey for Nigeria. *Abuja: Federal Government of Nigeria*.

Nriagu, J. (2011). Oil Industry and the Health of Communities in the Niger Delta of Nigeria. *Encyclopedia of Environmental Health*, 240-250.

Numbisi, F. N., Van Coillie, F. M., & De Wulf, R. (2019). Delineation of Cocoa Agroforests Using Multi-2 Season Sentinel-1 SAR Images: Low Grey Level 3 Range Reduces Uncertainties in GLCM Texture-4 Based Mapping 5.

O'Rourke, D., & Connolly, S. (2003). Just oil? The distribution of environmental and social impacts of oil production and consumption. *Annual Review of Environment and Resources, 28*(1), 587-617.

Obida, C. B., Blackburn, G. A., Whyatt, J. D., & Semple, K. T. (2018). Quantifying the exposure of humans and the environment to oil pollution in the Niger Delta using advanced geostatistical techniques. *Environment international, 111*, 32-42.

Ochege, F. U., George, R. T., Dike, E. C., & Okpala-Okaka, C. (2017). Geospatial assessment of vegetation status in Sagbama oilfield environment in the Niger Delta region, Nigeria. *The Egyptian Journal of Remote Sensing and Space Science, 20*(2), 211-221.

Ogbonna, G., & Ebimobowei, A. (2012). Petroleum income and Nigerian economy: Empirical evidence. *Oman Chapter of Arabian Journal of Business and Management Review, 34*(965), 1-27.

Ohsol, E. O., Pinkerton, J. W., Gillespie, T. E., & Laity, T. H. (1999). Process for upgrading heavy crude oil production. In: Google Patents.

Okoro, U., Chen, W., Chineke, T., & Nwofor, O. (2014). Recent monsoon rainfall characteristics over the Niger Delta Region of Nigeria: A casual link. *International Journal of Science, Environment, and Technology, 3*(2), 634-651.

Ologunorisa, T. E., & Adeyemo, A. (2005). Public perception of flood hazard in the Niger Delta, Nigeria. *Environmentalist, 25*(1), 39-45.

OLOGUNORISA*, T. E. (2004). An assessment of flood vulnerability zones in the Niger Delta, Nigeria. *International journal of environmental studies, 61*(1), 31-38.

Omo-Irabor, O. O., Olobaniyi, S. B., Akunna, J., Venus, V., Maina, J. M., & Paradzayi, C. (2011). Mangrove vulnerability modelling in parts of Western Niger Delta, Nigeria using satellite images, GIS techniques and Spatial Multi-Criteria Analysis (SMCA). *Environmental monitoring and assessment, 178*(1-4), 39-51.

Onojeghuo, A. O., Blackburn, G. A., Wang, Q., Atkinson, P. M., Kindred, D., & Miao, Y. (2018a). Mapping paddy rice fields by applying machine learning algorithms to multi-temporal Sentinel-1A and Landsat data. *International Journal of Remote Sensing, 39*(4), 1042-1067.

Onojeghuo, A. O., Blackburn, G. A., Wang, Q., Atkinson, P. M., Kindred, D., & Miao, Y. (2018b). Rice crop phenology mapping at high spatial and temporal resolution using downscaled MODIS time-series. *GIScience & Remote Sensing, 55*(5), 659-677.

Onuoha, F. (2007). Poverty, pipeline vandalisation/explosion and human security: Integrating disaster management into poverty reduction in Nigeria. *African Security Studies, 16*(2), 94-108.

Onwurah, I., Ogugua, V., Onyike, N., Ochonogor, A., & Otitoju, O. (2007). Crude oil spills in the environment, effects and some innovative clean-up biotechnologies.

Onyia, N. N., Balzter, H., & Berrio, J.-C. (2018). Normalized Difference Vegetation Vigour Index: A New Remote Sensing Approach to Biodiversity Monitoring in Oil Polluted Regions.

Orisakwe, O. E. (2009). Environmental pollution and blood lead levels in Nigeria: who is unexposed? *International journal of occupational and environmental health, 15*(3), 315-317.

Orlov, D., Ammosova, I., & Bocharnikova, E. (1993). *Remote monitoring of oil-polluted soils using soil reflectance spectra.* Paper presented at the CIS Selected Papers: Optical Monitoring of the Environment.

Osuji, L. C., & Opiah, U. C. (2007). Hydrocarbon contamination of a terrestrial ecosystem: the case of Oshire-2 oil spill in Niger Delta, Nigeria. *The Environmentalist, 27*(3), 337-340.

Osunmadewa, B. A., Gebrehiwot, W. Z., Csaplovics, E., & Adeofun, O. C. (2018). Spatio-temporal monitoring of vegetation phenology in the dry sub-humid region of Nigeria using time series of AVHRR NDVI and TAMSAT datasets. *Open Geosciences, 10*(1), 1-11.

Ouarzeddine, M., Souissi, B., & Belhadj-Aissa, A. (2007). *Unsupervised classification using Wishart classifier*. Paper presented at the Proceedings of the 3rd International Workshop on Science and Applications of SAR Polarimetry and Polarimetric Interferometry.

Pampaloni, P., & Paloscia, S. (1985). *Microwave emission from vegetation: general aspects and experimental results.* Paper presented at the Proc. EARSeL Workshop on Microwave Remote Sensing Applied to Vegetation.

Park, Y. S., Al-Qublan, H., Lee, E., & Egilmez, G. (2016). *Interactive Spatiotemporal Analysis of Oil Spills Using Comap in North Dakota.* Paper presented at the Informatics.

Puissant, A., Rougier, S., & Stumpf, A. (2014). Object-oriented mapping of urban trees using Random Forest classifiers. *International Journal of Applied Earth Observation and Geoinformation, 26*, 235-245. doi:10.1016/j.jag.2013.07.002

Qi, J., Chehbouni, A., Huete, A., Kerr, Y., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment, 48*(2), 119-126.

Rajah, P., Odindi, J., & Mutanga, O. (2018). Feature level image fusion of optical imagery and Synthetic Aperture Radar (SAR) for invasive alien plant species detection and mapping. *Remote Sensing Applications: Society and Environment, 10*, 198-208.

Ramsey, E., 3rd, Meyer, B. M., Rangoonwala, A., Overton, E., Jones, C. E., & Bannister, T. (2014). Oil source-fingerprinting in support of polarimetric radar mapping of Macondo-252 oil in Gulf Coast marshes. *Mar Pollut Bull, 89*(1-2), 85-95. doi:10.1016/j.marpolbul.2014.10.032

Ramsey, E., Rangoonwala, A., & Jones, C. (2015). Structural Classification of Marshes with Polarimetric SAR Highlighting the Temporal Mapping of Marshes Exposed to Oil. *Remote Sensing, 7*(9), 11295-11321. doi:10.3390/rs70911295

Ramsey III, E., & Rangoonwala, A. (2015). 8 Radar and Optical Image Fusion and Mapping of Wetland Resources. *Remote Sensing of Wetlands: Applications and Advances*, 155.

Ramsey III, E., Rangoonwala, A., Suzuoki, Y., & Jones, C. E. (2011). Oil Detection in a Coastal Marsh with Polarimetric Synthetic Aperture Radar (SAR). *Remote Sensing, 3*(12), 2630-2662. doi:10.3390/rs3122630

Reséndez-Hernández, L. A., Prudencio-Csapek, D., & Lozano-García, D. F. (2018). Hyperspectral Analysis of soil polluted with four types of hydrocarbons. *Geocarto International*, 1-18.

Ribeiro, H., Almeida, C. M. R., Mucha, A. P., & Bordalo, A. A. (2013). Influence of different salt marsh plants on hydrocarbon degrading microorganisms abundance throughout a phenological cycle. *International journal of phytoremediation, 15*(8), 715-728.

Rodrigue, J.-P. (2009). The Geography of Transport Systems. *[online] Available at:* <*http://people.hofstra.edu/geotrans/index.html*>.

Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. (2015). Machine learning predictive models for mineral prospectivity: An

evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews, 71*, 804-818.

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing, 67*, 93-104. doi:10.1016/j.isprsjprs.2011.11.002

Rosso, P. H., Pushnik, J. C., Lay, M., & Ustin, S. L. (2005). Reflectance properties and physiological responses of Salicornia virginica to heavy metal and petroleum contamination. *Environmental pollution, 137*(2), 241-252.

Roujean, J.-L., & Breon, F.-M. (1995). Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sensing of Environment, 51*(3), 375-384.

Rouse Jr, J. W., Haas, R., Schell, J., & Deering, D. (1974). Monitoring vegetation systems in the Great Plains with ERTS.

Roy, D. P., Wulder, M., Loveland, T. R., Woodcock, C., Allen, R., Anderson, M., . . . Kennedy, R. (2014). Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment, 145*, 154-172.

Rüetschi, M., Schaepman, M. E., & Small, D. (2017). Using Multitemporal Sentinel-1 C-band Backscatter to Monitor Phenology and Classify Deciduous and Coniferous Forests in Northern Switzerland. *Remote Sensing, 10*(1), 55.

Samat, A., Gamba, P., Liu, S., Miao, Z., Li, E., & Abuduwaili, J. (2018). Quad-PolSAR data classification using modified random forest algorithms to map halophytic plants in arid areas. *International Journal of Applied Earth Observation and Geoinformation, 73*, 503-521.

Sarma, P. C. (2012). Degradation of Petroleum Fractions in Soil Under Natural Environment: A Gravimetric and Gas Chromatographic Analysis. In *Crude Oil Exploration in the World*: InTech.

Schlund, M., von Poncet, F., Kuntz, S., Boehm, H.-D. V., Hoekman, D. H., & Schmullius, C. (2016). TanDEM-X elevation model data for canopy height and aboveground biomass retrieval in a tropical peat swamp forest. *International Journal of Remote Sensing, 37*(21), 5021-5044.

Schmidt, J., Fassnacht, F. E., Förster, M., & Schmidtlein, S. (2017). Synergetic use of Sentinel-1 and Sentinel-2 for assessments of heathland conservation status. *Remote Sensing in Ecology and Conservation*.

Schumacher, D. (1996). Hydrocarbon-induced alteration of soils and sediments.

Shang, J., McNairn, H., Champagne, C., & Jiao, X. (2009). Application of multi-frequency synthetic aperture radar (SAR) in crop classification. In *Advances in geoscience and remote sensing*: IntechOpen.

Shapiro, K., Khanna, S., & Ustin, S. L. (2016). Vegetation impact and recovery from oil-induced stress on three ecologically distinct wetland sites in the Gulf of Mexico. *Journal of Marine Science and Engineering, 4*(2), 33.

Shi, L., Zhang, X., Seielstad, G., Zhao, C., & He, M.-X. (2007). *Oil spill detection by MODIS images using fuzzy cluster and texture feature extraction.* Paper presented at the OCEANS 2007-Europe.

Shittu, W. J. a. (2014). *Mapping oil spill human health risk in rivers state, Niger Delta, Nigeria.* University of Nottingham,

Siegel, F. R. (1974). *Applied geochemistry*.

Singh, A. (1989). Review article digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing, 10*(6), 989-1003.

Smith, M. O., Ustin, S. L., Adams, J. B., & Gillespie, A. R. (1990). Vegetation in deserts: I. A regional measure of abundance from multispectral images. *Remote Sensing of Environment, 31*(1), 1-26.

Solberg, S., May, J., Bogren, W., Breidenbach, J., Torp, T., & Gizachew, B. (2018). Interferometric SAR DEMs for Forest Change in Uganda 2000–2012. *Remote Sensing, 10*(2), 228.

Son, N.-T., Chen, C.-F., Chen, C.-R., & Minh, V.-Q. (2018). Assessment of Sentinel-1A data for rice crop classification using random forests and support vector machines. *Geocarto International, 33*(6), 587-601.

Sorgwe, C. (1997). *A Textbook of Niger Delta History:(from the Earliest Times to the Present)*: Rescue Publications.

Spies, B., Lamb, A., Brown, S., Balzter, H., & Fisher, P. (2013). *Comparison Of Multi-Frequency SAR Land Cover Signatures For Multi-Site Semi-Arid Regions Of Africa.* Paper presented at the ESA Living Planet Symposium.

Sportouche, H., Tupin, F., Nicolas, J.-M., Perciano, T., & Deledalle, C.-A. (2012). *How to combine TerraSAR-X and Cosmo-SkyMed high-resolution images for a better scene understanding?* Paper presented at the Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International.

Stagakis, S., Markos, N., Sykioti, O., & Kyparissis, A. (2010). Monitoring canopy biophysical and biochemical parameters in ecosystem scale using satellite hyperspectral imagery: An application on a Phlomis fruticosa Mediterranean ecosystem using multiangular CHRIS/PROBA observations. *Remote Sensing of Environment, 114*(5), 977-994.

Stankevich, S. A., Kozlova, A. A., Piestova, I. O., & Lubskyi, M. S. (2017). *Leaf area index estimation of forest using sentinel-1 C-band SAR data.* Paper presented at the Microwaves, Radar and Remote Sensing Symposium (MRRS), 2017 IEEE.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC bioinformatics, 9*(1), 307.

Sun, S., Hu, C., Feng, L., Swayze, G. A., Holmes, J., Graettinger, G., . . . Leifer, I. (2016). Oil slick morphology derived from AVIRIS measurements of the Deepwater Horizon oil spill: Implications for spatial resolution requirements of remote sensors. *Mar Pollut Bull, 103*(1-2), 276-285. doi:10.1016/j.marpolbul.2015.12.003

Susantoro, T., Wikantika, K., Saepuloh, A., & Harsolumakso, A. (2018). *Utilization of Vegetation Indices to Interpret the Possibility of Oil and Gas Microseepages at Ground Surface.* Paper presented at the IOP Conference Series: Earth and Environmental Science.

Taiwo, E., Otolorin, J., & Afolabi, T. (2012). Crude oil transportation: Nigerian Niger delta waxy crude. In *Crude oil exploration in the world*: IntechOpen.

TeamR, R. C. (2017). A language and environment for statistical computingVienna. *AustriaR Foundation for Statistical Computing*.

Tedesco, S. A. (2012). *Surface geochemistry in petroleum exploration*: Springer Science & Business Media.

Thenkabail, P. S., Smith, R. B., & De Pauw, E. (2000). Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sensing of Environment, 71*(2), 158-182.

Therneau, T., Atkinson, B., & Ripley, B. (2015). rpart: Recursive partitioning and regression trees. *R package version, 4*, 1-9.

Tolulope, A. O. (2004). Oil exploration and environmental degradation: the Nigerian experience. *Environmental Informatics Archives, 2*, 387-393.

Tucker, C. J. (1980). A spectral method for determining the percentage of green herbage material in clipped samples. *Remote Sensing of Environment, 9*(2), 175-181.
Tuttle, M. L., Charpentier, R. R., & Brownfield, M. E. (1999). *The Niger Delta Petroleum System: Niger Delta Province, Nigeria, Cameroon, and Equatorial Guinea, Africa*: US Department of the Interior, US Geological Survey.
Twumasi, Y., & Merem, E. (2006). GIS and remote sensing applications in the assessment of change within a coastal environment in the Niger Delta region of Nigeria. *International journal of environmental research and public health, 3*(1), 98-106.
Ugochukwu, C. N., & Ertel, J. (2008). Negative impacts of oil exploration on biodiversity management in the Niger De area of Nigeria. *Impact assessment and project appraisal, 26*(2), 139-147.
Ukoli, M. (2005). Environmental factors in the management of the oil and gas industry in Nigeria. *Available on http://www. warmofloor. co. uk/pages/environmenta lpdf [Accessed on 2/12/2005]*.
Ulaby, F. T., Moore, R. K., & Fung, A. K. (2015). Microwave remote sensing active and passive.
UNEP. (2011). Environmental Assessment of Ogoniland. In: UNEP Nairobi.
van Beijma, S., Comber, A., & Lamb, A. (2014). Random forest classification of salt marsh vegetation habitats using quad-polarimetric airborne SAR, elevation and optical RS data. *Remote Sensing of Environment, 149*, 118-129.
Van der Werff, H., Noomen, M., Van der Meijde, M., & Van der Meer, F. (2007). Remote sensing of onshore hydrocarbon seepage: problems and solutions. *Geological Society, London, Special Publications, 283*(1), 125-133.
Vreugdenhil, M., Wagner, W., Bauer-Marschallinger, B., Pfeil, I., Teubner, I., Rüdiger, C., & Strauss, P. (2018). Sensitivity of Sentinel-1 Backscatter to Vegetation Dynamics: An Austrian Case Study. *Remote Sensing, 10*(9), 1396.
Wang, Y., Feng, J., Lin, Q., Lyu, X., Wang, X., & Wang, G. (2013). Effects of crude oil contamination on soil physical and chemical properties in Momoge wetland of China. *Chinese geographical science, 23*(6), 708-715.
Waske, B., van der Linden, S., Oldenburg, C., Jakimow, B., Rabe, A., & Hostert, P. (2012). imageRF – A user-oriented implementation for remote sensing image analysis with Random Forests. *Environmental Modelling & Software, 35*, 192-193. doi:10.1016/j.envsoft.2012.01.014
Whanda, S., Adekola, O., Adamu, B., Pandey, P., Ogwu, F., & Yahaya, S. (2016). Geo-spatial analysis of oil spill distribution and susceptibility in the Niger Delta region of Nigeria. *Journal of Geographic Information System, 8*, 438-456.
Whanda, S., Sani, Y., & Bulus, G. (2015). Modelling of Potential Pipeline Impact Radius and High Consequence Area in a Wetland Sub-Region of Nigeria. *Journal of Geographic Information System, 7*(06), 692.
White, D., Williams, M., & Barr, S. (2008). Detecting sub-surface soil disturbance using hyperspectral first derivative band rations of associated vegetation stress. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci, 27*, 243-248.
Whyte, A., Ferentinos, K. P., & Petropoulos, G. P. (2018). A new synergistic approach for monitoring wetlands using Sentinels-1 and 2 data with object-based machine learning algorithms. *Environmental Modelling & Software, 104*, 40-54.
Woodhouse, I. H. (2005). *Introduction to microwave remote sensing*: CRC press.
Xiao, Q., & McPherson, E. G. (2005). Tree health mapping with multispectral remote sensing data at UC Davis, California. *Urban Ecosystems, 8*(3-4), 349-361.
Xue, J., & Su, B. (2017). Significant remote sensing vegetation indices: a review of developments and applications. *Journal of Sensors, 2017*.

Yajima, Y., Yamaguchi, Y., Sato, R., Yamada, H., & Boerner, W.-M. (2008). POLSAR image analysis of wetlands using a modified four-component scattering power decomposition. *IEEE Transactions on Geoscience and Remote Sensing, 46*(6), 1667-1673.

Yamaguchi, Y., Moriyama, T., Ishido, M., & Yamada, H. (2005). Four-component scattering model for polarimetric SAR image decomposition. *IEEE Transactions on Geoscience and Remote Sensing, 43*(8), 1699-1706.

Yang, H., Meer, F., Zhang, J., & Kroonenberg, S. (2000). Direct detection of onshore hydrocarbon microseepages by remote sensing techniques. *Remote sensing reviews, 18*(1), 1-18.

Ye, X., Zhang, H., Wang, C., Wu, F., Zhang, B., & Tang, Y. (2013). *Classification of high-resolution SAR imagery by Random Forest classifier.* Paper presented at the Synthetic Aperture Radar (APSAR), 2013 Asia-Pacific Conference on.

Younes, M. (2012). Crude Oil Geochemistry Dependent Biomarker Distributions in the Gulf of Suez, Egypt. In *Crude Oil Exploration in the World*: InTech.

Zabbey, N., & Uyi, H. (2014). Community responses of intertidal soft-bottom macrozoobenthos to oil pollution in a tropical mangrove ecosystem, Niger Delta, Nigeria. *Mar Pollut Bull, 82*(1-2), 167-174. doi:10.1016/j.marpolbul.2014.03.002

Zhou, R., Qin, X., Peng, S., & Deng, S. (2014). Total petroleum hydrocarbons and heavy metals in the surface sediments of Bohai Bay, China: long-term variations in pollution status and adverse biological risk. *Mar Pollut Bull, 83*(1), 290-297. doi:10.1016/j.marpolbul.2014.03.003

Zhu, L., Zhao, X., Lai, L., Wang, J., Jiang, L., Ding, J., . . . Xiao, N. (2013). Soil TPH concentration estimation using vegetation indices in an oil polluted area of eastern China. *PLoS One, 8*(1), e54028.

Zhu, L., Zhao, X., Lai, L., Wang, J., Jiang, L., Ding, J., . . . Rimmington, G. M. (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

Zlinszky, A., Mücke, W., Lehner, H., Briese, C., & Pfeifer, N. (2012). Categorizing wetland vegetation by airborne laser scanning on Lake Balaton and Kis-Balaton, Hungary. *Remote Sensing, 4*(6), 1617-1650.

Zuhlke, M., Fomferra, N., Brockmann, C., Peters, M., Veci, L., Malik, J., & Regner, P. (2015). *SNAP (sentinel application platform) and the ESA sentinel 3 toolbox.* Paper presented at the Sentinel-3 for Science Workshop.