Knowledge and Data Driven Approaches for Hydrocarbon Microseepage Characterizations: An Application of Satellite Remote Sensing

Ali Al Farid

Ali Al Farid. Knowledge and Data Driven Approaches for Hydrocarbon Microseepage Characterizations: An Application of Satellite Remote Sensing

Master degree thesis, 30 credits in Master in Geographical Information Science
Department of Physical Geography and Ecosystem Science, Lund University

Knowledge and Data Driven Approaches for Hydrocarbon Microseepage Characterizations: An Application of Satellite Remote Sensing

—————————————————————————————————

Ali Al Farid

Master Thesis, 30 credits, in Geographical Information Science

Supervisor: Jonas Ardö

Lund University

Abstract

Methane is a potent greenhouse gas that plays a major role in climate change. Significant uncertainty exists in the estimates of emissions from natural methane sources. This uncertainty in data is one of the primary scientific challenges in climate model. Geological seepage considered as the second largest natural source of methane after wetlands and today is recognized as a major contributor to atmospheric methane. This work focuses on positive methane fluxes to the atmosphere from sedimentary basins hosting natural gas and oil reservoirs termed as microseepage. Microseepage contributes to the global atmospheric methane budget and creates large uncertainties in the global methane atmospheric budget estimates (sources and sinks). The global coverage of methane microseepage is unknown, and data available today is based on estimates. With respect to global and regional estimates, the level of microseepage emission was established by assuming, a priori, that the full area of petroleum basins in dry climate produces positive fluxes of methane into the atmosphere. This assumption is subject to considerable uncertainties because microseepage does not occur throughout the entire petroleum field area. In this context, satellite remote sensing imagery was used to investigation areas affected by natural hydrocarbon microseepage using knowledge and data-driven based approaches. Knowledge-based approaches constructed based on the theoretical model established from literature targeting specific minerals and surface manifestations. The specific mineral groups or features were determined based on their reflectance and absorption characteristics. Methods in knowledge-driven approach include Band Ratio (BR), Principle Component Analyses (PCA) with Crosta technique and Mixture Tuned Matching Filtering (MTMF) classification. The data-driven approach used Support Vector Machine (SVM) algorithm. The SVM model was purely estimated from the multispectral image data based on occurrence and abundance of oil and dry hole wells and without any prior assumption of area mineralogical assemblage. The mapped microseepage extent exhibit some level of consistency between all models. Results were satisfactory enough judged by the level of consistency realized between all models for the mapped microseepage extent, which indicates statistical significance. The data-driven model yields best results, the gain in performance from using data-driven approach as compared to knowledge-based was relatively small. However, the long processing steps and time in the knowledge-based approaches gives merits to the data driven approach. The work demonstrated the potential of satellite remote sensing and its analysis in mapping hydrocarbons microseepage extent on regional or even global scale.

Keywords: Geography, Geographical Information Systems, Hydrocarbon Microseepage, PCA, Crosta, Band Ratio, MTMF, Weight of Evidence, SVM, Methane Fluxes, Remote Sensing, Satellite Image Classification, Earth Observation, Mineral Classifications, Machine Learning.

Acknowledgements

I wish to express my sincere appreciation and gratitude to my supervisor Prof. Jonas Ardö, for his guidance, advice, great patience, encouragements and insight throughout the research.

I am very grateful to Lund University for the great opportunity and to all my teachers, I would like to thank each person for their generous help.

Finally, I want to thank my family for their non-stop love, constant care, patience and support over the years.

Table of contents

List of tables

List of figures

Acronyms and abbreviations

AdaBoost	Adaptive Boosting
AEA	Albers Equal Area
AR5	Assessment Report 5
ARD	Analysis Ready Data
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
AUC	Area Under the Curve
BR	Band Ratio
BRT	Boosted Regression Tree
CAPP	Cumulative Area-Posterior Probability
CONUS	Conterminous United States
GBRT	Gradient Boosted Regression Tree
IPCC	Intergovernmental Panel on Climate Change
LTS	Land surface temperature
MBOGC	Montana Board of Oil and Gas Conservation
ML	Machine learning
MNF	Minimum Noise Fraction
MODIS	Moderate Resolution Imaging Spectroradiometer
MTMF	Mixture Tuned Matching Filtering
nDV	n-Dimensional Visualizer
NDVI	Normalized Difference Vegetation Index
OLI	Operational Land Imager
PCA	Principle Component Analysis
PPI	Pixel Purity Index
ROC	Receiver Operating Characteristic
ROI	Region of Interest
SCIAMACHY	Scanning Imaging Absorption Spectrometer for Atmospheric Cartography
SVM	Support Vector Machines
SWIR	Short-wavelength infrared
Tg yr	Teragrams of million tonnes per year
TIRS	Thermalinfrared Sensor
TM	Thematic Mapper
USGS	United States Geological Survey
VNIR	Visable and Near-Infrared
VOG	Vogelmann Indeces
WELD	Web-Enabled Landsat Data
WGS84	World Geodetic System 84
WofE	Weight of Evidence

1   Introduction

1.1   Background and motivation

Methane (CH_4) is a potent greenhouse gas that plays a major role in climate change. It is the second most important greenhouse gas in terms of radiative forcing after carbon dioxide (CO_2). Methane is released into the atmosphere through natural and anthropogenic sources. Significant uncertainty exists in the estimates of emissions from natural methane sources (Etiope, 2004, Etiope et al., 2008, Christensen et al., 2019, Van Amstel, 2012, Turner et al., 2019). This uncertainty in data is one of the primary scientific challenges in climate change model. Geological seepage considered as the second largest natural source of methane after wetlands (Etiope, 2015, Tang et al., 2017), and estimated to account for 15% of global natural methane emission (Ciais et al., 2013). Geological sources have wider classes including mud volcanoes, macro and micro seepage, geothermal and marine seepage (Etiope and Klusman, 2010a, Etiope, 2015, Etiope et al., 2019). This work focuses on positive methane fluxes to the atmosphere from sedimentary basins hosting natural gas and oil reservoirs termed as microseepage. Drylands which are part of petroleum basins and sedimentary basins may not show methane uptake or sink but positive exhalation into the atmosphere ranging from a few units to tens of mg m⁻² d⁻¹ (Etiope and Klusman, 2010a). Microseepage contributes to the global atmospheric methane budget and creates large uncertainties in the global methane atmospheric budget estimates both sources and sinks (Etiope and Klusman, 2010b). The global coverage of methane microseepage is unknown (Etiope, 2005). The data available today is based on estimates driven by gridded database or by extrapolating the global estimates through multiplying average field measurements by estimates of area covered (Potter et al., 1996, Etiope and Klusman, 2010b, Etiope et al., 2019). With respect to global and regional estimates, the level of microseepage emission was established by assuming, a priori, that the full area of petroleum basins in dry climate produce positive fluxes of methane into the atmosphere (Etiope, 2005, Etiope and Klusman, 2010a). This assumption is subject to considerable uncertainties because microseepage does not occur throughout the entire petroleum field area. Therefore a correct assessment of the natural release of methane from different oil/gas basins are still of great value for improving the existing source-sink balance model. Here it is propose to use knowledge-driven and data-driven mapping approaches based on satellite remote sensing to provide improved overall regional estimates of microseepage extent. Mapping the boundary of a microseepage can help in a better understanding of methane positive fluxes spatial distribution, and then facilitates for more accurate predictions of soil uptake and sink which in turn leads to an improved atmospheric CH_4 modelling.

Hydrocarbon microseepages from Earth's subsurface to the surface and their chemical, physical and biological interactions are well documented (Brown, 2000, Etiope, 2015). The chemical and physical interactions have always been an area of interest to geologists in the field of petroleum explorations as an important driver for new hydrocarbon potential. These interactions are studied under surface geochemistry (chemical) and structure geology (physical).

The fluxes of gases in the atmosphere resulting from hydrocarbons microseepage were always ignored and considered a negligible source of atmospheric methane. Hydrocarbon microseepage significance as atmospheric CH_4 source was initially reported by Klusman et al. (1998) and Klusman and Jakel (1998). Until the fifth Intergovernmental Panel on Climate Change (IPCC) assessment reports (AR5) that was issued in 2013 (Ciais et al., 2013) hydrocarbons microseepage was only considered as minor contributor to atmospheric methane, and was categorized under geological seepage. More recently, microseepage from hydrocarbons become an area of interest to the climate change and atmospheric science community.

It has been difficult to determine the extent of microseepage and the rate of methane emission from dry lands. Uncertainties arise due to incomplete knowledge of the actual area of microseepage. Accurate and more reliable estimates can be obtained by performing as many as possible ground measurements from many different petroleum basins (Tang et al., 2017), an exercise that pose lots of challenges at regional or global scales.

Methane emission from petroleum basins in dry climate remain ill-quantified (Reay et al., 2010, Scafutto et al., 2018, Etiope and Klusman, 2010b) despite decades of research. For instance, the global potential microseepage has been projected to be 7 Tg yr^{-1} (Klusman et al., 2000, Klusman et al., 1998) and 25 Tg yr^{-1} was suggested by Etiope and Klusman (2010a). Recent evaluations based on grid data modelling estimated the global gridded CH_4 emission was 37.4±17.6 Tg yr^{-1} with hydrocarbon microseepage account for a total of 24(±9) Tg yr^{-1} (Etiope et al., 2019). The wide range of projection indicates that there is still substantial uncertainty of microseepage estimates.

In this context, remote sensing methods have been modelled and tested to identify hydrocarbon microseepage affected areas through the characterization of spectral features of microseepage in image data. The objective of the study is to investigate a comprehensive integrated multi-method approach using satellite remote sensing in an attempt to improve the success rate of finding a localized alteration induced by onshore hydrocarbon microseepage. Such model can then facilitate evaluating the contribution of every sedimentary basin to natural methane inventory.

1.2 Research questions and Aims

The global emission from microseepage remains ill-quantified proven by wide range of projections and conflicting estimates. The discrepancies between estimates have been described and discussed in a wide body of literature (for example, (Reay et al., 2010, Scafutto et al., 2018, Etiope and Klusman, 2010b, Asadzadeh and de Souza Filho, 2017)). Uncertainties arise due to incomplete knowledge of the actual area of microseepage (Etiope and Klusman, 2010b). As it is hard to measure every drylands, this thesis aims to develop methods in satellite remote sensing that can assist in the detection of regional areas affected by hydrocarbon microseepage.

The following research questions form the basis of the thesis:

1- Comparing knowledge-driven based methods PCA with Crosta technique and Band Ratio; which method is the most adopted to the determination of microseepage-induced surface alterations?

2- Are there any significant difference between knowledge-driven and data-driven approaches in hydrocarbon microseepage characterizations? How do microseepage extent vary?

3- Mixed spectral response from different materials on the ground is a common problem with low/moderate spatial resolution images. What possible unmixing approaches that can decompose the mixed pixels into a collection of pure spectral signatures?

4- What is applicability of Landsat-8 OLI/TIRS data on the application of hydrocarbon microseepage characterization?

1.3 Thesis outline

The rest of this paper is organized as follows:

- Section 2: provides background about the study area including some case histories.
- Section 3: reviews the formation of seepage and the role of remote sensing in mapping alterations.
- Section 4: describes the used dataset.
- Section 5: explains the methods used.
- Section 6: shows analysis results and describes models validations
- Section 7: provides general discussions.
- Section 8: provides a conclusions, as well as suggestions for future work.

2 Study Area

2.1 Geographic position

This study was conducted on the Fort Peck Indian Reservation located in the north-eastern part of Montana State in United State at 48°21′40″N 105°32′18″W (Figure 1). The reservation is 177 km long (east to west) and 64 km wide (north to south). It extends over four counties and occupies around 8287 km². The reservation altitude ranges from about 930 meters in the north-western part of the reservation to less than 579 meters in the south-eastern part. The Missouri River is the largest stream in the area and it extends over the southern boundary of the reservation. Creeks cross the reservation: Porcupine Creek flows on the reservation western boundary and Big Muddy Creek on the east boundary. At the central of the reservation the Poplar River flows south across the reservation to join the Missouri River.
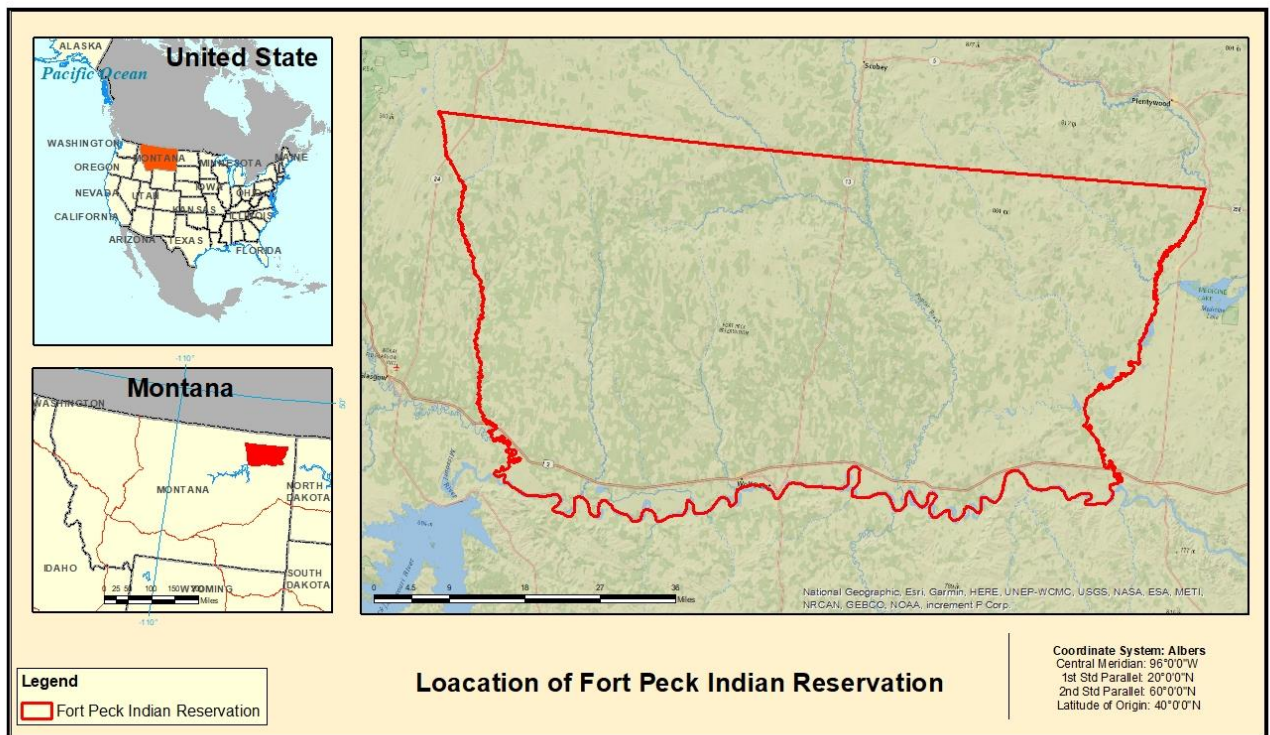
Figure 1 – Geographical location of the study area.

2.2 Climate and Topography

The reservation has a semi-arid, continental climate classified as BSk (cold semi-arid) under the Koppen climate classification system. It is situated within north-western Great Plains and is typically dominated by shrub and dwarf-shrub species. The region is characterized by three terrain types. The general terrain of the area is rolling to rough plains, Badlands broken by hills and isolated mountain ranges primarily used for livestock grazing. The second type of terrain found on the southern part of the reservation near the major Missouri river tributaries. It is composed of a high bench lands encompassing farmland areas characterized by relatively fertile soils used for production of small grains and cattle grazing. The third terrain type is

situated towards the centre of the reservation on the Missouri River valley bottom and it is characterized by irrigated farmland with deciduous trees (Thamke, 1991).

2.3 Geology

The Fort Peck Indian Reservation is located on the western part of the Williston Basin. The outcropping sedimentary rocks on reservation are mostly shale and siltstone with some bids of sandstone, salt, bentonite and lignite (Mudge et al., 1977). Major sedimentary strata underlying the area are: Bearpaw Shale which is only exposed at the southern and western part of the reservation and it forms a terrain of small hills and Badlands topography. The Fox Hills sandstone which overlies the Bearpaw and it is exposed along the south-central part of the reservation. The Hell Creek formation which overlies the Fox Hills sandstone and it is exposed along the central and north central part of the reservation. The Fort Union formation which overlies the Hell Creek and forms a terrain of rolling landscape covered by grass with some sandstone benches at the eastern part of the reservation. Deposits of sand and gravel are widespread in the reservation. However, the Flaxville Gravel caps highland areas extended in much of the central and north-western parts of the reservation and contains the largest and most extensive deposit of sand and gravel (Mudge et al., 1977).

Hydrocarbons in Williston basin province is mostly produced from three formations that belong to Mississippian – Devonian period. The lowermost is Three Forks formation that is immediately overlain by Bakken shale formation which underlies the Lodgepole formation. These formations cross the entire area of the reservation area and produce petroleum in the Williston Basin (Gaswirth and Marra, 2015).The main structural feature within the reservation include Poplar Dome, a northwest-trending asymmetrical anticline. The western part of the reservation is dominated by the Eastern Flank of the Bowdoin Dome. The Brockton-Froid fault zone extends northeast across the southeast corner. Other structural features include the Wolf Creek Nose, the Oswego, and the Bredette Nose (Mudge et al., 1977).

2.4 Previous investigations and case histories

Monson and Shurr (1993) mapped circular satellite tonal anomalies features also called curvilinears on Landsat images and suggested that these could represent micro-seepage chimneys. The subtle tonal anomalies is a concept used in hydrocarbon satellite remote sensing literatures (Zhang et al., 2011, Almeida-Filho et al., 1999, Schumacher, 1996, Yang et al., 2000) to explain a phenomena or unusual colour tone observed on satellite images resembling blurring of the image at petroleum concentrations areas. The phenomena was also described as a peculiar hazy anomaly (Asadzadeh and de Souza Filho, 2017, Simpson, 1978). These curvilinears were the primary reason for surface exploration work in the Fort Peck Reservation (Monson, 2000, 2001, 2003). Different surface hydrocarbon exploration techniques have been tested and the results documented hydrocarbon seepage on the Fort Peck Reservation. The work progressed in phased approach and assessed three separate study areas on the reservation: the Palomino Oil Field, the Wicape Prospect area and the Smoke Creek area. The largest study area was the Smoke Creek area which was further divided into three sub areas based on magnetic susceptibility measurements values: Smoke Creek Core, Lobo West and Site

26. Complete sets of data and maps for each of the three study areas were compiled and documented in (Monson, 2000, 2001, 2003).

Among the direct surface exploration techniques that have been employed are: soil gas survey, soil head gas analysis of basic hydrocarbon, soil acid extraction of gases, soil UV fluorescence and thermal desorption. The indirect methods are: soil microbial measurement, soil iodine and measurement of soil supplemental indicator such as redox potential (Eh), pH and conductivity. Two additional techniques (magnetic surveys) have been employed in phase II of the project: soil magnetic susceptibility and magnetometer.

3 Background

3.1 Conceptual model of hydrocarbon microseepage

Hydrocarbon in subsurface and due to high pressure at depth and weak areas in geological structure can escape and migrate to the surface (Salati, 2014, Brown, 2000, Etiope and Klusman, 2010b). Surface gas seepage is classified into different types based on its significance and implications. The names that normally appear in petroleum geology literatures are Seeps, Macroseeps, Microseepage, Microseeps and Miniseepage. These names are sometimes used improperly in scientific literatures which in turn creates some confusion (Etiope, 2015). The terms seeps or macro-seeps, also sometimes called macroseeps, refer to the visible focused manifestations. Microseepage, rarely written as micro-seepage, refers to invisible but detectable widespread exhalations from the soil. The seepage in all its types is traditionally restricted to the hydrocarbon-rich gas that is composed of methane and light alkenes, ethane, propane and butane with methane as the most predominant gas seep (Etiope, 2015).

The natural origin of methane is classified into geological and modern methane. Modern methane (sometimes called recent gas) is the gas produced through the microbial activity in wetlands. The geological methane is the gas that is produced in deep source rocks with carbon older than 50,000 years. This is the gas that is used today for energy and termed as fossil gas (Etiope, 2015). Methane of geological origin is further classified into biotic and abiotic. The abiotic methane is produced by chemical reactions that do not require the presence of organic matter and is not related to hydrocarbons generations whereas biotic methane required the presence of organic matters. The biotic methane can originate from two types of processes: microbial (biogenic) and thermogenic. Microbial methane is the gas that is produced in sediment by bacterial action at relatively low temperatures up to 60–80 °C. Microbial gas indicate shallow gas source rocks. The thermogenic gas is produced by the thermal cracking of organic matter or oil at higher temperatures relatively up 190–200 °C. Thermogenic gas indicate deeper source rocks. The origin of both gases can be distinguished on the basis of chemical and carbon isotopic composition analysis (Etiope, 2015). In this study, when we investigate methane from hydrocarbon, we are concerned of investigating geological methane (i.e. biotic) and its associated alterations.

3.2 Geological methane seepage and the role of atmospheric methane concentrations

It has long been thought that methane flux in dry lands is generally negative, indicating that it flows from the atmosphere to the soil. The process derived and controlled by methylotrophic bacteria that consume methane from the soil (Striegl, 1993, Etiope and Klusman, 2010a, Etiope, 2015). Later in 1988, unexpected emissions of methane to the atmosphere were reported by Hao et al. (1988) indicating that the soil was not a sink, but a small source for atmospheric methane. The measurements were criticized as there was no explanation for the positive methane flux (Etiope and Klusman, 2010a). Later several studies (Klusman and Jakel, 1998, Klusman et al., 1998, Klusman et al., 2000, Brown, 2000) have reported that dry lands which are part of petroleum basin can produce positive methane fluxes to the atmosphere.

Nevertheless, the attention to geological seepage in the scientific literature remained within the domain of interest of the oil industry. The fluxes of gases in the atmosphere resulting from hydrocarbons microseepage continued to be ignored and considered as a negligible source to atmospheric methane concentrations by atmospheric science and climate change community (Etiope, 2015). The reason behind that is due to the lack of seepage flux data and lack of the knowledge of the role of geological processes in the global methane budget.

The air pollutant emission inventory guidebook released in 2009 (EEA, 2009) was the first to report comprehensively the geological seepage as natural CH_4 source. The Intergovernmental Panel on Climate Change (IPCC) assessment reports (AR) did not consider geological seepage as natural CH_4 source until the fifth assessment report (AR5) that was issued in 2013 (Ciais et al., 2013). However, AR5 did not properly classify the geological seepage nor comprehensively described the process instead it included it only as submarine seeps and related gas hydrates (Etiope, 2015).

Today, it is known that the long term hydrocarbon microseepage in dry soil can overcome the methanotrophic consumptions and produce positive methane fluxes in the atmosphere in all petroleum basins (Schumacher, 2012, Etiope, 2015). These gases interact with soil and rocks, and bring about compositional changes. The surface manifestations of hydrocarbon microseepage can take many forms, including:

- Direct alteration in soil, formation of new minerals
- Indirect affect in vegetation's (geobotanical anomalies) due to change in environment.
- Thermal anomalies.
- Specific spatial pattern and fragmentation in the surrounding area due to the changed environment.

3.3 Methodologies for detection of gas Microseepage on Earth's surface

Methods for detecting gas microseepage are divided into direct and indirect methods (Etiope, 2015). Direct methods relay on the in situ sampling of air and soil collected and stored for laboratory analyses using traditional surface geochemistry methods such as closed-chambers. Indirect methods are based on the recognition of chemical, physical, and biological changes in soils, sediments, rocks and vegetation using remote sensing methods. In this work the focus is on the onshore indirect gas detection using above the ground (atmospheric measurements) and in land (soils, rock and vegetation's) satellite remote sensing methods.

3.3.1 Measurements in soil and sediments

Microseepage from hydrocarbon is known to cause anomalies in surface sediments and soils. Schumacher (1996) has documented different hydrocarbon-induced alterations and surface expressions. These include geobotanical anomalies, formation of paraffin dirt, mineralogical changes such as formation of calcite, pyrite, iron oxides and enrichment of ferrous iron, bleaching of red beds and clay mineral alteration.

Almeida-Filho et al. (1999), in one of the early satellite remote sensing studies for hydrocarbon microseepage detection, had used band ratio and band ratio difference using Landsat-Thematic Mapper images. The model used was based on the measurement of bleached material and

healthy vegetation cover. Bleaching of red beds occurs when microbial activities caused by microseepage cause oxidation reduction reactions on ferric oxide (Fe2O3) also abbreviated (Fe3+) rich soil and convert it to ferrous oxide (FeO or Fe2+). The process takes place through the removal of ferric oxide (hematite) to produce altered minerals in the surface (Schumacher, 1996). This in turns leads to visible change of terrain pigmentation from red to green/grey.

Mapping the bleached materials was successful using the band ratio TM2/3 where TM2 has the spectral range of 052-0.60 µm and TM3 has range of 0.63-0.69 µm. The high value of the band ratio TM2/3 indicate a poor ferric iron material. However, in presence of vegetation, healthy vegetation will also have high reflectance value in band ratio TM2/3. Thus, the ratio of TM4/3 was used to differentiate areas of ferric iron oxides from areas of vegetation. TM4 has a spectral range of 076-0.90 µm.

It was found that the pixels linked to the areas of bleached materials are better distinguished with the ratio TM2/3 versus the ratio difference (TM2/3±TM4/3). The pixels with high reflectance value in band ratio TM2/3 and high reflectance value in the ratio difference of TM2/3 -TM4/3 would exhibit the spectral signature of bleached materials. On the other hand high responses in TM2/3 and TM4/3 ratios would exhibit the spectral signature of vegetation cover.

Van der Meer et al. (2002) documented from different sources (e.g.,Segal and Merin, 1989, Elvidge, 1990, Schumacher, 1996) the spectral signature of different surface alterations induced by hydrocarbon microseepage. The most extensive studies were concerned with bleached red beds phenomena and ferric iron reduction, formation of carbonate content, clay mineral alteration and the anomalous spectral reflectance of vegetation.

The bleached areas were diagnosed by the ferric iron which exhibits an absorption feature at 0.9 µm and a sharp fall-off in reflectance to the blue from approximately 0.80 µm. As a result, the bleached areas were identified spectrally by the ratio of TM bands 2/3 which delineated the variations in ferric iron content. A very low 2/3 ratio value indicate a ferric iron rich areas, while high value indicated poor ferric iron rock (Van der Meer et al., 2002).

Segal and Merin (1989) defined the variations in clay mineral content caused by microseepage. The variations were attributed to the conversion of normally stable illite clays to kaolinite as a result of the diagenetic weathering of feldspars. Kaolinite exhibits a strong absorption feature centred at 2.2 µm, and its subordinate (doublet) absorption feature centred at 2.16 µm. Kaolinite-rich rocks can be identified spectrally by high ratio of TM bands 5/7 associated with bleached exposures.

Production of carbonate is one of the most known hydrocarbon induced alterations. Formation of the carbonates occur as a result of petroleum oxidation, stimulates particularly from methane (Schumacher, 1996). An increase in total carbonate in soil can be discriminated using TMS band-ratios for the wavelengths range of 1.550 - 1.650 µm, 1.985 - 2.085 µm, 2.037 - 2.137 µm, and 2.039 - 2.193 µm. Calcite is the most common obvious carbonate, the bands mostly used to map its anomaly are from 1.8 to 2.6 µm - 1.8 µm, 2.0 µm, 2.16 µm, 2.35 µm and 2.55 µm (Van der Meer et al., 2002).

Shi et al. (2010) found the spectral reflectance feature of the measured microseepage alerted rocks comes from the absorption spectrum of gypsum, ferric ion, hydroxyl ion and clay minerals. In comparison of altered and unaltered samples, it was found that the unaltered rocks have an intense absorption at 0.56 µm with relative high reflectance near 0.7 µm, while the altered rocks have an absorption at 0.7 µm. This was attributed to the presence of ferric oxide in the unaltered red bed. The ratio of band 2 to band 1 on ASTER sensor was used to extract related information since ferric oxide shows high reflectance in band 2. This ratio delineate red beds in white or bright pixels in the grayscale image.

The ASTER band ratio of band 5 and band 4 was used to map ferrous oxide alterations. However, the presence of hydroxyl ion on the altered greyish green rocks shows an intense absorption near 1.4 µm, 1.95 µm and 2.21 µm, while those with the present of carbonate ion show absorption at 2.35 µm. Some altered rocks shows also an absorption characteristic of gypsum, and multiple sets of gypsum veins were identified in those samples.

The analysis of the alerted rocks shows high value at ASTER band 4, which was attributed to the formation of carbonate minerals such as calcite, dolomite and siderite. The band ratio band 4 to band 9 was used to extract the information of carbonate minerals. Combining the analysis of spectroscopy and X-ray diffraction for the altered sample, it was found that carbonate minerals and ferrous oxide mineral are directly related to hydrocarbon seepage. Results demonstrated that mineralogical alterations induced by hydrocarbon seepage can be successfully revealed by using ASTER band ratios 2/1 for bleached red bed i.e. iron oxide bearing rocks. The band ratios 4/9 was used for mapping secondary carbonates.

Salati et al. (2014) and (Wang, 2016) have conducted two different studies in Dezful Embayment, in south-west Iran. Both have used ASTER and World View-2 satellite sensors to identify and map hydrocarbons microseepage related surface alterations. Comparing both studies, both were able to map hydrocarbon alteration, however there are several differences in their observations mainly due to the different image processing and classification methods that were used.

The alteration model in both studies was based on the result of (Salati, 2014) geochemical analysis. The analysis indicate that sulphur and gypsum have high concentration while calcite has low concentration in altered samples. On the other hand, the absences of gypsum, sulphur and high concentration of calcite was the characteristic of the unaltered samples. Gypsum was the main indicator of hydrocarbon microseepage surface alteration zone in the study area.

On ASTER image the band ratios:

- b2/b1 was used to indicate ferric iron
- b5/b4 was used in (Wang, 2016) and b3/b1 was used in (Salati et al., 2014) to indicate ferrous iron
- (b5+b7)/ b6 was used to indicate clays
- (b7+b9)/b8 was used to indicate calcite
- b4/(b9+b6) was used to indicate gypsum
- NDVI was used to map vegetation's health

The ferric oxide index, ferrous iron index and calcite index are higher in unaltered samples than altered samples while clay index and gypsum index are lower in in unaltered samples than altered field samples. Wang (2016) has identified that the band ratio b5/b4 can map ferrous iron in silicates and carbonates but it has limitation if it is associated with oxide and sulphate. In (Wang, 2016) much attention was given to the formation lithology types as testing same index in different lithology types were found to produce different results. For instance, gypsum is common mineral in evaporate formation, therefore it cannot be used as indicator for hydrocarbon microseepage while it is a good indicator in marly limestone formation. The altered field samples in the study area located within marly limestone formation which contains less carbonates. That explains why band ratio b5/b4 that was used to map ferrous iron index exhibit low value in altered samples than in unaltered samples.

On the WorldView-2 image the band ratio below were selected to map the surface alterations:

- b2/b1 was used to indicate sulphur
- b5/b3 was used to indicate ferric iron in (Wang, 2016) while band ratio b4/b7 in (Salati et al., 2014)
- (b3*b4)/b2 was used to indicate iron oxide
- (b3+b5)/b1 was used to indicate ferrous iron in (Salati et al., 2014)

Salati et al. (2014) explains that hydrogen sulphide will be produced as a result of sulphur minerals reduction resulted from microseepage. Reaction of hydrogen sulphide with calcite in limestone will produce gypsum and native sulphur. Therefore sulphur absorption feature at 0.4 μm can be used to map sulphur in WorldView-2 image for the altered samples. Accordingly, (Wang, 2016) used band ratio b2/b1 to map sulphur. On the other hand, and in reference to (Kalinowski and Oliver, 2004) the band ratio b5/b3 was used to indicate ferric iron and (b3*b4)/b2 was used to indicate iron oxide in (Wang, 2016). Both studies have used same satellite sensors in the same study area, however they have reported different results and observations

Chen et al. (2017) have conducted a study using airborne hyper-spectral data on Xifeng Oilfield on China to detect hydrocarbon microseepage through the identification of altered mineral. In their study they proposed to conduct mapping of hydrocarbons altered minerals based on previous studies (Defang, 1995, Saunders et al., 1999). The availability of Fe2+ (ferrous), carbonate, and clay minerals in the overlying soil were used as an important indicator of microseepage alterations.

The mapping result shows much siderite and calcite, but little illite and kaolinite were seen in microseepage area. It was found that the content of carbonate always distribute at the edges of hydrocarbon area which represent the scope of hydrocarbon microseepage. The measurement was based on the calculation of spectra absorption depths difference at 2150 -2220 nm and 2300 nm and 2350 nm. These calculation was based on the assumption that at the edges of hydrocarbon accumulations, the contents of carbonate are higher while the contents of clay minerals are lower. The opposite condition is true at the centre of hydrocarbon area. The vegetation spectral index, fat index (FI) was used to simulate the detection of spectral absorption depths.

$$FI = \frac{(H1 + H2)}{|H1 - H2|}$$

Where H1 represent the absorption depth of spectral feature of clay at 2218 nm (illite, kaolinite)

H2 represent the absorption depth of spectral feature of carbonate minerals at 2350 nm (calcite, siderite)

The larger the value of FI the higher the indication of surface anomaly as a result of hydrocarbon microseepage. The validation of field samples using X-ray diffraction (XRD) and spectrophotometer in laboratory showed there were higher calcium carbonate and lower illite concentrations, also higher concentration of ferrous oxide ($Fe2+$) and lower concentration of ferric oxide ($Fe3+$). This is the same phenomena explained earlier by (Shi et al., 2010) as bleaching of red beds, however it is never been described as such by (Chen et al., 2017).

3.3.2 Geobotanical anomalies

Noomen (2007) studied the influence of gas seepage on vegetation. The study classified the effect of hydrocarbon gases (i.e., methane, ethane and natural gas). Also, it tried to segregate between the effects of gases and carbon dioxide concentration that is known to be accompanied with these gases. More importantly, the effect was studied to determine if it was caused by gases themselves or by oxygen shortages. The experiment also simulated small and large gas leak effects. The researcher has tested different vegetation indices such as NDVI, LIC, and VOG. The result that was collected from the simulated gas leaks was used in a field study in an area that was known with a natural hydrocarbon seepage. Along with reflectance measurements, (Noomen, 2007) has studied vegetation diversity and spatial pattern. The result shows that the total vegetation cover increased with distance from the seep, the distribution and diversity of vegetation around the seep were clustered differently and created what the researcher has named it as "green ring".

Noomen (2007) concluded that natural gas and methane have not caused any significant changes to the morphology of the studied plants (maize and wheat) whereas ethane has affected the maize morphology. The effect of ethane and methane on maize was determined by leave reflectance. The main conclusion of the thesis is the change in plant growth or reflectance is attributed to the soil oxygen shortage that accompanies the leak.

Noomen (2007) work focused on the concentrations of hydrocarbon gases on maize and wheat and the effect on their growth and reflectance. However, those plants are known of their tolerance to methane and might not be the best for the case study. Although there is much debate that terrestrial plants can produce CH_4 under aerobic conditions. Methane during drought stress might have a beneficial response. It enhances the osmoti1c stress tolerance in maize by modelling sugar and ascorbic acid metabolism (Han et al., 2017). This point was unnoticed by Noomen (2007).

Elvidge (1990) explains that changes in chlorophyll concentration would affect the reflectance properties of vegetation and produce a spectral shifts from visible part with absorption at 0.66 and 0.68 µm to the red-edge absorption near 0.7 µm. Arellano et al. (2017) indicate that

hydrocarbon pollution will cause degradation in plants biophysical and biochemical parameters such as chlorophyll, pigment concentration, leaf and canopy index. Different plant species respond differently. The stress level on vegetation will vary based on intensity and the length of exposure. Response in plants can be explained by the reduction in plant cell activity, reduced plant growth or plant mortality caused by reduced rate of photosynthetic activity and plant transpirations (Arellano et al., 2017). Robson et al. (2004) found that some species will have certain stress tolerance to hydrocarbon pollution, however the general trend shows decrease in vegetation cover and species richness near to hydrocarbon seepage.

Krupnik and Khan (2017) studied hydrocarbon microseepage related geobotanical anomalies around different US oil fields that have been confirmed with the presence of hydrocarbon microseepage. They have used Normalized Difference Vegetation Index (NDVI) based on Moderate Resolution Imaging Spectroradiometer (MODIS) to examine the impact of hydrocarbon seepage on vegetation. Depending on the season, the typical NDVI value in desert environments can range from 0.18 to 0.49. In (Krupnik and Khan, 2017) the average NDVI was calculated for the different oil fields. It was found that there is a trends in oil fields for the average NDVI value. NDVI decreases inside oil fields and increases in surrounding areas. NDVI is lower inside oil field may be an indication of an anaerobic rich environment explained by lower abundance of vegetation or stressed vegetation.

3.3.3 Fugitive methane direct measurements

Fugitive methane is the fluxes of methane in the atmosphere that is released and elevated from different sources either biogenic or thermogenic origin. Here we are more interested in thermogenic sources that are linked to the natural degassing from the subsurface of oil and gas basins.

The Scanning Imaging Absorption Spectrometer for Atmospheric Cartography (SCIAMACHY) running on board of ENVISAT satellite is one of the primary space-based imaging spectrometer used globally for observations and measurement of atmospheric compositions. These include trace of gases, aerosols and clouds. The instrument records the transmitted, backscattered and reflected radiation at high spectral resolution between 0.2 nm to 0.5 nm in wide wavelength range between 240 and 2400 nm (Buchwitz et al., 2005).

Despite the high spectral resolution of SCIAMACHY, it has a very low spatial resolution. On the other hand the instrument measures the total concentration of gases in earth atmosphere. This makes it difficult for such spaceborne sensor to identify and distinguish between the natural and anthropogenic sources. Indeed, it demonstrates a challenge to identify methane emission from hydrocarbon microseepage.

3.3.4 Hydrocarbon thermal anomaly

Muhammad and Alki (2015) studied hydrocarbon microseepage related surface thermal anomalies in different oil fields in Baghdad, Iraq. The study assumed that the ambient temperatures on the earth's surface is 300 ˚K (26.85 Celsius) with a dominant wavelength of 9.66 μm. This was based on Wein's Displacement Law, though it was not stated in (Muhammad and Alki, 2015) work.

The wavelength in the range between 8 - 14 μm was considered as the most appropriate wavelength to study microseepage related surface thermal anomalies. Landsat Enhancement Thematic Mapper plus (ETM+) also named Landsat 7 was selected for the study. Landsat band 6 outputs two versions of the thermal data that are provided as two separate band files, high and low gain at 60m spatial resolution. The high gain is used to improve the radiometric sensitivity while low gain is used to prevent over bright areas. The estimated minimum and maximum emitted surface radiance in the study area on band 6 low gain band (0- 17.04 w/m ^{2}/μm/sr) and high gain (3.2-12.65 w/m ^{2}/μm/sr). The thermal anomaly was determined as a function of change of the local areas from their surrounding radiation temperature. It was found that thermal anomaly between 1 and 5 ˚K was related to hydrocarbons microseepage (Muhammad and Alki, 2015).

Muhammad and Alki (2015) methodology was not clearly described, that might be due to the translation from Arabic language to English, besides it lacked proper referencing. It is necessary to do a time series analysis rather than looking at just one time stamp when monitoring the thermal anomaly. Other strategy is to view night-time infrared imagery of the study area to avoid any generated heat from the sun during daylight hours which was not considered in (Muhammad and Alki, 2015).

Suherman et al. (2014) investigated the influence of hydrocarbon seepage on land surface albedo and land surface temperature (LTS). The study focused on the statistical comparison analysis of albedo and LST between hydrocarbon microseepage area and non microseepage areas. Their theoretical base, hydrocarbon seepage can cause intrinsic change in thermodynamic properties of interacted surface material. This can leads to: loose of radiative transfer, reduce of convective overturning energy at the surface, decrease in emissivity and increase in surface albedo and land surface temperature. Therefore, areas with hydrocarbon seepage will have:

- Lower vegetation density or vegetation stress that generally associates with an increase in the albedo.
- Higher LST than the surrounding area with no hydrocarbon seepage.

The visible to near infrared spectral regions were used for the determination of the increase or shift on albedo. The thermal infrared range was used for determination of land surface temperature, additionally, NDVI was calculated for vegetation stress analysis. The result shows that, albedo was positively shifted on seepage potential area. The positive shift in farm and forest areas was attributed to the developed bare land or stress vegetation as a result of hydrocarbon seepage and in turn cause the increase in the albedo. The no or slight negative shift for bare and town areas was attributed to mixed surface materials with hydrocarbon seepage that reduced the reflectivity and reduced the albedo. The comparison of land surface temperature shows slight positive shift in the forest and plantation areas, while farm, bare and town areas were negatively shifted. Shift on albedo and LST was positively correlated with NDVI value (Suherman et al., 2014).

3.4 Image and spectral processing methods

There are variety of approaches for hydrocarbon microseepage mapping and characterizations. These approaches are normally categorized into data-driven and knowledge-driven methods (Asadzadeh and de Souza Filho, 2016).

The data-driven approaches are mathematically based approaches derived by reference data. They are empirical models completely estimated from measured data. Therefore, modelling process is more objective. The data-driven approaches are sub-divided into learning based approaches and geostatistical based approaches.

Weight of evidence (WofE) is an example of geostatistical based data driven approach that is widely applied in mineral exploration. For example (Hartley, 2014, Yuan et al., 2014, Zeghouane et al., 2016) used WofE model for mineral prospectivity mapping. Allek et al. (2016) used WofE to test the spatial association between known hydrocarbon fields and sedimentary residual magnetic anomalies.

Machine learning (ML) is a learning based data driven approach. It is a field of computer science and an application of artificial intelligence that uses mathematical models and algorithms to allow the systems to automatically learn and improve from experience without being explicitly programmed. Machine learning algorithms can offer solutions for prediction, approximation and classification problems. These approaches have achieved promising performance in mineral prospectivity mapping and hydrocarbon microseepage classifications.

For instance, Wang (2016) has tested three different machine learning classifiers for finding hydrocarbon surface alterations. The tested classifiers are Support Vector Machines (SVM), Random Forest (RF) and Gradient Boosted Regression Tree (GBRT). Salati et al. (2014) have used Boosted Regression Tree (BRT) to map hydrocarbon alteration zones. The result showed BRT provided the best discrimination between altered and altered areas when compared to band ratio and PCA methods. The AdaBoost algorithm the so-called Adaptive Boosting was used by (Avcioglu, 2010) to map hydrocarbon microseepage related surface manifestations. AdaBoots has been shown to be effective in improving the accuracy of microseepage mappings, when compared to the results from Crosta technique.

Principle Component Analysis (PCA), Band Ratio (BR) are band calculation knowledge based approaches used in remote sensing. These approaches depend on spectral/absorption characteristics of the mapped feature (Asadzadeh and de Souza Filho, 2016). The knowledge based approaches are physically based approaches, modelling process using these approaches is more subjective as it is centred at expert opinion.

Below is a brief description of some of the ML algorithms and spectral processing methods used in the application of remote sensing image classification:

3.4.1 Support Vector Machines (SVM)

Support Vector Machine (SVM) is a supervised classification method which use a non-parametric statistical learning algorithm derived from statistical learning theory (Vapnik, 2013). In its simplest form, SVM is a binary (two-class) classifier but it can function as a

multiclass classifier by combining several binary classifiers. SVM works by defining optimal hyperplane that provide the maximum margin to separate two classes (Cortes and Vapnik, 1995, Huang et al., 2002). The data points that lie on the margin of the hyperplane called support vector and these are the points that define the optimal hyperplane of maximum margin. SVM can adopts nonlinear classification by using various types of kernels (linear, polynomial, radial basis function, and sigmoid) to define optimal hyperplane. Kernels are used to construct a linear separation between two points of different classes by embedding the data into high dimensional space. The main challenge in using SVM is to select the most appropriate kernel function type and its parameters. Kernel configuration plays a major role on SVM performance and classification accuracy (Cristianini and Scholkopf, 2002, Huang et al., 2002). SVM is not sensitive to overfitting and works well with few or small training samples (Mountrakis et al., 2011). SVM includes a number of parameters (Zhu et al., 2009) these including kernel type, gamma and C or penalty, these parameter are used to tailor the behaviour of the algorithm to a specific dataset to attain better classification performance.

3.4.2 Weight of Evidence (WofE)

Weight of Evidence (WofE) is a geostatistical data-driven approach that use Bayesian statistics to measure the strength of the spatial association between a particular condition and a set of data points. The core of WofE is the use of conditional, prior and posterior probabilities. WofE is a quantitative method that is relatively easy to interpret. Details of the mathematical explanation are thoroughly covered in (Bonham-Carter, 1990, Bonham-Carter, 1994).

3.4.3 Mixture Tuned Matching Filtering (MTMF)

The Mixture Tuned Matching Filtering (MTMF - Boardman, 1998) is a target detection algorithm used for spectral mixture analysis. It was developed based on both Matched Filter (MF) model and Liner Spectral Mixing (LSM) model. MTMF takes the advantage of MF by means of performing partial unmixing and subpixel target abundance estimation without knowing the other background endmembers signatures.

MTMF include three main steps: First, Minimum Noise Fraction (MNF) transformation is performed for dimension reduction. Then, matched filtering is performed through maximizing the response of a known endmember while suppressing the response of the background. Finally, the liner spectral mixing is used to avoid the drawbacks in MF by adding infeasibility images to the result to reduce the number of false positives found on MF (Boardman, 1998, Goodarzi Mehr et al., 2013). The result of MTMF is a set of rule images given as MF and infeasibility scores. These are usually represented as a greyscale image with values ranging from 0 to 1 for each pixel related endmember. Pixels with low infeasibility and high MF score (i.e., value of 1) indicate a very high degree of matching and correctly mapped pixels (Zadeh et al., 2014).

3.4.4 Spectral transformations methods

The spectral transformation methods are used for spectral enhancement. These methods can emphasis certain features in the images and help in the identification of endmembers of the altered minerals that are associated with hydrocarbon microseepage.

3.4.4.1   Principle Component Analysis (PCA)

The PCA is a spectral decomposition or transformation method that reduce the dimension, remove redundancy and minimize noise in multispectral data (Singh and Harrison, 1985). PCA converts correlated variables (original bands) on multispectral images into uncorrelated variables (principal component) by selecting uncorrelated linear combinations (so-called eigenvector loadings) of variables in such a way that each successively extracted linear combination, or principal component (PC), has a smaller variance (Singh and Harrison, 1985). PCA has been widely used to map mineral potential and hydrocarbon microseepage related alteration, see for example (Zhang et al., 2007, Avcioglu, 2010, Tangestani and Moore, 2000, Zhang et al., 2011, Liu et al., 2016)

3.4.4.2   Band Ratio (BR)

Band ratio is a technique that is used widely in remote sensing for spectral enhancement (Almeida-Filho et al., 1999, Almeida-Filho et al., 2002, Van der Meer et al., 2002, Zhang et al., 2007, Pour and Hashim, 2015). It works by dividing the digital numbers (brightness values) of the pixels in a spectral band by digital numbers from another band which results in new image. Band ratio has the ability to transform the data and reduce or minimize the effects of different environmental factors like topographic slope and shadows. It has successfully been used in mapping alteration zones (e.g.,Segal, 1983, Poormirzaee and Oskouei, 2010, Zhang et al., 2011, Mia and Fujimitsu, 2012, Hajibapir et al., 2014).

3.4.4.3   Minimum Noise Fraction (MNF)

MNF is another transformation method that is used to segregate noise in the data and also used to reduce the computational requirements for subsequent processing (Boardman and Kruse, 1994). It is two cascade PCA transformations. The first transformation de-correlate and rescale the noise in the data by performing noise whiting based on the noise covariance matrix. This step will result in no band-to-band correlations and the noise in every band has unit variance. The second step is a standard principal components transformation that determine the inherited dimensionality through the examination of the final eigenvalues (Green et al., 1988).

The image pixels are presented by eigenvalues in every MNF transformed band. The eigenvalues provides a measure of MNF information content, with progressively noisier bands approaching eigenvalues near zero. MNF eigen-images with values close to 1 contain mostly noise (Green and Boardman, 2000), the common practice is to exclude MNF components with eigenvalues less than 1.25 (Green and Boardman, 2000, Jensen, 2015). The number of MNF output bands are generally equivalent to the number of spatially coherent eigen channels i.e. number of input bands (Boardman, 1993).

3.4.4.4   The Pixel Purity Index (PPI)

The pixel purity index (PPI) is supervised endmember extraction algorithms. It identifies and locates the most spectrally pure or extreme pixels in multispectral and hyperspectral images (Boardman et al., 1995). As a pre-step, dimensionality reduction transformation is generally required using either the Principal Components Analysis (PCA) or Maximum Noise Fraction (MNF). The image spectra are treated as points in n-dimensional spectral space, where n is the

number of bands. PPI is computed by repeatedly projecting the data onto a random unit vector called skewers (Boardman, 1993, Boardman et al., 1995). Using a threshold value specified by the analyst the PPI can determines how many pixels are marked as extreme at the ends of the projected vector. PPI with large thresholds value can result in finding more extreme pixels which are less likely be pure endmembers. A good threshold value should be approximately two to three times more than the noise level of the data (Stein et al., 2006). After many repeated projections to different random skewers, pixels selected above a certain cut-off threshold for a number of times are recorded and the total number of times each pixel is marked as extreme is noted (Wu et al., 2013).

When the iterations are complete, a PPI image is created in which the value of each pixel corresponds to the number of times that pixel was recorded as extreme. Bright pixels in the PPI images represent the more spectrally pure pixels and generally are the image endmembers, whereas darker pixels represent the less spectrally pure pixels (Boardman, 1993, Boardman et al., 1995, Stein et al., 2006).

3.4.4.5   N-Dimensional Visualizer

The n-Dimensional visualizer is multi-dimensional scatter plot chart, which has spectral brightness values plotted on its axes and allows the visualization and extraction of endmembers in multi- or hyperspectral imagery (Harris, 2006, Ahmad, 2012). Each graph axis refers to an individual spectral band, analysing two bands from the image will result into 2-D scatter plot (two axes). Analysing more than two image bands will form points cloud in the multi-dimensional space.

3.4.5   Endmembers extraction methods

Endmembers are the spectra that are chosen to represent pure surface of materials in image cube. Surface on earth is not homogeneous and composed of multiple mixed materials. A single pixel in a given scene from hyperspectral or multispectral image typically contains a mixture of reflectance of many distinct materials. The process that decomposes the mixed pixels into a pure pixels (endmembers) and determine their corresponding proportions (abundances) called spectral mixing models. There are two approaches to select spectra as endmembers from hyperspectral or multispectral images:

3.4.5.1   Selection of endmembers from a spectral library

In this method, spectral signatures are derived from spectral libraries built from field or laboratory measurements. Collected spectra should be then resampled and scaled to an input image wavelength. The advantage of this method is that reference endmember (library spectra) can be easily obtained. However, this approach might carry out some risk due to the fact that different materials may have similar spectral characteristic. In addition, these spectra are acquired under different conditions. An intermediate step is also required to calibrate the reference library spectra to the surface reflection (Roberts et al., 1998, Plaza et al., 2004, Harris, 2006).

3.4.5.2 Selection of endmembers from the image

Endmember reference spectra can be derived directly from the image data themselves. Steps to define endmembers include Minimum Noise Fraction (MNF) transformation, Pixel Purity Index (PPI) and n-Dimensional Visualization (nDV). In-scene endmembers have an advantage over spectral library endmember in that they have been obtained and collected at same scale. However, other external factors such as atmospheric contamination, radiometric calibration and elevation angle can influence the endmember selections (Roberts et al., 1998, Plaza et al., 2004, Harris, 2006).

4 Data

Due to the lack and absence of gas flux data and in order to assess the spatial distribution of the seepage area, this work has utilized oil and gas wells data to classify the areas with microseepage. Statistics for more than 2700 drilled wells around the world have shown that 82% of wells drilled on prospect associated with microseepage anomalies resulted in discoveries compared to an average of 11% for those drilled without any associated anomalies (Schumacher, 2010). This suggests that the location of oil and gas wells corresponds closely to the spatial distribution of microseepage from hydrocarbon reservoirs. Because of the wealth of oil and gas wells data around the world that are made publicly available, the attempt to re-assess the extent of microseepage using oil and gas wells data and satellite remote sensing can facilitate a more rapid and comprehensive evaluation of every sedimentary or petroleum basin contribution to atmospheric methane budget.

4.1 Wells data

Oil and gas data in the Fort Peck reservation was downloaded from Montana Board of Oil and Gas Conservation website (MBOGC, 2019). The data include all wells that were drilled in the region since 1900 until today, including the abandoned and currently producing wells. To prepare data for classifications, wells dataset were further categorized into two classes: 340 oil wells and 451 dry holes, where the oil indicate wells that have hydrocarbon fuel and dry hole indicate wells drilled for oil and gas but yielded none. Wells data were used as training and validation data for the classification. Assumption was built based on the hypothesis that, sedimentary basins hosting natural gas and oil reservoirs are frequently leaking, even in large and productive fields, causing gas to migrate vertically to the surface and in turn produce positive fluxes of methane to the atmosphere (Etiope and Klusman, 2010a, Etiope, 2015). Accordingly, in this study oil wells were used as an indicator of methane fluxes and dry hole as indicator of no fluxes.

4.2 Satellite data

4.2.1 Landsat-8 OLI/TIRS data

Landsat-8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) Analysis Ready Data (ARD) image generated over conterminous United States (CONUS) covering the study area (ARD Grid Tile horizontal = 12, vertical = 2 and 3) was chosen and downloaded from US Geological Survey Earth Resources Observation and Science Centre (EarthExplorer, 2018). Landsat ARD tile grid system is modified from the Web-Enabled Landsat Data (WELD) and based on the World Geodetic System 84 (WGS84) (Roy et al., 2010). Each tile contains 5000 x 5000 30-meter pixels include all pixels acquired in a given day within its extents. ARD tile can intersect more than one row along a WRS-2 path (Dwyer et al., 2018). Table 1 and Table 2 list down the WRS-2 Path/Rows that intersected the study area ARD tiles.

Table 1 - WRS-2 Path/Row intersected with ARD tile h=12 V=2.

WRS-2 Path	WRS-2 Row
35	25
35	26
36	25
36	26
37	25
37	26

Table 2 -WRS-2 Path/Row intersected with ARD tile h=12 V=3

WRS-2 Path	WRS-2 Row
34	26
34	27
35	26
35	27
36	26
36	27

The Landsat scenes were acquired on 10th of November 2016. Landsat 8 satellite carries two sensors, the Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS). The spectral bands of the OLI sensor include nine visible, near-infrared and shortwave infrared bands. The TIRS instrument collects data in two long wavelength thermal bands. Landsat 8 have an improved radiometric precision (signal-to-noise) performance when compared to Landsat 7 (Roy et al., 2016) and provides moderate resolution imagery of 30 meter for multispectral bands 1 to 7 and 9 (Zanter, 2019). The thermal bands 10 and 11 are collected at 100 meters resolution and resampled to 30 meter in the delivered data product. The Landsat ARD product delivered in Albers Equal Area (AEA) conic map projection based on World Geodetic System 84 (WGS84) datum (Barsi et al., 2014, Dwyer et al., 2018). The characteristic of Landsat 8 OLI/TIRS are shown in Table 3.

Table 3- Landsat 8 OLI/TIRS spectral characteristics.

Bands	Wavelength (micrometers)	Resolution (meters)
Band 1 - Ultra Blue (coastal/aerosol)	0.435 - 0.451	30
Band 2 - Blue	0.452 - 0.512	30
Band 3 - Green	0.533 - 0.590	30
Band 4 - Red	0.636 - 0.673	30
Band 5 - Near Infrared (NIR)	0.851 - 0.879	30
Band 6 - Shortwave Infrared	1.566 - 1.651	30
Band 7 - Shortwave Infrared	2.107 - 2.294	30
Band 8 - Panchromatic	0.503 - 0.676	15
Band 9 - Cirrus	1.363 - 1.384	30
Band 10 - Thermal Infrared (TIRS) 1	10.60 - 11.19	100 * (30)
Band 11 - Thermal Infrared (TIRS) 2	11.50 - 12.51	100 * (30)

4.3   Spectral library

The spectra for the identified minerals that are associated with hydrocarbon gas induced alteration have been obtained from U.S Geological Survey Spectral Library Version 7 (Kokaly et al., 2017) and resampled to Landsat image wavelength. The spectral library contains spectra measured with laboratory, field and airborne spectrometers with wavelength coverage spans the ultraviolet, visible, near-infrared, mid-infrared and far-infrared regions (0.2 to 200 microns). It includes samples of minerals, rocks, soils, liquids, vegetation and other biologic materials and man-made materials. Along with spectra information, the library includes results of X-ray Diffraction, Electron Probe Micro-Analysis, and other analytical methods (Kokaly et al., 2017).

5 Methods

5.1 Satellite data Image pre-processing

Landsat-8 OLI ARD data for the study area was obtained from US Geological Survey Earth Resources Observation and Science Centre (EarthExplorer, 2018). Data package contains set of images. The analysis was mainly performed on two layers: surface reflectance and surface temperature layers. As pre-processing step and because the study area was covered by two image scenes, image mosaicking was performed on each band to create one scene of the study area. Subsequently, layer stacking was performed to combine all the mosaicked bands into a single image. The Landsat data includes a Quality Assessment (QA) band that includes all the contaminated pixels such as clouds, shadows, snow/ice, water, terrain occlusion and fill values. A mask was built from QA band. This mask is a binary image with values 0 and 1 representing cloudy/water pixels or no cloud/no water pixel respectively. By means of raster calculation all the contaminated pixels like clouds and water were masked out from the image. Finally, spatial image subset was performed to clip the image to study area for further image analysis.

Flowchart depicted in Figure 2 provides a summary of the different steps implemented to classify hydrocarbon related surface alteration. Image pre-processing is the initial step that was used to prepare the image for analysis. In the second step, two spatial modelling techniques were developed and applied: knowledge-driven methods and data-driven method. The knowledge-driven model depends on expert opinions. The data-driven model is estimated from measured data. It uses statistical methods to quantify the spatial relationships between predictor maps (factors) and set of locations of discrete events, also called training site or ground truth points.

The applied knowledge-driven model used different spectral transformation methods such as principal component analysis (PCA), minimum noise fraction (MNF) or band ratio. It also included a step for endmember determination and identification. Classification to estimate the fractional abundance of each mixed pixel was performed using mixture-tuned matched-filtering (MTMF). Weights of evidence was performed on PCA and band ratio results to quantify the spatial associations between hydrocarbon microseepage and mineral anomalies.

The data-driven model was performed in parallel using Support Vector Machine (SVM) classification to create the spatial relationship based on possible hydrocarbon gases fluxes and image spectral characteristics.
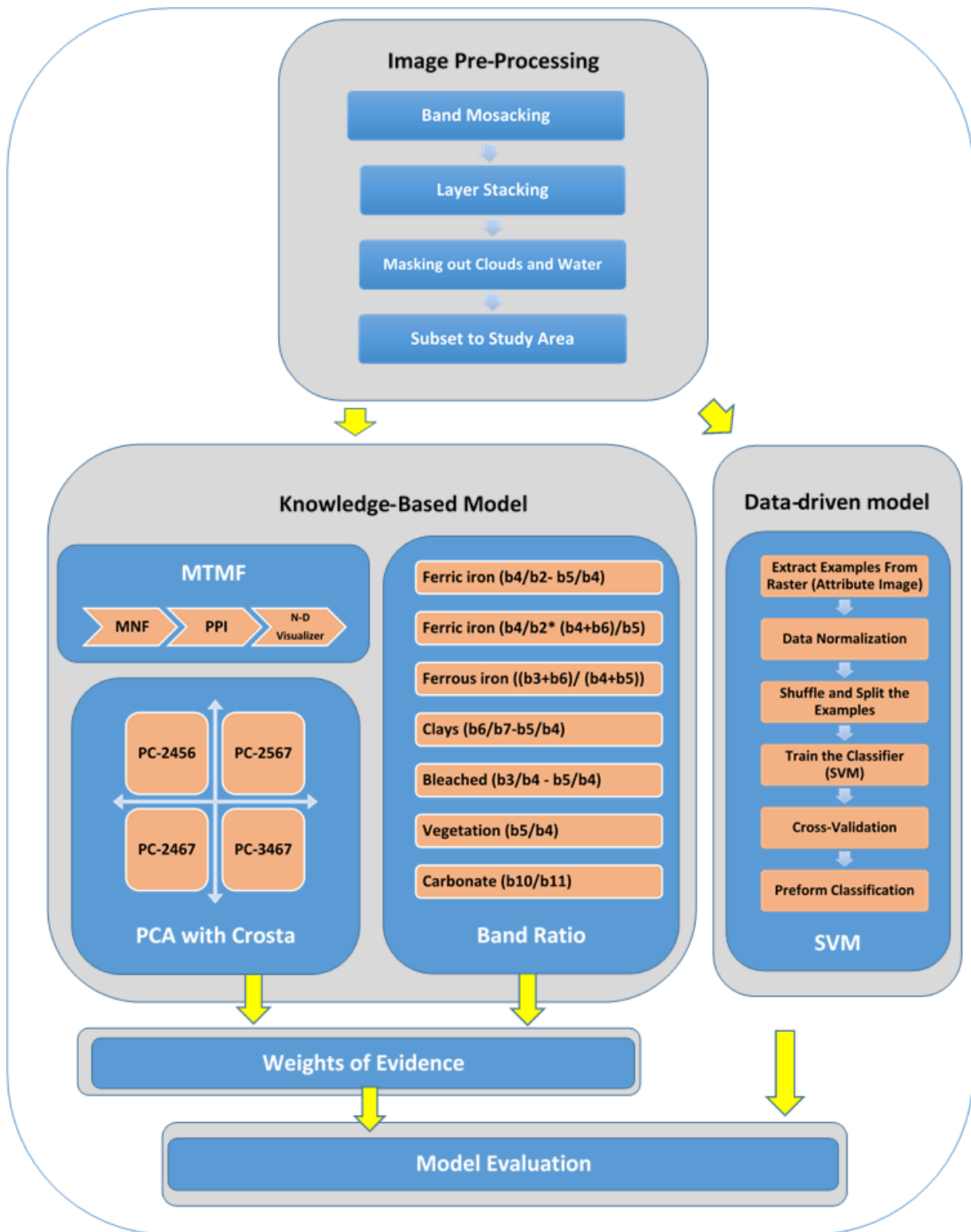
Figure 2- General workflow of methodology used for image calibration and mapping model (b is denoting image band, and PC is denoting PCA image).

5.2 Knowledge-based models

5.2.1 Principle Component Analysis (PCA)

The principal component analysis (PCA) transformation was applied on Landsat-8 OLI data obtained for the study area. The first principal component (PC1) is composed of negative weighting from all bands (Table 4). This PC accounts for 96 percent of the total variance of the data. It does not contains any relevant spectral feature related to minerals; instead it contains information mainly related to albedo and topography information. PC2 contains the spectral feature relevant to vegetation indicated by the positive values in the green (band 3) and near-infrared region (band 5). PC3, PC4 and PC5 can highlight minerals related spectral features because of the different contribution of positive and negative weights of eigenvector in each band specially band 6 and band 7. The subsequent PCs normally are excluded due to the noise as they contain very little data variance which can explain little or no variability of the data (Crósta and Moore, 1989, Loughlin, 1991).

Table 4- Principal component (eigenvectors) of Landsat image covering the study area.

Eigenvectors	Band 1	Band 2	Band 3	Band 4	Band 5	Band 6	Band 7
PC 1	-0.113	-0.144	-0.201	-0.256	-0.428	-0.670	-0.478
PC 2	-0.107	-0.086	0.004	-0.038	0.833	-0.108	-0.524
PC 3	-0.153	-0.236	-0.412	-0.531	-0.122	0.638	-0.219
PC 4	-0.004	0.160	0.223	0.528	-0.318	0.361	-0.643
PC 5	-0.721	-0.489	-0.139	0.434	-0.010	-0.022	0.184
PC 6	0.361	0.079	-0.824	0.420	0.078	-0.048	0.011
PC 7	0.550	-0.803	0.205	0.089	-0.027	0.025	-0.032

The Crosta method (Crósta and Moore, 1989) is commonly used with PCA and allows identifying principal components that contain spectral information about specific minerals such as clay minerals and iron oxides.

In order to suppress the interference and concentrate the information on specific targeted minerals, Crosta technique was applied by reducing the number of PCA input bands to four bands. Those bands were chosen based on the spectral characteristics of hydrocarbon alteration associated minerals in VNIR and SWIR regions. Details of the selection are shown below and summarized in Table 5.

Table 5- Summary of PCs selected from PCA transformation based on Crosta technique for Landsat 8.

Minerals	PC Bands
Ferric and Ferrous Iron	PC-2456
Clay and Carbonate Minerals	PC-2567
	PC-2467
	PC-3467

5.2.1.1   PC to indicate the presence of ferric and ferrous iron minerals and bleaching of red beds

The OLI bands 2, 4, 5 and 6 were chosen to enhance the possibilities to detect ferric iron oxide. Ferric oxide is characterized by strong absorption in blue region measured by OLI band 2 and high reflectance in band 4. In the PCA, it can be examined by the eigenvector loading of band 4 and band 2. Band 7 was omitted deliberately to avoid hydroxyl mapping.

The areas with abundance of iron oxide or ferric rich surfaces are of less important for hydrocarbon microseepage anomalous concentration. Ferric oxide and ferrous oxide are two different types of iron oxides. However, iron oxide in remote sensing studies such as hydrothermal alteration mapping is always referred to as ferric ions and not ferrous ions. Ferrous oxide most of the times is referred to independently as ferrous minerals or ferrous iron. Presence of hydrocarbon microseepage result in removal or conversion of ferric ions ($Fe3+$) to ferrous ions ($Fe2+$), causing red bed bleaching (Schumacher, 1996, Van der Meer et al., 2002). The PC2456 is known in literature (Tangestani and Moore, 2000, Zhang et al., 2011, Mia and Fujimitsu, 2013, Liu et al., 2016) for mapping of ferric oxide. However, the same has also been used in number of studies to indicate the presence of ferrous iron minerals and beaching of red beds (Zhang et al., 2007, Zhang et al., 2011). The importance of this PC is in highlighting ferrous iron minerals and beaching of red beds apparently because it uses band 5 and 6 where ferrous iron has spectral characteristics in these two bands.

5.2.1.2   PC to indicate the presence of clay and carbonate minerals

Clays (so called Hydroxyl-bearing minerals) and carbonates have characteristic absorption in the SWIR range at wavelength of 2.1-2.4 µm measured in Landsat OLI band 7 and at the same time they have a very high reflectance in the wavelength range of 1.55-1.75 µm measured in band 6. Therefore, PCA can be examined by the eigenvector loading of band 6 and band 7 to determine which PC image will enhance the presence of clay and carbonates.

The PCA on bands (2,5,6,7), (2,4,6,7) and (3,4,6,7) are typically used for clay and carbonate mapping (Tangestani and Moore, 2000, Zhang et al., 2011, Mia and Fujimitsu, 2013, Liu et al., 2016). These three PCs were chosen and applied to Landsat-8 OLI images.

5.2.2   Band Ratio

Band ratio technique has been applied through dividing the digital numbers (brightness values) of the pixels in a spectral band by another band which results in a new image. The selection of bands for highlighting specific minerals group or features were determined based on reflectance and absorption characteristics. Table 6 summaries the spectral characteristic of different minerals assemblage on Landsat OLI.

Table 6 - Spectral characteristics of mineral assemblage on Landsat OLI bands.

Mineral	Band 2 (Blue)	Band 3 (Green)	Band4 (Red)	Band 5 (NIR)	Band 6 (SWIR)	Band 7 (SWIR)
Ferric Iron	Absorb		Reflect	Absorb	Reflect	
Ferrous Iron	Reflect	Reflect	Reflect	Absorb	Reflect	Absorb
Bleached Red Bed	Reflect		Reflect	Absorb		
Clay and Carbonates	Reflect			Reflect	Reflect	Absorb
Vegetation	Reflect	Reflect	Absorb	Reflect	Reflect	Absorb

5.2.2.1 Ratios to indicate the presence of ferric, ferrous minerals and bleached red beds

Ferric iron oxide-bearing mineral has high reflectance between 0.63 and 0.69 μm corresponding to band 4 (red) on OLI and high absorption between 0.45 and 0.52 μm corresponding to band 2 (blue) on OLI (Shi et al., 2010, Pour and Hashim, 2015, Ducart et al., 2016). Therefore, ferric iron mineral can be detected by band 4/band 2 ratio on OLI. On the other hand, a decrease in these ratios in conjunction with an increase in total reflected radiance will indicate bleached areas (Van der Meer et al., 2002, Zhang et al., 2007, Ducart et al., 2016). However, in presence of vegetation, this band can get confused and create vegetation interference. In the spectral region covered by Landsat-8/OLI the chlorophyll of the green vegetation absorbs strongly radiation of the red wavelengths at band 4 and reflects the near infrared at band 5. Thus, bands 4 and 5 can be used to differentiate areas of ferric iron oxides from areas of vegetation. Band ratio of (band 4/band 2) – (band 5/band 4) was used to map ferric iron oxide. The areas of interest for the study are the areas that are lacking the ferric iron which will be the inverse of this band ratio on OLI.

The band ratio (band 4/band 2) x ((band 4 + band 6)/ band 5)) converted from Landsat 7 ETM+ documented in (Rockwell, 2013) has also been applied to highlight ferric iron bearing areas. This band is suggested to identify areas in which ferric iron are present with great abundance because it uses the near infrared band (OLI band 5). The near infrared is more sensitive to ferric iron crystal field absorption as well as grain size variations (Rockwell, 2004). As grain size increases, the absorption feature increase and reflectance level decrease, this in turn makes shifting in apparent reflectance minima to longer wavelengths.

Bleached red beds can be mapped using band ratio band 3/band 4. The high value of the band ratio band 3/band 4 indicate a poor ferric iron material. Subsequently, healthy vegetation will also have high reflectance value in band ratio band 3/band 4. Therefore band ratios (band 3/band 4) – (band 5/band 4) in Landsat 8 OLI was used to indicate the presence of bleached red beds subtracting the effect of vegetation interference. High value of this ratio indicate ferric iron poor rocks and low value indicate ferric iron rich rocks. Ferric iron poor rocks was the indirect indicator for the bleached material.

The band ratio (band 3/band 6) + (band 4/band 5) has been used to map ferrous iron. The band 6/band5 is widely used in literature (e.g., Pour and Hashim, 2014, Banerjee et al., 2019, Soe et al., 2005, Mia and Fujimitsu, 2012), however because ferric iron can have an absorption on

NIR and reflections on SWIR this band ratio was not used as it might create some interference between ferrous and ferric iron mapping. Because ferrous iron can produce a broad absorption on red and NIR in relative to green and SWIR, the band ratio (band 3/band 6) + (band 4/band 5) was applied in this study. Areas that contain coarse-grained ferric iron might still get mapped using this band ratio index (Rockwell, 2013).

5.2.2.2   Ratios to indicate the presence of clays and carbonates

Clay or hydroxyl-bearing minerals and carbonates are characterized by strong absorption feature near 2.1 – 2.4 μm correspond to OLI band 7, and reflectance of 1.55–1.75 μm correspond to band 6 of Landsat-8 (Segal and Merin, 1989, Shi et al., 2010).

Band ratio band 6/ band 7 is commonly used to map clay minerals in Landsat 8 (Zhang et al., 2007, Zhang et al., 2011, Mia and Fujimitsu, 2013, Pour and Hashim, 2015, Liu et al., 2016). Similarities in the spectral feature and reflectance curves of vegetation and clay minerals in Landsat-8/ OLI bands 6 and 7 make their differentiation difficult. Vegetation have a high reflectance on band 6 and low reflectance in band 7. Band ratio band 5/ band 4 can highlights areas of green vegetation with abundant chlorophyll content. Therefore, on Landsat-8/OLI the band ratio band 5/ band 4 was subtracted from band ratio band 6/ band 7 to generate the clay index.

5.2.2.3   TIR Ratios to indicate the presence of carbonates

Carbonates with its two major mineral type, calcite and dolomite have spectral feature in the thermal infrared region. Simulating work in (Ninomiya, 2002) using Landsat-8 the band ratio b10/b11 was used to map carbonates. Areas with high carbonate content expected to show high value in this band ratio.

Boolean logic (Table 7) was applied to combine the results from the different band ratios into a single classified map.

Table 7- Boolean algebra expressions used to generate mineral groups image class

Class Name	Class Color	Ferric iron b4/b2- b5/b4	Ferric iron b4/b2* (b4+b6)/b5	Ferrous iron (b3+b6)/ (b4+b5)	Clays b6/b7- b5/b4	Bleached b3/b4 - b5/b4	Vegetation b5/b4	Carbonate b10/b11
Minor Ferric	cyan	X						
Major Ferric	red	X	X					
Ferrous	yellow			X				
Ferric + Ferrous	blue	Not X	X	X				
Bleached Materials	light orange					X		
Clays	orange				X			
Clays and Ferrous	dark cyan			X	X			
Clays and Minor Ferric	magenta	X			X			
Clays and Major Ferric	purple	X	X	Not X	X			
Dark vegetation's	dark green						X	
Carbonate	Light green							X

5.2.3 Weights of evidence

Visual analysis is widely used in microseepage satellite remote sensing literatures to characterize the spatial correlation between minerals and hydrocarbon microseepage. Thus, prediction models results are more of subjective decisions by the analysts. Here, weight of evidence was used to quantitatively investigate the spatial association between hydrocarbon basin, indicated by oil wells, and mineral anomalies resulted from PCA with Crosta technique and band ratio analysis. PCA and band ratio mineral maps were treated as evidential theme or predictor of hydrocarbon microseepage occurrence. These maps were reclassified and assigned an integer value. This process was deemed necessary because weight of evidence calculation requires integer raster. These raster's were then combined into one evidential theme with 30 classes for PCA and 19 classes for band ratio, each represent one class of mineral (See Appendix C). The cells that were highlighted by different minerals in different PC or band ratio images were lumped into a new category. For example, the cells that were marked in PC2456 as ferric and in PC3467 as a bleached area, were assigned a class called ferric bleached areas.

An import assumption in weight of evidence modelling is that each unit cell should contain at most one training site. The training sites of oil wells were reduced from 340 to 337 to make sure each training point is occupying a single unit cell, no reduction on number of training points were required for dry hole wells. A unit cells size of 900 m2 (30 m cell size) was selected based on map scale and after considering training points distributions. As the generated

evidential maps are of nominal data type, they were evaluated using categorical weights. The confidence level was arbitrarily chosen to be 80%, which roughly equals to Studentized contrast of 0.842.

The training sites consist of two types of wells. The weights were calculated twice: one time for oil wells to quantify the spatial relationship between the evidential theme and presence of hydrocarbon microseepage (positive evidence) and the other time for dry hole wells to quantify the spatial relationship between the evidential theme and absence of hydrocarbon microseepage (negative evidence). These evidences were combined to provide support of hydrocarbon microseepage evidence. Weight of evidence analysis result in a set of statistical measures which reflect the degree of spatial association including weights (W+ and W-), contrast (C) and Studentized contrast (CS).

Weights (W+, W-) provide spatial association between the training points and the evidential theme. When the value of the weight W+ is positive and W- is negative, it indicates there are more anomalies on that particular class than would occur due to random chance. Conversely, when the weight of W- is positive and W+ is negative, it indicates that the training sites unlikely to occur within the pattern or theme class indicating absence of the mapped feature. The difference between the weights is known as the Contrast (C) which measures the overall correlation between the training site and the theme. A large positive contrast indicates positive correlation; whereas negative contrast indicates negative correlation. The Studentized Contrast (CS) is a student t-test that provides a measure of the uncertainty in C value and the degree of the significance. It is defined as the ratio of the contrast to its standard deviation. A large positive contrast with large Studentized contrast means the contrast is large compared to its standard deviation, implying the class or pattern is a useful predictor of the training sites.

The statistics obtained from weight of evidence calculation are shown in Table 14 and Table 15 in result section. Contrast value helped in inferring the strength of the spatial correlations. The weights were analysed to determine the classes with positive contrast. These evidence maps were reclassified to combine the selected classes. Weights were recalculated and the reclassified evidence maps were combined to generate a relative ranking of microseepage favorability areas for each PCA model and band ratio model.

The reclassified two evidence maps resulted from weight calculation for PCA and band ratio models were then combined to generate a continuous scale probability map. To decide on the threshold that can define classes' breaks, the Cumulative Area-Posterior Probability (CAPP) curve was created by plotting posterior probability of the Y-axis and cumulative area on the X-axis Figure 3. Breakpoint was selected at (0.440968) to separate microseepage from non-microseepage areas. Breakpoint (0.644347) was selected to separate permissive microseepage from favorable microseepage areas. To validate the combined model result, the map was reclassified into binary form. Everything below (0.440968) was assigned zero indicating non-microseepage areas and everything above that point was assigned one indicating potential microseepage areas.
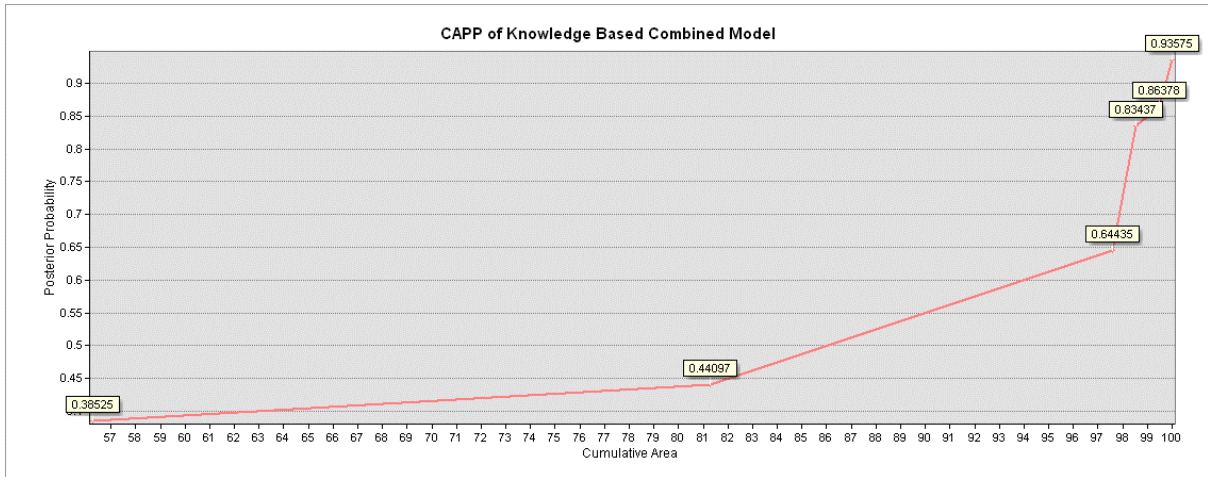
Figure 3 - Cumulative area-posterior probability (CAPP) curve for KW based combined model showing class breaks

5.2.4   Mixture Tuned Matching Filtering (MTMF)

MTMF algorithm was used to generate map of the distribution and abundance of image end-member minerals. In order to extract the image endmembers that will be used for classification, minimum noise fraction (MNF), pixel purity index (PPI) and n-dimensional visualization were performed on the multispectral image to decompose the mixed pixels into spectra of known mineral. Figure 4 shows the workflow that summaries the end-to-end tasks performed to complete the classification.

35
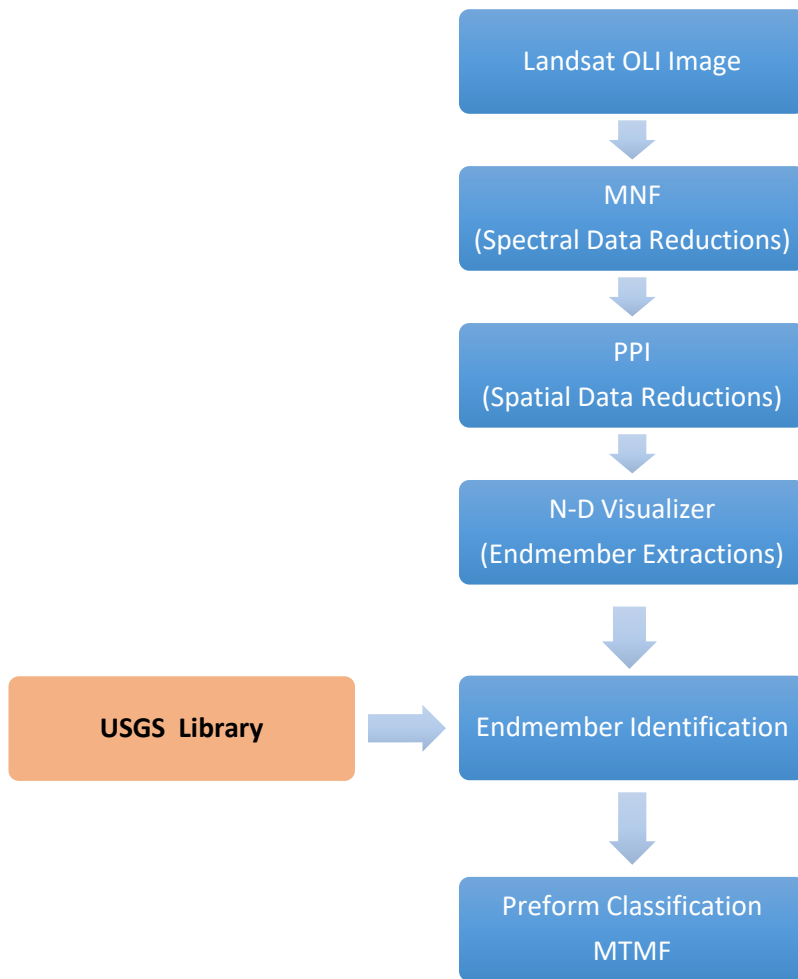
Figure 4 - End-to-end processes performed to complete MTMF classification

The first step was to perform MNF transformation. The MNF is used segregate and equalize the noise in the input image. It additionally reduces the computational requirements for subsequent processing. This is a mandatory step for MTMF classification because it requires an MNF transform input file. The input spectra need to be in MNF space, too. Examination of the eigenvalues and eigenimages (Figure 5) shows that the resulted MNF bands eigenvalues were all above the defined threshold (>1.25). Therefore, the seven MNF bands were used for the subsequent processing.

Next, the Pixel Purity Index (PPI) technique was applied on the 7 MNF images to identify the most spectrally pure pixels (extreme). 11,045,424 pixels were operated. PPI threshold value was set to 2.50, the unit of this threshold value is the noise standard deviation. The MNF transformed image is already in noise standard deviation. This means that the tolerance factor was set to 2.50 times the noise of the data which corresponds to 2.5 times the pixels value. A PPI image was created in which the digital number of each pixel corresponds to the number of the times that pixel was recorded as extreme. A region of interest was created from the PPI

image by applying a threshold value to select only pixels that were flagged 10 times or more as extreme or pure.
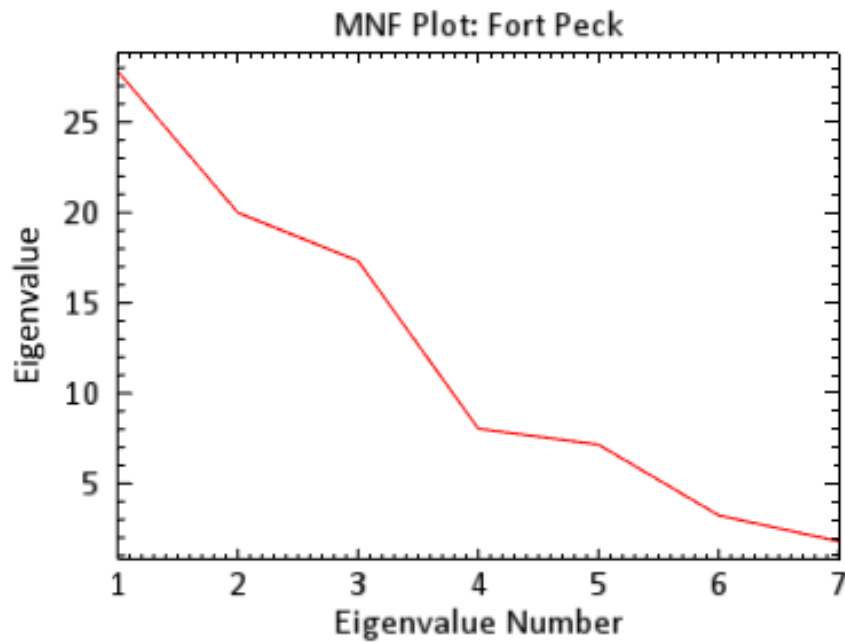
Figure 5 - Landsat 8/OLI Eigenvalues Plots (MNF)

To separate those identified pure pixels into classes corresponding to the spectral endmembers of the image, the PPI ROI pixels were used as input to the n-Dimensional visualizer. The n-D visualizer is like 2d scatter plot but it can plot the image data in more than two dimensions (i.e., in all available dimensions). The scatter plot will show pixels that were plotted against MNF bands but only the purest pixels determined from the PPI (i.e., those in the ROI). Four to five bands were chosen each time and then their axes got rotated to examine the shape of the formed data cloud. After several rotations with different MNF bands, the pixels that formed corners in the data cloud and remained together in all projections got selected. Mean spectra were then calculated and exported into classes in both reflectance and MNF space shown in Figure 6 and Figure 7.
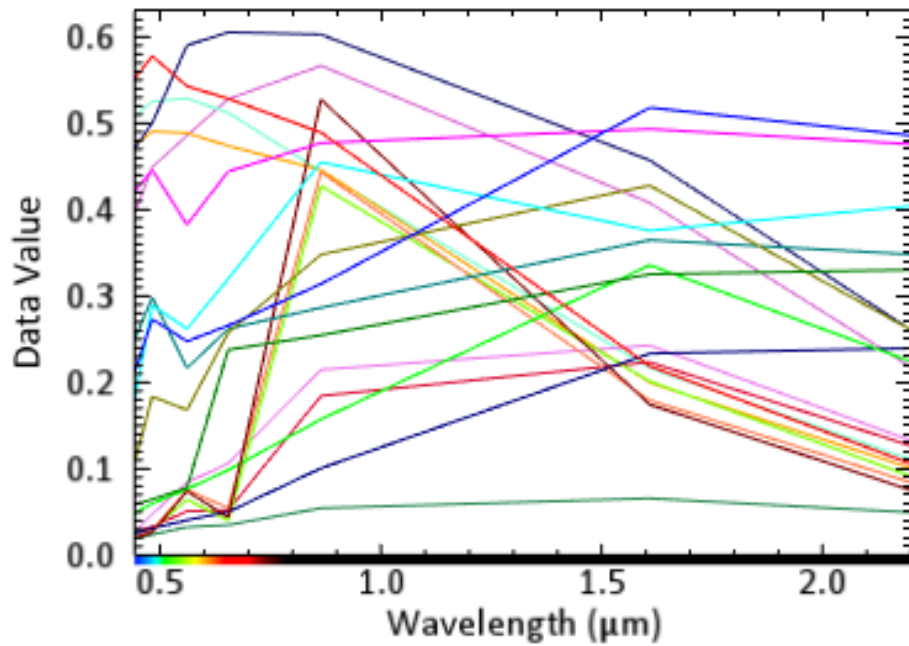
Figure 6- Result of n-D Visualizer - Spectral signatures show the mean value of reflectance data for each class.
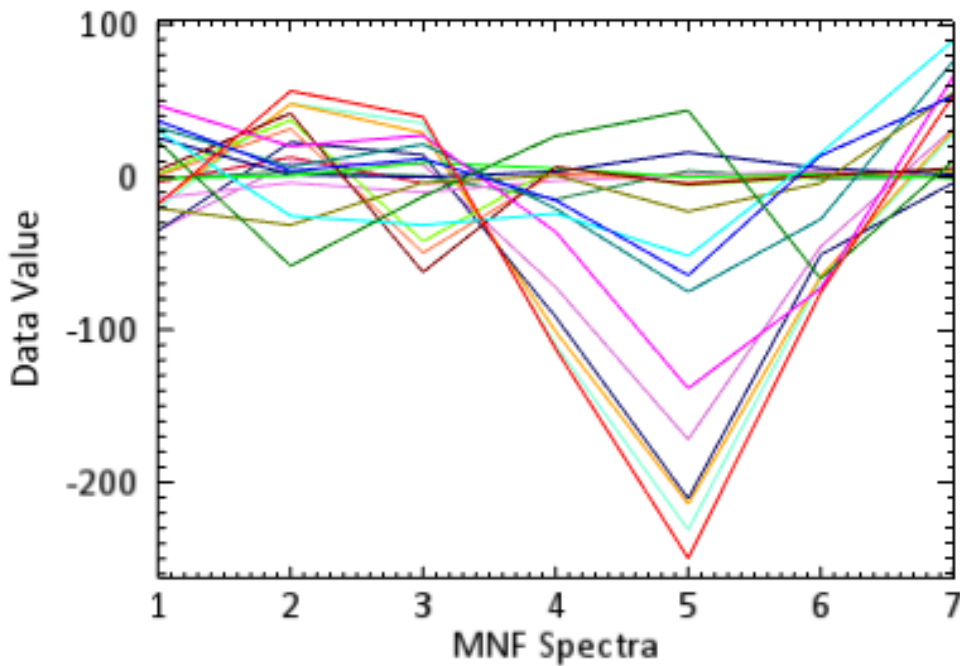
Figure 7 - Result of n-D Visualizer - Spectral signatures show the mean value of MNF spectra for each class.

Spectral analysis was performed on nDV visualizer exported classes in a process to match image spectra to USGS library spectra in order to identify the material that comprised the image endmembers. The USGS spectral library was downloaded and resampled to Landsat 8 image resolution and used for comparison purpose. Each image endmember spectral shape was compared to the spectral shape of known materials in the spectral library. The best match was selected for each image endmember. Figure 8 shows an example of image endmember and USGS spectral library spectral comparison.
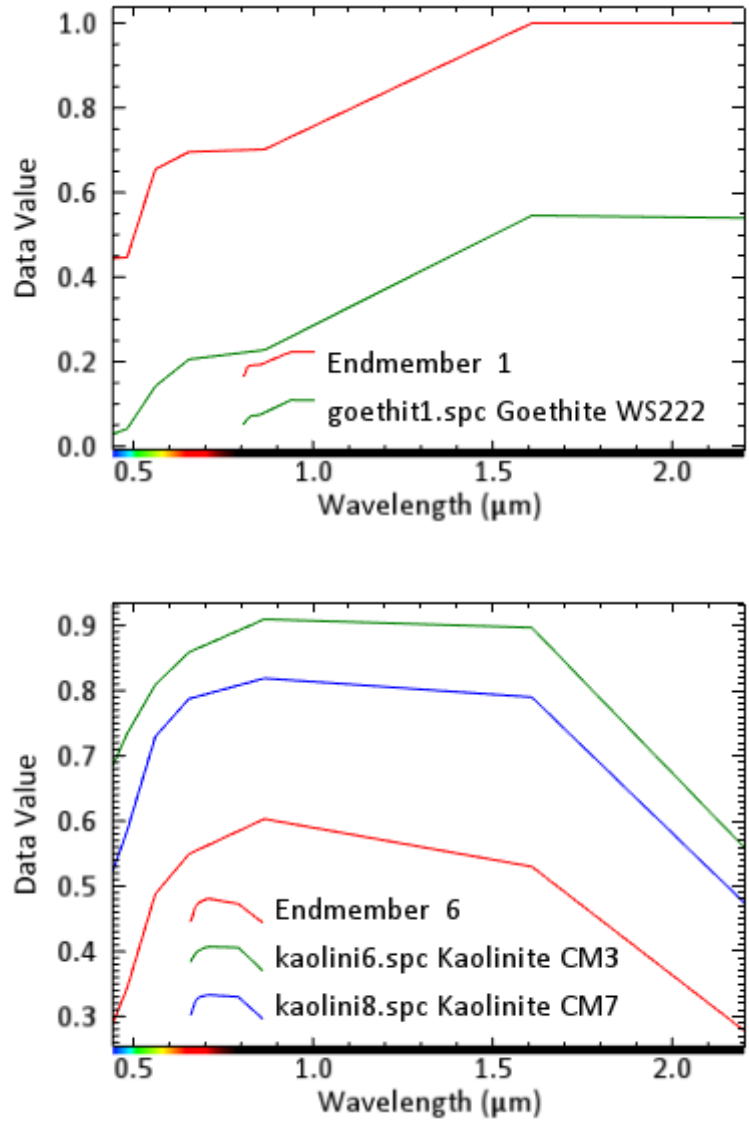
Figure 8 – Examples of image endmember and USGS spectral library spectral comparison

Identified minerals groups were used for the subsequent classification. MTMF was applied to the MNF transformed image. MTMF requires also the endmembers to be in the same MNF space. For this reason, the MNF spectra were used instead of reflectance spectra as the input endmembers. MTMF outputs two sets of grey scale images that consist of MF scores image and infeasibility scores image for each endmember. Brighter pixels in the image represent higher mineral abundance. MF score values indicate the relative degree of match to the reference spectrum and range from 0 to1, where 1 is a perfect match. The infeasibility score images measures the mixing between the composite background and the target. High values are likely to be MF false positives.

In order to determine the purest pixels that have close matching to the target mineral spectra, a two dimensional scatter plot was used to identify the pixels with low infeasibilities and high MF scores. The best match to the target spectra can be obtained when the MF score is high

(near 1) and the infeasibility score is low (near 0). Figure 9 shows an example of the 2d scatter plot used to identify the pixels with low infeasibilities and high MF scores. The resulted abundance pixels were combined as color-coded thematic image maps to show the distribution of minerals.
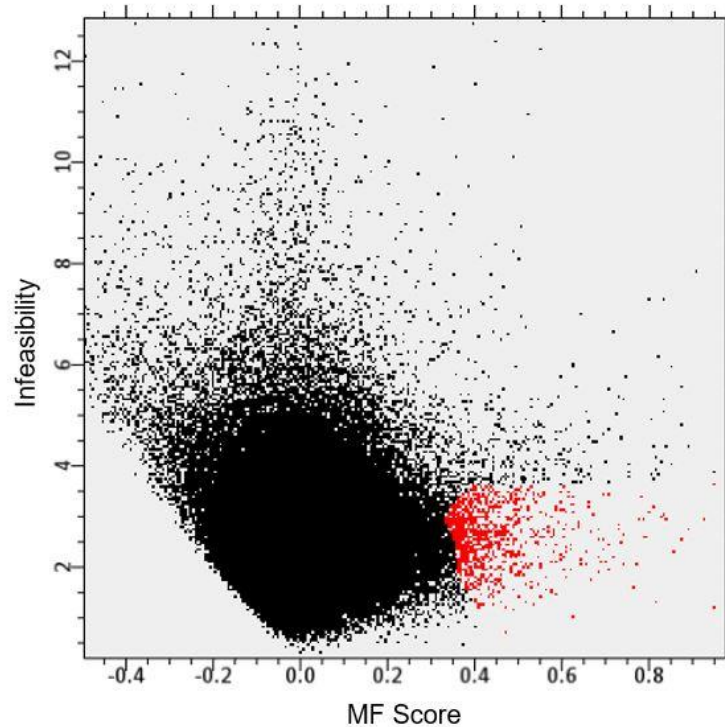
Figure 9 - MF Score vs. Infeasibility plot used to determine purest pixels matching target mineral spectra.

5.3 Data-driven model

5.3.1 Support Vector Machine (SVM)

A data driven approach was used to explore the possibility of mapping potential area of microseepage. The data driven model is not built based on expert opinions such as knowledge based models instead its parameters are estimated from measured data. Since it is purely estimated from data, the data driven approach might help redefine or adjust the previously held conceptual model for hydrocarbon microseepage.

Because the available dataset of microseepage fluxes is rather poor, the known oil, gas and dry hole wells in Fort Peck reservation were used as training points for the model. Assumption was that areas with surface microseepage are linked to subsurface hydrocarbons accumulations. In other words, positive hydrocarbon microseepage fluxes from surface occur over the areas that host natural gas and oil reservoirs indicated by oil wells.

Oil and gas data in the Fort Peck reservation was downloaded from Montana Board of Oil and Gas Conservation (MBOGC) web site (MBOGC, 2019). To prepare the data for classifications, wells dataset were further categorized into two classes: oil and dry hole, where the oil wells indicate potential gas fluxes and dry hole wells indicate no fluxes. These were used as training and validation data for the classification. PCA was performed on the Landsat OLI image. The

classification occurs on the PCA image and the training data guides the classification to identify areas on the image that belong to certain categories. In this case either oil or dry hole. Figure 10 shows a complete workflow that summaries end-to-end tasks performed to complete the classification.
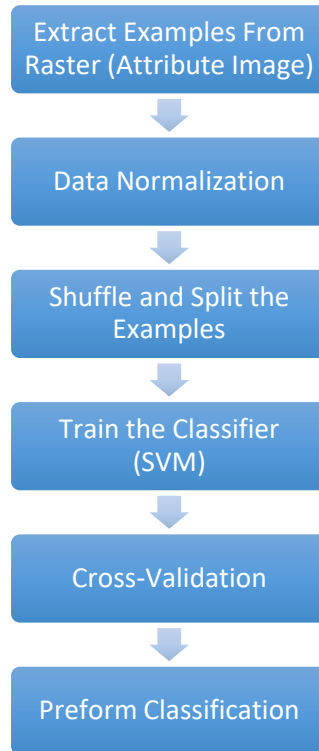
Figure 10 - End-to-end processes performed to complete the data-driven classification using SVM.

As an initial step, pixels from the attribute image that fall within the training data boundaries were extracted and assigned an integer class value based on the classes that were defined in the training dataset as follows.

Class Name	Value
Dry hole	1
Oil	2

The next step was to normalize the data (i.e., the examples obtained from the first step). The purpose of this step is to scale the data into a consistent range of values prior to classification. Normalization can help in speeding up the learning phase. Additionally, it will help in avoiding any numerical problems such as precision loss from arithmetic overflows. Furthermore, it prevents attributes with a large original scale from biasing or dominating the solution. For data normalization, a linear normalization (linear scaling by unit variance and offset by mean value) was performed using below formula:

DataOut = Gain * INPUT_EXAMPLES + Offset

Where

Gain$_x$ = 1/ (Max$_x$-Min$_x$) one for each attribute in **INPUT_EXAMPLES** where **Max$_x$** is the maximum example value and **Min$_x$** is the minimum example value for the attribute **x**.

Offset = - Mean$_x$ one for each attribute in **INPUT_EXAMPLES** where **Mean$_x$** is the mean of all example values for attribute **x**.

Because we need to train the classifier on one set of examples and evaluate the classifier with another set, the next step in the classification process was to shuffle the examples to create a random distribution of data and then split the data. The data were split into two-element arrays where the first array element contains the examples that were used to train the classifier. The second array element (the unknown dataset to the classifier) contains the examples used to evaluate the classifier. This is a common strategy in machine learning to obtain the prediction accuracy from an independent data set (the unknown). The split fraction was set to 0.8 which means that 80 percent of the original examples will be used for training and the other 20 percent will be used for evaluation.

Next, SVM classifier was defined and trained. This step is only to train the classifier. The examples and classes' values are passed to the trainer that iteratively trains the classifier in order to minimize its errors. The trainer iterates to adjust the classifier internal parameters and it decreases or falls until the error (also called loss) meets a specified threshold below which iterations can stop. If the specified threshold was not met, it will work until it reaches a specified maximum number of iterations. The threshold (also called convergence criterion) value was set to 0.001 below which the iteration will stop. The maximum number of iteration value was set to 100.

The calculated loss is a unit less number that indicates how closely the classifier fits the training data. A value of zero represents a perfect fit and when the value is further from zero, it is considered less accurate fit.

By examining different configurations, the SVM parameters that were found to produce the highest accuracy are:

Kernel Type: Radial Basis Function (RPF)

RBF kernel has two parameters: **gamma** which can be thought of as the spread of the kernel and C which is the penalty for misclassifying a data point, they were set as:

Gamma: 0.1	**C:** 100

To determine the baseline of the classifier performance and to find out what would be the success rate of the model, an evaluation step was performed. Using the cross-validation technique, a confusion matrix and accuracy metrics between the true class values and the predicted class values were calculated. In this step, only the dataset that was designated for

evaluation and was not used to train the classifier has been used for the evaluation. Because of the random nature resulted from shuffling the examples, the resulting confusion matrix and accuracy metrics will vary. Classification was conducted 10 times and the most accurate classification was recoded. Table 8 and Table 9 exhibit details of accuracy metrics that reveal how well the classifier has performed.

Table 8 - Confusion Matrix

Truth

		Oil	Dry Hole	Total
Predicted	Oil	26	3	29
	Dry Hole	33	96	129
	Total	59	99	158

Table 9 - Accuracy Metrics derived from the confusion matrix

Overall accuracy The sum of correctly classified values divided by the total number of values 122/158	0.772152
Kappa coefficient Represent the agreements between classification and truth values. A kappa value of 1 represent perfect agreements, while value of zero represent no agreements	0.457355

	Dry Hole	Oil
Error of commission (false positive) Fraction of values that were predicted to be in a class but belong to another class.	0.255814	0.103448
Error of omission (false negative) Fractions of values that belong to the class but were predicted in different class (misclassified)	0.0303030	0.559322
Producer accuracy (Recall) A probability that a value in a given class was classified correctly	0.969697	0.440678
User accuracy (Precision) Probability that a value predicted in a certain class really is that class	0.744186	0.896552
F1 score Harmonic mean of user accuracy and producer accuracy	0.842105	0.590909

The final step was to run the classification, prior to that, the source raster was normalized using the gains and offsets that were computed earlier for the examples. Then classification was performed on the normalized raster using the trained classifier.

6 Results

6.1 Knowledge-based models

Below results demonstrate the performance of applying knowledge-based methods in discriminating hydrocarbon gases related surface anomalies.

6.1.1 Principle Component Analysis (PCA)

6.1.1.1 PC to indicate the presence of ferric and ferrous iron minerals and bleaching of red beds.

Table 10 describes the principle component transformation on OLI bands 2, 4, 5 and 6. The rules for ferric iron oxide mapping can be defined by the magnitude of eigenvectors loading for band 2 and band 4 as moderate or strong loading with opposite sign in either PC3 or PC4.

By looking for the moderate or large eigenvector loading for band 2 and band 4 where these loading are also opposite in sign, we can predict that ferric iron can be discriminated by bright pixels in PC4, moderate loading of band 4 (0.4248) and high loading of band 2 (-0.9024) but with opposite sign corresponding to absorption in band 2 and reflectance in band 4 for ferric iron oxide. Therefore, iron oxide will be discriminated by bright pixels in PC4.

However, in PC4 band 6 has small positive loading and band 5 has small negative loading. This implies that ferrous iron will be mapped by bright pixels too. This in turn might create random noise or enhance the false anomalies. Ferric iron oxides are shown in green colour in Figure 11.

Table 10 -Landsat 8 PC-2456 for Ferric and Ferrous Iron Mapping

Eigenvectors	Band 2	Band 4	Band 5	Band 6
PC 1	0.1833	0.3652	0.5072	0.7587
PC 2	0.1277	0.1101	-0.8561	0.4884
PC 3	-0.3683	-0.8209	0.0837	0.4282
PC 4	-0.9024	0.42489	-0.0523	0.0484

Band ratio band6/band5 can highlights ferrous oxide bearing minerals. Looking for moderate or large eigenvector loading for band 5 and band 6 where these loading are also opposite in sign, we can predict that ferrous iron can be discriminated by bright pixels in PC2 (Table 10), high loading of band 5 (-0.8561) and moderate loading of band 6 (0.4884) with opposite in sign. The negative loading of band 5 in this PC implies that vegetation, if present, in the area will be mapped by dark pixels.

6.1.1.2 PC to indicate the presence of clay and carbonate minerals.

PCA for clays and carbonates was examined by the eigenvector loading for band 6 and band 7 to determine which PC image will enhance the presence of clay and carbonates.

Table 11 describes the principle component transformation on OLI bands 2, 5, 6 and 7. The rules for clay and carbonate mapping can be defined by the magnitude of eigenvectors loading for band 6 and band 7 as moderate or strong loading with opposite sign in either PC3 or PC4. In this PC image, the visible band 4 was omitted deliberately to avoid ferric iron oxide mapping.

Table 11 -PC-2567 for Clays and Carbonates Mapping

Eigenvectors	Band 2	Band 5	Band 6	Band 7
PC 1	-0.1683	-0.3940	-0.6861	-0.5878
PC 2	-0.0643	0.8783	-0.0903	-0.4649
PC 3	0.1531	0.2654	-0.7217	0.6206
PC 4	0.9716	-0.0519	-0.0111	-0.2304

Looking for moderate or large eigenvector loading for band 6 and band 7 where these loading are also opposite in sign, we can predict that clay and carbonate minerals can be discriminated by dark pixels in PC3, high loading of band 6 (-0.7217) and moderate loading of band 7 (0.6206) with opposite signs. Clays also show reflectance in band 5 and 2. Therefore, in order to get accurate mapping of clays, band 5 and band 2 should follow the sign of band 6. However, that is not the case with the current analysis for PC3-2567, which means it might map other materials with clays and carbonate. For instance and since band 5 was used in this PC, then loading of band 5 and band 6 can be used to map ferrous iron. Band 6 and band 5 follow the sign of band 7 which indicate that PC3 will map both ferrous iron and clays minerals in dark pixels. Figure 11 shows PC3-2567 mapped pixels above the threshold value (mean + 1 standard deviation).

Table 12 describes the principle component transformation on OLI bands 2, 4, 6 and 7. Looking for moderate or large eigenvector loading for band 6 and band 7 where these loading are also opposite in sign, we can predict that clay and carbonate minerals can be discriminated by dark pixels in PC2 or PC4. PC2 has moderate loading of band 6 (-0.6183) and band 7 (0.7018) with opposite sign. PC4 has small loading of band 6 (-0.0270) and band 7 (0.1105) with opposite sign. However, the moderate to high loading of band 2 and band 4 in PC4 with opposite sign would help in suppressing the ferric iron from the result. Therefore, PC4 was selected to map clay and carbonate minerals. PC4 is shown in Figure 11 for pixels above the threshold value (mean + 1 standard deviation).

Table 12 - PC-2467 for Clays and Carbonates Mapping

Eigenvectors	Band 2	Band 4	Band 6	Band 7
PC 1	-0.1765	-0.3278	-0.6923	-0.6180
PC 2	0.1319	0.0754	-0.6959	0.7017
PC 3	0.2938	0.8745	-0.1886	-0.3364
PC 4	-0.9300	0.3492	-0.0270	0.1105

Table 13 describes the principle component transformation on OLI bands 3, 4, 6 and 7. This PC also used to map clays and carbonates.

Table 13 – PC3467 for Clays and Carbonates Mapping

Eigenvectors	Band 3	Band 4	Band 6	Band 7
PC 1	0.2386	0.3268	0.6822	0.6088
PC 2	0.2310	0.2467	-0.7295	0.5945
PC 3	-0.4739	-0.7070	0.0373	0.5235
PC 4	0.8154	-0.5764	0.0287	-0.0423

Looking for moderate or large eigenvector loading for band 6 and band 7 where these loading are also opposite in sign, we can predict that clay and carbonate minerals can be discriminated by dark pixels in PC2, moderate loading of band 6 (-0.7295) and high loading of band 7 (0.5945) with opposite signs. Notice that PC4 has a loading for band 6 and band 7 with opposite signs which also can be used for clay and carbonates mapping however it has small magnitude. The opposite loading of band 3 and band 4 can help in mapping of bleached material. However, the result of PC4 shows the manifestation of vegetation due to strong reflections of band 3 and absorption of band 4. Since PC2-3467 yield similar result like PC3-2567, instead PC4-3467 will be used here to show bleached materials. Figure 11 shows the result of the analysis on PC4-3467 with threshold value greater that mean + 1 standard deviation. The PC was overlaid with band ratio b5/b4 to discriminate vegetation.
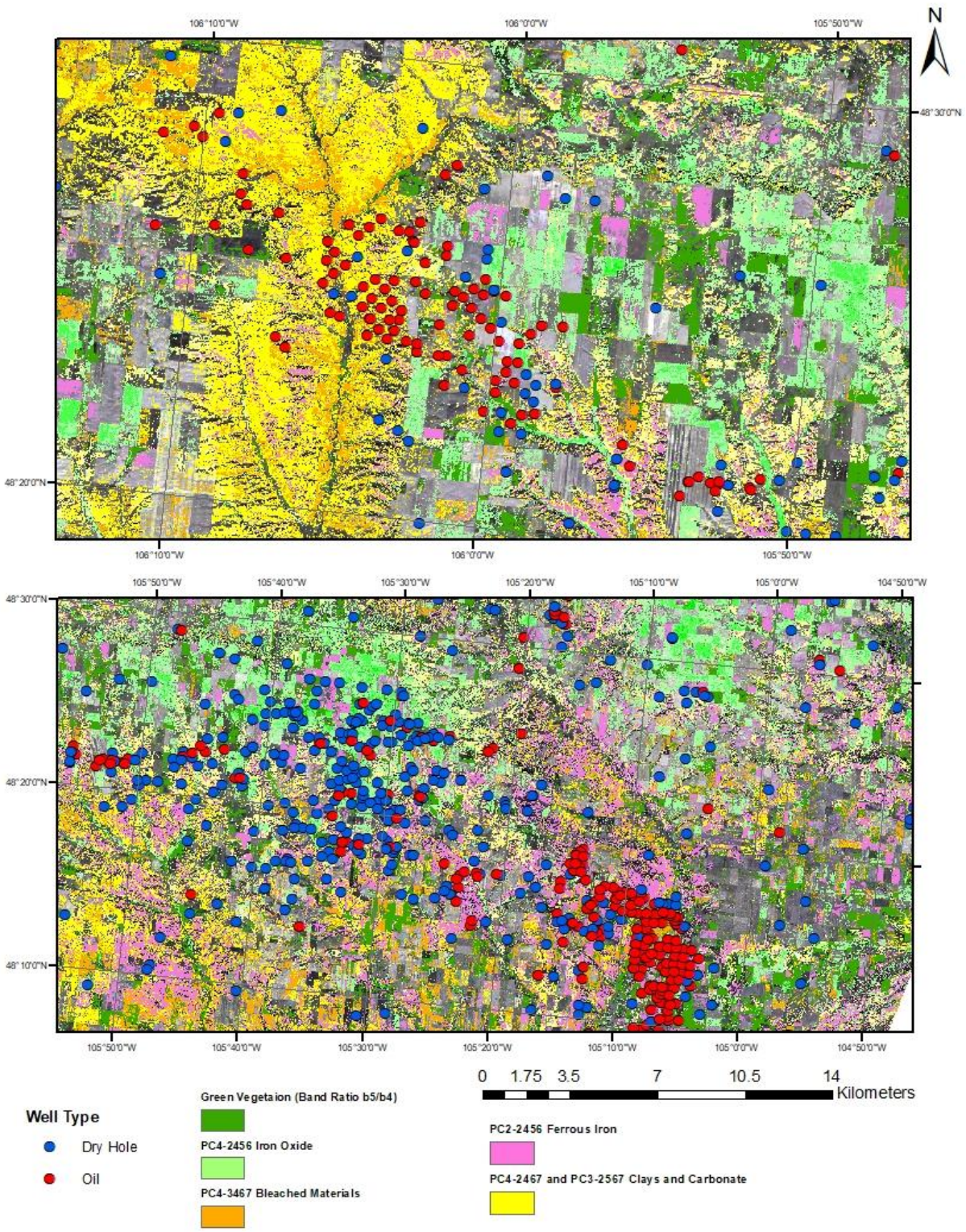
Figure 11- PCA crosta technique results showing the distributions of wells and minerals group

6.1.2 Band Ratio

6.1.3 Ratios to indicate the presence of ferric, ferrous iron minerals and bleached red beds

Bands 4 and 5 were used to differentiate areas of ferric iron oxides from areas of vegetation. Band ratio of (band 4/band 2) – (band 5/band 4) was used to map ferric iron oxide.

Band ratio (band 3/band 4) – (band 5/band 4) was used to indicate the presence of ferrous iron minerals and bleached red beds subtracting the effect of vegetation interference. High value of this ratio indicate ferric iron poor rocks and low value indicates ferric iron rich rocks. Ferric iron poor rocks were the indirect indicator for the bleached material.

Because ferrous iron can produce a broad absorption on red and NIR relative to green and SWIR, the band ratio OLI (3+6)/(4+5) adopted from Rockwell (2013) was used to highlight the area of ferrous iron. However, this index can highlight areas that contain coarse-grained ferric iron.

6.1.4 Ratios to indicate the presence of clays and carbonates

Band ratio band 6/ band 7 was used to map clay minerals in Landsat 8. The result of applying this ratio shows clearly clays spectra are contaminated with vegetation spectra. Therefore, on Landsat-8/OLI, the band ratio 5/4 was subtracted from band ratio 6/7 to generate the clay index.

6.1.5 TIR Ratios to indicate the presence of carbonates

The band ratio b10/b11 was used to map carbonates. Areas with high carbonate content expected to show high value in this band ratio (Ninomiya, 2002).

Figure 12 illustrate the result from band ratio technique showing different minerals groups generated using band algebra with wells distributions.

Legend

					Well Type
Green Vegetaion	Ferric and Ferrous Iron	Clay and Ferric	Major Ferric Iron	Bleached Materials	● Dry Hole
Carbonate	Clays and Major Ferric Iron	Clays	Ferric Iron	Ferrous Iron	● Oil
Clays and Ferrous Iron					

0   2.5   5   10   15   20 Kilometers
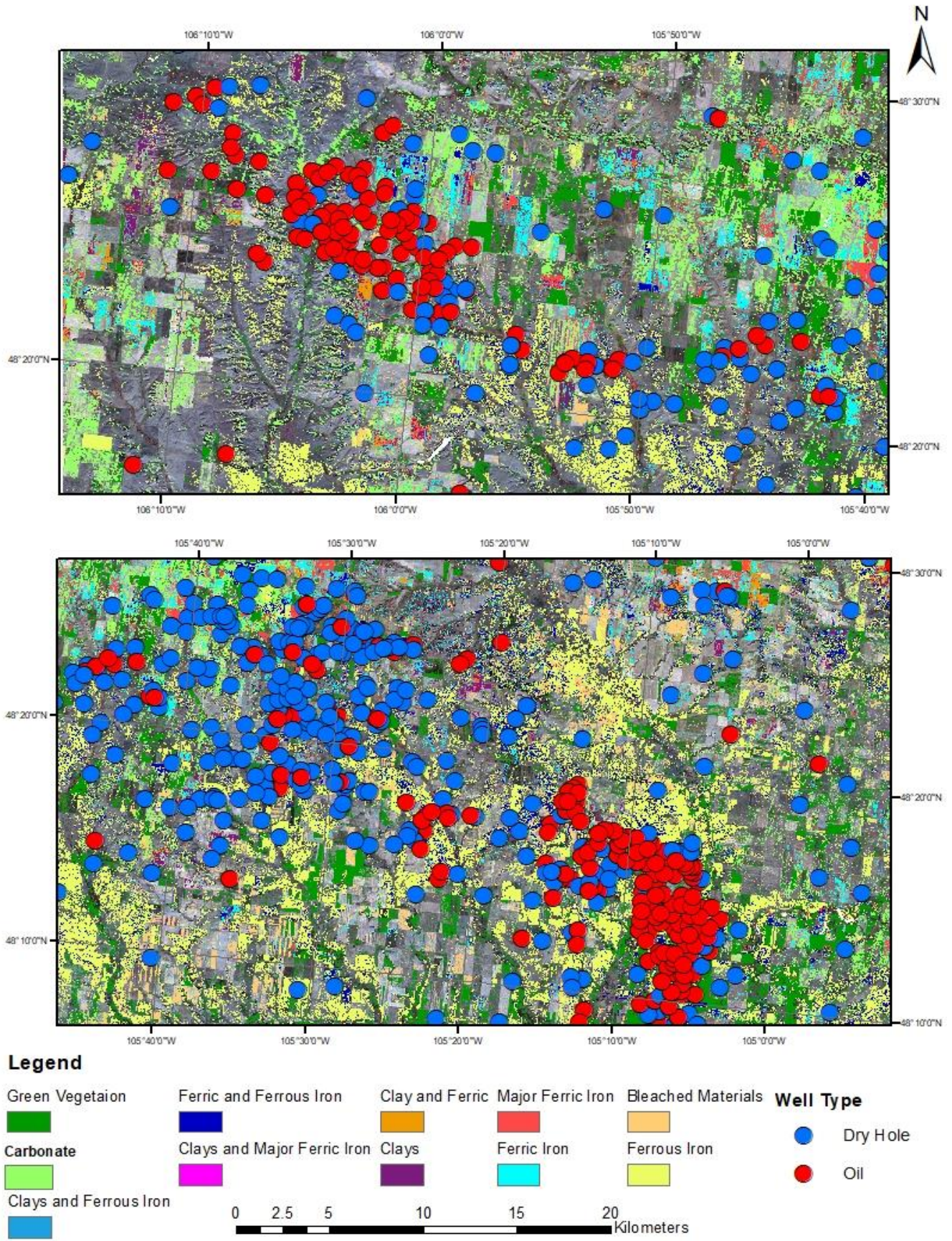
Figure 12- Band ratio technique results showing different minerals group generated using band algebra.

50

6.2 Weights of evidence

From literature review, the conceptual model for recognizing locations of microseepage was identified. A list of predictor maps were developed based on this model. PCA and band ratio assembled to produce evidential themes. Weights of evidence was performed to quantify the spatial associations between hydrocarbon microseepage and mineral anomalies. Statistical values obtained from weight of evidence analysis are shown in Table 14 and Table 15 for PCA and in Table 18 and Table 19 for band ratio. These statistics were used to select and adjust the most appropriate groupings of the evidential themes and provide measure of the model confidence.

6.2.1 Weights of evidence for PCA

As can be seen by the contrast values in Table 14, there is a strong spatial correlation between oil wells and clays carbons, bleached, bleached-ferric and bleached-clays carbons minerals. Clay's carbons and bleached-clays carbons show positive W+ and C, but as it can be seen the confidence is lower than the defined threshold (0.842) for Studentized contrast. Mineral classes 4, 11 and 27 show positive contrast as well. However positive contrast here was the result of one oil well in a small area. One point should not be considered sufficient for the spatial correlation. In contrast, clays carbons, bleached and bleached-ferric were chosen as predictor for oil basin. These were extracted and combined into a binary evidence layer and statistics were re-calculated and shown in Table 16. Class 2 refers to the selected minerals that correlate with oil wells; everything else was grouped into class 1.

Table 15 measures the spatial correlation between dry hole wells and the different minerals. Dry hole wells, and as explained by C value, have strong spatial correlation with the classes 1, 2, 3, 9 and 18 representing respectively background, ferric, ferrous, bleached and clays ferrous - ferric minerals (See Appendix C). These were extracted, except bleached, and combined into a binary evidence layer. Statistic shows that bleached areas correlate with both oil and dry hole wells. However, as per the defined conceptual model, bleached is correlated with oil therefore it was discarded for dry hole wells. Class 2 was assigned to the minerals that correlate with dry hole wells and everything else was grouped into class 1. Statistics were re-calculated and shown in Table 17.

Table 14 - Weights and statistics for PCA minerals evidential theme with oil wells

CLASS	AREA_SQ_KM	NO_POINTS	W+	S_W+	W-	S_W-	CONTRAST	S_CONTRAST	STUD_CNT	WEIGHT
1	2982.79	126	0.0484	0.0891	-0.0278	0.0688	0.0762	0.1126	0.6765	-0.0439
2	1119.05	37	-0.1966	0.1644	0.0272	0.0577	-0.2238	0.1742	-1.2844	-0.1966
3	308.09	9	-0.3205	0.3333	0.0104	0.0552	-0.3309	0.3379	-0.9793	-0.3205
4	20.64	1	0.1855	1.0000	-0.0005	0.0546	0.1860	1.0015	0.1858	-0.0439
5	737.18	34	0.1362	0.1715	-0.0142	0.0574	0.1504	0.1809	0.8316	-0.0439
6	6.99	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0439
7	172.93	2	-1.2471	0.7071	0.0149	0.0546	-1.2620	0.7092	-1.7794	-1.2471
8	0.82	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0439
9	648.17	67	0.9433	0.1222	-0.1411	0.0609	1.0844	0.1365	7.9446	0.9433
10	25.65	9	2.1657	0.3334	-0.0240	0.0552	2.1897	0.3379	6.4798	2.1657
11	1.89	1	2.5775	1.0002	-0.0027	0.0546	2.5803	1.0017	2.5758	2.5775
12	0.09	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0439
13	467.00	20	0.0621	0.2236	-0.0038	0.0562	0.0659	0.2306	0.2858	-0.0439
14	2.29	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0439
15	30.70	1	-0.2117	1.0000	0.0007	0.0546	-0.2124	1.0015	-0.2121	-0.0439
16	0.10	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0439
17	568.04	12	-0.6446	0.2887	0.0340	0.0555	-0.6786	0.2940	-2.3085	-0.6446
18	377.28	3	-1.6217	0.5774	0.0372	0.0547	-1.6589	0.5799	-2.8604	-1.6217
19	313.80	9	-0.3388	0.3333	0.0111	0.0552	-0.3500	0.3379	-1.0358	-0.3388
20	48.02	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0439
21	244.79	1	-2.2877	1.0000	0.0267	0.0546	-2.3144	1.0015	-2.3110	-2.2877
22	2.44	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0439
23	217.11	3	-1.0691	0.5774	0.0173	0.0547	-1.0865	0.5799	-1.8734	-1.0691
24	2.64	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0439
25	18.09	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0439
26	0.83	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0439
27	1.33	1	2.9294	1.0003	-0.0028	0.0546	2.9322	1.0018	2.9268	2.9294
28	0.07	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0439
29	41.45	1	-0.5117	1.0000	0.0020	0.0546	-0.5137	1.0015	-0.5129	-0.0439
30	0.11	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0439
31	12.61	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0439
32	0.08	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0439

Table 15 -Weights and statistics for PCA minerals evidential theme with dry holes wells

CLASS	AREA_SQ_KM	NO_POINTS	W+	S_W+	W-	S_W-	CONTRAST	S_CONTRAST	STUD_CNT	WEIGHT
1	2982.79	178	0.1025	0.0750	-0.0616	0.0605	0.1641	0.0963	1.7029	0.1025
2	1119.05	79	0.2705	0.1125	-0.0491	0.0518	0.3196	0.1239	2.5802	0.2705
3	308.09	25	0.4098	0.2000	-0.0195	0.0485	0.4294	0.2058	2.0864	0.4098
4	20.64	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0102
5	737.18	26	-0.4234	0.1961	0.0328	0.0485	-0.4562	0.2020	-2.2583	-0.4234
6	6.99	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0102
7	172.93	5	-0.6222	0.4472	0.0097	0.0474	-0.6319	0.4497	-1.4050	-0.6222
8	0.82	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0102
9	648.17	46	0.2758	0.1474	-0.0270	0.0497	0.3028	0.1556	1.9462	0.2758
10	25.65	1	-0.3232	1.0000	0.0008	0.0471	-0.3241	1.0011	-0.3237	0.0102
11	1.89	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0102
12	0.09	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0102
13	467.00	8	-1.1456	0.3536	0.0395	0.0475	-1.1851	0.3567	-3.3221	-1.1456
14	2.29	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0102
15	30.70	1	-0.5031	1.0000	0.0015	0.0471	-0.5045	1.0011	-0.5040	0.0102
16	0.10	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0102
17	568.04	34	0.1055	0.1715	-0.0081	0.0490	0.1136	0.1784	0.6370	0.0102
18	377.28	31	0.4223	0.1796	-0.0251	0.0488	0.4474	0.1861	2.4040	0.4223
19	313.80	10	-0.5249	0.3162	0.0158	0.0476	-0.5406	0.3198	-1.6906	-0.5249
20	48.02	2	-0.2573	0.7071	0.0013	0.0472	-0.2586	0.7087	-0.3649	0.0102
21	244.79	1	-2.5791	1.0000	0.0275	0.0471	-2.6066	1.0011	-2.6037	-2.5791
22	2.44	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0102
23	217.11	3	-1.3605	0.5774	0.0196	0.0472	-1.3801	0.5793	-2.3825	-1.3605
24	2.64	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0102
25	18.09	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0102
26	0.83	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0102
27	1.33	1	2.6380	1.0003	-0.0021	0.0471	2.6400	1.0014	2.6362	2.6380
28	0.07	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0102
29	41.45	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0102
30	0.11	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0102
31	12.61	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0102
32	0.08	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0102

Table 16 - Weights and statistics for minerals associated with oil wells from PCA WofE analysis.

CLASS	AREA_SQ_KM	NO_POINTS	W+	S_W+	W-	S_W-	CONTRAST	S_CONTRAST	STUD_CNT	WEIGHT
1	6956.34	230	-0.1975	0.0750	0.6334	0.0605	-0.8309	0.1170	-7.1001	-0.1975
2	1409.97	107	0.6334	0.1125	-0.1975	0.0518	0.8309	0.1170	7.1001	0.6334

Table 17 - Weights and statistics for minerals associated with dry hole wells from PCA WofE analysis.

CLASS	AREA_SQ_KM	NO_POINTS	W+	S_W+	W-	S_W-	CONTRAST	S_CONTRAST	STUD_CNT	WEIGHT
1	3582.98	141	-0.3147	0.0842	0.1842	0.0568	-0.4989	0.1016	-4.9113	-0.3147
2	4783.33	310	0.1842	0.0568	-0.3147	0.0842	0.4989	0.1016	4.9113	0.1842

6.2.2 Weight of evidence for band ratio

Weights were calculated to test band ratio spatial association with oil wells. The resulted statistical measures are shown in Table 18.The results show that there are few classes of minerals that positively correlate with oil wells. Bleached areas (class 7) show positive W+ and C, the Studentized contrast are markedly greater than the defined threshold (0.842) indicating that the spatial association are more likely to be real. Mineral classes 8 and 10 show positive contrast as well. However, positive contrast here was the result of few oil wells in a small area. These were extracted and combined into a binary evidence layer and statistics were re-calculated and shown in Table 20. Class 2 refers to the selected minerals that correlate with oil wells; everything else was grouped into class 1 including classes with positive contrast with one or two points which considered uncertainties due to the data.

Table 19 measures the spatial correlation between dry holes and the different minerals anomalies resulted from band ratio analysis. Dry hole wells, as explained by C value, have strong spatial correlation with classes 2, 3, 7 and 8. The corresponding Studentized contrast are greater than the specified threshold of 0.842 (approximately 80%) indicating the contrasts are more certain for those minerals. Similar to PCA analysis statistics, the band ratio analysis also shows that bleached areas correlate with both oil and dry hole wells. However, as per the defined conceptual model, bleached is correlated with oil therefore it was discarded for dry hole wells. Correlated minerals were extracted, except bleached, and combined into a binary evidence layer. Class 2 was assigned to the minerals that correlate with dry holes and everything else was grouped into class 1 including classes with positive contrast with one or two points. Statistics were re-calculated and shown in Table 21.

Table 18 -Weights and statistics for band ratio minerals evidential theme with oil wells

CLASS	AREA_SQ_KM	NO_POINTS	W+	S_W+	W-	S_W-	CONTRAST	S_CONTRAST	STUD_CNT	WEIGHT
0	5835.99	234	-0.0035	0.0654	0.0080	0.0985	-0.0115	0.1182	-0.0977	0.1677
1	747.26	16	-0.6309	0.2500	0.0448	0.0558	-0.6757	0.2562	-2.6378	-0.6309
2	138.13	3	-0.6166	0.5774	0.0077	0.0547	-0.6243	0.5799	-1.0765	-0.6166
3	38.60	1	-0.4403	1.0000	0.0016	0.0546	-0.4419	1.0015	-0.4412	0.1677
4	56.57	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1677
5	79.97	3	-0.0701	0.5774	0.0007	0.0547	-0.0707	0.5799	-0.1220	0.1677
6	13.18	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1677
7	169.70	46	1.9078	0.1475	-0.1263	0.0586	2.0341	0.1587	12.8183	1.9078
8	28.28	3	0.9696	0.5774	-0.0056	0.0547	0.9751	0.5800	1.6813	0.9696
9	9.14	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1677
10	2.59	5	3.8720	0.4476	-0.0146	0.0549	3.8867	0.4510	8.6187	3.8720
11	1.78	2	3.3319	0.7075	-0.0057	0.0546	3.3376	0.7096	4.7037	3.3319
12	971.72	23	-0.5306	0.2085	0.0526	0.0564	-0.5833	0.2160	-2.7000	-0.5306
13	188.43	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1677
14	63.48	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1677
15	3.28	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1677
16	4.17	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1677
17	18.31	1	0.3053	1.0000	-0.0008	0.0546	0.3061	1.0015	0.3056	0.1677
18	3.54	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1677
19	0.62	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1677

Table 19 - Weights and statistics for band ratio minerals evidential theme with dry holes wells

CLASS	AREA_SQ_KM	NO_POINTS	W+	S_W+	W-	S_W-	CONTRAST	S_CONTRAST	STUD_CNT	WEIGHT
0	5835.99	319	0.0150	0.0560	-0.0353	0.0870	0.0503	0.1035	0.4856	0.0257
1	747.26	37	-0.0839	0.1644	0.0079	0.0491	-0.0918	0.1716	-0.5348	0.0257
2	138.13	15	0.7015	0.2582	-0.0172	0.0479	0.7187	0.2626	2.7366	0.7015
3	38.60	5	0.8779	0.4472	-0.0065	0.0474	0.8844	0.4497	1.9665	0.8779
4	56.57	3	-0.0153	0.5774	0.0001	0.0472	-0.0154	0.5793	-0.0266	0.0257
5	79.97	3	-0.3615	0.5774	0.0029	0.0472	-0.3644	0.5793	-0.6291	0.0257
6	13.18	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0257
7	169.70	19	0.7320	0.2294	-0.0226	0.0481	0.7546	0.2344	3.2190	0.7320
8	28.28	6	1.3714	0.4083	-0.0100	0.0474	1.3814	0.4110	3.3608	1.3714
9	9.14	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0257
10	2.59	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0257
11	1.78	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0257
12	971.72	33	-0.4610	0.1741	0.0473	0.0489	-0.5083	0.1808	-2.8113	-0.4610
13	188.43	5	-0.7077	0.4472	0.0116	0.0474	-0.7193	0.4497	-1.5995	-0.7077
14	63.48	3	-0.1306	0.5774	0.0009	0.0472	-0.1315	0.5793	-0.2270	0.0257
15	3.28	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0257
16	4.17	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0257
17	18.31	1	0.0139	1.0000	0.0000	0.0471	0.0139	1.0011	0.0139	0.0257
18	3.54	2	2.3509	0.7073	-0.0040	0.0472	2.3549	0.7089	3.3221	2.3509
19	0.62	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0257

Table 20 -Weights and statistics for minerals associated with oil wells from band ratio WofE analysis.

CLASS	AREA_SQ_KM	NO_POINTS	W+	S_W+	W-	S_W-	CONTRAST	S_CONTRAST	STUD_CNT	WEIGHT
1	8174.76	283	-0.1504	0.0594	1.9010	0.1361	-2.0514	0.1485	-13.8127	-0.1504
2	200.57	54	1.9010	0.1361	-0.1504	0.0594	2.0514	0.1485	13.8127	1.9010

Table 21 - Weights and statistics for minerals associated with dry hole wells from band ratio WofE analysis.

CLASS	AREA_SQ_KM	NO_POINTS	W+	S_W+	W-	S_W-	CONTRAST	S_CONTRAST	STUD_CNT	WEIGHT
1	8170.33	425	-0.0346	0.0485	0.8567	0.1961	-0.8913	0.2020	-4.4116	-0.0346
2	205.01	26	0.8567	0.1961	-0.0346	0.0485	0.8913	0.2020	4.4116	0.8567

Because the number of dry hole wells that fall into the area covered by the theme were relatively small compared to the total dry holes wells (26 out of 451), another model was tested. The new model added the background class to the classes used previously for dry holes theme. Statistics were re-calculated and shown in Table 22.

Table 22 - Weights and statistics for minerals associated with dry hole wells from BR WofE analysis 2nd model.

CLASS	AREA_SQ_KM	NO_POINTS	W+	S_W+	W-	S_W-	CONTRAST	S_CONTRAST	STUD_CNT	WEIGHT
1	2334.34	106	-0.1705	0.0971	0.0588	0.0538	-0.2293	0.1111	-2.0645	-0.1705
2	6041.00	345	0.0588	0.0538	-0.1705	0.0971	0.2293	0.1111	2.0645	0.0588

The results were combined to produce relative favorability ranking map. The study area was ranked on a relative scale representing high to low potential microseepage area: favorable, permissive and not permissive. Results are shown in Figure 14 and Figure 15.
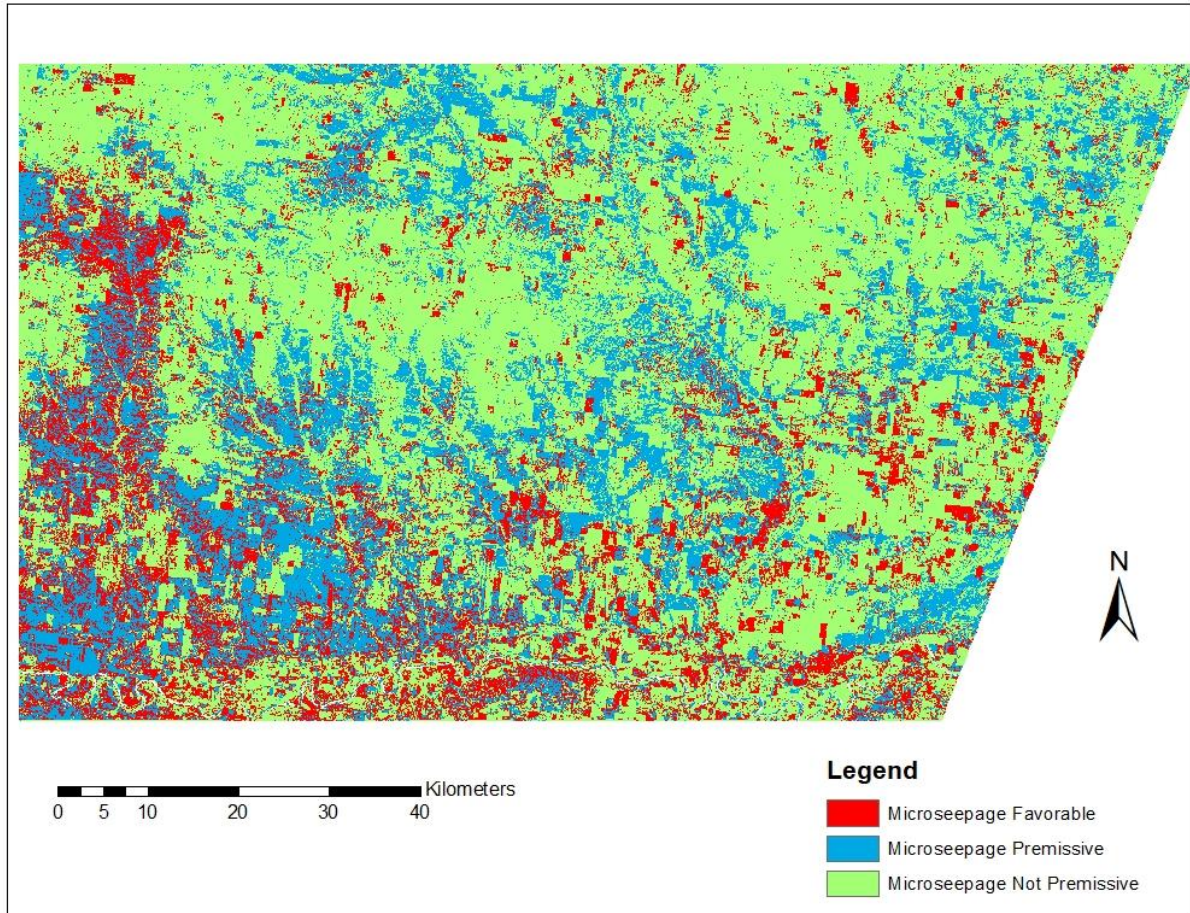
Figure 13 - Relative favourability ranking map of microseepage derived from WofE analysis on PCA images
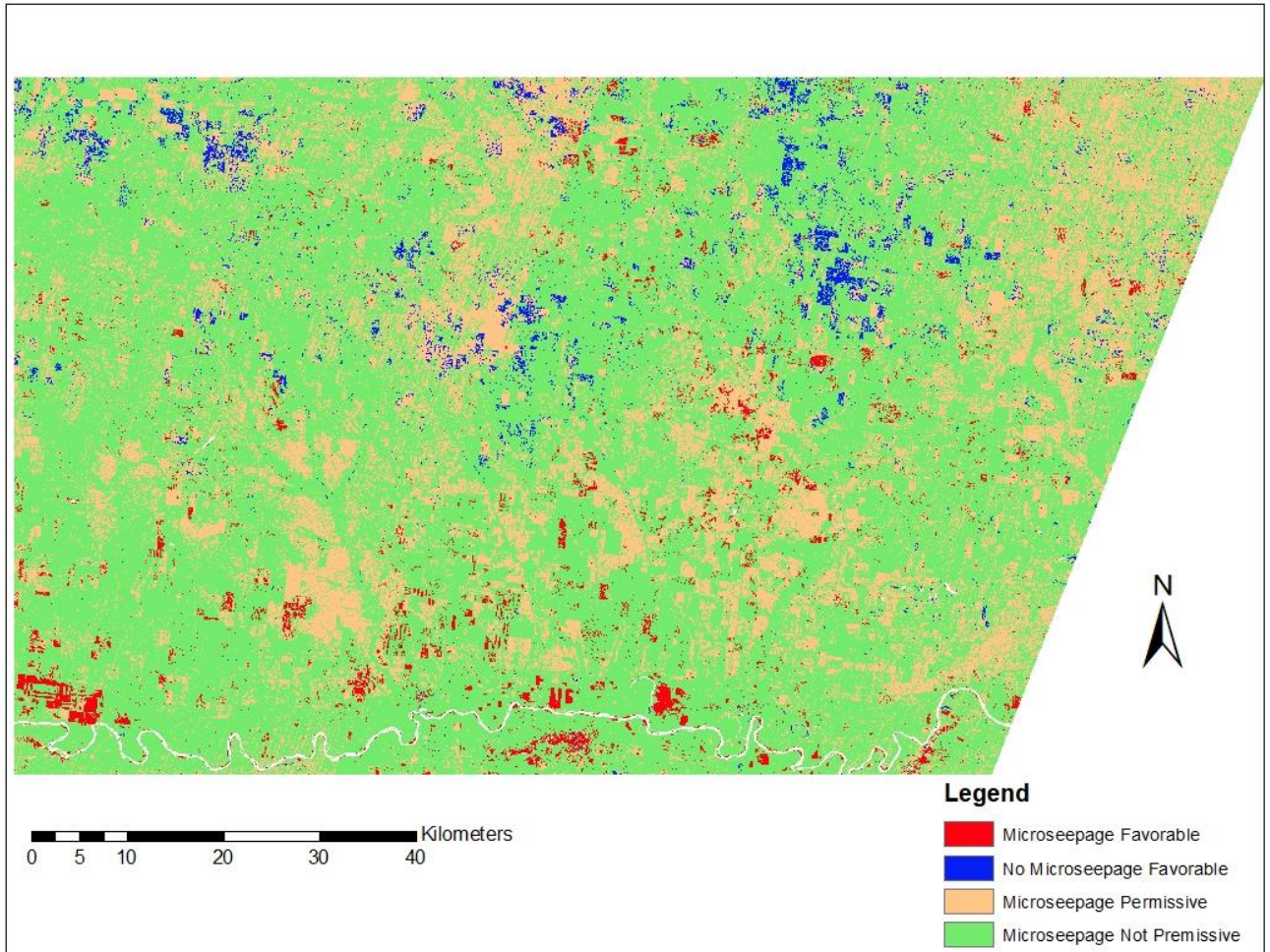
Figure 14 - Relative favourability ranking map of microseepage derived from WofE analysis on band ratio images

The reclassified evidence maps resulted from the combined knowledge based model for PCA and band ratio models is shown in Figure 15.
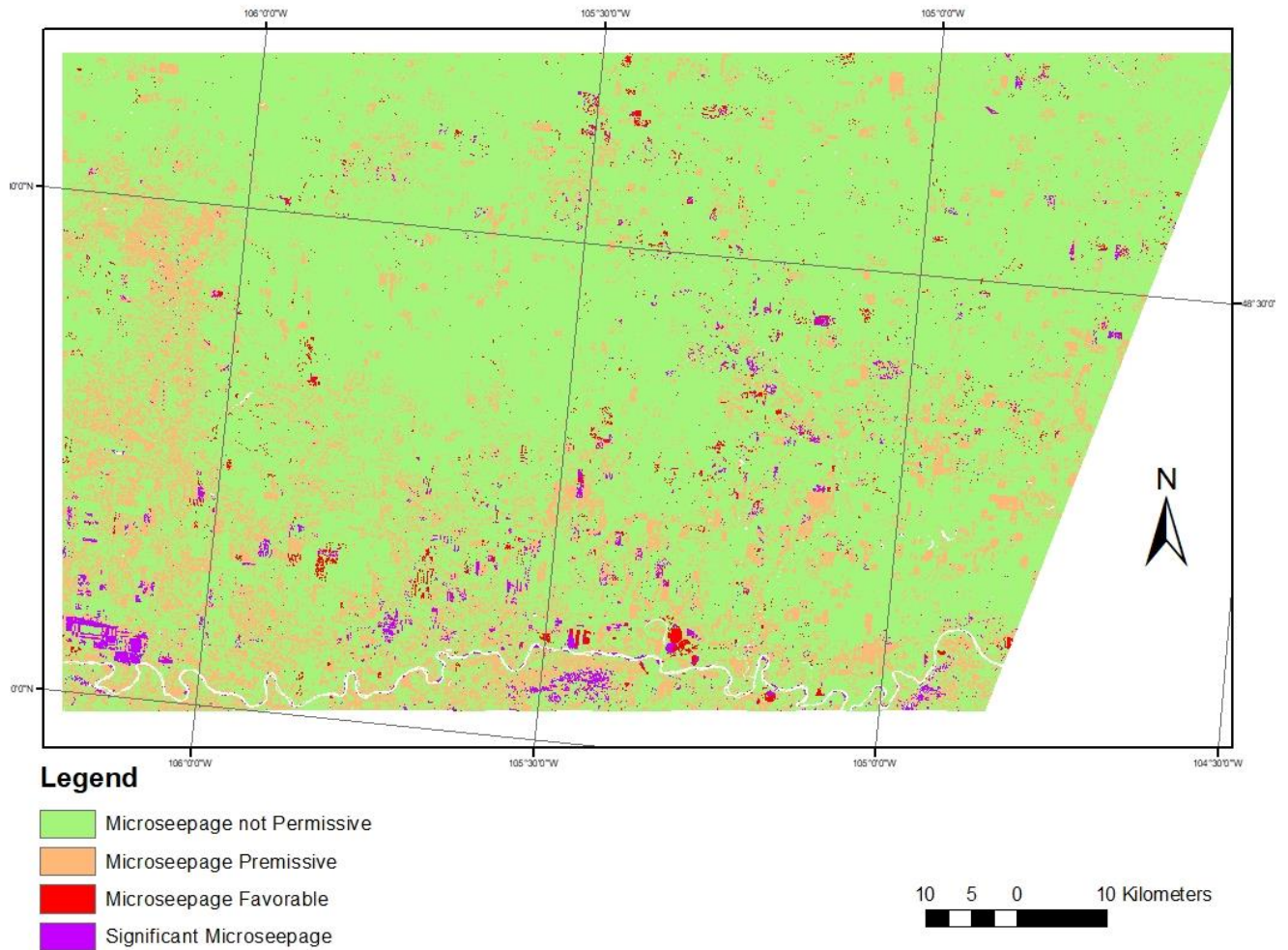
Figure 15 - Relative favourability ranking map of microseepage for the combined evidence layers of PCA and BR.

6.3 Mixture Tuned Matched Filtering (MTMF)

Figure 16 show the results of MTMF classification for ferrous iron, clays and ferric iron shown respectively in RGB for MTMF bands 18, 15 and 19. These bands have been selected because they were better delineating the chosen minerals and have less noise or vegetation interference.
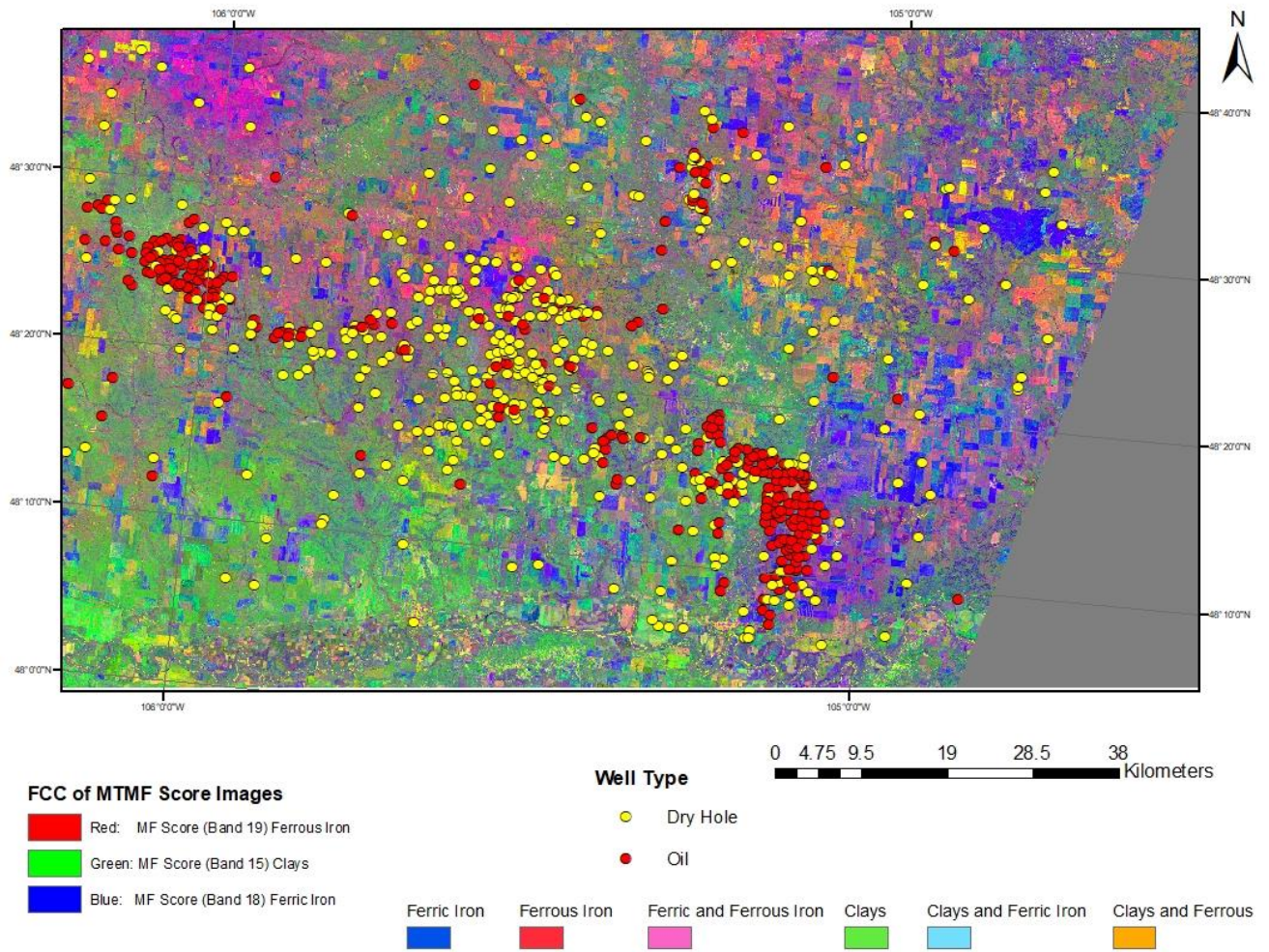
Figure 16 -MTMF image in RGB colour with wells distributions - ferrous iron (band 19) - clays area (band 15) – ferric iron area (band 18)
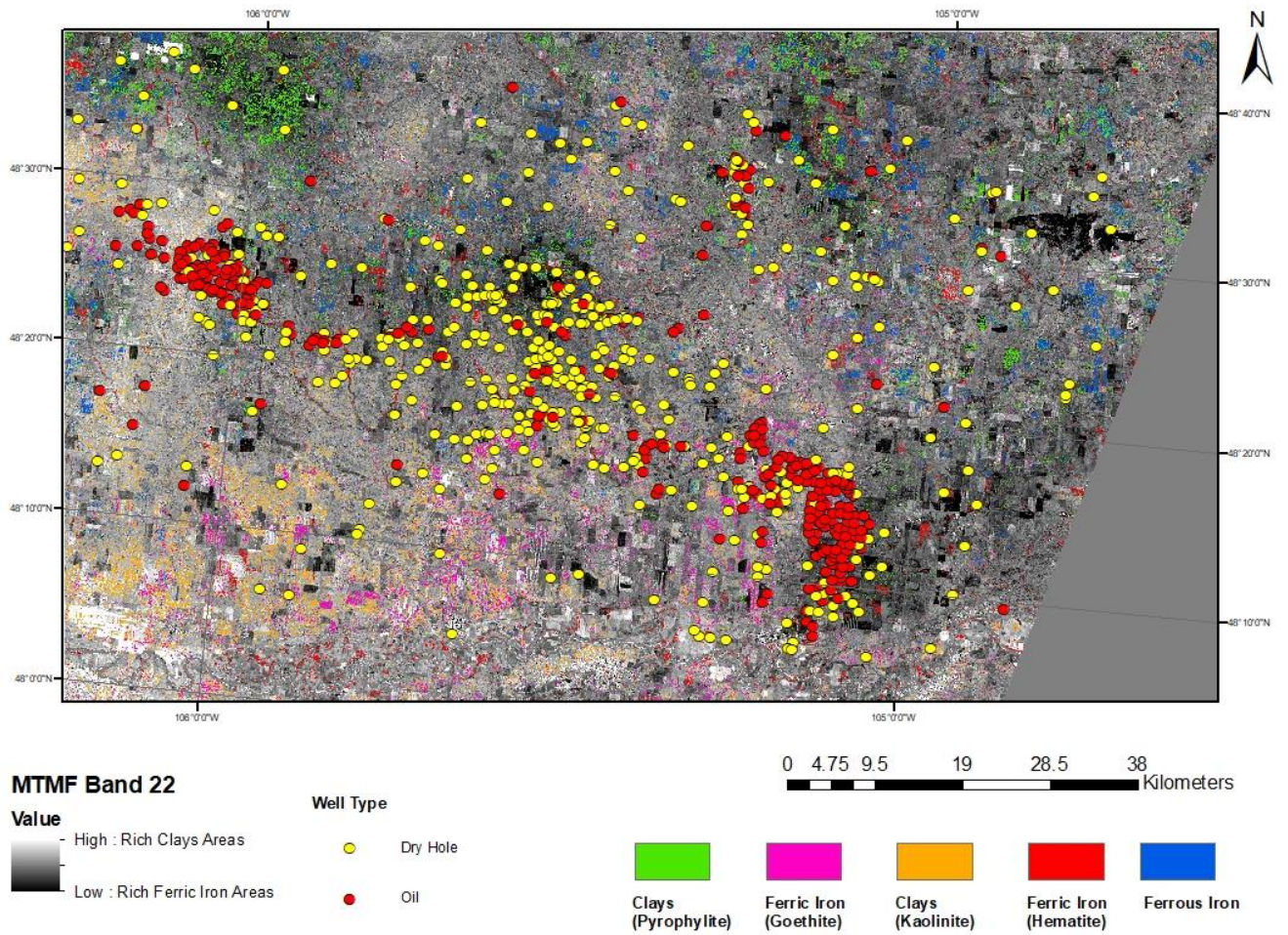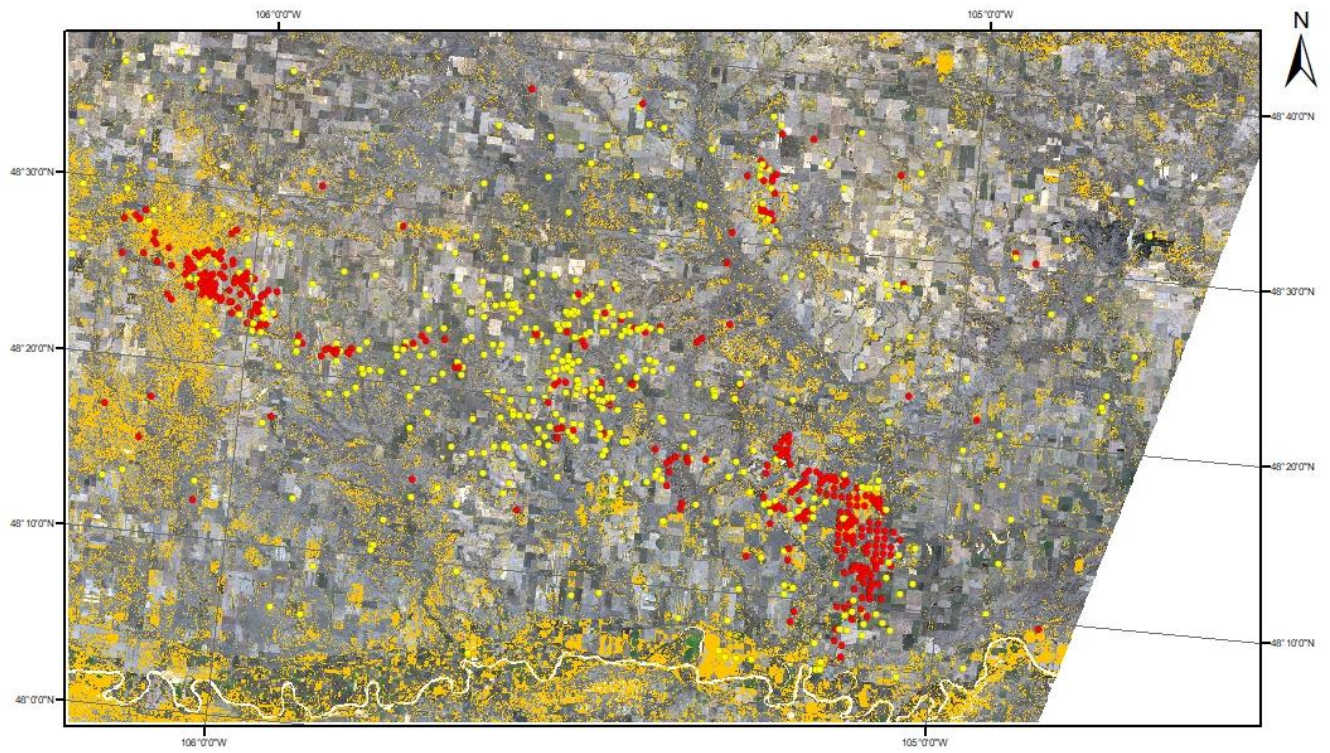
Figure 17 - MTMF result of band 22 delineate clays in bright pixels and ferric iron in dark pixels

6.4 Data-driven model

The results that were obtained from SVM classification are shown in Figure 18 and Figure 19.

Figure 18 - Results from SVM classification with wells distributions over Fort Peck Reservation
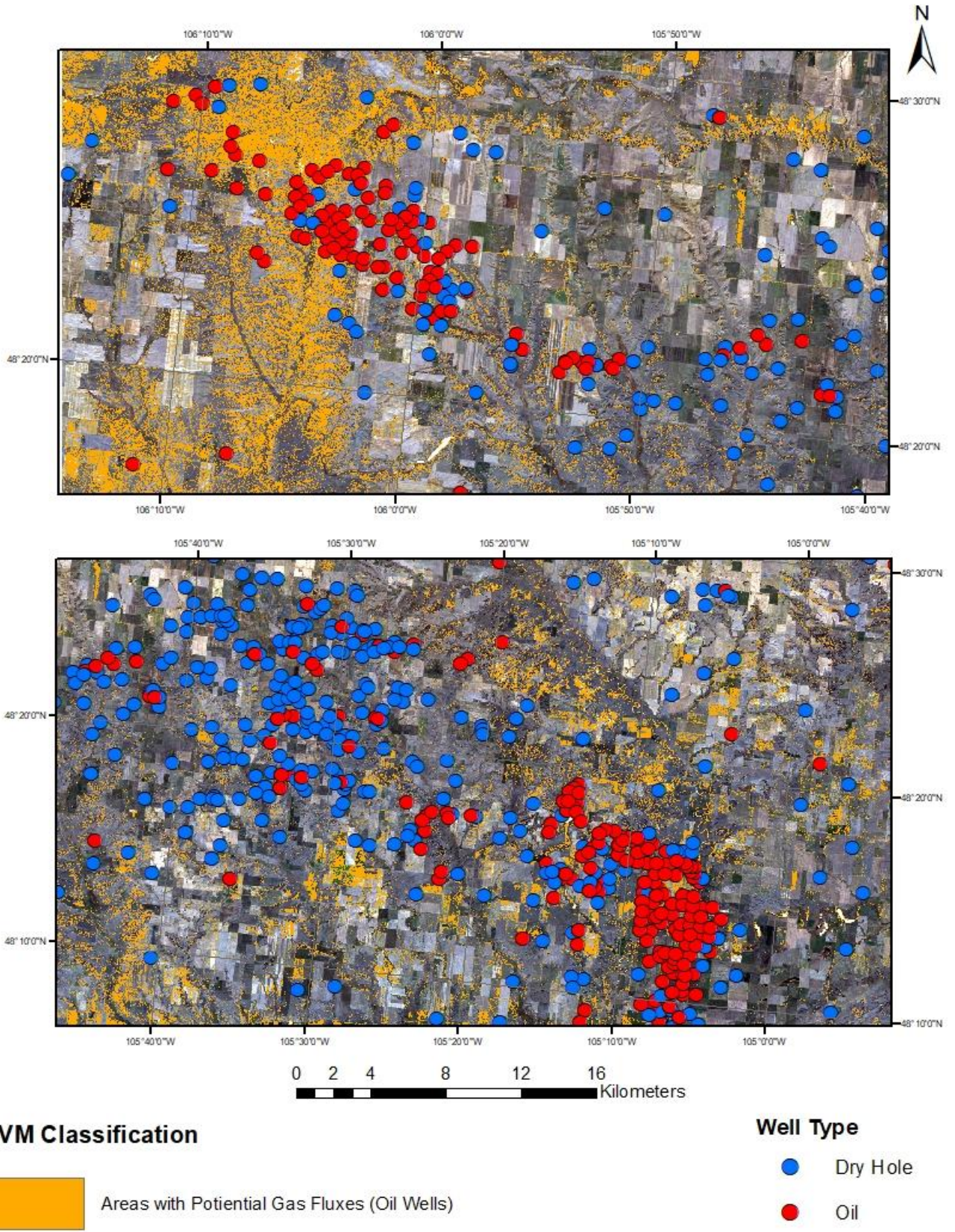
Figure 19 - A close look of SVM classification results with wells distribution over Fort Peck Reservation

6.5   Models evaluation

The models classification accuracy performance were evaluated using area under the curve (AUC) of a receiver operating characteristic (ROC) plot. ROC curve was constructed twice for each model, one time with oil wells and one time with dry hole wells. A good model should have high value (AUC > 0.5) with oil wells and low value (AUC < 0.5) with dry hole wells.

Figure 20 - ROC curve shows the performance of PCA model for hydrocarbon microseepage. On the top, ROC curve testing the association between oil wells and PCA evidential theme. Below ROC curve testing the association between dry hole wells and PCA evidential theme
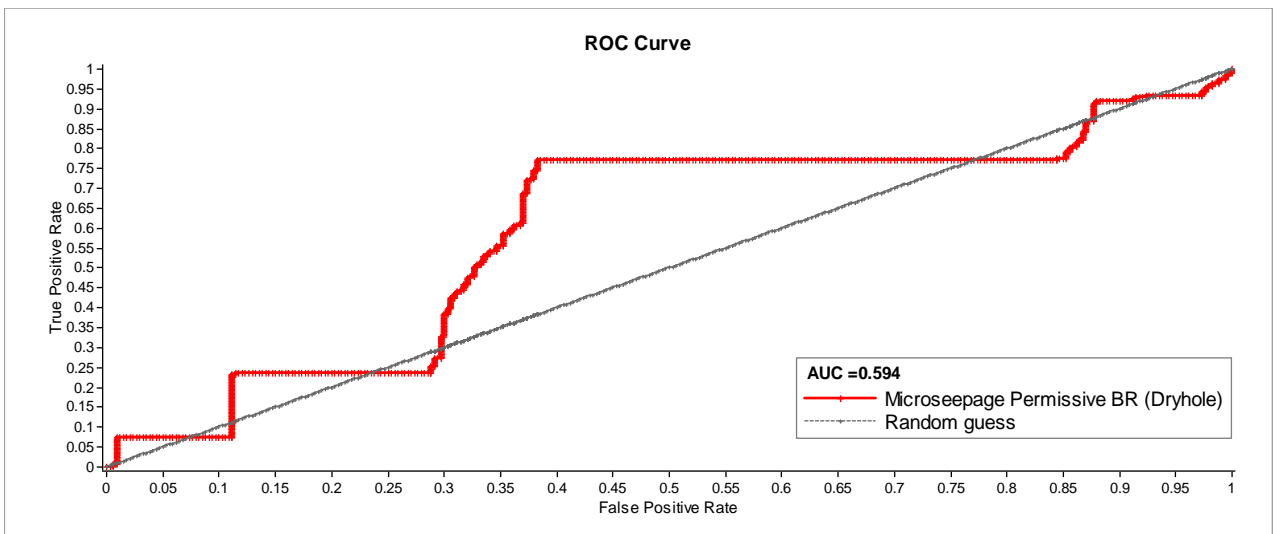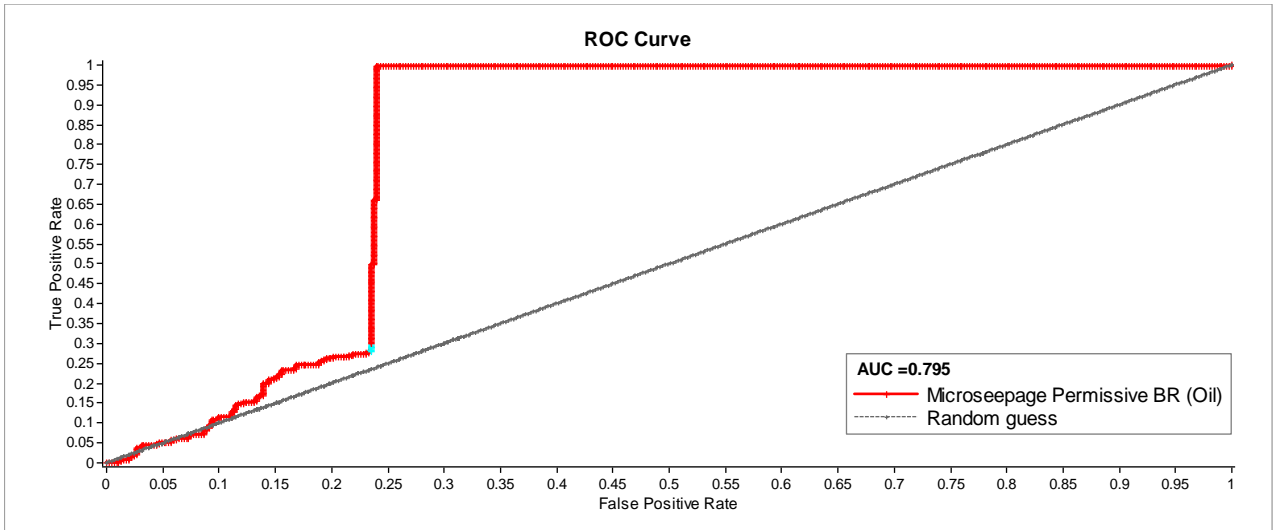
Figure 21 - ROC curve shows the performance of band ratio model for hydrocarbon microseepage. On the top, ROC curve testing the association between oil wells and band ratio evidential theme. Below ROC curve testing the association between dry hole wells and band ratio evidential theme

Figure 22 - ROC curve shows the performance of the combined knowledge based model for hydrocarbon microseepage. On the top, ROC curve testing the association between oil wells and knowledge based evidential theme. Below ROC curve testing the association between dry hole wells and knowledge based evidential theme
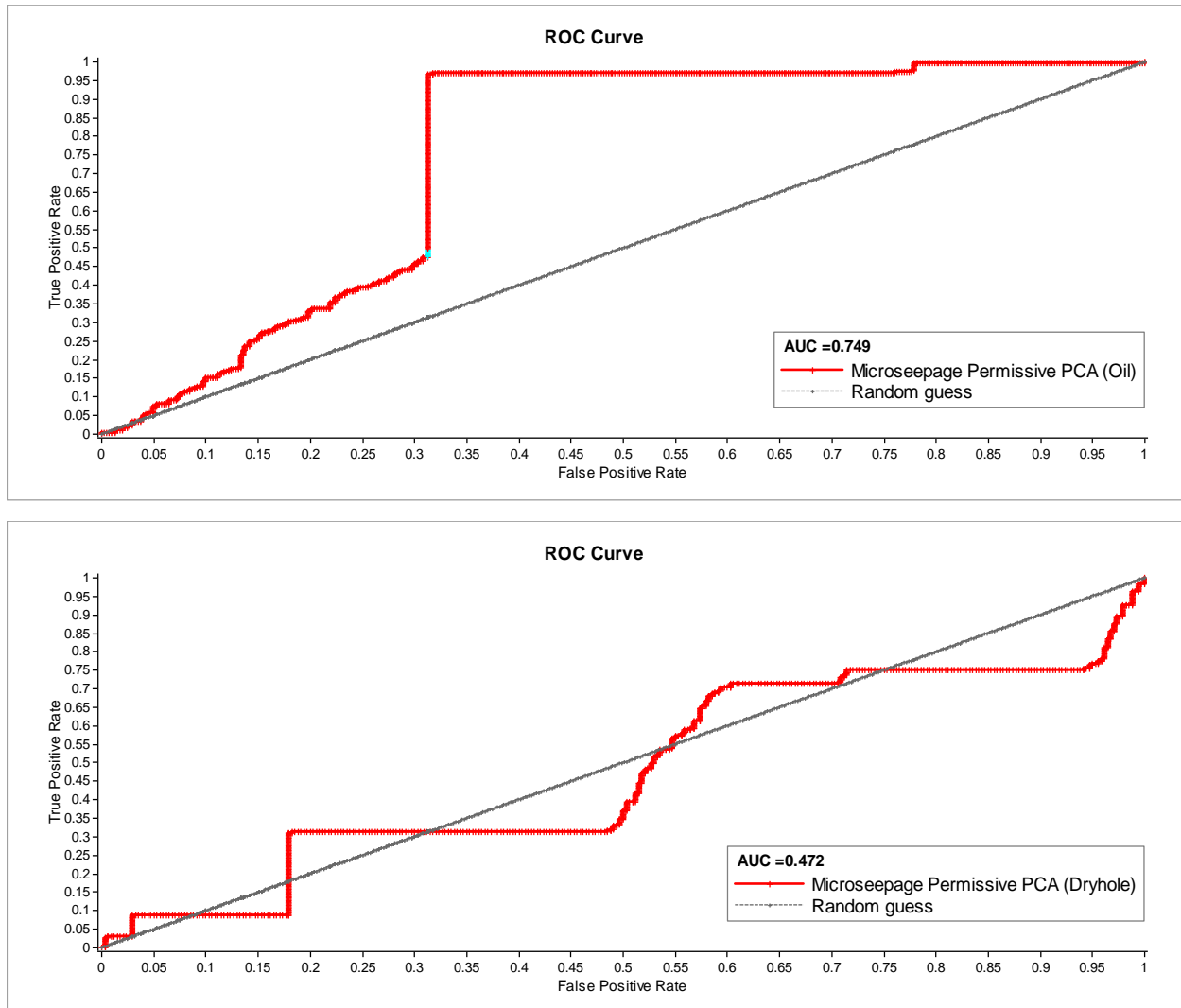
Figure 23 - ROC curve shows the performance of the SVM model for hydrocarbon microseepage. On the top, ROC curve testing the association between oil wells and SVM evidential theme. Below ROC curve testing the association between dry hole wells and SVM evidential theme
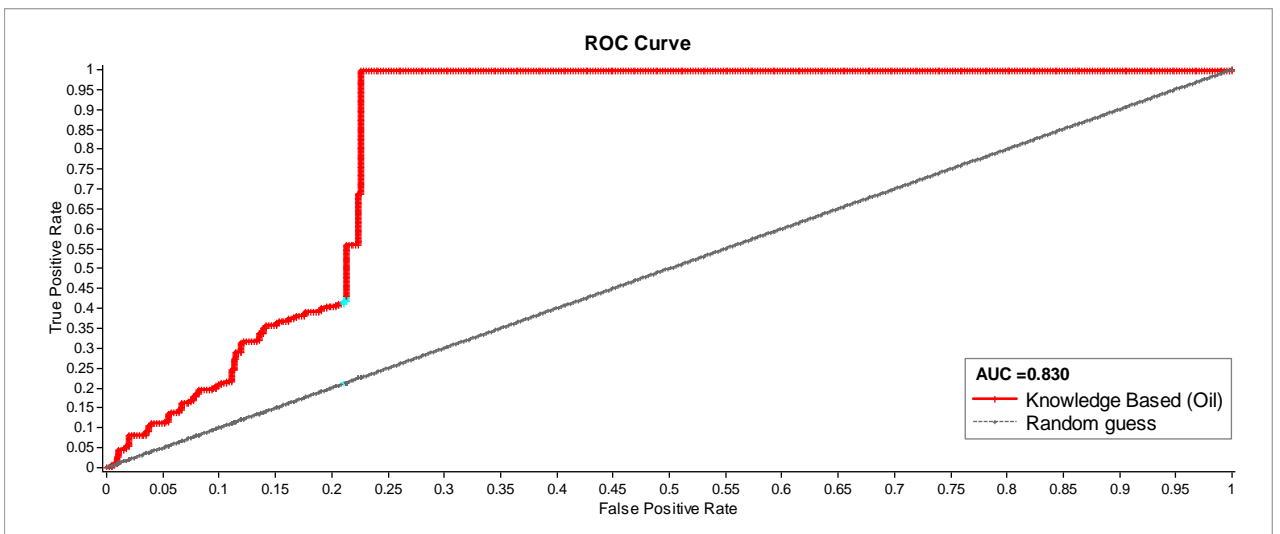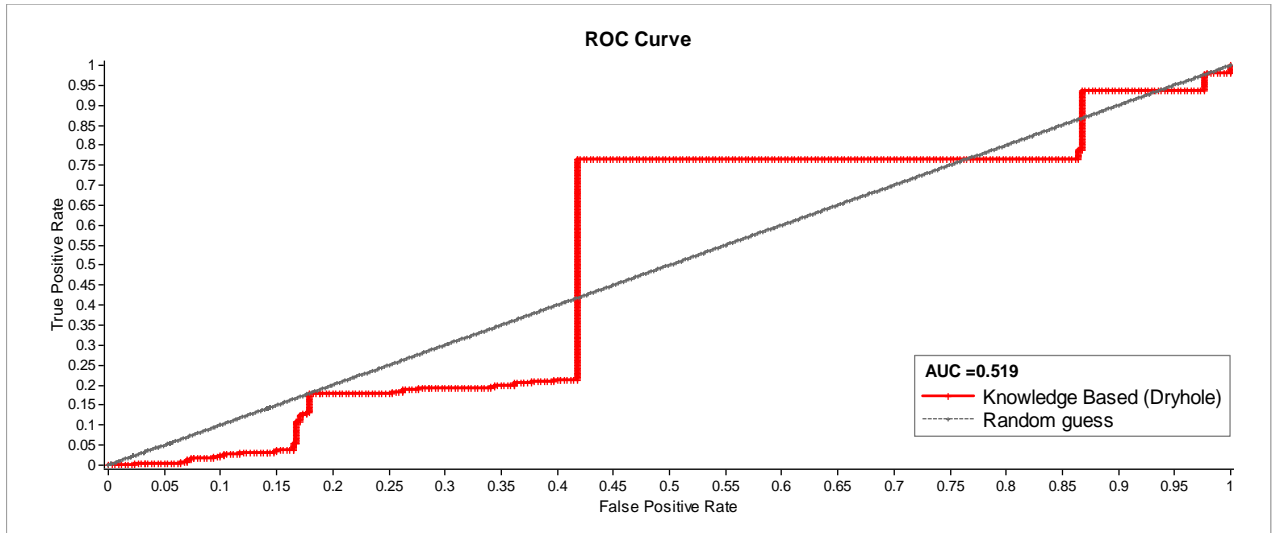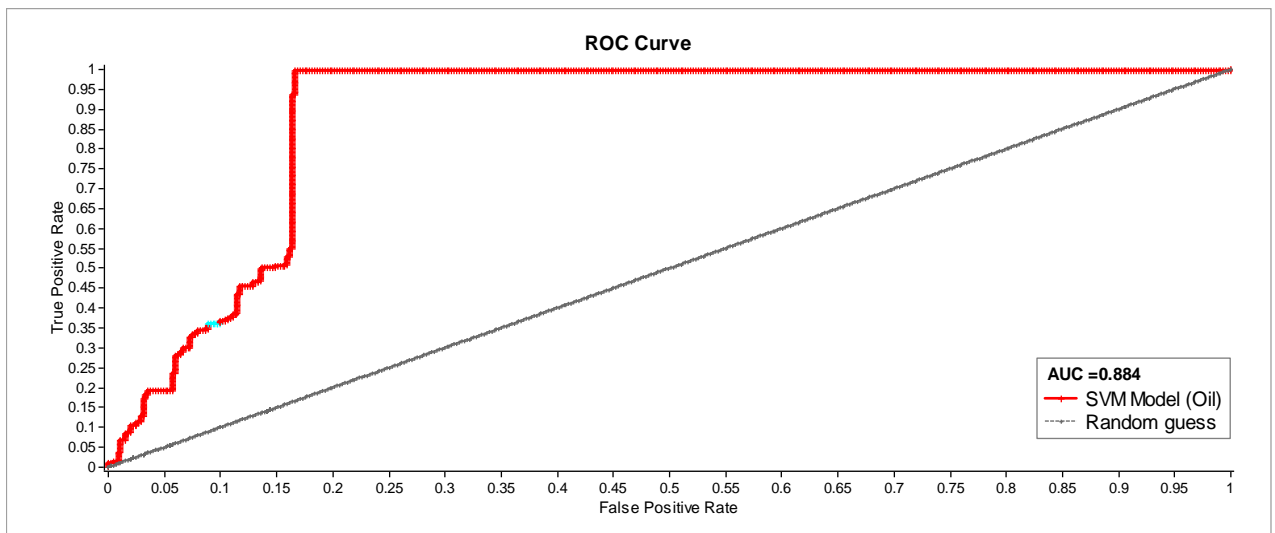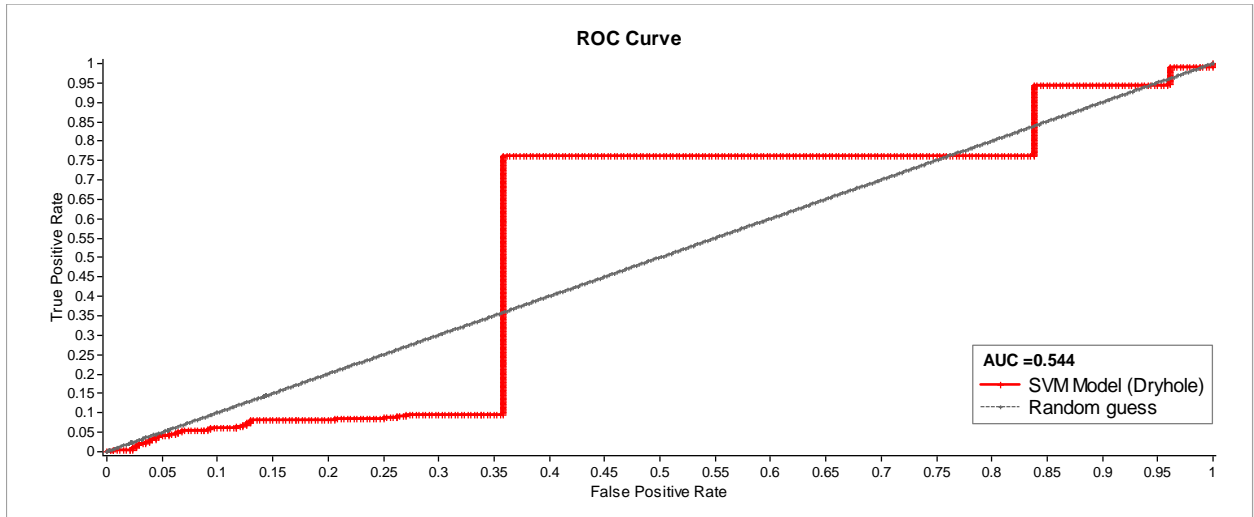
7 Discussion

Satellite remote sensing methods were used to investigation areas affected by natural hydrocarbon microseepage using knowledge and data driven based approaches. Knowledge based approaches constructed based on the theoretical model established from literature targeting specific minerals and surface manifestations related to hydrocarbon microseepage. A series of multiple image processing techniques were developed based on the spectral characteristics of the minerals. These were derived from two main methods PCA with crosta technique and band ratio technique.

7.1 Knowledge based models

The PCA and band ratio minerals evidential themes were tested with weight of evidence to quantify their spatial association with hydrocarbon accumulations. Presence and absence was determined based on the correlations with oil and dry hole wells. Statistics obtained from weight of evidence analysis (Table 14, Table 15 for PCA) and (Table 18 and Table 19 for band ratio) show both models agrees that oil wells correlated with bleached, carbonate and clays areas while dry hole wells are more correlated with ferric iron areas. This is to some extent agrees with the established theoretical detection model, the areas affected by hydrocarbon microseepage are lacking ferric iron and have proximity to bleached areas. However, bleached areas also shows significant correlation with dry hole wells. In order to follow the conceptual models, the bleached areas evidential theme was included in the model for microseepage and arbitrary excluded from non-microseepage model for both PCA and band ratio.

The Area under the curve (AUC) was used to evaluate PCA and band ratio models shown in Figure 14 and Figure 15. PCA resulted in AUC mean value of 0.749 using ROIs of oil wells and AUC mean value of 0.472 using ROIs of dry hole wells (Figure 20). On the other hand, band ratio model resulted in AUC mean value of 0.795 using ROIs of oil wells and AUC mean value of 0.594 using ROIs of dry hole wells (Figure 21). For the microseepage area, a good model should have high AUC using ROIs of oil wells and low AUC value below 0.5 with dry hole wells ROIs. The predicted power of both model are fair, both have mean value around 0.7 with oil wells and around/below 0.5 for dry hole.

Both PCA and band ratio models were combined into one model so-called knowledge based model (Figure 15). This model has an improved classification performance. It has resulted in AUC mean value of 0.830 using ROIs of oil wells and AUC mean value of 0.519 using ROIs of dry hole wells (Figure 22).

7.2 Data-driven model

The model created using data driven approach (Figure 18) was purely estimated from the multispectral image data based on occurrence and abundance of oil and dry hole wells and without any prior knowledge of the area minerals assemblages or other surface manifestations. The model yields best results. The AUC mean value is 0.884 using ROIs of oil wells and AUC mean value of 0.544 using ROIs of dry hole wells (Figure 23).

The main challenge with data driven approach and SVM classification in particular is parameter turning (also called hyper-parameters) which are internal coefficients or weights that allow the analyst/scientist to tailor the behaviour of the algorithm to a specific dataset. It was challenging to know what values to use for the hyper-parameters on the given dataset. Multiple values and different combination were tested for kernel types, gamma and penalty parameter in SVM. The highest attained classification performance on the test dataset was 77% of accuracy (Table 9). The strategies on selecting the parameters values were more of random search, tuning by means of such technique was time consuming since with every new parameters combination the classification has to restart. The more hyper-parameters need to be tuned and the larger the search space, the slower the tuning process. The performance achieved with SVM classification was the highest when compared to the other methods. However, performance is highly dependent on the selected parameters. This implies that performance could be enhanced further. Hyper-parameters tuning by means of systematic techniques can bring down the time spent to develop the optimal set of parameters and bring better classification performance.

7.3   Surface manifestations of hydrocarbon microseepage

In terms of visual interpretations, the knowledge based methods (Figure 11 and Figure 12) have good agreements of the location of ferric iron and clays minerals. In fact, there is a consistency between all models in regards to the areas of clays. The mapped clays areas in both PCA and band ratio models are identical with the areas that were mapped in MTMF classification as rich clays areas which was revealed as kaolinite mineral based on the spectral shape of the image endmembers (Figure 17). The same extent area was mapped with SVM classification as area with potential gas fluxes (Figure 19).

Ferric iron rich areas correlated with dry hole locations. Dry hole wells were found to cluster inside and around ferric iron rich area. Clay mineral alteration, particularly kaolinite mixed layer at the edge of rich ferric iron area, was the strongest evidence of hydrocarbon microseepage. Results in MTMF classification (Figure 17) indicate that pyrophyllite clay type is abundant in non-oil regions and correlate more with areas with dry hole wells.

Ferrous iron rich areas in Figure 11, Figure 28 and Figure 17 were confusing, and no clear pattern could be identified either for oil or dry hole wells. Statistically, weight of evidence calculations (Table 15) on PCA evidential theme shows relatively few dry hole wells have positive spatial associations with ferrous iron rich areas. The conceptual model obtained from literature for the recognition of the locations of microseepage suggest that that microseepage-induced sediments will have higher ferrous content and lesser ferric iron content. Therefore ferrous iron was tested as an indicator of microseepage alterations.

However, in wide body of literature the mineralogical and geochemical properties of the sediments affected by hydrocarbon microseepage have defined explanations with numerous and sometimes conflicting hypotheses. For instance, the reducing environments and bleaching of red beds have been widely used as indicators of hydrocarbon microseepage induced alterations. Both conceptually explain the same geochemical process, however in literature the terms sometimes used improperly which in turn creates some confusion. The terms that

normally appear are: abundance of ferrous iron, reduction of ferric iron or red bed bleaching. Reduction of ferric iron to ferrous iron as result of microseepage will cause red bed bleaching in rocks and sediments. The affected zone becomes deficient of ferric iron minerals but not necessarily become enriched of ferrous iron mineral. This is because the developed ferrous iron minerals may precipitate or mobilize to other places by groundwater (Garain et al., 2020). This brings to attention that magnetic anomalies which measures the ability of soil minerals to be magnetized because of the abundance of ferrous iron might not have any merit toward hydrocarbons potential area. Results of previous studies (Monson, 2000, 2001, 2003) conducted in Fort Peck Reservations show that interpretation of high magnetic susceptibility prove poor gas fluxes. On the other hand, (Wang, 2016) found that ferrous index was not a good indicator for mapping hydrocarbon alteration. Indeed, the result here concluded the same findings. Thus, microseepage alteration should be referred to as reduction of ferric iron and bleaching of red beds but not as ferrous iron abundance.

The results of band ratio and Crosta technique show that the red bed bleaching occurs within areas with no ferric iron and found to be correlated with oil wells. However, pixels containing vegetation can show anomalous value which means that the result might not show the true characteristic of the bleached red beds. Results from weight of evidence calculations (Table 14, Table 15, Table 18, Table 19 demonstrate that bleached areas have significant association with oil wells but also correlate with dry holes wells as well in both model PCA and band ratio. This indicated that bleached areas might not always be hydrocarbon related, but could also be associated with the presence of waterlogging, cropping of annual vegetation, leaching of nutrients or as a result of weathering processes. In the study, bleached area was arbitrarily excluded from non microseepage potential area (i.e., areas associated with dry hole wells) to match the conceptual model which might have affected the result. Therefore, the model developed could be biased in both PCA and band ratio.

7.4 Current models and challenges

Despite the pre-processing steps that were preformed to atmospherically correct the image and masking out the clouds and water bodies, random noise signals were contained in knowledge-based approaches. The influence of random noise was obvious when using Crosta technique. The extent of minerals has larger distribution in Crosta compared to other techniques. The concept behind Crosta technique is to reduce the number of input bands for PCA, normally to four bands in multispectral image, to facilitate the manifestation of certain minerals. However, the interpretation of eigenvector matrix, carried based on the diagnostic reflectance and absorption of two bands only while the other two bands get usually ignored in the analysis. For example, when using the bands (2, 5, 6, and 7) to map clays, the mapped area will be defined by the sign and magnitude of eigenvectors and loading for band 6 and band 7. This was the case in almost all the reviewed literatures. Though clays also have reflectance in band 5 and 2, these two bands should follow the sign of band 6 in order to get accurate mapping of clays. Otherwise, noise is enhanced from these two bands and creates interference-induced false anomalies. This is a key part and major deficiency of Crosta technique that appears to be ignored by many researchers. Limitation could also arise from uncorrelated noise in principle component (PC) images. The diagnostic reflectance and absorption of minerals usually

contained in the noisiest PCs, the third or fourth PC, which in turn creates random noise and enhances the false anomalies.

Band ratio technique can overcome the deficiency in Crosta technique by allowing independent control over band selection to enhance the spectral contrast of specific absorption features between multiple bands. Nevertheless, different bands ratio combination may result in error values of duplicate pixel mapping. Inconsistency arise from mapping one pixel as two different minerals. Band ratio despite its flexibility in band selection, has a major drawback in that it cannot account for endmember variability. Therefore it cannot handle overlapping and mixing absorption features. Boolean algebra was modelled to combine different mineral indices to generate uniform map of minerals.

MTMF classification was used to identify individual minerals however results were restrictive. The extent of the mapped minerals was the least among other used knowledge based techniques. Classification result of MTMF (Figure 17) was consistent with the other methods. The MTMF needs complex hand crafted spectral feature, the pure image endmembers were obtained after long processing using the MNF transformation, pixel purity index and n-D visualization procedures.

To eliminate some of the interference of random noise and to extract the altered minerals, a threshold value was applied in all the images produced with knowledge based approach. The question that arises is what a good threshold values is. There is no theoretical base in determining the optimal threshold value. In this study value greater than $\mu + \sigma$ was used to extract the distribution of minerals, where μ represent mean and σ represent standard deviation. In normal distribution, about 68 % of data lies within $\mu \pm \sigma$, above that were considered anomalous. Threshold was chosen through a trial and error approach by testing that the chosen threshold is able to minimize the scattered results and can highlight contiguous pixels. Therefore, the selected threshold for separating the concentration mineral anomalies from background might be subjective and not the optimal threshold. Benefits and costs for setting the threshold should be taken into consideration.

Class imbalance is when the classes are not represented equally. It can provide poor predictive performance and misleading classification accuracy. Standard classification algorithms have a bias towards majority class. Therefore there is a high probability of misclassification of the minority class. In this study, the distribution of classes was slightly biased towards dry holes wells. The number of classes used for dry holes is 451 whereas number of classes for oil wells is 340. This technically imbalanced dataset therefore results might have little skewed. However, slight class imbalance is often not a concern and can be treated as a normal distribution.

Balanced class distribution could be generated by down sampling the majority class, however this was not considered here in this study due to the following facts. First, oil wells or area of microseepage are naturally smaller than the area of no microseepage or dry holes wells. In contrast the imbalanced classes here might have represented actual imbalance population. Thus, the data was reflecting the correct ratio of distribution. Second, randomly eliminating majority class might potentially discard useful information and in contrast lead to information

loss. Finally, the random chosen sample could be a biased sample which might not represent the accurate population, thereby resulting in inaccurate classification results.

The results from the methods used were spatially consistent throughout the study area. The methodologies used are suitable for the detection and visualization of overall spatial distribution of minerals surface anomalies. The knowledge based methods are sensitive to the data type and the pre-processing techniques because they rely on the spectral characteristics. Therefore, it is more vulnerable to the noise. For the same reason, this type of method is feasible to regional scale mineral mapping. Obtaining an accurate atmospheric correction is a critical prerequisite for the knowledge based methods. The data driven methods are more of scene dependent because they relies directly on the reference data. Therefore, it is more difficult to apply on the large scale mineral mapping. It demand more computing power and accuracy is impacted by the availability of the training data and the chosen parameters. Despite that challenge, the data driven methods are more robust and less sensitive to noise and yield the best results. Relationship between microseepage and the induced mineralogical assemblage has been wide controversial among wide literature which gives less credibility to the knowledge based approaches. The application of Landsat 8 in non-arid environment is capable of delineating the surface of different minerals groups. Nonetheless, it is limited in targeting specific mineral.

8 Conclusion

Knowledge based and data driven methods to Landsat 8 image characterization were developed and tested for delineating hydrocarbon microseepage surface manifestations. The results were satisfactory enough and comparable to each other indicating results were statistically significant. The data driven approach using SVM outperforms other methods. It has the highest AUC of 0.884, followed by the knowledge based approaches, weight of evidence (WofE) with combined knowledge-based (AUC of 0.830), WofE with band ratio (AUC of 0.795) and WofE with PCA (AUC of 0749).

Comparing knowledge-driven based methods, the random noise signals were contained in knowledge-based approaches. The influence of random noise was obvious when using PCA with Crosta technique. The extent of minerals has larger distribution in PCA with Crosta technique when compared to band ratio technique. Band ratio technique can overcome the deficiency in Crosta technique by allowing independent control over band selection to enhance the spectral contrast. Nevertheless, band ratio despite its flexibility in band selection, has a major drawback in that it cannot account for endmember variability. Therefore it cannot handle overlapping and mixing absorption features. Statistically, the predicted power of both PCA and BR models were fair, both had mean value around 0.7 for microseepage potential area and around/below 0.5 for no potential microseepage area.

The data-driven model yields best results, the gain in performance from using data-driven approach as compared to knowledge-based was relatively small. However, the long processing steps and time in the knowledge-based approaches gives merits to the data driven approach. The mapped clays and bleached areas in knowledge-driven models were identical with the areas that were mapped with data-driven approach as area with potential gas fluxes.

Mixture Tuned Matching Filtering (MTMF) was effective in decomposing the mixed pixels into a collection of pure pixels (endmembers) and determining their corresponding proportions (abundances). However, it has required complex hand crafted spectral feature that required to be manually labelled by the analyst. The pure image endmembers were obtained after long processing steps using various algorithms such as MNF transformation, pixel purity index and n-D visualization procedures.

Landsat 8 the Operational Land Imager (OLI) was used as data source of imagery for the study. However, it has some limitation in mineral exploration because of its spatial and spectral resolution. Based on the result obtained from the developed models, there are four fundamental issues to consider:

1)   Different data have their own advantage and challenges. High resolution and hyperspectral data can provide opportunity to maximize the accuracy of mineral mapping. Attention needs to be paid to the trade-off between obtaining high mineral mapping precision and the increase of the dimensionality and non-linearity in the data inhered from using hyperspectral data.

2)   Deficiencies in the methods also may arise from the random noise signals contained in the remote sensing data such as water, cloud, shadow and vegetation. Proper calibration and

pre-processing of data is important to minimize any interference noise to make sure the data yield the true diagnostic spectral shape. Atmospheric correction, noise reduction, masking out water bodies and clouds can help in eliminating some of false anomalies.

3)   Hydrocarbon microseepage is a complex problem. The uses of remote sensing images to investigate hydrocarbon microseepage related surface manifestation typically require careful knowledge of several different disciplines including geology and geochemistry. Considerable research and review is needed before we can understand the many factors affecting the formation of the surface anomalies and then make sense of the theoretical bases in regards to the distinction of geochemical behaviour, false anomalies characteristics and real anomaly manifestation in remote sensing images. Contradictory or conflicting background knowledge of hydrocarbon microseepage can pose a risk of producing inaccurate results. It is necessary to pay more cognitive attention and efforts to the different claims. The data-driven methods present an alternative approach as it can drive information from the data itself using statistical methods. Advancement of such methods nowadays such as machine learning and deep learning can function as help to redefine or adjust the previously held conceptual models. Yet, this approach strives for sufficient training or measured data. Availability of these data is an absolute necessity for having optimal classification results.

4) Disciplines such as geoscience and environmental science are based primarily on data that are often collaboratively generated and analysed. Hence, open data policies and practices expected to accelerate scientific research and increase its quality and quantity. The more data that are openly accessible the more success data driven science can generate.

9 References

AHMAD, F. 2012. Pixel purity index algorithm and n-dimensional visualization for ETM+ image analysis: A case of district vehari. *Global Journal of Human Social Science Arts & Humanities,* 12**,** 76-82.

ALLEK, K., BOUBAYA, D., BOUGUERN, A. & HAMOUDI, M. 2016. Spatial association analysis between hydrocarbon fields and sedimentary residual magnetic anomalies using Weights of Evidence: An example from the Triassic Province of Algeria. *Journal of Applied Geophysics,* 135**,** 100-110.

ALMEIDA-FILHO, R., MIRANDA, F., YAMAKAWA, T., BUENO, G. V., MOREIRA, F. R., CAMARGO, E. G. & BENTZ, C. 2002. Data integration for a geologic model of hydrocarbon microseepage areas in the Tonã Plateau region, North Tucano basin, Brazil. *Canadian journal of remote sensing,* 28**,** 96-107.

ALMEIDA-FILHO, R., MIRANDA, F. P. & YAMAKAWA, T. 1999. Remote detection of a tonal anomaly in an area of hydrocarbon microseepage, Tucano basin, north-eastern Brazil. *International Journal of Remote Sensing,* 20**,** 2683-2688.

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**,** e0169867.

ASADZADEH, S. & DE SOUZA FILHO, C. R. 2016. A review on spectral processing methods for geological remote sensing. *International journal of applied earth observation and geoinformation,* 47**,** 69-90.

ASADZADEH, S. & DE SOUZA FILHO, C. R. 2017. Spectral remote sensing for onshore seepage characterization: A critical overview. *Earth-Science Reviews,* 168**,** 48-72.

AVCIOGLU, E. 2010. Hydrocarbon Microseepage Mapping via Remote Sensing for Gemrik Anticline, Bozova Oil Field, Adiyaman, Turkey. *Ankara*.

BANERJEE, K., JAIN, M. K., JEYASEELAN, A. & PANDA, S. 2019. Landsat 8 OLI data for identification of hydrothermal alteration zone in Singhbhum Shear Zone using successive band depth difference technique-a new image processing approach. *Current Science (00113891),* 116.

BARSI, J., LEE, K., KVARAN, G., MARKHAM, B. & PEDELTY, J. 2014. The spectral response of the Landsat-8 operational land imager. *Remote Sensing,* 6**,** 10232-10251.

BOARDMAN, J. & KRUSE, F. 1994. Automated spectral analysis: a geological example using AVIRIS data, north Grapevine Mountains, Nevada: in Proceedings, ERIM Tenth Thematic Conference on Geologic Remote Sensing. *Environmental Research Institute of Michigan, Ann Arbor, MI, pp. I-407-I-418*.

BOARDMAN, J. W. 1993. Automating spectral unmixing of AVIRIS data using convex geometry concepts.

BOARDMAN, J. W. Leveraging the high dimensionality of AVIRIS data for improved sub-pixel target unmixing and rejection of false positives: mixture tuned matched filtering. Summaries of the seventh JPL Airborne Geoscience Workshop, JPL Publication, 1998, 1998. NASA Jet Propulsion Laboratory, 55-56.

BOARDMAN, J. W., KRUSE, F. A. & GREEN, R. O. 1995. Mapping target signatures via partial unmixing of AVIRIS data.

BONHAM-CARTER, G. 1990. Weights of evidence modelling: A new approach to mapping mineral potential.: In Statistical Applications in the Earth Sciences. *Geological Survey of Canada Paper,* 89**,** 171-183.

BONHAM-CARTER, G. 1994. Geographical information systems for geoscientists: modeling with GIS. *Computer Methods in the Geosciences,* 13.

BROWN, A. 2000. Evaluation of possible gas microseepage mechanisms. *AAPG bulletin,* 84**,** 1775-1789.

BUCHWITZ, M. D., BEEK, R. D., BURROWS, J., BOVENSMANN, H., WARNEKE, T., NOTHOLT, J., MEIRINK, J., GOEDE, A., BERGAMASCHI, P. & KÖRNER, S. 2005. Atmospheric methane and carbon dioxide from SCIAMACHY satellite data: initial comparison with chemistry and transport models. *Atmospheric Chemistry and Physics,* 5**,** 941-962.

CHEN, S., ZHAO, Y., ZHAO, L., LIU, Y. & ZHOU, C. 2017. Hydrocarbon micro-seepage detection by altered minerals mapping from airborne hyper-spectral data in Xifeng Oilfield, China. *Journal of Earth Science,* 28**,** 656-665.

CHRISTENSEN, T. R., ARORA, V. K., GAUSS, M., HÖGLUND-ISAKSSON, L. & PARMENTIER, F.-J. W. 2019. Tracing the climate signal: mitigation of anthropogenic methane emissions can outweigh a large Arctic natural emission increase. *Scientific reports,* 9**,** 1146.

CIAIS, P., SABINE, C., BALA, G., BOPP, L., BROVKIN, V., CANADELL, J., CHHABRA, A., DEFRIES, R., GALLOWAY, J. & HEIMANN, M. 2013. The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change. *IPCC Climate Change***,** 465-570.

CORTES, C. & VAPNIK, V. 1995. Support-vector networks. *Machine learning,* 20**,** 273-297.

CRISTIANINI, N. & SCHOLKOPF, B. 2002. Support vector machines and kernel methods: the new generation of learning machines. *Ai Magazine,* 23**,** 31-31.

CRÓSTA, A. & MOORE, J. M. 1989. Enhancement of Landsat Thematic Mapper imagery for residual soil mapping in SW Minais Gerais State, Brazil: A prospecting case history in Greenstone belt terrain. *Proceedings of the 7th (ERIM) Thematic Conference: Remote Sensing for Exploration Geology*, 1173-1187.

DEFANG, G. 1995. DIRECT SEARCHING FOR OIL AND GASBY REMOTE SENING TECHNOLOGY [J]. *Acta Petrolei Sinica,* 4.

DUCART, D. F., SILVA, A. M., TOLEDO, C. L. B. & ASSIS, L. M. D. 2016. Mapping iron oxides with Landsat-8/OLI and EO-1/Hyperion imagery from the Serra Norte iron deposits in the Carajás Mineral Province, Brazil. *Brazilian Journal of Geology,* 46**,** 331-349.

DWYER, J., ROY, D., SAUER, B., JENKERSON, C., ZHANG, H. & LYMBURNER, L. 2018. Analysis ready data: enabling analysis of the Landsat archive. *Remote Sensing,* 10**,** 1363.

EARTHEXPLORER. 2018. USGS. Available: https://earthexplorer.usgs.gov [Accessed 22 February 2018].

EEA, E. 2009. EEA air pollutant emission inventory guidebook—2009. *European Environment Agency (EEA), Copenhagen*.

ELVIDGE, C. D. 1990. Visible and near infrared reflectance characteristics of dry plant materials. *Remote Sensing,* 11**,** 1775-1795.

ETIOPE, G. 2004. GEM—Geologic emissions of methane, the missing source in the atmospheric methane budget. *Atmospheric Environment*.

ETIOPE, G. 2005. Mud volcanoes and microseepage: The forgotten geophysical components of atmospheric methane budget. *Annals of Geophysics,* 48.

ETIOPE, G. 2015. *Natural gas seepage: The Earth's hydrocarbon degassing*.

ETIOPE, G., CIOTOLI, G., SCHWIETZKE, S. & SCHOELL, M. 2019. Gridded maps of geological methane emissions and their isotopic signature. *Earth System Science Data*.

ETIOPE, G. & KLUSMAN, R. W. 2010a. Methane microseepage in drylands: soil is not always a CH4 sink. *Journal of Integrative Environmental Sciences,* 7**,** 31-38.

ETIOPE, G. & KLUSMAN, R. W. 2010b. Microseepage in drylands: flux and implications in the global atmospheric source/sink budget of methane. *Global and Planetary Change,* 72**,** 265-274.

ETIOPE, G., LASSEY, K. R., KLUSMAN, R. W. & BOSCHI, E. 2008. Reappraisal of the fossil methane budget and related emission from geologic sources. *Geophysical Research Letters,* 35.

GARAIN, S., MITRA, D. & DAS, P. 2020. Mineralogical and geochemical characterization of hydrocarbon microseepage-induced sediments in part of Assam-Arakan Fold Belt, Cachar area, NE India.

GASWIRTH, S. B. & MARRA, K. R. 2015. US Geological Survey 2013 assessment of undiscovered resources in the Bakken and Three Forks Formations of the US Williston Basin ProvinceResource Assessment of the Bakken and Three Forks Formations. *AAPG Bulletin,* 99**,** 639-660.

GOODARZI MEHR, S., AHADNEJAD, V., ABBASPOUR, R. A. & HAMZEH, M. 2013. Using the mixture-tuned matched filtering method for lithological mapping with Landsat TM5 images. *International journal of remote sensing,* 34**,** 8803-8816.

GREEN, A. A., BERMAN, M., SWITZER, P. & CRAIG, M. D. 1988. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on geoscience and remote sensing,* 26**,** 65-74.

GREEN, R. O. & BOARDMAN, J. 2000. Exploration of the relationship between information content and signal-to-noise ratio and spatial resolution in AVIRIS spectral data. *Spectrum,* 7.

HAJIBAPIR, G., LOTFI, M., ZARIFI, A. Z. & NEZAFATI, N. 2014. Application of different image processing techniques on aster and ETM+ images for exploration of hydrothermal alteration associated with copper mineralizations mapping Kehdolan Area (Eastern Azarbaijan Province-Iran). *Open Journal of Geology,* 4**,** 582.

HAN, B., DUAN, X., WANG, Y., ZHU, K., ZHANG, J., WANG, R., HU, H., QI, F., PAN, J. & YAN, Y. 2017. Methane protects against polyethylene glycol-induced osmotic stress in maize by improving sugar and ascorbic acid metabolism. *Scientific reports,* 7**,** 46185.

HAO, W., SCHARFFE, D., CRUTZEN, P. & SANHUEZA, E. 1988. Production of N 2 O, CH 4, and CO 2 from soils in the tropical savanna during the dry season. *Journal of Atmospheric Chemistry,* 7**,** 93-105.

HARRIS, A. T. 2006. Spectral mapping tools from the earth sciences applied to spectral microscopy data. *Cytometry Part A: The Journal of the International Society for Analytical Cytology,* 69**,** 872-879.

HARTLEY, B. K. 2014. *Evaluation of Weights of Evidence to Predict Gold Occurrences in Northern Minnesota's Archean Greenstone Belts*, University of Southern California.

HUANG, C., DAVIS, L. & TOWNSHEND, J. 2002. An assessment of support vector machines for land cover classification. *International Journal of remote sensing,* 23**,** 725-749.

JENSEN, J. R. 2015. *Introductory digital image processing: a remote sensing perspective*, Prentice Hall Press.

KALINOWSKI, A. & OLIVER, S. 2004. ASTER mineral index processing manual. *Remote Sensing Applications, Geoscience Australia,* 37**,** 36.

KLUSMAN, R., JAKEL, M. & LEROY, M. 1998. Does microseepage of methane and light hydrocarbons contribute to the atmospheric budget of methane and to global climate change. *Association of Petroleum Geochemical Explorationists Bulletin,* 11**,** 1-56.

KLUSMAN, R. W. & JAKEL, M. E. 1998. Natural microseepage of methane to the atmosphere from the Denver-Julesburg basin, Colorado. *Journal of Geophysical Research: Atmospheres,* 103**,** 28041-28045.

KLUSMAN, R. W., LEOPOLD, M. E. & LEROY, M. P. 2000. Seasonal variation in methane fluxes from sedimentary basins to the atmosphere: Results from chamber measurements and modeling of transport from deep sources. *Journal of Geophysical Research: Atmospheres,* 105**,** 24661-24670.

KOKALY, R. F., CLARK, R. N., SWAYZE, G. A., LIVO, K. E., HOEFEN, T. M., PEARSON, N. C., WISE, R. A., BENZEL, W. M., LOWERS, H. A. & DRISCOLL, R. L. 2017. USGS spectral library version 7. US Geological Survey.

KRUPNIK, D. & KHAN, S. 2017. Hydrocarbon microseepage-related geobotanical analysis in and around oil fields. *The Leading Edge,* 36**,** 12-23.

LIU, L., ZHOU, J., JIANG, D., ZHUANG, D., MANSARAY, L. R., HU, Z. & JI, Z. 2016. Mineral resources prospecting by synthetic application of TM/ETM+, Quickbird and Hyperion data in the Hatu area, West Junggar, Xinjiang, China. *Scientific reports,* 6**,** 21851.

LOUGHLIN, W. 1991. Principal component analysis for alteration mapping. *Photogrammetric Engineering and Remote Sensing,* 57**,** 1163-1169.

MBOGC, M. B. O. O. A. G. C. 2019. Available: http://www.bogc.dnrc.mt.gov/gisdata/WellSurface/ [Accessed 10/30/2019].

MIA, B. & FUJIMITSU, Y. 2012. Mapping hydrothermal altered mineral deposits using Landsat 7 ETM+ image in and around Kuju volcano, Kyushu, Japan. *Journal of earth system science,* 121**,** 1049-1057.

MIA, M. B. & FUJIMITSU, Y. 2013. Landsat thermal infrared based monitoring of heat losses from Kuju fumaroles area in Japan. *Procedia Earth and Planetary Science,* 6**,** 114-120.

MONSON, L. & SHURR, G. Remote sensing applications in the assessment of natural resources on the Fort Peck Reservation, Montana. PROCEEDINGS OF THE THEMATIC CONFERENCE ON GEOLOGIC REMOTE SENSING, 1993. ENVIRONMENTAL RESEARCH INSTITUTE OF MICHIGAN, 431-431.

MONSON, L. M. 2000. Phase I interim report 1. *Fort Peck Reservation assessment of hydrocarbon seepage: Semi-annual Technical Progress Report, Department of Energy Grant Award# DE-FG26-00BC15192.*

MONSON, L. M. 2001. Phase I report 2. *Fort Peck Reservation assessment of hydrocarbon seepage: Semi-annual Technical Progress Report, Department of Energy Grant Award# DE-FG26-00BC15192.*

MONSON, L. M. 2003. ASSESSMENT OF HYDROCARBON SEEPAGE DETECTION METHODS ON THE FORT PECK RESERVATION, NORTHEAST MONTANA. Fort Peck Reservation (US).

MOUNTRAKIS, G., IM, J. & OGOLE, C. 2011. Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing,* 66**,** 247-259.

MUDGE, M., SOKASKI, M., RICE, D. & MCINTYRE, G. 1977. Status of mineral resource information for the Fort Peck Indian Reservation, northeastern Montana. *US Bureau of Indian Affairs Administrative Report BIA-28.*

MUHAMMAD, I. A. & ALKI, H. D. 2015. Hydrocarbon Exploration Using Thermal Images West Baghdad. *Engineering and Technology Journal,* 33**,** 307-317.

NINOMIYA, Y. Mapping quartz, carbonate minerals, and mafic-ultramafic rocks using remotely sensed multispectral thermal infrared ASTER data. Thermosense XXIV, 2002. International Society for Optics and Photonics, 191-202.

NOOMEN, M. F. 2007. *Hyperspectral reflectance of vegetation affected by underground hydrocarbon gas seepage.*

PLAZA, A., MARTÍNEZ, P., PÉREZ, R. & PLAZA, J. 2004. A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data. *IEEE transactions on geoscience and remote sensing,* 42**,** 650-663.

POORMIRZAEE, R. & OSKOUEI, M. M. 2010. Use of spectral analysis for detection of alterations in ETM data, Yazd, Iran. *Applied Geomatics,* 2**,** 147-154.

POTTER, C. S., DAVIDSON, E. A. & VERCHOT, L. V. 1996. Estimation of global biogeochemical controls and seasonality in soil methane consumption. *Chemosphere,* 32**,** 2219-2246.

POUR, A. B. & HASHIM, M. Alteration mineral mapping using ETM+ and hyperion remote sensing data at Bau Gold Field, Sarawak, Malaysia. IOP Conference Series: Earth and Environmental Science, 2014. IOP Publishing, 012149.

POUR, A. B. & HASHIM, M. 2015. Hydrothermal alteration mapping from Landsat-8 data, Sar Cheshmeh copper mining district, south-eastern Islamic Republic of Iran. *Journal of Taibah University for Science,* 9**,** 155-166.

REAY, D., SMITH, P. & VAN AMSTEL, A. 2010. *Methane and climate change*, Earthscan.

ROBERTS, D., BATISTA, G. T., PEREIRA, J. L., WALLER, E. K. & NELSON, B. W. 1998. Change identification using multitemporal spectral mixture analysis: Applications in eastern Amazonia. *Remote sensing change detection: Environmental monitoring methods and applications,* 9**,** 137-159.

ROBSON, D. B., KNIGHT, J. D., FARRELL, R. E. & GERMIDA, J. J. 2004. Natural revegetation of hydrocarbon-contaminated soil in semi-arid grasslands. *Canadian Journal of Botany,* 82**,** 22-30.

ROCKWELL, B. W. 2004. *Spectral variations in rocks and soils containing ferric iron hydroxide and (or) sulfate minerals as seen by AVIRIS and laboratory spectroscopy*, DIANE Publishing.

ROCKWELL, B. W. 2013. Automated mapping of mineral groups and green vegetation from Landsat Thematic Mapper imagery with an example from the San Juan Mountains, Colorado. *US Geological Survey Scientific Investigations Map,* 3252.

ROY, D. P., JU, J., KLINE, K., SCARAMUZZA, P. L., KOVALSKYY, V., HANSEN, M., LOVELAND, T. R., VERMOTE, E. & ZHANG, C. 2010. Web-enabled Landsat Data (WELD): Landsat ETM+ composited mosaics of the conterminous United States. *Remote Sensing of Environment,* 114**,** 35-49.

ROY, D. P., KOVALSKYY, V., ZHANG, H., VERMOTE, E. F., YAN, L., KUMAR, S. & EGOROV, A. 2016. Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. *Remote sensing of Environment,* 185**,** 57-70.

SALATI, S. Characterization and Remote detection of onshore hydrocarbon seep-induced alteration. 2014. ITC.

SALATI, S., VAN RUITENBEEK, F., VAN DER MEER, F. & NAIMI, B. 2014. Detection of alteration induced by onshore gas seeps from ASTER and worldview-2 Data. *Remote sensing,* 6**,** 3188-3209.

SAUNDERS, D. F., BURSON, K. R. & THOMPSON, C. K. 1999. Model for hydrocarbon microseepage and related near-surface alterations. *AAPG bulletin,* 83**,** 170-185.

SCAFUTTO, R. D. M., DE SOUZA FILHO, C. R., RILEY, D. N. & DE OLIVEIRA, W. J. 2018. Evaluation of thermal infrared hyperspectral imagery for the detection of onshore methane plumes: Significance for hydrocarbon exploration and monitoring. *International journal of applied earth observation and geoinformation,* 64**,** 311-325.

SCHUMACHER, D. 1996. Hydrocarbon-induced alteration of soils and sediments.

SCHUMACHER, D. 2010. Integrating hydrocarbon microseepage data with seismic data doubles exploration success.

SCHUMACHER, D. 2012. Hydrocarbon microseepage—A significant but underutilized geologic principle with broad applications for oil/gas exploration and production. *American Association of Petroleum Geologists***,** 27.

SEGAL, D. & MERIN, I. 1989. Successful use of Landsat Thematic Mapper data for mapping hydrocarbon microseepage-induced mineralogic alteration, Lisbon Valley, Utah. *Photogrammetric Engineering and Remote Sensing,* 4**,** 1137-1145.

SEGAL, D. B. 1983. Use of Landsat multispectral scanner data for the definition of limonitic exposures in heavily vegetated areas. *Economic Geology,* 78**,** 711-722.

SHI, P., BIHONG, F. & NINOMIYA, Y. 2010. Mapping hydrocarbon seepage-induced anomalies in the arid region, west china using multispectral remote sensing. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science*.

SIMPSON, C. 1978. LANDSAT: developing techniques and applications in mineral and petroleum exploration. *BMR Journal of Australian Geology and Geophysics,* 3**,** 181-191.

SINGH, A. & HARRISON, A. 1985. Standardized principal components. *International journal of remote sensing,* 6**,** 883-896.

SOE, M., KYAW, T. A. & TAKASHIMA, I. 2005. Application of remote sensing techniques on iron oxide detection from ASTER and Landsat images of Tanintharyi coastal area, Myanmar.

STEIN, A., VAN DER MEER, F. D. & GORTE, B. 2006. *Spatial statistics for remote sensing*, Springer Science & Business Media.

STRIEGL, R. G. 1993. Diffusional limits to the consumption of atmospheric methane by soils. *Chemosphere,* 26**,** 715-720.

SUHERMAN, A., RAHMAN, M. & BUSU, I. Albedo and land surface temperature shift in hydrocarbon seepage potential area, case study in Miri Sarawak Malaysia. IOP Conference Series: Earth and Environmental Science, 2014. IOP Publishing, 012148.

TANG, J., XU, Y., WANG, G., ETIOPE, G., HAN, W., YAO, Z. & HUANG, J. 2017. Microseepage of methane to the atmosphere from the Dawanqi oil-gas field, Tarim Basin, China. *Journal of Geophysical Research: Atmospheres,* 122**,** 4353-4363.

TANGESTANI, M. H. & MOORE, F. 2000. Iron oxide and hydroxyl enhancement using the Crosta Method: A case study from the Zagros Belt, Fars Province, Iran. *International Journal of Applied Earth Observation and Geoinformation,* 2**,** 140-146.

THAMKE, J. N. 1991. Reconnaissance of ground-water resources of the Fort Peck Indian Reservation, northeastern Montana.

TURNER, A. J., FRANKENBERG, C. & KORT, E. A. 2019. Interpreting contemporary trends in atmospheric methane. *Proceedings of the National Academy of Sciences,* 116**,** 2805-2813.

VAN AMSTEL, A. 2012. Methane. A review. *Journal of Integrative Environmental Sciences,* 9**,** 5-30.

VAN DER MEER, F., VAN DIJK, P., VAN DER WERFF, H. & YANG, H. 2002. Remote sensing and petroleum seepage: a review and case study. *Terra Nova,* 14**,** 1-17.

VAPNIK, V. 2013. *The nature of statistical learning theory*, Springer science & business media.

WANG, W. 2016. *Comparing three classifiers for detecting hydrocarbon seepage alteration.* Master's Thesis, University of Twente, Enschede, The Netherlands.

WU, X., HUANG, B., PLAZA, A., LI, Y. & WU, C. 2013. Real-time implementation of the pixel purity index algorithm for endmember identification on GPUs. *IEEE Geoscience and Remote Sensing Letters,* 11**,** 955-959.

YANG, H., VAN DER MEER, F. & ZHANG, J. 2000. Aerospace detection of hydrocarbon-induced alteration. *Handbook of Exploration Geochemistry.* Elsevier.

YUAN, F., LI, X., ZHANG, M., JOWITT, S. M., JIA, C., ZHENG, T. & ZHOU, T. 2014. Three-dimensional weights of evidence-based prospectivity modeling: A case study of the Baixiangshan mining area, Ningwu Basin, Middle and Lower Yangtze Metallogenic Belt, China. *Journal of Geochemical Exploration,* 145**,** 82-97.

ZADEH, M. H., TANGESTANI, M. H., ROLDAN, F. V. & YUSTA, I. 2014. Sub-pixel mineral mapping of a porphyry copper belt using EO-1 Hyperion data. *Advances in Space Research,* 53**,** 440-451.

ZANTER, K. 2019. Landsat 8 (L8) Data Users Handbook, Department of the Interior. *US Geological Survey,* 4.

ZEGHOUANE, H., ALLEK, K. & KESRAOUI, M. 2016. GIS-based weights of evidence modeling applied to mineral prospectivity mapping of Sn-W and rare metals in Laouni area, Central Hoggar, Algeria. *Arabian Journal of Geosciences,* 9**,** 373.

ZHANG, G., SHEN, X., ZOU, L., LI, C., WANG, Y. & LU, S. 2007. Detection of hydrocarbon bearing sand through remote sensing techniques in the western slope zone of Songliao basin, China. *International Journal of Remote Sensing,* 28**,** 1819-1833.

ZHANG, G., ZHENG, Z., SHEN, X., ZOU, L. & HUANG, K. 2011. Remote sensing interpretation of areas with hydrocarbon microseepage in northeast China using Landsat-7/ETM+ data processing techniques. *International journal of remote sensing,* 32**,** 6695-6711.

ZHU, H., YANG, X. & LUO, Y. Classification of urban remote sensing image based on support vector machines. 2009 17th International Conference on Geoinformatics, 2009. IEEE, 1-6.

10 Appendices

10.1 Appendix 1: Software's

Table 23 - Software used in this study

SOFTWARE	PURPOSE
ArcMap 10.6.1	Overlay Analysis Boolean logic Thematic map generation
ENVI 5.5	Pre-processing of Landsat image PCA Band Ratio MNF PPI N-D Visualizer MTMF Spectral analysis of samples
Interactive Data Language (IDL) 8.7	SVM
ArcSDM 5	Weights of evidence Calculate ROC curves and AUC values

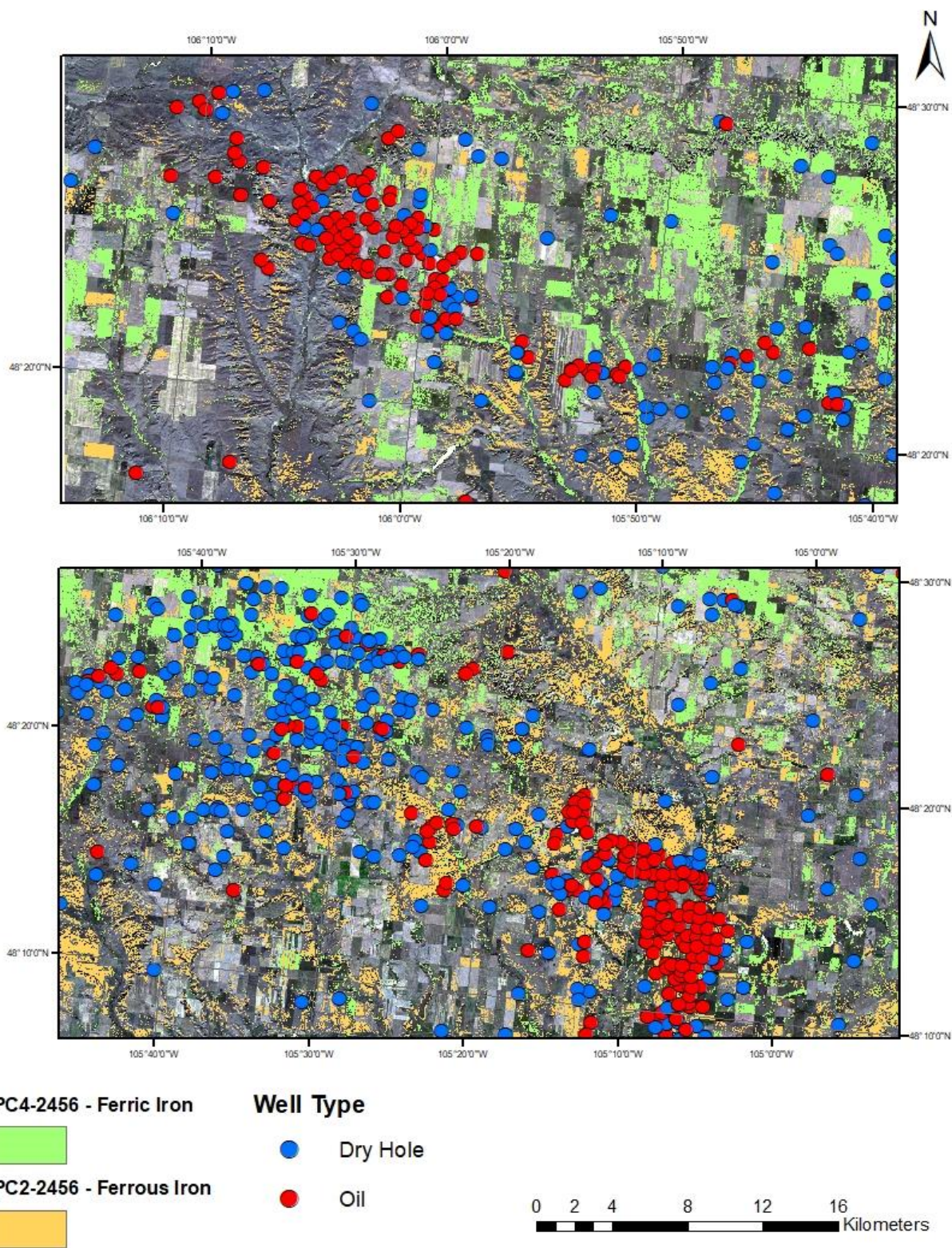## 10.2 Appendix 2: Additional maps

Figure 24 – Ferric iron and ferrous iron from PC2-2456 with wells distribution in Fort Peck Reservation.

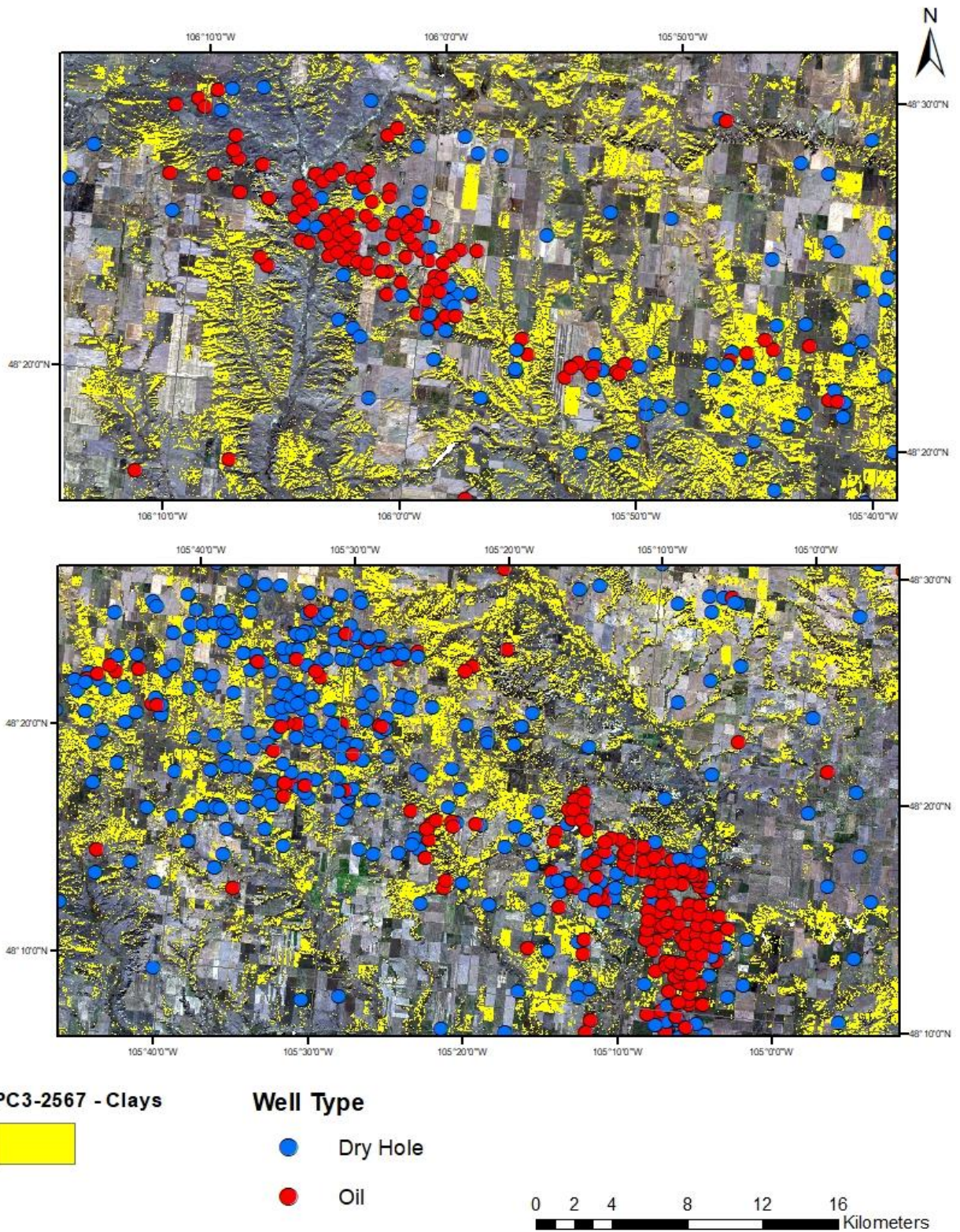*Figure 25 - Clays and ferrous iron mapped using PC3-2567 with wells distribution in Fort Peck Reservation*

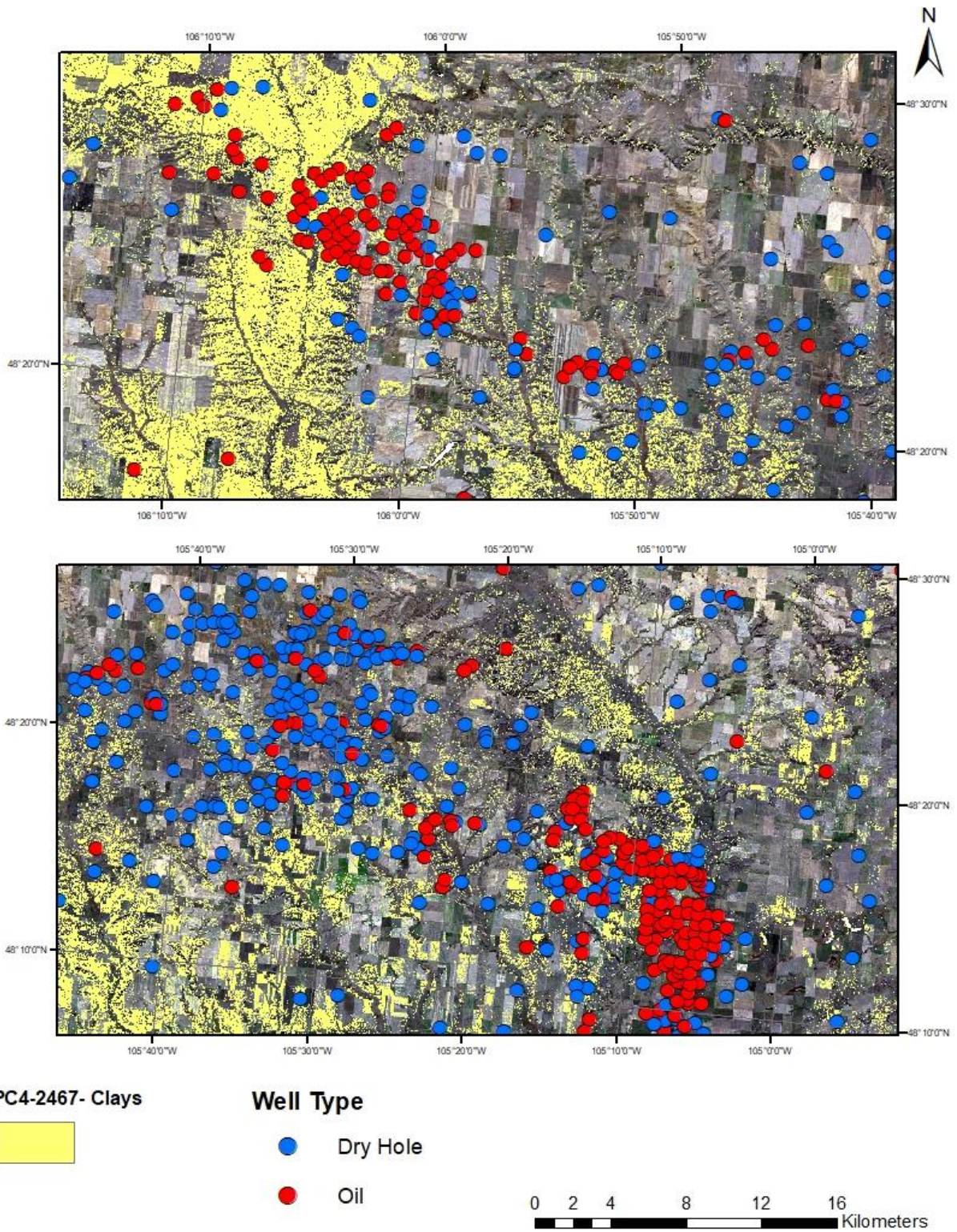*Figure 26- clays and carbonate minerals mapped using PC4-2467 with wells distribution in Fort Peck Reservation*
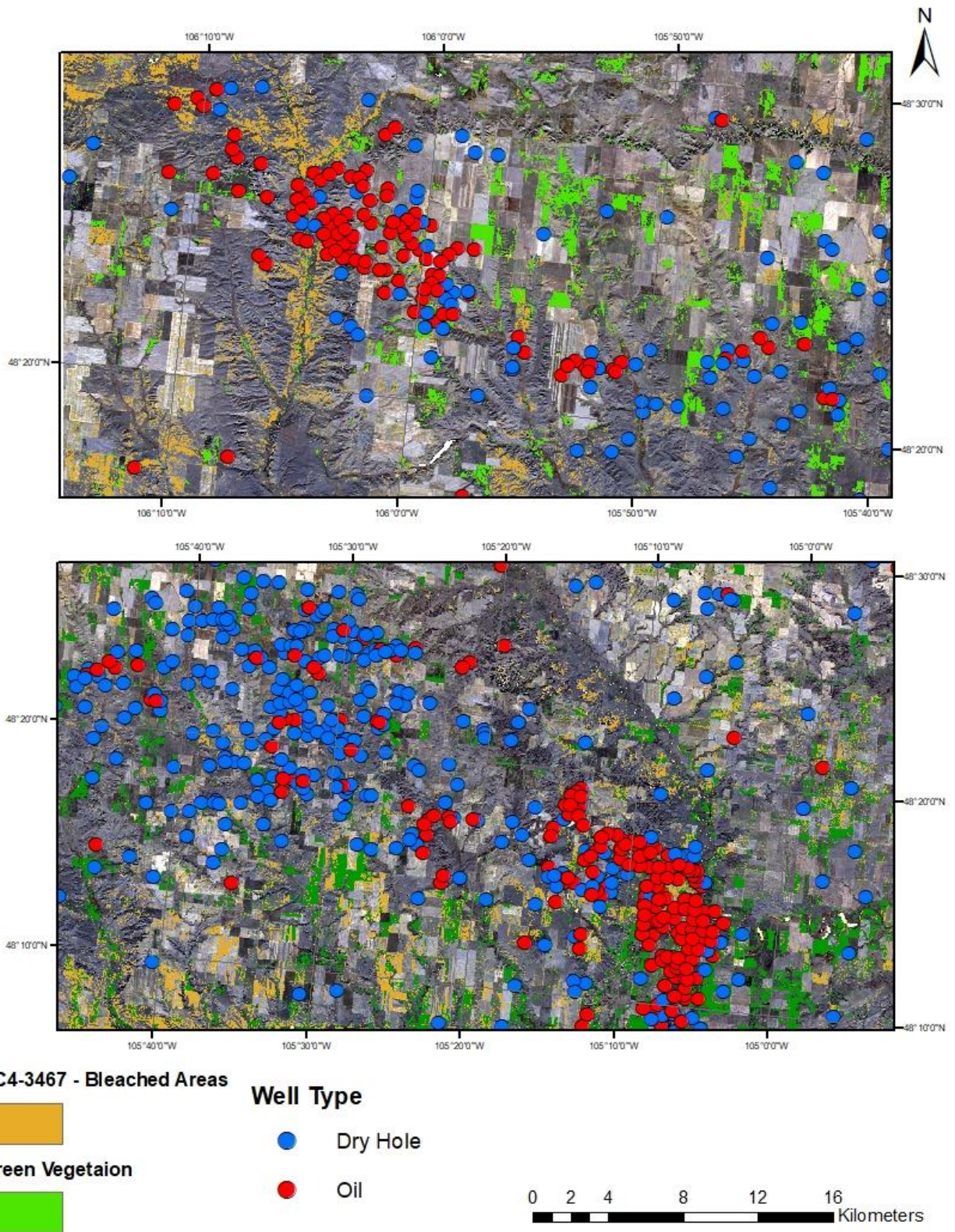
Figure 27- Bleached materials mapped using PC4-3467 with wells distribution in Fort Peck Reservation
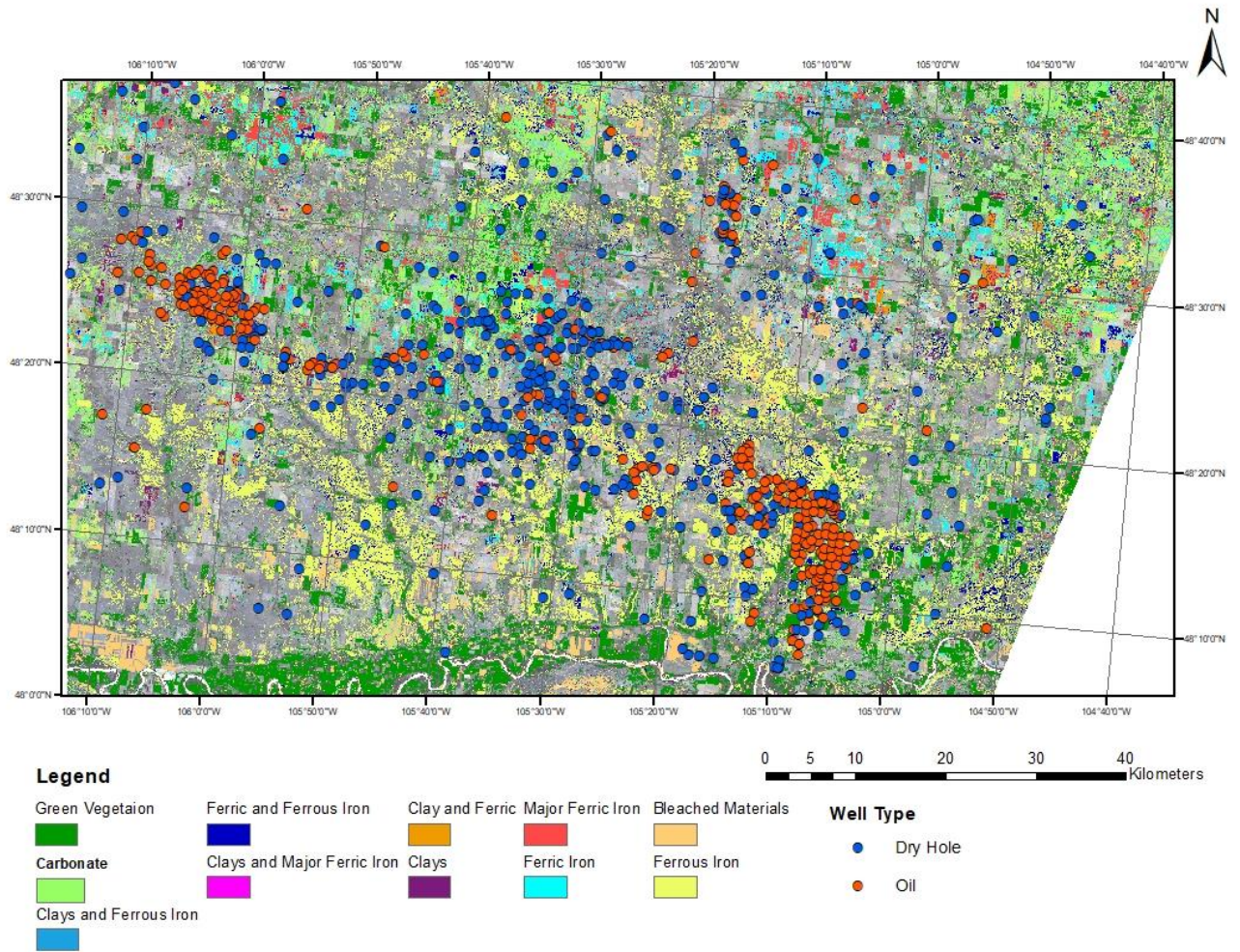
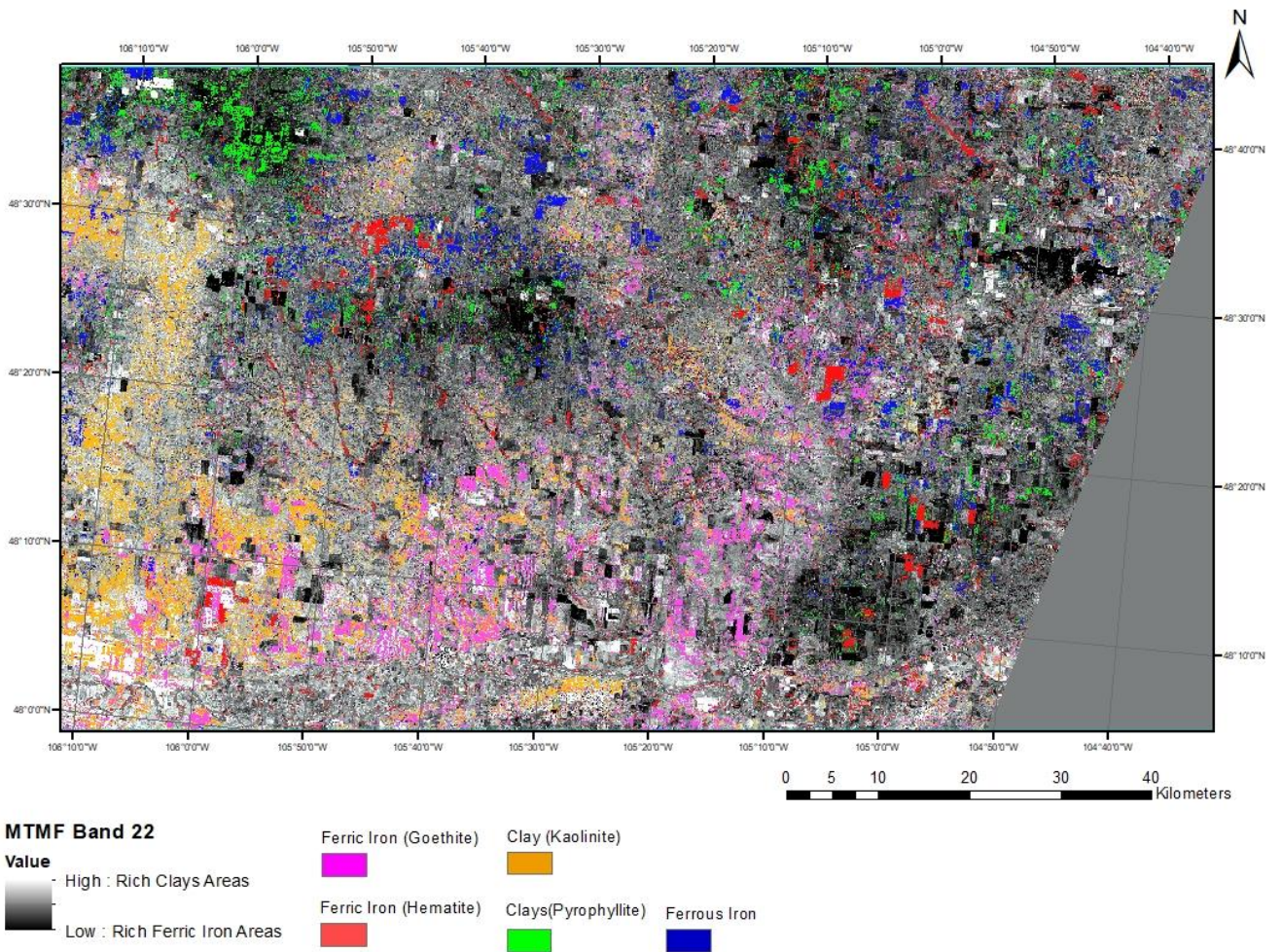*Figure 28-Result from BR technique showing different minerals groups generated using band algebra*

Figure 29 - MTMF result of band 22 delineate clays in bright pixels and ferric iron in dark pixels

10.3 Appendix 3: Weight of evidence calculation classification scheme

Table 24 - Classification scheme for evidential theme used for WofE analysis obtained from reclassifying PCA.

Class	Description
1	Background
2	Ferric
3	Ferrous
4	Ferric - Ferrous
5	Clays and Carbons
6	Clays and Carbons - Ferric
7	Clays and Carbons - Ferrous
8	Clays and Carbons - Ferric - Ferrous
9	Bleached
10	Bleached - Ferric
11	Bleached - Ferrous
12	Bleached - Ferric - Ferrous
13	Bleached - Clays and Carbons
14	Bleached - Clays and Carbons - Ferric
15	Bleached - Clays and Carbons - Ferrous
16	Bleached - Clays and Carbons - Ferric - Ferrous
17	Clays and Ferrous
18	Clays and Ferrous - Ferric
19	Clays and Ferrous - Ferrous
20	Clays and Ferrous - Ferric - Ferrous
21	Clays and Ferrous - Clays and Carbons
22	Clays and Ferrous - Clays and Carbons - Ferric
23	Clays and Ferrous - Clays and Carbons - Ferrous
24	Clays and Ferrous - Clays and Carbons - Ferric - Ferrous
25	Clays and Ferrous - Bleached
26	Clays and Ferrous - Bleached - Ferric
27	Clays and Ferrous - Bleached - Ferrous
28	Clays and Ferrous - Bleached - Ferric - Ferrous
29	Clays and Ferrous - Bleached - Clays and Carbons
30	Clays and Ferrous - Bleached - Clays and Carbons - Ferric
31	Clays and Ferrous - Bleached - Clays and Carbons - Ferrous
32	All

Table 25 - Classification scheme for evidential theme used for WofE analysis obtained from reclassifying BR.

Class	Description
0	Background
1	Carbonates
2	Ferric Iron
3	Ferric and Carbonates
4	Clays
5	Ferric and Clays
6	Clays and Carbonates
7	Bleached
8	Bleached and Carbonates
9	Bleached and Clays
10	Bleached and Ferric
11	Bleached and Ferric and Clays
12	Ferrous
13	Ferrous and Ferric
14	Ferrous and Carbonates
15	Ferrous and Clays
16	Ferrous and Ferric and Clays
17	Ferrous and Bleached
18	Ferrous and Ferric and Bleached
19	Ferrous and Clays and Bleached

Department of Physical Geography and Ecosystem Science

Master Thesis in Geographical Information Science

1. *Anthony Lawther:* The application of GIS-based binary logistic regression for slope failure susceptibility mapping in the Western Grampian Mountains, Scotland (2008).

2. *Rickard Hansen:* Daily mobility in Grenoble Metropolitan Region, France. Applied GIS methods in time geographical research (2008).

3. *Emil Bayramov:* Environmental monitoring of bio-restoration activities using GIS and Remote Sensing (2009).

4. *Rafael Villarreal Pacheco:* Applications of Geographic Information Systems as an analytical and visualization tool for mass real estate valuation: a case study of Fontibon District, Bogota, Columbia (2009).

5. *Siri Oestreich Waage:* a case study of route solving for oversized transport: The use of GIS functionalities in transport of transformers, as part of maintaining a reliable power infrastructure (2010).

6. *Edgar Pimiento:* Shallow landslide susceptibility – Modelling and validation (2010).

7. *Martina Schäfer:* Near real-time mapping of floodwater mosquito breeding sites using aerial photographs (2010).

8. *August Pieter van Waarden-Nagel:* Land use evaluation to assess the outcome of the programme of rehabilitation measures for the river Rhine in the Netherlands (2010).

9. *Samira Muhammad:* Development and implementation of air quality data mart for Ontario, Canada: A case study of air quality in Ontario using OLAP tool. (2010).

10. *Fredros Oketch Okumu*: Using remotely sensed data to explore spatial and temporal relationships between photosynthetic productivity of vegetation and malaria transmission intensities in selected parts of Africa (2011).

11. *Svajunas Plunge:* Advanced decision support methods for solving diffuse water pollution problems (2011).

12. *Jonathan Higgins:* Monitoring urban growth in greater Lagos: A case study using GIS to monitor the urban growth of Lagos 1990 - 2008 and produce future growth prospects for the city (2011).

13. *Mårten Karlberg:* Mobile Map Client API: Design and Implementation for Android (2011).

14. *Jeanette McBride:* Mapping Chicago area urban tree canopy using color infrared imagery (2011).

15. *Andrew Farina:* Exploring the relationship between land surface temperature and vegetation abundance for urban heat island mitigation in Seville, Spain (2011).

16. *David Kanyari*: Nairobi City Journey Planner: An online and a Mobile Application (2011).

17. *Laura V. Drews:* Multi-criteria GIS analysis for siting of small wind power plants - A case study from Berlin (2012).

18. *Qaisar Nadeem:* Best living neighborhood in the city - A GIS based multi criteria evaluation of ArRiyadh City (2012).

19. *Ahmed Mohamed El Saeid Mustafa:* Development of a photo voltaic building rooftop integration analysis tool for GIS for Dokki District, Cairo, Egypt (2012).

20. *Daniel Patrick Taylor*: Eastern Oyster Aquaculture: Estuarine Remediation via Site Suitability and Spatially Explicit Carrying Capacity Modeling in Virginia's Chesapeake Bay (2013).

21. *Angeleta Oveta Wilson:* A Participatory GIS approach to *unearthing* Manchester's Cultural Heritage '*gold mine'* (2013).

22. *Ola Svensson:* Visibility and Tholos Tombs in the Messenian Landscape: A Comparative Case Study of the Pylian Hinterlands and the Soulima Valley (2013).

23. *Monika Ogden:* Land use impact on water quality in two river systems in South Africa (2013).

24. *Stefan Rova:* A GIS based approach assessing phosphorus load impact on Lake Flaten in Salem, Sweden (2013).

25. *Yann Buhot:* Analysis of the history of landscape changes over a period of 200 years. How can we predict past landscape pattern scenario and the impact on habitat diversity? (2013).

26. *Christina Fotiou:* Evaluating habitat suitability and spectral heterogeneity models to predict weed species presence (2014).

27. *Inese Linuza:* Accuracy Assessment in Glacier Change Analysis (2014).

28. *Agnieszka Griffin:* Domestic energy consumption and social living standards: a GIS analysis within the Greater London Authority area (2014).

29. *Brynja Guðmundsdóttir:* Detection of potential arable land with remote sensing and GIS - A Case Study for Kjósarhreppur (2014).

30. *Oleksandr Nekrasov:* Processing of MODIS Vegetation Indices for analysis of agricultural droughts in the southern Ukraine between the years 2000-2012 (2014).

31. *Sarah Tressel:* Recommendations for a polar Earth science portal in the context of Arctic Spatial Data Infrastructure (2014).

32. *Caroline Gevaert:* Combining Hyperspectral UAV and Multispectral Formosat-2 Imagery for Precision Agriculture Applications (2014).

33. *Salem Jamal-Uddeen:* Using GeoTools to implement the multi-criteria evaluation analysis - weighted linear combination model (2014).

34. *Samanah Seyedi-Shandiz:* Schematic representation of geographical railway network at the Swedish Transport Administration (2014).

35. *Kazi Masel Ullah:* Urban Land-use planning using Geographical Information System and analytical hierarchy process: case study Dhaka City (2014).

36. *Alexia Chang-Wailing Spitteler:* Development of a web application based on MCDA and GIS for the decision support of river and floodplain rehabilitation projects (2014).

37. *Alessandro De Martino:* Geographic accessibility analysis and evaluation of potential changes to the public transportation system in the City of Milan (2014).

38. *Alireza Mollasalehi:* GIS Based Modelling for Fuel Reduction Using Controlled Burn in Australia. Case Study: Logan City, QLD (2015).

39. *Negin A. Sanati:* Chronic Kidney Disease Mortality in Costa Rica; Geographical Distribution, Spatial Analysis and Non-traditional Risk Factors (2015).

40. *Karen McIntyre:* Benthic mapping of the Bluefields Bay fish sanctuary, Jamaica (2015).

41. *Kees van Duijvendijk:* Feasibility of a low-cost weather sensor network for agricultural purposes: A preliminary assessment (2015).

42. *Sebastian Andersson Hylander:* Evaluation of cultural ecosystem services using GIS (2015).

43. *Deborah Bowyer:* Measuring Urban Growth, Urban Form and Accessibility as Indicators of Urban Sprawl in Hamilton, New Zealand (2015).

44. *Stefan Arvidsson:* Relationship between tree species composition and phenology extracted from satellite data in Swedish forests (2015).

45. *Damián Giménez Cruz*: GIS-based optimal localisation of beekeeping in rural Kenya (2016).

46. *Alejandra Narváez Vallejo:* Can the introduction of the topographic indices in LPJ-GUESS improve the spatial representation of environmental variables? (2016).

47. *Anna Lundgren:* Development of a method for mapping the highest coastline in Sweden using breaklines extracted from high resolution digital elevation models (2016).

48. *Oluwatomi Esther Adejoro:* Does location also matter? A spatial analysis of social achievements of young South Australians (2016).

49. *Hristo Dobrev Tomov:* Automated temporal NDVI analysis over the Middle East for the period 1982 - 2010 (2016).

50. *Vincent Muller:* Impact of Security Context on Mobile Clinic Activities
A GIS Multi Criteria Evaluation based on an MSF Humanitarian Mission in Cameroon (2016).

51. *Gezahagn Negash Seboka:* Spatial Assessment of NDVI as an Indicator of Desertification in Ethiopia using Remote Sensing and GIS (2016).

52. *Holly Buhler:* Evaluation of Interfacility Medical Transport Journey Times in Southeastern British Columbia. (2016).

53. *Lars Ole Grottenberg:* Assessing the ability to share spatial data between emergency management organisations in the High North (2016).

54. *Sean Grant:* The Right Tree in the Right Place: Using GIS to Maximize the Net Benefits from Urban Forests (2016).

55. *Irshad Jamal:* Multi-Criteria GIS Analysis for School Site Selection in Gorno-Badakhshan Autonomous Oblast, Tajikistan (2016).

56. *Fulgencio Sanmartín:* Wisdom-volkano: A novel tool based on open GIS and time-series visualization to analyse and share volcanic data (2016).

57. *Nezha Acil:* Remote sensing-based monitoring of snow cover dynamics and its influence on vegetation growth in the Middle Atlas Mountains (2016).

58. *Julia Hjalmarsson:* A Weighty Issue: Estimation of Fire Size with Geographically Weighted Logistic Regression (2016).

59. *Mathewos Tamiru Amato:* Using multi-criteria evaluation and GIS for chronic food and nutrition insecurity indicators analysis in Ethiopia (2016).

60. *Karim Alaa El Din Mohamed Soliman El Attar:* Bicycling Suitability in Downtown, Cairo, Egypt (2016).

61. *Gilbert Akol Echelai:* Asset Management: Integrating GIS as a Decision Support Tool in Meter Management in National Water and Sewerage Corporation (2016).

62. *Terje Slinning:* Analytic comparison of multibeam echo soundings (2016).

63. *Gréta Hlín Sveinsdóttir:* GIS-based MCDA for decision support: A framework for wind farm siting in Iceland (2017).

64. *Jonas Sjögren:* Consequences of a flood in Kristianstad, Sweden: A GIS-based analysis of impacts on important societal functions (2017).

65. *Nadine Raska:* 3D geologic subsurface modelling within the Mackenzie Plain, Northwest Territories, Canada (2017).

66. *Panagiotis Symeonidis*: Study of spatial and temporal variation of atmospheric optical parameters and their relation with PM 2.5 concentration over Europe using GIS technologies (2017).

67. *Michaela Bobeck:* A GIS-based Multi-Criteria Decision Analysis of Wind Farm Site Suitability in New South Wales, Australia, from a Sustainable Development Perspective (2017).

68. *Raghdaa Eissa*: Developing a GIS Model for the Assessment of Outdoor Recreational Facilities in New Cities Case Study: Tenth of Ramadan City, Egypt (2017).

69. *Zahra Khais Shahid*: Biofuel plantations and isoprene emissions in Svea and Götaland (2017).

70. *Mirza Amir Liaquat Baig*: Using geographical information systems in epidemiology: Mapping and analyzing occurrence of diarrhea in urban - residential area of Islamabad, Pakistan (2017).

71. *Joakim Jörwall*: Quantitative model of Present and Future well-being in the EU-28: A spatial Multi-Criteria Evaluation of socioeconomic and climatic comfort factors (2017).

72. *Elin Haettner*: Energy Poverty in the Dublin Region: Modelling Geographies of Risk (2017).

73. *Harry Eriksson*: Geochemistry of stream plants and its statistical relations to soil- and bedrock geology, slope directions and till geochemistry. A GIS-analysis of small catchments in northern Sweden (2017).

74. *Daniel Gardevärn:* PPGIS and Public meetings – An evaluation of public participation methods for urban planning (2017).

75. *Kim Friberg:* Sensitivity Analysis and Calibration of Multi Energy Balance Land Surface Model Parameters (2017).

76. *Viktor Svanerud:* Taking the bus to the park? A study of accessibility to green areas in Gothenburg through different modes of transport (2017).

77. *Lisa-Gaye Greene*: Deadly Designs: The Impact of Road Design on Road Crash Patterns along Jamaica's North Coast Highway (2017).

78. *Katarina Jemec Parker*: Spatial and temporal analysis of fecal indicator bacteria concentrations in beach water in San Diego, California (2017).

79. *Angela Kabiru*: An Exploratory Study of Middle Stone Age and Later Stone Age Site Locations in Kenya's Central Rift Valley Using Landscape Analysis: A GIS Approach (2017).

80. *Kristean Björkmann*: Subjective Well-Being and Environment: A GIS-Based Analysis (2018).

81. *Williams Erhunmonmen Ojo*: Measuring spatial accessibility to healthcare for people living with HIV-AIDS in southern Nigeria (2018).

82. *Daniel Assefa*: Developing Data Extraction and Dynamic Data Visualization (Styling) Modules for Web GIS Risk Assessment System (WGRAS). (2018).

83. *Adela Nistora*: Inundation scenarios in a changing climate: assessing potential impacts of sea-level rise on the coast of South-East England (2018).

84. *Marc Seliger*: Thirsty landscapes - Investigating growing irrigation water consumption and potential conservation measures within Utah's largest master-planned community: Daybreak (2018).

85. *Luka Jovičić*: Spatial Data Harmonisation in Regional Context in Accordance with INSPIRE Implementing Rules (2018).

86. *Christina Kourdounouli*: Analysis of Urban Ecosystem Condition Indicators for the Large Urban Zones and City Cores in EU (2018).

87. *Jeremy Azzopardi*: Effect of distance measures and feature representations on distance-based accessibility measures (2018).

88. *Patrick Kabatha*: An open source web GIS tool for analysis and visualization of elephant GPS telemetry data, alongside environmental and anthropogenic variables (2018).

89. *Richard Alphonce Giliba*: Effects of Climate Change on Potential Geographical Distribution of Prunus africana (African cherry) in the Eastern Arc Mountain Forests of Tanzania (2018).

90. *Eiður Kristinn Eiðsson*: Transformation and linking of authoritative multi-scale geodata for the Semantic Web: A case study of Swedish national building data sets (2018).

91. *Niamh Harty*: HOP!: a PGIS and citizen science approach to monitoring the condition of upland paths (2018).

92. *José Estuardo Jara Alvear*: Solar photovoltaic potential to complement hydropower in Ecuador: A GIS-based framework of analysis (2018).

93. *Brendan O'Neill*: Multicriteria Site Suitability for Algal Biofuel Production Facilities (2018).

94. *Roman Spataru*: Spatial-temporal GIS analysis in public health – a case study of polio disease (2018).

95. *Alicja Miodońska*: Assessing evolution of ice caps in Suðurland, Iceland, in years 1986 - 2014, using multispectral satellite imagery (2019).

96. *Dennis Lindell Schettini*: A Spatial Analysis of Homicide Crime's Distribution and Association with Deprivation in Stockholm Between 2010-2017 (2019).

97. *Damiano Vesentini*: The Po Delta Biosphere Reserve: Management challenges and priorities deriving from anthropogenic pressure and sea level rise (2019).

98. *Emilie Arnesten*: Impacts of future sea level rise and high water on roads, railways and environmental objects: a GIS analysis of the potential effects of increasing sea levels and highest projected high water in Scania, Sweden (2019).

99. *Syed Muhammad Amir Raza*: Comparison of geospatial support in RDF stores: Evaluation for ICOS Carbon Portal metadata (2019).

100. *Hemin Tofiq*: Investigating the accuracy of Digital Elevation Models from UAV images in areas with low contrast: A sandy beach as a case study (2019).

101. *Evangelos Vafeiadis*: Exploring the distribution of accessibility by public transport using spatial analysis. A case study for retail concentrations and public hospitals in Athens (2019).

102. *Milan Sekulic*: Multi-Criteria GIS modelling for optimal alignment of roadway by-passes in the Tlokweng Planning Area, Botswana (2019).

103. *Ingrid Piirisaar*: A multi-criteria GIS analysis for siting of utility-scale photovoltaic solar plants in county Kilkenny, Ireland (2019).

104. *Nigel Fox*: Plant phenology and climate change: possible effect on the onset of various wild plant species' first flowering day in the UK (2019).

105. *Gunnar Hesch*: Linking conflict events and cropland development in Afghanistan, 2001 to 2011, using MODIS land cover data and Uppsala Conflict Data Programme (2019).

106. *Elijah Njoku*: Analysis of spatial-temporal pattern of Land Surface Temperature (LST) due to NDVI and elevation in Ilorin, Nigeria (2019).

107. *Katalin Bunyevácz*: Development of a GIS methodology to evaluate informal urban green areas for inclusion in a community governance program (2019).

108. *Paul dos Santos*: Automating synthetic trip data generation for an agent-based simulation of urban mobility (2019).

109. *Robert O' Dwyer*: Land cover changes in Southern Sweden from the mid-Holocene to present day: Insights for ecosystem service assessments (2019).

110. *Daniel Klingmyr*: Global scale patterns and trends in tropospheric NO2 concentrations (2019).

111. *Marwa Farouk Elkabbany*: Sea Level Rise Vulnerability Assessment for Abu Dhabi, United Arab Emirates (2019).

112. *Jip Jan van Zoonen*: Aspects of Error Quantification and Evaluation in Digital Elevation Models for Glacier Surfaces (2020).

113. *Georgios Efthymiou*: The use of bicycles in a mid-sized city – benefits and obstacles identified using a questionnaire and GIS (2020).

114. *Haruna Olayiwola Jimoh*: Assessment of Urban Sprawl in MOWE/IBAFO Axis of Ogun State using GIS Capabilities (2020).

115. *Nikolaos Barmpas Zachariadis*: Development of an iOS, Augmented Reality for disaster management (2020).

116. *Ida Storm*: ICOS Atmospheric Stations: Spatial Characterization of CO2 Footprint Areas and Evaluating the Uncertainties of Modelled CO2 Concentrations (2020).

117. *Alon Zuta*: Evaluation of water stress mapping methods in vineyards using airborne thermal imaging (2020).

118. *Marcus Eriksson*: Evaluating structural landscape development in the municipality Upplands-Bro, using landscape metrics indices (2020).

119. *Ane Rahbek Vierø*: Connectivity for Cyclists? A Network Analysis of Copenhagen's Bike Lanes (2020).

120. *Cecilia Baggini*: Changes in habitat suitability for three declining Anatidae species in saltmarshes on the Mersey estuary, North-West England (2020).

121. *Bakrad Balabanian*: Transportation and Its Effect on Student Performance (2020).

122. *Ali Al Farid*: Knowledge and Data Driven Approaches for Hydrocarbon Microseepage Characterizations: An Application of Satellite Remote Sensing (2020).