Evaluating the Performance of the Resampled nSight-2 Sensor's Spectral Configuration in Discriminating Wetland Plant Species Using Advanced Classifiers and Spectroscopy.

Mchasisi Gasela

GSLMCH001

Supervisor: Professor Gerhard De Jager

Co-supervisor: Mr. Mahlatse L. Kganyago

Dissertation in Partial Fulfilment of the Master of Philosophy in Space Studies

Department of Electrical Engineering

University of Cape Town

FEBRUARY 2022

DECLARATION

I know the meaning of plagiarism and declare that all the work in the document, save for that which is properly acknowledged, is my own. It has not been previously submitted, in part or whole, to any university of institution for any degree, diploma, or other qualification. This thesis/dissertation has been submitted to the Turnitin module (or equivalent similarity and originality checking software) and I confirm that my supervisor has seen my report and any concerns revealed by such have been resolved with my supervisor.

Signed: Signed by candidate

Date: 14/02/2022

Mchasisi Gasela, MSc. Environmental Management, Unisa

PUBLICATIONS AND MANUSCRIPTS

The following papers are under peer-review or being prepared for publication. I substantially contributed to the study design, analysis, interpretation and discussion of the results and the overall preparation of the manuscripts; hence I am the appropriate first author in both papers.

1 **Gasela, M**., Kganyago, M. and De Jager, G., 2022. Testing the utility of the resampled nSight-2 spectral configurations in discriminating wetland plant species using Random Forest classifier. *Geocarto International*, pp.1-16.

2 **Gasela, M.**, Kganyago, M. and De Jager G. **(In Preparation)** Comparison of Machine Learning Classifiers for Discriminating Wetland Plant Species Using Resampled nSight 2 Spectral Settings. GIScience and Remote Sensing for peer review.

ABSTRACT

Accurate and reliable information about wetland plant species is critical, as it informs improved preservation, conservation and management of wetland ecosystems. Well managed ecosystems guarantee achieving Sustainable Development Goals. Therefore, remote sensing technique has gained prominence in providing such information. However, broadband sensors are affected by effects of soil and water reflectances associated with wetlands hence cannot adequately discern subtle differences among wetland plant species. On the other hand, hyperspectral sensors allow for an in-depth examination of plant leaf and canopy biochemical traits such as lignin, cellulose, nitrogen, chlorophyll, carotenoids, anthocyanin and water content through spectral measurements which is critical for plant species discrimination. This study sought to test the capability of the forthcoming nSight-2 hyperspectral sensor in discriminating among four dominant wetland plant species. To accomplish this, the performance of nSight-2 spectral settings were compared with those of the upcoming EnMap hyperspectral satellite and an already established Worldview-2 multi-spectral sensor that carries strategic wavebands for vegetation studies, i.e. red-edge and near-infrared. The study also evaluated the performances of non-parametric machine learning algorithms in classifying wetland plant species using nSight-2 spectral configuration. The results showed a high discrimination accuracy by nSight-2 spectral settings with an overall accuracy of 84.09%, followed by Worldview-2 i.e. 81.82% while EnMap was the worst i.e. 77.77%. The most important bands for vegetation analysis were within the visible (VIS), Red-edge (RE) and near infrared (NIR) regions of the electromagnetic spectrum. The study also demonstrated that within these spectral bands, the four dominant Verloren Vallei Nature Reserve wetland plant i.e. Crocosmia sp., Grasses, Agapanthus sp. and Cyperus sp. could be differentiated using the spectral settings of these sensors. Furthermore, the results showed a superior performance of Support Vector Machine (SVM) with overall accuracy of 93.18%, compared with the RF and Partial Least Squares-Discriminant Analysis (PLS-DA) that had overall accuracies of 84.09% and 83.63% respectively. In summary, the study demonstrated that the spectral configuration of nSight-2 hyperspectral sensor can discriminate among the wetland plant species with comparable accuracy to that of a state-of-the-art sensor, i.e. Worldview-2 and better than the upcoming EnMap.

ACKNOWLEDGEMENTS

First and foremost, I send my sincere gratitude to my supervisors: Prof. Gerhard De Jagger and Mr. Mahlatse Kganyago for their unwavering tutelage and dexterous guide towards the production of this academic work. I cannot fail to marvel their skill in helping me knit this research work.

Special recognition also goes Prof. Peter Martinez who introduced me to the Space Studies programme at the University of Cape Town and Mr. Martin Jacobs for organizing funding for this work from the Space Advisory Company. This research work would not have been complete without their efforts.

I also thank SANSA as an institute that relentlessly provided me with tools and necessary material required for this work. A special thank you to Ms Andiswa Mlisa and Asanda Sangoni for mobilizing the resources desperately needed from SANSA to accomplish this project.

Lastly, but not least a special appreciation to the Mpumalanga Tourism and Parks Agency for allowing me a chance to undertake my academic research within their nature reserve. A thank you further goes to Ms. Shirley Sibiya the manager at Verloren Vallei Nature Reserve for the support she gave in demarcating the study area and identification of various plant species. Her continued guide and support throughout the fieldwork period cannot go unrecognized.

I would like also to thank the Department of Electrical Engineering at the University of Cape Town for affording me an opportunity to further my academic career and mingle with other academics in the Space Lab.

To my family who provided utmost support and had to cotton to the idea of an absent father, let me say, you are such special people and this work is a testimony of the beauty that you are.

DEDICATION

This research report is dedicated to my late parents Gibson "Gibbis" Gasela and Fildah "Manina" Gasela, my children Khayelihle, Khanyile, Kholekile and Khethulwazi.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

ACRONYMS

AD	Allocation Difference
ANN	Artificial Neural Network
IVM	Import Vector Machine
AISA	Airborne Imaging Spectrometer for Applications
APEX	Airborne Prism Experiment
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
AVHRR	Advanced Very High Resolution Radiometer
CE	Commission Error
CI	Confidence Interval
DLR	Germany Space Agency
GSD	Ground Sampling Distance
EnMap	Environmental Monitoring and Analysis Programme
Envisat	Environmental Satellite
DT	Decision Trees
VNIR	Visible Near Infrared
GBA	Gradient Boosting Algorithm
NASA	National Aeronautics and Space Administration
GIS	Geographical Information Systems
EO	Earth Observation
ETM	Enhanced Thematic Mapper
FWHM	Full Width Half Maximum

Landsat	Land Remote-Sensing Satellite
MERIS	Medium Resolution Imaging Spectrometer
ML	Maximum Likelihood
MLA	Machine Learning Algorithm
MODIS	Moderate Resolution Imaging Spectroradiometer
MSI	Multi Spectral Imager
MTPA	Mpumalanga Tourism and Parks Agency
mtry	Number of variables
NDVI	Normalized Difference Vegetation Index
nEVI	narrow Enhanced Vegetation Index
NIR	near Infrared
nNDVI	narrow Normalized Difference Vegetation Index
NOAA	National Oceanic and Atmospheric Administration
nSight 2	A forthcoming hyperspectral satellite
n-tree	Number of trees
nTVI	narrow Transformed Vegetation Index
OA	Overall Accuracy
OE	Omission Error
OLI	Operational Land Imager
PA	Producer Accuracy
PLS-DA	Partial Least Squares Discriminant Analysis
QD	Quantity Difference
RBF	Radial Basis Function

RE	Red-Edge
RF	Random Forest
SAC/SCS	Space Advisory Company
SANBI	South African National Biodiversity Institute
SANSA	South African National Space Agency
ASD	Analytical Spectral Device
SPOT	Satellite Pour l'Observation de la Terre
SVM	Support Vector Machine
SWIR	Short Wave Infrared
TM	Thematic Mapper
TVI	Transformed Vegetation Index
UA	User Accuracy
UN-SDG	United Nations Sustainable Development Goals
VIS	Visible
SR	Simple Ratio
VVNR	Verloren Vallei Nature Reserve
WRI	World Resources Institute
WV 2	WorldView 2
WV 3	Worldview

CHAPTER ONE

1 GENERAL INTRODUCTION

1.1.1 Background

Wetland ecosystems are among the most productive systems in the world offering great ecological, social and economic benefits. They support over one billion people globally through provision of various ecosystems goods and services (Amler, Schmidt and Menz, 2015). Global statistics reveal that the estimated financial value of the ecosystem goods and services provided by wetlands is US$4.9 trillion annually across the globe (Walter and Mondall, 2019). Their ecological benefits include, water purification (through filtering agricultural and industrial waste), aquifer recharging and surface water run-off velocity retardation (Guo, et al. 2017; Samiappan, et al. 2017). They protect human life from flooding and drought. They act as sponges that are pivotal in flood attenuation (Ramsey and Rangoonwala, 2012; Adam, Mutanga and Rugege, 2010). Wetlands are an essential habitat for various flora and fauna housing over 40% of global plant and animal species, providing food, water, shelter and breeding nests among other services. They are a sanctuary for endangered and threatened flora and fauna (Dalponte, et al. 2012). Healthy wetlands are also critical players in carbon sequestration, as are an important sink of carbon, assisting in combating climate change (Samiappan, et al. 2017; Guo, et al. 2017). Wetland ecosystems offer great socio-economic opportunities through outdoor activities such as recreation, education, scientific research, photography, fishing, hunting and bird-viewing (Ola and Benjamin, 2019). As such there is an increased demand for accurate, reliable and up-to-date information about their status at any given time.

Despite their huge ecological, social and economic value, wetlands continue to be degraded, drained and transformed (Assessment, 2005; Randin, et al. 2020). Land use and land cover changes, climate change, pollution, over-exploitation, biological invasions, urban sprawl, industrialization, agricultural expansion and wildfire events are some of the anthropogenic and natural drivers of wetlands transformation (Randin, et al. 2020). Furthermore, the World Resources Institute (WRI) noted in the Millennium Ecosystem Assessment Report of 2005 that little research work is done about this critical ecosystem compared with forest ecosystems (Amler, Schmidt and Menz, 2015).

While efforts directed at monitoring, conservation and management of wetland ecosystems exist at varying scales, these are hampered by the scarcity of accurate, reliable and up-to-date spatial information about wetland ecosystems in some parts of the world particularly in Africa (Adam et al. 2010). For example, at the international level, wetland ecosystems conservation measures have culminated in the demand for large integrated monitoring and reporting frameworks such as the Ramsar Convention of 1971. This Convention aims at wise use of wetlands with emphasis on the wetland ecosystems and protection of Red Data species within this ecosystem (Dixon, et al. 2016). Moreover, wetland ecosystems functions are critical in supporting eleven of the seventeen United Nations Sustainable Development Goals (UN-SDG) also known as Agenda 2030. Other efforts include the Strategic Plan for Biodiversity 2011 – 2020, Post 2020 Global Biodiversity Framework of the Convention on Biological Diversity (Rebelo et al, 2018). The success in meeting these targets hinges on a thorough understanding of the current and emerging pressures on wetland ecosystems and their various components through a robust suite of monitoring strategies.

One of the critical components of wetland ecosystems is their plants. They are critical in the development of nutrient cycles, pollutants filtration, carbon storage and nitrogen cycles (Jiao, et al. 2019; Munoz, Cissell and Moftakhari, 2019; Amani, et al. 2018). In essence, they are important in carbon sequestration since they have been reported to contain about 12% of the global carbon pool, thus assisting in combating climate change (Guo, et al. 2017; Samiappan, et al. 2017). Wetland plants purify water, thus improving its quality. They retard surface water run-off velocity allowing industrial and agricultural pollutants to settle in wetlands soils where they are taken up in wetland plant tissues before they reach aquatic ecosystems (Jiao, et al. 2019; Munoz, Cissell and Moftakhari, 2019). Moreover, wetland plants play a critical role in storm surge dissipation, attenuating floods and mitigating soil erosion (Munoz, Cissell and Moftakhari, 2019; Zhou, Wong and Zhao, 2018). They recharge ground water and store water like sponges and play a crucial role in alleviating effects of drought (Ramsey and Rangoonwala, 2012; Adam, Rugege and Mutanga, 2010). Wetland plants are a habitat for animal and other plants species. Thus, wetland plants species play an important ecological role in biodiversity assessment and monitoring, hazard and stress monitoring and management, monitoring of invasive species, wildlife habitat assessment and are a surrogate measure of the healthy status of the wetland ecosystem

(Sun, et al. 2019). Given the above, robust and accurate methods of assessing the distribution, quality and quantity of wetland plant species in a timely manner at different spatio-temporal scales is mandatory.

Traditionally, mapping of wetland plant species was achieved via a combination of point source field surveys, aerial image interpretation, taxonomical and visual estimation (Adam et al. 2010; Taddeo, Dronova and Depsky, 2019). A collection of wetland plant species information using these techniques is challenging due to terrain challenges, inaccessibility and the presence of dangerous aquatic animals (Jiao, et al. 2019; Mahdianpari, et al. 2017). Consequently, these traditional methods are impractical, labour intensive, expensive and time consuming (Taddeo, Dronova and Depsky, 2019). Accordingly, scientists advocate for contemporary earth observation technologies such as spaceborne remote sensing.

Satellite Earth Observation (EO) technologies offer a lucrative and reliable alternative primary source of spatial data for wetland plant species mapping with several advantages compared with the conventional methods (Siebrits and Gasela, 2019). Satellite EO data are relatively cheap, offer synoptic coverage and comes in digital format which can be integrated with ancillary data within a geographic information system (GIS) environment. (Mather and Koch, 2011). They are a great source of spatial information on landscapes that are remote and inaccessible and considerably vast to survey using field-based techniques. Remote sensing provides consistent repeat coverage at relatively frequent intervals. Their data is delivered at varying temporal, spatial and spectral resolutions with ample opportunities for analyzing change over time (Rapinel, et al. 2019; Ayanlade, 2017; Stratoulias, et al. 2018).

In remote sensing, spectral, temporal and spatial resolutions are critical technical considerations, particularly in vegetation studies. Within the optical remote sensing domain, there are multispectral and hyperspectral sensors. They both offer remote sensing data in the visible, near infrared and shortwave channels of the electromagnetic spectrum but at varying spectral resolution. While multispectral sensors have a few broad spectral channels with spectral resolutions typically over 100 nm, the hyperspectral sensors have many narrow spectral channels (over 100) with a spectral resolution less than 10 nm.

Currently, most vegetation studies are conducted using freely available multispectral sensors, such as Moderate Resolution Imaging Spectroradiometer (MODIS), Landsat Thematic Mapper (Landsat TM), Landsat Enhanced Thematic Mapper (Landsat ETM, Landsat 8 Operational Land Imager (Landsat 8 OLI) and Sentinel-2 Multispectral Imager (MSI) (Slagter, et al. 2020; Bhatnagar, et al. 2020; Rapinel, et al. 2019; Richter, et al. 2016). However, their wide spectral bandwidths impede their capability for discerning subtle differences among individual plant species, hence sensors with improved spectral resolution are desirable (Adam, Mutanga and Rugege, 2010).

The need for improved sensor capability has led to the emergence of spaceborne hyperspectral imaging (Rast and Painter, 2019). Hyperspectral sensing also known as imaging spectroscopy entails quantitative sensing of the earth surface targets in the optical range with contiguous, spectral sampling from visible (VIS) to the shortwave infrared (SWIR) portion of the electromagnetic spectrum (Verrelst, et al. 2018; Rast and Painter, 2019; Lanaras, Baltsavias and Schindler, 2017). Spaceborne hyperspectral remote sensing sensors are proving to have a great potential in wetland plant species discrimination owing to their narrowband spectral characteristics and improved signal to noise ratio (Stratoulias, et al. 2018: Ballanti, et al. 2016). However, imaging spectroscopy data is scarce since most of the hyperspectral sensors are still under construction. For example, the National Aeronautics and Space Administration (NASA) is working on the Surface Biology and Geology (SGB) replacing the concept of Hyperspectral Infrared Imager (HyspIRI) (Green, et al. 2013) and Germany is building the DLR Earth Sensing Imaging Spectrometer (DESIS) (Heiden, et al. 2017) and EnMap (Gaunter, et al. 2015). Also, India is manufacturing GISAT (Misra, 2017), and Space Advisory Company (SCS) is building the nSight-2 in South Africa. Before their launch, these need to be tested for their capability and utility for various applications. Their spectral band settings need to be tested, evaluated and their performance needs to be benchmarked against the existing new generation of sensors like Worldview 2 and other forthcoming hyperspectral sensors like Environmental Monitoring and Analysis Programme (EnMap) that are also used in vegetation studies particularly in the wetland plant species.

Despite their vast potential for many environmental applications, the classification of hyperspectral data is challenging. Classification of plant species using remotely sensed hyperspectral data is affected by multidimensionality, multi-modality and multicollinearity. Moreover, numerous spectral bands which do not add value creates redundancy (Hennessy, Clarke and Lewis, 2020). Specifically, the addition of dimensions when keeping a small training sample results in a decrease in the classification accuracy (Hughes, 1968). This Hughes effect or 'the curse of dimensionality" causes instability in parameter estimates which increases the generalization errors by the classifier. For example, parametric classifiers such as Maximum Likelihood (ML) have yielded low classification results since data over-fits and hence they are unable to estimate mean and covariance statistics from multidimensional data (Lim, Kim and Jin, 2019; Zafari, Zurita-Milla and Izquierdo-Verdiguier, 2019; Raczko and Zagajewski, 2017). Consequently, machine learning algorithms, such as Artificial Neural Networks (ANN), Support Vector Machine (SVM), Import Vector Machines (IVM), Gradient Boosting Algorithm (GBA), Random Forest (RF), Decision Trees (DT) and Partial Least Squares-Discriminant Analysis (PLS-DA) have become instrumental in the automated digital analysis of hyperspectral Earth observation data (Bhatnagar, et al. 2020; Slagter, et al. 2020).

Among these machine learning algorithms, RF and SVM have gained prominence in the classification of hyperspectral data (Transon, et al. 2018). RF requires less time to classify hyperspectral data (Hastie, et al. 2001), while SVM has a capability of handling high dimensionality data, eliminating the challenge of Hughes effect or "curse of dimensionality" (Mountrakis, Im and Ogole, 2011) with a minimal training data (Stratoulias, et al. 2018). On the other hand, PLS-DA which is a powerful multivariate supervised pattern recognition method has a great potential in the classification of hyperspectral data owing capability, in noise reduction, to show probability of a sample belonging to the class being modelled and in the selection of variables, although it has been limitedly used in remote sensing. The previous performance of SVM, RF and PLS-DA classifiers, indicate their potential as ideal candidates for classification of nSight-2 data and accurately discriminating wetland plant species. Moreover, to the best of our knowledge, no study has ever compared their performances using the resampled spectral bands of the nSight-2 satellite sensor.

1.1.2 Research Problem

Discrimination of wetland plant species is a challenging task, since their spectral signatures similar for different plant species yet may be different for similar plant species owing to the influence of soil and plant moisture and dead plant matter. As such there is a search for improved sensors spectral configurations with a capability of using biochemical and biophysical properties of plants such as chlorophyll content, nitrogen, carbon, phosphorus, water content, carotene, xanthophyll, leaf area index, leaf dry biomass, leaf orientation in order to discern among the subtle differences among wetland plant species. Existing satellite earth observation sensors which are predominantly multispectral, such as Satellite Pour l'Observation de la Terre (SPOT 5), Sentinel 2 MSI and Worldview-2 have broad spectral bands that limit their performance on such tasks, yet accurate, reliable and up-to-date spatial information about plant species is needed as it provides useful insights for wetland ecosystem management. While hyperspectral sensors have shown potential in effectively discriminating among plant species (Adam, et al. 2012; Rast and Painter, 2019), there is a limited number of spaceborne hyperspectral sensors in orbit. Many are lined up for design and launch, however, they need to be tested for their spectral settings capability. One of the planned hyperspectral sensors is the nSight-2. Its spectral bands have not been tested for their robustness in classifying plant species. Furthermore, the potential use of its spectral bands in mapping and monitoring terrestrial vegetation is also not yet known. Classification of hyperspectral data is a challenge since it is characterized by multi-dimensionality and multi-collinearity. Machine learning algorithms have shown a better capability in handling such data. However, none of the machine learning statistical classifiers has been proven to produce higher classification accuracies for all environments. In this study, Random Forest, Support Vector Machine and Partial Least Squares-Discriminant Analysis were compared and used to test the capability of nSight-2 spectral bands in discriminating plant species.

1.1.3 Aims and Objectives

1.1.3.1 Aim of the dissertation

The study aims to evaluate the utility of the forthcoming hyperspectral nSight-2 satellite sensor in classifying vegetation species in a conserved wetland nature reserve using robust machine learning classifiers.

1.1.3.2 Objectives of the dissertation

- To compare the performance of the spectral configuration of nSight-2 with those of EnMap and WorldView-2 in classifying wetland plant.
- To test the capability of the proposed nSight-2 spectral bands for discriminating wetland vegetation species using Random Forest (RF), Support Vector Machine (SVM) and Partial Least Squares-Discriminant Analysis (PLS-DA) classifiers in discriminating wetland vegetation species.

1.1.4 Scope of the study

Upcoming satellites need to be tested for their capabilities in various applications to inform further design and development prior to launch. This study intends to evaluate the performance of the upcoming n-sight 2 satellite in wetland plant species classification by comparing its spectral bands settings with those of the existing new generation of multispectral sensors, i.e. WorldView-2 which contain novel spectral bands such as the red-edge and have been proven important in plant species identification using the Random Forest classifier. The spectroscopy measurements over dominant wetland plant species were resampled to the spectral bands of the upcoming hyperspectral sensors, nSight-2 and EnMap and the existing multispectral, WorldView-2 earth observation sensors. The measured vegetation spectra will also be used to compare the performances of the robust machine learning algorithms, i.e. RF, SVM and PLS-DA.

1.1.5 Study area

The study was conducted in Verloren Vallei Nature Reserve (VVNR), i.e., an international Ramsar environmentally protected subtropical inland grassland and wetland ecosystem, located 15 km north of Dullstroom, Mpumalanga province, South Africa (Mpumalanga Tourism and Parks Agency, 2020) (Figure 1). The study area measures approximately 6000 ha and consists of more than 30 different wetlands of all types. It received its status

of International Importance on 16 October 2001. The topography is undulating, characterized by hills and rocky outcrops, with an elevation above 2000 m, making it one of the highest and coldest sites in South Africa (Mpumalanga Tourism and Parks Agency, 2020). The area is characterized by an annual rainfall of more than 800 mm, and the average temperature ranges from –13ºC in winter to 29ºC in summer. Like most of the global wetlands, the Verloren Vallei Nature Reserve (VNR) is threatened while it is critical in provision of essential ecosystem services such as hydrological regulation, carbon sequestration at the same supporting crucial biodiversity of several red data plants, birds and animal species surviving in the land-water interface. There are limited studies which have targeted this nature reserve to aid its effective management.

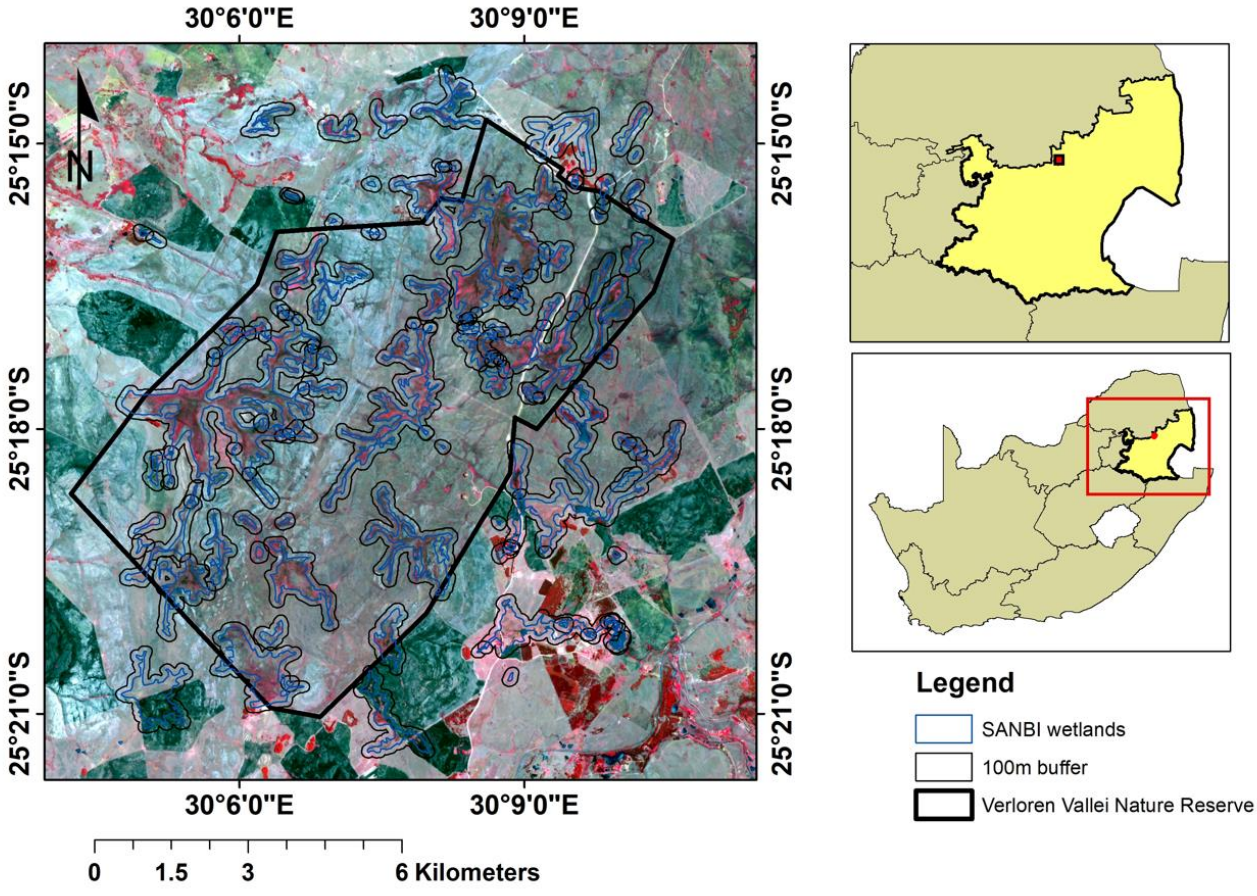

Figure 1.1 Map of the study area showing location of study area in the context of Mpumalanga province and South Africa, the extent of the wetlands (shown in blue) from the South African Biodiversity Institute (SANBI) and a 100 m buffer used for sampling.

1.1.6 Dissertation outline

Chapter 1: General Introduction

This chapter introduces the research study by highlighting the importance of wetlands, their ecological, social and economic functions and value. It also highlights the constraints of the lack of spatial data for the implementation of conservation and management strategies. The chapter also specifies the aim, objectives, scope of study, research problem and outlines the layout of chapters of this dissertation.

Chapter 2: Overview of Literature

This chapter discusses the issues of remote sensing of plant species identification in a wetland environment. Various optical remote sensing datasets, i.e. multispectral-moderate resolution spatial/spectral resolution, high spatial resolution and high temporal resolution sensors and the potential of hyperspectral sensors commonly used for plant species identification were discussed. Performances of SVM, RF and PLS-DA classification algorithms were compared in plant species identification. It also shows that this research intended to address the gap in understanding wetland plant species classification using n-Sight-2 hyperspectral satellite sensor.

Chapter 3: Methodology

This section describes the methodology used in this work. Firstly, it describes the resampling of laboratory-acquired spectral measurement of various wetland plant species to nSight-2, EnMap and Worldview-2 spectral band settings. Secondly, machine learning algorithms i.e. RF, SVM and PLS-DA are described.

Chapter 4: Results

This chapter presents the findings. It shows the strength of spectral bands of nSight-2 in discriminating plant species in a wetland environment. This chapter also presents the results of the comparison in performances among the upcoming space-borne hyperspectral sensors, i.e. nSight-2 and the new generation of multispectral sensor i.e. WorldView-2. These results are presented graphically using the resampled spectral data and SVM, RF and PLS-DA.

Chapter 5: Discussions, Conclusions and Recommendations

This chapter discusses the findings in relation to the performance among sensors and algorithms. It also closes the research, by highlighting major aspects of the results and gaps that could not be addressed in the current study and suggests the direction for future research work.

CHAPTER TWO

2. OVERVIEW OF LITERATURE

2.1 Introduction

Preservation, conservation and sustainable management of wetland ecosystems have become important due to the critical role they play in providing essential services and goods (Amani, et al. 2017). Such programmes hinge on the availability of accurate, reliable and up-to-date information on their condition and the status of their components. The main focus of these programmes should endeavour to monitor of wetland plant species composition and changes over time, their spatial distribution condition. Consequently, such information is needed hence the development of methods and techniques to acquire and deliver it timeously. Traditional methods and techniques have been proven ineffective in this regard since they are time-consuming, expensive and limited in spatial extent. An important source of this information that would target the wetland plant species is remote sensing. In this chapter, an overview of literature related to optical remote sensing of wetland plant species is presented. An overview of South African wetland classes will be given followed by a description of optical remote sensing datasets commonly used in wetland plant species discrimination. After this, machine learning classification techniques are presented, wherein, their advantages and challenges in their use are discussed.

2.2 An Overview of South African Wetlands

Wetlands are transitional land between aquatic and terrestrial systems and are generally classified as perennial or intermittent due to the moisture (Ollis, et al. 2013). In South Africa, there are three classes of wetlands, i.e. marine wetlands, estuarine wetlands and inland wetlands. Ollis, et al. (2013) further describes these marine wetlands as a part of the open ocean overlying the continental shelf and associated with coastline, while estuarine is a body of surface water that is part of a water course that is permanently/periodically open to the sea. Inland wetland is not linked to the ocean and is not influenced by tidal and marine activities. Of these classes of wetlands, twenty-three have been declared wetlands of international importance covering 557 028 ha (Malherbe, 2018). They are important hubs for biodiversity and perform important ecological roles

such as flood attenuation, filtering water and carbon sequestration. Major threats to South African wetlands are urbanization, mining, veld fires, pollution and draining and conversion to agricultural land use (Mitchell, 2013).

2.3 Drivers of Spectral Variability of Wetland Vegetation

Wetlands are characterized by the high rate of biological activity, strong natural selection pressures and diverse aquatic environments resulting in many species of plants (Xu, et al. 2019). Mapping and monitoring plant species for sustainable management of wetlands using remote sensing is a challenge, yet it is the most viable source of data. Remote sensing endeavours to extract information about the condition of an object without physical contact through capturing of reflectance measurements of the electromagnetic energy (Mather and Koch, 2011). However, spectral reflectance measurements of vegetation at plant species-level in wetland landscapes captured in this manner are a challenge (Adam, et al. 2010). According to Adam et al. (2010), it is difficult to accurately create a clear boundary among vegetation communities due to the high spectral and high spatial variability of these vegetation communities. This is caused by sharp ecological gradients that create short ecotones (Adam and Mutanga, 2009). Also, it is difficult to get pure plant species spectra since spectral reflectance in wetlands is a combination of spectra from underlying wet soils, water and plant species. A search for remote sensing techniques with the capability of detecting subtle differences among wetland plant species spectra is on-going.

High spatial heterogeneity and temporal dynamics, resulting from seasonal and daily changes in water levels make spectral signatures of wetlands highly dynamic (Ludwig et al., 2019; Adam, Mutanga and Rugege, 2010). Also, Amani, et al. (2017) observed that in a wetland environment, spectral signatures of the same plant species can vary within and between the years. Moreover, the spectral properties of plants are an outcome of the influence of biochemical and physical properties of plants such as chlorophyll content, nitrogen, carbon, phosphorus, water content, carotene, xanthophyll, leaf area index, leaf dry biomass, leaf orientation and soil background, thus it tends to be difficult to use the same properties to differentiate among plant species. Specifically, the same plant species can give different spectral signatures owing to the seasonal and daily changes in water

levels, while on the other hand, some different plant species will be spectrally similar to differentiate (Amani, et al. 2017). Dead plant matter also attenuates the spectral signal of vegetation, while inundation may affect plant signal in the red and near-infrared regions of the electromagnetic spectrum (Dronova and Tadeo, 2016). For accurate mapping of plant species under such conditions, it is important to search for appropriate remotely sensed datasets with the capability of discerning subtle differences among plant species.

Other environmental factors that influence spectral variability among wetland vegetation include plant phenology, shadows from tall objects, hydrological regime, cloud cover and length of the growing season (Tiner, 2015). Phenological stages of plant exhibit varying reflectance, absorbance and transmittance of light solar radiation. Tiner (2015) argues that during spring, leaf-off and wet season are the best phenological stages for the remote sensing wetland plants. Specifically, inundation may affect plant signal in the red and near infrared regions of the electromagnetic spectrum recorded remotely by satellite sensors (Dronova and Tadeo, 2016). Again, cloud cover, tall objects like trees and buildings also impact on the spectral variability in a wetland ecosystem (Tiner, 2015). Accordingly, it is difficult to accurately create a clear boundary among vegetation communities due to the high spectral and high spatial variability of these vegetation communities (Adam et al. 2010). Also, dead plant matter attenuates the spectral signal of vegetation. Despite these challenges, satellite EO remains a viable alternative in vegetation mapping and monitoring. Given the above, there is a lack of space-borne satellites which possess sensors with fine spectral resolution suitable for this task.

2.3 Commonly used optical datasets and their characteristics

Optical satellite remote sensing provides data from various parts of the electromagnetic spectrum including the visible through near-infrared to shortwave infrared (Amani et al. 2018; Adam et al. 2012). These can be grouped as multispectral sensors, (i. e. low, moderate and high-resolution sensors) and hyperspectral sensors (characterized by many narrow bands, usually over a hundred and improved spectral resolution of less than 10 nm. (Table 2.2 shows the differences between the multispectral and hyperspectral satellite sensors) (Mather and Koch, 2011).

Table 2. 1 Characteristics of commonly (potentially) used data for wetland assessment

	Data	Spatial Resolution	Spectral Coverage	Temporal Resolution
High Temporal	MODIS	250m – 1000m	36 bands	1 – 2 days
	AVHRR	1000m	6	daily
	MERIS	1150	15	3
	Sentinel 3	300m	21	1.9 days
Medium Resolution Spatial/Spectral	Landsat	30	9	16 days
	SPOT	5 -30	5	1 - 5 days
	Sentinel 2	10 – 60m	13	5 days
	ASTER	15 – 90 m	15	16 days
High Resolution	QuickBird	2.44m	5	1 – 3 days
	Ikonos	3.2m	5	1 – 3 days
	RapidEye	5	5	1 – 3 days
	WorldView 1 WorldView 2 WorldView 3 WorldView 4	0.5 m 0.50 m 0.30 m 0.30 m	1 8 8 5	1.7 days 1.1 days 1 day 4.5 days
	OrbView	1.65m	4	1 – 3 days
Hyperspectral	Hyperion 1	30m	220	16 days

2.3.1 Medium to High-Resolution Multispectral Sensors

Moderate to high resolution multispectral sensors provides remotely sensing data with a spectral resolution of less than 10 spectral channels and a spatial resolution of between 10m and 100m. Examples of these include ASTER, Landsat TM, Landsat ETM+, SPOT 1 – 5, Sentinel 2A and 2B. Literature shows that these satellite sensors have a capability in

vegetation classification. For instance, Wright and Gant (2007) used Landsat TM to model palustrine wetland in Yellowstone National Park with an overall accuracy of 86%. The same authors also used SPOT 5 imagery to classify tree species in aquatic marsh vegetation in Southern France and produced an overall accuracy of 80%. Among the attractiveness of these multispectral moderate to high resolution is their wide swath width, for example, Sentinel 2 has a wide swath width of 290km enabling regional assessments (Slagter, et al. 2020). Bhatnagar, et al. (2020) used Sentinel 2 MSI in mapping vegetation communities inside wetlands, in Ireland, achieving an overall accuracy of 87%. Sentinel MSI offers data for free and also carries important novel red-edge wavebands that are effective in vegetation species discrimination. However, most of the multispectral sensors are hindered by their broad spectral bands in plant species discrimination. Satellite sensors with an improved spectral resolution such as n-Sight 2 hyperspectral sensors offer a great potential in plant species differentiation. Nevertheless, the performance of n-Sight 2 sensor versus these multispectral moderate to high sensors is still not known since no study has compared their performances.

2.3.2 Very High Resolution Multispectral Sensors

Another set of remote sensing data that has become attractive for plant species identification is the very high spatial resolution sensors, such as WorldView-2, Ikonos, OrbView, QuickBird and RapidEye. They are attractive to the remote sensing community for plant species classification owing to their very high spatial resolution of less than 5m with some dropping to sub-metre, which enables spatial detail. Also, these high spatial resolution sensors have a high temporal resolution of between 1 – 4 days such as Worldview 2, suitable for plant species mapping in wetland environments since their reflectances are highly dynamic. Literature reveals its use in plant species classification. For instance, Mahdavi, et al. (2019) employed RapidEye satellite data to map vegetation in spectrally similar wetlands in Newfoundland and Labrador in Canada and yielded an overall accuracy of 93%. Berhane, et al. (2018) used WorldView 2 for wetland inventorying, in Selenga River Delta of Lake Baikal in Russia and achieved an overall accuracy of 81%. However, these are commercial remote sensing datasets are costly and prohibit their use in resource-constrained regions like Africa. Moreover, they have a limited number of spectral channels usually less than 5. Despite these drawbacks, a new

generation of high spatial resolution sensors like WorldView-2 boast of very high spatial resolution and an important red-edge waveband that has become prominent in the discrimination of vegetation species in different environments (Mutanga, Adam and Cho, 2012).

2.3.3 High temporal/Low Spatial Resolution

Another group of optical remote sensing satellite sensors include the National Oceanic and Atmospheric Administration (NOAA) Advanced Very High-Resolution Radiometer (AVHRR), Moderate Resolution Imaging Spectroradiometer (MODIS), Medium Resolution Imaging Spectrometer (MERIS), Environmental Satellite (Envisat), Sentinel 3. These sensors have a short revisit time of 1 – 2 days and their spectral channels amount to just under 40 bands. The spatial resolution is over 250m to over a kilometre. Their strength lies in the short revisit time and wide swath width that enables tracking of quick vegetation changes at regional to global scale. Moreover, they have produced considerable results in vegetation mapping and monitoring across the globe. For instance, Marcheti et al. (2016) used MODIS time series of the spectral vegetation indices in Middle Parana River floodplains to monitor vegetation, while De Almeida et al. (2015) also used these MODIS time series in Pantanal macro system for monitoring vegetation. Similarly, AVHRR data has been used in combination with the spectral indices in mapping vegetation attributes. However, their spatial resolution limits their applicability on wetland vegetation monitoring at a local scale. Also, they have few broad spectral bands and cannot discern subtle differences among plant species.

2.3.4 Hyperspectral sensors

Limitations of the broadband multispectral sensors in spectrally discriminating subtle differences among individual plant species have prompted the demand for technological advancement in sensor capability. This has resulted in the debut of hyperspectral sensors that contain multiple contiguous narrow bands with the capability of providing remotely sensed data with a significant level spectral detail (Hill, Buddenbaum and Townsend, 2019). A considerable body of literature demonstrates use of hyperspectral sensors in plant species classification, proving their capability in producing invaluable plant species maps. For instance, Ballanti, et al. (2016) used hyperspectral data acquired using an Airborne

Imaging Spectrometer for Applications (AISA) Eagle sensor in conjunction with SVM and RF machine learning algorithms to classify tree plant species in heterogeneous forests of Muir Woods National Monument and Kent Creek in Marin County, California to obtain an overall accuracy of 90%. In another study, Lim, Kim and Jin (2019) applied Hyperion spaceborne archive data to classify tree species in South Korea and China. They produced impressive results of 99% and 97% of RF and SVM classifiers, respectively. Stratoulias, et al. (2018) also used airborne hyperspectral data to classify emergent wetland vegetation at Lake Balaton, Hungary and concluded that de-noised hyperspectral information in the visible and near infrared bands (400 – 1000nm) can accurately map vegetation in a wetland. Despite the great potential exhibited by the hyperspectral sensors, most of these datasets are derived from airborne campaigns that have been proven to be expensive for resource-starved regions of Africa. Use of handheld spectrometers that are limited in spatial extent and are time consuming make the situation worse. Moreover, these campaigns are subject to airspace regulatory imperatives. Again, since the decommissioning of Hyperion on-board Earth Observer 1 in 2017, there has been a few spaceborne hyperspectral sensors in orbit. Spaceborne hyperspectral sensors are relatively cheaper than airborne hyperspectral sensors. A lot of these spaceborne sensors are lined up for manufacture and launch. Their spectral bands configuration needs to be tested for capability and this study therefore, attempts to close this knowledge gap.

2.4 Spectral bands used for vegetation mapping and monitoring

Satellite EO provides a detailed earth surface information through capturing electromagnetic energy measurements of vegetation species' biochemical and biophysical properties (Jensen, 1983). The interaction of electromagnetic radiation with plants, i.e. reflectance, absorption and transmittance depends on these properties. For example, carotenoids, xanthophyll, anthocyanin and chlorophyll pigments are absorbed in the blue region (i.e. 450 – 520nm) and the red region (i.e. 600 – 630nm) while the mesophyll cells show high reflectance in the near-infrared region (Im and Jensen, 2008) (see Figure 2.1).

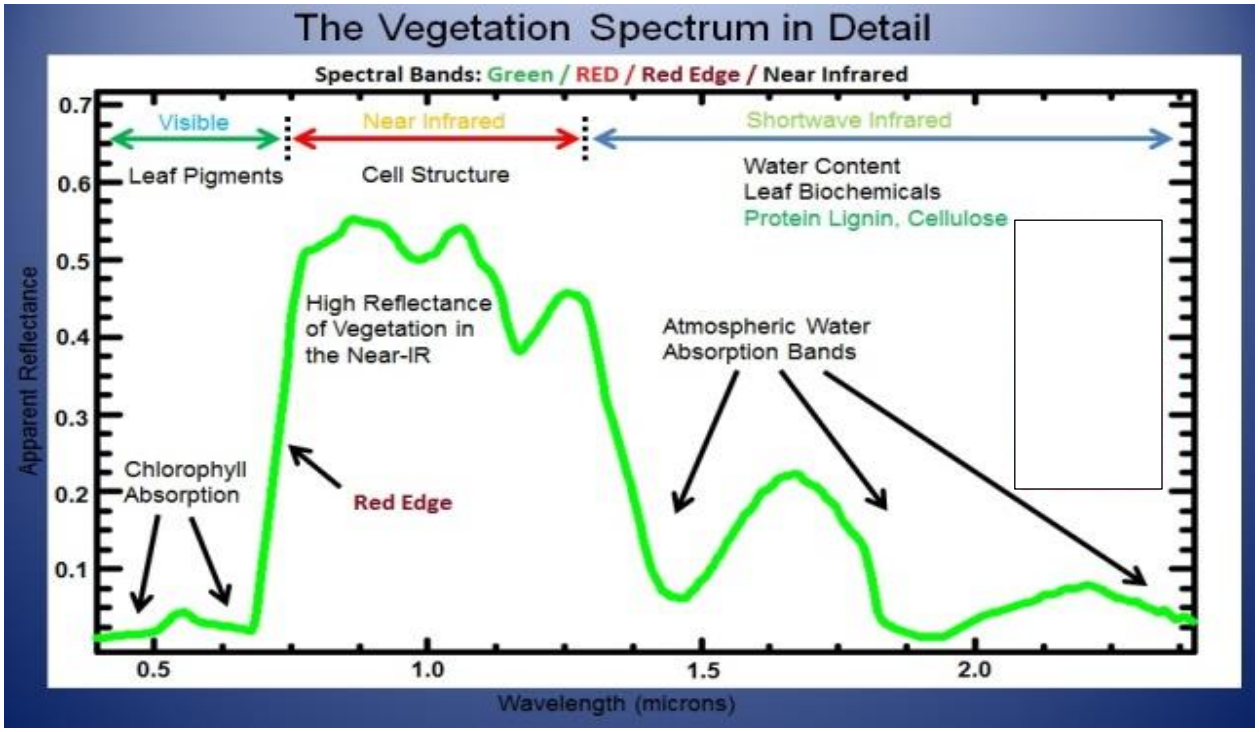

Figure 2. 1 Adopted from: Corrigan, F. (2019) Multispectral Imaging Camera Drones in Farming Yield Big Benefits

Plant species' leaves interact with incoming radiation differently owing to biochemical and biophysical make-up, creating a spectral footprint that has become critical in discriminating among them in the four regions of the electromagnetic spectrum (Table 2.2). The behavior of solar radiation in these four portions of the electromagnetic spectrum upon interaction with different biochemical components of plants, creates spectral signatures which are important for plant species discrimination. For instance, there is low reflectance and transmittance of visible light due to chlorophyll and carotene absorption, while in the near-infrared there is high reflectance and transmittance and low absorption owing to the internal leaf structures (Rosso, et al. 2005; Kumar, et al. 2001). However, inconsistencies on the shape of these spectral signature curves make research on which portions of the electromagnetic spectrum will give optimal plant separability on going.

Table 2. 2 The spectral reflectance of green vegetation on the four regions of the electromagnetic spectrum according to Kumar, et al. (2001)

Spectral Channels	Description	Spectral Reflectance of Vegetation
400 nm – 700 nm	Visible	Low reflectance and transmittance due to chlorophyll and carotene absorption
680 nm – 750 nm	Red-edge	The reflectance is strongly correlated with plant biochemical and biophysical parameters
700 nm – 1300 nm	Near-infrared	High reflectance and transmittance, very low absorption. The physical control is internal leaf structures
1300 nm – 2500 nm	Mid-infrared	Lower reflectance than other spectrum regions due to strong water absorption and minor absorption of biochemical content

Previous studies (Kumar, et al. 2001; Aneece and Epstein, 2017) indicate that VIS, NIR AND SWIR are the most important portions of the electromagnetic wavebands important for plant species studies. Leaf level differences in pigments, water content, cell size, intercellular space, cell wall thickness and other biochemical and structural elements assist in differentiating species at the leaf level (Aneece and Epstein, 2017). The reflectances in the visible (VIS), near infrared (NIR) and short wave infrared (SWIR) regions are influenced by pigment content, leaf structure and water content and structural compound content respectively. For instance, in the VIS region, Schmidt and Skidmore (2003) proved the blue absorption maximum at 404nm usefulness in species identification in coastal saltmarshes in Netherlands, while Thenkabail (2014) appraised the 375nm band for estimating the fraction of photosynthetically active radiation and leaf water content. The 500nm – 549nm wavebands were used by Hanson et al. (2018) to discriminate plant species in a forest in Panama and the 501nm in classifying coastal salt marshes.

In a study by Masaitis, Mozgeris and Augustaitis (2013) the blue and NIR regions were found to be important for deciduous tree species discrimination at the onset of greening phenological stage. At the peak of the season, it was found that Red and SWIR regions were critical in differentiating among the deciduous tree species. Roth et al. (2015) argue that the most useful regions of the electromagnetic spectrum for plant species discrimination are the visible and NIR since most of the variables that influence i.e. chlorophyll a and b are concentrated in these regions. Others, (Peerbhay, Mutanga and Ismail, 2013; Fassnach, et al. 2016) have demonstrated the importance of the SWIR region (1000 – 2500nm) in plant species discrimination. Specifically, Adam and Mutanga (2009) showed the benefit of using the SWIR in the discrimination of *papyrus sp.* in a coastal wetland in South Africa.

Several research works (Mutanga, Adam and Cho, 2012; Adjorlolo, Mutanga and Cho 2014) have demonstrated the usefulness of the red-edge band in vegetation studies as an important input variable due to its sensitivity to biochemical and biophysical parameters. For example, Otunga et al. (2019) and Adjorlolo, Mutanga and Cho (2013) used the red-edge to discriminate grasses. In another study, Sibanda, Mutanga and Rouget (2017) tested the capabilities of WorldView 3's red-edge spectral band in discriminating and mapping complex grassland management treatments in KwaZulu-Natal, South Africa, and concluded that it was important for improving the overall accuracy from 67% to 70%. However, the red-edge of WorldView 2 sensor considered an unimportant input variable by Pu and Cheng, (2015) in vegetation mapping. As argued by Zhu et al. (2017) these inconsistences warrant further research on the performances of spectral bands among different satellite sensors. However, concentrations of individual biochemical components may vary producing distinct spectral signature curves (Ustin and Jacquemoud, 2020).

Vegetation spectral measurements in wetland environments tend to be obfuscated by canopy geometry, soil background, sun view angles and atmospheric conditions (Lefebvre et al. 2019). As such, use of traditional sensors like Landsat, MODIS tended to saturate leading to poor classification accuracies (Adam et al. 2010). Thus, a lot of spectral indices, such as Normalized Difference Vegetation Index, Ratio Vegetation Index, Soil Adjusted vegetation Index and Enhanced Vegetation Index were developed to improve accuracies

in vegetation mapping (Xue and Su, 2017). Precisely, most of these vegetation indices are based on the VIS and NIR spectral bands. The broadband spectral vegetation indices like Normalized Difference Vegetation Index (NDVI) that utilize the red and near infrared wavebands, tend to saturate with heavy vegetation cover and are thus not effective in discriminating vegetation species in wetlands. Spectral indices derived from the narrow wavebands are instead preferred. For example, Elvidge and Chen (1995) used various combinations in the red and near infrared wavebands of Landsat TM and PS-2 Analytical Spectral Device (ASD) to compare the performance of broadband and narrow band vegetation indices in Nevada, United States of America and concluded that narrow band vegetation indices were more effective in estimating Leaf Area Index and green cover. In another study, Adam et al. (2014) compared the narrow band Enhanced Vegetation Index (nEVI) and narrow band Normalized Difference Vegetation Index (nNDVI) from different wavebands combinations involving the red-edge waveband in iSimangaliso Wetland Park in South Africa to predict papyrus above ground biomass and concluded that nEVI outperformed nNDVI. Similarly, Mutanga and Adam (20011) compared the performance of imaging spectroscopy vegetation indices (narrow Normalized Different Vegetation Index (NDVI), Simple Ratio (SR) and Transformed Vegetation Index (TVI)) using all possible waveband combinations between 350nm and 2500nm to estimate Blue Buffalo Grass biomass in Southern Africa and found that they performed better than the commonly used broadband indices and raw spectral bands. In the same study they concluded that the SR outperformed the nNDVI and nTVI and recommend application of SR in dense vegetation cover. From the preceding, it can be learnt that the spectral indices derived from VIS and NIR cannot be effectively used in wetland plant species discrimination, as such there is need to evaluate spectral indices derived from the spectral bands important for vegetation mapping such as the RE and SWIR.

2.5 The performance of MLA in plant species discrimination

2.5.1 Machine Learning Algorithms

Machine Learning Algorithms (MLA) refers to statistical methods that apply artificial intelligence in to analyse information, recognise patterns and improve prediction accuracy

through automated repeated learning from training data (Maxwell, Warner and Fang, 2018; Rogan *et al.*, 2008). The interest in MLAs is based on the following:

- They can deal with multi-modal, noisy and missing data, since they are non-parametric (Hastie *et al*., 2001)
- They are able to deal with large and complex data measurement spaces within reduced computational demand (Foody, 2004)
- They are accommodative to both categorical and continuous ancillary data (Lawrence & Wright, 2001);
- Users are able to investigate the relative importance of input variables in terms of contribution to classification accuracy (Hansen, Dubayah and DeFries, 1996; Foody & Arora, 1997)
- Are flexible, and can be adapted to improve performance for particular problems (Lees & Ritman, 1991), and are able to accommodate multiple subcategories per response variable (Gopal, Woodcock & Strahler, 1999).

Commonly used MLA in plant species with optimal accuracies include SVM, RF, NN, DT and PLS-DA (Mountrakis, Im and Ogole, 2011). Among these machine learning algorithms, RF and SVM have gained prominence in the classification of hyperspectral data (Transon, et al. 2018). RF is an ensemble-based statistical machine learning algorithm that is robust in terms of using a few parameters N *tree* (number of trees) and m *try* (number of predictor features) hence it is fast in computing numerous spectral data channels such as hyperspectral data (Hastie, et al. 2001). On the other hand, SVM a kernel-based classifier has a capability of handling data with high dimensionality, eliminating the challenge of the Hughes effect (Camps-Valls and Bruzzone, 2005; Mountrakis, Im and Ogole, 2011) with a very limited number of training samples (Stratoulias, et al. 2018). PLS-DA is a discrimination and classification technique based on latent variables defined in a classical partial least squares regression approach with categorical response variables (Richter, 2016). Despite its limited use in remote sensing, PLS-DA, has been proven as one of the robust statistical approaches in dealing with multi-collinearity and imbalanced ratios between training samples and feature space (Peerbhay, Mutanga and Ismail, 2013). The above-mentioned strengths of these MLA makes them ideal classifiers for processing

hyperspectral data such as in discriminating wetland plant species. To the best of our knowledge, there is no study that has compared the performances of these MLA on the data resampled to n-Sight 2 satellite sensor's spectral band settings for wetland plant species classification.

2.5.2 Comparing performances of the MLA in plant species discrimination

From the available literature, the performances of these machine learning algorithms result in varying classification accuracies. For instance, Adam et al. (2014) compared the performances of RF and SVM in a heterogeneous South African coastal wetland using multispectral RapidEye data and found a comparative similarity in their performances. In another study, Ballanti et al. (2016) used hyperspectral imagery to compare the performances of SVM and RF and concluded that there was no significant difference between their performances since they both had overall accuracies over 90%. They further observed that they increased the training samples, SVM outperformed RF. In another study, Lim, Kim and Jin (2019) classified tree species using Sentinel 2 MSI and Hyperion in South Korea and China to compare MLA and found that SVM and RF had accuracies over 90% and concluded that these are well established in hyperspectral imagery classification for tree species classification. In a different environment, Peerbhay, Mutanga and Ismail (2013) conducted another study to test the PLS-DA in classifying commercial tree species in KwaZulu Natal. They found that PLS-DA can significantly discriminate tree species with an overall accuracy of 88.8% using AISA Eagle bands. Although this was in a different environment, their studies ascertains that the PLS-DA can be successfully used in with hyperspectral data in vegetation mapping and monitoring. Richter, et al. (2016) used airborne hyperspectral in a heterogenous mixed forest in Central European to compare performances of SVM, RF and PLS-DA in discriminating tree species, interestingly PLS-DA outperformed both SVM and RF. In another study, Raczko and Zagajewski (2017) compared the performances SVM, RF and ANN algorithms to classify tree species in the northwestern part of the Karkonosze National Park, Poland, south of Szklarska Poręba town and airborne hyperspectral APEX images. They concluded that ANN yielded 77% overall accuracy outperforming SVM and RF with 68% and 62% respectively.

Given the above, it can be concluded that MLA have a potential of being used in a wetland ecosystem with various hyperspectral data sets and produce optimal accuracies. However, to the best of our knowledge no study has compared the performances of RF, SVM and PLS-DA MLA on wetland plant species discrimination using remotely sensed data resampled to nSight 2 spectral band setting.

2.6 Summary of gaps in literature and way forward

In summary, the multispectral sensors with their broad bandwidths cannot effectively differentiate among the wetland plant species because of signal attenuation by underlying spectra. Lack of hyperspectral sensors with an improved capability is a challenge while these tend to be associated with exorbitant acquisition and computational cost to process. Also, to the best of our knowledge, there is no study that has compared the performances of WV2, EnMap and nSight 2 in the discrimination of wetland vegetation, let alone the using the advanced machine learning algorithms. Literature too, has shown that there is no classification algorithm that has been proven to be superior on various datasets and environments. In this regard, this study is an attempt to evaluate nSight 2 as an option in providing the hyperspectral dataset needed for assessing wetland vegetation attribute characterization.

CHAPTER THREE

3 METHODOLOGY

3.1 Experimental Design

3.1.1 Field Survey

A field survey was conducted between 14 and 18 December 2020. Thirty 40 m x 40 m plots were randomly selected across the study area and tagged with a GPS coordinate with ±3 m accuracy based on the distribution of wetlands, as guided by the ecologist and through use of their field guide booklet (van Ede, 2016). The plots were dominated by plant species: *Crocosmia paniculata*, *Agapanthaceae, Grasses* and *Cyperus sp* (Figure 3.1). The grid strategy was preferred where five sub-plots of 1 m x 1m representing homogeneous species were generated.

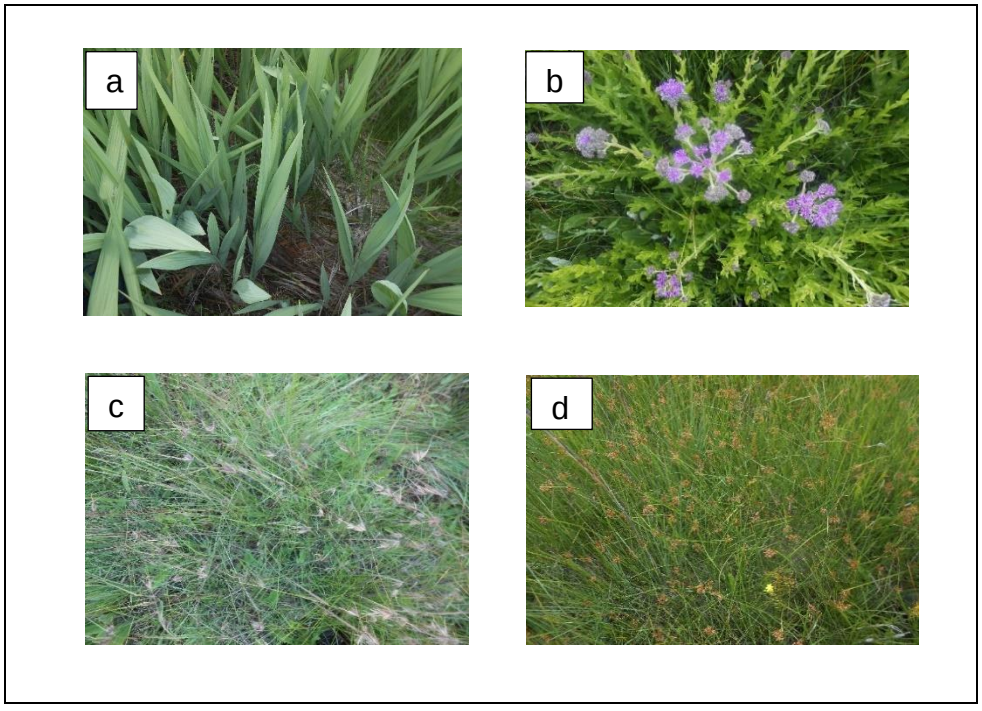

Figure 3. 1 pictures of the dominant plant species identified in the Verloren Vallei Wetland, a) *Crocosmia sp.* b) *Agapanthus sp*. c) *Grasses* and d) *Cyperus sp.*

Due to overcast conditions during fieldwork, we destructively collected leaves from the dominant plant species. We placed these leaves in sealed plastic bags and stored them

in a cooler box to preserve their quality (Masemola, Cho and Ramoelo, 2019). The leaf spectral reflectance measurements were acquired in the laboratory following the procedure by Masemola, Cho and Ramoelo (2019). A Spectral Evolution PSR-3500 spectrometer (Spectral Evolution, Inc. © 2014), with a spectral range of 350 nm – 2500 nm, spectral resolutions of 3.5 nm, 10 nm, and 7 nm at 350 nm – 1000 nm, 1500 nm and 2100 nm, respectively, was used. The spectral bands 350 nm – 1000 nm, at 1500 nm and at 2100 nm have nominal spectral sampling intervals of 1.5 nm, 3.8 nm and 2.5 nm, respectively (Kganyago et al. 2017). A bifurcated cable attached to a leaf-clip was used to produce light using Fibre Optic Illumination Module with 5 watt tungsten halogen source. This lamp produces a continuous spectrum of light in the range from near-ultraviolet to the infrared with a shift towards the blue. They produce light with higher effective colour temperature and high power efficiency. Halogen lamps are a light source with black-body radiation spectrum similar to that of the Sun. This set-up mimics the natural light. The experimental setting used a leaf-clip device that provided a direct-contact probe which prevented the influence of ambient effects. Spectral measurements were collected per leaf by repositioning the clip at five different positions for each scan after optimizing and calibrating using a white reference panel of approximately 100% reflectance. Poor quality measurements were discarded. We then averaged the collected spectra reflectance per plant species. Table 3.1 shows a summary of the spectral measurements collected.

Table 3. 1 Number of plant species spectra, training and testing samples per plant species.

Name of Plant Species	Number of Spectra collected	Training Data (70%)	Validation Data (30%)
Crocosmia sp.	210	147	63
Agapanthus sp.	210	147	63
Grasses	210	147	63
Cyperus sp.	210	147	63

3.2 Data Processing

3.2.1 Spectral sub-setting and noisy spectral removal

The measured spectral data had 2151 spectral bands sampled at an interval of 1 nm. Since nSight-2 has a spectral range of 400 nm – 900 nm, we resampled the data to match the nSight-2 spectral range. Consequently, commonly excluded spectral regions due to noise, i.e., below 400 nm, 1350 – 1465 nm, 1790 – 1960 nm, and 2350 – 2500 nm, were removed by default (Thenkabail et al. 2004).

3.2.2 Spectral resampling

The spectral reflectance samples measured in the 30 plots were then resampled to nSight-2, Worldview-2 and EnMAP spectral settings.

The nSight-2 is a forthcoming earth observation hyperspectral nano-satellite sensor from South Africa by the Space Advisory Company (SAC). This satellite sensor will have a spectral setting that ranges from 400 nm to 900 nm with 160 linear filtered and pre-selected spectral bands. Its ground sampling distance will be 20 m and will have a swath-width of 14 km. The spectral reflectance samples measured in the 30 plots were then resampled into nSight-2's spectral settings (i.e., FWHM and band centres) using the Hsdar R package (Lehnert et al., 2018). The response of each band was estimated using Gaussian distribution.

Worldview-2 (WV-2) satellite is the first high resolution 8-band multispectral commercial satellite sensor developed and launched by Digital Globe Inc. in 2009. It is a sun-synchronous satellite at 770 km altitude with a Ground Sampling (GSD) of 0.46 m and 0.52 m at nadir and 20° off-nadir in the panchromatic band, respectively. It has a swath width of 14.6 km at nadir and a dynamic range of 11-bit per pixel. Its spectral configuration is as follows: panchromatic band (450 – 800 nm), the coastal band (400 – 450 nm), Blue band (450 – 510 nm), Green band (510 – 580 nm), Yellow band (585 – 625 nm), Red band (630 – 690 nm), Red-Edge (705 – 745 nm), NIR-1 (770 – 895 nm), NIR-2 (860 – 1040 nm).

Environmental Monitoring and Analysis Programme (EnMAP) is a German Space Agency (DLR) small satellite. It is a polar-orbiting earth observation satellite with a dedicated

hyperspectral push-broom imager with a GSD of 30 m and a swath width of 30 km. The EnMAP has a spectral sampling in two regions of the electromagnetic spectrum, i.e., VNIR and SWIR at 6.5 nm and 10 nm spectral resolutions, respectively and the range of 420 nm – 2450 nm. This spectral range is divided as follows: VNIR (420 nm – 1000 nm); SWIR-1 (900 – 1390 nm); SWIR-2 (1480 – 1760 nm), and SWIR-3 (1950 – 2450 nm). The resampled EnMap data had 242 bands with a mean bandwidth of 10.47 nm and a spectral range of 423.03 nm – 2438.6 nm. To facilitate comparison with nSight-2 spectral settings, a spectral subset of 78 bands with a range of 400 nm – 900 nm was used instead of the entire VNIR and SWIR dataset.

3.3 Machine Learning

Traditional parametric image classification algorithms have been found insufficient in digital processing of hyperspectral data (Pham, et al. 2019). As such, advances in computer vision, pattern recognition and artificial intelligence technologies have resulted in the new developments like machine learning algorithms that have become so instrumental in classification of remote sensing data. Among these machine learning algorithms, Random Forest (RF), Support Vector Machine (SVM) and Partial Least Squares Discriminant Analysis (PLS-DA) have been proven robust in producing higher overall mapping accuracy (Zafari, Zurita-Milla and Izquierdo-Verdiguier, 2019).

3.3.1 Random Forest

Random forest (RF), as introduced by Breinman (2001), is an ensemble machine learning classifier that works through training of several decision trees to individually provide a classification for input data via the bagging process. Each decision tree in the classification takes input from samples in the initial dataset. Randomly selected features are then used to grow the tree at each node without pruning until the prediction is finally reached. The RF classifier bootstraps random samples where the prediction with the majority vote from all trees is selected (Igual and Segui, 2017).

It usually uses three hyper-parameters that need tuning before training, i.e. node size, the number of trees and the number of samples. In each bootstrapped sample, about two-thirds of the input data (i.e., in-bag samples) are randomly selected using bagging to grow an unpruned classification tree. In other words, RF uses recursive partitioning of data into

nodes until each of the trees contains similar samples or reaches one stopping condition (Pal, 2005). In this process, two parameters are used, i.e., the number of trees (*n-tree*) and the number of variables randomly selected at each split (*mtry*). Then, the Shanon entropy or *Gini coefficient* is used as the splitting function to measure the impurity of an attribute concerning the classes (Lim, Kim and Jim, 2019). Each classification tree votes for a class membership for each test sample, and the class with maximum votes will be considered the final class. The remaining one-third (i.e., out-of-the-bag) is used for internal cross-validation to estimate out-of-the-bag errors (Lim, Kim and Jin, 2019; Zafari, Zurita-Milla and Izquierdo-Verdiguier, 2019). In this study, the n-tree and mtry parameters were tuned using the grid-search strategy in 'random Forest' R-Statistics package. The n-tree values from 100 to 1000 and mtry values from 1 to $\sqrt{m}$ (i.e., square root of the number of bands) were tested.

Random Forest is known for producing higher classification accuracy and runs efficiently on large databases. It does not require much parameter tuning, hence it is easy to determine which parameter to use. It also has been proven to be robust in handling noisy data and therefore effective for a wide variety of classification and regression tasks (Golrang, et al. 2020). Overfitting is easily dealt with since there is no need for pruning the random forest (Breinman, 2001). RF algorithm is readily available from recommendable open-source data software like Python and R Statistics. It makes use of all samples in a dataset and hence avoids overfitting (Pal, 2005). Random Forest is simple and easy to implement since it uses a few lines of code. Moreover, it is computationally fast saving on time. RF is versatile as it does not require input preparation and is capable of handling numerical, binary and categorical features with scaling, transformation or modification. It is also implicit as it estimates on what variables are important in the classification (James, 2021).

However, theoretical analysis is difficult with RF. It is affected by a large number of decision trees in the RF as they slow down the real-time predictive power of the algorithm (Yiu, 2019). Random Forest is highly sensitive to categorical variables with different number of levels since the algorithm becomes biased towards those attributes with more levels resulting in unreliable variable importance scores.

In this study, RF was tuned for *n-tree* and *mtry***.** For tuning, the grid-search strategy was adopted at 5-fold cross validation. The optimal parameters were 500 and 2 for ntree and mtry respectively. This contributed to a training accuracy of 78.29% (Kappa=70.18)./

3.3.2 Support Vector Machine

Support Vector Machines, as first developed by Vapnik (1999), are supervised machine learning algorithms that classify data through finding a hyperplane, i.e. a function used to separate the features into different domains (Yadav and Mather, 2020). The SVM algorithms is founded on the idea that the greater the margin between the support vectors and the hyperplane, the higher the chances of correctly classifying the points in their respective classes (Mountrakis *et al*., 2011). In this, the SVM algorithm uses the kernel functions i.e. the way of computing the dot product of two vectors X and Y in a feature space. There are four types of kernel functions used in the SVM algorithm i.e. linear kernel, polynomial kernel, sigmoid kernel and the radial basis function kernel (RBF) also known as the Gaussian kernel.

SVM defines decision boundaries using geometrical characteristics of data through a hyper plane based on support vectors (Melgani & Bruzzone, 2004). It does not require *a priori* knowledge about the statistical distribution of data, can reduce classification errors while increasing resolution, and uses kernels which are effective in their accuracy (Mountrakis *et al*., 2011). Its strength lies in that it does not employ density estimation to discriminate classes; instead, it utilizes the geometrical characteristics of data to define decision boundaries by assessing only support vectors (Melgani & Bruzzone, 2004). SVM is highly sensitive to the choice of kernel, size of kernel and parameter *C* (Hsu, Chang & Lin, 2010). Furthermore, SVM is computationally fast and easy to manipulate.

In this study, the Gaussian kernel was used. It uses a formula:

$$K(X - X) = \text{exponent}(-\text{gamma}||X1 - X2||) \qquad \textit{Equation 1}$$

Gamma specifies how much a single training point has on the other data points around it and $||X1 - X2||$ is the dot product between features.

The SVM-RBF kernel requires parameterization of *C,* i.e. the inverse of the strength of regularization. As value of C increases the model over-fits and as the value of C

decreases the model under-fits. Also the RBF employs the *Gamma (Y*) parameter which also when the value increases, the model tends to over-fit and when the value decreases, the under-fits. Using the grid-search strategy, the two parameters were set as *Gamma (Y)* at 32 and cost parameter *C* at 0.01136596. This yielded a training accuracy of 90.54% (Kappa= 0.87).

3.3.3 Partial Least Squares Discriminant Analysis

Partial Least Squares Discriminant Analysis (PLS-DA) is a multivariate supervised clustering statistical machine learning algorithm that finds a linear regression model through constructing predictive variables and response variables into a new space (Chauhan et al. 2020). Although the PLS-DA model was technically designed to solve data problems in chemo metrics and related subjects, it has become instrumental in classification problems in remote sensing (Richter et al. 2016). Its versatility in handling multi-collinear data, information redundancy and missing data, has prompted its utility in many applications including remote sensing.

The major tenets of PLS-DA are that it creates predictive variables and response variables which produces few eigenvectors from spectral matrices (Peerbhay, et al. 2013). This results in data that is correlated and characterized by predictor variables that are more than observation. PLS-DA then finds optimum components which improves its classification performance (Peerbhay, et al. 2013; Peerbhay et al. 2016). The commonly used method for optimization of training data is cross-validation (Sibanda, et al. 2015; Peerbhay et al. 2016). The cross-validation process allows for the selection of noise-free components, reduces multicollinearity, and handles information redundancy (Yuan et al. 2020). To enhance PLS-DA classification accuracy, variable importance in the projection (VIP) score is generated (Palemo et al. 2017). This is achieved through the selection of relevant response variables. Scores of importance generated from each spectral band act as a representative measure of importance between the spectral bands (Peerbhay et al. 2013).

PLS-DA has been proven important and has gained momentum in the classification of remote sensing data due to a number of advantages. It reduces noise in the dataset, shows the probability of a sample belonging to the class being modelled and can select

best spectral variables (Li et al. 2016). Furthermore, it can handle multicollinearity, missing data and information redundancy which is a common occurrence with hyperspectral remote sensing data (Peerbhay, et al. 2016). PLS-DA has the power in determining the important variables that are explaining the discrimination outcome. It also determines the appropriate number of components for the model, which helps to prevent overfitting (Jiang, et al. 2014). However, the higher number of predictor variables than observations and multicollinearity of the wavebands runs the risk of overfitting (Richter et al. 2016). In remote sensing, it has been applied as an important tool for feature reduction, noise reduction, solving multicollinearity and overfitting which is common place (Sibanda et al. 2015; Peerbhay et al. 2016)

3.4 Accuracy Assessment

Accuracy assessment determines the magnitude of performance of a specific classification model or input data (Hasmadi, Pakhriazad & Shahrin, 2009). In this study, accuracy assessment was performed overall accuracy (OA), producer's accuracy (PA) and user's accuracy (UA) derived from a confusion matrix (Iqbal & Khan, 2014). From PA and UA, omission errors (OE) and commission errors (CE) were calculated by $1 - PA$ and $1 - UA$, respectively. The OA is the ratio of correctly classified sampled to the total samples. It represents the probability that a randomly selected point is classified correctly on the map (Foody, 2004). To compute the OA, the sum of correct classifications is divided by the total of reference map classes as shown in equation 2.

$$OA = \frac{\sum_{i=1}^{r} n_{ii}}{n} \text{ X } 100 \qquad 2$$

where *r* is the number of classes, n_{ii} are the diagonal elements and *n* represents the total number of considered pixels.

On the other hand, the PA (Equation 3) indicates the classification accuracy from the view of the map maker. It shows the probability that the classifier has correctly labelled the pixels and is computed from the following equation. It is calculated by taking the total number of correct classifications for particular class i.e. n_{ii} and divide by n_{icol} as shown in equation 3 (Verma et al. 2020). The UA shows the classification accuracy from the map

user's perspective. It denotes the extent to which the selected class labels to inform the classifier was correct (Adam *et al.*, 2014). It is calculated dividing the total number of correct classifications for a particular class i.e. n_{ii} by the row total i.e n_{irow} as shown in equation 4.

$$\mathrm{PA} = \left(\frac{n_{ii}}{n_{icol}}\right) \quad (3)$$

$$UA = \left(\frac{n_{ii}}{n_{irow}}\right) \quad (4)$$

n_{ii} is the number of correctly classified pixels and n_{icol} and n_{irow} are the column and row total, respectively (Verma et al. 2020)

Alongside these metrics, we used quantity and allocation differences ahead of the traditional Kappa coefficient. Use of Kappa coefficient ($\hat{\kappa}$) has been discredited by the contemporary accuracy revisionists (Pontius Jr and Millones, 2011, Pontius Jr and Santacruz, 2014), because it is redundant and misleading for practical applications. These authors argue that Kappa's proportion correct is not helpful in improving accuracy since it does not give the sources of disagreement. Also, Kappa coefficient is based on randomness which is meaningless and misleading. Furthermore, Kappa provides uninterpretable conclusions regarding agreement and disagreement between reference and classified maps (Pontius Jr and Millones, 2011). Consequently, they suggest two mutually exclusive measures, i.e., quantity and allocation difference are gaining prominence, hence were used in this study. Quantity difference or disagreement (QD) refers to the imperfect match in the class proportions between the classification and reference data (Equation 5). In contrast, allocation difference refers to the imperfect match in the class allocations between the reference data and the classification given their quantities (Equation 6).

$$QD = \frac{\sum\left|\frac{n_{+i}}{n} - \frac{n_{i+}}{n}\right|}{2} \text{ X } 100 \quad (5)$$

$$AD = \frac{\sum(\ 2*\ min\left(\frac{n_{+i}}{n} - \frac{n_{ii}}{n}, \frac{n_{i+}}{n} - \frac{n_{ii}}{n}\right))}{2} \text{ X } 100 \qquad (6)$$

n_{+i} and n_{i+} and represent the marginal sum of the columns and the marginal sum of rows, respectively.

This measure is divided into Shift (SD) and Exchange (ED) to account for differences by pairwise and non-pairwise class confusions, respectively (Pontius Jr and Santacruz, 2014). These measures were calculated using the Pontius matrix (Warrens, 2015).

$$Q = \frac{\sum_{j=1}^{J} q_j}{2} \qquad (7)$$

$$E = \frac{\sum_{j=1}^{J} e_j}{2} \qquad (8)$$

Q is the overall quantity difference, computed from the sum of class-wise quantity difference and divided by two, as shown in equation 7, since this process double-count the class-wise quantity difference, while E is the overall exchange difference and is divided by two as shown in equation 8, since summation of the numerator double-counts the class-wise exchange difference.

CHAPTER FOUR

4. RESULTS

4.1.1 Introduction

This study intended to test the capability of the spectral configuration of the forthcoming nSight 2 hyperspectral satellite sensor in discriminating among plant species in an inland wetland. Also, the study sought to test the performance of the advanced machine learning classifiers in handling simulated remotely sensed data derived from the nSight 2 configuration. This was achieved by comparing its performance with those of WV 2 and EnMap. This chapter details the results of the methods documented in Chapter 3. For comparison purposes, a spectral subset of 78 bands with a range of 400 nm – 900 nm was used (instead of the entire VNIR and SWIR dataset). This will be followed by a discussion of the results.

4.1.2 Comparison of simulated nSight 2, EnMap and WV 2 datasets

The results of the resampled spectra of nSight 2, EnMap and WV-2 datasets estimated using Gaussian distribution are shown in Figures 4.1, 4.2 and 4.3. The resulting resampled spectra had a mean bandwidth of 6.08 nm and a spectral range of 467 nm – 901 nm. The mean and ±SD of the resampled WV-2 spectral signatures for different wetland species are shown in Figure 4.2 while the mean and the ±SD of the resampled EnMap spectral signatures for different wetland species are shown in Figure 4.3.

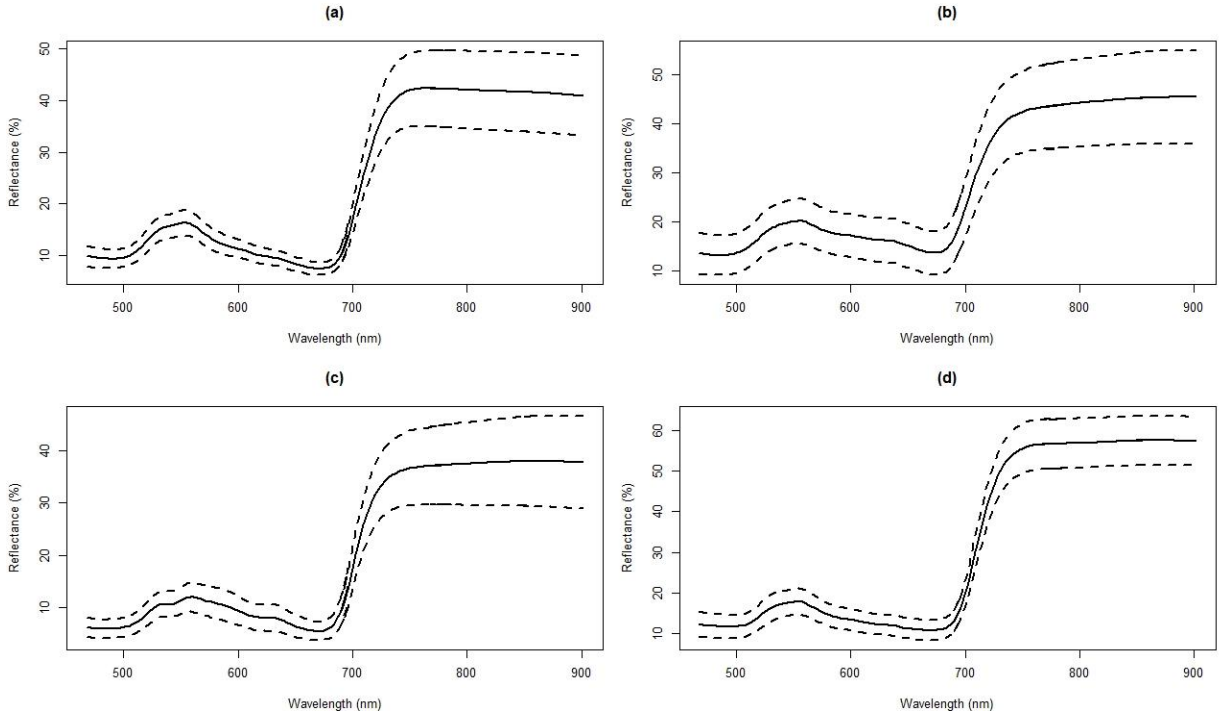

Figure 4. 1 The mean and ±SD of resampled nSight-2 spectra for (a) Crocosmia sp., (b) Grasses, (c) Agapanthus sp., and (d) Cyperus sp.

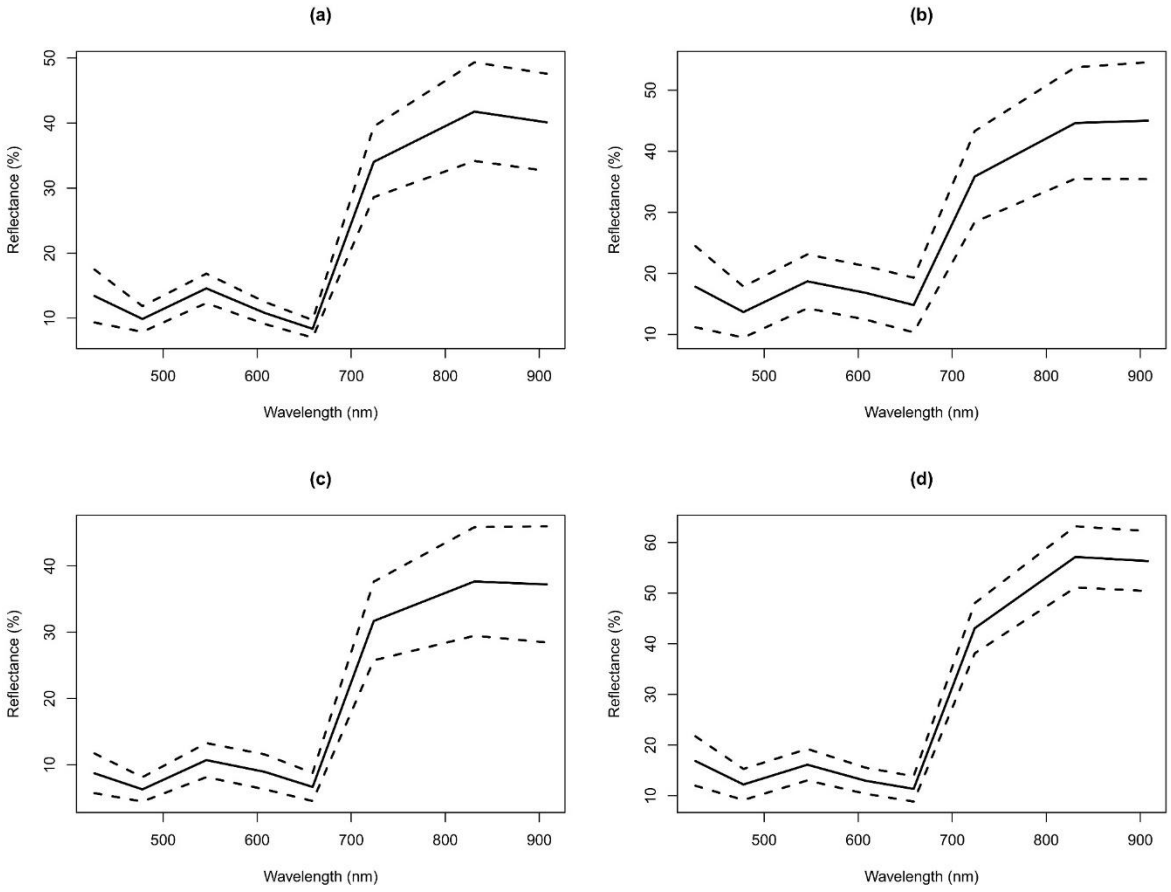

Figure 4. 2 The mean and ±SD of resampled Worldview-2 spectra for (a) Crocosmia sp., (b) Grasses, (c) Agapanthus sp., and (d) Cyperus sp.

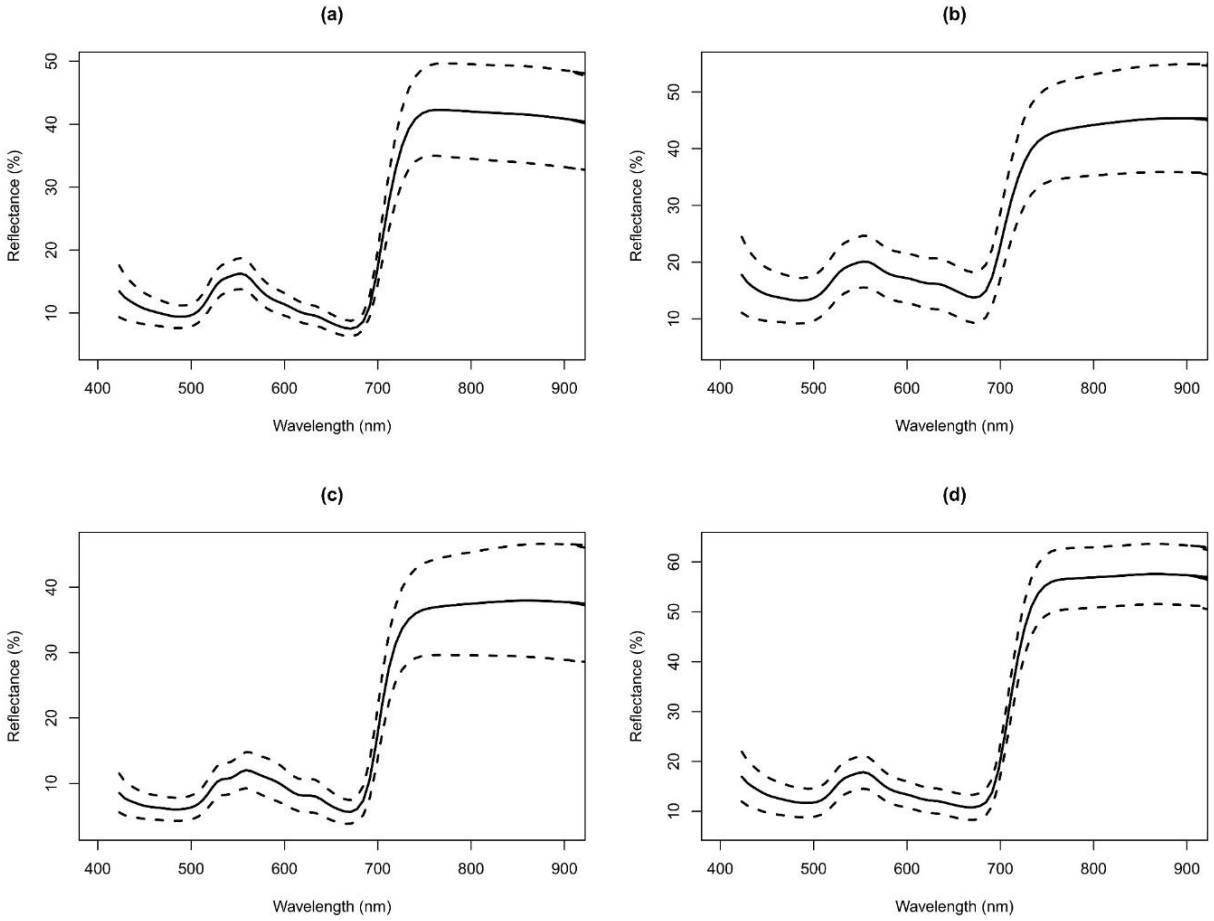

Figure 4. 3 The mean and ±SD of resampled EnMap spectra (subset up to 902 nm for (a) Crocosmia sp., (b) Grasses, (c) Agapanthus sp., and (d) Cyperus sp.

The results (Table 4.1 and Figure 4.2) show a superior performance of nSight 2 (93.36%) in the discrimination of plant species, followed by WV 2 (91.67%) and lastly the EnMap (77.77%). The class-wise accuracies (i.e., PA and UA) were relatively high when using nSight 2 band settings. However, for most cases, the class-wise accuracies were identical either between nSight 2 and WV-2 (e.g. Grasses) or WV-2 and EnMap (*Crocosmia sp.*) or across all sensors. The latter is particularly true for *Agapanthus sp.*, where PA and UA were approximately 67% and 100% throughout, respectively. Similarly, Grasses had similar PA across all sensors (i.e. approximately 92%), while nSight 2 and WV-2 had equivalent UA which was better than EnMap. In classes where there are no equivalent accuracies, EnMap was worse than other sensors evaluated in this study, i.e. nSight 2 and WV-2. Regarding the class-wise errors (i.e. OE and CE), Figure 4.4 reveals high OE

for *Agapanthus sp.* across all sensors, i.e. greater than 30%, followed by Grasses which showed high CE, i.e. greater than 20% for nSight 2 and WV-2, and greater than 30% for EnMap. The nSight 2 sensor had the lowest errors, i.e. less than 20%, for *Crocosmia sp.* and *Cyperus sp.* while WV-2 had similarly low errors for *Cyperus sp.* only.

Table 4.1: The performance of nSight 2, Worldview-2 and EnMAP in classifying wetland species using an independent validation sample. OA, CI, AD and QD denote overall accuracy, confidence intervals, allocation difference, and quantity difference.

	nSight 2		Worldview-2		EnMap	
	PA	UA	PA	UA	PA	UA
Crocosmia sp.	84.62%	84.62%	76.92%	83.33%	76.92%	83.33%
Grasses	91.67%	73.33%	91.67%	73.33%	91.67%	64.71%
Agapanthus sp.	66.67%	100.0%	66.67%	100.0%	66.67%	100.0%
Cyperus sp.	84.62%	91.67%	84.62%	84.62%	71.43%	83.33%
OA (95% CI)	84.09% (69.93%, 93.36%)		81.82% (67.29%, 91.81%)		77.77% (64.70%, 90.20%)	
AD (%)	9.09		11.36		11.11	
Shift (%)	4.54		6.82		6.66	
Exchange (%)	4.54		4.54		4.44	
QD (%)	6.81		6.82		11.11	

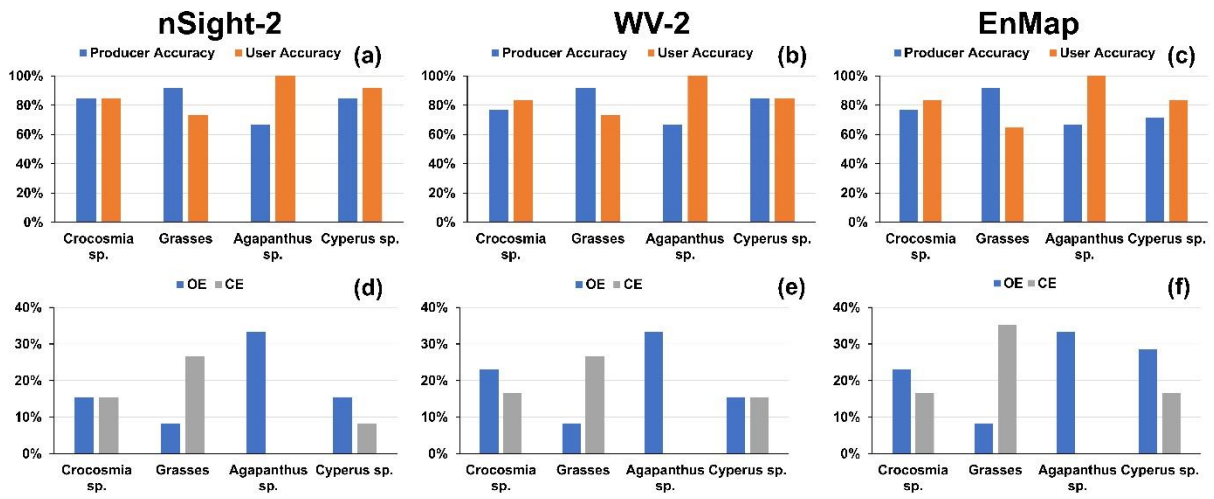

Figure 4. 4 The accuracy metrics (i.e. Producer's and User's accuracy) and class-wise omission (OE) and commission errors (CE) for nSight 2 (a, d), WV-2 (b, c) and EnMap

As indicated in Table 4.2, the most influential spectral bands among the datasets of the three sensors, i.e. nSight 2, EnMap and WV-2 were chosen in the visible, red-edge and near-infrared bands of the electromagnetic spectrum in the range of 400 nm to 900 nm, common amongst the three models, SVM, RF and PLS-DA. A total of 20 spectral bands from both nSight 2 and EnMap were selected while a total of 8 were selected from WV 2. In the VIS region, the nSight 2 had four spectral bands, while WV 2 had three and EnMap had 2. For the RE region, there were 11 spectral bands for nSight 2, 10 for EnMap data and 3 for WV-2. All the remaining were selected from the NIR region.

Table 4.2 Comparison of optimal spectral bands common among the algorithms in each dataset.

Electromagnetic Spectrum Region	nSight 2	Worldview-2	EnMap
	479.64 nm	427 nm	537.72 nm
	482.28 nm	478 nm	542.87 nm
VIS (400nm – 600nm)	540.58 nm	546 nm	
	573.31 nm		
	620.70 nm	608 nm	620.84 nm
	656.84 nm	659 nm	633.02 nm
	659.54 nm	724 nm	639.21 nm
	665.37 nm		645.47 nm
	668.29 nm		651.80 nm
RE (650nm – 750nm)	670.72 nm		658.20 nm
	675.18 nm		664.47 nm
	681.09 nm		671.21 nm
	683.60 nm		677.83 nm
	698.84 nm		684.51 nm
	747.25 nm		747.68 nm
	826.38 nm	831nm	755.01 nm
	848.82 nm	908nm	762.41 nm
	856.69 nm		769.86 nm
NIR (750 nm – 1200 nm)	869.93 nm		784.93 nm
	892.58 nm		871.05 nm
			887.16 nm
			903.36 nm

4.1.3 Performances of machine learning algorithms for classifying wetland plant species

We compared the achieved overall accuracies for RF, SVM and PLS-DA classifiers on the simulated nSight 2 sensor spectral bands, and found that the nSight 2 SVM model had the highest overall classification accuracy (93.18%), followed by the nSight 2 RF model with overall accuracy of 84.09%, and then the nSight 2 PLS-DA model with an overall accuracy of 83.63% (Table 4.3). Correspondingly, the AD and QD were found to be proportionately low for the nSight 2 SVM model, i.e. approximately 2 and less than 5 respectively, juxtaposed with nSight 2 RF model, i.e. approximately 9 and 7 respectively, and nSight 2 PSL-DA model less than 7 for the AD and 10% for QD. Although the nSight 2 PSL-DA model had an appropriately lower OA than nSight 2 RF model, its AD was better at 6.36% compared to the 9.09% of the RF model. To compare the differences caused by pair-wise and non-pair-wise class confusions, the Shift and Exchange metrics as contemplated in the AD, were used. The nSight 2 RF model and nSight 2 PLS-DA had a similar Shift at 4.54%, while the nSight 2 SVM model had the lowest Shift approximately 2. The Exchange was 0% for the nSight 2 SVM model, and less than 2% for the nSight 2 PLS-DA classification model, while it was 4.54% for the nSight 2 RF model.

Table 4.3: The performance of RF, SVM and PLS-DA in classifying wetland species. The terms OA, CI, AD and QD denote overall accuracy, confidence intervals, allocation difference and quantity difference.

	RF		SVM		PLS-DA	
	PA	UA	PA	UA	PA	UA
Crocosmia sp.	84.62%	84.62%	100	100	84.85	90.32
Grasses	91.67%	73.33%	100	86.67	95.83	74.19
Agapanthus sp.	66.67%	100.0%	84.62	100	93.33	87.5
Cyperus sp.	84.62%	91.67%	91.67	91.67	93.10	84.38
OA	84.09%		93.18%		83.63%	
(95% CI)	± 0.1089		± 0.0753		± 0.0694	
AD (%)	9.09		2.27		6.36	
Shift (%)	4.54		2.27		4.54	

Exchange (%)	4.54	0	1.81
QD (%)	6.81	4.54	10.00

The class-wise PA and UA metrics were fairly high across all models. The class of grasses had greater than 90% PA across all the classification models, while its UA was less than 75% for both nSight 2 RF and PLS-DA, compared with greater than 85% for the nSight 2 SVM model (Figure 4.3). Similarly, the class of *Crocosmia* also had fairly high PA approximately 85% for the nSight 2 RF and PLS-DA models, and 100% for the nSight 2 SVM model, while its UA metric was greater than 85% across the classification models. The only class that had the lowest class-wise accuracy metric was *Agapanthus sp*. at approximately 67% PA for the nSight 2 RF model. In using class-wise errors (i.e. OE and CE) as depicted in Figure 4.5, the class of *Agapanthus sp*. had the highest OE in the nSight 2 RF and SVM models greater than 30%, while it had OE approximately 6% for the nSight 2 PSL-DA model. Interestingly, the class of grasses, despite their higher PA accuracy, had the highest CE, i.e. greater than 25% across the classifiers, while the *Agapanthus sp* had 0% CE for nSight 2 RF and SVM and approximately 12 % for the nSight 2 PLS-DA model.

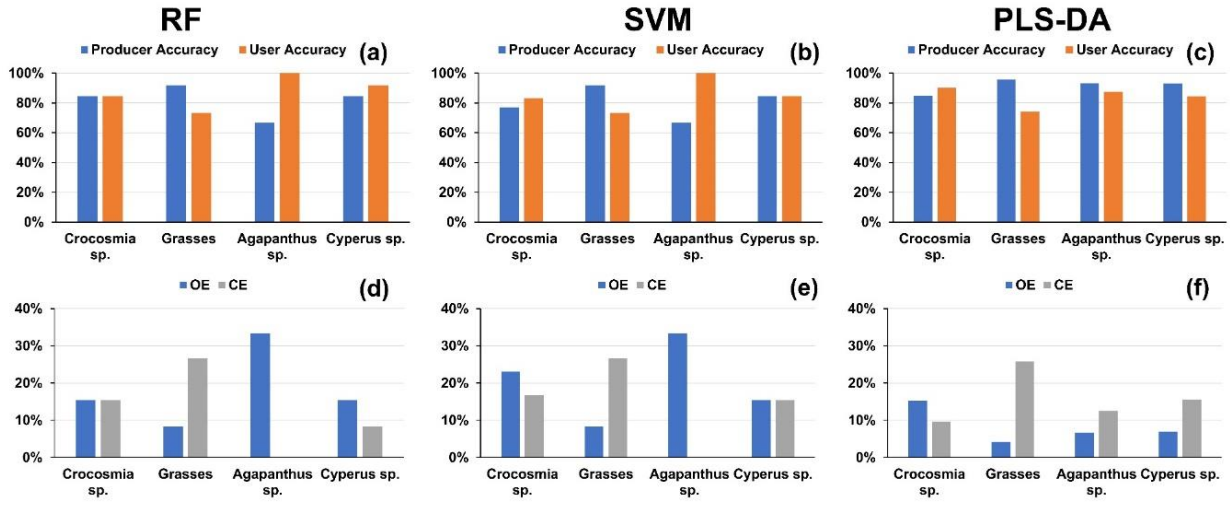

Figure 4. 5: The accuracy metrics (i.e. Producer's and User's accuracy) and class-wise omission (OE) and commission errors (CE) for RF Model (a, d), SVM Model (b, c) and PLS-DA Model (c, f).

Figure 4.6 shows that the RE region of the electromagnetic spectrum had the highest contribution in terms of the number of predictive features: 11, 10, 5 for nSight 2, EnMap and WV-2, respectively, and in terms of mean importance percentages (greater than 70 %) in each sensor dataset. This was followed by NIR region 5, 8, 2 and VIS 4, 3, 2 for nSight 2, EnMap and WV-2, respectively. For nSight 2, out of the 11 selected RE spectral bands, the band located at 665.37 nm had the highest contribution greater than 87%, and this was followed by the band located at 681.09 nm. For the EnMap, the RE also had the highest contribution, particularly for the band located at 664.67 nm, while the WV-2 RE band located at 659 nm had the highest contribution approximately 70% among the selected bands in the WV-2 sensor. The NIR spectral bands located at 848.82 nm and 869.93 nm had almost the same contribution approximately 75% for the nSight 2, while the EnMap NIR spectral bands had a contribution of approximately 65%, with a spectral band located at 762.41 nm having the highest contribution of 65%. The 831 nm band of the WV-2 sensor had a contribution of approximately 57%.

Furthermore, the results show variations in the variable importance by species. Across all sensors, the RE had the highest contribution for the *Crocosmia sp.* between 90% and 100%, while *Agapanthus sp.* had the highest VIS spectral bands contribution across all the sensors. Also, the grass species had high reflectance in the red to RE spectral bands across the sensors. The plant species that dominated the NIR spectral region were the *Crocosmia sp*. and *Cyperus sp*.

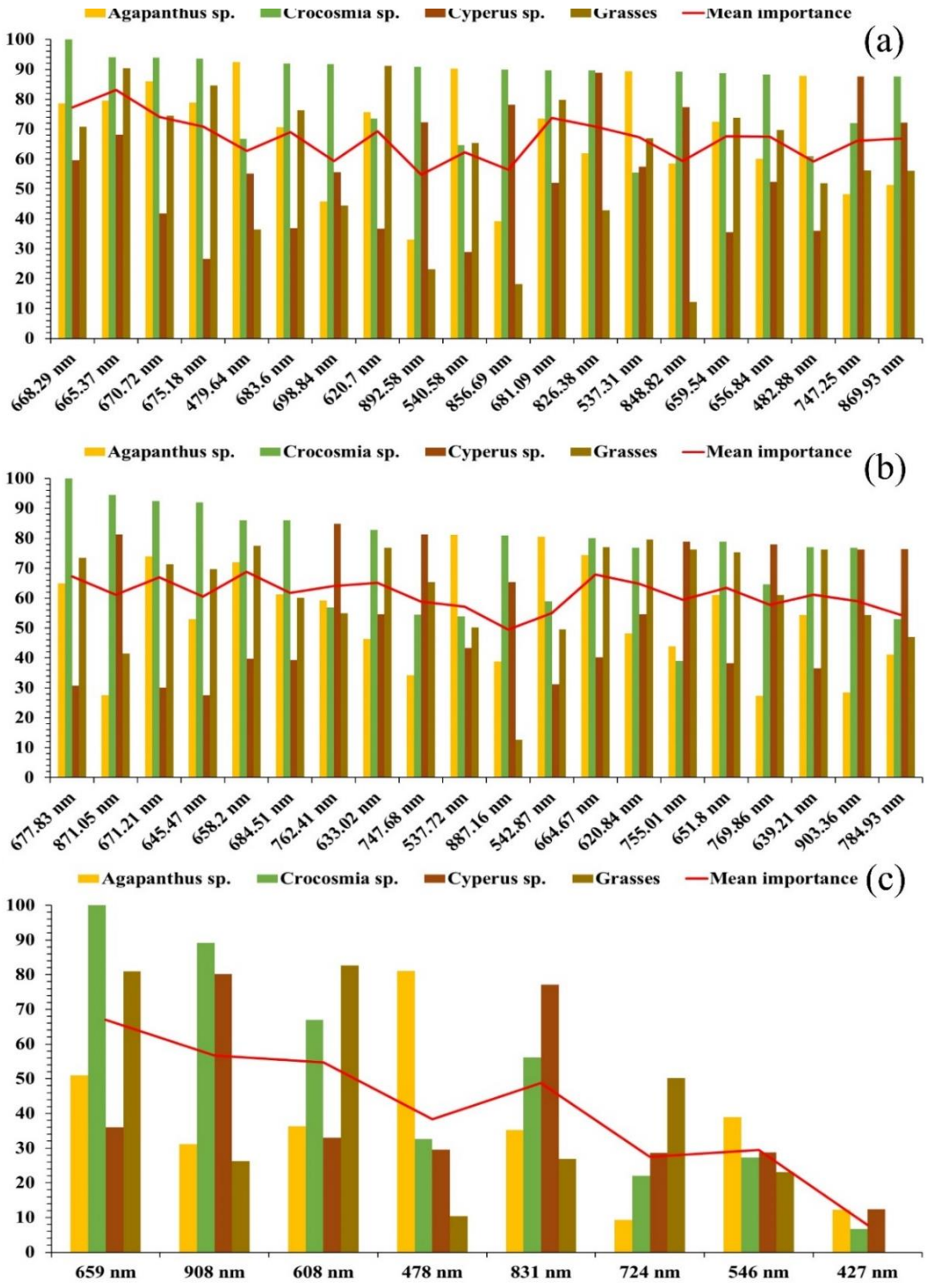

Figure 4. 6 Support Vector machine variable importance using (a) nSight-2 bands, (b) EnMAP bands, and (c) Worldview-2 bands. The y-axis indicate scaled SVM variable importance (%).

CHAPTER FIVE

5 DISCUSSION, CONCLUSION AND RECOMMENDATIONS

5.1 Discussion

5.1.1 Comparison of simulated nSight 2, WV 2 and EnMap

The results demonstrate the potential of the nSight 2 spectral band settings in discriminating different wetland plant species in Verloren Vallei Nature Reserve. Specifically, the spectral configuration on nSight 2 hyperspectral sensor discriminated *Crocosmia sp*., *Agapanthus sp*., *Themeda triandra sp*. and *Cyperus sp*. to an overall accuracy of 0.84 based on the resampled wavebands in the red edge and the NIR regions of the electromagnetic spectrum. The discriminative power of nSight 2 could be attributed to the narrow configuration of its wavelength bands which makes them to be more sensitive to minute variations in the reflectance of different vegetation species considered in this study. Specifically, both nSight 2 and EnMap contributed twenty optimal spectral bands as top most contributing wavebands. As aforementioned, their narrow contiguous spectral bands are able to discern on the spectral variations of vegetation species in this study when compared to only eight broad bands chosen from the WV 2, a multispectral satellite sensor. Broadband sensors' wavebands tend to mask out critical information that is required in discriminating different vegetation traits such as species variabilities.

While both nSight 2 and EnMap selected an equal number of spectral bands, i.e. 20, the nSight 2 had a higher overall accuracy i.e. 93.36% compared to the 77.77% of EnMap map. This can be explained by the fact that nSight 2 has a limited number of variables. The sensor will provide a smaller number of narrow spectral bands, i.e. 160, that are less susceptible to the Hughes phenomena hence they could be better handled by machine learning classifiers. Furthermore, nSight 2 will have an improved spectral bandwidth of 6 nm compared with the 10 nm of EnMap. This could then explain the high performance of nSight 2 in relation to EnMap. Meanwhile, EnMap covers a wider section of the electromagnetic spectrum such that it has many redundant variables that do not contribute much information for discrimination purposes, instead, they have an effect of decreasing the classifiers accuracies. Even though nSight 2 had a better performance than the two other sensors, there was no significant difference between the performance

of nSight 2 and WV 2 simulated datasets. This may be because the most influential spectral sections derived from these sensors was similar, such as RE bands.

Specifically, the RE wavebands (620.70 nm; 656.84; 659.54; 665.37; 668.29 nm; 670.72 nm; 675.18 nm; 681.09 nm; 683.60 nm; 698.84 nm; 747.25) and NIR (826.38 nm; 848.82 nm; 856.69 nm; 869.93 nm; 892.58 nm) were important for discerning subtle differences among the four dominant plant species. This can be explained by the fact that plant species differ in concentrations of chlorophyll and nitrogen contents in the plant species considered in this study. The RE is highly sensitive to these components, while the NIR spectral region is highly sensitive to tannin content which indicates the differences in structural and physiological properties of plant species, enabling their differentiation (Mutanga et al. 2015; Adam et al. 2012). These results concur with the findings of Otunga, et al. (2019) who evaluated the potential of RE using the Sentinel 2 MSI, RapidEye data and diiscriminant analysis and Maximum Likelihood classifiers in the discrimination of C3 (Festuca spp) in grassland in KwaZulu-Natal in South Africa and concluded that satellite sensors which integrate the strategically located RE bands could offer information that is critical for plant species discrimination and valuable for sustainable rangeland management. Again, these results support the findings by Sibanda et al. (2017) who tested the capability of Worldview 3 satellite data in discriminating grasses under different rangeland management practices in South Africa using raw spectral bands and spectral indices and found that incorpration of RE, yellow (57nm to 590 nm) red (620 nm to 750 nm) and the NIR (780 nm to 1300 nm) bands in plant species discrimination improved the overall accuracies. It can be concluded that the simulated spectral bands of EnMap and WV 2 also have a potential of discriminating wetland plant species with relatively high accuracies although relatively lower in comparison to nSight 2. This could be because they have spectral bands such as RE which are important for vegetation analysis.

5.1.2 Performances of RF, SVM and PLS-DA algorithms in classification of resampled nSight 2 dataset.

We compared the achieved overall accuracies for RF, SVM and PLS-DA classifiers on the simulated nSight 2 sensor spectral bands, and found that the nSight 2 and SVM model had the highest overall classification accuracy (93.18%), followed by the nSight 2 and RF

model with overall accuracy of 84.09%, and then the nSight 2 and PLS-DA model with an overall accuracy of 83.63%. The superior performance of SVM model in the classification of resampled nSight 2 dataset, can be explained by the fact that it can keep a balanced ratio in validation and training samples and parameter tuning (gamma and penalty) in the SVM model. SVM was able to reduce overlap in the spectra of the plant species understudy and the extraction of useful spectral bands for plant species discrimination. Even though both SVM and RF are machine learning algorithms, there are other studies that have illustrated that RF performs less than SVM (Rackzo and Zagajeweski, 2017; Dabija, et al. 2021). Specifically, these findings are in agreement with the findings by Adelabu et al. (2013) who compared the performances of RF and SVM in tree species classification in a Savanna woodland in Southern Africa and concluded that the SVM algorithm was a better classifier than RF.

Again, there was no significant difference between the performances of RF and PLS-DA algorithms, i.e. 84.09% for RF and 83.63% PLS-DA. This can be explained by the fact that although PLS-DA is a machine learning algorithm, it is a multivariate method that is slightly passive in feature selection. While this helps it to improve classification on one hand, it leads to overfitting on the other. Interestingly, these results are in disagreement with the findings by Richter et al. who compared the performances of PLS-DA, RF and SVM in a forest in Central Europe and found that the PLS-DA model outperformed both SVM and RF in tree species discrimination. Furthermore, the machine learning algorithms, RF, SVM and PLS-DA, can successfully handle the data from these upcoming sensors and WV-2.

5.2.3 Influence of species varieties

The class-wise PA and UA metrics were fairly high across all models. The class of grasses had greater than 90% PA across all the classification models, while its UA was less than 75% for both nSight 2 RF and PLS-DA, compared with > 85% for the nSight 2 SVM model. This confirms the superiority of the nSight 2 dataset used in conjunction with SVM classifier in discriminating wetland plant species. While the per class accuracies differed from plant to plant and algorithm to algorithm, this can be explained by the fact plant species differ interms of foliage, structural and density differences. The spectra

under these differing attributes result in different classification accuracies. Also there is a great influence of leaf area, size and texture, which greatly afftects the reflectance of these plant species. Another important contributing factor to the differences in the reflectances of plant spcies could be perhaps the phenological stage at which these plant species were spectrally measured (Deepak et al. 2019; Prospere, McLaren and Wilson, 2014).

5.1.4 Implications of the results

While these results reveal a great potential of nSight 2 spectral band setting as suitable for plant species discrimination, it should be noted that resampled data was used in this study. Actual dataset may be affected by atmospheric and processing impurities which will need to be considered. Specifically, the satellite imagery may also be inluenced by sun angle, sensor geometry, illumination and atmospheric conditions.

5.2 Conclusion

This study sought to test the capability of the spectral configuration of the forthcoming nSight 2 hyperspectral satellite sensor in discriminating among plant species in an inland wetland. Also the study sought to test the performance of the advanced machine learning classifiers in handling simulated remotely sensed data derived from the nSight 2 configuration. Grounded on the findings of this study, it can be concluded:

- That nSight 2 spectral configuration is suitable for plant species discrimination in a wetland land environment.
- That the RE region of the nSight 2 is critical for plant species discrimination in a wetland environment.
- That use of SVM classification algorithm with nSight 2 is suitable for plant species discrimination in a wetland area.

5.3 Limitations of the study

This study encountered a number of shortcomings that can be leveraged for further investigation. Among these were the fact that the data is not actual spectra; instead, it is resampled from spectral band settings of the sensors, overcast weather conditions hindered fieldwork, and the study was based on laboratory experiments after harvesting

leaves from the dominant plant species. More experiments, using other classification algorithms such as NN, deep learning and gradient boosting, are required. Moreover, the influence of actual conditions of capturing remotely sensed data, i.e. atmospheric conditions, influence of underlying spectra, spatial resolution and radiometric noise, was not considered under these circumstances, and could have an impact on the performance of the actual spectra.

5.4 Recommendations

The aim and results of this study accentuate the importance of the upcoming hyperspectral sensors, with their fine spectral resolution, in discriminating among the wetland plant species. The study also demonstrates the comparative performance of the spectral settings of the newer generation of multispectral sensors, in discriminating among the wetland plant species. These are critical for reserve management, in effectively using data derived from these sensors in developing strategies for monitoring and managing wetland plant species. The findings of this study are a comprehensive foundation for grounding future studies. Based on the findings of this study, we recommend that future studies should do the following:

- Compare the performance of the spectral bands and spectral indices derived from the spectral configuration of the nSight 2 in discriminating wetland plant species.
- Test the capability of spectral configuration of nSight 2 in the field, as compared to the laboratory conditions.
- Consider upscaling the results to satellite sensor data for the understanding of the performance of the nSight 2 spatially.
- Evaluate the influence of the atmospheric and underlying conditions under sensor data conditions.
- Consider fusing nSight 2 satellite sensor data with those of unmanned aerial vehicles (drones).
- For operational use and application of remote sensing for ecosystem management, it would be ideal to have the band settings of nSight 2 programmable like the Compact High Resolution Imaging Spectrometer

(CHRIS) a European Space Agency satellite. This satellite sensor has user-defined spectral band setting which involves programming the spectral bands for specific targets on the earth's surface, rather than collection of voluminous amounts of data that may not be useful all the time. This will help in dealing with multi-dimensionality and multi-collinearity which is associated with hyperspectral data and poses a challenge.

REFERENCES

Adam, E. and Mutanga, O., 2009. Spectral discrimination of papyrus vegetation (Cyperus papyrus L.) in swamp wetlands using field spectrometry. *ISPRS Journal of Photogrammetry and Remote Sensing*, *64*(6), pp.612-620.

Adam, E., Mutanga, O., & Rugege, D. (2010). Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: A review. *Wetlands Ecology and Management*, *18*(3), 281-296.

Adam, E., Mutanga, O., Abdel-Rahman, E. M., & Ismail, R. (2014). Estimating standing biomass in papyrus (*Cyperus papyrus L*.) swamp: Exploratory of in situ hyperspectral indices and random forest regression. *International Journal of Remote Sensing*, *35*(2), 693-714.

Adelabu, S., & Dube, T. (2015). Employing ground and satellite-based QuickBird data and random forest to discriminate five tree species in a Southern African Woodland. *Geocarto International*, *30*(4), 457-471.

Adelabu, S., Mutanga, O., Adam, E. E., & Cho, M. A. (2013). Exploiting machine learning algorithms for tree species classification in a semi-arid woodland using RapidEye image. *Journal of Applied Remote Sensing*, *7*(1), 1-14.

Adjorlolo, C., Mutanga, O., & Cho, M. A. (2014). Estimation of canopy nitrogen concentration across C3 and C4 grasslands using WorldView-2 multispectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(11), 4385-4392.

Adjorlolo, C., Mutanga, O., Cho, M. A., & Ismail, R. (2013). Spectral resampling based on user-defined inter-band correlation filter: C3 and C4 grass species classification. *International Journal of Applied Earth Observation and Geoinformation, 21*, 535-544.

Amani, M., Salehi, B., Mahdavi, S., & Granger, J. (2017). Spectral analysis of wetlands in Newfoundland using Sentinel 2A and Landsat 8 imagery. *Proceedings of the IGTF*.

Amler, E., Schmidt, M., & Menz, G. (2015). Definitions and mapping of East African wetlands: A review. *Remote Sensing*, *7*(5), 5256-5282.

Aneece, I., & Epstein, H. (2017). Identifying invasive plant species using field spectroscopy in the VNIR region in successional systems of north-central Virginia. *International Journal of Remote Sensing*, *38*(1), 100-122.

Ayanlade, A. (2017). Remote sensing vegetation dynamics analytical methods: A review of vegetation indices techniques. *Geoinformatica Polonica*, *2017*, 7-17.

Ballanti, L., Blesius, L., Hines, E., & Kruse, B. (2016). Tree species classification using hyperspectral imagery: A comparison of two classifiers. *Remote Sensing*, *8*(6), 445.

Belle, J. A., Jordaan, A., & Collins, N. (2018). Managing wetlands for disaster risk reduction: A case study of the eastern Free State, South Africa. *Jàmbá: Journal of Disaster Risk Studies*, *10*(1), 1-10.

Berhane, T. M., Lane, C. R., Wu, Q., Autrey, B. C., Anenkhonov, O. A., Chepinoga, V. V., & Liu, H. (2018). Decision-tree, rule-based, and random forest classification of high-resolution multispectral imagery for wetland mapping and inventory. Remote sensing, *10*(4), 1-26.

Bhatnagar, S., Gill, L., Regan, S., Naughton, O., Johnston, P., Waldren, S., & Ghosh, B. (2020). Mapping vegetation communities inside wetlands using Sentinel-2 imagery in Ireland. *International Journal of Applied Earth Observation and Geoinformation*, *88*, 1-13.

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

Campbell, J. B., & Wynne, R. H. (2011). *Introduction to remote sensing*. New York: Guilford Press.

Camps-Valls, G., & Bruzzone, L. (2005). Kernel-based methods for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *43*(6), 1351-1362.

Chauhan, S., Darvishzadeh, R., Boschetti, M. and Nelson, A., 2020. Discriminant analysis for lodging severity classification in wheat using RADARSAT-2 and Sentinel-1 data. *ISPRS journal of photogrammetry and remote sensing*, *164*, pp.138-151.

Corrigan, F. (2019). Multispectral imaging camera drones in farming yield big benefits. From: https://www.dronezon.com/learn-about-drones-quadcopters/multispectral-sensor-drones-in-farming-yield-big-benefits/

Dabija, A., Kluczek, M., Zagajewski, B., Raczko, E., Kycko, M., Al-Sulttani, A.H., Tardà, A., Pineda, L. and Corbera, J., 2021. Comparison of support vector machines and random forests for corine land cover mapping. *Remote Sensing*, *13*(4), p.777.

Dalponte, M., Bruzzone, L., & Gianelle, D. (2012). Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data. *Remote Sensing of Environment*, *123*, 258-270.

Dixon, M. J. R., Loh, J., Davidson, N. C., Beltrame, C., Freeman, R., & Walpole, M. (2016). Tracking global change in ecosystem area: The Wetland Extent Trends Index. *Biological Conservation*, *193*, 27-35.

Dronova, I., & Taddeo, S. (2016). Canopy leaf area index in non-forested marshes of the California Delta. *Wetlands*, *36*(4), 705-716.

Dube, T., Mutanga, O., Sibanda, M., Shoko, C., & Chemura, A. (2017). Evaluating the influence of the Red Edge band from RapidEye sensor in quantifying leaf area index for hydrological applications specifically focusing on plant canopy interception. *Physics and Chemistry of the Earth, Parts A/B/C*, *100*, 73-80.

Elvidge, C. D., & Chen, Z. (1995). Comparison of broad-band and narrow-band red and near-infrared vegetation indices. *Remote Sensing of Environment*, *54*(1), 38-48.

Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., Straub, C., & Ghosh, A. (2016). Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, *186*, 64-87.

Foody, G. M. 2004. Thematic map comparison. *Photogrammetric Engineering & Remote Sensing*, *70*(5), 627-633.

Foody, G. M., & Arora, M. K. (1997). Evaluation of some factors affecting the accuracy of classification by an artificial neural network. *International Journal of Remote Sensing, 18*, 799-810.

Foody, G. M., & Mathur, A. (2004). A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on Geoscience and Remote Sensing, 42*(6)*,* 1335-1343*.*

Furey, T. S. et al. (2000). Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*, *16*(10), 906-914.

Gaunter, L. (2015). The EnMAP Spaceborne Imaging Spectroscopy Mission for Earth Observation. *Remote Sensing, 7*(7), 8830- 8857.

Ghimire, B., Rogan, J., & Miller, J. (2010). Contextual land-cover classification: incorporating spatia dependence in land-cover classification models using random forests and the Getis statistic. *Remote Sensing Letters*, *1*(1), 45-54.

Golrang, A., Golrang, A. M., Yayilgan, S. Y., & Elezaj, O. (2020). A novel hybrid IDS based on modified NSGAII-ANN and random forest. *Electronics*, *9*(4), p.577.

Gopal, S., Woodcock, C. E., & Strahler, A. H. (1999). Fuzzy neural network classification of global land cover from a 1 AVHRR data set. *Remote Sensing of Environment*, *67*(2), 230-243.

Green, E., Mumby, P., Edwards, A. and Clark, C., 2000. *Remote sensing: handbook for tropical coastal management*. United Nations Educational, Scientific and Cultural Organization (UNESCO).

Green, R. O. et al. (2013). The HyspIRI Decadal Survey Mission: Update on the Mission Concept and Preparatory Airborne Science Campaign. In: *Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS' 13)*, Melbourne, Australia, p.4.

Guo, G., & Li, S. Z. (2003). Content-based audio classification and retrieval by support vector machines. *IEEE Transactions on Neural Networks*, *14*(1), 209-215.

Guo, M., Li, J., Sheng, C., Xu, J., & Wu, L. (2017). A review of wetland remote sensing. *Sensors*, *17*(4), 777.

Hansen, M., Dubayah, R., & DeFries, R. (1996). Classification trees: an alternative to traditional land cover classifiers. *International Journal of Remote Sensing*, *17*(5), 1075-1081.

Harrison, D., Rivard, B., & Sanchez-Azofeifa, G. A. (2018). Classification of tree species based on longwave hyperspectral data from leaves, a case study for a tropical dry forest. *International Journal of Applied Earth Observation and Geoinformation,* 66, 93-105.

Hasmadi, M., Pakhriazad, H. Z., & Shahrin, M. F. (2009). Evaluating supervised and unsupervised techniques for land cover mapping using remote sensing data. *Geografia: Malaysian Journal of Society and Space*, *5*(1), 1-10.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning: Data mining, inference and prediction.* New York: Springer-Verlag.

Heiden, U., Mueller, A., Guanter, L., Storch, T., Fischer, S., Godela, R., Habermeyer, M., Foerster, S., Segl, K., Chlebek, C. and Kaufmann, H., 2017, July. Preparatory activities for the German spaceborne imaging spectrometer mission EnMAP. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 439-442). IEEE.

Hennessy, A., Clarke, K., & Lewis, M. (2020). Hyperspectral classification of plants: A review of waveband selection generalizability. *Remote Sensing*, *12*(1), 113.

Hill, J., Buddenbaum, H., & Townsend, P. A. (2019). Imaging spectroscopy of forest ecosystems: perspectives for the use of space-borne hyperspectral earth observation systems. *Surveys in Geophysics*, *40*(3), 553-588.

Hsu, C., Chang, C., & Lin, C. (2010). *A practical guide to support vector classification*. Taipei: National Taiwan University, Department of Computer Science.

Hughes, G. F., 1968. On the mean accuracy of statistical pattern recognizers, IEEE Trans. Inform. Theory, IT-14, pp. 55-63

Igual, L., & Seguí, S. (2017). Introduction to data science. In *Introduction to Data Science* (pp. 1-4). Cham: Springer.

Im, J., & Jensen, J. R. (2008). Hyperspectral remote sensing of vegetation. *Geography Compass*, *2*(6), 1943-1961.

Iqbal, M. F., & Khan, I. A. (2014). Spatiotemporal land use land cover change analysis and erosion risk mapping of Azad Jammu and Kashmir, Pakistan. *Egyptian Journal of Remote Sensing and Space Science*, *17*(2), 209-229.

Isnard, T., Penatti, N. C., Ferreira, L. G., Arantes, A. E., & Do Amaral, C. H. (2015). Principal component analysis applied to a time series of MODIS images: The spatio-temporal variability of the Pantanal wetland, Brazil. *Wetlands Ecology and Management*, *23*(4), 737-748.

Jensen, J. R. (1983). Biophysical remote sensing. *Annals of the Association of American Geographers*, *73*(1), 111-132.

Jiang, M., Wang, C., Zhang, Y., Feng, Y., Wang, Y. and Zhu, Y., 2014. Sparse partial-least-squares discriminant analysis for different geographical origins of Salvia miltiorrhiza by 1H-NMR-based metabolomics. *Phytochemical Analysis*, *25*(1), pp.50-58.

Jiang, M., Wang, C., Zhang, Y., Feng, Y., Wang, Y., & Zhu, Y. (2014). Sparse partial-least-squares discriminant analysis for different geographical origins of Salvia miltiorrhiza by 1H-NMR-based metabolomics. *Phytochemical Analysis*, *25*(1), 50-58.

Jiao, S., Xu, Y., Zhang, J., & Lu, Y. (2019). Environmental filtering drives distinct continental atlases of soil archaea between dryland and wetland agricultural ecosystems. *Microbiome*, *7*(1), 15.

Jinru. X., & Su, B. (2017). Significant remote sensing vegetation indices: A review of developments and applications. *Journal of Sensors*, *2017*(1), 1-17.

Kalantari, Z., Seifollahi-Aghmiuni, S., Ferreira, C. S. S., Nockrach, M., & Pereira, P. (2019). January. The potential of wetland ecosystem services for achieving the Sustainable Development Goals. In: *Geophysical Research Abstracts*, Vol. 21.

Kaur, S., & Josan, G. S. (2011). Gurmukhi text extraction from Image using Support Vector Machine (SVM). *International Journal of Engineering Science and Technology*, *1*(3), 2977-2991.

Kganyago, M., Odindi, J., Adjorlolo, C., & Mhangara, P. (2017). Selecting a subset of spectral bands for mapping invasive alien plants: a case of discriminating *Parthenium hysterophorus* using field spectroscopy data. *International Journal of Remote Sensing*, *38*(20), 5608-5625.

Kumar, L., Schmidt, K.S., Dury, S. and Skidmore, A.K., 2001. Review of hyperspectral remote sensing and vegetation science. *Imaging spectrometry: Basic principles and prospective applications*, pp.111-155.

Lanaras, C., Baltsavias, E., & Schindler, K. (2017). Hyperspectral super-resolution with spectral unmixing constraints. *Remote Sensing*, *9*(11), 1196.

Lawrence, R. L., & Wright, A. (2001). Rule based classification systems using classification and regression trees (CART) analysis. *Photogrammetric Engineering and Remote Sensing*, *67*(10), 1137-1142.

Lees, B. G., & Ritman, K. (1991). Decision-tree and rule-induction approach to integration of remotely sensed and GIS data in mapping vegetation in disturbed or hilly environments. *Environmental Management*, *15*(6), 823-831.

Lefebvre, G., Redmond, L., Germain, C., Palazzi, E., Terzago, S., Willm, L., & Poulin, B. (2019). Predicting the vulnerability of seasonally-flooded wetlands to climate change across the Mediterranean Basin. *Science of the Total Environment*, *692*, 546-555.

Lehnert, L. W., Meyer, H., Obermeier, W. A., Silva, B., Regeling, B., & Bendix, J. (2018). Hyperspectral data analysis in R: The Hsdar package. *ArXiv preprint arXiv: 1805.05090*.

Li, X., Wang, S., Shi, W., & Shen, Q. (2016). Partial least squares discriminant analysis model based on variable selection applied to identify the adulterated olive oil. *Food Analytical Methods*, *9*(6), 1713-1718.

Lillesand, T., Kiefer, R. W., & Chipman, J. (2015). *Remote sensing and image interpretation*. London: John Wiley & Sons.

Lim, J., Kim, K. M., & Jin, R. (2019). Tree species classification using Hyperion and Sentinel-2 Data with Machine Learning in South Korea and China. *ISPRS International Journal of Geo-Information*, *8*(3), 150.

Ludwig, C., Walli, A., Schleicher, C., Weichselbaum, J., & Riffler, M. (2019). A highly automated algorithm for wetland detection using multi-temporal optical satellite data. *Remote Sensing of Environment*, *224*, 333-351.

Mahdavi, S., Salehi, B., Granger, J., Amani, M., Brisco, B., & Huang, W. (2018). Remote sensing for wetland classification: A comprehensive review. *GIScience Remote Sensing*, *55*, 623-658.

Mahdianpari, M., Salehi, B., Mohammadimanesh, F., Homayouni, S., & Gill, E. (2019). The first wetland inventory map of Newfoundland at a spatial resolution of 10m using sentinel-1 and sentinel-2 data on the google earth engine cloud computing platform. *Remote Sensing*, *11*(1), 43.

Malherbe, W. (2018). Ramsar wetlands in South Africa: Historic and current aquatic research. *South African Journal of Science and Technology*, *38*(1), 1-13.

Marchetti, Z. Y., Minotti, P. G., Ramonell, C. G., Schivo, F., & Kandus, P. (2016). NDVI patterns as indicator of morphodynamic activity in the middle Paraná River floodplain. *Geomorphology*, *253*, 146-158.

Masaitis, G., Mozgeris, G., & Augustaitis, A. (2013). Spectral reflectance properties of healthy and stressed coniferous trees. *Iforest-biogeosciences and Forestry*, *6*(1), 30.

Masemola, C., Cho, M. A., & Ramoelo, A. (2019). Assessing the effect of seasonality on leaf and canopy spectra for the discrimination of an alien tree species, Acacia Mearnsii,

from co-occurring native species using parametric and nonparametric classifiers. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(8), 5853-5867.

Mather, P. M., & Koch, K. (2011*). Computer processing of remotely-sensed images: An introduction.* London: John Wiley and Sons.

Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, *39*(9), 2784-2817.

McCoy, R. M. (2005). Field methods in remote sensing. London: Guilford Press.

Meerdink, S. K., Roberts, D. A., Roth, K. L., King, J. Y., Gader, P. D., & Koltunov, A. (2019). Classifying California plant species temporally using airborne hyperspectral imagery. *Remote Sensing of Environment*, *232*, 111308.

Melgani, F., & Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote sensing, 42*, 1778-1790*.*

Millennium Ecosystem Assessment. (2005). *Ecosystems and human well-being: wetlands and water*. Washington, DC: World Resources Institute. From: https://www.millenniumassessment.org/documents/document.47.aspx (accessed on July 26, 2020).

Misra, T. (2017). Indian remote sensing sensor system: Current and future perspective. In: *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences*, 87(4), 473-486.

Mitchell, S. A. (2013). The status of wetlands, threats and the predicted effect of global climate change: The situation in Sub-Saharan Africa. *Aquatic Sciences*, *75*(1), 95-112.

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing 66*(3), 247-259.

Mpumalanga Tourism and Parks Agency Verloren Nature Reserve. (http://www.mpumalanga.com/parks/verloren.asp) Accessed on the 16th of October 2020.

Mpumalanga Tourism and Parks Agency Verloren Vallei Nature Reserve (http://www.mpumalanga.com/) Accessed on the 16th of October 2020.

Muñoz, D. F., Cissell, J. R., & Moftakhari, H. (2019). Adjusting emergent herbaceous wetland elevation with object-based image analysis, random forest and the 2016 NLCD. *Remote Sensing*, *11*(20), 2346.

Mutanga, O., & Adam, E. (2011). High density biomass estimation: Testing the utility of Vegetation Indices and the Random Forest Regression algorithm. In: *34th International Symposium for Remote Sensing of the Environment (ISRSE*), Sydney, Australia.

Mutanga, O., Adam, E. and Cho, M.A., 2012. High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. *International Journal of Applied Earth Observation and Geoinformation*, *18*, pp.399-406.

Ola, O., & Benjamin, E. (2019). Preserving biodiversity and ecosystem services in West African forests, watersheds, and wetlands: A review of incentives. *Forests*, *10*(6), 479.

Ollis, D. J., Snaddon, C. D., Job, N. M., & Mbona, N. (2013). Classification system for wetlands and other aquatic ecosystems in South Africa. User manual: inland systems. Pretoria: South African National Biodiversity Institute. (SANBI Biodiversity Series 22.)

Otunga, C., Odindi, J., Mutanga, O., & Adjorlolo, C. (2019). Evaluating the potential of the red edge channel for C3 (*Festuca spp*.) grass discrimination using Sentinel-2 and Rapid Eye satellite image data. *Geocarto International*, *34*(10), 1123-1143.

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, *26*(1), 217-222.

Pal, M., & Mather, P. M. (2005). Support vector machines for classification in remote sensing. *International Journal of Remote Sensing, 26*(5), 1007-1011.

Palermo, A., Botrè, F., De la Torre, X., & Zamboni, N. (2017). Non-targeted LC-MS based metabolomics analysis of the urinary steroidal profile. *Analytica Chimica Acta*, *964*, 112-122.

Peerbhay, K., Mutanga, O., Lottering, R., & Ismail, R. (2016). Mapping *Solanum mauritianum* plant invasions using WorldView-2 imagery and unsupervised random forests. *Remote Sensing of Environment*, *182*, 39-48.

Peerbhay, K.Y., Mutanga, O., & Ismail, R. (2013). Commercial tree species discrimination using airborne AISA Eagle hyperspectral imagery and partial least squares discriminant analysis (PLS-DA) in KwaZulu-Natal, South Africa. *ISPRS Journal of Photogrammetry and Remote Sensing*, *79*, 19-28.

Pérez-Enciso, M., & Tenenhaus, M. (2003). Prediction of clinical outcome with microarray data: A partial least squares discriminant analysis (PLS-DA) approach. *Human Genetics*, *112*(5-6), 581-592.

Pham, T. D., Xia, J., Ha, N. T., Bui, D. T., Le, N. N., & Tekeuchi, W. (2019). A review of remote sensing approaches for monitoring blue carbon ecosystems: Mangroves, sea-grasses and salt marshes during 2010-2018. *Sensors*, *19*(8), 1933.

Pignatti, S. (2013). The PRISMA Hyperspectral Mission: Science activities and opportunities for agriculture and land monitoring. *Proceedings of the International Geoscience and remote Sensing Symposium (IGARSS'13)*, Melbourne, Australia.

Pontius Jr, R. G., & Millones, M. (2011). Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing, 32***,** 4407-4429.

Pontius Jr, R. G., & Santacruz, A. (2014). Quantity, exchange, and shift components of difference in square contingency table. *International Journal of Remote Sensing, 35***,** 7543-7554.

Pride, M., Priscilla, M. T., & Gara, T. (2020). Dominant wetland vegetation species discrimination and quantification using in situ hyperspectral data. *Transactions of the Royal Society of South Africa*, *75*(3), 229-238.

Pu, R., & Cheng, J. (2015). Mapping forest leaf area index using reflectance and textural information derived from WorldView-2 imagery in a mixed natural forest area in Florida, US. *International Journal of Applied Earth Observation and Geoinformation*, *42*, 11-23.

Raczko, E., & Zagajewski, B. (2017). Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. *European Journal of Remote Sensing*, *50*(1), 144-154.

Ramsey III, E., & Rangoonwala, A. (2015). 8 Radar and Optical Image Fusion and Mapping of Wetland Resources. *Remote Sensing of Wetlands: Applications and Advances*, p. 155.

Randin, C. F., Ashcroft, M. B., Bolliger, J., Cavender-Bares, J., Coops, N. C., Dullinger, S., Dirnböck, T., Eckert, S., Ellis, E., Fernández, N., & Giuliani, G. (2020). Monitoring biodiversity in the Anthropocene using remote sensing in species distribution models. *Remote Sensing of Environment*, *239*, 111626.

Rapinel, S., Mony, C., Lecoq, L., Clément, B., Thomas, A., & Hubert-Moy, L. (2019). Evaluation of Sentinel-2 time-series for mapping floodplain grassland plant communities. *Remote Sensing of Environment*, *223*, 115-129.

Rast, M., & Painter, T. H. (2019). Earth Observation Imaging Spectroscopy for Terrestrial Systems: An overview of its history, techniques, and applications of its missions. *Surveys in Geophysics*, *40*(3), 303-331.

Rebelo, L. M., Finlayson, C. M., Strauch, A., Rosenqvist, A., Perennou, C., Tøttrup, C., Hilarides, L., Paganini, M., Wielaard, N., Siegert, F., & Ballhorn, U. (2018). The use of Earth Observation for wetland inventory, assessment and monitoring: An information source for the Ramsar Convention on Wetlands. In: *Ramsar Technical Report No. 10*. Gland, Switzerland: Ramsar Convention Secretariat.

Richter, R., Reu, B., Wirth, C., Doktor, D., & Vohland, M. (2016). The use of airborne hyperspectral data for tree species classification in a species-rich Central European forest area. *International Journal of Applied Earth Observation and Geo-information*, *52*, 464-474.

Rogan, J., Franklin, J., Stow, D., Miller, J., Woodcock, C., & Roberts, D. (2008). Mapping land-cover modifications over large areas: A comparison of machine learning algorithms. *Remote Sensing of Environment*, *112*(5), 2272-2283.

Rosso, P. H., Pushnik, J. C., Lay, M., & Ustin, S. L. (2005). Reflectance properties and physiological responses of Salicornia virginica to heavy metal and petroleum contamination. *Environmental Pollution*, *137*(2), 241-252.

Roth, K. L., Roberts, D. A., Dennison, P. E., Alonzo, M., Peterson, S. H., & Beland, M. (2015). Differentiating plant species within and across diverse ecosystems with imaging spectroscopy. *Remote Sensing of Environment*, *167*, 135-151.

Samiappan, S., Prasad, S., Bruce, L. M., & Robles, W. (2010). July. NASA's upcoming HyspIRI mission—precision vegetation mapping with limited ground truth. In: *2010 IEEE International Geoscience and Remote Sensing Symposium* (pp. 3744-3747). IEEE.

Samiappan, S., Turnage, G., Hathcock, L., Yao, H., Kincaid, R., Moorhead, R., & Ashby, S. (2017). July. Classifying common wetland plants using hyperspectral data to identify optimal spectral bands for species mapping using a small unmanned aerial system – A case study. In: *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 5924-5927). IEEE.

Schmidt, K. S., & Skidmore, A. K. (2003). Spectral discrimination of vegetation types in a coastal wetland. *Remote Sensing of Environment*, 85, 92-108.

Sibanda, M., Mutanga, O., & Rouget, M. (2015). Examining the potential of Sentinel-2 MSI spectral resolution in quantifying above ground biomass across different fertilizer treatments. *ISPRS Journal of Photogrammetry and Remote Sensing*, *110*, 55-65.

Sibanda, M., Mutanga, O., & Rouget, M. (2017). Testing the capabilities of the new WorldView-3 space-borne sensor's red-edge spectral band in discriminating and mapping complex grassland management treatments. *International Journal of Remote Sensing*, *38*(1), 1-22.

Siebrits, A., & Gasela, M. (2019). Part 1: Sustainable Development Goals and Space in Africa: Introduction. In: Froehlich, A. (ed.) Integrated space for African society, legal and policy implementation of space in African countries. Cham, Switzerland: Springer Nature.

Slagter, B., Tsendbazar, N. E., Vollrath, A., & Reiche, J. (2020). Mapping wetland characteristics using temporally dense Sentinel-1 and Sentinel-2 data: A case study in

the St. Lucia wetlands, South Africa. *International Journal of Applied Earth Observation and Geoinformation*, *86*, p.102009.

Stratoulias, D., Balzter, H., Zlinszky, A., & Tóth, V. R. (2018). A comparison of airborne hyperspectral-based classifications of emergent wetland vegetation at Lake Balaton, Hungary. *International Journal of Remote Sensing*, *39*(17), 5689-5715.

Sun, Y., Xin, Q., Huang, J., Huang, B., & Zhang, H. (2019). Characterizing tree species of a tropical wetland in Southern China at the individual tree level based on Convolutional Neural Network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12*(11), 4415-4425.

Taddeo, S., Dronova, I., & Depsky, N. (2019). Spectral vegetation indices of wetland greenness: Responses to vegetation structure, composition, and spatial distribution. *Remote Sensing of Environment*, *234*, p.111467.

Thenkabail, P. S., Enclona, E. A., Ashton, M. S., & Bauke, V. D. M. (2004). Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote Sensing of Environment*, *91* (3): 354-376. doi:10.1016/j.rse.2004.03.013.

Tiner, R. W., Lang, M. W., & Klemas, V. (2015). Remote sensing of wetlands: Applications and advances. Boca Raton, FL: CRC Press/Taylor & Francis.

Transon, J., d'Andrimont, R., Maugnard, A., & Defourny, P. (2018). Survey of hyperspectral earth observation applications from space in the sentinel-2 context. *Remote Sensing*, *10*(2), 157.

Trustorff, J. H., Konrad, P. M., & Leker, J. (2011). Credit risk prediction using support vector machines. *Review of Quantitative Finance and Accounting*, *36*(4), 565-581.

Ustin, S. L., & Jacquemoud, S. (2020). How the optical properties of leaves modify the absorption and scattering of energy and enhance leaf functionality. In: *Remote Sensing of Plant Biodiversity* (pp. 349-384). Cham, Switzerland: Springer.

Van Ede, G. 2016. Flowers of Verloren Vallei: Field guide to the orchids and selected flowers of Verloren Vallei Nature Reserve, Wild Orchids Southern Africa, Bramely.

Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, *10*(5), 988-999.

Verma, N., Mishra, P., & Purohit, N. (2020). Development of a knowledge based decision tree classifier using hybrid polarimetric SAR observables. *International Journal of Remote Sensing*, *41*(4), 1302-1320.

Verrelst, J., Malenovský, Z., Van der Tol, C., Camps-Valls, G., Gastellu-Etchegorry, J. P., Lewis, P., North, P., & Moreno, J. (2019). Quantifying vegetation biophysical variables from imaging spectroscopy data: a review on retrieval methods. *Surveys in Geophysics*, *40*(3), 589-629.

Walter, M., & Mondal, P. (2019). A rapidly assessed wetland stress index (RAWSI) using Landsat 8 and Sentinel-1 radar data. *Remote Sensing*, *11*(21), 2549.

Warrens, M. J. (2015). Relative quantity and allocation disagreement measures for category-level accuracy assessment. *International Journal of Remote Sensing*, *36*(23), 5959-5969.

Wright, C., & Gallant, A. (2007). Improved wetland remote sensing in Yellowstone National Park using classification trees to combine TM imagery and ancillary environmental data. *Remote Sensing of Environment*, *107*(4), pp.582-605.

Xu, F., Ouyang, D. L., Rene, E. R., Ng, H. Y., Guo, L. L., Zhu, Y. J., Zhou, L. L., Yuan, Q., Miao, M. S., Wang, Q., & Kong, Q. (2019). Electricity production enhancement in a constructed wetland-microbial fuel cell system for treating saline wastewater. *Bioresource Technology*, *288*, p.121462.

Yadav, B. and Mathur, S., 2020. River discharge simulation using variable parameter McCarthy–Muskingum and wavelet-support vector machine methods. *Neural Computing and Applications*, *32*(7), pp.2457-2470.

Yuan, Z., Zhang, L., Wang, D., Jiang, J., Harrington, P.D.B., Mao, J., Zhang, Q. and Li, P., 2020. Detection of flaxseed oil multiple adulteration by near-infrared spectroscopy and nonlinear one class partial least squares discriminant analysis. *LWT*, *125*, p.109247.

Zafari, A., Zurita-Milla, R., & Izquierdo-Verdiguier, E. (2019). Evaluating the performance of a random forest kernel for land cover classification. *Remote Sensing*, *11*(5), 575.

Zhou, C., Wong, K. and Zhao, J., 2018. Coastal wetland vegetation in response to global warming and climate change. In *Sea Level Rise and Coastal Infrastructure* (pp. 61-81). Intech Open.

Zhu, G., Huang, Y., Li, S., Tang, J., & Liang, D. (2017). Hyperspectral band selection via rank minimization. *IEEE Geoscience and Remote Sensing Letters*, *14*(12), pp.2320-2324.

APPENDICES

Appendix 1 Adapted from McCoy (2005)

Spectral Data Collection- Field Notes Sheet

General Information

Plot No: _____ Lat: _____

Date: _____ Long: _____

Time: _____ Photo: _____

Operator: _____ Voice: _____

Other: _____

Environmental Information

LU Type: _____ Slope: _____

Soil Type: _____ Aspect: _____

Soil Moisture: _____ Geology: _____

Vegetation Information

Spp 1: _____ % Cover: _____ Phenology: _____

Spp 2: _____ % Cover: _____ Phenology: _____

Spp 3: _____ % Cover: _____ Phenology: _____

Spp 4: _____ % Cover: _____ Phenology: _____

Spectrometer Readings

Lens: _____ Date: _____ Solar Azimuth: _____

# of Scans averaged per scan: _____

Sensor Height above Target/FOV: _____

Scan #	Target	Time	Sky	Solar Z	Sensor A	Sensor Z	FOV	Comments

Comments:……
……
……