Estimation of nitrogen content across grass communities at Telperion Nature Reserve using Sentinel-2

by

Chabalala Yingisani Winny (1251589)

Supervisor: Dr Elhadi Adam

Co-supervisor: Dr Zakariyyaa Oumaer

A thesis report submitted in partial fulfilment of the requirements for the degree in Master of Science in GIS and Remote Sensing

Faculty of Science

University of the Witwatersrand

March 2017

Johannesburg, South Africa

Abstract

Grass nitrogen is the main indicator of forage conditions in a rangeland environment. The main objectives of the research were to map the quality and quantity of common grass communities and to predict Nitrogen (N) content across different grass communities. A machine-learning algorithm of Support Vector Machines (SVM) was tested in the mapping of grass quality and quantity. An overall accuracy of 72.68% was achieved for the mapping analysis which demonstrated the capability of the Sentinel-2 10m resolution in discriminating the spectral properties of different grass communities.

The foliar nitrogen was predicted using univariate regression, stepwise multiple linear regression (SMLR), multivariable regression methods, partial least square regression (PLSR) and random forest (RF). Foliar N was predicted using multivariate regression models; the best model was selected based on the highest coefficient of determination (R^2) value, and the low root mean square error (RMSE). The best RF model for foliar N estimation was based on the simple ratio (SR) index because the model attained the highest prediction accuracy of 35%. The study demonstrates the applicability of Sentinel-2 MSI utility in mapping and estimation of leaf N at a landscape scale .The results of both regression models (univariate and multivariate) such as random forest and partial least squares indicated that the inclusion of the Sentinel-2 MSI red edge bands provides an opportunity to accurately map and estimate leaf bio-chemical composition using remote sensing techniques.

Dedication

I dedicate this research to my husband and my children, George, Vuyelo and Vutivi Chabalala. Thank you for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of writing this thesis. This accomplishment would not have been possible without you.

Many thanks go to my father; Phineas Mathonsi and my late mum & my late sister-in -law; N'wa Ben Mathonsi and N'wa Mhlongo Mathonsi. Thank you for your support and encouragements.

ACKNOWLEDGEMENTS

All the honour and glory belong to the man above for and his tender mercies during the course of my studies.

I also express my gratitude to my supervisors Dr Elhadi Adam and Dr Zakkariyya Oumaer for all their guidance and support offered throughout my research. Dr Adam your door was always open whenever I had a question about my research. Your consistency and support steered me to the right direction and provided me with an opportunity to learn and improve my knowledge.

I also extend my gratitude to my master's lecturers (Dr Stefania Merlo and Dr Elhadi Adam) from the school of Geography, Archaeology and Environmental Studies. Thank you for all your persistence and willingness to guide and instil scientific knowledge on me.

Declaration 1 – Plagiarism

I, Chabalala Yingisani Winny (Student number: 1251589), hereby declare that:

- The "Estimation of nitrogen content across grass communities at Telperion Nature Reserve using Sentinel-2'' is my own work.
- All the sources quoted in this research have been re-rewritten and acknowledged in the reference list.
- This research report has never been submitted at any other university for examination or any degree.

Signature: Date: 30-03-2017

Chabalala Yingisani Winny

Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ABREVIATIONS

N=Nitrogen

RS=Remote Sensing

ESA= European Space Agency

MSI=Multispectral Instrument

REP=Red Edge Position

SA=South Africa

PLSR= Partial Least Square Regression

RF=Random Forest

SMLR=Simple Multiple Linear Regression

SWIR= Short Wave Infrared Red

Nm=Nanometre

GPS= Global Positioning System

SVM=Support Vector Machine

Ha= Hectares

M=Metre

Km=Kilometre

CHAPTER ONE

Introduction

1.1. General introduction

The interaction between the grassland and grazing pattern has a significant influence on grassland process through nutrient redistribution and cycling (Ling *et al.*, 2014a). Grazing affects the nourishing value of forage quality which changes the grazing density within the rangelands environments (Guo *et al.*, 2010). Nitrogen (N) is considered the greatest essential bio-chemical components that determines plants quality (Zhai *et al.*, 2013; Mutanga & Skidmore, 2004) and helps to understand the vegetation quality (Cho *et al.*, 2015a). Furthermore, it is also one of the principal nutrients limiting productivity, that affects species, determines the grazing and migration patterns of wildlife in the grassland ecosystem (Mutanga *et al.*, 2003). Nitrogen is a needed nutrient for plant development and its concentration helps to understand the operational dynamics of the natural environment (Wang *et al.*, 2010; Mutanga & Skidmore, 2004; Shi *et al.*, 2015). Leaf N indicates the quality of forage distribution, which determines the habitat selection by wildlife (Adjorlolo *et al.*, 2015; Ling et al., 2014; Shi *et al.*, 2015).

There is a need to develop remote sensing techniques to help understand the dynamics of vegetation species, the distribution of the grass quality and quantity at appropriate scales for rangeland management. Traditional techniques such as field-based method for mapping grass species quality and quantity have long been used at a local scale (Ramoelo *et al.*, 2015). However, the use of traditional methods comes with limitations of being laborious in nature, costly, complex time consuming and at times difficult to achieve due to inaccessibility which results in inadequate data collection and analysis (Ling *et al.*, 2014; Ramoelo *et al.*, 2013).

Conversely, remote sensing (RS) are the most preferred techniques over traditional methods because they provide timely and economical ways of vegetation monitoring by reducing the intensive field work sampling and laboratory analysis (Ramoelo *et al.*, 2013; Darvishzadeh *et al.*, 2008). RS is an innovative technique that can be used to assess forage nutrients at larger extents and used to access areas that are inaccessible such as mountains and densely vegetated areas (Running *et al.*, 1993).

The vegetation indices from broadband sensors such Landsat has been successfully applied to predict various grass biophysical factors (Darvishzadeh *et al.*, 2008; Ramoelo *et al.*, 2013a). One of the limitations of multispectral data in estimating N is that they lack appropriate spectral and spatial resolution for detecting and estimating the grass bio-chemical contents.

This was attributed to the fact that the optimal period for estimating the Leaf N is the peak production period during which the Normalised Difference Vegetation Index (NDVI) normally experience a problem of signal saturation (Mutanga & Skidmore, 2007a).

The potential of hyperspectral instruments with very high spectral resolution (10nm) has also been investigated in spectral discrimination and mapping of vegetation species (Adam *et al*., 2010), mapping and detecting of foliar nutrients (Mutanga *et al*.,2005b).However, the use of hyperspectral data in mapping and estimation of grass nutrients comes with limitations such as high cost particularly in the developing countries, unavailability and high data dimensionality that require special skills to process and analyze (Sibanda *et al*., 2016).

The arrival of the new generation sensors with appropriate resolution for land cover monitoring is perceived as an improvement to multispectral and overcoming the limitation of hyperspectral data (Ramoelo *et al*., 2012; Fu *et al*., 2014; Karlson & Ostwald, 2016). Imageries such as Worldview-2 & 3 and RapidEye incorporate the red edge bands which have been identified as key bands for predicting and mapping of grass quality and quantity (Ramoelo *et al*., 2012; Ramoelo *et al*., 2015; Sibanda *et al*., 2017; Huang *et al*., 2017). A study by Mansour *et al* (2016) used SPOT 5 image to map the grassland degradation and achieved an overall accuracy of 75.30%. The use of commercial data in mapping grass species was successful but also comes with limitation of cost which hindered their usability in forage nutrient analysis of larger areas especially in developing countries (Cho *et al*., 2015b).

In June 2015, a new high spectral resolution sensor (Sentinel-2 MSI) was launched by the European Space Agency (ESA) with spectral bands similar to those of WorldView-2 and RapidEye. Sentinel-2 is a multispectral instrument with 13 spectral channels and has a wide swath of 290 kilometres which will enable researchers to map vegetation species and estimate nitrogen at a wider area or even at country level (Drusch *et al*., 2012). Its configuration provides enhanced continuity to SPOT and Landsat, the inclusion of the short wave infrared bands will enable it to accurately estimate leaf N even during the wet season (Drusch *et al*., 2012). The image has a repeat cycle of five days and is available free of charge which will facilitate the use of earth observation for different applications especially in the developing countries where the accessibility to the high resolution remote sensing data has been limited due to the high cost.

1.2. Problem statement

The Telperion Game Reserve currently uses field based methods to collect data on different grass communities and wildlife which is time consuming and laborious. The effective management of the game reserve rely on up-to-date, accurate and highly detailed spatial information of species diversity and the distribution of grass quality to understanding the feeding pattern of the wildlife (Adam *et al*., 2010). Grass N is considered as one the most essential bio-chemical constituent of vegetation organic matter (Zhai *et al*., 2013) and a key indicator of rangeland quality (Song *et al*., 2016; Cho *et al*., 2015a). Browsers select forage with high quality and the quality of the forage has been reported to affect the behaviour of animals by various researchers but regardless, the mapping of grass quality at appropriate scales for management of animals has been a problem even though it can help understand the grazing patterns of wildlife (Mutanga *et al*., 2004).

The nutrients concentrations determine the grass quality and considering the importance that leaf N plays in understanding the distribution of wildlife, the mapping and estimating of leaf N content is crucial at rangeland management level. Effective management of the game reserves relies on accurate spatial information about the grass concentration that is up-to-date in order for the game managers to understand the movement, feeding pattern and the distribution of wildlife across the whole nature reserve. Estimation of the main bio-chemical nutrients such as leaf N is required at various spatial scales to monitor rangeland resources accurately (He & Mui, 2010).

The grassland ecosystems are characterised by various grass species that provide habitat to biodiversity. Assessing the geographic distribution of these grasses is important for conservation efforts because they provide habitat and pasture to biodiversity. The game rangers at Telperion game reserve currently use field-based approach for data collection and this method requires intensive labour and lab analysis (Mutanga *et al*., 2004). Remote Sensing is an effective innovation tool that can provide a synoptic spatial distribution of grass community distribution and the nutrient content (Darvishzadeh *et al*., 2008).

Remote Sensing methods for mapping grass quantity and quality are well established. However, these methods were developed using hyperspectral and commercial data, which is expensive especially for non-profit organization such as game reserves. Over the last few years, European Space Agency has released Sentinel-2 MSI data that therefore opens great opportunity for the use of earth observation data for different applications. Sentinel-2 MSI is an open source data that is freely available and therefore allows monitoring of the earth surface and provides more details at different scales which was otherwise impossible with the use of hyperspectral and commercial multispectral data due to cost. To our knowledge, there is no study in South Africa that was undertaken using Sentinel-2 in the grassland environment and it is, therefore, important to test its applicability in estimating the quality of different grass communities in the grassland environment. This could be of considerable benefit to nature conservationist, ecologists and resource managers because the results of this study will assist them in developing appropriate effective strategies suitable for wildlife production and management of grazing areas.

1.3. Aim and objectives

- The main aim of this study was to explore the use of Sentinel -2 MSI data in mapping the Nitrogen (N) concentration across different grass communities at Telperion Game Reserve.

1.3.1. Objectives

The specific objectives are to:

1. To map the common grass communities in the game reserve using Sentinel 2
2. Investigate the use of different vegetation indices calculated from Sentinel bands in estimation the nitrogen concentration across different grass communities at Telperion Game Reserve.

CHAPTER TWO

Literature review

2.1. Mapping grass communities quality and quantity

2.1.1. Introduction

The rangeland ecosystems accommodate different biodiversity including the endangered species. These ecosystems are characterized by various spatial variations within the grazing lands that need to be acknowledged by the managers and conservators for better planning of conservation strategies (Zheng *et al.*, 2015). Species based quality is more important for resource management because this information acts as a bench mark in understanding the distribution and densities of herbivores in protected areas (Ramoelo *et al.*, 2015).

Several studies have indicated that wildlife feeding and distribution patterns are largely controlled by the concentration of leaf nitrogen (N) content (Mutanga & Skidmore, 2007b). Therefore, the remote estimation of leaf N concentration is essential within the conservation areas to inform effective management strategies for wildlife browsing areas (Cho *et al.*, 2015; Mutanga & Skidmore, 2007). This will additionally assist in monitoring of the grazing processes (Zhai *et al.*, 2013), to understand the browsers feeding patterns as well as the health of the rangeland ecosystems (Mutanga *et al.*, 2004). Furthermore, the mapping of grass species will help in determining and understanding the forage quality and quantity spatial distribution, habitat selection as well as the migration patterns of some of the animals (Wang *et al.*, 2015; Mutanga & Skidmore, 2004). This is crucial because the geographic distribution of forage quality influences the movement and interaction of wildlife within the rangeland, the productivity of wildlife as well as the livestock populations (Mutanga & Skidmore, 2004).

The grassland ecosystems are characterised by various vegetation species that need to be preserved as they provide various important ecosystem services. Mapping the quality and quantity of these vegetation variations as a way of understanding how grazers change the structure and the normal functioning of the rangeland ecosystems is important for the prediction of future foraging patterns of browsers (Wang *et al.*, 2015) because this has been identified as one of the crucial components that is perplexing to the resource managers (Wang *et al.*, 2007). Studies in vegetation mapping have become an important theme today (Cingolani *et al.*, 2004) as they contribute valuable spatial information necessary for conservation management (Zhang *et al.*, 2016).

The mapping of vegetation quantity and quality has been traditionally based in the past and previous studies have been successful in estimating leaf N at local scale by using methods such as image spectroscopy and field surveys (Mutanga & Skidmore, 2004). For example, a study by (Mutanga & Skidmore, 2004) modelled the relationship between the concentration and reflectance of grass in the Kruger National Park and surrounding region in the laboratory, while Skidmore *et al* (2010) established the relationship between foliar nitrogen of grass and trees reflectance, again for the Kruger National Park. A study by (Goecker *et al*., 2005) evaluated how the concentration of foliar nitrogen content can be influenced by grazing on turtle grass (*Thalassia testudinal*) by the generalist invertebrate herbivore in the north-eastern Gulf of Mexico. The problem with the field techniques is that they are laborious, time-consuming and sometimes unviable to large areas containing poor accessibility (Adam *et al*., 2012). Predicting leaf N based on laboratory methods is complex and expensive (Ling *et al*., 2014 and Mutanga & Skidmore, 2004) while resulting in inadequate data collection (Ramoelo *et al*., 2012 and Cho *et al*., 2015).

Remote sensing (RS) methods provide reliable valuable information about the distribution of the forage quality which can be used to inform decisions about rangeland conservation (Zheng *et al*., 2015). Remote Sensing methods provide a synoptic view about the spatial distribution of grass which is necessary to understand the distribution and feeding of animals (Dzikiti *et al*., 2015; Running *et al*., 1993 and Adam *et al*., 2010). Previous work has indicated that, the survival of wildlife largely depends on the quality of vegetation as it is known to relate more to plants protein content which is the major nutrient requirements (Cho *et al*., 2015). The use of RS methods reduce the intensive field surveys and laboratory analysis required by traditional mapping techniques (Darvishzadeh *et al*., 2008) by allowing for access of inaccessible areas (Running *et al*., 1993). Mapping of the distribution of foliar N from remotely sensed images also emerges patterns, which assist in understanding the dominant drivers causing vegetation patterns (Skidmore *et al*., 2010).

2.1.2. Mapping grass quality and quantity using multispectral data

The mapping of grass and vegetation species has been achieved using multispectral data in the past. A study by Paudel & Anderson (2010) mapped and discriminated rangeland degradation in the Upper Mustang of the Trans Himalaya in Nepal using Landsat data.

Additionally, Boegh et al (2013) successfully managed to quantify the concentration of nitrogen of agricultural lands in Denmark using multispectral data. A similar dataset approach was used by (Belluco *et al.*, 2006; Song *et al.*, 2016; Adam *et al.*, 2010) when mapping vegetation communities and obtained an overall accuracy of 86% and 98% spatial variation while Wang *et al* (2007) used Quickbird-2 to map the terrestrial and submerged aquatic vegetation and achieved a total classification accuracy of approximately 82% and 75% respectively.

However, multispectral sensors such as Landsat and SPOT have not been widely used in mapping grass and vegetation communities because they lack the suitable spectral and spatial resolution (Ramoelo *et al.*, 2012). Adam *et al* (2012) stated that, multispectral data are inadequate in discriminating vegetation species because of their broad spectral wavebands and hence results in inaccurate and low classification accuracy (Belluco *et al.*, 2006). Furthermore, mapping vegetation and grass communities using multispectral data has been proven to be a difficult process in areas that are heterogeneous in nature such as mountainous areas (Feng *et al.*, 2015; Zhang *et al.*, 2016).

2.1.3. The mapping of grass quality and quantity using hyperspectral data

The utility of hyperspectral data has been successful in mapping grass species at a local level because of their narrow bands. Hyperspectral data has been widely used in mapping grass species communities because their narrow spectral bands are able to uniquely detect small variations that mostly exist in grassland ecosystem (Sibanda *et al.*, 2015). For example, a study by Mutanga *et al* (2004) discovered the use of continuum absorption approach as an improved method for estimating grass nutrients composition when predicting the quality of grass in the Kruger National Park of South Africa. On the other hand, a study by Sibanda *et al* (2015) managed to quantify the above ground biomass with an R^2 of 81% using hyperspectral data while Abbasi *et al* (2015) mapped tree species in the boreal forests using hyperspectral data and achieved 89% overall accuracy. Related outcomes were achieved by (Dalponte *et al.*, 2014; Anderson *et al.*, 1976; Guo *et al.*, 2010) when mapping vegetation species and delineating tree crown using hyperspectral data. (Mutanga & Skidmore, 2004; Skidmore *et al.*, 2010; Mutanga & Kumar, 2007) also used hyperspectral data to map grass bio-chemical nutrients (nitrogen and phosphorus) .

Although the results were successful, the use of hyperspectral data is often expensive for the mapping of large areas and is only applicable to local scales. The variation within vegetation classes has also made it difficult to accurately predict foliar N using hyperspectral data (Anderson *et al*., 1976). The use of hyperspectral data is hindered by cost, high data dimensionality caused by the narrow bands which makes data processing difficult (Belluco *et al*., 2006) and a need for a large training size training data to successfully discriminate mixed vegetation species (Adjorlolo *et al*., 2012).

2.1.4. The mapping of grass quality and quantity using new generation multispectral data

The arrival of new generation satellites consisting of suitable resolution is perceived as an enhancement to hyperspectral and multispectral data for predicting foliar nitrogen estimates (Schlemmera *et al*., 2013, Mittapalli *et al*., 2014; Antonio, 2014; Li *et al*., 2014). Worldview-2 and RapidEye have increased spectral and spatial resolution with their bands combinations allowing for grass quality estimation at various scales (Ramoelo *et al*., 2012). Both the image include the red edge position which is sensitive to plant materials and important for vegetation assessment (Ramoelo *et al*., 2015a). Several studies have been able to map grass species in the past using these new sensors. A study by (Adjorlolo & Botha, 2015) evaluated the applicability of the Worldview -2 image in predicting and mapping the African grass foliar nitrogen concentration. Recently, Lück-Vogel *et al* (2016) mapped estuarine vegetation in St Lucia using multispectral data with satisfactory results. However, the use of commercial satellites comes with the limitation of cost which has resulted in their limited usage in forage nutrient analysis of larger areas (Ramoelo *et al*., 2015a).

The effective management of conservation areas depends largely on the availability of well-timed accurate spatial datasets and identification of techniques that can be used to assess different conservation management practices at a regional level (Zheng *et al*., 2015; Sibanda *et al*., 2017). The availability of Sentinel-2 MSI free of charge provides an opportunity to map vegetation species quality and quantity because of its spectral and spatial characteristics making it suitable for rangeland management at the regional level, which was not feasible before.

2.2. Conclusion

The mapping of different grass species has been successful in the past using various remote sensing techniques mostly at a local scale because remote sensing images suitable at the regional level were commercialized. The availability of the new Sentinel-2 MSI image at no cost provides an opportunity to map grass species at a regional and national scale, which was not possible before. The sensor has not been fully utilized in grassland studies but its application in forestry has indicated that it is stronger in estimating grass quantity such as biomass (Frampton et al. 2013).

Therefore, it is important to examine the applicability of the new (Sentinel-2 MSI) in mapping grass bio-chemical and biophysical composition at rangeland level, to see whether the inclusion of the unique spectral bands located within the red edge position can improve the mapping of grass quality and quantity as it has been shown with previous existing commercial multispectral sensors (RapidEye, Worldview, SPOT, etc.).

CHAPTER THREE

Mapping grass communities using Sentinel-2 MSI

3.1. Introduction

The rangeland environments are characterized by different grass communities that provide different nutrients to the wildlife community. The animal survival in the rangelands ecosystems depends largely on the conservation and management strategies applied by the game managers. Therefore, discriminating these environs in order to understand their spatial distribution is crucial for the normal functioning of the natural ecosystem which supports biodiversity (Ramoelo *et al.*, 2012). Deriving accurate thematic information of grass species is essential for effective natural resource management and conservation practises *(*Corbane *et al.*, 2014; Burai *et al.,* 2015; Ramoelo *et al.*, 2012). Field based survey for mapping grass communities requires intensive fieldwork and laboratory analysis for grass identifications. This often results in the collection and analysis of data that is not generally representative of the grass communities due to the large area, poor accessibility and high cost (Burai *et al.*, 2015).

Remote sensing potentially offers cost effective methods to map grass communities by reducing the intensive field sampling and laboratory analysis required by field-based mapping techniques (Lewis *et al.*, 2013). Both multispectral and hyperspectral data have been successfully used in mapping grass communities (Luther *et al.*, 2006). However, mapping grass species and communities is challenging with multispectral data, because they lack suitable spatial and spectral resolutions (Sibanda *et al.*, 2017).

Hyperspectral data has been widely used in mapping vegetation species and communities across different ecosystems (Schmidt & Skidmore, 2003; Sibanda *et al.*, 2015). The data has been found to be useful in providing accurate and detail detection of grass species because of their narrow bands which are sensitive to vegetation characteristics (Si *et al.*, 2012). Recently, a study by Skowronek *et al* (2017) mapped invasive species using airborne imaging spectroscopy in the United States and achieved 74% overall classification accuracy. However, there is still a challenge in processing and analysis of hyperspectral data because of the high data dimensionality. Furthermore, the high cost of data have limited the use of hyperspectral data particularly in the developing countries (Sibanda *et al.*, 2017, Ramoelo *et al.*, 2015; Van Deventer *et al.*, 2015).

Over the last decade, a collection of new-generation multispectral sensors has emerged. These new sensors are seen as an improvement to hyperspectral and traditional multispectral data. The new multispectral sensors such as Worldview-2 and 3 (WV) and RapidEye, were launched to provide high spatial resolution with unique spectral bands which are arranged within the red edge position (Adelabu *et al*., 2015). Different studies have investigated the use of the new generation sensors in mapping vegetation species and communities (Baumstark *et al*., 2016; Rapinel *et al*., 2014; Omer *et al*., 2015; Odindi *et al*., 2014). However, despite the high precisions achieved on these studies, their application in mapping and monitoring the ecosystems raises concern because of the high cost involved when acquiring this remotely sensed data.

The launch of the new Sentinel-2 MSI image available free of charge offers high spatial and spectral resolutions. Sentinel-2 has high spatial resolution with unique spectral bands strategically positioned in the red edge position which is important for mapping vegetation at species level (Cho *et al*., 2012; Hedley *et al*., 2012). As far as we know, the sensor has never been tested in discriminating grass species at landscape level. Therefore, the main aim of this study is to evaluate the capabilities of the new sentinel-2 MSI in mapping different grass communities at Telperion Game Reserve.

3.2.Materials and methods

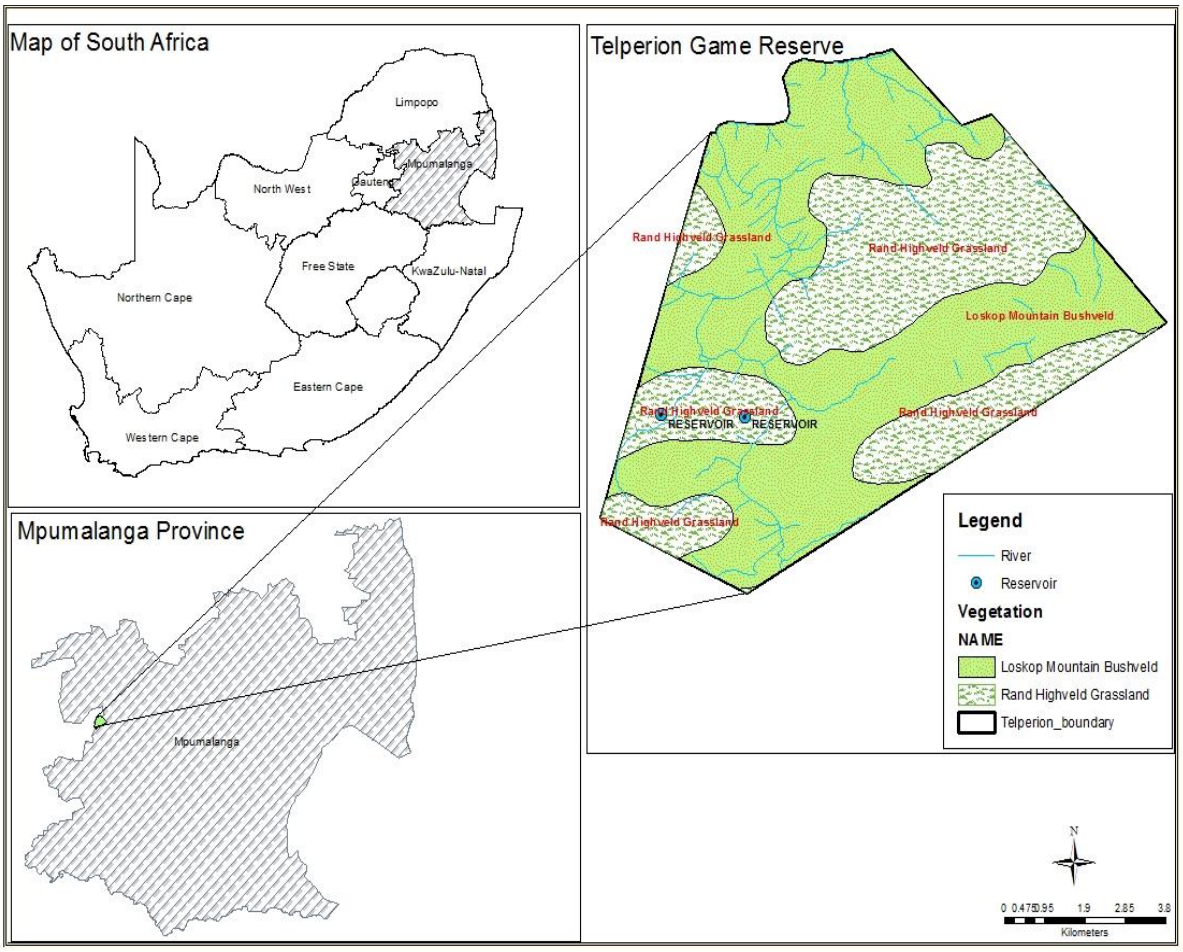
3.2.1. Study area

The Telperion Game Reserve depicted on (Figure 1) below is part of the eZemvelo Nature Reserve. The study area has area coverage of approximately 11 000 hectares (ha) and forms part of the 13000 ha of eZemvelo Nature Reserve. Telperion is located at 25^0 41’ South and 28^0 56 East at the border between Gauteng and Mpumalanga Provinces (Mucina & Rutherford, 2006). The reserve was established in 2008 and still belongs to the E Oppenheimers and Son Private Limited. Although fences do not separate them, the Wilge River geographically separates the two reserves.

Today eZemvelo and Telperion Nature Reserve house 33 game species and 300 bird species. The Ezemvelo Nature Reserve is located about 24 kilometres northeast of Bronkhorstspruit on the border of Gauteng and Mpumalanga provinces in South Africa. Telperion falls within Mpumalanga and is located between southern latitudes 25^0 38’ and $25^0$45’, and eastern longitudes 28^0 55’ and 29^0 03’. The game reserve is surrounded by farmlands practicing mainly crop, agriculture and cattle production.

The vegetation type is classified as Sourveld Mixed bushveld and Bankenveld by Mucina and Rutherford (2006). These vegetation types are listed as endangered species and protected in the conservation legislative of private reserves. The vegetation species prevalent in the sampled plots include wetland grass, mixed grasslands, woody vegetation, *Hyparrhenia hirta Eragrostis chlomeras, Serephium plumosum, Aristida congesta; Eragrostis gummiflua,* and *Cynodon dactylon.* The temperatures generally range from 4^{0}C to 18°C during winter and 14^{0}C to 26 °C in summer. Telperion Game Reserve lies at the altitude of between 1240 to 1500 metres above sea level. The landscape in the eastern side is characterized by the highlands while the western section is comprised of flat to undulating terrain characterized by valleys and ridges. The eZemvelo Nature Reserve lies at the altitude of between 1240 to 1500m above sea level. The landscape in the eastern side is characterized by mountains as depicted by image one below while the western section is comprised of flat to undulating terrain.

Figure 1: Map of the Telperion Game Reserve**.**



3.3. Remote sensing data acquisition and pre-processing

3.3.1. Sentinel -2 Multi Spectral Instrument image acquisition

One scene of Sentinel-2 MSI that covers the whole of eMalahleni area was downloaded from the ESA website (earthexplorer.usgs.gov). Sentinel-2 is a multispectral high-resolution image that is used for land monitoring. The image has a twin satellite capability with a frequent and systematic coverage that allows for mapping and prediction of grass biophysical and bio-chemical parameters (ESA, 2016). Sentinel-2 was designed to offer data continuity to previous missions such as LANDSAT and SPOT. The sensor has a spatial resolution and a repeat cycle of five days that allows for the provision of geospatial information at local, regional, national and even international scales. The sensor used a push broom scanner to record scenes and has a swath width of 290 kilometres (km). It also contains 13 spectral bands with spatial resolutions ranging from 10, 20 and 60 metres (Table 1). Its benefits are attributed to its finer spatial resolution as compared to previous sensors as well as the four narrow bands that are strategically positioned in the red edge region as well (Majasalmi & Rautiainen, 2016). Based on its unique spectral and resolution configurations to the previous sensors, Sentinel-2 is expected to significantly contribute to climate and biodiversity variables (Ramoelo *et al*., 2015).

Table 1: Spectral and spatial resolution of Sentinel-2 MSI.

Band Name	**Band Central wavelength (nm)**	**Band description**	**Band Width(nm)**	**Spatial resolution**
Band 1	443nm	Aerosols	20	60
Band 2	490nm	Blue	65	10
Band 3	560nm	Green	35	10
Band 4	665nm	Red	30	10
Band 5	705nm	RedEdge 1	15	20
Band 6	740nm	RedEdge 2	15	20
Band 7	783nm	RedEdge 3	20	20
Band 8	842nm	NIR	115	10
Band 8A	865nm	RedEdge 4	20	20
Band 9	945nm	Water vapor	20	60
Band 10	1375nm	Cirrus	30	60
Band 11	1610nm	SWIR1	90	20
Band 12	2190nm	SWIR2	180	20

3.2.2. Image pre-processing

The Sentinel-2 image was geo-referenced and geo-coded by selecting the correct UTM zone 35S and resampled to 10m spatial resolution. The study is one of the applications that require the image to be corrected atmospherically before running the analysis to get the true reflectance values of the objects on Sentinel-2 data. Therefore, the image was atmospherically corrected using Level 2A Prototype processer (Sen2Cor) as described in the ESA Snap software package. Sen2cor is an atmospheric correction developed by TPZV on behalf of ESA that works in conjunction with Anaconda (Python 2.7) and the correction process is performed using a python script (ESA 2015). The corrected image was loaded into ESA Snap software and converted to ENVI format, which resulted in 10 separate bands. The converted bands were then displayed on ENVI classic, stacked chronologically into one image again converted to ArcView raster format for further analysis. The spectral reflectances from the Sentinel-2 image that correspond to each sampled GPS points were extracted. Vegetation indices consisting of the SRs and NDVIs were computed from the reflectance values using the R programming language.

3.3. Field data collection

3.3.1. Ground control points/ Field data collection

The field grass survey was carried out to identify the common grass communities in the game reserve with the help of ecologists who used the field guide to grass species for Southern Africa (Van Oudtshoorn, 2014). Fieldwork was carried out from the 21st to 23rd of May 2016, which was within the window period of the image. The grass species prevalent in the sampled plots include wetland grass, mixed grasslands, woody vegetation, *Hyparrhenia hirta Eragrostis chlomeras, Serephium plumosum, Aristida congesta; Eragrostis gummiflua,* and *Cynodon dactylon.*

The sample plots were randomly established and spread evenly across the study area (Ramoelo *et al*., 2012) using a 10 metres x 10 metres to account for the Sentinel-2 pixel size of image (10m). A global positioning system receiver (GPS was used to record the latitude and longitude coordinates at each sample plot and for each common grass species, 80 GPS points in total were recorded. Prior classification, the dataset was splinted randomly into 70% training and 30% training data sets using a python script in ArcGIS 10.3. The regions of interest (ROIs) were created by overlaying the splinted datasets (training and testing) separately on Sentinel-2 image in ENVI 5.3. (Table 2).

Table 2: Training and testing datasets collected grass species mapping in the study area**.**

Species	**Code**	**Test samples (70%)**	**Training samples (30%)**	**Total samples**
Serephium plumosum	SP	15	53	68
Hyparrhenia hirta	HH	14	44	58
Mixed grassland	MG	102	186	228
Cynadon dactylon	CD	17	30	47
Eragrostis chloromelas	EC	21	53	74
Woody vegetation	WV	41	94	135
Wetland grass	WG	19	41	60
Aristida congesta	AC	13	46	59
Eragrostis gummiflua	EG	13	51	64

3.4. Image classification

Image classification is considered as one of the important steps in remote sensing studies because it retrieves and transforms vegetation information from satellite images and group them into thematic layers for easy interpretation (Cingolani *et al*., 2004; Schmidt & Skidmore, 2003). The algorithm is a non-parametric supervised classifier (Cortes and Vapnik, 1995a) that was developed by Vapnik (1979) as a binary classifier that uses a hyperplane to discriminate classes and identify patterns between separate classes, which in turn minimises misclassifications.

The SVM algorithm uses a kernel function as an optimizer of the non-linear procedure to map the input data space into a high-dimensional feature space while avoiding over-fitting and multi-dimensional problems that mostly exist in remotely sensed data (Ramoelo *et al*., 2012). Support Vector Machine is a machine learning technique that is widely applied to classify remote sensing images (Ramoelo *et al*., 2012; Corbane *et al*., 2014). A study by Darvishzadeh *et al* (2008) investigated the potential of the SVM classifier by estimating several physiological parameters (leaf nitrogen content, leaf chlorophyll content, etc.) of crops using hyperspectral data. (Yuan & Zhang, 2008) successfully applied SVM and spectral data to estimate the phosphorus content present in cucumber leaves. Therefore, this algorithm

(SVM) was selected as a classifier for this study to map the different grass communities at Telperion Game Reserve. The SVM classification was performed using the Support Vector Machine classification tool in ENVI 5.3.

3.5. Accuracy assessment

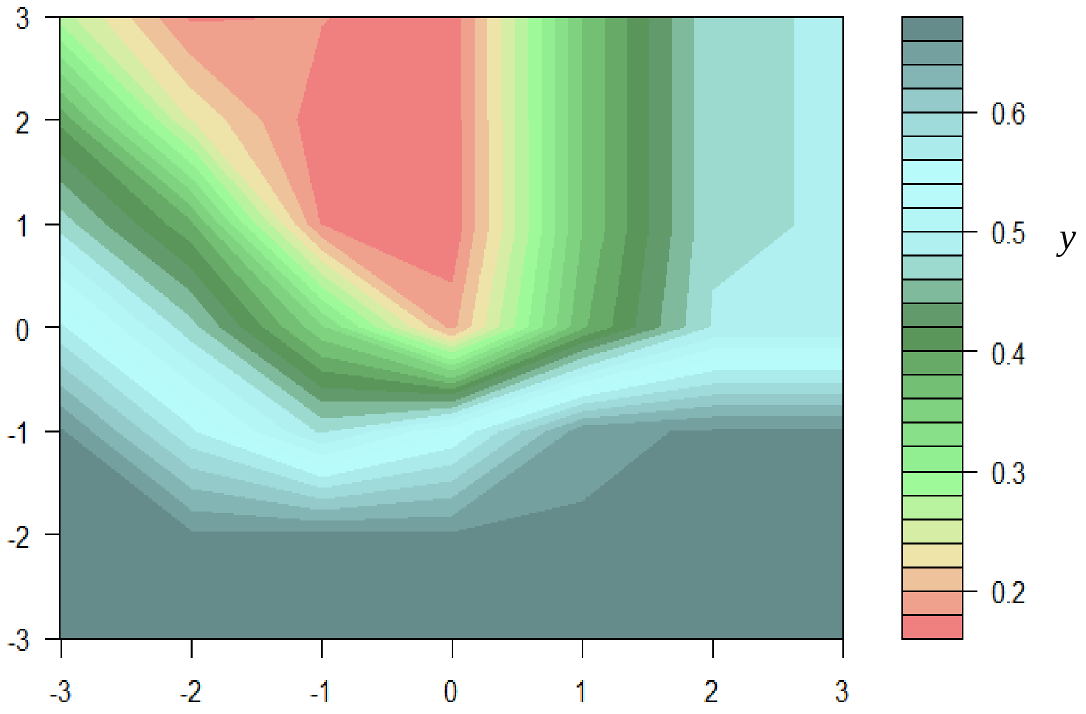
The accuracy of the SVM classifier was assessed using the 30% test data (Table 2). This was done to analyse the mapping predictive power of the grass species by using the SVM algorithm on Sentinel-2 image. A confusion matrix was generated to compare the producer and user accuracies. The producer's accuracy corresponds to the error of omission and demonstrates the likelihood that a specific grass species types on the map are correctly classified, while the user's accuracy corresponds to the error of inclusion and demonstrate the probability that a pixel labelled as specific grass species in the map is the definite class on the ground. The total accuracy of the grass species was calculated based on the total number of correctly classified pixels divided by the total number of pixels.

4. RESULTS

4.1. Optimization of the SVM Parameters

Optimization of the SVM parameters used for classification the SVM-Kernel radial function. The SVM classification model was optimized to select the best parameters to use in training the classification algorithm of nine grass species. A 10-fold cross-validation was used which yielded a gamma (Y) of 1, the lowest error was produced from the combination of *g*amma (γ) value of 0.1 and cost (C) value of 100 (Figure 2).

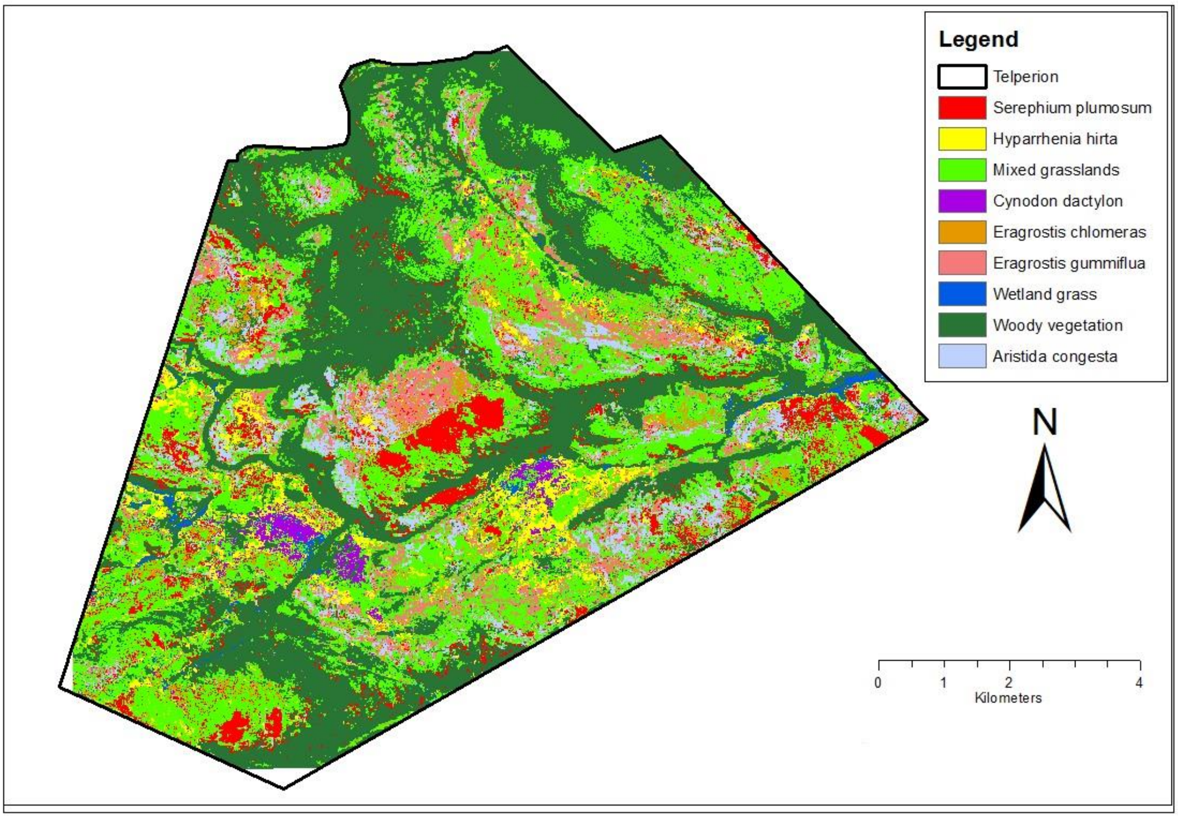
Figure 2: Parameter optimization for the Support Vector Machines using the 10-fold search radius.



4.2. The performance of the Support Vector Machine in grass species classification.

The SVM classification was executed using 10 bands of the Sentinel-2 image acquired over Telperion Game Reserve. The image was classified using EnMAp-Box version 2.2.1 software, which is an IDL Virtual Machine application. (Figure 3) shows that the SVM classifier has managed to successfully provide the geographic distribution of different grass species in the study area. The classification statistics indicates that the most dominate species are grasslands followed by the woody vegetation, the least occurring being the *Eragratis Chloromelas*.

Figure 3: Map illustration different grass communities at Telperion Game Reserve using SVM algorithm.



4.3. Accuracy assessment

The performance of the SVM classifier was evaluated by using the test dataset. The SVM algorithm yielded an overall classification accuracy of 72.68% with kappa coefficient of 67.78% meaning that there is 68% better agreement on the overall classification achieved than by chance alone. The grass species achieve a user's accuracy of 80%. The user's accuracy indicates that over 90% accuracy was achieved in all grass species. A misclassification can be noted between *Cynadon dactylon* (CD), *Eragrostis gummiflua (EG)* and *Hyparrhenia hirta* (HH) which yielded lower user accuracies of 59, 52%, 53, 45% and 64,29% respectively. This means that spectral similarities exist between these three grass species. The SVM classification yielded an overall accuracy of 72, 66% with a kappa coefficient of 67.78% (Table 3). The grey diagonals represent pixels that were correctly classified according to reference data while the off-diagonals represent the mis-classified pixels.

Table 3: A confusion matrix using Support Vector Machine for *Serephium plumosum* (SP), *Aristida congesta* (AC), Mixed grassland (MG), *Eragrostis gummiflua* (EG), *Cynadon dactylon* (CD) ,Woody vegetation (WV), Eragrostis chlomeras (EC), and Wetland grass (WG) and *Hyparrhenia hirta* (HH). User's accuracy (UA) and Producer's accuracy (PA).

Class	**SP**	**AC**	**EC**	**CD**	**WV**	**WG**	**EG**	**HH**	**MG**	**Total**	**UA**	**PA**
SP	36	5	5	0	0	0	2	0	0	48	82,35%	84,00%
AC	4	20	7	0	1	0	4	3	3	42	64,29%	54,55%
EC	9	2	20	0	7	0	1	0	0	39	86,36%	57,58%
CD	0	0	0	10	0	2	0	2	0	14	92,31%	46,15%
WV	1	0	0	0	75	2	1	0	1	80	96,10%	88,10%
WG	1	0	0	0	0	20	0	1	0	21	95,83%	92,00%
EG	0	1	1	8	1	0	22	3	0	36	59,52%	78,13%
HH	0	2	0	8	0	1	2	24	0	37	53,45%	93,94%
MG	0	3	0	0	0	0	0	0	20	23	80,00%	83,33%
Total	50	33	33	26	84	25	32	33	24	283		
Overall Accuracy=72,66% Kappa coefficient =67.78%.												

5. Discussion

The mapping of vegetation species plays a major role to biodiversity in rangeland ecosystems. Obtaining up-to-date information about the spatial distribution of different grass communities is important to game managers because it helps them to understand the movement, feeding pattern and the distribution of wildlife across the whole Game Reserve. The availability of sentinel-2 MSI with high spatial resolution provides a great potential of achieving better classification results for grass species mapping.

The results of the study have demonstrated the ability of the Sentinel-2 in discriminating the spectral characteristics of different grass species. The importance of the red edge bands was tested in this study by classifying the image with and without the red edge bands and it has been noted that the inclusion of the red edge bands increased the classification accuracy.

A study by Ramoelo *et al* (2012) indicated that the red edge bands are the most important for mapping grass quality and quantity because they are sensitive to plant bio-chemical materials (Cho *et al*., 2015). The unique spectral regions (red edge) of the image make it suitable to discriminate the different vegetation communities because the red edge region makes it tolerant of the effects of the soil (Clevers *et al*., 2001). The literature reviewed showed that a high correlation exists between the red edge region with grass bio-chemical nutrients (Delegido *et al*., 2013).

The overall classification accuracy for the producer and user accuracies that were obtained through the execution of the SVM in this study were significantly higher for all classes, which indicate the ability of the Sentinel-2 in discriminating different grass communities. The results of this study are synonymous with the result achieved by (Zheng *et al*., 2015) when mapping vegetation using multispectral sensor (RapidEye) and the SVM classification algorithm. The results highlight the capability of Sentinel-2 data, in discriminating different grass communities and highlight the predictive power of the SVM algorithm. The Support Vector Machines have the ability to classify a large number of classes using limited support vectors as trainings without compromising the accuracy (Zheng *et al*., 2015).

6. Conclusion

The aim of this study was to map different vegetation species communities at Telperion Game Reserve using the newly released generation Sentinel 2 MSI utility data. The results of the study have shown that the combination of Sentinel-2 MSI and SVM classifier was successful in mapping vegetation at Telperion Game Reserve.

The misclassifications seen in this study can be attributed to the spectral similarities between the grass communities. Consequently, in situ spectral measurement analysis should be further explored to better understand the spectral signature between the grass communities. These promising results are encouraging to conservators to test the use of the free data to better the conservation management as compared to the traditional methods that is expensive and time-consuming.

The aim of this study was to investigate the use of Sentinel-2 data in mapping vegetation species and communities at Telperion Game Reserve. The results of the study have shown that the SVM classifier was successful in mapping vegetation at Telperion Game Reserve as it produced an overall classification accuracy of 72.6% with kappa coefficient of 0.68%. Support Vector Machines require limited training sites (support vectors) which are used to optimally separate hyperplanes that differentiate land cover classes. Furthermore, the study has demonstrated the capability of the Sentinel-2 data in mapping the vegetation species of the rangeland ecosystems. The results are encouraging to conservationists and are an alternate method of mapping vegetation as compared to the traditional methods that are expensive and time-consuming.

CHAPTER FOUR

Mapping grass nitrogen (N) across different grass communities

4.1. Introduction

The concentration of foliar nitrogen directly indicates the rangeland quality and provides imperative information for effective wildlife and livestock management. In the rangeland ecosystem, the spatial concentration patterns of grass nitrogen (N) is found to be one of the key indicators of nutrient limitation that influence grazing behaviour of livestock and wildlife (Ramoelo *et al*., 2013). For example, the large herbivores in South Africa are known to be found around nutrient rich areas such as termite moulds and beneath large trees (Treydte *et al*., 2007).

The survival of animal largely depends on the quality of vegetation as it is known to relate more to plants protein content which is the major nutrient requirement (Ramoelo *et al*., 2015). Mapping the vegetation quality as a way of understanding how grazers change the structure and normal functioning of the rangeland ecosystems is important since these activities affect the future foraging pattern of browsers (Wang *et al*., 2015). Integrating this information with RS measurement and analysis will enable managers to attain a landscape-wide view of the nutrient distribution and identify areas with nutrients limitations. Studies have also shown that herbivore diversity increase with increasing grass diversity level (Feng *et al*., 2016). Therefore, accurate and reliable assessment of grass quality across different grass communities plays a vital role in effective management of grassland for sustainable grazing production and understanding the feeding patterns of wildlife (Beeri *et al*., 2007; Knox *et al*., 2010).

Remote sensing images have been widely used to estimate foliar N, especially in the grassland ecosystems to help determine the carrying capacity of various forage types depending on the quality of the available grass communities (Beeri *et al*., 2007).The estimation of foliar bio-chemical has been done in the past using traditional vegetation indices calculated from multispectral broadband such as SPOT and Landsat (Darvishzadeh *et al*., 2008, Ramoelo *et al*., 2013b and Mutanga *et al*., 2004).

However, the NDVI derived from the Landsat images were found to be problematic as they were only applicable to grass biophysical mapping because they lack good spectral resolution such as red edge position, which is sensitive and widely used to estimate foliar bio-chemical (Mutanga & Skidmore, 2004). The NDVI from Landsat were mostly used to estimate biomass but cannot be used to estimate grass N because they are reported to be unstable and their broader bands mask away various important bio-chemical nutrients. As indicated by (Mutanga & Skidmore, 2007; Cho *et al*., 2007; Mutanga *et al*., 2003) leaf N can be estimated successfully during the peak production and the NDVI method is known to have problems of signal saturation during this period because of the absence of the red edge band which has the ability to highlights detailed bio-chemical features (Ramoelo *et al*., 2013). This has also hindered the use of the traditional multispectral data in the production of forage N maps at a country scale (Ramoelo *et al*., 2013b).

Hyperspectral data has been applied in estimating foliar bio-chemical and N because they contain the red edge band, which is sensitive to foliar bio-chemical nutrients. The literature reviewed show that, hyperspectral data has a strong ability of discriminating small grass variations because of their narrow bands (Mutanga *et al*., 2003; Adam *et al*., 2012). Various studies have recently shown that, the red edge has the ability to provide spectral information that are missing in the Landsat visible and the near infrared bands because the red edge band is less delicate to atmospheric effects and soil characteristics (Mutanga & Skidmore, 2007). A study by Mutanga *et al* (2003) obtained promising results when discriminating tropical grass nutrients using hyperspectral data. On the other hand, a study by (Beeri *et al*., 2007), used hyperspectral images to estimate forage quality of northern mixed-grass prairie in the Unites States (US) and obtain an accuracy of 80%. Similar results were obtained by (Knox *et al*., 2010) when mapping savannah forage quality during the dry season using hyperspectral Carnegie Airborne Observatory Sensor. However, the cost and high dimensionality of hyperspectral data have hindered their usage in mapping of grass N particularly in the developing countries, limited spatial coverage, unavailability, high dimensionality which require special skills to process and analyse (Sibanda *et al*., 2016).

The arrival of the new generation satellites with reasonable resolution is perceived as an improvement to multispectral and hyperspectral data (Clevers & Gilten, 2013; Li *et al*., 2014) in predicting foliar nitrogen estimates. Imageries such as Worldview-2 and RapidEye both incorporate the red edge band, which is very much important for vegetation assessment and monitoring (Mutanga *et al*., 2015). Several researchers have been able to produce nitrogen maps in the past using these new sensors for capturing the variation of various biophysical parameters such as Nitrogen. A study by Mutanga *et al* (2015) evaluated WorldView-2 image in predicting the African grass foliar nitrogen concentration. On the other hand, (Sibanda *et al*., 2017) tested the capabilities of Worldview data red edge band in mapping nitrogen in the grasslands and obtained an overall accuracy of 78%. This is the widely used sensor because it offers new wavebands (red edge and near infrared) with a 2 metre spatial resolution than the traditional multispectral and hyperspectral satellites images such as SPOT and Hyperion. However, the limitation of new generation sensors is that they are commercial and their cost has limited their utilisation in forage nutrient analysis of larger areas especially in developing countries (Ramoelo *et al.*, 2015). The authors further mention that their progress is also hindered by the expensive data that is not even usable in the long term.

The launch of high spectral resolution sensor (Sentinel-2 MSI) by the European Space Agency (ESA) with spectral bands similar to those of RapidEye and WorldView-2 has brought back hope to remote sensing of vegetation at no cost. The sensor has spectral bands that are strategically placed in the red edge position, which will allow leaf N to be mapped at a national scale (Ramoelo *et al.*, 2015). Sentinel-2 is a multispectral instrument with 13 spectral channels in the visible, red edge, near infrared and shortwave infrared spectral range (Drusch *et al.*, 2012). According to Drusch *et al* (2012), Sentinel-2 MSI has additional channels in the red edge bands, which incorporate the short wave infrared bands, and this will enable it to accurately estimate leaf N even during the wet season (Drusch *et al.*, 2012). The image has a repeat cycle of five days with free open access to the data, which will make it possible to map the nitrogen content of grasslands in southern Africa cost effectively. The study entailed testing Sentinel-2 red edge bands for grass bio-chemical mapping at rangeland level, which was never possible before using Random Forest (RF) technique. The main aim of this study was therefore to investigate the use of Sentinel-2 red edge bands in mapping vegetation species and communities at Telperion Game Reserve. The specific objectives of the study were to (a) map the common grass communities in the game reserve using Sentinel; (b) investigate the use of different vegetation indices calculated from Sentinel bands in estimation the nitrogen concentration across different grass communities in the game reserve.

4.2.Materials and methods

4.2.1. The study area

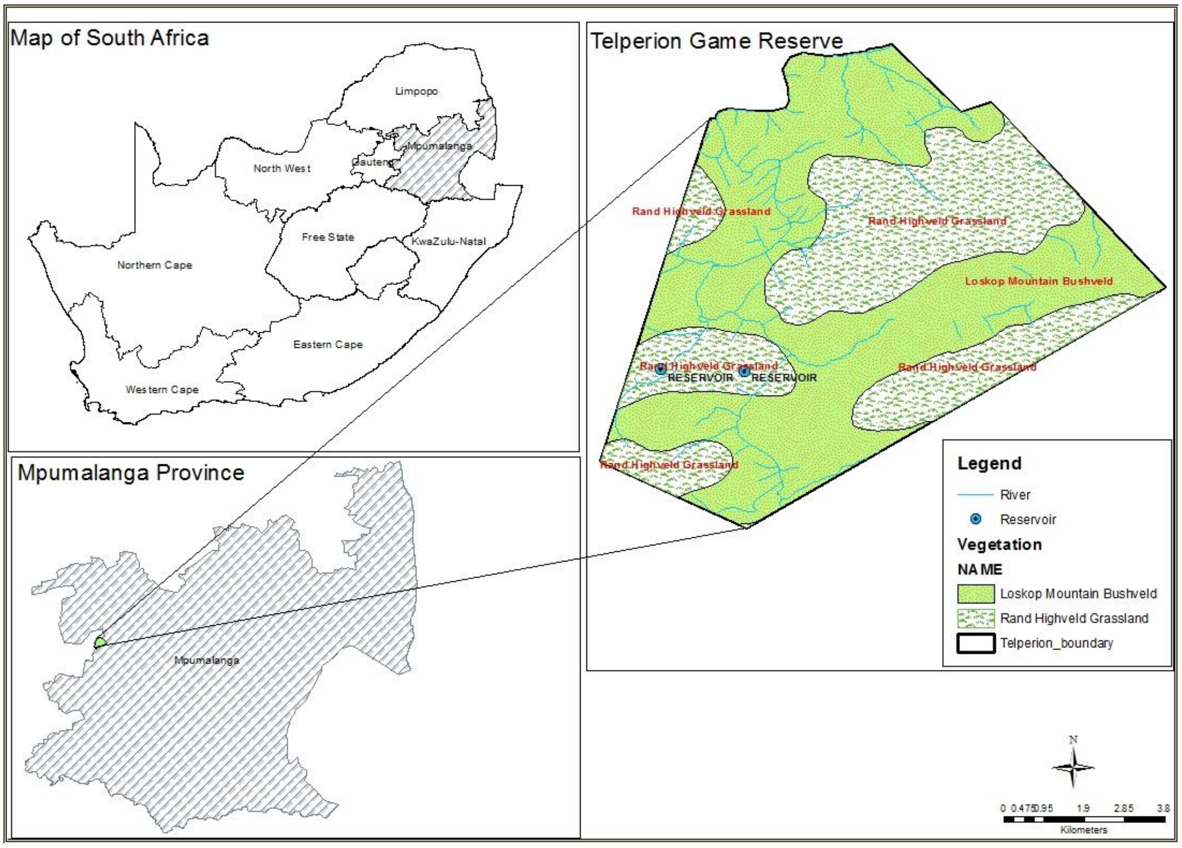
The study area is situated at Telperion Game Reserve which is part of eZemvelo Nature Reserve and has an area coverage of approximately 11 000 hectares (ha) and forms part of the 13000 hectares of eZemvelo Nature Reserve. Telperion is located at 25^0 41’ South and 28^0 56 East at the border between Gauteng and Mpumalanga Provinces (Mucina Rutherford, 2006). The game reserve was established in 2008 and still belongs to the E Oppenheimers and Son

Private Limited. The game reserve falls within the grassland biome in the far north and is part of the 13000ha eZemvelo Nature Reserve. Although fences do not separate them, the Wilge River geographically separates the two reserves.

Today eZemvelo Nature Reserve and Telperion Game Reserve house 33 game species and 300 bird species. The ezemvelo Nature Reserve is located about 24 kilometres northeast of Bronkhorstspruit on the border of Gauteng and Mpumalanga provinces in South Africa. Telperion falls within Mpumalanga and is located between southern latitudes 250 38’, 25045’, and eastern longitudes 28055’ and 29003’ (Figure 4). The game reserve is surrounded by farmlands practicing mainly agriculture and cattle production.

The vegetation type is classified as Sourveld Mixed bushveld and Bankenveld by Mucina and Rutherford (2006). These vegetation types are listed as endangered species and are protected in the legislative conservation of private reserves. The vegetation species prevalent in the sampled plots include wetland grass, mixed grasslands, woody vegetation, *Hyparrhenia hirta Eragrostis chlomeras, Serephium plumosum, Aristida congesta; Eragrostis gummiflua,* and *Cynodon dactylon.* The temperatures generally range from 4^{o}C to 18°C during winter and 14^{o}C to 26°C in summer. Telperion Game Reserve lies at the altitude of between 1240 to 1500m above sea level. The eZemvelo Nature Reserve lies at the altitude of between 1240 metres to 1500 metres above sea level. The landscape in the eastern side is characterized by the highlands while the western section is comprised of flat to undulating terrain.

Figure 4: Map showing the study area.



4.3. Data collection

4.3.1. Field data collection and chemical analysis

The fieldwork was undertaken in May 2016 which is considered early grass dry season and the grass samples were collected in-situ. In this study, 80 experimental plots with a length of 10m x 10m were used to account for the 10m geometric resolution of Sentinel-2 image. In each plot, four to seven grass samples were collected to represent dominant species within the plot. The samples were then oven-dried at 70˚C for 48 hours. The dried samples were then bagged and transported to the Department of Agriculture and Rural Development laboratory in Cedara, Kwa-Zulu Natal, South Africa for chemical analysis and extraction of grass N. To determine the nitrogen concentrations, Dumas method was used which is based on dry combustion. During the process, the grass samples were burned horizontally at 1 350^{0}C and then passed through the infrared detection cells to quantitatively determine the level amount of nitrogen concentration.

4.3.2. Computation of vegetation indices

Vegetation indices have been widely used in the field of environment to estimate grass bio-chemical composition (Ramoelo *et al*., 2015; Mutanga *et al*., 2004). The vegetation indices for this study were calculated using the normalized difference formula that is commonly used to estimate nitrogen and chlorophyll for agricultural and ecological applications (Zengeya *et al*., 2012). The normalized difference index (NDVI) and simple ratio index (SRI) formulas were adopted to calculate the indices using all possible two-band combinations of the Sentinel-2 MSI consisting of 10 spectral bands (Mutanga *et al*., 2015; Sibanda *et al*., 2015). These bands allow for a computation of 100 NDVIs and 100 SRIs.

4.4.Image acquisition and pre-processing

The Sentinel-2 multispectral data was acquired on the 17th of May 2016 from the ESA website (earthexplorer.usgs.gov). The sensor is a multi-spectral push broom scanner. Sentinel-2 MSI has a high spatial resolution ranging from 10m to 60m and acquires all its images at 13(thirteen) spectral bands ranging from visible, passing through the red edge and the near infrared regions to the shortwave lengths (Table 4)

Table 4: Sentinel-2 MSI bands spectral and spatial resolution.

Band Name	**Band Central wavelength (nm)**	**Band description**	**Band Width(nm)**	**Spatial resolution**
Band 1	443nm	Aerosols	20	60
Band 2	490nm	Blue	65	10
Band 3	560nm	Green	35	10
Band 4	665nm	Red	30	10
Band 5	705nm	RedEdge 1	15	20
Band 6	740nm	RedEdge 2	15	20
Band 7	783nm	RedEdge 3	20	20
Band 8	842nm	NIR	115	10
Band 8A	865nm	RedEdge 4	20	20
Band 9	945nm	Water vapor	20	60
Band 10	1375nm	Cirrus	30	60
Band 11	1610nm	SWIR1	90	20
Band 12	2190nm	SWIR2	180	20

Sentinel-2 MSI has a wide swath of 290 kilometres which will enable estimation of grass nitrogen at a larger scale or even at a country level. Sentinel-2 configuration provides enhanced continuity to SPOT and Landsat, and consists of additional unique bands in the red edge position which will enable accurate estimates of leaf N during the wet season (Drusch *et al*., 2012). The image has a repeat cycle of five days and is available for use at no cost. The image will contribute more to monitoring climate and biodiversity variables which include leaf area index at local and global scales (Zengeya *et al*., 2012).

The Sentinel-2 image was geo-referenced and geo-coded by selecting the correct UTM zone, which was 35^0S followed by resampling the image to 10m spatial resolution. This study is one of the many of the applications that require the image to be corrected atmospherically before running the analysis to ensure the extraction of quality reflectance values on Sentinel-2 data. Therefore, a Level 2A Prototype processer (Sen2Cor) was used to atmospherically correct the image as described in the ESA Snap software package. Sen2cor is an atmospheric correction developed by TPZ V on behalf of ESA that works in conjunction with Anaconda (Python 2.7) and the correction process is performed using a python script (ESA 2015). The corrected image was loaded into ESA Snap software and converted to ENVI format, which resulted in 10 separate bands. The converted bands were then displayed on ENVI classic, stacked chronologically into one image again converted to ArcView raster format.

4.5.Data analysis

Prior to statistical analysis, the reflectance values for each grass plots (n=80) were then extracted from the Sentinel-2 MSI spectral bands using the GPS points (grass sample plot); this was achieved using the Extract to Many from ArcGIS spatial analyst extension and overlay analysis in ArcGIS 10.3.

4.5.1. Vegetation indices

Vegetation indices are a method that is commonly used to describe the quality and health condition of vegetation at a pixel level and it is the mostly widely used index for estimating vegetation parameters. The approach of using vegetation indices has been selected for this study because they have been widely used in the field of environment to estimate grass bio-chemicals (Ramoelo *et al.*, 2015; Mutanga *et al.*, 2004). The studies by (Ramoelo *et al.*, 2012; Mutanga & Skidmore, 2007) respectively indicated that NDVIs and SRs indices computed from the red edge bands from hyperspectral and multispectral data, provide accurate estimates of leaf N as compared to the conventional NDVI derived from the visible and NIR bands (680nm and 800nm). A study by (Mutanga & Skidmore, 2004; Mutanga & Kumar, 2007) proved that grass nutrient (N) can be mapped using hyperspectral data during wet seasons because the successful estimation of leaf N using the red edge vegetation indices, depends mainly on chlorophyll concentration (Mutanga & Skidmore, 2004; Cho & Skidmore, 2006).

The vegetation indices for this study were calculated using the normalized difference formula that is commonly used to estimate nitrogen and chlorophyll for agricultural and ecological applications (Zengeya *et al*., 2012). The normalized difference index (NDVI) and ratio index (SRI) were adopted to calculate the indices using all possible two-band combinations of Sentinel-2 MSI consisting of 10 spectral bands after atmospheric correction. These bands allow for a computation of 100 NDVIs and 100 SRIs (Sibanda *et al.*, 2015).

4.5.2. Univariate and multivariate analysis

A univariate analysis (Stepwise Multiple Linear Regression) was used in this study to estimate the linear relationship that exists between the dependent variable (Nitrogen) and the explanatory variables (Sentinel-2 MSI bands and vegetation indices). This was done at 95% confidence level ($p<0.05$). The univariate analysis involved bootstrapping the linear models to check for the linear relationship between Sentinel-2 MSI reflectance, vegetation indices. The vegetation indices for this study were calculated using the normalized difference formula that is commonly used to estimate nitrogen and chlorophyll for agricultural and ecological applications (Zengeya *et al*., 2012). The normalised difference index (NDVI) and ratio index (SRI) were adopted to calculate the indices using all possible two-band combinations of Sentinel-2 MSI consisting of 10 spectral bands after atmospheric correction (SR & NDVI) and foliar N (Ramoelo *et al*.,2012) and the process produces a regression fit line in a form of a scatter plot. The best vegetation indices were selected based on the highest R-squared value (R^2) and the lowest root mean square error (RMSE).

4.5.3. Multivariate analysis

Multivariate statistical techniques are applied when there are several measurements that are made on one variable in one or more samples. In this study, a two-way approach was adopted, Partial Least Square (PLSR) and Random Forest (RF) regression models. All these models were tested using original reflectance data and vegetation indices (SR & NDVI) and their predictive performances were then determined and compared using calibration data set (n = 56) and validated using the test data (n = 24) data sets. The performance of the models were compared using the (R^2) and the standard error from the calibration dataset.

All the analysis were carried out using R Studio (R studio,inc) and Minitab (Minitab.Inc). Random Forest was used to determine whether it can be used to improve grass N estimation as compared to the stepwise multiple linear regression (SMLR) and partial least squares regression (PLSR).

4.5.3.1.Partial least squares regression (PLSR)

Partial least squares regression has been widely used in remote sensing studies to estimate bio-chemical contents of plants (Chun & Keles, 2010). The PLSR method is a full-spectrum method that is an extension of MLR and principal component analysis that convert correlated variables into linearly uncorrelated variables. The model is a multiple linear regression that builds a linear regression using a regression equation:

$$Y=Xb+E \qquad (3)$$

Where Y denotes to the response variable (N) and $\boldsymbol{X}$ denoting to the explanatory variables (in this study that is the spectral bands), $\boldsymbol{b}$ refers to the regression coefficients and $\boldsymbol{E}$ refers to the regression residuals.

Partial Least Squares has a capability of selecting important variable while eliminating redundancy in the data (Cho *et al*., 2007).The model is regressed using various independent variables against one dependent variable while transforming the variables into components by selecting important variables that are later used in the estimation models (Li *et al*., 2011).This is important because spatial data is mostly correlated which causes multi-collinearity which causes model overfittings (Sibanda *et al*., 2015).

The Partial least square (PLS) regression models were used to assess the relationship that exist between foliar N and that of Sentinel-2 reflectance (Li et al. 2016). Preceding the PLS regression, all the datasets used were cross-validated using a leave-one-out method, this was done to select the most optimal variables that can be used to estimate foliar N in the PLS model (Cho *et al.*, 2007). The best model was selected based on the lowest Akaike Information Criterion (AIC) (Sakamoto *et al.*, 1986). The PLSR process was implemented in R- Studio using R statistical programming language.

4.5.3.2.Random Forest regression model (RF)

RF is a machine-learning algorithm that improves the regression and classification trees by combining a large set of decision trees and the method established by Brieman (2001). Furthermore, it is also a multivariate and nonparametric regression algorithm which is incorporated within the ‘’Random Forest’’ package of the R environment software *(*Odindi *et al.*, 2014). This technique was selected and used on this research because of its ability to select the most optimal variables that can later be used in estimating foliar grass N distribution. The RF model was implemented on R-Studio using R statistical programming language. During the RF execution, a set of variables is selected randomly from the training dataset (70%). Three parameters were optimized in RF model, that is the (***ntree)*** which was 500 in this study, the number of regression trees to be drawn based on the values of observations called the bootstrap sample and the (***mtry)*** which is the number of predictors to be tested at each node which should not be more than three in total, a default value of 1 ***mtry*** was used for this study. The retrieved leaf nitrogen was used to develop prediction models in combination with remote sensing data using random forest (RF) regression models. The RF was used to estimate the amount of variation that exists in leaf nitrogen concentration that was explained by the calculated coefficient of determination (r). This technique was chosen for this study as it gives a good indicator of each band by indicating their importance in estimating grass N.

The best RF model with highest R^2 was selected and used for mapping of grass N. This was done using EnMap-box software, which is an IDL, based tool specifically designed for regression and classification of hyperspectral data. The German Hyperspectral Environmental Mapping developed the tool and analysing Programme specifically for analyzing hyperspectral data and it can be fully integrated into ENVI/IDL environment or used as a separate tool. The RF regression based on Sentinel-2 MSI provides a better alternative to multivariate regression involving multispectral indices for grass N estimation at Telperion Game Reserve.

4.6.The calibration and validation dataset

The validations of the models were done using the bootstrap technique, which was implemented in R statistical programming language. Bootstrapping technique is a non-parametric method that validates models by randomly selecting representative samples repetitively while estimating standard error from the population (Fox, 1997).

This was done to test the predictive capability of the models developed using field reflectance training data. The spectral dataset (n = 80) was randomly split into 70% (n = 56) and 30% (n = 24) for calibration and validation datasets. RF and PSLR models from the training data set were used to predict grass nitrogen using the R^2 values and the root mean square errors values were computed. Both the models were validated using the test dataset. The RMSE of the test data were calculated and recorded. The R^2 and RMSEs for the test data set and that of the image test data set were compared statistically using the RMSE value from the test data set.

4.7.Descriptive statistics

The statistical analysis started with the test for normality which was done using Kolmogorov Simonov test (Table 5). The null hypothesis tested was that data does not significantly deviate from the normal distribution versus data that normally distributed at 95% confidence interval. This was followed by the one-way variance (ANOVA) based on the 95% confidence level (p-value<0.05) and the null hypothesis tested was that there is no significant difference in the leaf nitrogen concentration among different grass communities.

Table 5: Descriptive statistics of measures nitrogen concentration

	N	Mean	Minimum	Maximum	Standard deviation
Nitrogen (%)	80	0.36000	2.0200	0.7836.3	0.29693

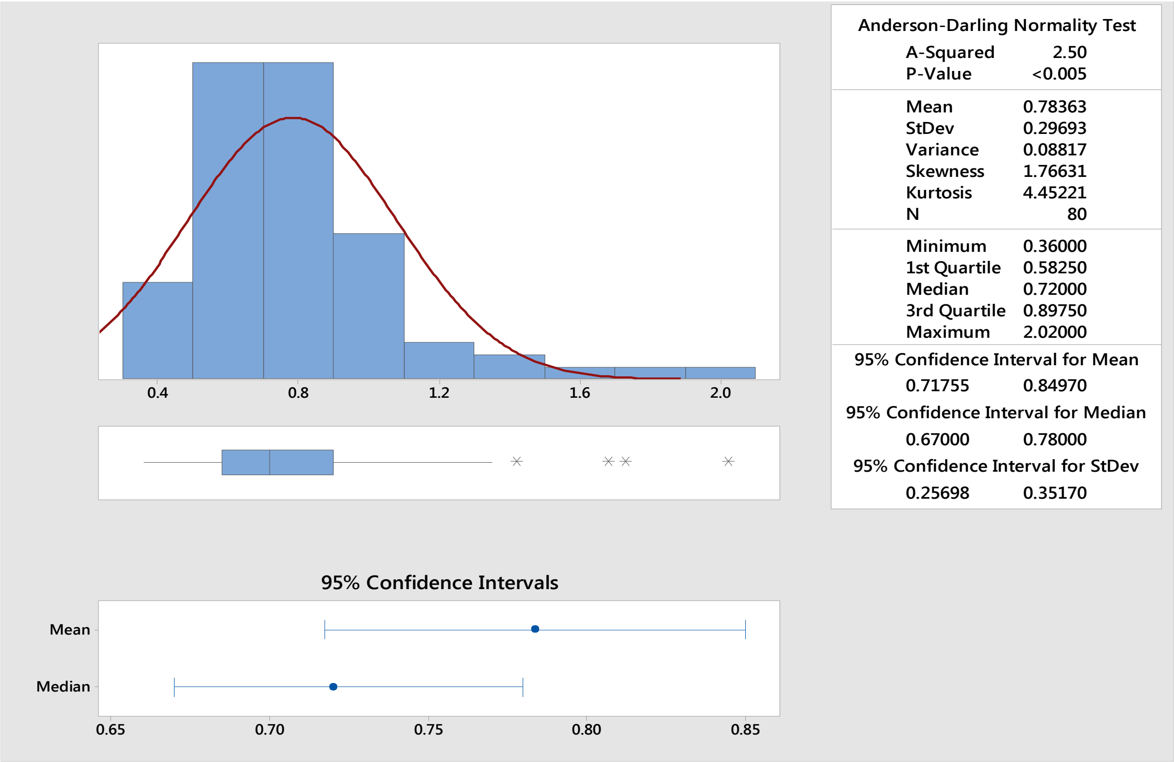
5. Results

5.1.Descriptive and exploratory statistics

The descriptive statistics of the data used in (Table 2) showed that the mean grass N for the 80 samples is 0.78%. The minimum and maximum values of the leaf N were 0.36% and 2.02% respectively. The variability of nitrogen concentration was not significantly high as demonstrated by the variance of 0.089 % but they do indicate an interesting variability that exists in grass N concentration across different grass communities at Telperion Game Reserve.

The normality test was used to test the null hypothesis that data is normally distributed. This was done using the Minitab software and the test indicated that the grass N data did not significantly deviate from the normal distribution (Figure 5). The one-way ANOVA results from XLSTAT (Excel Add-ins) yielded an (*F*= 4.11, P= 0.000) indicating that the spectral reflectance of different grasses are statistically significant which means different grasses reflect differently across the study area.

Figure 5: The distribution of Grass N data.

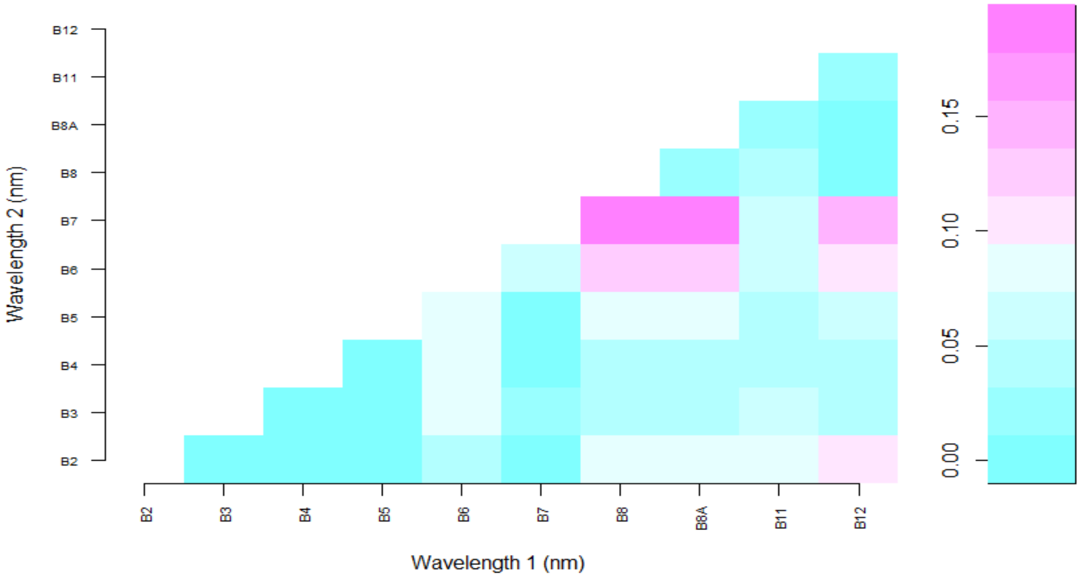


5.2.Correlation between foliar nitrogen and vegetation indices.

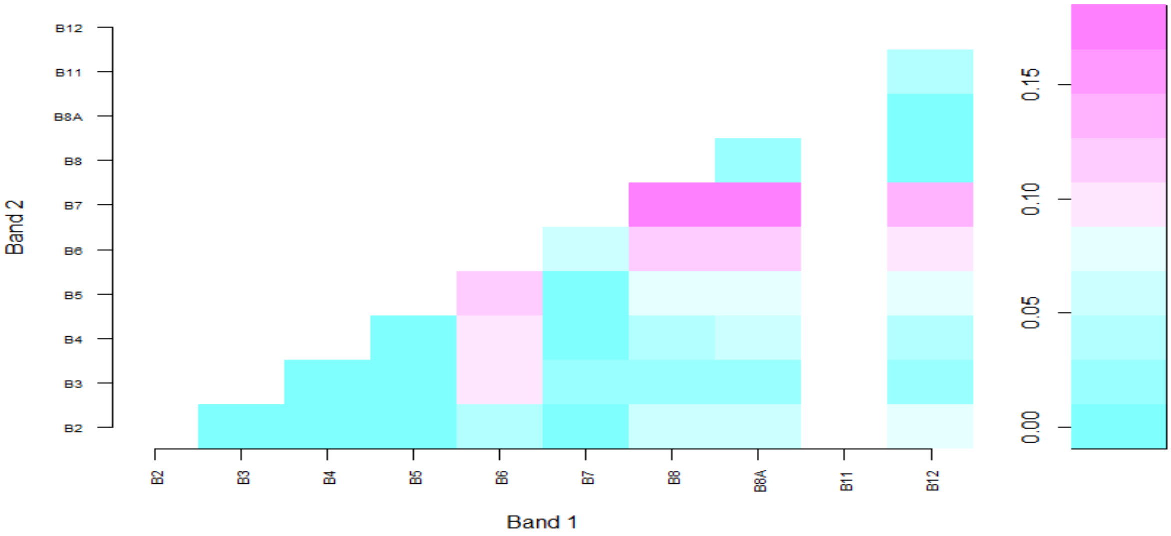
Figure 6 depicts a correlogram demonstrating the correlation between the Sentinel-2 MSI bands and the leaf N. The Sentinel-2 MSI bands allows for the computation of 100 NDVI and 100 SR indices respectively. The cut-off values selected as an indicator of strong correlation between grass N and Sentinel-2 MSI bands was 0.10 and 0.15.

Figure 6: The correlation coefficients between foliar N concertation and vegetation indices NDVIs (a), SRs (b) calculated from all possible two-band combinations of Sentinel-2 image.

(a) NDVI indices



(b) SR indices



5.3.Univariate analysis (Stepwise Multiple Linear Regression)

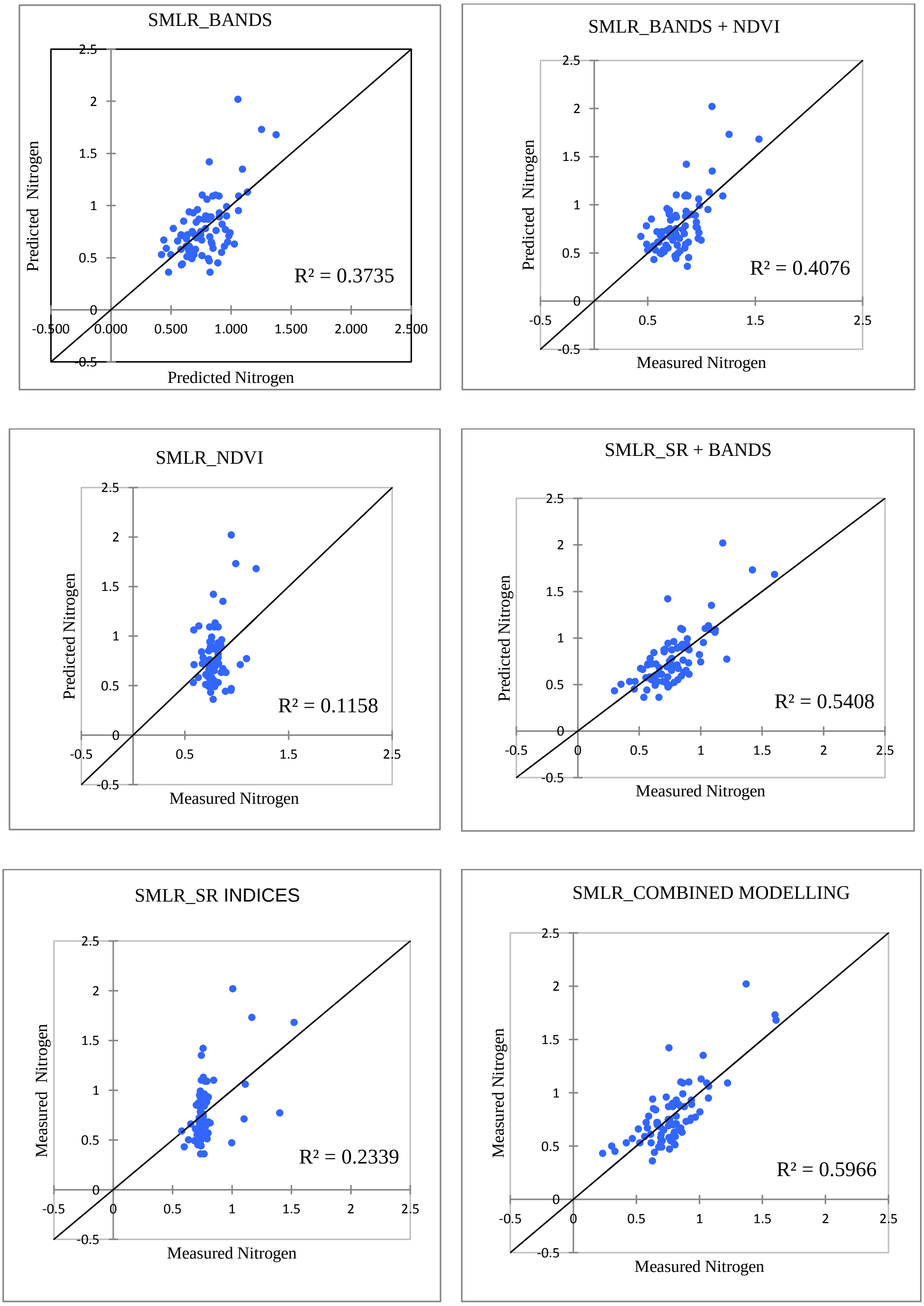
The results of the stepwise linear regression are presented in Table 5 for all the 6 models used. The univariate statistical method showed that the highest estimation accuracy was based on the combined modelling of optimal bands and indices (bootstrapped: R^2 =59.7% & RMSE =46.7%). For example, using Sentinel-2 MSI bands only, it can be observed that the leaf N model yielded a (bootstrapped: R^2 = 37.4% & RMSE= 28%). The NDVI indices yielded the lowest (bootstrapped: R^2 = 14.4%, RMSE=7%) and the model was insignificant. It is known that NDVIs are very unstable and their performance varies depending on the various environmental factors such as soil colour and atmospheric condition. The SR indices produced (bootstrapped: R^2=23.4%, RMSE =16%). The NDVI seem to perform better when combined with Sentinel-2 MSI bands as they produced a (bootstrapped: R^2 = 44.4%, RMSE=30%) while the SR combined with the Sentinel-2 MSI bands produced (bootstrapped: R^2=54.1%, RMSE=41%). Although the combined modelling optimally estimated leaf N, it can be observed that the prediction errors of the model are more comparable to those produced by SR combined with Sentinel-2 MSI bands only.

Table 6: Univariate analysis (Linear regression) using Sentinel-2 MSI bands and vegetation indices. Sig=Significant and insig= insignificant.

MSI bands	**R^2 (%)**	**Adjusted R^2(%)**	**Std error**	**Akaike information criterion**	**No of variables**	**F statistics**	**P- value**
NDVI	14.4	7.3	0.28	31.45	6	7.4%	7.4%(insig)
SR INDICES	23.4	15.9	0.27	26.41	7	3.14	0.6% (sig)
BANDS	37.4	28.3	0.25	18.35	10	4.11	0.0%(sig)
NDVI+BANDS	44.4	30.1	0.25	26.77	16	3.10	0.0(sig)
SR+BANDS	54.1	41.5	0.23	15.73	17	4.30	0.0%(sig)
COMBINED	59.7	46.7	0.21	37.67	22	2.81	0.1% (sig)

The scatter plot in (Figure 6) indicates the linear relationship that exists between nitrogen and sentinel-2 MSI bands and vegetation indices. The performances of the univariate models indicate that the red edge bands (6 and 7) and the indices (SR and NDVI) including the red edge bands performed better as compared to the others. The NDVI model yielded the lowest R^2 (0.11) as compared to an R^2 value of 0.59 when the entire bands and indices were used.

Figure 6: Predicted versus observed nitrogen using simple multiple linear regression (SMLR).



5.4.Multivariate analysis

5.4.1. Partial Least square regression (PLSR)

The multivariate regression technique of PLSR was used to predict leaf N models and to test the performance of all 10 Sentinel-2 MSI spectral bands in combination with various vegetation indices (NDVI & SR). Sentinel-2 MSI bands including the red edge position (REP) bands were used to estimate leaf N in combination with vegetation indices (NDVI & SR) using six different models as shown in (Table 6). The top PLSR model was chosen based on the R^2 value and a lowest root mean square error (RMSE).

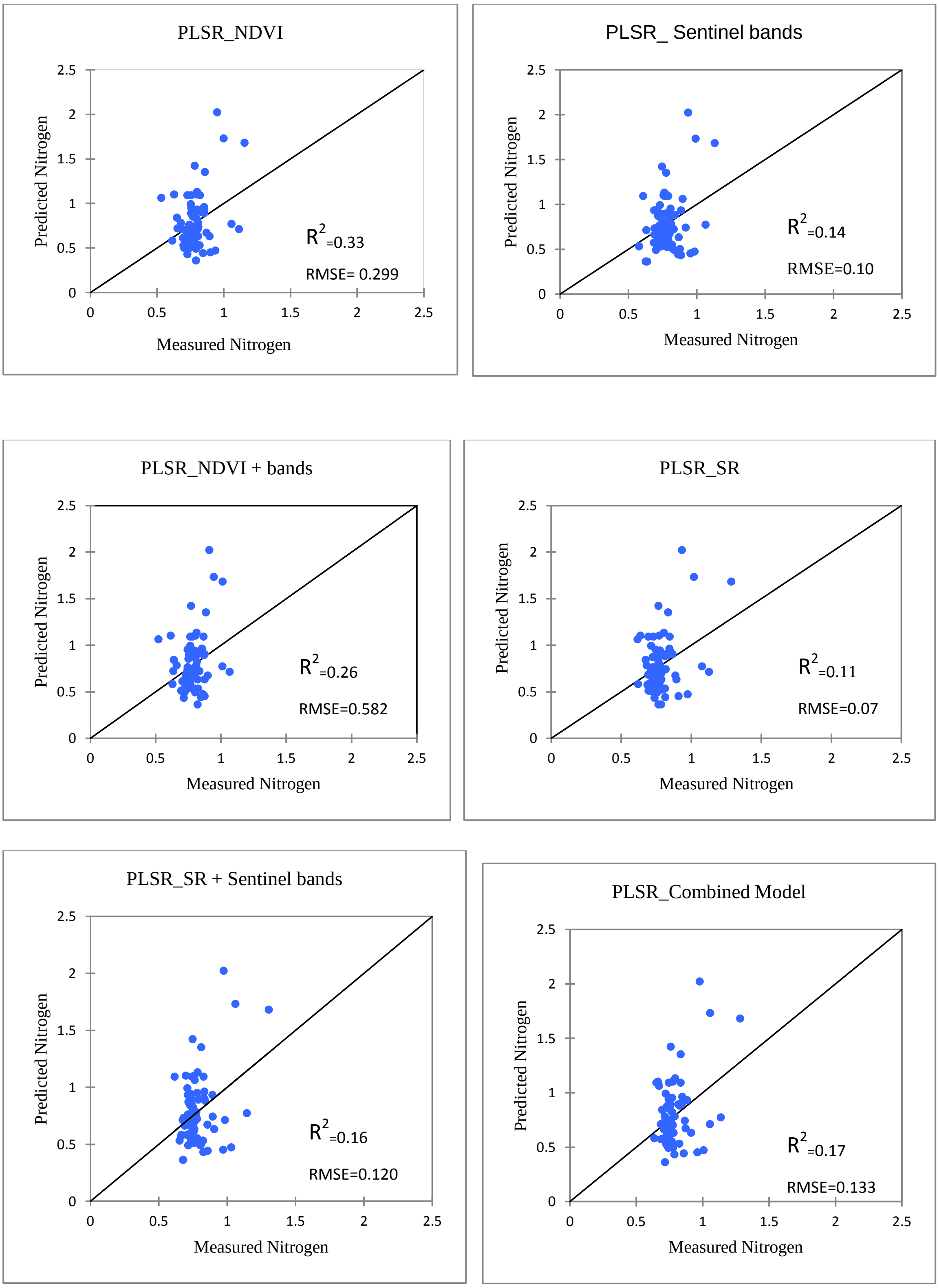
Table 4: Performance of the Partial least square models used validated by bootstrapping.

Dataset	**R^2**	**RMSE**	**Mean of Squared Error**	**No of Variables used**	**% of RMSE to MEAN**	**AIC**
BANDS	0.3301303	0.2996817	0.02309568	10	261.48	3.763871
NDVI	0.1436171	0.1046906	0.2599522	6	784.58	12.5972
NDVI+BANDS	0.261945	0.05823818	0.2413259	16	1345.56	9.440691
SR	0.1145842	0.0743379	0.8281574	8	1054.14	0.287772
SR+BANDS	0.1716237	0.1339703	0.2783535	12	854.92	12.7236
SR+NDVI+BANDS	0.1601599	0.1201675	0.2248618	25	652.11	2.626781

5.4.1.1.Predicted nitrogen versus measured of the PLSR models.

The predicted versus observed leaf N are depicted in (Figure 7). The plots indicate a one on one relationship that exists between nitrogen, bands and vegetation indices (NDVI & SRI). Out of the six PLSR models used, the model based on SR (n=8) indices yielded the lowest R^2 (0.11) as compared to the R^2 value of 0.33 when using the Sentinel bands (n=10).

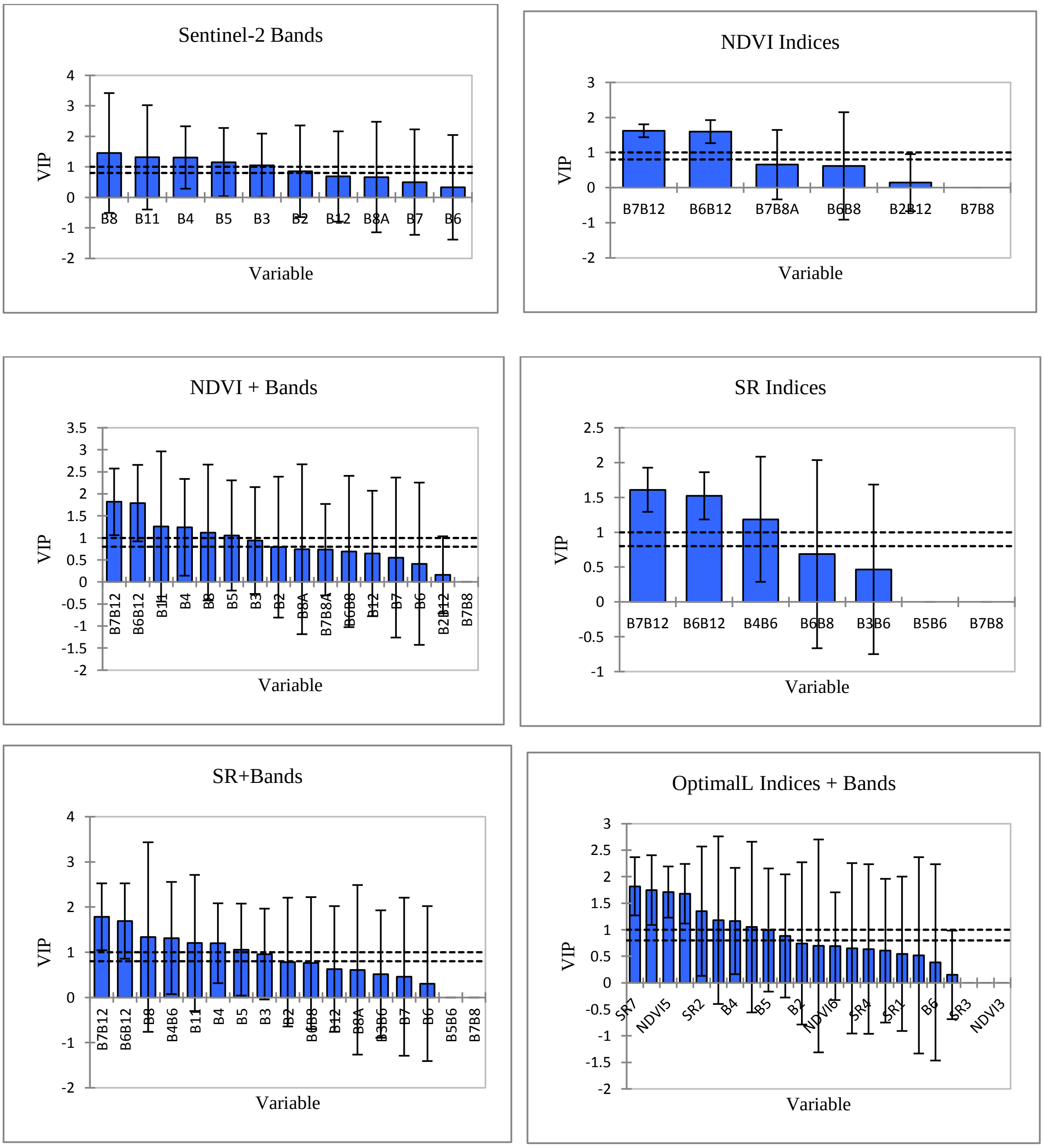
Figure 7: Predicted versus measured nitrogen using Sentinel-2 bands and indices



5.4.1.2.Variable importance for the PLSR models.

The graphs in (Figure 8) below indicate the very important variables that were selected using the PLSR model in XLSTAT. When looking at the performances of all the red edge and SWIR bands (8 and 11) and the indices (SR and NDVI) including the red edge bands are performing better compared to the others.

Figure 7: The performance of the bands and vegetation indices in estimating leaf N



5.4.2. Random forest

The random forest regression was used to estimate leaf N and to test the performance of all ten bands in combination with best performing vegetation indices (NDVI & SR). Again, five models were tested and the results of the multivariate regression analysis using random forest are shown in (Table 7). The table shows their R^2, root mean square error (RMSE), the F-value and the probability-value (p-value) of each model. The graphs in (Figure 9) indicates the predicted versus observed values of nitrogen. Based on all the models used; the predicted and observed nitrogen does not deviate from each other, which is good.

The first model was a regression of leaf N against the 10 Sentinel-2 MSI bands, out of the 10 independent variables used, only two bands were selected as optimal variable (B8A and B6). The bootstrapped model produced an R^2 of 0.29% and a RMSE of 0.26%. Leaf N was regressed against six NDVIs, only two indices were selected as optimal variables, and all these indices included the bands from the red edge position (B7B8A and B6B8). Here the bootstrapped model produced an R^2 of 0.26% with a RMSE of 0.22%. On this model, leaf N was regressed against seven SR indices and only three indices were selected as the optimal variables (i.e. B3B6, B4B6 and B7B12). This model yielded an R^2 of 0.29% with an RMSE of 0.26%. The combination of bands plus NDVIs regressed against leaf N. 6 variables were selected as being the best variables (i.e. B8A, B8, B7, B6) and 1 NDVI (BB78A). The model yielded an R^2 of 0.26% and a RMSE of 0.22%. The combination of SR indices and bands regressed against leaf N, seven variables were selected as the optimal variables after model optimization and this are (B3, B2, B4, B5, B7, B12, B4B6 and B6B8). Here the model produced an R^2 of 0.35% and an RMSE of 0.32%. The last model was based on the combined modelling of optimal bands and indices selected from the five models used. On this model, 3 indices plus two bands were selected as the optimal variables. The model produced an R^2 of 0.23% and an RMSE of 0.20%.

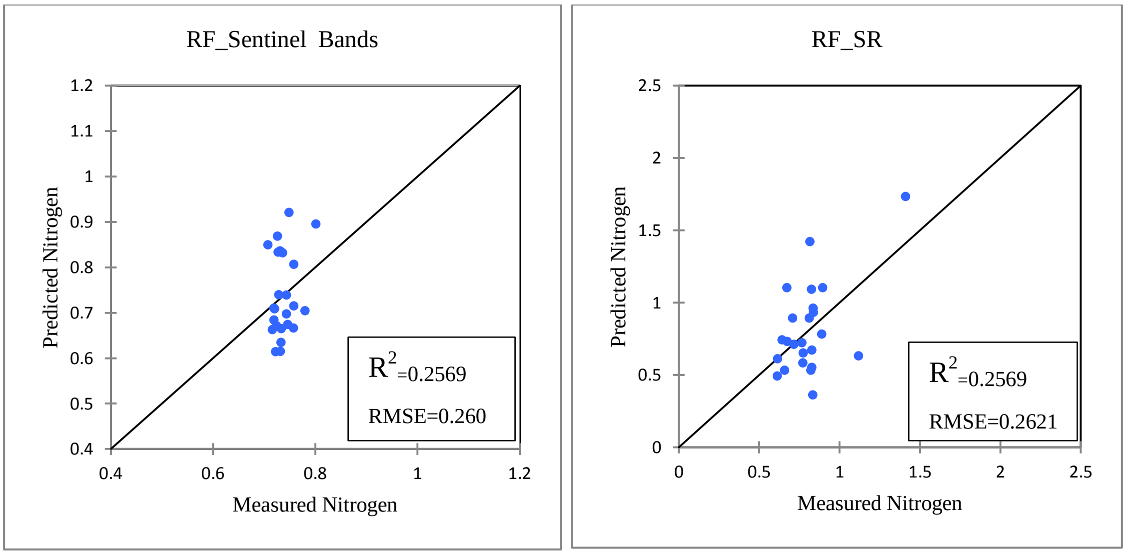
Table 5: The performance of random forest models used validated by bootstrapping

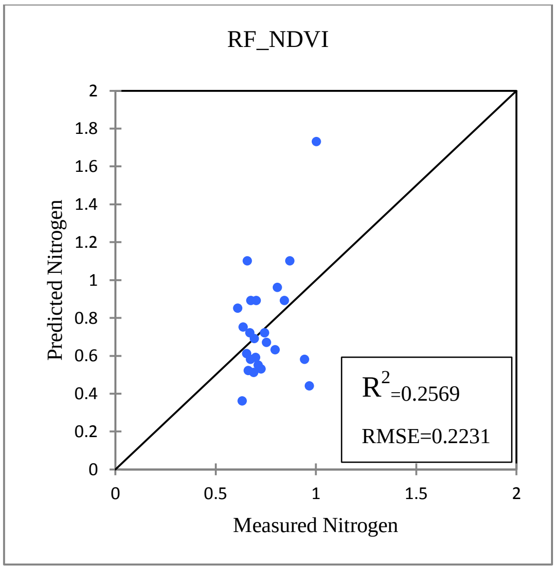
Dataset	**R2**	**RMSE**	**Mean of squared residuals**	**Number of Variables used**	**% of RMSE to MEAN**	**F-Value**	**P-Value**
Bands + N	0.2928	0.2607	0.1172655	10		9.11	0.006322
NDVI + N	0.2569	0.2231	0.09438078	30		7.606	0.01148
SR + N	0.2942	0.2621	0.08624484	7		0.006178	0.006178
NDVI+BANDS+ N	0.2535	0.2196	0.08910324	40		7.472	0.01213
SR+BANDS+N	0.3529	0.3235	0.09802823	12		12	0.002208
NDVI+SR+BANDS+N (combined)	0.2351	0.2003	0.09619631	15		6.761	0.01634

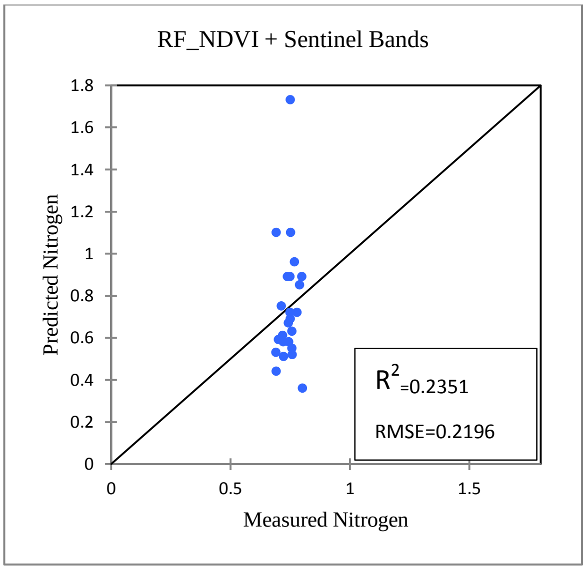
5.4.2.1.Predicted nitrogen versus observed using random forest models.

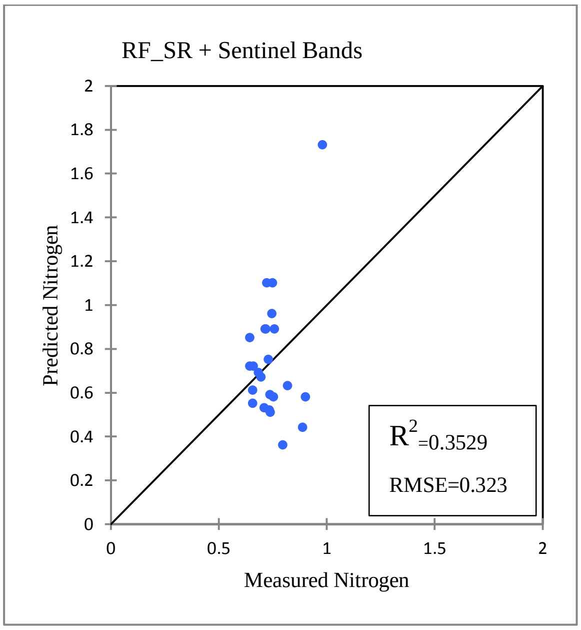
Figure 9 shows the predicted versus observed nitrogen for the calibration dataset derived from Sentinel-2 image. The R^2 and RMSE of each model are shown and the Random Forest results shows that the concentration of nitrogen at Telperion Game Reserve was best predicted based on SR indices and bands model, which yielded the highest R^2 of 0.35%. The detailed results of the models are shown in Table 7.

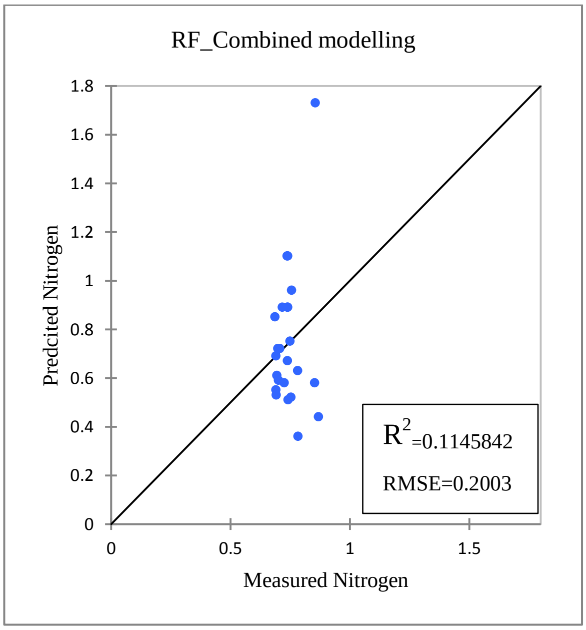
Figure 9: Scatter plots from Random forest analysis using Sentinel-2 MSI bands and vegetation indices.







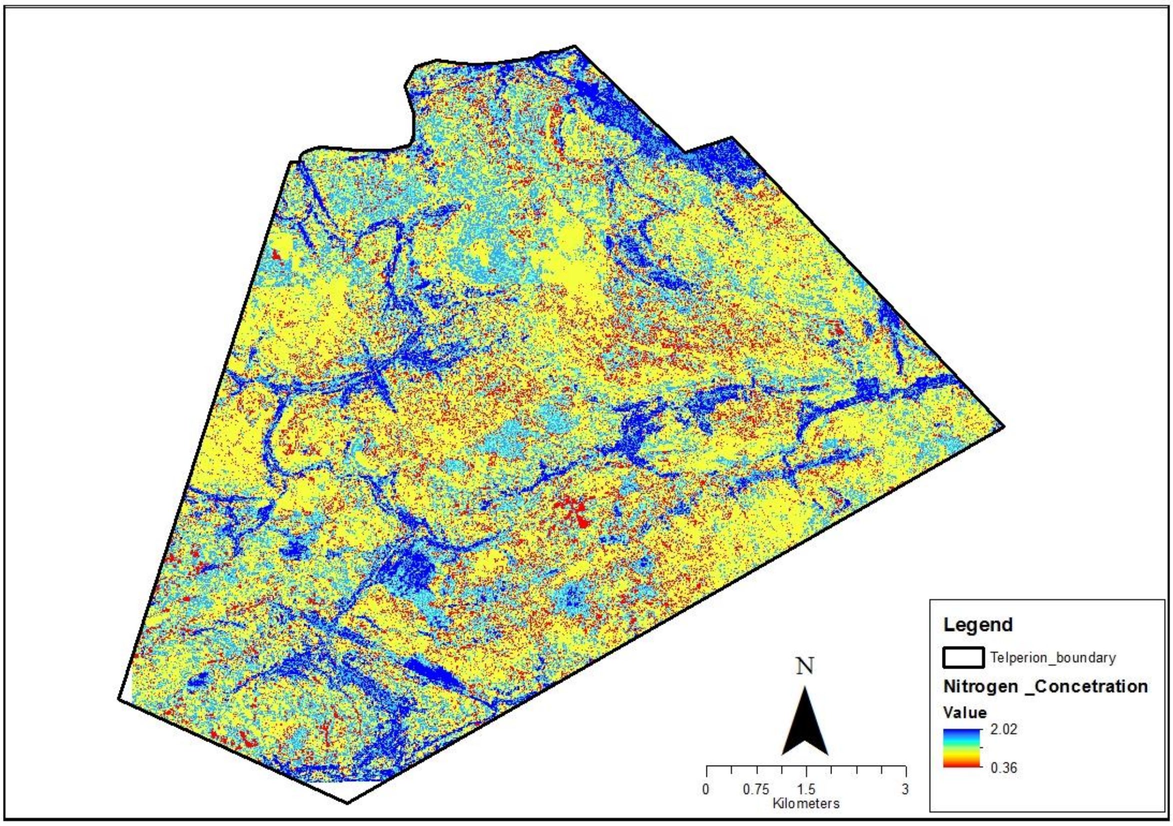




3.2. The distribution of foliar nitrogen

RF model based on bands and SR indices was selected as the best RF model as it produced the highest R^2 values as compared to the other models. This model contained (n=7) SR and (n=5) bands (6, 7, 8, 8A and 12) and all these variables were used to map foliar nitrogen concentration across different grass communities at Telperion Game Reserve. The nitrogen mapping was done using EnMAP-Box software version 2.2.1. which is an IDL Virtual Machine Application and shows the spatial distribution of grass nitrogen content at Telperion Game Reserve as depicted in (Figure 10).

Figure 8: The spatial distribution of grass N in May 2016.



4. Discussion

The aim of the research was to examine the ability of Sentinel-2 MSI sensor in predicting leaf N across different grass communities at Telperion Game Reserve, Mpumalanga, South Africa. The results of the study highlight the importance of the new generation multispectral data in remote sensing of foliar N because of their spatial resolution accompanied by the unique spectral bands (REP) that allows for species discrimination in a rangeland ecosystem. Furthermore, the results also highlight the significance of the red edge position (REP) in estimating foliar N because the red edge bands (6 & 8A) and vegetation indices because they exhibited a strong correlation to foliar N. The results of this study are in agreement with those attained by Sibanda *et al* (2015) and Cho & Skidmore (2006) who discovered that leaf N correlates more with the bands located within the REP compared to the other spectral bands which proves consistency of the use of multispectral remote sensing in leaf N estimation reported.

Band 12, which falls within the short infrared regions, also exhibited a strong correlation to leaf nitrogen. It can also be observed that the bands from the visible spectrum exhibited a very poor relationship to grass N meaning that nitrogen can be optimally estimated using four bands (which are band 6, 8,8A and 12) because leaf N is more correlated and sensitive to these bands as compared to the other bands. The findings of this study support the results achieved by Mutanga & Skidmore (2007) who discovered a strong correlation between nitrogen and REP when investigating the relationship between the red edge position (REP) and the bio-chemical contents. The sensitiveness of the REP to foliar N has been well documented in previous studies (Cho & Skidmore, 2006). However, the results of this study is in contrast with many previous studies because it discovered that a high correlation exists between foliar N and the new incorporated Sentinel-2 band 12 from the short wave region. This outcome agrees with a study by (Knox *et al.*, 2010) who discovered that the SWIR contributes more when mapping foliar bio-chemicals using dry grass samples. This is attributed to the fact that the grass samples were collected in May which is considered the early dry grass season which means that the grass was water stressed.

4.1.Vegetation indices

The results of the study highlight the importance of the new generation multispectral data in remote sensing of foliar N because of their spatial resolution accompanied by the unique spectral bands (REP) that allows for species discrimination at landscape level. Furthermore, the results also highlight the significance of the red edge position (REP) in estimating foliar N because the vegetation indices including the REP bands yielded the highest correlation coefficients (Figure 6). These results concur with a study by Cho and Skidmore (2006) who discovered that, the leaf N correlates more with the bands located within the REP as compared to the other spectral bands, which reiterate the importance of the red edge bands in estimating foliar nitrogen and proves consistency in the use of multispectral remote sensing in leaf N estimation.

4.2.The univariate and multivariate analysis

The applicability of all ten MSI bands and the effect of the red edge bands were tested using six different SMLR models. The estimation of leaf N based on SMLR combined modelling yielded the highest leaf N estimation as compared to the other models. The PLSR variable importance in (Figure 8) shows that, band 8 and vegetation indices including the red edge bands (6, 7) and band 12 from the shortwave region (SWIR) were selected as the best performers in estimating leaf N for this study. These results are similar to what was achieved by Sibanda *et al* (2015) when quantifying the above ground biomass using Sentinel-2 image. This reiterates the important role that the REP and SWIR play in the estimation of foliar N during wet and dry the season.

A similar approach was used for the partial least square regression (PLSR). The best vegetation index selected was NDVIB7-B8A and band 8 and 8A. The highest results were obtained by using the Sentinel-2 MSI bands followed by the NDVI model. For the Random Forest (RF) regression, the top five vegetation indices selected are the ones including the red edge bands. This means that the univariate (SLMR) and multivariate (RF & PLSR) models both selected vegetation indices that includes the red edge bands as the best variables for predicting leaf N concentration. The capability of the REP in improving the performance of the models in leaf N estimation has once again been demonstrated in this study (Sibanda *et al*., 2015).

The results of the multivariate analysis (PLSR and RF) were lower than that of SMLR but the results of the stepwise linear regression cannot be trusted because the model fits every variable which produces noise. Therefore, the Random Forest model was selected as the best model that can be used to predict nitrogen because of their stability and the capability to deal with over fitting and multicollinearity that mostly exists in remote sensing data (Lawrence *et al*., 2006). The same R^2 value was achieved for both training and testing which confirms the stability of the RF regression model.

Furthermore, it has been discovered in previous studies that PLSR mostly performs better than the RF model (Cho *et al*., 2007; Si *et al*., 2012), but it should be noted that, the PLSR model performs well when the data is linearly related which was never the case in this study. The observation that the inclusion of the red edge bands in regression models improves the estimation of leaf N has been noted in previous studies (Ramoelo *et al*., 2012 ; Cho *et al*., 2015). The result of the PLSR models showed minor improvement because the highest PSLR model yielded a correlation coefficient of 0.33% based on bands only, which was lower compared to the stepwise linear regression. This is attributed to the fact that the PSLR model pulls out only important variables from the sample and uses those to run the regression which decreases model errors whereas simple linear regression uses all the variables which yield better correlation coefficient consisting of noise and high model error.

This means that the effect of red edge using the PLSR model was not as significant as compared to univariate regression, meaning that the red edge bands did not improve leaf N estimation as it was expected. It should be noted that the PSLR model is still the better model as compared to SMLR regardless of the low coefficient of determination based on the reason explained above.

In the case of RF regression, optimal variables for each model were selected and recorded as the best variables to predict leaf N. The bands and indices selected as optimal variable falls within the visible spectrum(490-665nm), the NIR (842), the red edge position (705-865) and the SWIR (2190).

From all the models, B8A and B6 were selected as the top 2 variables. The RF model based on combined optimal bands and indices yielded an overall accuracy of 35%. The studies by (Oumar & Mutanga, 2014; Ramoelo *et al.*, 2011; Ramoelo *et al.*, 2012) integrated environmental variables in their models to increase grass nitrogen prediction accuracy. The difference in this study is that it only used the vegetation indices and spectral bands calculated from Sentinel-2 image to estimate Leaf N while other studies employed the integrated modelling to improve the estimation accuracy. It was discovered in this study that, leaf N is also sensitive to band 12 which is situated in the shortwave region including the well- known red edge bands, which is in agreement with the findings obtained by (Knox *et al.*,2011).

(Kalcska, 2014) obtained an R^2 of 0.3 when estimating foliar chlorophyll and nitrogen content in an ombrotrophic bog using hyperspectral data. A study by (Knox *et al.*, 2011) mapped the forage grass bio-chemical during the dry season and obtained a variation of only 41% when using the spectral features only. According to (Knox *et al.*, 2011), the concentration of limiting nutrients (such as N) are known to decline during the dry seasons due to spectral changes in the red edge region which increase the importance of the shortwave infrared in return. The author further indicated that out of the three bio-chemical estimated, nitrogen was the only nutrient that had the lowest predictive power for both training and validation dataset.

4.3.The distribution Leaf N

The distribution of leaf N as indicated in (Figure 9) which was created using the best models indicate a significant variation in N concentration across different grass communities within the study area. The areas covered by grass exhibits low leaf N concentration and the areas comprising of wood vegetation surrounded by the drainage lines displays high leaf N content. This means that grasses in the periphery of drainage supplies are well nourished as compared to the other areas because they all display the highest leaf N concentration. The results of this study concur with the findings by (Knox *et al.*, 2011; Ramoelo *et al.*, 2014) who monitored plants stress and mapped savanna forage quality using RS for wet and dry seasons. A study by (Ramoelo *et al.*, 2014) revealed that the spatial distribution of plant stress is lower in the areas along the water course and under stress, than in the areas characterized by grass only.

CHAPTER FIVE

Conclusion and recommendations

The study has demonstrated the applicability of Sentinel-2 MSI utility in mapping leaf N at a landscape scale. The effect of the red edge in mapping and estimating leaf N has been highlighted through the methods of univariate regression, multivariate methods such as random forest and partial least squares. The vegetation indices calculated from the red edge should then be used for the mapping and estimation of leaf N, because the red edge bands produce better and accurate results for bio-chemicals, as they are not sensitive to background effects.

The objectives of the study were reached as leaf N was successfully mapped and estimated using vegetation indices through univariate and multivariate regression methods except for the NDVI model, which was insignificant. The concentration of Leaf N was recognized to be dissimilar across different grass communities. The use of simple ratios analysis has indicated that the use of vegetation indices can be used in leaf N estimation because they enhanced the mapping and prediction of foliar N (Mutanga *et al*., 2003). The extraction of vegetation indices offers good possibilities of mapping rangeland quality because they increased the capability of Sentinel-2 MSI in discriminating leaf N specifically in the red edge position (REP).

This study highlights two critical aspects in remote sensing of foliar nitrogen, which is the time and spectral reflectance because the spectral reflectance of the image varies based on time. A study by (Ramoelo *et al*., 2012) recommended a concurrent prediction of leaf N using remote sensing in order to determine the variability that exists in leaf N concentration during various seasons because rainfall variability disturbs the normal functioning of rangeland ecosystems which causes variations of the vegetation cover (Paudel & Anderson, 2010).

Therefore, it is important to know that the mapping of foliar bio-chemical should be timed but not specifically limited to the wet season only, because knowing the geographic concentration of grass bio-chemicals is also crucial especially during the dry season when the leaf nutrients are limited (Knox *et al*., 2011). What is crucial is to know which spectral regions to consider based on the mapping season. Furthermore, it is recommended that, the integration of ancillary data and spectra from the shortwave infrared region (SWIR) is necessary if one needs to improve the accuracy estimation of leaf N during dry season. Additionally, if the grass samples are collected during the wet season, the estimation capability of the models can be enhanced by integrating ancillary data such as soils, species, fire and phenological maps. Sentinel-2 MSI offers possibilities to map and estimate leaf quality at a rangeland level.

5. REFERENCES

A, K.D.Z. et al., 2012. Remote sensing of foliar nitrogen in cultivated grasslands of human dominated landscapes. *ISPRS Journal of Photogrammetry and Remote Sensing*, 5(1), pp.88–97.

Abbasi, B. et al., 2015. Automatic Generation Of Training Data For Hyperspectral Image Classification Using Support Vector Machine. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-7/W3(May), pp.575–580.

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), pp.281–296.

Adam, E.M. et al., 2012. Discriminating the papyrus vegetation (Cyperus papyrus L.) and its co-existent species using random forest and hyperspectral data resampled to HYMAP. *International Journal of Remote Sensing*, 33(2), pp.552–569.

Adelabu, S., Mutanga, O. & Adam, E., 2015. Testing the reliability and stability of the internal accuracy assessment of random forest for classifying tree defoliation levels using different validation methods. *Geocarto International*, 30(7), pp.810–821.

Adjorlolo, C. et al., 2012. Challenges and opportunities in the use of remote sensing for C-3 and C-4 grass species discrimination and mapping. *African Journal of Range & Forage Science*, 29(2), pp.47–61.

Adjorlolo, C. & Botha, J.O., 2015. Integration of remote sensing and conventional models for modeling grazing/browsing capacity in southern African savannas. *Journal of Applied Remote Sensing*, 9(August).

Adjorlolo, C., Mutanga, O. & Cho, M.A., 2015. Predicting C3 and C4 grass nutrient variability using in situ canopy reflectance and partial least squares regression. *International Journal of Remote Sensing*, 36(6), pp.1743–1761..

Anderson, B.J.R. et al., 1976. A Land Use And Land Cover Classification System For Use With Remote Sensor Data. *Development*, 2001(964), p.41.

Baumstark, R., Duffey, R. & Pu, R., 2016. Mapping seagrass and colonized hard bottom in

Springs Coast, Florida using WorldView-2 satellite imagery. *Estuarine, Coastal and Shelf Science*, 181, pp.83–92.

Beeri, O. et al., 2007. Estimating forage quantity and quality using aerial hyperspectral imagery for northern mixed-grass prairie. *Remote Sensing of Environment*, 110(2), pp.216–225.

Belluco, E. et al., 2006. Mapping salt-marsh vegetation by multispectral and hyperspectral remote sensing. *Remote Sensing of Environment*, 105(1), pp.54–67.

Boegh, E. et al., 2013. Remote sensing of LAI, chlorophyll and leaf nitrogen pools of crop- and grasslands in five European landscapes. *Biogeosciences*, 10(10), pp.6279–6307.

Burai, P. et al., 2015. Classification of herbaceous vegetation using airborne hyperspectral imagery. *Remote Sensing*, 7(2), pp.2046–2066.

Cho, M.A. et al., 2007. Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares regression. *International Journal of Applied Earth Observation and Geoinformation*, 9(4), pp.414–424.

Cho, M.A. et al., 2012. Potential utility of the spectral red-edge region of SumbandilaSat imagery for assessing indigenous forest structure and health. *International Journal of Applied Earth Observation and Geoinformation*, 16(1), pp.85–93.

Cho, M.A. & Skidmore, A.K., 2006. A new technique for extracting the red edge position from hyperspectral data: The linear extrapolation method. *Remote Sensing of Environment*, 101(2), pp.181–193.

Clevers, J.G.P.W. et al., 2001. MERIS and the red-edge position. *International Journal of Applied Earth Observation and Geoinformation*, 3(4), pp.313–320.

Corbane, C. et al., 2014. Remote sensing for mappong natural habitats and their conservation status – New opportunities and challenges. *International Journal of Applied Earth Observation and Geoinformation*, 37, pp.7–16.

Darvishzadeh, R. et al., 2008. LAI and chlorophyll estimation for a heterogeneous grassland using hyperspectral measurements. *ISPRS Journal of Photogrammetry and Remote Sensing*, 63(4), pp.409–426.

Delegido, J. et al., 2013. A red-edge spectral index for remote sensing estimation of green LAI over agroecosystems. *European Journal of Agronomy*, 46, pp.42–52.

Van Deventer, H. et al., 2015. Capability of models to predict leaf N and P across four seasons for six sub-tropical forest evergreen trees. *ISPRS Journal of Photogrammetry and Remote Sensing*, 101, pp.209–220.

Drusch, M. et al., 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, 120, pp.25–36.

Feng, W. et al., 2015. Remote estimation of above ground nitrogen uptake during vegetative growth in winter wheat using hyperspectral red-edge ratio data. *Field Crops Research*, 180, pp.197–206.

Frampton, W.J. et al., 2013. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 82, pp.83–92.

Fu, Y. et al., 2014. Winter wheat biomass estimation based on spectral indices, band depth analysis and partial least squares regression using hyperspectral measurements. *Computers and Electronics in Agriculture*, 100, pp.51–59.

Goecker, M.E., Heck, K.L. & Valentine, J.F., 2005. Effects of nitrogen concentrations in turtlegrass Thalassia testudinum on consumption by the bucktooth parrotfish Sparisoma radians. *Marine Ecology Progress Series*, 286, pp.239–248.

Guo, X., Wilmshurst, J.F. & Li, Z., 2010. Comparison of laboratory and field remote sensing methods to measure forage quality. *International Journal of Environmental Research and Public Health*, 7(9), pp.3513–3530.

He, L., Zhang, H.Y., et al., 2016. Estimating canopy leaf nitrogen concentration in winter wheat based on multi-angular hyperspectral remote sensing. *European Journal of Agronomy*, 73, pp.170–185.

He, L., Song, X., et al., 2016. Improved remote sensing of leaf nitrogen concentration in winter wheat using multi-angular hyperspectral data. *Remote Sensing of Environment*, 174, pp.122–133.

Hedley, J. et al., 2012. Capability of the Sentinel 2 mission for tropical coral reef mapping and coral bleaching detection. *Remote Sensing of Environment*, 120, pp.145–155.

Huang, S. et al., 2017. Potential of RapidEye and WorldView-2 Satellite Data for Improving Rice Nitrogen Status Monitoring at Different Growth Stages. *Remote Sensing* , 9(3), p.227.

Karlson, M. & Ostwald, M., 2016. Remote sensing of vegetation in the Sudano-Sahelian zone: A literature review from 1975 to 2014. *Journal of Arid Environments*, 124, pp.257–269.

Knox, N.M. et al., 2011. Dry season mapping of savanna forage quality, using the hyperspectral Carnegie Airborne Observatory sensor. *Remote Sensing of Environment*, 115(6), pp.1478–1488.

Knox, N.M. et al., 2010. Nitrogen prediction in grasses: effect of bandwidth and plant material state on absorption feature selection. *International Journal of Remote Sensing*, 31(915935530), pp.691–704.

Lewis, D., Phinn, S. & Arroyo, L., 2013. Cost-effectiveness of seven approaches to map vegetation communities - A case study from Northern Australia’s tropical savannas. *Remote Sensing*, 5(1), pp.377–414.

Li, F. et al., 2014. Improving estimation of summer maize nitrogen status with red edge-based spectral vegetation indices. *Field Crops Research*, 157(February 2014), pp.111–123.

Li, L. et al., 2016. Methods for estimating leaf nitrogen concentration of winter oilseed rape (Brassica napus L.) using in situ leaf spectroscopy. *Industrial Crops and Products*, 91, pp.194–204.

Li, L.J. et al., 2011. Foliar N/P ratio and nutrient limitation to vegetation growth on Keerqin sandy grassland of North-east China. *Grass and Forage Science*, 66(2), pp.237–242.

Ling, B. et al., 2014a. Estimating canopy nitrogen content in a heterogeneous grassland with varying fire and grazing treatments: Konza Prairie, Kansas, USA. *Remote Sensing*, 6(5), pp.4430–4453.

Ling, B. et al., 2014b. Estimating canopy nitrogen content in a heterogeneous grassland with varying fire and grazing treatments: Konza Prairie, Kansas, USA. *Remote Sensing*, 6(5), pp.4430–4453.

Lück-Vogel, M. et al., 2016. Vegetation mapping in the St Lucia estuary using very high-resolution multispectral imagery and LiDAR. *South African Journal of Botany*, (August).

Luther, J.E. et al., 2006. Biomass mapping using forest type and structure derived from Landsat TM imagery. *International Journal of Applied Earth Observation and Geoinformation*, 8(3), pp.173–187.

Mansour, K. et al., 2016. Multispectral remote sensing for mapping grassland degradation using the key indicators of grass species and edaphic factors. *Geocarto International*, 6049(April), pp.1–15.

Mittapalli, B., Desai, M.D. & Antonio, S., 2014. Multispectral Data. , 7(11), pp.2322–2325.

Mutanga, O. et al., 2004. Explaining grass-nutrient patterns in a savanna rangeland of southern Africa. *Journal of Biogeography*, 31(5), pp.819–829.

Mutanga, O. & Kumar, L., 2007. Estimating and mapping grass phosphorus concentration in an African savanna using hyperspectral image data. *International Journal of Remote Sensing*, 28(February 2015), pp.4897–4911.

Mutanga, O. & Skidmore, A.K., 2004. Hyperspectral band depth analysis for a better estimation of grass biomass (Cenchrus ciliaris) measured under controlled laboratory conditions. *International Journal of Applied Earth Observation and Geoinformation*, 5(2), pp.87–96.

Mutanga, O. & Skidmore, A.K., 2004. Integrating imaging spectroscopy and neural networks to map grass quality in the Kruger National Park, South Africa. *Remote Sensing of Environment*, 90(1), pp.104–115.

Mutanga, O. & Skidmore, A.K., 2007a. Red edge shift and bio-chemical content in grass canopies. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(1), pp.34–42.

Mutanga, O. & Skidmore, A.K., 2007b. Red edge shift and bio-chemical content in grass canopies. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(1), pp.34–42.

Mutanga, O., Skidmore, A.K. & Prins, H.H.T., 2004. Discriminating sodium concentration in a mixed grass species environment of the Kruger National Park using field spectrometry. *International Journal of Remote Sensing*, 25(20), pp.4191–4201.

Mutanga, O., Skidmore, A.K. & Prins, H.H.T., 2004a. Predicting in situ pasture quality in the Kruger National Park, South Africa, using continuum-removed absorption features. *Remote Sensing of Environment*, 89(3), pp.393–408.

Mutanga, O., Skidmore, A.K. & Prins, H.H.T., 2004b. Predicting in situ pasture quality in the Kruger National Park, South Africa, using continuum-removed absorption features. *Remote Sensing of Environment*, 89(3), pp.393–408.

Mutanga, O., Skidmore, A.K. & Van Wieren, S., 2003. Discriminating tropical grass (Cenchrus ciliaris) canopies grown under different nitrogen treatments using spectroradiometry. *ISPRS Journal of Photogrammetry and Remote Sensing*, 57(4), pp.263–272.

Odindi, J. et al., 2014. Comparison between {WorldView}-2 and {SPOT}-5 images in mapping the bracken fern using the random forest algorithm. *Journal of Applied Remote Sensing*, 8(1), p.83527.

Omer, G. et al., 2015. Performance of Support Vector Machines and Artifical Neural Network for Mapping Endangered Tree Species Using WorldView-2 Data in Dukuduku Forest, South Africa. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(10), pp.4825–4840.

Oumar, Z. & Mutanga, O., 2014. Integrating environmental variables and WorldView-2 image data to improve the prediction and mapping of Thaumastocoris peregrinus (bronze bug) damage in plantation forests. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87, pp.39–46.

Ramoelo, A. et al., 2012. Estimating grass nutrients and biomass as an indicator of rangeland (forage) quality and quantity using remote sensing in Savanna ecosystems. *Researchspace*, pp.1–8.

Ramoelo, A. et al., 2011. Integrating environmental and in situ hyperspectral remote sensing variables for grass nitrogen estimation in Savannah ecosystems. *Proceedings of ISRSE34 : 34th international symposium on remote sensing of environment, Sydney, Australia, 10-15 April 2011*, p.4.

Ramoelo, A., Cho, M.A., et al., 2015a. Monitoring grass nutrients and biomass as indicators of rangeland quality and quantity using random forest modelling and WorldView-2 data. *International Journal of Applied Earth Observation and Geoinformation*, 43, pp.43–54.

Ramoelo, A., Cho, M.A., et al., 2015b. Monitoring grass nutrients and biomass as indicators of rangeland quality and quantity using random forest modelling and WorldView-2 data. *International Journal of Applied Earth Observation and Geoinformation*, 43, pp.43–54.

Ramoelo, A. et al., 2013. Non-linear partial least square regression increases the estimation accuracy of grass nitrogen and phosphorus using in situ hyperspectral and environmental data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 82, pp.27–40.

Ramoelo, A., Cho, M., et al., 2015. Potential of Sentinel-2 spectral configuration to assess rangeland quality. *Journal of Applied Remote Sensing*, 9(1), p.94096.

Ramoelo, A., Dzikiti, S., et al., 2015. Potential to monitor plant stress using remote sensing tools. *Journal of Arid Environments*, 113, pp.134–144.

Ramoelo, A., Skidmore, A.K., Cho, M.A., Schlerf, M., Mathieu, R. & Heitkönig, I.M.A., 2012. Regional estimation of savanna grass nitrogen using the red-edge band of the spaceborne rapideye sensor. *International Journal of Applied Earth Observation and Geoinformation*, 19(1), pp.151–162.

Ramoelo, A., Skidmore, A.K., Cho, M.A., Schlerf, M., Mathieu, R. & Heitk??nig, I.M.A., 2012. Regional estimation of savanna grass nitrogen using the red-edge band of the spaceborne rapideye sensor. *International Journal of Applied Earth Observation and Geoinformation*, 19(1), pp.151–162.

Ramoelo, A. et al., 2013a. Savanna grass nitrogen to phosphorous ratio estimation using field spectroscopy and the potential for estimation with imaging spectroscopy. *International Journal of Applied Earth Observation and Geoinformation*, 23(1), pp.334–343.

Ramoelo, A. et al., 2013b. Savanna grass nitrogen to phosphorous ratio estimation using field spectroscopy and the potential for estimation with imaging spectroscopy. *International Journal of Applied Earth Observation and Geoinformation*, 23(1), pp.334–343.

Rapinel, S. et al., 2014. Identification and mapping of natural vegetation on a coastal site using a Worldview-2 satellite image. *Journal of Environmental Management*, 144, pp.236–246.

Schlemmera, M. et al., 2013. Remote estimation of nitrogen and chlorophyll contents in maize at leaf and canopy levels. *International Journal of Applied Earth Observation and Geoinformation*, 25(1), pp.47–54.

Schmidt, K.S. & Skidmore, A.K., 2003. Spectral discrimination of vegetation types in a coastal wetland. *Remote Sensing of Environment*, 85(1), pp.92–108.

Shi, T. et al., 2015. Estimating leaf nitrogen concentration in heterogeneous crop plants from hyperspectral reflectance. *International Journal of Remote Sensing*, 36(18), pp.4652–4667.

Si, Y. et al., 2012. Mapping spatio-temporal variation of grassland quantity and quality using MERIS data and the PROSAIL model. *Remote Sensing of Environment*, 121, pp.415–425.

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, pp.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), pp.1–22.

Skidmore, A.K. et al., 2010. Forage quality of savannas - Simultaneously mapping foliar protein and polyphenols for trees and grass using hyperspectral imagery. *Remote Sensing of Environment*, 114(1), pp.64–72.

Skowronek, S., Asner, G.P. & Feilhauer, H., 2017. Performance of one-class classifiers for invasive species mapping using airborne imaging spectroscopy. *Ecological Informatics*, 37, pp.66–76.

Wang, C. et al., 2007. Mapping mixed vegetation communities in salt marshes using airborne spectral data. *Remote Sensing of Environment*, 107(4), pp.559–570.

Wang, J. et al., 2015. Evaluating different methods for grass nutrient estimation from canopy hyperspectral reflectance. *Remote Sensing*, 7(5), pp.5901–5917.

Wang, X. et al., 2010. Evaluation of soil nitrogen emissions from riparian zones coupling simple process-oriented models with remote sensing data. *Science of the Total Environment*, 408(16), pp.3310–3318.

Wang, Y. et al., 2007. Terrestrial and Submerged Aquatic Vegetation Mapping in Fire Island National Seashore Using High Spatial Resolution Remote Sensing Data. *Marine Geodesy*, 30(1–2), pp.77–95.

Yuan, L. & Zhang, L.Q., 2008. Mapping large-scale distribution of submerged aquatic vegetation coverage using remote sensing. *Ecological Informatics*, 3(3), pp.245–251.

Zengeya, F.M., Mutanga, O. & Murwira, A., 2012. Linking remotely sensed forage quality estimates from worldview-2 multispectral data with cattle distribution in a savanna landscape. *International Journal of Applied Earth Observation and Geoinformation*, 21(1), pp.513–524.

Zhai, Y. et al., 2013. Estimation of nitrogen, phosphorus, and potassium contents in the leaves of different plants using laboratory-based visible and near-infrared reflectance spectroscopy: comparison of partial least-square regression and support vector machine regression met. *International Journal of Remote Sensing*, 34(7), pp.2502–2518.

Zheng, B. et al., 2015. Support vector machines in remote sensing: A review. *International Journal of Applied Earth Observation and Geoinformation*, 34(1), pp.66–75.