

https://dspace.org/

Mapping grass nutrient phosphorus (P) and sodium (NA) across different grass communities using Sentinel-2 data

Mashamba, Tendani

2017

http://hdl.handle.net/10539/23495

Mashamba, Tendani (2017) Mapping grass nutrient phosphorus (P) and sodium (NA) across different grass communities using Sentinel-2 data, University of the Witwatersrand, Johannesburg, <http://hdl.handle.net/10539/23495>

Downloaded from a DSpace repository. https://dspace.org/

UNIVERSITY OF THE WITWATERSRAND

JOHANNESBURG



UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

FACULTY OF SCIENCE

MAPPING GRASS NUTRIENT PHOSPHORUS (P) AND SODIUM (NA) ACROSS DIFFERENT GRASS COMMUNITIES USING SENTINEL-2 DATA

BY

TENDANI MASHAMBA (0604731R)

A research report submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in partial fulfilment of the requirement for the degree of Master of Science (Environmental Sciences) at the *School* of Geography, Archaeology & Environmental Studies

Supervisor: Dr Elhadi Adam

Co-Supervisor: Dr Zakariyya Oumar

MARCH 2017

DECLARATION

I, Tendani Mashamba, declare that this research is my own unaided work. It is being submitted to the Degree of Master of Science in Environmental Sciences to the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other university

……

Signature of candidate

…….day of …….20…… in……

ABSTRACT

Accurate estimates and mapping of grass quality is important for effective rangeland management. The purpose of this research was to map different grass species as well as nutrient Phosphorus (P) and Sodium (Na) concentration across grass communities using Sentinel-2 imagery in Telperion game reserve.

The main objectives of the study were to: map the most common grass communities at the Telperion game reserve using Sentinel-2 imagery using artificial neural network (ANN) classifier and to evaluate the use of Sentinel-2 (MSI) in quantifying grass phosphorus and sodium concentration across different grass communities. Grass phosphorus and sodium concentrations were estimated using Random Forest (RF) regression algorithm, normalized difference vegetation index (NDVI) and the simple ratios (SR) which were calculated from all two possible band combination of Sentinel-2 data.

Results obtained demonstrated woody vegetation as the dominant vegetation and *Aristida congesta* as the most common grass species. The overall classification accuracy = 81%; kappa =0.78 and error rate=0.18 was achieved using the ANN classifier. Regression model for leaf phosphorus concentration prediction both NDVI and SR data sets yielded similar results (R^2 =0.363; RMSE=0.017%) and (R^2 =0.36 2; RMSE=0.0174%). Regression model for leaf sodium using NDVI and SR data sets yielded dissimilar results (R^2 =0.23; RMSE=16.74 mg/kg) and (R^2 =0.15; RMSE =34.08 mg/kg). The overall outcomes of this study demonstrate the capability of Sentinel 2 imagery in mapping vegetation quality (phosphorus and sodium) and quantity.

The study recommends the mapping of grass communities and both phosphorus and sodium concentrations across different seasons to fully understand the distribution of different species across the game reserve as well as variations in foliar concentration of the elements. Such information will guide the reserve managers on resource use and conservation strategies to implement within the reserve. Furthermore, the information will enable conservation managers to understand wildlife distribution and feeding patterns. This will allow integration of effective conservation strategies into decisions on stocking capacity.

ACKNOWLEDGEMENTS

I am very grateful for my supervisor Dr Elhadi Adam for the support, guidance and inputs he provided during my research. I am also thankful to the contributions of the co-supervisor Dr Zakariyya Oumar.

Table of Contents

DECLARATION i
ABSTRACT ii
ACKNOWLEDGEMENTS iii
List of Figures vi
List of Tables vii
LIST OF ABBREVIATIONS viii
Chapter 1: Introduction 1
1.1Background 1
1.2 Problem Statement 3
1.3 Aims and objectives 4
CHAPTER 2 5
2 LITERATURE REVIEW 5
2.1Mapping grass communities quality and quantity 5
2.2Mapping grass quality and quantity using multispectral data 6
2.3 Mapping grass quality and quantity using hyperspectral data 7
2.4 Mapping grass quality and quantity using new generation multispectral 8
2.5 Classification of vegetation 9
2.5.1 Support Vector Machine 9
2.5.2 Random Forest 10
2.5.3 Artificial Neural Network 10
2.5 Conclusion 10
Chapter 3 12
3 Mapping grass communities using Sentinel -2 image 12
3.1Introduction 12
3.2 Methodology 15
3.2.1Study Area 15
3.2.2Image acquisition and pre-processing 16
3.2.3 Ground reference data collection 17
3.3 Image classification 18
3.3.1 Artificial Neural Network (ANN) Classifier 18
3.3.2 Accuracy assessment 19
3.4. Results 20
3.4.1 Classification 20
3.4.2 Accuracy Assessment 20
3.5 Discussion and conclusion 23

Chapter 4 25
4 Mapping grass phosphorus and sodium concentration using Sentinel-2 imagery 25
4.1Introduction 25
4.2 Materials and Methods 28
4.2.1 Study area 28
4.2.2 Remote sensing data acquisition and processing 29
4.2.3 Field data collection 31
4.2.4 The calibration and validation dataset 31
4.2.5 Chemical analysis 31
4.2.6 Data Analysis 32
4.2.7 Random forest regression 32
4.2.8 Selecting the important band indices 34
4.3 Results 34
4.3.1 Descriptive statistics 34
4.3.2 Relationship between the vegetation indices and the phosphorus and sodium concentrations 35
4.3.3 Selecting important predictor variable 38
4.3.4 Spatial distribution of foliar P and Na 44
4.4 Discussion and conclusion 45
4.4.1 Prediction of P and Na 45
4.4.2 Relationship between vegetation indices (NDVIs & SRs) and prediction of P and Na concentrations 47
4.4.3 Mapping of P and Na 48
CHAPTER 5 49
5 CONCLUSION 49
REFERENCES 51

List of Figures

Figure 1: Study area map showing Loskop Mountain Bushveld and Rand Highveld vegetation 16
Figure 2: Classified vegetation of the study area using ANN classifier.....20
Figure 3: Study area map showing Loskop Mountain Bushveld and Rand Highveld vegetation 29
Figure 4: Contour plots representing the correlation (r) relationship between phosphorus NDVI (a) and SR (b) as well as the relationship between sodium NDVI (c) and SR (d) obtained from all two possible band combinations.....37
Figure 5: Measurement of variable of importance (a)-(b) NDVIs and SRs phosphorus (c)-(d) NDVIs and SRs sodium. Highest decrease accuracy represents the most important variables 40
Figure 6: The number of variables selected for prediction using phosphorus (a)(b) NDVIs and SRs (4 and 6) and Sodium NDVIs and SRs (c)(d) (6 and 7) 42
Figure 7: The actual and predicted phosphorus and sodium concentration showing one-one relationship developed using NDVIs and SRs best bands for prediction.....44
Figure 8 Spatial distribution maps of sodium (a) and phosphorus (b) 45
Figure 9: Spatial distribution maps of sodium (a) and phosphorus (b) 45

List of Tables

Table 1Sentinel 2 MSI spatial and spectral resolutions 177
Table 2 Training and test data set for grass species.....**Error! Bookmark not defined.**8
Table 3 Confusion matrix showing the overall classification and Kappa distinguishing among vegetation species..... 21
Table 4 Producer's accuracy (%) and user's accuracy (%) of nine classes (SP,HH,MG,CD,EC,EG,WGand AC) using all Sentinel-2 bands 22
Table 5 Kappa, overall accuracy and Error rate..... 22
Table 6 show Sentinel-2 MSI spatial and spectral resolutions 30
Table 7 Descriptive statistics of the measures phosphorus and sodium concentrations 35

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
AVHRR	Advanced Very High Resolution Radiometer
ESA	European Space Agency
GPS	Global Positioning System
MLP	Multilayer Perceptron
MSI	Multispectral Imager
Na	Sodium
NDVI	Normalised Difference Vegetation Index
NIR	Near Infrared Region
P	Phosphorus
RADAR	Radio Detection and Ranging
REP	Red Edge Position
RF	Random Forest
RMSE	Root Mean Square Error
ROI	Region of Interest
RS	Remote Sensing
SA	South Africa
SAR	Synthetic Aperture Radar
SPOT	Satellite Pour I 'Observation de la Terre
SR	Simple Ratio
SVM	Support Vector Machine
SWIR	Short Wave Infrared Red
VIS	Visible Infrared Spectrum

Chapter 1: Introduction

1.1Background

The feeding patterns of wildlife has been proved to be linked to the quality and quantity of the grass (Mutanga et al. 2004). Grass quality refers to characteristics that influence grass intake by wildlife such as crude protein (a measure of the total protein present in grass and is directly related to digestibility of grasses) and digestibility (a measure of the proportion of grass that can be used by animal). Grass quality is directly linked to the energy content of grass as well as energy needed for body functions (Mutanga et al. 2004). Grasses with higher digestibility are rapidly processed by wildlife resulting in increased production levels (Duru 2000).

Grass growth stages significantly influence digestibility, leafy, greener and actively growing grasses are associated with higher digestibility than aging grasses (Duru 2000). Wildlife require either micro or macro nutrients for physiological well-being (Zhu et al. 2011). Phosphorus (P) and sodium (Na) are important limiting nutrients for grazers, however these nutrients have been reported to be varied between seasons (Knox et al. 2011; Skidmore et al. 2010; McNaughton 1988) and different stages of plant growth (Duru 2000). Sodium (Na) has been found to be an important macro nutrient to mammals as it regulates chemical processes and physiological mechanisms (Mutanga & Skidmore 2004). However, few terrestrial plants require sodium and due to this as a result, herbivores can experience shortage in sodium concentrations.

Current techniques which are used to monitor rangeland management do not adequately take into consideration information on grass quality and are therefore limited in explaining wildlife and livestock distribution and feeding patterns (Mutanga et al. 2016; Adjorlolo et al. 2015). The development of techniques to map the grass diversity and quality such as foliar phosphorus and sodium concentrations helps to describe wildlife distribution and feeding behaviour (Mutanga & Skidmore 2004). Therefore, information on grass quality is essential to gain knowledge on the wildlife distribution and feeding behaviour (Ramoelo et al. 2013a).

Spatial patterns of grass quality and quantity are influenced by various interacting factors (biotic and abiotic) which also vary with spatial scale (Mutanga et al. 2005). To date, rangeland monitoring methods rarely show the nutrient quality of grasses, due to the fact that

many of these methods are limited in explain foraging patterns and behaviours (Mutanga et al. 2005). However, remote sensing application has a significant potential in predicting biochemical foliar concentration (Ramoelo et al. 2013a).

There are two approaches that have been widely used for vegetation estimation, namely: field-based traditional methods and remotely sensed methods (Dube et al. 2014). In order to distinguish the in situ reflectance vegetation measurement, field spectroscopy has been used over the years (Mutanga et al. 2015). This approach has helped researchers to understand the electromagnetic wavelength regions that are important for identifying the vegetation spectral features before the development of imaging spectroscopy (Mutanga et al. 2015). Imaging spectroscopy are field based vegetation mapping techniques which are time-consuming, costly, and sometimes difficult to execute if an area is inaccessible or if land coverage is extensive (Ramoelo 2013; Adam 2010). In addition, imaging spectrometers take measurements in visible and near –infrared regions (400-1000 nm) (Adam 2010). On the other hand, the freely available remotely sensed data presents an opportunity for researchers to effectively and economically generate timely and accurate data for vegetation mapping (Dube & Mutanga 2015; Adam 2010).

In the past, many researchers relied on the use of very high spectral resolution spectrometers to map vegetation species and foliar nutrients such as nitrogen, phosphorus and sodium (Mutanga & Skidmore 2004), biomass estimation (Dube et al. 2014) using remote sensing.

Mapping vegetation species occurs because vegetation such as grass species with different nutrient levels reflects differently at specific wavelengths (Mutanga et al. 2003). Effective monitoring and mapping techniques for grass species are significant for a better understanding of distribution of grass communities within a conservation area (Mutanga & Kumar 2007). However, in SA, the limitations associated with remote sensing applications are that most multispectral images are not freely available, they are expensive to acquire as well as to analyse them (Mutanga et al. 2016). Various vegetation monitoring studies have been limited by compartment levels due to their high acquisition and high spatial and spectral resolution (Dube & Mutanga 2015; Dube & Mutanga 2015). However, cheap and low resolution images are known to greatly compromise accuracies (Knox et al. 2012).

Traditional field techniques for mapping plant species are time consuming, require extensive field work as well as laboratory analysis to predict biochemical and biophysical properties of vegetation (Adam et al. 2012; Mutanga et al. 2003). Therefore, developments in new remote

sensing data sets that can satisfy the scope of work of botanist, ecologists and conservation managers for vegetation monitoring (Mutanga et al. 2016). This is extremely dependent on: development of sensors having decent spatial and spectral resolutions (Mutanga et al. 2016; Adjorlolo et al. 2015). However, the newly launched Sentinel-2 MSI satellite is likely to succeed in various applications. Sentinel-2 MSI has bands that are strategically positioned and is freely available (Mutanga et al. 2016). The spectral band of Sentinel-2 and its 290 km wide swath provides a detailed view of area under investigation which enables the assessment of foliar biochemicals over large protected and communal lands, as well as mapping of the distribution of vegetation.

1.2 Problem Statement

Modelling of vegetation quality is important to improve understanding of wildlife distribution and foraging patterns (Mutanga 2004). Both phosphorus and sodium are limiting nutrients for herbivores in Savana rangeland. Ruminants are known to be selective foragers which prefer plants with high sodium concentrations (Skidmore et al. 2005). The remote sensing of foliar phosphorus and sodium is important for describing wildlife distribution and foraging behaviour but has been poorly investigated as compared to nitrogen (Mutanga & Kumar 2010). Effective game reserve management needs up to date spatial data on the distribution of vegetation communities as well as vegetation quality and quantity (Adam et al. 2010).

In Telperion game reserve, traditional field techniques are solely relied on to gather information about grass quality and quantity. However, these methods are labour intensive; require long hours in field work and laboratory analysis and do not allow for generation of detailed vegetation maps (Mutanga & Skidmore 2004).

Remote sensing technique offer detailed and most recent information on vegetation quality and quantity. Hyperspectral and multispectral remotely sensed data has been successful in discriminating vegetation to species level due to the high spectral resolution (Schmidt & Skidmore 2003). The narrow bands in hyperspectral data allow for detailed discrimination of vegetation species which can be masked if multispectral data is used (Mutanga et al. 2003). The much complicated procedures of processing and analysis of hyperspectral data can result in poor classification or performance of the classifiers (Adam et al. 2012).

Recent vegetation studies in the field of Remote Sensing have used multispectral images such as Worldview-2 and RapidEye (Cho et al. 2015; Adelabu et al. 2014; Ramoelo et al. 2012;

Zengeya et al. 2012). These sensors offer good spectral resolutions but, the high cost associated with these commercial data has limited the accessibility specially in the developing countries such as South Africa (Cho et al. 2015).

Sentinel-2 Multispectral Imager (MSI) has great potential in effective protected area management due to its unique spatial and spectral resolutions as well as higher signal to noise ratio (Sibanda et al. 2015). The use of Sentinel-2 MSI is driven by the fact that it is freely available and has high temporal resolution with wide swath width making it more appropriate for regional applications (Mutanga et al. 2016; Sibanda et al. 2015). Freely available remote sensing data offer private game reserves such as Telperion to utilise remote sensing technology to map grass quality and quantity with no added costs to the reserve financial budget.

Based on the best available data, no study has tested the use of Sentinel-2 data in mapping grass communities and predicting vegetation quality such as P and Na. The results of the study will enable interpretation of the fundamental interactions of grass status with energy of the electromagnetic waves detected by the multispectral sensors. The resulting map of the study together with ground survey data will provide significant information upon which wildlife management decisions in Telperion game reserve can be based on.

1.3 Aims and objectives

The main aims of this study are to explore the utility of Sentinel 2 MSI data in mapping grass community quality and quantity at Telperion game reserve.

The specific objectives are:

- To map the most common grass communities at the game reserve using Sentinel-2 MSI and Artificial Neural Network classifier.
- Test the use of Sentinel-2 MSI in estimating foliar phosphorus and sodium concentration across different grass communities.

CHAPTER 2

2 LITERATURE REVIEW

2.1Mapping grass communities quality and quantity

Grassland ecosystems are estimated to cover up to 40.5% of the total land area on earth (White et al. 2000). However, biodiversity in grasslands is impacted upon by various factors such as type of land use practices, landscape structure and topography (Hubbell 2001). The existence of grass species that are significant for grazing in rangelands have been affected by overgrazing and unsustainable agricultural activities (Skidmore et al. 2010). For sustainable and effective national resource management, spatial data on grass species and communities distribution is essential to understand land productivity and the ecosystem functions and services (Ramoelo et al. 2015; Svoray et al. 2013; Cho et al. 2012; Adjorlolo et al. 2012) and livestock feeding patterns (Knox et al. 2012).

The estimation of vegetation traits (biophysical and biochemical) of vegetation is important in many ecological application (Asner 1998). In grasslands, accurate estimates of quality and quantity are significant as there is variation in space and time. Detailed maps of grass biophysical variables are helpful in managing and conserving conservation areas and endangered habitats.

In natural ecosystem, understanding of plant growth processes is guided by characterization of plant nutrients such as sodium and phosphorus (Ramoelo et al. 2013). Mineral nutrients such as phosphorus, sodium, calcium and potassium have been found to be limiting minerals for herbivores in savanna rangelands (McNaughton 1983).

Various studies have demonstrated that wildlife feeding behaviour and distribution is largely determined by the pasture quality and quantity (Mutanga & Kumar 2007; Adjorlolo et al. 2015). Estimation of grass P and Na are important in understanding wildlife feeding and migration patterns (Wang et al. 2015). Information on grass quality and quantity is essential for rangeland management (Kaszta et al. 2016) and can be used to inform conservation and

resource use within conservation areas (Mutanga & Skidmore 2007), monitoring of grazing patterns (Skidmore et al. 2005) and productivity of rangeland ecosystems (Mutanga 2004). East and Southern African studies have revealed nitrogen as the most restricting nutrient for herbivores and it was also found that migration patterns of wildebeest in the Tarangire ecosystem is determined by phosphorus (Mutanga & Kumar 2007; Mutanga 2004). Mapping of grass communities has been dependant on traditional techniques such as literature review, use of aerial photographs, field surveys and image spectroscopy (Adam et al. 2012; Youngentob et al. 2012). For example , Ramoelo et al (2013) determined the foliar nitrogen: phosphorus ratio using field spectroscopy, while Mutanga et al (2004) discriminated sodium concentration in mixed grass species using field spectroscopy. However, these techniques require intensive labour, complicated analysis and often data captured does not fully represent the landscape (Adam et al. 2014).

Remote sensing provides multiple advantages over traditional field mapping methods such as time efficient maps, accessibility to complex terrains and repeat mapping of same area over different periods (Mutanga et al. 2016).

2.2Mapping grass quality and quantity using multispectral data

Over the years, different vegetation types and ecosystems have been successfully mapped using multispectral data. Using integrated LIDAR data in combination with multispectral images to classify vegetation in Canada, Bork and Su (2007) reported an overall accuracy of 91%.

In Finland, Toivonen and Luoto (2003) used Landsat imagery in mapping grasslands and achieved high accuracy results of 89% ; South American grasslands were differentiated by Baldi et al (2006) with accuracies between 90% and 95%; (Masocha & Skidmore 2011) achieved an overall accuracy of 80% in their study when they discriminated warm and cold season grasslands in the USA. In Mangolia, Yokota et al (2005) assessed vegetation indices of MODIS with those of AVHRR while monitoring the forage quality and quantity.

The assessment of grassland use and monitoring of changes over time has been done in certain studies. In Peru, object based classification of Landsat image between 1987 and 2007 was used to investigate changes in grassland ecosystem, an annual decrease in the spatial extent of the grassland ecosystem was observed and an overall accuracy of 80.3% was achieved. Price et al (2002) investigated in USA the potential of Landsat and ERS-2 Synthetic Aperture Radar (SAR) imagery together with the dataset to discriminate among

three commonly used grass in areas with high cool and warm season grasses. An accuracy value of 91% was achieved using Landsat data and 73.2% using the SAR data. Masocha and Skidmore (2011) implemented a hybrid classification method in southern Zimbabwe to map invasive species.

2.3 Mapping grass quality and quantity using hyperspectral data

Vegetation quality and quantity has been successfully mapped through the use of hyperspectral data.

Cushnahan et al (2011) in New Zealand successfully identified grass species with a high accuracy using hyperspectral data. Boschetti et al (2007) used Landsat 7 to assess pasture production in the Italian Alps and produced a biomass mapping. Adam et al (2014) also followed the same methodology to estimate biomass using hyperspectral data. They successfully mapped and identified wetland vegetation species using hyperspectral data. They achieved high individual species class accuracies which ranged from 93.73% to 100% and kappa value of 0.87.

Mutanga and Kumar (2007) found a correlation between phosphorus and nitrogen in their study to estimate phosphorus and nitrogen in South African Savannas using hyperspectral data. However, limited studies to predict foliar phosphorus and sodium have been done in rangelands. Since there is a correlation between phosphorus and nitrogen, an assumption that through the prediction and mapping of foliar nitrogen, phosphorus has been indirectly predicted and mapped. Ramoelo et al (2013) determined a high phosphorus and nitrogen ratio in their study using hyperspectral data. Zhang et al (2011) also confirmed that there is a synergistic relationship between foliar nitrogen and phosphorus in grassland species in a study done in China.

Mutanga (2004) found variations of foliar sodium in different species ranging from 0.02 % to 0.2 %. However, Axelsson et al (2012) were unsuccessful in using hyperspectral data to map foliar phosphorus and sodium in mangrove species, as they only mapped and predicted nitrogen. Ferwerda and Skidmore (2007) successfully studied the foliar concentrations of phosphorus, sodium, calcium and magnesium in three tree species and one shrub. They obtained R^2 = 0.51 for phosphorus and sodium was best predicted using continuum-removed spectra (R^2 =0.02).

Different remote sensing techniques offer significant and reliable information about the distribution of forage quality which can be integrated in decision making in the management and conservation of rangelands (Svoray et al. 2013). There are challenges associated with the use of hyperspectral data such as high dimensionality caused by narrow bands, therefore, in order to discriminate vegetation species accurately, large training points must be created (C Adjorlolo et al. 2012).

Additional detailed investigations using hyperspectral remotely sensed data to assess grassland quality are needed. Such investigations will exploit the many bands of the vegetation indices at different wavelengths making it possible for researches to extract more vegetation parameters.

2.4 Mapping grass quality and quantity using new generation multispectral

The launch of the new generation sensors which have high spectral and spatial resolutions is seen as advancement to the hyperspectral and multispectral sensors. These new sensor offer un- matched spectral and spatial spectral resolution which offers researchers opportunity for much detailed monitoring and assessment of vegetation.

The ability of multispectral data to estimate and map grass biochemical content such as nitrogen in Africa has been investigated by Adjorlolo and Botha (2015). Furthermore, successful mapping of estuarine vegetation in St Lucia was achieved by Luck –Vogel et al (2016). However, there is a shortage of studies to map foliar phosphorus and sodium using new generation multispectral data.

Odindi et al (2014) was successful in evaluating the utility of the new generation WorldView-2 and SPOT -5 images in mapping the bracken fern in South Africa. The overall accuracy of 84.72% and 72.22% were achieved for the WorldView -2 and SPOT-5 images. A similar study in Australia by Rasel et al (2016) compared capability of WorldView-2 and Landsat 8 in saltmarsh mapping and accuracy assessment,WorldView-2 data produced a higher accuracy of 92.12% and Landsat data yielded overall accuracy of 82.04%. The high performance of the new generation sensors was illustrated in these studies.

In SA, Mansour and Mutanga (2012) successfully used Worldview data in mapping grass species in order to assess the grassland degradation. They successfully discriminated decreaser and increaser grass species and reported an overall accuracy of 90% and kappa of

0.87. However, they reported that the cost of image acquisition is too high and this can limit research in resource constrained countries.

Various remote sensing methods have been used to obtain information on grasslands biophysical parameters which include biomass, quality, and land degradation, grazing and stock.

Mapping of vegetation species using remote sensing data requires image classification techniques. However, in addition to the choice of the suitable remotely sensed data in terms of the cost and the resolutions (spectral and spatial), the choice of appropriate classification techniques is significant in order to produce reliable accurate vegetation mapping (Adam et al. 2014).

2.5 Classification of vegetation

Image classification is considered as one of the significant steps in remote sensing studies as it because it retrieves and converts vegetation information from satellites image into thematic layers that can be easily interpreted (Cingolani et al. 2004; Schmidt & Skidmore 2003). There are different classification algorithms that have been successfully applied to gather vegetation information from multispectral and hyperspectral remotely sensed data (Omer et al. 2015; Dube & Mutanga 2015a; Adelabu et al. 2014; Mutanga & Rugege 2006). However, much recognition has been given to the advanced methods of image classification such as support vector machines (SVM), Random Forest (RF) and artificial neural networks (ANN) due to their greater effectiveness in image processing. However, each of these methods has its own advantages and limitations and they are briefly described below.

2.5.1 Support Vector Machine

The SVM has been defined as a supervised classifier trained to find the most effective classification hyperlane through reducing the upper bound of the classification error (Cortes and Vapkin 1995; Dlamini et al. 2016). In SVM, the hyperlane leaves the maximum margin between two classes, where the margin has been defined as the sum of the distances of the hyperlane from the nearest point of the two classes, if there is inseparability between the classes the SVM searches for the hyperlane that maximizes the margin while minimizing a quantity proportional to the number of misclassification errors (Vapkin 1995; Dlamini et al. 2016).

According to Seetha et al (2017) the benefits of SVM are that: overfitting problem is unlikely to occur; it models nonlinear class boundaries; reduces computational complexity, easy to control complexity of decision rule and frequency of error.

Seetha et al (2017) further explained that the limitations associated with the use of SVM are: slow training; difficulty in determining optimal parameters when training data is not linearly separable and the structure of the algorithm is complex and users find it difficult to understand. However, SVM is believed to be competitive in the classification of high dimensionality data sets.

2.5.2 Random Forest

RF classifier is a combination of tree classifiers produced through the use of a random vector sampled independently from the input vector, each tree votes for the most popular class to classify an input vector (Breiman 1999). The major advantage of RF classifier is that it requires that two parameters be set; can handle categorical data, and can detect outliers through the use of proximity analysis.

2.5.3 Artificial Neural Network

In this study ANN was chosen because the textural features used in ANN assist in solving misclassification associated Seetha et al (2017). Furthermore, ANN has the ability to tolerate noisy inputs.

In their study, Seetha et al (2017) concluded that neural network approach of classification improves the accuracy and finer information from the individual class is obtained through the use of texture.

ANN performance has been reported to be higher that other traditional classifiers such as maximum likelihood, minimum distance to mean and decision trees (Atkinson & Tatnall 1997). Such classification methods are restricted due to the use of large datasets and use of limited training samples to generate a map (Omer et al. 2015).

2.5 Conclusion

Traditional methods for mapping vegetation quality and quantity have been successful used. However, these techniques only allow for assessments at a local scale. The advancement of remote sensing provided researchers with an opportunity to map vegetation quality at national and regional scale. However, this opportunity is costly as most of the remote sensing data are

commercialised and not freely available. The newly launched Sentinel-2 has red edge bands which allows for grass mapping and estimation of foliar parameters at a regional scale. The potential of Sentinel-2 mapping grass species has never been done but its application in forestry yielded satisfactory results on grass quality and biomass estimation.

Great outstanding progress in the field of remote sensing from the use of aerial photographs to the use of multispectral scanners (Mutanga et al. 2016) has been made. The importance of aerial photographs cannot be ignored but much attention has been focussed towards the use of multispectral scanners due to higher spatial and spectral resolution and low costs (Mutanga et al. 2016). Multispectral sensor technology has advanced from new generation multispectral scanner, second generation Landsat Thematic Mapper (TM) and Satellite Pour I'Observation de la Terre (SPOT) with higher accuracy (Mutanga et al. 2016).

Strategic bands within multispectral sensors allow for detection of foliar biochemicals and discrimination of vegetation at species level. The use of multispectral sensors such as Worldview-2 and RapidEye have produced similar results to those achieved using hyperspectral data (Mutanga et al. 2016; Adjorlolo et al. 2015).

Chapter 3

3 Mapping grass communities using Sentinel -2 image

3.1Introduction

Grassland biome is classified as being the second largest biome after Savanna biome in South Africa (Carbutt & Martindale 2014). The main threats within the Grassland Biome are agriculture, coal mining and alien invasive species (Carbutt et al. 2011).

The existence of grass species that are significant for grazing in rangelands have been affected by overgrazing and unsustainable agricultural activities (Skidmore et al. 2010). For sustainable and effective national resource management, spatial data about the distribution of the grass species and communities is essential to gain knowledge on land productivity and the ecosystem functions and services (Ramoelo et al. 2015; Svoray et al. 2013; Cho et al. 2012; Mansour et al. 2012; Adjorlolo et al. 2012) and livestock feeding patterns (Knox et al. 2012).

Traditional survey methods such as field observation, indigenous knowledge and photographs can yield better results of mapping grass communities at local scale (Adjorlolo et al. 2012). However, such traditional techniques involve intensive field work which is time consuming and costly (Adam et al. 2012).

Over the last three decades, remote sensing has increasingly been used in mapping vegetation species (Skidmore et al. 2010b; Adelabu et al. 2014; Thenkabail et al. 2016; Vaglio Laurin et al. 2016a; Curran 1989). Remote sensing has been reported to be timely, cost effective techniques in providing accurate spatial information on grass species quality and quality (Ramoelo et al. 2015).

Both multispectral and hyperspectral remote sensing data have been successfully used in mapping grass species and communities. Vegetation mapping studies have successfully been done using multispectral data (Odindi et al. 2014; Adam, et al. 2014; Mansour et al. 2012; Ingram et al. 2005).

Multispectral data are affordable and easy to acquire and to process (Mansour et al. 2016). There are limitations that have been identified to be associated with multispectral images such as Landsat and SPOT when mapping vegetation and plant species.

The effectiveness of multispectral data when mapping vegetation species is limited by their spectral and spatial resolutions, which are sometimes not useful for vegetation mapping and identification (Adam et al. 2017).

Spatial and temporal resolution limitations associated with Landsat imagery makes it difficult to map and distinguish species at species level in heterogeneous environments. Furthermore, Landsat image quality decreases when the interest of period is within a rainy season.

Hyperspectral remotely sensed data has been most commonly and successfully used in vegetation mapping as they possess advantageous narrow bands (Thenkabail et al. 2011; Zhang et al. 2011; Skidmore et al. 2005; Marshall & Thenkabail 2015). The high spectral resolution allows for identification of spectral features that are vital for discriminating different grass species and classifying them at species level which is otherwise be masked by multispectral data (Mansour et al. 2016; Cho et al. 2012; Vaglio Laurin et al. 2016b). However, whereas high spectral resolution of the hyperspectral data allows accurate detection and identification of vegetation species, many challenges such as data cost, poor accessibility to the data, image processing and analysis have limited the use of hyperspectral data particularly in developing countries (Ramoelo et al. 2013).

Over the last 10 years, there has been development of new-generation satellites such as RapidEye, SumbandilaSAT and WorldView2 (Adam et al. 2014) which are considered to be much more improved than first generation multispectral sensors (Ramoelo et al. 2015). These new sensors allow for more detailed classification of complex vegetation communities (Stratoulias et al. 2015). This is due to the fine spatial resolution coupled with a sensible number of spectral bands arranged in different regions of the electromagnetic spectrum (red edge) (Ramoelo et al. 2015).

Researchers have successfully investigated the use of these new multispectral data in different vegetation studies such as vegetation species mapping (Vaglio Laurin et al. 2016; Mansour et al. 2016; Odindi et al. 2014; Dalponte et al. 2012). However, since these are commercial sensors, they come with their own limitations in terms of cost and availability (Adam et al. 2014).

On the 23rd of June 2015, the European Space Agency (ESA) launched Sentinel-2A satellite which is one of the very first two of the Sentinel-2 mission (Navarro et al. 2017). Sentinel-2 has a high resolution multispectral Instrument (MSI) with a wide swath (Sibanda et al. 2015). The MSI has 13 spectral bands which stretch from visible infrared spectrum (VIS) and near

infrared region (NIR) to short wave infrared (SWIR) at different spatial resolution allows for measurement of earth's reflected radiance (Navarro et al. 2017). The integration of the four narrow bands in red edge region of Sentinel allows for enriched vegetation assessment (Stratoulias et al. 2015). The wide swath and frequent revisit time of Sentinel is advantageous in that it allows the sensor to provide continuous monitoring of vegetation (Sibanda et al. 2015). Sentinel was designed to outperform the Landsat mission (Stratoulias et al. 2015).

Classification methods are important for mapping of vegetation species using remotely sensed. Hence, in addition to the choice of the suitable remotely sensed data in terms of the cost and the resolutions, the selection of appropriate classification method is vital for reliable accurate vegetation mapping (Adam et al. 2014).

Numerous advanced classification algorithms such as classification trees, random forest (RF), artificial neural networks (ANN), support vector machines (SVM), partial least squares regression have been applied to derive vegetation data from multispectral and hyperspectral images (Omer et al. 2015; Dube & Mutanga 2015a; Adelabu et al. 2014; Mutanga & Rugege 2006). However, much attention has been given to RF, SVM and ANN due to their greater image processing effeciveness. The application of ANN algorithm has been widely used in classifying forest species using multispectral imagery (Atkinson & Tatnall 1997). ANN is a machine learning system which encompasses interrelated linkages of indeterminate processing elements (Atkinson & Tatnall 1997; Liu et al. 2013; Omer et al. 2015). These linkages are categorized by a robust pattern recognition ability allowing representation of complex multivariate data forms (Atkinson & Tatnall 1997). Advantages of ANN have been attributed to its statistical methods which include generalization for technique with multiple images (Wang et al. 2009), easy adaptation to different types of data and input and its ability to use limited training data to categorize data as compared with other traditional classifiers (Omer et al. 2015).

The ANN performance has been demonstrated in previous studies to be higher than that of traditional classifiers such as maximum likelihood, minimum distance to mean and decision trees (Atkinson & Tatnall 1997). These traditional classification methods are dependent on the assumptions which restrict their potential to use large datasets and map areas with restricted training samples (Omer et al. 2015)

On the basis of the best available knowledge, Sentinel-2 data has never been used in mapping of grass species. Therefore, the objective of this study was to examine the utility of the

Sentinel-2 satellite data for mapping grass species in Telperion nature reserve using ANN classification algorithm.

3.2 Methodology

3.2.1 Study Area

The study area is Telperion Nature Reserve which forms part of the eZemvelo nature reserve. The reserve is approximately 13000 hectares in size and situated at the border between Gauteng and Mpumalanga Provinces, Telperion falls within the far north of the Grassland Biome located in Mpumalanga Province of South Africa (25° 41′ S, 28° 56′ E) (Fig 1). Mucina and Rutherford (2006) classified the vegetation type as the Mixed Bushveld and Bankenveld Grassland vegetation type.

These vegetation types is considered to be endangered and small patches are protected in statutory conservation reserves and private reserves (Mucina and Rutherford, 2005) and the domininat grass species identified by (Coetzee 2012) are*: Themeda trianda, Eragrostis curvula, Sporobolus africanas* ; *Setaria sphacelata* ,*Hyparrhenia hirta* and *Cynodon dactylon*.

The study area is characterised by a mean annual temperature of and receives summer rainfall. The Wilge River which is a tributary of the Oliphant's River, flows northwards through the reserve, the river is fed by various wetlands and smaller streams.

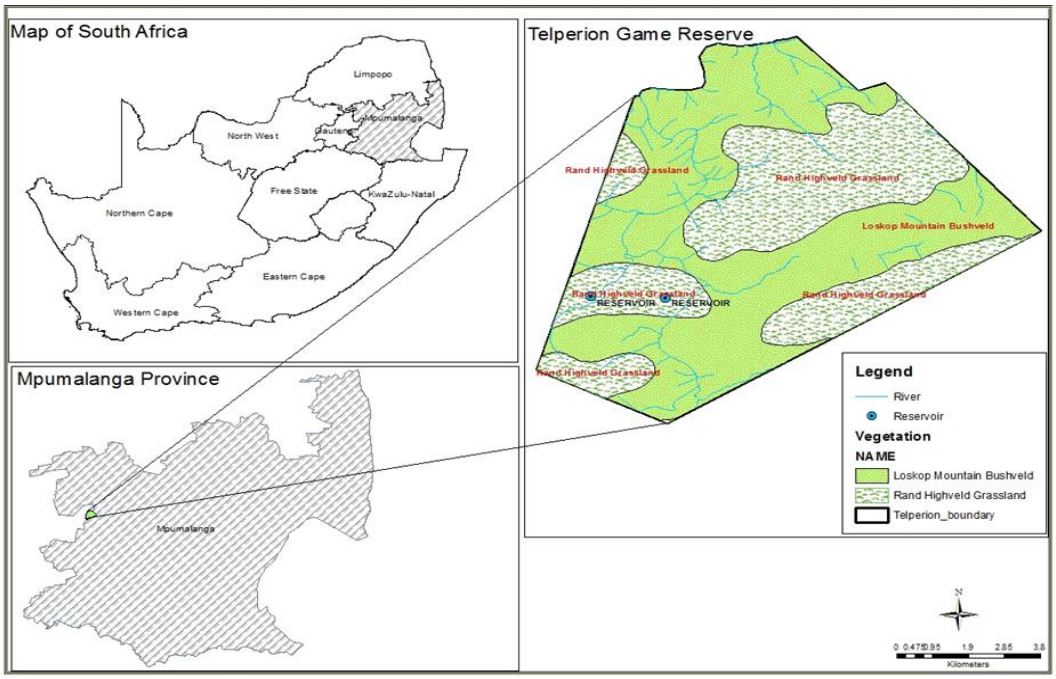


Figure 1: Study area map showing Loskop Mountain Bushveld and Rand Highveld vegetation

3.2.2Image acquisition and pre-processing

An image from Sentinel 2 sensor was used in this study. Sentinel-2 image was downloaded from the Sentinel's Scientific Data Hub (https:scihub.copernicus.eu/) on the 20th of May 2016. The data was radiometrically and geometrically corrected using the Top-Of Atmosphere (TOA) products. Additional corrections which were done include orthorectification as well as spatial registration on a global reference system. Sentinel 2 Toolbox was used to process the image to Level 2 data.

The image was atmospherically corrected using Sen2Cor plugin which is a prototype processor of Sentinel-2. Sen2Cor processor has been designed to execute functions such as atmospheric, terrain and cirrus correction as well as a scene classification. The image was re-sampled based on the sampling distance (10m,20m,60m) as per the spectral bands resolution as shown in Table 1 below.

Re-sampling of the image was done using a constant Ground Sampling Distance (GSD) of 10m, 20m and 60m based on the resolution of different spectral bands as shown in Table 1 below.

Table 1Sentinel 2 MSI spatial and spectral resolutions

Acronym	Band	Central Wave-length (nm)	Spatial resolution
Band 1	Violet	443	60
Band 2	Blue	490	10
Band 3	Green	560	10
Band 4	Red	665	10
Band 5	Red-edge 1	705	20
Band 6	Red-edge 2	740	20
Band 7	Red-edge 3	783	20
Band 8	NIR	842	10
Band 8a	NIR narrow	865	20
Band 9	NIR	945	60
Band 10	NIR	1380	60
Band 11	SWIR1	1610	20
Band 12	SWIR2	2190	20

(ESA, 2016)

3.2.3 Ground reference data collection

Sampling was done between the 21st and the 25th of May 2016 soon after remotely sensed data acquisition. This period is usually characterised by moderate to high rainfall and high biomass productivity, but in the year 2016, the country was experiencing severe drought which was declared as a natural disaster. As a result, the grass was drier than expected during the time of sampling.

In each identified sample point, vegetation plot (20m by 20m) was constructed comprising of a homogenous area of the grass communities. The ground reference data was then randomly partitioned for classification into training sites (70%) and test sites (30%) as shown in Table 2 below.

Using field sample points a point map was created. The map was then overlaid on the Sentinel-2 image to Region of Interests (ROIs) for training and test which were then used to gather image spectra from grass species.

Table 2.Training and test data set for grass species

Species	**Training samples (70%)**	**Test samples (30%)**	**Total samples**
Serephium plumosum	48	20	68
Hyparrhenia hirta	41	17	58
Mixed grassland	38	15	53
Cynadon dactylon	35	14	49
Eragrostis chloromelas	52	22	74
Woody vegetation	94	40	134
Wetland grass	42	18	60
Aristida congesta	42	17	59
Eragrostis gummiflua	45	19	64

3.3 Image classification

Artificial Neural Network algorithm was used to classify the Sentinel-2 image. A resultant map showed the spatial distribution of different vegetation communities in Telperion game reserve.

3.3.1 Artificial Neural Network (ANN) Classifier

Neural Networks are described as knowledge based artificial intelligence technique (Atkinson & Tatnall 1997). ANN is non-parametric mathematical model independent of assumptions of data normality (Omer et al. 2016). ANN attempts to imitate mimic neural linkages (Omer et al. 2015) based on biological and functional aspects by using a group of artificial neurons

which are connected in order to process the information (Omer et al. 2016; Atkinson & Tatnall 1997; Omer et al. 2015). ANN is better compared to other classification methods (Omer et al. 2015). The advantage is attributed to its nonparametric nature, as such; no previous information on the distribution of the input data is required (Omer et al. 2015). ANN has gain popularity and is used much more often than other statistical classifier (Omer et al. 2015). This is due to advantages such as high accuracy as compared other statistical classifiers (Omer et al. 2015) as well as and once they are trained; integrate unique data types including different sensor information into analysis, thus enabling collaboration in different field of studies (Atkinson & Tatnall 1997). ANN has various models that have been used in remote sensing (Omer et al. 2015), these have been identified as back propagation (BP), RBF and Multilayer perceptron (MLP) (Omer et al. 2015). The ANN commonly uses the MPL structure which is comprised of input layer and nonlinear activating nodes (Omer et al. 2015). When data has been processed, perceptron learning is achieved through modifications in linkages. An MPL has been described as a ANN model that maps input data based on suitable output (Omer et al. 2016). Various land cover classification studies have used ANN to classify different land cover as well as vegetation spectral discrimination. The most popular ANN structure of algorithm is MPL model that uses the BP algorithm. In this study, ANN classification algorithm was used for mapping different grass communities using Sentinel 2 data.

3.3.2 Accuracy assessment

In classification, accuracy assessments are important. In this study, a 70% subset of the reference data was used for training the model. The model validation was done using the 30% of the remaining data. In order to make a comparison of the true class and the predicted class, a confusion matrix was created (Omer et al. 2016). From the confusion matrix the overall accuracy (OA) which is the overall probability that test data has been accurately classified (Congalton & Green) and producer's accuracy (PA) defined as the possibility of a certain class being accurately classified and user's accuracy which is the reliability of the classes that are in the classified image (Congalton & Green 2008) were calculated. The kappa coefficient lies between 0 and 1 values and it represents the variation between the actual and expected agreement (Adam et al. 2014). When kappa coefficient is 1, it shows the best agreement between classification and ground pixels, non-agreement is represented by a value of 0 (Dorn et al. 2015).

3.4. Results

3.4.1 Classification

The below map shows distinct species that were classified using ANN classifier and their distribution across the study area. Woody vegetation was classified as the most dominant vegetation type within the study area.

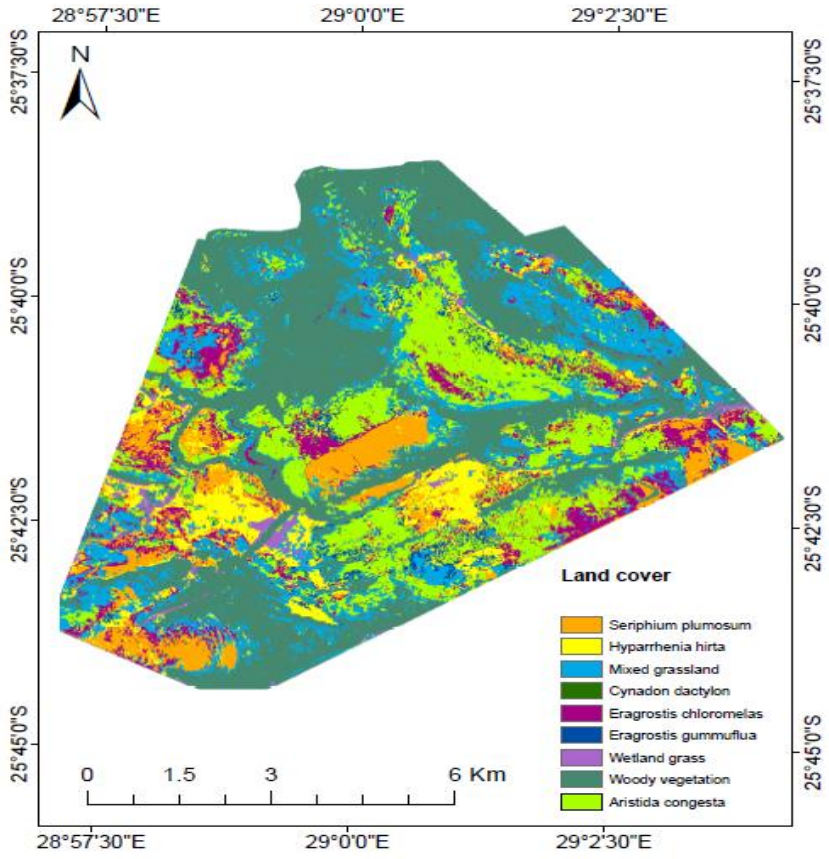


Figure 2: Classified vegetation of the study area using ANN classifier

3.4.2 Accuracy Assessment

The classification results showed the performance of ANN algorithm in distinguishing different species as shown in Table 3 below. The classification model achieved overall accuracy of 81% and Kappa Coefficient of 0.78% as shown in Table 4 below. Variations of User and producer accuracies of single species are depicted in Table 4.

Table 2 Confusion matrix showing the overall classification and Kappa distinguishing among vegetation species.

	AC	CD	EC	EG	HH	MG	SP	WG	WV	Row Total
AC	**11**	2	0	0	0	0	0	0	1	14
CD	0	**12**	0	2	0	1	0	0	0	15
EC	0	0	**19**	1	0	1	1	0	1	23
EG	0	0	0	**16**	0	0	1	0	0	17
HH	1	0	0	1	**16**	1	0	0	0	19
MG	1	0	2	0	0	**11**	1	1	0	16
SP	2	0	1	1	0	1	**13**	0	3	21
WG	0	2	0	0	0	0	1	**15**	0	18
WV	0	0	0	1	1	0	2	0	**35**	39
Column Total	15	16	22	22	17	15	19	16	40	**182**

The above confusion matrix shows the performance of the ANN model in classification. Different vegetation species were classified; *Serephium plumosum* (SP);*Hyparrhania hirta* (HH); Mixed grassland (MG) ;*Cynadon dactylon* (CD); *Eragrostis chloromelas* (EC);*Eragrostis gummiflua* (EG)Wetland grass (WG) , Woody vegetation (WV) and *Aristida Congesta* (AC) The error was determined based on the OOB method and test dataset (Mansour & Mutanga, 2012).

Table 3 Producer's accuracy (%) and user's accuracy (%) of nine classes (SP, HH, MG, CD, EC, EG, WG and AC) using all Sentinel-2 bands

	AC	CD	EC	EG	HH	MG	SP	WG	WV
Producer accuracy	73.33333	75	86.36364	72.72727	94.11765	73.33333	68.42105	93.75	87.5
User accuracy	78.57143	80	82.6087	94.11765	84.21053	68.75	61.90476	83.33333	89.74359

The accuracy achieved for all species (MG, HH, CD, EC, EG, WG, WV, AC, MG) was high. HH class had the highest 94% (HH) and the lowest accuracy was achieved was 64 (SP). The SP class was underestimated in the classified image.

The reliability of pixels that were correctly classified was higher in MG, HH, CD, EC, EG, WG, WV and AC. However, MG only achieved the lowest user's accuracy of 62% meaning that there were some pixels that were misclassified.

The achieved Kappa was 0.78 and the error rate of 0.18 was achieved in this study as shown in Table 5 below. The low error rate achieved by the classifier demonstrates the high performance of the classifier in correctly classifying the different grass species.

Table 4 Kappa, overall accuracy and Error rate

Kappa	0.786503
Overall accuracy	0.813187
Error rate	0.186813

3.5 Discussion and conclusion

The ability to accurately locate and map different grass species within rangeland provides valuable information (Ramoelo et al. 2013) which can help to guide effective management of natural resources in rangeland and give understanding on wildlife and livestock grazing and distribution patterns.

In this study the performance of ANN algorithm and the ability of the Sentinel-2 data to map the different grass species and their distributions in Telperion Nature Reserve was assessed. Woody vegetation was found to be the most widely distributed vegetation within the study area as shown in Fig. 1 with dominant grass species mapped as *Aristida congesta*.

Multispectral satellites with medium spatial resolution (30-100m) have produced grass species maps with lower accuracy for operational application. Imagery from sensors with very high spatial resolution (<5m) have been reported to have limitations such as cost, availability, high dimensionality and processing (Adam et al. 2014). Application of remotely sensed data combined with field data is restricted by the limitations that are associated with the fine and medium scale remote sensing data.

The high overall accuracy 81% (Table 4) achieved demonstrates the ability to map different vegetation types and species through the use of multispectral data. Such information provides more details on grass species distribution within conservation area. A kappa coefficient of 0.7 was achieved which shows that there was a perfect agreement between the classification and ground pixels (Table 5). A study by Omer et al (2015) yielded relatively lower accuracy (65%) when ANN algorithm was used.

The overall accuracy produced in this study is higher compared to overall accuracy achieved by Mansour et al (2015) (75%) using multispectral SPOT 5 imagery. Using WorldView-2 data, Mansour and Mutanga (2012) successfully mapped grass species and obtain overall accuracy of 90%.

Higher vegetation classification accuracies have been reported in studies using hyperspectral data (Vaglio Laurin et al. 2016a). However, the limitations associated with the use of hyperspectral data is the high cost, small spatial coverage, and the difficulties of image

processing (Thenkabail et al. 2016). Hyperspectral data can only be used for small scale surveys but can produce maps of exceptional thematic detail (Mutanga & Kumar 2007).

To increase accuracy, similar study using Sentinel-2 can be replicated but across different seasons. This study was done during a period of severe drought in South Africa. As a result, the grass was dry and at low productivity stage when image was acquired. As grasses senesce, they exhibit red/brown colour which create spectral confusion with the bare soil. Additionally, a much more advanced classification algorithm may be used to achieve higher accuracy results.

Classification information obtained from this study is useful for effective resource management and will help rangeland managers to identify the distribution most dominant grass species. Such information will give guidance to rangeland managers as to which species to introduce and as well as grazing/ feeding patterns of wildlife and livestock.

Chapter 4

4 Mapping grass phosphorus and sodium concentration using Sentinel-2 imagery

4.1 Introduction

South African grassland biome is considered to be the most agriculturally productive land and has numerous mineral deposits most of which are extracted for economic gains (SANBI 2012). One of the most challenges experienced in grassland ecosystem is the high levels of transformations as a result of overgrazing and overstocking (Ali et al. 2016). Understanding the grassland ecosystem quality and quantity is crucial for explaining the functions and dynamics of grassland.

Grass quality refers to the characteristics that influence grass intake by wildlife such as crude and digestibility which is a measure of the proportion of grass that can be used by animal (Mutanga 2004). The quantity and quality of vegetation determines the amount of material that can be digested by herbivores (Knox et al. 2011).

Grass nutrients such as phosphorus, sodium, magnesium and calcium have been identified as essential nutrients which determine foraging preference in herbivores (Knox et al. 2011; Prins 1996). Spatial information about grass nutrients is essential to advice rangeland managers to manage their grazing lands effectively. Studies done in savanna rangelands revealed that phosphorus is one of the limiting minerals in herbivores (MacNaughton & Banyikiwa, 1995) whereas sodium was found to influence the distribution of wildlife as herbivores prefer sodium enriched forage (Mutanga & Skidmore 2004).

Mapping of grass (phosphorus) (P) and sodium (Na) provides useful data for sustainable resource use, planning and management of wildlife and livestock by nature reserve managers, conservation planners and livestock farmers (Ramoelo et al. 2012b). Leaf phosphorus concentration determines the migration behaviour of wildebeest in the East African savanna (Voeten 1999).

Studies to estimate vegetation quality have mainly been focussed on mapping foliar concentration of nitrogen (Kaszta et al. 2016; Vaglio Laurin et al. 2016; Zhang et al. 2011; Marshall et al. 2016; Skidmore et al. 2005; Ling et al. 2014) and useful results have been achieved using different remote sensing data and techniques. However, only a handful of

studies have been done to map foliar phosphorus concentration even though it is considered to be of equivalent importance as nitrogen in establishing herbivore feeding patterns (Mutanga & Kumar 2007; Mahajan et al. 2014; Ramoelo et al. 2013).

Wildlife prefer specific feeding sites and plant species to maximise essential nutrient uptake and avoid the risk of malnutrition (Mutanga et al. 2004). The advancement of various pasture quality predicting techniques is important for explaining wildlife feeding and distribution patterns (Mutanga et al. 2004).

Traditional methods for detecting and mapping pasture quality are associated with high laboratory analysis costs and require comprehensive sampling. These methods have been found to be limited to small and local scale mapping (Adam et al. 2014).

The advancement of spatial, spectral and temporal resolution in remote sensors has awarded researchers with an opportunity to evaluate and monitor vegetation quality and quantity conditions (Curran et al. 2001). Satellite remote sensing offers cost effective methods for surmounting the challenge of field-based survey (Mutanga et al. 2003; Mutanga et al. 2005; Skidmore et al. 2010; Mutanga & Skidmore 2007; Mutanga & Skidmore 2004) Multispectral and hyperspectral remote sensing methods have been developed steadily (Adam & Mutanga, 2009) and successfully over time to predict vegetation quality and quantity (Liu et al. 2015; Atzberger et al. 2015; Zlinszky et al. 2015 Knox et al. 2011; Ausseil et al. 2011; Ferwerda et al. 2007).

The lack of the spectral resolution has been considered to be one of the limitations with the use of the vegetation indices calculated from broad multispectral bands (Blackburn 1998; Thenkabail 2000). However, Multispectral data has been found to be cost effective and easy to process (Mansour et al. 2016).

Various studies have successfully assessed vegetation quality through the use of hyperspectral remote sensing data (Thenkabail et al. 2016; Kaszta et al. 2016b; Zhang et al. 2011; Skidmore et al. 2005; Knox et al. 2011b; Wang et al. 2015; Mutanga & Skidmore 2004). The contiguous of hyperspectral narrow bands make it possible to understand and estimate the fluctuations in biochemical and biophysical features of plants as which has been difficult with the use of multispectral remote sensing (Zhang et al. 2011).

In the use of hyperspectral, studies have found that the shape and position of the “red edge” (the transition from visible to near-infrared portions of the spectrum - 690 to 800 nm) was

most sensitive to vegetation quality variations). The term " red edge" refers to the region of sudden change in leaf reflectance between 690 and 800 nm (Mutanga & Skidmore 2007) because of high chlorophyll absorption in red wavelengths and high reflectance in the NIR wavelengths due to internal leaf scattering (Horler et al. 1983). Multispectral sensors such as Landsat, SPOT, ASTER, and MODIS have coarse spatial resolution and do not possess the red edge region (Sahoo et al. 2015)

In the last 10 years, a number of space borne sensors such as RapidEye, SumbandilaSAT and WorldView2 have been developed for a wide range of applications including vegetation mapping (Adam et al. 2014). These remote sensing sensor are considered to significant improvement on multispectral satellites technology (Abel Ramoelo et al. 2015). The new generation multispectral sensors such as WorldView-2 and RapidEye offer more wavebands (up to 13 bands) and higher spatial resolution (1.5 m) than the traditional broadband satellite images such as SPOT and Landsat TM while reducing unnecessary loss of essential information as in hyperspectral image such as Hyperion (Stratoulias et al. 2015). The new generation sensors have been widely used over the last decade to estimate vegetation quality and quantity these include: nitrogen (Ramoelo et al. 2012b; Dube & Mutanga 2015a) biomass (Adam et al. 2014; Greaves et al. 2016; Jin et al. 2014; Lu 2005), leaf area index (Delegido et al. 2013; Haboudane 2004) and good results have been obtained.

In June 2015, Sentinel-2A satellite which is one of the very first two of the Sentinel-2 mission, was launched by the European Space Agency (ESA) (Navarro et al. 2017). Sentinel-2 has a wide swath and comprises of 13 spectral bands stretching from the VIS and NIR to SWIR at unique spatial resolutions allowing measurements of earth's reflected radiance (Navarro et al. 2017). One of the advantages of Sentinel-2 is that it offers a good temporal resolution with frequent revisit time of 5 days which allows continuous monitoring of vegetation condition (Sibanda et al. 2015).

Random forest (Breiman, 2001) has been increasingly used to improve accuracy in prediction accuracy in remote sensing applications (Grimm et al. 2008). Many researchers have used random forest as a classification algorithm (Mureriwa et al. 2016; Vincenzi et al. 2011; van Deventer et al. 2015; Mureriwa et al. 2016; Rodriguez-Galiano et al. 2015). However, less studies have used RF as a regression algorithm (Mureriwa et al. 2016; Vincenzi et al. 2011; Adam et al. 2014; Mutanga et al. 2012). Random forest has been found to be able in

handling complicated relationships between predictors caused by noise and large data volume (Vincenzi et al. 2011; Ismail et al. 2010).

The objective of this study was to evaluate the performance of Sentinel 2 imagery and random forest model to predict and map foliar P concentration in Telperion game reserve.

4.2 Materials and Methods

4.2.1 Study area

The study area is Telperion game reserve which forms part of the eZemvelo game reserve. The reserve is approximately 13000 hectares in size and situated at the border between Gauteng and Mpumalanga Provinces, Telperion falls within the far north of the Grassland Biome located in Mpumalanga Province of South Africa (25° 41' S, 28° 56' E) (Fig 4). Mucina and Rutherford (2006) classified the vegetation type as the Mixed Bushveld and Bankenveld Grassland vegetation type.

These vegetation types is considered to be endangered and small patches are protected in statutory conservation reserves and private reserves (Mucina and Rutherford, 2005) and the domininat grass species identified by (Coetzee 2012) are*: Themeda trianda, Eragrostis curvula, Sporobolus africanas ; Setaria sphacelata ,Hyparrhenia hirta* and *Cynodon dactylon*.

The study area experiences mean annual temperature 22 ° C and receives summer rainfall. The Wilge River which is a tributary of the Oliphant's River, flows northwards through the reserve, the river is fed by various wetlands and smaller streams.

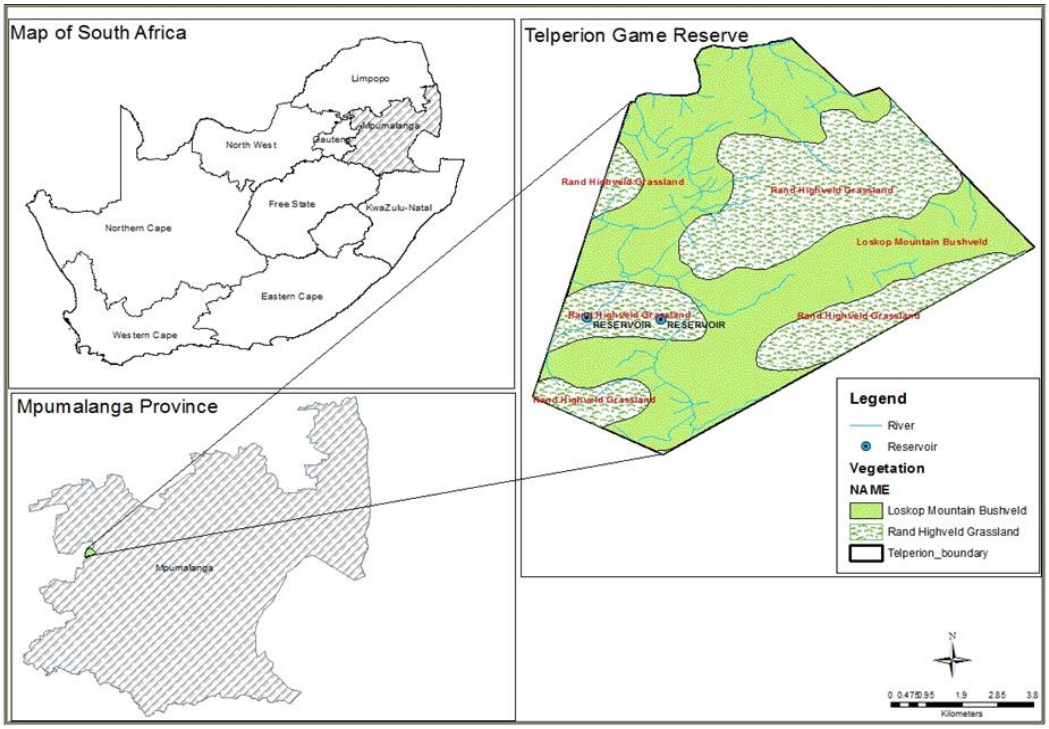


Figure 3: Study area map Study area map showing Loskop Mountain Bushveld and Rand Highveld vegetation

4.2.2 Remote sensing data acquisition and processing

Sentinel-2 innovative feature of having three red edge bands having the ability to distinguish differences in chlorophyll pigments (Vaglio Laurin et al. 2016). Higher chlorophyll content shows higher nitrogen and phosphorus content in plant tissue (Curran, 1989). Eventhough, sentinel has promising potential, there is insufficient use in monitoring vegetation mostly in grasslands.

Sentinel-2 image was downloaded from the Sentinel's Scientific Data Hub (https:scihub.copernicus.eu/) on the 20th of May 2016. The data was radiometrically and geometrically corrected using the Top-Of Atmosphere (TOA) products. Additional corrections were done included orthorectification as well as spatial registration on a global reference system. Sentinel 2 Toolbox was used to process the image to Level 2 data.

The image was atmospherically corrected using Sen2Cor plugin which is a prototype processor of Sentinel 2. Sen2Cor processor has been designed to execute functions such as atmospheric, terrain and cirrus correction as well as a scene classification (Anon 2017.). The

image was re-sampled with a constant Ground Sampling Distance (GSD) of 10m, 20m and 60m according to the resolution of different spectral bands (Table 1).

Field data and GPS information were used to produce a point map. The point map was overlain on Sentinel-2 image in order to produce region -of- interest (ROI) map using the GPS points (n=82). An 8 by 8 pixels window was then used to gather vegetation image spectra from each band (n=10) using ArcGIS 10.2 software. The reason for using 8 X 8 pixel windows was to eliminate pixels located outside the 20 m by 20m plot (Cho et al., 2007). The spectral data set was inclusive of only pixels which were entirely within the ROIs. Pixels which were partially inside the ROI were not considered. For each vegetation plot, spectra were collected and averaged.

The data set was partitioned randomly into 70% and 30% for training data and a test independent data respectively.

Table 5 show Sentinel-2 MSI spatial and spectral resolutions

Acronym	Band	Central Wave-length (nm)	Spectral Width	Spatial resolution
B1	Violet	443	20	60
B2	Blue	490	65	10
B3	Green	560	35	10
B4	Red	665	30	10
B5	Red-edge 1	705	15	20
B6	Red-edge 2	740	15	20
B7	Red-edge 3	783	20	20
B8	NIR	842	115	10
B8a	NIR narrow	865	20	20
B9	NIR	945	20	60
B10	NIR	1380	30	60
B11	SWIR1	1610	90	20
B12	SWIR2	2190	180	20

Source: ESA 2016

4.2.3 Field data collection

Sampling was done between the 21st and 23rd of May. This period is usually characterised by moderate to high rainfall and high biomass productivity, but in the year 2016, the country was experiencing severe drought which was nationally declared as a natural disaster. As a result, the grass was drier than expected and was already starting to senesce. Foliar nutrient quality is significant for herbivores throughout the year (Prins & Beekman 1989) . However, during the wet season, the grass is much greener and more nutritious whereas nutrient levels decreases in the dry season (Owen-smith & Novellie 1982). Mutanga (2004) emphasized that the nutrient levels become more significant in determining distribution of wildlife during the dry season.

Sampling technique used was random in this study, this was done across different grass communities. A total of 82 random samples were obtained. At each sample point, a 20 m X 20 m vegetation plot was generated. Grass representative samples from each plot (20 m X 20 m) were then randomly clipped. A Global Positioning System (GPS) was used to record the coordinates of the centre vegetation plot. The grass samples were kept and labelled in a paper for the chemical lab analysis.

4.2.4 The calibration and validation dataset

The data was randomly partitioned into 70% and 30% for training dataset (n = 57) and independent test dataset (n = 25) respectively. The random forest regression model was trained using the training data while the performance of the regression model was evaluated using the test data.

4.2.5 Chemical analysis

Chemical analysis of the field collected samples was done at the KwaZulu Natal Department of Agriculture and Rural Development. Dry combustion was done based on the Dumas (1831) method for phosphorus and sodium analysis. To extract the elements the Hunter's dry method was adopted (Hunter 1984). The samples were dried in oven at 70˚C for 48 hours. During chemical analysis, reagent blanks were used

4.2.6 Data Analysis

To estimate the grass P and Na concentration, Simple Ratio (SR) vegetation index and Normalised Vegetation Index (NDVI) were used. To explore the ability of all Sentinel-2 bands in estimating P and Na concentration, NDVI and SR were calculated from all the possible bands combination of Sentinel-2 bands. NDVI and SR were selected due to the fact that they have been used intensively in predicting vegetation biophysical and biochemical characteristics. The calculation of NDVI and SR were based on the below equations (1 and 2).

$$NDVI = \frac{R_{(i,n)} - R_{(j,n)}}{R_{(i,n)} + R_{(j,n)}} \quad (1)$$

$$SR = \frac{R_{(i,n)}}{R_{(j,n)}} \quad (2)$$

Where $R_{(i,\ n)}$ and $R_{(j,n)}$ are the reflectance at any two bands from the bands of the image spectra (n=10) (Mutanga et al. 2015).

The discrete 10 bands allowed a calculation of 100 NDVIs and 100 SRs. To simplify the analysis process, minimum number of NDVIs and SRs that yielded high prediction were identified based on their relationship with P and Na concentration as represented by the correlation coefficient (*r*). The NDVIs and RSs were ranked based on correlation coefficient and the top 20 indices for NDVI and SR were selected as input variables for further analysis using random forest regression to estimate vegetation features

4.2.7 Random forest regression

Random Forest was developed by Breiman (2001) to enhance the classification and regression tress (CART) method through the combination of large sets of decision trees. In this study Random Forest (RF) algorithm (Breiman 2000) which is a multivariate and non-parametric regression applied in the "randomForest" found in R software. It was used to estimate the amount of variations in grass phosphorus and sodium concentrations that can be derived from the NDVIs and SRs calculated. The R^2 values calculated from the NDVIs and SRs were noted and RF produced R2 and least RMSE values. In RF, multiple regression trees

(*ntree*) are produced through the generation of multiple bootstrap samples (Adam et al. 2014) from the original training data set with replacements (Adam et al. 2014). Regression trees are permitted to grow up to the maximum size without being pruned (Adam et al. 2014). In order to decide the top split at each node of tree, each tree is grown with random subset of predictors (*mtry*) (Grimm et al. 2008). The overall prediction accuracy is achieved by averaging the results from each aggregation. (Grimm et al. 2008). The larger the number of trees, the lower the biasness and variance is achieved by the RF model (Grimm et al. 2008).

Exclusion of about 33% of data from the sample is done during bootstrap sampling in RF model , while the rest of the data is replicated to achieve the full data size which is referred to as the 'in bag' data (Adam et al. 2014) while the data excluded is termed 'out-of-bag' (OOB) data (Adam et al. 2014; Breiman 1996). The OOB error which is the prediction variations derived from the OOB data and the 'in bag 'data is calculated for each tree in the RF ensemble (Pamer et al. 2007). The OOB estimate of error reflects reliable prediction accuracy as OOB data is not used in building or pruning any of the regression trees. OOB estimate of error has been regarded as a type of cross validation since reliable estimates of error are achieved by it. (Ismail & Mutanga 2010; Breiman 2001,). However, other researchers have argued on the insignificance of having independent validation data (Lawrence et al. 2006). OBB data makes it possible to evaluate the significance of each variable for prediction by calculating increase in error of prediction if OBB data of that variable were calculated (Prasad et al. 2006).

In RF, there is a need to optimise two parameters (i) *ntree* (default value 500) these are the number of trees grown from the bootstrap sample of observations and (ii) *mtry*, the number of different predictor variables (VIs) tested on each node (Adam et al. 2014). The *ntree* values were tested at 1000 intervals between 500 and 9500 (Prasad et al 2006), *mtry* was tested at a single interval from 1 to 25. Throughout the analysis, the node size was kept as the default size (Vincenzi et al. 2011). All analysis were performed using the randomForest library developed for the R statistical package.

Validation of the performance of the RF algorithm was done through data partitioning (training (70%) and test (30%)). The test samples respectively. Training was used to develop prediction model for phosphorus and sodium concentration respectively and to validate the predictive performances of both models, the test data was used (Ismail & Mutanga 2010;

Lawrence et al. 2006). A relationship between the measured and predicted phosphorus and sodium concentration values were fitted respectively and calculation of the R^2 and RMSEC was done (Adam et al. 2014).

4.2.8 Selecting the important band indices

The modelling process was simplified by identifying the number of NDVIs and SRs that were the highest predictors for both phosphorus and sodium concentration. NDVIs and SRs were firstly ranked in accordance to their relation to phosphorus and sodium concentrations as shown by the (*r*) (Mutanga & Skidmore 2004). After ranking, the top 20 indices for phosphorus and 25 for sodium for both NDVI and SRs data sets that produced the highest *r* values were chosen for further processing in the RF using the variable selection process.

The top 20 variables of importance phosphorus and 25 sodium concentrations were ranked by the RF algorithm from the NDVIs and SRs data, the mean decrease in accuracy determined from the OOB sample was used. This ranking was used to identify the way to eliminate the least important NDVIs and SRs data sets using the backward variable selection method (Ismail & Mutanga 2010). The backward variable selection model uses the whole data set first and then discards the least important variables based on RF ranking. To optimise the model at each iteration (n=20 phosphorus and n=25 sodium. In order to model phosphorus and sodium concentrations, the smallest set of vegetation indices with low RMSE were chosen for. The backward variable selection method was assessed through the use of a 10 fold cross validation (Mutanga et al., 2012; Kohavi and John 1997).

Using EnMAP- Box software, the best precise RF regression model was implemented to map the study area distribution of both phosphorus and sodium. EnMAP-Box is a freely available licence-free processing environment. RF regression and classification can be applied in ENMAP to analyse remote sensing imagery (Waske et al. 2012).

4.3 Results

4.3.1 Descriptive statistics

To determine the connection between the vegetation indices, phosphorus and sodium concentrations, data statistics were calculated as shown in Table 7 below. The variation of

phosphorus concentration was between 0.01 and 0.2 while sodium concentration varied between 20.2 and 774.

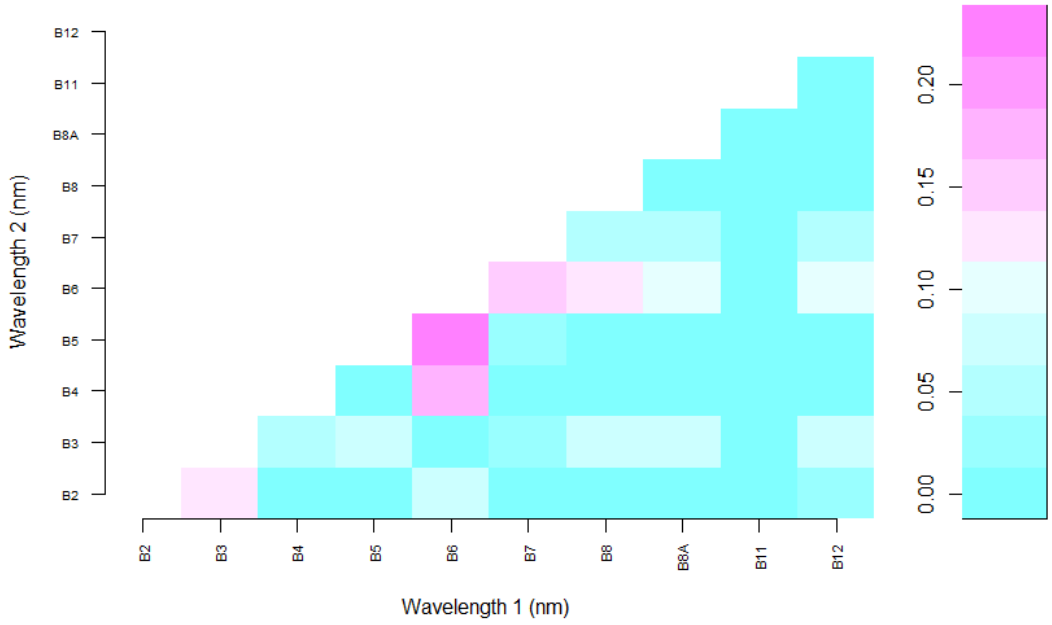
Table 6 Descriptive statistics of the measures phosphorus and sodium concentrations

Biochemical	Minimum	Median	Mean	Maximum	Standard deviation	Coefficient of variation
Phosphorus	0.01	0.05	0.053	0.19	0.0276	51%
Sodium	20.20	62.80	87.07	774.10	120.839	138%

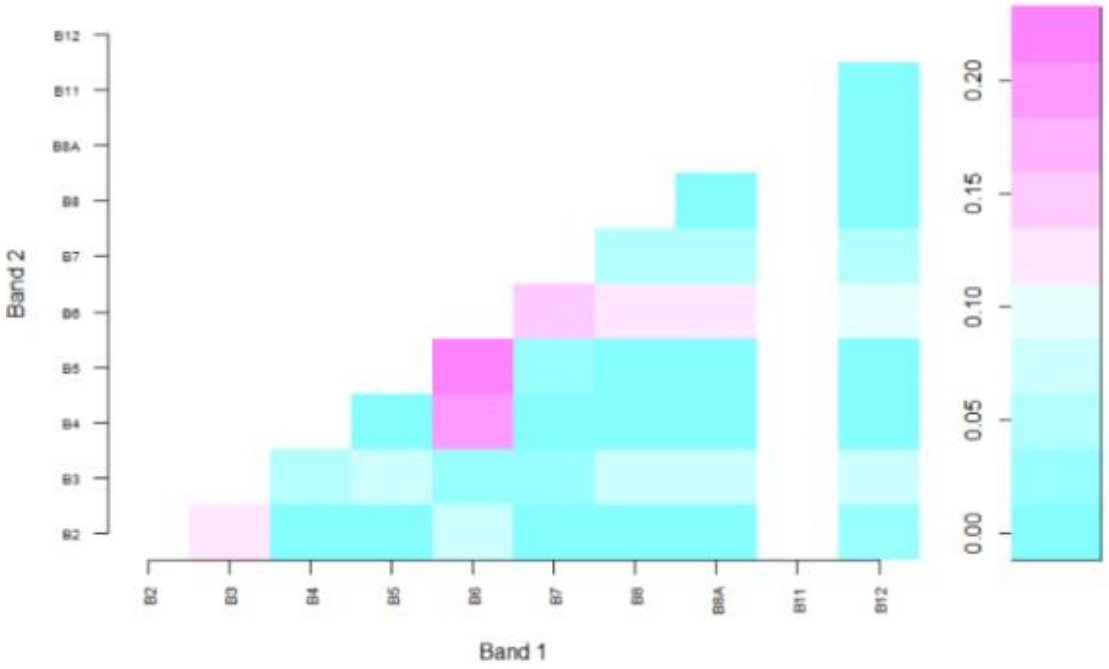
4.3.2 Relationship between the vegetation indices and the phosphorus and sodium concentrations

The correlation coefficient (r) between the NDVI and SR datasets demonstrated high dissimilarity in the relationship between the vegetation indices and concentration levels of phosphorus and sodium. The *r* values ranged between (0.0) and (0.20) for the NDVI phosphorus; (0.0) and (0.20) SR phosphorus; (0.0) and (0.20) NDVI sodium and (0.0) and (0.20) SR sodium datasets respectively. For the NDVI phosphorus dataset, the indices calculated using the Red, Red–edge 1 and Red edge 2 range produced the strongest correlations (0.15) and (0.20) (Fig 5) , SR phosphorus data set, the indices calculated using the Red, Red –edge 1 and Red edge 2 range produced the strongest correlations (0.13) and (0.2) for the sodium NDVI dataset , the indices calculated using the NIR, NIR narrow and SWIR2 range produced the strongest correlations (0.12) and (0.20) and in SR dataset indices calculated using the NIR, NIR narrow and SWIR2 also gave strong correlation values (0.11 and 0.25).

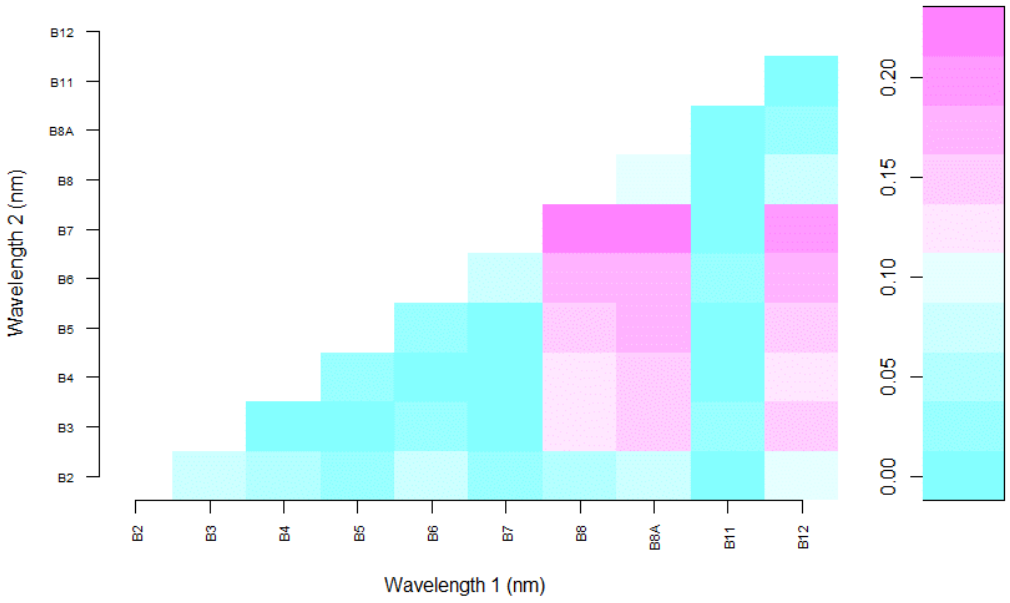
(a)


Wavelength 2 (nm)
B12
B11
B8A
B8
B7
B6
B5
B4
B3
B2
B2
B3
B4
B5
B6
B7
B8
B8A
B11
B12
Wavelength 1 (nm)
0.00
0.05
0.10
0.15
0.20

(b)


Band 2
B12
B11
B8A
B8
B7
B6
B5
B4
B3
B2
B2
B3
B4
B5
B6
B7
B8
B8A
B11
B12
Band 1
0.00
0.05
0.10
0.15
0.20

(c)



(d)

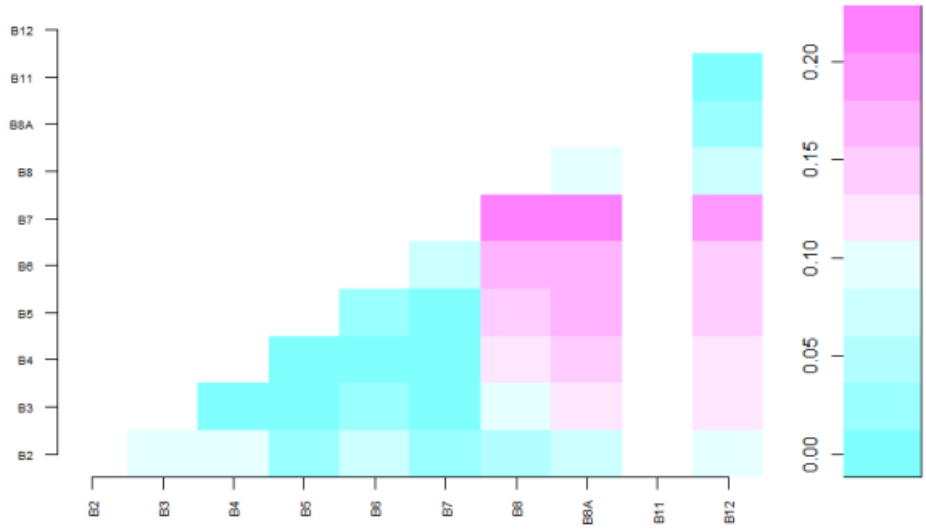


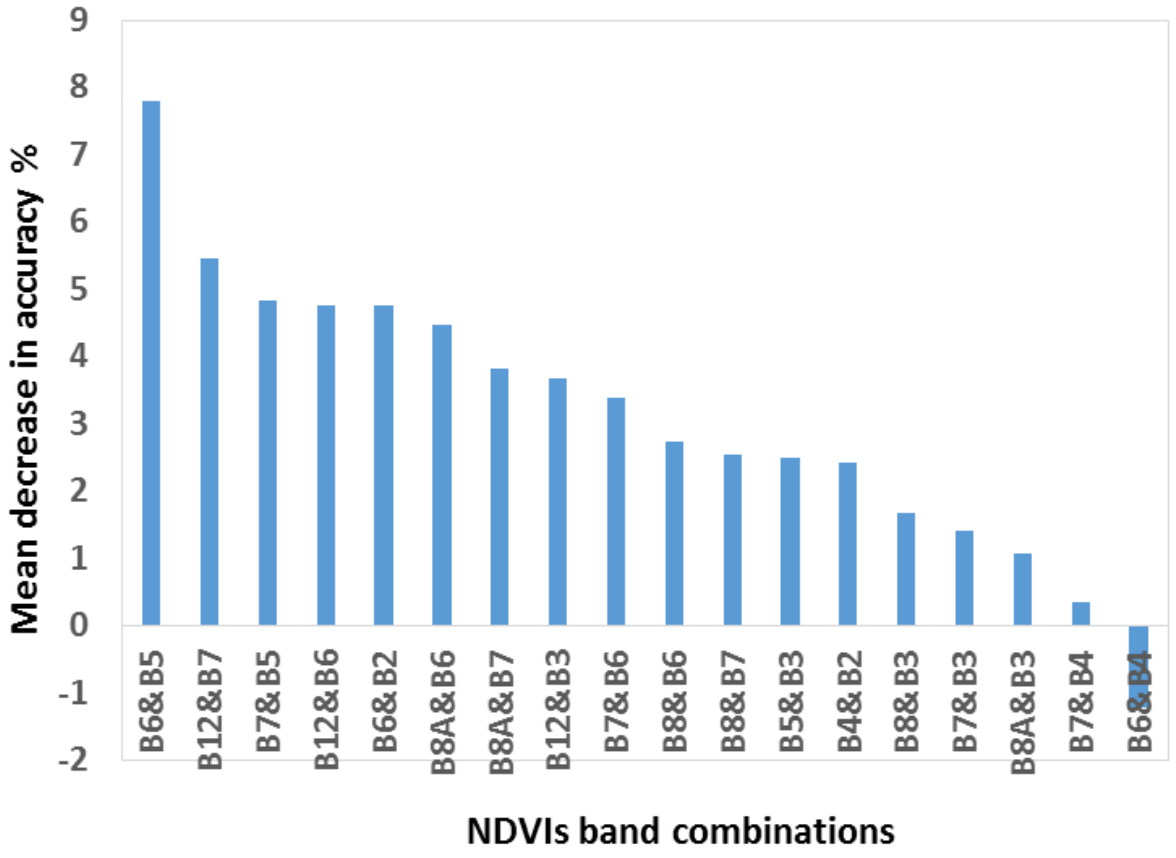
Figure 4: Contour plots representing the correlation (r) relationship between phosphorus NDVI (a) and SR (b) as well as the relationship between sodium NDVI (c) and SR (d) obtained from all two possible band combinations

4.3.3 Selecting important predictor variable

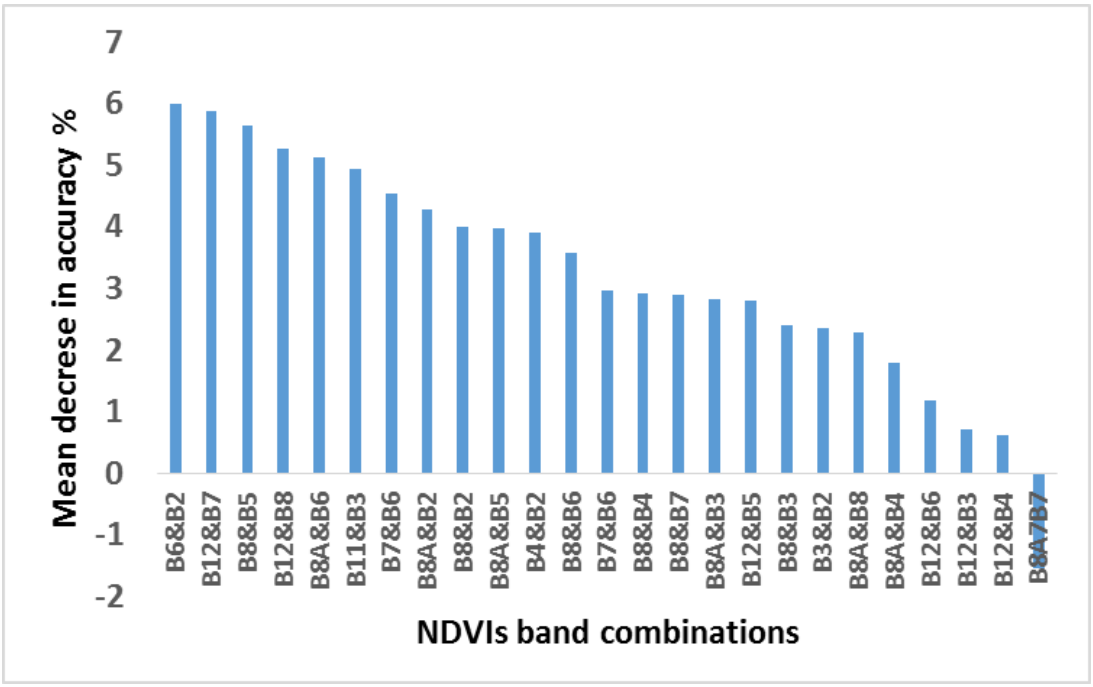
The RF model demonstrated the ability to measure and rank the importance of each top 20 NDVIs and 25 SRs in estimating P and Na concentration.

Fig 6 (a-b) shows the variable importance in estimating phosphorus and Fig (c-d) shows variable of importance measured based on the OOB error increase which demonstrates the weakness of prediction model when each variable is calculated. Less NDVI predictors contributed notably towards predicting phosphorus and sodium e.g. the red edge band combination

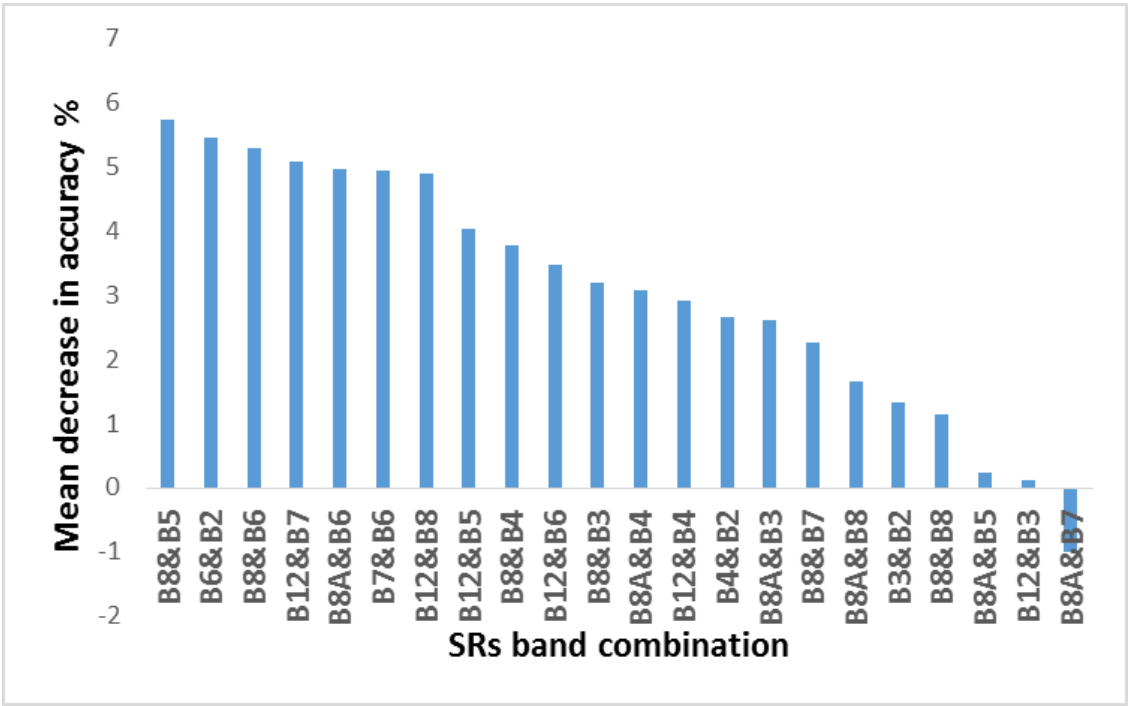
(a)



(b)


Mean decrese in accuracy %
7
6
5
4
3
2
1
0
-1
-2
B6&B2
B12&B7
B8&B5
B12&B8
B8A&B6
B11&B3
B7&B6
B8A&B2
B8&B2
B8A&B5
B4&B2
B8&B6
B7&B6
B8&B4
B8&B7
B8A&B3
B12&B5
B8&B3
B3&B2
B8A&B8
B8A&B4
B12&B6
B12&B3
B12&B4
B8A7B7
NDVIs band combinations

(c)


Mean decrease in accuracy %
7
6
5
4
3
2
1
0
-1
-2
B8& B5
B6& B2
B8& B6
B12& B7
B8A& B6
B7& B6
B12& B8
B12& B5
B8& B4
B12& B6
B8& B3
B8A& B4
B12& B4
B4& B2
B8A& B3
B8& B7
B8A& B8
B3& B2
B8& B8
B8A& B5
B12& B3
B8A& B7
SRs band combination

(d)

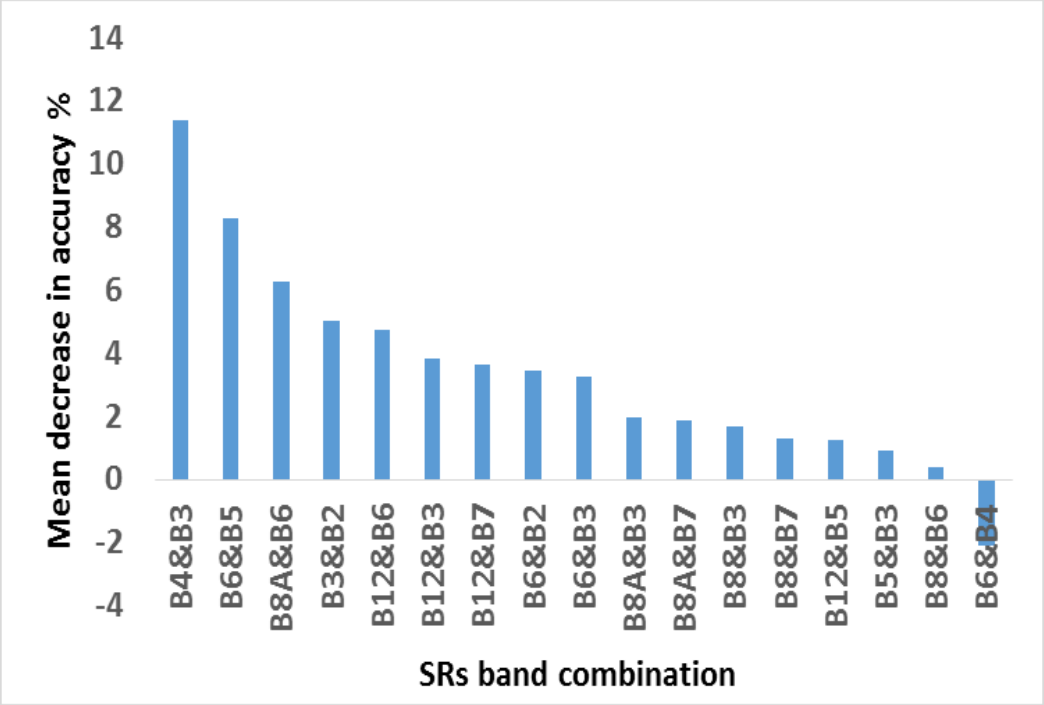
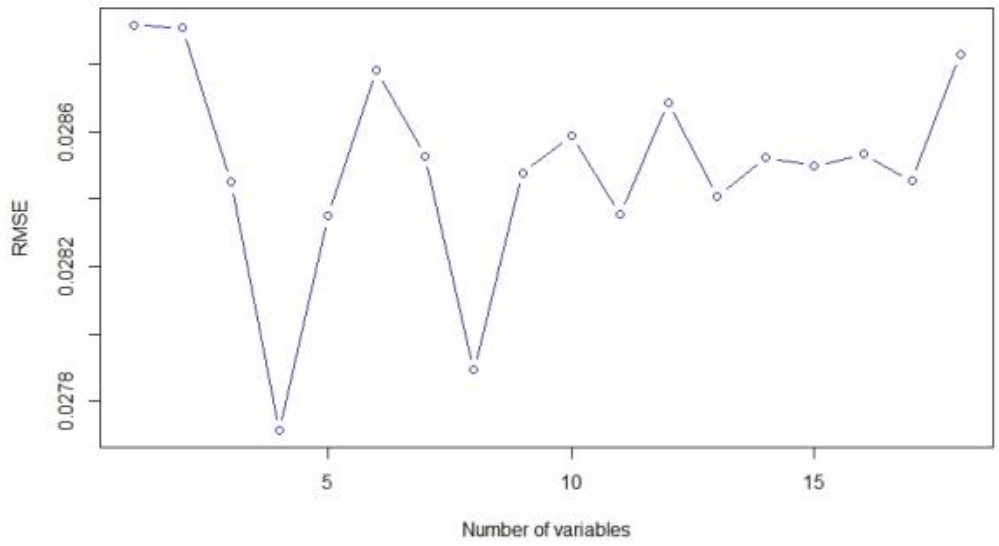


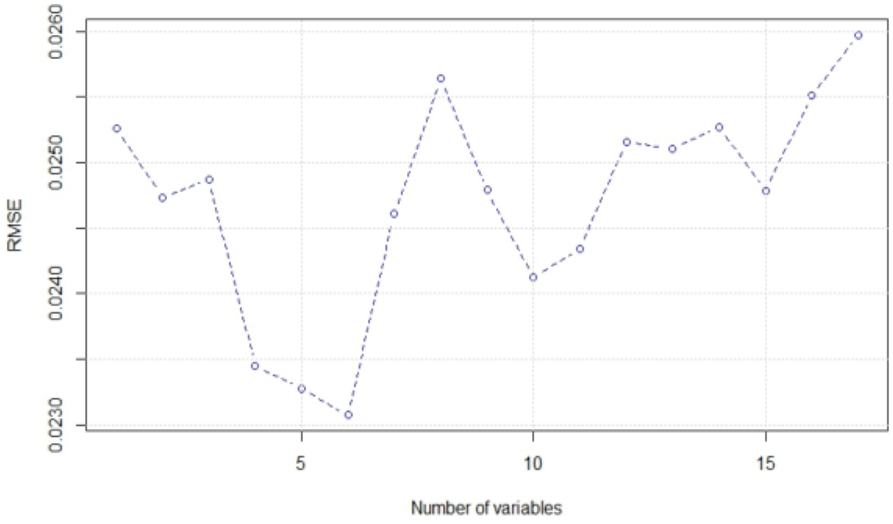
Figure 5: Measurement of variable of importance (a)-(b) NDVIs and SRs phosphorus (c)-(d) NDVIs and SRs sodium. Highest decrease accuracy represents the most important variables

Few NDVI and SR contributed noticeably towards predicting phosphorus and sodium. Using NDVIs, Red- edge 1&2; red edge 1&3,; Red -edge 2 &SWIR2 were the best predictors of phosphorus while using SRs the red &green ; Red- edge 2 & Red-edge 1; NIR narrow and Red -edge 2 were the best predictors for phosphorus. 19 important variables for the prediction of sodium using NDVIs were calculated by the model, the best predictors were found in Red edge 2 & Blue; SWIR2 & Red- edge 3; NIR &Red –edge 1 ; SWIR2&NIR ; NIR narrow &Red-edge 2. The model calculated 17 important variables for the prediction of sodium using SRs were calculated in NIR&Red-edge 1; Red-edge 2 &Blue; NIR&Red –edge 2

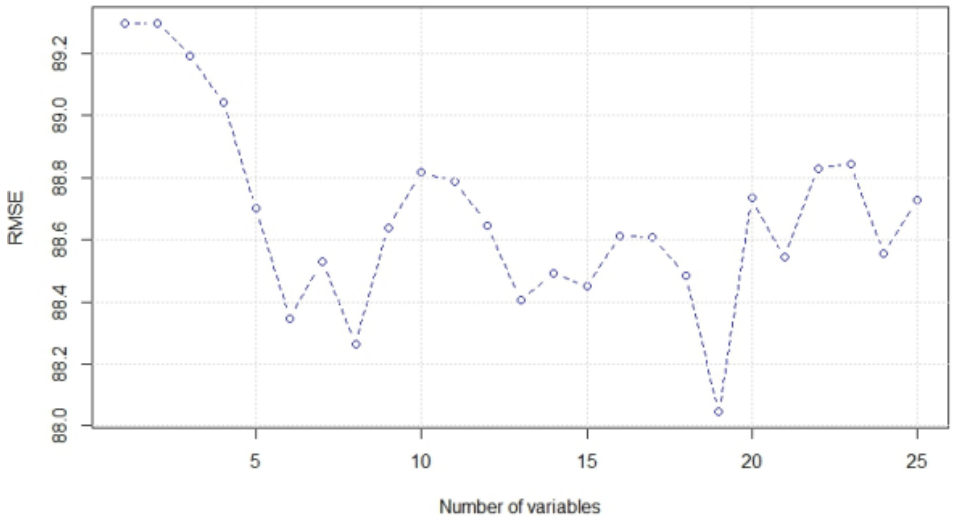
(a)


RMSE
0.0278
0.0282
0.0286
5
10
15
Number of variables

(b)


RMSE
0.0230
0.0240
0.0250
0.0260
5
10
15
Number of variables

(c)


RMSE
88.0
88.2
88.4
88.6
88.8
89.0
89.2
5
10
15
20
25
Number of variables

(d)

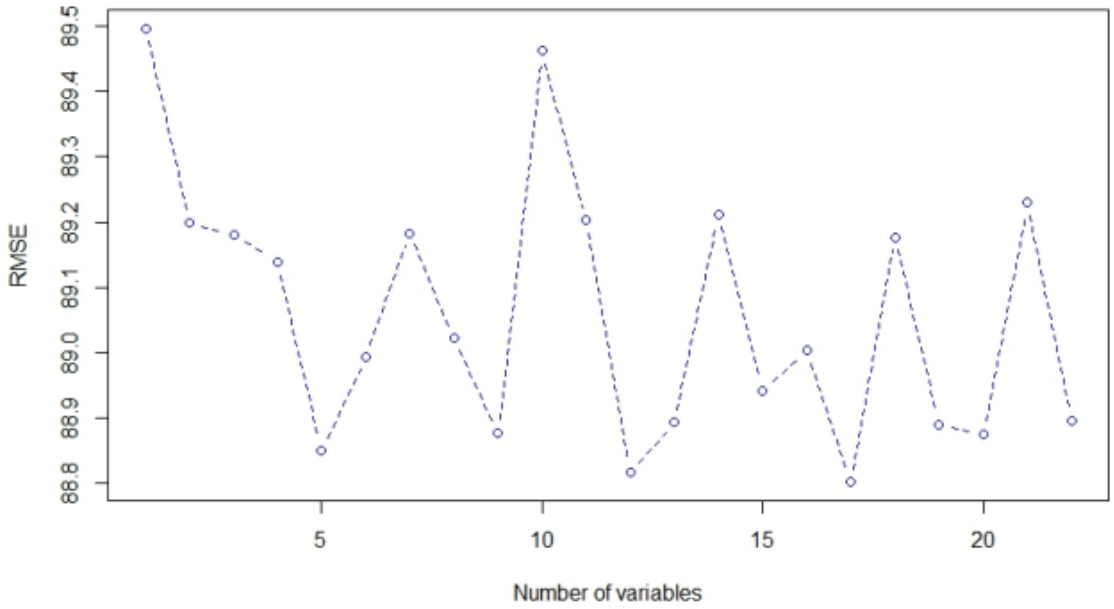
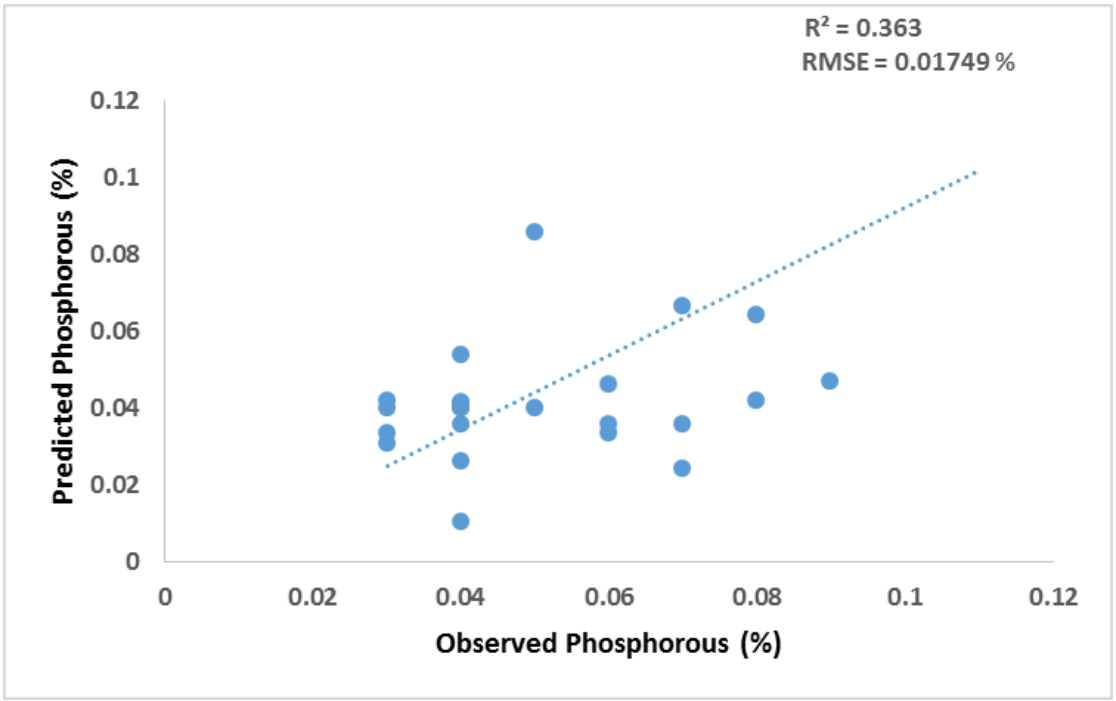
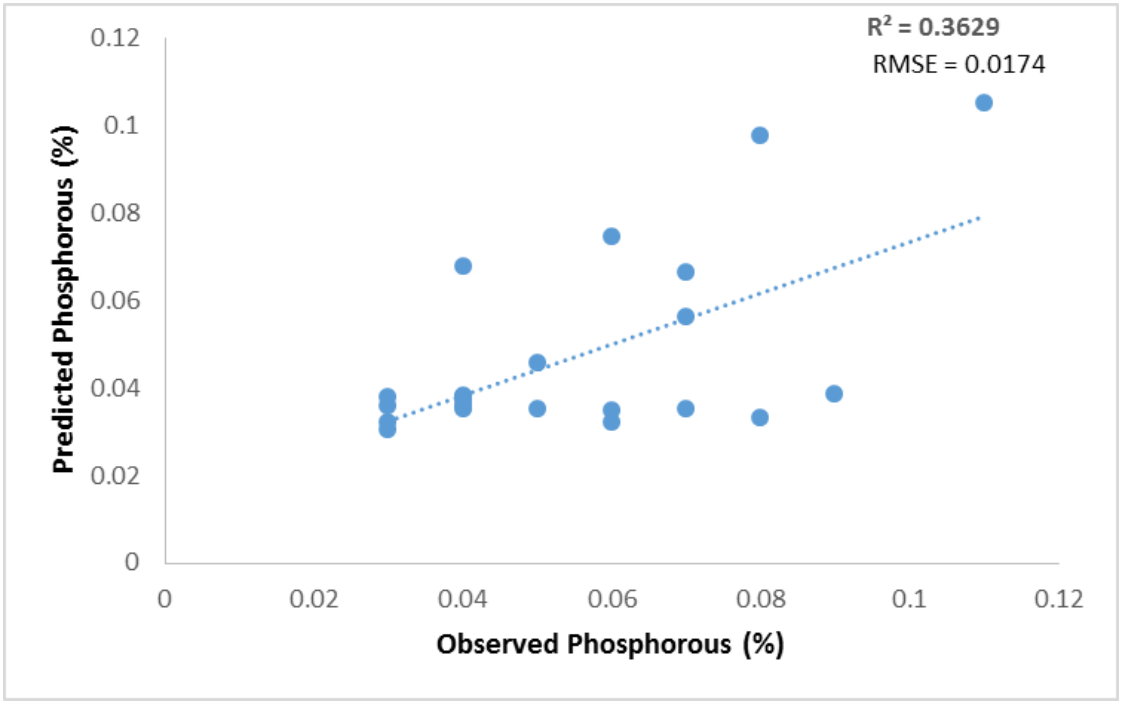


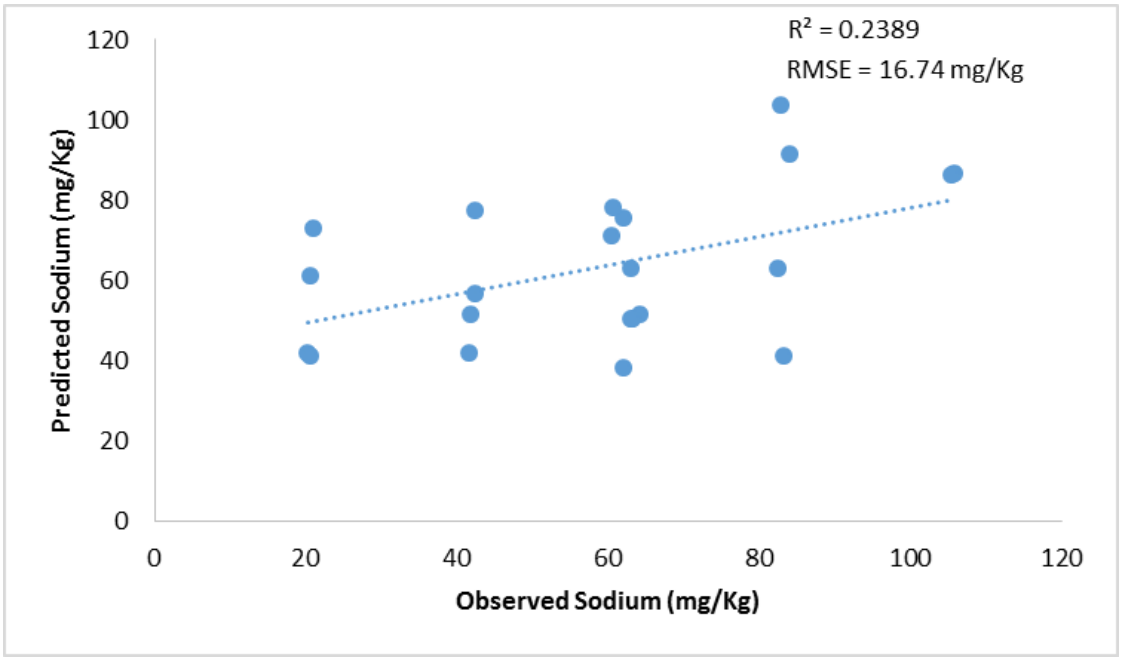
Figure 6: The number of variables selected for prediction using phosphorus (a)(b) NDVIs and SRs (4 and 6) and Sodium NDVIs and SRs (c)(d) (6 and 7)



(b)


R² = 0.3629
RMSE = 0.0174
Predicted Phosphorous (%)
Observed Phosphorous (%)
0
0.02
0.04
0.06
0.08
0.1
0.12

(c)


R² = 0.2389
RMSE = 16.74 mg/Kg
Predicted Sodium (mg/Kg)
Observed Sodium (mg/Kg)
0
20
40
60
80
100
120

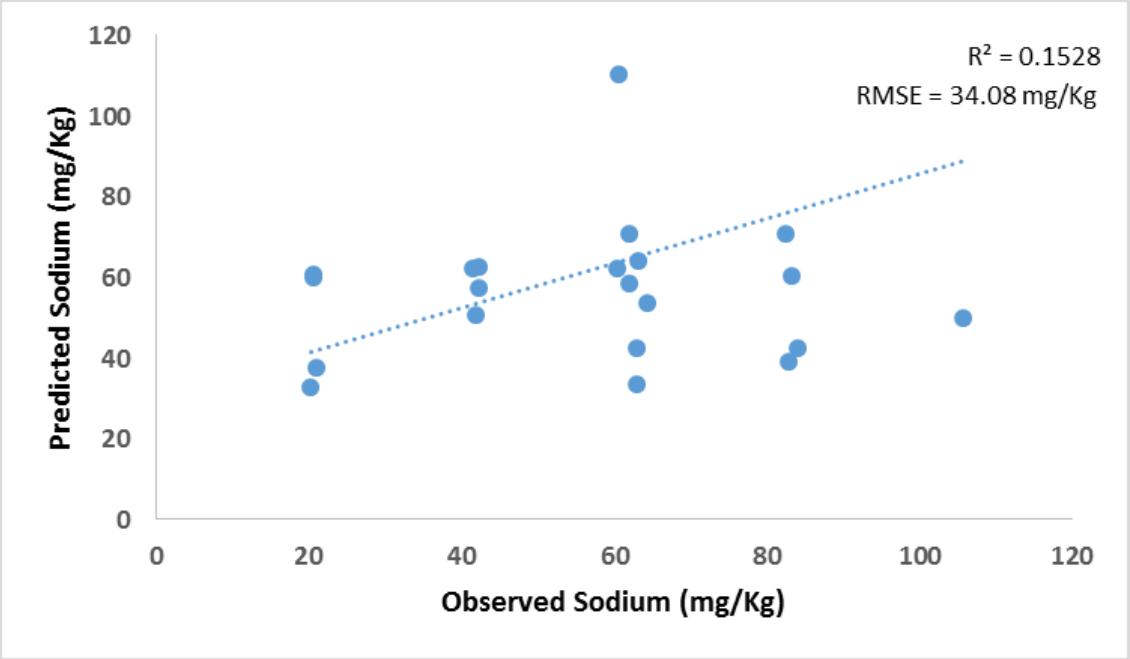


Figure 7: The actual and predicted phosphorus and sodium concentration showing one-one relationship developed using NDVIs and SRs best bands for prediction

4.3.4 Spatial distribution of foliar P and Na

RF model based on NDVIs and SR indices was selected as the best RF model as it produced the highest R2 values for phosphorus and as compared to the other models. This model was contained of (n=7) SR and (n=5) bands (6, 7, 8, 8A and 12) and all these variables were used to map foliar P and Na 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 Nature Reserve is depicted by (Fig 8).

(a) (b)

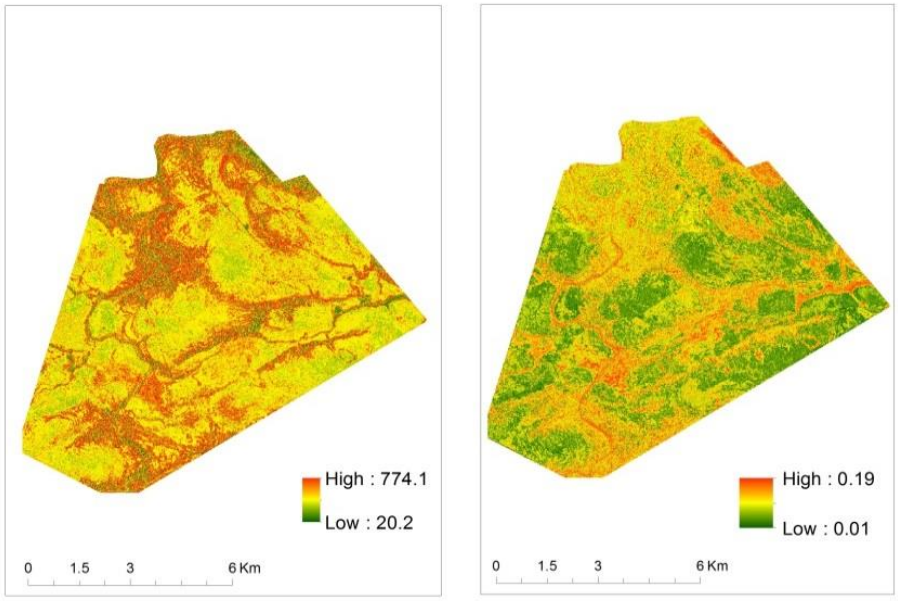


Figure 9: Spatial distribution maps of sodium (a) and phosphorus (b)

4.4 Discussion and conclusion

4.4.1 Prediction of P and Na

The research was done to determine the accuracy in prediction of foliar P and Na concentration when using Random Forest and to map foliar P and Na across the study area. The descriptive statistics of field data showed that the mean, min and max values of the grass phosphorus was 0.01, 0.05, 0.19 and sodium 87.07, 20.20 and 777 (Table 7). The coefficients of variance for phosphorus and sodium concentrations indicated that there is relatively high dispersion (50% and 137%) respectively as shown in Table 7. Mutanga and Kumar (2007) explained that variation in phosphorus and sodium can be caused by gradients of soil nutrient concentrations covering the rock formations. Different species yielded different foliar concentrations. Variations in sodium among different species can be due to differences in sodium root uptake and translocation to the shoots as well as soil sodium content (Mutanga et al. 2004).

Mutanga (2004) found high sodium foliar concentration variations (0% and 0.78%) in sites having high nutrient materials as a result of erosion and weathering of underlying dolerite and basalt intrusions. In order to estimate foliar biochemical, the red edge bands have been broadly used since it is a strong relationship to chlorophyll (Cho et al. 2012). In this study,

the RF model produced R^2 = 0.36 and RMSE = 0.017% for phosphorus using NDVI data set and R^2 = 0.362, RMSE = 0.0174% using SR data set (Fig 8). The SR and NDVI yielded similar results for phosphorus. Similar R^2 value (0.37) was achieved by Van Deventer et al. (2015) in their study to test the potential of models to predict foliar P between foliar P and leaf spectra within the red-edge region (680-760 nm). The achieved R^2 values demonstrated that there is a connection between carotenoids, chlorophyll and P concentrations during peak productive season. Furthermore, predictive models for leaf P showed low accuracy as compared to leaf N.

Cho et al (2012) investigated estimation of foliar P concentrations and significant sensitive bands were found to be located within the SWIR (Cho et al. 2012), a high grass P estimation accuracy R^2 = 0.80 and RMSE = 0.03 compared to using remote sensing variables only. This was higher than the achieved R^2 and RMSE values obtained in this study.

In their study using imaging spectroscopy and neural networks to map phosphorus in grass savanna, Mutanga and Kumar (2007) found that the best predictor of phosphorus distribution had a R^2 of 0.63 and a root mean square error (RMSE) of 0.07 (28%) of the mean observed phosphorus concentration based on the test data.

Although, there are very few studies that have been done to estimate leaf P, the estimation of leaf nitrogen has been widely done. However, Ramoelo et al. (2013) found R^2 =0.80 and RMSE of 0.03 for phosphorus in their study using hyperspectral data to estimate grass nitrogen and phosphorus using non–linear partial least squares regression. The higher prediction accuracy achieved by Ramoelo et al. (2013) using hyperspectral remote demonstrates the higher ability and performance of the many bands found in hyperspectral data in predicting foliar biochemical.

The RF model produced R^2 values = 0.23, RMSE = 16.74 mg/kg and R^2 = 0.15 and RMSE = 34.08 mg/kg for sodium using NDVI and SR data respectively. The model performed better using NDVI data as compared to SR data.

Mutanga (2004) found that sodium was best absorbed at the 589 nm. However, there is a lack of studies done estimating sodium leaf concentration using remote sensing. The obtained results could not be compared to other studies to assess the accuracy of the models used in this study in predicting sodium.

The lower prediction accuracy produced by the RF model for both phosphorus and sodium could be attributed to the dryness of the grass. The country experienced severe drought in the year that this study was commissioned, as a result, during the time of sampling the grass species were already dry. Kolkaky and Clark (1999) reported that the biochemical nutrient content and vegetation structure changes during dry season and these lead to spectral changes in vegetation. Spectral reflectance changes that occur as vegetation dries up include: change of red edge as well as increased significance of SWIR features associated with lignin and cellulose (Knox et al. 2011). The low results obtained are due to vegetation dryness. Further similar study to predict and map foliar P and Na can be done across different seasons to compare the differences in estimates obtained.

4.4.2 Relationship between vegetation indices (NDVIs & SRs) and prediction of P and Na concentrations

In this study, the calculation of NDVIs and SRs as well as the correlation of the indices to (P and Na) was done to improve the knowledge on the linkages between electromagnetic spectrum and biochemical prediction of both P and Na at the 10 Sentinel 2 bands as well as provided an opportunity to investigate data within the multispectral data. RF models were developed to model grass quality information in the NDVI and SR. The *r* values (0.0 and 0.20) demonstrate how different band combinations detect the grass variation in P and Na differently.

Using NDVI data, the RF model the most important variables for phosphorus were found to be located in the following regions: Red-edge2 & Red-edge 1; SWIR2 & Red-edge 3; Red-edge 3 & Red-edge 1; SWIR2 and Red-edge 2; Red-edge 2 & Blue and NIR narrow & Red edge 3. Using SR data, the most important variables for phosphorus were found to be located in the following regions: Red & Green; Red-edge 2 & Red-edge 1; NIR narrow & Red-edge 2; Green & Blue; SWIR2 &Red edge 2 and SWIR2 & Green.

In this study, using the NDVI data, the RF model the most important variables for sodium were found to be located in the following regions: Red-edge & Blue; SWIR2 & Red-edge 3; NIR & Red-edge 1; SWIR2 & NIR , NIR narrow &Red-edge 2; SWIR1 & Green; Red-edge 3 & Red-edge 2; NIR narrow &Blue and NIR &Blue. Using the SR data, the most important variables for sodium were found to be located in the following regions: NIR & Red-edge1; Red-edge 2 &Blue; NIR & Red edge 2; SWIR2 &NIR; SWIR2 & Red-edge 1. Mutanga et al. (2004) reported that significant variations between sodium classes (low and high).

The red edge position spectral features are linked to the chlorophyll content (Fiella & Penuelas 1994). There is a relationship between chlorophyll content and foliar nutrients. Based on this, one can expect foliar biochemical detection at the red edge position (Ali 2014; Skidmore et al. 2010).

Although, Sentinel-2 imagery, a new generation multispectral sensor, prediction of the foliar chemicals was low, the use of hyperspectral imagery can possibly give better estimates for grass P and Na due to the many narrow bands associated with hyperspectral.

4.4.3 Mapping of P and Na

The study showed a potential to estimate grass quality using new high spatial resolution remote sensing data (Sentinel-2) with red edge band. The spatial distribution maps of foliar P and Na demonstrated variation of both biochemicals across the grass communities within the study area as indicated in Fig (8). High concentrations of both P and Na is found to be along the drainage area and the wetland grass species may be linked with other environmental data such as soils and geology of the area which were not measured in this study. Ramoelo et al (2014). In their studies, Ramoelo et al (2013) reported that the underlying geology geological of an area contributes to the variations of grass nutrients. Drier grasses were reported to have lower photosynthetic activity as compare to fresh grasses due low water composition resulting in lower grass quality.

An integrated modelling approach that combines environmental variables and vegetation indices could potentially enhance the prediction of the foliar biochemical in the grasslands. This was successfully demonstrated in a study by Ramoelo et al (2012) where space borne multispectral remote sensing image was used to regionally map foliar and canopy nitrogen at high grass productivity period. In future studies, different regression models such as Stepwise Multiple Linear regression and SVM can be used to predict the foliar P and Na across the nature reserve. The results can be compared with the results achieved in this study to fully assess the usefulness of Sentinel-2 data in mapping vegetation quality.

CHAPTER 5

5 CONCLUSION

The objectives of the study were successfully reached. The different grass communities occurring across the study area were successfully classified and mapped. The classification results showed species diversity within the game reserve and variation in distribution.

This study showed the capability of Sentinel 2 in mapping leaf P and Na. Random Forest regression has demonstrated the effect of the red edge on predicting grass P and Na and mapping the variations of both biochemicals across the study area.

The individual classification accuracy depicted in Table 4 are considerably higher. The ANN produced individual class accuracy ranging from 64% to 94%. Classification accuracy obtained in this study for each grass species may be influenced by various environmental factors that may be linked to the spectral reflectance (Omer et al. 2015).

The RF regression model using NDVI and SR data for phosphorus prediction performed better compared to the sodium prediction. The regression algorithm identified the optimal number of vegetation indices which offered best predictive accuracy. However, the prediction for both P and Na were lower as compared to previous studies (Ramoelo et al. 2013; Mutanga & Skidmore 2004) . The lower predictions are because the study was done in the year 2016 where in the country experience severe drought. During the period of sampling the grasses were already dried up and senescing. The biochemical/ nutrient content of vegetation changes during dry seasons and this result in spectral variations (Kokaly &Clark 1999). The spectral changes that occur as vegetation dries up include change of red edge shoulder and increase of short wave infrared (SWIR) features associated with the cellulose and lignin of vegetation (Knox et al. 2011).

Future studies can explore the use of red edge bands in Sentinel-2 to predict grass phosphorus and sodium across all seasons. This can provide the assessment of the foliar biochemical as rangeland quality indicators

The study demonstrated the potential of Sentinel-2 data in predicting grass P and Na using. Furthermore, RF algorithm allowed for the identification and understanding of the Sentinel

red edge bands that are most sensitive to P and Na estimation using both NDVI and SR data sets. The study further showed that RF algorithm can be used as a regression method for multispectral remote sensing applications such as prediction of grass macro nutrients.

REFERENCES

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.

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

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

Adelabu, S., Mutanga, O. & Adam, E., 2014. Evaluating the impact of red-edge band from Rapideye image for classifying insect defoliation levels. *ISPRS Journal of Photogrammetry and Remote Sensing*, 95, pp.34–41.

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. et al., 2012. Optimizing spectral resolutions for the classification of C3 and C4 grass species, using wavelengths of known absorption features. *Journal of Applied Remote Sensing*, 6(1), pp.063560-1.

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.

Anon, Sen2Cor | STEP. Available at: http://step.esa.int/main/third-party-plugins-2/sen2cor/ [Accessed March 7, 2017b].

Asner, G.P., 1998. Biophysical and biochemical sources of variability in canopy reflectance. *Remote sensing of Environment*, 64(3), pp.234–253.

Atkinson, P.M. & Tatnall, A.R.L., 1997. Introduction Neural networks in remote sensing. *International Journal of Remote Sensing*, 18(4), pp.699–709.

Baldi, G., Guerschman, J.P. & Paruelo, J.M., 2006. Characterizing fragmentation in temperate South America grasslands. *Agriculture, Ecosystems & Environment*, 116(3–4), pp.197–208.

Bork, E.W. & Su, J.G., 2007. Integrating LIDAR data and multispectral imagery for enhanced classification of rangeland vegetation: A meta analysis. *Remote Sensing of Environment*, 111(1), pp.11–24.

Boschetti, M., Bocchi, S. & Brivio, P.A., 2007. Assessment of pasture production in the Italian Alps using spectrometric and remote sensing information. *Agriculture, Ecosystems & Environment*, 118(1–4), pp.267–272.

Carbutt, C. et al., 2011. The conservation status of temperate grasslands in southern Africa. *Grassroots*, 11(1), pp.17–23.

Carbutt, C. & Martindale, G., 2014. Temperate indigenous grassland gains in South Africa: Lessons being learned in a developing country. *Parks*, 20(1), pp.104–125.

Cho, M.A. et al., 2012. Mapping tree species composition in South African savannas using an integrated airborne spectral and LiDAR system. *Remote Sensing of Environment*, 125, pp.214–226.

Cho, M.A., Malahlela, O. & Ramoelo, A., 2015. Assessing the utility WorldView-2 imagery for tree species mapping in South African subtropical humid forest and the conservation implications: Dukuduku forest patch as case study. *International Journal of Applied Earth Observation and Geoinformation*, 38, pp.349–357.

Congalton, R.G. & Green, K., 2008. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, Second Edition*, CRC Press.

Cingolani, A.M. et al., 2004. Mapping vegetation in heterogenous mountain rangeland using Landsat data: an alternative method to define and classify land-cover units. *Remote Sensing of Environment*.92(1), pp.84-97

Curran, P.J., 1989. Remote sensing of foliar chemistry. *Remote sensing of Environment*, 30(3), pp.271–278.

Curran, P.J., Dungan, J.L. & Peterson, D.L., 2001. Estimating the foliar biochemical concentration of leaves with reflectance spectrometry: testing the Kokaly and Clark methodologies. *Remote Sensing of Environment*, 76(3), pp.349–359.

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.

Dlamini, M et al. 2016. Detecting and mapping the impacts of subsistence farming on flood plain wetlands using WorldView-2: A comparison of support vector machine and maximum likelihood classifier algorithm. *The centenary conference of society south african geographers*.

Dube, T. et al., 2014. Intra-and-Inter Species Biomass Prediction in a Plantation Forest: Testing the Utility of High Spatial Resolution Spaceborne Multispectral RapidEye

Sensor and Advanced Machine Learning Algorithms. *Sensors*, 14(8), pp.15348–15370.

Dube, T. & Mutanga, O., 2015a. Evaluating the utility of the medium-spatial resolution Landsat 8 multispectral sensor in quantifying aboveground biomass in uMgeni catchment, South Africa. *ISPRS Journal of Photogrammetry and Remote Sensing*, 101, pp.36–46.

Dube, T. & Mutanga, O., 2015b. Investigating the robustness of the new Landsat-8 Operational Land Imager derived texture metrics in estimating plantation forest aboveground biomass in resource constrained areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 108, pp.12–32.

Duru, M., 2000. Growth and Senescence of the Successive Grass Leaves on a Tiller. Ontogenic Development and Effect of Temperature. *Annals of Botany*, 85(5), pp.635–643.

Ferwerda, J.G. & Skidmore, A.K., 2007. Can nutrient status of four woody plant species be predicted using field spectrometry? *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(6), pp.406–414.

Greaves, H.E. et al., 2016. High-resolution mapping of aboveground shrub biomass in Arctic tundra using airborne lidar and imagery. *Remote Sensing of Environment*, 184, pp.361–373.

Haboudane, D., 2004. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment*, 90(3), pp.337–352.

Horler, D.N.H., DOCKRAY, M. & Barber, J., 1983. The red edge of plant leaf reflectance. *International Journal of Remote Sensing*, 4(2), pp.273–288.

Ingram, J.C., Dawson, T.P. & Whittaker, R.J., 2005. Mapping tropical forest structure in southeastern Madagascar using remote sensing and artificial neural networks. *Remote Sensing of Environment*, 94(4), pp.491–507.

Jin, Y. et al., 2014. Remote Sensing-Based Biomass Estimation and Its Spatio-Temporal Variations in Temperate Grassland, Northern China. *Remote Sensing*, 6(2), pp.1496–1513.

Kaszta, Ż. et al., 2016a. Bulk feeder or selective grazer: African buffalo space use patterns based on fine-scale remotely sensed data on forage quality and quantity. *Ecological Modelling*, 323, pp.115–122.

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., 2012. Remote sensing of forage nutrients: Combining ecological and spectral absorption feature data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 72, pp.27–35.

Liu, M. et al., 2013. Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and Chinese vinegar. *Sensors and Actuators B: Chemical*, 177, pp.970–980.

Lu, D., 2005. Aboveground biomass estimation using Landsat TM data in the Brazilian Amazon. *International Journal of Remote Sensing*, 26(12), pp.2509–2525.

Mahajan, G.R. et al., 2014. Using hyperspectral remote sensing techniques to monitor nitrogen, phosphorus, sulphur and potassium in wheat (Triticum aestivum L.). *Precision Agriculture*, 15(5), pp.499–522.

Mansour, K. et al., 2012. Discriminating indicator grass species for rangeland degradation assessment using hyperspectral data resampled to AISA Eagle resolution. *ISPRS Journal of Photogrammetry and Remote Sensing*, 70, pp.56–65.

Mansour, K. et al., 2016. Multispectral remote sensing for mapping grassland degradation using the key indicators of grass species and edaphic factors. *Geocarto International*, 31(5), pp.477–491.

Mansour, K. et al., 2016. Multispectral remote sensing for mapping grassland degradation using the key indicators of grass species and edaphic factors. *Geocarto International*, 31(5), pp.477–491.

Mashao, D.J., 2003. Comparing SVM and GMM on parametric feature-sets. In: Proceedings of the 14th Annual Symposium of the Pattern Recognition Association of South Africa. Citeseer.

Marshall, M. & Thenkabail, P., 2015. Advantage of hyperspectral EO-1 Hyperion over multispectral IKONOS, GeoEye-1, WorldView-2, Landsat ETM+, and MODIS vegetation indices in crop biomass estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 108, pp.205–218.

Masocha, M. & Skidmore, A.K., 2011. Integrating conventional classifiers with a GIS expert system to increase the accuracy of invasive species mapping. *International Journal of Applied Earth Observation and Geoinformation*, 13(3), pp.487–494.

McNaughton, S.J., 1988. Mineral nutrition and spatial concentrations of African ungulates. *Nature*, 334(6180), pp.343–345.

McNaughton, S.J., 1983. Serengeti Grassland Ecology: The Role of Composite Environmental Factors and Contingency in Community Organization. *Ecological Monographs*, 53(3), pp.291–320.

Mureriwa, N. et al., 2016. Examining the Spectral Separability of Prosopis glandulosa from Co-Existent Species Using Field Spectral Measurement and Guided Regularized Random Forest. *Remote Sensing*, 8(2), p.144.

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.

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. & Skidmore, A., 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., 2004. Narrow band vegetation indices overcome the saturation problem in biomass estimation. *International Journal of Remote Sensing*, 25(19), pp.3999–4014.

Mutanga, O., Skidmore, A.. & Prins, H.H., 2004. 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., 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. & 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. et al., 2005. Estimating tropical pasture quality at canopy level using band depth analysis with continuum removal in the visible domain. *International Journal of Remote Sensing*, 26(6), pp.1093–1108.

Mutanga, O. & Rugege, D., 2006. Integrating remote sensing and spatial statistics to model herbaceous biomass distribution in a tropical savanna. *International Journal of Remote Sensing*, 27(16), pp.3499–3514.

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(21), pp.4897–4911.

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

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

Mutanga, O. et al., 2015. Evaluating the robustness of models developed from field spectral data in predicting African grass foliar nitrogen concentration using WorldView-2 image as an independent test dataset. *International Journal of Applied Earth Observation and Geoinformation*, 34, pp.178–187.

Mutanga, O., Dube, T. & Ahmed, F., 2016. Progress in remote sensing: vegetation monitoring in South Africa. *South African Geographical Journal*, 98 (3), pp.461–471.

Navarro, G. et al., 2017. Evaluation of forest fire on Madeira Island using Sentinel-2A MSI imagery. *International Journal of Applied Earth Observation and Geoinformation*, 58, pp.97–106.

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.083527.

Omer, G. et al., 2015. Performance of support vector machines and artificial 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.

Omer, G. et al., 2016. Empirical Prediction of Leaf Area Index (LAI) of Endangered Tree Species in Intact and Fragmented Indigenous Forests Ecosystems Using WorldView-2 Data and Two Robust Machine Learning Algorithms. *Remote Sensing*, 8(4), p.324.

Prins, H.H.T. & Beekman, J.H., 1989. A balanced diet as a goal for grazing: the food of the Manyara buffalo. *African Journal of Ecology*, 27(3), pp.241–259.

Ramoelo, A. et al., 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, pp.151–162.

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. et al., 2013. 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, pp.334–343.

Ramoelo, A. et al., 2015. 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.

Rasel, S.M.M. et al., 2016. Comparative analysis of Worldview-2 and Landsat 8 for coastal saltmarsh mapping accuracy assessment. In M. S. Kim, K. Chao, & B. A. Chin, eds. p. 986409. Available at: http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2222960 [Accessed March 28, 2017].

Rodriguez-Galiano, V. et al., 2015. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews*, 71, pp.804–818.

Sahoo, R.N., Ray, S.S. & Manjunath, K.R., 2015. Hyperspectral remote sensing of agriculture. *Current Science*, 108(5), pp.848–859.

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.

Seetha, M. et al., 2017. Comparison of advanced techniques of image classification. *Map world forum*.Hyderab, India.

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.

Skidmore, A.K. et al., 2005. Modelling herbivore grazing resources using hyperspectral remote sensing and GIS. In *AGILE 2005: 8th conference on geographic information science, Lisboa, AGILE, Denver, USA*. Citeseer. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.540.2592&rep=rep1&type=pdf [Accessed March 8, 2017].

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.

Stratoulias, D. et al., 2015. Evaluating Sentinel-2 for Lakeshore Habitat Mapping Based on Airborne Hyperspectral Data. *Sensors*, 15(9), pp.22956–22969.

Svoray, T., Perevolotsky, A. & Atkinson, P.M., 2013. Ecological sustainability in rangelands: the contribution of remote sensing. *International Journal of Remote Sensing*, 34(17), pp.6216–6242.

Thenkabail, A., Lyon, P.S. & Huete, J.G., 2011. *Hyperspectral remote sensing of vegetation*, CRC Press. Available at: http://www.crcnetbase.com/doi/pdf/10.1201/b11222-1 [Accessed February 28, 2017].

Thenkabail, P.S., Lyon, J.G. & Huete, A., 2016. 1 Advances in Hyperspectral Remote Sensing of Vegetation and Agricultural Croplands. *Hyperspectral Remote Sensing of Vegetation*, p.1.

Vaglio Laurin, G. et al., 2016. Discrimination of tropical forest types, dominant species, and mapping of functional guilds by hyperspectral and simulated multispectral Sentinel-2 data. *Remote Sensing of Environment*, 176, pp.163–176.

Vincenzi, S. et al., 2011. Application of a Random Forest algorithm to predict spatial distribution of the potential yield of Ruditapes philippinarum in the Venice lagoon, Italy. *Ecological Modelling*, 222(8), pp.1471–1478.

Wang, Y. et al., 2009. Validation of artificial neural network techniques in the estimation of nitrogen concentration in rape using canopy hyperspectral reflectance data. *International Journal of Remote Sensing*, 30(17), pp.4493–4505.

Wang, J. et al., 2015. Evaluating Different Methods for Grass Nutrient Estimation from Canopy Hyperspectral Reflectance. *Remote Sensing*, 7(5), pp.5901–5917.

Waske, B. et al., 2012. Image RF – A user-oriented implementation for remote sensing image analysis with Random Forests. *Environmental Modelling & Software*, 35, pp.192–193.

Youngentob, K.N. et al., 2012. Using imaging spectroscopy to estimate integrated measures of foliage nutritional quality. *Methods in Ecology and Evolution*, 3(2), pp.416–426.

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.

Zhang, Y. et al., 2011. Forest leaf chlorophyll study using hyperspectral remote sensing. *Hyperspectral remote sensing of vegetation*, pp.167–186.

Zhu, X., Zhang, H. & Yan, L., 2011. VARIATION AND INTERRELATIONS AMONG NUTRIENT ELEMENTS IN WHEAT LEAVES USED FOR FORAGE. *Journal of Plant Nutrition*, 34(9), pp.1321–1329.