UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

M.Sc Dissertation

Estimation and monitoring of grass nitrogen and phosphorus using Sentinel-2 derived spectral information and vegetation indices in a Savanna ecosystem

by

Charity Lehlabule Tseka

Supervisor: Prof. Abel Ramoelo

Co-Supervisor: Dr. Philemon Tsele

submitted in fulfilment of the requirements for the degree of Master of Science in Geography

Depatment of Geography, Geoinformatics and Mateorology, Faculty of Natural and Agricultural Sciences

University of Pretoria

July 2023

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

DECLARATION

I, Charity Lehlabule Tseka declare that this dissertation was written by me. It is submitted to the Faculty of Natural and Agricultural Sciences at the University of Pretoria in fulfilment of the Degree for Master of Science in Geography. It has not been submitted before for any degree at any other University.

Signature: _____  _____ Date: ____28/07/2023_____

DEDICATION

This thesis is dedicated to my late mother Ngwatomosadi'a Lehumo and my late daughter Bohlokwa…

ACKNOWLEDGEMENTS

I would like to extent my sincere gratitude to my supervisors Prof. Ramoelo and Dr. Tsele for their time and supervision on this research. Both my supervisor and co-supervisor encouraged me through the hardest time of my life when I felt like I cannot do it anymore. Thank you for your understanding nature. I am grateful for your support.

I would like to thank my family; my brother Kanyane, my cousins (Lehumo, Mogoshadi and Pebetse), my uncle (Malope) and his beautiful wife (Boledi). Thank you for your constant love and always believing in me. I will always be indebted to you.

A special thanks to my grandmother Mogoshadi'a Gobetse whom I attribute all my success to the moral and intellectual education I receive from her.

I would like to thank my father Sekhutle Tseka for his love and believing in me, I am because you are Ngwato'a Ngwamorei'a Kanyane.

I would also like to thank my extended family and friends for their support, especially those who have supported me in different ways since the passing of my mother. A special thanks to the Maroga's family for providing me with a roof over my head throughout my high school days, I would not have made it this far.

A special thanks to my friend Lerato Molekoa, who was always willing to make time out of her busy schedule to assist me.

I would also like to thank fellow students at the CFES lab especially Mcebisi Qabaqaba for helping me with parts of my research.

I would also like to thank my friend Dr. Lerato Mokoloko for giving me courage to complete my studies and proof reading my work, you are appreciated.

I would like to send my deepest appreciation to SANSA (South African National Space Agency) for funding my studies for the first 2 years.

A special thanks to my Ancestors for having my back, I would not have made it this far without their guidance and protection.Most importantly, I would like to thank my creator God for this life, for showering me with wisdom and strengthening me when I was at my weakest.

ABSTRACT

Grass quality as measured by leaf nitrogen (P) and phosphorus (N) plays a major role in understanding the distribution, densities and population dynamics of herbivores including livestock and wild herbivores. The aim of this study was to estimate and monitor grass N and P using Sentinel-2 derived spectral information and vegetation indices during the wet season for the selected period between 2017 and 2022 spanning a savanna ecosystem in the Kruger National Park area. Sentinel-2 satellite images were used as they provide images with high spatial, spectral and temporal resolutions. Field data were used where grass samples were collected, and spectral reflectance measurements (400 and 2350 nm) were undertaken. Three analysis scenarios were employed to estimate grass N and P in conjunction with classical regression and machine learning techniques. The scenarios included: (i) specific spectral bands, (ii) conventional and red edge-based vegetation indices (VIs) and (iii) a combination of VIs and spectral bands using the Stepwise Multiple Linear Regression (SMLR), Random Forest (RF) and Support Vector Machine (SVM) statistical models. Results showed that SMLR yielded the highest estimation accuracy based on a combination of bands and VIs for leaf N (i.e. Coefficient of determination (R^2) = 0.69 and Root Mean Square Error (RMSE) = 0.14%, Relative Root Mean Square Error (RRMSE) = 6.73%, Mean Absolute Error (MAE) = 0.11) and for P based on a combination of VIs only (i.e. R^2= 0.40 and RMSE= 0.04%, RRMSE = 34.322% and MAE = 0.04). This study confirms that the combination of red edge bands and VIs of Sentinel-2 data are crucial for accurately estimating biochemical concentrations in a savanna ecosystem. This study has significant implications for mapping and monitoring grass quality over large spatial extents.

Key words: Sentinel-2, Red edge, Vegetation indices, Cross-validation, Random Forest

DRAFT MANUSCRIPT FROM THIS DISSERTATION

1. Tseka C, Ramoelo A, Tsele P. and Qabaqaba M. (Draft manuscript). Estimating and monitoring of grass nitrogen and phosphorus using Sentinel-2 derived spectral information and vegetation indices in a Savanna ecosystem. Remote Sensing Journal.

PRESENTATIONS

1. "An application of Remote Sensing in estimating grass quality (leaf N and P) using Sentinel-2 data sets: The case study of Kruger National Park area" (**Oral presentation**). SANSA Student workshop, Hermanus, Western Cape, October 2017.
2. "An application of Remote Sensing in estimating grass quality (leaf N and P) using Sentinel-2 data sets: The case study of Kruger National Park area" (**Oral presentation**). Department of Geography, Geoinformatics and Meteorology Mini-Conference, University of Pretoria, November 2018.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
DGPS	Differential Global Positioning System
IR	Infrared
KNP	Kruger National Park
LAI	Leaf Area Index
N	Nitrogen
NDVI	Normalized Difference Vegetation Index
NDNI	Normalized Difference Nitrogen Index
NRI	Normalized Ratio Indices
OOB	Out Of Bag
RF	Random Forest
SLR	Simple Linear Regression
SMLR	Stepwise Multiple Linear Regression
SR	Simple Ratio
SVM	Support Vector Machine
SWIR	Short Wave Infrared
P	Phosphorus
PCA	Principal Component Analysis
VNIR	Visible Near Infra-Red

CHAPTER 1: INTRODUCTION

1.1. Background

Terrestrial ecosystems are land-based population of species whose role is to sustain life and fuels ecosystem processes. Nitrogen (N) and phosphorus (P) are nutrients that are known to limit the growth of plants in terrestrial ecosystems (Güsewell, 2004). For example, leaf N concentration plays a key role in the physiological processes of the leaf such as photosynthesis, transpiration and respiration (Wang et al. 2016). Furthermore, it is important for understanding vegetation health and quality as it also relates to the protein content (Wang et al. 2004), which is a very important nutrient required by herbivores (Prins and Beekman, 1989). Phosphorus is a non-renewable resource that plays a major role in crop production and plant growth (Shen et al. 2011). It also plays a key role in different metabolic processes in plants such as respiration, protein formation, energy storage, photosynthesis and nutrient movement within the plant (Shen et al. 2011; Siedliska et al. 2021).

Livestock production is the dominating primary activity in most rural areas and it is dependent on grass quality (Shackleton et al. 2002). Several studies postulated that high-nutrient areas such as sodic sites, termite mount, valleys, underneath trees, and bottom lands have higher herbivore densities (Ferwerda et al. 2005; Grant and Scholes, 2006). Other factors which affect the distribution of herbivores include biomass availability (Bergman et al. 2001), soil and foliar nutrient status (McNoughton, 1990), the distance from water sources as well as threats to predation (Redfern et al. 2003). Consequently, leaf N plays a major role in planning and managing grazing areas by determining the carrying capacity, spatial zoning and grazing camps (Ramoelo et al. 2015). Nutrient-rich areas are characterized by the presence of large herbivores, and understanding the factors that affect the distribution of these animals will assist in implementing conservation strategies that will sustainably manage the ecosystems (Grant and Scholes, 2006).

There is a need to sustainably manage grazing camps to ensure an increase in livestock production and the development of rural communities (Ramoelo et al. 2015). This can be done by predicting the total number of animals that can be supported by the area using

management tools such as stocking rate and carrying capacity and developing monitoring programs for nutrient rich areas (Grant and Scholes, 2006). However, very few studies were carried out on understanding the factors that lead to spatial distribution of herbivores, because of the unavailability of the spatially explicit resource and quality indicators. Traditional methods of estimating grass nutrients were previously carried out by sampling leaves, drying them and weighing them to calculate the concentration of leaf N (Clevers et al. 2007; Mutanga et al. 2015). This is however a tedious, costly and time-consuming process at large spatial scales (Clevers et al. 2007).

Recent studies focus on relating remotely sensed data with vegetation leaf chemistry in predicting grass nutrients using the indices derived from a combination of bands from the red edge and infrared part of the electromagnetic spectrum (Ferwerda et al. 2005). Hyperspectral remote sensing has been instrumental in estimating Leaf N and P because of the full spectrum and the red edge inflection point (REP) (Knox et al. 2011; Tian et al. 2011). The REP is known to relate to chlorophyll and N (Knox et al. 2011). Today, satellites such as Sentinel-2 carry on-board a multispectral instrument that has the red edge band (among other bands) known to be critical for mapping chlorophyll and N (Wang et al. 2016). This satellite provides remotely-sensed data that is crucial for a wide range of applications related to e.g., land management, disaster control and agricultural purposes (food security) (Aschbacher and Milagro-Pérez, 2012). Statistical methods and machine learning techniques have been previously used to estimate leaf N and P (Curran, 1989; Mutanga et al. 2012; Manyashi, 2015; Ramoelo et al. 2015). Machine learning techniques have been widely used in remote sensing to estimate grass quality, however is it limited by the availability of ground truth data to reach its full potential. This is mainly because ground truth data collection is tedious, time consuming and has financial implications (Moghimi et al. 2020).

Accurate estimations of grass biomass offer crucial information about the functioning and productivity of rangelands (Cho et al. 2007; Ramoelo et al. 2015). Spectral bands of Sentinel-2 satellite imagery, traditional VIs and red edge based VIs were used to test the applicability of estimating the concentration of grass N and P. This study intends to use both statistical and machine learning approaches to predict leaf N and P. Furthermore, employed Sentinel-2

satellite imagery to map the distribution of these two variables over time, and such information is limited.

1.2. Problem Statement

Leaf N is an important factor that relates to protein content and it is an important nutrient required for herbivores, it is postulated that leaf nutrient affects the feeding and movement of herbivores (La Pierre and Smith, 2016). It is therefore important to map the spatial distribution of leaf N content to understand the feeding patterns of animals. Estimating leaf N concentration could assist in understanding the feeding and movement patterns of herbivores, which will contribute greatly to the environmental management of the rangeland. There are numerous multispectral satellite remote sensing data sets that have been used to estimate leaf N concentration, notwithstanding their diverse inherent limitations such as the absence of a red-edge band (Cho and Skidmore, 2006).

On the other hand, P is crucial for crop production and plant growth, it also plays a key role in different metabolic processes in plants such as respiration, protein formation, energy storage, photosynthesis and nutrient movement within the plant (Shen et al. 2011; Siedliska et al. 2021). Poor management of P is detrimental to plants and the environment with cost implications (Siedliska et al. 2021). Appropriate P management plays a major role in precision agriculture (Siedliska et al. 2021). The lack of P negatively affects the production of chlorophyll in plants (Choi and Lee, 2012). It is therefore important to manage the distribution of P. For example, Siedliska et al. (2021) conducted a study where P content was measured using visual inspection and chemical analysis but these methods are somewhat tedious, costly and time-consuming.

Previously mapping grass nutrients have been a challenge due to the lack of freely available high spatio-temporal resolution satellite data that incorporate the red edge band(s), known to be sensitive to chlorophyll and N (Ramoelo et al. 2011). However, the launch of satellite sensors such as RapidEye, SumbandilaSat and Multispectral Instrument (MSI) on-board Sentinel-2 allows for the regional mapping of leaf N (Ramoelo et al. 2011; Chabalala et al.

2020). Hyperspectral remote sensing provides a great tool for predicting leaf N due to the presence of red-edge bands (700 – 750nm). Few studies have shown the potential of red edge bands in estimating leaf N at a large scale (Haung et al. 2004; Cho and Skidmore, 2006). Sentinel-2 satellite is a new addition to the already existing remote sensing satellites and in particular, one of the new generations of terrestrial observing satellites which consist of the red edge spectral bands (705 nm and 740 nm) (Baillarin et al. 2012). Therefore, this study evaluated the use of various red edge based indices, and spectral bands to estimate grass N and P during the wet season spanning the period 2017 and 2022 over the savanna ecosystem in Kruger National Park (KNP) area. Furthermore, conventional statistical and machine learning techniques were tested in estimating grass N and P and to establish whether the spectral bands and vegetation indices (VIs) found to be critical in estimating grass N, are similar to those of grass P.

1.3. Research Aim and Objectives

The aim of this study was to estimate and monitor grass N and P using Sentinel-2 derived spectral information and vegetation indices during the wet season for the selected period between 2017 and 2022 spanning a savanna ecosystem (KNP).

The objectives of this research study were to:

(1) Estimate grass N and P using Sentinel-2 spectral bands coupled with statistical and machine learning methods,

(2) Estimate grass N and P using Sentinel-2 derived traditional and red-edge based vegetation indices coupled with statistical and machine learning methods,

(3) Estimate grass N and P using a combination of spectral bands and vegetation indices coupled with statistical and machine learning methods, and

(4) Use the optimal grass N and P estimation models to map and monitor the respective spatial distributions of these grass nutrients during the wet season of the period 2017 and 2022 over the savanna ecosystem in KNP.

CHAPTER 2: LITERATURE REVIEW

2.1.The importance of leaf N and P and the impact on the feeding and movement of herbivores

Nitrogen (N) plays a key role in the physiological processes of the leaf such as respiration, photosynthesis and transpiration (Wang et al. 2016). It is linked to the chlorophyll content as well as the net primary productivity. Nitrogen is one of the nutrients in plants that affect the ecosystem processes such as grazing behavior (Ferwerda et al. 2005). The disadvantage of leaf N is that its excess limits the growth of the plant (Tian et al. 2011; Ollinger and Smith, 2005). It is therefore important to manage the amount of N. The combination of chlorophyll and leaf area index (LAI) provides a good estimation for the canopy N concentration (Baret et al. 2007). Several studies have determined leaf N concentration in forests using satellite data (Coops et al. 2003, Chabalala et al. 2020).

Phosphorus is a non-renewable resource that plays a major role in crop yield (Shen et al. 2011). Phosphorus is crucial for crop production and plant growth, it also plays a key role in different metabolic processes in plants such as respiration, protein formation, energy storage, photosynthesis and nutrient movement within the plant (Shen et al. 2011; Siedliska etal. 2021). Poor management of P fertilizer is detrimental to plants and the environment with cost implications (Siedliska et al. 2021). It was previously shown thatthe relationship between P content and chlorophyll is inversely proportional (Siedliska etal. 2021).

The shift in plant tissue chemistry affects the availability of nutrients as a result herbivores are unable to meet their nutritional needs (La Pierre and Smith, 2016). This therefore affects the feeding of secondary consumers. A study conducted by Ben-Shahar and Coe (1992) showed that areas with a high diversity of grass have high nutrient levels.

2.2.The application of remote sensing in estimating grass nutrients

Remote sensing provides as opportunity to observe objects on the earth's surface without being in physical contact with them (Hall et al. 2002). These objects are observed using satellites or aircraft-mounted sensors which record the amount of electromagnetic energy reflected by the surface of the objects. Traditional methods of predicting leaf N are tedious and costly.

Remote sensing, therefore, serves as the most efficient way to predict leaf N over large scales (Ramoelo et al. 2015).

The ability of the sensor to detect these objects is dependent on the sensor's spectral, spatial, temporal and radiometric resolution. The size and number of spectral bands a sensor possesses across the electromagnetic spectrum affect the spectral resolution (Clark et al. 1990). By calculating the smallest object or feature that the sensor can detect at that specific place, spatial resolution establishes the sensor's position in relation to distance from the object or area of interest (Hall et al. 2002). Temporal resolution, commonly referred to as the revisit time, is the length of time it takes a sensor to collect data from a single item (Clark et al. 1990). A characteristic of radiometric resolution is the number of divisions that a sensor can divide or separate data derived from reflected irradiation, the higher the divisions the better the sensors (King,1992). Remote sensing satellite tools are improved yearly to improve the above-mentioned four resolutions (Hall et al. 2002). Remote sensing plays a critical role in assessing the condition of the vegetation over large spatial scales and aids in analyzing ecological issues over the areas (Serrano et al. 2002). Optical remote sensing techniques provide an approach to studying different phenomena on the earth through the acquisition of images at different spatial and temporal scales (Yu et al. 2014).

2.2.1.Hyperspectral sensors

Hyperspectral sensors are extremely spectrally advanced as they cover visible, near-infrared (IR) and short-wave infrared (SWIR) in many narrowly defined spectral channels of the electromagnetic spectrum (Campbell, 2007; Oldeland et al. 2010). Hyperspectral sensors provide 200 or more spectral channels (Campbell, 2007).

The ability of hyperspectral data to detect small changes in the biochemical content of the vegetation gives it an advantage in estimating foliar N concentration (Abdel- Rahman et al. 2013). Hyperspectral sensors can detect the narrow absorption features of N. This is due to the narrow spectral band information which provides an effective method for estimating leaf N (Wang et al. 2016). For example, Cho and Skidmore (2006) reported that WorldView-2 and

Rapid Eye are amongst the first satellite sensors with a red edge band and they have been used to estimate leaf N due to the relationship of a red edge band with N and chlorophyll. A study conducted by Chabalala et al. (2020) showed that the red edge bands and VIs derived from Sentinel-2 provides a greater potential for predicting leaf N as compared to Rapid Eye.

Phosphorus content can be predicted indirectly by predicting the content of a substance that relates to P such as chlorophyll, alternately P content can be estimated by combining different spectral bands (Siedliska et al. 2021). Hyperspectral remote sensing allows for the collection of canopy images which provides reflectance information that is crucial for the estimation of P in plants (Siedliska et al. 2021).

Studies have shown success in using hyperspectral datasets to estimate leaf N using VIs, full spectral, absorption features and the integrated method (Huang et al. 2004; Ramoelo et al. 2011; Ramoelo et al. 2013). Siedliska et al. (2021) employed hyperspectral imagery to detect leaf P in crops. Previous studies conducted by Christensen et al. (2004) showed high accuracy (74%) in predicting P content based on spectral canopy reflectance.

The drawback of using hyperspectral data is that it provides imagery that is redundant due to the large number of bands that are adjacent to each other, and therefore there is a need to select only the relevant spectral bands (Cho et al. 2010). In order to reduce the redundancy, principal component analysis (PCA) can be used (Kalacska et al. 2007) and feature selection techniques (Jackson, 2005). Another disadvantage of using hyperspectral data is that it is expensive and available on a small scale (Goetz, 2009; Knox et al. 2011). The different technologies that have been used to estimate nutrients in plants from are explained below:

2.2.1.1.Estimating grass nutrients using hyperspectral vegetation indices

The vegetation index (VI) (simple ratio) describes the greenness in vegetation and hence the indicating vegetation health (Tucker, 1979). Vegetation indices provide the easiest way of estimating leaf biochemicals including N content (Wang et al. 2016). Chlorophyll and N have

been correlated among different species in previous studies (Evans, 1989; Homolova et al. 2013; Wang et al. 2015). As a result, vegetation indices used for chlorophyll and are also used to estimate N concentration (Wang et al. 2016).

Hyperspectral remote sensing has been instrumental in estimating leaf N because of the full spectrum and also because of the red edge infliction point (REP) (Tian et al. 2011). A study conducted by Haboudane et al. (2002) employed the ratio TCARI/OSAVI to predict chlorophyll content using a hyperspectral Compact Airborne Spectrographic Imager (CASI). The Normalized Difference Vegetation Index (NDVI) represents the ratio between NIR and the red part of the electromagnetic spectrum (Meer et al. 2001). Several studies used hyperspectral data to estimate biochemical content (Curran et al. 2001, Mutanga et al. 2004). Tian et al. (2011) employed Hyperion image hyperspectral to estimate leaf N concentration using VIs. VIs have been proven to be successful in estimating vegetation parameters (Ramoelo et al. 2012; Manyashi, 2015). However, conventional VIs such as NDVI or simple ratio uses the visible and near-IR red bands; they do not include the SWIR and the red-edge bands (Ramoelo et al. 2015).

2.2.1.2.Estimating grass nutrients using absorption features

There are three main sections on the electromagnetic spectrum that relate to the estimation of chlorophyll and N content, namely the red edge (700-750 nm), red (630-690 nm) and the green band (500-580 nm) (Tian et al. 2011). The electromagnetic spectrum consists of absorption features that are associated with physical bond vibrations or electron transition of certain biochemical content (Darvishzadeh et al. 2008). For example, the red edge region of the electromagnetic spectrum is characterized by the concentration of chlorophyll (Lamb et al. 2002). Knox et al. (2010) showed that the following wavelengths in Table 1 are associated with N.

Table 1 Absorption features that have been identified in previous studies (Curran, 1992, Cho and Skidmore, 2006, Mutanga et al. 2004, Knox et al. 2010.

Wavelength (nm)	Bond variations	Biochemical
430	electron transition	chl a
460	electron transition	chl b
640	electron transition	chl b
660	electron transition	chl a
910[a]	C–H stretch, third overtone	protein
1020[a]	C–H stretch, second overtone	protein
1420	C–H stretch, C–H deformation	protein, lignin
1510[a]	N–H stretch, first overtone	protein, nitrogen
1520[a]		protein
1690[a]	C–H stretch, first overtone	protein, nitrogen, lignin, starch
1730[a]	C–H stretch	protein, cellulose, lignin
1940	O–H stretch, O–H deformation	protein, nitrogen, lignin, starch, water, cellulose
1950	O–H stretch, O–H deformation	protein, nitrogen, lignin, starch, water, cellulose
1960	O–H stretch, O–H bend	protein, sugar, starch
1980[a]	N–H asymmetry	protein, lignin
2060[a]	N–H bend, N–H stretch, second overtone	protein, nitrogen
2130[a]	N–H stretch	protein
2180[a]	N–H bend, second overtone, C–H stretch, C–O stretch, C–N stretch	protein, nitrogen

2200[a]		protein
2240[a]	C–H stretch	protein
2270	C–H stretch, O–H stretch, CH2 bend, CH2 stretch	protein, nitrogen, lignin, starch, sugar, cellulose
2290[a]		protein
2300[a]	N–H bend, C–O stretch, C–H bend, second overtone	protein, nitrogen, cellulose
2350[a]	CH2 bend, second overtone, C–H deformation, second overtone	protein, nitrogen, cellulose

[a]Wavelength shown to relate to nitrogen in fresh material (Cho and Skidmore, 2006, Mutanga, 2004, Huang et al. 2004)

The concentration of N is correlated with chlorophyll content, consequently the high concentration of N in plants results in chlorosis (yellowing) which is characteristic of the lack of chlorophyll (Lamb et al. 2002). Nitrogen concentration can be estimated using reflectance spectroscopy since reflectance in the visible near-IR is influenced by the pigmentation of chlorophyll and the leaf structure (Lamb et al. 2002). Ramoelo et al. (2011) used the known absorption features for chlorophyll, protein and N to predict leaf N. The advantage of using this method is that it avoids redundancy in the data (Ramoelo et al. 2012). Mapping grass nutrients using NIR and SWIR may not yield accurate results due to water absorption (Gao and Goetz, 1994).

2.2.2. Multispectral sensors

2.2.2.1. Estimating grass nutrients using specific bands

Fernandez-Habas et al. (2022) showed the importance of bands located in the red edge and NIR (700, 710, 1160 and 1170 nm) in predicting grass nutrients, whereas the bands from the SWIR showed low stability. Previous studies have shown that the 440, 445, 730 and 930 nm in the VNIR region play a crucial role in predicting phosphorus (Osborne et al. 2002; Knox

et al. 2011). A study conducted by Singh et al. (2017) investigated known absorption wavebands for detecting grassland nutrients in KwaZulu-Natal and the results showed that the spectral bands located within the blue, red and SWIR play a crucial role in determining forage quality.

2.2.2.2. Estimating grass nutrient using conventional and red edge based indices

Vegetation indices are commonly used as the predictors for grass variables in astatistical way (Clevers et al. 2007). Conventional and red edge based vegetation indices such as the NDVI and simple ratio (SR) can be used to estimate the concentration of N in the grass (Ramoelo et al. 2013; Mutanga et al. 2015). NDVI depicts areas of high photosynthetic activity and measures the greenness on the surface of the earth (Lück et al. 2010). Equation 1 below shows the equation for NDVI. The NDVI is the commonly used index, where the red reflectance and the NIR reflectance are used to estimate the concentration of N, chlorophyll as well as other crop variables (Hansen and Schjoerring, 2003).

$$NDVI = \frac{pNIR - pRed}{pNIR + pRed}$$ *(Equation 1)* (Campbell, 2007)

Where: $pNIR$ is the reflectance value in the near-IR on the EMS and $pRed$ is the reflectance value in the red region of the electromagnetic spectrum.

NDVI values range from -1 to +1, the values below 0.1 correspond to water bodies whereas higher values correspond to vegetation (Menses-Tovar, 2011).

This vegetation index is often used because it utilizes the red-edge effect to reduce background reflectance (Haung et al. 2004). The drawback of this method is that it only accommodates the NIR and the red bands which are only sensitive to leaf pigmentation (Haung et al. 2004). Ramoelo et al. (2014) also used several band combinations of NDVI which corresponded to a different combination of SR, the vegetation indices were then correlated to leaf N. The disadvantage of using NDVI is that it saturates at high biomass and it is prone to soil background effects (Bausch, 1993). NDVI uses only two bands of hyperspectral data and this could be a limitation to the amount of information that can be extracted (Cho et al. 2007).

Normalized Difference Nitrogen Index (NDNI) is calculated from the SWIR, it has a strong correlation with N concentration and therefore indicates N content (Yu et al. 2014). NDNI values range from 0-1 and the green vegetation NDNI is 0.02-0.15 (Yu et al. 2014). The normalized difference Nitrogen index (NDNI) was used at shrub ecosystems to estimate foliar N (Serrano et al. 2002). Ferwerda et al. (2005) conducted a study of comparing different band combinations across 300 to 2500 nm of normalized ratio indices (NRI) shown by equation 2 below.

$$NRI = \frac{Rband1 - Rband2}{Rband1 + Rband2} \qquad \text{(Equation 2)}$$

where the wavelength of band 1 is greater than band 2, correlation of N and NIR was proven to be high when band two was chosen between 500 and 750 nm (Ferwerda et al. 2005).

Very often, vegetation indices combine near-IR and visible spectral bands. The disadvantage of this combination is that these indices have less sensitivity due to the strong absorption by chlorophyll (Clevers and Gitelson, 2013). To avoid this problem, the red edge region is used to reduce the saturation effect as there is lower absorption of chlorophyll in the red edge. Red edge based NDVIs has been proven to show high accuracy in estimating foliar N as compared to conventional NDVIs derived from the visible and NIR (Zengeya et al. 2013).

The estimation of leaf N by using the red edge band and computing vegetation indices has been widely used during the wet season (Cho and Skidmore, 2006). A study conducted by Ramoelo et al. (2015) on monitoring grass nutrients using WorldView-2 showed that red-edge band indices play a pivotal role in predicting leaf N. A study focusing on the potential of sentinel-2 and sentinel-3 in estimating crop and grass chlorophyll and N content was conducted by Clevers and Gitelson (2013). This study confirmed the reliability of using the red-edge bands of Sentinel-2 for agricultural applications.

There is a large number of indices that are used to estimate chlorophyll, however only a limited number of indices are specifically for estimating N concentration (Wanget al. 2016). Chlorophyll determines the amount of reflection in the visible region of the electromagnetic

spectrum and therefore allows for the estimation of the chlorophyll and N content (Clevers and Gitelson, 2013).

Horler et al. (1983) were the first to show the significance of the REP in identifying plant stress; the REP was then used to estimatechlorophyll and N concentration. Equation 3 below shows the REP which is the maximum slope in the red-near IR region and has been successfully used to estimate grass biomass (Mutanga and Skidmore, 2004).

$$(REP = 705 + 35\ \frac{(R665+R783)/2-R705}{R740-R705}) \qquad \textit{(Equation 3)}$$

Where R665 is the red band with a wavelength of 665nm whereas R783, R705 and R740 are is the red edge bands at wavelengths of 783nm, 705nm and 740nm respectively.

REP index was also studied for estimating chlorophyll content (Guyot and Baret, 1988). The $CI_{red\text{-}edge}$ and CI_{green} index was regarded as the best for estimating both the chlorophyll and N content (Clevers and Gitelson, 2013). In contrast to the normalized NDVI, the REP is less likely to be affected by different soil and atmospheric conditions (Cho and Skidmore, 2006; Clevers et al. 2001).

2.3. Statistical and machine learning techniques for estimating grass nutrients

2.3.1. Univariate (simple linear regression)

Simple linear regression (SLR) (shown by equation 4 below) is a statistical method that is used to study the relationship between two variables (*x* and *y*), where *x* is an independent (predictor) variable and *y* is a dependent (response) variable (Dowdy et al. 2011). If there is a relationship between x and *y*, a model can be developed to predict new observations of *y* from *x* (Dowdy et al. 2011, He et al. 2021).

$$y = \beta_0 + \beta_1 x_1 \qquad \textit{(Equation 4)}$$

Where β_0 is the *y* intercept and β_1 is the slope.

Sudalaimuthu and Sudalayandi (2019) successfully used a linear regression model to predict ground elevation measurements from satellite elevation using Cartosat-1 imagery. Hihi et al. (2019) also employed linear regression to predict spatial differences in soil salinity using

multispectral Sentinel-2 satellite imagery. This analytical technique has the following advantages; it is relatively easy to understand, use and interpret. For data samples where the absolute value of the Pearson correlation coefficient, r is high (further from zero, and closer to 1 or -1) linear regression tends to have predictive results with high accuracy (Nicolic et al. 2012, Sudalaimuthu and Sudalayandi, 2019). Some of the disadvantages are; selecting the correct feature to yield optimum results is not always easy, and since it uses a single dependent variable, some problems may require parameterization, which is also not always easy (Dowdy et al. 2011, Kuchibhotla et al. 2019).

2.3.2. Multivariate regression

2.3.2.1 Stepwise Multiple Linear Regression (SMLR)

Stepwise Multiple Linear Regression (SMLR) is a generalization of Simple Linear Regression into multiple or higher dimensional spaces (Kuchibhotla, et al. 2019, He et al. 2021). It generalizes Equation 5 as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \qquad \textit{(Equation 5)}$$

Where y is the response variable, β_0 is the constant variable and β_1, β_2 *and* β_n are the coefficient of the control variables, x_1, x_2 *and* x_n *are* the controlled variables.

This SMLR technique is commonly used in remote sensing for predicting grass nutrients (Clevers et al. 2007, Ramoelo, 2012, Askari et al. 2019). For high dimensional data, a feature selection process is usually essential, which allows for the selection of the optimum set of spectral bands for predicting vegetation variables (Clevers et al. 2007). Variable selection was previously performed using the lowest Akaike Information Criterion (AIC) (Sakamoto et al. 1986). (Empirical techniques have been widely used in deriving the variables of vegetation such as N from remote sensing (Wang et al. 2016). Kokaly and Clark (1999) and Serrano et al. (2002) used SMLR, while Martin et al. (2008) used partial least squares regression (PSLR) to determine leaf biochemistry.

Several studies have used the SMLR statistical method to estimate leaf N (Johnson et al. 1994; Martin and Aber, 1997; Ramoelo et al. 2012).The problem of the number of wavebands being greater than the number of samples (multi-collinearity) is likely to be encountered when

using multiple regression with hyperspectral data (De Jong et al. 2003). The principal component analysis method (PCA) can be employed to reduce multi-collinearity by decomposing independent variables into uncorrelated components (Jain et al. 2007). Alternatively, the PLSR can also be used to avoid multi-collinearity because it is a full-spectrum method in contrast to SMLR (Kooistra et al. 2004; Darvishzadeh et al. 2008).

2.3.2.2. Support Vector Machines (SVMs)

The SVM requires a labeled training dataset with the correct categories (Steeb, 2015). New examples are classified into the identified data classes, making SVM classification a non-probabilistic task. Previously it has been shown that datasets of different classes tend to overlap, as a result, non-linear decision boundaries are used to map datasets into a three-dimensional space, called the feature space (Wang, 2005). This adjustment on the data vectors is usually applied, lifting each into a higher dimensional space *N+1*, which often leads to better separability (Steeb 2015).

A study conducted by Clevers et al. (2007) employed SVM band shaving to select thebest spectral bands as predictors in estimating grassland biomass using hyperspectral data. Hsu and Lin (2002) have shown that the directed acyclic graph SVM and the one-against-one procedure are suitable for multiclass classification. Another study conducted by Mashao (2003) compared the use of gaussian mixture modes (GMM) and SVM in response to spectral compression in the feature sets and it was evident that SVM performs better. Pal et al. (2005) compared SVM, maximum likelihood and ANNs in classifying multispectral and hyperspectral data, the results have shown thatSVM yields a higher accuracy with small training datasets as compared to other classification methods. The advantage of using SVM is that it can easily classify data with a high number of dimensions. Moreover, it is relatively easy to implement and interpret (Wang, 2005).The disadvantages of using SVM involve low-performance accuracy on larger datasets and the inability to operate with classes that are not distinct or labeled (Mashao, 2003).

2.3.2.3. Random Forest (RF)

The RF algorithm is an ensemble learning algorithm built using multiple decision trees,each trained on a subset of the feature space with a randomized selection to maximize variance, and the bagging technique for aggregation of solutions per subtree for optimal performance (Ho, 1998). It has shown great success over traditional decision trees, with reduced overfitting and an ability to train on high dimensional data with improved accuracy (Ho, 1998, Belgiu and Dragut, 2016). Given a dataset *D* with feature vectors *xi*, for each feature vector, a subset of feature is used, selected at random, to populate a decision tree classifier. Classification then follows the traditional decision tree algorithm for each partial or subtree. At the end of the training time, the results of each training subtree or partial tree are aggregated, and the aggregation which yields maximal accuracy is then selected. The randomized selection allows for a degree of parallel computation on the feature space by using multiple decision trees for the same feature vector, however, with a possible overlap offeatures per tree (Breiman, 2001, Belgiu and Dragut, 2016).

For classification problems, the dataset is split into training and validation sets, with the training set usually larger than the validation set. The more frequent ratio is a 70:30 split for training and validation, respectively. However, changing this ratio is also a part of algorithm tuning, which helps with convergence to better results (Belgiu and Dragut,2016). Random Forest can also be used for regression problems, where the initial computation is similar to that of classification, and the aggregation is then tweaked to return a number or aggregate with the least error based on the Euclidean error metricused, usually RMSE. However, Random Forest has a shortfall in performing unbounded or extrapolation regression, which makes it less effective for extrapolating since it can only compute into the range present in the feature space (Ho, 1998, Breiman, 2001, Belgiu and Dragut, 2016).

Together with other machine learning techniques, Random Forest has been used for the prediction of grass nutrients such as crude protein, leaf N and P (Skidmore et al. 2010; Knox et al. 2011; Fernández-Habas et al. 2022). It has also been found to be robust when predicting vegetation parameters (Kokaly et al. 2009). Other literature has suggested that larger feature vectors such as hyperspectral bands tend to have better classification accuracy with this algorithm when there are fewer classes (Breiman, 2001).

Random Forest has been used for monitoring leaf N and above-ground biomass using WorldView-2 satellite imagery (Ramoelo et al. 2015). Previous studies have shown that the RF regression model has a potential for predicting grass nutrients using hyperspectral data (Addel-Rahman et al. 2013; Punalekar et al. 2021; Fernández- Habas et al. 2022). Adam et al. (2012) also indicated that the RF model was more robust in terms of estimating vegetation parameters as compared to other parametric regression models. Although machine learning techniques have been widely used in remote sensing to estimate grass quality, is it limited by the availability of ground truth data to reach its full potential this is mainly because ground truth data collection is tedious, time consuming and has financial implications (Adjorlolo et al. 2012; Moghimi et al. 2020).

CHAPTER 3: MATERIALS AND METHODS

3.1. Study Area

3.1.1. Motivation for selecting study location

The KNP (Figure 1) located in the Savanna ecosystem, the area was selected for this study because it is very rich in biodiversity. Considering the various factors that affect the distribution of grass N and P, this location serves as a good example for testing whether the Sentinel-2 imagery can successfully estimate the concentration of grass N and P.

3.1.2. Description of study area

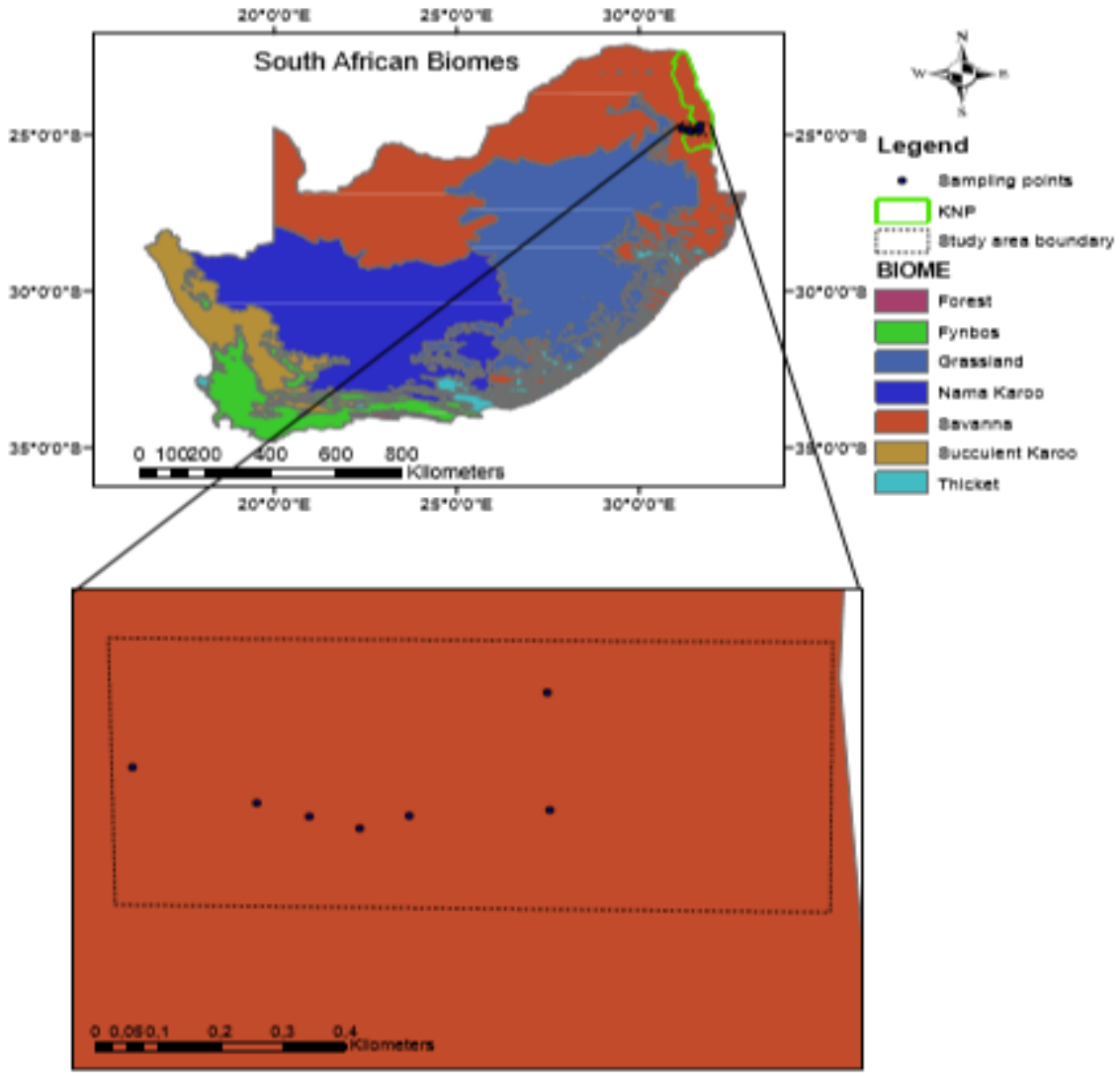


Figure 1 The study area in the Savanna ecosystem

KNP is a protected area that is geographically situated at 24° 00' 41" S and 31° 29'07" in the

eastern parts of Mpumalanga and Limpopo (Foxcroft et al. 2008). It is amongst the largest game reserves in Africa and is known for being rich in biodiversity. It was established in 1898 and covers an area of about 20 000 square kilometers (Foxcroft et al. 2008). The park is in the Lowveld region and forms part of the semi-arid Savanna biome as depicted in Figure 1 above (Low and Rabelo, 1998).

The geology of the park consists of granite rocks in the western part of the park and basaltic rocks in the eastern part of the park (Grant and Scholes, 2006). The Olifants River divides the park into North and South. The southern part is more humid with a mean annual rainfall of 500 – 700 mm in contrast to the northern part which is dry with a mean annual rainfall of 300 - 500 mm (Grant and Scholes, 2006). The park experiences humid summer days with hot temperatures, and the rainy season is from September until May (Foxcroft et al. 2008).

3.2. Data collection

3.2.1. Field data collection

I. Sample collection and spectral measurements

Several studies (Skidmore et al. 2010; Ramoelo et al. 2012 and Ramoelo et al. 2015) have shown that the best period to estimate grass nutrients is during peak productivity. Hence, the field spectral measurements were collected by Ramoelo et al. (2015) during peak productivity (March/April 2009). Eight sample sites were selected across three different land uses, henceforth, L denotes land use (Table 2), which are; KNP (L1 gabbro, L2 granite), Sabi-Sands Game Reserve (SGR)(L3 granite, L4 gabbro) and the communal lands (L5-6 gabbro, L7 -8 granite, at Bushbuckridge. 1:250 000 geology maps and SPOT 5 images were used to set the boundaries of the total area (35 000 ha) of the sample sites (Wessels et al. 2011). A line transects sampling design (Fewster et al. 2005) was employed to collect the field data in all the sample sites, however, due to the limitation of access to L3 this design was not applied to L3. 49 plots (30 m × 30 m) were created following a combination of the systematic and purposive method (Ramoelo et al. 2015). The plots were separated by a distance of 500 m to 1000 m depending on the homogeneity and the ease of access to the area. In each plot, three to four (0.5 m x 0.5 m) sub-plots were randomly created. The Leica®'s GS20 differential

geographic positioning System (DGPS) was used to capture the sample locations (Ramoelo et al. 2015). The DGPS points were then imported to Leica's GeoPro software for post-processing to produce not more than 1 m positional accuracy (Ramoelo et al. 2015). The dominating grass species were collected together with the grass samples. The grass samples were collected at the same time to avoid diurnal variations in the mineral contents (Kreulen and Hoppe, 1979).

Table 2 Field sampling and spectral measurement dates (Ramoelo et al. 2015)

Sample name	Date	Geology types	Land use	Measurements
L1(G1-5)	17 April 2009	Granite	Protected area	x
L2 (G1-4)	01 April 2009	Granite	Protected area	x
L2 (G4-5)	08 April 2009	Granite	Protected area	x
L4 (G1-6)	15 April 2009	Gabbro	Communal area	x
L4 (G7-10)	16 April 2009	Gabbro	Communal area	x
L5 (G1-5)	18 April 2009	Gabbro	Communal area	x
L6 (G1-5)	07 April 2009	Gabbro	Communal area	x
L7 (G1-6)	06 April 2009	Granite	Communal area	x
L8 (G1-4)	02 April 2009	Granite	Communal area	x
L8 (G5-6)	03 April 2009	Granite	Communal area	x
L8 (G7-9)	04 April 2009	Granite	Communal area	x

Before collecting the grass, spectral reflectance measurements were taken using the Analytical Spectral Device (ASD) FieldSpec®3 spectrometer (Clevers and Gitelson, 2013). The device has a 1 nm narrow band resolution between 350-2500 nm sample range and consists of a 25° Field of View (FOV) with the use of a permanent fibre optic pistol held at 1 m above the ground and nadir to accommodate the total area of the subplot. Multiple spectral measurements were recorded and then averaged to obtain the spectral measurement for each subplot (Adjorlolo et

al. 2012; Mutanga et al. 2015). The spectral measurements were taken on a sunny day (between 10h30 and 15h00) for increased illumination (Abdel-Rahman et al. 2010). The spectral radiance values were converted into reflectance prior to each measurement using the spectral on reference panel (Ramoelo et al. 2015).

II. Chemical analysis

The obtained grass samples were dried at 80˚C in an oven for 48 hours (Knox et al. 2011), and the measurements were then averaged for each plot (Ramoelo et al. 2015). Chemical analysis of the dried grass samples was performed at South Africa's Agricultural Research Council-Institute for Tropical and Subtropical Crops (ARC-ITSC) in Nelspruit (Ramoelo et al. 2015). Leaf N concentrations were retrieved from the leaf samples using the acid digestion technique in sulphuric acid whereas foliar P concentrations were retrieved using a mixture of perchloric acid and nitric acid (Mutanga et al. 2004; Ramoelo et al. 2011b and Zengeya et al. 2013).

III. Spectral resampling

The spectral resampling of the analytical spectral device (ASD) reflectance data was computed on R Studio. The data was simulated to Sentinel-2 sensor's configurations using Sentinel-2's spectral response function. The Savitzky-Golay (S-G) filter method was used for smoothing the spectral data by removing random noise and computing the first derivative of the spectral data. This method was first proposed by Savitzky and Golay (1964), it fits a high-order polynomial to the data and removes fluctuations until the optimum results are obtained (Savitzky & Golay, 1964). The method uses a least-squares fitting and requires a fixed number of sample points to give a smooth value (Bian et al. 2010). The disadvantage of this method is that it reduces the resolution of the data (Pres & Teukolsky, 1990). The Shapiro-Wilk normality test was conducted in R Studio to test the normality of the data.

IV. First-derivative

A first-derivative transformation is commonly used to improve absorption features that might be masked by interfering background absorption (Curran et al. 1990, Dawson et al. 1990).

Spectral derivatives also help in reducing the continuum resulting from other leaf biochemical (Curran et al. 1990). In this study the reflectance spectra were transformed into a first derivative spectrum to determine the slope of the spectrum using Equation 7 below.

$$D_{\lambda(i)} = (R_{\lambda(j+1)} - R_{\lambda(j)})/\Delta\lambda \qquad \textit{(Equation 7)}$$

Where $D_{\lambda(i)}$ is the first derivative transformation at a wavelength *í* midpoint between wavebands *j* and *j* + 1. $R_{\lambda(j)}$ is the reflectance at the *j* waveband, $R_{\lambda(j+1)}$ is the reflectance at he *j* + 1 and the $\Delta\lambda$ is the difference in wavelengths between *j* and *j* + 1 (Dawson et al. 1990).

3.2.2. Sentinel-2 data

3.2.2.1. Image acquisition

The sentinel-2 imagery covering the study area was downloaded for free at the Google Earth engine home page https://code.earthengine.google.com (Accessed: 05 August 2022). Five images were downloaded from 2017 to 2022 for the months between January and April (averaged into one image for each year) to correspond to the dates of the field data collection. The year 2018 was omitted due to the imagery having overcast clouds.

Sentinel-2 multispectral satellite is an addition to the already existing remote sensing satellites, it was first launched in 2014 by the European Space Agency (ESA) (Cleversand Gitelson, 2013). The sensor consists of 13 spectral bands, of which 4 bands have 10 m, 6 bands have 20 m and 3 bands have 60 m spatial resolutions, respectively (Clevers and Gitelson, 2013). The 3 bands at 60 m spatial resolution are very crucial for atmospheric corrections (Drusch et al. 2012). The two red-edge bands are at 705 nm and 740 nm with a spatial resolution of 20 m (Clevers and Gitelson, 2013). The sensor has a swath width of 290 km and a temporal resolution of 5 days and a radiometric resolution of 12 bits. Table 3 below summarises the band properties of the Sentinel-2 sensor.

Table 3 Spectral position (λ) and bandwidth (Δλ), spatial resolution, heritage and purpose of Sentinel-2 bands (Druschet al. 2012; Clevers and Gitelson, 2013)

Band	Name	λ(nm)	Δλ (nm)	Res (m)	Heritage	Purpose

B1	Coastal aerosol	443	20	60	Moderate Resolution Imaging Spectroradiometer (MODIS) satellite, Advanced Land Imager (ALI), LandSat-8 (LS8)	Atmospheric correction (aerosol scattering)
B2	Blue	490	65	10	Medium Resolution Imaging Spectrometer (MERIS), LS8, (LandSat-7) LS7	Vegetation senescing, carotenoid, browning and soil background; Atmospheric correction (aerosol scattering)
B3	Green	560	35	10	MERIS, LS8, LS7, SPOT5	Green peak, sensitive to total chlorophyll in vegetation
B4	Red	665	30	10	MERIS, LS8, LS7	Max. chlorophyll absorption
B5	Red edge	705	15	20	MERIS	Red edge position; consolidation of atmospheric corrections/fluorescence baseline.
B6	Red edge	740	15	20	MERIS	Red edge position; atmospheric correction; retrieval of aerosol load
B7	Red edge	783	20	20	MERIS, ALI	LAI; edge of the NIR plateau
B8	NIR	842	115	10	Landsat, SPOT5	LAI
B8a	Red edge	865	20	20	MERIS, ALI, LS8	NIR plateau, sensitive to total chlorophyll, biomass, LAI and protein; water vapour absorption reference; retrieval of aerosol

						load and type
B9	Water vapour	945	20	60	MODIS, MERIS	Atmospheric correction (water vapour absorption)
B10	SWIR-Cirrus	1375	30	60	MODIS, LS8	Atmospheric correction (detection of thin cirrus)
B11	SWIR	1610	90	20	LS8, SPOT5, LS7	Sensitive to lignin, starch and forest above ground biomass; snow/ice/cloud separation
B12	SWIR	2190	180	20	LS8, LS7	Assessment of Mediterranean vegetation conditions; distinction of clay soils for monitoring of soil erosion; distinction between live biomass, dead biomass and soil, e.g. for burn scars mapping

3.2.1.2.Image pre-processing

The imagery downloaded from the Google Earth Engine was already pre-processed (atmospheric and radiometric corrections) and ready to use (Earth Engine Data Catalog, 2022). The imagery was computed using sen2cor and the QA60 bitmask band, which contains cloud information, was used to mask out bad-quality observations caused by opaque and cirrus clouds in each image (Mahdianpari et al. 2018). The remaining pixels were reserved as good-quality observations. The number of good-quality observations in each location of an individual pixel is referred as good-quality observation number.

3.2.1.3.Image processing

ENVI 5.3 was used to create the optimal indices selected for both N and P from the Sentinel-2 bands. The SMLR model was then implemented in R Studio software to predict leaf N and P. ArcGIS 10.6 was used to create maps showing the spatial distribution of Leaf N and P.

3.3. Data analysis

Finding solutions through comprehension of the information gathered to create conclusions that explain the significance of the data is known as data analysis (Strydom et al. 2005). One step in data analysis is to summarize the raw data in order to provide an answer to the research question and reach a conclusion (Strydom et al. 2005). The Shapiro-Wilk normality test was performed to test the normality of the leaf N and P measurements by the ASD in R studio. The Shapiro-Wilk normality test is limited to samples sizes between 3 and 50 (Royston, 1982). In this study, the quantitative data (remotely sensed data) was analyzed in a form of maps. R Studio software was used for statistical analysis and a repeated k-fold cross validation technique was used for validating the models. The optimal bands and vegetation indices that were selected by the model which yielded high estimating accuracy were used to produce the variables (N and P) maps using the ENVI 5.3.

3.3.1. Univariate (Simple Linear Regression)

3.3.1.1.Various bands vs each variable

The Simple Linear Regression (SLR) technique was used to estimate leaf N and P using the bands of Sentinel-2. All the 13 spectral bands (Table 4) of Sentinel-2 were used. The coefficient of determination (R^2), p-value and Root mean square error (RMSE) was identified for each band. R^2 was used to identify the most accurate band to estimate each variable as it measures the correlation between the two variables. R^2 was chosen because it indicates the strength of a liner relationship in terms of the proportion of the variability (Richter et al. 2012).

3.3.1.2.Vegetation indices vs each variable

The following vegetation indices in Table 4 were computed to estimate leaf N and P using SLR technique. The coefficient of determination (R^2), p-value and RMSE was identified for each vegetation index. R^2 was used to identify the most accurate index to estimate each variable.

Table 4 List of vegetation indices used in this study

Index	Original formulae	Revised formulae	Reference
REP	700 + 40 *((R670-R780)/2-R700)/R740-R700)	705 + 35 *((R665-R783)/2-R705)/R740-R705)	Guyot and Baret (1988)
NDRE_1	(RRED_EDGE-RRED-EDGE)/(RREDEDGE+RRED-EDGE)	(R740-R705)/(R740+R705)	Gitelson and Merzlyak (1994), Sims and Gamon (2002)
NDRE_2		(R783-R705)/(R783+R705)	Barness et al. (2000)
MCARI	((RRED-EDGE-RRED-EDGE)-0.2(RRED-EDGE-RGREEN))/(RRED-EDGE/RRED-EDGE)	((R740-R705)-0.2*(R740-R560))*(R740/R705)	Daughtry et al. (2000)
MTCI	(RRED-EDGE-RRED-EDGE)/ (RRED-EDGE-RREED-EDGE)	(R740-R705)/ (R740-R665)	Dash and Curran (2004)
NRI	(RGREEN – RRED)/(RGREEN+RRED)	(R560 – R665)/(R560+R665)	Schleicher et al. (1998)
SAVI_1	((1+*L*)*RNIR-RRED)/((RNIR-RRED)+*L*)	((1+0.2)*R865-R665)/((R865R665)+0.2)	Huete (1988)
SAVI_2		((1+0.2)*R842-R665)/((R842R665)+0.2)	Huete (1988)
OSAVI	(1+0.16)*(RNIRRRED)/(RNIR+RRED+0.16)	(1+0.16)*(R842R665)/(R842+R665+0.16)	Rondeaux et al. (1996)
OSAVI1		(1+0.16)*(R865-R665)/(R865+R665+0.16)	Rondeaux et al. (1996)
OSAVI2	(1+0.16)*(RNIR-Rred-edge)/(RNIR+RRED-EDGE+0.16)	(1+0.16)*(R842-R705)/(R842+R705+0.16)	Wu et al. (2008)
OSAVI3		(1+0.16)*(R842R740)/(R842+R740+0.16)	Wu et al. (2008)
OSAVI4		(1+0.16)*(R865-R705)/(R865+R705+0.16)	Wu et al. (2008)
OSAVI5		(1+0.16)*(R865-R740)/(R865+R740+0.16)	Wu et al. (2008)

NDVI_1	$(R_{NIR}-R_{RED})/ (R_{NIR}+R_{RED})$	$(R_{842}-R_{665})/ (R_{842}+R_{665})$	Gitelson et al. (1996)
NDVI_2		$(R_{865}-R_{665})/ (R_{865}+R_{665})$	Gitelson et al. (1996)
GI	R_{GREEN}/R_{RED}	R_{560}/R_{665}	Smith et al. (1995)
TCARI	$3*((R_{700}-R_{670})-0.2*(R_{700}-R_{550})*(R_{700}/R_{670}))$	$3*((R_{740}-R_{705})-0.2*(R_{740}R_{560})*(R_{740}/R_{705}))$	Haboudane et al. (2002)
SR1	R_{NIR}/R_{RED}	R_{842}/R_{665}	Jordan (1969)
SR2		R_{865}/R_{665}	Jordan (1969)
SR705	R_{750}/R_{705}	R_{740}/R_{705}	Gitelson and Merzlyak (1994)

3.3.2. Multivariate regression

3.3.2.1. Stepwise Multiple Linear Regression (SMLR) implementation

SMLR is among the feature selection techniques which reduce the dimensionality of hyperspectral data (Clevers et al. 2007). SMLR was used to predict models for the independent variables (N and P). The model tested the applicability of all thirteen bands, twenty-one vegetation indices and a combination of all bands and vegetation indices. The process was done in R Studio. The Akaike Information Criterion (AIC) value, coefficient of determination (R^2), P-value and RMSE were identified for each model. The best performing model was selected using the lowest AIC (Akaike Information Criterion) value (Sakamoto et al. 1986). Repeated k-fold cross validation with 10 folds and 5 repeats were used to analyse the performance of this model by assessing model accuracy (Kim, 2009).

3.3.2.2. Random Forest (RF) implementation

RF method produces the measures of variable's importance (Grimm et al. 2008). RF draws *ntree* bootstrap samples and for each sample, it grows un-pruned treed by choosing the best split based on a random sample of *mtry* predictors at each node (Grimmet al. 2008). This assists

in identifying which indices or bands plays a great role in prediction. The *ntree* was tested at 100 and *mtry* values for all thirteen wavebands and twenty-one vegetation indices. RF ranked the wavebands and the vegetation indices according to their predictor importance. RF method was executed on R Studio to randomly select a set of variables from the training dataset (70%). Repeated k-fold cross validation with 10 folds and 5 repeats was used to analyze the performance of this model by assessing model accuracy (Kim, 2009; Richter et al. 2012). The 10 folds and 5 repeats were previously proven by Jiang and Wang (2017) to yield higher results.

3.3.2.3.Support Vector Machines (SVMs) implementation

SVM is relatively easy to implement and interpret (Wang, 2005). For the purpose of this study, the support vector machine recursive feature elimination (SVM-RFE) algorithm was used implemented in R Studio to select optimal bands and indices for predicting leaf N and P. Repeated k-fold cross validation with 10 folds and 5 repeats was used to analyze the performance of this model by assessing model accuracy (Kim, 2009).

3.3.3. Selection of the best model to predict leaf N and P

The best prediction model was chosen based on the RMSE to predict leaf N and P in KNP. RMSE indicates the magnitude or error in % (McKeen et al. 2005; Chai and Draxler, 2014). Relative root mean square error (RRMSE) is the RMSE divided by the mean of the observe variables and is a good indicator than RMSE because it gives a great indication of the error without being affected by the data unit (Ritchter et al. 2012). The mean absolute error (MAE) also assisted in evaluating model performance (Chai and Draxler, 2014). The optimal bands and indices were then used to produce the regional maps of leaf N and P.

CHAPTER 4: RESULTS

The findings of the multivariate analysis, which were based on SMLR, SVM, and RF, as well as the univariate analysis, which were based on simple regression, are presented in this chapter. Maps of the spatial distribution of Leaf N and P using the most effective model for

each variable are also included in the results. Prediction models with the highest accuracy were also developed to estimate each variable (N and P).

4.1. Descriptive statistics

Descriptive statistics for Leaf N and P are shown in Table 5 below, the mean for leaf N and P were 0.71(%) and 0.10 (%), respectively. The minimum values for N and P were 0.34 and 0.04, respectively, while the maximum values were 1.06 and 0.29, respectively.The Shapiro-Wilk normality test showed that the distribution of N was normal ($p = 0.63$ and $W = 0.98$) whereas P distribution was not normal ($p = 0.00$ and $W = 0.89$). The histograms in Figure 2 below shows that N is normally distributed and P is right skewed. Figure 3 and 4 show the spectral reflectance graph of dried leaf samples before and after applying the S-G filter respectively.

Table 5 Descriptive statistics for N and P

Variables	Number of observations	Minimum	Maximum	Mean	Standard Deviation	Coefficient of variation
Nitrogen (N %)	49	0.3425	1.0633	0.7075	0.1849	0.26
Phosphorus (P %)	49	0.0415	0.2903	0.1047	0.0510	0.49

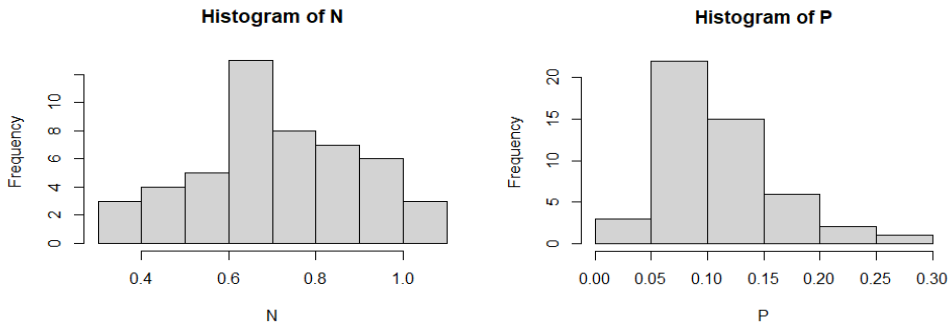


Figure 2 Histograms showing the distribution of N and P, respectively

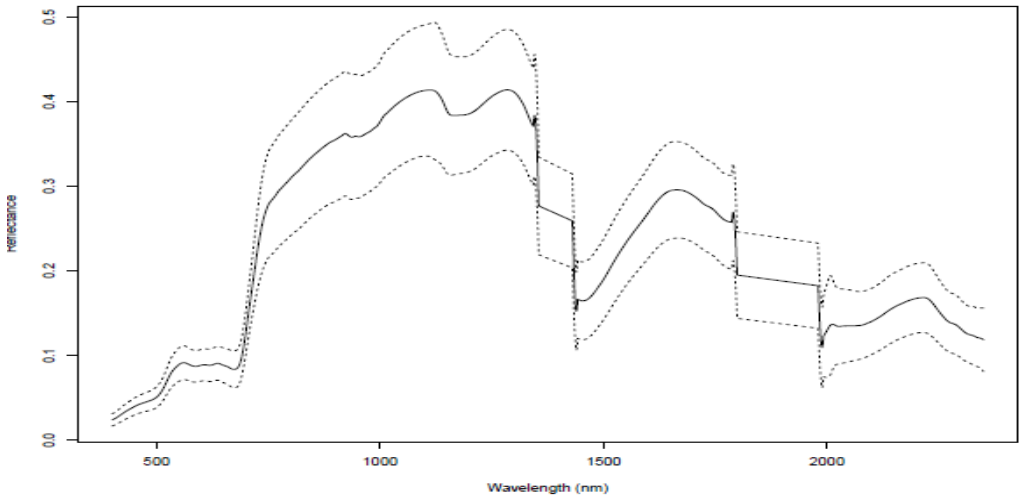


Figure 3 Spectral reflectance graph of leaf samples before applying the S-G filter

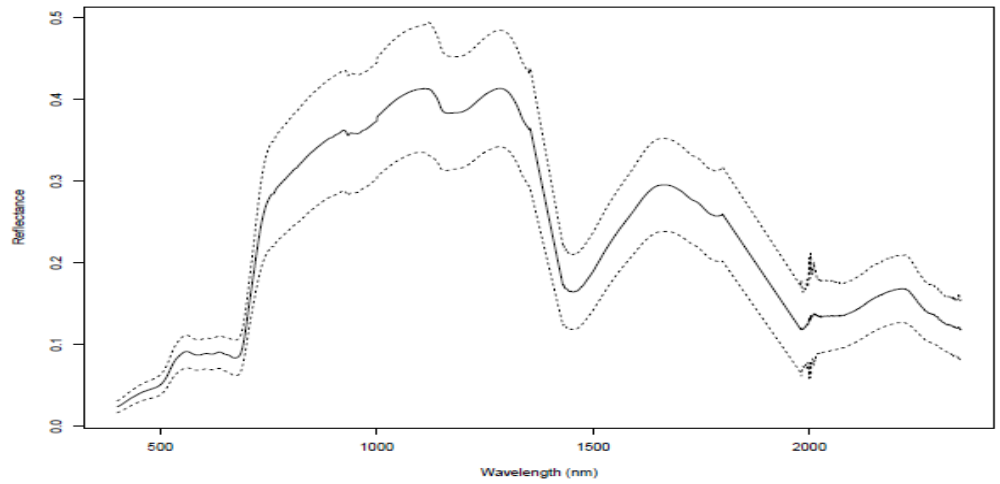


Figure 4 Spectral reflectance graph of leaf samples after noise removal (S-G filter)

4.2. Univariate analysis results: Simple linear regression (SLR)

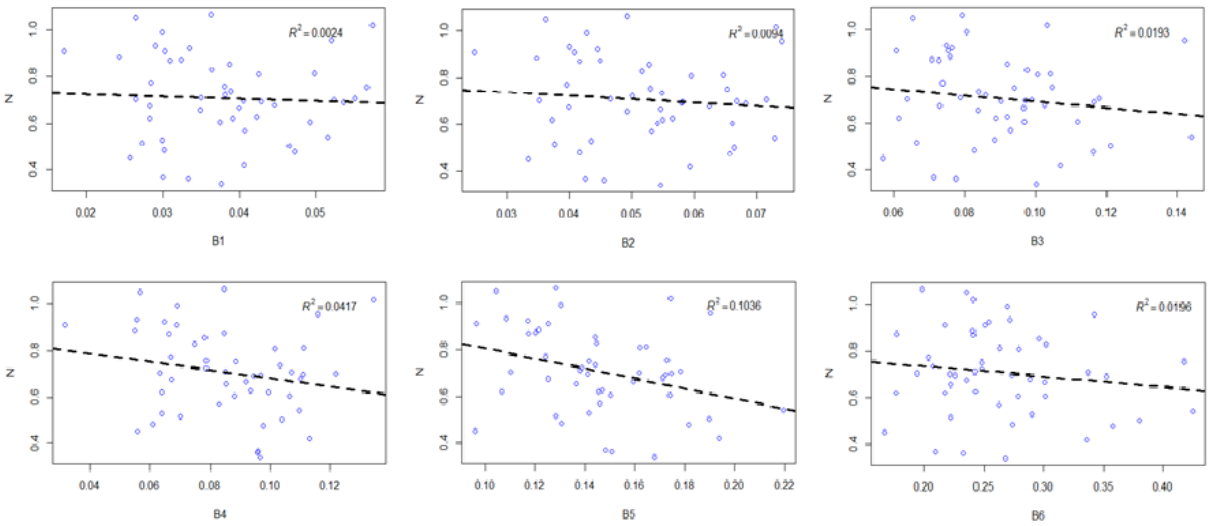
4.2.1. Results for estimating Leaf N using various bands

Table 6 shows the simple linear regression analysis result for leaf N versus various bands. These results showed that the red edge band at 705 nm (B5) yielded higher accuracy than other bands for predicting leaf N (R^2 = 0.1036, P-value = 0.0241, RMSE = 0.0258 %). This shows the importance of the red edge in estimating leaf N. Figure 5 illustrates the scatter plots of simple linear regression using different bands against leaf N. These plots below show that the relationship between bands and leaf N is poor.

Table 6 Results for simple linear regression for each band against leaf N

Band No	R^2	P-value	RMSE(%)
B1 – Coastal aerosol	0.0024	0.7366	0.0095
B2 - Blue	0.0094	0.5082	0.0120
B3 - Green	0.0193	0.3416	0.0192
B4 - Red	0.0417	0.1595	0.0206
B5 - Red edge	0.1036	0.0241	0.0258
B6 – Red edge	0.0196	0.3369	0.0570
B7 - Red edge	0.0164	0.3802	0.0652
B8 - NIR	0.0185	0.3508	0.0682
B8A – Red edge	0.0196	0.3380	0.0698
B9 – water vapour	0.0236	0.2919	0.0713
B10 - SWIR	0.0301	0.2330	0.0581
B11 - SWIR	0.0105	0.4826	0.0537
B12 - SWIR	0.0030	0.7105	0.0389

The values of R^2 which are closer to 1 indicate good leaf N estimation models whereas the values close to 0 indicate poor leaf N estimation models. P-values that are less than 0.05 means the leaf N estimation model is good at 95 % significance level. Low values of RMSE means the estimation model is good.



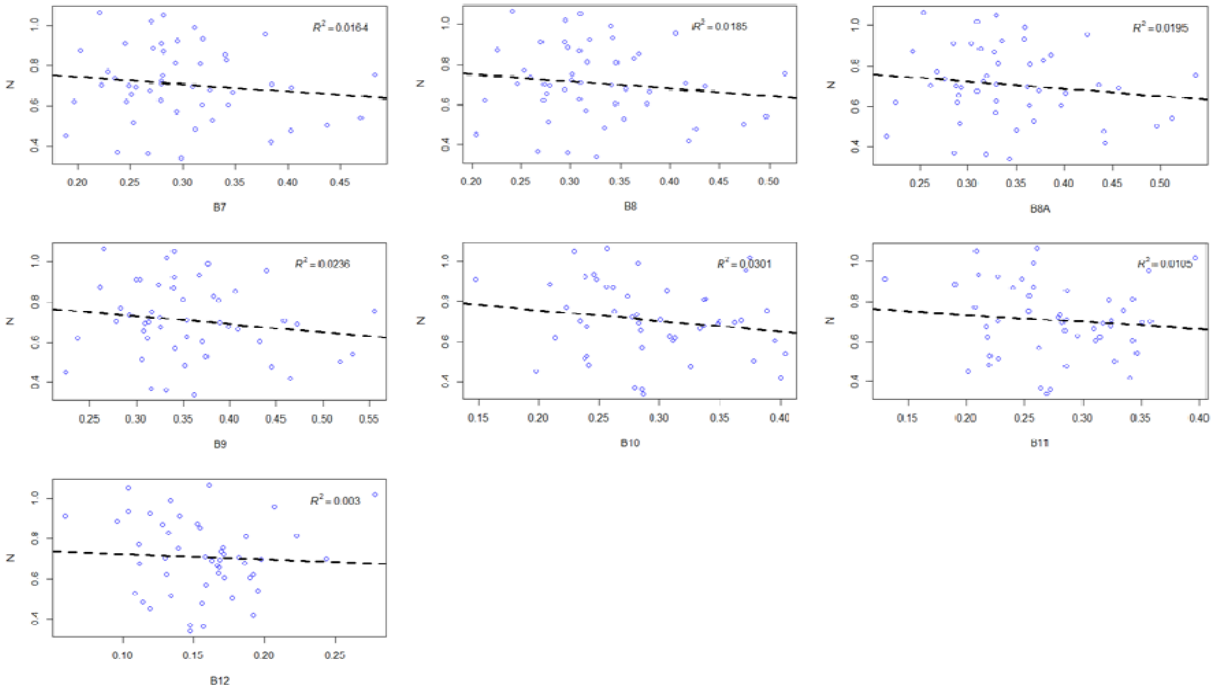


Figure 5 Scatter plot of different bands and leaf N (%) obtained from simple linear regression. X-axis = different bands, Y-axis=N

4.2.2. Results for estimating Leaf N using various vegetation indices

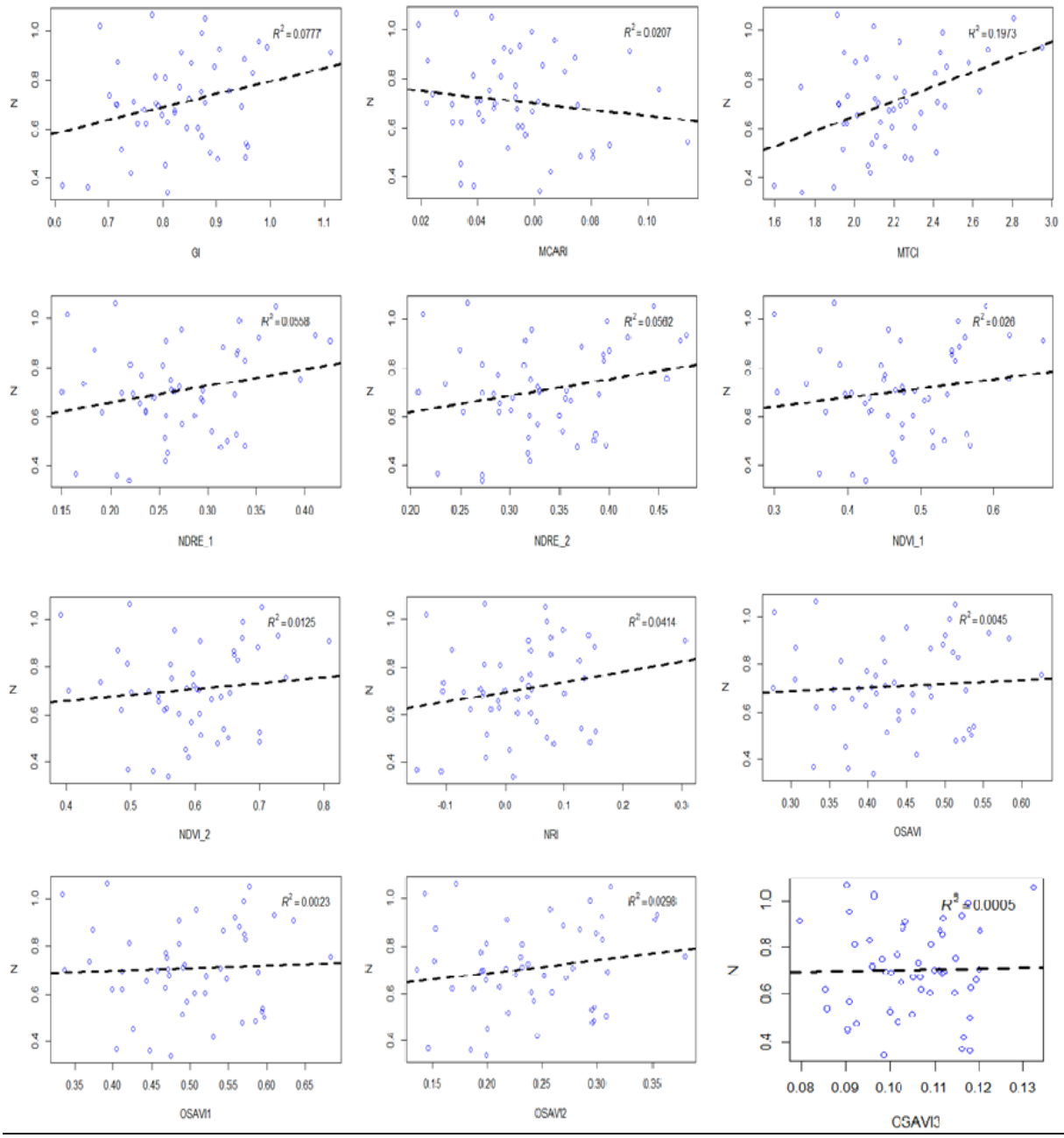
The results of the simple linear regression in Table 7 below showed that the vegetation index which yielded the highest accuracy in estimating leaf N was based on the red edge and is called MTCI (R^2 = 0.1973, P-value = 0.0014, RMSE = 0.2408%), the second index which yielded the highest accuracy was the REP (R^2 = 0.1720, P-value = 0.0030,RMSE = 1.1010%) which is also based on the red edge bands. This emphasized the importance of the red edge in estimating leaf N. Figure 6 illustrates the scatter plots of simple linear regression using various vegetation indices against leaf N. These plots show that the relationship between vegetation indices and leaf N is generally poor.

Table 7 Results for simple linear regression for leaf N against various vegetation indices.

Vegetation Index	R^2	P-value	RMSE (%)

GI	0.0777	0.0525	0.0935
MCARI	0.0207	0.3235	0.0204
MTCI	0.1973	0.0014	0.2408
NDRE_1	0.0558	0.1023	0.0623
NDRE_2	0.0562	0.1009	0.0632
NDVI_1	0.0260	0.2681	0.0791
NDVI_2	0.0125	0.4445	0.0843
NRI	0.0414	0.1606	0.0852
OSAVI	0.0045	0.6476	0.0805
OSAVI1	0.0023	0.7460	0.0788
OSAVI2	0.0298	0.2357	0.0577
OSAVI3	0.0005	0.8812	0.0111
OSAVI4	0.0279	0.2513	0.0550
OSAVI5	0.0001	0.9580	0.0128
REP	0.1720	0.0030	1.1010
SAVI_1	0.0001	0.9465	0.0756
SAVI_2	0.0009	0.8423	0.0754
SR_1	0.0512	0.1179	0.6210
SR_2	0.0425	0.1555	1.2228
SR705	0.0669	0.0727	0.1786
TCARI	0.0307	0.2288	0.0273

The values of R^2 which are closer to 1 indicate good leaf N estimation models whereas the values close to 0 indicate poor leaf N estimation models. P-values that are less than 0.05 means that the leaf N estimation model is good at 95 % significance level. Low values of RMSE means the estimation model is good.


R² = 0.0777
N
GI
R² = 0.0207
MCARI
R² = 0.1973
MTCI
R² = 0.0588
NDRE_1
R² = 0.0582
NDRE_2
R² = 0.026
NDVI_1
R² = 0.0125
NDVI_2
R² = 0.0414
NRI
R² = 0.0045
OSAVI
R² = 0.0023
OSAVI1
R² = 0.0298
OSAVI2
R² = 0.0005
OSAVI3

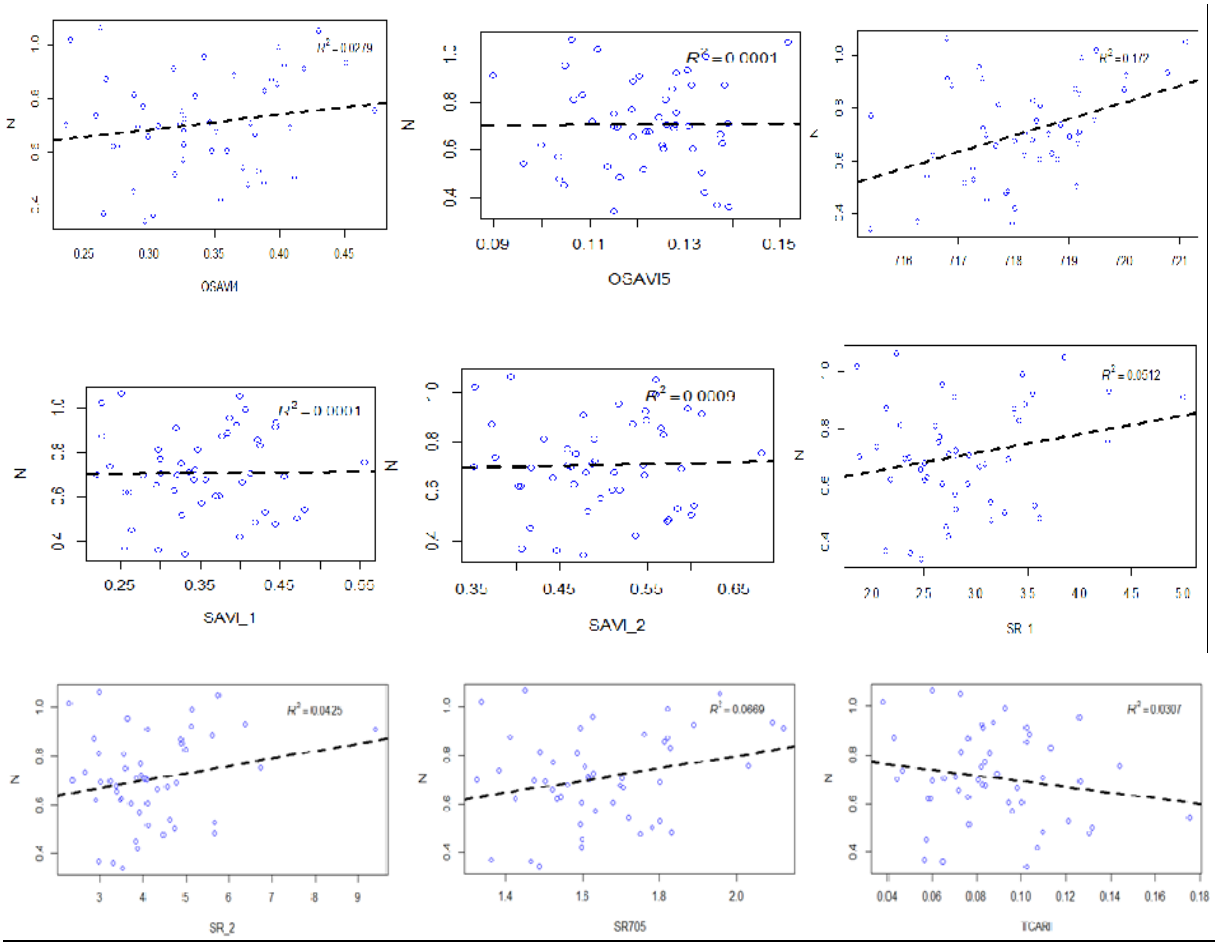


Figure 6 Scatter plot of different vegetation indices and leaf N (%) obtained from simple linear regression. X - axis = different vegetation indices, Y – axis=N

4.2.3. Results for estimating Phosphorus (P) using various bands

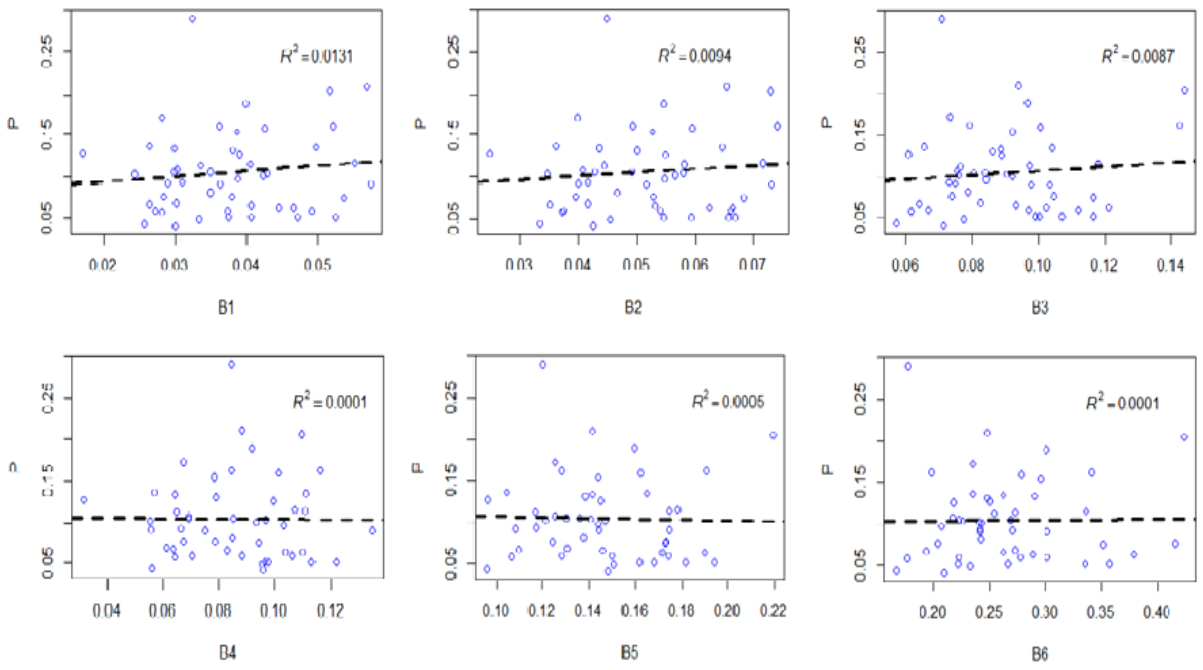
Table 8 below shows the simple linear regression analysis result for P versus various bands. These results showed that the blue band at 443 nm (B1) yielded the highest accuracy for predicting P (R^2 = 0.0131, P-value = 0.4329, RMSE = 0.0094%). Figure 7 illustrates the scatter plots of simple linear regression using different bands and P. These plots below show that the relationship between bands and P is very poor.

Table 8 Results for simple linear regression for each band against P.

Band No	R^2	P-value	RMSE (%)
B1	0.0131	0.4329	0.0094
B2	0.0094	0.5069	0.0120

B3	0.0087	0.5236	0.0193
B4	0.0001	0.9463	0.0210
B5	0.0005	0.8786	0.0273
B6	0.0001	0.9382	0.0576
B7	0.0000	0.9835	0.0657
B8	0.0000	0.9826	0.0688
B8A	0.0000	0.9718	0.0705
B9	0.0005	0.8735	0.0721
B10	0.0005	0.8732	0.0590
B11	0.0000	0.9801	0.0540
B12	0.0000	0.9854	0.0390

The values of R^2 which are closer to 1 indicate good P estimation models whereas the values close to 0 indicate poor P estimation models. P-values that are more than 0.05 means the P estimation model is poor at 95 % significance level. Low values of RMSE means the estimation model is good.



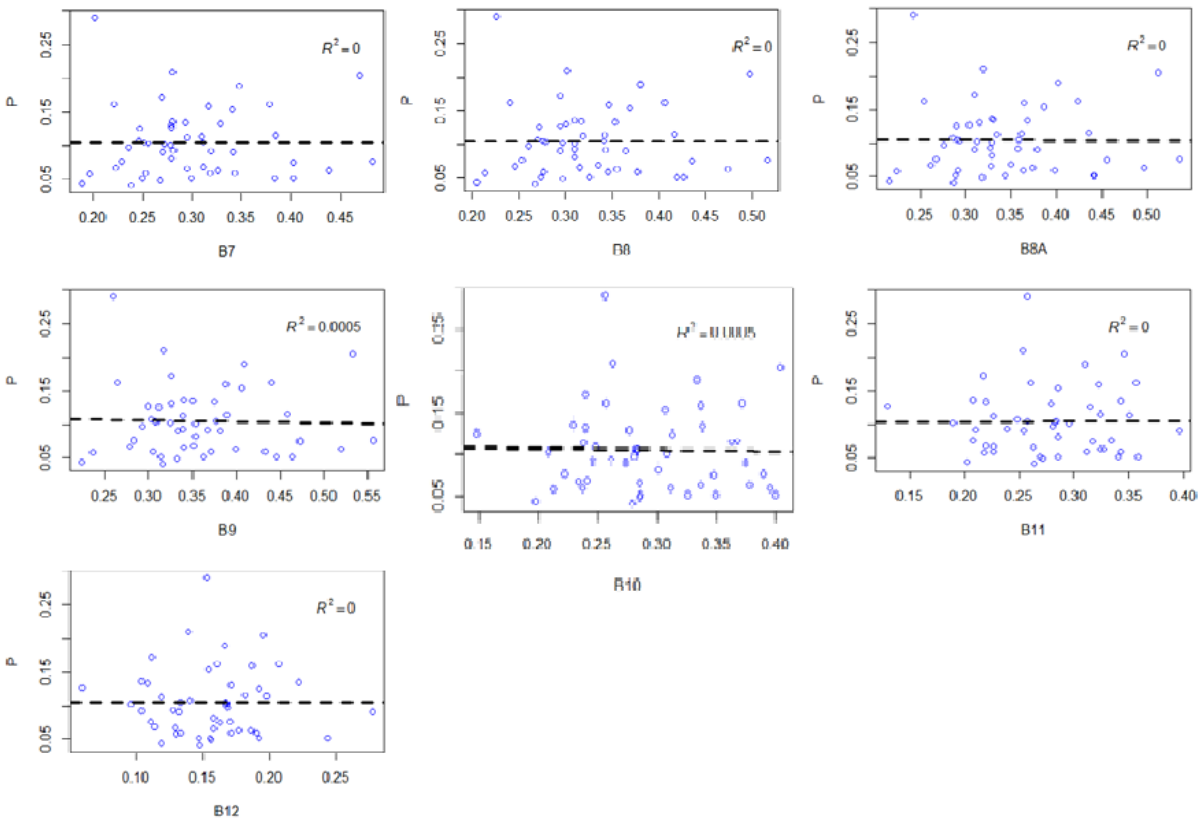


Figure 7 Scatter plot of different bands and leaf P (%) obtained from simple linear regression. X-axis = different bands, Y-axis=P

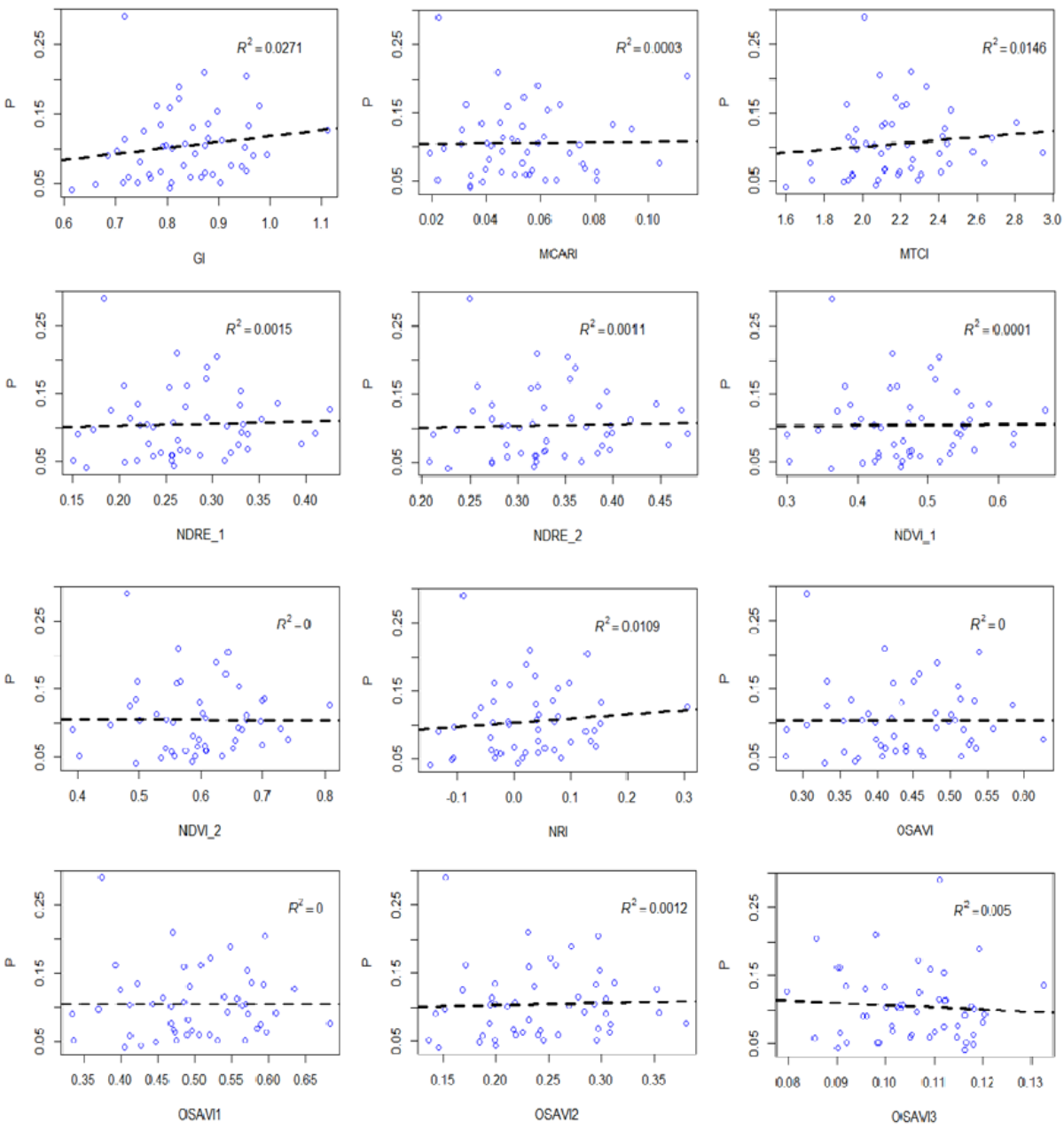
4.2.4. Results for estimating Phosphorus (P) using various vegetation indices

The results of the simple linear regression in Table 9 below showed that the vegetation index yielded the highest accuracy in estimating P was the red edge based index called the REP (R^2 = 0.0161, p-value = 0.3852, RMSE = 1.2002%). The second index which yielded the highest accuracy was based on the red edge and is called MTCI (R^2 = 0.0146, p-value = 0.4080, RMSE = 0.2668%). The results show the importance of the red edge in estimating phosphorus. Figure 9 illustrates the scatter plots of simple linear regression using various vegetation indices and P. These plots below show that the relationship between vegetation indices and P is poor.

Table 9 Results for simple linear regression for P against various vegetation indices.

Vegetation Index	R^2	P-value	RMSE (%)
GI	0.0271	0.2586	0.0960
MCARI	0.0003	0.8985	0.0206
MTCI	0.0146	0.4080	0.2668
NDRE_1	0.0015	0.7907	0.0640
NDRE_2	0.0011	0.8215	0.0650
NDVI_1	0.0001	0.9501	0.0801
NDVI_2	0.0000	0.9746	0.0848
NRI	0.0109	0.4753	0.0865
OSAVI	0.0000	0.9788	0.0807
OSAVI1	0.0000	0.9942	0.0788
OSAVI2	0.0012	0.8131	0.0585
OSAVI3	0.0050	0.6293	0.0111
OSAVI4	0.0006	0.8711	0.0558
OSAVI5	0.0038	0.6759	0.0128
REP	0.0161	0.3852	1.2002
SAVI_1	0.0000	0.9750	0.0756
SAVI_2	0.0000	0.9938	0.0754
SR_1	0.0002	0.9274	0.6375
SR_2	0.0001	0.9356	1.2495
SR705	0.0012	0.8154	0.1848
TCARI	0.0016	0.7876	0.0277

The values of R^2 which are closer to 1 indicate good P estimation models whereas the values close to 0 indicate poor P estimation models. P-values that are more than 0.05 means the P estimation model is poor at 95 % significance level. Low values of RMSE means the estimation model is good.


P
GI
$R^2 = 0.0271$
MCARI
$R^2 = 0.0003$
MTCI
$R^2 = 0.0146$
NDRE_1
$R^2 = 0.0015$
NDRE_2
$R^2 = 0.0011$
NDVI_1
$R^2 = 0.0001$
NDVI_2
$R^2 = 0$
NRI
$R^2 = 0.0109$
OSAVI
$R^2 = 0$
OSAVI1
$R^2 = 0$
OSAVI2
$R^2 = 0.0012$
OSAVI3
$R^2 = 0.005$

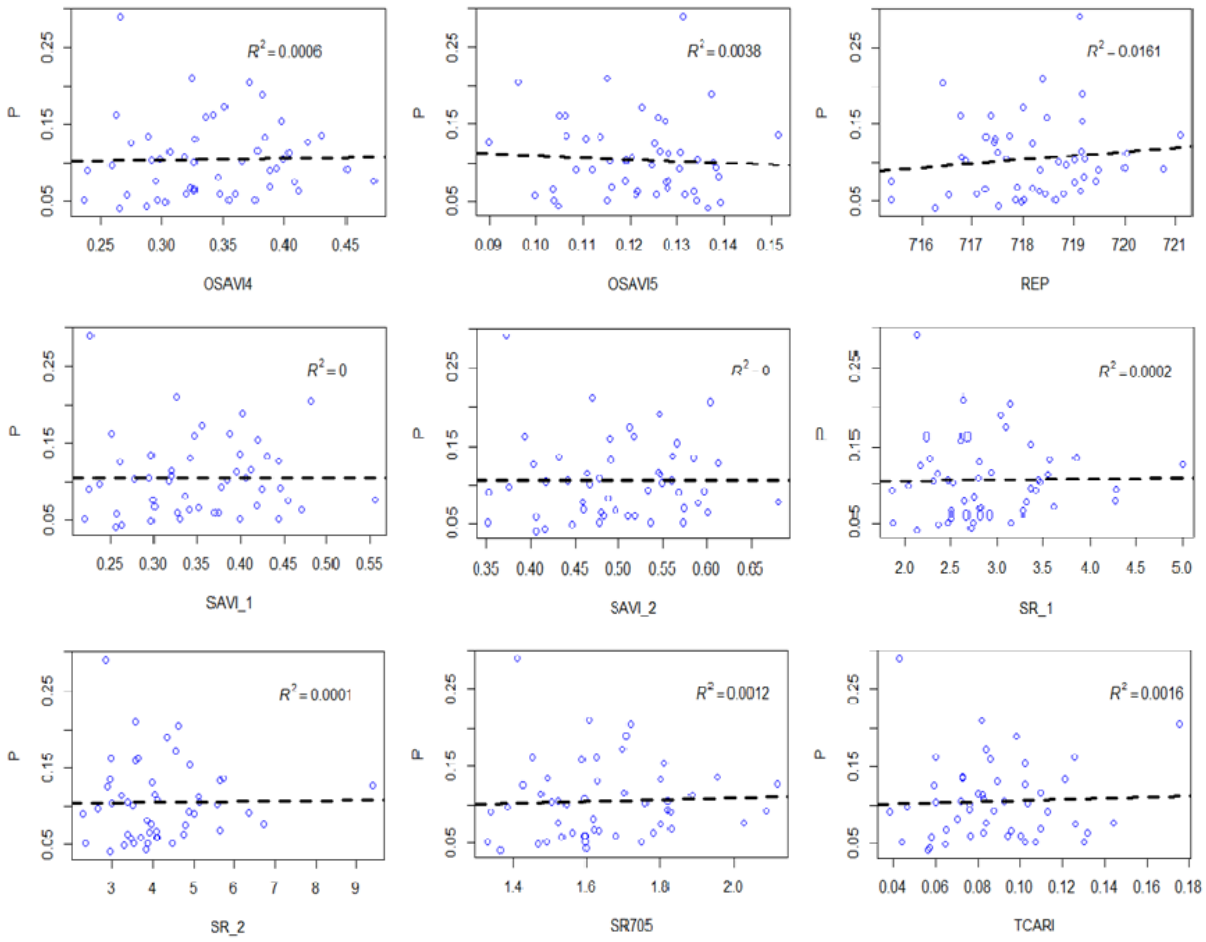


Figure 8 Scatter plot of different vegetation indices and P (%) obtained from simple linear regression. X - axis = different vegetation indices, Y-axis=P

4.3. Multivariate results for estimating leaf N and P

The machine learning techniques SMLR, SVM and RF were used to determine the best model in using the optimal bands and vegetation indices to predict leaf N and P. The AIC and the coefficient of determination (R^2) were used to select the best model which performed well in predicting leaf N and P concentration.

4.3.1. SMLR results for estimating leaf N and P

The best model using SMLR was selected based on the lowest AIC value (Table 10). According to the model the optimal bands for estimating leaf N are; bands 4, 5, 6, 7, 8,10 and 12. The overall model yielded AIC = -198.48, R^2 = 0.63 and RMSE = 0.11%. The optimal VIs for estimating leaf N are Modified Chlorophyll Absorption in Reflectance Index (MCARI),

MERIS Terrestrial Chlorophyll Index (MTCI), NDRE_2, NDVI_1, NDVI_2, OSAVI, OSAVI1, OSAVI3, OSAVI5, REP, SAVI_1, SAVI_2, SR_1, SR_2 and SR705. The overall model yielded AIC = -215.65, R^2 = 0.81 and RMSE = 0.08%. The optimal combination of bands and VIs for estimating leaf N are B1, B2, B3, B5, B6, B7, B8, B8A, B9, B11, GI, MCARI, MTCI, NDRE_1, NDRE_2, NDVI_1, NDVI_2, OSAVI, OSAVI1, OSAVI3, OSAVI4, OSAVI5, REP, SAVI_2, SR_1, SR_2 and SR705. The overall model yielded AIC= -242.37, R^2= 0.93 and RMSE= 0.05%.

The optimal bands for estimating P using the SMLR model are; bands 2, 5, 6, 7 and 8A. The overall model yielded AIC = -289.48, R^2 = 0.17 and RMSE = 0.05%. The optimal vegetation indices for estimating P are GI, NDRE_2, NDVI_1, NDVI_2, NRI, OSAVI2, OSAVI4, OSAVI5, REP, SAVI_1, SAVI_2 and TCARI. The overall model yielded AIC = -299.92, R^2 = 0.49 and RMSE = 0.04%. The optimal combination of bands and VIs for estimating P are B1, B4, B6, B7, B8, B8A, B9, B11, MCARI, NDRE_1, NDRE_2, NDVI_1, NDVI_2, NRI, OSAVI, OSAVI4, OSAVI5, SAVI_2 and TCARI. The overall model yielded AIC = -305.12, R^2 = 0.66 and RMSE = 0.03%.

Table 10 Summary of SMLR results for estimating leaf N and P

Variable	Scenario	Optimal bands/Vis	AIC	P-value <0.05	R^2	RMSE (%)
N	Bands	B4, B5, B6,B7,B8,B10 and B12	-198.48	Yes	0.6250	0.1121
	VIs	MCARI, MTCI, NDRE_2, NDVI_1, NDVI_2, OSAVI, OSAVI1, OSAVI3, OSAVI5, REP, SAVI_1, SAVI_, SR_1, SR_2 and SR705	-215.65	Yes	0.8095	0.0799

	Bands + VIs	B1, B2, B3, B5, B6, B7, B8, B8A, B9, B11, GI, MCARI, MTCI, NDRE_1, NDRE_2, NDVI_1, NDVI_2,OSAVI, OSAVI1, OSAVI3, OSAVI4, OSAVI5, REP, SAVI_2, SR_1, SR_2 and SR705.	-242.37	Yes	0.9323	0.0476
P	Bands	B2, B5, B6, B7 and B8A	-289.48	No	0.1661	0.0461
	VIs	GI, NDRE_2, NDVI_1, NDVI_2, NRI, OSAVI2, OSAVI4, OSAVI5, REP, SAVI_1, SAVI_2 and TCARI	-299.92	Yes	0.4935	0.0359
	Bands + VIs	B1, B4, B6, B7, B8, B8A, B9, B11, MCARI, NDRE_1, NDRE_2, NDVI_1, NDVI_2, NRI, OSAVI, OSAVI4, OSAVI5, SAVI_2 and TCARI	-305.12	Yes	0.6577	0.0296

The repeated k-fold cross validation results on Table 11 below confirm that the best model for estimating N using SMLR is a combination of bands and indices which yielded R^2 = 0.6931, RMSE = 0.1424%, RRMSE = 6.73% and MAE = 0.1122. Whereas the best model for estimating P is a combination on vegetation indices which yielded R^2 = 0.40, RMSE = 0.05%, RRMSE = 34.32% and MAE = 0.04.

Table 11 Cross-validated results for estimating leaf N and P using SMLR

Variable	Scenario	Optimal bands/VIs	R^2	RMSE (%)	RRMSE (%)	MAE
N	Bands	B4, B5, B6,B7,B8,B10 and B12	0.5916	0.1298	15.842	0.1102
	VIs	MCARI, MTCI, NDRE_2, NDVI_1, NDVI_2, OSAVI,	0.6864	0.1465	11.293	0.1167

		OSAVI1, OSAVI3, OSAVI5, REP, SAVI_1, SAVI_2, SR_1, SR_2 and SR705				
	Bands + VIs	B1, B2, B3, B5, B6, B7, B8, B8A, B9 and B11, GI, MCARI, MTCI, NDRE_1, NDRE_2, NDVI_1, NDVI_2,OSAVI, OSAVI1, OSAVI3, OSAVI4, OSAVI5, REP, SAVI_2, SR_1, SR_2, SR705.	0.6931	0.1424	6.73	0.1122
P	Bands	B2, B5, B6, B7 and B8A	0.3506	0.0480	44.041	0.0382
	VIs	GI, NDRE_2, NDVI_1, NDVI_2, NRI, OSAVI2, OSAVI4, OSAVI5, REP, SAVI_1, SAVI_2 and TCARI	0.4013	0.0468	34.322	0.0399
	Bands + VIs	B1, B4, B6, B7, B8, B8A, B9, B11, MCARI, NDRE_1, NDRE_2, NDVI_1, NDVI_2, NRI, OSAVI, OSAVI4, OSAVI5, SAVI_2 and TCARI	0.3878	0.0471	28.214	0.0399

4.3.2. Random Forest regression results for estimating leaf N and P

The RF regression results on Table 12 below show that the best model for estimating leaf N was a combination of bands only (B4, B5, B11 and B12) which yielded R^2 = 0.48, RMSE = 0.16%, RRMSE = 10.72% and MAE = 0.13. Whereas the best model for estimating P was a

combination of two VIs (NRI and NDRE_2) which yielded R^2 = 0.36, RMSE = 0.05%, RRMSE = 25.98% and MAE = 0.04.

Table 12 Cross-validated results for random forest.

Variable	Scenario	Optimal bands/VIs	R^2	RMSE (%)	RRMSE (%)	MAE
N	Bands	B4, B5, B11 and B12	0.4832	0.1554	10.723	0.1279
	VIs	REP	0.3874	0.1583	12.025	0.1343
	Bands + VIs	B5, B12, REP and MTCI	0.3864	0.1590	10.419	0.1340
P	Bands	B2, B4, B5 and B10	0.2282	0.0500	23.961	0.0424
	VIs	NRI and NDRE_2	0.3564	0.0531	25.976	0.0432
	Bands + VIs	B1, B2, B4, B5 and NRI	0.2555	0.0526	23.819	0.0440

4.3.3. Support vector machines (SVMs) results for estimating leaf N and P

The SVMs results on Table 13 below show that the best model for estimating leaf N was a combination of bands only (B5, B12, B4 and B11) which yielded R^2 = 0.14, RMSE = 0.18%, RRMSE = 17.54% and MAE = 0.15. Similarly, the best model for estimating P was a combination of bands only (B5, B2 and B4) which yielded R^2 = 0.08, RMSE = 0.06 (%), RRMSE = 40.78 (%) and MAE = 0.04.

Table 13 Cross-validated results for SVM

Variable	Scenario	Optimal bands/VIs	R^2	RMSE (%)	RRMSE (%)	MAE
N	Bands	B5, B12, B4 and B11	0.1432	0.1838	17.543	0.1532
	VIs	MCARI, NDRE_2, SAVI_2, MTCI and REP	0.0535	0.1899	17.543	0.1555
	Bands + VIs	B5, REP, MTCI,	0.0625	0.1875	18.119	0.1519

		NDRE_2 and B12				
P	Bands	B5, B2 and B4	0.0805	0.0567	40.783	0.0440
	VIs	GI, NRI, SR_2 and NDVI_1	0.0577	0.0497	45.681	0.0373
	Bands + VIs	B2	0.0550	0.0581	48.517	0.0446

4.4. Leaf N and P maps derived from Sentinel-2 bands and VIs

The spatial distribution of Leaf N (%) was demonstrated in Figure 9, 10, 11, 12 and 13 for the years 2017, 2019, 2010, 2021 and 2022, respectively. These maps were based on the SMLR model using a combination of the following bands and VIs; B1, B2, B3, B5, B6, B7, B8, B8A, B9 and B11, GI, MCARI, MTCI, NDRE_1, NDRE_2, NDVI_1, NDVI_2, OSAVI, OSAVI1, OSAVI3, OSAVI4, OSAVI5, REP, SAVI_2, SR_1, SR_2 and SR705. The concentration of leaf N is depicted as low, moderate and high on the maps. Figure 9-13 below depict that the spatial distribution of leaf N is evenly distributed in the eastern and western sides of the study area, however, it increases on the eastern side and decreases on the western side from 2017 to 2022.

The spatial distribution of phosphorus (%) was demonstrated in Figure 14, 15, 16, 17 and 18 for the year 2017, 2019, 2010, 2021 and 2022, respectively. These maps were based on the SMLR model using a combination of the following VIs GI, NDRE_2, NDVI_1, NDVI_2, NRI, OSAVI2, OSAVI4, OSAVI5, REP, SAVI_1, SAVI_2 and TCARI. The concentration of phosphorus is depicted as low, moderate and high on the maps. The Figure 14-18 below depict that the spatial distribution of phosphorus low in the eastern and high in the western side of the study area in 2017, the distribution reduces in 2019 and becomes moderate in 2020 and 2021. In 2022 the distribution of phosphorus becomes low again in the eastern side of the park.

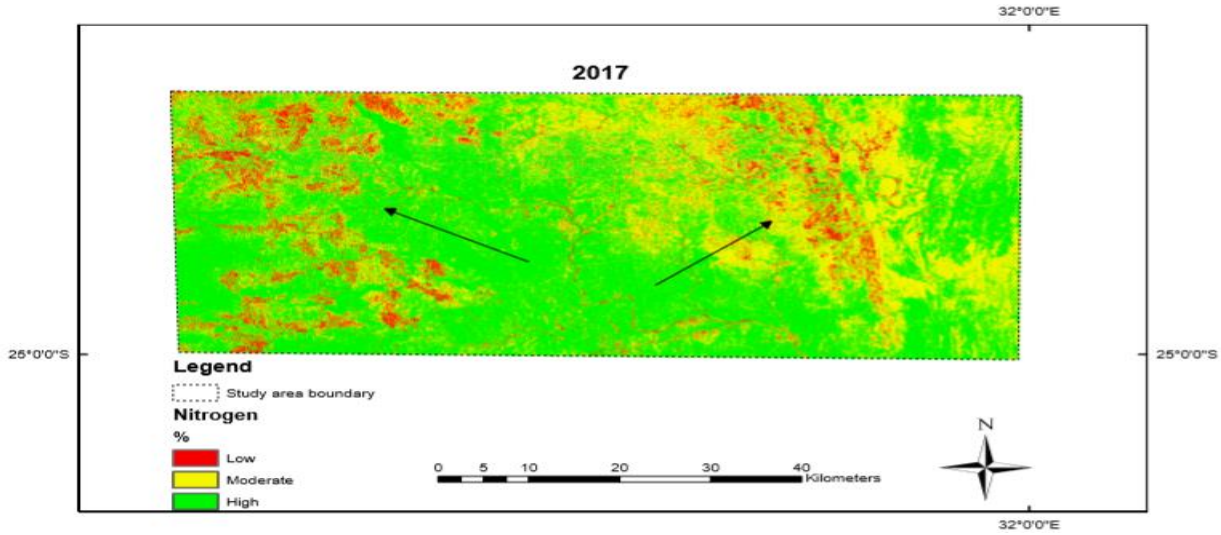


Figure 9 Spatial distribution of Leaf N in 2017 at the Savanna ecosystem for the wet season

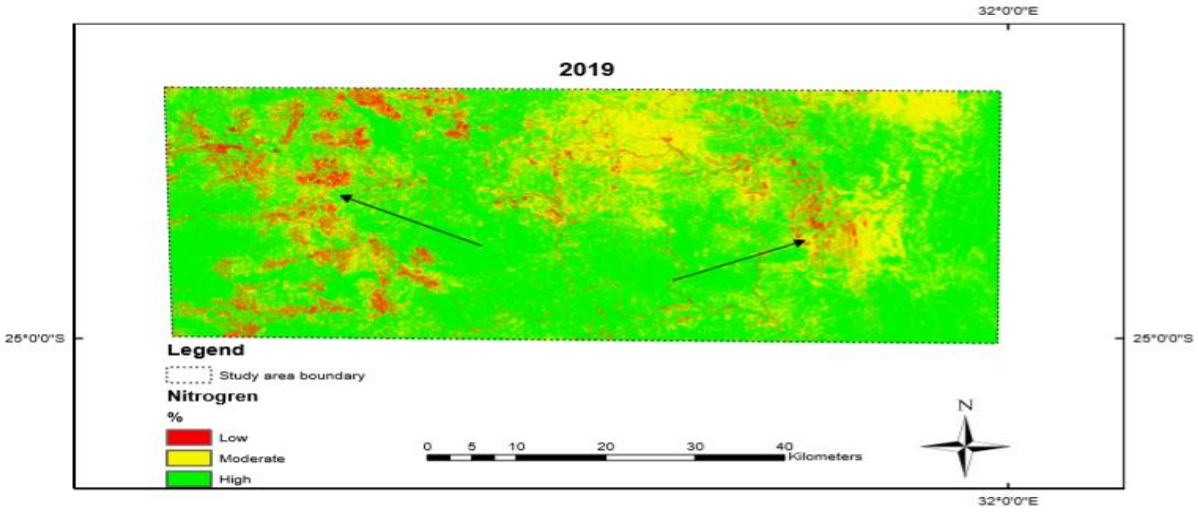


Figure 10 Spatial distribution of Leaf N in 2019 at the Savanna ecosystem for the wet season

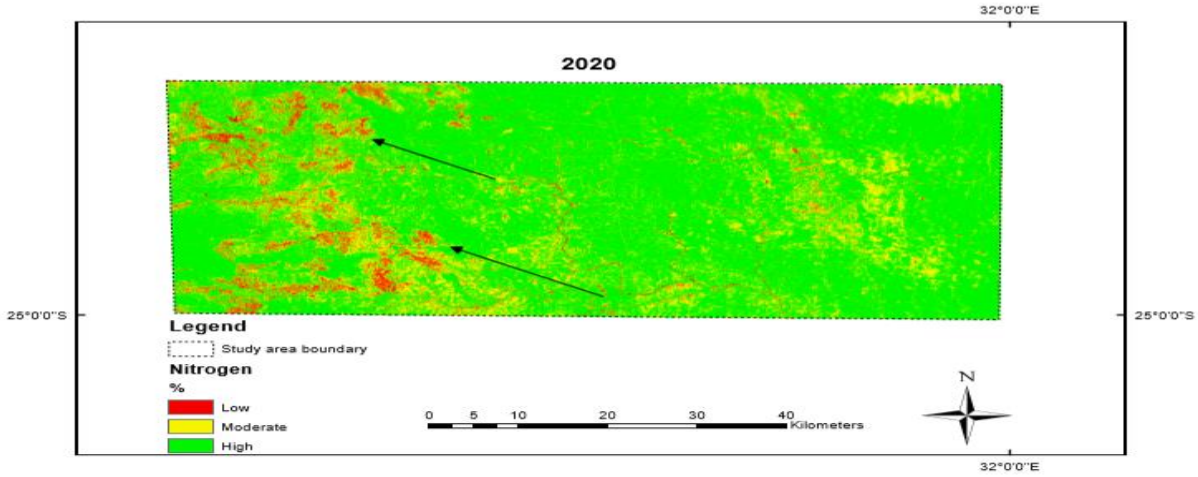


Figure 11 Spatial distributions of Leaf N in 2020 at the Savanna ecosystem for the wet season

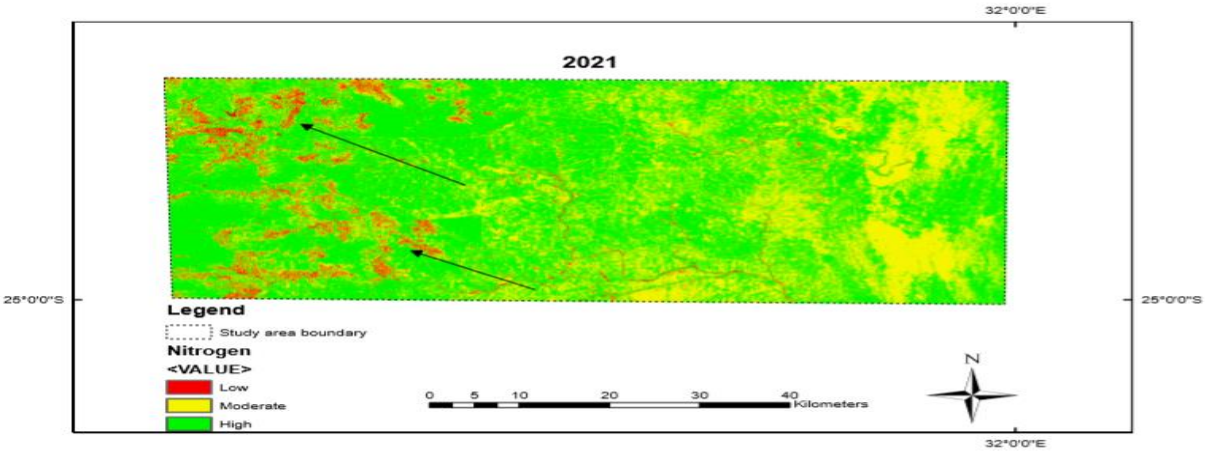


Figure 12 Spatial distribution of leaf N in 2021 at the Savanna ecosystem for the wet season

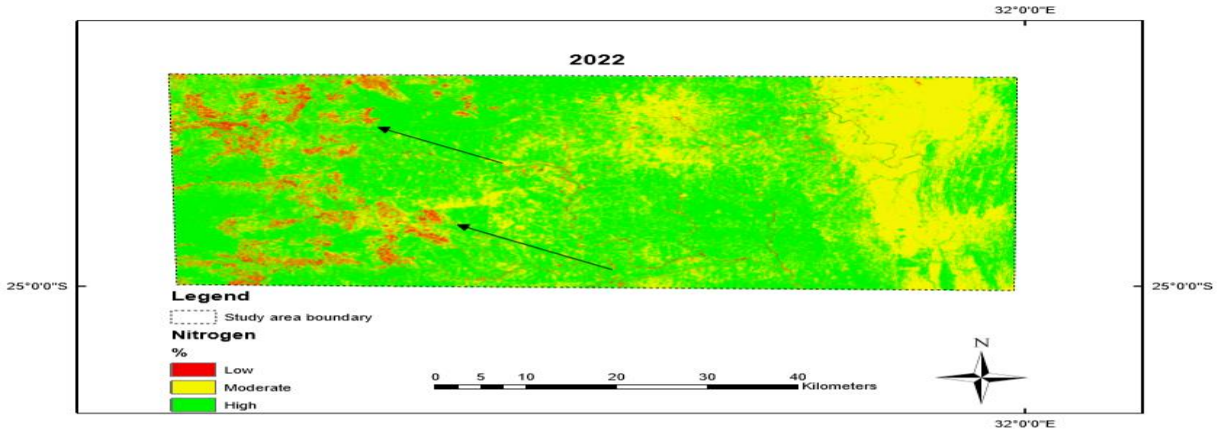


Figure 13 Spatial distribution of leaf N in 2022 at the Savanna ecosystem for the wet season

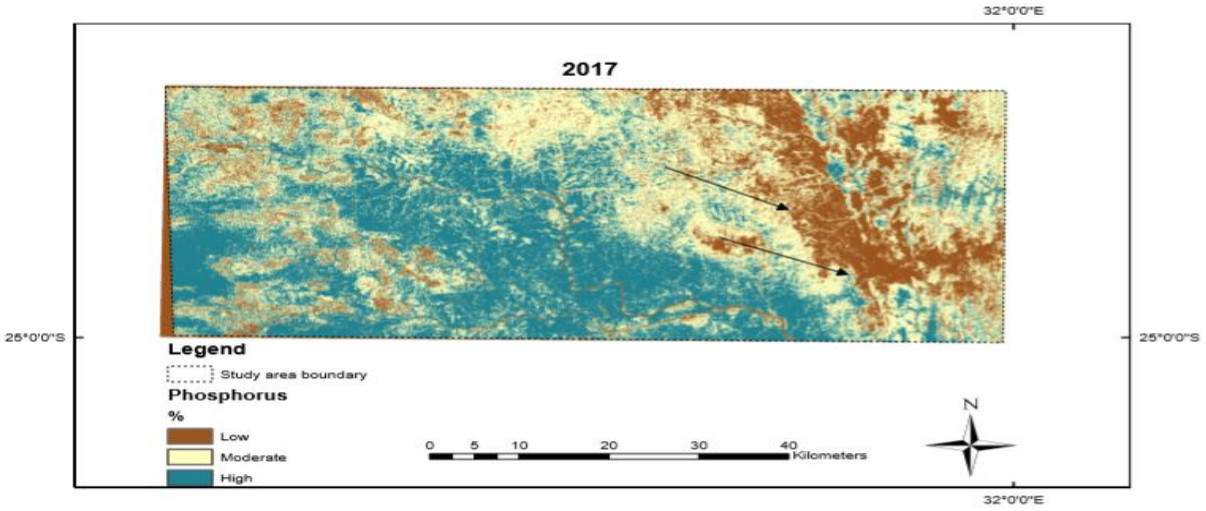


Figure 14 Spatial distribution of P in 2017 at the Savanna ecosystem for the wet season

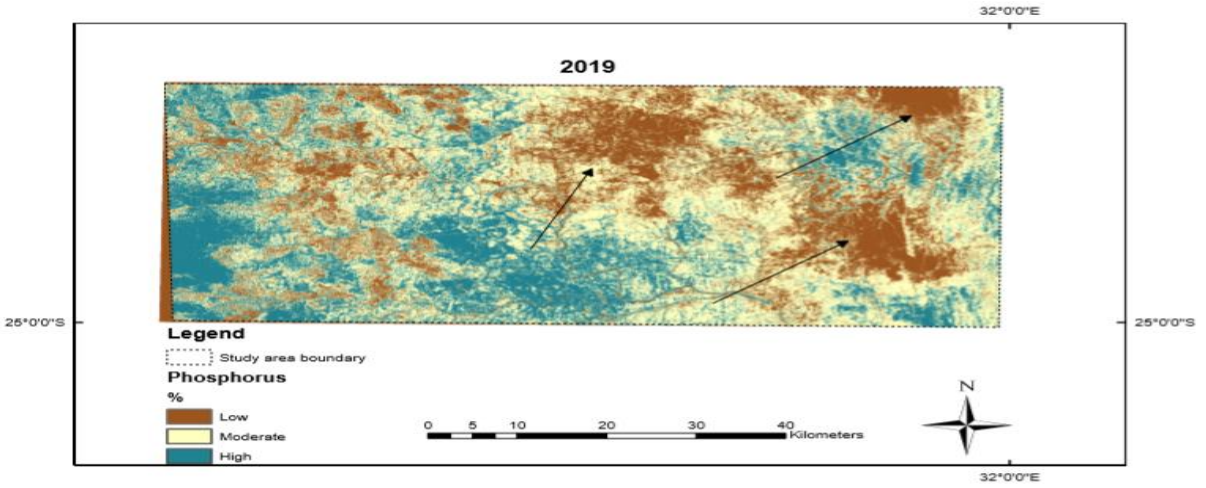


Figure 15 Spatial distribution of P in 2019 at the Savanna ecosystem for the wet season

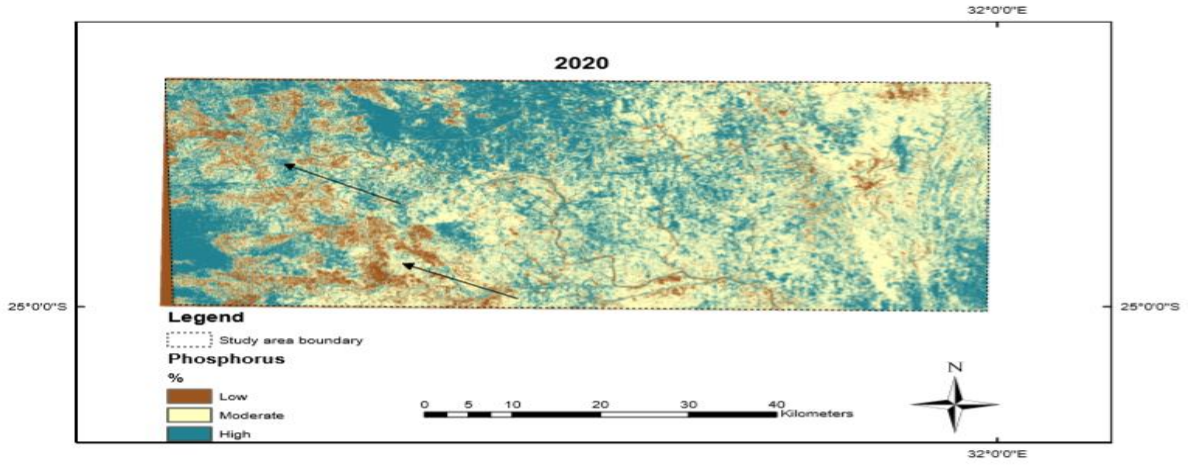


Figure 16 Spatial distribution of P in 2020 at the Savanna ecosystem for the wet season

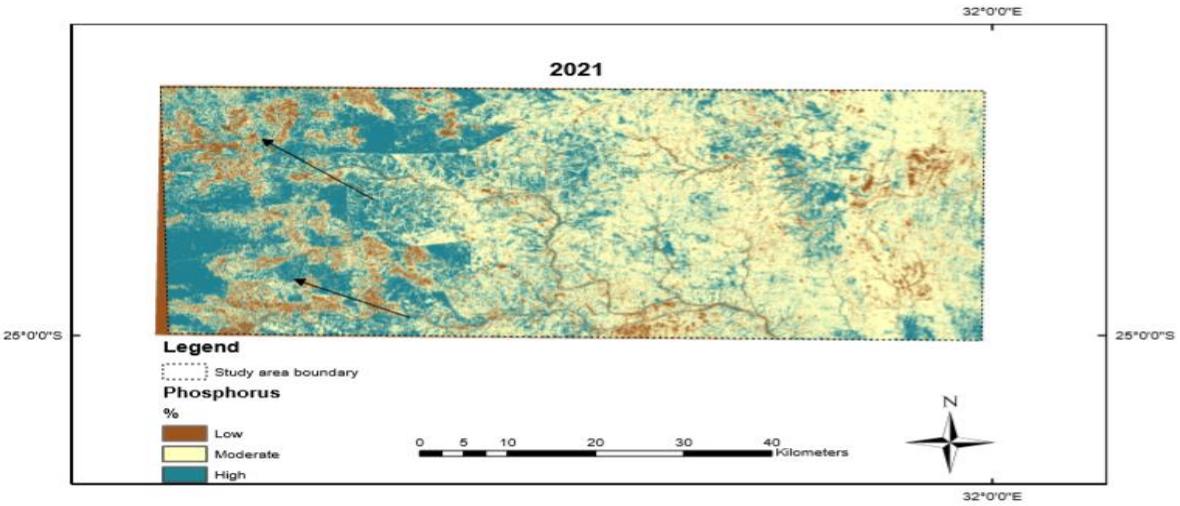


Figure 17 Spatial distribution of P in 2021 at the Savanna ecosystem for the wet season

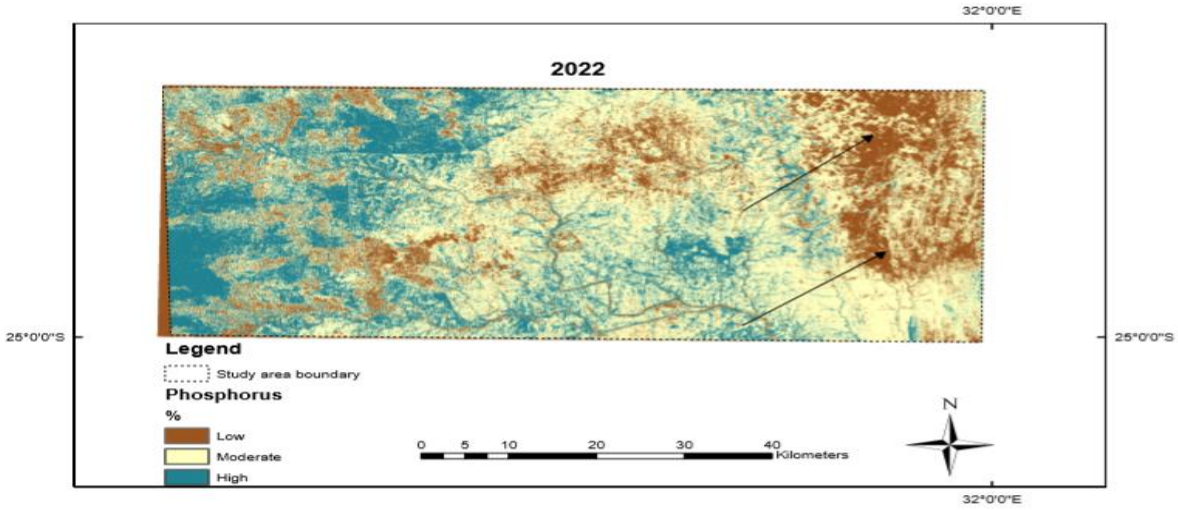


Figure 18 Spatial distribution of P in 2022 at the Savanna ecosystem for the wet season

CHAPTER 5: DISCUSSION

5.1. Estimating leaf N and P using simple linear regression

The results of SLR in Table 6 for estimating leaf N against different spectral bands of the Sentinel-2 satellite have shown that the red edge band at 705 nm (B5) produces the highest accuracy for predicting leaf N (R^2 = 0.1036, P-value = 0.0241, RMSE = 0.0258%). The importance of the red edge band has been shown in previous studies (Fernández-Habas et al, 2022). The overall results of predicting leaf N using various bands have shown that the

relationship between Leaf N and bands is poor as depicted by scatter plots in Figure 5. The results of estimating leaf N against different vegetation indices in Table 7 have shown that the high performing vegetation index was based on the red edge and is called MTCI (R^2= 0.1973, P-value = 0.0014, RMSE = 0.2408%). Similar results showing the importance of the MTCI were obtained in previous studies (Loozen et al, 2018). The second index which yielded the highest accuracy was the REP (R^2 = 0.1720, P-value = 0.0030,RMSE = 1.1010%) which is also based on the red edge. Studies conducted by Cho & Skidmore (2006) and Kanke et al. (2012) have shown that REP is highly correlated with chlorophyll and therefore leaf N.

The results of SLR as shown in Table 8 of estimating P against different spectral bands of Sentinel-2 satellite have shown a poor performance of bands in predicting P. The blue band at 443 nm (B1) produced the highest accuracy for predicting P (R^2 = 0.0131,P-value = 0.4329, RMSE = 0.0094%). The low coefficient of determination for the relationship between B1 and phosphorus could be due to the insensitivity of the 443 nm band to phosphorus. The results of estimating P using different vegetation indices in Table 9 have shown that the high-performing index was the REP (R^2 = 0.0161, p-value = 0.3852, RMSE = 1.2002%). Mutanga et al. (2007) successfully mapped grass phosphorus concentration in KNP using the REP derived from the HyMAP imagery. The second index which yielded the highest accuracy was based on the red edge and is called MTCI (R^2 = 0.0146, p-value = 0.4080, RMSE = 0.2668%). Clevers and Gitelson(2013) also conducted a study that showed the importance of MTCI in predicting P. The overall univariate results showed the importance of the red edge bands and the REP in estimating N and P, this was also previously shown by Horler et al. (1983).

5.2. Estimating leaf N and P using SMLR

For SMLR the best-performing model was selected based on the highest coefficient of determination (R^2). The results of the SMLR model in Table 11 have shown that for estimating leaf N, the high-performing model was based on a combination of the following optimal bands and VIs; B1, B2, B3, B5, B6, B7, B8, B8A, B9 and B11, GI, MCARI, MTCI, NDRE_1, NDRE_2, NDVI_1, NDVI_2, OSAVI, OSAVI1, OSAVI3, OSAVI4, OSAVI5, REP, SAVI_2, SR_1, SR_2 and SR705 (R^2 = 0.69). Whereas for estimating P the best performing model was based on a combination of the following VIs; GI, NDRE_2, NDVI_1,

NDVI_2, NRI, OSAVI2, OSAVI4, OSAVI5, REP, SAVI_1, SAVI_2 and TCARI (R^2 = 0.40). These results emphasize the importance of red edge bands as well as the red edge based vegetation indices in estimating grass nutrients. Similarly, Fernandez-Habas et al (2022) showed the importance of bands located in the red edge and NIR (700, 710, 1160 and 1170 nm) in predicting grass nutrients.

5.3. Estimating leaf N and P using RF

For RF the best-performing model was selected based on the highest coefficient of determination (R^2). The results of the RF model in Table 12 have shown that for estimating leaf N, the high-performing model was based on a combination of the following bands;B4, B5, B11 and B12 (R^2= 0.48). The results conquer with those obtained by Ramoeloet al. (2015) who showed that when using the RF technique, the spectral bands at 705 nm(B5) and the two SWIR bands at 1610 nm (B11) and 2190 nm (B12) were the most important bands in estimating leaf N using Sentinel-2. The results also correspond toprevious studies which highlighted the importance of the red edge bands and the SWIR bands in estimating leaf N (Ramoelo, 2012). Chabalala et al. (2020) also successfully predicted grass nitrogen using the RF model with a mean R^2 of 0.92 and 0.53 for Sentinel-2 and RapidEye respectively. Whereas for estimating P, the high-performing model was a combination of the following two VIs; NRI and NDRE_2 (R^2= 0.36). NRI is based on the red and green band whereas the NDRE_2 is based on the red edge bands. The selection of only two vegetation indices may be because phosphorus has few identifiable absorption features. Generally, the RF technique selected few optimal bands and vegetation indices as compared to SMLR and SVM.

5.4. Estimating leaf N and P using SVM

For SVM the best-performing model was selected based on the highest coefficient of determination (R^2). The results of SVM model in Table 13 have shown that for estimating leaf N, the high-performing model was based on a combination of the following bands B5, B12, B4 and B11 (R^2 = 0.14 and RMSE = 0.18%, RRMSE = 17.54% MAE = 0.15). The optimal bands selected are similar to those selected by RF technique for estimating leaf N. This again emphasizes the importance of red edge and SWIR bands in estimating leaf N (Ramoelo, 2012). Whereas for estimating P, the high-performing model was a combination

of the following bands B5, B2 and B4 (R^2 = 0.08 and RMSE = 0.06%, RRMSE = 40.78% MAE = 0.04).

5.5. Selection of the best model for predicting leaf N and P

SMLR outperformed RF and SVM in estimating leaf N and P in terms of accuracy and the certainty of the predictors. SMLR model using a combination of the following bands and VIs; B1, B2, B3, B5, B6, B7, B8, B8A, B9 and B11, GI, MCARI, MTCI, NDRE_1, NDRE_2, NDVI_1, NDVI_2, OSAVI, OSAVI1, OSAVI3, OSAVI4, OSAVI5, REP, SAVI_2, SR_1, SR_2, SR705 yielded the highest accuracy in predicting leaf N (R^2 = 0.69). On the other hand, the model which produced the highest accuracy for estimating phosphorus was based on a combination of the following vegetation indices; GI, NDRE_2, NDVI_1, NDVI_2, NRI, OSAVI2, OSAVI4, OSAVI5, REP, SAVI_1, SAVI_2 and TCARI (R^2 = 0.40). SMLR is capable of handling high dimensional datasets (Clevers et al. 2007) this may explain better performance as compared to random forest and SVM.

5.6. The spatial distribution of N and P across KNP

The distribution map of Leaf N as depicted in Figure 9, 10, 11, 12 and 13 for the years 2017, 2019, 2010, 2021 and 2022 respectively were created based on the SMLR model using a combination of bands and VIs. The maps show a great difference in the distributions of Leaf N across different sampling points in the study area. From the maps, it is clear that the concentration of leaf N is high alongside the water streams and decreases as you move away from the streams. Similar results were obtained in previous studies of mapping the spatial distribution of grass nutrients (Knox et al, 2011; Ramoelo et al. 2015; Chabalala et al. 2020). The concentration of leaf N increases on the eastern side and decreases on the western side from 2017 to 2022. This could be attributed to various environmental factors such as geology as shown in previous studies (Ramoelo and Cho, 2018; Manyashi, 2015).

The maps show the spatial distribution of P as depicted in Figure 14, 15,16, 17 and 18 for the years 2017, 2019, 2010, 2021 and 2022 respectively. The concentration of phosphorus is depicted as low, moderate and high on the maps. Figure 14-18 depict that the spatial distribution of phosphorus was low in the eastern and high in the western side of the study

area in 2017, the distribution reduces in 2019 and becomes moderate in 2020 and 2021. In 2022 the distribution of P becomes low again in the eastern side of the park and high in the western part. These variations might be due to the occurrences of veld fires. Few studies have successfully mapped P as compared to N (Mutanga & Kumar, 2007; Knox et al. 2011). The formulae for SAVI and MTCI both utilizes the red edge bands and hence they showe high significance in estimating leaf N and P.

CHAPTER 6: CONCLUSION AND RECOMMENDATIONS

The study investigated optimal spectral bands and VIs which were derived from the field spectral measurements for predicting leaf N and P. This study showed the importance of the red-edge derived vegetation indices and the red-edge bands of Sentinel-2 in estimating grass nutrients. The study demonstrated the importance of combining spectral bands and VIs for estimating grass nutrients. This study further demonstrated that SMLR outperformed RF and SVM techniques in predicting both leaf N and P. The study further demonstrated that the freely available Sentinel-2 datasets can be used for predicting grass nutrients to avoid the commercial implications of purchasing satellite data.

The spatial distribution of leaf N and P were successfully mapped. The concentration of Leaf N is high alongside the water streams and decreases as you move away from the streams. Whereas the concentration of phosphorus varies greatly between the eastern and western parts of the study area. This could be attributed to various environmental variables, veld fires and geology. The results of this study can be used to guide livestock farmers about pasture quality. Leaf N concentration has been estimated and mapped more than P and therefore future studies should explore the possibility of estimating foliar P using SWIR bands. The limitation of this study is that various environmental factors that could affect the distributions of Leaf N and P were not studied. Therefore, future studies should explore the impacts of environmental variables of these grass nutrients. The study further recommends future research should be conducted to compare the different machine-learning techniques to predict grass nutrients. Studies using drone remote sensing will be valuable for monitoring and estimating grass nutrients.

REFERENCES

Abdel-Rahman, E. M., Ahmed, F. B. & Van den Berg, M. 2010. Estimation of sugarcane leaf N concentration using in situ spectroscopy, International Journal of Earth Observation. 12; S52–S57.

Adam, E. M., Mutanga, O., Rugege, D. & Ismail, R. 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); 552–569.

Adjorlolo, C., Cho, M. A., Mutanga, O. & Ismail, R. 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; 1-15.

Aschbacher, J. & Milagro-Pérez, M. P. 2012. The European Earth Monitoring (GMES) Programme: status and perspectives. Remote sensing of the Environment. 120; 3-8.

Askari, M. S., McCarthy, T., Magee, A. & Murphy, D. J. 2019. Evaluation of grass quality under different soil management scenarios using remote sensing techniques. Remote Sensing. 11; 1-23.

Baillarin, S.J., Meygret, A., Dechoz, C., Petrucci, B., Lacherade, S., Trémas, T., Isola, C., Martimort, P. & Spoto, F. 2012. Sentinel-2 level 1 products and image processing performances. In 2012 IEEE international geoscience and remote sensing symposium. 7003-7006.

Baret, F., Houlès, V. & Guérif, M. 2007. Quantification of plant stress using remote sensing observations and crop models: the case of N management. Journal of Experimental Botany. 58; 869–880.

Barnes, E. M., Clarke, T. R., Richards, S. E., Colaizzi, P. D., Haberland, J., Kostrzewski, M., Waller, P., Choi, C., Riley, E., Thompson, T. & Lascano, R. J. 2000. July. Coincident detection of crop water stress, nitrogen status and canopy density using ground based multispectral data. In Proceedings of the Fifth International Conference on Precision Agriculture, Bloomington, MN, USA (Vol. 1619, p.6).

Bausch, W. C. 1993. Soil background effects on reflectance-based crop coefficients for corn.

Remote Sensing of Environment. 46; 213−222.

Belgiu, M. & Dragut, L. 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing. 144; 24-31.

Ben-Shahar, R. & Coe, M. J. 1992. The relationship between soil factors, grass nutrients and the foraging behaviour of wildebeest and zebra. Oecologia. 90; 422-428.

Bergman, C. M., Fryxell, J. M., Gates, C.C. & Fortin, D. 2001. Ungulate foraging strategies: energy maximizing or time minimizing? Journal of Animal Ecology. 70; 289-300.

Bian, J. H., Li, A., Song, M., Ma, L. & Jiang, J. 2010. Reconstruction of NDVI time-series datasets of MODIS based on Savitzky-Golay filter. Journal of Remote Sensing. 14(4); 725-741.

Breiman, L. 2001. Random Forests. Machine Learning. 45; 5-32.

Campbell, J. B. 2007. Introduction to remote Sensing. 4th edn. The Guilford Press. London.

Chabalala, Y., Adam, E., Oumar, Z. & Ramoelo, A. 2020. Exploiting the capabilities of Sentinel-2 and RapidEye for predicting grass nitrogen across different grass communities in a protected area. Applied Geomatics. 12(4); 379-395.

Chai, T. & Draxler, R. R. 2014. Root mean square error (RMSE) or mean absolute error (MAE)? –Arguments against avoiding RMSE in the literature. Geoscientific model development, 7(3). 1247-1250.

Chen, X., Zheng, H., Wanf, H. & Yan, T. 2022. Can machine learning algorithms perform better than multiple linear regression in predicting nitrogen excretion from lactating dairy cows. Scientific reports. 12(12478).

Cho, J. H., Min, B. J., Chen, Y. J., Yoo, J. S., Wang, Q., Kim, J. D. & Kim, I. H. 2007. Evaluation of FSP (fermented soy protein) to replace soybean meal in weaned pigs: growth performance, blood urea nitrogen and total protein concentrations in serum and nutrient digestibility. Asian-Australasian Journal of Animal Sciences. 20(12); 1874-1879.

Cho, M. A. & Skidmore, A. K. 2006. A new technique for extracting the red edge position from hyperspectral data: the linear extrapolation method. Remote sensing of the environment. 101(2); 181-193.

Choi, J. M. & Lee, C. W. 2012. Influence of elevated phosphorus levels in nutrient solution on

micronutrient uptake and deficiency symptom development in strawberry cultured with fertigation system. Journal of Plant Nutrition. 35(9); 1349–58.

Christensen, L. K., Bennedsen, B. S., Jørgensen, R. N. & Nielsen, H. 2004. Modelling nitrogen and phosphorus content at early growth stages in spring barley using hyperspectral line scanning. Biosystems Engineering. 88(1); 19–24.

Clark, R. N., King, T. V., Klejwa, M., Swayze, G.A. & Vergo, N. 1990. High spectral resolution reflectance spectroscopy of minerals. Journal of Geophysical Research: Solid Earth. 95(B8); 12653-12680.

Clevers, J. G. P. W, & Gitelson, A. A. 2013. Remote estimation of crop and grass chlorophyll and N content using red-edge bands on Sentinel-2 and-3. International Journal of Applied Earth Observation Geoinformation. 23; 344–351.

Clevers, J. G. P. W., De Jong, S. M., Epema, G. F., Van der Meer, F., Bakker, W. H., Skidmore, A. K. & Addink, E.A. 2001. MERIS and the red-edge position. International Journal of Applied Earth Observation and Geoinformation. 3 (4); 313-319.

Clevers, J. G. P. W., De Jong, S. M., Epema, G. F., Van Der Meer, F. D., Bakker, W. H., Skidmore, A. K. & Scholte, K. H. 2002. Derivation of the red edge index using the MERIS standard band setting. International Journal of Remote Sensing. 23(16); 3169-3184.

Clevers, J. G. P. W., van der Heljden, G. W. A. M., Verzakov, S. & Schaepman, M. E. 2007. Estimating Grassland Biomass Using SVM Band Shaving of Hyperspectral Data. Photogrammetric Engineering and Remote Sensing. 73; 1141-1148.

Clifton, K.E., Bradbury, J. W. & Vehrencamp, S. L. 1994. The fine-scale mapping of grassland protein densities. Grass Forage Science. 49 (1); 1-8.

Coops, N.C., Smith, M. L., Martin, M. E. & Ollinger, S. V. 2003. Prediction of eucalypt foliage N content from satellite-derived hyperspectral data. IEEE Trans. Geoscience Remote Sensing. 41; 1338–1346.

Curran, P. J., Dungan, J. L. & Gholz, H. L. 1990. Exploring the relationship between reflectance red edge and chlorophyll content in slash pine. Tree Physiology. 7; 33-48.

Curran, P., Dungan, J., Macler, B. & Plumber, S. 1992. Reflectance spectroscopy of fresh

whole leaves for the estimation of chemical concentration. Remote Sensing of Environment. 39; 153–166.

Curran, P. J. 1989. Remote sensing of foliar chemistry. Remote Sensing of Environment. 30; 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; 349–359.

Darvishzadeh, R., Skidmore, A., Schlerf, M., Atzberger, C., Corsi, F. & Cho, M. 2008. LAI and chlorophyll estimation for a heterogeneous grassland using hyperspectral measurements. ISPRS Journal of Photogrammetry & remote sensing. 63; 409-426.

Dash, J. & Curran, P. J., 2004, The MERIS Terrestrial Chlorophyll Index. International Journal of Remote Sensing. 25; 5003–5013.

Daughtry, C. S., Walthall, C. L., Kim, M. S., De Colstoun, E. B. & McMurtrey Iii, J. E. 2000. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. Remote sensing of Environment. 74(2); 229-239.

Dawson, T. P. & Curran, P. J. 1990. Technical note A new technique for interpolating the reflectance red edge position. International Journal of Remote Sensing. 19; 2133-2139.

De Jong, S. M., Pebesma, E. J. & Lacaze, B. 2003. Above-ground biomass assessment of Mediterranean forests using airborne imaging spectrometry: the DAIS Peyne experiment. International Journal of Remote Sensing. 24 (7); 1505-1520.

Dowdy, S., Weardon, S. C. & Chilko, D. 2011. Statistics for Research. New fdJersey: John Wiley & Sons.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F. & Bargellini, P. 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote Sensing of Environment. 120; 25–36.

Earth Engine Data Catalog Sentinel-2 MSI: MultiSpectral Instrument, Level-2A. Availableonline:https://developers.google.com/earthengine/datasets/catalog/COPERNICUS_

S2_SR (accessed on 05 August 2022)

Evans, J. 1989. Photosynthesis and N relationships in leaves of C3 plants. Oecologia. 78; 9–19.

Fernández-Habas, J., Cañada, M.C., Moreno, A.M.G., Leal-Murillo, J.R., González-Dugo, M. P., Oar, B. A., Gómez-Giráldez, P. J. & Fernández-Rebollo, P. 2022. Estimating pasture quality of Mediterranean grasslands using hyperspectral narrow bands from field spectroscopy by Random Forest and PLS regressions. Computers and Electronics in Agriculture. 192; 106614.
Ferwerda, J. G., Skidmore, A. K. & Mutanga, O. 2005. N detection with hyperspectral normalized ratio indices across multiple plant species. International Journal of Remote Sensing. 26; 4083–4095.

Fewster, R. M., Laake, J. L. & Buckland, S. T. 2005. Line transect in small and large regions, Biometrics. 61; 856–859.

Foxcroft, L. C., Richardson, D. M. & Wilson, J. R. 2008. Ornamental Plants as Invasive Aliens: Problems and Solutions in Kruger National Park, South Africa. Environmental Management. 41 (1); 32–51.

Gao, B. C. & Goetz, A. F. H. 1994. Extraction of dry leaf spectral features from reflectance spectra green vegetation. Remote Sensing of Environment. 47(3); 369-374.

Gitelson, A. & Merzlyak, M. N. 1994. Quantitative Estimation of Chlorophyll-a Using Reflectance Spectra: Experiments with Autumn Chestnut and Maple Leaves. Journal of Photochemistry and Photobiology. 13; 247-252.

Gitelson, A. A., Kaufman, Y.J. & Merzlyak, M. N. 1996. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. Remote sensing of Environment. 58(3); 289-298.

Goetz, A.F. 2009. Three decades of hyperspectral remote sensing of the Earth: A personal view. Remote sensing of environment, 113; S5-S16.
Grant, C. C. & Scholes, M. C. 2006. The importance of nutrient hot-spots in the conservation and management of large wild mammalian herbivores in semi-arid savannas. Biological Conservation. 130(3); 426–437

Grimm, R., T. Behrens, M. Märker, H. & Elsenbeer, H. 2008. Soil Organic Carbon Concentrations and Stocks on Barro Colorado Island – Digital Soil Mapping Using Random Forests Analysis. Geoderma. 146; 102–13.

Güsewell, S. 2004. N: P ratios in terrestrial plants: variation and functional significance. New Phythologist. 164; 243-266.

Guyot, G. & Baret, F. 1988. Utilisation de la haute resolution spectrale pour suivre l'etat des couverts vegetaux. In Proceedings of the 4th International colloquim on spectral signatures of objects in remote sensing, ESA SP-287 279–286. edited by, Assois, France, ESA, Paris. 279-286.

Haboudane, D., Miller, J. R., Tremblay, N., Zarco-Tejada, P. J. & Dextraze, L. 2002. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. Remote sensing of environment. 81(2-3); 416-426.

Hall, A., Lamb, D. W., Holzapfel, B. & Louis, J. 2002. Optical remote sensing applications in viticulture- a review. Australian Journal of Grape and Wine Research. 8; 36-47.

Hansen, P. M. & Schjoerring, J. K. 2003. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. Remote sensing of environment. 86, 542-553.

Hihi, S., Rabah, Z. B., Bouaziz, M., Ctourou, M. Y. & Bouaziz, S. 2019. Prediction of Soil Salinity Using Remote Sensing Tools and Linear Regression Model. Advances in Remote Sensing. 8; 77-88.

Homolova, L., Maenovsky, Z., Clevers, J., Garcia-Santos, G. & Schaeprnan, M. E. 2013. Review of optical-basedremote sensing for plant trait mapping. Ecological Complex. 15; 1–16.

Horler, D. N. H., Dockray, M. & Barber, J. 1983. The red edge of plant leaf reflectance. International Journal of Remote Sensing. 4; 273–288.

Hsu, C. W., & Lin, C. J. 2002. A comparison of methods for multiclass support vector machines. Neural Networks, IEEE Transactions on. 13; 415-425

Huang, Z., Turner, B. J., Dury, S. J. Wallis, I. R. & Foley, W. J. 2004. Estimating foliage N

concentration from HYMAP data using continuum removal analysis. Remote Sensing of Environment. 93; 18-29.

Huete, A.R. 1988. A soil-adjusted vegetation index (SAVI). Remote sensing of environment. 25(3); 295-309.

Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F. & King, D. 1992. Evaluation of radiometric quality, statistical characteristics and spatial resolution of multispectral videography. Journal of Imaging Science and Technology. 36; 394–404.

Jackson, J.E. 2005. A user's guide to principal components. Vol. 587. New York: John Wiley & Sons.

Jiang, G. & Wang, W. 2017. Error estimation based on variance analysis of k-fold cross-validation. Pattern Recognition. 69; 94-106.

Johnson, L. F., Hlavka, C. A. & Peterson, D. L. 1994. Multirative analysis of AVIRIS data for canopy biochemical estimation along the Oregon transect. Remote Sensing of the Environment. 47; 216-230.

Jordan, C. F. 1969. Derivation of leaf-area index from quality of light on the forest floor. Ecology. 50(4); 663-666.

Kalacska, M., Bohlman S., Sanchez-Azofeifa G. A., Castro-Esaum K. & Caelli, T. 2007. Hyperspectral discrimination of tropical dry forest lianas and trees: Comparative data reduction approaches at the leaf and canopy levels. Remote Sensing of Environment. 109 (4); 406–415.

Kanke, Y., Raun, W., Solie, J., Stone, M. & Taylor, R. 2012. Red edge as a potential index for detecting differences in plant nitrogen status in winter wheat. Journal of plant nutrition. 35(10); 1526-1541.

Kim, J. H. 2009. Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. Computational statistics & data analysis. 53(11); 3735-3745.

Knox, N. M., Skidmore, A. K., Prins, H. H. T., Asner, G. P., van der Werff, H. M. A., de Boer, W. F., van der Waal, C., de Knegt, H, J. & Kohi, E. M. 2011. Dry season mapping of savanna forage quality, using the hyperspectral Carnegie Airborne Observatory sensor. Remote Sensing of the Environment. 15 (6); 1478–1488.

Knox, N. M., Skidmore, A. K., Schlerf, M., De Boer, W. F., Van Wieren, S. E., Van Der Waal, C., Prins, H. H. T. & Slotows, R. 2010. Nitrogen prediction in grasses: effect of bandwidth and plant material state on absorption feature selection. International Journal of Remote Sensing. 31; 691-704.

Kokaly, R. F., Asner, G. P., Ollinger, S. V., Martin, M. E. & Wessman, C. A. 2009. Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies. Remote Sensing of Environment. 113 (1); S78–S91.

Kokaly, R. F. & Clark, R. N. 1999. Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. Remote Sensing of Environment. 67; 267–287.

Kooistra, L., Salas, E. A. L., Clevers, J. G. P. W., Wehrens, R., Leuven, R. S. E. W., Niehuis, P. H. & Buydens, L. M. C. 2004. Exploring field vegetation reflectance as an indicator of soil contamination in river floodplains. Environmental Pollution. 127; 281–290.

Kreulen, D. A. & Hoppe, P. P. 1979. Diurnal trends and relationship to forage quality of ruminal volatile fatty acid concentration, pH and osmolarity in wildebeest on dry range in Tanzania. African Journal of Ecology. 17(2);53-63.

Kuchibhotla, A. K., Brown, L. D., Buja, A. & Cai, J., 2019. All of linear regression. arXiv preprint arXiv:1910.06386.

La Pierre, K. J. & Smith, M. D. 2016. Soil nutrient additions increase invertebrate herbivore abundances but not herbivory, across three grassland systems. Oecologia. 180; 485-497.

Lamb, D. W., Steyn-Ross, M., Schaare, P., Hanna, M. M., Silvester, W. & Steyn-Ross, A. 2002. Estimating leaf N concentration in ryegrass (Lolium spp.) pasture using the chlorophyll red-edge: Theoretical modelling and experimental observations. International Journal of Remote Sensing. 23 (18); 3619-3648.

Loozen, Y., Rebel, K. T., Karssenberg, D., Wassen, M. J., Sardans, J., Peñuelas, J. & De Jong, S. M. 2018. Remote sensing of canopy nitrogen at regional scale in Mediterranean forests using the spaceborne MERIS Terrestrial Chlorophyll Index. Biogeosciences. 15(9); 2723-2742.

Low, A. B. & Rebelo, A.G. 1998. Vegetation of South Africa, Lesotho and Swaziland. Department of Environmental Affairs and Tourism, Pretoria.

Lück, W., Mhangara, P., Kleyn, L. & Remas, H. 2010. Land cover field guide CSIR satellite applications centre: Earth observation service centre. Chief Directorate: National Geospatial Information. Version 2.0.

Mahdianpari, M., Salehi, B., Mohammadimanesh, F., Homayouni, S., & Gill, E. 2018. The first wetland inventory map of Newfoundland at a spatial resolution of 10 m using sentinel-1 and sentinel-2 data on the google earth engine cloud computing platform. Remote Sensing. 11(1); 43.

Manyashi, E. K. Assessment of foliar nitrogen as an indicator of vegetation stress using remote sensing: The case study of Waterberg region, Limpopo Province [dissertation]. University of South Africa; 2015.

Martin, M. E. & Aber, J. D. 1997. High spectral resolution remote sensing of forest canopy lignin, N, and ecosystem processes. Ecological Applications. 7; 431-443.

Martin, M. E., Plourde, L. C., Ollinger, S. V., Smith, M. L. & McNeil, B.E. 2008. A generalizable method for remote sensing of canopy N across a wide range of forest ecosystems. Remote Sensing of Environment. 112; 3511–3519.

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.

McKeen, S., Wilczak, J., Grell, G., Djalalova, I., Peckham, S., Hsie, E. Y., Gong, W., Bouchet, V., Menard, S., Moffet, R. & McHenry, J. 2005. Assessment of an ensemble of seven real-time ozone forecasts over eastern North America during the summer of 2004. Journal of Geophysical Research: Atmospheres. 110(D21).

Meer, F. V. D., Bakker, W. K. L., Scholte, K., Skidmore, A. K., Jong, S. D., Clevers, J., Addinks, E. & Epema, G., 2001, Spatial scale variations in vegetation indices and above-ground biomass estimates: Implications for MERIS. International Journal of Remote Sensing. 22; 3381–3396.

Moghimi, A., Pourreza, A., Zuniga-Ramirez, G., Williams, L. E. & Fidelibus, M. W. 2020. A novel machine learning approach to estimate grapevine leaf nitrogen concentration using aerial multispectral imagery. Remote Sensing. 12(21); 3515.

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); 4897-4911.

Mutanga, O. & Skidmore, A. K. 2004. Narrow band vegetation indices overcome the saturation problem in biomass estimation. International Journal of Remote Sensing. 25; 1-16.

Mutanga, O., Adam, E., Adjorlolo, C. & Abdel-Rahman, E. M. 2015. Evaluating the robustness of models developed from field spectral data in predicting African grass foliar N concentration using WorldView-2 image as an independent test dataset. International Journal of Applied Earth Observation and Geoinformation. 34; 178-187.

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

Oldeland, J., WouterDorigo, W., Lieckfeld, L., Lucieer, A. & Jürgens, N. 2010. Combining vegetation indices, constrained ordination and fuzzy classification for mapping semi-natural vegetation units from hyperspectral imagery. Remote Sensing of Environment. 114; 1155–1166.

Ollinger, S. V. & Smith, M. L. 2005. Net primary production and canopy N in a temperate forest landscape: An analysis using imaging spectroscopy, modelling and field data. Ecosystems. 8; 760–778.

Osborne, S. L., Schepers, J. S., Francis, D. D., & Schlemmer, M. R. 2002. Detection of phosphorus and nitrogen deficiencies in corn using spectral radiance measurements. Agronomy Journal. 94; 1215−1221.

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

Press, W.H. and Teukolsky, S.A., 1990. Savitzky-Golay smoothing filters. Computers in Physics. 4(6);669-672.

Prins, H. H. T. & Beekman, J. 1989. A balanced diet as a goal for grazing: the food of the Manyara buffalo. African Journal of Ecology. 27; 241-259.

Punalekar, S. M., Thomson, A., Verhoef, A., Humphries, D. J. & Reynolds, C. K. 2021. Assessing suitability of Sentinel-2 bands for monitoring of nutrient concentration of pastures

with a range of species compositions. Agronomy. 11; 1-17.

Ramoelo, A. and Cho, M.A., 2018. Explaining leaf nitrogen distribution in a semi-arid environment predicted on Sentinel-2 imagery using a field spectroscopy derived model. Remote Sensing.10(2); 269.

Ramoelo, A., Cho, M., Mathieu, R. & Skidmore, A. K. 2015. Potential of Sentinel-2 spectral configuration to assess rangeland quality. Journal of Applied Remote Sensing. 9; 1-11.

Ramoelo, A., Cho, M. A., Mathieu, R., Skidmore, A. K, Sclerf, M., Heitkönig, I. M. A. & Prins, H. H. T. 2011. Integrating environmental and in situ hyperspectral remote sensing variables for the grass N estimation in savannah ecosystems. 34th International Symposium on the Remote Sensing of Environment (ISRSE 2011), The GEOSS Era: Towards Operational Environmental Monitoring. Sydney, Australia.

Ramoelo, A., Skidmore, A. K., Cho, M. A., Schlerf, M., Mathieu, R. & Heitkonig, I. M. A. 2012. Regional estimation of savanna grass N using the red-edge band of the spaceborne RapidEye sensor. International Journal of Earth Observation. 19; 151- 162.

Redfern, J. V., Grant, C. C., Biggs, H. C. & Getz, W. M. 2003. Surface water constraints on herbivore foraging in the Kruger National Park, South Africa. Ecology. 84; 2092-2107.

Richter, K., Atzberger, C., Hank, T. B. & Mauser, W. 2012. Derivation of biophysical variables from Earth observation data: validation and statistical measures. Journal of Applied Remote Sensing. 6(1); 063557-063557.

Rondeaux, G., Steven, M. & Baret, F. 1996. Optimization of soil-adjusted vegetation indices. Remote sensing of environment. 55(2); 95-107.

Royston, J. P. 1982. An extension of Shapiro and Wilk's test for normality to large samples. Journal of the Royal Statistical Society. Series C (Applied Statistics). 3; 115-124.

Sakamoto, Y., Ishiguro, M., & Kitagawa, G. 1986. Akaike information criterion statistics. The Netherlands: D. Reidel Publishing Company, Dordrecht.

Savitzky, A. & Golay, M. J. 1964. Smoothing and differentiation of data by simplified least squares procedures. Analytical chemistry. 36(8); 1627-1639.

Schleicher, T. D., Bausch, W. C., Delgado, J. A. & Ayers, P. D. 1998. Evaluation and

refinement of the nitrogen reflectance index (NRI) for site-specific fertilizer management. In 2001 ASAE Annual Meeting (p. 1). American Society of Agricultural and Biological Engineers.

Serrano, L., Peñuelas, J. & Ustin, S. L. 2002. Remote sensing of N and lignin in Mediterranean vegetation from AVIRIS data: Decomposing biochemical from structural signals. Remote Sensing of Environment. 81; 355–364

Shackleton, S. E., Shackleton, C. M., Netshiluvhi, P. R., Geach, B. S., Ballance, A. & Fairbanks, D. H. K. 2002. Use patterns and value of savanna resources in three rural villages in South Africa. Economic Botany. 56 (2); 130-146.

Shen, J., Yuan, L., Zhang, J., Li, H., Bai, Z. & Chen, X. 2011. Phosphorus dynamics: from soil to plant. Plant Physiology. 156; 997–1005.

Siedliska, A., Baranowski, P., Pastuszka-Woźnaik, J., Zubik, M. & Krzysczak, J. 2021. Identification of leaf P content at different growth stages based on hyperspectral reflectance. BMC Plant Biology. 21:28; 1-17.

Sims, D. A. & Gamon, J.A. 2002. Relationships Between Leaf Pigment Content and Spectral Reflectance across a Wide Range of Species, Leaf Structures and Developmental Stages. Remote Sensing of Environment. 81; 337-354.

Singh, L., Mutanga, O., Mafongoya, P. & Peerbhay, K. 2017. Remote sensing of key grassland nutrients using huperspectral techniques in KwaZulu-Natal, South Africa. Journal of Applied Remote Sensing. 11(3); 036005.

Skidmore, A. K., Ferwerda, J. G., Mutanga, O., van Wieren, S. E., Peel, M., Grant, R. C., Prins, H. H. T., Balcik, F. B. & Venus, V. 2010. Forage quality of savannas-simultaneously mapping foliar protein and polyphenols for trees and grass using hyperspectral imagery. Remote Sensing of the Environment. 114 (1); 64-72.

Steeb, W. H. 2015. The nonlinear workbook. 6th edn. World Scientific Publishing Singapore.

Strydom, H., Fouche C. B. & Delport C. S. L. 2005. Research at Grassroots for Social Sciences and Human Service Professions. 3rd edn.

Sudalaimuthu, K. & Sudalayandi, K. 2019. Development of linear regression model to predict

ground elevation from satellite elevation- statistical approach. AIP Conference Proceedings. 2112; 2444301.

Thissen, U., Pepers, M., Ustun, B., Melssen, W. J. & Buydens, L. M. C. 2004. Comparing support vector machines to PLS for spectral regression applications, Chemometrics and Intelligent Laboratory Systems. 73 (2); 169–179.

Tian, Y. C., Yao, X., Yang, J., Cao, W. X., Hannaway, D. B. & Zhu, Y. 2011. Assessing newly developed and published vegetation indices for estimating rice leaf N concentration with ground- and space-based hyperspectral reflectance. Field Crops Resolution. 120; 299–310.

Tucker, C. J. 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment. 8; 50-127.

Wang, L. (Editor). 2005. Support vector machine: theory and applications, Springer.

Wang, Z. J., Wang, J. H., Liu, L. Y., Huang, W. j., Zhao, C. J. & Wang, C. Z. 2004. The prediction of grain protein in winter wheat (Triticumaesetivum L.) using plant pigment ratio (PPR). Field Crop Research. 90; 311-321.

Wang, Z., Skidmore, A. K., Darvishzadeh, R., Heiden, U., Heurich, M. & Wang, T. 2015. Leaf N content in directly estimated by leaf traits derived from the PROSPECT model. IEEE Journal of Selected Topics in Applied Earth Observation Remote Sensing. 8; 3172–3182

Wang, Z., Wang, T., Darvishzadeh, R., Skidmore, A. K., Jones, S., Suarez, L., Woodgate, W., Heiden, U., Heurich, M. & Hearne, J. 2016. Vegetation Indices for Mapping Canopy Foliar N in a Mixed Temperate Forest. Remote Sensing. 8(491); 1-20.

Wessels, K. J., Mathieu, R. & Erasmus, B. F. N. 2011. Impacts of communal land use and conservation on woody vegetation structure in the Lowveld savannas of South Africa. Forest Ecology and Management. 26; 19-29.

Wu, C., Niu, Z., Tang, Q. & Huang, W. 2008. Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. Agricultural and forest meteorology. 148(8-9); 1230-1241.

Yu, Q., Wang, S., Shi, H., Huang, K. & Zhou, L. 2014. An Evaluation of Space borne Imaging Spectrometry for Estimation of Forest Canopy N Concentration in a Subtropical Conifer

Plantation of Southern China. Journal of Resources and Ecology. 5(1); 1-10.

Zarco-Tejeda, P. J., Miller, J. R., Morales, A., Berjon, A. & Aguera, J. 2004. Hyperspectral indices and model simulation for chlorophyll estimation in open canopy tree crops. Remote Sensing of Environment. 90; 463–476.

Zengeya, F. M., Mutanga, O. & Murwira, A. 2013. Linking remotely sensed forage quality estimates from WorldView-2 multispectral data with cattle distribution in a savannah landscape. International Journal of Applied Earth Observation and Geoinformation. 21; 513-524.