M.Sc. Dissertation

An application of multi-scale remote sensing in estimating grass nutrient limitation as measured by a ratio of Nitrogen and Phosphorus in a savanna ecosystem.

Nasiphi Ngcoliso

Supervisor: Prof. Abel Ramoelo

Co-supervisor: Dr. Philemon Tsele

Submitted in partial fulfillment of the requirements for the degree of Master of Science in Geoinformatics.

In Department of Geography, Geoinformatics and Meteorology

University of Pretoria

July 2023

Declaration of originality

I, Nasiphi Ngcoliso, declare that:

I declare that the dissertation, which I hereby submit for the degree in Master's Science in Geoinformatics at the University of Pretoria, is my work and has not previously been submitted by me or any other student for a degree at any other tertiary institution. All the data and graphics contained in this research are not from any other person unless indicated otherwise. The information obtained from other sources was paraphrased, and the authors were duly referenced.

Signed…….

Date…….

Nasiphi Ngcoliso (21753742)

ACKNOWLEDGEMENTS

Without God's mercy, this project would have been impossible; I thank God for giving me the strength and intellectual talents to continue my studies and for leading me throughout the two years of this endeavor. I would like to convey my heartfelt gratitude to my supervisors, Prof. Abel Ramoelo and Dr. Philemon Tsele; it was not easy, but this work was a success thanks to their guidance, counsel, encouragement, support, and critical comments. I appreciate their sharing of their remote sensing experience, knowledge, and expertise. I'd like to express my heartfelt appreciation to the National Research Foundation (NRF) for its financial assistance during my studies. A particular thanks to the KNP managers for granting permission to perform this research in the park.

I'd like to thank Mr. Mcebisi Qabaqaba for assisting me with the data analysis and sharing his coding experience, as well as for his guidance, suggestions, and critical remarks throughout the project. Finally, I'd want to express my heartfelt gratitude to my mother for her emotional support, love, prayers, and teachings, which kept me going during my studies.

Draft manuscript from this dissertation

Ngcoliso N, Ramoelo A, Tsele P (2023) An application of simulated Sentinel-2 data sets for predicting and mapping the distribution of nutrient limitation in a savanna landscape, *Applied Geoinformatics* (Submitted)

Abstract

Nutrient limitations may impact the ecosystem services the savanna biome provides. It may lead to degradation and, consequently, reduce the grazing capacity of the savannas if the necessary control measures are not implemented in time. The key indicator of the growth-limiting nutrients is the Nitrogen to Phosphorus (N:P) ratio. Grass foliar phosphorus content had rarely been investigated in African savannas, especially with remote sensing. Hence, information on the distribution of nutrient limitation is very limited. This study aimed to develop a Sentinel-2-based N:P predicting model and map the spatiotemporal variations of the N:P ratio in the Kruger National Park (KNP) area in the Northern part of the South African savanna biome. This was achieved by simulating the Analytical Spectral Device (ASD) reflectance data from 49 sampling points to Sentinel-2 MultiSpectral Instrument (MSI) configuration dataset. Laboratory-based chemical analysis was conducted to extract the concentrations of N and P from the grass samples. Partial least squares regression (PLSR) and random forest regression (RFR) techniques were used to develop the N:P prediction models from the simulated Sentinel-2 datasets. Results show that the best predicting RFR model explained over 80% of N:P variability with the lowest relative root mean square error (RRMSE) of 14%, with a p-value of less than 0.05. The optimal-predicting model was used to map the distribution of nutrient limitation using Sentinel-2 images across KNP and surroundings. Different parts of the KNP area are either N-limited or co-limited. The observed variations may result from varying environmental factors and anthropogenic activities. The Sentinel-2 N:P ratio estimation accuracies were then compared to the ratio of N:P of data from commercial multispectral (RapidEye and WorldView-2) and hyperspectral (Hyperion and EnMap) sensors. There is no vast difference between the estimation accuracy of these commercial sensors and that of the freely available Sentinel-2 when using RFR. However, when using PLSR, Sentinel-2 produced improved N:P ratio estimation accuracy than the commercial sensors with the highest R^2 value of 0.66 and an RRMSE of 20.696%. This makes Sentinel-2 a cost-effective means for estimating nutrient limitation in a heterogeneous savanna landscape. This study provides decision-makers with a cost-effective tool for managing, sustaining, and restoring the savanna biome. The inclusion of textural information is recommended for future research.

Keywords: Nutrient limitation, N:P ratio, ASD, Spectral simulation, Sentinel-2, PLSR, RFR, Model inversion, KNP.

Dedication

To my mother and my siblings

To my daughter, Luniko Ngcoliso

TABLE OF CONTENTS

TABLES

Figures

Acronyms and abbreviations

AF – Absorption Features

AGB – Above Ground Biomass

ANN- Artificial Neural Network

ASD – Analytical Spectrometer Device

CV – Cross-validation

CW – Water removal

FD – First derivative

LAI -Leaf Area Index

LOOCV – Leave-one-out Cross-validation

MAE – Mean absolute error.

MSI- Multi-Spectral Instrument

MLRA- Machine Learning Regression Algorithm

N – Nitrogen

NDVI – Normalised Difference Vegetation Index

NDRE – Normalised Difference Red-edge

P - Phosphorus

PSLR - Sparse Partial Least Squares Regression

RFR - Random Forest Regression

RMSE – Root mean square error.

RRMSE – Relative root mean square error.

RRI – Red-edge Ratio Index

REIs – Red-edge Indices

SMLR- Stepwise Multiple-Linear Regression

SVM - Support Vector Machine

SR–Simple Ratio

CHAPTER 1:

General introduction

1.1 Background

Over 52% of the land globally is covered by the savanna biome playing a crucial role in agricultural production, economy, and ecotourism (Ramoelo et al., 2013; Ali et al., 2016; Askari et al., 2019; Shoko et al., 2016). African savannas support a diverse range of wildlife through provision of connection and habitation (Sankaran et al., 2013), and rural communities depend on the savannas for livestock grazing. Considering the above factors, the role of savannas in the global economy through tourism and food security by providing feed for ruminants used for milk and meat production is remarkable (FAO, 2015). Moreover, the savanna biome is critical as carbon and groundwater storage. Effective functioning and productivity of the savannah biome depend on grass quality as defined by the availability of foliar nutrients.

Regardless of the importance of savannas in several human livelihoods, ecological and ecosystem activities mentioned earlier, the worsening land degradation is still a significant concern not only to the environment but a socio-economic challenge as well (Hammad & Tumeizi, 2012). Land degradation manifestations on dryland, as explained by Ponce-Hernandez, (2008), include reduced plant productivity, deteriorations in soil fertility and physical properties, unpleasant changes in biomass and ecosystem activities, the decline in vegetation's nutritional status, and reduced species diversity. Major concerns are food and energy security, rural economy, biodiversity loss, and land desertification, as these are the most threatened aspects (Salih et al., 2017; Ramoelo et al., 2014; Liu et al., 2008). The literature revealed that the savanna is one of the most devastated biomes due to the high sensitivity of these ecosystems to changes in climatic conditions and changes in soil properties (Zhao et al., 2017).

Other factors that contribute to land degradation include rapid population growth leading to extensive grazing to meet the increasing food demands and conversion of natural grassland to protected farmlands, residential and industrial areas, and the introduction of alien species (Xie et al., 2020). The key indicators of land degradation are changes in quality, quantity, and distribution of vegetation cover; hence, these changes

are used as a proxy to measure the extent to which the land is degraded (Safriel, 2007). Assessment and monitoring of vegetation cover are critical steps towards effective management and restoration to minimise the effect of land deterioration on biodiversity and agricultural and ecosystem services. The United Nations Sustainable Development Goal 15, “life on land” of 17 aims “to Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, halt and reverse land degradation as well as the biodiversity loss” by the year 2030, which prioritises environmental sustainability as one of its main priorities. This pillar aims to restore and protect the earth's balance, ensuring moderate use of natural resources without compromising the future. These include land, water, forest resources and biodiversity in terrestrial ecosystems. To achieve this goal, near real-time monitoring of the terrestrial ecosystem is required to provide regular information. Inclusive of this are natural rangelands found within the savannas, Kellner and de Wet., (2021) reported a 25% of South African rangelands are degraded. For the sustainability and continuous productivity of these rangelands to be ensured and, adequate monitoring, protection, and management of the savannas are necessary to inhibit this deterioration.

All this is possible by assessing the biome’s biophysical and biochemical conditions. However, the main challenge is usually the financial costs and unavailability of information required by decision-makers and policymakers to formulate risk management and coping strategies to mitigate the threatening conditions. Especially information on the biome’s foliar biochemical content, which defines the quality conditions of the savannas. Estimating grass quality at a larger scale is essential for management and wildlife conservation, and such information provides ecologists and nature conservationists with updated information on their conditions. This is a crucial step towards implementing strategic management techniques such as adequate fire regimes and rotational grazing which allow pastures to recover and minimise degradation. Moreover, assessing grass quality is essential for understanding the operational dynamics of rangelands, feeding patterns, distribution, and the density of herbivores across savanna biomes (Carlier et al., 2009; Gao et al., 2019; Reinermann et al., 2020). For example, several studies report that in South African Savannas, large herbivores, such as antelopes and buffalos, dominate in areas with high grass quality.

Foliar biochemical content nitrogen (N) and phosphorus (P) are the key determinants of grass quality status (Ramoelo et al., 2015; Mutanga et al., 2003; Ramoelo et al., 2013), and are linked with numerous essential plant processes. N is associated with chlorophyll content, carbon fixing enzymes, decomposition (Klodd et al., 2016; Gokkaya et al., 2015), and is associated with key nutrients herbivores require, such as protein (Clifton et al.,1994). P is essential for significant molecules within cells like the nucleic acids and Adenosine triphosphates (ATP) (Ramoelo et al., 2013; Epstein & Bloom, 2004). P is also related to other essential plant processes such as metabolism and respiration and is a requirement for lactation by animals (Loozen et al., 2019). These nutrients greatly influence the productivity of savannas and contribute equally towards the growth of vegetation. The ratio of N to P (N:P ratio) is the primary determinant of growth-limiting nutrients (Cech et al., 2008; Gusewell, 2004; Koerselman and Meuleman, 1996). The balance of N and/or P is evinced by the value of this ratio, which affects vegetations' reproduction and growth at all stages. N:P ratio is linked with species richness and composition, productivity, and functionality traits (Cech et al., 2008; Gussewell, 2004), playing a significant role in the sustainability, planning, and management of savanna ecosystems (Craine et al., 2008; Ramoelo et al., 2013). The application of N:P ratio as a proxy of nutrient limitation is supported in the literature (Tessier and Raynal, 2003; Guseweel, 2004). Threshold values of the N:P ratio differ in several studies under different experimental conditions and methods. Currently, there are no conclusions on the final threshold values due to limited research on the N:P ratio (Ramoelo et al., 2013).

Most of the present knowledge on the N:P ratio in savanna ecosystems is based on fertiliser experiments (Craine et al., 2008; Lebauer and Treseder, 2008), use of N:P stoichiometry (Gusewell, 2003), and factorial fertilisation (Olde Venterink et al., 2008). These include assessing the response of net primary productivity with an increase in the supply of a particular nutrient, either N or P, to identify the insufficient nutrient (Vitousek and Howarth, 1991). These traditional field-based and laboratory methods for quantifying grass foliar biochemical content have long been utilised locally (Ramoelo et al., 2012a; Loozen et al., 2019). It is often impossible and challenging to upscale these experiments to the regional level due to several constraints, including the unavailability of data covering larger areas, as well as time, labour, and financial limitations (Pullangari et al., 2013; Shoko et al., 2016; Lu et al., 2016; Karimi et al., 2018; Loozen et al., 2019).

1.2 The Remote Sensing of vegetation foliar biochemicals

The science of collecting information about an object, area, or phenomenon from a distance is known as remote sensing (RS) (Fussel *et al.,* 1986). The information is usually collected by a sensor mounted on handheld devices, spaceborne satellites, or aircrafts, which measures the electromagnetic energy reflected and emitted by these features (Read and Torrado, 2009, Prasad *et al.,* 2011). The remotely sensed data is crucial for understanding changes occurring in the surface of earth and the atmosphere at spatially and spectrally unlimited extents. This makes RS a viable tool for predicting both biochemical and biophysical indicators of savanna vegetation cover, density, and biomass. (Dube and Mutanga 2015; Naidoo, et al., 2012). The synoptic view of the surface and high revisit period offered by different airborne and spaceborne sensors allows for time-effective and dimensionally unlimited means for environmental observations (Cohen and Goward, 2004). RS techniques for grass foliar nutrients estimation and monitoring at a regional scale became popular over the last three decades, and this owes to the above-mentioned RS capabilities providing explicit data for analysis on a near-real time basis at reduced labour (Ramoelo et al., 2011; Loozen et al., 2019). Accurate estimation of foliar biochemicals provided by RS is pivotal for understanding ecosystem functioning at a regional scale. This is due to the relation of foliar biochemical contents to vegetation's chemical processes, including photosynthesis, respiration, evapotranspiration, and decomposition (Huber et al., 2008).

The spectral reflectance and absorption behaviour of electromagnetic energy by vegetation leaves is influenced by their biochemical composition. For example, N and chlorophyll, starch, proteins, phosphorus, oil, and leaf water content determine reflectance characteristics (Usha and Singh, 2013). Absorption of light by N, carotenoids, anthocyanin, and chlorophyll, for example, is high in the red (665nm) and blue (490 nm) regions of the electromagnetic spectrum (Kganyago et al., 2014; Singh et al., 2017), while water, and phosphorus greatly influence absorption in the SWIR (945nm) region (Ramoelo et al., 2013). Numerous biochemical constituents including N, protein, and chlorophyll influence light reflection at the NIR (865 nm) region. The spectral signature curve of healthy vegetation across different wavelengths is illustrated in figure 1 below, indicating the vegetation biophysical and biochemical components influencing absorption or reflection at particular wavelength regions.

Spectroscopy measures these reflectance and absorption features at the field level. This is significant for developing spaceborne and airborne RS instruments to understand the reflectance of different earth features at different wavelengths (Milton et al., 2009; Qiao et al., 2022). Unlike the space-borne sensors, the field spectroscopy instrument is held 0.5m above the surface. Therefore, it is assumed to detect the maximum energy reflected by the earth's features as the distance between the sensor and the object is minimised. Subsequently, the atmospheric effect on the reflected energy is reduced. This basic understanding of electromagnetic energy interactions with surface features (i.e., vegetation in this case) as influenced by vegetation biochemical and biophysical parameters essential for detecting various vegetation compositions from RS datasets. This enables the computation of numerical algorithms, for example, vegetation indices (VIs) that are essential for accentuating vegetation traits using the highly reflected and highly absorbed light. Therefore, identification of N:P ratio responsive wavelength regions from reflectance data prior to mapping is essential for identifying N:P ratio sensitive VIs .

Passive RS utilises the unique spectral signature of vegetation from hyperspectral and multispectral sensors to extract meaningful information explaining vegetation's physical conditions and biophysical traits (Zhang et al., 2015). The utility of hyperspectral and multispectral data for continuous monitoring and estimating grass quality measures are well documented in the literature (Askari et al., 2019; Ramoelo et al., 2015; Mutanga & Kumar, 2007). Even though the data from spaceborne and airborne hyperspectral datasets are known to generate accurate vegetation biochemical predicting models, the application of hyperspectral data is restricted by the unavailability of these datasets at a regional extent and the expensive costs associated with data them (Tong et al., 2014), especially in underdeveloped countries like South Africa (Adjorlolo et al., 2015). Consequently, a thorough investigation of the potential of cost-effective multispectral sensors offering large volumes of data at a regional scale for the estimation of grass micronutrients is a necessity.

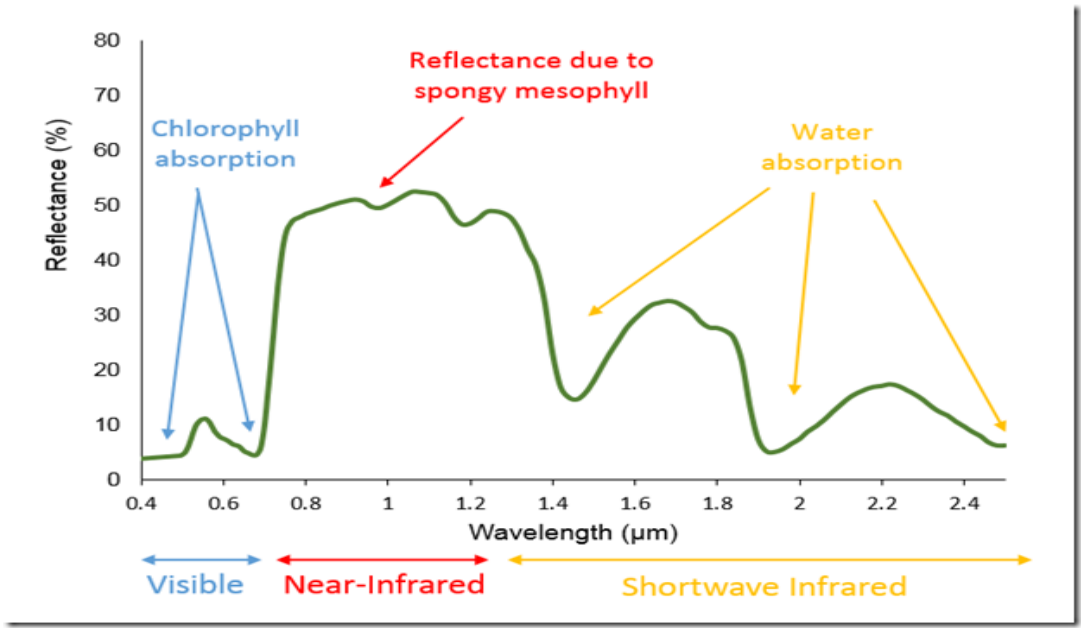

Figure 1:The spectral signature of vegetation, (source: gsp.humboldt.edu).

Retrieval of the N: P ratio relies on accurate foliar N and P prediction. However, P prediction becomes constrained by several challenges, including a few identified P absorption wavelengths (Ramoelo et al., 2013) and the small concentration of P in plants. P is 10 times lower than N, making it challenging to detect P from a satellite image. Hence, there are few RS studies on P estimation (Knox et al., 2010). The challenges experienced during P estimations consequently affect RS-based N:P ratio retrieval. Based on the reviewed literature, limited studies, e.g., Ramoelo et al. (2013), investigated the applicability of RS datasets in estimate the N:P ratio at a regional scale in South African Savannas. In recent times, most vegetation assessment studies utilised hyperspectral data owing to the presence of a considerable number of narrow bands, including the red edge providing information on even subtle variations in spectral reflectance of vegetation possible resulting from changes in vegetation biochemical and biophysical composition. However, models developed using hyperspectral data are highly subjected to data redundancy and overfitting (Mutanga & Kumar, 2007). In addition, most hyperspectral sensors are commercial; hence their availability to the public is limited, especially in underdeveloped countries.

Inclusion of the red-edge bands in broadband sensor like WorldView-2 and RapidEye have been proven effective in dealing with the issue of overfitting and presents an excellent opportunity for detecting changes in vegetation's nutrient content (Mutanga et

al., 2012; Clevers and Gitelson, 2013; Ramoelo et al., 2012). However, financial constraints are still a significant drawback, making Sentinel-2 the only financially viable sensor for the remote sensing of vegetation. Sentinel satellites developed by ESA Copernicus are meant to ensure the continuity and improvement of satellites assigned for earth observation. Sentinel-2 satellite mission contributes significantly to earth surface monitoring, and it is equipped with three red edge bands (705 nm, 740nm, and 783 nm) known to be effective in determining green vegetation health and status (Drusch et al., 2012). This satellite comprises a wide swath of 270 km and offers a high spatial resolution of 10 m and 20 m, enabling up to regional vegetation monitoring (Duraisamy, 2019). Owing to the inclusion of red-edge bands to Sentinel-2 Multispectral Instrument (MSI), which were initially available only with the commercial sensors. It can also be argued that Sentinel-2 is comparable with commercial sensors and is sufficiently great for assessing vegetation quality (Ramoelo et al., 2015).

Several studies investigated the capability of Sentinel-2 MSI absorption features and vegetation indices for estimating biochemical and biophysical traits. For example, Ramoelo et al. (2015), Ramoelo and Cho, (2018) successfully investigated the capabilities of Sentinel-2 simulated spectral bands in the estimation of leaf nitrogen, and Chabalala et al. (2020) used sentinel-2 MSI satellite images to map the spatial distribution of nitrogen. To date, the potential of sentinel-2 MSI spectral configuration for estimation of N:P ratio are seldomly investigated (Loozen et al., 2019). Sentinel-2 MSI is currently providing large volumes of data, which opens research opportunities to investigate the capabilities of these datasets in estimating and monitoring the grass N:P ratio. Regression techniques mainly used to develop vegetation assessing models are reportedly simple linear and parameterized regression techniques. These methods are successfully used to study several vegetation parameters with acceptable accuracies. However, their inability to handle large and complex datasets and the issue of overfitting is still a significant challenge making it difficult to upscale them for regional application. Parametric regression techniques perform best with normally distributed data. Their accuracy is limited when the case is the opposite (Zaetizadeh et al., 2011). In this regard, to fully exploit the potential of non-commercial sensors (like Sentinel-2 MSI) that produce data at high spatiotemporal resolution data, there is a necessity to compare its performance on both parametric and non-parameterized methods as the latter approach takes advantage of the full spectrum

and can handle complex datasets. The main objective of this study was to develop a parametric or non-parametric Sentinel-2-based N:P ratio estimation model using simulated *in situ* reflectance data obtained from hyperspectral Analytical Spectroscopy Device (ASD) and later invert the developed model to map the distribution of N:P ratio across Kruger National Park (KNP) area and surroundings.

1.3 Problem statement

Most of the world's Savanna biome, including South African Savannas, manifest signs of degradation in response to overgrazing (Ramoelo *et al.,* 2015). In addition to this soil erosion, conversion of open grasslands into crop production farms, weed and bush encroachment (Greer *et al.,* 2014; Symeonakis *et al.,* 2016), and changing climatic conditions (Palmer & Ainsile, 2005; Zhou *et al.,* 2015) are also factors. Symeonakis *et al.* (2016) reported that millions of hectares of South African land are degraded, and 1.1 million hectares is regarded as non-viable. This resulted in a 50 % decrease in the grazing capacity of the South African savanna in 2016 (Symeonakis *et al.,* 2016). The rapid incline in the degradation rate threatens the ability of the savanna to provide the ecosystem services mentioned above. When savanna services are disrupted, especially grazing capacity, the poor and vulnerable rural communities who rely solely on natural rangelands found within the savanna biome are disproportionately more affected (Symeonakis *et al.,* 2016).

For decades, farmers lost livestock owing to poor rangeland management techniques. This may result from a lack of information that enables farmers to implement relevant actions such as rotational grazing to allow pastoral recovery. Hence, there is a need for a reliable and financially affordable approach to savanna management to ensure its sustainability. The following core problems were addressed in this study: a) lack of information on the distribution of the N:P ratio across the South African Savannas, b) Lack of information on the potential freely available Sentinel-2 datasets for predicting foliar N:P ratio and, c) the accuracy of simulated Sentinel-2 datasets derived models for continuously mapping the spatiotemporal distribution of N:P ratio in heterogeneous natural environments.

1.4 Significance of the study

Monitoring the savanna biome is crucial in the agricultural economy, especially for the smallholder farmers in rural areas who depend mainly on natural grasslands for livestock grazing and productivity. The information is essential to farmers, ecologists, and nature conservationists as a decision-making support tool to manage and sustain protected areas such as national parks and farms. Regular information is required to ensure the development of risk management and coping strategies to avoid dealing with the consequences of nutrient limitation. Moreover, decision-makers and policymakers formulate, and update policies related to land management and food security as such actionable information that RS techniques provide is critical to them. The use of advancing technologies and innovative means for timely monitoring of vegetation is crucial for eradicating poverty and ensuring food security and sustainability and restoration of savanna biodiversity.

1.5 Hypothesis

Does the Sentinel-2 MSI offer comparable accuracy as the commercial Multispectral and hyperspectral sensors in the estimation of grass N:P ratio?

1.6 Aim

The aim of this study was to evaluate the capabilities of simulated Sentinel-2 models for N:P ratio estimation and assess their transferability to satellite images to map grass nutrient limitation.

1.7 The main objectives of this study were:

- To determine the optimal Sentinel-2 spectral bands and vegetation indices for retrieving foliar N:P ratio,
- To evaluate the transferability of FieldSpec 3-(ASD) derived N:P models to multispectral satellite images for regional mapping, and
- To compare N:P ratio predictive capability of Sentinel-2 to commercial multispectral and hyperspectral sensors.

1.8 Research questions

The research questions that this study addressed were:

i. What are the optimal Sentinel-2 bands and indices for predicting foliar N:P?

ii. Can ASD-derived models be transferred to satellite images for mapping the N:P ratio?

iii. Is Sentinel-2 N:P ratio predictive capability comparable to the commercial multispectral sensors equipped with the red-edge bands?

1.9 Scope of the study

This study explored the capability of simulated multispectral and hyperspectral datasets for estimating and mapping grass foliar N:P ratio in the savanna landscape of Limpopo province, in the Kruger National area. Simulated sentinel-2 datasets were used to develop an N:P ratio predicting model. The potential of sentinel-2 was explored using both parametric and non-parametric regression methods, partial least squares regression (PSLR), and random forest regression (RFR), respectively. The accuracy of Sentinel-2 was then compared to commercial multispectral (WorldView-2 and RapidEye) and hyperspectral sensors (Enmap and Hyperion). Sentinel-2 images were used to map the distribution of the N:P ratio by inverting the best-predicting model derived from the simulated Sentinel-2 dataset.

1.10 Research outline

This thesis consists of five chapters. The main purpose of this study is to detect grass nutrient limitation from simulated Sentinel-2 spectra. It is based on ASD field measurement and laboratory chemical analysis to extract grass N:P ratio. Regression analysis is be executed in chapter 3 to measure the relationship between the simulated spectra and the laboratory-extracted concentrations and finally investigate the transferability of Simulated ASD datasets to Sentinel-2 images for regional mapping of N:P ratio.

Chapter one

Chapter one is an introductory chapter outlining the background of the study, and it introduces the critical role of grassland in maintaining biodiversity and its role on the ecosystem, rural livelihoods, and the economy. It introduces the problem that has been

investigated and the importance of remote sensing-based nutrient estimation. This chapter presents the objectives to be executed proceeding with the aim of the study and the hypothesis to be tested.

Chapter two

Chapter two is a review of previous work related to the estimation of grass quality using remote sensing. The use of hyperspectral and multispectral data for grass nutrient mapping is discussed in this chapter, looking at the challenges and advantages of each dataset. It further reviews important predicting variables (i.e., vegetation indices, red-edge position (REP), and absorption feature) for the N:P ratio. Parametric and non-parametric regression techniques were reviewed, focusing on their pros and cons and their success in previous experiments based on obtained estimation accuracies and gaps in the literature in relation to N:P ratio estimation is identified in this chapter.

Chapter three

Chapter three provides a clear description of the study area, methods, and methodology used in this research. The chapter explains the acquisition of all the datasets used in this research, the sampling methods, and the procedure used to measure spectral reflectance. Lab-based chemical analysis of N:P from the grass samples was described, and the spectral resampling of the ASD hyperspectral data to the desired spectra was also explained. Chapter three clearly describes regression algorithms (SPLR and RF) used to estimate and map the N:P ratio, and the statistical analysis performed on the data is further explained.

Chapter four

Chapter four is the presentation of obtained results. Results were presented in scatter plots and graphs showing the relation between measured and predicted N:P ratio. The contributions of bands and VIs were analysed in this chapter. Furthermore, the performance of each sensor was analysed based on the estimation accuracies of the SMLR and RF models.

Chapter five

The results achieved in this study are discussed in chapter 5, the overall summary on the achievement of the objects and aim of this research study are also presented.

Chapter five also gives a discussion of the contribution of this study to scientific knowledge and the recommendation for future research to improve the prediction and mapping of N:P ratio to improve the results and the accuracy achieved in this study.

CHAPTER 2:

Literature Review

2.1 Chapter overview

This chapter provides review of previous RS applications for predicting and mapping grass quality. The drivers of variations in grass quality were reviewed. Followed by the applications of Hyperspectral and multispectral datasets in estimation of foliar biochemicals in different environments. A further review of bands and indices applicable to the prediction of foliar micronutrients, focusing mostly on N, P and N:P ratio is provided in this chapter. Lastly, both parametric and non-parametric regression algorithms are discussed in this chapter.

2.2 Background

Grass quality, as denoted by the concentration of grass biochemicals such as N and P is imperative for the health and development of vegetation. Grass N and P are associated with several plant processes, including light use efficiency, plant photosynthetic capacity and rate, leaf life span, chlorophyll content, and plants' amino, and nucleic acids (Walker *et al.,* 2014; Bakker *et al.,* 2011, Gusewell, 2004; Wright *et al.,* 2004). Availability of grass N alters dry matter, and protein production, which is the major nutrient required by plants (Prins & Langevelde, 2008), foliar P, on the other hand, is the major nutrient requirement for lactating animals (Mutanga & Kumar, 2007). Nutrient limitation is demonstrated by the N and P quotient (N:P). This ratio indicates the balance of foliar N and P in an ecosystem. Nutrient limitation is known to have implications on the feeding patterns of foragers as N and P limitation determines forage productivity and biomass. As such inadequate levels of these biochemicals will lead to low quality and reduced plant growth (Zhao & Zeng, 2009).

Different critical N:P ratio values are proposed in the literature. These values indicate whether the grass is N, P, or co-limited. Based on short-term fertilised plants, Gusewell, (2004) proposed that small N:P values (less than 10) denote N limitation and larger N:P values (greater than 20) indicate P limitation, and a condition where both Nand P are insufficient is known as co-limitation; this is denoted by any N:P values in between

10 and 20. Gusewell, (2004) further posits that these values are subject to change with different plant species. Koerselman and Meuleman (1996) initially proposed a narrow range of values based on a wetland's ecosystem experiment, with N deficiency indicated by N:P ratio < 14 while N:P ratio > 16 indicating P deficiency, and N:P values between 14 and 16 suggesting co-limitation. However, the later N:P ratio threshold values cannot be applied in savanna ecosystems due to differing grass species between the two ecosystems, wetlands are mainly made of C3-type grass, whereas grassland is made of C4-type (Craine *et al.,* 2008, Ludwig *et al.,* 2001). Among the various factors influencing foragers' feeding site selection preferences, the concentration of grass micronutrients stands out as a particularly significant factor. In addition to factors like predation risk, water availability, grass species type, age, and phenology, the spatial variability of nutrients has been found to have a substantial impact on foraging behavior and migration patterns of foragers (Prins & van Langevelde, 2008; Ludwig et al., 2008). It is postulated that large animals tend to be denser in nutrient-rich areas and less dense in nutrient-limited areas.

2.3 The influence of environmental factors on the distribution of grass quality

Understanding biotic and abiotic factors influencing the spatio-temporal variations in grass quality is of fundamental importance in descerning the feeding trends of animals in savanna biomes (Mutanga *et al.,* 2004). Several studies indicated that several environmental factors drive variability in grass quality, however, elucidating spatial patterns of vegetation health is still a complicated task because the spatial distribution of grass quality is not solely explained by one independent factor, but by an array of different interacting factors simultaneously (Venter *et al.,* 2003). Evidence of those inter-relationships is presented by the significant correlations between some of those variables. For review, see (Ramoelo *et al.,* 2013). Mutanga *et al.* (2004) used correlation analysis and Analysis of Variance (ANOVA) to examine the correlation of grass quality with environmental parameters at a local scale. Their study indicated an important relationship between environmental factors, focusing specifically on the percentage of grass cover, soil texture, slope, aspect, and altitude. Soil texture and percentage of grass cover was found to be crucial in explaining grass quality variability at a local scale.

Mutanga *et al.* (2004) further investigated interactions between biological and azoic factors affecting the spatial distribution of grass nutrients using multivariate analysis techniques. Noticeable interactions between plant species type with altitude, slope, and underlying geology were identified. Tlhone *et al.* (2018) investigated the importance of environmental variables for estimating grass quality (N), focusing mainly on climate, land use, slope, aspect, altitude, and geology using the spatial and least squares (Analysis of Covariance (ANACOVA)). From the study, the models demonstrated the significance of each factor in assessing grass quality variability. Ramoelo *et al.* (2013) integrated climatic, edaphic, and topographic environmental variables with *in situ* hyperspectral ASD data to predict grass quality (N and P) using non-linear partial least squares regression (PLSR). The results demonstrated a positive correlation between climate, geology, and topographic factors, as shown by positive PLSR weights of these parameters.

The influence of these environmental parameters on grass quality distribution varies with spatial scale. At a regional or continental scale, a noticeable influence on forage quality results from climatic conditions, particularly rainfall and temperature. There are no vast variations at the local level as climatic differences are not significant over a few kilometres except in the mountains (Pearson & Dawson, 2003). These affect the physical and chemical traits of the forage as well as the digestive processes of the foraging animals. It is understood that rainfall has a positive influence on plants as they require water to maintain soil moisture and nutrients. However, excessive rainfall than required (flooding) promotes diseases in plants. In the savanna biome, local differences in rainfall occur due to rainfall incidents, for example, thunderstorms. Moreover, rainfall varies on annual records, and there are no noticeable differences on monthly and daily basis (Venter *et al.,* 2003). In a three-year period Ferner *et al.* (2021) identified precipitation and plant processes to be the key factors deriving forage supply. Temperature is another climatic factor influencing plants' physiology. Drought intensity and warming level determine a plant's ecophysiological changes in response to heat stress. Forage quality strongly decreases with an extreme increase in heat due to tissue senescence, while average heat stress promotes plant maturation, the utmost increase in heat causes an increase in water-soluble carbohydrates and decreases plant tissues' water content (Dumont *et al.,* 2017). Furthermore, increased heat stress reduces dry matter digestibility and reduces voluntary intake by the feeding animals (Baetty *et al.,* 2008).

Distribution of grass health status at a local scale affected by several parameters such as soil physical and chemical characteristics, topography, underlying geology as well as fire frequencies are understood to have an impact (Pearson & Dawson, 2003). Soil's physical and chemical properties affect the nutritional status of vegetation. These properties are directly influenced by parent rock material. Soil originating from gabbro is more nutritious than basalt originating soils. Subsequently, plants occurring on higher-nutrient soils, such as basalt, have higher concentrations of leaf nutrients. Growth rates and forage quality differs significantly on sandy and clayey soils in response to water availability (Kumar *et al.,* 2002). This is due to the varying infiltration capacity of these soils; sandy soils are more permeable, allowing more water to pass through to reach vegetation roots hence the accelerated growth, whereas on clay soils growth rate is hampered by low permeability leading to low infiltration and high surface runoff. However, clay soil contains high mineral content and is more nutritious than sandy soil. Run-off in steeply sloped areas causes minerals and clay particles to be deposited into the valley leading to thinner and coarser texture soil layers, hence low nutrient concentrations. Valleys are associated with high nutrient concentrations owing to the high accumulation of water and nutrients from the mountainous areas (Grant *et al.,* 2000).

Fire frequency is another important factor influencing vegetation quantity as well as quality in savannas, resulting in decreased biomass production (Ben-Shahar, 1996), accelerated nitrogen mineralization (Giardina & Rhoades, 2001), and an increase in post-fire growth (Rieske, 2002). Fires remove built-up layers of decaying materials from past years and decapitate vying trees and shrubs, allowing grasses to grow in more space and with favourable light, water, and nutrient availability. (Bachelet *et al.,* 2000). However, increased fire frequency may lead to overgrazing in the unburnt areas, which may result in degradation and reduced rangeland quality. At Kruger National Park, the relationship between geology and fire frequency is explained by Venter *et al.* (2003) as follows: "Where there is high grass biomass and fires are more intense, such as on the basalt plains, the savannas are more open, and where fires are less intense, such as on the granites, they are more closed". Increased and intense fire frequencies stimulated the coalescence of shrubs through boscaging of the mopane veld (Bronn *et al.,* 2001). Therefore, leading grass condition influencers at KNP are fire frequencies and the parent material.

2.4 Foliar quality estimation using remote sensing.

Passive remote sensing is dependent on radiation from the sun to illuminate objects. It is based on detection of the interaction of light with features at a specific wavelength by multispectral and hyperspectral to detect and monitor changes occurring on the surface of the earth. These include water bodies, built-up and vegetation, the summary of passive remote sensing process is shown in figure 2 below and figure 3 demonstrates the interaction of these objects at different wavelength regions of the electromagnetic spectrum. Its primary spectral regions of interest are the visible, near infrared, and shortwave infrared (Zhang *et al.,* 2015). The RS of vegetation uses the reflected or absorbed energy to understand vegetation Physical and biochemical conditions. Multiple reflectance values are combined (i.e., computation vegetation indices) from the vegetation's distinct spectral reflectance curve to accentuate traits of interest. Both vegetation biophysical and biochemical parameters could be mapped using vegetation indices (VIs) (Zhang *et al.,* 2012). Developing high spectral resolution satellites with additional effective wavelength bands allows for a robust investigation of grass quality parameters (Oumar & Mutanga, 2007). The use of vegetation indices in estimating grass nitrogen and phosphorus is salient in remote sensing. Several airborne and spaceborne hyperspectral and multispectral sensors are used to estimate grass quality. Techniques developed using the large number of wavelengths recorded by hyperspectral imaging sensors allow for the development of multiple VIs, offering opportunities for exceptional investigation of vegetation conditions (Mutanga, 2004).

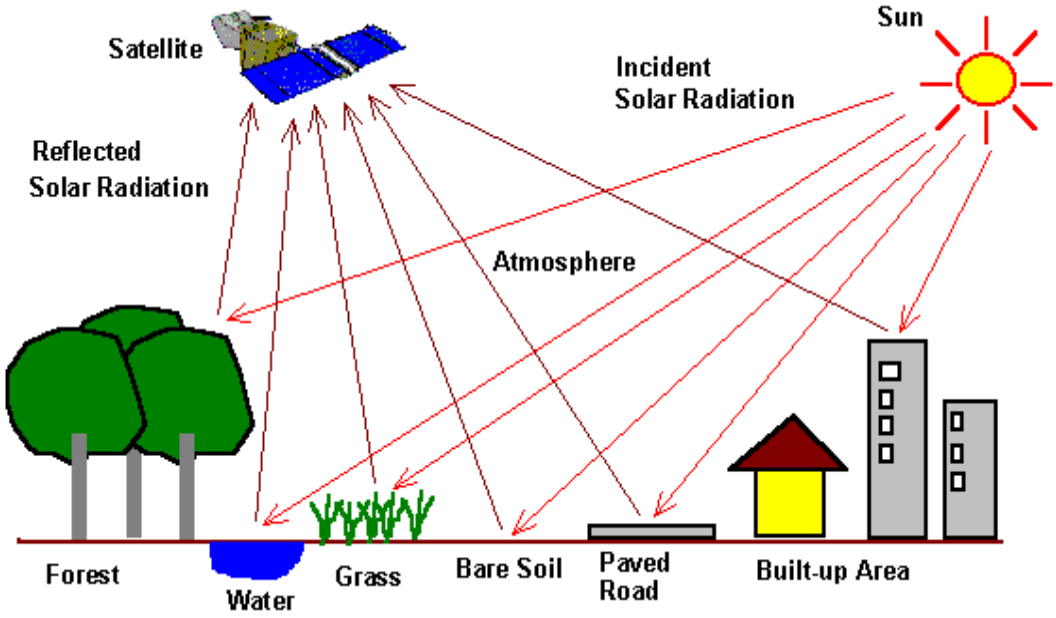

Figure 2: The process of passive remote sensing, . The diagram shows the illumination of the target by the source and the recording of the reflected electromagnetic energy by a spaceborne sensor mounted on a satellite (source: researchgate.net).

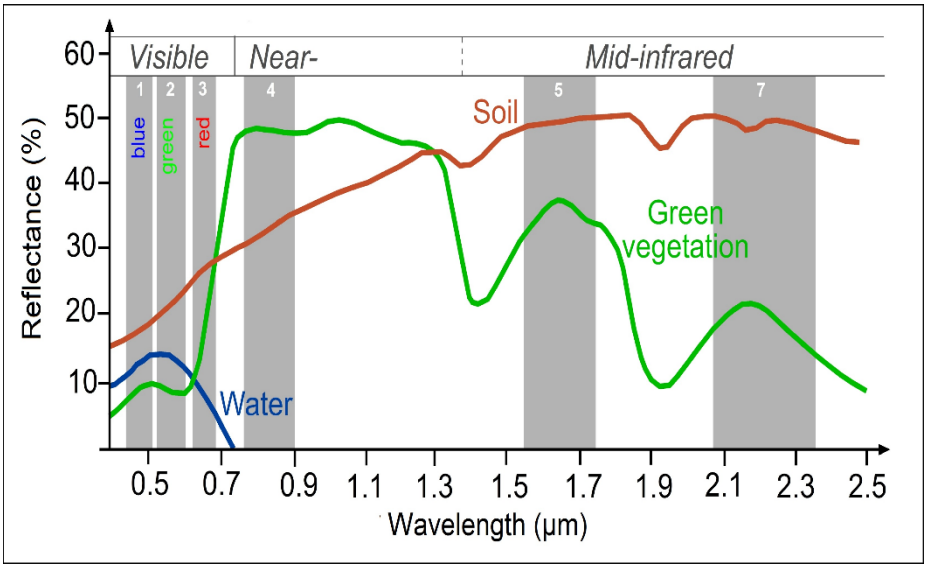

Figure 3: The spectral signatures of different objectsThe figure shows the reflection and absorption of light by different objects at different wavelengths. Soil (the brown line), water (the blue line) and vegetation (the green line) (Source: seos-project.eu.).

Regional assessment and mapping of grass nutrients using hyperspectral is hampared by the significant costs and unavailability of data at a larger scale (Tong *et*

*al.,*2014; Ramoelo *et al.,* 2015). In addition, overfitting resulting from the large number of wavelengths of hyperspectral images present a challenge to using hyperspectral data (Mutanga and Kumar, 2007). Additional effective bands in multispectral sensors such as the red edge band and the wide swath make these sensors robust for regional mapping of grass biochemical content (Ramoelo *et al.,* 2012, Cho *et al.,* 2006, Oumar & Mutanga, 2013; Li *et al.,* 2014). Presence of this red edge in the non-commercial sentinel-2 MSI makes this satellite comparable with commercial multispectral satellites (Sibanda *et al.,*2015; Loozen *et al.,*2019, Ramoelo., 2014). Loozen *et al.,* (2019) investigated the performance of vegetation indices developed from commercial and non-commercial sensors for N and P in estimating the N:P ratio using field spectroscopy datasets. Resampled Sentinel-2 derived VIs showed high correlation with the N:P ratio (R^2 of 0.72).

2.5 Grass quality estimation using simulated field spectroscopy data.

Field spectroscopy has long been utilised as a tool to explain *in situ* vegetation measurements. It contributed significantly to the understanding of vegetation spectral signatures and identifying important regions for the detection of vegetation spectral traits. This later led to the development of imaging spectroscopy (Goetz, 2009; Milton, 2009). The main advantage of these tools is obtaining accurate reflectance of objects as they are held at a shorter distance from the surface during spectral measurements, hence the impact of the atmosphere on radiation energy is minimal (scattering and absorption of light by atmospheric particles), resulting in almost the exact reflection of objects. Since the surfacing of field spectroscopy in optical RS, this has played a prominent role in scaling up radiation-object interaction from a finer scale of the Earth's surface to a coarser canopy level (Milton, 2009b). Field spectroscopy applications in vegetation remote sensing include the detection of nitrogen, biomass estimation, assessment of vegetation health, and species classification. A study by Ramoelo *et al.,* (2013) integrated hyperspectral field spectroscopy datasets with environmental variables to test the performance of nonlinear partial least squares regression (PSLR) for detecting grass N and P in savanna ecosystems. High estimation accuracies R^2 0.81 and root mean square error (RMSE)= 0.08 g/m^2 for N, R^2 = 0.80, and RMSE = 0.03 g/m^2 for P) were obtained from the non-linear PSLR model when integrating *in situ* hyperspectral and environmental variables (climatic, topographic and geology) compared to using spectral data only. This

emphasised the influence of different environmental factors on the spectral reflectance received at the canopy level.

Mutanga & Skidmore, (2010) used airborne HyMap image data to test the effectiveness of imaging spectroscopy and neural networks to map P concentration in savanna grasses. This study further attempted to find the key wavelengths to detect phosphorus using hyperspectral remote sensing. Continuum-removed absorption features and the red edge position (REP) were used as predictors in a propagation neural network. The results showed a significant potential of a best-trained neural network and HyMap data to estimate phosphorus distribution (R^2 = 0.63 and RMSE = 0.07 g/m^2). A similar experiment by Ramoelo *et al.* (2013) investigated the potential of field spectroscopy reflectance data and simulated HyMap spectra for measuring the N:P ratio. N:P predicting models were developed from original field spectral reflectance data and resampled or simulated HyMap data. Spectral transformation techniques were used to minimise the effect of other absorbing features in the SWIR region of the electromagnetic spectrum. The original field spectra demonstrated high N:P estimation potential when combining water removal (WR) and continuum removal (CR) spectra with the PSLR compared to other spectral transformation techniques (i.e., log 1/R and first derivative (FD)). RMSE value of 1.12 was obtained for both CR and WR with the highest R^2 values of 0.85 and 0.81, respectively. The simulated HyMap data predicted the N:P ratio satisfactorily. However, by comparing the obtained RMSE and R^2 values, the results show that original reflectance data outperformed the HyMap data for all the spectra transformation techniques with RMSE ranging from 1.12 to 1.50 for field spectra and 1.40 to 1.70 for HyMap spectra and R^2 values ranging from 0.69 to 0.85 for field spectra and (0.50 to 0.64) for HyMap spectra. WR showed high accuracy for HyMap spectra indicated by the lowest RMSE of 1.40.

Pang *et al*. (2022) used enhanced ASD spectrometer reflectance data to develop PLSR models for C, N, and P estimation in an open grassland, the models attained over 0.75 R^2 for all the biochemicals. The study further compared the results with raw Sentinel-2 images, and the RMSE of the models improved significantly. Qiao *et al*. (2022) used the ASD spectrometer data to identify P sensitive bands and develop P prediction in maize leaves to monitor growth under different P fertiliser levels. The study applied back propagation neural networks to develop predictive models, and identified P sensitive wavelengths at 763 nm, 815 nm, and 900–1000 nm. Sibanda *et al.,* (2015) assessed the

utility of in situ hyperspectral grass reflectance data and multivariate techniques in distinguishing the effect of different fertilisers on grass quality. Specifically, 12 complex ammonium nitrate, ammonium sulphate, lime, and phosphorus fertiliser combinations were investigated. Partial least squares regression discriminant analysis (PLSR-DA) and discriminant analysis (DA) were used to assess grass quality under fertilised and unfertilized plots. In situ hyperspectral data demonstrated great potential in detecting different fertilisers. Four effective bands were identified within the red edge (731 and 737 nm) and the shortwave infrared (1310 and 1777 nm).

Loozen *et al.* (2019) investigated the applicability of vegetation indices (VIs) developed for canopy N or P detection in estimating canopy N:P detection using *in situ* hyperspectral. To investigate the impact of band composition and bandwidth 60 VIs developed from the original *in situ* hyperspectral data and simulated the data to band characteristics of six multispectral satellites: sentinel-2 MSI, Sentinel-3 Ocean and Land Color Instrument (OLCI), RapidEye, Worldview-4, Moderate Resolution Imaging Spectrometer (MODIS)- Terra, and Aqua, and Landsat-8. Out of the 60 VIs, some were the existing VIs, others explicitly optimised for this study. Both VIs were significant for N:P detection, (R^2 = 0.16 – 0.48 and R^2 = 0.59 -0.72) for existing and optimised VIs respectively. Sentinel-2 MSI and Landsat-8 Operational Land Imager (OLI)-derived VIs yielded a high correlation with N:P for existing VIs; sensor parameters did not influence the optimisation. The capability of simulated Hyperspectral and multispectral datasets, particularly Hyperion and Sentinel-2 with PLSR is presented by Ferner *et al.* (2021). Their study further assessed the transferability of the models to satellite images, even though Sentinel-2 obtained lower accuracies than Hyperion, it produced more realistic maps. Accuracy of ASD data derived nutrients prediction models is affected by different vegetation phenological stages (Amaral *et al*., 2022). Their study estimated leaf P, K, Ca, and Zn at three different growth stages using spectrometer reflectance data, the results revealed that the accuracy of the models was varying with different stages.

2.6 Grass quality estimation using absorption features.

Absorption features for nitrogen, protein, chlorophyll, and starch were identified in several studies. Remote sensing of phosphorus mapping is lagging compared to nitrogen, subsequently, known absorption features sensitive to P are limited and phosphorus

estimation is still a challenging venture (Loozen *et al.,* 2019). Since phosphorus is essential in metabolism processes responsible for producing sugar, absorption sensitive to starch is used to detect phosphorus. Knox *et al.* (2006) used sugar absorption features to estimate P and obtained accurate results. The shortwave infrared (SWIR) capability for P estimation was demonstrated in RS studies, for example (Mutanga & Kumar, 2007). Mutanga and Kumar, (2007) identified absorption features located in the shortwave infrared (R2015 - R2199) to be robust for mapping phosphorus distribution. This region received little attention before in spectroscopic-based phosphorus experiments compared to the visible region. The reason for this was due to the interference of absorption radiation by several compounds in this region, for example, absorption by leaf water content is strong in this region thus making it difficult to interpret the SWIR absorption (Mutanga & Kumar, 2007). SWIR absorption features are enhanced with spectral transformation techniques such as water-removed spectra, see (Ramoelo *et al.,* 2011) and continuum removal technique (Mutanga & Kumar, 2007) to accentuate vegetation biochemicals in this wavelength region. For P estimation, Gao *et al.* (2019) indicated that first derivative (FD) and continuum removal (CR) spectral transformation techniques retrieve more absorption features located in the NIR and SWIR compared to log-transformed and Raw spectra. In multispectral sensors, the use of SWIR was restricted by a few bands in this region (Ramoelo *et al.,* 2013).

N and P estimation using RS was used to derive N:P, However, this approach is likely prone to errors. For instance, errors that occurred in N and P measurements are automatically carried to N:P measurements; therefore, identification of N:P absorption features is necessary for RS of nutrient limitation. Ramoelo *et al.* (2013) utilised protein, sugar, and nitrogen absorption features to estimate N:P ratio and attained meaningful results. The Blue region is sensitive to N:P; however, Raleigh scattering is challenging for satellite sensors. The position of absorption features changes with seasons. For instance, in the wet season, nitrogen features are in the visible, NIR, and SWIR and are found in the SWIR only in the dry season (Asner *et al.,*1998). Known absorption features sensitive to both N and P are listed in Table 1 below, these are the central wavelengths in which the absorption features are situated. Absorption of light in these wavelength regions indicate the presence of either N or P in the observed sample.

Table 1: N and P known absorption features.

Foliar biochemical	Absorption features/ wavelengths (nm)
N and P	430 nm, 460nm, 640 nm, 660nm, 910 nm, 1510 nm, 1690 nm, 1940 nm, 1950 nm, 2060 nm, 2180 nm, 2300 nm, 2350 nm, 2270 nm

2.7 Grass quality parameters using vegetation indices.

The oldest and most common technique used to study vegetation traits involves a combination of reflectance values from two or more wavelengths into a single numerical variable, i.e., computation of VIs enhancing spectral features sensitive to vegetation (Clevers, 2014). Application of VIs originated with the use of broadband sensors. Normalised difference vegetation index (NDVI) (Rouse *et al.,* 1974) and simple ratio (SR) (Jordan *et al.,* 1969) are one of the oldest optical VIs and the most used vegetation indices to study vegetation biochemical as well as biophysical traits for decades. The oldest technique of deriving these indices was based on the strongly absorbed red band and the highly reflected near the infra-red band. Net primary production is strongly associated with NDVI; therefore, in remote sensing, NDVI was widely used as the key measure of vegetation variations (Vlek., *et al.,* 2010; Field *et al*., 1995). Bio-indicators such as nitrogen and chlorophyll were extracted using the NDVI derived from narrowband hyperspectral and broadband multispectral satellites. However, broadbands derived VIs, which include indices derived from the broad NIR like NDVI are affected by soil brightness, illumination, viewing atmospheric geometry conditions as well as the leaf canopy (Huete *et al.,*1992), and therefore broad wavebands multispectral sensors are unsuitable for quantification of plant biochemicals (Cho *et al.,*2006; Chabalala *et al.,*2020). Consequently, Vis derived from broadbands is suitable for noticeable mapping differences, such as differentiating between the healthy green vegetation and yellow stressed vegetation.

2.8 Sensitivity of the red edge bands to vegetation parameters

The use of the red edge position for the detection of plant biochemicals became popular in the early 1980s (Guyot & Barret, 1988). The advent of new wavebands extremely sensitive to chemical changes in vegetation was motivated by the response of

plants' reflectance to changes in nitrogen and chlorophyll contents (Main *et al.,* 2011). Since then, several multispectral and hyperspectral satellite sensors with the red edge band were developed. The red-edge bands, firstly available with the RapidEye sensor are proved to be significant in quantifying the biochemical constituents of plants. The red-edge position located between 670 nm and 780 nm is presented by a sudden rise in the reflectance of vegetation, due to the combination of strong chlorophyll absorption in the red band and high reflection in the NIR band (Horler *et al.,* 1983). Leaf structure and chlorophyll content alter the shape and position of the red-edge. Mutanga *et al.* (2003) stated that increased chlorophyll and leaf water content results in expanded red band absorption, causing the red-edge to move towards the longer wavelengths. Whereas reduced leaf water content and chlorophyll cause a shift of the red edge towards a shorter wavelength (Rock *et al.,*1988). Horler *et al*. (1983) were one of the first studies to prove the significance of the position of the red-edge inflection point for detecting plant stress. Since these first publications, the red-edge position (REP) has become trendy for estimating chlorophyll content.

Estimating chlorophyll and nitrogen simple ratio and normalised difference vegetation indices derived from the REP perform better than the traditional VIs computed from the NIR and the visible region of the electromagnetic spectrum (Curran, 1989; Clevers *et al.,* 2001). Due to the saturation effect resulting from the strong absorption of chlorophyll in the red band, chlorophyll and nitrogen absorption are minimal in the red-edge region. Thus, reducing the saturation effect. Hence REP is sensitive to nitrogen and chlorophyll (Gitelson and Merzlyak, 1996). The efficacy of red-edge indices for grass quality and quantity were demonstrated as reported by Ramoelo *et al.* (2014). Red-edge derived VIs were found to be important predictors for leaf N. Both visible and red edge were crucial for biomass. A recent study by Gao *et al.* (2020) utilised Sentinel-2 MSI images to investigate the capabilities of Sentinel-2 red-edge position in predicting grass N:P ratio in alpine grasslands using a random forest regression method. The results revealed the promising potential of spaceborne Sentinel-2 data with fare estimation accuracies. R^2 values of 0.49 and 0.59 were obtained for vigorous growth and senescence stage consecutively. VIs derived from Sentinel-2 B5 and B8A had an outstanding contribution to N:P estimation. In addition, Sentinel-2 B9 and B12 were found to be effective.

Parametric regression techniques for grass quality estimation

Parametric regression techniques have long been the most prevalent in remote sensing. These methods assume a normal distribution of the datasets or datasets that can be satisfactorily modeled by a probability distribution with a constant set of parameters. Regression is based on systematically deriving specified formulas linking restricted spectral bands with the bio-geophysical parameter of interest (Verrelst *et al.,* 2015). Modest spectral characteristics are employed to minimise undesirable impacts, which are frequently linked to changes in leaf structure, soil reflectance, luminance, viewing geometry, and weather systems (Verrelst, 2010). The oldest and simplest approach employed by these regression algorithms to map vegetation parameters is vegetation indices (Glen *et al.,* 2008). Linear and non-linear fitting functions regress and express the relationship between vegetation indices and traits of interest. Parametric regression models are sensitive to environmental changes and lose their predictive power when sensor configuration is changed (Verrelst *et al.,* 2015). As such, the utility of this approach is only suitable for mapping vegetation traits at a smaller scale under fixed conditions. Moreover, parametric regression is limited to massive data with non-linear relationships (Zaetizadeh *et al.,* 2011). Attempts to mitigate these challenges were made by using large datasets created by leaf canopy radiative transfer models. Nevertheless, the challenge remains, data simulated from limited discontinuous bands do not represent a detailed reflection of ground variations, and some spectral information is left out when computing vegetation indices from simulated data.

Simple linear regression techniques present reliable estimation performance. Even though that is the case, curse dimensionality was incurred when samples are fewer than the predictors as these models depend on the estimation of co-variances (Hughes, 1968). Minimising dimensionality improves linear models' accuracy. Stepwise multiple linear regression (SMLR) and partial least square regression (PLSR are the most used linear regression techniques in mapping vegetation traits. SMLR is commonly used to select bands carrying the most important information (Dorigo *et al.,* 2007). The dominant application of these models includes estimating forage quality and quantity. Ramoelo *et al.* (2011) combined later models with a continuum removal spectral transformation technique to estimate grass nitrogen. PLSRs have proven to overcome the effect of multicollinearity and model overfitting encountered by the SMLR.

2.9 Nonparametric regression methods for grass quality estimation

Non- parametric regression methods are more robust than parametric methods. This approach does not rely on the prior selection of specific bands, and it applies the learning phase of training datasets to select highly performing bands taking advantage of the full spectrum (Verrelst *et al.,* 2015). Adjusted coefficients were used to minimise estimation errors. The flexibility of these methods allows them to handle complex datasets with non-linear structures. However, over-flexible models are subjected to overfitting training data (Verrelst *et al.,* 2015). Machine learning regression algorithms (MLRAs) were categorised as non-linear nonparametric regression models. These later methods outperform the linear models by applying non-linear transformations assuming that the relationship between datasets is unknown. Another superiority of non-linear models is their ability to detect non-linear relationships within datasets without knowledge about the data distribution (Verrelst *et al*., 2015). MLRAs include but are not limited to Artificial Neural Networks (ANN) (Abdipour *et al.,* 2019), Support Vector Machines (SVM) (Avitabile *et al.,* 2012), and Random Forest (RF) (Dube & Mutanga, 2015) regression models have long been utilised in remote sensing to map vegetation biophysical parameters.

Non-parametric methods have proven to be more accurate for vegetation assessment than parametric regression methods. For instance, the RF regression method outperformed all other non-parametric and simple regression models together with physical model retrieval methods with the R^2 value of 0.79 and an RMSE of 0.33 g/m^2 for nitrogen retrieval in a comparative model assessment study by Zheng *et al.* (2018). RF method utilises the full spectrum without overfitting or underfitting and can process large datasets faster (Zheng *et al*., 2018). The physical model, particularly the lookup table-based radiative transfer models (RTM), showed low estimation accuracy with 0.46 g/m^2 and 0.62 RMSE and R^2 values, respectively, and low processing speed. RF regression has been used in optical-based vegetation remote sensing and demonstrated a compelling potential for grass quality estimation due to the ability of this technique to handle the randomness and complexity of data. The use of random forest classification and regression models are extensively reported in the literature. RF applications include aboveground biomass (AGB), leaf water content (LWC), leaf area index (LAI) (Srinet *et al.,* 2019), and in the classification of wetlands species. Recent studies demonstrated the potential of random forest regression models in mapping grass quality and quantity. For

example, the estimation of leaf nitrogen content (LNC) (Liang *et al.,* 2018), N:P ratio (Gao *et al.,* 2020), grassland nutrients content (Singh *et al.,* 2017), grass nutrients and biomass (Ramoelo *et al.,* 2014) and light extinction coefficient (Srinet *et al.,* 2019). A time-series study using the RF regression and VIs by Ramoelo *et al*. (2014) explained over 89% of leaf N concentration and over 84% of above-ground biomass variability in both dry and wet seasons.

De Peppo *et al.* (2021) compared estimation accuracies of parametric and non-parametric regression methods for retrieving leaf area index (LAI) of three different crops (maize, wheat, and alfalfa) from sentinel-2 MSI images. Overall results showed high estimation accuracy on MLRAs, with Gaussian Process regression showing the highest r-square value of 0.80 for all the crops. An integrated modeling approach combining red-edge position, machine learning (artificial neural networks), and continuum-removed absorption features is developed by (Mutanga & Skidmore, 2004) to map grass nitrogen in savanna rangeland from hyperspectral imaging spectroscopy data. Acceptable accuracy with RMSE of 0.13 g/m^2 was better than results obtained from multiple linear regression (RMSE = 0.16 g/m^2). Zhou *et al.* (2019) compared the performance of parametric and non-parametric mathematical approaches to estimate forage crop quality based on nitrogen uptake (N_{up}), crude protein (CP), and dry matter yield in legume and grass mixtures during the development stage and harvest period. Support vector machine (SVM) and PLSR prediction models are also used to estimate N_{up}, CP, and dry matter yield using canopy spectra. Results indicated that SVM is more accurate compared to PLSR, with a mean absolute error (MAE) of 2.8% for PLSR and 1.8% for the SVM model for both the development and harvest stages. A combination of appropriate VIs, and machine learning algorithms could improve the estimation accuracies of Leaf N, P, and Potassium (K) content Peng *et al.,* (2022) derived VIs from UAV multispectral imagery to predict N, P, and K contents in grapes leaves, and the study used parametric and non-parametric algorithms, particularly the extreme learning machine (ELM), PLS, RF, and SVM prediction models. Based on R^2 and RRMSE% values, all the MLA outperformed the PLS, where the supreme model obtained an R^2 value over 0.65 and an RRMSE% less than 20%.

CHAPTER 3:

Methods and methodology

3.1 Introduction

This chapter describes the methods used in proceeding with the objectives to achieve the aim of this study. Data collection techniques and analysis processes are explained. Figure 6 presented at the end of chapter 3 is a flow chart diagram outlining a step-by-step summary of methods used in this research.

3.2 Description of the study area

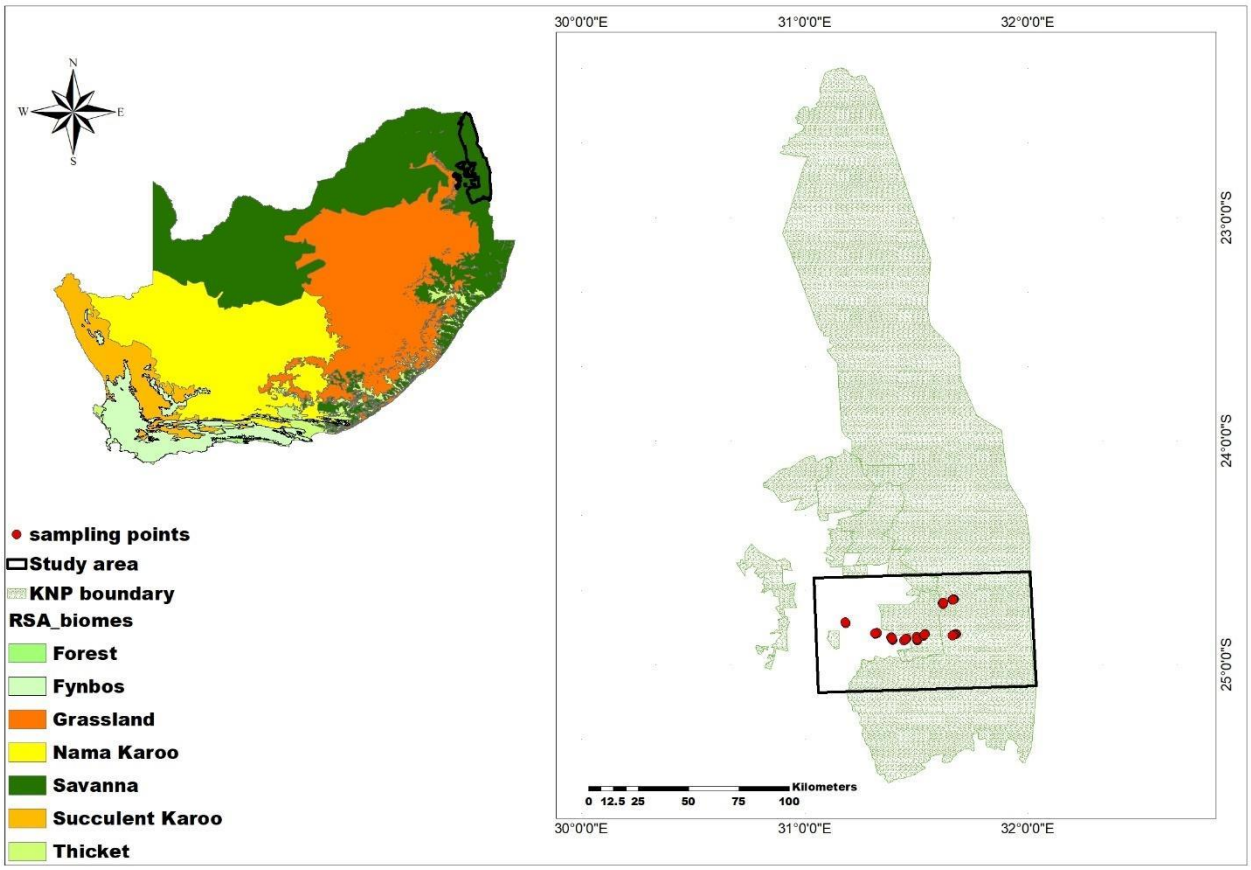

Figure 4: The study area map showing the location of the KNP and the biome in which it is located. The sample points were taken inside the central parts of the KNP and the surrounding communal areas (Source: http://download.geofabrik.de.).

The study area mapped in Fig 4 above is in the Lowveld in the northeastern part of Southern Africa. The lowveld savanna is connected to a low-lying area extending from the lower slope of the Drakensberg Great Escarpment to the west and the Mozambique coastal plain to the east (Venter *et al.,* 2003). The study area comprises land use types ranging from the state-owned KNP, the privately owned Sabi Sands Game Reserve (SSGR), and the communal Bushbuckridge region. Ecosystem functionalities are controlled by several variables, including topography, climate, geomorphology, water resources, and soil types as determined by the parent material (geology). In KNP, dominating geology is basalt on the eastern side of the park and granite on the western side of the park. Local intrusions of gabbro are also present (Venter *et al.,* 2003). The Lebombo mountain is the eastern border between Mozambique and the park, which is made of a thin band of rhyolite. Sand and shale deposits are relatively dispersed north and south of KNP (Siyabonga Africa, 2021). Dominant vegetation types are the granitic Lowveld and gabbro grassy bushveld (Musina *et al.,* 2006). At the crest of gabbro patches, dominant grass species include *Setaria sphacelata*, while *Urochloa mosambicensis* grass species dominate the valleys (Ramoelo *et al.,* 2015).

Varying soil moisture and soil nutrients result from the different geology types. The granitic western side of the park was dominated by less fertile, acidic soil; this is also due to the high rainfall (800 mm/ year) experienced by this side, leaching away all fine nutritious clay particles from the soil (Musina *et al*., 2006). Whereas the gabbro-derived soil on the Eastern side receives a low average annual rainfall of 580 mm/year; hence it is more fertile. The average temperature range is from 47 degrees Celsius in summer to 35 degrees Celsius in winter (Venter *et al*., 2023, Siyabonga Africa, 2021). Consequently, the gabbro patches comprise more productive grass species (for example, *Urochloa mosambicensis*) compared to the species (*Eragrostis rigidior* and *Pagonorthria squarrosa*) dominating the granitic side. Fine-leaved tree species, such as *Acacia* are found on gabbro-derived soil, and broad-leaved tree species, such as the *Combretum* are dominant in granitic soils (Ramoelo *et al.,* 2015). Water resources include the Crocodile River flowing through the south of KNP, the Sabie River rising from the Drakensberg Mountain, and the Letaba and Olifants river flowing through the centre of the park (Siyabonga Africa, 2021).

3.3 Data acquisition

3.3.1 N and P Field measurements

Field data used in this study is from Ramoelo *et al*., (2013). Measurements were taken in 2008 towards the end of the wet season when biomass productivity peaked. During this period, the interaction effect of N, P, and biomass is minimal (Skidmore *et al.,* 2010). Data was collected using the transect sampling method (Fewster *et al.,* 2005). To cover nutrient variability across the study area, eight sites were located along the land-use gradient from granitic to high gabbro-derived soils. The sites covered a total area of 35,000 hectares and were distributed as follows: two sites in KNP, two sites in SGR (L3 granite, L4 gabbro), and the other four (L5-6 gabbro, L7-8 granite) is in communal areas. A geology map with a 1:250000 scale was used to demarcate the sites and was further refined using 2008 SPOT-5 images (Wessel *et al.,* 2011). Measurements were taken from 49 plots sized 30 m x 30 m. Variations within each plot were measured by randomly selecting three to four smaller plots (0.5 m x 0.5 m) from each plot. Measurements of the dominant grass species and grass samples were collected from each sub-plot utilising the Leica® 's GS20 differential geographic positioning system (DGPS). Grass samples were dried at 80 °C for 24 hours, and measurements were averaged at plot level. Post-processing of DGPS points was implemented with the Leica's GeoPro software and GPS reference from Nelspruit to produce positional accuracy of less than 1 m. see (Ramoelo *et al*. 2013) for a detailed field data collection of the datasets used in the study.

3.3.2 Chemical analysis

According to Ramoelo *et al.* (2013), the chemical analysis of dried grass samples was done at the South African Agricultural Research Council-Institute for Tropical and Subtropical Crops (ARC-ITSC) lab in Nelspruit. The acid digestion methods used for chemical analysis in this study were successfully used by Mutanga *et al.* (2004a, b), Ramoelo *et al.* (2011), and Ramoelo *et al.* (2013). Phosphorus concentrations were retrieved using perchloric and nitric acid, and nitrogen was retrieved using sulphuric acid. The auto analyzer measured Foliar N and P using the colorimetric method (Technicon Indus-trial Method 329-74W; Technicon Industrial Systems, Farrytown, New York). The

chemical reaction of sodium nitroprusside, sodium salicylate, ammonia, and sodium hypochlorite for foliar N measurement was done to produce an ammonia-salicylate complex. Chemical measurement of phosphorus was achieved by the reaction of ortho-phosphate and the molybdate ion to produce phosphomolybdenum complex.

3.3.3 Spectral measurements

a) Analytical Spectral Device (ASD) spectroradiometer

An ASD FieldSpec spectroradiometer is a portable device with a wavelength range of 350 nm - 2500 nm (ASD Inc., 2010). The device remotely acquires data at the visible and near infrared (VNIR) and short-wave infrared (SWIR) region of the electromagnetic spectrum with a rapid acquisition period of 0.1 seconds per spectrum. Spectroradiometer is a unique type of spectrometer that can measure radiance, irradiance, transmitted, and reflected light. The fieldSpec spectroradiometer device offers high accuracy and precision because of its high signal-to-noise ratio and superlative repeatability of results for the detailed analysis of objects (ASD Inc., 2010). For this study, reflectance data collected by Ramoelo *et al* (2013) was used. An ASD spectroradiometer, Fieldspec 3®, consisting of a fibre optic with a field of view (FOV) of 25° was used to measure the spectral reflectance of grass from the 3-4 randomly selected subplots of 1 m x 1 m size. An example of FieldSpec 3 is shown in figure 5 below, the device has a spectral range of 350 nm – 2500 nm and a sampling interval of 1.4 nm from 350 nm - 1000 nm, 2 nm from 1000 nm – 2500 nm. The spectral resolution (FWHM) of FieldSpec 3 ranges from 3 nm at 700 nm, 10 nm at 1400 nm and 2100 nm (ASD Inc., 2010) The device was held at a nadir position, 1 m above the ground, creating a FOV covering the entire subplot. To rationalise the effect of illumination, structural differences in leaf canopy and bidirectional effect, the reflectance spectra for each subplot measurement were taken five times and then averaged, and the mean values were used as the representation of reflectance at each subplot (Mutanga *et al.,*2003; Wang *et al.,*2009). The sensor was calibrated before and after each subplot measurement using a spectral (a white reference panel), and the spectral radiance was then transformed to reflectance. All the spectral measurements were taken on cloud-free days between 10h00 am to 15h00 pm to reduce the atmospheric effect and ensure maximum radiant energy (Abdel-Rahman *et al.,*2010).

Figure 5: Fieldspec3; a device used to collect reflectance data, (Source: ASD Inc 2010).

3.4 Descriptive statistics

Descriptive statistics is a crucial element of quantitative research to organise and simplify the datasets and identify the relationships within the data. These statistics are also important to organise large datasets to few variables that provide meaningful information about the entire dataset. The following descriptive statistics were formulated to (1): measure the central tendency of the data, and summarise the mean, minimum, and maximum values of N:P concentrations, (2) compute variance and standard deviation to measure the dispersion of the data, and 3) calculate the Shapiro-wilk normality test to determine whether the data is normally distributed or not.

3.5 Data processing and analysis

3.5.1 Reflectance spectra pre-processing and spectral simulation.

The noisy bands outside the range of 400 nm to 2500 nm were removed prior to data analysis. Further noise removal was carried out to remove the noisy wavelength between 400 nm and 2350 nm, and we were left with a total of 1697 Wavelengths. The noise-free spectra were then projected to a Savitzky-Golay filter for further smoothing (Golay, 1964). The Savitzky-Golay filter fits a set of points to a polynomial in the least squares. This filter applies a digital filter with a length of 2m+1 to the input data computing convolution coefficient for all points in the dataset, A second-order polynomial (p = 3) and a window size of 13 was used in this research. The general mathematical algorithm applied by the filter is as follows:

$$\boldsymbol{H_{P^*} = \frac{1}{N}\sum_{L=-t}^{t} V_L H_P + L} \quad \textbf{(1)}$$

Where, H_{p*} = new value, N = normalising coefficient, t = number of neighbour values from each side of p and V_L = precomputed coefficients, as determined by the polynomial order and degree used. The fundamental interest of this study was to assess the spectral capabilities of Sentinel-2 datasets for estimating the foliar N:P ratio. The field spectra consisting of 1690 wavelength points were simulated to Sentinel-2 spectral configuration (see Table 2). The same field spectra data was then simulated to RapidEye, Worldview-2, EnMap and Hyperion configurations.

Spectral simulation refers to the process of generating synthetic spectral data that resembles the reflectance characteristics of real-world objects or scenes. This simulation is typically performed to study and understand the interaction of electromagnetic radiation with different materials or to model the behaviour of remote sensing instruments (Somers et al., 2013; Schott *et al.,* 1999). The simulation process involves defining the properties of the materials or objects of interest, such as their spectral signatures or reflectance properties across different wavelengths. This process is used to develop new wavelength bands from available observed bands, it considers factors such as the composition, geometry, and surface characteristics of the objects (Peddle *et al.,* 2001). Various techniques can be used for spectral simulation, including physics-based models, empirical or statistical models, and machine learning algorithms. These methods aim to accurately reproduce the spectral response of materials based on their physical properties (Adjorlolo *et al.,* 2013).

Spectral simulation has several applications, including the definition of future Earth Observation systems, the development and evaluation of instrument specifications, developing and testing new algorithms for image processing and classification, and the generation of synthetic datasets for training and validation purposes (Guanter *et al.,* 2009). The simulation of Hyperspectral ASD data to various Hyperspectral and multispectral datasets considered in this study was implemented using the “hsdar” package in R programming software. The package focuses on hyperspectral remote sensing data, allowing users to simulate spectra

and perform various spectral analysis tasks. The target multispectral and hyperspectral sensors considered in this study were Sentinel-2, RapidEye, WorldView-2, EnMap, and Hyperion. The spectral response function of each sensor available within R programming was used to resample the reflectance spectra to different sensor configurations. The hsdar package allows further analyses and processing of the simulated spectra, such as applying spectral indices, performing classification, etc.

RapidEye and WorldView-2 sensor specifications are presented in Table 3 and Table 4 respectively. For the Hyperspectral sensors, EnMap and Hyperion sensor specifications are not tabulated because of the very large number of bands these sensors have (up to hundreds of narrow bands). EnMap consists of 244 bands ranging from 420 nm to 2440 nm wavelengths with a spatial resolution of 30 m. The FWHM of EnMap is 6.5 nm from 420 nm to 999 nm and 10 nm from 900 nm to 2240 nm. Hyperion has a total of 242 bands ranging from 350 nm to 2582 nm and an FWHM ranging between 10 nm-12 nm. The 'hsdar' package in R statistical software was used for spectral resampling based on the Gaussian spectral response function using the full-width half maximum (FWHM) technique (band centre and bandwidth).

Table 2: Sentinel-2 MSI sensor specifications.

Band name	**Spectral regions (nm)**	**Central wavelength (nm)**	**FWHM (nm)**	**Spat.Res (m)**
B1	433 - 453 (Coastal)	443	20	60
B2	458 - 522 (Blue)	490	65	10
B3	543 - 577 (Green)	560	35	10
B4	650 - 680 (Red)	665	30	10
B5	698 - 712 (Red edge)	705	15	20
B6	733 - 747 (Red edge)	740	15	20

B7	773 - 793 (Red edge	783	20	20
B8	785 - 899 (NIR)	842	115	10
B8a	855 - 875 (NIR narrow)	865	20	20
B9	985 - 955 (Water vapour)	945	20	60
B10	1360 - 1390 (SWIR-cirrus)	1375	30	60
B11	1565 - 1655 (SWIR)	1610	90	20
B12	2100 - 2280 (SWIR)	2190	180	20

Table 3: RapidEye sensor specifications.

band name	spectral range (nm)	central wavelength (nm)	FWHM (nm)	spat. Res (m)
B1	440 -510 (Blue)	475	70	5
B2	520 -590 (green)	555	70	5
B3	630 -685 (red)	658	55	5
B4	690 -730 (Red edge)	710	40	5
B5	760 - 850 (NIR)	805	90	5

Table 4: WorldView-2 Sensor specifications.

band name	spectral range (nm)	central wavelength (nm)	FWHM (nm)	spat. Res (m)
B1	450-800 (Pan)	625	350	0.46
B2	400 - 450 (coastal)	425	50	0.52

B3	450 - 510 (blue)	480	60	0.52
B4	520 -600 (green)	560	80	0.52
B5	585 - 625 (yellow)	605	40	0.52
B6	620 - 700 (red)	660	80	0.52
B7	705 - 745 (Red edge)	725	40	0.52
B8	770 - 850 (NIR1)	810	80	0.52
B9	860 - 104 (NIR2)	950	180	0.52

Model development

To develop an N:P ratio predicting model, five modelling scenarios were developed from the simulated Sentinel-2 spectra to assess the predictive potential of Sentinel-2 bands and VIs. The VIs were divided into two categories, the traditional vegetation indices (TVIs) and the red-edge indices (REIs). The normalised difference (NDVI) and simple ratio (SR) vegetation indices were computed using the red-edge and the traditional bands. Both red-edge and traditional vegetation indices that are proven to be robust for N and P estimation were used to develop foliar N:P predicting models (Abdel-Rahman *et al.,* 2010 and Ramoelo *et al.,* 2018). In total, eighteen known and optimised indices from the red-edge (RE), red (R), near-infrared (NIR), and narrow infrared (NIR2) were used together with Sentinel-2 bands. The construction of the optimized VIs was motivated by the lack of known indices optimal for the prediction of N:P ratios, such possible band combinations from the known N and P absorption features was tested. The used known indices are either ideal for or P estimations. A list of Sentinel-2 indices is presented in Table 5 below.

Table 5: List of VIs used in this study.

Traditional Vegetation Indices		
Vegetation index	**formular**	**Reference**

NDVI	(NIR - R) / (NIR + R)	Rouse *et al.,* (1974)
NDVI1	(NIR2 - R) / (NIR + R)	
SR	NIR/R	Jordan, (1969)
SR1	NIR2/R	
Red-edge Vegetation Indices		
RRI	NIR2/RE1	Barnes *et al.,* (2000)
RRI1	RE3/RE1	Optimised
RRI2	RE2/RE1	Optimised
RRI3	RE1/R	Optimised
RRI4	NIR/RE1	Optimised
RRI5	NIR/RE2	Optimised
RRI6	NIR2/ RE2	Optimised
NDRE	(NIR2 – RE1)/ (NIR2 + RE1)	Barnes *et al.,* (2000)
NDRE2	(RE3 – RE1) / (RE3 + RE1)	Barnes *et al.,* (2000)
NDRE1	(RE2 - RE1) / (RE2 + RE1)	Sims & Gamons, (2002)
NDRE3	(RE1 - R) / (RE1 + R)	OptimiSed
NDRE4	(NIR -RE1) / (NIR +RE1)	Optimised
NDRE5	(NIR - RE2) / (NIR+ RE2)	Optimised
NDRE6	(NIR2 - RE2) / (NIR2+RE2)	Optimised

The following modelling combinations were developed for the study:

- ❖ Traditional vegetation indices, Red-edge indices, and bands (TVIs + REIs + band)
- ❖ Traditional vegetation indices and bands (TVIs + bands)
- ❖ Red-edge indices and bands (REIs + bands)
- ❖ Bands only
- ❖ Traditional vegetation indices and red-edge indices (TVIs +REIs)

3.6 Linear Partial Least Squares Regression and Leave one out cross validation.

The linear multivariate PLSR effectively analyses high dimensional, collinear, and noisy datasets (Wold *et al*., 2001). Similarly, to the components analysis (PCA) method, A linear PLSR extracts non-correlated features to explain the relationship between two variables. In the first set of data an N-dimensional vector of variables is denoted by x ∈ X ⊂ R_N, and a RM vector of variables of the second dataset is denoted by y ∈ Y ⊂. Scores vectors are used to model the relationship between these two datasets. The number of samples for both datasets is indicated by *n,* so mean-centred data matrices of X and Y are (*n**N and *n**M) sampled at X-space and Y-space, respectively. The PLS decomposes the X and Y matrices into the form:

$$X = TP^T + E \quad (1)$$

$$Y = UQ^T + E \quad (2)$$

In equations (1) and (2) *X* and *Y* are projected into low-dimensional matrices indicated by *U* and *T* for *X* and *Y* respectively. *T* and *U* represent the (n*p) matrices of the p extricated latent vectors. Matrices of loading are denoted by matrix (N*p) P and (M*p) Q, matrix (n*N) *E*, and matrix (n*M) F are the matrices of residuals. The PLS algorithm is originally founded on the iterative PSL (NIPALS) approach (Wold, 1975) acquires component vectors w, c such that,

$$[cov(t,u)]^2 = cov(Xw, Yc)]^2 = max_{|r|=|s|=1}[cov(Xr, Ys]^2 \quad (3)$$

where *cov (t, u*) = t^T *u/n* is the sample covariance between the score vectors *t* and *u*. The first step of the NIPALS algorithm is randomly initialising the *Y*-space score vector u and repeating a sequence of the following steps until convergence.

a. $W = X^T U/(U)^T U$

b. $||w|| \rightarrow 1$

c. $t = Xw$

d. $c = Y^T t/(t^T t)$

e. $||c|| \rightarrow 1$

f. $u = Yc$ (4)

where $||.|| \rightarrow 1$ is the transformation of a vector to a unit norm. After extracting the score vectors t and u, X and Y are regressed on t and u respectively and loading matrices p and q are reckoned through the regression process.

$p = X^T t/(t^t t)$ and $q = Y^T u/(u^T u)$ (5)

After extracting the score vectors t and u, rank-one matrix approximations of X and Y matrices are subtracted to deprecate X and Y. There are diverse ways of deprecating, see Rosipal and Kramer (2006) for review. The mostly used PSL technique is based on the following assumptions:

a) The score vectors $\{t_i\}^p_i = 1$ are good predictors of Y and

b) there is a linear relationship between t and u such that,

$U = TD + H$ (6)

D represents the (p × p) diagonal matrix and the matrix of residuals is indicated by H. The asymmetric (predictor–predicted) variables relationship is transformed into a deflation scheme where the input space score vectors $\{t_i\}^p_i = 1$ are good predictors of Y. at each iteration of the PLS, Y is deflated using the score vectors removing a regression component of Y on t

The score vectors are then used to deflate Y; that is, a component of the regression of Y on t is removed from Y at each iteration of PLS:

$X \leftarrow X - tt^T X/(t^T t)$ and $Y \leftarrow Y - tt^T Y/(t^T t) = Y - tc^T$

Based on the assumption of a linear relation between the scores' vectors t and u in equation (6), the decomposition of the Y matrix in equation (1) is written as:

$$Y = TDQ^T + (HQ^T + F$$

and the linear PLS regression model is:

$$Y = TC^T + F^* \quad (8)$$

$C^T = DQ^T$ represents the (p × M) matrix of regression coefficients and $F^* = HQ^T + F$ denotes

the residual matrix.

3.6.1 PLSR variable importance

Contribution of each band and index in the development N:P estimation model was determined by calculating variable importance for prediction (VIP), using PLSR weights in R statistical software. The PLSR weights uses the following formula to determine the contribution of each predicting variable to the entire model:

$$VIPK(a) = K \sum a\, w2ak\ (SSYt) \quad (9)$$

Where VIP*k*(*a*) is the importance of the k^{th} predictor variable based on a model with factors.

Wak correlates to the loading weight of the k^{th} variable in the a^{th} PLSR factor, SSY*a* is the explained sum of squares of *y* by a PLSR model with factors, SSY*t* is the total sum of squares of *y*, and *K* is the total number of predictor variables (Viscara & Rossel 2008).

3.6.2 Leave-One-Out-Cross-Validation (LOOCV)

Due to the fewer collected samples, the data were not divided to train and test data. Instead, the leave-one-out-cross-validation (LOOCV) was used; this is a special kind of K-folds cross-validation where K=N. This type of cross-validation repeatedly fits a model to a data set that contains an N-1 number of samples. As such, the measure of test MSE is less biassed. The LOOCV is suitable for small sample sizes as it maximises the number of training samples. The PLSR was implanted in the R statistical environment using ‘pls’ and ‘caret ‘packages’ (Bjorn-Helge *et al*., 2015, Kuhn *et al*., 2015).

3.7 Random forest regression

The random forest regression (RFR) method was initially developed to improve the classification and regression trees (CART) by (Breiman *et al.,* 1984). The RFR method is associated with bragging and bootstrapping (Breiman, 1996), which are essential for

minimising predictive variance (Hadi, 2015.) This technique generates multiple trees using a deterministic method to randomly select important variables from the calibration data while assuming no normal of the dataset. It stratifies the explanatory data into numerous decision trees. The single decision tree is then projected to bagging based on unanimity modeling, which averages prediction from the “forest” (many trees developed by bootstrapping the entire datasets). Random forest regression is practical when dealing with data that contains an enormous number of predictor variables for a few samples. According to Cutler *et al.* (2009), the great advantage of RFR is its ability to fit non-linear relationships and handle outliers in the explanatory variables. To improve the bagging, all the explanatory variables that make similar trees are removed by randomly sampling the regressors in each split to differentiate the trees (Hadi, 2015). A summarised description of the RF regression proposed by Breiman, (2001) is as follows:

I. Randomly draw ntree bootstrap samples *Xi* (*i* is the bootstrap iterations and *X* is the calibration datasets) from the calibration datasets. Each bootstrap sample contains one-third of the original datasets. Variables that are not incorporated at each bootstrap sample are called out-of-bag (OOB).

II. Grows an unpruned regression tree for each bootstrap sample such that it randomly selects one-third of the predicting variables and chooses the best split among those variables at each node.

III. The OOB response value is predicted and averaged at each bootstrap iteration for ntree.

IV. A percent increase in the root means square error is then calculated to measure the importance of each predictor. The relationship between the predictors and the dependent variables is ranked based on the value of variable importance. Predictors with higher variable importance values are regarded as the critical variables for the model.

In a random forest algorithm, two parameters ought to be adjusted, the number of trees (n) and the number of regressor variables randomly selected from each split (mtry). For this research, a “10 folds 2 repeats” cross-validation was used to tune these parameters. The data contained only 49 sample points, so the data could not be split to train and test data cross-validation was used instead. The number of folds and repeats was chosen

based on commonly used values in previous studies with a small number of samples (Han & Kim *et al.,* 2021). The RFR was executed in R statistical software using the 'random forest' and 'caret' packages (Kuhn *et al*., 2015).

3.7.1 Selection of significant variables

The importance of each variable is weighed based on the out-of-bag (OOB) error or permutation. For each bootstrapped split (70% of the training dataset), the remaining 30% that is not used in the training of the model is projected to the tree for prediction. The value of each variable is then randomly permuted while the explanatory variable remains constant. The modified OOB is passed through the tree to compute another set of predictions. Finally, the importance of the variables is calculated by computing the difference in OOB error between the permuted and non-permuted datasets throughout the random trees (Parasad *et al*., 2006). In this study, important RFR variables were selected using the Vsurf function in R statistical software. Vsurf algorithm creates nested and predictive models and evaluates each explanatory variable's importance by ranking them using an RF permutation-based score of importance. The nested models include variables with some redundancy. In the final predictive model, the algorithm tries to avoid redundancy and leave only unique best-predicting variables (Genuer *et al*., 2015) and ensure that the final model is significant with P-value less than 0.05.

3.7.2 Calibration and validation of the RFR models

Single random training-test partitioning might result in validation biases. In this research, such bias is dealt with by applying a systematic hierarchical 10-folds cross-validation (Kohavi, 1995) and two repeats to all the modeling scenarios to evaluate their maximum predictive accuracy. The data could not be divided into a train (70%) and test (30%) due to the fewer number of samples (n = 49) which could result in biased the validation.

3.8 Accuracy assessment and model comparison

The performance of each modeling scenario was determined using the following statistical metrics, coefficient of determination (R-squared), root mean square error (RMSE), and the relative root mean square error (RRMSE). The coefficient of determination (R^2) measures the strength at which an independent variable explains a

dependent variable in a model. In simple terms, R^2 measures how well a predicted value fit the measured values in a model, and this coefficient ranges 0 to 1, with values closer to one indicating a perfect fit model: R^2 is calculated as follows:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum\left(y_i - \underline{y}\right)^2} \quad (1)$$

Where,

The sum of the residual squares (RSS) is indicated by: $\sum(y_i - \hat{y}_i)^2$

y_i = i[th] value of the variable to be predicted

$\hat{y}_i$ = predicted value of *yi*

n = upper limit of summation

And,

the total sum of squares (TSS) is given by: $\sum\left(y_i - \underline{y}\right)^2$

y_i **=** value in a sample

$\underline{y}$ = mean value of a sample

RMSE is the measure of the standard deviation of the residuals from the mean, it measures the error rate by the square root of the MSE, calculated as follows:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(Y - \hat{Y}_i\right)^2 \quad (2)$$

Then,

$$\text{RMSE} = \frac{\sqrt{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}}{n} \quad (3)$$

Where Y*i* are the observed values at *n* data points

estimated values are indicated by $\hat{Y}_i$

Lower RMSE and MAE values indicate high accuracy in a regression model.

Relative Root Mean Square Error (RRMSE) is the root mean squared error normalised by the root mean square value where each residual is scaled against the actual value, it expresses the error in percentage form where values less than 10% indicate high accuracy

and values greater than 30% indicating the poor performance of a model (Khan & Noor, 2019). RRMSE is expressed as:

$$\text{RRMSE} = \sqrt{\frac{\frac{1}{n}\sum_{l=1}^{n}(y_l - \hat{y}_i)^2}{\sum_{l=1}(\hat{y}_i)^2}} \quad (4)$$

3.9 Mapping the distribution of foliar N:P ratio

The present study further investigated the possibilities of mapping the spatial distribution and variations of N:P ratio from simulated ASD datasets. That was achieved by inverting the most accurate model to Sentinel-2 MSI. Sentinel-2 Level-1C and Level-2C that perfectly match the season of data collection (end of wet season), model inversion was implemented for 4 years (2018 – 2021) to observe nutrient limitation spatiotemporal variations over the study area and to assess the possibility of continuously transferring a once off trained model over time. The varying pre-processing levels were due to the unavailability of Level-2C data for some years. The data acquisition and the necessary pre-processed was executed in Google Earth Engine. The best-predicting indices, NDRE2 and RRI were selected in the RFR variables importance, the indices were calculated in Google Earth Engine and then transferred to the R statistical environment for model inversion using the ‘random Forest’ and ‘raster’ packages.

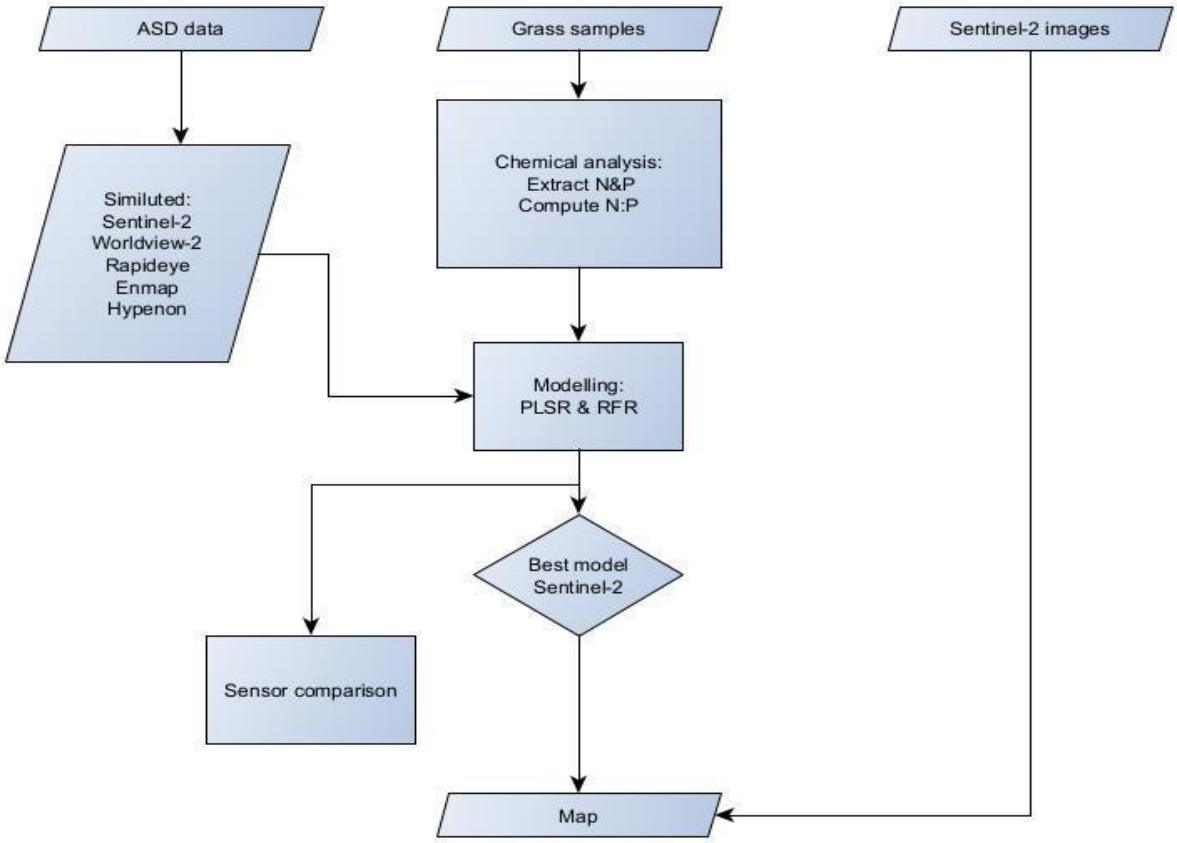

Figure 6: The methodology flow chart diagram summarising the overall methods executed this study.

CHAPTER 4:

Results

4.1 CHAPTER OVERVIEW

The results presented in this chapter are arranged into three sections: Firstly, we evaluated the potential simulated Sentinel-2 bands and indices to estimate nutrient limitation in a savanna environment and determined Sentinel-2 significant bands and indices for estimating foliar N:P ratio using the linear parametric PLSR and non-parametric RFR. Secondly, transferability of ASD models to spaceborne satellite images was investigated. Lastly, the predictive capabilities of the Sentinel-2 datasets were compared to commercial multispectral and hyperspectral sensors to assess the influence of different spatial and spectral configurations for retrieving foliar biochemical contents and further assessed the consistency of bands and indices selected in all sensors to understand N:P sensitive wavelengths. The scatter plots showing the regression between the observed N:P ratios and all the different modeling scenarios are presented in Appendix A of this thesis.

4.2 Descriptive statistics

The foliar N:P ratio is normally distributed across the study area as confirmed by the Shapiro-Wilk normality test (W=0.977, p=0.45) (Royston,1982); see Table 6 for a summarised descriptive statistic of N:P distribution. Based on the calculated low standard deviation (SD) and the variance (var) the datasets are clustered closer to the mean and the N:P ratio values are more identical with no outliers. Diagrams A and B in Fig 7 below are the visual demonstrations of the normality distribution of data. A is the quantile–quantile (*Q*-*Q*) plot, and all points lie closely, forming a diagonal line indicating the normal distribution of the data. Furthermore, the box plot B, the whiskers are closely symmetrical.

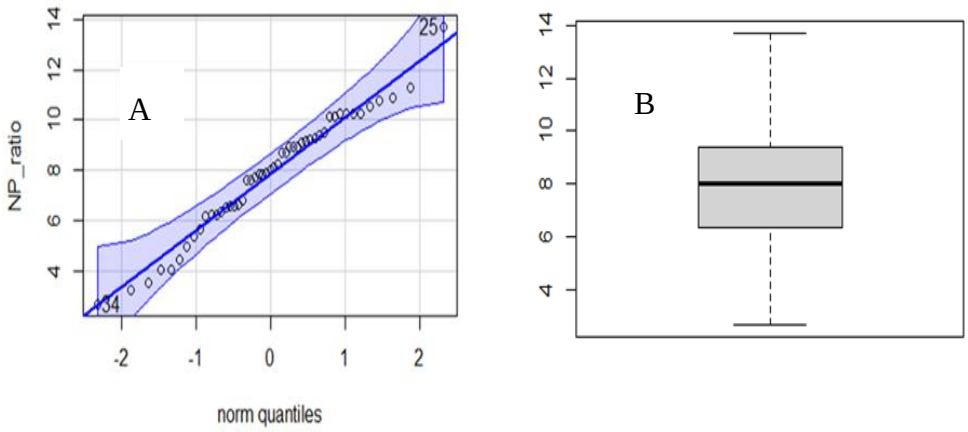

Figure 7: The normality test diagrams.

Table 6: N:P ratio descriptive statistics

BIO	min	median	mean	max	SD	var
N:P	2.683	7.987	7.863	13.699	2.353	5.538

4.3 Spectral Simulation

Figure 8 below outlines the field spectra data pre-processing and spectral simulation from ASD hyperspectral datasets to the configuration of the desired multispectral and hyperspectral datasets. The simulation resampled the hyperspectral data by average narrow wavelength to produce the 13 broad Sentinel-2 bands. The Savitzky-Golay-Filter smoothied the spectral curve while ensuring that the original shape was retained. This resampling procedure was applied to all the sensors used in this study.

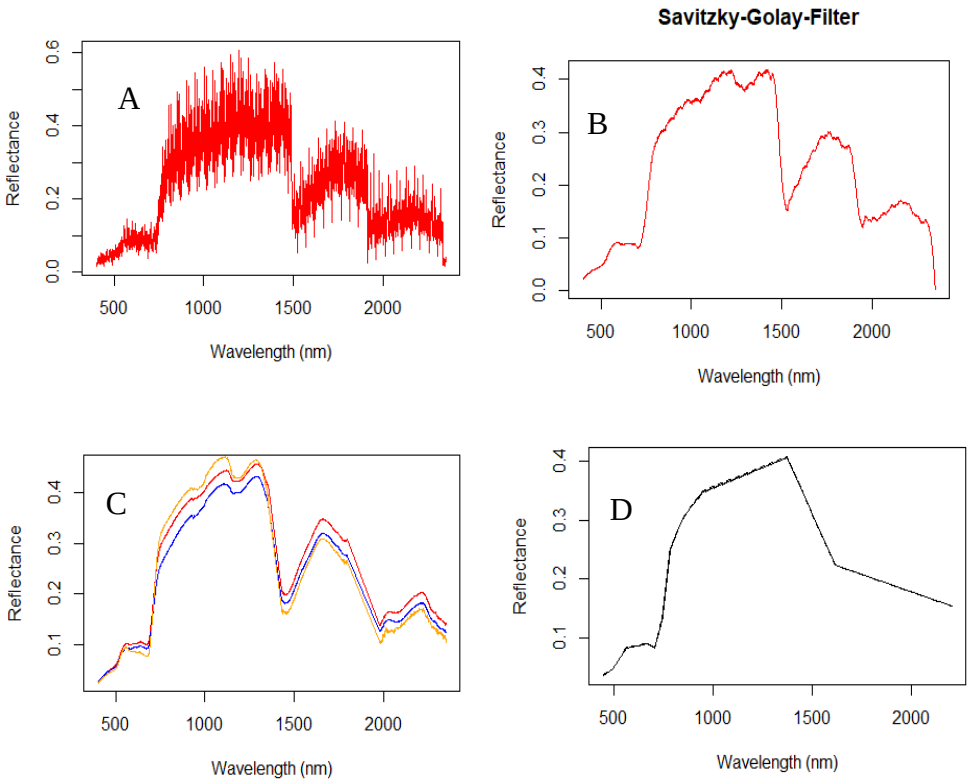

Figure 8: Spectral resampling sequence, where, A is the original spectra, B is filtered spectra, C is the original, min, and maximum spectra, D is the resampled Sentinel-2 spectra.

4.4 Modeling N:P ratio from Sentinel-2 datasets using PLSR.

The results presented in Table 7 is the Sentinel-2 N:P ratio estimation accuracy obtained from both parametric (PLSR) and non-parametric (RFR) algorithms. The accuracy of each modeling scenario is measured based on RMSE, RRMSE, and R^2 values. When using PLSR, the combination of TVIs, REIs, and Sentinel-2 bands explained 66% of N:P ratio variability (R^2 = 0.662), with the lowest RRMSE and RMSE values of 15.35% and 1.21, respectively. Sentinel-2 bands produced the poorest estimation accuracy with the lowest R^2 value of 0.26, the highest RMSE of 1.99, and the highest RRMSE of 25.42%. Higher N:P ratio estimation accuracy was obtained when the RFR was used. There are no significant differences in the R^2 values of all the RFR models. The best predicting RF model was selected based on RRMSE and RMSE. For this study these

accuracy parameters were interpreted as follows: RMSE measures the deviation of the residuals from the mean. A low RMSE value indicates that the predicted and observed data are close to each other indicating a better accuracy. As explained in chapter three of this thesis, RRMSE is the root mean squared error normalised by the root mean square value where each residual is scaled against the actual value. Measured in percentage, a lower RRMSE (≤ 10) signifies a smaller discrepancy relative to the predicted value. Therefore, the lower the RMSE the better is model performance. The RRMSE The combination TVIs and REIs produced high accuracy, as indicated by the lowest RRMSE value of 13.482%, and RMSE of 1.07 while explaining 89% of N:P variability. Sentinel-2 bands derived model produced high error terms, the RMSE, and RRMSE of the model are the highest compared to other models, with RMSE and RRMSE of 1.16 and 14.79%, respectively. The accuracy of the models could not be concluded based on the R^2 value only since the difference is not significant (less than 1 in all the models), and the R^2 value is influenced by the addition and removal of explanatory variables.

Vsurf variable selection was used to select the best predictor variables in each model, and the Vsurf-selected variables were then used to re-run random forest regression models. Results in Table 7 indicate that the accuracy of the random forest regression models improved the RRMSE decreased when using only Vsurf selected variables. The error term of the best-predicting model is slightly lower, with an RRMSE of 13.09% (a difference of 0.74% from the original RFR model), explaining 86% of N:P ratio variability at P<= 0.05. the selected optimal variables are the NDRE2 (R745nm vs. R705nm) and the RRI (R865 nm vs. R705 nm) (where R stands for the reflectance at a particular wavelength). In Table 7 below, RMSECV is the root mean square error for cross-validation, and ncomp is the number of components used for each PLSR model. P-value is defined to less than or equal to 0.05 for this study. The VSurf algorithms reduces the number of predicting variables while increasing the efficiency of the model and results in a statistically significant model (p-value less than 0.05) (Jiang et al., 202).

Table 7: Sentinel-2 regression accuracy.

PLSR Accuracy						
Modeling scenario	**RMSECV**	**RMSE**	**RRMSE%**	**R^2**	**ncomp**	**P-value ≤ 0.05**

TVIs +REIs +Bands	2.218	1.706	20.696	0.662	30	No
TVIs + Bands	2.396	1.973	25.088	0.337	16	No
REIs +Bands	2.155	1.714	21.803	0.653	27	No
Bands	2.365	1.998	25.416	0.264	13	No
REIs +TVIs	2.424	2.244	28.54	0.465	17	No

RFR Accuracy

Modeling scenario	**RMSECV**	**RMSE**	**RRMSE%**	**R^2**	**p-value ≤ 0.05**
TVIs + REIs + bands	2.414	1.138	14.467	0.927	No
TVIs + bands	2.486	1.13	14.371	0.924	No
REIs + bands	2.433	1.117	14.209	0.938	No
Bands	2.489	1.181	15.017	0.915	No
REIs + TVIs	2.344	1.107	14.073	0.901	No

RFR VSURF Accuracy

Modeling scenario	**RMSECV**	**RMSE**	**RRMSE%**	**R^2**	**Selected**	**p-value ≤ 0.05**
TVIs + REIs + bands	2.21	1.1	13.85	0.86	B2, B5, NDRE, NDRE2, RRI5	Yes
TVIs + bands	2.32	1.16	14.72	0.86	B2, B4, SR	Yes
REIs + Bands	2.32	1.12	14.19	0.87	NDRE2, NDRE, RI5, B2, B4, NDRE4	Yes
bands	2.38	1.09	13.99	0.88	B2, B4, B5	Yes
REIs + TVIs	2.22	1.09	13.10	0.86	NDRE2, RRI	Yes

4.4.1 Comparing the N:P prediction accuracy of PLSR and RFR

The Figure 9 below is a comparison of PLSR and RFR accuracies for all the modelling scenarios. PLSR has higher RRMSE and RMSE in the models with lower R^2 relative to the RFR. The combination of REIs and bands attained higher predictive accuracy for both PLSR and RFR compared to other modelling scenarios.

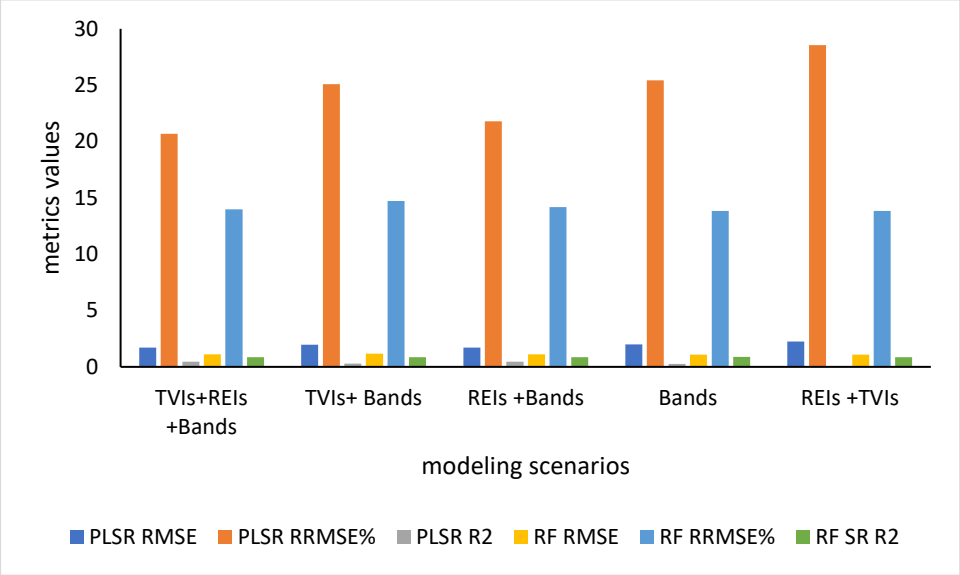

Figure 9: A bar graph showing PLSR and RFR N:P ratio estimation accuracy for all Sentinel-2 models for each modeling scenario.

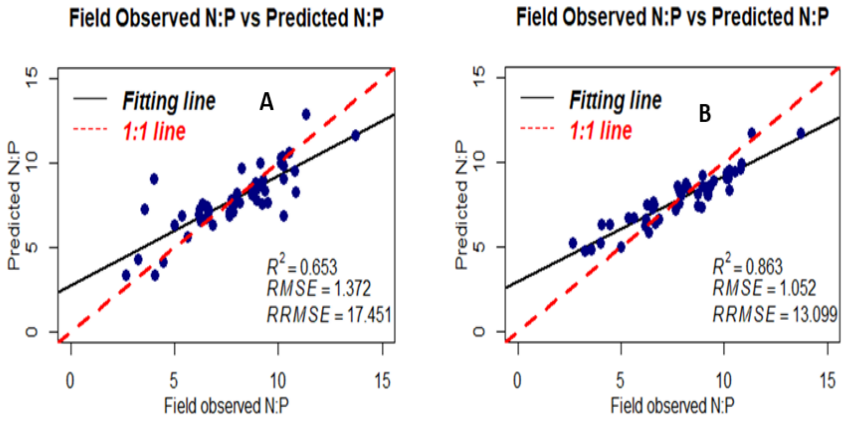

Figure 10: 1:1 regression plots for best PLSR (A) and RFR (B) models. The RFR final model improves N:P ratio estimation accuracy, with a noticeable decrease of 4.35% in RRMSE and a 0.21 increase in the R^2 value.

4.5 Mapping the spatial distribution Foliar N:P ratio

RFR algorithm predicts foliar N:P with higher accuracies than PLSR. There were no significant differences between the accuracies of all the RFR models. Therefore, any model could be used to map the distribution of nutrient limitation. Since there is a slight improvement in accuracy with the NDRE and RRI models, we used those two indices for model inversion. The distribution of foliar N:P value and their interpretation in terms of nutrient limitation status are presented in Figure 10 and Figure 11, respectively. Figure 10 shows the distribution of N:P ratio values across the central parts of KNP and surroundings from 2018 to 2021. From the maps, the N:P ratio values range from 4 to 12. In 2018 the dominating N:P values from 4 to 10, with a very small portion of the area on the south-western and north-eastern corners of the map showing high N:P ratio values above 10. There is a change in the range of N:P ratio values in the 2019, with a larger area covered by N:P values above 10 in the north-eastern and eastern sides of the map. In 2020, there is a drastic increase in N:P ratio values across the study area, the area is covered mostly by N:P values above 10. The N:P ratio values above 10 disappear completely in 2021, the entire area is mainly dominated by N:P ratio values ranging from 4 to 10 with very small patches of values ranging from 10 to12. Most of the areas with N:P ratio values above 11 and below 6 are gullies and bare soil with no vegetation cover, including roads, water bodies, and built-up areas. The vegetation density of the mapped area from 2018 to 2021 is shown in the false colour composite maps in Figure 12.

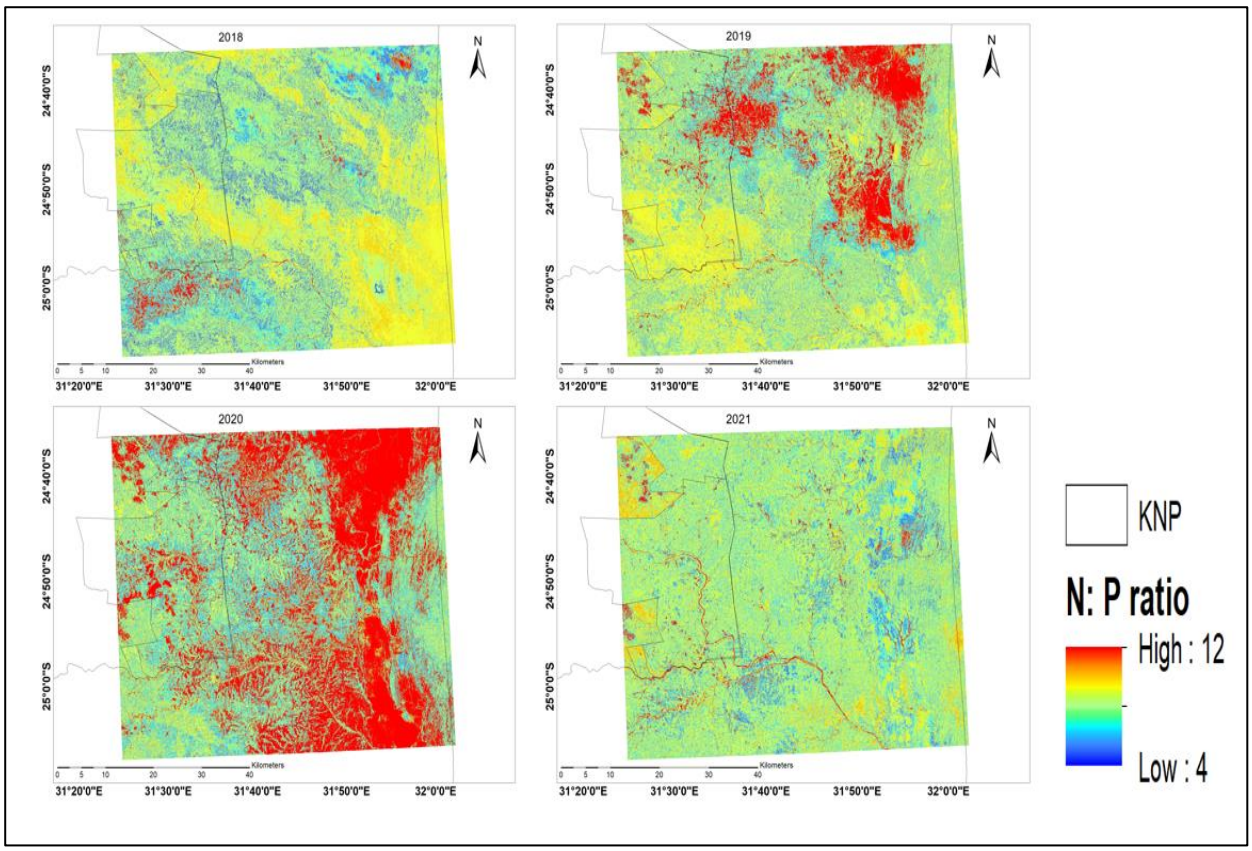

Figure 11: N:P ratio values from 2018-2021 at central parts KNP and surroundings mapped by inverting the best predicting RFR model to Sentinel-2 images acquired between April and May.

The maps in figure 11 are an interpretation of the nutrient limitation status of the KNP based on the N:P ratio critical values using Gusewell's thresholding. *Recap*: Gusewell, (2004) suggests that N:P ratio values from 0 to 10 indicates N-limitation, values from 10 to 20 suggest co-limitation, and lastly, P limitation is presented by N:P values from 20 and above. The highest distribution of co-limitation was observed in 2019 and 2020. From all these years, the central part of KNP and surroundings is affected by either N limitation or co-limitation, with N limitation being the most prevalent. There are no areas solely affected by P limitation (no N:P ratio values above 20). Nutrient co-limitation affected larger areas in the north and eastern parts of the park between 2019 and 2020, especially in 2020. However, as stated above, most of the areas affected by co-limitation (with N:P ratio values greater 10) are non-vegetated (either bare soil, roads, gullies, and rivers). Due to the unavailability of a cloud-free image between March and April 2020, a satellite image acquired in May was used instead. Even though it was outside the scope of

our study, the extensive distribution of nutrient co-limitation in 2020 could be associated with leaf dryness, as the co-limitation completely disappears in the 2021. In May, most vegetation is parched with smaller leaves, i.e., low leaf area index (LAI). The distribution of nutrients limitation is not uniform in the study area. In the savanna environment, a 10m area (Sentinel-2 spatial resolution) can contain various vegetation types, including different grass species, shrubs, and other leafy vegetation with different N and P concentrations or requirements. Vegetation type stratification was delimited in this study; however, it might have an impact on the foliar N:P ratio maps.

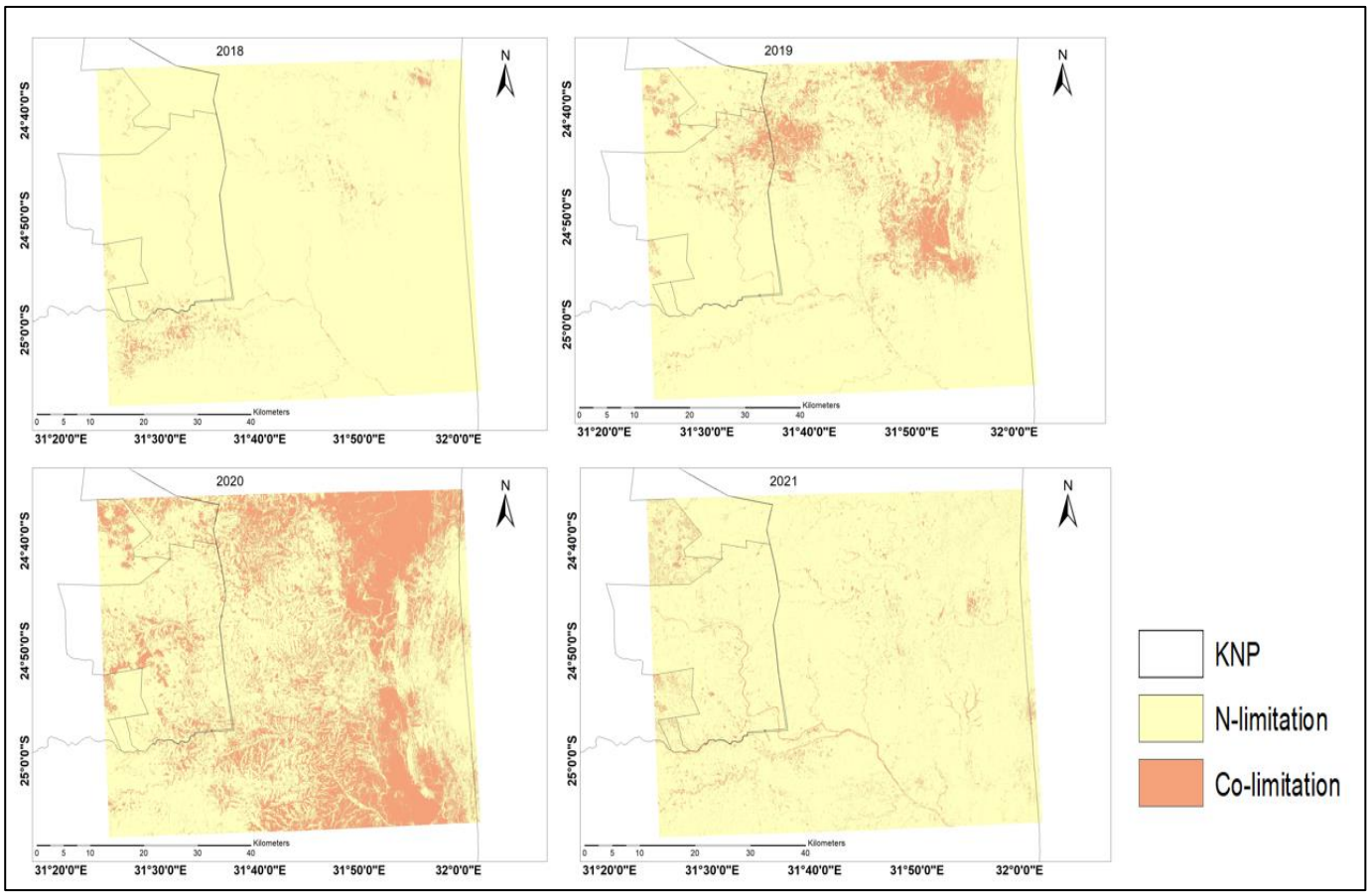

Figure 12: Nutrient limitation status based on the interpretation of the above N:P ratio threshold values.

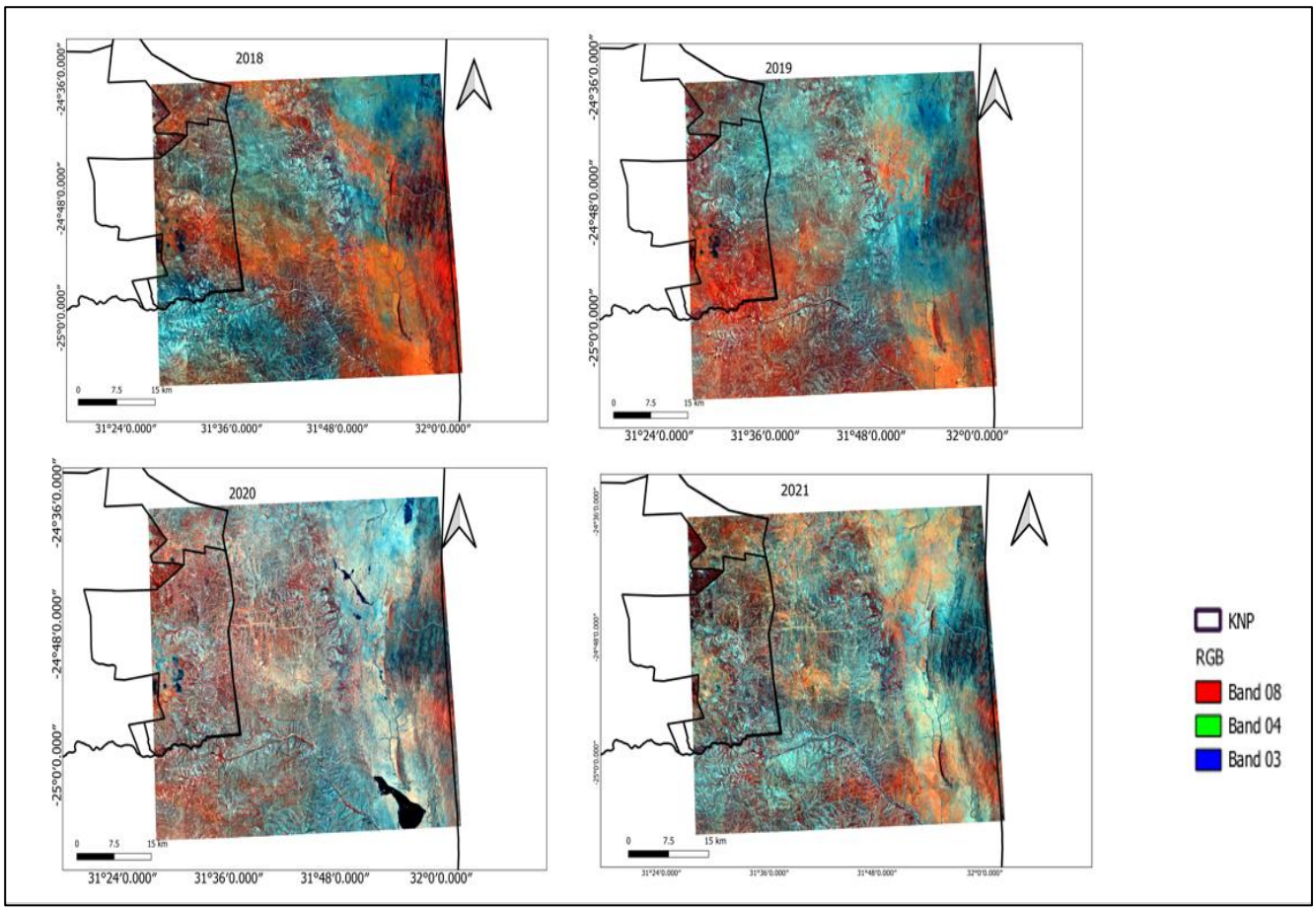

Figure 13: False-color composite of 2018 -2021 Sentinel-2 images RGB (843) showing vegetation density in the mapped area, where the red colour indicate areas with vegetation.

4.6 Comparing Sentinel-2 with Multispectral and Hyperspectral sensors

4.6.1 Estimating N:P ratio from RapidEye

A summary of RapidEye regression accuracy is presented in Table 8 below. When using the PLSR algorithm, RapidEye shows low accuracy in all the scenarios. A combination of TVIs, REIs, and RapidEye bands explained 20% of N:P variation with the lowest RRMSE of 26.397%. Poorest accuracy was observed when estimating the N:P ratio using RapidEye bands only, with the highest RRMSE of 28.03%. The estimation accuracy of RapidEye improved significantly when using the RF regression method as there are no significant differences in all the RapidEye RF models. A slightly low error term is produced by the REIs and bands model (RRMSE = 14.028 %). The accuracy of the RapidEye

decreased (i.e., increased RRMSE in all models) after selecting important variables using the VSURF method.

Table 8: RapidEye regression accuracy

PLSR accuracy						
Modeling scenario	**RMSECV**	**RMSE**	**RRMSE**	**R^2**	**ncomp**	**p-value ≤ 0.05**
TVIs +REIs + bands	2.36	2.08	26.39	0.21	10	No
TVIs + bands	2.33	2.16	27.48	0.14	6	No
REIs + bands	2.43	2.09	26.56	0.19	7	No
Bands	2.35	2.21	28.04	0.10	5	No
REIs + TVIs	2.43	2.09	26.56	0.20	7	No
RFR Accuracy						
Modeling scenario	**RMSECV**	**RMSE**	**RRMSE**	**R^2**	**p-value ≤ 0.05**	
REIs + TVIs + bands	2.24	1.11	14.12	0.89	No	
TVIs + bands	2.33	1.08	13.69	0.89	No	
REIs + bands	2.34	1.11	14.17	0.88	No	
Bands	2.35	1.14	14.46	0.91	No	
REIs + TVIs	2.25	1.25	15.88	0.80	No	
Vsurf Accuracy						
Modeling scenario	**RMSECV**	**RMSE**	**RRMSE**	**R^2**	**selected variables**	**p-value ≤ 0.05**
TVIs + REIs + bands	2.31	1.11	14.08	0.86	B2, B3, RRI	Yes
TVIS + bands	2.28	1.09	13.09	0.88	B2, B3, SR	Yes
REIs +bands	2.28	1.10	14.03	0.87	B2, B3, NDRE	Yes
Bands	2.36	1.36	14.44	0.87	B4,32	Yes

TVIs + REIs + bands	2.28	1.29	16.50	0.73	NDRE	Yes

4.6.2 Estimating N:P ratio from WorldView-2

Table 9 shows N:P ratio estimation accuracy obtained from WorldView-2 datasets. The models explained a maximum of 30% of N:P ratio variation when modeling the N:P ratio from REIs +TVIs + bands, with the lowest RRMSE of 24.65% and RMSE of 1.939. The lowest estimation accuracy was obtained when explaining N:P variabilities with WorldView-2 bands only, with the highest RRMSE of 28.033%, RMSE of 2.204, and the lowest R^2 value of 0.104. The estimation accuracy of WorldView-2 increased significantly when using RFR. The combination of REIs, TVIs, and bands modeling scenario improved to an RRMSE of 13.95% (i.e., an improvement of 10.69%). There are no significant differences in the accuracy of WorldView-2 after variable selection. The difference in RRMSE is less than 0.1%. The selected important variables from the best performing model are the NDRE, B6, B4, and B3.

Table 9: Regression accuracy for Worldview-2

PLSR Accuracy						
Modeling scenario	**RMSECV**	**RMSE**	**RRMSE**	**R^2**	**ncomp**	**P-value ≤ 0.05**
TVIs +REIs + bands	2.33	1.94	24.65	0.31	14	No
TVIS +bands	2.37	2.09	26.52	0.19	10	No
REIS +bands	2.33	2.00	25.44	0.26	12	No
Bands	2.44	2.20	28.03	0.10	7	No
TVIs +REIs	2.17	2.00	25.49	0.26	6	Yes
RFR accuracy						
Modeling scenario	**RMSECV**	**RMSE**	**RRMSE**	**R^2**		**P-value ≤ 0.05**
TVIs + REIs + bands	2.20	1.91	13.96	0.91		No
TVIs + bands	2.24	1.06	13.53	0.90		No

REIs +bands	2.19	1.90	13.89	0.90	No
Bands	2.40	1.13	13.39	0.90	No
TVIs + REIs	2.35	1.22	15.57	0.81	No

Vsurf Accuracy						
Modeling scenario	**RMSECV**	**RMSE**	**RRMSE%**	**R^2**	**Selected variables**	**p-value ≤ 0.05**
TVIs + REIs + bands	2.28	1.09	13.88	0.88	NDRE, B6, B4, B3	Yes
TVIs +bands	2.29	1.12	14.27	0.87	B4+B3+SR	Yes
REIs + Bands	2.26	1.09	13.94	0.87	NDRE, B4, B6	Yes
Bands	2.29	1.44	14.54	0.86	B3, B4, B6	Yes
TVIs +REIs	2.28	1.19	15.23	0.81	NDRE, SR, RRI1	Yes

Estimating foliar N:P ratio from EnMap

The overall PLSR results for EnMap in Table 10 indicate that a combination of TVI, REI and EnMap absorption features (all variables) explained 67% of N:P ratio variability with an RRMSE of 17%. The estimation accuracy of EnMap improved significantly when using the random Forest Regression. This is indicated by the notable decrease in the RRMS in all the scenarios. There are no significant differences in models' accuracy after selecting important variables. The absorption features at 637 nm, 1945 nm, and 2345nm and the SR are significant for N:P ratio retrieval from EnMap datasets.

Table 10: EnMap regression accuracy.

PLSR Accuracy						
modeling scenario	**RMSECV**	**RMSE**	**RRMSE%**	**R^2**	**ncomp**	**P-value ≤ 0.05**
TVIs + REIs + AF	2.34	1.34	17.02	0.67	32	No

TVIs + AF	2.10	1.42	18.71	0.60	24	No
REIs + AF	2.13	1.56	19.77	0.55	22	No
AF	2.08	1.73	22.04	0.45	14	yes
REIs + TVIs	2.16	1.69	21.48	0.47	18	No

EnMap RFR Accuracy

modeling scenario	**RMSECV**	**RMSE**	**RRMSE%**	**R^2**	**P-value ≤ 0.05**
TVIs + REIs +AF	2.18	0.99	12.69	0.92	No
AF	2.23	1.09	13.85	0.89	No
TVIs + REIs	2.17	1.03	13.08	0.88	No
REIs + AF	2.27	1.11	14.08	0.90	No
TVIs + AF	2.24	1.02	12.95	0.93	No

RF VSURF Accuracy

modeling scenario	**RMSECV**	**RMSE**	**RRMSE%**	**R^2**	**selected variables**	**P-value ≤ 0.05**
TVIs + REIs + AF	2.05	0.99	12.76	0.91	B34, B194, B234, SR	Yes
AF	2.27	1.06	13.44	0.87	B34, B194	Yes
TVIs + REIs	2.05	1.02	12.98	0.88	NDVI, SR, SR3, RRI	Yes
REIs + AF	2.09	1.00	12.53	0.89	B37, B193, B194, B234	Yes
TVIs + AF	2.11	1.05	13.34	0.89	B34, B194, B234, SR	Yes

[1]

[1] From Table 10 above the bands indicate the location of the absorption features, **B34** at 638 nm, **B37** at 657 nm, **B194** at 1945 nm, and **B234** at 2345 nm central wavelengths.

4.6.3 Estimating Foliar N:P ratio using Hyperion datasets

Results in Table 11 below show that Hyperion performs improves N:P predictions when using PLSR; it explains a maximum of 62% N:P variability with an RRMSE of 18% and an RMSE value of 1.43. The TVIs and absorption features produced higher accuracies than all other models; the model explained over 90% of N:P variability with the lowest RRMSE of 13%. The selected RF significant variables are NDVI4 (R1095 vs R671), SR (R803 vs R702), B29 (R640), B191(2062 nm) and B220(2355nm).

Table 11: Hyperion regression accuracy.

PLSR Accuracy						
Modeling scenario	**RMSECV**	**RMSE**	**RRMSE%**	**R^2**	**ncomp**	**P-value ≤ 0.05**
TVIs + REIs + AF	2.23	1.43	18.11	0.63	28	No
TVIs + AF	2.08	1.55	19.65	0.56	21	No
REIs + AF	2.48	1.88	23.88	0.35	20	No
AF	2.31	1.98	25.18	0.23	12	No
REIs + TVIs	2.24	1.78	22.64	0.42	17	No
RFR Accuracy						
Modeling scenario	**RMSECV**	**RMSE**	**RRMSE%**	**R^2**	**P-value ≤ 0.05**	
TVIs + REIs + AF	2.31	1.13	14.38	0.91	No	
AF	2.28	1.16	14.78	0.89	No	
REIs + TVIs	2.24	1.07	13.57	0.90	No	
REIs + AF	2.32	1.12	14.22	0.89	No	
TVIs + AF	2.24	1.08	13.68	0.88	No	
VSURF accuracy						
Modelling scenario	**RMSECV**	**RMSE**	**RRMSE%**	**R^2**	**selected variables**	**P-value ≤ 0.05**

TVI + REI +AF	2.12	1.09	13.88	0.89	B29, B191, B220, SR, RRI1	Yes
AF	2.28	1.09	13.93	0.88	B29, B191	Yes
TVIs + REIs	1.98	1.05	13.47	0.83	NDVI4	Yes
REIs + AF	2.21	1.08	13.67	0.90	B29, B191, RRI1, NDRE	Yes
TVIs + AF	2.24	1.06	13.82	0.90	B29, B191, B220, NDVI4, SR2	Yes

[1]

A summary of all the sensor comparisons for both PLSR and RFR are presented in figures 13 and 14 respectively. In the figures below only the best modeling scenarios are selected for each sensor, regression plots for other scenarios are presented in Appendix A at the end of this document.

[1] The bands in Table 11 indicate the location of absorption features, **B29** at 640 nm, **B191** at 2063 nm, and **B220** at 2355 nm central wavelengths.

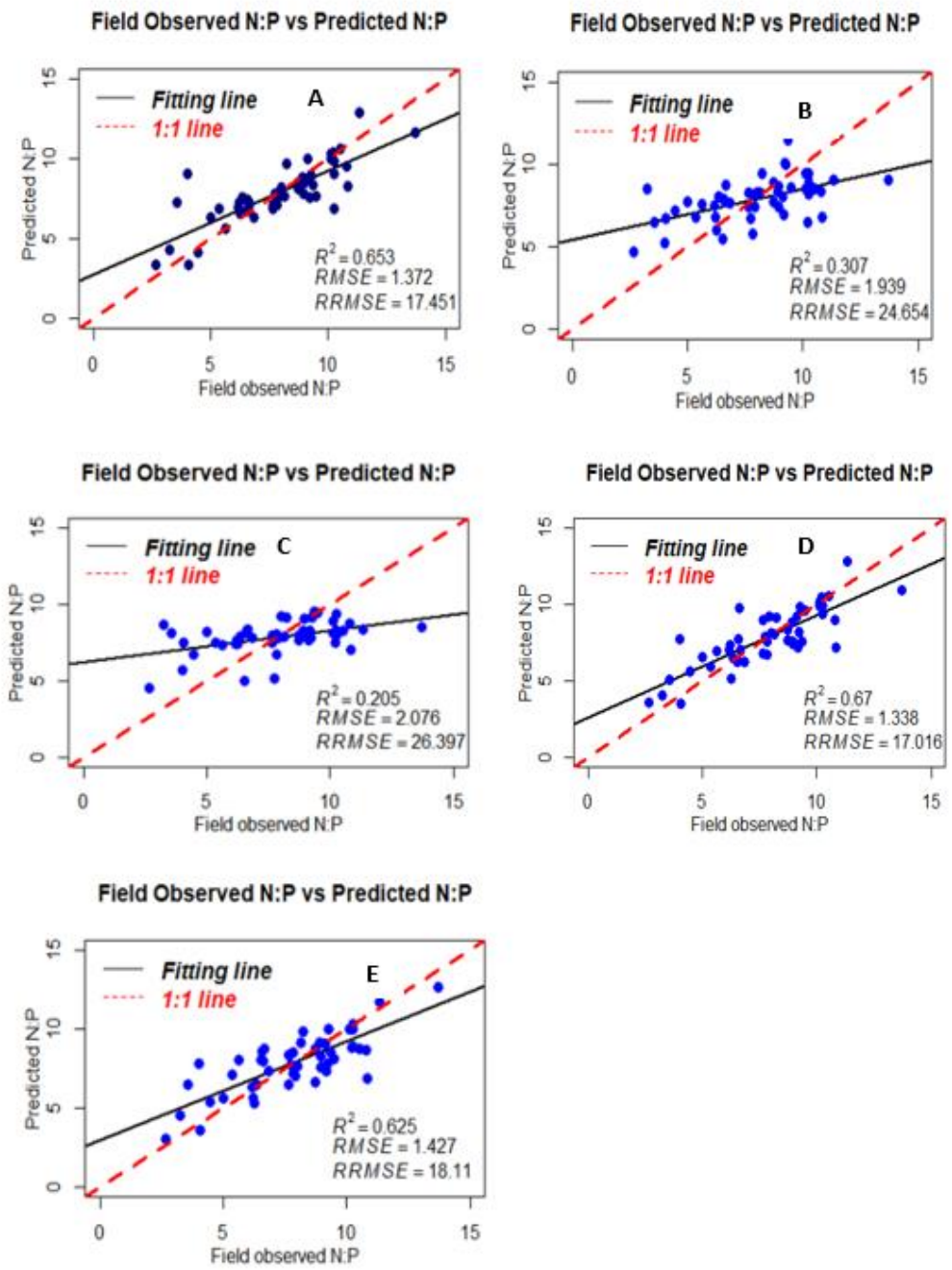

Figure 14: Final PLSR models for all the multispectral and hyperspectral sensors. A is Sentinel-2, B is WorldView-2, C is RapidEye, D is EnMap and E is Hyperion.

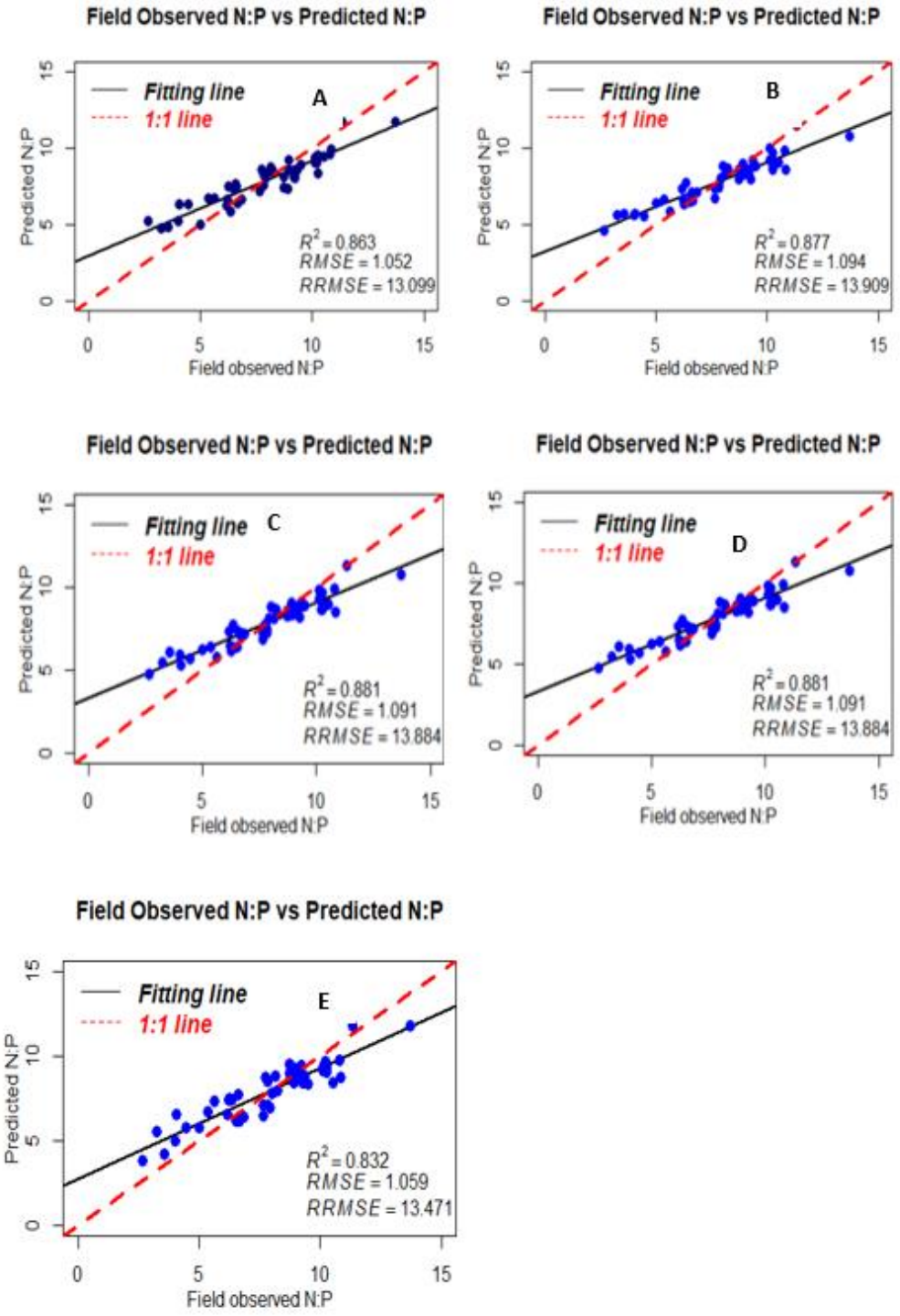

Figure 15: RF regression plots for all the multispectral and hyperspectral sensors. A is Sentinel-2, B is RapidEye, C is EnMap, D is WorldView-2, and E is Hyperion.

CHAPTER 5:

Discussion

5.1 The potential of Sentinel-2 for predicting variability of N:P ratio

ESA Copernicus Sentinel-2 missions have long been tested and validated for quantifying vegetation's biophysical and biochemical content and demonstrated a propitious opportunity for remote sensing of vegetation. This study evaluated the capabilities of simulated Sentinel-2 datasets for predicting and mapping the spatiotemporal distribution of foliar N:P ratio in a heterogeneous savanna system using PLSR and RFR. The PLSR and RFR showed different N:P ratio predictive capabilities; overall results indicated high prediction accuracies when using the RFR, which explained 87% of the N:P ratio variability with an error term of 13.8%. While the PLSR explained 67% of N:P ratio variability with a higher error of 21%, see Table 7 for reference. This work suggests that non-parametric regression techniques are more optimum for modeling foliar biochemical content than parametric regression algorithms, which corresponds to the findings of previous studies. For example, Ramoelo *et al.* (2015) demonstrated that random forest regression accurately predicts foliar N in a savanna landscape. The potential of multispectral data and RF for estimating nutrient limitation in an Alpine grassland is reported in Gao *et al.* (2020).

The results of the present study revealed that in a heterogenous savanna environment, Sentinel-2 bands located in the visible, red-edge, and NIR regions are significant for N:P ratio estimation at the end of the wet season. Gao *et al*. (2020) found a similar response in an alpine grassland during the senescence stage. However, their study also highlighted the importance of bands in the SWIR, which is not the case in our study. This might be the result of the different seasons in which the field measurements were taken which determines leaf water content of vegetation, the water content is highly associated with this region. In addition, most of the REIs, including the ones optimised for this study, demonstrated good potential for estimating the N:P ratio; some of these red-edge based VIs were reported to be ideal for N content in leaves (Ramoelo *et al*., 2015). The high correlation between foliar N and the red-edge and NIR wavelengths is well presented in the literature see (Lepine *et al*., 2016; Ramoelo *et al*., 2014). Foliar P

correlates significantly with NIR and SWIR regions (Mutanga & Kumar, 2007). Overall variable importance of this work indicated that Sentinel-2 spectral wavelengths at 490 nm, 665nm, and 705 nm are significant for N:P ratio prediction.

Normalised difference and simple ratio indices derived from the red-edge bands were consistently selected as important variables in all Sentinel-2 indices-based models, some conventional VIs were also selected in some scenarios. The consistent significance of red-edge derived indices reveals the importance of the red-edge region for the estimation of micronutrients. The combination of significant bands and indices further improves the prediction accuracy of the foliar N:P ratio. Past studies also demonstrated the effectiveness of combining Sentinel-2 MSI spectral and vegetation indices for predicting foliar biochemical content (Gao *et al.,* 2020). The overall results of this research validate the applicability of Sentinel-2 spectral bands and indices to retrieve foliar N:P ratio.

This study further demonstrated the transferability of ASD-derived models to satellite images. When the RF robust performing model was inverted to Sentinel-2 satellite images, it produced meaningful N:P ratio distribution maps with values falling within the observed N:P ratio range. The heterogeneous nature of the study area resulted in a slight over-prediction of the N:P ratio in the satellite images. The data used to calibrate the models was collected from grass samples. However, the satellite images also mapped the tree canopies. Trees have different biochemical concentrations than grass, thus could affect model transferability and hence the course texture of N:P ratio distribution maps in Fig 10 and 11 in chapter 4.

Based on Gusewell, (2003) N:P threshold values, it is evident that the study area is experiencing N-limitation and co-limitation with a foliar N:P ratio ranging from 4 to 12. Nutrient limitation is not consistent in the area, i.e., some areas are N-limited, and some are co-limited. The observed spatiotemporal dynamics in the N:P ratio could also result from environmental factors, such as geology, soil properties, and climatic conditions. Venter *et al*. (2003) highlighted that the gabbro and granite-derived soils covering different parts of the study area as explained in chapter 3 differ in their physical and chemical composition; therefore, their nutrient content differs. According to Ludwing *et al*. (2001), an average N:P ratio of six in an open grassland depicts N limitation in an open grassland. Craine *et al*. (2008) used a factorial fertilisation experiment to evaluate the use of the N:P

ratio as an indicator of nutrient limitation. Their study reported that at the KNP, unfertilized areas had an average N:P ratio of 5.8. Both these values correspond with the foliar N:P ratio average value of 7.86 observed in this study. Therefore, there is evidence of continuous nutrient limitation at KNP, and an incline in N limitation is observed from 2018 to 2021 (see fig 1, chapter 4). It was reported in the previous literature that the range of N:P ratio values is influenced by variation in grass species. Gao *et al*. (2020) found an N:P ratio range of 8.06 to 18.95 in alpine grassland, while Loozen *et al*. (2019) reported that the N:P ratio values of a *Holcus lanatus L*. (commonly known as Yorkshire fog) range from 6.1 to 75.

5.2 The Sentinel-2 N:P ratio Predictive capabilities compared to other sensors.

The statistical accuracy from this study indicates that there are no significant differences in accuracy for all the sensors when using RFR. However, when using the PLSR, the accuracy of the multispectral sensors is low compared to the hyperspectral sensors. The final accuracy is measured based on the RRMSE instead of the R^2 because the R^2 values, specifically for the RFR models are exaggerative. This could be associated with the low number of sampling points which led to the use of cross-validation instead of dividing the datasets to train and test datasets. Sentinel-2 accuracy is greater than that of all other multispectral sensors, however, the difference is marginal between the PLSR accuracy of hyperspectral datasets and Sentinel-2 datasets. All these sensors have varying spatial resolutions, and Sentinel-2 has the lowest resolution of 10 m. This could imply that the spatial resolution does not necessarily influence the capabilities of a sensor to retrieve foliar biochemicals, but the spectral resolution is influential. Even though the hyperspectral and the commercial multispectral sensors performed better than Sentinel-2 in PLSR models, their applicability is still limited by their financial costs.

Moreover, the limited area coverage and multicollinearity experienced with the hyperspectral sensors hinder their utility (Adjorlolo *et al*., 2015). The RFR used in this study demonstrated the comparability of the freely available Sentinel-2 MSI, which provides regional coverage with commercial sensors. The narrow intervals of 1.4 nm from 350-1000 nm and 2 nm from 1000-2500 nm of the hyperspectral sensors are possibly the cause of the high prediction accuracies achieved by these sensors (Sibanda *et al*., 2015;

Abdel-Rahman *et al*., 2014). Comparability of Sentinel-2 with the commercial multispectral and hyperspectral sensors demonstrates an immense potential for the remote sensing of foliar biochemicals, especially in developing countries where the accessibility of commercial sensors by the public is still a hitch.

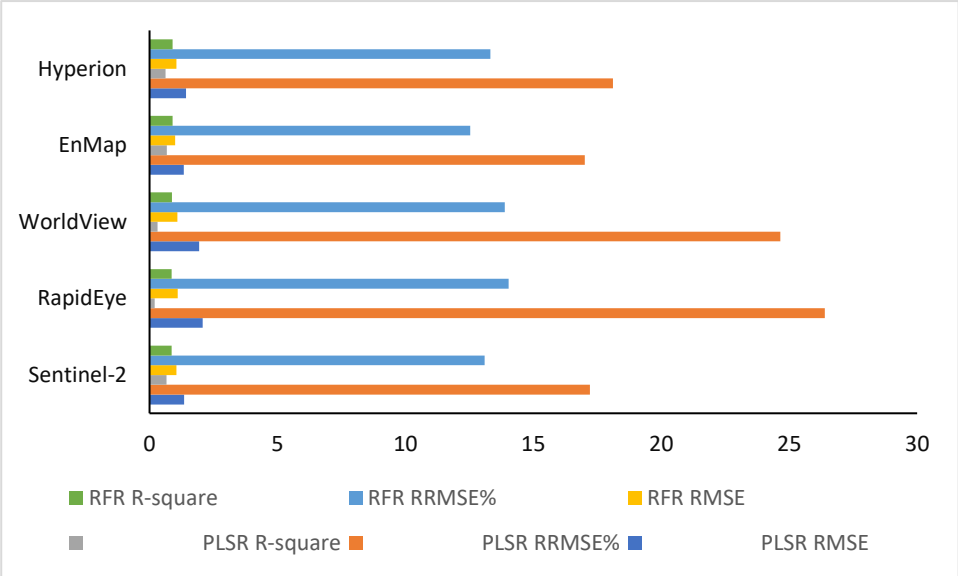

Figure 16: A bar graph showing a comparison of PLSR and RFR accuracies for each sensor.

From figure 13 above, we note that the estimation accuracy of the Sentinel-2 datasets is higher than the commercial sensors for both parametric and non-parametric regression algorithms. Sentinel-2 accuracy is closely the same as the accuracy of the hyperspectral sensors (i.e., EnMap and Hyperion). This emphasises the influence of spectral configuration and bandwidth for estimating biochemical content.

For the parametric algorithm, highest prediction accuracy (closest to the measure N:P values) is attained when using Hyperion datasets. Sentinel-2 PLSR predicted values are relatively close to the Hyperspectral predicted values. RapidEye predicted values, on the other hand, are much lower than all the sensors, and this could be associated the influence number of predictors used in the PLSR model, RapidEye has the lowest number of bands compared to all other sensors used. Therefore, few VIs could be optimized resulting in few input variables for modelling, the number of predictors greatly influence the accuracy of parametric regression algorithms. A comparison of PLSR predicted N:P ratio versus the measured values is presented in Fig 14 for all the sensors. There are no noticeable differences in all the predicted values when using RFR, see Fig 15.

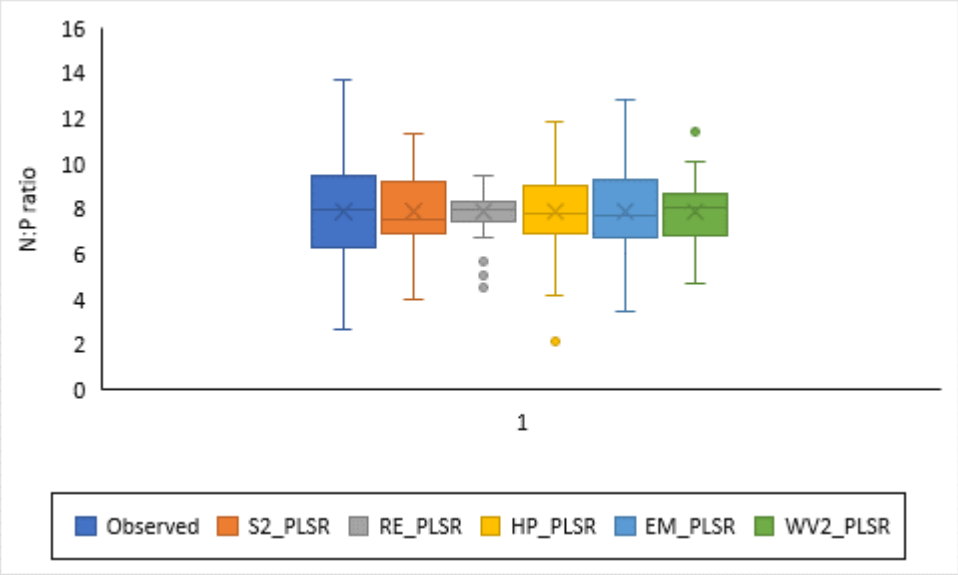

Figure 17: A box and whisker plot showing a comparison between field measured and the PLSR predicted N:P ratio values for all the simulated datasets from best predicting scenarios.

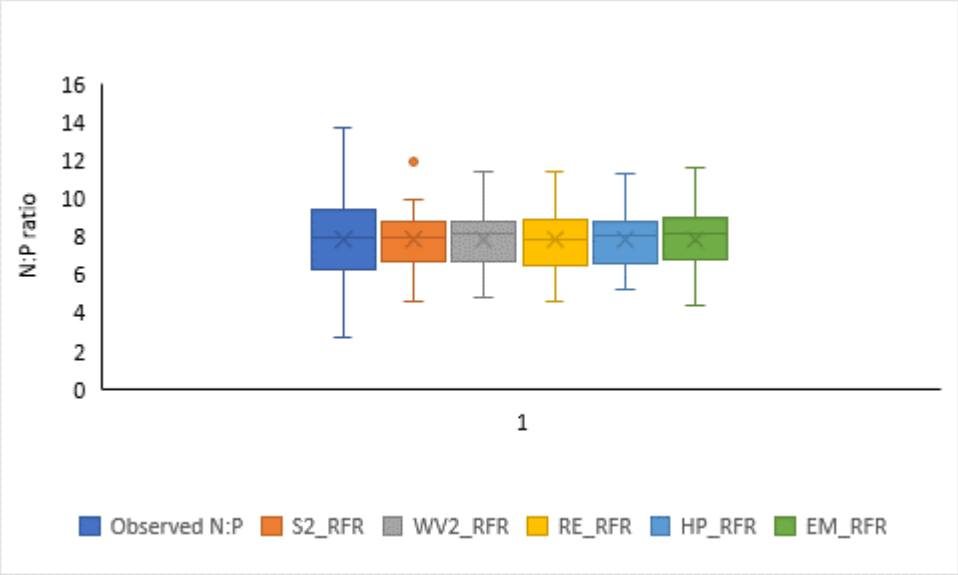

Figure 18: A box and whisker plot showing a comparison between field measured N:P ratio and the RFR predicted values for all the simulated datasets from best predicting scenarios.

5.3 Ideal bands and indices for N:P ratio estimation

The RFR significant variable selection indicated that N:P ratio sensitive bands are in the Visible, SWIR, and Red-edge regions. Even though the Sentinel-2 NIR band (B8) was

not selected in RF variable importance, it achieved a higher PLSR weight than all Sentinel-2 bands when using the parametric technique. The selected significant bands for all the sensors are associated with the known N and P absorption features, for example, 460 nm and 660 nm for WorldView-2, Sentinel-2, and RapidEye, and absorption features at 640 nm, 1940nm, 1950 nm, and 2060nm for Enmap and Hyperion. This may be attributed to the high correlation between the N:P ratio to N and P (Gao *et al.,* 2020). These absorption features and the Sentinel-2 red-edge band at 705 nm portray a potential for remote sensing-based N:P ratio retrieval. Surprisingly, in this study, the bands associated with the SWIR region reported to be sensitive to the foliar N:P ratio in the literature showed no significance in all the multispectral sensors when using the RFR. This could be linked to the broadness of the SWIR in Sentinel-2 MSI, the studies that indicated the importance of the SWIR were based on the use of hyperspectral datasets, for example, Ramoelo *et al.,* (2013). The results showed the consistency of the blue bands in the multispectral sensors. Loozen *et al*. (2019) also observed a consistent relationship between the foliar N:P ratio and the blue region. However, the concern is the misinterpretation that might result from the effect of Rayleigh scattering when using satellite images.

The combination of conventional and red-edge-based indices yielded higher estimation accuracy for Sentinel-2 and Hyperion models, whereas, for other sensors, higher estimation accuracy was attained by the models derived from REIs and bands/ or absorption features. The combination of NIR and RE1 (705 nm) indices selected in this study were also identified significant for N:P ratio estimation from Sentinel-2 by Gao *et al.* (2020). The performance of bands and indices varied for each sensor. The estimation uncertainty of indices was observed in Pacheco-Labrador *et al*. (2014); their study revealed that most known vegetation indices were not correlated with canopy N content in Holm oak leaves. This inconsistency of indices indicates the impact of sensor configurations in foliar micronutrient prediction accuracies. Regardless of the unavailability of specific N:P absorption features, the selected N and P absorption features and their associated indices demonstrate a promising potential for estimating nutrient limitation in a heterogeneous savanna landscape. A summary of RFR and PLSR selected variables is presented in Table 12, Table 13, and figure 16 respectively.

Table 12: RF selected significant VIs.

Table 13: RF selected important bands.

Spectral region (nm)	selected bands	sensor
BLUE	B2	Sentinel-2
	B3	WorldView-2
Green	B4	WorldView-2
	B2	RapidEye
Red	B4	Sentinel-2
	B6	WorldView-2
	B3	RapidEye
	B29	Hyperion
	B34	EnMap
Red edge	B5	Sentinel-2
SWIR	B194, B234	EnMap
	B191	Hyperion

Sensor	Known VIs	Optimised VIs
Sentinel-2	NDRE, NDRE2, RRI	RRI5, NDRE4
RapidEye	SR	NDRE, RRI
WorldView-2	SR	NDRE
EnMap	SR, NDVI	RRI
Hyperion	SR	RRI1, NDVI4, NDRE

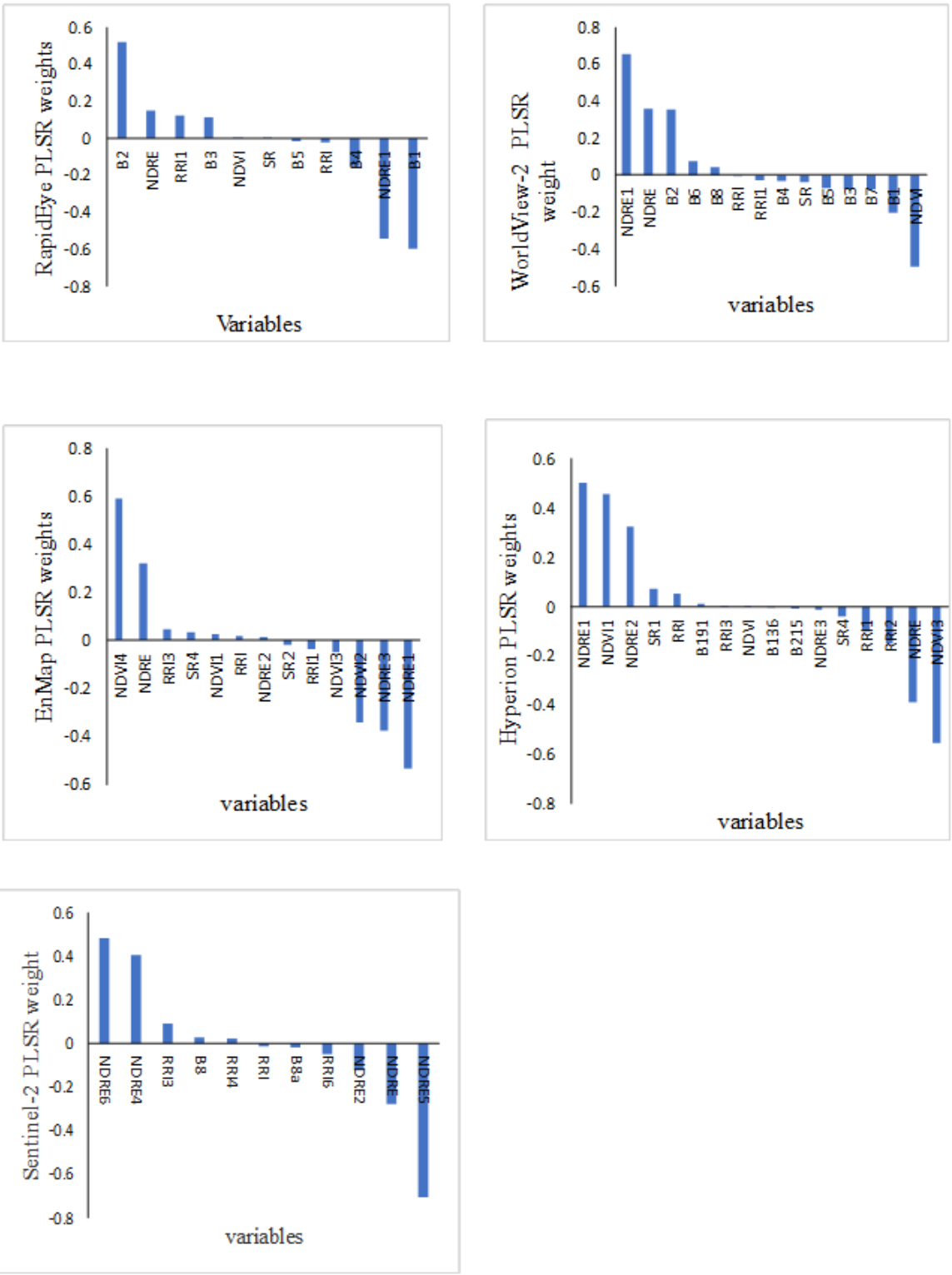

Figure 19: PLSR weights for the variables used in each sensor. The weights denotes the importance of each variable in the best modeling scenarios for each sensor.

CHAPTER 6: CONCLUSION

6.1 General Summary

Nutrient limitation is consistent in the central parts of the KNP area; prolonged limitation of nutrients poses a severe threat to ecosystem functioning, biodiversity, and ecotourism services rendered by the park. N seems to be more limited compared to P in the central to southern savanna landscape of KNP. Exploring the potential of freely available sensors for predicting and mapping foliar N:P ratio is necessary to monitor the nutritional status of the savanna environment on a near-real-time basis. This is a crucial step towards the implementation of ideal management and nutrient recovery strategies. The main aim of this study was to develop a Sentinel-2 N:P ratio prediction model and assess the transferability of simulated ASD datasets derived models to satellite images for regional mapping of nutrient limitation in the northern parts of the South African Savanna landscape. To achieve this aim, the following objectives were executed: identify N:P ratio sensitive bands and indices, develop an ideal model, and invert the model to satellite images.

The results of this study depict that the vegetation indices derived from Sentinel-2 red-edge and narrow NIR bands have the potential to detect foliar N:P ratio in a heterogeneous savanna system when using a suitable regression algorithm. Models derived from ASD datasets can be inverted to satellite images to upscale the prediction and mapping of nutrient limitation. Moreover, the study determined the prospective of simulated multispectral datasets for addressing the lack of remote sensing variables significant for estimating the foliar N:P, consequently attributed to the neglect of the distribution of nutrient limitation in natural environments. The transferability of these models to regional maps is critical for managing the savanna ecosystem. It can be concluded that the freely available Sentinel-2 MSI is compatible with high-resolution commercial sensors. Therefore Sentinel-2 can be adopted as an accurate, cost-effective approach to decision-makers and policymakers to address the issue of natural grassland degradation through the implementation of scientifically supported policies to ensure the management, sustainability, and restoration of savanna biodiversity.

6.2 Contribution to scientific knowledge for determining and mapping foliar N:P ratio.

The RF variable selection (Vsurf) in chapter 4 selected 3 optimal bands for predicting the N:P ratio from simulated Sentinel-2 r reflectance data. The ideal bands are located at the visible (490nm and 665nm) and red-edge (705 nm) regions of the electromagnetic spectrum. Red-edge derived vegetation indices; the RF also selected NDRE, NDRE2, NDRE4, RRI, and RRI5 based on Vsurf of all the modeling scenarios. An additional band in the NIR (B8) region and two additional REIs indices, NDRE6, and RRI3, were also selected as significant based on the PLSR weight. As part of our analysis in chapter 4, important N:P prediction bands and indices were also identified from the simulated reflectance data for other sensors for WorldView-2, bands located in the visible region (480 nm, 560 nm, and 660 nm together with SR and NDRE VIs were selected by VSurf. While the PLSR weights selected bands at visible and NIR regions, 425 nm and 810 nm, respectively, and the NDRE and NDRE1. Also, for the RapidEye, only bands at the visible region (555 nm and 658 nm) were selected, and the SR, NDRE, and RRI VIs. The RapidEye Vsurf and PLSR weight significant variables were similar. Only the RRI1 was selected by PLSR weight but not by the Vsurf.

For the hyperspectral datasets, Vsurf selected absorption features are in the visible, and the SWIR. Absorption features allocated at 638nm, 1945 nm, 2345 nm for EnMap; 640 nm, and 2063 nm for Hyperion. Vsurf-selected EnMap indices are SR, NDVI, and RRI, and Vsurf-selected indices for Hyperion are SR, RRI1, and NDRE. The PLSR weight selected only VIs, NDVI4, NDRE, RRI3, SR4 and NDVI1 for EnMap; and NDRE1, NDRE2, NDVI1, SR1 and RRI for Hyperion. Based on both Vsurf and PLSR selection, a general conclusion can be drawn that reliable N:P ratio estimation bands are in the visible, red-edge, and NIR for multispectral reflectance datasets, and for the hyperspectral datasets, some bands in the SWIR region are significant. Moreover, the narrower Sentinel-2 red-edge bands (705 nm) are more sensitive to the N:P ratio compared to the broad red-edge bands of WorldView-2 and RapidEye. Lastly, both red-edge based, and traditional indices normalised difference and simple ratio, are optimal for N:P ratio retrieval.

6.3 Recommendations for future research

The heterogeneity of the savanna system might cause inconsistencies in the wavelength sensitivity of different sensors to the N:P ratio, which could affect the transferability of ASD-derived models. To account for the different structural factors in future research, we recommend the following:

- The inclusion of textural information is recommended for future research. This could be achieved by integrating data from passive optical sensors with datasets from structurally sensitive sensors such as radio detection and ranging (RADAR) and light detection and ranging (LIDAR). Active sensors provide information on height, density, texture, and structure. As such, they are essential for developing stratified prediction models, which seems necessary in natural, non-uniform environments.
- The use of radiative transfer models (RTMs) originally developed to detect heterogeneous canopy structures in N:P ratio estimation models should be investigated.
- Lastly, to understand the observed spatio-temporal variations in nutrient limitation, the influence of ancillary variables such as soil physical and chemical properties, climatic conditions, underlying geology, and fire frequencies should be investigated. Moreover, grass species should also be included when modeling foliar N:P ratio, as the literature suggests varying N:P threshold values amongst grass species.

APPENDIX A:

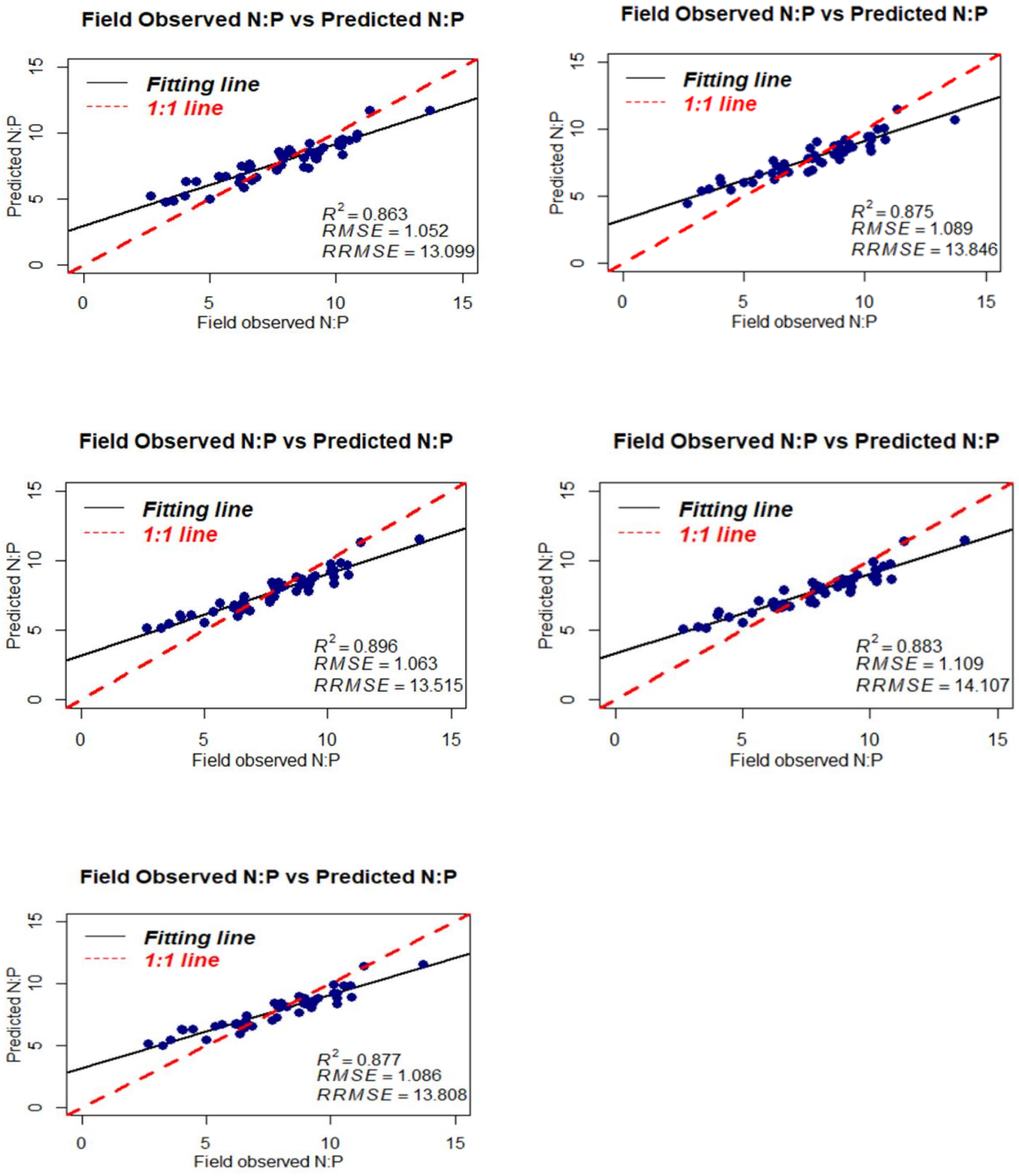

Fig 1A: Sentinel-2 RF regression plots depicting the relationship between the measured N:P ratio and the RF predicting.

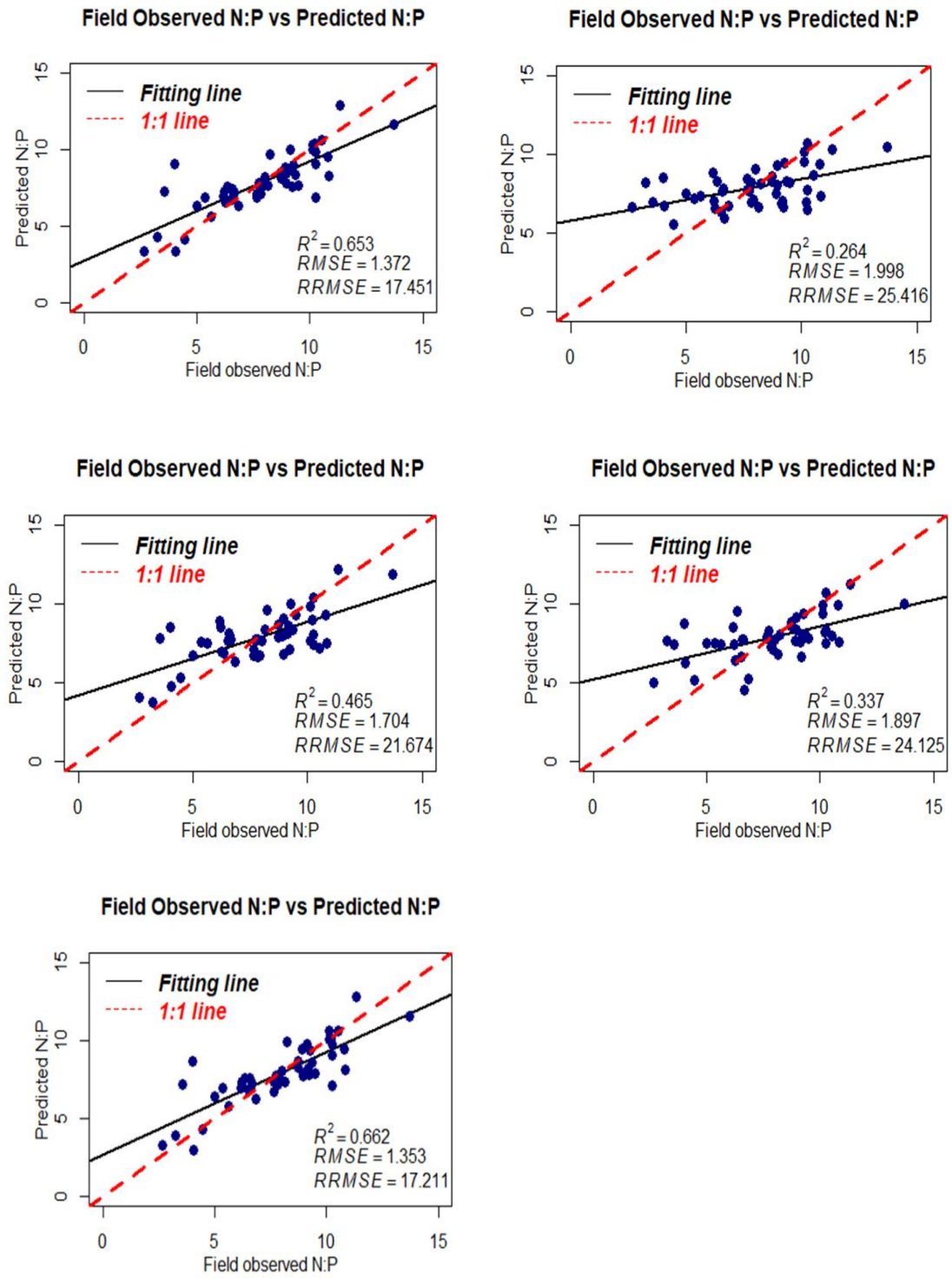

Fig 2A: Sentinel-2 PLSR plots depicting the relationship between the PLSR estimated and observed N:P ratio.

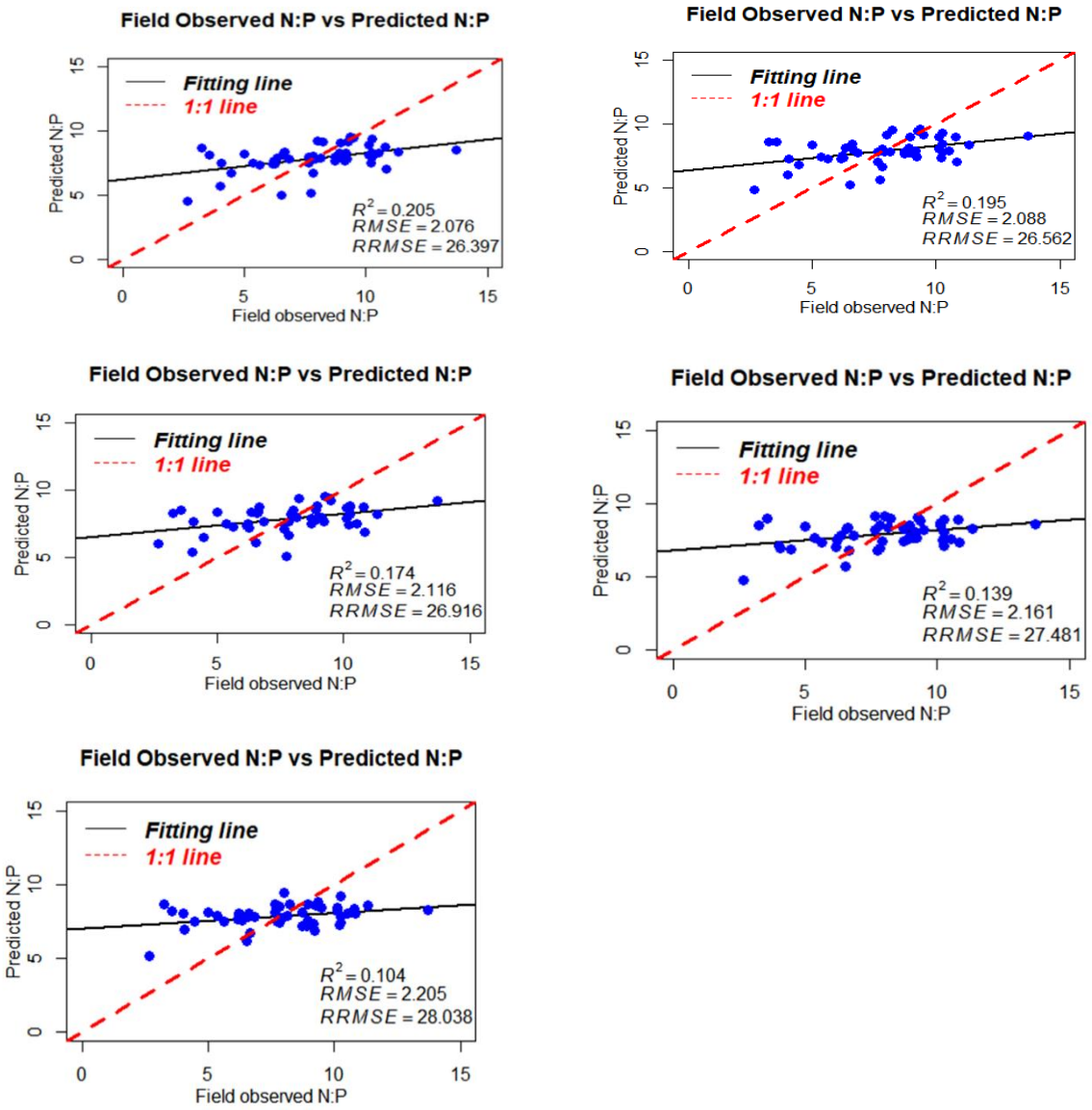

Fig 3A: RapidEye PLSR plots. Showing the relationship between the PLSR estimated and the observed N:P ratio.

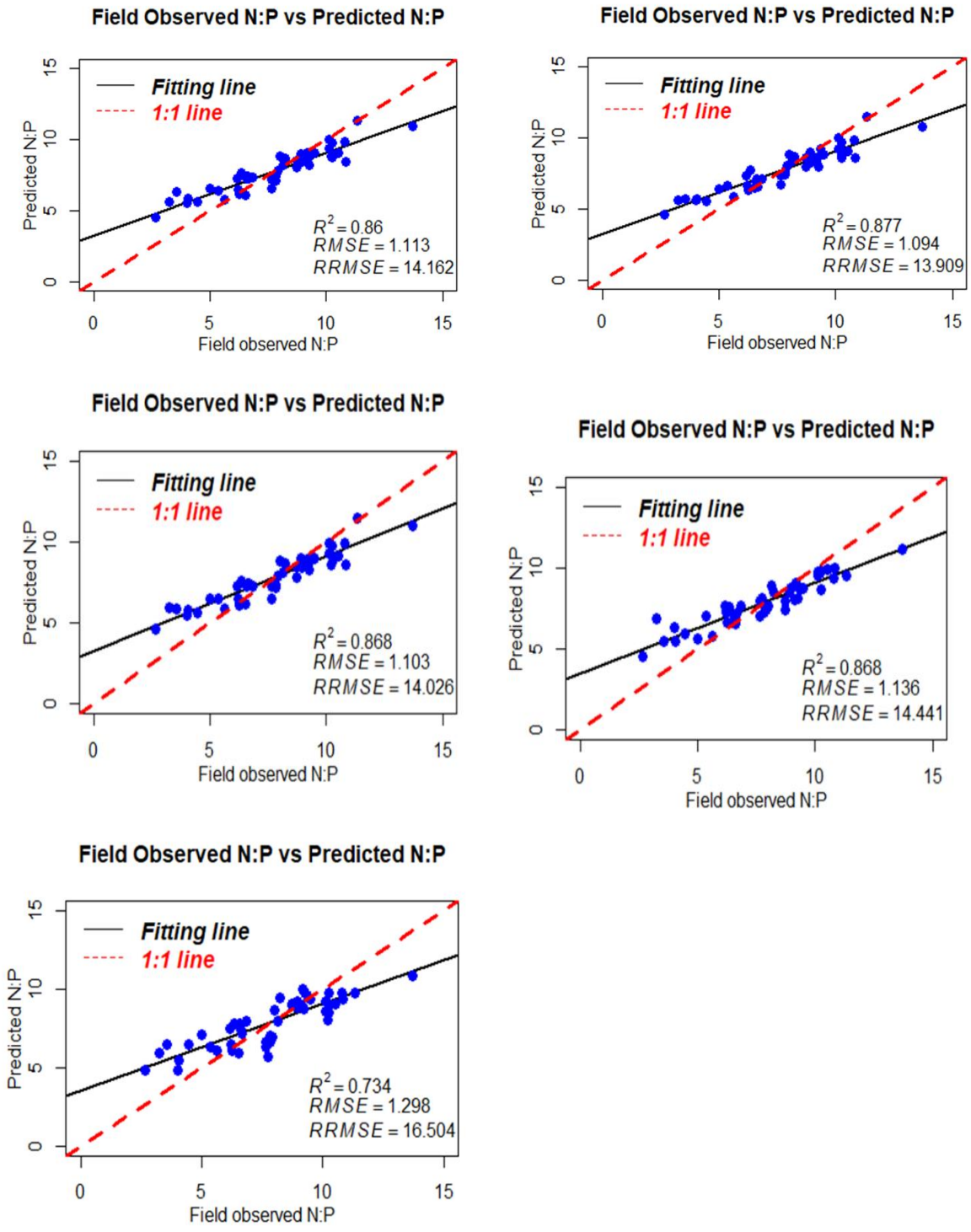

Fig 4A: RapidEye RF regression plots showing the relationship between the predicting and the field measured N:P ratio.

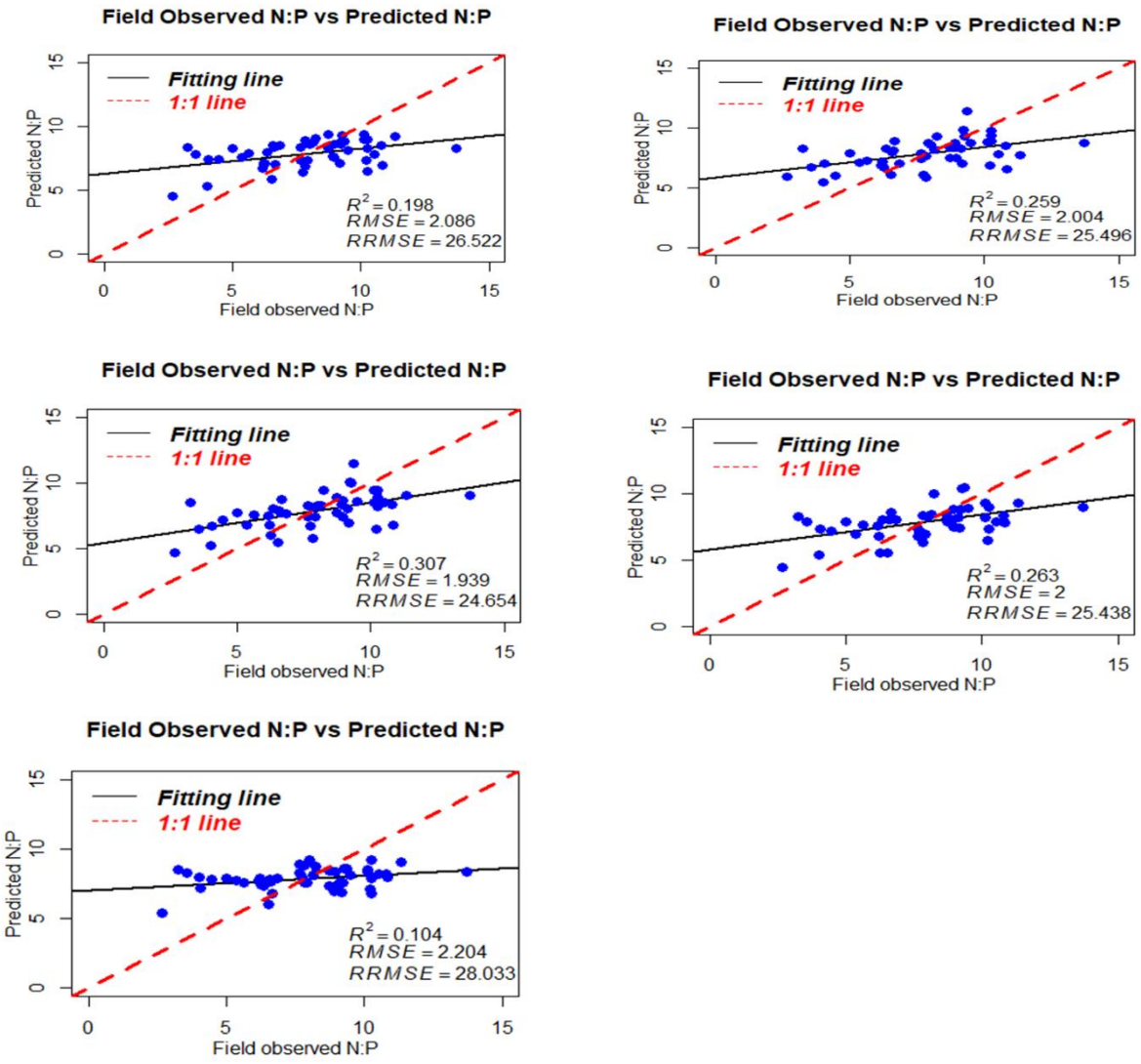

Fig 5A: WorldView-2 PLSR regression depicting the relationship between the observed and predicted N:P ratio.

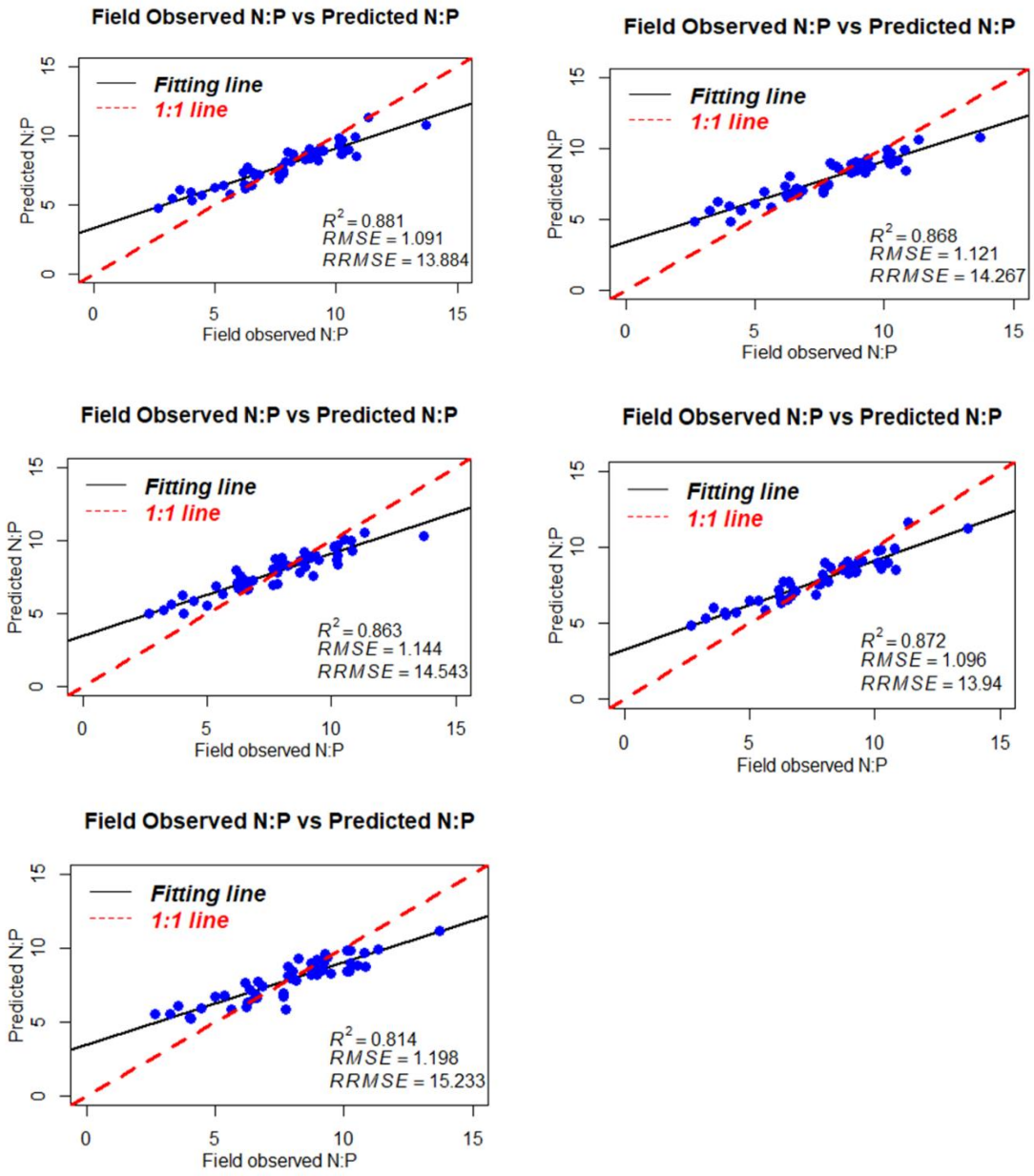

Fig 6A: WorldView-2 RF regression plots showing the relationship predicted and measured N:P ratio.

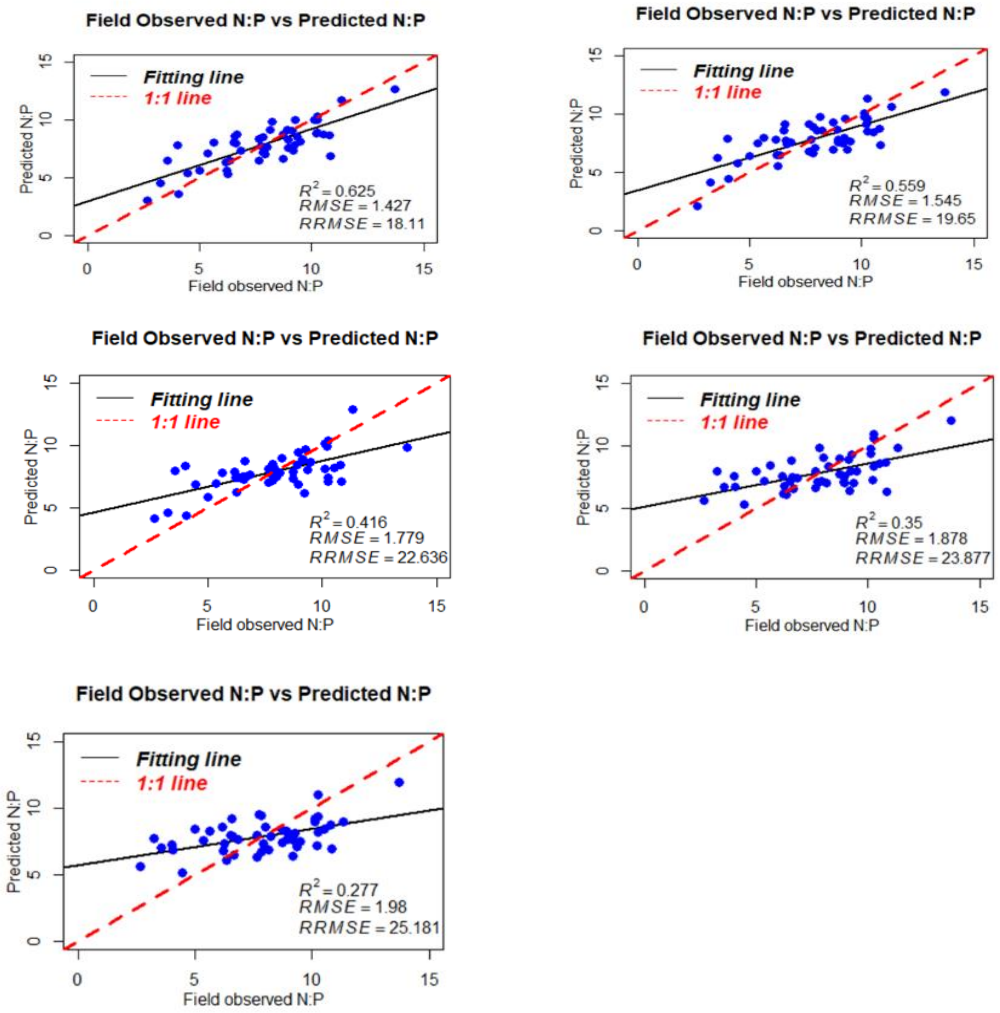

Fig 7A: Regression plots showing the relationship between the measured and the predicted N:P ratio from Hyperion datasets using PLSR.

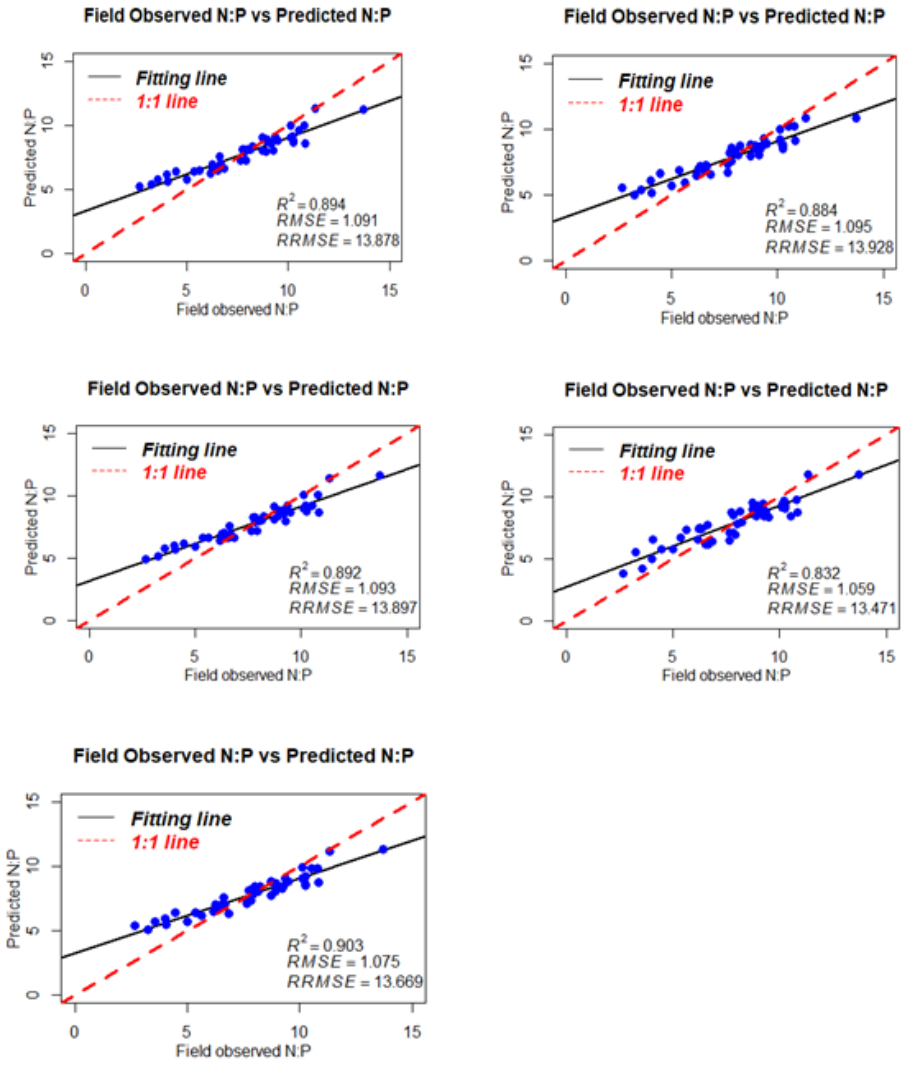

Fig 8A: The regression plots showing the relationship between the measured and the RFR estimated N:P ratio using Hyperion datasets.

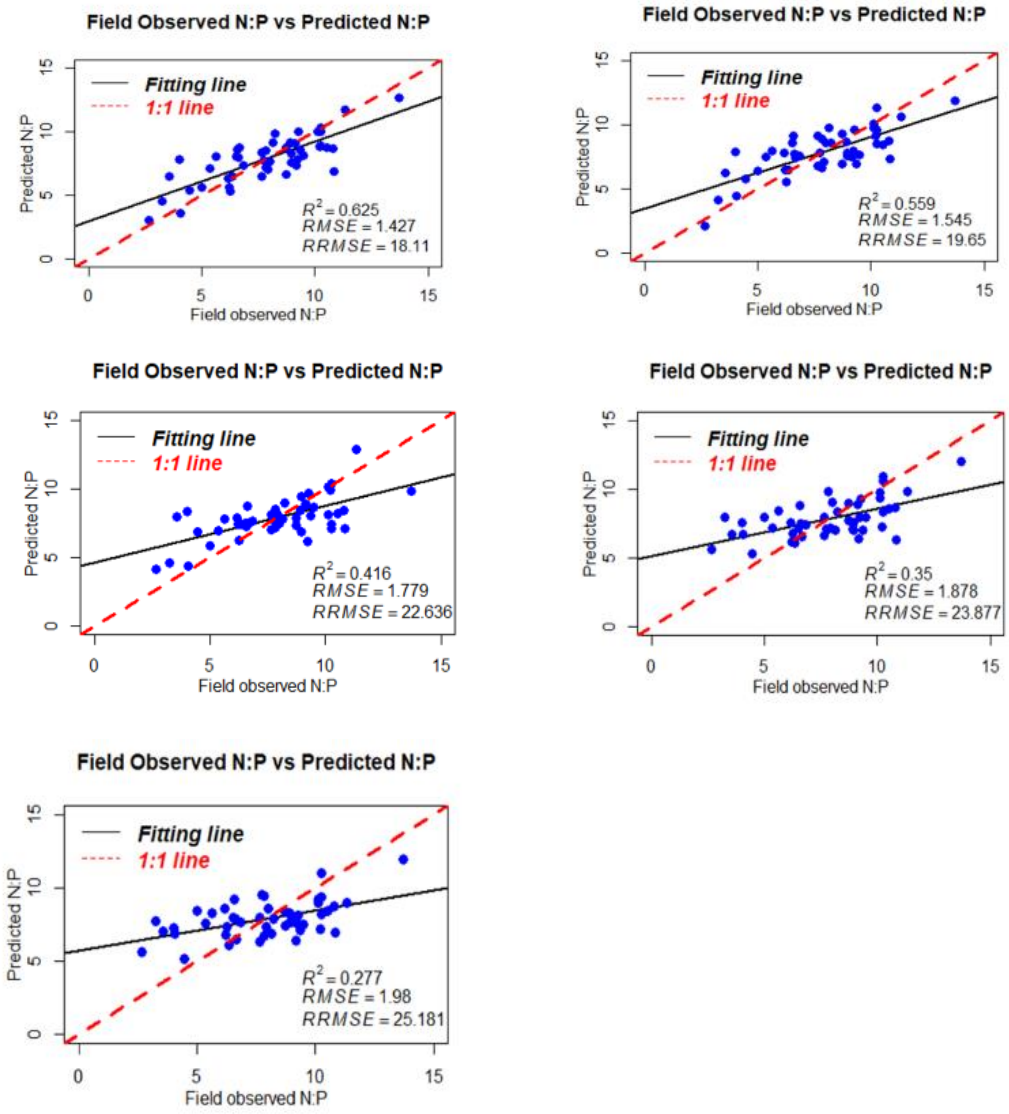

Fig 9A: The 1:1 regression plots showing EnMap PLSR predicted N:P ratio against the measured N:P ratio.

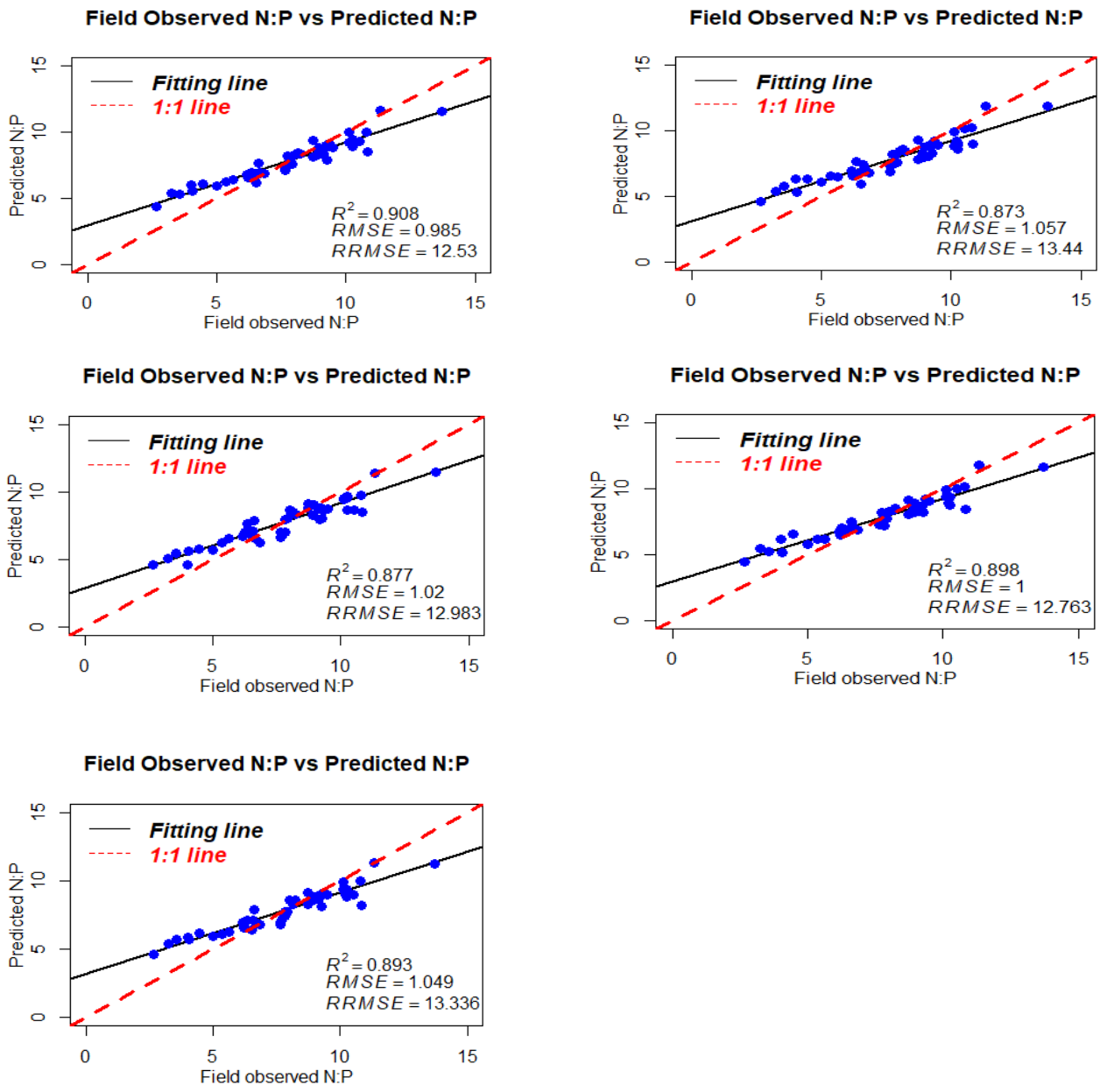

Fig 10A: EnMap PLSR regression plots depicting the relationship between the observed and the estimated N:P ratio.

References

1. Abdel-Rahman, E.M., Ahmed, F.B. and Van den Berg, M., 2010. Estimation of sugarcane leaf nitrogen concentration using in situ spectroscopy. International Journal of Applied Earth Observation and Geoinformation, 12, pp. S52-S57.
2. Abdipour, M., Younessi-Hmazekhanlu, M. and Ramazani, S.H.R., 2019. Artificial neural networks and multiple linear regression as potential methods for modeling seed yield of safflower (Carthamus tinctorius L.). Industrial crops and products, 127, pp.185-194.
3. Adjorlolo, C., Mutanga, O. and Cho, M.A., 2015. Predicting C3 and C4 grass nutrient variability using in situ canopy reflectance and partial least squares regression. International Journal of Remote Sensing, 36(6), pp.1743-1761.
4. Ali I., Cawkwe F., Dwyer E., Barrett B., Green S., 2016, Satellite remote sensing of grasslands: from observation to management. Oxford journals.
5. Askari M.S., McCarthy T., Magee A., Murphy D.J., 2019. Evaluation of Grass Quality under Different Soil Management Scenarios Using Remote Sensing Techniques. MDPI remote sensing
6. Asner, G.P., 1998. Biophysical and biochemical sources of variability in canopy reflectance. Remote sensing of Environment, 64(3), pp.234-253.
7. Avitabile, V., Baccini, A., Friedl, M.A. and Schmullius, C., 2012. Capabilities and limitations of Landsat and land cover data for aboveground woody biomass estimation of Uganda. Remote Sensing of Environment, 117, pp.366-380.
8. Bachelet, D., Lenihan, J.M., Daly, C. and Neilson, R.P., 2000. Interactions between fire, grazing and climate change at Wind Cave National Park, SD. Ecological modelling, 134(2-3), pp.229-244.
9. Beatty, D.T., Barnes, A., Taylor, E., and Maloney, S.K., 2008. Do changes in feed intake or ambient temperature cause changes in cattle rumen temperature relative to core temperature? Journal of Thermal Biology, 33(1), pp.12-19.

10. Bakker, M.A., Carreño-Rocabado, G. and Poorter, L., 2011. Leaf economics traits predict litter decomposition of tropical plants and differ among land use types. Functional Ecology, 25(3), pp.473-483.
11. Ben-Shahar, R., 1996. Woodland dynamics under the influence of elephants and fire in northern Botswana. Vegetation, 123(2), pp.153-163.
12. Mevik, B.H. and Wehrens, R., 2015. Introduction to the pls Package. Help Section of the “Pls” Package of R Studio Software, pp.1-23.
13. Breiman, L. and Ihaka, R., 1984. Nonlinear discriminant analysis via scaling and ACE. Davis One Shields Avenue Davis, CA, USA: Department of Statistics, University of California.
14. Breiman, L., 2001. Random forests. Machine learning, 45(1), pp.5-32.
15. Breiman, L., 1996. Bagging predictors. Machine learning, 24(2), pp.123-140.
16. Bronn, A. v. Z., Lombard, P. J. L. & Potgeiter, A. L. F. (2001). The effect of long-term fire regime and herbivory on the Colophospermum mopane shrub veld in the Kruger National Park. Poster presented at 36th Annual Congress of the Grassland Society of Southern Africa, Aldam, South Africa. Reported in Venter et al. (2003).
17. Carlier, L., Rotar, I., Vlahova, M. & Vidican, R., 2009, ‘Importance and functions of grasslands’, Notulae Botanicae Horti Agrobotanici Cluj-Napoca 37(1), 25–30.
18. Carlier, L., Rotar, I., Vlahova, M. & Vidican, R., 2009, ‘Importance and functions of grasslands’, Notulae Botanicae Horti Agrobotanici Cluj-Napoca 37(1), 25–30.
19. Chabalala, Y., Adam, E., Oumar, Z. and 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), pp.379-395.
20. Cho, M.A., and Skidmore, A.K., 2006. A new technique for extracting the red edge position from hyperspectral data: The linear extrapolation method. Remote sensing of environment, 101(2), pp.181-193.
21. Clevers, J.G. and Gitelson, A.A., 2013. Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and-3.

International Journal of Applied Earth Observation and Geoinformation, 23, pp.344-351.

22. Clevers, J.G.P.W., 2014. Beyond NDVI: extraction of biophysical variables from remote sensing imagery. In Land use and land cover mapping in Europe (pp. 363-381). Springer, Dordrecht.
23. Clifton, K.E., Bradbury, J.W. and Vehrencamp, S.L., 1994. The fine-scale mapping of grassland protein densities. Grass and Forage Science, 49(1), pp.1-8.
24. Cohen, W.B. and Goward, S.N., 2004. Landsat's role in ecological applications of remote sensing. Bioscience, 54(6), pp.535-545.
25. Craine, J.M., Morrow, C., and Stock, W.D., 2008. Nutrient concentration ratios and co-limitation in South African grasslands. New Phytologist, 179(3), pp.829-836.
26. Curran, P.J., 1989. Remote sensing of foliar chemistry. Remote sensing of environment, 30(3), pp.271-278.
27. Cutler, A., Cutler, D.R. and Stevens, J.R., 2009. Tree-based methods. In High-Dimensional Data Analysis in Cancer Research (pp. 1-19). Springer, New York, NY.
28. De Peppo, M., Taramelli, A., Boschetti, M., Mantino, A., Volpi, I., Filipponi, F., Tornado, A., Valentini, E. and Ragaglini, G., 2021. Non-Parametric statistical approaches for leaf area index estimation from Sentinel-2 Data: A multi-crop assessment. Remote Sensing, 13(14), p.2841.
29. Dorigo, W.A., Zurita-Milla, R., de Wit, A.J., Brazile, J., Singh, R. and Schaepman, M.E., 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. International journal of applied earth observation and geoinformation, 9(2), pp.165-193.
30. Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P. and Meygret, A., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote sensing of Environment, 120, pp.25-36.

31. Dube T., Mutanga O., 2015. Evaluating the utility of the medium-spatial resolution Landsat 8 OLI multispectral sensor in quantifying above-ground biomass in uMngeni catchment, South Africa. ISPRS journal of photogrammetry and remote sensing.
32. Martre, P., Reynolds, M.P., Asseng, S., Ewert, F., Alderman, P., Cammarano, D., Maiorano, A., Ruane, A.C., Aggarwal, P.K., Anothai, J. and Basso, B., 2017. The International Heat Stress Genotype Experiment for modeling wheat response to heat: field experiments and AgMIP-Wheat multi-model simulations.
33. Epstein, E., and Bloom, A.J., 2005. Mineral nutrition of plants: principles and perspectives, 2nd eds. *Sunderland, Mass Sinaur*, pp.292-305.
34. Ferner, J., Linstädter, A., Rogass, C., Südekum, K.H. and Schmidtlein, S., 2021. Towards Forage Resource Monitoring in subtropical Savanna Grasslands: going multispectral or hyperspectral? *European Journal of Remote Sensing*, *54*(1), pp.364-384.
35. Fewster, R.M., Laake, J.L. and Buckland, S.T., 2005. Line transect sampling in small and large regions. *Biometrics*, *61*(3), pp.856-859.
36. Gamon, J.A., Field, C.B., Goulden, M.L., Griffin, K.L., Hartley, A.E., Joel, G., Penuelas, J. and Valentini, R., 1995. Relationships between NDVI, canopy structure, and photosynthesis in three Californian vegetation types. *Ecological applications*, *5*(1), pp.28-41.
37. Gao, J., Meng, B., Liang, T., Feng, Q., Ge, J., Yin, J., Wu, C., Cui, X., Hou, M., Liu, J. and Xie, H., 2019. Modeling alpine grassland forage phosphorus based on hyperspectral remote sensing and a multi-factor machine learning algorithm in the east of Tibetan Plateau, China. *ISPRS Journal of*
38. Gao, L., Zhang, C., Yun, W., Ji, W., Ma, J., Wang, H., Li, C. and Zhu, D., 2022. Mapping crop residue cover using Adjust Normalized Difference Residue Index based on Sentinel-2 MSI data. *Soil and Tillage Research*, *220*, p.105374.
39. Genuer, R., Poggi, J.M. and Tuleau-Malot, C., 2015. VSURF: an R package for variable selection using random forests. *The R Journal*, *7*(2), pp.19-33.

40. Giardina, C.P. and Rhoades, C.C., 2001. Clear cutting and burning affect nitrogen supply, phosphorus fractions and seedling growth in soils from a Wyoming lodgepole pine forest. *Forest Ecology and Management*, *140*(1), pp.19-28.
41. Gitelson, A.A. and Merzlyak, M.N., 1996. Signature analysis of leaf reflectance spectra: algorithm development for remote sensing of chlorophyll. *Journal of plant physiology*, *148*(3-4), pp.494-500.
42. Glen, A.S., 2008. Population attributes of the spotted-tailed quoll (Dasyurus maculatus) in north-eastern New South Wales. *Australian Journal of Zoology*, *56*(2), pp.137-142.
43. Goetz, A.F., 2009. Three decades of hyperspectral remote sensing of the Earth: A personal view. *Remote sensing of environment*, *113*, pp. S5-S16.
44. Milton, E.J., Schaepman, M.E., Anderson, K., Kneubühler, M. and Fox, N., 2009.
45. Gökkaya, K., Thomas, V., Noland, T.L., McCaughey, H., Morrison, I. and Treitz, P., 2015. Prediction of macronutrients at the canopy level using spaceborne imaging spectroscopy and LiDAR data in a mixedwood boreal forest. *Remote Sensing*, *7*(7), pp.9045-9069.
46. Golay, M.J.E., 1964. Normalised equations of the regenerative oscillator—Noise, phase-locking, and pulling. *Proceedings of the IEEE*, *52*(11), pp.1311-1330.
47. Grant, C.C., Peel, M.J.S. and Van Ryssen, J.B.J., 2000. Nitrogen and phosphorus concentration in faeces: an indicator of range quality as a practical adjunct to existing range evaluation methods. *African Journal of Range and Forage Science*, *17*(1-3), pp.81-92.
48. Güsewell, S., Koerselman, W. and Verhoeven, J.T., 2003. Biomass N: P ratios as indicators of nutrient limitation for plant populations in wetlands. *Ecological Applications*, *13*(2), pp.372-384.
49. Güsewell, S., 2004. N: P ratios in terrestrial plants: variation and functional significance. *New phytologist*, *164*(2), pp.243-266.
50. Guyot, G., 1988. Evolution of research orientation in the domain of spectral signatures.

51. Guanter, L., Segl, K., Kaufmann, H. (2009): Simulation of Optical Remote-Sensing ScenesWith Application to the EnMAP Hyperspectral Mission. - IEEE Transactions on Geoscienceand Remote Sensing.
52. Hadi, H., 2015. Multivariate statistical analysis for estimating grassland leaf area index and chlorophyll content using hyperspectral data (Master's thesis, University of Twente).
53. Hammad, A. and Tumeizi, A., 2012. Land degradation: socioeconomic and environmental causes and consequences in the eastern Mediterranean. *Land Degradation & Development*, *23*(3), pp.216-226.
54. Horler, D.N.H., Dockray, M., Barber, J. and Barringer, A.R., 1983. Red edge measurements for remotely sensing plant chlorophyll content. *Advances in Space Research*, *3*(2), pp.273-277.
55. Huber, S., Kneubühler, M., Psomas, A., Itten, K. and Zimmermann, N.E., 2008. Estimating foliar biochemistry from hyperspectral data in mixed forest canopy. *Forest ecology and management*, *256*(3), pp.491-501.
56. Huete, A.R., Hua, G., Qi, J., Chehbouni, A. and Van Leeuwen, W.J.D., 1992. Normalisation of multidirectional red and NIR reflectances with the SAVI. *Remote Sensing of Environment*, *41*(2-3), pp.143-154.
57. Hughes, G., 1968. On the mean accuracy of statistical pattern recognizers. *IEEE transactions on information theory*, *14*(1), pp.55-63.
58. Jiang, F., Kutia, M., Sarkissian, A.J., Lin, H., Long, J., Sun, H. and Wang, G., 2020. Estimating the growing stem volume of coniferous plantations based on random forest using an optimized variable selection method. *Sensors*, *20*(24), p.7248.
59. Jordan, C.F., 1969. Derivation of leaf-area index from quality of light on the forest floor. *Ecology*, *50*(4), pp.663-666.
60. Karimi, M. and Moradi, K., 2018. The response of Stevia (Stevia rebaudiana Bertoni) to nitrogen supply under greenhouse conditions. *Journal of Plant Nutrition*, *41*(13), pp.1695-1704.
61. Klodd, A.E., Nippert, J.B., Ratajczak, Z., Waring, H., and Phoenix, G.K., 2016. Tight coupling of leaf area index to canopy nitrogen and phosphorus

across heterogeneous tallgrass prairie communities. *Oecologia*, *182*, pp.889-898

62. Kganyago, M.L., 2015. *An evaluation of hyperspectral and multispectral data for mapping invasive species in an African Savanna* (Masters dissertation).
63. Knox, A.S., Kaplan, D.I. and Paller, M.H., 2006. Phosphate sources and their suitability for remediation of contaminated soils. *Science of the Total Environment*, *357*(1-3), pp.271-279.
64. Koerselman, W. and Meuleman, AFM, 1995. The N:P ratio: a simple tool for ecological management of water catchment areas. *H2O*, *28*, pp.94-7.
65. Kohavi, R. and John, G.H., 1995. Automatic parameter selection by minimising estimated error. In *Machine Learning Proceedings 1995* (pp. 304-312). Morgan Kaufmann.
66. https://www.krugerpark.co.za/Maps_of_Kruger_Park-travel/kruger-geo-map.html
67. Williams, C.K., Engelhardt, A., Cooper, T., Mayer, Z., Ziem, A., Scrucca, L., Tang, Y., Candan, C., Hunt, T. and Kuhn, M.M., 2015. Package 'caret'.
68. LeBauer, D.S. and Treseder, K.K., 2008. Nitrogen limitation of net primary productivity in terrestrial ecosystems is globally distributed. *Ecology*, *89*(2), pp.371-379.
69. Li, Z.L., Tang, H., Li, Z.L., and Tang, H., 2014. Estimation and validation of evapotranspiration from thermal infrared remote sensing data. *Quantitative Remote Sensing in Thermal Infrared: Theory and Applications*, pp.145-201.
70. Liang, L., Di, L., Huang, T., Wang, J., Lin, L., Wang, L. and Yang, M., 2018. Estimation of leaf nitrogen content in wheat using new hyperspectral indices and a random forest regression algorithm. *Remote Sensing*, *10*(12), p.1940.
71. Liu, J., Xu, X. and Shao, Q., 2008. Grassland degradation in the "three-river headwaters" region, Qinghai province. *Journal of Geographical Sciences*, *18*, pp.259-273.
72. Loozen, Y., Karssenberg, D., de Jong, S.M., Wang, S., van Dijk, J., Wassen, M.J. and Rebel, K.T., 2019. Exploring the use of vegetation indices

to sense canopy nitrogen to phosphorus ratio in grasses. *International Journal of Applied Earth Observation and Geoinformation*, *75*, pp.1-14.

73. Lü, X.T., Reed, S.C., Yu, Q., and Han, X.G., 2016. Nutrient resorption helps drive intra-specific coupling of foliar nitrogen and phosphorus under nutrient-enriched conditions. *Plant and soil*, *398*, pp.111-120.
74. Ludwig, D., Mangel, M. and Haddad, B., 2001. Ecology, conservation, and public policy. *Annual review of ecology and systematics*, *32*(1), pp.481-517.
75. Ludwig, F., De Kroon, H. and Prins, H.H., 2008. Impacts of savanna trees on forage quality for a large African herbivore. *Oecologia*, *155*, pp.487-496.
76. Main, R., Cho, M.A., Mathieu, R., O'Kennedy, M.M., Ramoelo, A. and Koch, S., 2011. An investigation into robust spectral indices for leaf chlorophyll estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*(6), pp.751-761.
77. Milton, E.J., Schaepman, M.E., Anderson, K., Kneubühler, M. and Fox, N., 2009. Progress in field spectroscopy. *Remote Sensing of Environment*, *113*, pp. S92-S109.
78. Musina, L., Rutherford, M.C., Powrie, L.W., van Niekerk, A. and van der Merwe, J.H., 2006. Strelitzia.
79. Grant, R., 2006. *Sappi Tree Spotting: Lowveld, Including Kruger National Park*. Jacana Media.
80. Mutanga, O. and Kumar, L., 2007. Estimating and mapping grass phosphorus concentration in an African savanna using hyperspectral image data. *International Journal of Remote Sensing*, *28*(21), pp.4897-4911.
81. Mutanga, O., Ismail, R., Ahmed, F. and Kumar, L., 2007, November. Using in situ hyperspectral remote sensing to discriminate pest attacked pine forests in South Africa. In *Proceedings of the 28th Asian Conference on Remote Sensing, Kuala Lumpur, Malaysia* (pp. 12-16).
82. Mutanga, O. and Skidmore, A.K., 2004. Narrow band vegetation indices overcome the saturation problem in biomass estimation. *International journal of remote sensing*, *25*(19), pp.3999-4014.
83. Mutanga, O., Skidmore, A.K. and 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*(3), pp.393-408

84. Skidmore, A.K., Ferwerda, J.G., Mutanga, O., Van Wieren, S.E., Peel, M., Grant, R.C., Prins, H.H., Balcik, F.B. and Venus, V., 2010. Forage quality of savannas—Simultaneously mapping foliar protein and polyphenols for trees and grass using hyperspectral imagery. *Remote sensing of environment*, *114*(1), pp.64-72.
85. Mutanga, O., Skidmore, A.K. and Van Wieren, S., 2003. Discriminating tropical grass (Cenchrus ciliaris) canopies grown under different nitrogen treatments using spectroradiometry. *ISPRS Journal of Photogrammetry and Remote Sensing*, *57*(4), pp.263-272.
86. Mutanga, O., Prins, H.H., Skidmore, A.K., Van Wieren, S., Huizing, H., Grant, R., Peel, M. and Biggs, H., 2004. Explaining grass-nutrient patterns in a savanna rangeland of southern Africa. *Journal of biogeography*, *31*(5), pp.819-829.
87. Mutanga, O., Adam, E. and Cho, M.A., 2012. High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. *International Journal of Applied Earth Observation and Geoinformation*, *18*, pp.399-406.
88. Naidoo, L., Cho, M.A., Mathieu, R. and Asner, G., 2012. Classification of savanna tree species, in the Greater Kruger National Park region, by integrating hyperspectral and LiDAR data in a Random Forest data mining environment. *ISPRS journal of Photogrammetry and Remote Sensing*, *69*, pp.167-179.
89. Olde Venterink, H., Cech, P.G., Kuster, T., and Edwards, P.J., 2008. Effects of herbivory, fire, and N 2-fixation on nutrient limitation in a humid African savanna. *Ecosystems*, *11*, pp.991-1004.
90. Oumar, Z. and Mutanga, O., 2013. Using WorldView-2 bands and indices to predict bronze bug (Thaumastocoris peregrinus) damage in plantation forests. *International Journal of Remote Sensing*, *34*(6), pp.2236-2249.

91. Pacheco-Labrador, J., Ferrero, A., and Martín, M.P., 2014. Characterising integration time and gray-level-related nonlinearities in a NMOS sensor. *Applied Optics*, *53*(32), pp.7778-7786.
92. Palmer, A.R. and Ainslie, A.M., 2005. Grasslands of South Africa. *Grasslands of the World*, *34*, p.77.
93. Peddle, D.R., White, H.P., Soffer, R.J., Miller, J.R. and Ledrew, E.F., 2001. Reflectance processing of remote sensing spectroradiometer data. *Computers & geosciences*, *27*(2), pp.203-213.
94. Schott, J.R., Brown, S.D., Raqueno, R.V., Gross, H.N. and Robinson, G., 1999. An advanced synthetic image generation model and its application to multi/hyperspectral algorithm development. *Canadian Journal of Remote Sensing*, *25*(2), pp.99-111.
95. Somers, B., Tits, L. and Coppin, P., 2013. Quantifying nonlinear spectral mixing in vegetated areas: Computer simulation model validation and first results. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(6), pp.1956-1965.
96. Schwartz, M.W., Iverson, L.R., Prasad, A.M., Matthews, S.N. and O'Connor, R.J., 2006. Predicting extinctions as a result of climate change. *Ecology*, *87*(7), pp.1611-1615.
97. Berry, P.M., Dawson, T.P., Harrison, P.A., Pearson, R. and Butt, N., 2003. The sensitivity and vulnerability of terrestrial habitats and species in Britain and Ireland to climate change. *Journal for Nature Conservation*, *11*(1), pp.15-23.
98. Prins, H.H. and Van Langevelde, F. eds., 2008. *Resource ecology: spatial and temporal dynamics of foraging* (Vol. 23). Springer Science & Business Media.
99. 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), p.269.
100. Ramoelo, A., Skidmore, A.K., Cho, M.A., Mathieu, R., Heitkönig, I.M.A., Dudeni-Tlhone, N., Schlerf, M., and Prins, H.H.T., 2013. Non-linear partial least square regression increases the estimation accuracy of grass

nitrogen and phosphorus using in situ hyperspectral and environmental data. *ISPRS journal of photogrammetry and remote sensing*, *82*, pp.27-40.

101. Ramoelo, A., Stolter, C., Joubert, D., Cho, M.A., Groengroeft, A., Madibela, O.R., Zimmermann, I. and Pringle, H., 2018. Rangeland monitoring and assessment: a review.

102. Ramoelo, A., Skidmore, A.K., Schlerf, M., Mathieu, R. and Heitkönig, I.M., 2011. Water-removed spectra increase the retrieval accuracy when estimating savanna grass nitrogen and phosphorus concentrations. *ISPRS journal of photogrammetry and remote sensing*, *66*(4), pp.408-417.

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

104. Ramoelo, A., Cho, M.A., Mathieu, R.S., Skidmore, A.K., Schlerf, M. and Heitkönig, I.M.A., 2012. Estimating grass nutrients and biomass as an indicator of rangeland (forage) quality and quantity using remote sensing in Savanna ecosystems.

105. Ramoelo, A., Cho, M., Mathieu, R., and Skidmore, A.K., 2014, November. The potential of Sentinel-2 spectral configuration to assess rangeland quality. In *Remote Sensing for Agriculture, Ecosystems, and Hydrology XVI* (Vol. 9239, pp. 108-118). SPIE.

106. Ramoelo, A., Cho, M.A., Mathieu, R., Madonsela, S., Van De Kerchove, R., Kaszta, Z. and Wolff, E., 2015. Monitoring grass nutrients and biomass as indicators of rangeland quality and quantity using random forest modelling and WorldView-2 data. *International journal of applied earth observation and geoinformation*, *43*, pp.43-54.

107. Read, J.M., Torrado, m., 2009 in International Encyclopedia of Human Geography

108. Reed, S.C., Cleveland, C.C., and Townsend, A.R., 2011. Functional ecology of free-living nitrogen fixation: a contemporary perspective. *Annual review of ecology, evolution, and systematics*, *42*, pp.489-512.

109. Reinermann, S., Asam, S. and Kuenzer, C., 2020. Remote sensing of grassland production and management. A review. *Remote Sensing*, *12*(12), p.1949.

110. Rieske, L.K., 2002. Wildfire alters oak growth, foliar chemistry, and herbivory. *Forest Ecology and Management*, *168*(1-3), pp.91-99.

111. Rouse Jr, J.W., Haas, R.H., Deering, D.W., Schell, J.A. and Harlan, J.C., 1974. *Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation* (No. E75-10354).

112. Rossel, R.V. and McBratney, A.B., 2008. Diffuse reflectance spectroscopy as a tool for digital soil mapping. Digital soil mapping with limited data, pp.165-172.

113. Salih, A.A., Ganawa, E.T. and Elmahl, A.A., 2017. Spectral mixture analysis (SMA) and change vector analysis (CVA) methods for monitoring and mapping land degradation/desertification in arid and semiarid areas (Sudan), using Landsat imagery. *The Egyptian Journal of Remote Sensing and Space Science*, *20*, pp. S21-S29.

114. Sankaran, M., Augustine, D.J. and Ratnam, J., 2013. Native ungulates of diverse body sizes collectively regulate long-term woody plant demography and structure of a semi-arid savanna. *Journal of Ecology*, *101*(6), pp.1389-1399.

115. Shoko, C., Mutanga, O. and Dube, T., 2016. Progress in the remote sensing of C3 and C4 grass species aboveground biomass over time and space. *ISPRS Journal of Photogrammetry and Remote Sensing*, *120*, pp.13-24.

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

117. Singh, H.V., Kumar, S., Roy, A.K. and Singh, K.A., 2017. Growth and biomass production of fodder trees and grasses in a silvipasture system on non-arable land of semi-arid India. *Range Management and Agroforestry*, *38*(1), pp.43-47

118. Skidmore, A.K., Ferwerda, J.G., Mutanga, O., Van Wieren, S.E., Peel, M., Grant, R.C., Prins, H.H., Balcik, F.B. and Venus, V., 2010. Forage quality of savannas—Simultaneously mapping foliar protein and polyphenols for trees and grass using hyperspectral imagery. *Remote sensing of environment*, *114*(1), pp.64-72.

119. Srinet, R., Nandy, S., and Patel, N.R., 2019. Estimating leaf area index and light extinction coefficient using Random Forest regression algorithm in a tropical moist deciduous forest, India. *Ecological Informatics*, *52*, pp.94-102.

120. Symeonakis, E., Petroulaki, K. and Higginbottom, T., 2016. Landsat-based woody vegetation cover monitoring in southern African savannahs. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *41*, pp.563-567

121. Usha, K. and Singh, B., 2013. Potential applications of remote sensing in horticulture—A review. *Scientia horticulturae*, *153*, pp.71-83.

122. Venter, F., Randall, R., Hanekom, N. and Ruselli, I., 2003. RIVER, COASTAL. In South African National Parks: A Celebration Commemorating the Fifth World Parks Congress 2003 (p. 122). Horst Klemm Publications.

123. Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J.P., Veroustraete, F., Clevers, J.G. and Moreno, J., 2015. Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties–A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *108*, pp.273-290.

124. Verrelst, J., Schaepman, M.E., Malenovský, Z. and Clevers, J.G., 2010. Effects of woody elements on simulated canopy reflectance: Implications for forest chlorophyll content retrieval. *Remote Sensing of Environment*, *114*(3), pp.647-656.Verrelst,2010)

125. Vitousek, P.M. and Howarth, R.W., 1991. Nitrogen limitation on land and in the sea: how can it occur? *Biogeochemistry*, *13*, pp.87-115.

126. Vlek, P.L., Le, Q.B. and Tamene, L., 2010. Assessment of land degradation, it possible causes and threat to food security in Sub-Saharan

Africa. *Food Security and Soil Quality. CRC Press, Boca Raton, Florida*, pp.57-86.

127. Walker, A.P., Beckerman, A.P., Gu, L., Kattge, J., Cernusak, L.A., Domingues, T.F., Scales, J.C., Wohlfahrt, G., Wullschleger, S.D. and Woodward, F.I., 2014. The relationship of leaf photosynthetic traits–Vcmax and Jmax–to leaf nitrogen, leaf phosphorus, and specific leaf area: a meta-analysis and modeling study. *Ecology and evolution*, *4*(16), pp.3218-3235.

128. Wang, Z., Zhao, C., Yu, L., Zhou, W. and Li, K., 2009. Regional metabolic changes in the hippocampus and posterior cingulate area detected with 3-Tesla magnetic resonance spectroscopy in patients with mild cognitive impairment and Alzheimer disease. *Acta Radiologica*, *50*(3), pp.312-319.

129. Wessels, K.J., Mathieu, R., Erasmus, B.F.N., Asner, G.P., Smit, I.P.J., Van Aardt, J.A.N., Main, R., Fisher, J., Marais, W., Kennedy-Bowdoin, T., and Knapp, D.E., 2011. Impact of communal land use and conservation on woody vegetation structure in the Lowveld savannas of South Africa. *Forest Ecology and Management*, *261*(1), pp.19-29.

130. Wold, S., Trygg, J., Berglund, A. and Antti, H., 2001. Some recent developments in PLS modeling. *Chemometrics and intelligent laboratory systems*, *58*(2), pp.131-150.

131. Wright, A.L., Hons, F.M. and Rouquette Jr, F.M., 2004. Long-term management impacts on soil carbon and nitrogen dynamics of grazed bermudagrass pastures. *Soil Biology and Biochemistry*, *36*(11), pp.1809-1816.

132. Zhang, L., Zhang, L., Tao, D., Huang, X. and Du, B., 2015. Compression of hyperspectral remote sensing images by tensor approach. *Neurocomputing*, *147*, pp.358-363.