Mapping key fibre biochemical concentrations in KwaZulu-Natal grasslands using remotely sensed technologies

By

LEETH SINGH

210527571

Submitted in fulfilment of the academic requirements for the degree of Master of Science in the School of Agriculture, Earth and Environmental Sciences,

In the Discipline of Geography,

College of Agriculture, Engineering and Science

University of KwaZulu-Natal

Pietermaritzburg

South Africa

January 2017

DECLARATION 1

The research contained in this thesis was completed by the candidate while based in the Discipline of Geography, School of Agricultural, Earth and Environmental Sciences of the College of Agriculture, Engineering and Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa. The research was financially supported by the National Research Foundation (NRF).

This study was undertaken in fulfilment of a Master of Science Degree and represents the original work of the author. Where use has been made of the work of others, it is duly acknowledged in both the text and reference section of the dissertation.

……...……
Signed: Professor Onisimo Mutanga
Date: January 2017

……...……
Signed: Professor Paramu Mafongoya
Date: January 2017

DECLARATION 2: PLAGIARISM

I, Leeth Singh, declare that:

(i) the research reported in this dissertation, except where otherwise indicated or acknowledged, is my original work;

(ii) this dissertation has not been submitted in full or in part for any degree or examination to any other university;

(iii) this dissertation does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons;

(iv) this dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

a) their words have been re-written but the general information attributed to them has been referenced;

b) where their exact words have been used, their writing has been placed inside quotation marks, and referenced;

(v) where I have used material for which publications followed, I have indicated in detail my role in the work;

(vi) this dissertation is primarily a collection of material, prepared by myself, published as journal articles or presented as a poster and oral presentations at conferences. In some cases, additional material has been included;

(vii) this dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the References sections.

……..

Signed: Leeth Singh

Date: January 2017

DECLARATION 3: PUBLICATIONS

DETAILS OF CONTRIBUTION TO PUBLICATIONS that form part of and/or include the research presented in this thesis (includes publications in preparation and those that have been submitted, are in press or are published, and gives details of the contributions of each author to the experimental work and writing of each publication).

Publication 1: Singh, L.[1], Mutanga, O.[2], Mafongoya, P.[3] and Peerbhay, K. Y.[4] (in preparation). Remote sensing of key grassland nutrients using hyperspectral techniques in KwaZulu-Natal, South Africa.

Publication 2: Singh, L.[1], Mutanga, O.[2], Mafongoya, P.[3] and Peerbhay, K. Y.[4] (in preparation). Multispectral mapping of key grassland nutrients in KwaZulu-Natal, South Africa.

The work was done by the first author under the guidance and supervision of the second, third and fourth authors.

1, 2, 3, 4 School of Agriculture, Earth and Environmental Sciences, Faculty of Science and Agriculture, University of KwaZulu-Natal, Pietermaritzburg Campus, Private Bag X01, Scottsville, 3209, South Africa.

……..

Signed: Leeth Singh

Date: January 2017

ABSTRACT

Forage quality nutrient assessments are costly, labour intensive, time consuming, dangerous and usually results in spatially invariable data. These assessments are most commonly useful to policy makers, rangeland managers and farmers to obtain information relating to the concentrations of forage quality nutrients in a grassland ecosystem. Remote sensing offers a cost and time-effective alternative for obtaining accurate information relating to the concentrations of forage quality nutrients over local to broad scales. Hence, the aim of this research was to detect three forage fibre nutrients in KwaZulu-Natal, South Africa, using remotely sensed in-field hyperspectral data and satellite multispectral image data, in conjunction with the Random forest algorithm (RF).

The first part of this study examined the effectiveness of known absorption features for detecting forage fibre nutrients, neutral detergent fibre (NDF), acid detergent fibre (ADF) and Lignin using hyperspectral data. Results indicate successful correlations between the known absorption features and forage quality nutrients NDF, ADF and Lignin with coefficients (R^2) ranging between 0.57 and 0.81 using RF. In comparison, using the entire hyperspectral dataset, the study identified additional wavebands which contributes to the accurate detection of the forage quality nutrients in a grassland environment. Overall, RF was capable reducing problems of dimensionality and multicollinearity within the large contiguous wavebands of the hyperspectral dataset, providing an ideal framework to detect forage quality nutrients.

The second part of this study used high resolution RapidEye-5 multispectral image data to detect and map forage quality nutrients. More specifically, this study utilized the advanced sensor configurations of the RapidEye-5 image data to detect forage quality nutrients (NDF, ADF and Lignin) in conjunction with ancillary data and the Random forest algorithm. Results showed that the RF algorithm successfully mapped NDF, ADF and Lignin coefficients (R^2) ranging between 0.67 and 0.74. In comparison, the study utilized Stochastic Gradient Boosting (SGB) algorithm as an alternative modelling technique, which produced very similar results ranging between 0.65 and 0.72 (R^2). Overall, multispectral remote sensing in conjunction with a algorithm and ancillary data can detect and map forage quality nutrients NDF, ADF and Lignin in a grassland environment.

Overall, the results from this study indicate that (i) remote sensing can produce timely and accurate models for detecting and mapping forage quality nutrients, (ii) RF is an effective method for waveband reduction and the accurate discrimination of high dimensional hyperspectral datasets, (iii) the inclusion of ancillary data should always be considered.

ACKNOWLEDGMENTS

I would like to express my gratitude to the following persons and institutions for the support and knowledge that I received from them during my postgraduate research:

- First and foremost, I would like to thank God for making it all possible. My faith in him to guide me is evident in the entirety of this research.
- National Research Foundation (NRF) for funding and to Fort Nottingham Nature Reserve Manager Roy Tabernor, for providing me with access and field support during sampling.
- The impetus to carry out my research could not have been possible without the motivation and support of my supervisors, Professor Onisimo Mutanga and Professor Paramu Mafongoya. I am very grateful to you for your valuable guidance and input as my supervisors and mentors.
- A very special thank you to Dr Kabir Peerbhay for your steadfast guidance throughout my research. I am indebted to you for all the time and effort you put into helping me.
- I extend my gratitude to my dear parents for their unconditional love and support during my entire academic career.
- Finally, the Department of Geography at the University of KwaZulu-Natal supported me through my academic career. I would like to thank my mentor and friend, Dr Romano Lottering, for providing me with valuable support and advice.

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

LIST OF TABLES

Table **Page**

Chapter One

Introduction

1.1 Background

Accurate determination of forage quality nutrients within grasslands is important to scientific research as well as for effective rangeland management practices. Conventionally, the quality of forage is assessed and monitored using wet chemistry and Near infrared spectroscopy (NIRS). Here, many studies have used wet chemistry and NIRS to determine forage quality (Curran, 1989; Curran *et al.*, 2001; Elvidge, 1990; Kokaly & Clark, 1999; Norris *et al.*, 1976; Osborne & Fearn, 1986; Starks *et al.*, 2004; Stuth *et al.*, 2003). However, these assessments are labour intensive, costly, time consuming, dangerous and usually spatially invariable (Pullanagari *et al.*, 2013a). Remote sensing is a time-effective and cost-effective alternative for assessing and monitoring forage quality over a large area thereby providing valuable information on both local and global scales. More specifically, remote sensing can provide this information to policy makers, rangeland managers and farmers at a national level.

To date, hyperspectral data has been used in many studies to determine the quality of forage. Typically, several studies have used hyperspectral data for forage quality assessments (Albayrak, 2008; Kawamura *et al.*, 2008; Kawamura *et al.*, 2010; Knox *et al.*, 2011; Kokaly & Clark, 1999; Pullanagari *et al.*, 2013b; Zhao *et al.*, 2007). The use of high resolution spectral and spatial sensors contain a large number of contiguous wavebands that make it possible to determine discrete biochemical information with high precision (Schellberg *et al.*, 2008). Research has shown that hyperspectral remote sensing systems are capable of determining forage quality (Knox *et al.*, 2011; Pullanagari *et al.*, 2012; Zhao *et al.*, 2007). However, infield-hyperspectral studies are still point-based and require a large amount of time for processing a large amount of data.

A few studies have used multispectral data for the determination of forage quality (neutral detergent fibre (NDF), acid detergent fibre (ADF) and Lignin). Hence, researchers acknowledged that multispectral wavebands are too broad and generalized to determine forage quality, coupled with low spectral and spatial resolutions. Nonetheless, technological advancements have provided more specialized multispectral platforms capable of analyzing more detailed vegetation properties. The latest inventions have advanced spatial, spectral, radiometric and temporal resolutions capable of detecting more discernable information

(Mutanga *et al.*, 2016). New generation multispectral platforms are configured with the new Red-edge waveband, capable of potentially improving classification accuracies. Previous research suggests that a combination of high resolution and strategically placed wavebands, make it possible to accurately detect forage quality (Eitel *et al.*, 2007). Hence, the selection of most appropriate multispectral platform for the correct application is imperative to attaining accurate results.

New generation RapidEye-5 multispectral satellite is configured with a Red-edge waveband that has the potential to considerably improve classification accuracies within most vegetation applications (Adelabu *et al.*, 2014). RapidEye-5 imagery promises to advance the mapping of vegetation over a large area, at high resolution, at faster time (5 days), and with a lower cost (RapidEye, 2011). Researchers often use statistical approaches to improve the detection and prediction accuracies of forage quality. Hence, there are various statistical packages available, some which are popular in forage quality applications such as: partial least squares, neural networks, and random forest (Mutanga *et al.*, 2016). Furthermore, including ancillary data can improve the prediction accuracy of forage quality nutrients. For example, Knox *et al*., 2011 found phenology, soil type, species, and geology as the best input ancillary data variables that are related to variations in nutrient levels in the dry season, hence these variables were included in the overall prediction model (Knox *et al.*, 2011).

The first paper in this thesis assess the utility of hyperspectral remote sensing in determining key forage fibre grassland nutrients in Fort Nottingham Nature Reserve, KwaZulu-Natal, South Africa. This paper focuses on the effectiveness of using known absorption features to detect forage fibre nutrients (neutral detergent fibre (NDF), acid detergent fibre (ADF) and Lignin) using hyperspectral data and the random forest algorithm. Hyperspectral data can provide detailed information for the detection of forage fibre nutrients, however the technology is costly to small scale farmers, and time consuming. Problems of multicollinearity and dimensionality associated with hyperspectral data exist, providing a difficult task during analysis. Therefore, the second paper focuses on the effectiveness of RapidEye-5 multispectral imagery to detect and map forage fibre nutrients (NDF, ADF and Lignin) using a specialized algorithm and ancillary data. The RapidEye-5 multispectral sensor is a 5-waveband system including the red-edge waveband that promises to be an effective tool for mapping vegetation properties. Finally, the results from the hyperspectral analysis are compared with the results obtained from the RapidEye-5 image data. It was interesting to investigate the performance of both hyperspectral and multispectral remote sensing technologies in detecting and predicting forage fibre nutrients.

1.2 Aims and objectives

The aim of this research was to assess the utility of hyperspectral and multispectral remote sensing in predicting key forage quality nutrients. The main objectives were as follows:

- To assess the capability of hyperspectral remote sensing in detecting three forage fibre quality nutrients (NDF, ADF and Lignin) using Random Forest.
- To evaluate the effectiveness of using known absorption features for selection of the most optimal subsets of hyperspectral wavebands.
- To investigate the capability of high resolution RapidEye-5 multispectral imagery to detect and predict forage quality nutrients (NDF, ADF and Lignin) using a specialized algorithm and ancillary data.
- To test the capability of the Stochastic Gradient Boosting algorithms in identifying the most important wavebands for detecting forage quality nutrients NDF, ADF and Lignin.
- To compare the respective capabilities of in-field hyperspectral and multispectral remotely sensed data to detect and predict key forage quality nutrients.

1.3 Outline of thesis

This thesis consists of four chapters. The main structure of this thesis is within two core chapters (Chapter Two and Three), which form publishable papers and will be submitted to peer-reviewed journals. Since both these chapters have detailed sections covering the study area, literature review and methodology, these sections are not covered with in the introductory section of the thesis to avoid repetition.

Chapter Two assesses the capability of hyperspectral remote sensing to detect forage fibre nutrients NDF, ADF and Lignin using Random forest. This chapter primarily focuses on using known absorption features to detect these forage fibre nutrients and for the selection of the most optimal hyperspectral wavebands. The Random forest algorithm was used to analyze the large amount of hyperspectral contiguous wavebands (n = 1910), for variable selection and to produce the final prediction model.

Chapter Three evaluates the effectiveness of RapidEye-5 multispectral image data (n = 5) to detect and predict forage quality nutrients (NDF, ADF and Lignin) using a specialized algorithm and ancillary data. The Stochastic gradient boosting algorithm was tested to identify

the most important wavebands used for the detection of NDF, ADF and Lignin. The ancillary data (Leaf Area Index, Chlorophyll, species type, species Count) was used in the main dataset.

Chapter Four provides an overview of the study. Both aims and objectives are discussed in depth, and the most important findings are highlighted. A discussion of the most appropriate remote sensing approach for detecting and predicting forage quality nutrients is presented in this chapter. Lastly, the chapter examines the limitations and presents the recommendations for future research within this studies context.

Chapter Two

Remote sensing of key grassland nutrients using hyperspectral techniques in KwaZulu-Natal, South Africa

2.1 Abstract

The concentration of forage fibre content is critical in explaining the palatability of forage quality for livestock grazers in tropical grasslands. Low grass quality negatively impacts on the health of livestock, creating a chain of undesirable impacts on animal performance; biodiversity; food security and state economies. Traditional methods of determining forage fibre content are usually time consuming, costly and require specialized laboratory analysis. With the potential of remote sensing technologies and their application across broad spatial extents, determination of key fibre attributes can be made more accurately. This study aims to determine the effectiveness of known absorption wavelengths for detecting forage fibre biochemicals, neutral detergent fibre (NDF), acid detergent fibre (ADF) and Lignin using hyperspectral data. Hyperspectral reflectance spectral measurements (350-2500 nm) of grass were collected and implemented within the random forest (RF) ensemble. Results show successful correlations between the known absorption features and the biochemicals with coefficients (R^2) ranging from 0.57 to 0.81. In comparison, using the entire dataset, the study identified additional wavelengths which contributes to the accurate determination of forage quality in a grassland environment. Overall, the results showed that hyperspectral remote sensing in conjunction with the RF ensemble could discriminate each key biochemical evaluated. The study was successful in determining the effectiveness of using known absorption features for detecting fibre biochemicals NDF, ADF and Lignin using hyperspectral data. This study shows the potential to upscale the methodology to a space-borne multispectral platform with similar spectral configurations for an accurate and cost effective mapping analysis of forage quality.

Keywords: fibre biochemical, tropical grassland, NDF, ADF, Lignin

2.2 Introduction

Globally, grasslands comprise 26% of the total land cover; of which 80% is used for agricultural purposes and 68% are found in developing countries (Boval & Dixon, 2012). In South Africa, more than 70% of land surface comprises grassland vegetation which is mainly used for pastoral production (Mansour *et al.*, 2013). Here, grasslands provide the primary feed base for grazing livestock which in turn provides the majority of low-income people with food, goods and services that are essential for sustaining their livelihoods (Boval & Dixon, 2012). Forage quality in grasslands is therefore a vital component in the management of grazing lands and livestock (Georgiadis & McNaughton, 1990). However, inadequate veld management practices, result in grazing pastures being overgrazed and overstocked especially during the dry season (Muchenje *et al.*, 2008). It is evident in literature that pasture management is a major concern, particularly in tropical pasture systems where improper management can lead to unpalatable grasses. Unpalatable grasses have a ripple effect on the intake potential of livestock, later affecting livestock production levels and eventually food security in a country.

Traditionally, the productivity of rangelands has been assessed and monitored using field based point assessments which are spatially invariable, expensive, harmful and often time consuming (Dabasso *et al.*, 2012; Jančík *et al.*, 2011; Stuth *et al.*, 2003; Zhao *et al.*, 2007). These methods usually involve extensive laboratory chemical analysis for determining forage quality (Zhao *et al.*, 2007); and require personnel with specialized skills (Mutanga *et al.*, 2004a; Zhao *et al.*, 2007). Laboratory personnel are exposed to harmful chemicals and also, the hazardous waste produced from laboratory processes must be appropriately disposed of in order to reduce risk of environmental pollution (Zhao *et al.*, 2007). Furthermore, supervised laboratory experiments require a large amount of time for analysis that hinder immediate solutions and are not practical and logical for broad grassland ecosystems as is in many cases of grazing lands in South Africa (Mutanga *et al.*, 2004a; Stuth *et al.*, 2003). Nonetheless, this information becomes relevant when rangeland managers make informed decisions on the use of expensive dietary supplements to minimize grazing; nutrient deficiencies and when maintaining production levels (Mbatha & Ward, 2010). However, using conventional approaches over broad tropical grazing grasslands is challenging. To date, remote sensing is an efficient technology available for mapping land cover dynamics across broad geographic extents (Kuemmerle *et al.*, 2013) and may serve as a viable alternative.

Remote sensing offers possibilities to provide accurate information related to forage quality in grazing lands, thereby reducing the tedious procedure of rigorous sampling and laboratory analysis (Mutanga *et al.*, 2004b). The 1970s inventory of near infrared reflectance spectroscopy (NIRS) was used to assess forage quality and provided fairly accurate predictions of ADF, NDF and Lignin concentrations contained in dried ground forage (Kawamura *et al.*, 2008; Mutanga *et al.*, 2004b; Starks *et al.*, 2004; Stuth *et al.*, 2003; Zhao *et al.*, 2007). NIRS analysis provided a fairly cost effective and quick estimation of forage nutrient composition (Stuth *et al.*, 2003; Zhao *et al.*, 2007). In effect, many laboratories have replaced wet chemistry as the standard analytical procedure for assessing and estimating plant biochemicals (Mutanga *et al.*, 2004b). A study by Starks *et al*., (2004) successfully used in-field hyperspectral spectrometry for the prediction of nitrogen, NDF and ADF and compared the estimates obtained using NIRS and laboratory chemical methods. The study showed that forage composition estimates from the radiometer were equivalent to those from the NIRS. However, there are many challenges with the use of NIRS analysis. For example, when applying NIRS across different vegetation types, in conjunction with a multiple linear regression to predict canopy chemistry, it has yielded inconsistent results (Mutanga *et al.*, 2004b; Starks *et al.*, 2004; Stuth *et al.*, 2003). According to Mutanga *et al*., (2004b) problems of overfitting and spectral variability exist that are independent of the biochemical concentrations of forage. In addition, NIRS requires time for collecting, drying, and grinding vegetation samples. However, the advent of in-field hyperspectral remote sensing proved significant in successfully detecting and predicting forage quality biochemicals.

For example, Mutanga *et al*., (2004b) applied 'continuum removal on absorption features' to predict macronutrients (N; P; K; Ca and Mg) in tropical rangelands using in-field hyperspectral remote sensing. As a result, these absorption features accounted for 69% of the wavelengths selected using a stepwise regression and is an important step towards the remote sensing and mapping of rangelands (Mutanga *et al.*, 2004b). Similarly, Kawamura *et al*., (2010) successfully predicted pasture biomass and forage fibre quality variables NDF, ADF, Lignin in Hokkaido, Japan. Genetic algorithm - partial least squares (GA-PLS) successfully estimated pasture mass and quality parameters in a mixed grass pasture. Results of 77%, 76% and 62% were obtained for NDF, ADF and Lignin, respectively (Kawamura *et al.*, 2010). Zhao *et al*., (2007) successfully demonstrated the potential of hyperspectral remote sensing to estimate forage fibre quality variables; NDF, ADF and crude protein (CP) using canopy reflectance data of grass pastures in the wet season. The study produced accuracies of 52% and 20% for NDF and ADF, respectively through utilizing the MAXR regression algorithm, and found that forage

quality variables can be rapidly and non-destructively predicted using canopy reflectance data (Zhao *et al.*, 2007). Furthermore, a study by Pullanagari *et al*., (2012) successfully mapped forage quality variables using the PLS regression algorithm and yielded 75%, 82%, 71% accuracies for NDF, ADF and Lignin, respectively. The study stated that the information produced from using in-field hyperspectral remote sensing of forage quality could help pastoral farmers improve productivity and business resilience by enabling them to make more accurate and time effective decisions (Pullanagari *et al.*, 2012). In addition, Knox *et al*., (2011) successfully mapped forage quality variables (N; P and ADF) in the dry season on savanna grasslands using hyperspectral Carnegie Airborne Observatory sensor (CAO) in Kruger National Park, South Africa. The neural networks algorithm was used to generate forage quality maps and produced an R^2 of 62% for ADF. The frequency of fire appeared to impact nutrient levels when mapped on a g g−1 basis (Knox *et al.*, 2011). It is evident in literature that most studies use all wavebands as input in the model to predict biochemical "fishing expedition" and therefore results in selection of wavebands that are independent of the biochemical of concern. There are also problems of dimensionality and overfittings using many wavebands. A few studies focus on using known absorption features to detect forage quality nutrients.

In summary, research suggests the need for improved methods to evaluate the current status and the potential of grassland systems, to guide management decisions and yield better quality grasslands for livestock production (Ash *et al.*, 2011; Boval & Dixon, 2012; Fynn *et al.*, 2015; Kuemmerle *et al.*, 2013; Swanepoel *et al.*, 2015). Moreover, a small amount of rangeland managers utilizes remote sensing as a tool to inform their decision making. Whilst many studies focused on remote sensing of chlorophyll and nitrogen to predict pasture quality (Clevers & Gitelson, 2013; Clevers & Kooistra, 2012; Pellissier *et al.*, 2015; Ramoelo *et al.*, 2012a; Ullah *et al.*, 2012), not many focused on determining fibre content as the key limiting intake potential variable to ruminants. Within this context, this research aims to detect NDF, ADF and Lignin as some of the key limiting variables of forage quality and to determine their most important wavelengths using known absorption features on tropical mixed grass vegetation.

2.3 Materials and Methods

2.3.1 The study site

The study site is located in Fort Nottingham Nature Reserve, KwaZulu-Natal, South Africa (latitude 29° 24' 9.62" S and longitude 29° 53' 48.95" E) (Figure 1). The reserve occupies an area of 1227 ha and is characterized by an extensive area of the Eastern Mist belt Forest and Drakensberg Foothill Moist Grassland of the KwaZulu-Natal midlands. *Festuca costata* (blue fescue), *Tristachya leucothrix* (hairy trident grass) and *Themeda triandra* (red oat grass) are the dominant grass species occurring in the study area (Mucina & Rutherford, 2006). *Festuca costata* and *Tristachya leucothrix* are native to Southern Africa while *Themeda triandra* are native to Kenya. The study site consists of eleven cattle grazing camps with an annual spring and summer rainfall of approximately 950 mm per annum. The site topography is characterized by a diversity of habitat split between a moderately undulating Drakensberg Foothill Moist Grassland plateau (at approximately 1750m amsl) leading to densely forested south and east facing slopes that merge into lower lying remnant Drakensberg Foothill Moist Grassland flats. The most dominant soils found at the study area are: Clovelly; Hutton; Griffin and Oatsdale (Mucina & Rutherford, 2006). The study site is prone to wild fires during the late dry season that have a detrimental effect on the ecology of the grasslands and contributes to an altered vegetation species composition and abundance. Approximately 778.2 ha of the land is mainly used for cattle grazing that is currently leased from uMngeni Municipality. The study site was preferred due to heterogeneity of grass species and the importance of grazing in the local community's livelihood.

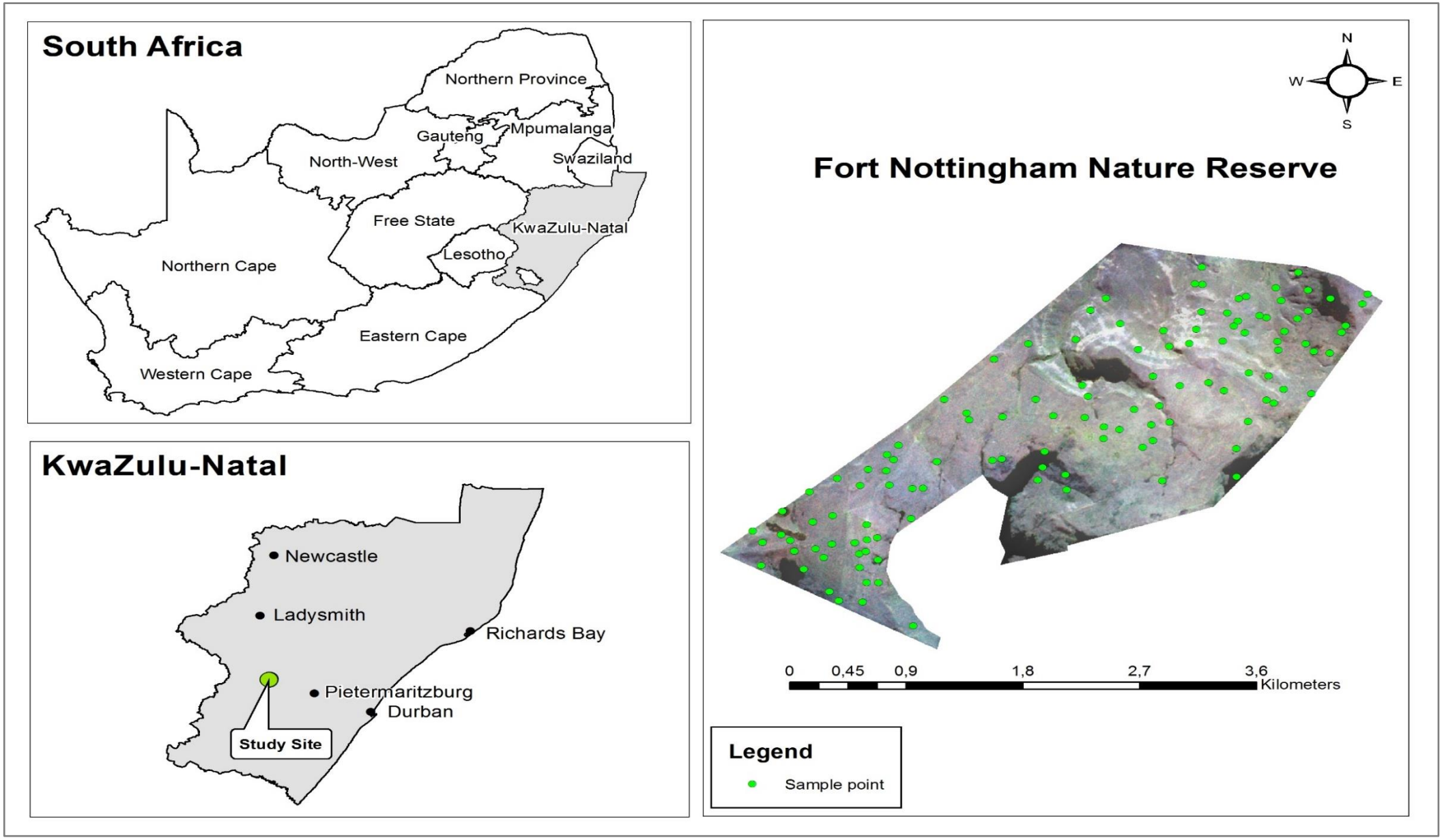


Figure 1. Location of the study site in Fort Nottingham Nature Reserve, KwaZulu-Natal, South Africa

2.3.2 Field sampling

A stratified random sampling method was used to select field sampling points using Hawths tool in ArcGIS 10.2. A total of 140 sample points were established across the eleven existing leased grazing camps. All sample points were uploaded into a Trimble GeoExplorer 6000 series GeoXH GPS with a centimeter-level accuracy of less than 10 cm and was used to locate each point. A 1 x 1m quadrant was overlaid at the center of each of the 140 points to declare subjectivity over the three existing grass species. Within each quadrant, two measurements were taken: a series of non-destructive spectral measurement using a spectrometer device and a destructive sample of the grass species present. The destructive sample was taken immediately after the spectral measurements were taken. Grass samples were clipped at 1 cm above ground, pooled and bagged (brown bag) for drying and later chemical analysis (Knox *et al.*, 2011).

2.3.3 Chemical analysis

Samples collected (n = 140) were transported to the Department of Animal Science laboratory at the University of KwaZulu-Natal for chemical analysis. All samples were oven dried at 65 degrees Celsius for 72 hours and then later, grinded to 1 millimeter using a sieve. Milled samples were analyzed for chemical composition on a dry matter (DM) basis. A series of three tests were performed on each sample and included NDF, ADF and Lignin. These key nutrients were then analyzed sequentially according to Van Soest *et al.*, (1991) (Van Soest *et al.*, 1991) using an ANKOM 220/200 Fibre Analyzer (Ankom Technology, New York, USA). NDF was determined with α-amylase.

2.3.4 In-situ field spectral measurements

Spectral reflectance measurements were carried out between the dates of 28th of April to the 04th of May 2015, using a hand-held field analytical spectral device (ASD) FieldSpec® 3 spectrometer (350 nm to 2500 nm). ASD reflectance measurements were made on clear days between 10:00 and 14:00 hours (CAT) for all 140 randomly selected sample points (Zhao *et al.*, 2007). To decrease atmospheric noise, the ASD was positioned to point on the representative grass species, one meter above ground at nadir position for each spectral measurement, yielding a ground field of view (FOV) of about 18 centimeters. The narrow FOV allows focus on a pure target (i.e. an undisturbed patch of grass) whereas the short distance

from the sensor minimizes atmospheric obscurities as well as signal-to-noise problems (Mutanga *et al.*, 2015). To derive the representative reflectance spectra for each sample point, 6 – 10 spectral measurements were made by moving over the 1 x 1m canopy (Adjorlolo *et al.*, 2012a; Mutanga *et al.*, 2015). Subsequently, spectral measurements at each sample point were averaged into a single spectral measurement and converted from radiance to reflectance using ViewSpec Pro software (ASD Inc. Boulder, CO, version 6.0.11). A Spectralon (Labsphere, Inc., Sutton, NH, USA) reference panel (white reference) was used to calibrate the sensor. Calibration was performed before each sample point to offset any change in irradiance of the sun and atmospheric condition (ASD, 2008).

2.4 Data analysis

2.4.1 Spectra evaluation and absorption features

Prior to algorithmic analysis, all sample spectra were visually analyzed for noise (Knox *et al.*, 2012). Noisy variables were consistently seen in the short-wave infrared (SWIR) spectral regions which could be accounted for by the seasonal variation of the grasses. Therefore, spectral regions between 1824-2064 nm of the electromagnetic spectrum were excluded from analysis due to high levels of noise associated within these regions. It is important to note that regression analysis was run on two sets of data (known absorption features dataset (table 1) and the entire hyperspectral dataset (350 – 2500 nm)) within this study. Only wavelengths that are physically linked to each fibre biochemical were selected as the known absorption features dataset. This physical link between biochemical and wavelengths are also known as absorption features. The establishment of these specific wavelengths have been extensively studied by (Curran, 1989; Curran *et al.*, 2001; Fourty *et al.*, 1996; Knox *et al.*, 2012; Kumar *et al.*, 2002; Mutanga & Kumar, 2007) (Table 1). These studies have determined these physically linked wavelengths through the association of excitation and reaction of molecular bonds at specific wavelengths (Knox *et al.*, 2012). A further regression analysis was run on a combined dataset (Known absorption features dataset + Variable selection dataset). The convergence of these two datasets were aimed at improving the models discrimination (R^2) of each biochemical within RF.

Table 1. Interpretation of known absorption features in reflectance of biochemicals NDF, ADF and Lignin from previous research

Wavelength (Nm)	Absorption Mechanism	Absorption Compounds		
			ADF	**NDF**
430	Electron transition	Chlorophyll a		**410–450**
460	Electron transition	Chlorophyll b		**455, 460**
700–800	Red edge	Nitrogen, protein		**720–745**
930	C-H stretch, 3rd overtone	Oil		**935**
970	O-H bend, 1st overtone	Water, starch		**960**
990	O-H stretch, 2nd overtone	Starch		**995**

1120	C-H stretch, 2nd overtone	**lignin**		**1115–1120**
1200	O-H bend, 1st overtone	Water, cellulose, starch, **lignin**	**1195**	
1420	C- H stretch, C--H deformation	**lignin**		
1450	O-- H stretch, 1st overtone, C-- H stretch, C-- H deformation	Starch, sugar, **lignin** water		
1580	O-H stretch, 1st overtone	Starch, cellulose	**1580**	
1670	Stretch OH	**lignin**, starch protein		
1685	C-H stretch, 1st overtone	**lignin**		**1675–1685**
1690	C-H stretch, 1st overtone	**lignin**, starch, protein, nitrogen		**1700**
1760		**lignin**		
1940	O- H stretch, O- H deformation	Water, **lignin**, protein, nitrogen, starch, cellulose		
1950	Stretch OH, deformation OH	Water, **lignin**, protein, nitrogen, starch, cellulose		
2050	N-H asymmetric stretch + Amide II	Protein		**2050**
2060	N = H bend, 2nd overtone/N = H bend/N-H deformation	Protein, nitrogen		**2060**
2080	O-H stretch/O-H deformation	Sugar, starch		**2070**
2100	2 × O-H deformation + 2 × CO stretch	Starch, cellulose		**2090**
2240	C-H stretch	Protein	**2245**	
2270	Stretch CH, stretch C = C	**lignin**		
2276	O-H stretch + C-C deformation	Cellulose, **lignin**		
2280	C-H stretch/CH2 deformation	Starch, cellulose		**2285**
2294	N-H stretch + C = O stretch	Protein		**2290**
2310	C-H bend, 2nd overtone	Oil	**2315**	**2310–2315**
2320	C-H stretch/CH2 deformation	Starch		**2320–2325**
2336	C-H stretch + C-H deformation	Cellulose, starch, **lignin**		**2335**
2340	C-H stretch/O-H deformation/ C-H deformation/O-H stretch	Cellulose		**2340**
2380	Stretch OH, aromatic deformation	**lignin**		

Source: (Curran, 1989) (Himmelsbach *et al.*, 1988) (Elvidge, 1990) (Osborne & Fearn, 1986) (Kumar *et al.*, 2002) (Kawamura *et al.*, 2008) (Fourty *et al.*, 1996). The known absorption compounds and wavelengths are depicted in bold font.

2.4.2 Random forest regression analysis

In this study, the random forest (RF) ensemble (Breiman, 2001) was implemented using the "randomForest" package (Venables & Smith, 2005) in R statistical software (Team, 2013). RF is a machine learning algorithm developed to enhance the classification and regression trees (CART) method. A large ensemble of un-pruned regression trees (n_{tree}) are constructed using a systematic approach called bootstrapping. Bootstrap samples are created from many regression trees (generally between 500 and 2000 trees) to create a final prediction that is taken from an average of all individual tree outputs (Dye *et al.*, 2011; Lawrence *et al.*, 2006; Prasad *et al.*, 2006). The resultant trees in the ensemble were used to assign each input spectral waveband to a class membership of the response variables (i.e. NDF, ADF and Lignin). When bootstrap samples are drawn, approximately one third of the samples are left out. The samples that are left out are called out-of-bag (OOB) samples (Breiman, 2001). OOB samples are used to calculate an unbiased assessment of the classification accuracy (i.e. OOB error) and to determine variable importance (Adjorlolo *et al.*, 2013; Breiman, 2001; Peerbhay *et al.*, 2015). Within the training data set (approximately 70%), trees are grown to a maximum length and based on the maximum number of votes that a class receives, the ensemble of trees assigns each waveband to a class (Adjorlolo *et al.*, 2012a; Breiman, 2001). In each tree, a randomly chosen subset of the total number of variables (m_{try}) is determined by the best split (the random selection of variables) at each node (Abdel-Rahman *et al.*, 2013; Breiman, 2001; Peerbhay *et al.*, 2015). It is important that the algorithm is well parameterized according to two main variables: n_{tree}, number of regression trees grown based on a bootstrap sample of observation; m_{try}, number of predictors tested at each node (Ramoelo *et al.*, 2015b). The m_{try} function is significant as it is a determinant of solving the 'small n large p' problem whereby the user is given the option to choose a small subset of variables (Dye *et al.*, 2011). Furthermore, studies constitute that the m_{try} default parameter value is confident in consistently yielding accurate results (Adjorlolo *et al.*, 2013; Dye *et al.*, 2011; Peerbhay *et al.*, 2015).

Furthermore, RF offers variable importance measures embedded in its system. Variable importance measures provide the researcher with good indicators of the effect of any predictor variable against the response variable. The permutation measure of variable importance is used to determine the importance of each predictor by measuring the percentage increase in the mean square error when the OOB data for each variable is permuted, while the others are unchanged (Breiman, 2001). In this study, the percentage increase in mean square error (MSE) was used to predict the most important wavelengths to detect NDF, ADF and Lignin. It is the increase in

mean square error of predictions (estimated with out-of-bag-CV) as a result of a variable being permuted. The higher percentage gives more importance to a specific waveband. High variable importance scores suggests high weighting on a particular variable in the model (Ramoelo *et al.*, 2015b). The RF algorithm was preferred due its robustness and accurate method for band selection, specifically when using high dimensional data (Breiman, 2001; Dye *et al.*, 2011; Mutanga *et al.*, 2015; Peerbhay *et al.*, 2015).

2.4.3 Accuracy assessment (R^2)

The R-square (R^2) (Albayrak, 2008; Knox *et al.*, 2011; Pullanagari *et al.*, 2012) was the preferred accuracy assessment used to assess the performance of the RF method. The final data set was split into training and test (Congalton & Green, 2008). The square of correlation between predicted and observed values was calculated, values closer to 1, predict better results.

2.5 Results

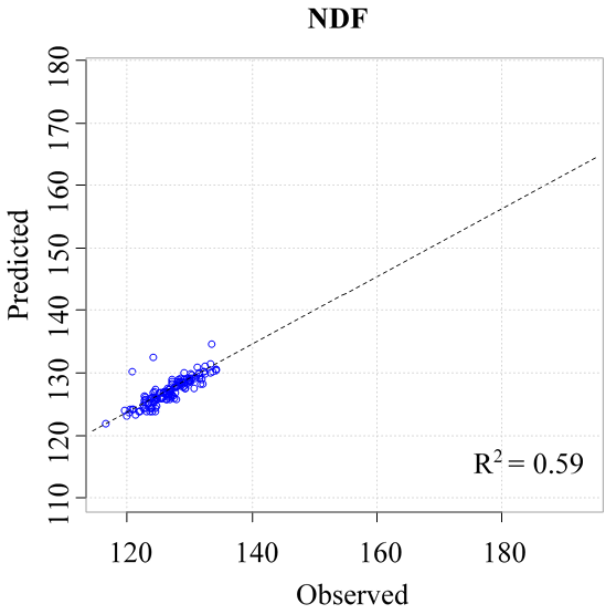
The results from chemical analysis performed on the grass samples collected in the field are presented in table 2.

Table 2. Summary of chemical analysis results obtained from 140 tropical mixed grass samples collected in the field

Nutrient	Mean (g/kg)	Range (g/kg)	Std. Dev. (g/kg)
NDF	127.48	116.48 – 193.16	6.623980369
ADF	91.80	85.06 – 100.86	3.427728864
Lignin	91.87	85.27 – 100.99	3.428042785

2.5.1 Detecting NDF, ADF & Lignin concentrations using known absorption features dataset

Using the known absorption features (Table 1) for NDF, ADF and Lignin, the random forest model (n_{tree} = 500 and m_{try} = 40) successfully determined NDF, ADF and Lignin biochemicals, however the fit of the model as measured by R^2 was low, the model still explains 59% of the variation in NDF. Lignin was successfully predicted with an R^2 of 0.81 and ADF an R^2 of 0.79, whilst NDF was predicted with an R^2 of 0.59 (Figure 2).



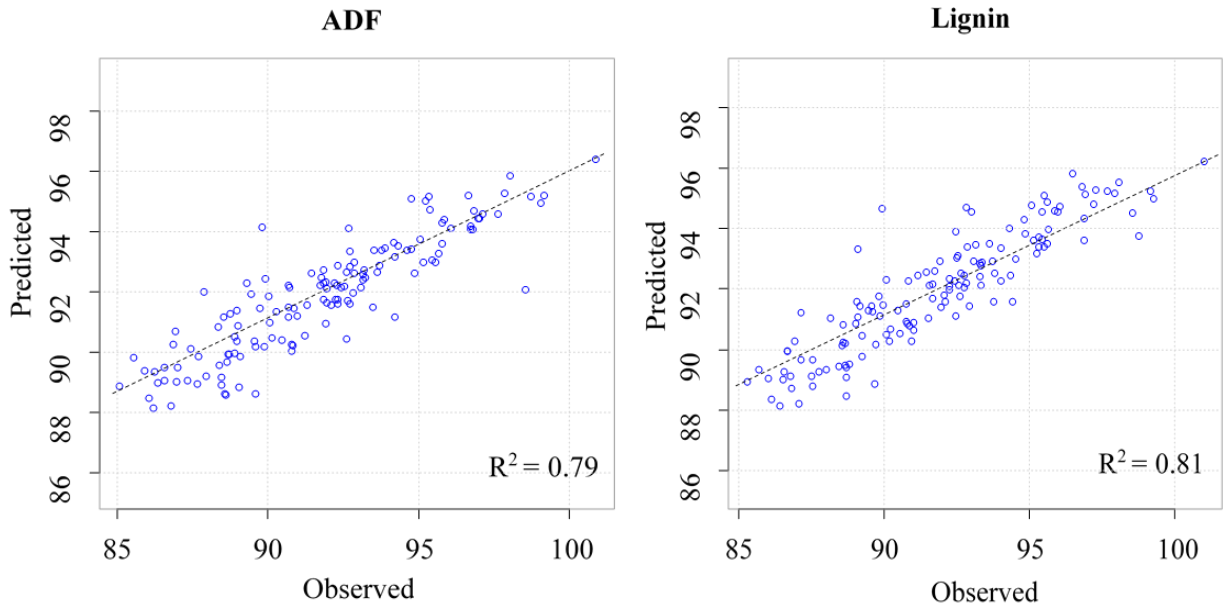
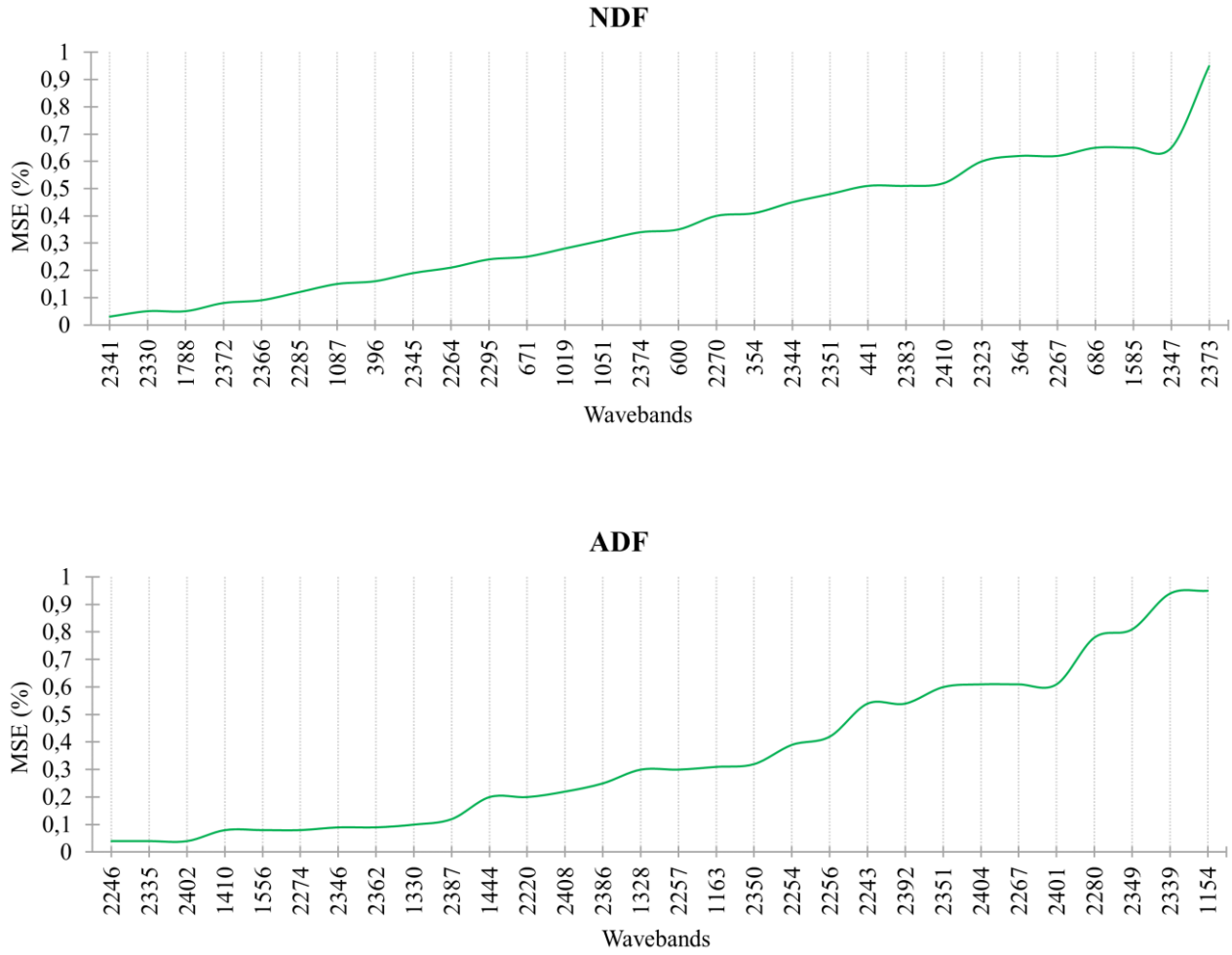


Figure 2. Regression results of NDF, ADF and Lignin biochemicals using the known absorption features



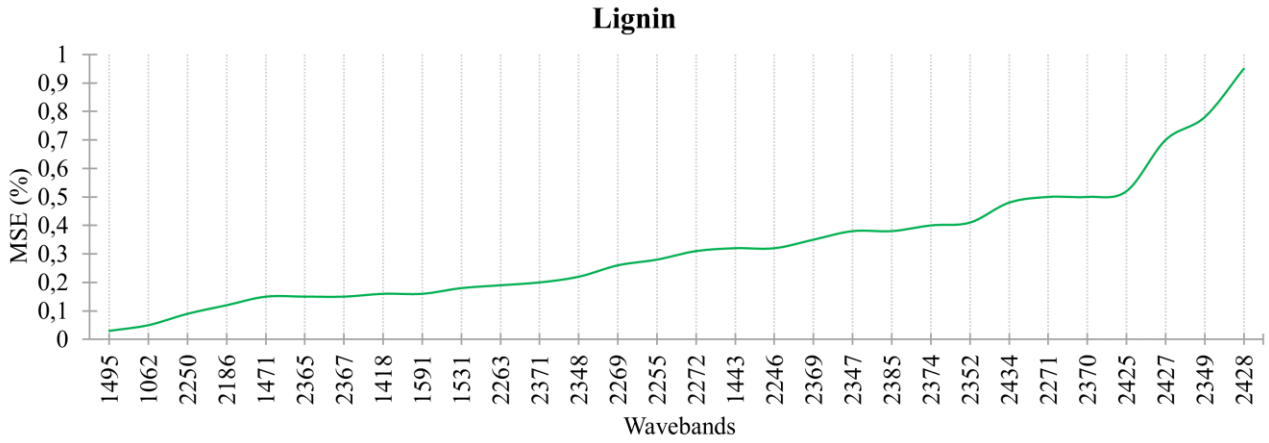
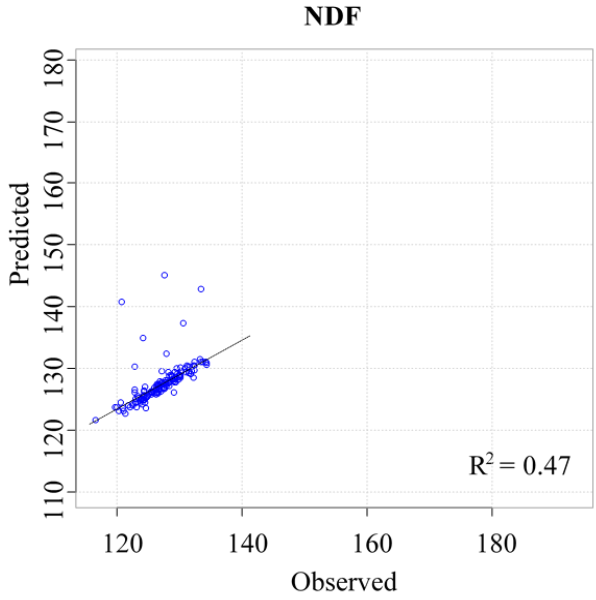


Figure 3. Variable importance of NDF, ADF and Lignin biochemicals using known absorption features

2.5.2 Detecting NDF, ADF & Lignin using the entire hyperspectral dataset

Whilst, the n_{tree} and m_{try} values remained consistent for each biochemical, the entire hyperspectral dataset was used in the final model to predict each biochemical. Using the entire hyperspectral dataset, the random forest model (n_{tree} = 500 and m_{try} = 40) successfully predicted the key ADF and Lignin biochemicals used in this study. Again, NDF was unsuccessfully predicted. More specifically, NDF was predicted with a low R^2 of 0.47, while ADF and Lignin were predicted with a high R^2's of 0.88 and 0.87, respectively (Figure 4).



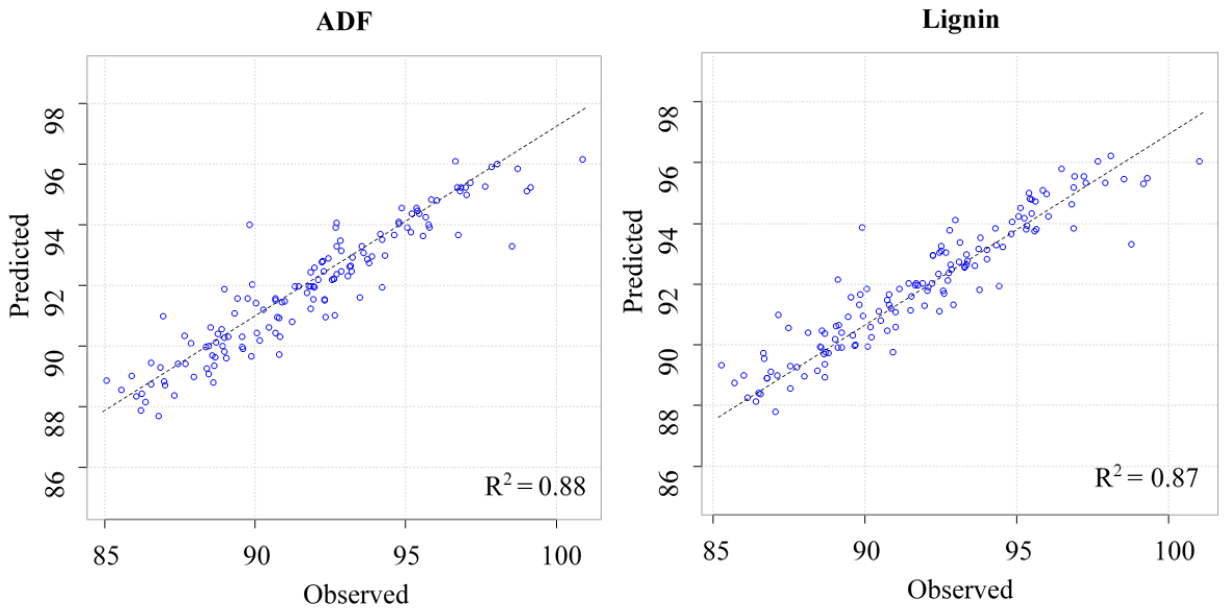


Figure 4. Regression results of NDF, ADF and Lignin using the entire hyperspectral dataset

2.5.3 Variable importance using the entire hyperspectral dataset

Figure 5 depicts the most effective variables used in the prediction model. For NDF, wavelengths in the blue (352, 368, 385, 386), red (696), and shortwave region (1137, 1344, 1708, 2470, 2495 and 2497) were highly important, while for ADF, wavebands specifically along the blue (403, 407, and 420), and shortwave region (1380, 2428, 2470, 2487 and 2497) showed the most importance. For Lignin, wavelengths located in the blue (407, 408) and shortwave region (2428, 2470 and 2487) showed the most influence (figure 5). Furthermore, RF modelling offers three measures of variable importance: (1) based on the number of times a candidate variable is selected; (2) the Gini index that was proposed by (Breiman *et al.*, 1984) in the original classification and regression trees method; (3) the permutation of a variable as an ensemble of variable importance (Strobl *et al.*, 2008). In this study, the permutation of a variable as an ensemble of variable importance, measured as a percentage increase in MSE was an accurate measure in determining variable importance and is illustrated in figure 5.

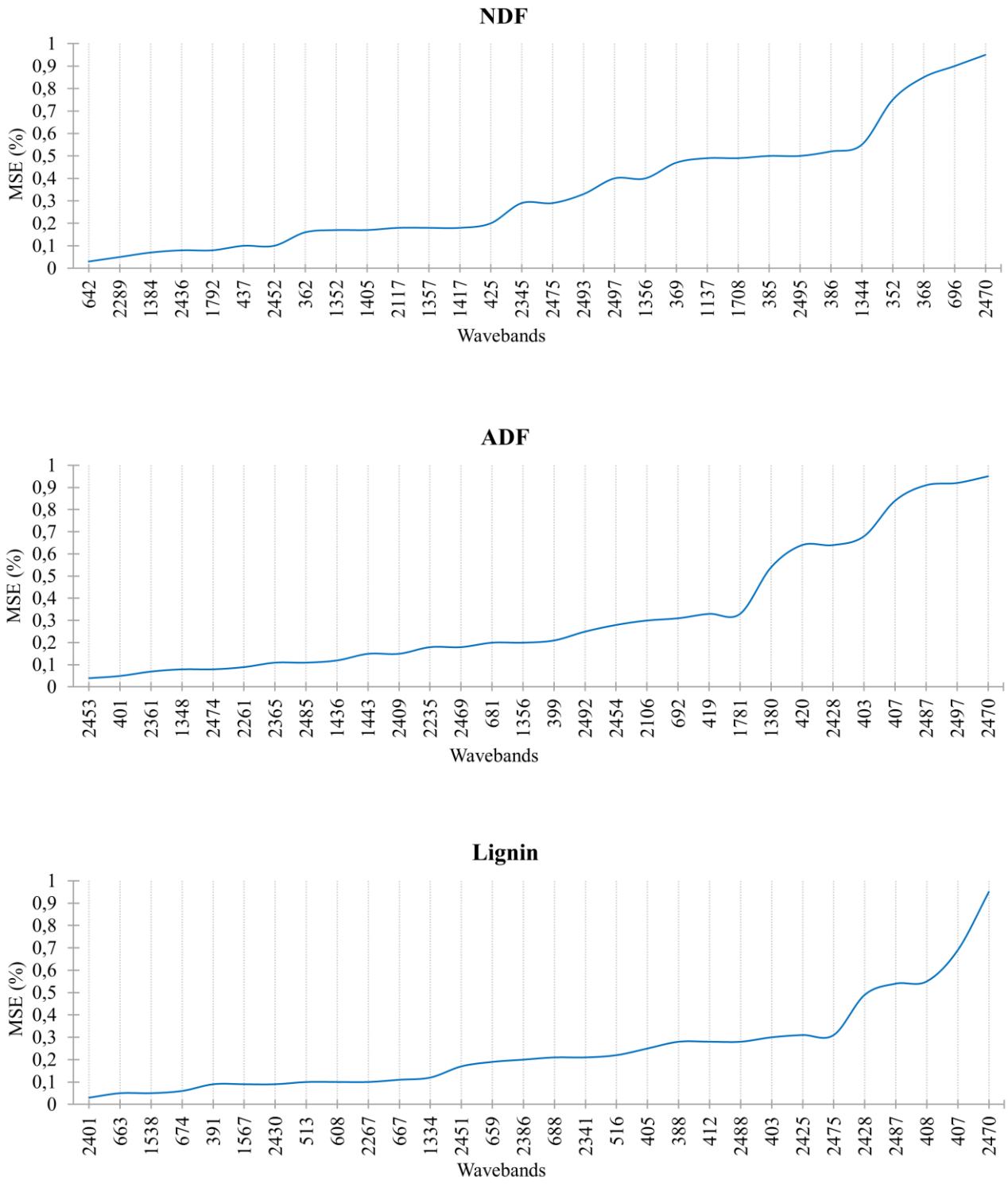


Figure 5. Variable importance of NDF, ADF and Lignin biochemicals using the entire hyperspectral dataset

For comparison purposes, we executed a combination of selected variable importance wavebands derived from the entire hyperspectral dataset and known absorption features

dataset. Both datasets were combined and implemented in the RF model (combined dataset: variable selection wavebands + known absorption feature wavebands for NDF, ADF and Lignin (Table 3)). Results reveal a small increase in R^2 values of 0.03% for ADF and Lignin and a small decrease of 3% for NDF, when compared with the known absorption features dataset (Table 3).

Table 3. R^2's of each biochemical using different datasets

Data sets	No. of wavebands	Biochemicals (%)		
		NDF	ADF	Lignin
Known absorption features	NDF: (110) ADF: (4) Lignin: (14)	R^2: 59	R^2: 79	R^2: 81
Entire hyperspectral	(1910)	R^2: 47	R^2: 88	R^2: 87
Variable selection	(30)	R^2: 40	R^2: 84	R^2: 85
Combined	NDF: (140) ADF: (34) Lignin: (44)	R^2: 54	R^2: 83	R^2: 84

2.6 Discussion

Forage fibre content is important in understanding the distribution patterns of livestock and the quality of their potential intake. Increases in forage fibre content decreases the palatability of grasslands. Thus, livestock move to more palatable areas where grasses are easily digested. This has negative effects on agricultural production which developing countries rely on, as it directly affects economic and socio-economic status of a country. This study aimed to determine the effectiveness of known absorption wavelengths for detecting forage fibre biochemicals, using hyperspectral data on a mixed tropical grassland in South Africa. Using the random forest regression, NDF, ADF and Lignin were successfully detected while effectively reducing the large number of wavelengths of the hyperspectral data set. It is important to use processing techniques that reduce the number of input wavelengths without decreasing the integrity of the model's accuracy (Dye *et al.*, 2011).

2.6.1 Determination of NDF, ADF & Lignin using known absorption features

Using known absorption features, this method effectively detected each biochemical, however it has shown to have produced slightly lower R^2 results for ADF and Lignin than when using the entire hyperspectral data set, which are contrary to most studies (Kawamura *et al.*, 2008; Kawamura *et al.*, 2010; Mutanga *et al.*, 2004b; Zhao *et al.*, 2007). Nonetheless, NDF increased by 0.12, and ADF and Lignin decreased by 0.09 and 0.06, respectively, still is reliable. The results in this study explain the findings of previous studies (Kawamura *et al.*, 2008; Pullanagari *et al.*, 2012), which detected biochemicals with similar R^2 accuracies. Although ADF, Lignin results were highest, NDF produced the lowest result. This could be a result of a small range of variation (Zhao *et al.*, 2007). Kawamura *et al* (2008) recommends linking these results to radiative transfer models that could further improve accuracies.

It is important to note the relationship between fibre dry matter and R^2 produced over the series of chemical measurements made to derive each biochemical from NDF to ADF to Lignin, respectively. The relationship evidently shows that with a decrease in dry matter, there is almost a directly proportional increase in R^2 (Table 3).

2.6.2 Determination of NDF, ADF & Lignin using the entire hyperspectral dataset

The results confirm that forage fibre biochemicals can be detected in a tropical mixed grassland using hyperspectral data and with the robustness of the random forest algorithm. Whilst, many

remote sensing studies (Kawamura *et al.*, 2008; Kawamura *et al.*, 2010; Knox *et al.*, 2011; Pullanagari *et al.*, 2012; Zhao *et al.*, 2007) focused on using PLS, MLR, MAXR, and neural network (NN) algorithms, RF proved to be a valuable and accurate measure for detecting forage fibre biochemicals in this study. For example, Kawamura *et al*., (2010) successfully detected NDF, ADF and Lignin using GA-PLS and obtained relatively high r^2 accuracies. In a subsequent study, Pullanagari *et al*., (2012) obtained similar NDF, ADF and Lignin r^2's through using PLS. However, this study achieved even higher accuracies when using the RF algorithm (Table 3). This success validates the robustness and effectiveness of the random forest ensemble to discriminate each biochemical especially when using highly dimensional data. When using the entire hyperspectral data set, once again NDF appeared to still be reasonably low due to the small range of variation and also, could be influenced by the difference in the solar zenith angle that could have affected the predictive accuracy of the biochemical (Kawamura *et al.*, 2008).

2.6.3 Variable importance using the entire hyperspectral dataset

The selection of the most important wavelengths was determined when using the entire hyperspectral data set. The results are related to the findings of previous studies (Curran, 1989; Elvidge, 1990; Himmelsbach *et al.*, 1988; Kawamura *et al.*, 2008) (Table 1), who determined the most important wavelengths for detecting NDF, ADF and Lignin. For instance, NDF wavebands (425, 437, 1708, 2289 and 2345) are closely linked to those found in previous studies. ADF waveband 1443 was the only closely linked waveband and Lignin wavebands (1334, 1538, 1567, 2267, 2386, 2341, 2430 2401, and 2451) were closely linked to previous studies. All the wavebands described above had a waveband-offset range of 0 to 137 nm of the known absorption features found in previous studies. The random forest algorithm could derive new detection wavelengths for each biochemical as shown in Figure 5. This is justifiable as most of the new wavelengths were found in the blue and shortwave regions, which relate to C-H, N-H and O-H bonds (Curran, 1989; Norris *et al.*, 1976). This study had the privilege of analyzing uncovered areas of the electromagnetic spectrum (i.e. shortwave regions) using advanced remote sensing technology, whereas previous studies (Curran, 1989; Elvidge, 1990; Himmelsbach *et al.*, 1988; Norris *et al.*, 1976) did not, and only analyzed areas within the visible and near infrared regions. Nonetheless, the random forest algorithm successfully recognized some of the wavelengths found in previous studies, and the model found new

wavelengths that also should be taken into consideration. Hence, these new wavelengths can be used and should be taken into consideration in later studies.

For example, improving the efficacy of these new wavelengths could result in better detection accuracies for future research. This will decrease computational time and increase the robustness of modelling especially when dealing with high dimensional data. Hence, the configuration of sensor-specific wavelengths to precisely detect NDF, ADF and Lignin, could lead to efficient and effective mapping resource for these forage fibre variables. Rapid detection of these forage fibre variables offers farmers and rangeland managers a time effective way to make informed decisions, especially in South Africa where land systems change is widespread. Future research could consider upscaling the results found in this study to regional scale mapping that will benefit farmers and rangeland managers on a greater scale. Furthermore, this application will benefit precision farming by effectively guiding decision making when considering the most important palatability variables prevalent in a management zone. More importantly, the results of this study will improve productivity and increase farm profits by sustainably improving management regimes of farm inputs (Mulla, 2013). In addition, current policy implications suggest that the biodiversity and ecosystems services of grasslands are threatened and under immense pressure in South Africa. Hereby, grasslands are host to a minimum of 45% of endemic animal species that depend on grasslands for good quality forage and for safety (WWF, Undated). As outcome of this research, policy will be more informed on the distribution and intake potential rates of these endemic species by creating quality and habitual parameter maps. Furthermore, remote sensing has the potential to create regional maps to inform policy of the quality of grassland vegetation, including the mapping potential fire danger areas. For example, the eThekwini municipality 'working on fire' programme can be more informed, thus, the programme will be proactively aware of potential fire risk areas. For these reasons, this research serves as valuable resource to society, the economy and the natural environment. Moreover, future research could focus on creating innovative modelling approaches to improve detection accuracies, hence providing assurance to rangeland managers and farmers. For example, innovative modelling approaches such as using a combination of datasets has the potential to improve prediction accuracies.

As a final step, this research used a combination of variably selected wavebands and known absorption features. This approach produced positive results which to our knowledge has never been done before. Increases in R^2 for both ADF and Lignin reveal the potential of this method to discriminate and predict more accurately. Hence, the models performed better with the inclusion of discriminately relevant wavebands (variable selected wavebands) found

in the entire hyperspectral dataset. However, a noticeable decrease in R^2 of NDF was predicted. The decrease in NDF was prone to the model reaching saturation point due to closely correlated predictor variables, caused by a multicollinearity problem. Nonetheless, majority of the biochemicals predicted with increased R^2's, hence the significance of converging datasets is a useful approach. Ultimately, fine tuning this approach could produce more positive findings when used across numerous datasets.

2.7 Conclusion

This study effectively determined the known absorption wavelengths for detecting forage fibre biochemicals, neutral detergent fibre (NDF), acid detergent fibre (ADF) and Lignin using hyperspectral data. The results showed that the random forest ensemble could discriminate each biochemical, demonstrating its robustness and effectiveness for this study. The study demonstrated the predictive ability of the RF ensemble and the importance of reducing the number of redundant wavelengths which is a necessary step for any hyperspectral application. The importance of using known absorption features should always be an integral part for improving the performance of regression models. This study showed that additional wavebands located within the blue, red and SWIR regions were also important for forage quality determination. These results can be used to provide livestock farmers and rangeland managers with an efficient tool for identifying pasture areas that contain high or low levels of fibre, hence guiding management strategies. Overall, the results are positive, although it is important to know that this study was carried out as a snapshot over one location. Future research should venture into applying this knowledge to various data sets and to new locations to detect foliar fibre biochemicals. Ultimately, future research could potentially upscale the results obtained in this study to a multispectral application.

Chapter Three

Multispectral mapping of key grassland nutrients in KwaZulu-Natal, South Africa

3.1 Abstract

RapidEye multispectral imagery promises to be effective in vegetation assessment. To date, the estimation of forage quality has been performed using hyperspectral systems. Even though some hyperspectral studies produced accurate results, they are costly, produce large datasets and require a large amount of time for data processing. This research intends to investigate the utility of using 5 m multispectral RapidEye-5 imagery combined with a specialized algorithm to detect and map important forage fibre biochemicals such as neutral detergent fibre (NDF), acid detergent fibre (ADF) and Lignin in an African tropical grassland. Analysis was conducted on 140 grass samples collected at plot level and then, correlated and mapped using the high-resolution image and two tree ensemble modelling techniques. Results showed that the random forest method successfully mapped NDF, ADF and Lignin with R^2 values ranging between 0.67 and 0.74. In comparison, the study used Stochastic Gradient Boosting (SGB) as an alternative method which produced very similar R^2 values ranging between 0.65 and 0.72. Overall, the results showed that multispectral remote sensing, in conjunction with ancillary data (Leaf Area Index, Chlorophyll, species type, species Count) can detect and map key forage fibre biochemicals in a grassland environment.

Keywords: RapidEye, multispectral, NDF, ADF, Lignin.

3.2 Introduction

Technological advancements have changed the science of remote sensing by providing more specialized platforms to analyze more detailed vegetation properties. Highly sophisticated satellite platforms can provide researchers with more detailed and accurate information. The latest inventions are equipped with specialized spatial, spectral, radiometric and temporal resolutions capable of detecting more discernible information (Mutanga *et al.*, 2016). More specifically, the launch of RapidEye-5 multispectral satellite has a combination of high resolution, multispectral capabilities, large-area coverage and frequent revisit intervals, offering researchers incomparable data resources. The RapidEye-5 platform configurations promises to have the potential to map vegetation over a large area, at finer resolution, at faster time, and with a lower cost (RapidEye, 2011). To date, not many vegetation applications have investigated the capabilities of remote sensing to perform detailed analysis on the quality, species and biomass of vegetation properties (Mutanga *et al.*, 2016). Hence, a research gap still exists in mapping the nutrient limiting foliar biochemicals, such as non-digestible fibre whereby the content of fibre can determine the intake of grazing animals and ultimately determine their quality of health.

In South Africa, more than 70% of land surface comprise grasslands, which are predominantly used for pastoral production (Mansour *et al.*, 2013). Grasslands provide the primary feed base for grazing livestock which in turn provides the majority of low-income people with food, goods and services that are essential for sustaining their livelihoods (Boval & Dixon, 2012). Problems exist when the palatability of the grasses are low, effectively the grasses become hard for livestock to digest. Hence, the palatability of grass directly affects the intake potential of livestock, later influencing livestock production levels and eventually food security in a country (Knox *et al.*, 2011). In this regard, fibre biochemicals are significant indicators of the palatability of grasses, and therefore knowledge regarding their concentrations is vital to rangeland managers and farmers (Mbatha & Ward, 2010). More specifically, accurate fibre content indicators can inform farmers and rangeland managers on livestock intake potential; stocking and distribution rates; and supplementation, whilst at policy level, information regarding the quality of grasslands; fibre content and the impacts it has on livestock grazing, can be used to make informed decisions related to ecosystem services and protection (Lamarque *et al.*, 2011; Loucougaray *et al.*, 2015). Therefore, improved knowledge of soil-plant-animal constraints allows for improved management techniques to meet the demands faced by farmers; rangelands managers; policy makers; commercial livestock production; and

the general economy (Boval & Dixon, 2012). However, current methods used to obtain this vital information is challenging. Hence the need to innovate automated, cost and time effective methods is vital.

In the past, determining or quantifying the fibre content of grasslands has mainly depended on two laboratory methods: wet chemistry and near-infrared spectroscopy, which are both Association of Agricultural Chemists (AOAC) approved methods (Pullanagari *et al.*, 2013b). However, both methods involve ground-based periodic surveys and tedious laboratory work, which are costly and time consuming. Generally, field hyperspectral remote sensing has proven successful when detecting fibre biochemicals by providing a more detailed spectral description for quantification, as a result of its high spectral and extremely fine radiometric resolutions (Jusoff, 2007). For example, Kawamura *et al*., (2008) successfully predicted above ground biomass (BM), crude protein (CP) and ADF from a mixed sown pasture using field hyperspectral systems. Thus, the PLS model produced the highest R^2 values of 0.72, 0.65, and 0.62 for BM, ADF and CP, respectively. Similarly, Pullanagari *et al*., (2012) successfully mapped pasture quality variables from four commercial farms across New Zealand. Thus, the PLS regression algorithm yielded R^2 accuracies of 0.75, 0.82, and 0.71 for NDF, ADF and lignin, respectively. Despite the abilities of laboratory and field hyperspectral techniques to effectively produce extremely accurate results, operations were highly expensive; technical; time consuming; point-based; dangerous and require optimal operating conditions. Furthermore, hyperspectral remote sensing data often represents an oversampled data which creates dimensionality and multicollinearity problems. Hence, on a local scale, the technology is neither feasible nor financially viable for farmers and rangeland managers (Pullanagari *et al.*, 2013b). With the advent of new space-borne multispectral remote sensing technologies, there is provision for an indirect, non-destructive, time efficient and cost-benefit alternative for determining or quantifying pasture quality over a large scale (Kawamura *et al.*, 2010; Peerbhay *et al.*, 2016; Pullanagari *et al.*, 2013b).

Recently, the latest advancements in space-borne sensor technologies allows researchers to investigate the potential of the new generation, high resolution multispectral satellites (e.g., RapidEye 1-5, Sentinel-2,WorldView-2, SPOT)(Peerbhay *et al.*, 2016). Most new generation satellites are configured with the new red-edge waveband, that has the potential to considerably improve classification accuracies within vegetation applications. Hence, several studies found strong correlations between the red-edge waveband and plant biochemical and biophysical parameters (Abdel-Rahman *et al.*, 2013; Adam *et al.*, 2010; Adelabu *et al.*, 2014; Adjorlolo *et al.*, 2012a; Delegido *et al.*, 2011; Laurin *et al.*, 2016;

Mutanga & Adam, 2011; Ramoelo *et al.*, 2012b; Schuster *et al.*, 2012; Zengeya *et al.*, 2013). Evidently, a combination of high resolution and strategically placed wavebands across the electromagnetic spectrum, make it possible to accurately predict chemical properties within the grassland ecosystem (Eitel *et al.*, 2007). For example, Zengeya *et al*., (2013) successfully estimated and mapped nitrogen concentrations using WorldView-2 (WV-2) imagery in savanna grasslands. A result of (R_{cv}^2 = 0.66, relative error = 0.13%) was obtained using partial least squares regression (PLSR). Similarly, Mutanga *et al*., (2015) successfully modelled nitrogen using the random forest (RF) model and obtained R^2's of 0.74 and 0.65 on an independent field spectral test dataset and on the actual WV-2 image, respectively. Here, many studies have focused on remote sensing of nitrogen as a limiting nutrient using mainly field hyperspectral remote sensing. However, to date, few studies have explored the capability of remote sensing to map fibre biochemicals (NDF, ADF, Lignin) as non-digestible and unpalatable fibre to grazing ungulates using high resolution space-borne multispectral remote sensing technology (Pullanagari *et al.*, 2013b).

Research suggests the need for more reliable results in the estimation of pasture quality (crude protein, fibre etc.); in instrumentation (spectral, spatial and temporal resolutions) and in the various challenges associated with the complexity of pastures (invasive weeds, animal grazing and trampling, presence of many different species, changing stages of maturity and varying proportions of green and dead material). Therefore, the need to establish improved methods to accurately map fibre content is vital for the sustainability of land management practices (Pullanagari *et al.*, 2012). Recent research recommends the use of high resolution sensors to detect plant foliar biochemicals with higher precision within the grassland ecosystem (Knox *et al.*, 2011; Schellberg *et al.*, 2008). In a rapidly changing climate with an increasing rate of uncertainty, this study aims to detect and map fibre content within a tropical grassland of KwaZulu-Natal, South Africa using 5 m RapidEye-5 multispectral imagery combined with a specialized algorithm, and ancillary data.

3.3 Materials and Methods

3.3.1 Study area

The study was conducted in Fort Nottingham Nature Reserve, located in KwaZulu-Natal, South Africa (latitude 29°24'9.62"S and longitude 29°53'48.95"E) (Figure 6). The study area resides within a reserve occupying an area of 1227 ha and is characterized by an extensive area of the Eastern Mistbelt Forest and Drakensberg Foothill Moist Grassland of the KwaZulu-Natal, Midlands. Approximately 778.2ha in total land use is currently leased from uMngeni Municipality for grazing of cattle to local commercial farmers. The study area is characterized by mainly grassland vegetation, that is categorized by *Festuca costata* (blue fescue), *Tristachya leucothrix* (hairy trident grass) and *Themeda triandra* (red oat grass) that are the most dominant grass species type (Mucina & Rutherford, 2006). *Festuca costata* and *Tristachya leucothrix* is native to Southern Africa while *Themeda triandra* is native to Kenya. The study area receives an annual spring and summer rainfall of approximately 950 mm/year. The most dominant soils found at the study area are: Clovelly; Hutton; Griffin and Oatsdale (Mucina & Rutherford, 2006). The topography is characterized by a diversity of habitats split between a moderately undulating Drakensberg Moist Foothill grassland plateau (at approximately 1750m amsl) leading to densely forested south and east facing slopes, (between 1700 and 1400 amsl) that merge into lower lying remnant Drakensberg Foothill Moist Grassland flats. This study area was preferred due to heterogeneity of grass species and the influence of grazing.

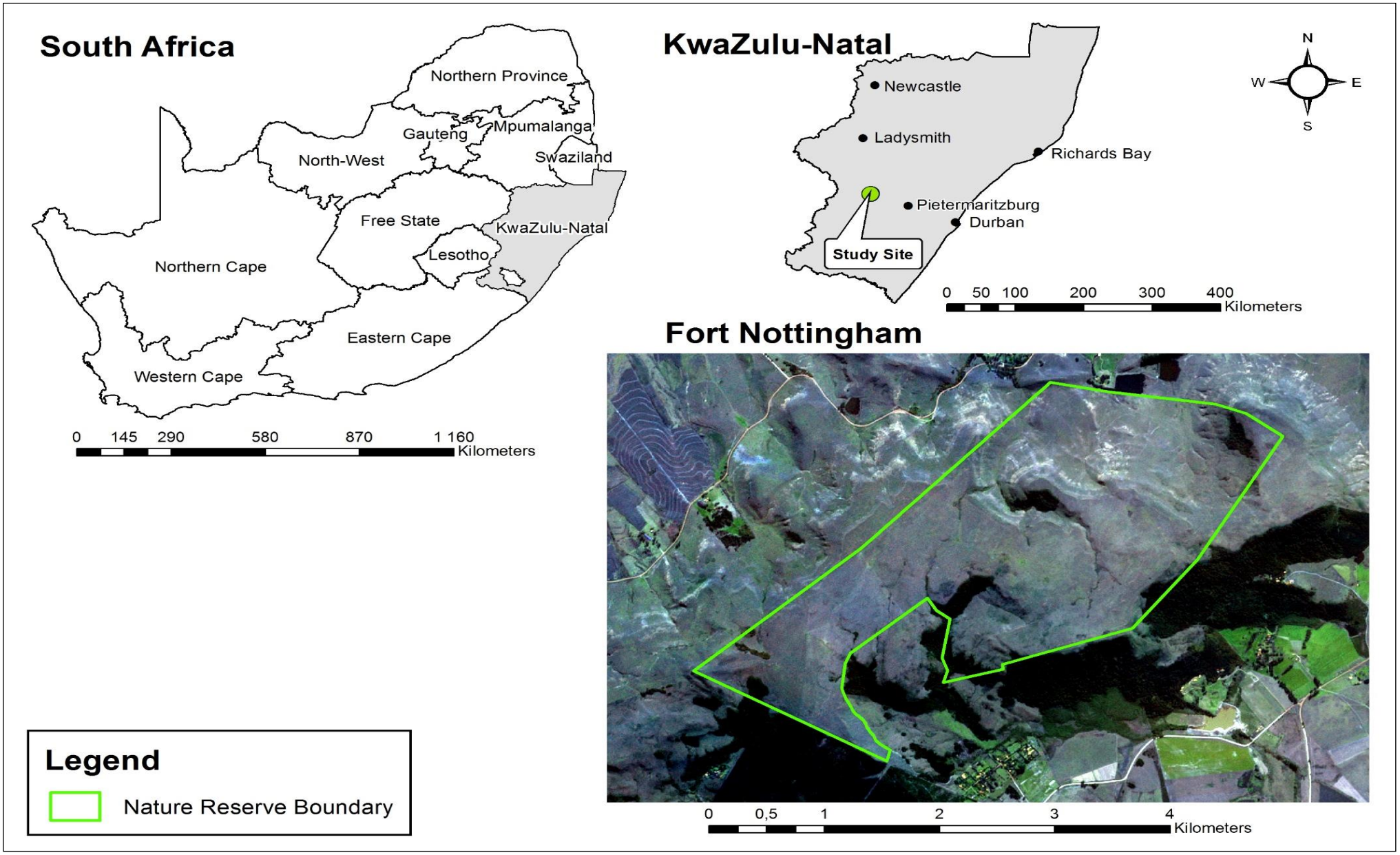


Figure 6. Location of the study site in the Fort Nottingham Nature Reserve, KwaZulu-Natal, South Africa

3.3.2 Field data collection

Field data collection was carried out between the dates of 28th of April to the 04th of May 2015. A stratified random sampling method was used to select field sampling points using the Hawths tool in ArcGIS 10.2. A total of 140 sample points were established across the eleven existing leased grazing camps. All sample points were uploaded into a Trimble GeoExplorer 6000 series GeoXH GPS with a centimeter-level accuracy of less than 10cm and was used to locate each sample point. To declare subjectivity over the three dominant grass species, a 1m x 1m quadrant was overlaid at the center of each of the 140 sample points (Adjorlolo *et al.*, 2012a). Within each quadrant, two non-destructive measurements were taken: a series of leaf area index (LAI) measurements, and chlorophyll (CHL) measurements; and one destructive sample of the grass species present. The destructive sample was taken immediately after LAI and CHL measurements were taken. Grass samples were clipped using a scissors at 1 cm above ground, pooled and bagged (brown bag) for drying and later chemical analysis (Knox *et al.*, 2011).

3.3.3 Ancillary data

Within each quadrant LAI and CHL measurements were taken. LAI measurements were taken using the LI-COR (LAI-2200) plant canopy analyzer under cloudless sky conditions using a 270^{o} view cap for use in short canopies and to hide the operator from being potentially in the field-of-view (FOV). One above canopy reading (*A reading*) was taken for four or more below (*B reading*) canopy readings in each quadrant, *B readings* were taken within the grass leaf structure. To reduce uncertainty, more *B readings* were taken (approx. 8) as recommended in the LI-COR instruction manual (LI-COR, 2009). Furthermore, a series of chlorophyll measurements were taken using a Konica-Minolta SPAD-502 Plus chlorophyll meter. Nine SPAD measurements were taken at the four corner points and center point of the 1m x 1m quadrant. The nine readings were then averaged to produce a single SPAD measurement of the amount of chlorophyll content present in the leaves at each sample point (Friedman *et al.*, 2016; Teoh *et al.*, 2015). In addition, grass species type, and species count (SC) were visually identified and recorded within each sample point and used as input variables in modelling. In this study, the following ancillary data were used as input variables in the model: LAI; CHL; species type; and SC Previous research recommends the incorporation of ancillary data with multispectral imagery as it improves classification accuracy (Lawrence & Wright, 2001).

3.3.4 Laboratory analysis

All samples collected (n = 140) were analyzed at the University of KwaZulu-Natal laboratory. Samples were oven dried at 65 ^{o}C for 72 hours. Samples were then grinded to 1 millimeter using a sieve. The milled samples were analyzed for chemical composition on a dry matter (DM) basis. The following foliar fibre biochemicals: NDF, ADF and Lignin, were derived from the samples. These three foliar fibre biochemicals (NDF, ADF and Lignin) were then analyzed sequentially using an ANKOM 220/200 Fibre Analyzer (Ankom Technology, New York, USA).

3.3.5 Image acquisition and pre-processing

A RapidEye-5 (level 3A) multispectral image of the study area was acquired on 27 May 2015 under cloudless (0%) conditions. The RapidEye-5 sensor comprise five wavebands namely: blue (440-510 nm), green (520-590 nm), red (630-685 nm), red-edge (690-730 nm) and, near-infrared (NIR) (760-850 nm) which sense in the 440 nm to 850 nm spectral range with a spatial resolution of 5 meters. The image was orthorectified as individual 25 by 25 kilometer tiles and was projected to Universal Transverse Mercator (UTM) using the WGS-84 Geodetic System by the supplier. The image was processed to remove distortions caused by terrain and it was geometrically, sensor, radio-metrically corrected and aligned to a cartographic map projection by the supplier. In addition, the image was atmospherically corrected to surface reflectance using the dark object subtraction method (DOS) using ENVI 4.7 (ENVI, 2009). The RapidEye-5 image was produced using Global Reference 2.0 GCPs and the World30 DEM that has a locational accuracy of 10 m RMSE.

3.4 Statistical analysis

3.4.1 Random forest

Random forest is an accurate data mining technique that is designed to enhance the classification and regression trees (CART) method (Breiman, 2001). Simply, RF combines many regression trees to produce more accurate predictions (Cutler et al., 2007). These classifications are made using regression trees, hence the more trees built, classifications can be made more accurately. Bootstrap samples are created from many un-pruned trees (generally between 500 and 2000 trees) to create a final prediction that is taken from an average of all

individual tree outputs (Dye et al., 2011; Lawrence et al., 2006; Prasad et al., 2006). Trees are grown to a maximum length and based on the maximum number of votes that a class receives, the ensemble of trees assigns each waveband to a class (Adjorlolo et al., 2012a; Breiman, 2001). The addition of ancillary data (LAI; CHL; species type; SC) were used as additional input variables within the RF model. In this study, The RF algorithm was preferred due its robustness and accurate method for waveband selection specifically when using highly dimensional data as it does not over fit the data (Breiman, 2001; Dye et al., 2011; Mutanga et al., 2015; Peerbhay et al., 2015). Furthermore, RF is the most competitive and most reliable available algorithm compared to others in common use, specifically for ecological applications (Cutler et al., 2007). A more detailed description of the RF algorithm is provided in Chapter 2 of this thesis.

3.4.2 Stochastic gradient boosting algorithm and variable importance

Recently, remote sensing classification procedures are increasingly becoming useful. Hence, the ability to accurately detect features is achieved by producing better results. Classification trees are generally produced from two methods: (i) 'boosting methods' that are based on the results obtained from previous classification trees and (ii) 'bagging methods' that are based on the subsets of the training data (Lawrence *et al.*, 2004). Stochastic gradient boosting (SGB) is a relatively new method that is a hybrid of 'bagging' and 'boosting' approaches developed to improve classification accuracy (Dube *et al.*, 2015; Friedman *et al.*, 2000; Lawrence *et al.*, 2004; Ye *et al.*, 2009). SGB is a procedural based algorithm: Firstly, at each step of the boosting process, a random sample of the data is selected. Secondly, the random sample is separated by a steepest gradient algorithm, with gradient defined by deviance (twice the binomial negative log-likelihood) as a replacement for misclassification rates. This produces relatively small trees (6 terminal nodes are a common size) at each iteration instead of developing full classification trees at each stage. Lastly, all the small trees are stacked together as a weighted sum of terms (often 100-200 trees), and each observation is classified according to the most common classification among the trees (Dube *et al.*, 2015; Friedman, 2002; Lawrence *et al.*, 2004; Moisen, 2008). SGB has a greater accuracy in classifying data compared to other conventional classification techniques (Lawrence *et al.*, 2004; Moisen, 2008). It has been established that with each iteration, the model gets progressively stronger, hence in this study the model was run 100 times. (Elith *et al.*, 2008; Friedman *et al.*, 2000; Friedman, 2002; Moisen, 2008). Hence, as small classification trees are formed at each iteration of the boosting process, the

algorithm introduces an element of stochasticity that reduces overfitting, which improves the model accuracy, (Dube & Mutanga, 2015; Lawrence *et al.*, 2004). Furthermore, it is important that the algorithm is well parameterized according to the complexity of data used in the model.

There are three important user-defined parameters that govern the output of the SGB model: (i) The learning rate (r_1), (ii) tree complexity (c_t), (iii) the number of regression trees (n_{tree})(Dube *et al.*, 2015; Friedman, 2002; Hastie *et al.*, 2001). The most important parameters are r_1 and c_t, as r_1 determines the contribution of each tree has to the growing model, whereas c_t is the number of samples used in the final node (Dube *et al.*, 2015; Hastie *et al.*, 2001). The SGB algorithm was implemented (Venables & Smith, 2005) in R statistical software (Team, 2013). In this study, r_1 was set between 0.0001 and 0.1, whereas c_t was set between 1 and 5. Furthermore, during analysis a bagging fraction needs to be determined which was set to 0.3 in this study. The subsequent model settings where summarized in literature (De'Ath, 2007; Elith *et al.*, 2008; Friedman, 2002; Hastie *et al.*, 2001), hence formed suitable guidelines used in this study. In this study, SGB was the most preferred algorithm to be compared with RF in the prediction of fibre biochemicals NDF, ADF and Lignin, as it assumes to have a better predictive accuracy when using high resolution multispectral imagery. The addition of ancillary data (LAI; CHL; species type; SC) were used as additional input variables within the SGB model. More importantly, SGB provides information on the relative importance of the variables in predicting each class.

3.4.2 Variable importance

Stochastic gradient boosting variable importance utilizes variable selection by successively removing redundant predictor variables that do not improve the accuracies of the model parameterization (Dube *et al.*, 2015). As a result, individual predictor variables can be determined by measuring the proportion of variability accounted for by the splits (redundant and important variables) associated with each predictor variable (Moisen, 2008). The model then scales (i.e. so that the sum totals 100) the relative contribution of individual variables, the higher number indicate a strong influence on the response variable and vice versa (Dube *et al.*, 2015). In this study, variable relative importance was calculated as a percentage.

3.4.3 Accuracy assessment (R^2)

The R-square (R^2) (Albayrak, 2008; Dube *et al.*, 2015; Knox *et al.*, 2011; Pullanagari *et al.*, 2012) was the preferred accuracy assessment used to assess the performance of both the RF

and SGB method. The final dataset was split into 70% training and 30% test (Congalton & Green, 2008). In this study, the square of correlation between predicted versus observed values was calculated whereby values closer to 1, predict better results.

3.5 Results

Table 4, Table 5 and Table 6 displays summary statistical information related to the distribution of the chemical analysis and ancillary datasets. The results from the chemical analysis performed on grass samples collected in the field are presented in Table 4. Biochemical concentrations were compared with LAI measurements to ensure that predictions were related to only biochemical concentrations and not standing biomass (Knox *et al.*, 2011). Here, the quantity of biochemical concentration was measured as a percentage DM (Van Soest *et al.*, 1991). Table 5 shows LAI and CHL ancillary data collected in the field. Table 6 shows the three types of dominant grass species type and, the diversity (count) of species within the 140 samples collected in the field.

Table 4. Chemical analysis results obtained from 140 tropical mixed grass samples collected in the field

Nutrient	Mean (g/kg)	Range (g/kg)	Std. Dev. (g/kg)
NDF	127.48	116.48 – 193.16	6.623980369
ADF	91.80	85.06 – 100.86	3.427728864
Lignin	91.87	85.27 – 100.99	3.428042785

Table 5. Ancillary data obtained from 140 tropical mixed grass samples collected in the field

Variable	Abbr.	Unit	Mean	Std. Dev.	Min.	Max.
Leaf Area Index	LAI	$m^2\ m^{-2}$	1.52	0.93	0	3.68
Chlorophyll	CHL	SPAD	19.65	11.99	0.7	49.3

Table 6. Species type and species count out of the total number of samples obtained in the field

Species Type	*Tristachya leucothrix* (TL)	*Festuca costata* (FC)	*Themeda triandra* (TT)
Species Count	79 of 140	58 of 140	2 of 140

3.5.1 Predicting fibre biochemicals using RF

Results show, the RapidEye-5 multispectral sensor has the capability to successfully detect forage fibre biochemicals. More specifically, NDF was the highest predicted by RF with an R^2 of 0.74, whilst ADF produced an R^2 of 0.68 and Lignin produced the lowest R^2 of 0.67 (Figure 7).

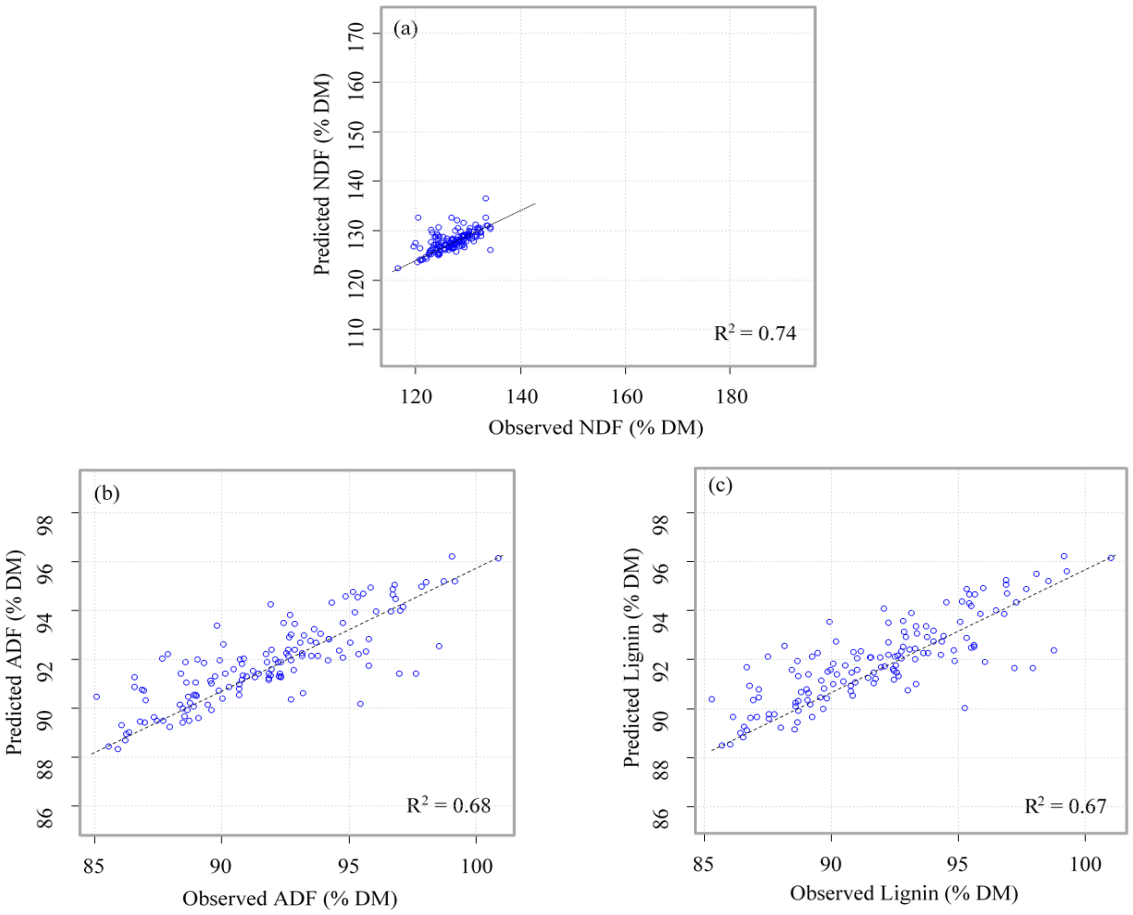


Figure 7. Random forest prediction of (a) NDF, (b) ADF, and (c) Lignin using RapidEye-5 waveband imagery and ancillary data

3.5.2 Predicting fibre biochemicals using SGB

Results show that the SGB model performed successfully when predicting the fibre biochemicals collected in field. More specifically, NDF was the highest predicted by SGB with an R^2 of 0.72, followed by ADF with an R^2 of 0.69 while lignin being the lowest predicted at 0.65 (Figure 8).

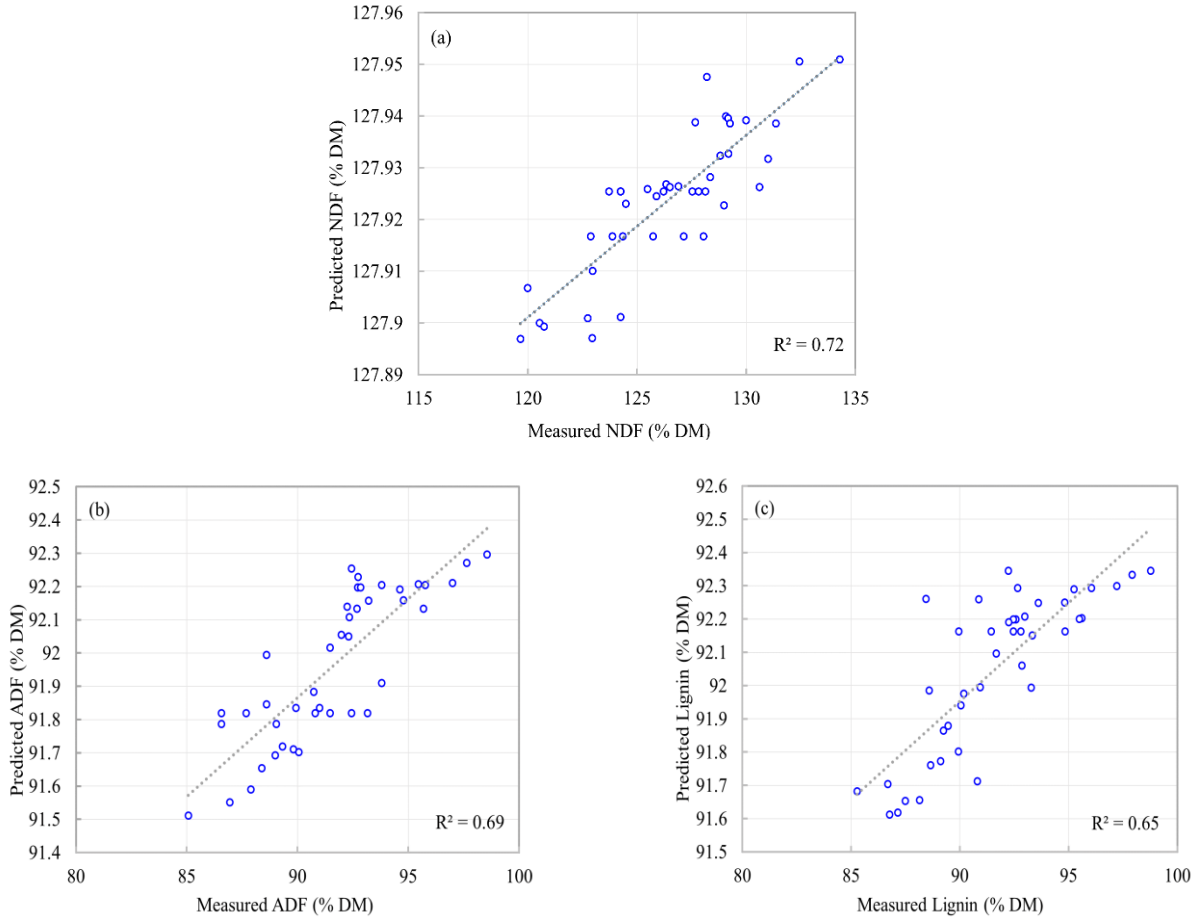


Figure 8. Stochastic gradient boosting prediction of (a) NDF, (b) ADF, and (c) Lignin using RapidEye-5 waveband imagery and ancillary data

Table 7. Main results achieved using RF and SGB modelling

Input data set	Statistical methods	NDF R^2	ADF R^2	Lignin R^2
All wavebands (with ancillary)	RF	0.74	0.68	0.67
All wavebands (with ancillary)	SGB	0.72	0.69	0.65
All wavebands (without ancillary)	RF	0.70	0.58	0.58

Table 7 indicates the performance of each statistical model when using the RapidEye-5 wavebands to predict and map all fibre biochemicals. More specifically, the performance of each model was calculated with and without using ancillary data for comparing whether the inclusion of ancillary data improves model performance. By comparing using RF shows that model performance does improve significantly when including ancillary data (Table 7). The model performance improved for all fibre biochemicals, ADF improved by 0.10 and Lignin improved by 0.09, whilst NDF improved by 0.04. Hence, there is a relationship with the fibre biochemicals and ancillary data used within this study.

3.5.3 Variable importance

Since the RF model combined with ancillary data produced the best accuracies for predicting NDF, ADF and Lignin (Table 7), this section only focuses on the most important variables selected within the RF model. In comparison, this section also assesses the variables selected within the SGB model. Figure 9 depicts the most effective and ineffective variables for detecting and predicting forage fibre biochemicals using the RapidEye-5 wavebands. Using RF, the Red, NIR, and Red-edge variables best predicted NDF (1a), ADF (1b) and Lignin (1c), hence produced correlations ranging between 40% and 80%, whilst CHL and LAI variables had weak correlations with all three fibre biochemicals ranging between -10% and -35%. Using SGB, species type, NIR and CHL variables best predicted ADF (2b) and Lignin (2c) ranging between 9% and 23%, whilst LAI, SC and NIR were better predictors for NDF (2a) ranging

between 7% and 27%. In contrast, the Blue and Green variables had weak correlations with ADF (2b) and Lignin (2c) ranging between 3% and 4.5%, whilst Red-edge, CHL and species type variables had weak correlations with NDF (2a) of 4% using SGB (Figure 9). Overall, when predicting ADF and Lignin similarities were found between the most important and unimportant variables, however when predicting NDF irregularities were found for both models (RF and SGB).

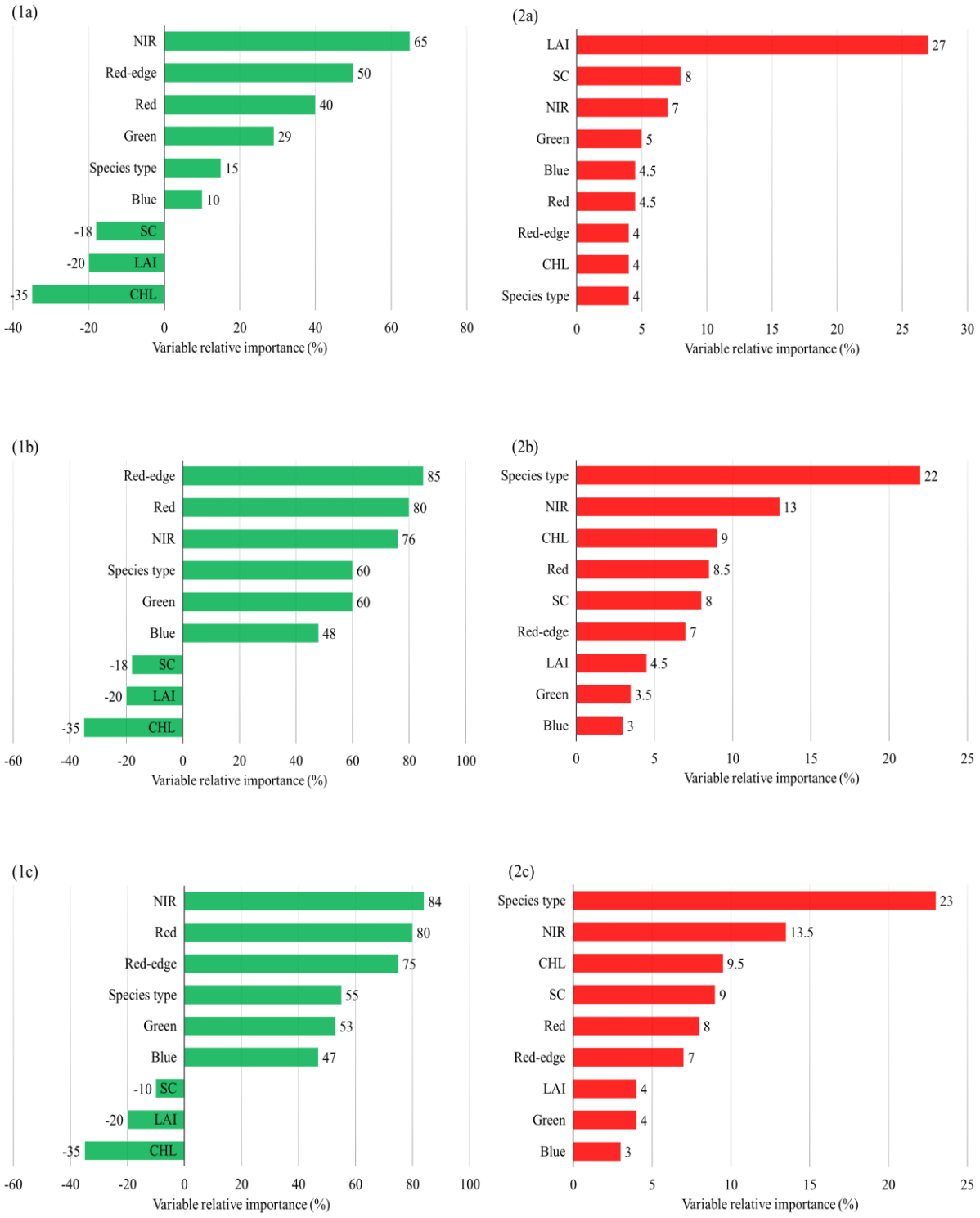


Figure 9. The relative importance of response variables using RapidEye-5 waveband imagery and ancillary data set (n = 9). The variables are ranked based on their contribution to the random forest and stochastic gradient boosting models. (1a), (1b), and (1c) represent the relative importance of the variables when predicting NDF, ADF, and Lignin, respectively using random forest. (2a), (2b), and (2c) represent the relative

importance of the variables when predicting NDF, ADF, and Lignin, respectively using stochastic gradient boosting

Figure 10 depicts prediction maps of the three key fibre biochemicals (NDF, ADF and Lignin) used in this study and that are primary determinants of grassland palatability. The maps show: low, intermediate and high levels of biochemical concentrations of all three biochemicals whereby green depicts low levels, yellow depicting intermediate levels and red depicting higher levels of biochemical concentration. Figure 10 (a) shows a clustered distribution of very high and very low levels of NDF concentration, whereas (b) ADF and (c) Lignin show a well distributed spread of high and low biochemical concentrations.

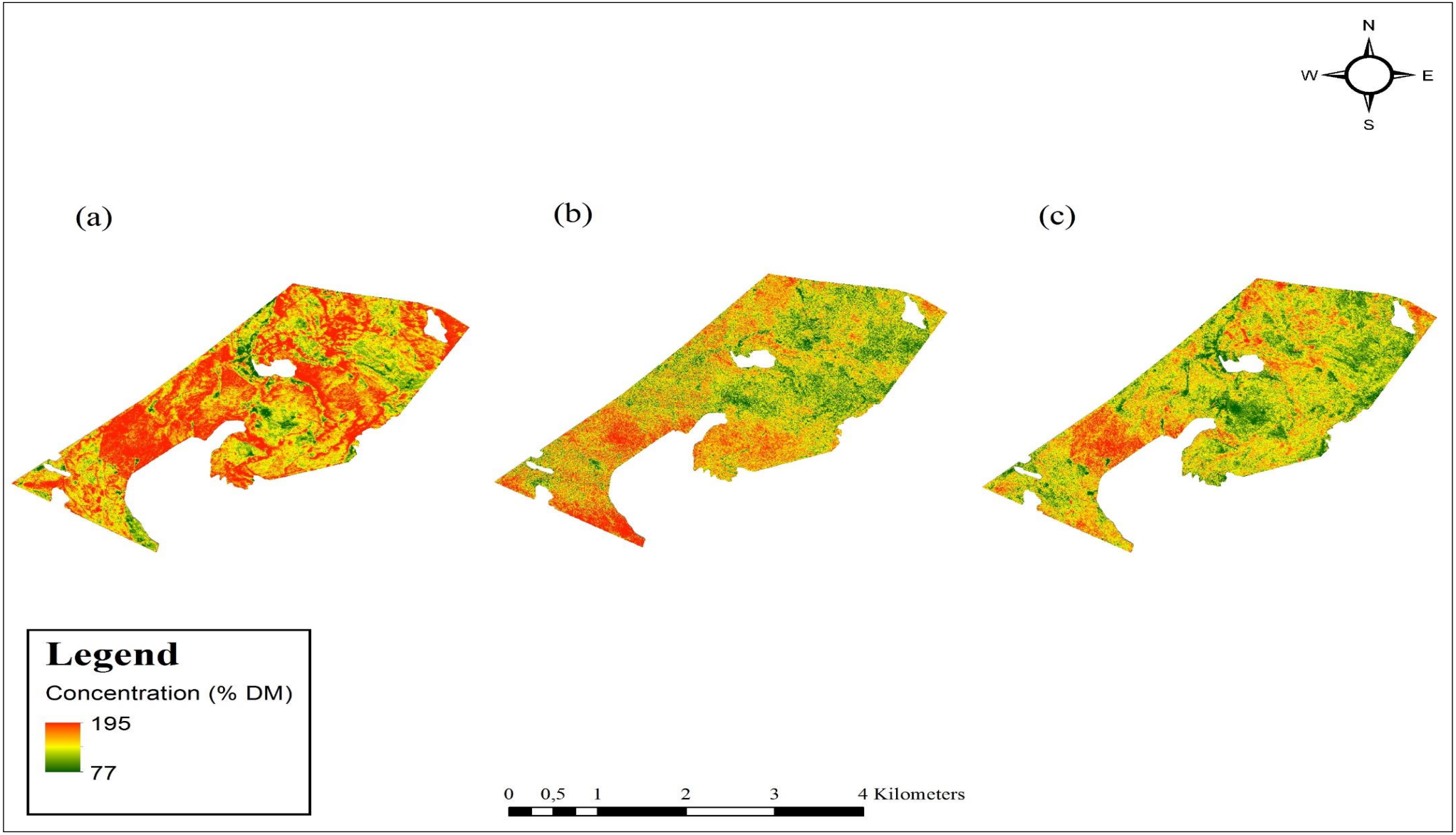


Figure 10. Prediction map of biochemical concentrations using RapidEye-5 in Fort Nottingham Nature Reserve, KwaZulu-Natal, South Africa

3.6 Discussion

Grasslands have a greater diversity than other ecosystems, offering complex challenges relating to species diversity, morphology, interactions between the grazing ungulates and including climatic factors (e.g. temperature, humidity, altitude, and rainfall) coupled with different management intensities (Adjorlolo *et al.*, 2012b; Darvishzadeh *et al.*, 2011; Pullanagari *et al.*, 2012). Understanding these challenges and then, to develop new models aimed at resolving some of these challenges is key for producing more plausible and reliable results (Schellberg *et al.*, 2008). Using a satellite sensor and a strategic modelling approach, this study assesses the capability of the RapidEye-5 multispectral sensor to effectively detect and map the content of important fibre biochemicals within a grassland ecosystem in KwaZulu-Natal, South Africa. Additionally, the utility of adding ancillary data collected in the field improves results and therefore, should always be considered an essential step for producing more effective and accurate predictions when mapping fibre biochemicals.

3.6.1 Prediction of fibre biochemicals using RF and SGB

In comparison to the RF model, the SGB model performed considerably well with a 0.01 increase in R^2 for ADF, however a 0.02 decrease in R^2 for NDF and Lignin effectively refutes the SGB model the best algorithm for application in this study. These results justify previous research recommendations of combining multispectral wavebands, ancillary data, and a algorithm to provide an ideal framework for mapping forage fibre biochemicals (Knox *et al.*, 2011; Ramoelo *et al.*, 2015a). It is important to note that selection of these two modelling algorithms was based on previous research recommendations. RF and SGB algorithms perform better than conventional regression models, such as linear, multiple-linear and stepwise regression models as they have elements of stochasticity that improve model accuracy and to reduce overfitting (Dube & Mutanga, 2015).

3.6.2 Combining ancillary data with multispectral imagery and specialized algorithms

RF and SGB successfully predicted NDF, ADF and Lignin with the inclusion of ancillary data. When examining the performance of the RapidEye-5 image with the inclusion of ancillary data: LAI, CHL, species count and species type, results indicate an improvement in final prediction. The combination of ancillary data with the multispectral imagery yielded better results compared to when executed separately (Table 7). Here, a 0.04 – 0.10% R^2 range of

improvement in accuracy was achieved when using ancillary data, deeming its contribution to the overall results substantially worthwhile. Knox *et al*., 2011 included ancillary data: phenology, soil type, species, and geology to map nitrogen, phosphorous and fibre in their study and found that it improved model predictions, however their study does not express and compare the actual increase in prediction accuracy. Similarly, Ramoelo *et al*., 2012 found a 0.27 R^2 increase in prediction accuracy when applying environmental ancillary data (mean annual temperature; precipitation; altitude; distance to rivers; slope and aspect) to their model to map biomass. Both studies have shown improvements in prediction accuracy when including ancillary data, hence the ancillary variables: LAI, CHL, species count and species type did well in predicting NDF, ADF and Lignin in this study.

3.6.3 Variable importance

Since the RF model produced the highest accuracies compared to SGB, the selection of the most important variables was determined using RF. The Red, Red-edge and NIR variables were the most effective contributors to the models performance and is confirmed by previous studies relating to forage quality biochemicals (Mutanga *et al.*, 2015; Pullanagari *et al.*, 2013b). In contrast, ancillary data variables: LAI and CHL were the lowest contributing variables to the overall model. To date, very few studies have successfully mapped forage fibre biochemicals NDF, ADF, and Lignin using high resolution multispectral imagery (Figure 10). Therefore, not many differences and similarities could be compared with the results found in this study. At this point, the most important wavebands discovered in this study should be taken into consideration when mapping forage fibre biochemicals in future studies. The remote sensing of these forage fibre biochemicals can contribute to understanding the impact of climatic change, as changes in fibre biochemical concentrations are influenced by climate change (Adjorlolo *et al.*, 2012b). Hence, this study shows the potential for retrieving information related to grassland quality in a timely and cost effective manner. This information could enable rangeland managers to influence policy and make current decisions regarding the health and conservation of grassland systems (Mansour *et al.*, 2016).

Future studies may also consider the possibilities of creating a remote sensing framework, that can be implemented as a toolkit for performing rapid spectroscopy on grasslands. This framework must be tested using other wavebands in other portions of the electromagnetic spectrum and in other locations and heterogeneous environments. Furthermore, freely available imagery such as Sentinel and Landsat can be used for mapping

at broader scales, however pixel resolution suffers due to increasing radiometric resolution and swath width with multispectral sensors. Very high resolution satellites such as IKONOS, QuickBird etc. supply under 10 cm pixel resolution imagery, capable of enhancing information at farm level for detailed analysis.

3.7 Conclusion

In this study, our main findings were:

- The results show that multispectral imagery is successful when mapping grassland biochemicals such as NDF, ADF and Lignin.
- The inclusion of ancillary data used in this study improves regression accuracy.
- The implementation of a data mining algorithms (RF or SGB) can detect NDF, ADF and Lignin and, avoids problems of multicollinearity and overfitting.

Overall this study provided an alternative framework to hyperspectral remote sensing for mapping important biochemicals across a grassland study site. Future studies should also investigate techniques for improving the prediction of vital biochemicals in grassland environments.

Chapter Four

Conclusion

4.1 Introduction

The mapping of forage quality nutrients is important for effective rangeland management practices, hence the information obtained is used to inform policy, rangeland managers and farmers (Skidmore *et al.*, 2010). Traditional methods commonly use wet chemistry and NIRS, however these methods are time consuming, costly and spatially invariable (Pullanagari *et al.*, 2013a). Remote sensing is a time effective and cost efficient method of gathering this information of a both local and broad scales, hence proposing an alternative for mapping forage quality nutrients. The aim of this research was to assess the utility of hyperspectral and multispectral remote sensing in predicting key forage quality nutrients (NDF, ADF and Lignin) using RF. The main objectives were: (i) to assess the capability of hyperspectral remote sensing in detecting three forage fibre quality nutrients (ii) to evaluate the effectiveness of using known absorption features for selection of the most optimal subsets of hyperspectral wavebands, (iii) to investigate the capability of high resolution RapidEye-5 multispectral imagery to detect and predict forage quality nutrients using a specialized algorithm and ancillary data, (iv) to test the capability of the SGB in identifying the most important wavebands for detecting forage quality nutrients, (v) to compare the respective capabilities of hyperspectral and multispectral remotely sensed data to detect and predict key forage quality nutrients. Each of these objectives will be discussed in the section below.

4.2 Assessing the capability of hyperspectral remote sensing in detecting forage fibre quality nutrients using RF.

Results from the study showed that the RF model that utilized all hyperspectral wavebands (n = 1910) successfully detected forage quality nutrients. More specifically, the detection of ADF and Lignin were successfully predicted, hence produced high R^2's of 88% and 87%, respectively, while NDF predicted a low R^2 of 47%. Here, majority of the RF model performed exceptionally well despite NDF predicting a slightly low result. This could be a result of a small range of variation (Zhao *et al.*, 2007). Overall, the results show that hyperspectral remote sensing in conjunction with a statistical approach (RF) can detect most of the forage quality nutrients present in this study by overcoming problems of dimensionality and multicollinearity.

4.3 Evaluating the effectiveness of using known absorption features for selection of the most optimal subsets of hyperspectral wavebands.

The study compared the use of known absorption features and using the entire hyperspectral dataset in detecting the most accurate results. When using the known absorption features dataset, this study produced slightly lower results than when using the entire hyperspectral dataset. Hence, the selection of the most optimal subsets of wavebands was taken from the variable importance of the entire hyperspectral dataset. It was found the detection of NDF was related to the blue, red and shortwave regions, and for ADF and Lignin wavebands only were related to the blue and shortwave regions of the electromagnetic spectrum. These findings are related to the findings of previous studies (Curran, 1989; Elvidge, 1990; Himmelsbach *et al.*, 1988; Kawamura *et al.*, 2008).

A final step was executing a combination of selected variable importance wavebands from using the entire dataset and known absorption features. Results showed an increase in R^2 of 0.03% for ADF and Lignin, however a decrease of 0.03% for NDF was predicted. In this study, it is shown that stronger predictions can be made using the known absorption features in combination with the most important wavebands derived from the original entire hyperspectral dataset. Hence, it can be concluded that the effectiveness of using known absorption features, when used in conjunction with the most important variables from the entire hyperspectral dataset and RF, improves overall results. This provides an ideal framework for reducing the number of wavebands in a hyperspectral dataset and can be used to determine the optimal subset of wavebands used to produce the highest prediction accuracies.

4.4 Investigating the capability of high resolution multispectral imagery to detect and predict forage quality nutrients using a specialized algorithm and ancillary data.

The results from this study confirm the effectiveness of RapidEye-5 multispectral data to accurately detect forage quality nutrients. Results show that the RapidEye-5 imagery produced R^2 predictions within the range of 0.67 and 0.74 using RF. Hence, these results confirm the effectiveness of the RapidEye-5 multispectral image sensor to detect and predict forage quality nutrients whilst dealing with the complexity of a grassland ecosystem. More specifically, results show the potential of the RapidEye-5 multispectral data to accurately detect forage quality nutrients (NDF, ADF and Lignin) in KwaZulu-Natal, South Africa.

4.5 Testing the capability of the SGB in identifying the most important wavebands for detecting forage quality nutrients.

This study compared the capability of both RF and SGB in detecting and predicting forage quality nutrients. It was envisaged that the SGB algorithm would produce better predictions than the RF algorithm, as SGB promises to yield better results when used ideally with high resolution data. However, the confidence of the RF algorithm is still prevalent as results show better predictions than when using the SGB algorithm. Hence, the predictions made using the RF algorithm was used to select the most important wavebands for detecting and predicting forage quality nutrients. Using RF wavebands in the Red, Red-edge and NIR regions most effectively contributed to the models performance, hence they were the most important wavebands used to predict forage quality nutrients (NDF, ADF and Lignin) and is confirmed by previous studies relating to forage quality biochemicals (Mutanga *et al.*, 2015; Pullanagari *et al.*, 2013b).

Furthermore, this study used ancillary data within the final prediction models to predict forage quality nutrients. Results indicate that the inclusion of ancillary data increases the performance of the overall model, hence a 0.04 to 0.10 range of increased R^2's was predicted for the forage quality nutrients in this study. Out of the four different types of ancillary data (LAI; CHL; species type; and SC) input into the final prediction model, species type predicted highest. More specifically, species type (*Festuca costata, Tristachya leucothrix,* and *Themeda triandra*) was the most important predictor ancillary data variable used to predict forage quality variables (NDF, ADF and Lignin) in this study. It can be concluded that the RF algorithm when used in conjunction with ancillary data, provides an ideal detection and prediction framework for successfully overcoming grassland complexities, which is a primary impediment in forage quality nutrient mapping.

4.6 A comparison between the respective capabilities of in-field hyperspectral and multispectral remotely sensed data to detect and predict key forage quality nutrients.

Overall, in-field hyperspectral remote sensing resulted in the best detection accuracies (ranging from 0.57 to 0.81 in R^2), outperforming the multispectral dataset image data (0.67 and 0.74 in R^2). However, it is important to understand the differences in the methodological approaches used for both in-field hyperspectral and multispectral remote sensing. The capabilities of satellite multispectral remote sensing used in this study, did exceptionally well to detect and map forage quality nutrients (NDF, ADF and Lignin). In literature, it is envisaged that in-field

hyperspectral remote sensing will produce more accurate results due to the proximity of the sensor to the grass leaf in the field. However, this is still a time consuming, expensive and laboring method.

It is important to note the complexity of the task when dealing with the large number of contiguous wavebands using in-field hyperspectral data that records between 350-2500 wavelengths, whilst satellite multispectral data only uses few wavelengths. Hence, computational time is reduced as well as problems of multicollinearity and dimensionality when using satellite multispectral data. A trade-off exists between the use of in-field hyperspectral and satellite multispectral data. In-field hyperspectral is more labour intensive and time consuming, however it produces more accurate results, whilst satellite multispectral remote sensing is inversely proportional. The differences in accuracy are not too far apart and predict successfully for both. Considering the two different methodological approaches, it was interesting to study the difference in detection accuracies and hence, it was important comparing the use of multispectral remote sensing using the RapidEye-5 image data, to in-field hyperspectral remote sensing.

4.7 Recommendations for future research

Several recommendations for future research are suggested below:

- Duplication of the same methodological approach over numerous locations and seasons, overall producing a synoptic scale map of forage quality nutrients in KwaZulu-Natal, South Africa, where land systems change is widespread. It will be interesting to see the change in biochemical concentrations over the different seasons. Hence, improving the efficacy of remote sensing in detecting and predicting forage quality nutrients and, the importance of informing policy, rangeland managers and farmers.
- Future research should improve prediction accuracies using very high resolution imagery with spatial resolutions below 5m such as IKONOS and QuickBird. Also, the use of freely available imagery such as Landsat and SPOT imagery in detecting forage quality nutrients, that will be inexpensive and useful to small scale farmers.
- Creating an easy to use operational framework for detecting and predicting forage quality nutrients using remote sensing for small and commercial farmers, conservation and policy makers. The production of forage quality nutrient maps will assist and inform these stakeholders to make better decisions regarding the effect of climate

change on forage quality, the effect of changing nutrient thresholds on grazing ungulates, and how these effects affect the policy and the economy.

4.8 Conclusion

The aim of this thesis was to assess the utility of remote sensing technologies to detect and predict forage quality nutrients. The research undertaken in this study showed that it is possible to detect and predict forage quality nutrients using both in-field hyperspectral and multispectral satellite imagery. The conclusion was based on the following observations as made in this dissertation:

1. In-field hyperspectral remote sensing is capable of accurately detecting forage quality nutrients. Results show that the random forest algorithm could discriminate each forage quality nutrient (NDF, ADF and Lignin). It is important to note the predictive ability of the RF algorithm and its ability to effectively reduce the number of redundant wavebands, which is necessary step when dealing with large amounts of data. Additional wavebands (blue, red and SWIR) were reported as effective regions for detection of forage quality nutrients in this study.
2. Multispectral remote sensing using a spatial resolution of 5m is capable of detecting and mapping forage quality nutrients. The RapidEye-5 image data successfully predicted and mapped NDF, ADF and Lignin with using RF and SGB algorithms. Both RF and SGB can discriminate the most important wavebands for the detection of forage quality nutrients. The inclusion of ancillary data should always be a vital step to yielding more accurate results.
3. The most important variables for the detection of forage fibre quality nutrients using multispectral satellite image data (NDF, ADF and Lignin) are prevalent in the red, red-edge and NIR regions of the electromagnetic spectrum.

REFERENCES

Abdel-Rahman, E. M., Ahmed, F. B., & Ismail, R. (2013). Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data. *International Journal of Remote Sensing, 34*(2), 712-728

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

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

Adjorlolo, C., Cho, M. A., Mutanga, O., & Ismail, R. (2012a). Optimizing spectral resolutions for the classification of C3 and C4 grass species, using wavelengths of known absorption features. *Journal of Applied Remote Sensing, 6*(1), 063560-063561-063560-063515

Adjorlolo, C., Mutanga, O., Cho, M., & Ismail, R. (2012b). Challenges and opportunities in the use of remote sensing for C3 and C4 grass species discrimination and mapping. *African Journal of Range & Forage Science, 29*(2), 47-61

Adjorlolo, C., Mutanga, O., Cho, M. A., & Ismail, R. (2013). Spectral resampling based on user-defined inter-band correlation filter: C 3 and C 4 grass species classification. *International Journal of Applied Earth Observation and Geoinformation, 21*, 535-544

Albayrak, S. (2008). Use of reflectance measurements for the detection of N, P, K, ADF and NDF contents in sainfoin pasture. *Sensors, 8*(11), 7275-7286

ASD. (2008). FieldSpec® 3 User Manual, 1-106

Ash, A. J., Corfield, J. P., McIvor, J. G., & Ksiksi, T. S. (2011). Grazing management in tropical savannas: utilization and rest strategies to manipulate rangeland condition. *Rangeland Ecology & Management, 64*(3), 223-239

Boval, M., & Dixon, R. (2012). The importance of grasslands for animal production and other functions: a review on management and methodological progress in the tropics. *Animal, 6*(05), 748-762

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

Breiman, L., Friedman, J., OHLSEN, R., & Stone, C. (1984). Classification and Regression Trees (Belmont: Wadsworth International Group). *Clark, RG, Henderson, HV, Hoggard, GK, Ellison, RS, and Young, BJ (1987),"The Ability of Biochemical and Haematological Tests to Predict Recovery in Periparturient Recumbent Cows," New Zealand Veterinary Journal, 35*, 126-133

Clevers, J. G., & 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*, 344-351

Clevers, J. G., & Kooistra, L. (2012). Using hyperspectral remote sensing data for retrieving canopy chlorophyll and nitrogen content. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of, 5*(2), 574-583

Congalton, R. G., & Green, K. (2008). *Assessing the accuracy of remotely sensed data: principles and practices*: CRC press.

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

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

Dabasso, B. H., Oba, G., & Roba, H. G. (2012). Livestock-based knowledge of rangeland quality assessment and monitoring at landscape level among borana herders of northern Kenya. *Pastoralism, 2*(1), 1-13

Darvishzadeh, R., Atzberger, C., Skidmore, A., & Schlerf, M. (2011). Mapping grassland leaf area index with airborne hyperspectral imagery: A comparison study of statistical approaches and inversion of radiative transfer models. *ISPRS Journal of Photogrammetry and Remote Sensing, 66*(6), 894-906

De'Ath, G. (2007). Boosted trees for ecological modeling and prediction. *Ecology, 88*(1), 243-251

Delegido, J., Verrelst, J., Alonso, L., & Moreno, J. (2011). Evaluation of sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content. *Sensors, 11*(7), 7063-7081

Dube, T., & Mutanga, O. (2015). Evaluating the utility of the medium-spatial resolution Landsat 8 multispectral sensor in quantifying aboveground biomass in uMgeni

catchment, South Africa. *ISPRS Journal of Photogrammetry and Remote Sensing, 101*, 36-46

Dube, T., Mutanga, O., Abdel-Rahman, E. M., Ismail, R., & Slotow, R. (2015). Predicting Eucalyptus spp. stand volume in Zululand, South Africa: an analysis using a stochastic gradient boosting regression ensemble with multi-source data sets. *International Journal of Remote Sensing, 36*(14), 3751-3772

Dye, M., Mutanga, O., & Ismail, R. (2011). Examining the utility of random forest and AISA Eagle hyperspectral image data to predict Pinus patula age in KwaZulu-Natal, South Africa. *Geocarto International, 26*(4), 275-289

Eitel, J., Long, D., Gessler, P., & Smith, A. (2007). Using in-situ measurements to evaluate the new RapidEye™ satellite series for prediction of wheat nitrogen status. *International Journal of Remote Sensing, 28*(18), 4183-4190

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology, 77*(4), 802-813

Elvidge, C. D. (1990). Visible and near infrared reflectance characteristics of dry plant materials. *Remote Sensing, 11*(10), 1775-1795

ENVI. (2009). *ENVI Reference Guide*. [ENVI Version 4.7 August, 2009 Edition]. Retrieved from http://www.harrisgeospatial.com/portals/0/pdfs/envi/Reference_Guide.pdf

Fourty, T., Baret, F., Jacquemoud, S., Schmuck, G., & Verdebout, J. (1996). Leaf optical properties with explicit description of its biochemical composition: direct and inverse problems. *Remote sensing of Environment, 56*(2), 104-117

Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The annals of statistics, 28*(2), 337-407

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis, 38*(4), 367-378

Friedman, J. M., Hunt, E. R., & Mutters, R. G. (2016). Assessment of Leaf Color Chart Observations for Estimating Maize Chlorophyll Content by Analysis of Digital Photographs. *Agronomy Journal, 108*(2), 822-829

Fynn, R. W., Augustine, D. J., Peel, M. J., & Garine-Wichatitsky, M. (2015). Strategic management of livestock to improve biodiversity conservation in African savannahs: A conceptual basis for wildlife–livestock co-existence. *Journal of Applied Ecology*

Georgiadis, N. J., & McNaughton, S. J. (1990). Elemental and fibre contents of savanna grasses: variation with grazing, soil type, season and species. *Journal of Applied Ecology*, 623-634

Hastie, T., Tibshirani, R., & Friedman, J. (2001). The elements of statistical learning: data mining, inference and prediction. *New York: Springer-Verlag, 1*(8), 371-406

Himmelsbach, D., Boer, H., Akin, D., & Barton, F. (1988). Solid-state 13C NMR, FTIR, and NIRS spectroscopic studies of ruminant silage digestion. *Analytical applications of spectroscopy/edited by CS Creaser and AMC Davies*

Jančík, F., Koukolová, V., Homolka, P., & Haman, J. (2011). Comparison of analyses to predict ruminal fibre degradability and indigestible fibre in temperate grass silages. *South African Journal of Animal Science, 41*(3), 297-308

Jusoff, K. (2007). Advanced processing of UPM-APSB's AISA airborne hyperspectral images for individual timber species identification and mapping. *International Journal of Systems Applications, Engineering & Development, 2*(1), 21-26

Kawamura, K., Watanabe, N., Sakanoue, S., & Inoue, Y. (2008). Estimating forage biomass and quality in a mixed sown pasture based on partial least squares regression with waveband selection. *Grassland Science, 54*(3), 131-145

Kawamura, K., Watanabe, N., Sakanoue, S., Lee, H. J., Inoue, Y., & Odagawa, S. (2010). Testing genetic algorithm as a tool to select relevant wavebands from field hyperspectral data for estimating pasture mass and quality in a mixed sown pasture using partial least squares regression. *Grassland science, 56*(4), 205-216

Knox, N. M., Skidmore, A. K., Prins, H. H., Asner, G. P., van der Werff, H. M., de Boer, W. F., . . . Slotow, R. (2011). Dry season mapping of savanna forage quality, using the hyperspectral Carnegie Airborne Observatory sensor. *Remote Sensing of Environment, 115*(6), 1478-1488

Knox, N. M., Skidmore, A. K., Prins, H. H., Heitkönig, I. M., Slotow, R., van der Waal, C., & de Boer, W. F. (2012). Remote sensing of forage nutrients: Combining ecological and spectral absorption feature data. *ISPRS Journal of Photogrammetry and Remote Sensing, 72*, 27-35

Kokaly, R. F., & Clark, R. N. (1999). Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote sensing of environment, 67*(3), 267-287

Kuemmerle, T., Erb, K., Meyfroidt, P., Müller, D., Verburg, P. H., Estel, S., . . . Kastner, T. (2013). Challenges and opportunities in mapping land use intensity globally. *Current opinion in environmental sustainability, 5*(5), 484-493

Kumar, L., Schmidt, K., Dury, S., & Skidmore, A. (2002). Imaging spectrometry and vegetation science *Imaging spectrometry* (pp. 111-155): Springer.

Lamarque, P., Tappeiner, U., Turner, C., Steinbacher, M., Bardgett, R. D., Szukics, U., . . . Lavorel, S. (2011). Stakeholder perceptions of grassland ecosystem services in relation to knowledge on soil fertility and biodiversity. *Regional Environmental Change, 11*(4), 791-804

Laurin, G. V., Puletti, N., Hawthorne, W., Liesenberg, V., Corona, P., Papale, D., . . . Valentini, R. (2016). Discrimination of tropical forest types, dominant species, and mapping of functional guilds by hyperspectral and simulated multispectral Sentinel-2 data. *Remote Sensing of Environment, 176*, 163-176

Lawrence, R., Bunn, A., Powell, S., & Zambon, M. (2004). Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. *Remote sensing of environment, 90*(3), 331-336

Lawrence, R. L., Wood, S. D., & Sheley, R. L. (2006). Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest). *Remote Sensing of Environment, 100*(3), 356-362

Lawrence, R. L., & Wright, A. (2001). Rule-based classification systems using classification and regression tree (CART) analysis. *Photogrammetric engineering and remote sensing, 67*(10), 1137-1142

LI-COR, I. (2009). *LAI-2200 Plant Canopy Analyzer Instruction Manual*. Retrieved from https://www.licor.com/documents/6n3conpja6uj9aq1ruyn

Loucougaray, G., Dobremez, L., Gos, P., Pauthenet, Y., Nettier, B., & Lavorel, S. (2015). Assessing the effects of grassland management on forage production and environmental quality to identify paths to ecological intensification in mountain grasslands. *environmental Management, 56*(5), 1039-1052

Mansour, K., Everson, T., & Mutanga, O. (2013). Evaluation of potential indicators for payment of environmental services on the impact of rehabilitation of degraded rangeland sites. *African Journal of Agricultural, 8*(15), 1290-1299

Mansour, K., Mutanga, O., Adam, E., & Abdel-Rahman, E. M. (2016). Multispectral remote sensing for mapping grassland degradation using the key indicators of grass species and edaphic factors. *Geocarto International, 31*(5), 477-491

Mbatha, K. R., & Ward, D. (2010). The effects of grazing, fire, nitrogen and water availability on nutritional quality of grass in semi-arid savanna, South Africa. *Journal of Arid Environments, 74*(10), 1294-1301

Moisen, G. (2008). Classification and regression trees

Muchenje, V., Dzama, K., Chimonyo, M., Raats, J., & Strydom, P. (2008). Meat quality of Nguni, Bonsmara and Aberdeen Angus steers raised on natural pasture in the Eastern Cape, South Africa. *Meat Science, 79*(1), 20-28

Mucina, L., & Rutherford, M. (2006). Strelitzia 19. *South Afircan National Bioidiversity Institute, Pretoria*

Mulla, D. J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering, 114*(4), 358-371

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

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

Mutanga, O., Dube, T., & Ahmed, F. (2016). Progress in remote sensing: vegetation monitoring in South Africa. *South AfricAn GeoGrAphicAl JournAl, 98*(3), 461-471

Mutanga, O., & Kumar, L. (2007). Estimating and mapping grass phosphorus concentration in an African savanna using hyperspectral image data. *International Journal of Remote Sensing, 28*(21), 4897-4911

Mutanga, O., Skidmore, A., & Prins, H. T. (2004a). Discriminating sodium concentration in a mixed grass species environment of the Kruger National Park using field spectrometry. *International journal of remote sensing, 25*(20), 4191-4201

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

Norris, K., Barnes, R., Moore, J., & Shenk, J. (1976). Predicting forage quality by infrared replectance spectroscopy. *Journal of animal science, 43*(4), 889-897

Osborne, B. G., & Fearn, T. (1986). *Near infrared spectroscopy in food analysis*: Longman.

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

Peerbhay, K. Y., Mutanga, O., & Ismail, R. (2015). Random Forests Unsupervised Classification: The Detection and Mapping of Solanum mauritianum Infestations in Plantation Forestry Using Hyperspectral Data

Pellissier, P. A., Ollinger, S. V., Lepine, L. C., Palace, M. W., & McDowell, W. H. (2015). Remote sensing of foliar nitrogen in cultivated grasslands of human dominated landscapes. *Remote Sensing of Environment, 167*, 88-97

Prasad, A. M., Iverson, L. R., & Liaw, A. (2006). Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems, 9*(2), 181-199

Pullanagari, R., Yule, I., Tuohy, M., Hedley, M., Dynes, R., & King, W. (2012). In-field hyperspectral proximal sensing for estimating quality parameters of mixed pasture. *Precision Agriculture, 13*(3), 351-369

Pullanagari, R., Yule, I., Tuohy, M., Hedley, M., Dynes, R., & King, W. (2013a). Proximal sensing of the seasonal variability of pasture nutritive value using multispectral radiometry. *Grass and Forage Science, 68*(1), 110-119

Pullanagari, R. R., Dynes, R. A., King, W. M., Yule, I. J., Thulin, S., Knox, N. M., . . . Badgery, W. (2013b). Remote sensing of pasture quality *Revitalising Grasslands to Sustain our Communities: Proceedings of the 22nd International Grassland Congress, Sydney, Australia* (pp. 633-638).

Ramoelo, A., Cho, M., Mathieu, R., Madonsela, S., Van De Kerchove, R., Kaszta, Z., & Wolff, E. (2015a). Monitoring grass nutrients and biomass as indicators of rangeland quality and quantity using random forest modelling and WorldView-2 data. *International Journal of Applied Earth Observation and Geoinformation, 43*, 43-54

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

Ramoelo, A., Cho, M., Mathieu, R., & Skidmore, A. K. (2015b). Potential of Sentinel-2 spectral configuration to assess rangeland quality. *JOURNAL OF APPLIED REMOTE SENSING, 9*

Ramoelo, A., Skidmore, A. K., Cho, M. A., Schlerf, M., Mathieu, R., & Heitkönig, I. M. (2012b). 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*, 151-162

RapidEye, A. (2011). Satellite imagery product specifications: Version.

Schellberg, J., Hill, M. J., Gerhards, R., Rothmund, M., & Braun, M. (2008). Precision agriculture on grassland: Applications, perspectives and constraints. *European Journal of Agronomy, 29*(2), 59-71

Schuster, C., Förster, M., & Kleinschmit, B. (2012). Testing the red edge channel for improving land-use classifications based on high-resolution multi-spectral satellite data. *International Journal of Remote Sensing, 33*(17), 5583-5599

Skidmore, A. K., Ferwerda, J. G., Mutanga, O., Van Wieren, S. E., Peel, M., Grant, R. C., . . . 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), 64-72

Starks, P. J., Coleman, S. W., & Phillips, W. A. (2004). Determination of forage chemical composition using remote sensing. *Journal of range management, 57*(6), 635-640

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC bioinformatics, 9*(1), 307

Stuth, J., Jama, A., & Tolleson, D. (2003). Direct and indirect means of predicting forage quality through near infrared reflectance spectroscopy. *Field Crops Research, 84*(1), 45-56

Swanepoel, P. A., Sanderson, M. A., Bell, L. W., Thomas, D. T., & Bennett, J. E. (2015). Southern African pasture science in the 21st century. *African Journal of Range & Forage Science, 32*(2), iii-iv

Team, R. C. (2013). R: A language and environment for statistical computing

Teoh, C. C., Daud, A. H., Mispan, M. R., & Jiken, J. J. (2015). Prediction of SPAD chlorophyll meter readings using remote sensing technique. *Journal of Tropical Agriculture and Food Science, 40*, 127-136

Ullah, S., Si, Y., Schlerf, M., Skidmore, A. K., Shafique, M., & Iqbal, I. A. (2012). Estimation of grassland biomass and nitrogen using MERIS data. *International journal of applied earth observation and geoinformation, 19*, 196-204

Van Soest, P. v., Robertson, J., & Lewis, B. (1991). Methods for dietary fiber, neutral detergent fiber, and nonstarch polysaccharides in relation to animal nutrition. *Journal of dairy science, 74*(10), 3583-3597

Venables, W. N., & Smith, D. M. (2005). the R development core team. *An Introduction to R. Notes on R: A Programming Environment for Data Analysis and Graphics*

WWF. (Undated). *Covering nearly a third of South Africa, the Grasslands region is the largest of South Africa's biomes, and its levels of biodiversity are only second to that of the Cape Floristic Region.* Retrieved from http://www.wwf.org.za/what_we_do/land/grasslands/

Ye, J., Chow, J.-H., Chen, J., & Zheng, Z. (2009). Stochastic gradient boosted distributed decision trees *Proceedings of the 18th ACM conference on Information and knowledge management* (pp. 2061-2064): ACM.

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

Zhao, D., Starks, P. J., Brown, M. A., Phillips, W. A., & Coleman, S. W. (2007). Assessment of forage biomass and quality parameters of bermudagrass using proximal sensing of pasture canopy reflectance. *Grassland Science, 53*(1), 39-49