REMOTE SENSING OF THE DISTRIBUTION AND QUALITY OF SUBTROPICAL C3 AND C4 GRASSES

A thesis submitted in fulfillment of the academic requirements

for the degree of Doctor of Philosophy (PhD) in Environmental Science

Clement Adjorlolo

School of Agricultural, Earth and Environmental Sciences

University of KwaZulu-Natal,

Pietermaritzburg, South Africa

March 2013

PREFACE

I, **CLEMENT ADJORLOLO,** declare that:

(i) The research reported in this thesis, except where otherwise indicated, is my original work.

(ii) This thesis has not been submitted for any degree or examination at any other university.

(iii) This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

(iv) This thesis 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 reproduced a publication of which I am an author, co-author or editor, I have indicated, in-detail, which part of the publication was actually written by myself alone and have fully referenced such publications.

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

Signed:

Supervisor:

Professor O. Mutanga

Co-supervisor:

Dr. M.A. Cho

PUBLICATION AND MANUSCRIPTS

The following published papers and manuscripts include other authors. My contribution was greatest and appropriate to being the first author in all cases and in the order that they are presented.

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

Adjorlolo, C., Mutanga, O., Cho, M.A., & 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, 47-61

Adjorlolo, C., Mutanga, O., Cho, A.M., & Ismail, R. (**2013**). Spectral resampling based on user-defined inter-band correlation filter: C3 and C4 grass species classification. *International Journal of Applied Earth Observation and Geoinformation*, 21, 535-544

Adjorlolo, C., Mutanga, O., & Cho, A.M. (**In preparation**). Mapping and characterizing C3 and C4 grass distribution in the subtropical montane rangeland of Southern Africa. *ISPRS Journal of Photogrammetry and Remote Sensing*

Adjorlolo, C., & Mutanga, O. (**In review**). Predicting C3 and C4 Grass Nutrient Variability, using in-situ Canopy Reflectance and Partial Least Squares Regression. *International Journal of Remote sensing*

Adjorlolo, C., Mutanga, O., & Cho, A.M. (**In review**). Predicting canopy nitrogen and fibre variability across C3 and C4 grasslands, using new generation multispectral imagery. *Remote Sensing of Environment*

Key

In review– paper submitted to a refereed journal and is currently with the anonymous reviewers.

In preparation – papers written by the authors and ready to be submitted to the respective journals.

ABSTRACT

Global climate change is expected to be accompanied by changes in the composition of plant functional types. Such changes are predicted to follow shifts in the percentage cover and abundance of grass species, following the C3 and C4 photosynthetic pathways. These two groups differ in a number of physiological, structural and biochemical aspects. It is important to measure these characteristic properties because they affect ecosystem processes, such as nutrient cycling. High spectral and spatial resolution remote sensing systems have been proven to offer data, which can be used to accurately detect, classify and map plant species. The major challenge, however, is that the spectral reflectance data obtained over many narrow contiguous channels (i.e. hyperspectral data) represent multiple classes that are often mixed for a limited training-sample size. This is commonly referred to as the Hughes phenomenon or "the curse of dimensionality". In the context of hyperspectral data analysis, the Hughes phenomenon often introduces a high degree of multicollinearity, which is caused by the use of highly-correlated spectral predictors. Multicollinearity is a prominent problem in processing hyperspectral data for vegetation applications, due to similarities in the spectral reflectance properties of biophysical and biochemical attributes. This study explored an innovative method to solve the problems associated with spectral dimensionality and the related multicollinearity, by developing a user-defined inter-band correlation filter function to resample hyperspectral data. The proposed resampling technique convolves the spectral dependence information between a chosen band-centre and its shorter and longer wavelength neighbours. The utility of the new resampling technique was assessed for discriminating C3 (*Festuca costata*) and C4 (*Themeda triandra* and *Rendlia altera*) grasses and for predicting their nutrient content (nitrogen, protein, moisture, and fibre), using partial least squares and random forest regressions. In general, results obtained showed that the user-defined inter-band correlation filter technique can mitigate the problem of multicollinearity in both classification and regression analyses. Wavebands in the shortwave infrared region were found to be very important in regression and classification analyses, using field spectra-only datasets. Next, the analyses were up-scaled from field spectra to the new generation multispectral satellite, WorldView-2 imagery, which was acquired for the Cathedral Peak region of the Drakensberg Mountains. The results obtained, showed that the WV2 image data contain useful information for classifying the C3 and C4 grasses and for predicting variability in their nitrogen and fibre concentrations. This study makes a contribution by developing a user-defined inter-band correlation filter to resample hyperspectral data, and thereby mitigating the high dimensionality and multicollinearity problems, in remote sensing applications involving C3 and C4 grass species or communities.

DEDICATION

To my grandparents, Matilda Nyeha and Joseph Dormefa Adjorlolo. I will never forget the values they taught me: their wisdom and virtue of humility will be always part of my life.

To my dearly loved wife, Zeria, and my sons, Selorm and Mawuli.

ACKNOWLEDGEMENTS

I honour You, God Almighty for a day, such as this and for bringing me this far. I would like to express my gratitude to the South African National Research Foundation (NRF) who provided me with support. I sincerely thank my supervisor and co-supervisor, Professor Onisimo Mutanga and Dr. Moses A. Cho (CSIR- Council for Scientific and Industrial Research), respectively, for their confidence, scientific guidance, commitment, critical comments and moral support during my PhD, and most of all, for being such wonderful friends in both academic and non-academic affairs. To Dr. Riyad Ismail and Dr. Elhadi Adam, I really appreciated all the effort you dedicated to my research, my thesis would not be the same without your ideas and comments. Many thanks to my colleagues, Natural Resources staff of the KwaZulu-Natal (KZN) Department of Agriculture and Environmental Affairs (DAEA) for their absolute support, particularly with fieldwork and specialist advice. I would like to thank Ezemvelo KZN Wildlife for providing me with the necessary permit to conduct this research within the Cathedral Peak World Heritage site. I would like to extend my gratitude to the Soil Analytical Services of the KZN DAEA, Cedara, for analyzing soil and plant samples. To Professors Trevor Hill and Heinz Beckedahl, I thank you for affording me the opportunity to study at the University of KwaZulu-Natal for the first time in 2005. I am deeply grateful for the confidence you have shown in me. I thank all staff and post-graduate students from the Department of Geography: Mrs. Shanitha Ramroop, Donavan de Vos, Victor Bangamwabo, Brice Gijsbertsen, Zakariyyaa Oumar, Khalid Manssour and Liandra, for their assistance and friendship throughout my study. My appreciation extends to Mrs. Sharon Rees for proofreading all the stand alone peer review papers, as well as the whole thesis. To all my friends in South Africa, Ghana, Nigeria, Zimbabwe, Cameron, Kenya, Tanzania, Zambia and the United Kingdom, it was a pleasure to have spent time with you. My immense gratitude to my family in Ghana, South Africa and Zambia for their support and encouragement: you all made me believed that a day such as this would come. And most of all, to my wife, Zeria. I will be always glad that the LOVE of GOD binds us together. Thank you for your love, for being at my side, for being such a caring person. And to my two very smart, lovely boys, Selorm and Mawuli: life has never been the same, because you have brought such joy to my life, and it is much more colourful with you around.

TABLE OF CONTENTS

CHAPTER 1: GENERAL INTRODUCTION

This Chapter provides a synoptic view of the research background and outlines the objectives and the structure of the thesis.

1.1 Remote sensing of plant functional types

Ecologists and plant geographers have historically described patterns and relationships between different groups of plants and their environment. Such groupings have been based on the functional type attributes of plants that play a common ecological role in ecosystems (Asner et al. 1998b). The functional type characteristics include distinct physiological, structural and/or biochemical properties of the plant species. The remote sensing measurement capabilities has provided the opportunity to classify and spatially characterize plant functional types (Ustin and Gamon 2010). Specifically, remote sensing data have been used to assess the composition of grass species following the C3 and C4 photosynthetic pathways (e.g. Goodin and Henebry, 1997; Davidson and Csillag, 2001; Foody and Dash, 2007). The C3 or C4 type of photosynthesis primarily produces 3-carbon and 4-carbon compounds, respectively. The two groups of grasses have different environmental requirements for growth and often exhibit asynchronous seasonal phenology, mainly in temperate ecosystems. The asynchronous seasonal profile exhibited by C3 and C4 grasses and related biomass production affects the spectral response of grasslands (Davidson and Csillag 2003). Since landscapes with the C3 and C4 grass composition are sensitive to even subtle changes in environmental conditions, the use of remote sensing methods to detect and monitor such changes should play a critical role in the management of landscapes, with specific reference to global climate change.

1.2 Role of remote sensing in vegetation-climate studies

There is strong evidence that global climate change and the associated elevated levels of greenhouse gases, particularly CO_2, affect the functioning of terrestrial ecosystems (Rock et al. 1988; Vandewalle et al. 2010; Winslow et al. 2003). Such changes could manifest as shifts in the composition of C3 and C4 grassland communities, with each species responding on the basis of its ecological adaptability (Bremond et al. 2012; Lattanzi 2010). On the other hand, the observed distribution and abundance of C3 and C4 grasses have been found to vary geographically (Bremond et al. 2012; Ehleringer et al. 2001). In addition, the growth responses of these two groups have been strongly correlated with changes in environmental resources, such as soil moisture content and nutrient cycling (Ehleringer and Monson 1993; Tieszen et al. 1979; Ustin and Gamon 2010). In this context, the use of remote sensing techniques to investigate C3 and C4 grass characteristics in relation to environmental factors, has been predominantly done in the

high latitude temperate regions, with limited focus on the subtropics (Davidson and Csillag 2001, 2003; Foody and Dash 2007, 2010; Goodin and Henebry 1997; Guan et al. 2012; Liu and Cheng 2011; Pearcy and Ehleringer 1984; Ricotta et al. 2003; Tieszen et al. 1997).

Climate change is predicted to result in a warmer and wetter environment along the eastern seaboard of Southern Africa, including the subtropical montane region of the KwaZulu-Natal Province (Hewitson et al. 2005). If this holds, the composition of C3 and C4 grass species in this region will be affected. Generally, the C3 type of grasses show higher positive responses to elevated carbon dioxide and low levels of warming, while the C4 type shows higher positive responses to higher levels of warming (Bremond et al. 2012; Lattanzi 2010). There is an increasing evidences that extra warming often negatively affect growth yields and forage quality (Chen et al. 1996). Forage quality can differ between these two groups of species (Barbehenn et al. 2004). In general, the C3 and C4 type of grasses differ in their forage nutrient status and will respond differently to changing climatic factors (Barbehenn et al. 2004; von Fischer et al. 2008). Important variables include limiting forage nutrients such as nitrogen crude protein, moisture, and nutrients (non-digestible fibre) that constrain the intake rate of herbivores (Grant et al. 2010; Knox et al. 2011). Quality in rangelands generally declines with increasing carbon dioxide concentration because of the effects on plant nitrogen and protein content (Sage and Pearcy 1987). The anticipated increasing CO_2 concentrations in the Southern African rangelands (Chamaillé-Jammes and Bond 2010; Stock et al. 2004) will favour undesirable C3 plants which thrive in soil and air with elevated carbon dioxide in the region (Adjorlolo et al. 2012b).

At the global scale, warming will cause the optimum latitude for C3 grass plants to move northward from the equator while the vice versa is known for C4 grasses (Bremond et al. 2012; Lattanzi 2010). At the scale of a region, the composition of plant species (specifically grass species) is predicted mostly by factors such as temperature, precipitation (water) and soils. Water is the primary factor controlling the distribution and abundance of plants: both the amount of water plants use and water availability over time and space (Winslow et al. 2003). The ability to determine changes in vegetation at local scales and over shorter periods is however, limited (Wand et al. 1999). This is because at these scales the response of vegetation to global- and regional-scale changes depends on a variety of local processes including the rate of disturbances such as fire and grazing (Winslow et al. 2003), and the rate at which different plant species adapt across complex and sometimes-fragmented landscapes (Lattanzi 2010).

Remote sensing applications offer data that can be used to study the composition and dynamics of C3 and C4 grasses, considering the challenges of field-based surveys, specifically in montane landscapes. The major challenges in the use of remote sensing for C3 and C4 grass species discrimination and mapping in mountainous environments are: (i) the high frequency of cloud cover, which limits low temporal resolution space-borne sensors, (ii) variations in the anisotropic reflectance signature of vegetation, which are controlled by topographic elements, such as, slope and aspect that render nadir sensing less efficient in capturing vegetation structure, and (iii) the coarse spectral and spatial resolutions of traditional broadband remote sensors, which are not fine enough to capture subtle variations in vegetation characteristics (Kumar et al. 2001; Mutanga et al. 2009). Traditional broadband remote scanners are mainly limited, because the spectral information in each channel is averaged over broadband widths. This results in a loss of critical information, which is often available in specific narrow-bands, for specific applications (Blackburn 1998; Cho and Skidmore 2009; Mutanga and Skidmore 2004c; Thenkabail et al. 2000).

Despite these challenges, remote sensing systems with many narrow-contiguous channels (i.e. hyperspectral data) are expected to offer a better capability to detect and differentiate between the grasslands with a C3 and C4 composition (Adjorlolo et al. 2012b; Liu and Cheng 2011). Hyperspectral data have been proposed for the characterization of the physiological and biochemical properties in vegetation that result in a common ecological role (Ustin and Gamon 2010). Important variables in this context are canopy nutrient parameters such as nitrogen (N), protein, vegetation moisture and fibre content. The percentage concentration of these variables has been found to vary between C3 and C4 grasses, making it possible to use spectral information obtained from remote sensing methods, to assess grasslands with the C3 and C4 composition.

1.3 Hyperspectral remote sensing

Unlike broadband systems, hyperspectral sensors, such as the narrow-band Hyperion, on board the Earth Observing satellite-1 (EO-1), can provide hundreds of bands within the visible, near infrared and shortwave infrared (SWIR) regions of the electromagnetic spectrum. Most hyperspectral sensors acquire radiance information with an average of 10 nm bandwidths from the visible to the SWIR (400-2500 nm). The major challenge, however, is that the resultant data

can represent multiple classes that are often mixed for a limited training-sample size (Dalponte et al. 2009). Specifically, this is a problem when parametric methods, which require first and other second order variations are used, whereby the training dataset is less than the number of input spectral bands (Chi and Bruzzone 2007). The problem with the use of very large number of predictors, compared to the number of observation for the response variable(s), is well-described in literature as the "curse of dimensionality" or the Hughes phenomenon of dimensionality (Hughes 1968). In addition, the Hughes phenomenon often leads to a high degree of multicollinearity, which is caused by the use of highly-correlated spectral predictors. This is a prominent problem in the statistical analysis of hyperspectral data (Clevers et al. 2007). The problem of multicollinearity in the matrix of input spectral bands often leads to highly unstable parameter estimates and thus high generalization error for most regression and classification algorithms (Bruzzone and Serpico 2000; Clevers et al. 2007).

1.4 Multivariate regression and classification techniques

Different statistical techniques that integrate information from several hyperspectral band values have been applied to distinguish between plant species. However, a survey of current literature showed that for any specific problems, different statistical algorithms have been investigated. This suggests that there is still no standard technique applicable to all problems. Data analysis techniques based on multivariate, non-parametric methods have been reported to provide better capability in processing hyperspectral data (Chan and Paelinckx 2008; Dalponte et al. 2009; Irisarri et al. 2009; Lawrence et al. 2006; Vyas et al. 2011). For example, Lawrence et al. (2006) mapped invasive plants by integrating hyperspectral image data and the random forest classification (RF) algorithm. Random forest (Breiman 2001a) is an ensemble classifier or regression algorithm that consists of many regression or decision trees. It uses the majority vote from individual trees to assign a class. RF is a non-parametric algorithm and it has been reported to outperform simple parametric techniques, such as stepwise linear regression (Kokaly and Clark 1999) and maximum likelihood classification (Chan and Paelinckx 2008; Lawrence et al. 2006), which require normal distribution of the data.

Although the random forest algorithm can deal with a very large number of input predictor variables, it has been recently reported to be susceptible to the multicollinearity problem (Strobl et al. 2008). In this respect, the authors proposed the use of the conditional

variable importance measure for random forest analysis, which has not been tested on a hyperspectral dataset. Therefore, it is envisaged that by mitigating against the multicollinearity problem in processing hyperspectral data, algorithms such as random forests can be aided to yield improved classification and regression accuracy in estimating the biophysical and biochemical parameters of vegetation, compared to inputting raw hyperspectral band features.

There has been a new push towards the development of multispectral sensors with far fewer bands than hyperspectral sensors but with bands strategically located in the absorption features of biochemical compounds such as WorldView-2. It is assumed that plant biochemistry and biophysical properties can be adequately modelled by capturing the spectral response related to these plant properties in the most relevant spectral regions rather than using the full spectral range of hyperspectral data. Several studies have highlighted specific wavelengths for vegetation biochemical and biophysical properties (e.g. Curran 1989; Thenkabail et al. 2004; Cho et al. 2007). Several hyperspectral data resampling techniques have been used to capture the spectral response around the specific bands related to biochemical and biophysical parameters including the application of symmetrical bandwidth integration (Thenkabail et al. 2000). The question arises whether the application of a user-defined response function that accounts for inter-band correlation could provide a more adequate means of resampling hyperspectral data around known bands related to biochemical and biophysical parameters?

1.5 Aim and objectives

The overall aim of the study was to identify optimal hyperspectral wavebands to detect, classify and map the C3 and C4 grass composition, as well as to predict variability in their nutrient concentration. The following objectives were set:

- To develop a hyperspectral resampling technique capable of characterizing the spectral response of vegetation around known wavelengths for biochemical and biophysical parameters and to assess the utility of the proposed technique for (i) classifying C3 and C4 grasses and (ii) remote sensing of nutrient concentration in C3 and C4 grasslands;
- To ascertain the utility of WorldView-2; a relatively new multispectral sensor designed with specific more bands in the visible to near-infrared (VNIR) related to biochemical and biophysical parameters when compared with conventional multispectral sensors, to (i) classify and (ii) map variability in nutrient quality in the C3 and C4 grasslands within

the Cathedral Peak World Heritage site, situated in the KwaZulu-Natal Drakensberg mountain range, South Africa.

1.6 Study area

The research was carried out within the Cathedral Peak World Heritage site, situated in the KwaZulu-Natal Drakensberg mountain range (Figure 1.1). The specific study site is located at NW Lat. = -28.97360039; NW Long. = 29.20739937; SE Lat. = -29.01429939; SE Long. = 29.26700020. The Cathedral Peak area comprises 15 catchments delineated for research and management purposes. These catchments are mainly situated on the flat terrace known as the Little Berg (foothills), which occurs below the main Drakensberg Escarpment.

The field investigations were conducted to collect canopy spectra of some key C3 and C4 grasses, which were chosen, based on preliminary findings from field reconnaissance survey and considering a classification of the major vegetation communities in previous studies in the area (Hill 1996; Killick 1963). Hill (1996) identified thirteen (13) vegetation groupings, in which each community either represents a climax community, within a particular attitudinal belt or progressing towards a climax community, or existing as a consequence of distinct topographical and/or geomorphological features. Further details on the specific vegetation communities are discussed in each stand-alone paper, forming Chapters 3 to 7.

In general, the specific sampling site, as indicated by the coverage of the WorldView-2 image in Figure 1.1 (25 km^2), is underlain by relatively homogeneous Drakensberg basalt formations of the Stormberg series (Killick 1963). With its diverse vegetation communities and associated functional type of C3 and C4 grasses, the Drakensberg mountain range is an ideal region in which to study possible vegetation changes or shifts, as a consequence of a changing environmental condition (Adjorlolo et al. 2012b). In the context of the remote sensing of the different C3 and C4 grassland communities in the area, a number of techniques were adopted. These included the use of an Analytical Spectral Device (ASD), a FieldSpec spectrometer (FieldSpec®3, Analytical spectral Device, Inc., USA) to collect canopy spectral reflectance data, a high accuracy global positioning device (Trimble GPS) to collect georeferenced data and a new generation multispectral satellite sensor, WorldView-2 imagery, which was acquired for the Cathedral Peak study area.

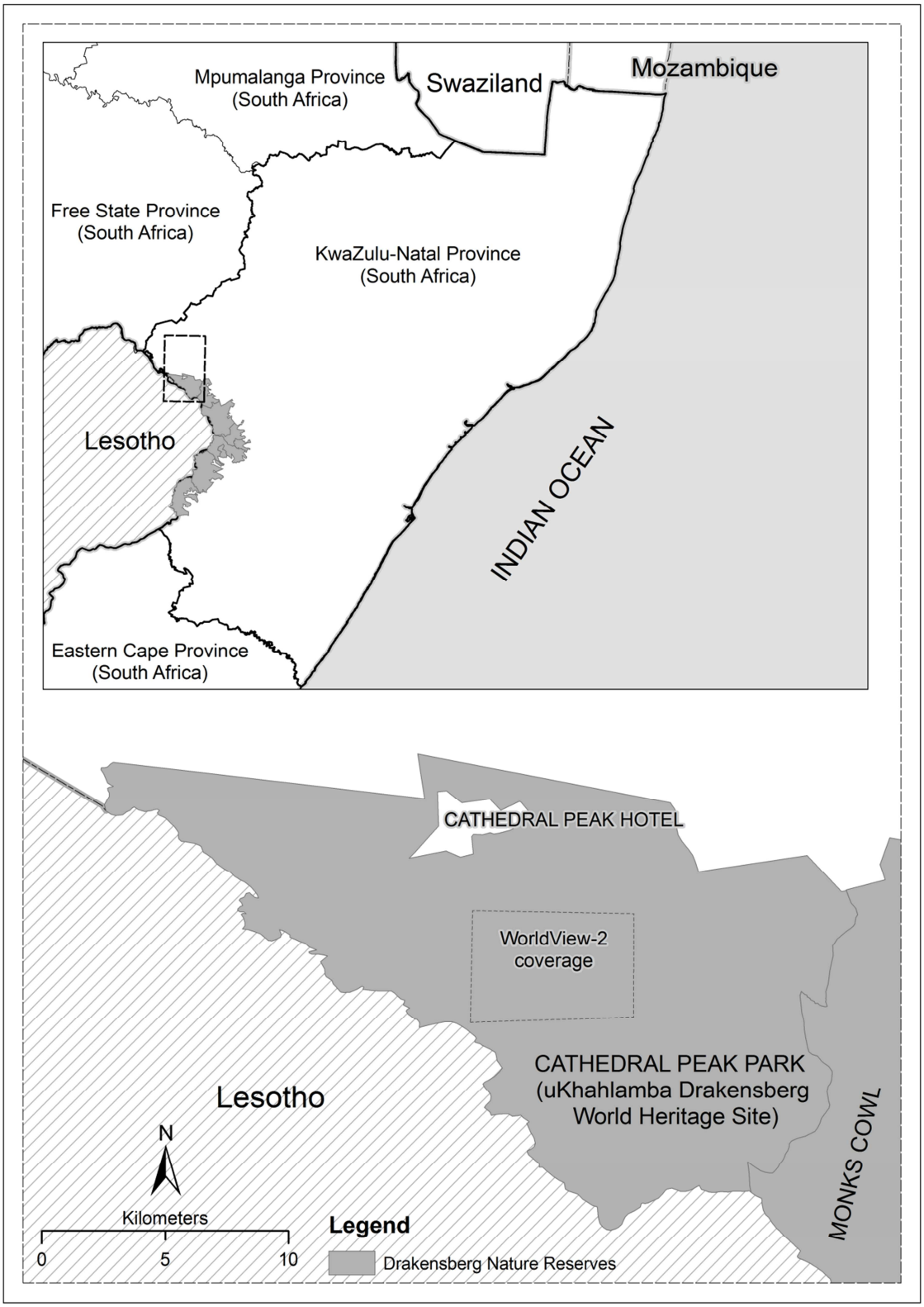

Figure 1.1 Location of the study area in the Cathedral Peak World Heritage Site.

1.7 Outline

This thesis comprises eight Chapters. The first chapter is the general introduction, followed by six chapters that form stand-alone papers, while the last chapter contains a synthesis of the research work. The stand-alone papers constitute three published, one in review and two in preparation. Each comprises an individual introduction, material and methods, results and discussion section. The stand-alone chapters have their own style, according to the corresponding journal. Although attempts were made to conform to a general style in the thesis, there may be some overlapping and repetition in some of the sections. Although the published papers, as well as manuscripts include other authors, my contribution was greatest and appropriate to being the first author in all cases.

Chapter 1

Provides the synoptic view of the research background and outlines the objectives and the structure of the thesis.

Chapter 2

Addresses the review of literature regarding the fundamentals of remote sensing, with its available tools, and explains the analysis methods for both multispectral and hyperspectral data. It provides an account of the opportunities and challenges, such as the spectral dimensionality and multicollinearity problem, in the use of remote sensing for C3 and C4 grass species discrimination and mapping.

Chapter 3

In this chapter, a user-defined inter-band correlation filter approach to resample hyperspectral data was proposed, using the asymmetrical spectral response feature of vegetation reflectance values. It addresses the fundamental problems of spectral dimensionality and multicollinearity in the statistical analysis of hyperspectral data. The random forest classification algorithm was used to analyze the resultant datasets.

Chapter 4

Focuses on optimizing the spectral resolution of the wavelengths of known reflectance/absorption features to discriminate between the C3 and C4 grass species and compares the developed user-defined inter-band correlation filter approach to the traditional approach of resampling hyperspectral data to the spectral response function of some commercial satellites sensors.

Chapter 5

Up-scales the capability of discriminating between the C3 and C4 groups of grasses from the spectral only dataset to the new generation multispectral sensor, WorldView-2 (WV-2) image data. The random forest classification algorithm was used to analyze the raw bands and vegetation indices developed from WV-2 data.

Chapter 6

Correlates the user-defined wavebands, developed from *in situ* spectra, against nutrient parameters (nitrogen, protein, moisture, and fibre) across the C3 and C4 grasslands, using the partial least squares regression analysis.

Chapter 7

Up-scales the capability of estimating variability in nutrient quality (nitrogen and fibre) across the C3 and C4 grasslands, to WorldView-2 (WV-2) bands and vegetation indices developed from the band combination of WV-2, using the Partial least squares and random forest regression algorithms.

Chapter 8

Provides a synthesis of the findings of this research and provides overall conclusions to the work.

CHAPTER 2: LITERATURE REVIEW

The chapter is based on:

Adjorlolo, C., Mutanga, O., Cho, A.M., & Ismail, R. (2012). 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, 47–61

A research poster from this review paper has been presented at the first, Academy of Science of South Africa (ASSAf), Young Scientists Conference, 2010, Pretoria, South Africa.

Abstract

Changes in the composition of plant functional type (PFT) activities are expected to accompany a changing climate. In tropical montane grasslands, such changes are predicted to follow shifts in the percentage cover and abundance of species following the C3 and C4 photosynthetic pathways. Reliable methods of detecting impacts of such changes on biomass and forage nutrient quality will likely provide a synoptic scale link between carbon sequestration, nutrient cycles and environmental change. Remote sensing approaches provide efficient methods to assess changes in vegetation composition quickly and are efficient for assessing vegetation in large areas. Multitemporal data obtained using broadband multispectral instruments have been used inconsistently to discriminate C3 and C4 grasslands and to draw some inferences. Advancements in narrowband hyperspectral systems are expected to offer greater potential for accurate mapping of C3 and C4 composition in grasslands. This paper presents an overview of the uses of optical remote sensing for C3 and C4 discrimination, which is consistent with the PFT concept used in land-surface modeling schemes. This concept uses eco-physiological characteristics, structural and/or phenological features to group species, in response to environmental conditions. This review describes the structural properties and biochemical characteristics that affect C3 and C4 grass reflectance. It highlights critical limitations and evaluates the potential of remote sensing approaches used for C3 and C4 studies.

Keywords: grassland, natural resource management, montane, remote sensing

2.1 Introduction

Discriminating plant species based on their functional types, rather than their higher taxonomic identity, is a promising way forward to understanding local vegetation and ecosystem dynamics (Ustin and Gamon 2010). Studies of plant functional type (PFT) responses to changing environmental conditions, namely precipitation, temperature, atmospheric CO_2 and land-use history, suggest important factors controlling rangeland vegetation (Bredenkamp et al. 2002; Ehleringer and Monson 1993). The concept of PFT used in modeling vegetated land surfaces uses morphological, structural and/or physiological attributes to group species: forbs, sedges, woody plants and grass functional forms. The grass form includes species incorporating CO_2 into an initial three-carbon or four-carbon compound referred to as C3 and C4 photosynthetic plants, respectively. C3 and C4 grasses exhibit a cosmopolitan distribution and occupy a wide range of habitats (Osborne 2011; Sage et al. 2011). The distribution of C3 and C4 grass types is closely linked to environmental conditions (Ehleringer and Monson 1993; Teeri and Stowe 1976). There is also accumulating evidence that climate change will possibly influence shifts in C3 and C4 grass compositions and that such changes will not be geographically uniform (Davidson and Csillag 2001; Ehleringer et al. 1997).

The composition of C3 and C4 grasses occurring in higher latitudes and high altitudinal gradients is of key research concern. Particularly in zones of biological transitions, C3-dominant grasslands can have C4 grass weeds or vice versa and, depending upon the species involved, this could favour or hamper their primary productivity (Winslow et al. 2003). A biological transition zone includes mosaics of C3 and C4 functional types occupying the intermediate temperate-tropical climatic conditions (Bredenkamp et al. 2002; Edwards et al. 2010; Neilson 1993). Ehleringer et al. (1997) describe transition zones as the point (crossover temperature) at which the quantum yields of C3 and C4 grasses are equal. The tropical montane grasslands of southern Africa are examples of zones of biological transition. In this transition zone, C3 and C4 grasses grow under conditions that are marginal to their growth. Thus, these types of grasslands can be highly responsive to even subtle change in environmental conditions. Reliable methods of detecting such changes are necessary if these grasslands are to be properly managed.

The use of field-based methods has traditionally yielded precise information about species relative cover and composition (Reich et al. 2001; Stock et al. 2004; Teeri and Stowe 1976). However, the use of field methods alone is limited in its provision of quick and regularly

updated spatial information. The methods are limited because they can be labour-intensive, time-consuming and impractical for inaccessible landscapes. Therefore, alternative methods capable of providing timely, up-to-date and accurate information are required. Such methods need to provide information that is useful for making important inferences about PFT responses to environmental change, without detailed information about the higher taxonomic classification of each species.

The advent of broadband multispectral sensors has provided useful information for broad-scale remote sensing of functional types of C3 and C4 grass communities (Davidson and Csillag 2003; Foody and Dash 2007; Goodin and Henebry 1997). Such studies have typically applied multi-temporal remote sensing approaches, taking advantage of the well-described asynchronous seasonal activities of these grasses. Key environmental gradients, such as precipitation and temperature, affect the asynchronous seasonal profile and related biomass productivity of C3 and C4 grasses (Davidson and Csillag 2003; Purevdorj et al. 1998; Ricotta et al. 2003). The environmental niche separations between C3 and C4 grasses have been based on observed contrasting seasonal dynamics, including photosynthetic activities (Kemp and Williams III 1980; Niu et al. 2008). This phenomenon has allowed the estimation and mapping of the relative abundance and composition of C3 and C4 grasses, predominantly in temperate regions (Davidson and Csillag 2003; Foody and Dash 2007; Goodin and Henebry 1997).

In tropical montane grasslands, the dynamics of C3 and C4 grasses are not controlled only by seasonality. A host of additional factors, such as fire, herbivory, site-specific edaphic conditions and topography, strongly influence the growth patterns of these types of grass species (Bond et al. 2004; Winslow et al. 2003). In general, empirical studies have shown that the composition of C3 and C4 types in tropical montane grasslands can vary across topographic (slope, aspect and altitude) and wetness gradients (Kotze and O'Connor 2000; Morris et al. 1993; Sieben et al. 2010; Tieszen et al. 1979). Morris et al. (1993) found that slope orientation dictates the proportion of C3 species present in the eastern alpine vegetation belts of the Drakensberg mountain range in Lesotho. Their study showed that topographic factors, rather than seasonality, directly influence C3 and C4 plant dynamics by modifying moisture regimes, solar radiation patterns and temperature at the landscape scale. As such, the floristic composition of tropical montane C3 and C4 grass communities is more dynamic than many other grassland landscapes.

Although multitemporal data have been interpreted in ways that correspond to the asynchronous seasonal activities of C3 and C4 grasses, such data are of limited use when applied to tropical montane environments, in which seasonality plays a minor role. For example, in the Drakensberg montane grasslands of South Africa, C3 grass communities, such as the *Festuca costata* subalpine vegetation belt, remain green in the winter weather conditions. *Festuca costata* grows as aggregated large tussocks and are taller than their coexisting C4 grasses, and the distribution of the species is concentrated on south-facing, cooler and moister slope aspects (Kotze and O'Connor 2000; Sieben et al. 2010). Thus, the remote sensing approaches that incorporate topographic factors and detailed optical features of C3 and C4 grass types could provide sensitive information for the accurate assessment of vegetation dynamics.

Advancements in hyperspectral sensors offer contiguous spectral data for discrimination and mapping of C3 and C4 grass types. This generation of instruments offers n-dimensional narrow-spectral band information for the identification of 'optical types' of C3 and C4 grass species (Ustin and Gamon 2010). The concept of 'optical type' could be modeled around basic ecological principles, characterized by species-specific optical properties detectable from remote measurements (Gamon 2008). Such properties include those of absorption and scattering features of plant canopies, based on the principles of spectroscopy (Kumar et al. 2001). However, as with any new technology, the application of current hyperspectral remote measurement capabilities comes with a number of challenges. These include cost and data processing complexities, namely hyper-dimensionality, low signal-to-noise ratios, and limiting spatial and temporal resolutions. In spite of these limitations, substantial research has been done in the fields of forestry and agriculture (Zwiggelaar 1998). In rangeland ecology, only a few studies have specifically applied hyperspectral remote sensing data for mapping C3 and C4 grass composition (Liu and Cheng 2011). This paper presents an overview of the basic optical and structural properties, and biochemical characteristics that affect C3 and C4 grass spectra. Key limitations are highlighted and new perspectives in remote sensing are discussed for the detection and mapping of C3 and C4 grass species. This is to highlight growing research interest in remote sensing that focuses on PFTs of C3 and C4 grass species and implications in assessing the impacts of environmental change.

2.2 Functional-type classification of C3 and C4 grasses

C3 and C4 grass species discrimination corresponds with PFT niche separation used in evaluating landscape vegetation and ecosystem dynamics (Black et al. 1969; Chazdon 1978; Chen 1996; Ehleringer and Monson 1993). In general, the functional types of C3 and C4 grasses represent a classification scheme that lies between species and broad vegetation types (Ehleringer and Monson 1993; Ustin and Gamon 2010). Such an approach to vegetated landscape classification offers a promising way forward to answering important ecological questions. These include those of C3 and C4 grass tolerance to environmental gradients or disturbances. By degree of tolerance, C3 and C4 grass species have been classified broadly as being representative of temperate (cold-season grassland; CSG) and tropical (warm-season grassland; WSG) plant communities, respectively (Davidson and Csillag 2001; Foody and Dash 2010; Goodin and Henebry 1997). The grasses of CSG and WSG have clear phenological differences. Generally, the latter type of grassland starts growth in mid-spring and grows rapidly through late spring to early summer. The WSG type has little or no growth in winter and therefore senesces, whereas the CSG type tolerates the winter weather.

It is widely accepted that key physiological differences between C3 and C4 functional types characterize their ecological adaptations. C4 grasses are more tolerant of tropical or subtropical climatic conditions where they dominate floral diversity and net primary productivity (Chamaillé-Jammes and Bond 2010; Davidson and Csillag 2001; Foody and Dash 2010; Goodin and Henebry 1997). Inferences drawn from the photosynthetic mechanisms of functional types suggest that C4 grass species have a number of competitive advantages over grasses with the C3 pathway under elevated solar irradiance and higher atmospheric temperatures. These include higher water-use efficiency and CO_2 fixation rates. Some empirical evidence further suggests that at high leaf temperatures, the C4 pathway suppresses photorespiration by concentrating CO_2 internally, through a series of biochemical and anatomical adaptations (Ehleringer et al. 1991). Other studies note that C4 plants can conserve moisture and nutrients in dry summers and increase productivity because of their extensive root system (Wang et al. 2010). This highly coordinated physiological mechanism of C4 species, with reference to resource-use efficiency, suggests significant controls that certain environmental conditions have on the PFTs of C3 and C4 grass distributions, from poles to tropics and along altitudinal gradients (Ehleringer and Monson 1993).

The C3 photosynthetic pathway is thought to have evolved during periods of high CO_2 concentrations and lower atmospheric temperatures (Edwards et al. 2010; Ehleringer et al. 1991). This so-called primitive photosynthetic mechanism is limited when it comes to CO_2 uptake under very high atmospheric temperatures and moisture deficient environments (Ehleringer and Monson 1993; Foody and Dash 2007; Goodin and Henebry 1997; Poorter 1993). Generally, because the C3 pathway is susceptible to the higher rate of photorespiration, grass species following this pathway are more limited in their uptake of CO_2, and with a net loss of photosynthetic carbon in hot dry weather conditions. Amongst others, Ehleringer et al. (1991) note that C3 grasses are widely distributed, with a marked dominance in higher latitudes and temperate grasslands, largely due to their environmental requirements for growth. Similarly, the temperate climatic conditions of high altitude, tropical montane grasslands are consistent with the occurrence of C3 grass species in these grassland landscapes.

It is generally accepted that because C3 and C4 grasses have markedly different environmental requirements for growth, they differ in their responses to changes in the environmental conditions (Ehleringer et al. 1997; Hill et al. 1997; Peters 2002; Tieszen et al. 1997). For example, Poorter (1993) found that the C3 species showed a larger increase (41%) in biomass production, compared to the C4 species (22%), in response to increasing CO_2 levels. Furthermore, Reich et al. (2001) note that C3 grasses show higher total biomass and higher forage nutrient (N) concentrations under elevated CO_2, compared to the C4 grass species. However, empirical examinations have shown that, although the C3 species generally store higher levels of protein, non-structural carbohydrates and moisture under elevated CO_2, they have lower total carbohydrate/protein ratios than C4 species (Barbehenn et al. 2004). Ehleringer and Monson (1993) discussed in detail the photosynthetic pathway variations in the C3 versus the C4 grass species and their ecological adaptations. For the purpose of the current review, a summary of key distinguishing attributes of C3 and C4 PFTs is presented in Table 2.1. Some examples of tropical (WSG) and temperate (CSG) species from the African tropical montane rangelands are presented (Kotze and O'Connor 2000; Schmidt and Skidmore 2001).

Table 2.1 Features of C3 and C4 grasses

Characteristics	C3 Grass species	C4 Grass species	Reference
Initial molecule formed during photosynthesis	3-carbon	4-carbon	Foody and Dash (2007)
Growth period	cool season or year long	warm season	Tieszen et al. (1997)
Solar radiation requirements	lower	higher	Gamon et al. (1997), Guo and Trotter (2004)
Temperature requirements	lower	higher	Corson et al. (2007)
Moisture requirements	higher	lower	Foody and Dash, (2007)
Frost sensitivity	lower	higher	Sieben et al. (2010)
Production	lower	higher	Peters, (2002)
Examples from African Montane rangelands	*Agrostis eriantha, Festuca costa, Festuca caprina, Koeleria capensis, Leersia hexandra, Merxmuellera macowanii*	*Cynodon dactylon, Harpochloa falx, Hyparrhenia hirta, Sporobolus africanus, Themeda triandra*	Kotze and O'Connor (2000), Schmidt and Skidmore (2001)

2.3 Implications of C3 and C4 grass species assessment in climate change studies

Increasing levels of greenhouse gases, coupled with climate change, are expected to have major impacts on vegetation and thereby affect the functioning of grassland ecosystems (Rock et al. 1988; Walther 2010; Walther et al. 2002). Changes in vegetation composition resulting from the impacts of climate change increasingly raise policy debates and scientific research interest (Stock et al. 2004; Wylie et al. 2003). For example, researchers have shown that increases in temperature and atmospheric CO_2 can alter the distribution and relative abundances of grass species, with critical implications (Ficken et al. 2002; Williams and Hunt 2004). Furthermore, physical examinations suggest that, because C3 and C4 grasses have different carbon sink capabilities, large-scale changes in their relative abundance may have significant feedback effects on carbon sequestration, which in turn will affect future atmospheric CO_2 concentrations (Corson et al. 2007; Tieszen et al. 1997; Wand et al. 1999).

In southern Africa, for instance, although the dynamics of C3 and C4 grasses are not fully known, the extent of the grassland biome is predicted to decrease under climate change scenarios. It is also predicted that southern Africa will generally experience a 3 °C warmer and much drier climate, with January temperatures increasing the most in the interior and rainfall decreasing by between 5% and 25% in the north and south, respectively (Lumsden et al. 2009; Schulze et al. 2005). Other modelling scenarios predict that climate change could result in a warmer and wetter environment along the eastern montane regions of South Africa (Rutherford et al. 1999).

Increases in temperature should favour C4 species, and C3 species could possibly move up to higher altitudes or into south-facing cooler and moister slopes in response to increasing temperatures. On the other hand, there is a growing concern that increasing concentrations of atmospheric CO_2 will favour the C3 species, and that the distribution and abundance within their current range will increase, possibly at the expense of C4 grasses (Bond et al. 2004; Winslow et al. 2003). Chamaillé-Jammes and Bond (2010) recently noted that increasing levels of CO_2 might be driving the southern African C4 grass flora towards an uncertain future, and that C3 species might become more abundant. There is also the hypothesis that the C3 species will respond to changes in site edaphic factors (e.g. increases in soil carbon and moisture contents), as they do for increasing atmospheric CO_2 concentrations. These changes could be as a result of increases in rainfall or cooler micro-temperatures (due to persistent cloud cover). However, because there is insufficient empirical evidence to support the extent of such impacts, the responses of C3 and C4 grasses to a changing environment is unclear (Chamaillé-Jammes and Bond 2010).

At the continental scale, climate models, such as the global circulation models, have provided predictions of significant climate warming in response to increasing CO_2 and other greenhouse gases. However, climate by its nature varies from place to place, driven by, inter alia, the variation in the distribution of solar irradiance. Global models show decreases in confidence level and are generally not sufficiently sensitive for monitoring changes at smaller scales (Fischlin et al. 2007). Therefore other techniques, such as the use of regional climate models (RCMs) or downscaling methods have been developed specifically for the study of regional- and landscape-scale climate change. These models rely heavily on species and long-term average temporal and spatial climate records that are themselves highly variable (Hewitson et al. 2005).

In addition, there is a general lack of reliable observational data or regularly updated detailed information on the taxonomic identification of individual species. However, the approach of PFT used in land-surface modeling schemes is consistent with C3 and C4 grassland classification using remote sensing techniques. Specifically, experimental approaches have been tested using remotely sensed data integrated with multifactorial treatments of ecosystem responses of temperate C3 and C4 grasslands to climate change (Tieszen et al. 1997; Wang et al. 2010).

Remotely sensed data have been increasingly analyzed to reflect the abundance of the C3 and C4 PFTs, which in turn indicate their response to resources or environmental conditions as well as to improve vegetation models. For example, dynamic global vegetation models rely on the concept of PFT to group plant species, based on shared traits, including photosynthetic pathway differences in plant communities (Poulter et al. 2011). These models offer scientists the opportunity to integrate remote measurements of vegetated surfaces in RCMs (Myoung et al. 2011). However, it is also pertinent to evaluate the potential contribution of remote sensing approaches in monitoring tropical montane C3 and C4 grassland dynamics. The contribution of remote sensing to studies of climate change in tropical montane systems is critical, since these particular systems are expected to be much affected by changing climatic conditions (Hewitson et al. 2005; Scholes 2006).

2.4 Reflectance properties of vegetation

2.4.1 Leaf and canopy reflectance characteristics

Measurements of major light absorbing compounds in leaves, such as plant pigment and water content, show how the concentration of these biochemical constituents is a key regulator of leaf or canopy reflectance (Gates et al. 1965; Knipling 1970). Using leaf and canopy optical radiative transfer models (RTMs), researchers have explored the complex set of factors that control vegetation reflectance. Thus RTMs, such as the PROSPECT model (Jacquemoud and Baret 1990) and SAIL (Verhoef 1984), have been used to explain the mechanisms that control the spectral properties of individual leaves and canopy dynamics. The PROSPECT model, for instance, is based on the concepts of multiple scattering within leaves, caused by the refractive index discontinuities between cells and intercellular air spaces, and leaf level absorption by the various biochemical components of vegetation (Jacquemoud and Baret 1990; Jacquemoud et al. 1995).

The effects of absorption by leaf biochemicals, such as starch, lignin, leaf nitrogen and protein, are increasingly recognized as having measurable effects on the spectral reflectance of leaves (Curran 1994; Kumar et al. 2001; Mutanga et al. 2003; Mutanga et al. 2009). The PROSPECT model uses leaf chlorophyll content, leaf water content and leaf structural parameters to predict leaf reflectance in the 400–2 500 nm range.

However, because leaf size is large compared with the optical wavelengths with which measurements are made, vegetation canopies cannot be modeled as one large unit. There is a range of structural variables, in addition to those of internal leaf properties that affect canopy reflectance. These include background substrates such as soil, leaf area index (LAI) of canopies and leaf angle distribution with the canopy. The canopy reflectance model must consider these spectral characteristic features. Hence, models such as the SAIL algorithm use a combination of spectral and directional reflectance properties to characterize the radiative transfer of canopy dynamics (Jacquemoud et al. 1995).

Topographic elements, directional reflectance effects of viewing and illumination angles influence variations in the anisotropic reflectance spectra of canopies. Specifically, topographic elements, such as slope and aspect, affect the remote sensing spectra of grass canopies and stand structure (Asner et al. 1998a). Consequently, these factors render nadir or single-view remote sensing less efficient, particularly in mountainous environments. It has been envisaged, although not fully developed, that multi-angular high-resolution data will provide a unique opportunity to detect and discriminate terrestrial vegetation in mountainous landscapes (Chopping et al. 2006). The development of this concept and generalized algorithms, based on multi-angular reflectance modeling, will offer more successful C3 and C4 grass canopy mapping in montane environments.

Spectral and angular vegetation reflectivity, which changes as a function of leaf and/or canopy structural characteristics, is commonly described as the bidirectional reflectance distribution function (BRDF; Asner et al. 1998). Models predicting the BRDF for canopy conditions and the effects of topographic elements are widely applicable for the assessment and monitoring of C3 and C4 grasses in mountainous landscapes. Multi-angular remote sensing approaches thus offer the promise of identifying combinations of viewing and illumination angles that would minimize the influence of background signals, such as soil-water substrates. As noted by Chopping et al. (2006), the Multi-angle Imaging Spectroradiometer (MISR)

instrument onboard NASA's Terra platform and the multi-angle satellite Compact High Resolution Imaging Spectrometer (CHRIS) are among the few sensor systems available for exploration.

2.4.2 Influence of morphology on reflectance

A number of factors influence the reflectance spectrum of C3 and C4 grasses. These include the optical, structural and biochemical characteristics of vegetation. Physical examination has shown that the molecules and pigments forming part of the complex structure of plant cells reflect, scatter and/or absorb radiation differently (Asner 1998; Asner et al. 1998a; Blackburn 2007a; Blackburn 2007b; Kumar et al. 2001). A number of experimental studies have examined vegetation reflectance properties and shown that leaf epidermal cells focus radiation, palisade cells act as radiation pipes, and spongy mesophyll cells act as efficient radiation scatterers (Brodersen and Vogelmann 2007; Knapp and Carter 1998; Myers et al. 1994). A number of characteristic features differ, with the C4 grass anatomy having more compact mesophyll cells, a higher proportion of leaf vascular tissue and a smaller interveinal distance than the C3 grass anatomy (Dengler et al. 1994; Ogle 2003). In addition, the C4 leaf structure has fewer intercellular air spaces, less water and therefore reduced scattering from the air-water interface than the C3 leaf, which has less compact and thinner cell structures (Lee et al. 1990; Ueno et al. 2006; Vogelmann and Martin 1993). The attenuation of radiation as it transmits through the different cell layers, the gradient of reflectance that results and irradiance scattering within the internal cell structures, have been well described and differ between C3 and C4 functional types, both at the leaf and canopy levels (Lichtenthaler et al. 1981; Vogelmann 1993).

Variations in internal cell structure result in different spectral reflectance of C3 and C4 grasses, providing the potential to discriminate between these functional types. As an example, Slaton et al. (2001), and many other studies, have demonstrated the substantial influence that a leaf's internal structure has on plants' optical reflectance. The cognition of leaf internal structure contributions to the reflectance spectrum is critical for understanding the drivers of spectral variation in C3, C4, or C3 and C4 dominated grasslands. Moreover, empirical evidence has shown that the intra- and inter-specific differences in leaf internal structure correspond significantly to reflectance differences, even when leaf water and pigment concentrations of five species remain identical (Carter 1991; Sinclair et al. 1971).

2.4.3 Reflectance in different wavelength regions

Incoming solar radiation interacting with any vegetated surface may be partly reflected, transmitted and/or absorbed. The nature of interaction in the different wavelength regions of the electromagnetic spectrum can vary between species at leaf or canopy level (Slaton et al. 2001). The fundamental principle is that reflectance, which is generated when radiation interacts with vegetation, is dependent both on the properties of the incoming radiation (e.g. conditions of radiation, angle of incidence, and wavelength) and on the properties of the intercepting vegetation (Kokaly et al. 2003; Ross 1981). These studies indicate that biochemical contents, including plant pigments (chlorophylls, carotenes, xanthophylls and tannins), nutrients (mainly nitrogen and phosphorous) and other carbon compounds (such as lignin, cellulose, starch and proteins), are spectrally detectable. In addition, several studies of vegetation of spectroscopy have suggested that these biochemical constituents absorb mainly incident radiation, with the nature and the amount varying in specific wavelength regions (Blackburn 2007a; Ferwerda et al. 2005; Gamon et al. 1997; Hansen and Schjoerring 2003; Mutanga and Skidmore 2004c; Mutanga et al. 2003). Recent studies have shown that there are coefficients of determination (r^2) between species-specific light-response and photosynthetic parameters (Guo and Trotter 2004). In their study, Guo and Trotter (2004) observed significant r^2 values between the photochemical reflectance index and several photosynthetic parameters, including ratio of carotenoid to chlorophyll content, for a group of species that differ in photosynthetic light use efficiency and net CO_2 uptake rate.

The reflectance in the visible (400–700 nm) region of the spectrum is determined mostly by the selective absorption of photosynthetic pigments, whereas leaf or canopy constituents, such as strong intercell scattering and leaf water absorption, affect the near-infrared region (NIR; 700–1 100 nm). Shortwave infrared (SWIR; 1 100–2 500 nm) reflectance is affected by strong leaf or canopy liquid water absorption, LAI, fraction of vegetation and other macronutrients (Ceccato et al. 2001; Ghulam et al. 2007). For a more detailed discussion on the influence of vegetation water content on reflectance, we refer the reader to Carter (1991).

At the area of transition from red into the NIR wavelength region (690–720 nm), empirical studies have shown that vegetation reflectance rapidly increases, forming a distinct feature referred to as red-edge features (Kumar et al. 2001). The red-edge features have provided important information for the assessment of a number of vegetation biochemical compositions

(Liu et al. 2004). In this study, the authors found negative and positive correlation coefficients for the red-edge width and position, respectively, for the estimation of leaf/canopy chlorophyll density, LAI and water content for six different growth stages. Mutanga et al. (2003), among others, found that reflectance centred at the red-edge region is useful for the detection of differences in foliar biochemical (nitrogen and phosphorus) concentrations in grassland canopies (Kumar et al. 2001).

Although it is possibly one of the most researched regions and proposed as the spectroscopic biosignature, the potential of the red-edge region to differentiate C3 and C4 grass species, based on differences in biochemicals, has not been fully exploited. Hansen and Schjoerring (2003) calculated normalized spectral indices based on the red-edge region (680–750 nm). They correlated such indices with chlorophyll and nitrogen density for the estimation of the C3 winter wheat. Mutanga et al. (2004) predicted canopy nitrogen and phosphorus concentrations in several C4 grass species, using the red-edge region (550–750 nm). In addition, Smith and Blackshaw (2003) found that reflectance in the red-edge region was consistently required to discriminate two crops and five weed species of C3 and C4 composition.

A number of studies demonstrate that the spectral variability between C3 and C4 grass species at both leaf and canopy level is greater than the intragroup variation (Daughtry et al. 1992; Trenholm et al. 2000). For example, reflectance at 531 and 570 nm has been used to estimate photosynthetic radiation use efficiency and proposed as a set of sensitive bands for discriminating between C3 and C4 grass species (Gamon et al. 1997). Slaton et al. (2001), among others, studied the NIR region and found that reflectance at 800 nm is significantly different among species that differ in intercellular structure.

In Figure 2.1, average reflectance curves measured at full canopy, under common field conditions, are shown to illustrate differences in reflection and absorption peaks between C3 (*Festuca costata*) and C4 (*Themeda triandra*) dominant canopies in the montane grasslands of southern Africa (CA and OM unpublished data). These grass canopies were sampled from the Cathedral Peak region of the Drakensberg mountain range, South Africa. The region consists of vegetation divided into altitudinal zones, corresponding closely with the physiographic features of the Drakensberg mountain range. Conspicuously, there are strong consociations of *T. triandra* and *F. costata* dominance on warm, northerly and cool, southerly slopes, respectively, within the subalpine vegetation belt (Granger and Schulze 1977; Hill 1996; Killick 1963).

The spectral dynamics of C3 and C4 grass reflectance have important implications for species detection and discrimination. As an example, the spectral differences in reflectance curves between *F. costata* and *T. triandra* can be used to detect changes in landscapes composed of these functional types. That is, any shift in abundance or composition, particularly in regions of biological transitions, would have far-reaching impacts that are indicative of changing environmental conditions (Chamaillé-Jammes and Bond 2010).

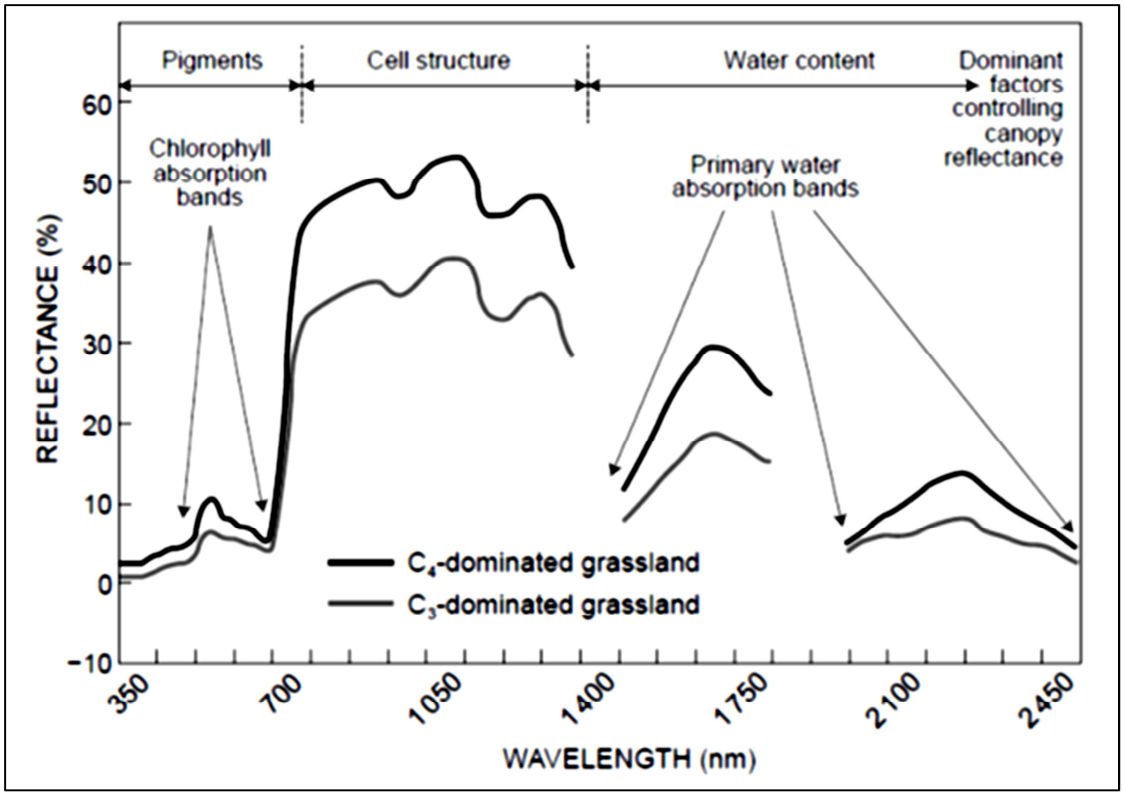

Figure 2.1 Average spectral canopy response curves for C3 (Festuca costata) and C4 (Themeda triandra) dominated grass patches in the Drakensberg montane grasslands of South Africa (n = 110 for each representative grass species). The canopy reflectance spectrum is significantly controlled by leaf optical properties: leaf area index (LAI), biomass quantity and leaf angle distribution, and soil background reflectance. Noisy Atmospheric water absorption bands (1350-1460 nm and 1790-1960 nm) removed.

2.5 Role of remote sensor systems in the C3 and C4 grass study

2.5.1 Use of broadband multispectral sensors

Broadband remote sensing has been a valuable source of multispectral data. However, these data have been inconsistent in mapping C3 and C4 grasslands at various spatial and temporal scales. For example, the supervised classification of Landsat TM data (bands 2–5) yields an overall accuracy of 91% in mid-summer (July) images, but could only attain 73% in autumn (September) images, whereas the use of multitemporal images could not improve accuracy levels (Peterson et al. 2002). A number of vegetation indices have been developed to achieve the normalization of spectral differences and the reduction of soil/water substrate effects (Lillesand et al. 2008; Price et al. 2002). Price et al. (2002) found that the normalized difference vegetation index (NDVI) in mid-summer outperformed raw-image bands classification up to 10% when discriminating six grassland types.

Temporal trajectory indices of sensor-derived NDVI have been assessed for mapping C3 and C4 grass relative abundances, on the basis of the phenological difference between these types of grasses (Goodin and Henebry 1997; Tieszen et al. 1997; Wang et al. 2010). These studies suggest that the aseasonal phenological difference between C3 and C4 grasslands could be enhanced using trajectory based indices derived from multitemporal image data. For example, Davidson and Csillag (2003) estimated C3 and C4 grass relative abundance using the temporal trajectory derived NDVI index for the Advanced Very High Resolution Radiometer (AVHRR) instrument. The authors recorded an overall accuracy of 74% (Table 2.2) when discriminating among grasslands of different C3 and C4 grassland composition. Similarly, the MERIS terrestrial chlorophyll index (MTCI) (Dash and Curran 2004; Harris and Dash 2010) derived from Envisat MERIS images showed that the cumulative MTCI corresponds with the percentage cover of cold-season grasses. However, the week of maximum MTCI indices correlated negatively with the percentage grass cover in the mixed grasslands of C3 and C4 species composition (Foody and Dash 2007). This study reported 77.8% accuracy for the mapping of C3/C4 grass community compositions in the temperate northern Great Plains.

The normalized difference water index (NDWI) (Bowyer and Danson 2004), computed using NIR and shortwave infrared (SWIR) reflectance, has been tested for the detection of vegetation water stress and for the delineation of grasslands differing in the compositions of PFT (Gao 1996; Wang et al. 2010; Wang et al. 2007). These studies exploit the principle that C4

grasses contain lower leaf moisture in spring and higher moisture in summer compared to C3 grasses. Other studies have shown that differences in seasonal leaf or canopy moisture content between C3 and C4 plants are spectrally distinguishable (Guo et al. 2003; Peterson et al. 2002). These studies show that environmental factors, such as temperature and precipitation, can have significant effects on the seasonal variation in reflectance of C3 and C4 grasses. In addition, land-use activities, such as grazing and fire regime, contribute to uncertainties in regular image interpretation and classification with reference to grass species (Wang et al. 2010).

Vegetation indices derived from multispectral sensors have been found to saturate asymptotically at a certain biomass density, LAI or percentage cover as the growing season progresses. Thus, vegetation indices used for assessments of grasslands could be limited with reference to prediction or classification accuracy (Ricotta et al. 2003). Specifically, NDVIs computed from Landsat Thematic Mapper (Landsat TM) have shown substantial saturation effects, both for predictive and discriminant analysis (Gao et al. 2000; Mutanga and Skidmore 2004a; Thenkabail et al. 2000). Kumar et al. (2001) comprehensively discussed the critical limitations that arise from using broadband multispectral sensor types for vegetation assessment. The authors demonstrate that these sensors are limited, mainly because of the lower number of spectral channels, averaged for discrete sets of wavebands. The consequence of this in C3 and C4 grass species discrimination is that fine spectral features between these functional types could be masked (Davidson and Csillag 2003; Tieszen et al. 1997).

Furthermore, spatial data obtained from multispectral sensors are generally limited in providing sufficient information that can be used to discriminate between the mixed pixels of C3 and C4 canopies. Davidson and Csillag (2003) note this as a major challenge that remains, when it comes to inferring C3 and C4 grass composition and dynamics from coarse resolution (1 km) AVHRR pixels of mixed grasslands. In this regard, the utility of approaches that use temporal trajectory indices has been suggested to take advantage of the asynchronous seasonal growth pattern exhibited by C3 and C4 grasses, although the majority of studies were confined mainly to modelling the ‘pure’ temperate communities of C3 and C4 grass composition (Guo et al. 2003, Ricotta et al. 2003, Price et al. 2004, Wang et al. 2010).

Table 2.2 Examples of some major sensors, their spectral/spatial properties, and application in C3 and C4 species prediction, classification and mapping.

Sensor/Dataset	Platform	Spectral Range	Maximum number of channels	Spectral Resolution	Spatial Resolution	Application	Techniques	Accuracy	References
NOAA AVHRR multispectral data	Satellite	0.58 –12.5 µm	6 Discrete wavebands	Variable	1.09 km at Nadir	Detection and prediction of C4 grass species cover in mixed C3 and C4 grass prairie	Spectral discrimination of the temporal trajectory indices	overall accuracy of 74%	Davidson and Csillag (2003)
Landsat TM/ multispectral sensor	Satellite	0.45 – 12.5µm	7 Discrete wavebands	Variable	30 m (120 m - thermal)	Spectral discrimination between cool season (predominantly C3) and warm season (mixed C3 and C4 grass species) grassland cover types	Multiple analysis of variance (MANOVA) and Discriminant analysis	overall accuracy of i) 81.8% and overall accuracy of ii) 70.4%	Guo et al. (2003), Peterson et al. (2002)
Envisat/ MERIS: Research satellite	Satellite	390–1040 nm	15 bands programmable in position and width	1.8 nm	300 m (1200 m-thermal)	Prediction and mapping of broad classes consisting of C3 and C4 grass species composition	MLP ANNs	$R^2 \sim 0.60$-0.65 of the variance in grassland composition and an overall accuracy of 77.8%	Foody and Dash (2007), Harris and Dash (2010)
AVIRIS: Hyperspectral sensor	Airborne	404–2400 nm	224	10 nm	19 m	Dominant plant species richness detected to precisions of 6 to 7 species within a range of 16 to 61 species per the study transect	Simple linear regression of species richness	R^2 =0.41 $P \leq 0.0001$	Carter et al. (2005)
AISA Eagle Hyperspectral sensor	Airborne	398 – 984 nm	68	9 nm	1.5 m	Estimation of abundance and distribution of C4 (*Miscanthus sacchariflorus)* and C3 (*Phragmites australis*)	Stepwise multiple linear regression	90 & 87.6 %, respectively, for two species prediction and distribution mapping	Lu et al. (2009)
OMIS II Hyperspectral sensor	Airborne	470 – 1100 nm	64	10 nm	3 m	Detection and Discrimination of C3 and C4 plants: including *Tamarix chinensis Lour* and *Zea mays* (maize), respectively.	simple decision tree-based classification	overall classification accuracy of 92%	Liu and Cheng (2011)

Temporal trajectory indices are expected to be relatively less useful in habitats where the drivers of variation in C3 and C4 dynamics may not only be controlled by seasonal activity. Such habitats may include montane grasslands, especially those in the tropical regions where C3 and C4 grass phenology relates not only to seasonality (Kotze and O'Connor 2000; Stock et al. 2004). Sieben et al. (2010) studied changes in plant functional types and the compositions of assemblages across altitudinal and wetness gradients in the Maloti-Drakensberg grasslands of South Africa. The authors found that, besides the seasonal profile, topographic elements, such as aspect, elevation and wetness gradient, control the variation in C3 and C4 grass species distribution and abundance. Their findings are similar to those of Morris et al. (1993), where the discriminant analysis of five vegetation communities (temperate and tropical) in the alpine catchments (above 2 950 m) of Lesotho indicated that these communities occupy particular topographic positions and solar radiation patterns in the grass sward.

Substantial advances in remotely sensed data analysis have led to the development of improved image interpretation and data analysis techniques. These include the minimisation of the impact related to the mixed-pixel problem. Initially, a spectral mixture analysis method was used to decompose the reflectance values into proportions of selected components (Lu et al. 2003). The use of the spectral mixture analysis (SMA) technique has been proposed to solve some of the problems that arise when differentiating vegetation types within heterogeneous landscapes, using pixels of mixed grasses (Goodin and Henebry 1997; Smith et al. 1990; Wessman et al. 1997). For example, in mapping C4 grass weed infestations in annual C3 crop fields, Smith and Blackshaw (2003) reported significant improvement (89%) for the identification of seven species, using classification rules derived from discriminant function analysis of SMA data. However, considerable limitations remain when using SMA data derived from broadband sensors for species discrimination. Hence, an alternative improved classification method is proposed. This involves the implementation of data fusion techniques, using higher spatial resolution multispectral data to enhance the information contents from coarser resolution data sets (e.g. Landsat TM or MODIS satellite images).

2.5.2 Narrowband hyperspectral sensors

Unlike the broadband multispectral sensors, hyperspectral sensors collect narrowband contiguous vegetation reflectance spectra across the range, which are visible to NIR to SWIR regions of the

electromagnetic spectrum. Figure 2.2 shows the typical narrowband, contiguous reflectance data and the discrete broad bands of hyperspectral and multispectral sensors, respectively. It should be noted that considerable data are lost across the spectral range (400–2 500 nm) and that spectral fine features characteristic of vegetation is not discernible from broadband sensors (CA and OM unpublished data). Further detail can be found in Kumar et al. (2001). Hyperspectral information can be acquired in less than 10 nm spectral intervals across hundreds of channels. Hyperspectral sensors used in collecting vegetation reflectance include handheld and field spectrometers, such as the Analytical Spectral Devices (ASD) (FieldSpec®, ASD, Inc., USA) and several airborne and space-borne spectrometer sensors (Kawamura et al. 2008). To date, the most commonly used hyperspectral sensors are the airborne imaging spectrometers, such as the Airborne Visible Infrared Imaging Spectrometer (AVIRIS), Compact Airborne Spectrographic Imager (CASI), HyMAP and Airborne Imaging Spectroradiometer for Applications (AISA) Eagle (Carter et al. 2005; Guanter et al. 2007; Kokaly et al. 2009; Lu et al. 2009; Schlerf et al. 2005).

The Hyperion sensor has been the only widely used space-borne hyperspectral sensor, though classical problems of high-dimensional inputs and low signal-to-noise ratio complexities have compounded and affected its utility (Rama Rao et al. 2008; Vyas et al. 2011). The CHRIS-PROBA and MSMI176 sensors are also used and recommended for applications in mountainous landscapes (Mutanga et al. 2009). The only other space-borne hyperspectral sensor, though not commercially available, is the US Air Force Research Laboratory's MightySat II, a spatially modulated Fourier Transform Hyperspectral Imager. The sensor was designed mainly for detailed terrain classification (Wang et al. 2010). Although these space-borne sensors are capable of providing systematic and regional, to near global, coverage, with relatively high temporal resolutions at regular intervals, they also offer the opportunity for obtaining relatively cost-effective data. The sensors' utility for mapping and monitoring grasslands at the landscape-scale is practical, yet research gaps remain. In recent reviews, the history of remote sensing of vegetation function (Ustin and Gamon 2010) and the main advantages of hyperspectral data for studies on functional types (Wang et al. 2010) have been discussed. Hence, the intention in the current review is not to provide an exhaustive discussion on the available sensors, but rather to highlight fundamental information that is useful for discriminating and mapping C3 and C4 grasses.

Many researchers who have proposed the use of hyperspectral data have also called for the development of robust statistical algorithms, capable of modeling the characteristic narrowband spectral features inherently associated with grass species response in mixed canopies (Carter and Miller 1994; Cochrane 2000; Fava et al. 2009; Yang et al. 2010). The potential of hyperspectral remote sensing and the associated data analysis techniques has already been realized and the capability is widely used in many fields, including forestry (Alvarez-Añorve et al. 2008) and agricultural investigations (Miglani et al. 2008; Smith and Blackshaw 2003). Remote sensing hyperspectral systems have been used to detect and discriminate between C3 and C4 plants, with an overall accuracy of 92% (Liu and Cheng 2011). Lu et al. (2009) previously reported accuracies of 90% and 87.6% for estimation of abundance and distribution of C4 (*Miscanthus sacchari-florus*) and C3 (*Phragmites australis*), respectively. Table 2.2 presents a brief summary of the application of some major sensors, including the multispectral and hyperspectral systems used for C3 and C4 grass species detection, classification, and estimation. However, this potential is yet to gain full acceptance in the respective ecological remote sensing field as a technique to facilitate the discrimination and mapping of species or communities of different functional types (Ustin and Gamon 2010).

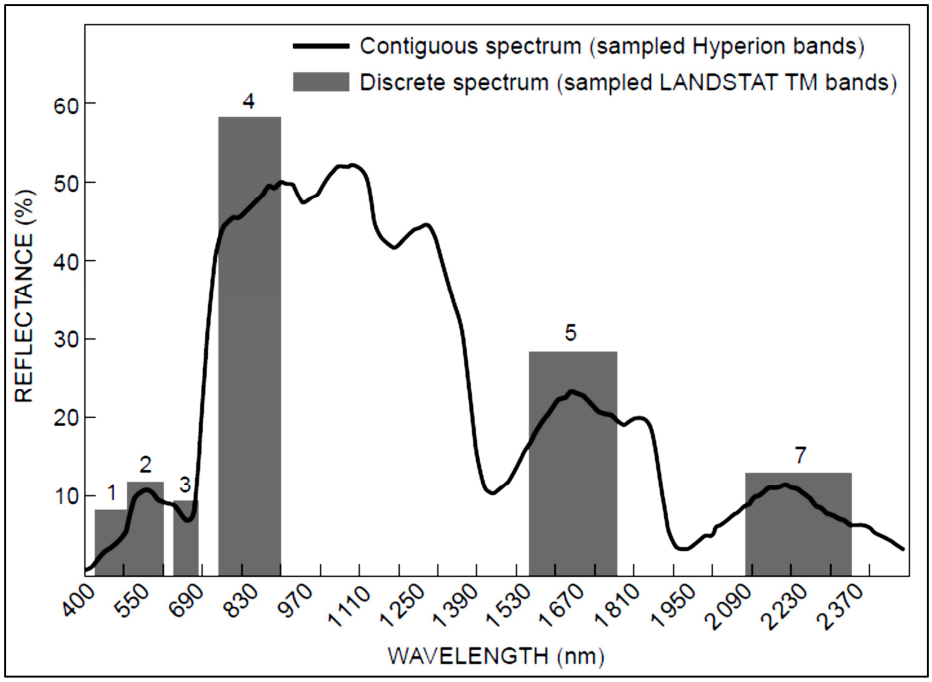

Figure 2.2 Data content of narrow-band Hyperion and broadband Landsat TM sensors. Fine absorption or reflectance spectral features characteristic of vegetation is lost from broadband sensors.

2.6 New opportunities for remote sensing of the functional types of C3 and C4 grasslands

Mapping plant functional types using remote sensing is based on the assumption that each species has a unique spectral signature that is defined by its characteristic biochemical and biophysical make-up (Kumar et al. 2001; Rosso et al. 2005). The characteristics of species reflectance spectra originating from foliar biochemicals have been extensively studied. Biochemicals such as nitrogen and lignin (Curran 1994; Mutanga et al. 2003), canopy moisture (Peñuelas et al. 1993) and plant pigments (Blackburn 2007a; Blackburn 2007b; Blackburn and Steele 1999) are major contributors of spectral variability in grassland canopy reflectance. However, within-species spectral variability poses a great challenge for the remote sensing of species. For example, it has been reported that the overlap in spectral signatures amongst functional types makes differentiation difficult because of the complex factors influencing the relatively small differences in intra-species spectral responses (Price 1992). Furthermore, Portigal et al. (1997) showed that spectral variance among species can be due to microclimatic factors, background substrates and several other environmental factors. For example, stress factors, including air pollution, heavy metals and drought, have been described that affect the spectral response of vegetation and which make discrimination between species very difficult (Portigal et al. 1997; Price 1994).

The phenological characteristics of a species, such as foliage age, flowering patterns (Gausman 1985; Roberts et al. 1998) and the position of a species in mixed canopies (Danson 1995), are also found to significantly affect vegetation reflectance. For this reason, Price (1994) concluded that spectral signatures may not be unique after all and that it was impossible to discriminate between individual plant species using remote sensing. However, with the rapid pace of advancements in vegetation spectroscopy and the increasing availability of detailed spectral information and the development of robust data mining algorithms, it becomes possible to use the large amount of information from remote sensing data to detect and discriminate between individual species.

Progress in hyperspectral sensor technology allows for the measurement of radiation in many narrow-spaced spectral wavebands. However, the high-dimensional nature of the data is a major limitation in statistical analysis. In this regard, traditional methods designed for a modest number of input spectral bands will struggle in the new high-dimensional feature space environment. Hence, advancement in non-parametric statistical data-mining algorithms, such as

artificial neural networks (Bishop 1995), random forests (Breiman 2001a) and support vector machines (Vapnik 1995; Vapnik 1998), have become attractive for selecting hyperspectral features. The random forests and support vector machines are suited for hyperspectral data mining and analyses, with some assertions that the methods are insensitive to the dimensionality problem and, therefore, may not require a dimensionality-reduction analysis in preprocessing (Melgani and Bruzzone 2004; Pal and Foody 2010).

In several application domains, studies show that support vector machines and random forests usually produce higher classification accuracy than conventional parametric statistical classifiers (Chan and Paelinckx 2008; Dalponte et al. 2009; Ismail and Mutanga 2010; Lawrence et al. 2006; Vyas et al. 2011). These techniques offer innovative ways to analyze hyperspectral data by virtue of their strong non-parametric statistical foundation and reliable empirical performance. In addition, studies have demonstrated that these algorithms are less affected by problems associated with multicollinearity and the so-called curse of dimensionality or the Hughes phenomenon (Hughes 1968) in statistical analysis (Breiman 2001b; Pal and Foody 2010). From this background, these algorithms are expected to provide a new way to produce maps that are robust to variations in class reflectance, caused by high intra- and inter-species variability and influenced by the varying environmental, biophysical and biochemical characteristics of C3 and C4 grasses. Although the algorithms have not been specifically tested for C3 and C4 grass species detection and classification, their reported generality and variable selection capabilities in other application domains offer great potential for the analysis of complex C3 and C4 grass reflectance data.

2.7 Further research needs in the remote sensing of C3 and C4 grass types

The application of techniques using hyperspectral data to discriminate between C3 and C4 grass species is generally lacking in many aspects. The reasons extend beyond lack of cost-effective data or low temporal–spatial resolutions from currently available sensors. As Price (1994) notes, there still remains an urgent need to expand vegetation spectral signature libraries so that they can account for variability in intra-species or vegetation communities. There is an urgent need to improve upon existing techniques in order to achieve the goal of a spectrally distinguishable continuum of C3 and C4 grass functional types. Recent trends in research and remote sensing

products are pushing the frontier of new algorithms being developed (Cochrane 2000; Ustin and Gamon 2010).

However, with the exception of a few closely related field-level studies (Foody and Dash 2007; Schmidt and Skidmore 2001), the evaluation of techniques that use hyperspectral data has been conducted mainly under controlled or semi-controlled experiments. For example, Irisarri et al. (2009) evaluated reflectance spectra collected from a garden experiment, under controlled nutrient and light conditions, to differentiate between C3 and C4 grasses. The authors concluded that C3 and C4 grass spectra are dependent on seasonality and it requires ad hoc sensor calibrations in order to select a specific model or set of bands sensitive enough to discriminate between mixed and pure C3 and C4 grass canopies. This potential to characterize fine spectral features, using narrowband spectral signatures developed for the species of interest, thus offers a good precedent for further research. Therefore a number of researchers have investigated the possibilities of analyzing specific spectral regions that are most sensitive to the structural and biochemical constituents of individual species (Asner 1998; Blackburn 1998; Gao 1996; Kokaly et al. 2009; Kokaly et al. 2003; Lu et al. 2009; Peñuelas et al. 1993; Thenkabail et al. 2004a; Thenkabail et al. 2000).

The challenge, however, is to investigate whether the techniques applied, mostly in the laboratory under controlled or semi-controlled conditions, can be up-scaled to assess C3 and C4 grasslands using field data and commercially available satellite data. For example, in a bid to evaluate the relationships between vegetation indices (VIs) and the relative contributions of C3 and C4 grasses to aboveground biomass, Davidson and Csillag (2001) computed a suite of vegetation VIs at various field sampling resolutions (0.5, 2.5, 10 and 50 m). The authors tested the performance of these calculated VIs, using radiometric spectra recorded in the field to estimate the relative C4 to C3 species coverage, for the aboveground live biomass and cover density. The results suggest that the commercially available satellite data may offer the potential for estimating the coverage of C4 species, though gaps in the applicability of techniques, using remotely sensed data, remain to be filled.

2.8 Conclusions

C3 and C4 grass classification represents a scheme that is consistent with the PFT approach used for the assessment of broad vegetation types, rather than the higher taxonomic identification of

individual species. Initially, the concept of PFT uses differences in plant physiology, structure and morphology. However, it has become evident in recent times that the C3 and C4 PFTs reflect responses of vegetation dynamics to changes in environmental conditions. This paper argues that the ability to monitor such changes is an important way of identifying the impacts of climate change on C3 and C4 vegetation. It is clear that, although remote sensing techniques have the potential to interpret spectral information in ways that correlate field observations with C3 and C4 grass composition, challenges still remain to effectively link physical observations made in specific wavebands to ecosystem functional processes (Schaepman et al. 2009; Ustin and Gamon 2010). However, the rapidly developing products of hyperspectral sensing offer the potential to detect and map PFTs of C3 and C4 grasslands. In summary, the evidence presented in this paper indicates that:

- differences in the structural composition between C3 and C4 grasses affect the amount and shape of radiation reflected, absorbed and/or transmitted in the wavelength spectrum;
- leaf/canopy structural and biochemical properties relate differently and affect different aspects of the electromagnetic spectrum;
- leaf and canopy radiative transfer modeling approaches can effectively measure species or communities of C3 and C4 grass reflectance characteristics;
- C3 and C4 grass canopies can be discriminated by using reflectance and absorption features and a combination of temporal trajectory of vegetation indices and;
- the integration of remote sensing techniques, solar radiation patterns, wetness gradients, and topographic elements such as slope orientation and altitudinal gradients, could improve the detection and mapping of C3 and C4 grass canopies in tropical montane grassland landscapes.

It has been made clear from the review of literature that the application of remote sensing systems, specifically, hyperspectral sensors, has been hampered by challenges of high spectral dimensionality and multicollinearity. Thus, in the following Chapter 3, an approach is proposed to overcome these problems in the classification analysis of C3 *and* C4 *grass species.*

CHAPTER 3: OVERCOMING THE HIGH SPECTRAL DIMENSIONALITY AND MULTICOLLINEARITY PROBLEMS IN CLASSIFICATION ANALYSIS

The chapter is based on:

Adjorlolo, C., Mutanga, O., Cho, A.M., & Ismail, R. (2013). Spectral resampling based on user-defined inter-band correlation filter: C3 and C4 grass species classification. *International Journal of Applied Earth Observation and Geoinformation*, 21, 535-544

Abstract

In this paper, a user-defined inter-band correlation filter function was used to resample hyperspectral data and thereby mitigate the problem of multicollinearity in classification analysis. The proposed resampling technique convolves the spectral dependence information between a chosen band-centre and its shorter and longer wavelength neighbours. Weighting threshold of inter-band correlation (WTC, Pearson's r) was calculated, whereby r = 1 at the band-centre, and bands with coefficients beyond a chosen threshold were assigned r = 0. Various WTC (r = 0.990 to r = 0.999) were analyzed based on pre-selected 13 band-centres. An optimum WTC threshold of 0.992 was obtained by scanning all the r- values between 0.900 and 1.The resultant, optimum WTC dataset was used in the random forest analysis to classify in-situ C3 and C4 grass canopy reflectance. The optimum WTC r = 0.992 dataset yielded improved classification accuracy (kappa = 0.81) with less correlated wavebands when compared to spectra resampled respectively to 5-nm and 10-nm fixed band-widths across 400 – 2500 nm spectrum (Kappa = 0.70 and 0.76) and, resampled-Hyperion bands (kappa = 0.76). In addition, the optimum WTC dataset yielded better classification results when compared to spectra resampled to the 13 pre-selected band-centres without using the WTC filter. Overall, the results obtained from this study suggested that resampling of hyperspectral data should account for the spectral dependence information to improve overall classification accuracy as well as reducing the problem of multicollinearity.

Keywords: Spectral resampling; Inter-band correlation; Grass species classification; Random forests

3.1 Introduction

Discriminating grass species, which correspond to the 3-carbon (C3) or 4-carbon (C4) photosynthetic pathways, is consistent with the plant functional type (PFT) approach used in land surface modelling schemes (Tieszen et al. 1997; Ustin and Gamon 2010). In general, C3 and C4 grasses differ significantly in a number of physiological and anatomical characteristic features. The C4 type of grass species has more compact leaf mesophyll, higher proportion of vascular tissue and a lower interveinal distance, than those of C3 grasses. In addition, several biochemicals such as intercellular air-moisture and nitrogen concentration are relatively lower in C4 grass, compared to C3 grass species (Oyarzabal et al. 2008). Such differences can manifest in the composition of C3 and C4 grasslands, with the dominant species strongly constituting the canopy reflectance. The fundamental principle is that vegetation canopy reflectance is directly dependent on the spectral properties, which are in turn, controlled by the biophysical and biochemical characteristics (Blackburn 1998; Curran 1994; Mutanga and Skidmore 2004a). Several empirical evidences have shown that the spectral variability between C3 and C4 grass species or groups of grasses is greater than the within group spectral information (Irisarri et al. 2009; Liu and Cheng 2011; Noble et al. 2002; Smith and Blackshaw 2003). For example, reflectance at 531 and 570 nm have been proposed as a set of spectral bands sensitive to differences in C3 and C4 species (Gamon et al. 1997). Slaton et al. (2001) (Slaton et al. 2001) modelled the near infrared (nearIR) region and found that the reflectance around 800 nm is significantly different among species, which differ in intercellular structure. In a common garden experiment, Irisarri et al. (2009) demonstrated that it is possible, using reflectance centred around 820 nm to differentiate between C3 and C4 grass compositions.

The major challenge, however, is that spectral reflectance data obtained over many narrow contiguous channels (i.e. hyperspectral data) can represent multiple classes that are often mixed for a limited training-sample size (i.e. $n < P$: Chi and Bruzzone (2007). The problem of $n < P$ is associated with the well-described “curse of dimensionality” or the Hughes phenomenon (Hughes 1968). This phenomenon causes a decrease in a classifier ability to generalize accurately (Ham et al. 2005; Pal and Foody 2010). Hence, very large training-samples are required to achieve a good description of data distribution (Dalponte et al. 2009). Besides, the Hughes phenomenon often introduces high degree of multicollinearity, caused by the use of highly-correlated predictor variables (Clevers et al. 2007). Multicollinearity is a prominent

problem in processing hyperspectral data for vegetation applications, due to similarities in the reflectance properties of biophysical and biochemical characteristics (Ferwerda et al. 2005; Knox et al. 2010; Zhang et al. 2011). The problem of multicollinearity in the matrix of input spectral bands often leads to highly unstable parameter estimates and thus generalization error for a classifier (Bruzzone and Serpico 2000; Clevers et al. 2007).

Attempts to solve the problems associated with spectral dimensionality and the related co-linearity include the use of feature reduction and feature selection techniques. The feature selection approach includes those based on a search strategy and on a separability measure. The Sequential Forward Floating Selection (Pudil et al. 1994) and the Steepest Ascent (Serpico and Bruzzone 2001) are commonly used search strategy techniques, whereas the Bhattacharyya distance (Djouadi et al. 1990), Jeffries-Matusita distance (Bruzzone et al. 1995) and the transformed divergence distance (Su et al. 1990) are examples of the separability measures used in processing hyperspectral data. However, these feature selection techniques require estimation of some statistical properties at full dimensionality, in order to select optimum subset of the input spectral bands for a given classification task. If the training samples are insufficient, the parameterization may not be reliably adequate for the feature selection process (Chi and Bruzzone 2007).

The studies by Schmidt & Skidmore (2003) and Becker et al. (2007) used the approach of analyzing the most sensitive wavebands, considering the physical or spectroscopic meaning of each band across the spectrum. This approach often involves the resampling of high-dimensional spectra to wider bandwidth intervals, around a few chosen band-centres or to the spectral configuration of existing sensors. In the case of resampling spectra to some existing sensors, their respective spectral response functions or spectral resolutions (i.e. Full Width at Half Maximum, FWHM) are simulated. The major limitation of existing resampling methods is that an inherent property of vegetation spectral response is not considered. That is, the asymmetrical nature of correlation between a given waveband (λ) and its shorter and longer wavelength neighbours are not fully accounted for. In this regard, a more innovative approach can be followed, whereby the researchers consider the inter-band correlations around each band centre of interest. The approach has the advantage of linking the physical properties of the target vegetation and its characteristic spectral response features across the spectrum (Schmidt and

Skidmore 2003). When hyperspectral data are processed in this way, classifications are based on the spectroscopic interpretability of each set band (Becker et al. 2007; Fourty and Baret 1997).

From this background, the present study sought to classify C3 and C4 grass canopies using resampled hyperspectral data obtained through an approach that convolves the spectral information around a chosen given band-centre. The resultant datasets were analyzed using the random forest (Breiman 2001a) algorithm. Random forests are advanced non-parametric classifiers, which are increasingly becoming recognized in remote sensing applications that classify vegetation (Chan and Paelinckx 2008; Ghimire et al. 2010; Ismail and Mutanga 2010; Lawrence et al. 2006). Included in the random forest computation are embedded methods of assessing the generalization error, variable importance measures and the computation does not require tuning of many parameters (Breiman and Cutler 2004).

3.2 Data acquisition and methods

3.2.1 Field spectral data measurements

Field data collection was conducted in the Cathedral Peak region of the Drakensberg Mountain Range, South Africa. The region consists of vegetation divided into altitudinal zones, which correspond closely with the physiographic features of the Drakensberg Mountains (Hill 1996; Killick 1963). Three zones namely, the Montane belt (1280–1829 m), the Sub-alpine belt (1930–2865 m) and the Alpine belt (2866–3353 m) above mean sea level, are defined. These zones also coincide with three terraces in the Drakensberg. These include the river valley system, the foothills (also known as the Little Berg), and the summit areas, respectively. The sub-alpine belt, which is composed of C3 and C4 grass species, is also known as the *Themeda-Festuca* sub-alpine grassland (Hill 1996). This zone is further divided into three grass communities denoted as the *Themeda triandra*, *Festuca costata* and the 'Mixed' grasslands. The so-called mixed grasslands community consists mainly of variable proportions of C4 grasses, although the occurrence of *Rendlia altera* (C4 grass) seems prevalent. Nonetheless, within the sub-alpine zone there are correlational occurrences of the *T. triandra* (C4) and the *F. costata* (C3) grass species on warm, northerly and cool, southerly slopes, respectively.

Reflectance measurements were collected during the December 2010 summer growing season, using a 2150 band (350–2500 nm resolution) Analytical Spectral Device (ASD), field spectroradiometer (FieldSpec®3 ASD, Inc., Boulder, CO, USA). This device uses a fibre optic

cable set at 25^{o} field of view (FOV) to record reflected canopy radiation, which was individually calibrated against a barium sulphate ($BaSO_4$) white reference panel. Canopy reflectance measurements were collected to characterize the spectral separability among 1 m × 1 m sample plots, represented by *F. costata*, *R. altera* and *T. triandra* dominant grass species. Spectral reflectance data for the grasses (i.e. in the 1 m × 1 m plots) were measured at full canopy cover. Although the dominant grasses co-exist with other species, their respective canopy cover was consistently estimated at ≥ 80% in each target 1 m × 1 m plot.

The field spectral measurements were consistently recorded, considering the recommendations in Thenkabail et al. (2000). The ASD optic sensor was held at about 1.5 m directly above the sampling plots, generating an instantaneous field of view of about 0.35 m^2. A minimum of three positions were randomly chosen within each 1 m ×1 m plot and five spectral measurements were consistently acquired for each one of these positions. This process resulted in a minimum of 15 reflectance measurements per plot. There were no major issues with background effects, since average spectral ($\overline{r}_i$) measurements for each plot (i ≥ 15) were taken at full canopy cover. A total of 110 plots were measured for each of the three categories of grass species. This process resulted in 330 sample plots, which were considered representative of the spectral variability within and among the grass species under investigation.

3.2.2 Data transformation

Spectral noise results partly from limitations of the sensor's (i.e. ASD field spectroradiometer) registration of radiation within its operating range, multiple reflections from the feature space, inconsistent atmospheric and illumination conditions or possible lower signal to noise ratio of the sensing instrument (van der Meer et al. 2001). In order to ensure significantly high quality spectra, noise removal is often necessary before spectral analysis of hyperspectral data. In this regard, different techniques have been applied to smooth noisy hyperspectral data, for applications involving vegetation species classification analysis (Irisarri et al. 2009; Koger et al. 2003; Thenkabail et al. 2000). For example, techniques such as the least squares polynomial fitting approach of Savitzky-Golay (King et al. 1999; Ruffin and King 1999; Savitzky and Golay 1964), mean filters, cubic spline and wavelet-transform (Koger et al. 2003) methods have been commonly used. In this paper, the Savitzky-Golay algorithm for removing noise was implemented to smooth the ASD spectra ($\overline{r}_i$), in order to avoid noise associated with specific

bands. The Savitzky-Golay smoothing approach is a widely used technique in absorption spectroscopy studies that focus on filtering and differentiating vegetation reflectance spectra (Chen et al. 2004; Shafri and Yusof 2009). Further explanation on the Savitzky-Golay smoothing/filtering method for hyperspectral data analysis can be found in King et al. (1999).

3.2.3 Resampling the spectral data

The proposed inter-band correlation filter technique was compare to traditional of spectral resampling methods. In this experiment, the ASD reflectance data were processed using ENVI's spectral resampling routine (ENVI Version 4.7, 2009 Edition, Copyright © ITT Visual Information Solutions). Initially, an ASCII file containing 10 nm band-widths was created and used to aggregate the 1-nm-wide $\bar{r}_i$ spectra across the 400–2500 nm spectrum. The ENVI's resampling routine fits a Gaussian model with an FWHM equal to a specified band spacing to resample the data. This initial sampling of the data was carried to aid calculation of the inter-band correlation coefficient matrix of the input spectral bands. The degree of linear relationship between a band and its shorter and longer wavelength neighbours was calculated, using the well-known Pearson's r coefficient of correlation. The linear spectral dependence between two sample wavebands (Xi, Yi) was assessed, resulting in values between 0 and 1:

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{s_X}\right)\left(\frac{Y_i - \bar{Y}}{s_Y}\right) \qquad \text{(Eq. 3.1)}$$

where $\frac{X_i - \bar{X}}{s_X}, \bar{X}, \text{ and } s_X$ are the standard score, sample mean, and sample standard deviation, respectively (Eq. 3.1). The Pearson's r correlation coefficient was used, since in this experiment the data assume multimodal normal distribution of the response variables (Chi et al. 2008). The inter-band r correlation coefficient matrix was computed using the R statistical software (R Development Core Team 2010). The R routines output a spreadsheet file format and an x : y axis correlation plot of the inter-band r coefficient values.

Vegetation spectral response curves were simulated for predefined thirteen (13) band-centres. The band-centres were chosen on the basis of their known sensitivity to the biophysical and biochemical characteristics of vegetation. Table 3.1 shows some causal reflectance or

absorption features of interest, associated with the chosen bands-centres. The choice of these band locations, among many other wavebands of known effects, was partly guided by the hotspots observed in the coefficients of correlation. Significant number of the selected band-centres has been reported in previous studies (e.g. Irisarri et al. 2009; Liu and Cheng 2011; Noble et al. 2002; Smith and Blackshaw 2003) that investigated the utility of remotely sensed data for C3 and C4 grass species classification. In addition, the band-centres were selected, considering the observed pattern in the matrix of the data points, as depicted in Figure 3.1a. Moreover, the chosen band-centres have been reported in the literature to be useful for C3 and C4 grass species discrimination. For example, Smith and Blackshaw, (2003) reported high frequencies (>15 out of 20 times) for the band-centres chosen within the visible and nearIR regions. Irisarri et al. (2009) found that the spectral channels around 820 nm were important for differentiating C3 and C4 grass compositions. Further, Noble et al. (2002) noted that the chosen bands-centres in the shortwave region are useful for C3 and C4 crop/weed species discrimination.

The inter-band Pearson's r correlation coefficient between each of the chosen band-centres and their shorter and longer waveband neighbours was calculated. The band-centres were located at the meeting point of the x : y axis (Figure 3.1a), where r = 1, and bandwidths were estimated by considering the vertical or the horizontal distance across a given band-centre, as a function of wavelength. It is important to note that the inter-band correlation, r values are asymmetrical across the horizontal or the vertical lines. They are only symmetrical across the diagonal. Therefore, it was possible to capture the spectral response information around each band-centre, using a chosen weighting threshold of inter-band correlation. Various weighting thresholds of user-defined inter-band correlation filters were assessed by scanning WTC r = 0.990 to WTC r = 0.999, with an increment of r = 0.001. Figure 3.1b illustrates three sizes of the user-defined inter-band correlation filter, using the 660 nm band-centre as an example. The shorter and longer wavelength sides of the band-centre (i.e. 660 nm) were calculated on the basis of $r \geq 0.99$, $r \geq 0.95$ and $r \geq 0.90$, and bands with coefficients beyond a specified WTC were assigned r = 0. It is important to note that r = 1 at the band-centre and largely decreases to the left or the right sides of the band-centre. The resultant user-defined inter-band correlation filter functions, contained in an ASCII file format, were used to implement the spectral resampling. ENVI 4.8 (RSI 2010) routine was use to resample the hyperspectral ASD data. To resample the

spectral response from the ASD to the 13-preselected band-centres, the spectral integration was calculated as:

$$R(w_i) = \frac{\int R(w) \cdot S(w) dw}{\int S(w) dw} \quad \text{(Eq. 3.2)}$$

whereby $\boldsymbol{R(w_i)}$ is the ASD sensors response of the i-th channel, $\boldsymbol{S}$ is the spectral sensitivity of the channel (presented by the correlation coeeficients), **R(w)** is the ASD reflectance for each wavelength $\boldsymbol{w}$, which is considered continuous. The spectral response is assumed to be a Gaussian distribution (van der Meer et al. 2001) for the 13 band settings, which is modelled as:

$$R(w) = \frac{1}{\sigma \cdot \sqrt{2\pi}} e^{-\frac{(w-\mu)^2}{2\sigma^2}} \quad \text{(Eq. 3.3)}$$

where $\boldsymbol{w}$ is the continuous spectrum of the ASD spectroradiometer, while μ and σ are the band-centre and the standard deviation for the i-th multispectral band setting.

Table 3.1 Selected wavelengths corresponding to known absorption features, as described in previous studies to be highly sensitive to the properties of reflection or absorption of vegetation structural and biochemical characteristics. These band centres are compiled from the list of literature sources.

No.	Band (nm)	Known causal compound/ feature	Source
1	470	Total plant pigment concentration	Blackburn, 1998
2	530	Chlorophyll a absorption	Gamon et al., 1997
3	600	Nitrogen	Fourty and Baret, 1997
4	660	Nitrogen	Carter, 1994
5	700	Total chlorophyll, nitrogen	Carter, 1994
6	720	Total chlorophyll, leaf mass	Horler et al., 1983
7	820	Leaf mass, leaf area index	Curran, 1994
8	1540	Cellulose, vegetation water content	Carter, 1994
9	2060	Protein	Carter, 1994
10	2280	Cellulose, sugar, starch, leaf mass	Curran, 1994
11	2300	Leaf mass, vegetation water content	Carter, 1994
12	2450	Cellulose, protein, nitrogen	Carter, 1994
13	2470	Cellulose, protein	Kumar et al., 2001

In addition, the WTC method is compared against resampling reflectance data to a fixed band-width without weighting. This involved resampling the smoothed ASD data to 5- and 10-nm-bandwith interval across the 400 – 2500 nm spectrum and, to match the Hyperion sensor's

(on-board the Earth Observing-1 Satellite) spectral resolution or FWHM function. Furthermore, the smoothed data were resampled based on the 13 preselected band centres with a symmetrical integration of fixed band-widths (5, 10, 30, 50 and 100 nm). The spectral resampling procedure for the additional datasets is analogous to that of the user-defined spectral resampling described above, with the ENVI routine assuming FWHM value equivalent to the specified band-widths. In the case of the resampled-Hyperion, the ENVI routine uses the pre-defined spectral library developed for the Hyperion sensor to resample the data. Since the canopy reflectance measurements were conducted under field conditions, the strong noisy incident radiation in 1350-1460 nm, 1790 - 1960 nm and inconsistent spectra below 400 nm were removed from all analysis (Thenkabail et al. 2004b).

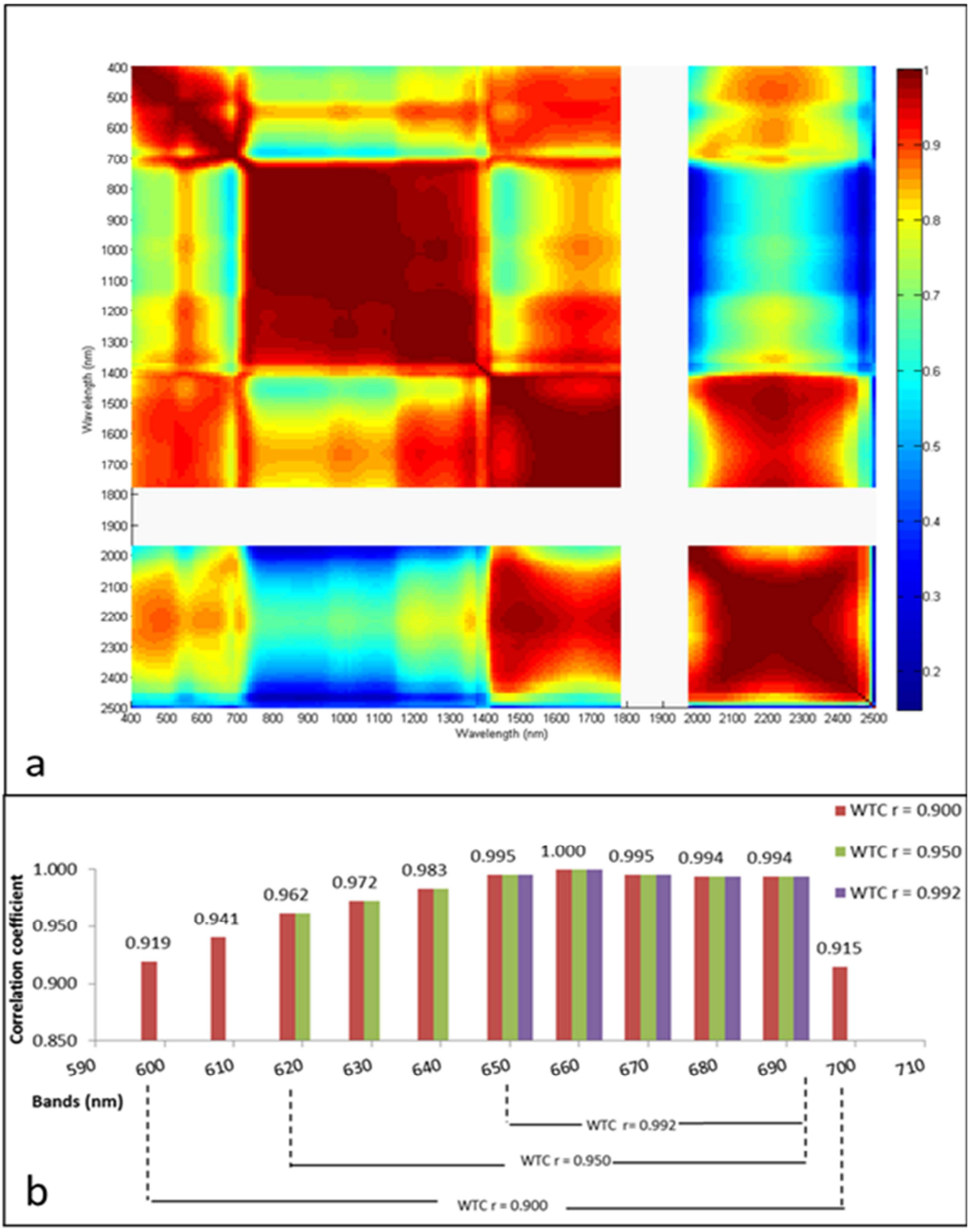

Figure 3.1Pearson's r correlation coefficients matrix (plot) of the input spectral bands, calculated using reflectance data aggregated into 10-nm-wide band intervals (a) and an illustration of the user-defined inter-band correlation filter for 660 nm band-centre (b). The white space (Figure a) represents removed noisy bands.

3.2.4 The Random forest's variable importance, classification and accuracy assessment

All the datasets were randomly split into 70% training and 30% holdout test sets (n = 77 and n = 33 subsets, respectively). The random forest algorithm was constructed to grow a large ensemble of classification trees. The resultant trees in the ensemble were used to assign the test observation, a class membership of the response variables: *F. costata*, *R. altera* or *T. triandra*. Each tree is grown from a randomly and independently selected bootstrap sample of the training data, and about one-third, excluded samples, called the out of bag (OOB) samples were used to calculate an unbiased assessment of the classification accuracy (i.e. the OOB error). Since the OOB error is an unbiased assessment of the classification accuracy (Breiman 2001a; Prasad et al. 2006), it provides theoretical guarantee for the groups of C3 and C4 grass species detection and classification. Further to using the OOB error samples to assess the overall classification accuracy, the kappa analysis was performed. This was necessary, since the study involves a multiclass application and the goal is to account for actual agreement specified by each class versus the chance agreement. That is, it was important to determine if one OOB error matrix is significantly different from another (Stehman 1997).

The random forest algorithm is easy to implement, because the user tunes only two parameters: (i) the number of trees (*ntree*) to grow and (ii) the number of variables to split at each node (*mtry*). The default value of the mtry parameter in the context of classification applications is denoted by the square root of the total number of input variables (Liaw and Wiener 2002). In the current analysis, the OOB error samples for each class membership of the input spectral bands were used to optimize the ntree and mtry hyper-parameters (Ismail and Mutanga 2011), based on the specified ntree (i.e.10 000 for all datasets) value.

The random forest mean decrease in accuracy (Strobl and Zeileis 2008) was used to calculate the importance of each predictor variable. The variable rankings were calculated using all variables (i.e. 70% training set) and optimize mtry and mtry values. To decrease computing time the routine starts with the default mtry value for each dataset and then calculates a multiplicative factor of the default mtry to the right and then to the left of the value. For example, the default mtry for the resampled Hyperion dataset (n = 197) was 14, so the routine uses mtry values of 14 (× 1/2), 14 (× 1/3)… to the left (deflate) of the default value and mtry values of 14 (× 1), 14 (× 2) … to the right (inflate) of the default mtry value. It then runs the random forest

based on the optimized mtry and ntree values and determines variable rankings and the test dataset error.

3.2.4.1 Random forest-based fast forward variable selection

To calculate the greedy fast forward variable selection (FvS) using the OOB error rates (Adam et al. 2009; Dye et al. 2011), the routine uses the optimized random forest variable rankings calculated above to create different subsets of variables. It involves iteratively fitting the random forest model on the 70% training datasets, and at each iteration building a new model by adding the band with highest importance. Consequently, the routine optimizes the mtry and ntree values for each step of the variable selection process. To decrease computing time, the routine was set to terminate at the iteration with subset OOB error less than the overall OOB error calculated when using all the variables.

Several empirical evidences from other application domains have shown that the random forest algorithm shows significant preference towards highly correlated predictor variables (Nicodemus et al. 2010; Strobl et al. 2008). The authors reported that traditional random forests preference for highly-correlated predictor variables can be carried forward to any significance test or variable selection processes constructed from the importance measures. In this respect, researchers have suggested the use of conditional variable importance approach to mitigate the problems associated random forest variable selection process.

3.3 Results

3.3.1 Optimizing the WTC-derived features

The results obtained indicate that large portions of the C3 and C4 grass canopy reflectance exist in highly correlated wavelengths. In general, an increase in spectral resolutions (i.e. band-widths) was observed for each of the 13 band-centres with an increasing WTC r = 0.990 to r = 0.999, using increment interval of 0.001 (n = 100). For each derived waveband, the inter-band correlation coefficient r = 1 at the band-centres generally decreases across the shorter or longer wavelengths neighbours, as quantified by a chosen WTC filter. The overall classification error calculated based on the OOB samples for the WTC r = 0.992 was significantly lower, among the n = 100 thresholds assessed. Figure 3.2 shows the OOB value obtained for all WTC datasets, with an indication of the WTC r = 0.992 dataset which yielded the best OOB error = 0.185.

Thus, WTC with r = 0.992 limit was considered to be the optimum threshold for obtaining spectral resolutions, in this study. In addition, it appeared that the band-centres of specific regions (i.e. the visible, red-edge, near infrared and shortwave infrared spectra) showed varying degrees of inter-band correlations, resulting in variable spectral resolutions, for each of the spectral regions. Table 3.2 shows the spectral resolutions obtained for the optimum WTC and those obtained for spectra resampled, using the chosen band-centres, but without WTC inter-band correlation filtering. The resampling approach without WTC filtering involved smoothing the input spectral bands to increase signal-to-noise ratio (SNR). Figure 3.3 illustrates the average reflectance for the target grass species before and after smoothing spectra with the Savitzky-Golay algorithm for the 10-nm-wide spectral data, illustrated using a 3-moving window.

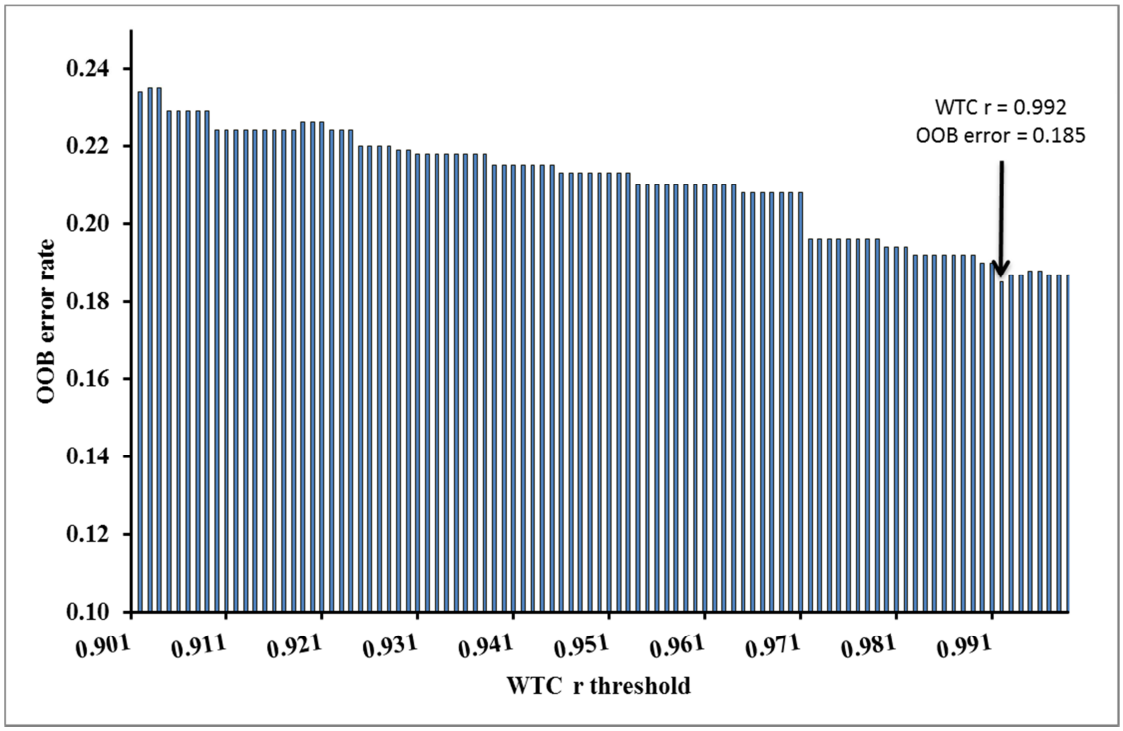

Figure 3.2 WTC feature selection for the classification of the three grass species.

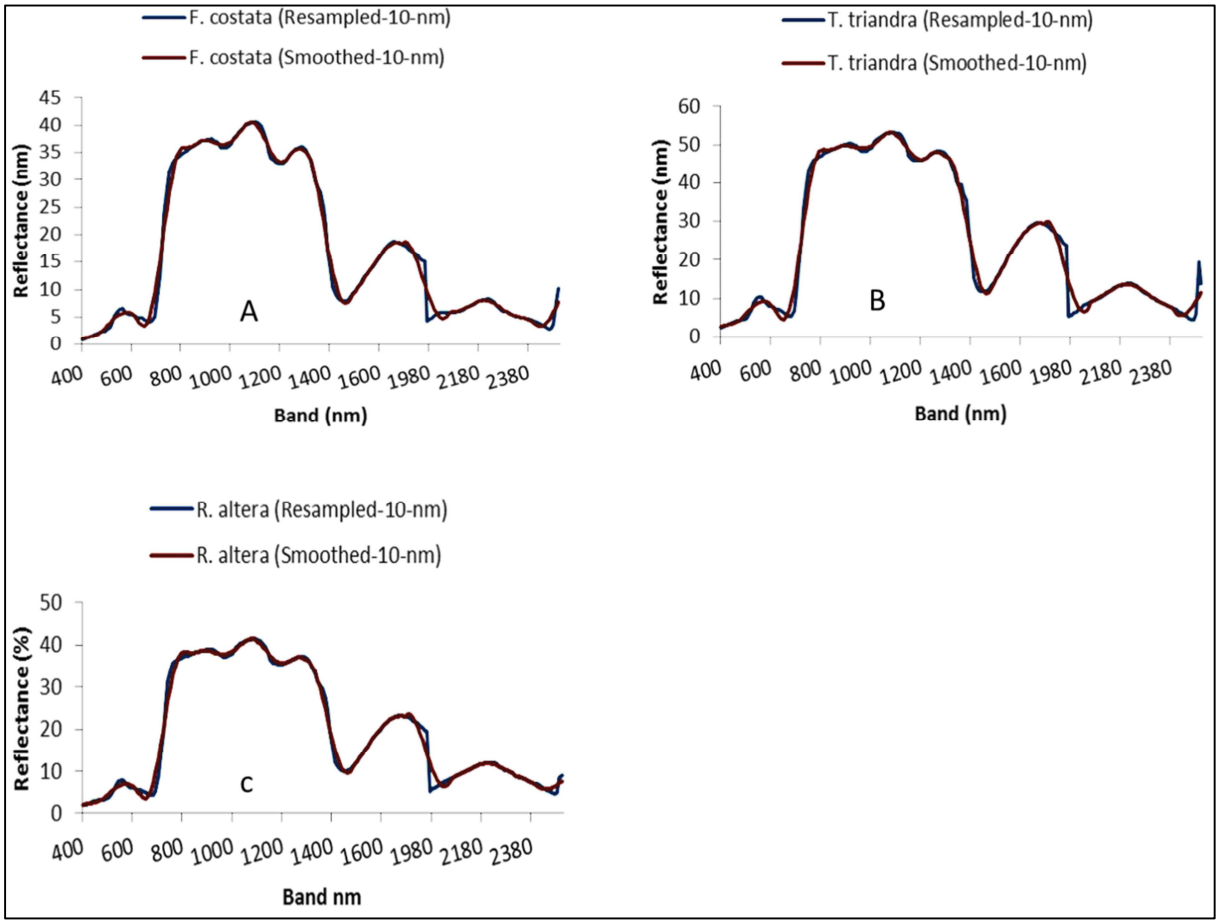

Figure 3.3 An illustration of resampled 10-nm band ASD data and, resample-10nm Savitzky-Golay-smoothed data with a 3-band window filtering. A, B and C are average spectrum for F. costata, T. triandra and R. altera, respectively.

Table 3.2 Spectral resolutions obtained from the analysis of the user-defined waveband datasets.

Band-centre (nm)	WTC-waveband (WTC r = 0.992)	Band-width					
		WTC-resampled (nm)	Resampled-5 (nm)	Resampled-10 (nm)	Resampled-30 (nm)	Resampled-50 (nm)	Resampled-100 (nm)
470	440 – 500	60	5	10	30	50	100
530	520 - 560	40	5	10	30	50	100
600	580 - 610	30	5	10	30	50	100
660	650 - 690	40	5	10	30	50	100
700	700*	10	5	10	30	40	100
720	720*	10	5	10	30	50	100
820	750 - 1000	250	5	10	30	50	100
1540	1510 - 1590	80	5	10	30	50	100
2060	2050 - 2080	30	5	10	30	50	100
2280	2110 - 2300	190	5	10	30	50	100
2300	2280 - 2340	60	5	10	30	50	100
2450	2450*	10	5	10	30	50	100
2470	2470*	10	5	10	30	50	100

* Indicates wavelength = 10 nm of the input spectral band.

3.3.2 WTC-resampled data and C3 *and* C4 *grass species classification*

The random forest hyper-parameters were optimized using the OOB error rates. WTC r = 0.992 filter yielded the highest classification accuracy. Table 3.3 shows the overall accuracies calculated based on the OOB samples and the kappa coefficients obtained for all developed datasets. The OOB error rates and kappa coefficients increased substantially when the resampled-100-nm dataset was analyzed. In general, the results showed that the proposed user-defined inter-band correlation filter approach can be used to resample hyperspectral data, and can yield similar or higher classification accuracies, compared to data obtained using conventional techniques of resampling spectra to fixed band-widths or to match the spectral resolution of existing sensors (e.g. the resampled-Hyperion).

The variable importance ranking obtained for the WTC r = 992 dataset is presented in Figure 3.4. The random forest algorithm was initially run using the full spectrum data for the fixed band-width datasets: Resampled-5-nm and Resampled-10-nm data (n = 359 and n = 205, respectively), as well as the resampled Hyperion bands (n =197). Next, the variable importance measure was exploited to evaluate whether the FvS process could improve the classification accuracy for these datasets. The results obtained showed that the three sets of data generally yielded a similar pattern with regard to variable importance for the input spectral regions. Figure

3.5 illustrates the random forest-based variable importance measures obtained for the resampled Hyperion dataset. The optimal subsets of the resampled-Hyperion bands, obtained from the random forest-based FvS procedure, are presented in Table 3.4. The band B7 of the resampled-Hyperion, which yielded the best mean decrease in accuracy also ranked highest in the subsequent variable selection process. The inter-band correlation coefficient matrix of the optimal subset selection through the random forest-based FvS process was exploited. As expected, the results showed that the random forest-based FvS procedure can yield highly correlated bands. Table 3.5 illustrates matrix of selected bands (n = 22) for the resampled Hyperion. The majority of the selected bands were concentrated in specific regions of the spectrum.

Table 3.3 Random forest (RF) model optimization and accuracy measures using OOB samples on the training dataset (70% sample). The Kappa-test set statistics were calculated using 30 % holdout samples.

Category	Resampled datasets	Number of bands	Optimized mtry	Optimized ntree	OOB error rate	Kappa-training set	Kappa-test set
A	**WTC-based resampling***						
	WTC-resampled	13	4	4000	0.14	0.80	0.81
B	**Smoothed-1-nm spectra**						
	Smoothed-3-band-window	13	3	4000	0.25	0.62	0.67
	Smoothed-5-band-window	13	3	4000	0.19	0.71	0.69
	Smoothed-7-band-window	13	4	4500	0.18	0.68	0.70
	Smoothed-9-band-window	13	4	4000	0.18	0.68	0.71
C	**Resampling based on preselected band-centres**						
	Resampled-5-nm	13	3	4500	0.19	0.68	0.68
	Resampled-10-nm	13	3	4000	0.14	0.76	0.80
	Resampled-30-nm	13	4	3500	0.16	0.75	0.79
	Resampled-50-nm	13	6	4000	0.21	0.69	0.71
	Resampled-100-nm	13	3	5000	0.23	0.69	0.69
D	**Resampling based on full-spectrum data (400 – 2500 nm)**						
	Resampled-5-nm	359	19	500	0.22	0.71	0.68
	RF-subset selection (Resampled-5-nm)	29	10	1000	0.16	0.73	0.70
	Resampled-10-nm	205	14	500	0.22	0.71	0.71
	RF-subset selection (Resampled-10-nm)	19	4	3500	0.18	0.73	0.76
E	**Resampled-Hyperion band settings**						
	Resampled-Hyperion	197	56	500	0.19	0.71	0.73
	RF-subset selection	22	20	1500	0.19	0.72	0.76

* Random forest analysis of the optimum WTC resampling without spectral smoothing.

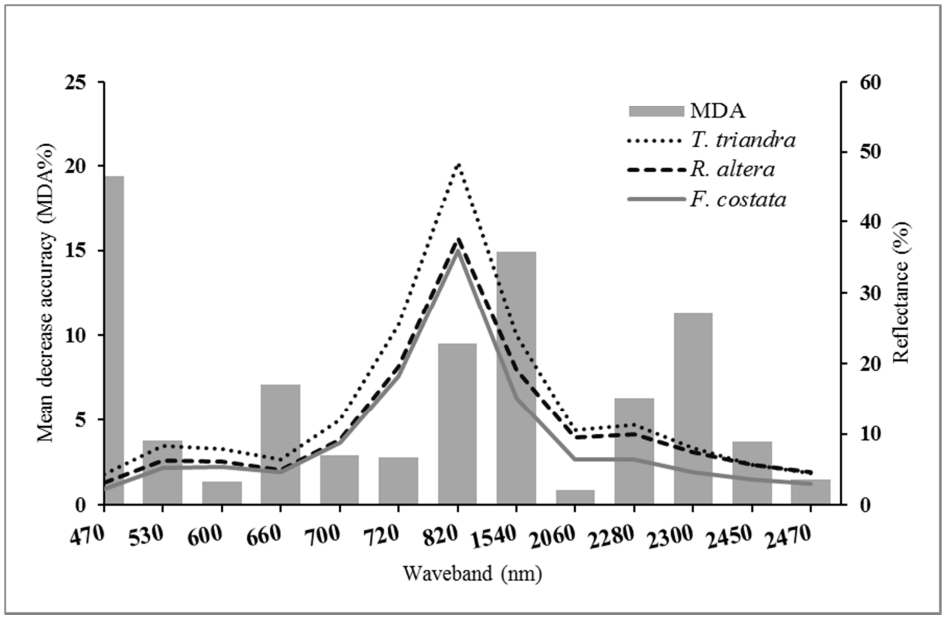

Figure 3.4 Random forests variable importance for WTC r = 0.992 dataset (n = 13 bands), based on the mean decrease in accuracy (MDA) values. The reflectance spectrum of the target grass species is shown in their respective line as in the legend.

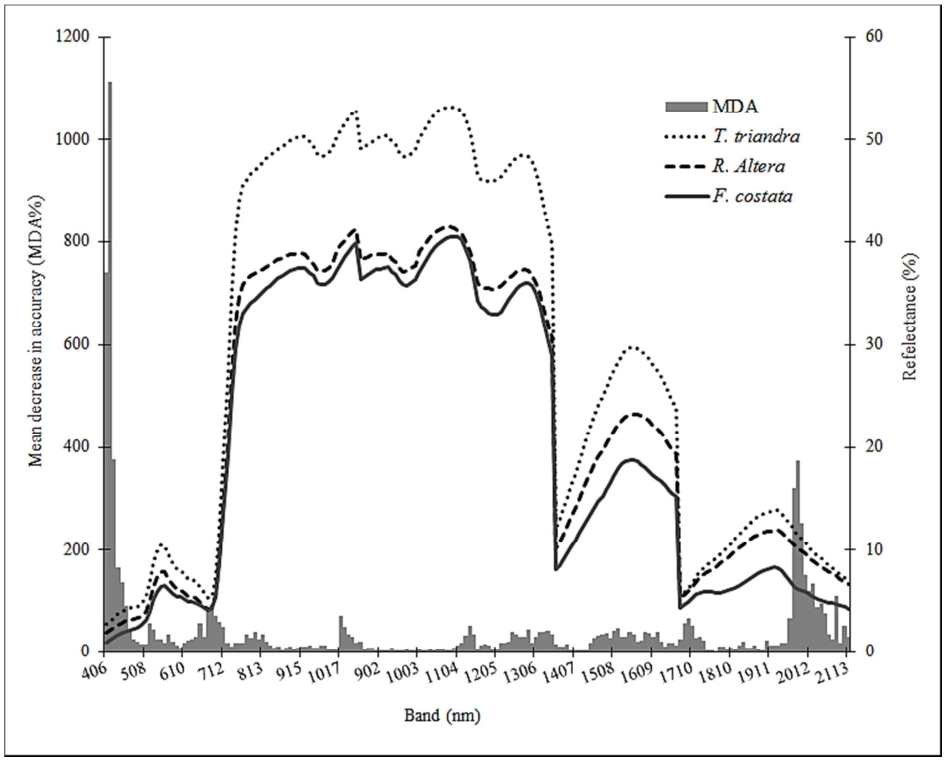

Figure 3.5 Random forests variable importance for the resampled Hyperion bands (n = 197), based on the mean decrease in accuracy (MDA: the values are scaled by a factor of 1000). The reflectance spectrum of the target grass species is shown in their respective line as in the legend.

Table 3.4 Random forest-based forward best ranked band selection on resampled-Hyperion spectral resolution. The Kappa statistics were calculated on 70 % training sample.

Rank	Hyperion band	Average wavelength (nm)	FWHM (nm)	Optimized mtry	Optimized ntree	Accuracy: cumulative OOB
1	B7	416.64	11.39	1	500	0.518
2	B6	406.46	11.39	2	10000	0.595
3	B212	2274.42	10.43	3	1000	0.693
4	B211	2264.32	10.44	4	1500	0.68
5	B8	426.82	11.39	4	500	0.693
6	B213	2284.52	10.42	4	7500	0.699
7	B9	436.99	11.39	7	1500	0.706
8	B216	2314.81	10.41	1	7000	0.699
9	B214	2294.61	10.41	6	500	0.706
10	B215	2304.71	10.41	6	2500	0.693
11	B217	2324.91	10.41	9	1000	0.693
12	B10	447.17	11.39	12	500	0.706
13	B210	2254.22	10.46	1	500	0.699
14	B12	467.52	11.39	8	500	0.712
15	B218	2335.01	10.41	2	1000	0.725
16	B11	457.34	11.39	4	500	0.725
17	B219	2345.11	10.41	4	6500	0.732
18	B33	681.2	10.33	18	500	0.771
19	B198	2133.24	10.73	8	500	0.771
20	B200	2153.34	10.68	4	1000	0.764
21	B158	1729.7	11.56	20	1000	0.803
22	B13	477.69	11.39	20	1500	0.81

Table 3.5 High correlated variables from random forest's importance rank and selection process.

	Resampled Hyperion Wavelength (nm)																					
	406.46	**416.64**	**426.82**	**436.99**	**447.17**	**457.34**	**467.52**	**477.69**	**681.2**	**1729.7**	**2133.24**	**2153.34**	**2254.22**	**2264.32**	**2274.42**	**2284.52**	**2294.61**	**2304.71**	**2314.81**	**2324.91**	**2335.01**	**2345.1**
406.46	1																					
416.64	0.99	1																				
426.82	0.99	0.99	1																			
436.99	0.99	0.99	0.99	1																		
447.17	0.99	0.99	0.99	0.99	1																	
457.34	0.98	0.99	0.99	0.99	0.99	1																
467.52	0.98	0.98	0.99	0.99	0.99	0.99	1															
477.69	0.97	0.98	0.99	0.99	0.99	0.99	0.99	1														
681.2	0.74	0.77	0.79	0.81	0.82	0.83	0.85	0.87	1													
1729.7	0.87	0.88	0.89	0.9	0.9	0.9	0.91	0.91	0.78	1												
2133.24	0.85	0.85	0.86	0.86	0.87	0.87	0.87	0.87	0.78	0.93	1											
2153.34	0.85	0.86	0.86	0.87	0.87	0.88	0.88	0.88	0.79	0.94	0.99	1										
2254.22	0.86	0.87	0.87	0.87	0.88	0.88	0.88	0.88	0.77	0.95	0.99	0.99	1									
2264.32	0.86	0.86	0.87	0.87	0.87	0.87	0.87	0.87	0.76	0.94	0.99	0.99	0.99	1								
2274.42	0.86	0.86	0.86	0.87	0.87	0.87	0.87	0.87	0.76	0.93	0.99	0.99	0.99	0.99	1							
2284.52	0.85	0.86	0.86	0.86	0.86	0.87	0.87	0.87	0.76	0.93	0.99	0.99	0.99	0.99	0.99	1						
2294.61	0.85	0.85	0.86	0.86	0.86	0.86	0.86	0.86	0.77	0.93	0.99	0.99	0.99	0.99	0.99	0.99	1					
2304.71	0.84	0.84	0.85	0.85	0.85	0.85	0.85	0.85	0.76	0.92	0.99	0.99	0.99	0.99	0.99	0.99	0.99	1				
2314.81	0.83	0.84	0.84	0.84	0.84	0.85	0.85	0.85	0.76	0.91	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	1			
2324.91	0.83	0.83	0.84	0.84	0.84	0.84	0.84	0.84	0.76	0.9	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	1		
2335.01	0.82	0.82	0.83	0.83	0.83	0.83	0.84	0.84	0.76	0.9	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	1	
2345.11	0.81	0.82	0.82	0.82	0.83	0.83	0.83	0.83	0.76	0.89	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	1

3.4 Discussion

Hyperspectral data are suitable for C3 and C4 grass species classification, since spectral features characteristic of vegetation are more discernible from narrowband sensors. However, it can be challenging to classify C3 and C4 grass species using the full spectrum of hyperspectral data, due to the problem of hyper-dimensionality and associated multicollinearity phenomena (Pal and Foody 2010). Among other factors (e.g. phenology and solar illumination conditions), spectral similarity between C3 and C4 grasses and their co-existing species can have significant impacts on the inter-band correlation and thus on the classification capability of canopy reflectance (Schmidt and Skidmore 2001). Despite these challenges, recent studies have shown that subtle differences in structural and physiological properties, such as those described for C3 and C4 grasses can still be detected by leaf or canopy reflectance (Irisarri et al. 2009; Liu and Cheng 2011). Although previous studies have used narrow-band spectral data to classify grasslands of C3 and C4 species composition, the present investigation explores the potential use of a user-defined inter-band correlation filter to resample hyperspectral data, for subsequent classification analysis.

This study demonstrated the trade-offs between retaining narrow bands spectral data versus the optimal reduction in spectral dimensionality for improved classification. The results obtained explained the spectroscopic interpretability of the chosen band-centres, in reference to their sensitivity to leaf or canopy surface properties, internal structure and biochemical concentrations. These characteristic features are known to significantly vary between C3 and C4 grass species. Hence, variations in pigments content, nitrogen, carbon compounds (lignin and fibre) and water components (inter-cellular air-moisture or leaf liquid water content) can be attributed to the good spectral separability obtained for the target grasses assessed in this study. In general, the results have shown that the proposed user-defined inter-band correlation filter technique yielded improved classification of *F. costata*, *R. altera* and *T. triandra* grass canopies. More detailed analyses of the results are presented as follows (i) the weighting thresholds of inter-band correlation filter approach to resampling hyperspectral data (ii) the random forest classification and band subset selection of resampled Hyperion dataset and (iii) the implications of the present investigation for applications involving C3 and C4 grass species classification.

3.4.1 The weighting thresholds of inter-band correlation filter approach

On the basis of the proposed resampling approach, this study has shown that highly correlated hyperspectral wavebands in specific regions can be optimally aggregated to reduce spectral dimension of the input spectral bands. The proposed spectral resampling technique takes advantage of the inherent property of vegetation reflectance, the asymmetrical nature of the inter-band correlation matrix of the collected wavebands. The results presented in Table 3.2 show the WTC-derived spectral resolution constructed using highly correlated wavelengths around each of the pre-selected 13 band-centres and those of fixed band-width across the 400 - 2500 nm spectrum, without WTC.

The vegetation spectral response property used to calculate the various WTC r values can be attributed to reflectance or absorption features characteristic of the target C3 and C4 grasses (Ferwerda et al. 2005; Knox et al. 2010). In a previous study, Slanton et al. (2001) found that 800 nm wavelength contained very strong discriminating power for plant species at the level of leaf internal structure. Further, Irisarri et al. (2009) reported that vegetation reflectance at the 820 nm spectral range is sensitive to even subtle differences among grass species or between groups of C3 and C4 grasses. Hence, in the present experiment, the user-defined inter-band WTC filters were used to assess the optimal spectral resolutions around the chosen band-centres, including the 820 nm wavelength. In this regard, the proposed resampling procedure offers the potential to reduce data dimension and it eliminates redundant spectral information by means of weighting the inter-band correlations. Therefore, it is worth noting that the newly introduced resampling approach not only reduces dimensionality in hyperspectral data, it also preserves the most relevant spectral information in the output waveband for improved classification of the target C3 and C4 grass species.

3.4.2 Random forest-based band subset selection vs. prior dimensionality-reduction

Random forests have been found attractive for the analysis of remotely sensed data for ecological applications (Chan and Paelinckx 2008; Ham et al. 2005; Lawrence et al. 2006; Prasad et al. 2006). A number of studies have asserted that the method is insensitive to high-dimensionality and, therefore, does not require a dimensionality-reduction analysis in pre-processing (Breiman and Cutler 2004; Ham et al. 2005). However, the assessment of random forest's variable importance measure in high-dimensional spectral space, has revealed that the algorithm thus

show a preference to highly correlated predictor variable. Such a preference was also found to manifest in the subsequent subset band selection process. The results from the present experiment thus reaffirm the findings of the recent studies, which investigated random forests variable importance under predictor correlation and the generalization of parameter estimates (Nicodemus and Shugart 2007; Strobl et al. 2008). In their studies, the authors recommended conditional variable importance approach for random forest-based variable selection. However, the conditional variable importance has to be further investigated within a remote sensing application.

The spectral features of the nearIR region (i.e. 820 nm band-centre), which have been widely used in remote sensing applications involving vegetation assessment yielded moderate variable importance. This can be attributed to the spectral similarity, based on multiple-scattering effects in leaf or canopy mass, leaf area index, canopy moisture content and structure (Curran 1994; Horler et al. 1983). These features of vegetation can have significant influence on the nearIR reflectance (Kumar et al. 2001) and mask small spectral differences among the target grass species. However, results obtained from the present study showed that WTC r = 0.992 yielded the highest classification accuracy (kappa = 0.82) among the datasets assessed. This superior accuracy demonstrates clearly, the role of band positions and their corresponding spectral resolutions on C3 and C4 grass classification and the classifier accuracy. The larger decrease in classification accuracy obtained for the lower limits (e.g. WTC r = 0.900) could be attributed to the very larger increase in spectral resolution for each individual waveband. The effect of coarse spectral resolutions could have possibly masked the relatively fine spectral feature space variation amongst the target C3 and C4 grasses in the study area. The results obtained for the fixed band-width datasets (the resampled smoothed-13-band-centres) confirm the effects of increasing spectral resolution on a classifier performance.

The classification accuracies obtained in the present experiment for both the WTC dataset and that of spectra resampled using the traditional methods compares well with those reported by Dalponte et al. (2009), who investigated the effect of changing spectral resolutions upon a classifier for forest applications. In their study, the authors found that as spectral resolutions were degraded from 4.6 nm to 36.8 nm, overall kappa accuracies dropped from ~ 89 % to ~ 84 %, respectively, using Support vector machines (Vapnik 1998). Compared with classification involving a simple parametric classifier such as Linear Discriminant Analysis (Fisher 1936),

Dalponte et al. recorded inferior kappa accuracies which also dropped from ~ 77 % to ~54 %, respectively. The authors concluded that advanced non-parametric classifiers are more applicable for classifications involving complex vegetation feature spaces. In general, the result obtained from this study demonstrated the relationship among classifier sensitivity to data dimensionality, spectral resolution and classification accuracy, under conditions of multicollinearity among the input spectral bands (Gomez-Chova et al. 2003). This suggests the concept of using spectral resampling techniques capable of reducing multicollinearity in the input spectral space, for applications involving C3 and C4 grass species.

As depicted on Table 3.5, the random forest band selection process showed a significant bias toward the highly correlated Hyperion bands (e.g. B6 - B13 and B198 - B219). However, the random forest analysis on the prior reduced dimension datasets offered a distinct technique, using the optimum WTC filter to aggregate the majority of the highly correlated wavebands, according to their linear relationship with a chosen band-centre. This forms the novelty of the proposed user-defined inter-band correlation filter approach, which considerably negates the problems associated with spectral redundancy and thereby mitigates against the multicollinearity phenomenon (Gomez-Chova et al. 2003).

In summary, the current study demonstrated the superiority of the WTC method to other traditional approaches for increasing signal-to-noise ratio (SNR). This was demonstrated by inferior results obtained by; 1) resampling all bands at various resolutions, following smoothing; and 2) testing the pre-selected band centres on smoothed data at various band widths. In addition, an optimum threshold of (WTC = 0.992) was obtained by scanning all the r- values between 0.900 and 1.

3.4.3 Implications of the present investigation and conclusion

The primary purpose of this study was to assess the spectral separability among C3 and C4 grasses, sampled from the Drakensberg Mountains of South Africa. The secondary goal was to address the issue of multicollinearity effect on the performance of the random forest variable importance and the subsequent band subset selection process under predictor correlation. The performance of the method, when applied to data derived by resampling spectra to the Hyperion sensor's spectral resolution, was compared to that of spectra resampled by weighting the inter-band correlations, as a function of wavelength. The overall implications for this investigation are

related to various hyperspectral data application constraints: i) the trade-off between the number of spectral bands and the resolution of remotely sensed imagery; ii) the trade-off between higher spectral resolution and reduced signal-to-noise ratio and iii) challenges associated with the optimal configuration of wavebands capable of providing sensitive information about a target vegetation (Price 1994; Thenkabail et al. 2004b). Therefore, in the present experiment, a technique has been proposed to reduce dimensionality, while preserving relevant spectral information for posterior classification task. It has been observed that the proposed resampling technique represents a potential method of reprogramming hyperspectral resolutions for applications involving C3 and C4 grass species. The results obtained in this study suggested that further studies addressing multicollinearity problem should consider techniques that account for the spectral dependence information contained in vegetation reflectance data. In summary, the current technique described in this paper yields the following distinct benefits:

- Reduces data dimensionality by accounting for the inter-band correlations around specific band-centres of interest and to mitigate against multicollinearity phenomenon caused by highly correlated input spectral bands.
- Optimizes the spectral resolutions useful for the separability among dominant C3 and C4 grass species.
- Assists the random forest to achieve improved classification accuracy, thereby providing the potential to link each individual input band to the physical meaning of interaction effects in the structure of the acquired data.
- Provides an alternative to the remote sensing community on data reduction techniques, which may yield similar or better classification results under certain environmental conditions than the traditional approaches.

The experimental results obtained in this study suggest that resampling hyperspectral data for applications involving C3 *and* C4 *grass species should account for the spectral dependence information contained in vegetation reflectance data. However, an important question is how the results of this investigation compare, with those of spectra, resampled to match the waveband settings, of some commercially available space-bone sensors. In that regard, Chapter 4 is a follow-up from Chapter 3 and it focuses on optimizing wavebands for the classification of the* C3 *and* C4 *grasses in sampled from the study area.*

CHAPTER 4: OPTIMIZING SPECTRAL RESOLUTIONS FOR THE CLASSIFICATION OF C3 AND C4 GRASS SPECIES

The chapter is based on:

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

Abstract

Hyperspectral remote sensing approaches are suitable for detection of the differences in C3 and C4 grass species phenology and composition. However, the application of hyperspectral sensors to vegetation has been hampered by the high-dimensionality, spectral redundancy and multicollinearity problems. In this experiment, resampling of hyperspectral data to wider wavelength intervals, around a few band-centres sensitive to the biophysical and biochemical properties of C3 or C4 grass species is proposed. The approach accounts for inherent property of vegetation spectral response: the asymmetrical nature of the inter-band correlations between a waveband and its shorter and longer wavelength neighbours. It involves constructing a curve of weighting threshold of correlation or WTC (Pearson's *r*) between a chosen band-centre and its neighbours, as a function of wavelength. In addition, data were resampled to some multispectral satellite sensors: ASTER, GeoEye-1, IKONOS, QuickBird, RapidEye, SPOT 5 and WorldView-2 (including 6 WorldView-2-derived vegetation indices) for comparative purposes, with the proposed method. The resulting datasets were analyzed, using the random forest algorithm. The proposed resampling method (optimum WTC r = 0.992-derived dataset) achieved improved classification accuracy (Kappa = 0.81), compared to the resampled multispectral bands and vegetation indices (Kappa = 0.78, 0.65, 0.62, 0.59, 0.65, 0.62, 0.76 and 0.78, respectively). Overall, results from this study demonstrated that spectral resolutions for C3 and C4 grasses can be optimized and controlled for high dimensionality and multicollinearity problems, yet yielding high classification accuracies. The findings also provide a sound basis for programming wavebands for future sensors.

Keywords: inter-band correlation, Spectral band configurations, Grass species discrimination, Random forests, Remote sensing.

4.1 Introduction

There has been an increasing concern that global climate change is affecting the function and dynamics of vegetation assemblages (Craine et al. 2009; McMahon et al. 2011). As an example, projections based on scenarios of climate change suggest an increased impact on the southern African rangelands productivity and community composition (Chamaillé-Jammes and Bond 2010; Hoffman and Vogel 2008). Depending on the ecological and physiological adaptive capacity, some species are expected to increase in abundance or migrate to new habitats, which provide favourable conditions for growth. Specifically, grasslands composed of species which correspond to 3-carbon (C3) or 4-carbon (C4) photosynthetic pathways, differ greatly in their environmental requirements and will respond differently to subtle changes in climatic conditions (Tieszen et al. 1997). The C3 and C4 groups of grasses also differ in many anatomical and physiological aspects. For example, leaves of C3 grasses have a relatively lower compartment of mesophyll cells, a lower proportion of vascular tissues, and a higher inter-veinal distance than those of C4 grasses (Dengler et al. 1994; Ogle 2003). In addition, biochemical constituents, such as intercellular air-moisture and nitrogen concentrations are generally higher in C3, compared to C4 grasses, while leaf area and specific leaf area ($cm^2\ g^{-1}$) are higher in C4 than C3 grass species (Ghannoum 2008; Oyarzabal et al. 2008; Sage and Pearcy 1987).

A number of studies have demonstrated that the spectral variability between C3 and C4 grass species is greater than that within each group (Doughty et al. 2010; Trenholm et al. 2000). The reflectance at 531–570 nm is sensitive to photosynthetic radiation use efficiency (Gamon et al. 1997), and is therefore proposed as a sensitive region capable of discriminating between C3 and C4 grasses (Irisarri et al. 2009). Leaf nitrogen content has been predicted using combination of reflectance measurements centred at 1770 nm and 693 nm (Ferwerda et al. 2005). The red-edge region (680–750 nm) correlates strongly with chlorophyll and nitrogen content in C3 winter wheat discrimination, from C4 grass weeds (Hansen and Schjoerring 2003). In this study, the authors in addition found that indices calculated based on blue wavebands significantly correlate with the wheat pigments density. (Slaton et al. 2001), among others, studied the near infrared (NearIR) region and found reflectance at 800 nm significantly different among species that differ in intercellular structure, which, as indicated above, widely differ between C3 and C4 grass canopies. Moreover, in a common garden experiment, (Irisarri et al. 2009) demonstrated that the green, red and the NearIR (above 820 nm region) were consistently required to

differentiate between C3 and C4 grass compositions, both as “pure” or “mixed” canopies. The fundamental principle is that the nature and amount of reflection, absorption or emission are controlled by the biophysical (e.g. canopy coverage, LAI and biomass) and biochemical properties of vegetation spectral response features (Cho et al. 2010; Cho and Skidmore 2009; Mutanga et al. 2003).

The spectral differences between C3 and C4 grass species manifest in the composition of grasslands, with the dominant species strongly constituting the optical properties of the canopies (Davidson and Csillag 2001; Foody and Dash 2007; Goodin and Henebry 1997). Although not consistent, these studies have demonstrated the potential use of remote sensing approaches to discriminate between C3 and C4 grass canopies, at the landscape level of assessment. One of the major limitations is the use of broadband sensor systems, which are susceptible to masking the narrow spectral diagnostic features of C3 and C4 grasses. Thus, narrowband hyperspectral approaches provide more detailed spectral information to detect and discriminate C3 and C4 grasses (Irisarri et al. 2009). However, the widespread applications of data acquired through hyperspectral sensors have been hampered by the increased dimensionality. Specifically, the data present difficult challenges in supervised classification environment, when small training sample sizes are used in the high dimensional feature space (Crawford et al. 2003). In this experiment, spectral resampling of hyperspectral data to wider bandwidths around a few sensitive wavelengths (i.e. band-centres) is proposed as a means to reduce dimensionality, as well as controlling for multicollinearity problems. The approach explores the inherent spectral response property of vegetation that is not considered in existing resampling methods. In addition to the data obtained by means of the proposed User-defined inter-band correlation filter technique, some existing multispectral sensors’ spectral band configurations were simulated for comparative purposes. The resulting datasets were analyzed, using the random forest algorithm (Breiman 2001a). The random forest algorithm was considered on account of its superior generalization capacity, insensitivity to limited number of empirical training samples and its ability to reduce classification error on a wide array of data structures (Dye et al. 2011; Ghimire et al. 2010; Ismail and Mutanga 2011).

4.2 Methodology

4.2.1 Study area

Key C3 and C4 grass canopies considered in this investigation were sampled from the Cathedral Peak region of the Drakensberg Mountain range, South Africa. This region consists of vegetation divided into altitudinal zones, corresponding closely with the physiographic features of the Drakensberg. Three zones, (i) the Montane belt (1280–1829 m), (ii) the Sub-alpine belt (1930–2865 m), and (iii) the Alpine belt (2866–3353 m) above mean sea level, are defined. These zones also coincide with the three terraces in the Drakensberg, namely, the river valley system, the foothills and the summit areas, respectively (Hill 1996; Killick 1963). The Sub-alpine belt is also known as *Themeda-Festuca* sub-alpine grassland, composed of dominant C3 and C4 grass species. This zone can be further divided into two grassland communities described as the Red grass, *Themeda triandra* and the tussock fescue, *Festuca costata* sub-alpine grasslands. Notably, there are strong consociations of *T. triandra* and *F. costata* occurrence on warm, northerly and cool, southerly slopes respectively, within the sub-alpine zone (Granger and Schulze 1977; Hill 1996).

Although it is predominantly a grassland landscape, patches of forests occur in the sheltered ravines, gorges and valleys of the Drakensberg Mountain range. In this study, sampling was confined to the grassland areas. The study area is characterized by high rainfall (> 900 mm) and mist plays an important role, while low temperatures, seasonal frosts, snow and fire, including topographic features are also important determinants of the C3 and C4 vegetation dynamics in the Drakensberg ecosystem. In general, vegetation coverage on the greater study site was a mixture of C3 and C4 grass species with multiple patches dominated by the Toothbrush grass, *Rendlia altera* and *T. triandra* grasslands. The C3 grass, *F. costata* grassland patches are sparse and in some areas they coexist with *Merxmuellera macowanii* (an evergreen coarse C3 grass) and C3 Bracken (*Pteridium aquilinum*); a large, deciduous, rhizomatous, perennial fern that has almost horizontal leaves and can grow taller than stands of the *F. costata*. The target species, *F. costata*, is a coarse, unpalatable species that tends to remain evergreen throughout the dry season (unlike the *P. aquilinum*) and can therefore relatively easily be identified. *F. costata* normally grows as large tussocks that are taller than surrounding C4 grassland. It often grows in an aggregated pattern, which is generally easily distinguishable in the field. Notably, the C4 dominant grass communities made up 80% of the composition and *R. altera* grassland consists of

relatively narrow leaves which tend to fold and break up into fibres when they reach maturity. There is generally low leaf production in this grassland patches, manifesting in the canopy density and structure. The other dominant C4 grassland of *T. triandra* forms dense and tufted stands that are distinctly red-coloured mature plants, distinguishable from surrounding grasslands on the basis of its characteristic canopy composition and appearance. There is no specific study focused on estimation of the variability between C3 and C4 grassland biomass or community compositions in the Cathedral peak study area. However, unpublished work by Adjorlolo and Mutanga (in prep.) shows that at the peak of the growing season, there is significantly higher mean biomass production in *F. costata* grasslands, compared to their surrounding C4 (i.e. *R. altera* and *T. triandra*) grasslands. This finding was based on analysis of 1×1 m sample plots (n = 330) to estimate above-ground green biomass.

4.2.2 Field data acquisition

The field data acquisition was conducted during the December 2010 summer growing season. Three sets of reflectance measurements were conducted on (i) *R. altera*, (ii) *T. triandra* and (iii) *F. costata* species. Reflectance data were collected at the recommended optimum conditions for *in situ* field measurements (Thenkabail et al. 2004b). The grass canopy spectra were measured using an Analytical Spectral Device (ASD), FieldSpec spectrometer (FieldSpec®3, Analytical spectral Device, Inc., USA). This device registers radiation at 1.4 nm intervals for the spectral region 350-1000 nm and 2 nm for the spectral region 1000–2500 nm. Data were interpolated to 1 nm spectral resolution across the spectrum. The ASD device uses a fibre optic cable set at a 25^{o} field of view, to record the reflected radiation. The fibre optic sensor was held at about 1.5 m, directly above the sampling plots, generating an instantaneous field of view of about 0.35 m^2. Reflectance was calculated as the ratio between the reflected energy from the grass canopy and the incident energy on the barium sulfate ($BaSO_4$) white reference panel. The canopy reflectance data were collected to classify the 1×1 m sample plots represented by 80% or more *R. altera*, *T. triandra* and *F. costata* dominant grasses. Figure 4.1 illustrates the average spectral curves for C3 and C4 types of grass species canopy reflectance measurements. A total of 110 plots were sampled from large and fairly homogeneous patches consisting of each representative grass species.

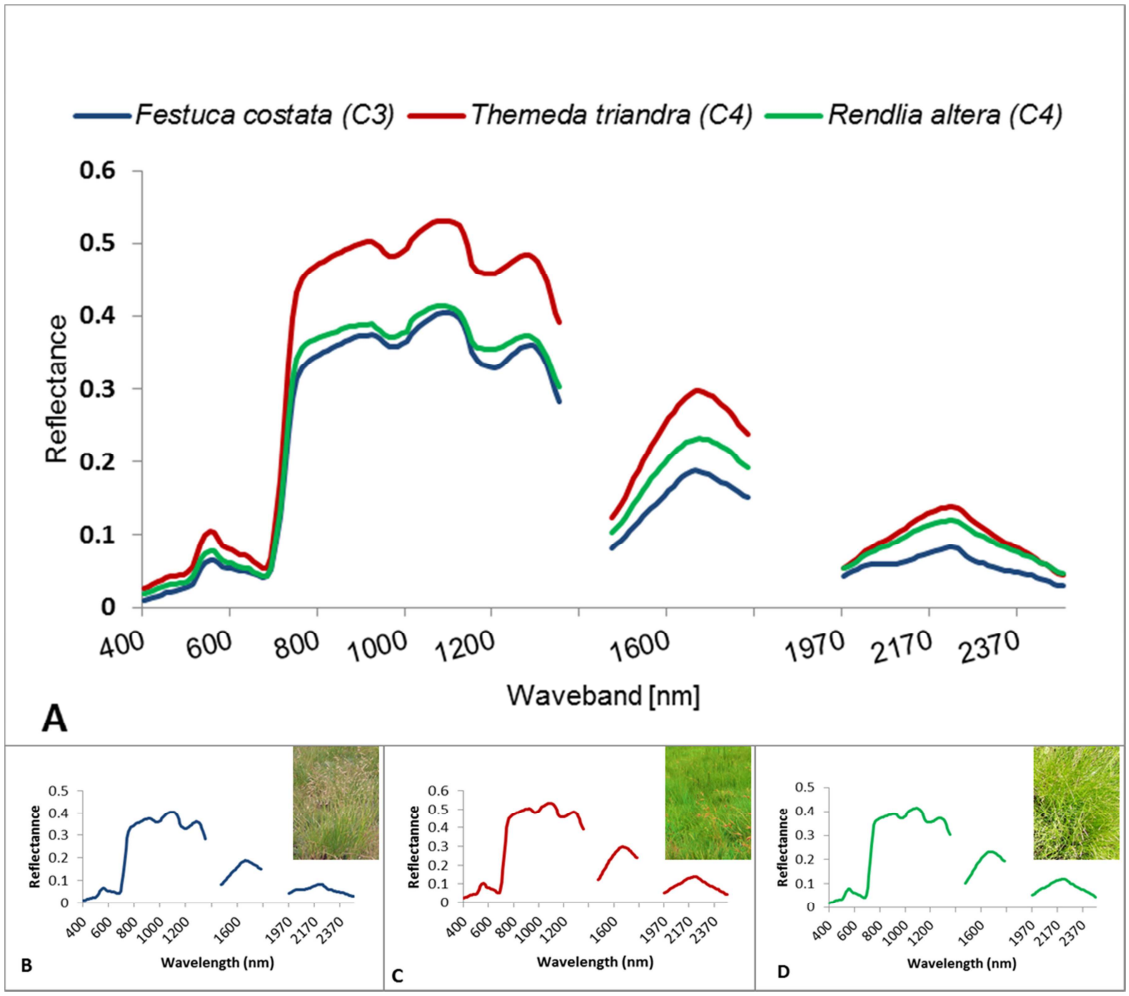

Figure 4.1 Average spectral canopy response curves for (A) three grass species compared. The inserts illustrate the spectral curves of each species: (B) Festuca costata, (C) Themeda triandra and (D) Rendlia altera dominant grasslands in the Drakensberg montane landscape of South Africa (n = 110 for each representative grass species). The original ASD spectra were aggregated to 10-nm-wide bandwidths (N = 197 bands) and reflectance measurements are scaled between 0 and 1. Noisy atmospheric water absorption bands (1350–1460 nm and 1790–1960 nm) were removed from all analysis.

4.2.3 The spectral resampling approaches

Processing of high-dimensional spectra often aims at dimensionality reduction of the data, without losing the spectral separability of a considered feature space (Irisarri et al. 2009; Schmidt and Skidmore 2003; Schmidt et al. 2004; Taylor et al. 2012; Thenkabail et al. 2000). The problem of dimensionality has been dealt with using a number of techniques, including resampling of spectra to wider bandwidths around a few band centres or to the spectral configuration of existing sensors, using their respective spectral response functions or spectral

resolution (i.e. Full Width at Half Maximum, FWHM/ λ). However, the major limitation is that an inherent property of the spectral response of vegetation, the asymmetrical nature of inter-band correlations between a given waveband and its shorter and longer wavelength neighbours is not fully accounted for in existing resampling methods (Guizhong and Fan 2007; Xiaolin and Memon 2000).

The user defined inter-band correlation filter technique of spectral resampling, which is adopted in this study convolves the spectral response information around a given band-centre. This was done by weighting the influence of the shorter and longer wavelength neighbours on the chosen band-centre. It involves constructing vegetation spectral response curves, by assigning weights (Pearson's r correlation coefficient) to a given band-centre and its shorter and longer wavelength neighbours, whereby r = 1 at the band-centre. Bands beyond the chosen threshold (i.e. r = 0.992) are given a value of r = 0 (Adjorlolo et al. 2013). This is an innovative spectral resampling approach, whereby the researchers consider the inter-band correlations around individual band centres of interest (Adjorlolo et al. 2013). The approach accounts for spectral dependence information in the matrix of acquired data. It should be noted that the Pearson's r correlation between bands was calculated on spectra, initially aggregated in 10-nm-wide (FWHM) across the spectrum. This was done in order to aid computational efficiency, thereby smoothing the acquired 1-nm-wide spectra collected in the field. The 10-nm-wide, FWHM was chosen, to mimic the average spectral resolution of the existing hyperspectral sensor, Hyperion, aboard the Earth Observing satellite.

The vegetation spectral response curves were calculated using thirteen (13) chosen band-centres. These band-centres (Table 4.1) were chosen based on the known vegetation spectral response properties of absorption: plant pigments (e.g. chlorophyll), micro and macro nutrients (e.g. proteins and plant nitrogen concentration), vegetation water content and reflectance properties of leaf and canopy architecture. The spectral response curves are calculated, based on the known characteristics of vegetation, which significantly vary between the C3 and C4 grass species or canopies (Asner 1998; Irisarri et al. 2009). The purpose in the current study was to optimize spectral resolutions and band configuration that are sensitive to the known vegetation absorption features across the spectrum. Thus, the proposed User-defined inter-band correlation approach of spectral resampling has distinct benefits of yielding reduced spectral dimension (i.e. wider bandwidths) sensitive to the physical interaction effect in the C3 and C4 grass space.

In addition, data were resampled to simulate the spectral response function of some multispectral remote sensing sensors, as listed in Table 4.2. Their simulation involves a conventional spectral resampling approach that uses the sensors FWHM function of wavelength (λ). The Environment for Visualizing Images, ENVI 4.7 software (Copyright © ITT Visual Information Solutions, 2009) spectral analysis routine was used to resample the *in situ* field measured grass canopy reflectance data. It should be noted that, since measurements were conducted under field conditions, the strong noisy incident radiation in 1350–1460 nm and 1790–1960 nm, and highly inconsistent spectra in 350–399 nm regions were removed from the analysis (Thenkabail et al. 2000). The resampled datasets were analyzed using the random forest classification algorithm.

Table 4.1 Wavelengths corresponding to known absorption features

No.	Band centre (nm)	Known Causal compound/ feature	Source
1	470	Total plant pigment concentration	Blackburn 1998
2	530	Chlorophyll a absorption	Gamon et al. 1997
3	600	Protein, Nitrogen	Fourty and Baret 1997
4	660	Plant nitrogen concentration	Carter 1994
5	700	Total Chlorophyll, Nitrogen	Carter 1994
6	720	Total Chlorophyll, Leaf mass	Horler et al. 1983
7	820	Leaf mass, Leaf area index	Curran 1989
8	1540	Cellulose, Leaf water content	Carter 1994
9	2060	Protein, Nitrogen	Carter 1994
10	2280	Cellulose, Leaf mass	Carter 1994
11	2300	Leaf mass, Leaf water content	Carter 1994
12	2450	Cellulose, Protein, Nitrogen	Carter 1994
13	2470	Cellulose, Protein	Kumar et al. 200

Table 4.2 Some existing multispectral sensors and their spectral properties

Sensor name	Band description	Spectral range (nm)	Bandwidth
ASTER	Band 1 (Green)	520–600	80
	Band 2 (Red)	630–690	60
	Band 3 (NearIR; Nadir view)	760–860	100
	Band 4	1600–1700	100
	Band 5	2145–2185	40
	Band 6	2185–2225	40
	Band 7	2235–2285	50
	Band 8	2295–2365	70
	Band 9	2360–2430	70
GeoEye-1	Blue	450–510	60
	Green	510–580	70
	Red	655–690	35
	NearIR	780–920	140
IKONOS	Blue	445–516	71
	Green	506–595	89
	Red	632–698	66
	NearIR	757–853	96
QuickBird	Blue	450–520	70
	Green	520–600	80
	Red	630–690	60
	NearIR	760–900	140
RapidEye	Blue	440–510	70
	Green	520–590	70
	Red	630–685	55
	Red-edge	690–730	40
	NearIR	760–850	90
SPOT 5	Green	500–590	90
	Red	610–680	70
	NearIR	780–890	110
	ShortwaveIR	1580–1750	170
WorldView-2	Coastal	400–450	50
	Blue	450–510	60
	Green	510–580	70
	Yellow	585–625	40
	Red	630–690	60
	Red-edge	705–745	40
	NearIR-1	770–895	125
	NearIR-2	860–1040	180

4.2.3.1 WorldView-2 feature extraction

In addition to the resampled 8-bands multispectral data of WorldView-2, a total of 6 normalized vegetation indices (NDVIs) were calculated:

$$NDVI1 = (NIR\text{-}1 - green)/(NIR\text{-}1 + green) \quad \text{(Eq.1)}$$

$$NDVI2 = (NIR\text{-}1 - red)/(NIR\text{-}1 + red) \quad \text{(Eq.2)}$$

$$NDVI3 = (NIR\text{-}2 - red\text{-}edge)/(NIR\text{-}2 + red\text{-}edge) \quad \text{(Eq.3)}$$

$$NDVI4 = (NIR\text{-}2 - yellow)/(NIR\text{-}2 + yellow) \quad \text{(Eq.4)}$$

$$NDVI5 = (red\text{-}edge - coastal\ blue)/(red\text{-}edge + coastal\ blue) \quad \text{(Eq.5)}$$

$$NDVI6 = (red\text{-}edge - red)/(red\text{-}edge + red) \quad \text{(Eq.6)}$$

The inclusion of these vegetation indices was based on recent studies by Mutanga et al. (2012) and, Pu and Landry (2012) who reported that these NDVIs are useful for improving the estimation of biomass and mapping of plant species. In the context of the current study, analysis of the resampled worldview-2 together with the additional NDVI features is to explore their potential in the C3 and C4 grass environment.

4.2.3.2 Random forest algorithm

The random forest algorithm utilizes bootstrap samples with no replacement to grow a large ensemble of trees. The ensemble of trees assign each waveband to a class, based on the maximum number of votes that a class receives from the collection of trees (Breiman 2001a). Each tree is grown from randomly and independently selected, bootstrap samples of the training data (i.e. 66 % random sample). About one-third excluded samples, called the out-of-bag (OOB) cases, are used to evaluate the model (Prasad et al. 2006). Researchers have shown that the OOB samples offer unbiased estimates of training error (Chan and Paelinckx 2008; Ismail and Mutanga 2011; Lawrence et al. 2006; Pal 2005). The algorithm is easy to implement because only two parameters are tuned: (i) the number of trees to grow (ntree); and (ii) the number of variables to split at each node (mtry). Another important feature of random forest is that it offers variable importance measures embedded in its computation and it can perform both binary and multiclass classification tasks (Breiman and Cutler 2004). The variable importance measures provide the researcher valuable insight to explore the effect of any predictor variable on the

response variable (Strobl et al. 2008). In this study, the random forest algorithm was implemented, using the R statistical software (R Development Core Team 2010).

4.2.3.3 Variable importance

The random forest algorithm returns three variable importance measures: (i) the selection rate of each candidate variable (Breiman 2001a), (ii) the Gini index of impurity reduction (Pal 2005), and (iii) the permutation of the predictor variables as an estimate of importance (Strobl et al. 2007a). Among these variable importance measures, the Gini index has been found to show substantial bias when predictor variables vary in their number of categories or scale of measurement (Strobl et al. 2007a; Strobl et al. 2007b). This is because the underlying Gini gain splitting criterion is a biased estimator that can be affected by multiple testing effects (Strobl et al. 2008). The alternative variable importance concepts are based on the impact of a predictor variable, denoted as "mean decrease in accuracy" and compare each candidate predictor variable with the magnitude of its impact in predicting the response variable (Strobl et al. 2008).

These later variable importance measures have been successfully calculated by means of randomly permuting each predictor variable's association with the response variables (Strobl et al. 2008). The permutation-based variable importance follows the rationale that random permutation of predictor variable represents the absence of the variable from the model. Hence, the difference in prediction accuracy prior and after permuting a variable (i.e. the class membership of a permuted variable, together with the remaining non-permuted predictor variables), is used to predict the response for the OOB observations as the measure of importance (Strobl et al. 2008). In this regard, the number of observations classified correctly decreases substantially, if the permuted variable was strongly associated with the response variables (Strobl et al. 2008).

4.3 Results and Discussion

Hyperspectral datasets have been investigated for detailed characterization of the spectral response of C3 and C4 grass species differentiation (Schmidt and Skidmore 2003; Thenkabail et al. 2000). In multivariate classification environment, the use of very large number of input spectral bands, compared to the number of training samples often resulted in the Hughes phenomenon (Hughes 1968) of dimensionality. The Hughes phenomenon leads to violation of

non-multicollinearity assumption in statistical analysis (Chi and Bruzzone 2007; Melgani and Bruzzone 2004). In large-area mapping objectives, seldom are there adequate training samples to support the use of less rigorous classification algorithms, under very high predictor dimensionality (Chan and Paelinckx 2008; Lawrence et al. 2006). That is, when the number of training samples is limited in multivariate statistical environment, classification techniques that are capable of reducing the error of non-generalization in modeling the response of C3 and C4 grasses are desirable (Liu and Cheng 2011). Therefore, the use of non-parametric algorithms such as random forests have become very attractive for classification analysis involving remotely sensed data, for vegetated landscape applications (Chan and Paelinckx 2008; Ismail and Mutanga 2011; Lawrence et al. 2006). In the present study, the random forest was used both as a measure of predictor variable importance and as the classifier for C3 and C4 grass reflectance data. Moreover, the cross-correlation property and the asymmetrical nature of the vegetation spectral response underline the interest and the importance of this study. Analyses of the results are organized in the following sections: optimizing the spectral resolutions for C3 and C4 grass classification, random forest-based variable importance measure, and utility of the proposed User-defined inter-band correlation filter, compared to those of the resampled multispectral sensors.

4.3.1 Optimizing spectral resolutions for C3 *and* C4 *grass classification*

Accurate information on the limits of spectral resolution and waveband configuration is critical for measurement of the spectral differences between C3 and C4 grass types (Irisarri et al. 2009). In the present study, optimized configuration of wavebands was achieved using User-defined inter-band correlation filter approach. The results obtained revealed that sizeable portions of the original data exist in varying cross-correlated spectral reflectance measurements, across the 400–2500 nm spectrum. The thirteen chosen band-centres clearly depicted varying spectral resolutions (Figure 4.2). This showed the spectral characteristics of C3 (*F. costata*) and C4 (*R. altera* and *T. triandra*) grass species response in specific regions, across the measured spectrum. Using the weighting threshold (r = 0.992) of inter-band correlation filter, wavebands optimal for the classification of these grass species were achieved (Adjorlolo et al. 2013). Further analysis of Figure 4.2 showed that spectral resolutions in the visible and red-edge regions are relatively narrow, compared to bands centred in NearIR (i.e. 820 nm) and shortwave Infrared (i.e. 1540,

2060, 2280, and 2300 nm) regions. The band-centres that were chosen for shortwaveIR (i.e. 2450 and 2470 nm) region also yielded narrow wavebands, comparable to those in the red-edge region (i.e. 700 and 720 nm band-centres). This demonstrated the varying degrees of inter-band correlations in specific wavelengths in reference to the C3 and C4 grass feature space. In previous studies, narrow-bands centred around 800 or 820 nm were identified to be useful for C3 and C4 grass species discrimination (Irisarri et al. 2009; Slaton et al. 2001). However, results obtained in the present study showed that highly correlated bands around 820 nm band-centre can be aggregated (750–1000 nm) to form a single broad but sensitive waveband. The underlying physical property explaining the wide-ranging waveband in this region can be attributed to the multiple interaction characteristics of C3 and C4 grass biochemical and structural properties (Curran 1989; Kumar et al. 2001). The simulation of spectral curves around the thirteen chosen band-centres demonstrated trade-offs between retaining narrow-waveband information versus optimized reduction in spectral resolutions for the classification of C3 and C4 grass species (Taylor et al. 2012).

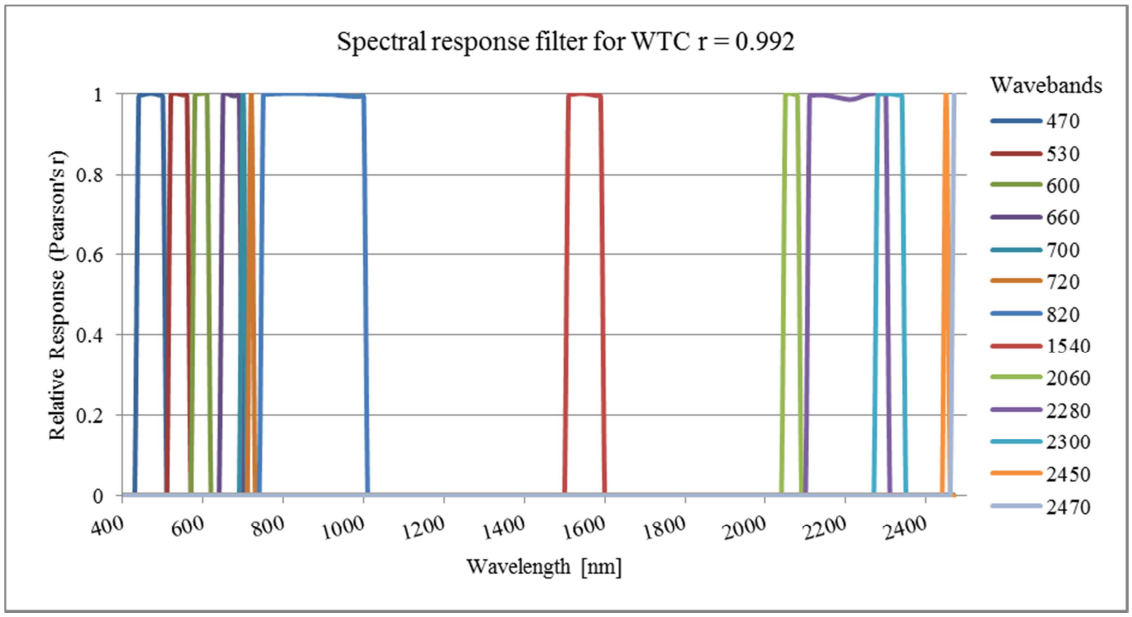

Figure 4.2 Vegetation spectral response filter simulated using the inter-band Pearson's r correlation coefficient calculated on sampled 10-nm-wide spectra.

4.3.2 Random forest-based variable importance

The random forest variable importance process offered a valuable screening tool for identification of the most sensitive wavebands or regions (Adam et al. 2009; Ismail and Mutanga 2011). The algorithm was run using all bands for all the datasets. The User-defined inter-band correlation filter dataset yielded good variable importance for wavelengths for C3 and C4 grass types differentiation, comparable to that achieved in (Irisarri et al. 2009). Table 4.3 presents the results of the ranked wavebands, the user-defined inter-band correlation filter wavebands (i.e. optimized spectral resolutions) obtained as well as the permutation-based mean decrease in accuracy (Ismail and Mutanga 2011; Strobl et al. 2008). It is interesting to note that the wavebands centred at 470 nm (440–500 nm) and 1540 nm (1510–1590 nm) have the highest impacts on the C3 and C4 grass response variables. As determined by the random forest variable importance, these two wavebands have significant lowering effect on the "mean decrease accuracy" values: 0.194% and 0.149%, respectively. Further analysis of Table 4.3 showed that wavebands centered in the blue and shortwave regions were more important for classification involving the target C3 and C4 grass species. Such an observation is consistent with previous reports that wavebands in the visible region exhibit superior power in discrimination between groups of C3 crops and C4 weed species, or vice versa, using canopy reflectance data (Noble et al. 2002; Smith and Blackshaw 2003; Vrindts et al. 2002). In their study, (Smith and Blackshaw 2003) recorded higher frequencies (17 to 18 out of 20 times) for wavebands located in the visible blue and shortwaveIR regions respectively. As expected, the authors found that 10-nm-wide wavebands in the red-edge (700 to 730 nm) region exhibited the greatest power in separating C3 (*Brassica napus*, *Chenopodium album*, *Sinapsis arvensis*) and C4 (*Amaranthus retroflexus*) plants. The high frequency of the red-edge selection was attributed to variability in plants phenology and pigment concentrations, in reference to known absorption features in the red wavelength (Blackburn 1998; Gamon et al. 1997; Sims and Gamon 2002) and leaf scattering at the start of the NearIR regions. However, in the current experiment, it was observed that wavebands located in the shortwaveIR region as opposed to the red-edge and the NearIR regions were consistently important and showed greater discriminating power, for those datasets that contain wavebands in the shortwaveIR region. The persistent high importance obtained for the shortwaveIR wavebands (e.g. 1540, 2280 and 2300 nm) for separability of the target C3 and C4

grasses is consistent with the findings by (Noble et al. 2002; Smith and Blackshaw 2003), who also reported high importance for 1930, 2110 and 1990 nm wavebands.

As stated by (Smith and Blackshaw 2003), the reason for the good variable importance yielded for the shortwaveIR region is seemingly not as apparent as those of the visible and red-edge regions. Related to this, (Price 1994) reported that there are a number of limitations that can be encountered when working with spectral data for species level or top of the canopy reflectance data. These include problems related to the intra-species spectral variability, inter-species spectral similarity, and the effects of spectral mixing (i.e. in situations where multiple objects are within a particular detection space, resulting in their superposition). Nonetheless, the higher importance observed for the visible blue and the shortwaveIR bands in the current study can be attributed to the sensitivity of such wavebands to canopy biochemical and biophysical characteristics (e.g. stronger influence of leaf or canopy water absorption, leaf scattering and biomass), which can vary significantly between the C3 and C4 grasslands investigated. In general, results obtained from the current study thus confirm the relevance of the visible and shortwaveIR regions for applications involving the classification of tropical montane C3 and C4 types of grasslands. The moderate variable importance obtained for the red-edge region across datasets can also be attributed to inconsistent biochemical absorption and multiple-scattering effects in the red-edge region (Price 1994), among the types of grass species assessed. In addition, leaf or canopy mass, moisture content and canopy structure can significantly mask subtle spectral differences among the grass species assessed (Ceccato et al. 2001; Fensholt et al. 2010). In this regard, the results presented in Tables 4.3 to 4.5, although they clearly show that specific spectral regions have consistent variable importance rank for the C3 and C4 classification, further testing of the spectral utility is critical. However, the results obtained from the random forest-based variable importance highlighted the inherent spectral response characteristics of C3 and C4 grass species and the potential to programme wavebands for specific applications. The detailed variable importance results presented in Tables 4.3-5 are only for the top three datasets, which yielded the best overall classification accuracies.

Table 4.3 Random forest-based variable importance calculated on resampled inter-band r filter bands. Mean decrease in accuracy measures were initially calculated, using variable permutation based on the 70% training samples.

Rank	Band centre	Spectral range	Wavelength description	Mean decrease accuracy (MDA, %)
1	470	440–500	Visible ultraviolet	0.194
2	1540	1510–1590	ShortwaveIR-1	0.149
3	2300	2280–2340	ShortwaveIR-2	0.113
4	820	750–1000	NearIR	0.095
5	660	650–690	Red	0.071
6	2280	2110–2300	ShortwaveIR-2	0.063
7	530	520–560	Visible Blue	0.038
8	2450	2450*	ShortwaveIR-2	0.037
9	700	700*	Red-edge	0.029
10	720	720*	Red-edge	0.028
11	2470	2470*	ShortwaveIR-2	0.015
12	600	580–560	Red	0.014
13	2060	2050–2080	ShortwaveIR-2	0.009

* Indicates wavelength = 10 nm of the input spectral band, at the $r \geq 0.99$.

Table 4.4 Random forest-based variable importance calculated on resampled ASTER dataset.

Rank	Band description	Spectral range (nm)	Wavelength description	MDA (%)
1	Band 7	2235–2285	ShortwaveIR-2	0.158
2	Band 4	1600–1700	ShortwaveIR-1	0.140
3	Band 8	2295–2365	ShortwaveIR-2	0.135
4	Band 1	520–600	Green	0.086
5	Band 2	630–690	Red	0.081
6	Band 3	760–860	NearIR	0.065
7	Band 5	2145–2185	ShortwaveIR-2	0.058
8	Band 6	2185–2225	ShortwaveIR-2	0.044
9	Band 9	2360–2430	ShortwaveIR-2	0.037

Table 4.5 Random forest-based variable importance calculated on resampled WorldView-2 dataset.

Rank	Band description	Spectral range (nm)	Wavelength description	MDA (%)
1	Band 1	400–450	Coastal	0.118
2	Band 8	860–1040	NearIR-2	0.109
3	Band 5	630–690	Red	0.107
4	Band 2	450–510	Blue	0.106
5	Band 7	770–895	NearIR-1	0.106
6	Band 3	510–580	Green	0.105
7	Band 6	705–745	Red-edge	0.103
8	Band 4	585–625	Yellow	0.033

4.3.3 The relative importance of resampled WV-2 bands and derived indices

Using the resampled WorldVview-2 bands (n = 8) and 6 NDVIs developed from two band combination, the target C3 and C4 grass species' spectral responses were classified. The random forest model provided variable importance scores, which were used to explore the relevance of the input spectral bands and NDVIs in classifying the three categories of species. Table 4.6 shows the variable importance scores measured in terms of the mean decrease in the OOB accuracy. Notably, NDVIs 5 and 4 respectively ranked best among the input 13 WorldView-2 features and have more decreasing effect on the random forest model classification accuracy. The WordView-2 bands yielded similar performance, which is consistent with the MDA obtained for the analysis of only the resampled WorldView-2 bands. For example, the WorldView-2 Band1 (coastal blue) yielded the most influence on the overall performance of the random forest model in both cases. In addition, the NDVI5 which was calculated from combination of the coastal blue and the red-edge (Band6) produced a relatively high mean decrease in accuracy on the random forest model classification analysis.

Table 4.6 Random forest-based variable importance calculated on resampled WorldView-2 + 6 NDVIs.

Rank	WorldView-2 Band/NDVI	MDA (%)
1	NDVI5	0.219
2	NDVI4	0.128
3	Band1	0.051
4	Band8	0.049
5	NDVI2	0.034
6	NDVI6	0.029
7	NDVI3	0.025
8	Band2	0.025
9	Band7	0.023
10	NDVI6	0.023
11	Band6	0.022
12	Band5	0.020
13	Band3	0.015
14	Band4	0.010

4.3.4 Utility of the inter-band correlation filter spectra versus the resampled multispectral sensors

Classification accuracies and the random forest's ntree and mtry optimized hyper-parameters are shown in Table 4.6. The random forest-based OOB samples provided accurate assessment of the error rates (Adam et al. 2009; Dye et al. 2011). The algorithm yielded better OOB error rates for

the User-defined inter-band correlation filter dataset, compared to those of the resampled multispectral sensors. The dataset yielded respective classification accuracies of kappa = 0.79 and 0.82 for training and test data. The proposed User-defined inter-band correlation filter approach is a simple method of resampling hyperspectral data, yet it can yield superior classification accuracies.

Amongst the resampled multispectral sensors, the ASTER dataset which consisted of six ShortwaveIR wavebands in the 400–2500 nm spectrum yielded the highest classification accuracy. The ShortwaveIR bands of the ASTER data were highly ranked, as observed for those of the User-defined inter-band correlation filter dataset. It is interesting to note that the resampled multispectral datasets (Table 4.2), which do not contain wavebands in the shortwaveIR region or their spectral resolutions, are not optimally configured for C3 and C4 grass spectral separation yielded relatively lower classification accuracies. Therefore, higher classification accuracies obtained for the resampled ASTER dataset can be attributed to the position of its wavebands, specifically, the shortwaveIR bands.

The resampled WorldView-2 dataset yielded classification accuracies, comparable to those of the simulated ASTER dataset. It is important to note that although, the resampled WorldView-2 does not consist of ShortwaveIR wavebands, its 8-bands can be considered optimally positioned to provide good classification accuracies, compared to those of QuickBird or SPOT 5. With reference to the positioning of wavebands, it is should be noted that the 8-bands of WorldView-2 yielded equal mean decrease accuracy values, which confirms that the WorldView-2 wavebands are strategically positioned for applications involving C3 and C4 grass species. In contrast, results obtained in this study have shown that existing multispectral sensors such as QuickBird or SPOT 5 wavebands are not optimally positioned and are therefore less suited for the classification of C3 and C4 groups of grass species. Overall, analysis of Table 4.7 showed important relationship amongst the random forest accuracy and spectral resolutions for classification application involving C3 and C4 grass species. This underlines the complexity and importance of this study.

Table 4.7 Random forest model optimization and accuracy measures, using all 70% training sample datasets

Resampled spectra	Number of Bands	Optimized mtry	Optimized ntree	OOB error rate	Kappa: Training set	Kappa: Test set
Inter-band r = 0.992	13	4	4000	0.14	0.80	0.81
ASTER	9	3	2000	0.20	0.70	0.78
GeoEye-1	5	4	1500	0.25	0.62	0.65
IKONOS	4	4	2500	0.26	0.61	0.62
QuickBird	4	2	1500	0.27	0.60	0.59
RapidEye	5	5	500	0.26	0.62	0.65
SPOT 5	4	1	4000	0.27	0.60	0.62
WorldView-2	8	3	5000	0.21	0.72	0.76
WorldView-2 + 6-NDVIs	13	8	5000	0.17	0.75	0.78

4.4 Conclusions

An innovative resampling approach using a User-defined inter-band correlation filter approach was evaluated in this study and optimal spectral resolutions obtained were compared with those of some existing multispectral sensors' waveband configuration. An understanding of the cross-correlated information in the matrix of the data was gained, providing the potential to optimize spectral resolutions for specific regions of the measured spectrum. The approach of this study can assist rangeland ecologists who are involved in modeling plants functional types of C3 and C4 grasses, to select programmable sensor types for specific applications. Specifically, this study emphasizes the call for optimized spectral resolutions and band configuration, taking into cognizance that future satellites might be designed to carry specialized sensors, targeting selected wavebands for specific applications (Dalponte et al. 2009; Taylor et al. 2012). From this background, the following highlights summarize the findings of this study:

- Spectral resampling of hyperspectral data to a reduced number of bands should account for the asymmetrical nature of inter-band correlation curves around a few chosen band-centres of interest.
- Accounting for the asymmetrical nature of inter-band correlations between a given band-centre of interest and its short and longer wavelength neighbours can result in an optimal spectral resolution for the concerned band-centre.
- Wavebands located above 2000 nm of the ShortwaveIR region are critically required for the classification of C3 and C4 grass canopies: wavebands from this region are conspicuously absent for some existing multispectral sensors assessed. Hence, the

proposed method achieved improved classification accuracy (Kappa = 0.81), when compared to the resampled ASTER, GeoEye-1, IKONOS, QuickBird, RapidEye, SPOT 5 and WorldView-2 satellites (Kappa = 0.78, 0.65, 0.62, 0.59, 0.65, 0.62 and 0.76, respectively).

- NDVIs computed from a two band combination of resampled WorldView-2 band offer further capability to achieve improved classification of C3 and C4 grasses.
- Robust, non-parametric algorithms, such as the random forest was used both, as a screening tool for selecting predictor variable importance and as a classification technique for complex C3 and C4 grass applications, following its successful applications in other studies.

It is important to note that subtle similarities in plant phenology, seasonality and canopy geometrical properties can significantly affect the uniqueness of the spectral characteristics of C3 and C4 grass signatures. Therefore these should be considered as a significant limitation to the utility of the resampling approach adopted in this research (Price 1994). Despite the limitations, recent paper by (Taylor et al. 2012) and the present research offer a platform to optimize spectral resolutions and band configuration for applications involving specific grass species. Although the spectral resampling and classification approaches in the present study are general and can be applied in other remote sensing domains, further validation of the User-defined inter-band correlation filter technique at various growth stages, in other plant forms and, from a variety of environments need to be established.

The experimental results reported in this paper offer the potential of field spectral dataset in the classification of C3 *and* C4 *grasses. In addition, the newly proposed user-defined inter-band correlation filtered spectra were also resampled to the band configuration of the various space-bone sensors (i.e. ASTER, GeoEye-1, IKONOS, QuickBird, RapidEye, SPOT 5, WorldView-2), using the conventional approach. It is worth to note that field spectrometer data have been widely used to measure vegetation biophysical and biochemical properties across different environments, for specific applications, in situations without available Hyperspectral or multispectral imagery. Researchers often use this approach in testing newly developed ideas and then scale-up to aerial or satellite platforms. Thus the following Chapter 5 evaluates the capability of actual image spectra of the new generation Worldview-2 imagery collected for the Cathedral Peak region of the Drakensburg area of KwaZulu-Natal. It focuses on mapping the spatial distribution of the key* C3 *(F. costata) and* C4 *(T. triandra and R. altera) grassland communities, including the fern, Pteridium aquilinum, commonly known as bracken fern.*

CHAPTER 5: MAPPING AND CHARACTERIZING THE DISTRIBUTION OF C3 AND C4 GRASSLANDS

The chapter is based on:

Adjorlolo, C., Mutanga, O., & Cho, A.M. (in preparation). Mapping and characterizing C3 and C4 grass distribution in the subtropical montane rangeland of Southern Africa. *ISPRS Journal of Photogrammetry and Remote Sensing*

Abstract

There is uncertainty regarding the extent and distribution of grasslands following the C3 and C4 photosynthetic pathways, specifically, in the subtropical montane rangelands of Southern Africa. These two groups of grasses differ in a number of physiological, structural and biochemical characteristic properties and will respond differently to subtle changes in climatic conditions. These properties affect the canopy spectral response of grasslands differently, depending on their C3/C4 composition. Thus, it is possible to estimate and map the C3/C4 composition in grasslands, using remote sensing data. However, studies on rangeland ecosystems, which focus on the assessment of the composition of the C3 and C4 grasses, have inconsistently benefited from traditional broadband multispectral systems. In that regard, the use of higher spectral and spatial resolution data should offer the improved capability of studying the C3 and C4 composition in grasslands. This was evaluated, using a new generation satellite, WorldView-2 (WV-2) imagery, acquired for the Cathedral Peak region of the Drakensberg Mountain range, South Africa. The raw WV-2 bands, as well as six (6) normalized vegetation indices, were integrated in the random forest classification to map the C3 and C4 vegetation cover types in the area. The WV-2 8-bands combined with 6 NDVIs yielded the best overall classification accuracy of 82.26% and kappa = 0.76 when the RF model was validated on an independent dataset. The results from this study highlighted the potential of WV-2 image in assessing the C3 and C4 species in a tropical montane environment.

Keywords: Grass species classification; natural resources; worldview-2; random forests

5.1 Introduction

An accurate prediction of the effect of global change on grasslands requires the grouping of the species with similar organismic characteristics that result in a common ecological role (Ehleringer et al. 2001; Ustin and Gamon 2010). Thus, based on their photosynthetic pathways, grassland species can be broadly classified into the C3 and C4 functional type of species. The C3 or C4 denotes the initial photosynthetic products of 3-carbon and 4-carbon compounds, respectively. The C3 and C4 grasses differ greatly in their environmental requirements and respond differently to changes in the environmental conditions (Ehleringer and Björkman 1977; Tieszen et al. 1997). In general, the C4 type of plants are more adapted to conditions of high incident radiation (Tieszen et al. 1997), elevated temperature (Schuster and Monson 1990), limited soil moisture (Niu et al. 2005), and low CO_2 concentrations (Ehleringer and Björkman 1977), than are the C3 type of grass species.

Controlled by the environment, the C3 and C4 grasses within the montane region of Southern Africa are expected to be more sensitive to global change, including changes in land-use. The temperate C3 grasses in this region exist under conditions that are sub-tropical and marginal for growth. Therefore, the C3 type of grasses is expected to be highly responsive to even subtle changes in the global climate, including changes in temperature, atmospheric CO_2 and moisture concentrations. Consequently, the floristic composition of C3 grassland communities are expected to be potentially more dynamic than other grasslands in the region (Adjorlolo et al. 2012b). Both the C3 and C4 grasses in the montane grasslands play an important role in the carbon cycle, provide forage for herbivores and influence a host of catchment dynamics. Therefore, it is critical to accurately characterize and map the spatial distribution of these two groups of grasses, in order to make inferences on the possible impacts of global change on the montane grasslands in the region.

It is well-known that remote sensing systems provide spatial data on vegetation taxa, where ground-based knowledge is insufficient for mapping rangeland vegetation cover. Remote sensing techniques have increasingly become the method of choice to obtain quantitative data for the classification of plant species or communities. Specifically, traditional remote sensing data have been used to assess the relative proportions of C3 and C4 grass communities, mainly in the temperate regions (Davidson and Csillag 2001, 2003; Foody and Dash 2007; Goodin and Henebry 1997). These studies focused on discriminating and mapping the C3 and C4 grassland

communities, using multi-temporal remote sensing techniques. For example, Foody and Dash (2007) discriminated and mapped the composition of grasslands in the northern Great Plains of USA, based on the characteristic asynchronous seasonal profiles exhibited by the C3 and C4 grasses in the region. The authors evaluated weekly composites of the normalized difference vegetation index (NDVI) calculated from MERIS MTCI (Dash and Curran 2004), by deriving relationships between the remotely sensed features against the C3 and C4 grasslands. Foody and Dash (2007) reported R^2 ~0.6 and an overall classification accuracy of 77.8% for a three class (high, medium and low) C3 cover in the area.

In another study, Liu and Cheng (2011) showed that the chlorophyll fluorescence (ChlF) signal for the C4 species was about 2.2 times greater than that of the C3 species. Using a simple decision tree, the C3 and C4 species composition in the Chinese Linze Oasis, in the Gansu Province, was discriminated, based on the differences in the ChlF signal, with an overall classification accuracy of 92% and a kappa coefficient of 0.84. The results of these studies are better, when compared to those of Davidson and Csillag (2001), who obtained an overall classification accuracy of 56% for a five broad-ordered classes of differing C3 and C4 grassland communities. The authors attributed the weak classification accuracy to the pixel mixing phenomenon of the remotely-sensed spectral responses in the heterogeneous environment.

The main challenge in using remote sensing data lies in inferring the relative abundances or distributions of the C3 and C4 species in mixed canopies (Goodin and Henebry 1997). This is because, at the canopy level, both the C3 and C4 photosynthetic activities would have potentially contributed collectively, in characterizing the reflectance values of the input spectral bands. Despite the challenges, remote sensing applications involving C3 and C4 grasses continue to yield encouraging results (Adjorlolo et al. 2012a; Guan et al. 2012). The present study sought to discriminate and map the distribution of the C3 and C4 grass communities, representative of the sub-tropical montane grasslands of the Southern African environment. This was done by evaluating the capability of new generation multispectral satellite, WorldView-2 (WV-2) imagery. The WV-2 instrument is configured to acquire spectral data in both visible and near-infrared (400–1040 nm) regions of the spectrum, with a spatial resolution of 2 m. In this study, the responses of the 8 spectral bands of WV-2, as well as normalized difference vegetation indices (NDVIs) were integrated, with the widely-used random forest (Brieman, 2001) algorithm, to classify the C3 and C4 communities in the study area.

5.2 Methods

5.2.1 Considered vegetation communities in the study area

The study area is located in the Cathedral Peak region of the Drakensberg Mountains, South Africa (Figure1.1). The specific study site is located at 29°14'10.347"E and 28°59'27.383"S. The proportion of C3 and C4 grasslands in the greater Drakensberg Mountain range varies with altitude. For example, the C3 community can vary from a near complete absence of C3 grasses in the lower altitudes, to over 80% C3 cover at about 2800 m above mean sea level (Killick 1963; Morris et al. 1993).

The area of interest for the current study is underlain by relatively homogeneous, Drakensberg basalt formations and is characterized by several mountain peaks. Patches of the C3 and C4 type of grasslands were sampled across the Montane and sub-Alpine vegetation belts (1280–2865 m above mean sea level). The sub-Alpine belt has been described as *Themeda-Festuca,* sub-Alpine grassland (Killick 1963). It is further divided into two grassland communities, known as the *Themeda triandra* (red grass), and *Festuca costata* (tussock fescue) sub-Alpine grasslands. There are strong associations of *T. triandra* and *F. costata* occurrence on warm, northerly and cool, southerly slopes, respectively (Granger and Schulze 1977; Hill 1996; Killick 1963). The area is also characterized by high rainfall (> 900 mm), with misty conditions, low temperatures, seasonal frosts and snow. Fire activities, herbivory and topographic factors, play an important role in influencing the composition of the C3 and C4 grasses in the area.

The C4 grassland communities consist mainly of the *T. triandra* and *Rendlia altera* (Toothbrush grass) species, while the C3 community consists mainly of the unpalatable *F. costata*, with its co-existing C3 species: *Poa binate*, *Agrostis barbulegera*, and *Merxmuellera macowanii*. The *F. costata* and *M. macowanii* remain evergreen throughout the dry (winter) season. They normally grow as large tussocks that are taller than the surrounding C4 grasslands. They also grow in an aggregated pattern, which makes them easily distinguishable in the field. The fern *Pteridium aquilinum* (Bracken), which is of key research interest in the region, is considered in this study. Bracken is a cosmopolitan invasive weed that presents a serious environmental problem in many regions of the world. It is a poison to herbivores (Blackburn and Pitman 1999). Similar to *F. costata*, bracken often grows in dense stands, but may also display scattered distribution of individual plants in grassland, likely when it is invading. In the current

study area, bracken occurs on all topographic aspects, with local dominance in areas of elevated soil moisture and unlike the *F. costata*, it is not evergreen (Killick 1963).

The dominant C4 grassland communities can make up 70%, or more, of the composition in a patch. The *R. altera* is characterized by relatively narrow leaves, which tend to fold and break-up into fibres in mature plants. This results in low leaf production and in a relatively low canopy density. The *T. triandra* forms dense and tufted stands that are distinctly red-coloured, when they are fully mature. Differences in canopy density, structure and geometry, as well as species physiological pathways, underlie the importance of this study.

5.2.2 Image acquisition, pre-processing and feature extraction

The WV-2 satellite imagery was acquired on 28 April 2011. It is the first commercial eight multispectral bands high resolution sensor (DigitalGlobe, Inc., USA), operating over the wavelength range 400–1040 nm and it has variable spectral resolutions. The satellite has a swath width of 16.4 km, a revisit time of 1.1 average day and a spatial resolution of 2 m for 8 multispectral (MS) channels: coastal blue (400–450 nm), blue (450–510 nm), green (510–580 nm), yellow (585–625 nm), red (630–690 nm), red-edge (705–745 nm), NIR-1 (770–895 nm) and NIR-2 (860–1040 nm). These 8 bands are uniquely spectrally positioned to meet the needs of a variety of applications, including natural resources management, vegetation mapping, environmental monitoring and several other applications (DigitalGlobe, 2009).

The mean in-track, cross-track and off-nadir view angles were -3.1°, 14.9° and 15.3°, respectively. The WV-2 satellite's mean azimuth and elevation angles were 112.1° and 72.9°, respectively, and the sun azimuth and sun elevation angles were 32.6° and 40.8°, respectively. The image data used in this study cover an area of 25 km^2, which were delivered as product level LV3D (DigitalGlobe, Longmont, Colorado, USA). That means the image had been radiometrically corrected and orthorectified by the supplier. The registration mean error of the image was 0.15 m.

The digital numbers were converted to WV-2 radiance, using the absolute radiometric calibration factors and effective bandwidths for each band in the ENVI routine (ENVI 4.8). The resultant radiance image was then atmospherically corrected, in order to reduce haze and other atmospheric influences. Since the area consists of rugged terrain, the ATCOR 3 module in ERDAS Imagine 2011 (Erdas, 2011) was used. This module is capable of controlling

topographically varying optical effects on the input image data. Atmospheric correction functions were calculated, using the MODTRAN 4 code developed for the WV2 sensor. Topographic corrections were calculated with the only available 20 m digital elevation model (DEM). However, it turned out that the 20 m DEM is insufficient for the effective correction of topographic effects on the WV-2 dataset for the landscape under investigation. This was the case, even when the 20 m DEM was re-sampled, step by step, from 20 m to 10 m and 5 m, in order to match the WV-2 spatial resolution. The output images were inaccurate, with a substantial positional shift between the DEM and the WV-2 pixels, showing a chessboard pattern of the output image. Therefore, topographic correction was excluded from the analysis, in order not to introduce an additional error to the reflectance values. In this regard, steep slopes and sheltered areas were masked from all analyses, which is an area for future investigation.

In addition to the raw 8 multispectral bands of WV-2, a total of 6 normalized vegetation indices (NDVIs) were calculated:

NDVI1 = (NIR-1−green)/(NIR-1+green) (1)

NDVI-1 = (NIR-1−red)/(NIR-1+red) (2)

NDVI-2 = (NIR-2−red-edge)/(NIR-2+red-edge) (3)

NDVI-3 = (NIR-2− yellow)/(NIR-2+ yellow) (4)

NDVI-4 = (red-edge− coastal blue)/(red-edge + coastal blue) (5)

NDVI-5 = (red-edge −red)/(red-edge +red) (6)

The inclusion of these vegetation indices was based on previous studies (Mutanga et al. 2012; Pu and Landry 2012), in which their utility in other remote sensing applications was shown, involving the estimation, discriminating and mapping of plant species.

5.2.3 Collecting ground truth data

Although an account of the plant ecology of the Cathedral Peak region of the Drakensberg is well-documented (Granger and Schulze 1977; Hill 1996; Killick 1963), there is no accurate map depicting the C3 and C4 grass composition in the grasslands, for use as a reference dataset. The current study forms part of a series of field investigations conducted in the area, whereby ground truth data were collected, based on some key C3 and C4 species in the area. A total of (n = 352) ground truth data, including those reported in Adjorlolo et al. (2012a) were collected from homogeneous large patches to characterize the selected vegetation classes in the current study.

The sample size was increased by obtaining data from surrounding pixels. Table 5.1 shows the resultant data (n = 435) used to derive a point distribution map of the species considered in this investigation. Other vegetated cover types were also recorded, including natural forests and other woody bushes in the area. Pixels representing these cover types, including bare surfaces, were used to estimate the minimum and maximum NDVI threshold values. These were then used to create a region-of-interest (ROI) map, using ENVI v4.8 software. Next, the ROI map was used to create a spectral mask of the woody vegetation and bare surfaces. The pixels of WV-2 image bands, as well as the developed NDVIs, were then extracted, using the zonal extract tool in ArcGIS 10 software.

Table 5.1 Training and validation samples used in analysis of the four dominant grasslands and Bracken communities.

Cover class	Number of pixels	
	Training	Validation
F. costata grassland community	70	28
P. aquilinum stands	54	34
T. triandra dominant-grassland	95	22
R. altera mixed-grassland	92	40
Sum of input raw WV-2 bands/NDVIs pixel values	311	124

5.2.4 Data analysis procedure

In this study, the random forest (RF) algorithm was implemented to utilize bootstrap samples with no replacement, in order to grow a large ensemble of classification trees. The ensemble of trees assign each waveband to a class, based on the maximum number of votes that a class receives from the collection of trees (Breiman 1996, 2001a). Each tree is grown from randomly- and independently-selected bootstrap samples of the training or calibration data (n = 311). About one-third excluded samples (out-of-bag or OOB cases) are used to evaluate the model. The OOB samples have been reported, to offer unbiased estimates of training error (Ismail and Mutanga 2011; Lawrence et al. 2006; Pal 2005; Prasad et al. 2006). The algorithm is easy to implement, since only two parameters are tuned: (i) the number of trees to grow (*ntree*); and (ii) the number of variables to split at each node (*mtry*).

RF also has variable importance measurements embedded in its computation, and it is capable of performing both binary and multiclass classification tasks (Breiman and Cutler 2004). These variable importance measures provide researchers with valuable insights, to explore the effect of any predictor variable on the response variable (Lawrence et al. 2006; Pal 2005). In the current study, the parameterisation of the algorithm was done, using the RF tuning provided in the EnMAP-Box v1.3 programming software. This was done, using the randomly-split training or calibration dataset and the model validation was carried out, using the independent test or validation dataset (Table 5.1). The EnMAP-Box is open source license software, which was developed, implemented and validated by the German Hyperspectral Environmental Mapping and Analysing Program (EnMAP) mission at the Geomatics Laboratory, Humboldt University, Berlin (www.enmap.org). This software outputs a spatial map, depicting the distribution pattern of the input vegetation classes.

5.2.4.1 Random forest-based variable importance

The RF variable importance was assessed, based on the random permutation of the predictor variables (Strobl et al. 2008; Strobl et al. 2007b). The algorithm randomly permutes each WV-2 feature's association with the input classes, to identify the magnitude of its impact in classifying the response variables (Adjorlolo et al. 2013; Lawrence et al. 2006). The random permutation of a predictor variable represents the absence of the variable from the model. The difference in prediction accuracy prior to, and after, permuting a variable, is used to predict the OOB observations. Thus, the number of observations classified correctly decreases substantially, if the permuted variable was strongly associated with the response variables (Cutler et al. 2007; Strobl et al. 2008).

5.3 Results and analysis

5.3.1 Classifying the C3 *and* C4 *plants/communities in the area*

Using the 14 features extracted from the WV-2 image (i.e. the 8 bands only and, integration of the 8-bands and the 6 NDVIs), each C3 or C4 grassland community was classified using the calibration datasets (Table 5.1). The RF model tuned with an mtry = 3 and the default ntrees = 500 for the analysis involving only the WV-2 bands and, tuned with an mtry = 4 and default ntrees = 500. The default 500 ntrees was used on the basis that ntree values larger than the

default are reported to have little influence on the overall classification accuracy (Breiman and Cutler 2004). The OOB samples for both sets of data analyzed in the current study were used to assess the internal cross validation of the models construction. Table 5.3 shows accuracies obtained using the training datasets (8-bands WV-2 only dataset and, the 8-bands of WV-2 combined with 6-NDVIs). The classification models validated on an independent test dataset and yielded overall classification accuracies of 74.19% with kappa = 0.66 and 82.26% with Kappa = 0.76 for the two sets of input spectral data (Table 5.2).

The individual class identification results, based on the model training and validation datasets are presented in Tables 5.3 and 5.4, including the confusion matrices obtained from the RF analysis. From these matrices it is clear that the number of correctly classified grassland or *P. aquilinum* stands is high. The Producer's and User's accuracies reveal that all the input classes produced satisfactory identification results. Interestingly, the C3, *F. costata* community and the fern, *P. aquilinum* stands have a relatively higher error of omission and commission, respectively, compared to that of the dominant C4 grass, *T. triandra* and *R. altera* mixed-grassland. Marginal differences between omission and commission errors were obtained for the C4 dominant *T. triandra* and *R. altera* mixed-grassland canopies. Lower confusion between *T. triandra* and all other canopy classes was obtained. On the other hand, the *P. aquilinum* stands has produced the lowest omission error (6.9%).

Table 5.2 Classification accuracies obtained using the WV-2 bands only dataset and WV-2 bands and the 6 NDVIs computed.

	Training		Test set	
Data set	Over all accuracy (%)	Kappa	Overall accuracy (%)	Kappa
WV-2 bands only	73.95	0.65	74.19	0.66
WV-2 bands + 6NDVIs	80.39	0.74	82.26	0.76

Table 5.3 Confusion matrices reflecting results of the input spectral features extracted from the WV-2 multispectral 8-bands image only.

	Predicted class				
Observed Class	*F. costata* grassland community	*P. aquilinum* stands	*T. triandra* dominant-grassland	*R. altera* mixed-grassland	Sum of reference
Training dataset					
F. costata grassland community	54	14	5	4	77
P. aquilinum stands	8	55	5	2	70
T. triandra dominant-grassland	2	9	71	13	95
R. altera mixed-grassland	5	0	14	50	69
Sum of estimation	**69**	**78**	**95**	**69**	**311**
Producer's Accuracy (%)	**78.26**	**70.51**	**74.74**	**72.46**	
User's Accuracy (%)	**70.13**	**78.57**	**74.74**	**72.46**	
Validation dataset					
F. costata grassland community	*20*	*2*	2	2	26
P. aquilinum stands	*5*	*23*	2	4	34
T. triandra dominant-grassland	*4*	*4*	21	6	35
R. altera mixed-grassland	*0*	*1*	0	28	29
Sum of estimation	**29**	**30**	**25**	**40**	**124**
Producer's Accuracy (%)	**68.97**	**76.67**	**84.00**	**70.00**	
User's Accuracy (%)	**76.92**	**67.65**	**60.00**	**96.55**	

Table 5.4 Confusion matrices reflecting results of input spectral features extracted from WV-2 8-bands and the 6 NDVIs combined.

	Predicted class				
Observed Class	*F. costata* grassland community	*P. aquilinum* stands	*T. triandra* dominant-grassland	*R. altera* mixed-grassland	Sum of reference
Training dataset					
F. costata grassland community	59	5	5	1	70
P. aquilinum stands	8	49	3	3	63
T. triandra dominant-grassland	2	5	71	11	89
R. altera mixed-grassland	4	0	14	71	89
Sum of estimation	**73**	**59**	**93**	**86**	**311**
Producer's Accuracy (%)	**80.82**	**83.05**	**76.34**	**82.56**	
User's Accuracy (%)	**84.29**	**77.78**	**79.78**	**79.78**	
Validation dataset					
F. costata grassland community	20	2	0	0	22
P. aquilinum stands	5	27	0	5	37
T. triandra dominant-grassland	1	0	20	5	26
R. altera mixed-grassland	2	0	2	35	39
Sum of estimation	**28**	**29**	**22**	**45**	**124**
Producer's Accuracy (%)	**71.43**	**93.10**	**90.91**	**77.78**	
User's Accuracy (%)	**90.91**	**72.97**	**76.92**	**89.74**	

5.3.2 The relative importance of WV-2 derived features

The OOB sample estimation accuracies were used to measure the importance of the input WV-2 8 bands and the 5 computed NDVIs, combined. The RF model provided the variable importance scores, to explore the relevance of the input predictor in classifying the four canopy classes. Figure 5.1 shows the variable importance scores measured in terms of the decrease of OOB accuracy, which represents the deterioration of the classification performance of the model, when each predictor is permuted. Notably, four of the predictors (Bands 5, 6 and 6 and NDVIs 3 and 5) have a 10%, or more, decreasing effect on the model. As was expected, WV-2 band5 yielded the most influence on the overall performance of the RF model, in discriminating and mapping dominant vegetation canopies. Noticeably, the NDVI calculated from combination of the coastal blue (band1), red-edge (band6), produced a relatively high mean decrease in accuracy of 12% influence on the RF model classification analysis.

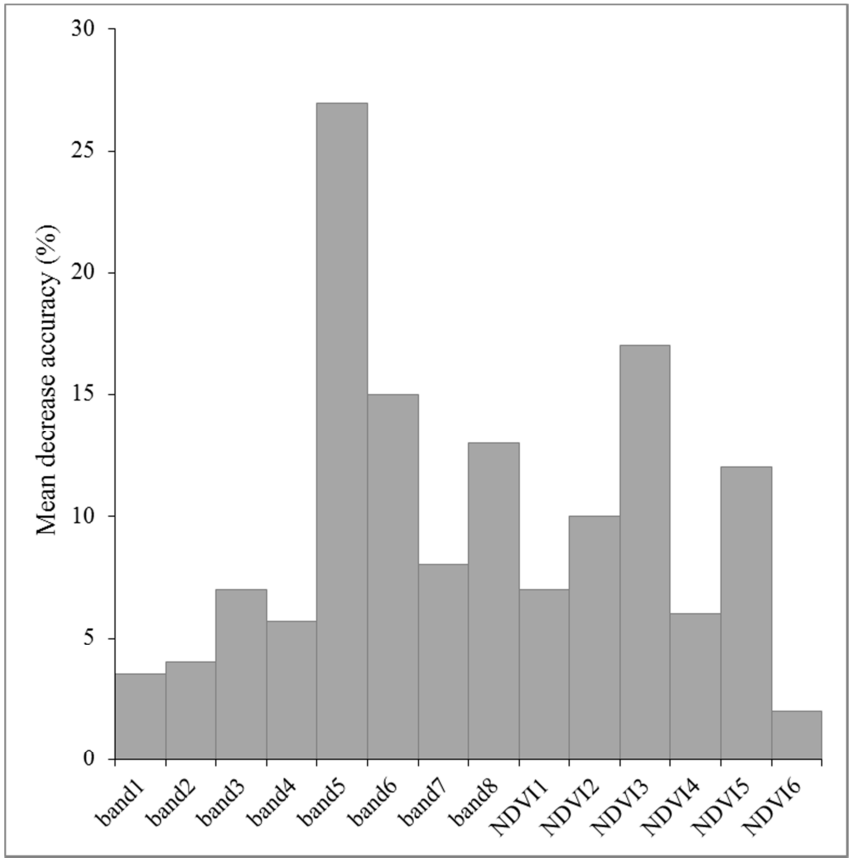

Figure 5.1The relative importance of the raw multispectral bands and 6 NDVIs features derived from the WV-2 image, which yielded the best overall classification and kappa accuracies.

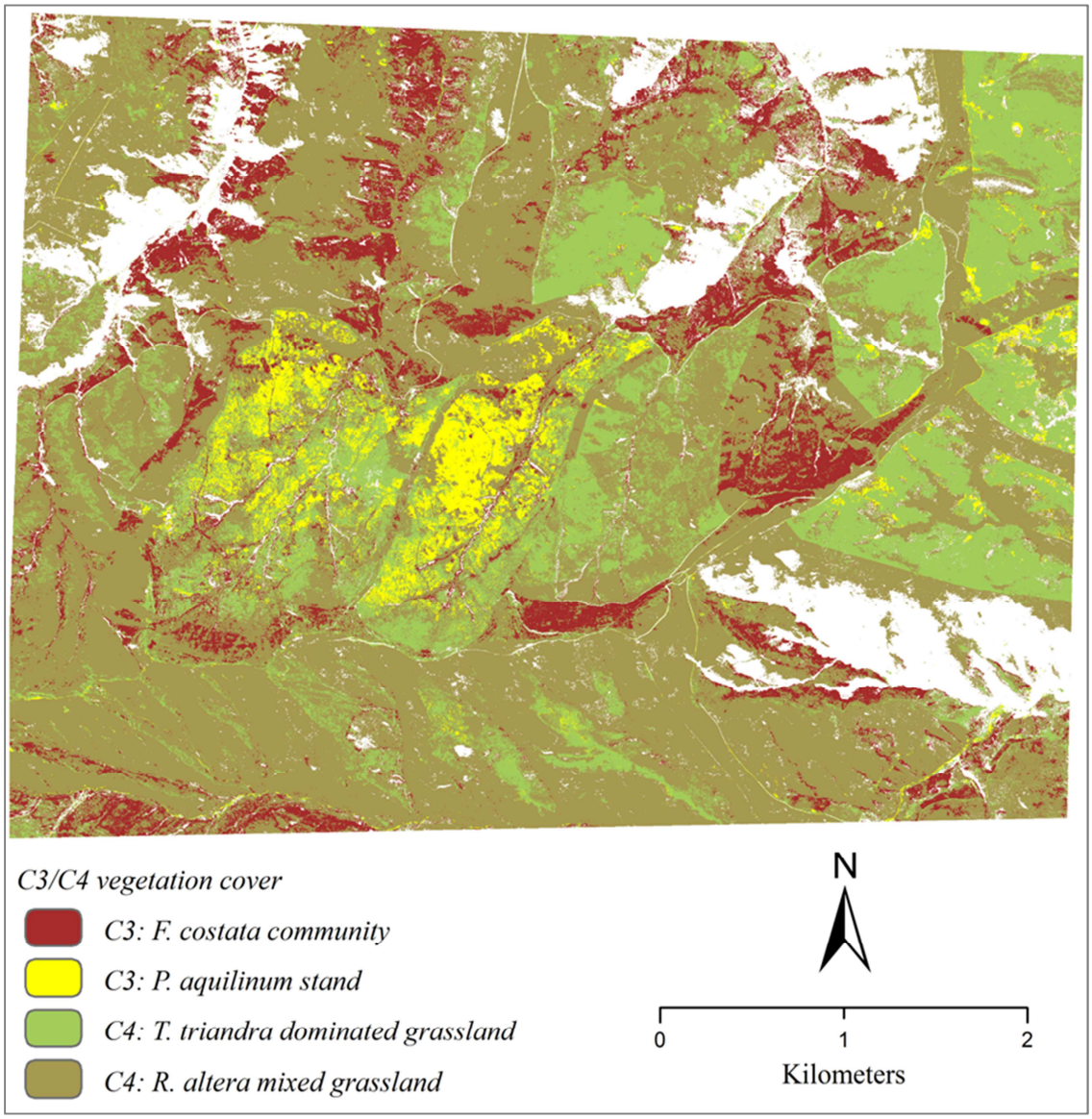

Figure 5.2 Distribution of the C3 and C4 grass communities, including the bracken fern (Pteridium aquilinum), mapped using the 8-bands and 6 NDVIs developed from the 2 m spatial resolution images of WorldView-2. The white patches depict masked natural forests, tracks and sheltered relief.

5.4 Discussion

The classification produced for the C3 and C4 plant communities in this study demonstrates that WV-2 sensor has the capability to identify and map these communities in a subtropical montane environment. This general conclusion confirms the previous field research that evaluated the potential of the WV-2 sensor, by simulating its spectral bands to classify the C3 and C4 grasses in the area (Adjorlolo et al. 2012a). Other studies evaluated WV-2 image data, to identify individual tree species in a savannah environment (Cho et al. 2012) or mapping the spatial distributions of tree species (Ozdemir and Karnieli 2011; Pu and Landry 2012). For example, Pu and Landry (2012) found out that WV-2 bands yielded a higher overall accuracy, compared to the same bands of a traditional multispectral sensor, IKONOS imagery. In a previous study by Adjorlolo et al. (2012a), the coastal, red and NIR-2 bands of WV-2 were the top three ranked bands, using the RF model, to classify individual C3 and C4 grass species in the area, at the canopy level. Noticeably in their study, the RF model's mean decreases in accuracy values were substantially marginal.

For the current study, the lower class accuracy obtained for the C3, *F. costata* community and the fern, *P. aquilinum* stands, is explained by the general patterns observed in the field. This is in consideration of the characteristic seasonal/phenological effect on the *F. costata* and *P. aquilinum* in April, when the WV-2 imagery was acquired. The evergreen *F. costata* is a tussock-forming grass, which very rarely forms a pure sward in the area (Killick 1963). The tussocks of this species are separated by other grasses, such as the *T. triandra*, *R. altera*, *Poa binate*, *Agrostis barbulegera*, etc., and these possibly caused confusion in classifying pixels representing the *F. costata* tussocks. The higher commission error obtained for the C4 dominant *T. triandra* and *R. altera* mixed-grassland canopies is explained by the spectral similarity, due to canopy structural attributes. This was expected, on the basis that the so-called *R. altera* mixed-grassland generally comprises *T. triandra* tufted stands, in the grass sward.

On the other hand, the relatively lower confusion obtained among *T. triandra* grassland, *F. costata* community and *P. aquilinum* stands, is likely to be explained by the characteristic red-coloured pigmentation of the matured plants of the *T. triandra* canopy. This canopy characteristic is very different from the surrounding grasslands in April, when the WV-2 imagery was acquired. Furthermore, the *P. aquilinum* stands are characterized by localized, distinctly dense, canopy composition, which are spectrally different from the surrounding grasslands. It is

therefore concluded that the resultant spectra of the *P. aquilinum* stands are likely not to be significantly affected by background strata, such as understory grasses and soil, compared to the surrounding grasslands, considering the sensitivity of the WV-2 bands. For example, the red band of WV-2, which yielded the highest variable importance value (27%, mean decrease in accuracy) is sensitive to bare soils, roads and underlying geological features (Cho et al. 2012; Ozdemir and Karnieli 2011; Pu and Landry 2012; Toutin et al. 2012).

Overall, the variable importance results obtained have shown that the red, red-edge, NIR-2 bands, as well as NDVIs, calculated from a combination of the coastal (band1) and red-edge (band6) bands, yielded the highest accuracy in classifying and mapping the C3 and C4 grass canopies in the area. The very high importance for the combination of the coastal and red-edge bands (Figure 5.1) is explained by the variation of pigment concentration (e.g. chlorophyll) in the canopies of the different C3 and C4 grasses, as well as the fern *P. aquilinum* stands. The results of this study provide additional evidence to suggest the potential of the strategically-positioned (Mutanga et al. 2012) WV-2 bands for improved vegetation classification, even in subtropical montane grassland environments.

There are at least a few factors that may have contributed to the accuracy (Table 5.2) differences between the individual vegetation classes considered in this study. First, the small differences between the spectral response curves of the input spectral bands or developed vegetation indices (Adjorlolo et al. 2012a). Second, despite an attempt to control the topographic influence on pixel reflectance values, by restricting the analysis to fairly flat and non-shady areas, this study is limited, due to the effect of sun-target-sensor geometry, resulting partially from the differences in both slope and aspect across the landscape under assessment. These topographic factors and WV-2 sensor's off-nadir configuration have a substantial effect on the image acquisition. In general, these can cause variability in spectral and angular vegetation reflection, which changes as a function of canopy structural characteristics, and is commonly described as the bidirectional reflectance distribution effect (Asner 1998). The significant image displacement of the canopies and larger shadow/sheltered areas for this study is also explained by the large view angle of WV-2 sensor. This was a possible source of error, when WV-2 image data were utilized for classifying the C3 and C4 grasslands in a montane environment. Nonetheless, results obtained in this study, provided valuable evidence to suggest that, in

grasslands dominated by C3 and C4 species, differences due to seasonal phenology can be detected and mapped, using the WV-2 sensor, with satisfactory accuracy.

The findings in this study also suggest that, on the basis of achieving acceptable overall classification accuracy, the RF classification is a useful statistical modelling approach. The RF model was implemented, using 311 pixel representations of the input of the vegetation communities considered in this study, to produce the classification map. The spatial distribution of the four vegetated communities is depicted in Figure 5.2. Compared with previous study by Adjorlolo et al. (2012a; b), who evaluated the potential of detecting individual C3 and C4 grasses, using spectra only datasets, the classification accuracies obtained in the current study are limited. Although in an ideal situation, it is very difficult to map individual grass species, the current study has demonstrated that satisfactory accuracy can be obtained by using only one image snapshot from the Worldview sensor. The classification accuracy may, however, be improved by incorporating a time series approach that accounts for the phenological aspect in C3 and C4 grass studies (Davidson and Csillag 2003; Foody and Dash 2010; Goodin and Henebry 1997; Guan et al. 2012; Tieszen et al. 1997).

In summary, the map produced, depicting the spatial distributions of the C3 and C4 grass communities, should offer rangeland managers and ecologists a baseline to conserve and understand the grassland montane ecological system. This study confirms the potential to derive estimates of the C3 and C4 distribution in grasslands from satellite data. It has shown that the satisfactory overall accuracy obtained is attributable to the high spatial and spectral resolution of WV-2 image, as well as the time of the WV-2 image acquisition, which coincide with the characteristic features of the mature plant canopies in the area. However, the impact of topography, background spectrum of the swards, shadow/sheltered areas (missed by the developed spectral mask) should be factored in, to improve the classification accuracy.

5.5 Conclusions

The significance of this study was to demonstrate the potential of newly-developed high resolution WV-2 imagery for identifying and mapping subtropical montane C3 and C4 vegetation, with the state-of-the-art classifier, RF modelling approach. Using raw bands and NDVIs features extracted from WV-2 imagery, the RF learning algorithm was explored to create

a classification of C3 and C4 grasslands, with an overall accuracy of 82.26% and kappa = 0.76. The conclusions that were drawn from the experimental results are as follows:

- The WV-2 sensor has the capability of identifying and mapping broad groups of C3 and C4 grass communities, as well as mapping the spatial distribution of the fern, *P. aquilinum* stands in a montane environment.
- The new bands of WV-2, especially, the coastal-blue, red-edge and NIR-2 bands, were important predictor variables for mapping the four classes considered in this study.
- Overall, the WV-2 sensor has the potential to improve the accuracy of C3 and C4 grass classification in subtropical montane environments.
- To improve the accuracy of identification and mapping of these two groups, future studies should consider reflectance responses from other regions of the electromagnetic spectrum, including bands in the shortwave infrared region, which were found to be useful in previous studies (Adjorlolo et al. 2012a; Adjorlolo et al. 2013).

A full understanding and importance of the C3 *and* C4 *grasses can only be achieved by evaluating their distribution as well as their quality in the light of land use and climate change. The objective of the following Chapter 6 is to determine whether the estimation of canopy concentration of these nutrient variables nitrogen (N), crude protein (CP), moisture and non-digestible fibre can be improved by using the developed user-defined inter-band correlation filter approach of resampling hyperspectral data, in a* C3 *and* C4 *grassland environment.*

CHAPTER 6: PREDICTING C3 AND C4 GRASS NUTRIENT VARIABILITY

This chapter is based on

Adjorlolo, C., & Mutanga, O. (in review). Predicting C3 and C4 Grass Nutrient Variability, using in-situ Canopy Reflectance and Partial Least Squares Regression. *International Journal of Remote sensing*.

A paper from this study has been presented at the 47th Annual GSSA (Grassland Society of Southern Africa) Congress, 2012, Cape Town, South Africa.

Abstract

Climate-change induced alterations in the composition of C3 and C4 grass species will affect the carbon cycle and forage nutrient quality of rangelands. A probable increase in the abundance of the C3 grass, *Festuca costata* and other plants, which are relatively poor in nutrient quality, is a key concern within the African montane grasslands. Forage quality in these grasslands depends on limiting nutrients such as nitrogen (N), crude protein (CP), moisture, and non-digestible fibres that constrain the intake rate of herbivores. Using in-situ reflectance data resampled to 13 preselected band-centres, WorldView-2 band settings and to 10-nm-bandwidths (400 – 2500 nm), variability in nutrient concentration was predicted at full canopy cover, across C3 and C4 type of grasses. Results obtained from the partial least squares analyses indicate prediction accuracies ranging from 32% to 66% of the variance in N, CP, moisture, and fibre concentrations can be achieved, using spectral information. This study provides an appropriate platform for further studies in establishing the variance in C3 and C4 grass nutrient concentrations, focusing on various growth stages, a variety of environments and the level of canopy.

Keywords: spectral resampling, grassland, nutrient variability, partial least squares, remote sensing

6.1 Introduction

Shifts in the composition of plants following the C3 and C4 photosynthetic pathways are expected to impact on the global carbon cycle and have differing effects on forage production. Such shifts are presumed to be accompanied by global climate trends such as elevated temperatures and atmospheric CO_2 concentrations (Chamaillé-Jammes and Bond 2010; Davidson and Csillag 2001; Tieszen et al. 1997; Winslow et al. 2003). In general, the C3 and C4 type of grasses differ in their forage nutrient status and will respond differently to changing climatic factors (Barbehenn et al. 2004; von Fischer et al. 2008). Important variables include limiting forage nutrients such as nitrogen (N), crude protein (CP), moisture, and neutral fibre (non-digestible fibre) that constrain the intake rate of herbivores (Grant et al. 2010; Knox et al. 2011). Differences in forage nutrient concentration between the C3 and C4 grass species will affect the condition and capacity of rangelands to support herbivores (Lattanzi 2010; McNaughton 1990).

Most field-based techniques for predicting forage nutrient concentration are destructive and cannot be repeated often enough for high spatial resolution investigations. Remote sensing techniques have proved useful to overcome these challenges. Specifically, hyperspectral remote sensing systems, which consist of narrow contiguous wavebands, provide a continuous reflectance spectrum of the target vegetation. Hyperspectral systems have been used more successfully to detect and measure forage nutrient levels, compared to the limited, very broadband (>100 nm) multispectral systems (Hansen and Schjoerring 2003; Kruse et al. 2006; Mitchell et al. 2011; Townsend et al. 2003). Hyperspectral data acquired through field spectroscopy measurements offer unceasing means to investigate vegetation characteristics, as pursued in this study.

Although data acquired through hyperspectral systems provide information that is more sensitive to specific plant nutrient characteristics, the data are very high in dimension, with hundreds of input spectral bands. Therefore, the selection of a subset is regularly performed to mitigate against the negative effects of high dimensionality of the input hyperspectral bands. The selection of a subset has to be optimal, in order to retain the sensitivity of the input spectral bands, for the specific response variables (Dalponte et al. 2009). In that regard, a number of approaches have been evaluated, including the feature selection and feature reduction technique of analyses. These have been used to reduce dimensionality in the input hyperspectral data,

satisfying some statistical conditions (Bruzzone et al. 1995; Bruzzone and Serpico 2000). However, there is possibly no universal applicable technique, as most feature selection or reduction techniques reported in literature showed considerable disagreement in the selection of optimal subsets for specific applications (Chan and Paelinckx 2008; Hansen and Schjoerring 2003).

Several other studies have analyzed the location of the most informative spectral bands, considering the physical or spectroscopic meaning of each waveband (Adjorlolo et al. 2012a; Adjorlolo et al. 2013; Becker et al. 2007; Schmidt and Skidmore 2003). In order to mitigate against the multicollinearity problem often caused by the Hughes phenomenon (Hughes 1968), commonly referred as the curse of dimensionality, Adjorlolo et al. (2012a) and Adjorlolo et al. (2013) proposed a user-defined inter-band correlation filtering technique to resample hyperspectral data. Their studies indicated that wavelengths of known absorption and reflectance characteristics can improve the ability to classify the C3 and C4 grass species. In the present study, the variability in forage nutrients concentration was measured, using the wavebands of known spectral absorption features in a partial least squares (PLS) regression analysis. The PLS model was constructed to estimate variability in N, CP, moisture, and neutral detergent fibre (hereafter simply referred to as fibre), across C3 grass, *Festuca costata*, and C4 grasses, *Themeda triandra* and *Rendlia altera*. This study provides an opportunity for a deeper understanding of the variability in C3 and C4 grass nutrient composition and ecosystem functioning in the African montane rangelands.

6.2 Methodology

6.2.1 The study area

Field work was carried out in January 2011 to sample some key grass species (*F. costata*, *T. triandra* and *R. altera*) in the southern Cathedral Peak region of the Drakensberg Mountain range, South Africa. This region consists of vegetation divided into altitudinal zones, which correspond with the physiographic features of the Drakensberg montane landscape (Hill 1996). The specific study site is located at NW Lat = -28.97360039; NW Long = 29.20739937; SE Lat = -29.01429939; SE Long = 29.26700020. The study area is characterized by a high mean annual precipitation of more than 900 mm. This area experiences frequent mist, seasonal frosts and snow conditions, in the Sub-alpine (1930–2865 m.s.l [meters above mean sea level]) and Alpine

(2866–3353 m.s.l) vegetation belts. Fire management regime and topographic factors play an important role in influencing the distribution of C3 and C4 grasses, their composition and dynamics. The area under investigation experiences one controlled burning in a cycle of two years. There was significant delay in rainfall for the 2010/2011 growing season, which resulted in low growth.

Spectral reflectance measurements and green biomass sampling were confined to large and homogeneous patches of the candidate grass species. The C3 grassland included *F. costata*, which commonly co-exists with other C3 grasses such as *F. caprina* and *Merxmuellera macowanii* and, to some extent, *Pteridium aquilinum* (bracken fern). The *F. costata* community grows as large tussocks that are taller than the surrounding C4 grasslands. The *F. costata*, *F. caprina* and *M. macowanii* are coarse, unpalatable species that remain green throughout the dry winter season and are generally easily distinguishable in the field.

The C4 grasses can make up 70% or more of the composition in a patch. The *R. altera* plant is characterized by relatively narrow leaves, which tend to fold and break up into fibres when they reach maturity. There is generally low leaf production in the patches of *R. altera* grassland, manifesting in relatively low canopy density and structure. The candidate C4 grass, *T. triandra*, forms dense and tufted stands that are distinctly red-coloured for mature plants. It is also distinguishable from the surrounding grasslands on the basis of its characteristic red-coloured canopy appearance.

6.2.2 In-situ hyperspectral measurements

Grass canopy reflectance measurements were done in-situ, using an Analytical Spectral Device (ASD) FieldSpec spectrometer (FieldSpec®3, ASD, Inc., USA). The measurement range of the ASD device is 350–2500 nm, registering radiation at 1.4 nm intervals for 350–1000 nm and 2 nm for 1000–2500 nm spectral regions. The resultant spectral data are interpolated to 1 nm resolution across the spectrum.

In this study, the ASD device's fibre optic cable was set at a 25° field of view, held at about 1.5 m directly above the sampling plots to create an instantaneous field of view of about 0.35 m^2. The barium sulphate ($BaSO_4$) white reference panel was used to measure the incident radiation. The target reflectance was calculated as the ratio between the reflected energy from the grass canopy and the incident energy on the $BaSO_4$ white reference panel. The reflectance data

for the grasses in the 1×1 m plots were measured at full canopy cover. This involved randomly moving the spectrometer around three or more positions within each 1×1 m plot and five spectral measurements were consistently recorded for each one of these positions. The process resulted in a minimum of 15 reflectance measurements per plot. The reflectance data were collected to correlate with concentration of forage nutrient variables (N, CP, moisture, and fibre), sampled in 1×1 m plots represented by 80% or more *F. costata*, *T. triandra* and *R. altera* grass species. The low signal-to-noise ratio bands in both ends of the spectrum (350–359 nm, and 2480–2500 nm), as well as the strong noisy incident radiation in 1350–1460 nm and 1790–1960 nm were removed from all analysis (Thenkabail et al. 2004b).

6.2.3 Chemical analysis

Above-ground green grass biomass for each candidate grass species was sampled by clipping plants as low as possible to the base, in the sample plots, where reflectance measurements were performed. The green grass samples were dried at 70°C for 48 hours in an oven. The dried samples were weighed to determine the percentage moisture content. Next, samples were mill-crushed to 1 mm, using facilities at the KwaZulu-Natal Department of Agriculture and Environmental Affairs Feed laboratory, Cedara, South Africa. The dried and crushed samples were processed for complete feed analyses, calculated on a 100% dry-matter (DM) basis, for mid-summer plants.

The Kjeldahl procedure was used for the determination of N concentration and other biochemicals. Prior digestion of the samples was done in a mixture of sulphuric acid, selenium and salicylic acid (Novozamsky et al. 1983). These samples were then analyzed using a colorimetric, continuous flow analyzer (SKALAR SAN plus) instrument. Crude protein was determined as a measure of the nitrogen content of each sample, including both true protein and non-protein nitrogen: N x 6.25 (average 16% nitrogen), on the basis of 100% DM/100g (Conklin-Brittain et al. 1999; Douglas J. Levey et al. 2000).

The determination of fibres, such as, neutral detergent fibre (NDF) and acid detergent fibre (ADF), was done according to the ANKOM (http://www.ankom.com/product/ankom-2000-automated-fiber-analyzer,-120v.aspx) filter bag procedure (Hansey et al. 2010).The moisture content for plants sampled from each plot was calculated using the wet-dry weight method. The

information obtained on N, CP, moisture, and fibre was pooled for use in the statistical analysis phase.

6.2.4 Preprocessing of the hyperspectral data

A total of 44 plots were measured for each of the three categories of grass species. This resulted in a total of 132 samples, which was considered to be representative of the spectral variability within and among the candidate grass species in the landscape under investigation. The WTC (Adjorlolo et al. 2013) approach of spectral resampling is compared with spectra resampled to WorldView-2 multispectral band setting and data resampled to 10-nm band-widths, across the 400 – 2500 nm spectrum. In addition, spectra were resampled to the 13 preselected bands and smoothed, but without applying the WTC filter.

Table 6.1 shows the 13 user-defined or pre-selected wavebands (Adjorlolo et al. 2013) of known absorption and reflectance characteristic features. The varying spectral resolution around each of the user-defined band-centres was estimated, using the WTC filter (Adjorlolo et al. 2013). In the current study, the WTC r = 0.992 (WTC r =0.992) was used to resample the original, unsmoothed, ASD data. The procedure involved calculating a linear inter-band correlation on a 10 nm spectral resolution bands. The WTC technique accounts for the asymmetrical inter-band information contained between all band combinations. The spectral resolution at each band centre was then calculated by weighting the r values to the left and to the right of each of the band centre of interest, within the limits of WTC r = 0.992 (Adjorlolo et al. 2013). The spectral resampling procedure was performed using ENVI v4.8 software (RSI 2010).

Table 6.1 Wavelengths of known absorption and reflectance properties, which are highly sensitive to the structural and biochemical characteristic features of vegetation.

No.	Band Centre(nm)	Bandwidth	Known Causal compound/ feature	Source
1	470	430 - 500	Total plant pigment concentration	Blackburn 1998
2	530	520 - 570	Chlorophyll a absorption	Gamon et al. 1997
3	600	580 - 630	Nitrogen	Fourty and Baret 1997)
4	660	650 - 690	Nitrogen	Carter 1994
5	700	700*	Total Chlorophyll, Nitrogen	Carter 1994
6	720	720*	Total Chlorophyll, Leaf mass	Horler et al. 1983
7	820	740 - 1110	Leaf mass, Leaf area index	Carter 1994
8	1540	1500 - 1630	Cellulose, vegetation water content	Carter 1994
9	2060	2040 - 2090	Protein, Nitrogen	Carter 1994
10	2280	2100 - 2300	Cellulose, Sugar, Starch, Leaf mass	Curran, 1994
11	2300	2300 - 2350	Leaf mass, vegetation water content	Carter 1994
12	2450	2450*	Cellulose, Protein, Nitrogen	Carter 1994
13	2470	2470*	Cellulose, Protein	Kumar et al. 2001

* Indicates wavelength = 10 nm of the input spectral band.

6.2.5 Data analysis

This study assesses whether reflectance differences due to variability in forage nutrient concentrations among the three grass species can be predicted. Thus, it was tested whether the reflectance differences in the means of forage nutrients among the C3 and C4 grass species were significant. This was statistically tested using the one-way analysis of variance (ANOVA). From those tests, it was possible to establish whether there were differences in N, CP, moisture, and fibre concentrations between the C3 and C4 grass species under investigation. The statistical tests also included correlating the various forage nutrient values with the spectra measured for all species combined (n = 132). This was done in order to assess the strength of the relationship between the nutrient variables and reflectance spectra, across the candidate grass canopies that were assessed.

6.2.6 Partial least squares regression and model accuracy

PLS is suitable for analyzing multi-collinear spectral datasets (Richter et al. 2012a). Increasingly, the remote sensing applications that have used the PLS model have done so by investigating its potential in estimating vegetation biochemical properties (Darvishzadeh et al. 2008; Hansen and

Schjoerring 2003) and biophysical properties (Cho et al. 2007; Darvishzadeh et al. 2008). The PLS algorithm uses data compression capability to mainly reduce the number of input collinear spectral variables to a few non-correlated latent variables (NLV) or principal components (Bastien 2005; Liu and Rayens 2007). The component projection helps to identify the relevant structural information contained in the reflectance values of the input wavebands. The NLVs explain the variation in both the predictors and response variables. This makes PLS model desirable, given that the information content of remote sensing data can be highly correlated (Cho et al. 2007). The algorithm projects the predictor (X) and response (Y) variables as:

$$\mathbf{X = TP' + E}$$
$$\mathbf{Y = UQ' + F} \quad (1)$$

Where X = the predictors matrix; Y = the response matrix; T = X-scores; U = Y-scores; P = X-loadings; Q = Y-loadings; E = X-residuals; and F = Y-residuals (Ye et al. 2008). An absolute coefficient (βw) indicates the importance of the input spectral variable, X and thereby identifies wavelengths with the most informative data. The βw matrix is calculated directly from the PLS' factor loadings corresponding to the model with the optimum NLV, according to the following equation:

$$\beta w = \mathbf{W} \times (\mathbf{P}^{\mathrm{T}} \times \mathbf{W})^{-1} \times \mathbf{Q}^{\mathrm{T}} \quad (2)$$

W is the X-weight loading matrix, P is the X loading matrix and Q is the Y loading matrix.
The algorithm searches for NLV in the spectral space that explains variability in both the input spectral data, X and the dependent variable, Y. A more elaborate description of the PLS theory and computational approach can be found in Bastien (2005) and Liu and Rayens (2007).

The minimum value of the root mean squared error from cross validation (RMSECV) criterion was used to select the optimum NLV (Cho et al. 2007) . The RMSECV was calculated from leave-one-out (LOO) cross validation as:

$$\mathrm{RMSECV} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i^c - y_i^c)^2}{n}} \quad (3)$$

whereby $\hat{y}_i^c$ and y_i^c denote the LOO cross-validation of the predicted and measured nutrient values, respectively, and n is the number of observations. The RMSECV provides a direct estimate of the model error expressed in the original measurement units.

In the present study, the PLS regression was performed on the WTC-derived dataset for comparison with spectra resampled to WorldView-2 band settings and spectra resampled to 10 nm wide band-widths across 400-2500 nm spectrum. The data were divided into 70% training (n = 31 and 30% test (n = 13) for each species (Duro et al. 2012). The samples were then pooled for all species to create the all-combined-dataset. PLS regression was performed using the training dataset and v-fold cross validation was used to select the optimal number of components and was repeated 10 fold. This was repeated for all possible permutations of the training and testing segments in order to minimize the total prediction error in identifying the significant components (Cho et al. 2007; Hoskuldsson 2003).

The PLS variable importance (VIP) scores were used to assess the predictive ability of each input spectral band, for all forage nutrient variables. The VIP scores of the input spectral bands are a summary of the importance of the projections to find n latent variables(Pullanagari et al. 2012). The magnitude of each input spectral band (x) in modelling a response nutrient (y) was calculated using PLS-regression weighted coefficients, which were represented by the VIP-score for the projection (Wold et al. 2001). A larger score indicates the corresponding spectral band had greater importance in building a model that predicts the y-response variable while the band having a lower score contributed less in developing a model. Further explanations and formula of the VIP scores calculation can be found in Pullanagari et al. (2012).The analysis was carried out using STATISTICA 8 software (StatSoft, 2007).

The predictive performance of the PLS models were estimated with the independent test data using the coefficient of determination (r^2) and the root mean square error of prediction (RMSEP). The RMSEP calculation is analogous with Equation 3; whereby $\hat{y}_i^c$ and y_i^c denote the predicted and measured forage nutrient values, respectively. The PLS model was chosen for the current study because it has been shown to be useful in a number of case studies (Cho et al. 2007; Hansen and Schjoerring 2003; Kawamura et al. 2008; Kruse et al. 2006; Thulin et al. 2012). All data processing and regression analyses were performed using Statistica software (Statistica 8.0, Statsoft, Inc, Tulsa, USA).

In order to evaluate the robustness of the developed PLS models' goodness-of-fit in predicting the four nutrient variables, the Nash-Sutcliffe efficiency (NSE) index (Nash and Sutcliffe 1970) was used. This was done to quantify the developed models' contribution within a class. NSE, dimensionless in index, gives improved information on the model performance,

compared to statistical measures of r^2 or RMSE (CV) (Richter et al. 2012a). Use of the NSE index in the current experiment provides an additional information (Richter et al. 2012b) on the PLS models' performance in detecting each species and their linked concentration of nutrients: N, CP, moisture, and fibre. The NSE index was calculated according to Richter et al. (2012b):

$$NSE = 1 - \frac{\sum_{i=1}^{n} (V_{obs}^{i} - V_{est}^{i})^2}{\sum_{i=1}^{n} (V_{obs}^{i} - \overline{V}_{obs})^2} \quad (4)$$

Where V_{obs}^{i} is the measured nutrient *i* and V_{est}^{i} the predicted value. $\overline{V}_{obs}$ is the mean value of all measured nutrients.

6.3 Results

6.3.1 Variation in forage nutrient values

Table 6.2 shows mean, minimum, maximum, standard deviation (SD) and coefficients of variation (CV) for N, CP, moisture, and fibre, calculated for each category of the target species (n = 44), including the pooled or the all-combined datasets (n = 132). Overall, the means of N and CP concentration were largest (1.53 % and 9.32%, respectively) for C4, *R. altera*. Moisture and fibre concentration were highest (64.61% and 45.58%, respectively) for *F. costata*.

One-way ANOVA was used to test if variations in the means of N, CP, moisture, and fibre concentrations between the species were significant. The results obtained, using the 'Scheffe probabilities for post hoc tests', indicated that there are significant differences ($P < 0.001$) in N and CP concentration among all species. On the other hand, it was established from these tests that differences in the means for moisture and fibre concentrations were not significant ($P < 0.05$) between *T. triandra* and *R. altera* samples. Notably, the CV values for all nutrients variables across the three species are low, indicating small variation in the concentration of nutrients within the classes.

Table 6.2 Descriptive statistics for the C3 and C4 grass species sampled in the Cathedral Peak study area. Measurements were on 100% DM basis.

Dependent variables	No. of samples	Mean	Minimum	Maximum	SD	CV (%)
Nitrogen (%)						
F. costata	44	0.74	0.46	1.07	0.11	14.86
T. triandra	44	1.19	0.74	1.53	0.24	20.17
R. altera	44	1.53	0.96	1.89	0.29	18.95
All combined	132	1.15	0.46	1.89	0.39	33.91
Crude protein (%)						
F. costata	44	5.8	2.74	7.04	1.2	20.69
T. triandra	44	8.43	5.16	12.7	1.1	13.05
R. altera	44	9.32	5.93	10.28	0.86	9.23
All combined	132	7.85	2.74	12.7	1.83	23.31
Moisture (%)						
F. costata	44	64.61	43.48	68.72	4.18	6.47
T. triandra	44	61.98	52.47	68.59	2.94	4.74
R. altera	44	52	27.51	67.84	10.95	21.06
All combined	132	59.5	27.51	68.72	8.83	14.86
Fibre (%)						
F. costata	44	45.58	35.25	55.24	3.86	8.47
T. triandra	44	41.44	33.52	45.15	2.16	5.21
R. altera	44	40.75	39.11	45.22	1.28	3.14
All combined	132	42.59	33.52	55.24	3.4	7.98

6.3.2 Correlation between nutrient variables and reflectance data across species

Figure 6.1 shows the mean reflectance spectra for *F. costata*, *T. triandra* and *R. altera* grass canopies, using spectral data resampled to the different band setting or configurations. In general, the spectral response curves were higher for *T. triandra* at all wavelengths, compared to those of *R. altera* and *F. costata* in all cases. One-way ANOVA analysis showed that differences in the mean reflectance values between species were significant ($P < 0.001$).

The 13 wavebands of known reflectance characteristic features and the resampled WorldVied-2 were statistically analyzed to assess the strength of their relation with each of the forage nutrient variables, across all the species. Although all the bands for the 10-nm resampled-full-spectrum data (n= 210) were analyzed, only the top 10-ranked bands selected through the PLS model are presented in Table 6.2, to illustrate PLS' variable importance. All the PLS-selected bands significantly correlate ($P < 0.05$) with the response variables. Notably, the coefficients of inter-correlation are highest between fibre and CP, which also showed an inverse trend around the near-infrared region of the spectrum.

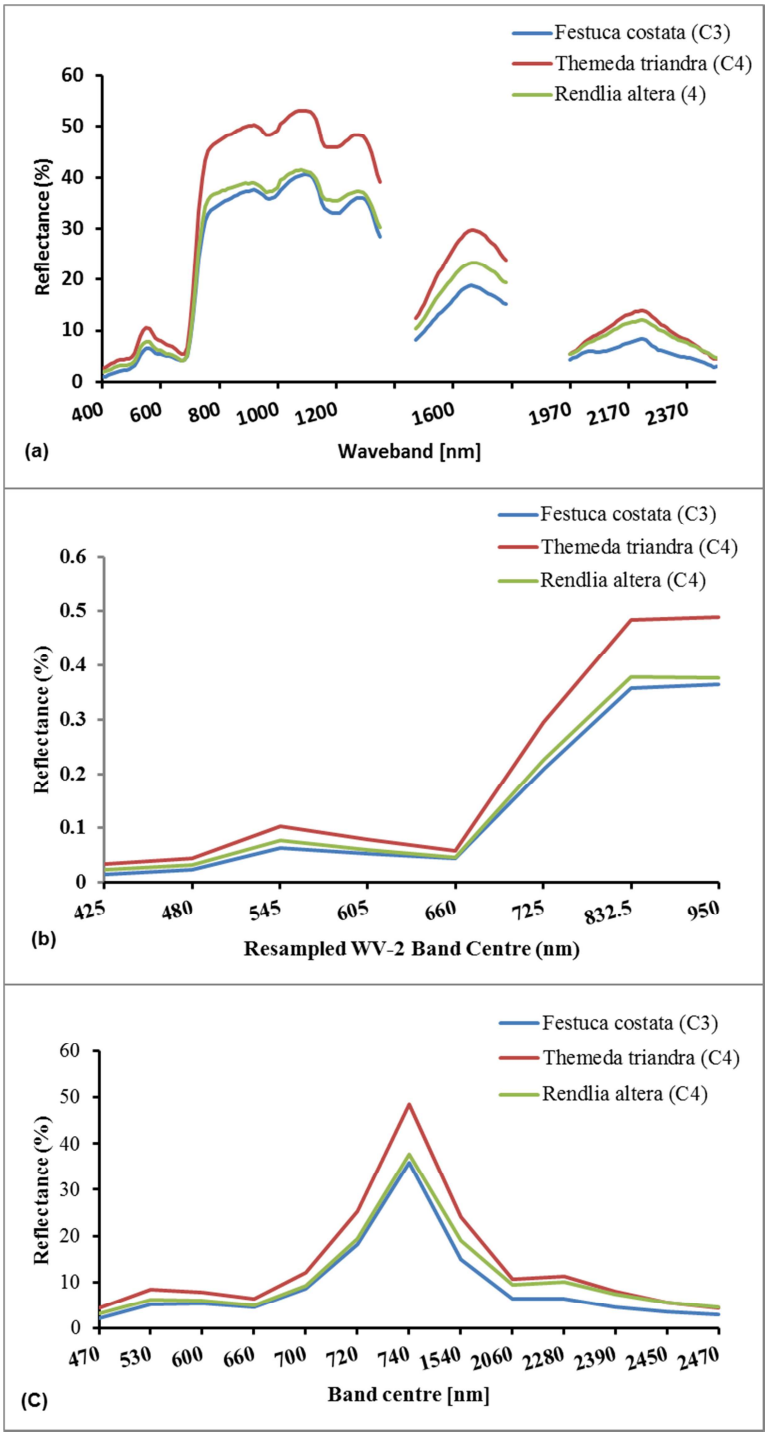

Figure 6.1 Average spectral response curves of the dominant grass canopies (n = 44 for each representative species) measured: (a) shows the original ASD data resampled to 10-nm-wide spectral resolution, (b) the spectral curves for resample WorldView-2 multispectral sensor and (c) illustrates the spectral response curves, based on the 13 band centres of known absorption and reflectance features, using the user-defined inter-band correlation filter technique of resampling hyperspectral data (Adjorlolo et al. 2013). Further detail about the size of the filter and the spectral resolutions are shown in Table 6.1. Noisy wavelengths at 1350–1460 nm and 1790– 1960 nm, due to atmospheric water absorption, were removed from all analysis.

Table 6.2 Correlation coefficients (r2) for the resampled WV-2 bands, WTC-resampled bands and PLS-selected to 10-ranked features, across the three species (F. costata, T. triandra and R. altera)

Variable	Moisture	Fibre	CP	N
Inter-correlation				
Moisture	1	0.56	0.53	0.01
Fibre	0.56	1	0.63	0.42
CP	0.53	0.63	1	0.26
N	0.01	0.42	0.26	1
Resampled WordView-2				
425	0.27	0.31	0.39	0.46
480	0.32	0.2	0.36	0.15
545	0.29	0.11	0.15	0.39
605	0.29	0.25	0.42	0.44
660	0.29	0.36	0.47	0.26
725	0.22	0.29	0.39	0.47
832	0.12	0.2	0.31	0.34
950	0.12	0.24	0.36	0.34
Resampled 13 bands with WTC filter				
470	0.28	0.26	0.39	0.46
530	0.29	0.26	0.4	0.45
600	0.31	0.28	0.45	0.49
660	0.29	0.25	0.47	0.44
700	0.29	0.26	0.47	0.46
720	0.2	0.29	0.29	0.33
820	0.2	0.27	0.31	0.34
1540	0.27	0.24	0.38	0.42
2060	0.3	0.26	0.43	0.47
2280	0.29	0.25	0.4	0.45
2300	0.32	0.28	0.45	0.5
2450	0.3	0.26	0.42	0.47
2470	0.3	0.27	0.42	0.47
Top 10-selected PLS variables				
410	0.28	0.26	0.39	0.46
430	0.29	0.26	0.4	0.45
660	0.27	0.24	0.38	0.42
710	0.29	0.25	0.47	0.44
810	0.29	0.25	0.4	0.45
1060	0.2	0.29	0.29	0.33
1560	0.31	0.28	0.45	0.49
2030	0.29	0.26	0.47	0.46
2300	0.32	0.26	0.43	0.47
2310	0.32	0.27	0.31	0.34

6.3.3 Performance of the partial least squares models

PLS models constructed based on the 13 preselected band-centres with WTC filtering yielded similar or better prediction errors and accuracies for each response variable, compared to other datasets assessed, across all species. Figures 6.3 and 6.4 show scatter plots for the measured and predicted concentration of the nutrient values obtained, using the WTC dataset. Noticeably, the within species prediction models were weaker than the combined data sets as reflected by the low or negative NSE index values obtained (Table 6.3).

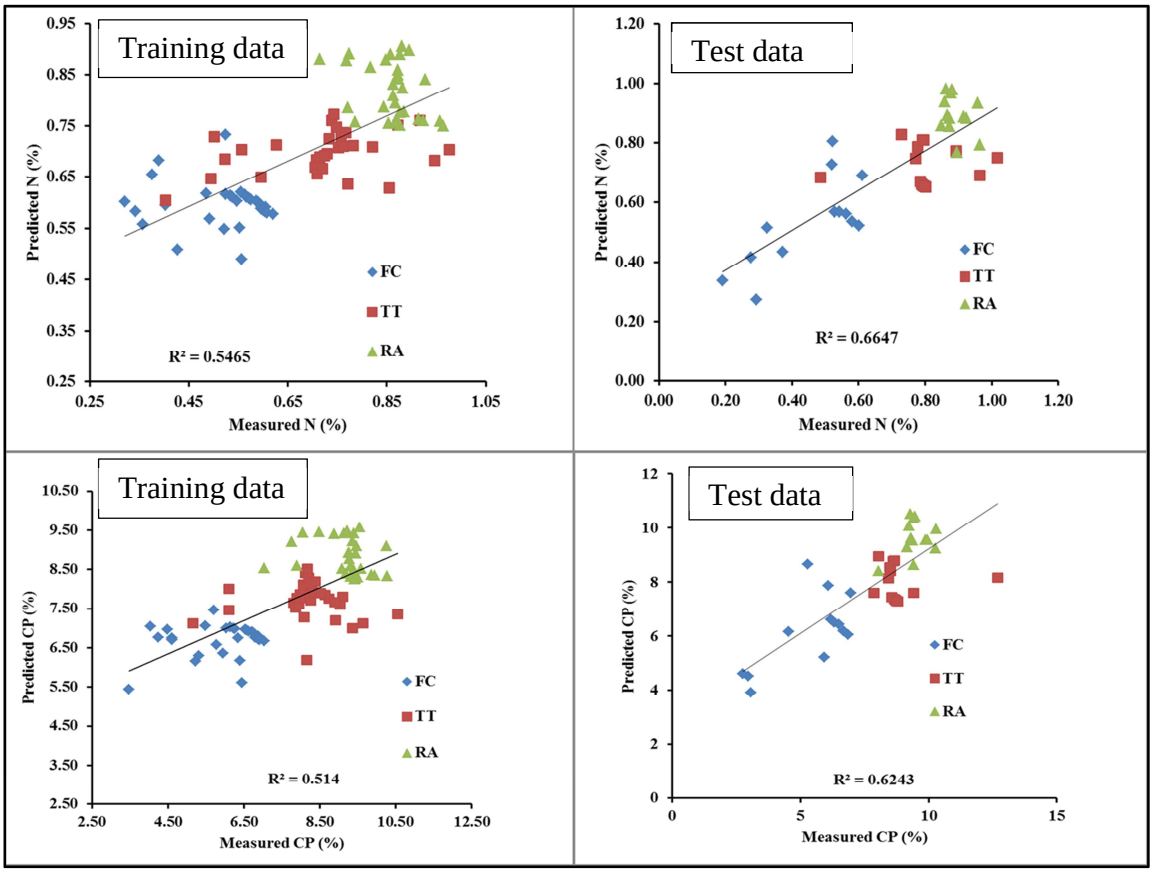

Figure 6.3 PLS regression showing observed and predicted (N and CP) values for the training and test datasets.

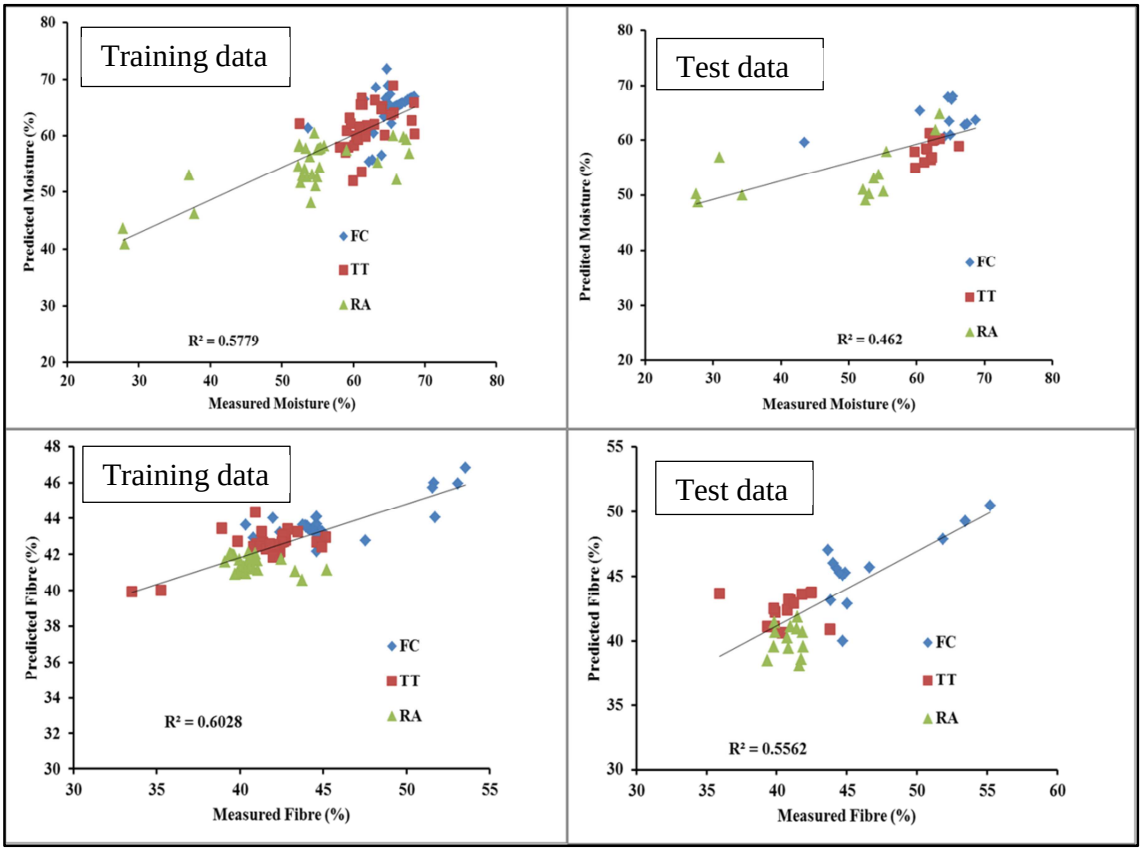

Figure 6.4 PLS regression, showing observed and predicted (Moisture and Fibre) values, for the training and test datasets.

Table 6.3 Percentage variation accounted for by PLS models. All samples (n =44) were used to construct PLS models for individual species, in the case of the WTC dataset and thus, testing of models with independent observations was not implemented. Within class analysis was not implemented for resampled-WV-2, resampled 10-nm spectral datasets and resampled 13 bands (without WTC filter).

		Training Dataset (n =92)					Test Dataset (n =40)		
Dataset	Species	Response Class	PLS Factors (CV)	R^2 (CV)	RMSECV (%)	NSE	RMSECV (%)	NSE	R^2
(A) Resampled 13 bands with WTC filter	*F. costata*	Nitrogen	4	0.32	0.9	0.48	-	-	-
		CP	4	0.21	1.04	0.21	-	-	-
		Moisture	4	0.03	15.6	-13.27	-	-	-
		Fibre	4	0.38	2.80	0.38	-	-	-
	T. triandra	Nitrogen	3	0.32	0.11	0.28	-	-	-
		CP	3	0.23	0.97	0.31	-	-	-
		Moisture	3	0.29	0.24	0.29	-	-	-
		Fibre	3	0.05	2.17	0.04	-	-	-
	R. altera	Nitrogen	4	0.51	0.04	0.47	-	-	-
		CP	4	0.52	0.48	0.55	-	-	-
		Moisture	4	0.19	14.99	-0.90	-	-	-
		Fibre	4	0.35	0.97	0.34	-	-	-
	All combined	Nitrogen	3	0.55	0.12	0.53	0.12	0.66	0.66
		CP	3	0.51	1.16	0.50	1.29	0.62	0.62
		Moisture	3	0.58	5.05	0.58	7.99	0.42	0.46
		Fibre	3	0.60	2.3	0.45	2.4	0.56	0.56
(B) Resampled 10-nm for 13 bands, without WTC	All combined	Nitrogen	3	0.50	0.18	0.50	0.18	0.60	0.59
		CP	3	0.47	0.15	0.46	0.17	0.46	0.60
		Moisture	3	0.28	6.2	0.45	9.00	0.32	0.32
		Fibre	3	0.53	3.1	0.46	3.2	0.50	0.56
(C) Resampled WV-2	All combined	Nitrogen	2	0.52	0.16	0.50	0.14	0.56	0.56
		CP	2	0.50	1.12	0.50	1.30	0.60	0.61
		Moisture	2	0.43	7.2	0.38	8.05	0.28	0.41
		Fibre	2	0.64	1.86	0.51	1.60	0.58	0.52
(D) Resampled 10-nm spectra	All combined	Nitrogen	2	0.38	0.19	0.39	0.19	0.50	0.50
		CP	2	0.49	1.80	0.44	1.62	0.40	0.51
		Moisture	2	0.43	7.1	0.35	7.9	0.46	0.50
		Fibre	2	0.53	2.11	0.54	1.89	0.59	0.60

Table 6.4 An illustration of the PLS X-loading for the three component projections used to analyse the WTC-bands

Band	Component 1	Component 2	Component 3
470	0.25	0.32	0.92
530	0.28	0.07	-0.22
600	0.29	0.08	-0.02
660	0.28	0.08	-0.26
700	0.29	0.05	-0.19
720	0.27	-0.65	0.20
820	0.26	-0.65	0.44
1540	0.28	-0.14	-0.10
2060	0.29	0.00	-0.04
2280	0.29	-0.03	-0.13
2300	0.29	0.13	-0.07
2450	0.28	0.11	-0.16
2470	0.28	0.16	-0.23

6.3.4 Variable importance

For each of the forage nutrient variables predicted, the PLS algorithm reported VIP scores for the input spectral bands. Figure 6.5 illustrates the PLS' VIP scores for each of the forage nutrient variables assessed, using the WTC dataset. It is clear from the PLS' VIP scores that the various forage nutrients differ in sensitivity, with reference to specific regions of the spectrum.

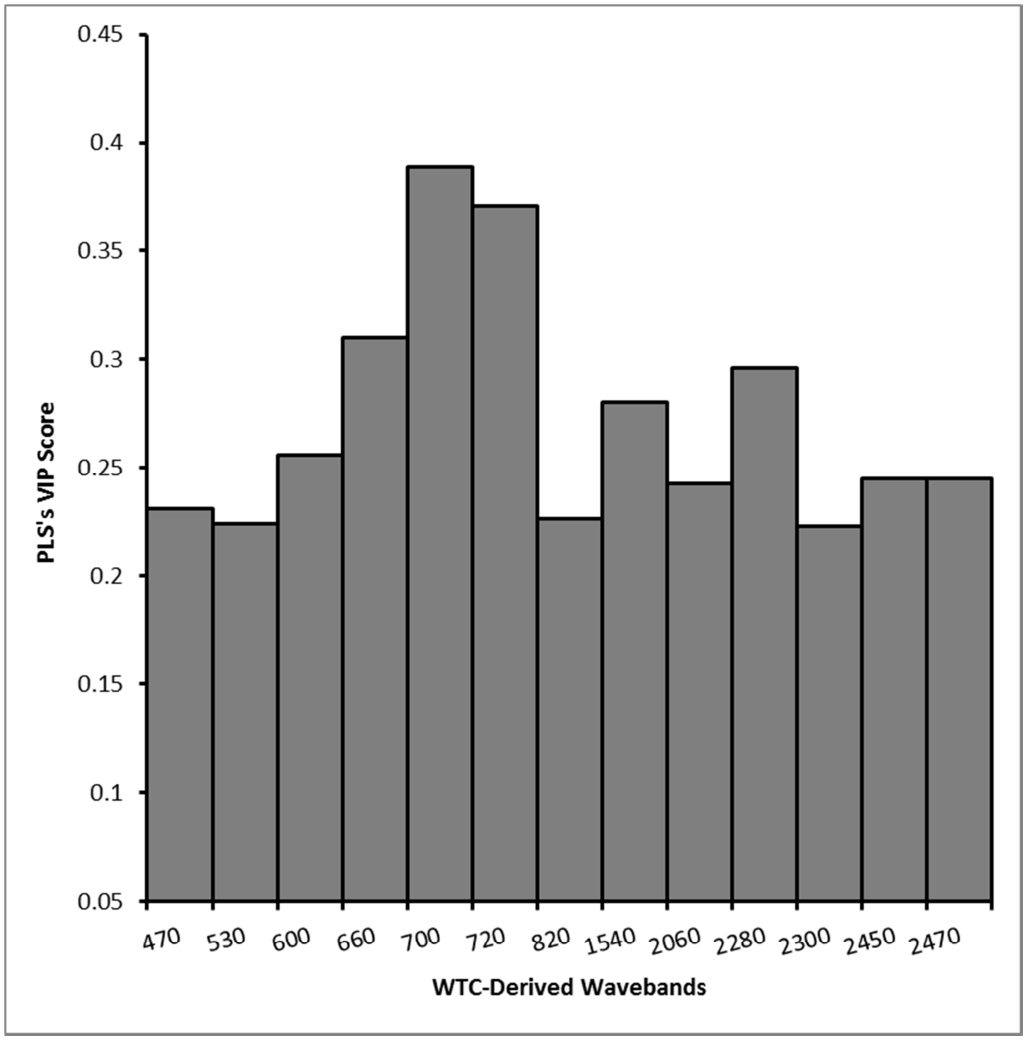

Figure 6.5 illustration of VIP scores obtained from PLS regression analysis, using the WTC-derived dataset.

6.4 Discussion

Recent studies have shown that there are clear patterns in the distribution of forage nutrient concentrations in the African rangelands (Knox et al. 2011; Ramoelo et al. 2011). In general, biophysical, physiological and biochemical characteristic features of rangeland vegetation have been linked to the variability of spectral reflectance (Knox et al. 2012; Knox et al. 2010; Mutanga et al. 2012; Mutanga et al. 2004; Serrano et al. 2000; Thenkabail et al. 2000). In the current study, variability in N, CP, moisture, and fibre concentrations was estimated, using in-situ reflectance measurements for the C3 and C4 type of grasses in the PLS modeling environment.

The effectiveness of the spectral variables for predicting variability in the respective forage nutrients concentration relied on the wavebands of known spectral reflectance or absorption features in the spectrum assessed. This study has demonstrated the unceasing importance of in-situ remote sensing data to assess the variability of forage nutrient concentrations across C3 and C4 grasses. There were species type-independent characteristic features that added extra variation to the spectral features. These additional variations in spectral features can be attributed to differences in the C3 and C4 grass fractional coverage, phenology and amount of standing biomass, and structural multi-scattering by mesophyll cells. The observed variation in N, CP, moisture, and fibre concentrations among the grasses, reflected differences of known spectral characteristic features between C3 and C4 grass species (Adjorlolo et al. 2012b).

The integration of the PLS and the optimized spectral resolutions added another useful dimension to yield high prediction accuracies that are consistent with those of other studies that predicted biochemical concentrations in rangelands (Cho et al. 2007; Hansen and Schjoerring 2003; Kawamura et al. 2008). For example, Kawamura et al. (2008) reported that optimized wavebands calculated from in-situ canopy reflectance spectra improved the predictive accuracy of PLS models for above-ground biomass, CP, ADF and NDF concentrations in cover crops.

The current study has also demonstrated that variability in N, CP, moisture, and fiber concentrations across C3 and C4 grasses can be determined using remote sensing. It showed that PLS regression analysis of the WTC wavebands (Table 6.1) yield good prediction accuracies compared to those obtained using traditional methods of resampling spectra. Studies such as this should provide valuable information for rangeland ecologists to better understand the biochemical composition in C3 and C4 grassland environment.

6.4.1 PLS model performance and important spectral bands

To predict variability in forage nutrient concentration across the groups of C3 and C4 species, this study chose wavebands that should be species-independent. It took into account the species-specific differences in reflectance due to biochemical absorption, internal structure and canopy scattering effects, which generally differ between C3 and C4 grasses (Peterson et al. 2002; Ustin and Gamon 2010). However, validation of the PLS performance within each class indicated that the PLS models contributed marginally in predicting nutrient concentrations in the study area.

This can be confirmed from the low or insignificant NSE values obtained, as reported by Richter et al. (2012). In their study, the authors noted that a prediction model should yield an NSE ≥ 0.9 and $0.5 \leq$ NSE ≤ 0.8 values to be considered excellent or good, respectively. Nonetheless, the results obtained from this study reaffirmed the capability of spectral data for predicting plant nutrient characteristics even in complex rangelands (Beeri et al. 2007; Blackburn 1998; Mutanga et al. 2004; Schmidt and Skidmore 2001; Slaton et al. 2001). These studies have invariably recommended specific spectral regions that are sensitive to different forage nutrient parameters. For example, Kawamura et al. (2008) found that narrow reflectance (< 10 nm) at 720 nm, 2280 nm, and 2300 nm have the strongest relation to canopy fibres, such as, neutral detergent fibre (NDF) and acid detergent fibre (ADF), and CP concentrations in forage, respectively.

The current study innovatively constructed spectral resolutions around some wavelengths with known vegetation reflectance effects, to predict variability in forage nutrient concentrations and across key C3 and C4 grasses in the area. The results obtained indicated that such an approach has the potential to not only reduce spectral dimensionality based on specific wavelengths of interest but also provided opportunity to directly or indirectly link nutrient variables to known spectral regions. For example, spectral information in the visible blue/green (430–570 nm) and red-edge (700–720 nm) regions have higher predictive power for foliar chlorophylls that are in turn highly correlated with macro nutrients such as N and CP concentrations. This study showed that variation in N concentration could be predicted using bands in the shorter wavelengths of the red edge (over 35% of the PLS' VIP score were obtained using only the 700 nm waveband, across all the target species). Several researchers have also found that the red-edge position is strongly related to chlorophyll concentration and this has been used to estimate N concentration in grass species (Knox et al. 2010; Mutanga et al. 2004; Ramoelo et al. 2011). In the current study, CP, which is closely related to N, also yielded strong relationship with the reflectance at the red-edge position (720 nm waveband). As expected, the most highly-ranked waveband for predicting CP was in the short-wave infra-red region, centered at 2280 nm. This can be attributed to the absorption due to C-H, N-H and O-H bonds (Curran 1994). Also, the waveband centered at 820 nm, which is sensitive to plant's internal structure (Slaton et al. 2001), was found to be useful for predicting variability in fiber concentration.

6.5 Conclusion

The current study assessed wavebands obtained through a user-defined inter-band correlation filter technique of resampling hyperspectral data. The adopted spectral resampling technique was useful for optimizing the spectral resolutions. The resultant wavebands were then integrated with the PLS regression model to predict the variability in N, CP, moisture, and fibre concentrations, across key C3 and C4 type of grasses in the Drakensberg montane landscape. Results from the PLS analysis showed that variability in forage nutrient concentration can be rapidly and non-destructively predicted at acceptable accuracies. Using the wavebands of known reflectance effects with the PLS regression analysis, variability in N, CP, moisture, and fibre concentration could be predicted better across all species as compared to predicting individual species. WTC approach to resampling can yield similar or even better results with less dimensional and uncorrelated data.

This study concludes that the integration of PLS algorithm and wavebands obtained through the user-defined inter-band correlation filter technique is promising for the assessment of a wide range of vegetation nutrient parameters. However, further studies are needed to establish the variation of forage nutrient concentrations in the C3 and C4 grasses, focusing on various growth stages, a variety of environments and the level of canopy.

In chapters 4 and 5 it has been demonstrated that the new generation satellite, WorldView-2 sensor, contains spectral information useful for the separation of C3 *and* C4 *grasses or communities, in the montane environment. The experimental results reported in Chapter 6 have also shown that there is potential for the estimation of the variability of forage nutrients concentration across* C3 *and* C4 *grasslands. Thus, the objective of the following Chapter 7 is to evaluate the capability of Worldview-2 image data in mapping the variability in canopy concentration of N and fibre variables across* C3 *and* C4 *grasslands in the Cathedral Peak study area. The following chapter also evaluates, whether the integration of WorldView-2 bands, calculated vegetation indices and the wildly used partial least squares regression, or random forest regression modeling, would improve the estimation accuracies.*

CHAPTER 7: PREDICTING CANOPY NITROGEN AND FIBRE VARIABILITY ACROSS C3 AND C4 GRASSLANDS

The chapter is based on

Adjorlolo, C., Mutanga, O., & Cho, A.M. (in preparation). Predicting canopy nitrogen and fibre variability across C3 and C4 grasslands, using new generation multispectral imagery. *Remote Sensing of Environment*.

Abstract

This paper assesses the potential of WorldView-2 (WV2) satellite imagery to estimate and map the spatial distribution of nitrogen (N) and fibre concentration in grasslands. The image was acquired for the Cathedral Peak region of the Drakensberg Mountain range, South Africa. The partial least squares (PLS) and random forest (RF) regression algorithms were implemented on the 8-multispectral bands, as well as normalized difference vegetation indices (NDVIs) developed from the WV2 imagery. The PLS accounted for 61% and 63% of the variation and the RF explained 71% and 66% of the variation in N and fibre concentration, respectively. Both the PLS and RF variable importance scores showed that the red-edge and near-infrared wavebands are the most important when predicting N and fibre concentrations across C3 grass, *Festuca costata* and C4 grasses, *Themeda triandra* and *Rendlia altera*. Differences in root-mean-square error prediction (≤ 0.58%/100g of dry matter) for the PLS and RF models were marginal. Thus, only the RF, which was available in the EnMAP-Box software, was implemented to map the spatial distribution of the N and fibre concentrations in the area. Overall, results from this study demonstrated that the WV2 image data contain useful information to achieve improved N and fibre mapping, across landscapes composed of the C3 and C4 grass species.

Keywords: Grassland nutrient content; worldview-2; partial least squares; random forests; remote sensing

7.1 Introduction

There is accumulated evidence that global change (e.g., elevated CO_2 and temperature) impacts on growth response of C3 and C4 grasses, specifically, in nutrient-limited ecosystems (Barbehenn et al. 2004; Bremond et al. 2012; Niu et al. 2008; Sage and Pearcy 1987; Wand et al. 1999; Winslow et al. 2003). The C3 or C4 denotes an initial photosynthetic product of 3-carbon and 4-carbon compounds, respectively (Ehleringer and Björkman 1977). These two groups differ, with the C3 grass type having a lower compartment of mesophyll cells, a lower proportion of vascular tissues and a higher inter-venal distance, than those of C4 grasses (Ogle 2003). Intercellular air-moisture and nitrogen (N) concentrations are also relatively higher in C3, compared to C4 grasses (Ghannoum 2009). Therefore, any shifts in composition between C3 and C4 grasses will impact on the condition of rangelands or their carrying capacity, depending upon the species involved (Winslow et al. 2003). The range condition refers to the state of the vegetation in relation to some functional characteristic features (e.g. nutrient cycling), of sustained forage status and resistance to degradation (Trollope 1990; Vetter 2005). Important variables in this context are limiting nutrients such as N, acid detergent fibre (ADF) or neutral detergent fibre (NDF) (hereafter broadly referred to as fibre), which constrains the intake rates of herbivores. The measurement of N and fibre provides critical information for assessing the quality of rangelands and the dynamics of the ecosystem functioning, such as forage nutrient-herbivores interactions (Grant et al. 2010; Knox et al. 2011; Ramoelo et al. 2012; Treydte et al. 2009). These biophysical and biochemical characteristics manifest in the canopy spectral response properties of vegetation (Curran 1994).

Different remote sensing data have been used to estimate several canopy variables related to biophysical and biochemical characteristics (Blackburn 1998, 2007a; Blackburn and Steele 1999; Carter et al. 2005; Mutanga et al. 2004; Peñuelas et al. 1993; Serrano et al. 2000; Thenkabail et al. 2004a; Thenkabail et al. 2000). For example, the relationship between forage nutrient content (e.g. N and fibre) and hyperspectral data has been established with demonstrated success (Beeri et al. 2007; Fava et al. 2009; Hansen and Schjoerring 2003; Knox et al. 2010; Kokaly et al. 2009; Kokaly and Clark 1999; Kruse et al. 2006; Mutanga et al. 2004; Rama Rao et al. 2008; Serrano et al. 2002; Ullah et al. 2012). Hyperspectral remote sensing systems acquire reflectance data in hundreds of narrow (< 10 nm) contiguous spectral channels over the visible, near-infrared and shortwave-infrared portions of the electromagnetic spectrum (400–2500 nm).

The major challenge is that, although hyperspectral systems provide a continuous spectrum (radiance) for each pixel of the scene, the data are high in dimensionality, not easily accessible and can be relatively costly, compared to the traditional broadband (>100 nm) multispectral data. Even more challenging is to simultaneously estimate, using remote sensing, a combination of canopy concentration of nitrogen and fibre (Knox et al. 2011; Skidmore et al. 2010), and use this information to explain the variation in forage quality, across nutrient limited grassland ecosystems with C3 and C4 composition.

The advent of new generation space-borne multispectral sensors, such as DigitalGlobe's WorldView-2 (WV2), is sought to offer new opportunities for measuring vegetation characteristics, using multispectral data (Mutanga et al. 2012; Zengeya et al. 2012). The WV2 sensor consists of eight (8) multispectral channels. These include four (4) new wavebands strategically positioned at the visible blue, yellow, red-edge and near infrared (NIR) regions, and 4 traditional bands (blue, green, red and NIR), present in sensors, such as the IKONOS satellite (Pu and Landry 2012). In this study, Pu and Landry demonstrated that the spectral configuration of the WV2, particularly, the additional 4 bands offered a better capability in classifying individual plants. In another study, Cho et al. (2012) has shown that the new bands of WV2 were relatively more sensitive to the spectral response of individual plants, than that of the traditional bands, in classifying individual plant species in the canopy, in a savannah environment. On the other hand, the study by Adjorlolo et al. (2012a) demonstrated that hyperspectral data resampled to WV2 band settings could mitigate some of the problems of high dimensionality and multicollinearity in statistical analysis, with reference to C3 and C4 grass species classification. In addition, the relationship between forage nutrient content and remote sensing data has been modeled, with demonstrated success in linking forage N concentration to WV2 image data (Zengeya et al. 2012). However, it remains to be established whether the WV2 sensor provides additional information to simultaneously map canopy N and fibre contents in C3 and C3 grassland environment.

Several studies that have used remotely sensed data (i.e. hyperspectral wavebands) for estimating biochemical parameters have done so mainly, using techniques based on partial least squares regression methods (Asner and Martin 2008; Cho et al. 2007; Doughty et al. 2010, 2011; Hansen and Schjoerring 2003; Kawamura et al. 2008; Kruse et al. 2006). In relatively recent applications, advanced techniques based on machine learning methods: artificial neural networks

(Fourty and Baret 1997; Kavzoglu and Mather 2002), support vector machines (Clevers et al. 2007; Mountrakis et al. 2010) and random forests (Abdel-Rahman et al. 2009; Dye et al. 2011; Ismail and Mutanga 2010), which make no assumption of the data distribution have been exploited. However, it remains to be established whether the integration of the WV2 data and the widely-used parametric techniques, such as partial least squares (PLS) regression (Hansen and Schjoerring 2003; Martens and Naes 1989), or the relatively new, non-parametric techniques, such as the random forest (Breiman 2001a), improves the accuracy in estimating forage nutrient variables for various ecosystems, including C3 and C3 grassland environments.

The PLS have been traditionally used to combine the spectral information from hundreds of narrowband features of hyperspectral data and it has the capability for performing dimensionality reduction of the input spectral bands. On the other hand, spectral transformation of multispectral data have been commonly-used to combine the spectral information from different regions, e.g. Green Normalized Difference Vegetation Index [(NIR – green) / (NIR + green)] (Buschmann and Nagel 1993), Normalized Difference Vegetation Index [(NIR – red) / (NIR + red)] (Rouse et al. 1973), Normalized Green [Green / (NIR + red + green)], Normalized Red [Red / (NIR + red + green)] and Normalized Near Infrared [NIR / (NIR + red + green)] (Sripada et al. 2006). The current study attempts to assess the capability in estimating and mapping N and fibre concentrations across an area by combining information from raw bands and NDVIs computed from WV2 image data. In addition, this study assesses the utility of two data processing methods such as the PLS and RF regression in predicting an N and fibre concentration across a landscape with composition of C3 and C4 grass species.

7.2 Methods

7.2.1 Vegetation in the Study area

The study area is located in the Cathedral Peak World Heritage site in the KwaZulu-Natal province, South Africa (Figure1.1). The specific study area (29°14'10.347"E and 28°59'27.383"S) is underlain by relatively homogeneous Drakensberg basalt formations of the Stormberg series and is characterized by several mountain peaks. Patches of C3 and C4 type of grasslands in moderate slopes, crests and open valleys were sampled across the Montane and sub-alpine vegetation belts (1280–2865 m above mean sea level). The sub-alpine belt has been

described as *Themeda-Festuca,* sub-alpine grassland (Killick 1963) and is composed of the C3 and C4 type of grass species. It can be further divided into two grassland communities known as the *Themeda triandra* (red grass) and *Festuca costata* (tussock fescue) sub-alpine grasslands. Notably, there are strong associations of *T. triandra* and *F. costata* occurrence on warm, northerly and cool, southerly slopes, respectively (Granger and Schulze 1977; Hill 1996; Killick 1963). The area is also characterized by high rainfall (> 900 mm). Misty conditions, low temperatures, seasonal frosts, snow, fire, herbivory and topographic factors play an important role in influencing the composition of the C3 and C4 grasses in the area.

The C4 grassland communities consist mainly of the *T. triandra* and *Rendlia altera* (Toothbrush grass) species, while the C3 community consists mainly of unpalatable *F. costata* and its co-existing, coarse C3 species, *F. caprina*, *Merxmuellera macowanii* and *Pteridium aquilinum* (Bracken fern). The *F. costata*, *F. caprina* and *M. macowanii* remain evergreen throughout the dry (winter) season. They normally grow as large tussocks that are taller than the surrounding C4 grasslands. They also grow in an aggregated pattern, which makes them easily distinguishable in the field. Figure 7.1a and b characterize a site of typical *F. costata* community, showing distinct phenology of this photosynthetic type in autumn (April) and winter (late June), respectively. This should make the *F. costata* relative contributions to pixel spectra to be extracted on the basis of its evergreen trajectory, which is an area for future investigation.

The dominant C4 grassland communities can make up 70%, or more, of the composition in C4 community or a patch. The *R. altera* is characterized by relatively narrow leaves, which tend to fold and break-up into fibres in mature plants. This results in low leaf production and in a relatively low canopy density. The *T. triandra* forms dense and tufted stands that are distinctly red-coloured, when they are fully mature.

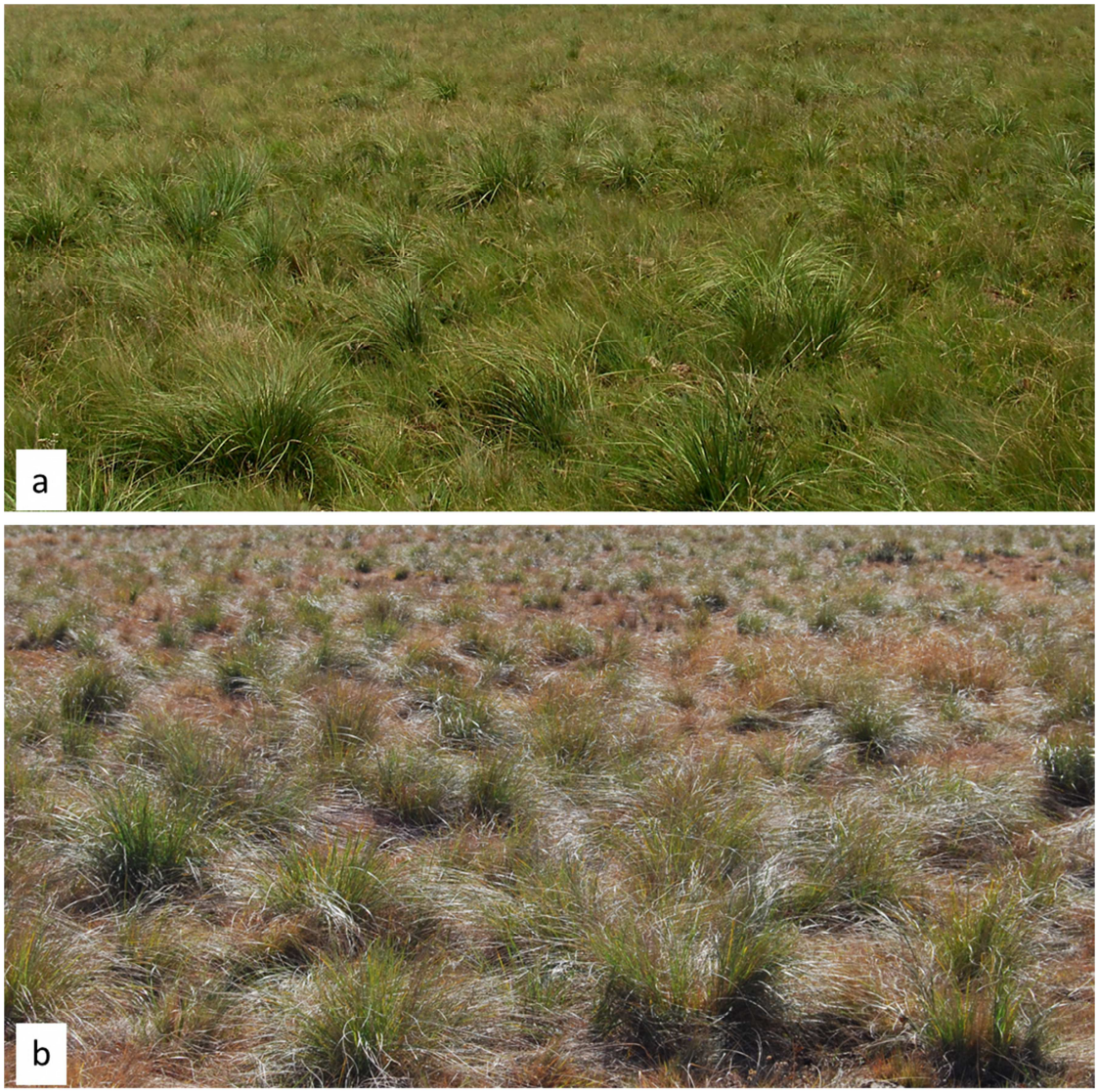

Figure 7.1 Typical field site showing the relatively evergreen, *Festuca* community in April (a) and in late June (b), after the area had frosted. In between the tussocks are C4 grasses, including the characteristic res-coloured *T. triandra*, which is clearly visible in winter (June).

7.2.2 Plant sample collection

The field campaign was conducted to coincide with the date (28 April 2011) of the WV2 image data acquisition, when all plants were at a mature state and thus full canopy cover. The stratified random sampling approach was adopted in this study, to place sample plots in each of the target grassland patches. Since there was no existing map depicting the spatial distribution of the C3 and C4 grassland in the area, the determination of the various patches was based on a preliminary classification of the major vegetation communities (Adjorlolo et al. in preparation).

A total of 64 plots were sampled for *F. costata* grassland, 49 plots for *R. altera* and 101plots for *T. triandra*. The number of samples per class was considered to be representative of the grasslands under investigation. The Trimble GPS device was used to locate the position of all sampled plots. The GPS data were differentially corrected to an accuracy of 1.89 m of the rectified WV2 image pixels.

The targets, namely the mature grass species in each of the 1×1 m sample plots, were clipped with a pair of scissors, just above the ground. All the plots sampled had 60% or more of the candidate grass species. Dry, undergrowth materials were removed from the clipped samples because these were not an important source of grazing material and that the undergrowth dry materials may not have adversely affected the spectra measured at full canopy. The remaining green material was weighed, *in situ*, and then transported in brown paper bags placed in a cooler box to the feed lab, within two to four hours of harvesting. A total of 214 plot samples were obtained.

7.2.3 Chemical analysis

The green grass samples were dried at 70°C for 48 hours in an oven and subsequently weighed. The dried samples were then mill-crushed to about 1 mm, using facilities at the KwaZulu-Natal Department of Agriculture and Environmental Affairs Feed laboratory in Cedara, South Africa. These dried and crushed samples were processed for complete feed analyses, including N and fibre concentration, calculated on a 100% dry-matter (DM) basis for mature plants. The Kjeldahl procedure was used to determine the concentration of nitrogen (Knox et al. 2011). The prior digestion of the samples was done in a mixture of sulphuric acid, selenium and salicylic acid (Novozamsky et al. 1983). The samples were then colorimetrically measured, using a continuous flow analyser (SKALAR SAN plus). The determination of fibre concentration (Neutral Detergent Fibre, NDF and Acid Detergent Fibre, ADF) was done according to the ANKOM filter bag procedure, using an ANKOM200/220 fibre analyser (ANKOM Technology, Macedon, NY, USA). The resultant data were pooled for use in the statistical analysis phase.

7.2.4 Image acquisition and pre-processing

The WV2 image was obtained for 28 April 2011. The WV2 satellite comprises eight (8) spectral channels, operating over the wavelength range 400–1040 nm and has variable spectral resolutions (Table 7.1). The spatial resolution of the 8-bands used in this study was 2 m, covering an area of 25 km^2. The data were delivered at product level LV3D (DigitalGlobe, Longmont, Colorado, USA). That means the image had been radiometrically corrected and orthorectified by the supplier. The mean in-track, cross-track and off-nadir view angles were -3.1°, 14.9° and 15.3°, respectively. The WV2 satellite's mean azimuth and elevation angles were 112.1° and 72.9°, respectively, and the sun azimuth and sun elevation angles were 32.6° and 40.8°, respectively.

Table 7.1 Spectral properties of the WorldView-2 satellite (Toutin et al. 2012)

Band no.	Band description	Spectral range (nm)	Bandwidth (nm)
B1	Coastal	400 – 450	50
B2	Blue	450 – 510	60
B3	Green	510 – 580	70
B4	Yellow	585 – 625	40
B5	Red	630 – 690	60
B6	Red-edge	705 – 745	40
B7	NearIR-1	770 – 895	125
B8	NearIR-2	860 – 1040	180

The registration mean error of the image was 0.15 m (DigitalGlobe, 2011). The data aligned well with georeferenced SPOT 5 imagery, with a geometric accuracy of < 1 m. The digital numbers were converted to WV2 radiance, using the absolute radiometric calibration factors and effective bandwidths for each band in the ENVI routine (ENVI 4.8). The resultant WV2 radiance image was then atmospherically corrected, in order to reduce haze and other atmospheric influences. ATCOR 3 module in ERDAS Imagine 2011 (Erdas, 2011) was used. This module is capable of controlling topographically varying optical effects on the input image data. Atmospheric correction functions were calculated using the MODTRAN 4 code developed for WV2 sensor. Considering that the WV2 image was acquired for a mountainous landscape, it was necessary to process the data, in order to correct for the topographic effects of the rugged or undulating terrain, and Sun geometry. Topographic corrections were calculated with the only

available 20 m digital elevation module (DEM). However, it turns out that the 20 m DEM is insufficient for effective correction of topographic effects on the WV2 dataset for the landscape under investigation. This was the case, even when the 20 m DEM was re-sampled step by step from 20 m to 10 m, 5 m, and to match the WV2 spatial resolution. The output images were inaccurate, with a substantial positional shift between the DEM and the WV2 dataset, showing a chessboard pattern of the output image. Therefore, topographic correction was excluded from the analysis, in order not to introduce an additional error to the reflectance values. In this regard, steep slopes and sheltered areas were masked from all analysis.

7.2.5 Collecting image spectra for grass plots

A 1×1 pixel (4 m^2) window corresponding to the sampled plot (1×1 m) locations was used to collect image spectra for all sample plots. The plots were located in homogenous sites of more than 4 × 4 m, given the rectification accuracy of the input image. This resulted in a total of 214 plots spectra extracted from the raw WV2 bands. In addition, spectral information was extracted from the developed normalized difference vegetation indices calculated, using ENVI 4.8 (Environment for Visualising Images, Research System, Inc.) software:

NDVI1 = (NIR-1−green)/(NIR-1+green) (Eq.1)

NDVI2 = (NIR-1−red)/(NIR-1+red) (Eq.2)

NDVI3 = (NIR-2−red-edge)/(NIR-2+red-edge) (Eq.3)

NDVI4 = (NIR-2− yellow)/(NIR-2+ yellow) (Eq.4)

NDVI5 = (red-edge− coastal blue)/(red-edge + coastal blue) (Eq.5)

NDVI6 = (red-edge −red)/(red-edge +red) (Eq.6)

These comprised combinations of the 3-traditional bands (Eq. 1 and 2) and combinations of the 4-new or additional bands (Eq.3 to 6) of WV-2. Studies by Mutanga et al. (2012) and Pu and Landry (2012) have assessed these NDVIs for other applications involving plant species analysis.

The study area under investigation comprises other vegetated surfaces such as natural forests, savanna or woodland, the fen, *Pteridium aquilinum* (Bracken fern) stands and other bush clumps (Hill 1996), for which foliar N and fibre concentration was not sampled. It was therefore necessary to exclude these vegetation cover types from all analysis in order to resolve the mixing of spectral information and related error in estimating the response variables (i.e. N and fibre

concentration). This was done by creating a point map of the sample plots developed and overlaid these onto the NDVI images. Pixels representing the dominant grasses were used to estimate the minimum and maximum NDVI threshold values for the target grassland areas. Next, the target grasslands region-of-interest map was created, using the NDVI threshold values. This was tested against the available, very high spatial resolution (50 cm) WV2 panchromatic image data. Subsequently, the region-of-interest map was used to subset the preferred grassland areas from the image, and pixels representing other vegetated surfaces and bare soils were masked out, using the mask tool in the ENVI software.

7.2.6 Data analysis approaches

The two main data analysis techniques implemented in this study included the PLS and RF regression models. All datasets were randomly split into training (n = 150) and test (n = 64) datasets. The training (calibration) set was used to construct the regression models and model validation was carried out, using the test or validation datasets. Both the PLS and RF models were implemented, using Statistica 8 software (Statsfort, 2010) to predict the variability of N and fibre concentrations, across the candidate C3 and C4 grass species. The VIP scores from the PLS and RF models were then assessed to establish the relevance of each of the input WV2 data. Several studies have suggested that it is better to use a basket of different methods for VIP measurements, rather than a single one (Chan and Paelinckx 2008; Guo et al. 2006; Melgani and Bruzzone 2004). Therefore, the PLS and RF algorithms were assessed in the current study to uphold the idea that there is possibly no universally applicable algorithm that is suitable for all problems. The performances of the PLS and the RF algorithms were assessed, based on the coefficients of determination (R^2) and on the root mean square errors cross-validation (RMSECV).

The PLS model was chosen for the current study because it has been shown to be useful in a number of case studies, in which a large number of spectral predictors have been assessed for grass biomass estimation (Cho et al. 2007; Hansen and Schjoerring 2003; Kawamura et al. 2008; Kruse et al. 2006; Thulin et al. 2012). The PLS algorithm is based on multivariate parametric technique. It is however, considered interesting, to also evaluated the capability of the non-parametric technique, RF regression model, which has been increasingly shown to have superior power in a number of classification and regression case studies (Ismail and Mutanga

2010; Lawrence et al. 2006; Mutanga et al. 2012; Prasad et al. 2006). The RF model has been validated against several statistical modelling approaches for ecological applications (Chan and Paelinckx 2008; Cutler et al. 2007).

The RF algorithm was further implemented, using the EnMAP-Box v1.3 software to spatially estimate the canopy N and fibre distributions in the study area. The EnMAP-Box is an open source license software developed, implemented and validated by the German hyperspectral Environmental Mapping and Analysing Program (EnMAP) mission at the Geomatics Lab, Humboldt University, Berlin (www.enmap.org). In the present study, the EnMAP-Box implementation of the RF model was based on the parameterisation obtained, using Statistica 8 software. That means no further optimisation was done at this phase of the analysis. The following sub-sections describe the PLS and RF regressions in the context of the present study.

7.2.6.1 Partial least squares regression and variable importance

PLS model was constructed as a bilinear calibration of the dependent N and fibre variables against the independent WV2 derived input spectral data. The PLS algorithm optionally has the capability to compress data, mainly to reduce the number of input collinear spectral variables to a few non-correlated latent variables (NLV) or factors. It uses such factors to represent the relevant structural information contained in the input data. It also uses component projection to find the number of NLV to construct the model (Bastien 2005; Liu and Rayens 2007). It assigns the measured N and fibre variables to a vector, combined with matrices of the WV2 spectral data to determine the coefficients matrix and the PLS error vector. The algorithm searches for NLV in the independent spectral space that explains variability in both the input WV2 data and the dependent variables (N and fibre concentrations). The PLS algorithm has been widely used in both the ecological and remote sensing community for several applications. It is therefore only briefly described in the current study and a more elaborate description of the PLS theory and computation can be found in Bastien (2005) and Liu and Rayens (2007).

The PLS regression was calculated on a calibration dataset, using all the WV2 bands, including the derived vegetation indices. This consisted of a 70% (n = 150) randomly split portion of the whole data. In this study, subset variable selection was not exploited, since the purpose was to evaluate the relative importance of the WV2 bands and NDVIs developed from

the WV2 data in predicting N and fibre in the C3 and C4 grassland environment. In that regard, the PLS- and RF-based variable importance (VIP) scores were assessed, to determine the predictive ability of each input WV2 8 multispectral bands, as well as the developed NDVIs for predicting N and fibre concentrations among the groups of C3 and C4 type grasses. The model with the minimum RMSECV was used to select the optimum NLV (Cho et al. 2007). RMSECV was calculated using the leave-one-out (LOO) cross-validation criterion:

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i^c - y_i^c)^2}{n}} \qquad \text{(Eq.7)}$$

whereby $\hat{y}_i^c$ and y_i^c respectively denote the LOO cross-validation predicted and measured N and fibre concentrations and *n* is the number of observations (n = 150). The RMSECV provides a direct estimate of the model error matrix, expressed in the original units of measurement. The predictive ability of the PLS regression models was assessed, using the independent test or holdout-validation dataset (n = 64). The holdout-validation datasets were analyzed, using the model parameterization for the calibration datasets to calculate the coefficients of determination (r^2) and root mean square error-prediction (RMSEP). The calculation of the RMSEP is analogous with Equation 3, with $\hat{y}_i^c$ and y_i^c, respectively, denoting the predicted and measured values for the N and fibre concentrations. In addition, the performance of the models was measured by calculating Nash–Sutcliffe efficiency (NSE). The NSE (Equation 3) index validates the model efficiency by assessing the relative magnitude of the residues compared to the variance in measured nutrients concentration (Pullanagari et al. 2012; Richter et al. 2012a) as:

$$NSE = 1 - \left(\frac{\sum_{i=1}^{n}(\hat{y} - y)^2}{\sum_{i=1}^{n}(\hat{y} - \bar{y})^2}\right) \qquad \text{(Eq.8)}$$

where ŷ indicates predicted N and fibre values and y was the measured data, ȳ represents the mean of measured values and *n* is the number of samples.

7.2.6.2 Random forest regression and variable importance

Random forests (Breiman 1996, 2001a) are ensemble classifiers or regression algorithms that utilize bootstrap samples with no replacement to grow a large number of decision trees. RF

bootstraps and sub-samples the training data (about 66% randomly and independently-selected samples) and then combines the predictions from the resulting ensemble of trees. These trees assign each WV2 band or vegetation index variables to a response (N and fibre) value, based on the maximum number of votes that the value receives from the collection of all trees (Breiman 2001a). About one-third of the data, which were not included in the bootstrapped training sample, called out-of-bag (OOB) cases, were used to evaluate the FR model. A number of researchers have shown that the OOB samples offer unbiased estimates of the training error matrix (Chan and Paelinckx 2008; Ismail and Mutanga 2011; Pal 2005; Prasad et al. 2006).

The RF algorithm is easy to implement because only few parameters are tuned: (i) the number of trees to grow (*ntree*) and (ii) the bootstrapped number of variables to split at each decision tree node (i.e. *mtry*). The default *mtry* value, in the case of RF regression analysis, is calculated as $p/3$, whereby p is the number of predictors (Breiman 2001b). In the current study, the *mtry* values were optimized using the tuning function provided in the Statistica software (Statistica 8.0, Statsoft, Inc, Tulsa, USA) implementation of random forests.

The RF algorithm offers VIP matrices in its computation, providing researchers with valuable insights to explore the effect of each predictor variable on the response variable (Grömping 2009). The VIP measure was used in this study to assess the contribution made by each of the WrldView-2 bands or vegetation indices in the resultant RF models. VIP in RF is the normalized difference prediction accuracy of each predictor variable, both when it is included as an observation and when it has been randomly permuted (Strobl et al. 2008). The permutation-based variable importance follows the rationale that the random permutation of a predictor variable represents the absence of the variable from the model. Hence, the difference in prediction accuracy prior and after permuting a variable is used to predict the response for the OOB observations and as the measure of importance (Diaz-Uriarte and Alvarez de Andres 2006). The number of observations predicted correctly, decreases substantially if the permuted variable is strongly associated with the response values (Breiman and Cutler 2004; Strobl et al. 2008). Grömping (2009) provided a more detailed account of the random forest's variable importance matrices, both from the theoretical understanding and from the perspective of computational advantages.

7.3 Results

7.3.1 Descriptive statistics

Table 7.2 shows the descriptive statistics for N and fibre concentration, for each category of the target species, as well as the whole-datasets (n = 214). The mean N concentration (1.54%) was largest for *R. altera*, while average concentration of N is 1.11% for *T. triandra*, which is the most utilizable grass species by herbivores in the area. The mean fibre concentration (64.61%) was highest for the *F. costata*, a C3 grass weed in the study area (Adjorlolo et al. 2012a; Adjorlolo et al. 2013). The standard deviations (SD) for all nutrient variables, across all categories of species, were low, compared to the sample mean, among the grass species under investigation.

Table 7.2 N and Fibre Concentrations (%) in C3 and C4 grass species sampled in the Cathedral Peak study area. Data on 100% DM basis

Dependent variables	No. of samples	Mean	Minimum	Maximum	SD	CV (%)
Nitrogen (%)						
F. costata	64	0.73	0.41	1.56	0.25	34.42
T. triandra	101	1.11	0.54	1.89	0.37	33.11
R. altera	49	1.54	0.53	1.86	0.24	15.26
All combined	214	1.09	0.41	1.89	0.43	38.87
Fibre (%)						
F. costata	64	72.25	49.35	83.44	9.33	12.91
T. triandra	101	62.85	39.72	81.65	11.78	18.74
R. altera	49	56.31	40.35	70.81	10.54	18.72
All combined	214	64.16	39.72	83.44	12.27	19.12

7.3.2 Correlation between N and fibre versus WV2 bands and indices

The WV2 bands and derived vegetation indices were statistically analysed, in order to assess the strength of their relationship with the N and fibre concentration across all the species. Table 7.3 shows the correlation coefficients of the input data. As expected, varying strengths of correlation ($P < 0.05$) were obtained for N and fibre. Overall, the WV2 band B6 yielded the highest coefficient of correlation with both the N and fibre. In general, all the normalized vegetation indices, computed, using the WV2 image yielded some improvements for the N and fibre concentration among the three species investigated. Markedly, the NDVI3 and NDVI6 showed the best improvements in the correlation for the N and fibre concentrations, respectively, compared to that of the two traditional indices (NDVI1 and NDVI2). Overall, N and fibre

showed an inverse correlation coefficient pattern across the spectrum of the input WV2 wavebands as well as the developed vegetation indices.

Table 7.3 Correlation (R^2) among canopy N and fibre concentration (n = 214), raw 8-bands of WV2 and NDVIs derived from the WV2 image.

Variable	N	Fibre
Response variables		
N	1	0.46
Fibre	0.46	1
Raw bands		
B1	0.03	0.09
B2	0.06	0.18
B3	0.21	0.48
B4	0.01	0.33
B5	0.15	0.20
B6	0.48	0.53
B7	0.35	0.47
B8	0.35	0.49
Traditional band		
NDVI1	0.36	0.50
NDVI2	0.24	0.53
Additional 4-bands		
NDVI3	0.50	0.52
NDVI4	0.29	0.54
NDVI5	0.30	0.55
NDVI6	0.41	0.55

7.3.3 Optimization of the PLS and RF regression models

The PLS' NLV and the RF's mtry and ntree parameters optimization was done using the calibration datasets (n = 150 for each variable) and with the RMSECV. The results obtained indicated that the PLS and the RF tuning parameters affected the error of prediction. For the PLS models, the optimal NLV were 4 and 5 for N and fibre, respectively, using the 10-fold cross-validation of the RMSECV. The RF models yielded optimal ntree of 5500 and 5000, and mtry of 3 and 5 for the lowest RMSECV values, respectively, for N and fibre concentrations. Differences in the optimized number of components in both the PLS and RF models, were found to be marginal. Therefore, the influence of the number of components between the two regression approaches was considered to be low.

7.3.4 Predictive performance of the PLS and the RF regression

Scatter plots for the observed versus predicted N or fibre values, as well as model accuracies for the PLS and RF regression, are shown in Figures 7.2 and 7.3, respectively. These show the cross-validated prediction for the dependent variables across the C3 grass, *F. costata* and C4 grasses, *T. triandra* and *R. altera*. Both the PLS and RF estimations exhibit a considerably larger scattering around the two extremes of the 1:1 line, but are at least of acceptable quality. This is the general pattern observed in similar studies of the African grasslands (Knox et al. 2011; Ramoelo et al. 2011).

Overall, the PLS and RF regression models yielded similar accuracies in terms of the RMSE prediction values for both the N (0.25% and 0.26%, respectively) and for the fibre (8.14 and 7.56%, respectively). Marginally, the RF regression produced the highest R^2 (0.71) of prediction for N concentration, while the PLS produced the highest R^2 (0.70) for predicting fibre concentrations. There were less significant differences in the R^2, RMSECV and /RMSE prediction, which were obtained, using the calibration and validation datasets for the RF models, compared to those of the PLS (Figure 7.2 and 7.3). It was therefore concluded that the performance of the RF model was relatively more-stable, compared to that of the PLS.

Table 7.4 Percentage variation accounted for by PLS and RF regression models

	Training Dataset (n =92)					**Test Dataset (n =40)**		
Model	Nutrient	PLS component	R^2	RMSECV (%)	NSE	RMSECV (%)	NSE	R^2
(A) PLS	**Nitrogen**							
	F. costata	2	0.53	0.16	0.13	0.20	0.10	0.18
	T. triandra	2	0.62	0.33	0.18	0.26	0.59	0.60
	R. altera	3	0.74	0.27	-0.43	0.24	-0.24	0.19
	All combined	**3**	**0.51**	**0.29**	**0.51**	**0.25**	**0.61**	**0.61**
	Fibre							
	F. costata	2	0.44	7.36	0.36	6.18	0.52	0.52
	T. triandra	2	0.43	8.73	0.42	7.10	0.66	0.68
	R. altera	3	0.55	6.69	0.42	5.82	0.51	0.62
	All combined	**3**	**0.56**	**8.11**	**0.56**	**8.49**	**0.61**	**0.63**
(B) RF	**Nitrogen**							
	F. costata		0.45	0.14	0.42	0.19	0.18	0.34
	T. triandra		0.56	0.26	0.50	0.26	0.61	0.60
	R. altera		0.51	0.20	0.18	0.20	0.13	0.47
	All combined		**0.71**	**0.23**	**0.70**	**0.24**	**0.68**	**0.71**
	Fibre							
	F. costata		0.57	6.39	0.51	7.39	0.31	0.34
	T. triandra		0.52	7.98	0.51	7.88	0.58	0.61
	R. altera		0.66	0.72	0.57	6.28	0.60	0.88
	All combined		**0.66**	**7.22**	**0.65**	**9.55**	**0.62**	**0.66**

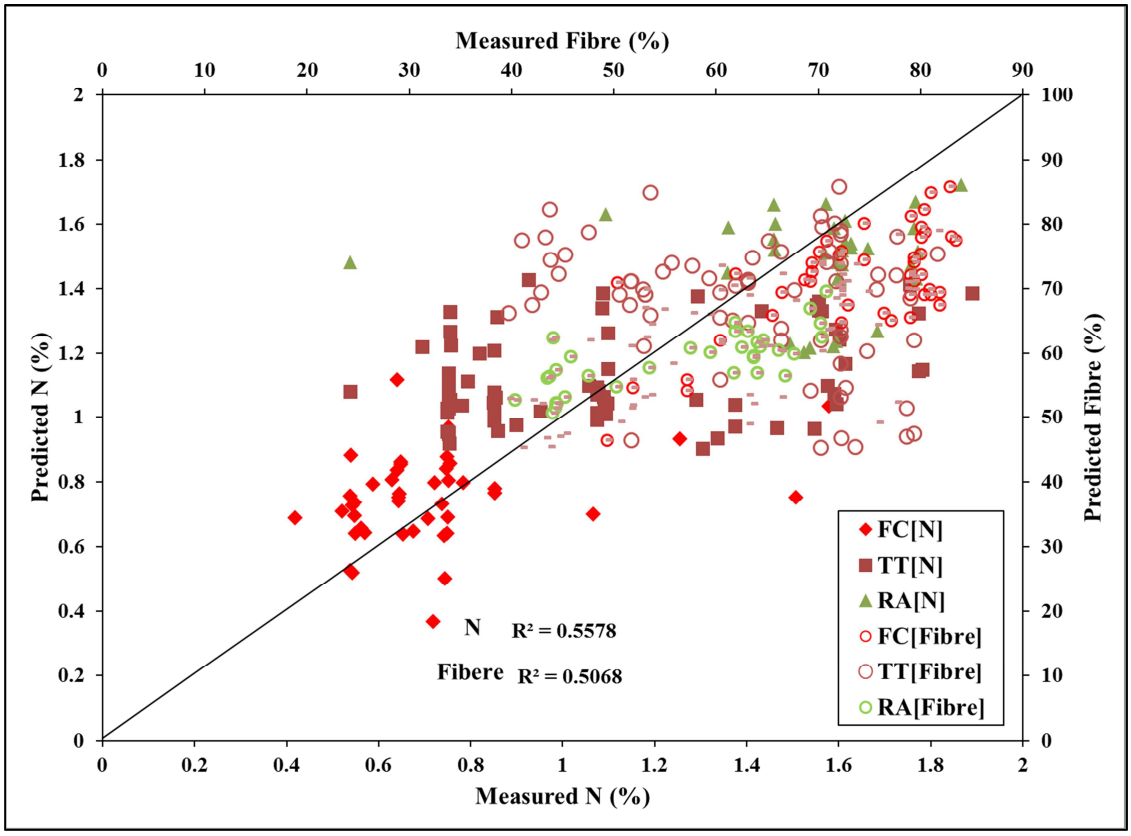

Figure 7.2 PLS regression using the training dataset: relationship between observed and predicted N and fibre concentration for the C3 and C4 type grasses in the Cathedral Peak study area. The regression models were developed, using all the WV2 bands and derived NDVIs (Eq. 1–6). Key: F. costata (FC), T. triandra (TT), R. altera (RA) and nitrogen (N).

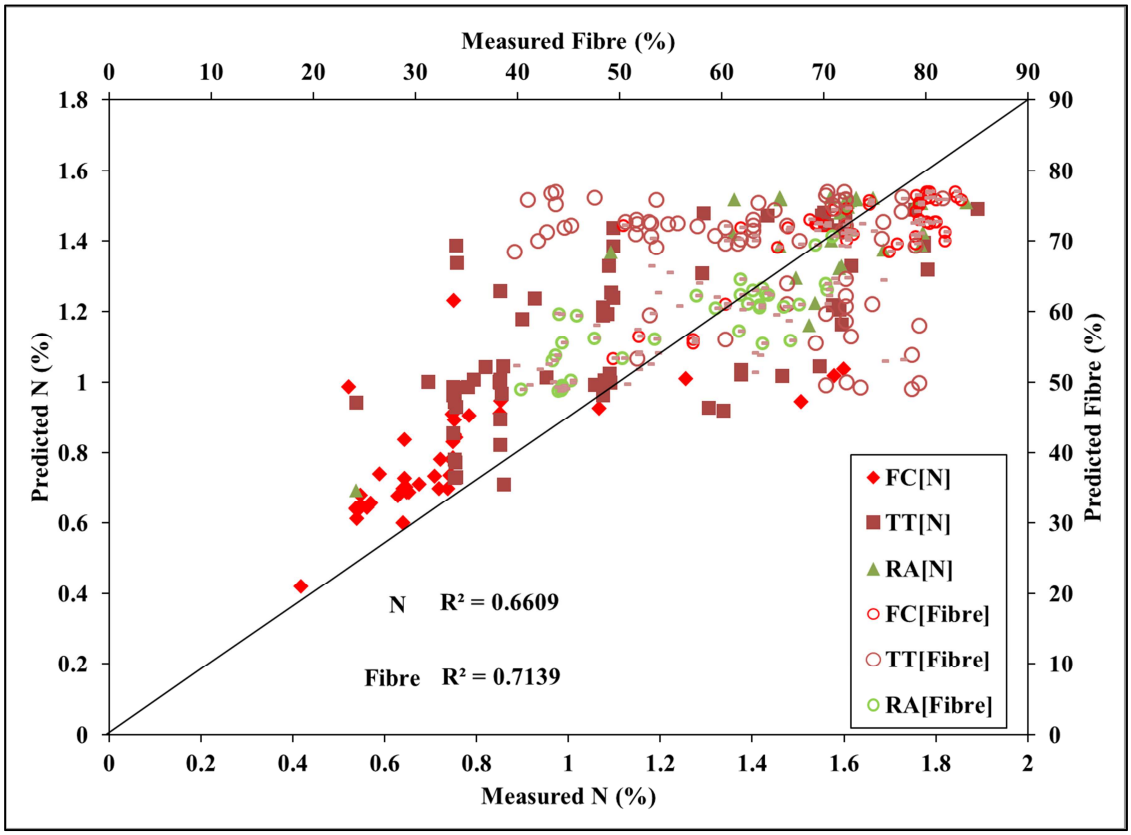

Figure 7.3 RF regression using the training dataset: relationship between observed and predicted N and fibre concentration for the C3 and C4 type grasses in the Cathedral Peak study area. The regression models were developed, using all the WV2 bands and derived NDVIs (Eq. 1–6).

7.3.5 Variable importance measures

The PLS and RF variable importance for the input WV2 spectral bands and the derived NDVIs were used to explore the relevance of the predictor variables. Figure 7.4 shows the variable importance scores (VIPs) obtained for the estimation of the variability of N and fibre concentrations among the investigated species. In the case of the PLS model, the VIPs were calculated in terms of the mean increase in RMSECV matrix for each of the predictor variables. Similarly, the RF model's VIPs were measured in terms of the average increase in the OOB error, which represented the decrease in the predictive performance of the model, when each predictor variable was both included as an observation and when it was randomly permuted (Mutanga et al. 2012). In general, the input predictors of WV2 spectral bands 3, 5–8, contributed highest, in predicting the dependent N content, using the RF model. On the hand, bands 3, 6–8 yielded the best VIP for predicting fibre content in the area, using the RF model.

Notably both PLS and RF model predicted the nutrients using similar combination of WV2 bands as illustrated by their VIP. The key issue with the VIP results obtained in this study thus affirms the notion that canopy radiometric response, can be partially linked to any set wavebands in predicting vegetation biochemicals that are in turn influenced by the canopy biophysical variables such as LAI, leaf angle distribution, or biomass. Possible reasons for the same bands selected in predicting the two nutrients can be attributed for example to the characteristic nature of canopy N concentration which may also be a function of biomass density (Baret et al. 2007). In this regard there is need to further investigate the effect of other variables on spectral reflectance of the montane c3 and c4 grasslands and develop techniques to quantify the contribution of N.

The NDVI3 and NDVI4 contributed best in predicting N and fibre, respectively. Markedly, the traditional NDVI1 and 2 yielded relatively lower VIP scores for the prediction of N concentration in the area. Noticeably, the two regression models depicted similar VIP patterns, reflecting the differences in the relevance of the input WV2 spectral bands as well as the NDVIs variables, in predicting N and fibre concentrations in the study are.

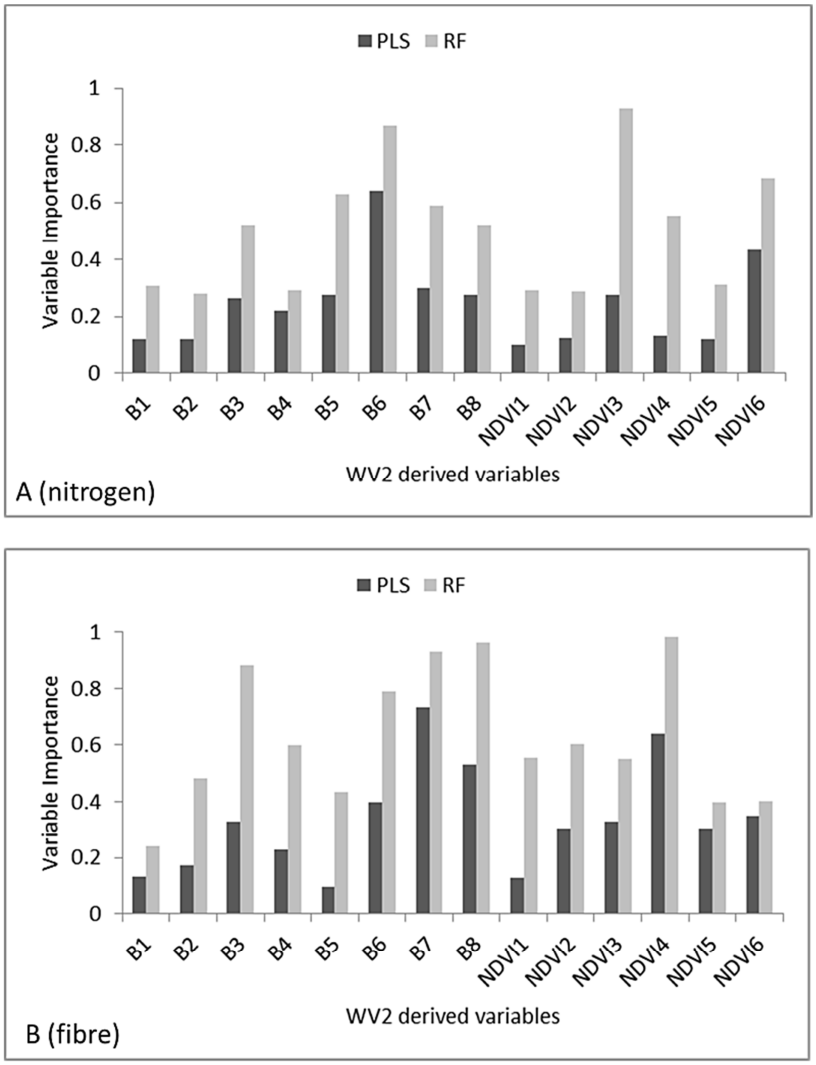

Figure 7.4 Measured variable importance for both the WV2 raw bands and the derived NDVIs for predicting N and fibre, using the PLS and RF regression models. The VIP scores were based on the models developed, using the respective optimized tuning parameters for the PLS and the RF algorithms.

7.4 Discussion

The accurate mapping of rangeland vegetation composition is critical for a better understanding of the range condition and for monitoring the dynamics of C3 and C4 type of grass species, especially in the light of climate change (Davidson and Csillag 2001; Foody and Dash 2007; Goodin and Henebry 1997; Tieszen et al. 1997). However, the complexity associated with the abundance and composition of the C3 and C4 grasses in montane environments, introduces difficult challenges for remote sensing (Adjorlolo et al. 2012b). Despite the challenges, the

current study investigated the use of the new generation multispectral image data to predict and map the variability of N and fibre concentrations across some key C3 and C4 grasses, in an African montane rangeland. The performances of vegetation indices, PLS and RF methods were investigated with the WV2 data, to assess the two forage parameters in the Cathedral Peak region of the Drakensberg mountains. The results obtained showed that variability in N and fibre concentrations can be rapidly and non-destructively predicted and mapped, with considerable accuracies using WV2 imagery.

7.4.1 The relationships among WV2 derive data, N and fibre concentrations

The results obtained from the PLS and RF regression models have shown that the red (630–690 nm) spectral absorption maxima for plant pigment content correlated reasonably with the N and fibre concentrations. In the electromagnetic spectrum, vegetation reflectance is focused on the absorption of red light by leaf pigment (e.g. chlorophyll) content, which correlates positively, with N concentrations, at the canopy scale.

The NDVI3 calculated (Eq. 3), using the red-edge and the NIR-2 (860–1040 nm) bands of WV2, correlated the best with the canopy N and fibre concentration. This was expected, since the red-edge reflectance is known to be very effective for measuring the health of plants, which is strongly linked to their physiological activities, relating to N status as well as structural composition. The red-edge reflectance is less affected by the underlying canopy biophysical complexes. It has proved to respond more linearly to canopy parameters, such as leaf area index (LAI) and plant pigment (e.g. chlorophyll) compared to traditional NDVI, which in most cases are limited by the saturation problems, even at relatively low LAI values (Cho et al. 2008; Danson and Plummer 1995; Hansen and Schjoerring 2003). In addition, the new NIR-2 band is configured, such that it is less affected by atmospheric influence (Ozdemir and Karnieli 2011) and thus combines very well with the red-edge band to predict N concentration. In this regard, the good results can be explained on the basis that the red-edge and NIR-2 bands combined more consistently in measuring the vegetation vigour, which is strongly related to the plants' N content (Kokaly et al. 2009; Kokaly and Clark 1999).

Nitrogen is a relatively small component of leaf mass and it is largely related to leaf pigments and protein content, which are in turn, indicative of plants heath. There is a reasonable amount of uncertainty regarding the direct effects that N content may have on the red reflectance,

at the canopy scale of measurement. For example, several studies that investigated the remote sensing capabilities to predict variability in canopy N concentrations, using the red region, have obtained inconsistent results thus far (Asner and Martin 2008; Hansen and Schjoerring 2003; Kawamura et al. 2008; Mutanga et al. 2004; Ollinger 2011). The relatively weak correlation coefficient results obtained for the red band, in predicting the variability of N concentrations can be attributed to a host of uncertainties in the grasslands under investigation. A number of remote sensing studies have found that the biophysical and biochemical complexities in grasslands rise large uncertainties in plant N content mapping, by using the spectral response properties of the vegetation canopy (Carter et al. 2005; Carter and McCain 1993; Curran 1989, 1994; Thenkabail et al. 2004b; Thenkabail et al. 2000). These complexities include the pixel mixing problem (Davidson and Csillag 2003; Tieszen et al. 1997) and/or the saturation problem (Grant and Scholes 2006; Mutanga and Skidmore 2004c; Skidmore et al. 2010). This is likely the case in late April, when the WV2 image was acquired for the current study area.

On the other hand, results from the current study showed that the standard NDVI2 (Eq. 2) calculated, using the red and the NIR-1(770–895 nm) bands, did not improve the prediction of N in the area. It is clear from Table 7.3 that the low coefficient ($R^2 = 0.15$) for the red band affected the performance of the output NDVI2. This can be attributed to the strong correlation between foliar N and healthy plant materials, in which under high canopy biomass or cover condition, the traditional NDVI turns to saturate (Mutanga and Skidmore 2004c; Mutanga et al. 2004; Skidmore et al. 2010). The results obtained in the current study have, in addition, indicated that all the NDVIs correlated strongly with the concentration of fibre in the grass species under investigation. The red-edge band is one of the additional narrowband (< 50 nm; Thenkabail et al. 2001)) of WV2, which correlated best with the fibre concentration in the area. The reason for the strong correlation between the NDVIs and fibre is not readily discernible. However, inferences can be made from the characteristics related to the canopy structural scattering and to the large range of variation in the observed fibre status. These could have contributed in influencing the spectral reflectance values across the C3 and C4 grasses, which differ in leaf or canopy structural composition (Slaton et al. 2001).

In this study, a number of the WV2 (the Green, Red-edge, and the NIR-1and 2) bands contributed independently, to predict variability in canopy fibre concentration across the grasses investigated in this study. The visible Green (510–580 nm) region has been reported to be useful

for the precise measurement of reflectance maxima in healthy vegetation (Carter et al. 2005). Adjorlolo et al. (2012a) reported that spectral reflectance information in the shortwave infrared (SWIR) region was critically important for the remote sensing applications in the C3 and C4 grass environment. This was the case in another study by Peterson et al. (2002), who reported that the greatest spectral differences between cold season C3 and warm season C4 grasslands occur in both NIR and SWIR. The SWIR spectra are sensitive to vegetation moisture status because of water absorption properties in this spectral region, which often differ between the two groups of grasses (Niu et al. 2008; Niu et al. 2005; Ricotta et al. 2003). In this regard, it is worth noting that the additional 8 SWIR bands information in the upcoming DigitalGlobe's WorldView-3 (http://www.digitalglobe.com/content/worldview3/) should offer improved capability for the estimation of folia nutrient concentration, in the C3 and C4 grass environment.

Overall, these studies, including the current investigation, demonstrate the usefulness of new generation remote sensors (e.g. WoldView-2 satellite) or anticipated sensors, containing critical wavebands information. Such sensors are expected to offer the potential to overcome difficulties associated with the use of traditional multispectral systems. Nonetheless, the findings in the current study were based purely on empirical models. Therefore, it should be noted that the results could inherently reflect the specific environment or the statistical structure in the dataset. Hence, these experimental results may not be applicable in different environments, seasons and vegetation types.

7.4.2 Prediction performance of PLS and RF on WV2 spectral data

The PLS method has been widely used in the regression type of remote sensing applications (Asner and Martin 2008; Cho et al. 2007; Doughty et al. 2011; Hansen and Schjoerring 2003; Kawamura et al. 2008; Kruse et al. 2006; Zengeya et al. 2012). On the other hand, remote sensing applications involving the classification of vegetation characteristics have been increasingly implemented, using the RF modelling approach (Adjorlolo et al. 2012a; Chan and Paelinckx 2008; Ham et al. 2005; Pal 2005; Rodriguez-Galiano et al. 2012). In comparison, only a few remote sensing studies have investigated RF regression, for the prediction of vegetation biophysical and biochemical parameters (Abdel-Rahman et al. 2009; Grömping 2009; Ismail and Mutanga 2010; Liaw and Wiener 2002; Mutanga et al. 2012).

In the present study, the PLS and RF were assessed and both models effectively provided similar VIP measurements for the input WV2 spectral bands, as well as the developed NDVIs. Both regression models showed some level of under-estimation and over-estimation of the N and fibre values, particularly at the two extremes (Figures 7.2 and 7.3). However, the predicted values did not fall below the range of the input data. The significantly positive NSE index (Richter et al. 2012) values (NSE $\geq$ 0.9 and 0.5 $\leq$ NSE $\leq$ 0.8) obtained for model predictions across all species confirms that the PLS and RF algorithms contributed to improve the detection of the response variables.

The results obtained were similar to those of previous studies that sought to predict nutrient quality in the African rangeland environment (Knox et al. 2011; Ramoelo et al. 2011). In the case of RF regression, such a limitation can be attributed to the manner in which the regression trees are constructed (Abdel-Rahman et al. 2009; Mutanga et al. 2012). Despite the known advantages of the PLS regression (Hansen and Schjoerring 2003; Kawamura et al. 2008), the present study has shown that the RF regression model yielded relatively better stability in its predictive performance, with reference to the differences between calibration and validation results. This demonstrated that RF regression is a useful and robust technique for remote sensing applications, involving the use of remote sensing data (e.g. WV2 spectral bands) for predicting the variability of N and fibre concentration, across C3 and C4 type of grasslands.

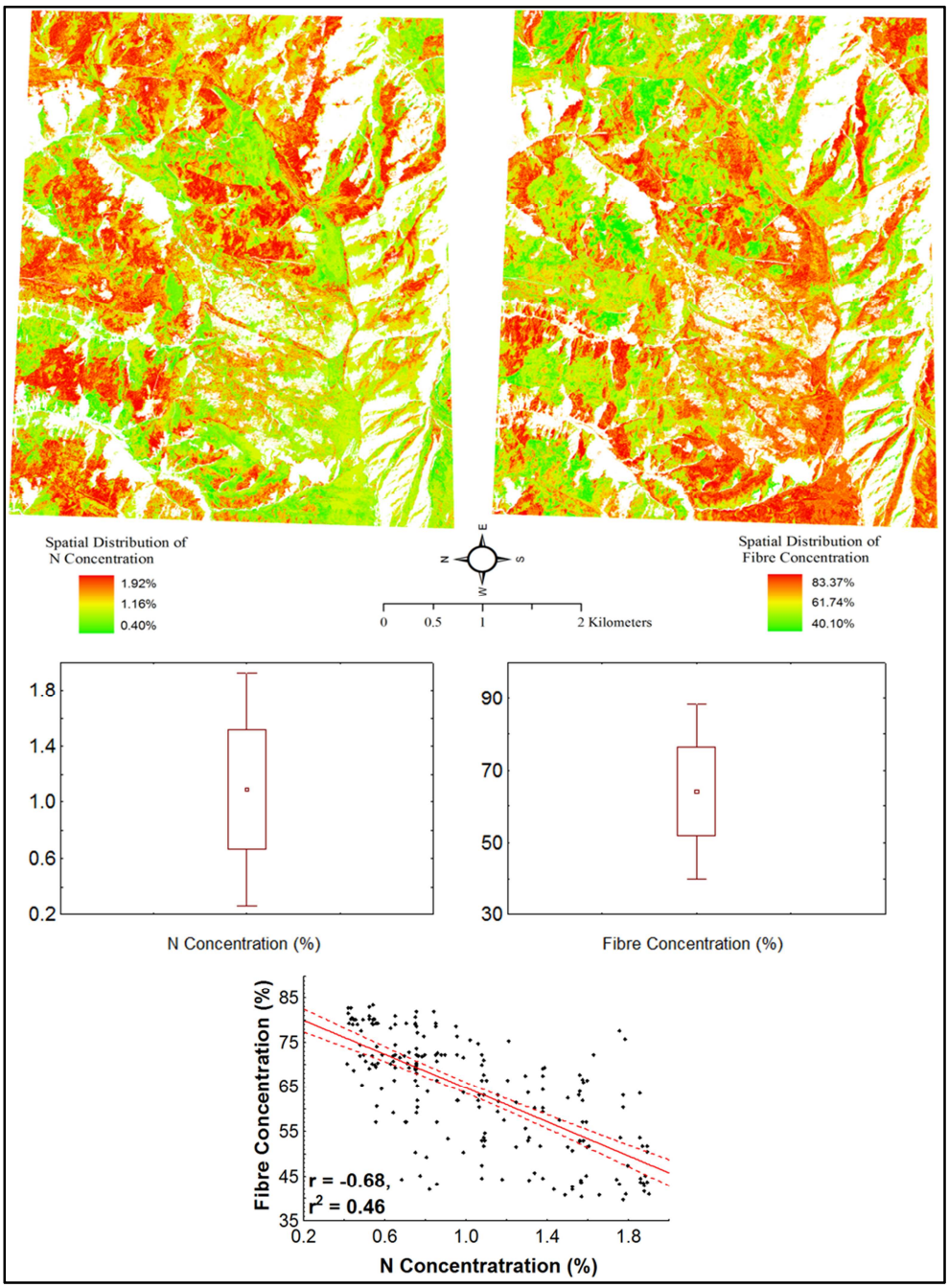

Figure 7.5 EnMAP-Box implementation of the RF spatial estimation of (a) nitrogen and (b) fibre distribution map (2 m spatial resolution), with box plots showing the mean and spread of N/fibre concentration (%) across all the target plant species. The scatter plot shows the inter-correlation (coefficient confidence P = 0.95) between N and fibre concentration in the area. For further details on the utility of the EnMAP-Box software, the reader is kindly referred to www.enmap.org. For all the maps, the white pixels are those that were masked out and therefore not modelled, and the light green pixels represent the lowest N and fibre concentration values on the respective scale bar. The colour scale bar represents the % dry matter of the respective nutrients.

7.4.3 Mapping nitrogen and fibre concentrations

In this study, the RF regression was considered to be more accurate in estimating variability in the N and fibre concentration. Hence, the RF model was implemented in the EnMAP-box software to map the spatial distributions of N and fibre concentrations across the landscape, which was composed of the C3 and C4 type of grasses under investigation. Figure 7.5 depicts the spatial distribution of the canopy N and fibre concentration in the study area. The most striking difference between the maps was the markedly inverse patterns for the N and fibre concentrations. These tend to mirror the composition of the species, as observed in the field and shown by the inter-correlation between N and fibre, across all species. Although the area mapped is underlain by relatively homogeneous Drakensberg basalt formations of the Stormberg series, there are, undoubtedly, spatial variations in canopy N and fibre concentrations, as depicted in the final map. Notably, the N and fibre maps inversely show high and low concentrations, with reference to the different slope, aspect and the type of dominant grassland patches in the area. Previous studies conducted in the area have shown that there are strong consociations of *T. triandra* and *F. costata* occurrence on warm northerly and cool southerly slopes, respectively (Granger and Schulze 1977; Hill 1996). Strikingly, the final maps of N and fibres tend to reflect the mean nutrient concentration between the *T. triandra* and *F. costata*, as observed in the study area. In addition, the valley areas, which are relatively colder, due to slope, aspect or shading, tend to have lower and higher concentrations of N and fibre, respectively. This corresponds well with the C3 grass community characterisation in the area (Killick 1963). In contrast, the spatial distribution of the nutrient variability, as predicted for the high-lying crests, which were dominated by the *T. triandra* and R. altera, showed higher and lower concentrations of N and fibre, respectively. Although the study offers a step towards mapping forage nutrient concentration, the results should be treated with some level of caution. The contribution of other canopy factors such as biomass and species variability should be taken into account in future research (Baret et al. 2007).

7.5 Conclusion

The potential of WV2 image data was investigated, using the PLS and RF regression methods, to predict the variability of N and fibre concentration across a landscape with C3 and C4 grass composition. The results obtained from this study showed that canopy N and fibre concentrations

can be rapidly and non-destructively predicted or mapped. The predictive accuracies of both the PLS and RF methods were improved considerably through the inclusion of the developed NDVIs, for N and fibre concentration. Using the most stable regression algorithm in the RF method, the spatial distributions of both N and fibre concentrations were mapped with 22% and 11.25% of the sample mean for N and fibre, respectively. The use of PLS and RF with WV2 waveband data is promising for a wide range of spectral assessment of ecosystem parameters. However, the effect of other canopy properties remains a subject of investigation.

CHAPTER 8: SUBTROPICAL C3 AND C4 GRASSES AND REMOTE SENSING

This chapter summarizes the major findings of the research, with some final reflections and suggestions for future work.

8.1 Introduction

The findings in the present research support previous studies that, based on the differences in physiological features, plant species following the C3 or C4 photosynthetic pathway can be characterized into different plant functional types (PFT) (Adjorlolo et al. 2012b; Ehleringer and Monson 1993; Ustin and Gamon 2010). Ecologists and plant geographers have historically described vegetation patterns and relationships between individual plants and their environments by establishing the PFT grouping of species. In that regard, recent ecological research has been driven by the need to determine functional type classification, in an effort to develop predictive capability, that will help researchers to better understand how ecosystems respond to the effect of global climate change and its impacts on vegetation communities (Ustin and Gamon 2010). However, there is a need to develop rapid assessment methods in the light of a rapid changing climate or environmental conditions. Of importance in the current investigation is to classify some major plant functional types (C3 and C4 grasses) and estimate the variability of nutrient quality (foliar concentration of nitrogen, proteins, plant moisture content and fibre) across C3 and C4 grasslands. These foliar nutrients have been strongly linked to ecosystem productivity (Beeri et al. 2007) and to the condition of rangelands (Trollope 1990). The variability in foliar nutrient concentration influences the distribution patterns of herbivores (Knox et al. 2012; McNaughton 1990; Ramoelo et al. 2012; Skidmore et al. 2010). In this regard, remote sensing technology has offered possibilities for a time- and cost-efficient way of characterizing the biophysical and biochemical parameters of rangeland vegetation, especially in large and inaccessible landscapes.

However, the capability provided by various (multispectral and hyperspectral) remote sensing techniques, invariably comes with a number of challenges. Generally, the traditional broadband multispectral systems average spectra in broad bandwidths. This limits their ability to detect fine spectral features, which are often masked by spectrally similar canopy biophysical and biochemical parameters (Kumar et al. 2001). On the other hand, although hyperspectral sensors provide narrowband spectral information that can be extracted for a detailed assessment of vegetation, the data is high in dimensionality. The high spectral dimensionality introduces the multicollinearity problem in statistical analysis (Adjorlolo et al. 2013). In particular, these problems limit the utility of conventional statistical methods in processing hyperspectral data. On the other hand, the extraction of hyperspectral information from remote sensors often results in

the use of sophisticated statistical filtering or data reduction methods. These include machine learning algorithms, such as artificial neural networks, support vector machines and random forests, which make no assumptions of the distribution of the input data.

However, these approaches are limited, because a fundamental spectral property of target vegetation, the asymmetrical nature of the highly correlated hyperspectral reflectance information is not accounted for. Therefore, an important question in this thesis is whether, by using a technique that accounts for the inter-band spectral correlation information, the problems of dimensionality and related multicollinearity in hyperspectral sensing can be mitigated? Such a technique needs to consider, in more detail, how radiation interacts (absorption and reflectance properties) with the target vegetation canopy (Curran 1994). From this background, the following objectives were set: (i) to identify optimal spectral bands for the classification and mapping of C3 and C4 grass species or communities, and (ii) to predict variability in the canopy concentration of nutrient variables across these two groups. The analyses were up-scaled from field spectral measurements, to evaluate the capability of a new generation multispectral sensor (WorldView-2 image data).

The overall challenges and opportunities in the use of remote sensing for C3 and C4 grass species discrimination and mapping was reviewed, with specific emphasis on the subtropical montane grassland environments in the Southern African region.

8.2 Challenges and opportunities in the use of remote sensing for C3 and C4 grass studies

The classification of C3 and C4 grasses represents a scheme that is consistent with the PFT approach used for the assessment of broad vegetation types, rather than the higher taxonomic identification of individual species (Ustin and Gamon 2010). The concept of PFT depends on the physiological, structural and morphological characteristics to group species that perform a common ecological role. In recent times, the C3 and C4 types of grass have been described, reflecting responses of vegetation to changes in environmental resources or conditions. Since C3 and C4 grasses respond differently to changes in environmental conditions, the ability to monitor such changes is an important way of identifying the impacts of climate on C3 and C4 grassland environments. Although remote sensing techniques have offered the possibility to interpret spectral information in ways that correlate field observations with C3 and C4 grass dynamics, the major challenge is linking physical observations made in specific wavebands to ecosystem

functional processes (Schaepman et al. 2009; Ustin and Gamon 2010). In spite of the limitations, remote sensing techniques continue to offer the potential to detect and map the PFTs of C3 and C4 grasslands. In summary, the following evidences were gathered from the review of current literature (Chapter 2):

- Differences in the structural composition between C3 and C4 grasses affect the amount and shape of radiation reflected, absorbed and/or transmitted in the wavelength spectrum.
- Canopy structural and biochemical properties relate differently and affect different aspects of the electromagnetic spectrum.
- C3 and C4 grass canopies can be discriminated by using reflectance and absorption features and a combination of temporal trajectory vegetation indices and.
- The integration of remote sensing techniques, solar radiation patterns, wetness gradients, and topographic elements, such as slope orientation and altitudinal gradients, could improve the detection and mapping of C3 and C4 grass canopies in tropical montane grassland landscapes.

8.3 Overcoming the high spectral dimensionality and multicollinearity problems in the classification analysis of C3 and C4 grasses

8.3.1 Spectral resampling based on user-defined inter-band correlation filter function

The high spectral dimensionality and associated multicollinearity challenges in the statistical analysis of hyperspectral data were addressed in this thesis (Chapter 3) by developing the user-defined inter-band correlation filter of the WTC (the weighting thresholds of correlation). Figure 8.1 shows: (a) the magnitude of the inter-band correlation matrix of the input spectral bands, and (b) an illustration of the WTCs, which mitigated the problem of collinearity in the input spectral bands in various ways, depending on a specific threshold of inter-band correlation. These WTCs were used to assess the spectral separability among the C3 and C4 grasses investigated and to address the effect of multicollinearity on the performance of the random forest-based variable and on the subsequent band-subset selection, under predictor correlation. The performance of the random forest model was also evaluated by applying data derived from the process of resampling spectra to the Hyperion sensor's spectral resolution.

Overall, the results obtained from this study suggested an improved classification of the response variables, using less correlated input spectral bands information. This study demonstrated the trade-off between the number of spectral bands and the resolution of remotely sensed imagery, and the trade-off between the higher spectral resolution and reduced signal-to-noise ratio wavebands. It was therefore concluded that the proposed resampling technique is potential step forward to reprogramming hyperspectral data for applications involving C3 and C4 grass species or communities.

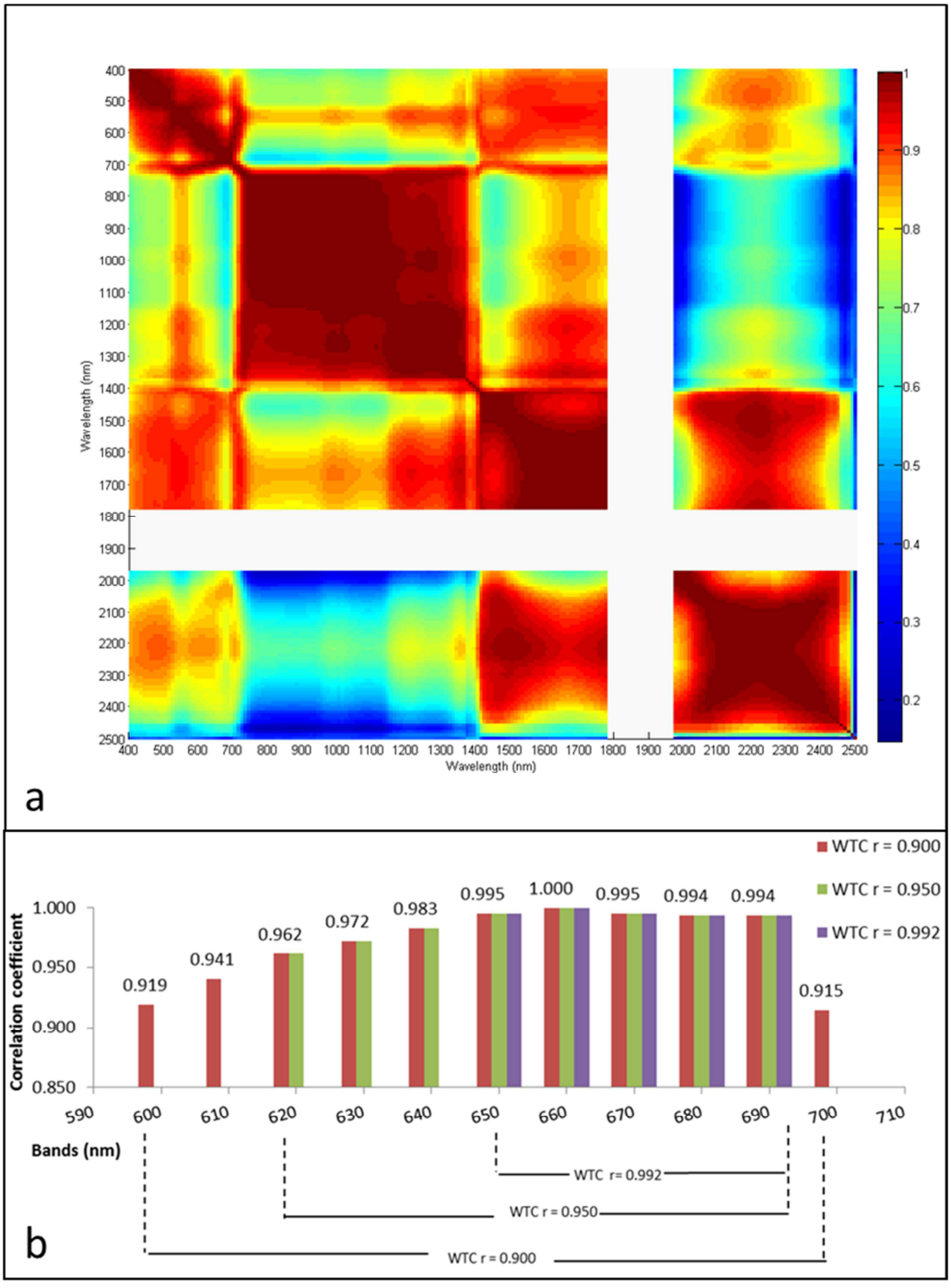

Figure 8.1 Pearson's r correlation coefficients matrix (plot) of the input spectral bands, calculated using reflectance data aggregated into 10-nm-wide band intervals (a) and an illustration of the user-defined inter-band correlation filter for 660 nm band-centre (b). The white space (Figure a) represents removed noisy bands.

8.3.2 Optimizing the spectral resolutions for C3 *and* C4 *grass species classification*

This experiment focused on evaluating the spectral resolutions of the developed wavelengths of known reflectance/absorption features, to discriminate between the C3 and C4 grass species, and compared the results to those obtained, using the traditional approach of resampling hyperspectral data, to the spectral response of some commercial satellites sensors (Chapter 4). This study has highlighted the critical importance of optimizing spectral resolutions for specific regions of the measured spectrum. In a previous study, Dalponte et al. (2009) suggested that the spectral resampling approach, such as the technique used in the current study, would offer the best method in optimizing spectral resolutions for various remote sensing applications involving the classification of vegetation cover types. Thus, the approach of the current study should assist rangeland ecologists who are involved in modeling the functional type of C3 and C4 grasses, to select programmable sensor types for specific applications. Specifically, the current study emphasized the call for optimized spectral resolutions and band configuration, taking into cognizance future satellites, such as WorldView-3, which are designed to include strategically-selected wavebands in the shortwave infrared region for specific application (Cho et al. 2012; Dalponte et al. 2009; Taylor et al. 2012). The following highlights summarize the findings of Chapter 4:

- Spectral resampling of hyperspectral data to a reduced number of bands should account for the asymmetrical nature of inter-band correlation curves around a few chosen band-centres of interest.
- Accounting for the asymmetrical nature of inter-band correlations between a given band-centre of interest and its short and longer wavelength neighbours can result in an optimal spectral resolution for the concerned band-centre.
- Wavebands located above 2000 nm of the ShortwaveIR region are critically required for the classification of C3 and C4 grass canopies: wavebands from this region are conspicuously absent for some existing multispectral sensors assessed. Hence, the proposed method achieved improved classification accuracy (Kappa = 0.81), when compared to the resampled ASTER, GeoEye-1, IKONOS, QuickBird, RapidEye, SPOT 5 and WorldView-2 satellites (Kappa = 0.78, 0.65, 0.62, 0.59, 0.65, 0.62 and 0.76, respectively).

- NDVIs computed from a two band combination of resampled WorldView-2 band offer further capability to achieve improved classification of C3 and C4 grasses.
- Robust, non-parametric algorithms, such as the random forest was used both, as a screening tool for selecting predictor variable importance and as a classification technique for complex C3 and C4 grass applications, following its successful applications in other studies.

8.4 Mapping and characterizing C3 and C4 grass communities

The significance of this study (Chapter 5) is to evaluate the potential of the new generation, high spatial resolution, WorldView-2 imagery, for the identifying and mapping of subtropical montane C3 and C4 grassland communities, using the random forest modelling approach. Using the raw 8-bands and NDVIs extracted from the WV-2 imagery, the random forest learning algorithm was exploited, to classify the C3 and C4 grasslands. An overall accuracy of 82.26% and kappa = 0.76. The conclusions that were drawn from the experimental results are as follows:

- The WV-2 sensor has the capability of identifying and mapping broad groups of C3 and C4 grass communities, as well as mapping the spatial distribution of the fern, *P. aquilinum* stands in a montane environment (Table 8.1).
- The new bands of WV-2, especially the coastal-blue, red-edge and NIR-2 bands, were important predictor variables for mapping the four classes considered in this study.
- Overall, the WV-2 sensor has the potential to improve the accuracy of C3 and C4 grass classification in subtropical montane environments.
- To improve the accuracy of identification and mapping of these two groups, future studies should consider reflectance responses from other regions of the electromagnetic spectrum, including bands in the shortwave infrared region, which were found to be useful in previous studies (Adjorlolo et al. 2012a; Adjorlolo et al. 2013).

Table 8.1 Classification accuracies obtained using the WV-2 bands only dataset and WV-2 bands and the 6 NDVIs computed.

	Training		Test set	
Data set	Over all accuracy (%)	Kappa	Overall accuracy (%)	Kappa
WV-2 bands only	73.95	0.65	74.19	0.66
WV-2 bands + 6NDVIs	80.39	0.74	82.26	0.76

Table 8.2 Confusion matrices reflecting results of the input spectral features extracted from the WV-2 multispectral 8-bands image only.

	Predicted class				
Observed Class	*F. costata* grassland community	*P. aquilinum* stands	*T. triandra* dominant-grassland	*R. altera* mixed-grassland	Sum of reference
Training dataset					
F. costata grassland community	54	14	5	4	77
P. aquilinum stands	8	55	5	2	70
T. triandra dominant-grassland	2	9	71	13	95
R. altera mixed-grassland	5	0	14	50	69
Sum of estimation	69	78	95	69	311
Producer's Accuracy (%)	78.26	70.51	74.74	72.46	-
User's Accuracy (%)	70.13	78.57	74.74	72.46	-
Validation dataset					
F. costata grassland community	*20*	*2*	2	2	26
P. aquilinum stands	*5*	*23*	2	4	34
T. triandra dominant-grassland	*4*	*4*	21	6	35
R. altera mixed-grassland	*0*	*1*	0	28	29
Sum of estimation	29	30	25	40	124
Producer's Accuracy (%)	68.97	76.67	84.00	70.00	-
User's Accuracy (%)	76.92	67.65	60.00	96.55	-

Table 8.3 Confusion matrices reflecting results of input spectral features extracted from WV-2 8-bands and the 6 NDVIs combined.

	Predicted class				
Observed Class	*F. costata* grassland community	*P. aquilinum* stands	*T. triandra* dominant-grassland	*R. altera* mixed-grassland	Sum of reference
Training dataset					
F. costata grassland community	59	5	5	1	70
P. aquilinum stands	8	49	3	3	63
T. triandra dominant-grassland	2	5	71	11	89
R. altera mixed-grassland	4	0	14	71	89
Sum of estimation	73	59	93	86	311
Producer's Accuracy (%)	80.82	83.05	76.34	82.56	
User's Accuracy (%)	84.29	77.78	79.78	79.78	
Validation dataset					
F. costata grassland community	20	2	0	0	22
P. aquilinum stands	5	27	0	5	37
T. triandra dominant-grassland	1	0	20	5	26
R. altera mixed-grassland	2	0	2	35	39
Sum of estimation	28	29	22	45	124
Producer's Accuracy (%)	71.43	93.10	90.91	77.78	
User's Accuracy (%)	90.91	72.97	76.92	89.74	

8.5 Predicting and mapping canopy nutrient variability across C3 and C4 grasslands

The traditional forage nutrient analysis from common laboratory spectroscopy procedures provides accurate, point-based information, but often does not provide it in a timely way, to allow changes in forage quality across large and inaccessible landscapes to be measured. The extension of the laboratory-based spectrometry to canopy level has been made possible through the use of remote sensing approaches at the field level, aerial platform and space-bone imaging spectrometers. Remote sensing hyperspectral systems have been more successfully used in predicting and mapping foliage concentration of nutrients, such as nitrogen, proteins, canopy moisture content, and fibre (Knox et al. 2011; Mutanga and Skidmore 2004b). The major problems, however, are that hyperspectral data are high in dimensionality, with the related problem of multicollinearity, which poses a limitation in statistical analysis. On the other hand, the hyperspectral data are relatively inaccessible, compared to multispectral data, and can be more costly. In that regard, there is a need to identify simple, but efficient, data analysis methods to process hyperspectral data or to evaluate the additional capability provided by the newer generation of remote sensors. Considering these observations, Chapter 6 and 7 focused on

evaluating the developed user-defined inter-band correlation function to resampling spectra, for forage nutrient estimation and on assessing the capability of the new generation WorldView-2 image data, in a C3 and C4 grassland environment.

8.5.1 Predicting variability, using in-situ canopy reflectance

In this experiment (Chapter 6), wavebands obtained through the developed user-defined inter-band correlation filter technique of resampling hyperspectral data (Chapter 3 and 4) were assessed. The results obtained showed that the adopted spectral resampling technique was useful for optimizing the spectral resolutions. The adopted wavebands were then integrated with the PLS regression model to predict the variability in N, CP, moisture, and fiber concentrations, across key C3 and C4 type of grasses in the Drakensberg montane landscape. Results from the PLS analysis showed that variability in forage nutrient concentration can be rapidly and non-destructively predicted at acceptable accuracies. Using the wavebands of known reflectance effects with the PLS regression analysis, variability in N, CP, moisture, and fibre concentration could be predicted at 80%, 79%, 71% and 71%, respectively. This study concluded that the integration of the PLS algorithm and wavebands obtained through the user-defined inter-band correlation filter technique, is promising for the assessment of a wide range of vegetation nutrient parameters. However, further studies are needed, to establish the variance of forage nutrient concentrations in the C3 and C4 grasses, focusing on various growth stages, a variety of environments and the level of canopy.

8.5.2 Predicting canopy nitrogen and fibre variability across C3 *and* C4 *grasslands*

In this study (Chapter 7), the potential to estimate the variability of forage nutrient concentration across C3 and C4 grassland environments was assessed from analysis of *in situ* reflectance measurements to actual imagery, using data derived from WorldView-2 satellite image. The input spectral data were analyzed, using the partial least squares and random forest regression methods. The results obtained from this study showed that canopy N and fibre concentrations can be rapidly and non-destructively predicted or mapped. The predictive accuracies of both the PLS and RF methods were improved considerably through the inclusion of the developed NDVIs. However, in Chapter 3, 4 and 6, it has been reported that wavebands in the shortwave infrared (SWIR) region were critically important in explaining the variability biophysical and

biochemical features in the C3 and C4 grassland environment. This was also the case in another study by Peterson et al. (2002), who reported that the greatest spectral differences between cold season C3 and warm season C4 occur in both NIR and SWIR. The SWIR spectra are sensitive to vegetation moisture status because of water absorption properties in this spectral region, which often differ between these two groups of grasses (Niu et al. 2008; Niu et al. 2005; Ricotta et al. 2003). It was therefore concluded that the additional 8 SWIR band information in the upcoming DigitalGlobe's WorldView-3 (http://www.digitalglobe.com/content/worldview3/) should offer an improved capability for estimating folia nutrient concentration in the C3 and C4 grass environments.

In addition, this study demonstrated that the systematic assessment of variable importance by PLS and RF models, offer the opportunity to identify the most useful WV2 wavebands for estimating N and fibre concentration in the context of C3 and C4 grasses. Using the most stable regression algorithm in the RF method, the spatial distributions in the variability of both N and fibre concentrations were mapped (Figure 8.2) with acceptable accuracies (22% and 11.25%) of the sample mean for N and fibre, respectively). The use of PLS and RF with WV2 waveband data is promising for a wide range of spectral assessments of ecosystem parameters. The study concluded that empirical models can be site- or sensor-specific and might be less useful for application in large areas or in different environments (Cho et al. 2007; Curran 1994; Ollinger 2011).

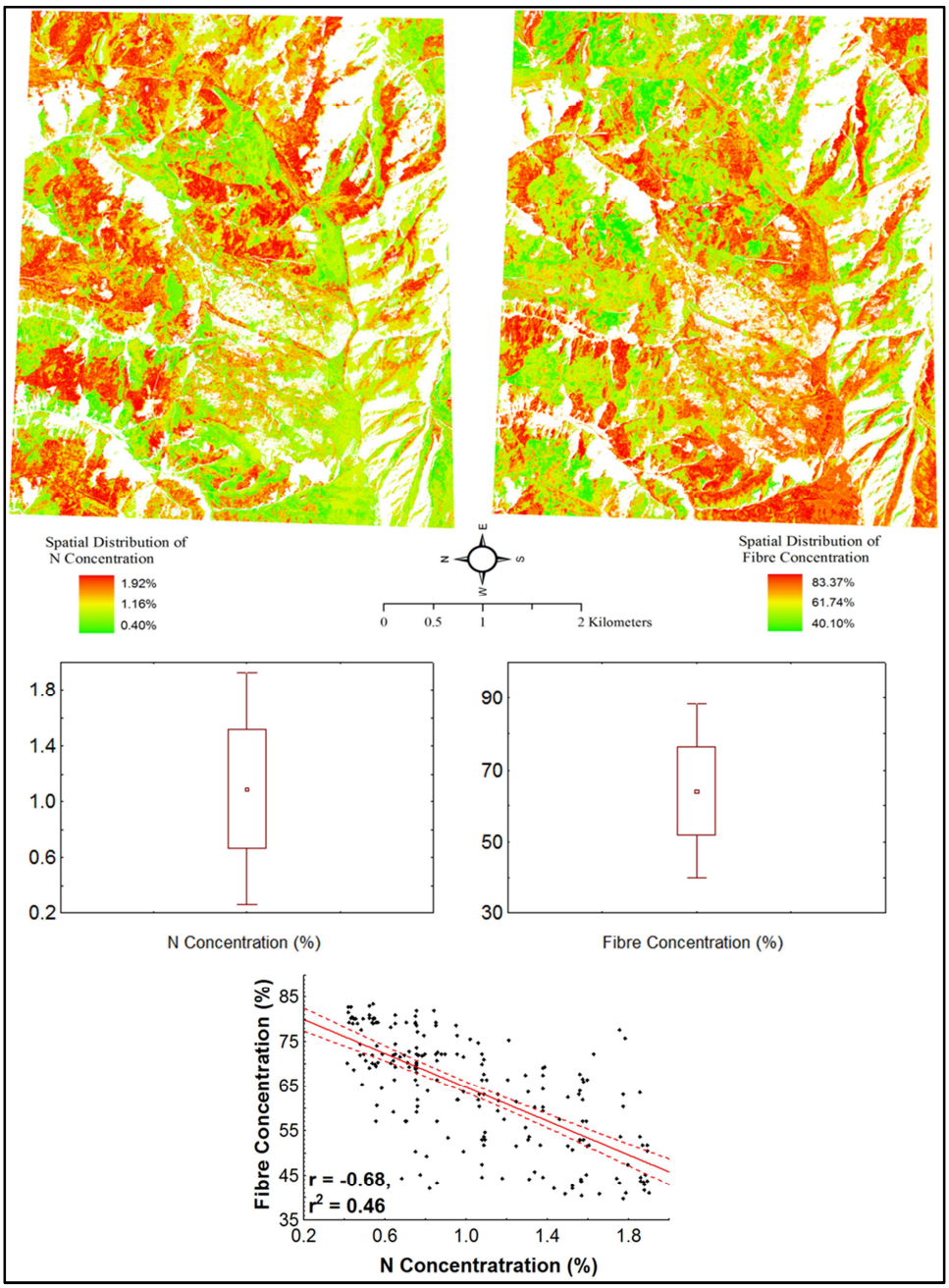

Figure 8.2 EnMAP-Box implementation of the RF spatial estimation of (a) nitrogen and (b) fibre distribution map, with box plots showing the mean and spread of N/fibre concentration (%) across all the target plant species. The scatter plot shows the inter-correlation (coefficient confidence P = 0.95) between N and fibre concentration in the area. For further details on the utility of the EnMAP-Box software, the reader is kindly referred to www.enmap.org. For all the maps, the white pixels are those that were masked out and therefore not modelled, and the light green pixels represent the lowest N and fibre concentration values on the respective scale bar. The colour scale bar represents the % dry matter of the respective nutrients.

8.6 Overall conclusions and the future

In this thesis, it has been shown that the information contained in hyperspectral data can be used to discriminate C3 and C4 grass species or community composition. It has also been shown that the information extracted from hyperspectral data can be used to predict the variability of foliar nutrient concentration of N, protein, canopy moisture, and fibre content in C3 and C4 grassland environments. However, an important lesson from this thesis is that the high spectral dimensionality and the related multicollinearity problems in the use of data acquired through hyperspectral systems can be mitigated by using the developed user-defined inter-band correlation filter technique, which optimizes the spectral information content of the input spectral bands. Having established that spectral resolutions could be optimized for the classification of C3 and C4 grasses using *in situ* spectra only datasets (Chapter 3, 4 and 6), the analyses involving use of the new generation multispectral sensor, WorldView-2 imagery, were done (Chapter 5 and 7, respectively). It was therefore concluded that potential exists for the use of remote sensing data for applications involving C3 and C4 grasses in the subtropical montane grassland environments. For this purpose, it has shown that the utility of the developed vegetation indices, integrated with advance statistical techniques, including the partial least squares and random forests, offer improvement in the classification and regression analyses of the input spectral data.

For the classification analyses, calibrated datasets were used to discriminate the C3 and C4 grass response variables in the study area. It has been generally reported in literature that specific wavebands or regions of spectral reflectance are very sensitive to differences in the C3 and C4 type of grass species, which differ in a number of biophysical (e.g. leaf or canopy structural) and biochemical (e.g. percentage N content) features. Consequently the variability in the concentration of biochemicals (e.g. foliar nutrient variables) across these two groups could be predicted, based on spectral data. This capability was observed in this present research, by applying the developed user-defined inter-band correlation filter function to the absorption features, at wavelengths presented in Table 8.2.

Table 8.4 Wavelengths corresponding to known absorption features, as described in previous studies to be highly sensitive to the properties of reflection or absorption of vegetation structural and biochemical characteristics.

No.	Band (nm)	Known causal compound/ feature	Source
1	470	Total plant pigment concentration	Blackburn, 1998
2	530	Chlorophyll a absorption	Gamon et al., 1997
3	600	Nitrogen	Fourty and Baret, 1997
4	660	Nitrogen	Carter, 1994
5	700	Total chlorophyll, nitrogen	Carter, 1994
6	720	Total chlorophyll, leaf mass	Horler et al., 1983
7	820	Leaf mass, leaf area index	Curran, 1994
8	1540	Cellulose, vegetation water content	Carter, 1994
9	2060	Protein	Carter, 1994
10	2280	Cellulose, sugar, starch, leaf mass	Curran, 1994
11	2300	Leaf mass, vegetation water content	Carter, 1994
12	2450	Cellulose, protein, nitrogen	Carter, 1994
13	2470	Cellulose, protein	Kumar et al., 2001

Since the C3 and C4 photosynthetic type of grasses in the current study area exhibit a distinct phenology (e.g. Figure 8.2a and b), it would be interesting to test these wavelengths, using the user-defined inter-band technique and to apply this to spectra representing the variation found in the field. It is expected that the differences in phenology, which affect the relative contribution of the C3 or C4 grass types to the pixel spectra to be extracted, should improve the estimation of percentage folia concentration of nitrogen, proteins, canopy-moisture or fibre. It is recommended here that future work, focusing on this, could lead to a greater utility of spectral data for C3 and C4 grass studies in the subtropical montane environments.

In general, the calibration models derived from the WorldView-2 image acquired for the Cathedral Peak study area in April, predicting forage N and fibre, were satisfactory. However, it still remains to be tested whether a similar performance of the WorldView-2 bands would be obtained for various phenological stages, which invariably affect forage nutrient status. In that regard, further research is necessary to clarify the issue of phenology and its influence on the spectra. It would also be interesting to include the C3 and C4 grass canopy biomass as a variable in the partial least squares and random forests analyses for temporal spectral datasets, to test if the classifications and predictions could be improved for the periods when there is the greatest difference in C3:C4 biomass, especially in mixed grasslands (e.g. *Themeda-Festuca* mixed grasslands in the sub-alpine vegetation belt in this current study area).

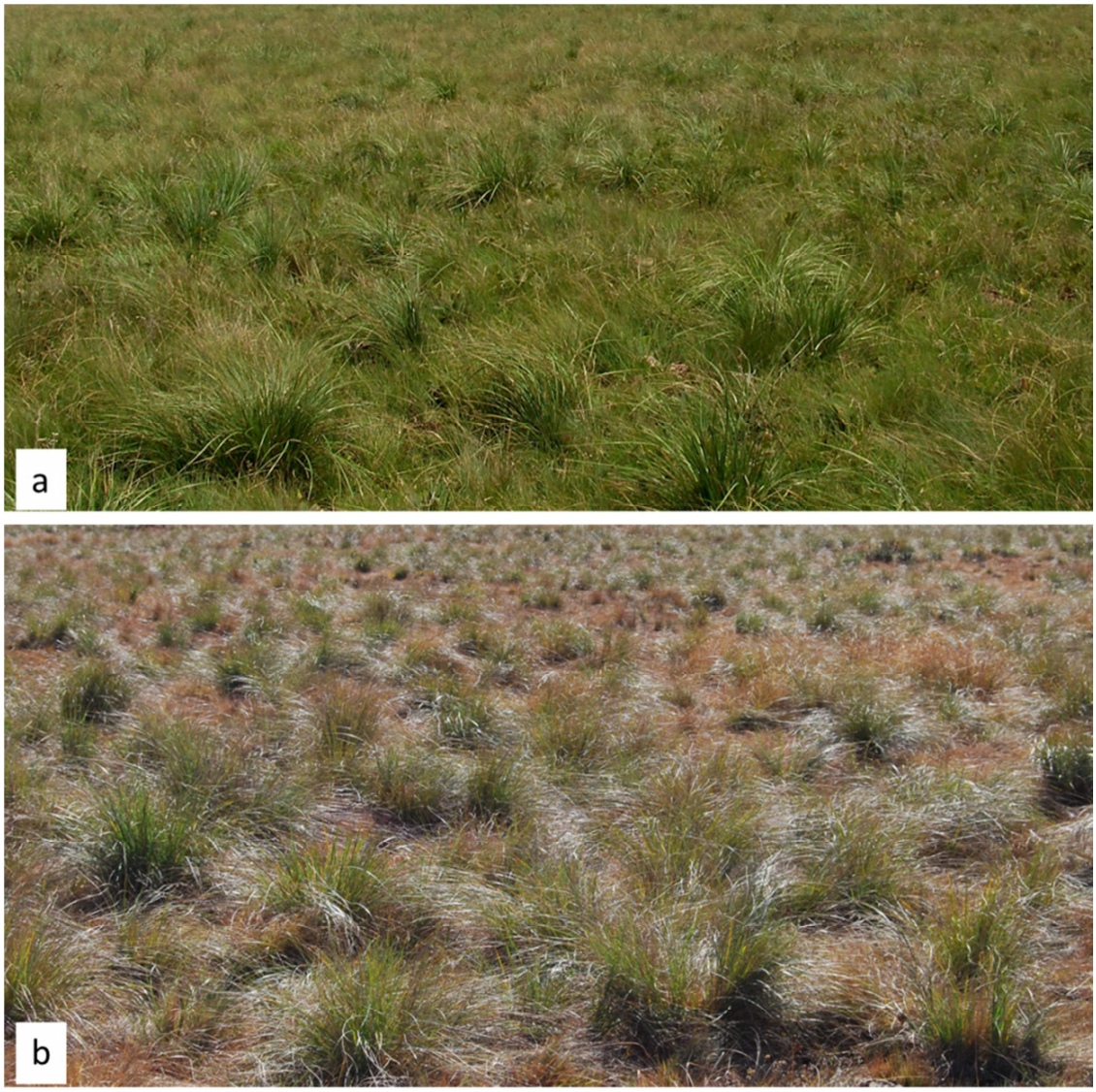

Figure 8.3 Typical field site showing the relatively evergreen, *Festuca* species in April (a) and in late June (b), after the area had frosted. In between the tussocks are C4 grasses, including the characteristic red-coloured *T. triandra*, which is clearly visible in winter (June).

It has been widely reported in literature that a good spectral dataset should consider samples of grass species, as well as the quality and quantity that represent all the variation found in the field. Therefore, the objective of a future study may be to collect canopy biomass, in order to ascertain its effect on the C3 and C4 grass spectra. In the case of collecting data to explore field spectroscopy for the estimation of the C3 and C4 forage nutrients, it is recommended that the researcher: (i) is aware of all structural and biochemical features that affect grass sward characteristics (e.g. canopy structure differences due to variability in inter-cellular scattering,

variation in percentage concentration of foliar N and moisture.); and (ii) is aware of the bio-physiological differences in grass swards caused by seasonality (e.g. evergreen C3 grasses, such as *F. costata*, versus C4, *T. triandra*, which frosts in winter).

In this current research and in previously reported studies, the greatest spectral differences between the C3 and C4 grasses occur in the shortwave infrared region of the electromagnetic spectrum (1195–2245 nm). This is sensitive to vegetation moisture status, which can differ between the C3 and C4 groups of grasses. It would therefore be interesting to suggest that the additional 8-shortwave infrared bands in the upcoming DigitalGlobe's WorldView-3 (http://www.digitalglobe.com/content/worldview3/) should offer the improved capability for studying the C3 and C4 grasses in the subtropical montane grasslands of the Southern Africa.

REFERENCES

Abdel-Rahman, E.M., Ahmed, F.B., & Ismail, R. (2009). 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*, 712-728

Adam, E., Mutanga, O., Rugege, D., & Ismail, R. (2009). Field spectrometry of papyrus vegetation (*Cyperus papyrus L.*) in swamp wetlands of St Lucia, South Africa In, *Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009* (pp. IV- 260 – IV- 263). Cape Town: IEEE

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*

Adjorlolo, C., Mutanga, O., Cho, A.M., & Ismail, R. (2013). Spectral resampling based on user-defined inter-band correlation filter: C3 and C4 grass species classification. *International Journal of Applied Earth Observation and Geoinformation, 21*, 535-544

Adjorlolo, C., Mutanga, O., Cho, M.A., & 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*, 47-61

Alvarez-Añorve, M., Quesada, M., & de la Barrera, E. (2008). Remote Sensing and Plant Functional Groups: Physiology, Ecology, and Spectroscopy in Tropical Systems. In M. Kalacska & G. Arturo Sanchez-Azofeifa (Eds.), *Hyperspectral Remote Sensing of Tropical and Sub-Tropical Forests* (pp. 27-45). Boca Raton, FL 33487-2742: CRC Press

Asner, G.P. (1998). Biophysical and Biochemical Sources of Variability in Canopy Reflectance. *Remote Sensing of Environment, 64*, 234-253

Asner, G.P., Bateson, C.A., Privette, J.L., El Saleous, N., & Wessman, C.A. (1998a). Estimating vegetation structural effects on carbon uptake using satellite data fusion and inverse modeling. *Journal of Geophysical Research, 103*, 28839-28853

Asner, G.P., & Martin, R.E. (2008). Spectral and chemical analysis of tropical forests: Scaling from leaf to canopy levels. *Remote Sensing of Environment, 112*, 3958-3970

Asner, G.P., Wessman, C.A., & Archer, S. (1998b). Scale Dependence of Absorption of Photosynthetically Active Radiation in Terrestrial Ecosystems. *Ecological Applications, 8*, 1003-1021

Barbehenn, R.V., Chen, Z., Karowe, D.N., & Spickard, A. (2004). C3 grasses have higher nutritional quality than C_4 grasses under ambient and elevated atmospheric CO_2. *Global Change Biology, 10*, 1565-1575

Baret, F., Houlès, V., & Guérif, M. (2007). Quantification of plant stress using remote sensing observations and crop models: the case of nitrogen management. *Journal of Experimental Botany, 58*, 869-880

Bastien, P. (2005). PLS generalised linear regression. *Computational Statistics & Data Analysis, 48*, 17-46

Becker, B.L., Lusch, D.P., & Qi, J. (2007). A classification-based assessment of the optimal spectral and spatial resolutions for Great Lakes coastal wetland imagery. *Remote Sensing of Environment, 108*, 111-120

Beeri, O., Phillips, R., Hendrickson, J., Frank, A.B., & Kronberg, S. (2007). Estimating forage quantity and quality using aerial hyperspectral imagery for northern mixed-grass prairie. *Remote Sensing of Environment, 110*, 216-225

Bishop, C.M. (1995). *Neural Networks for Pattern Recognition*: Oxford University Press

Black, C.C., Chen, T.M., & Brown, R.H. (1969). Biochemical Basis for Plant Competition. *Weed Science, 17*, 338-344

Blackburn, G.A. (1998). Quantifying chlorophylls and caroteniods at leaf and canopy scales: an evaluation of some hyperspectral approaches. *Remote Sensing of Environment, 66*, 273-285

Blackburn, G.A. (2007a). Hyperspectral remote sensing of plant pigments. *Journal of Experimental Botany, 58*, 855-867

Blackburn, G.A. (2007b). Wavelet decomposition of hyperspectral data: a novel approach to quantifying pigment concentrations in vegetation. *International Journal of Remote Sensing, 28*, 2831-2855

Blackburn, G.A., & Pitman, J.I. (1999). Biophysical controls on the directional spectral reflectance properties of bracken (Pteridium aquilinum) canopies: Results of a field experiment. *International Journal of Remote Sensing, 20*, 2265-2282

Blackburn, G.A., & Steele, C.M. (1999). Towards the Remote Sensing of Matorral Vegetation Physiology: Relationships between Spectral Reflectance, Pigment, and Biophysical Characteristics of Semiarid Bushland Canopies. *Remote Sensing of Environment, 70*, 278-292

Bond, W., J., Geldenhuys, C.J., Everson, T.M., Everson, C.S., & Calvin, M.F. (2004). Fire ecology: characteristics of some important biomes of sub-Sahara Africa. In J.G. Goldammer & C. de Ronde (Eds.), *Wildland Fire Management Handbook for Sub-Sahara Africa* (pp. 11-26): Global Fire Monitoring Center

Bowyer, P., & Danson, F.M. (2004). Sensitivity of spectral reflectance to variation in live fuel moisture content at leaf and canopy level. *Remote Sensing of Environment, 92*, 297-308

Bredenkamp, G.J., Spada, F., & Kazmierczak, E. (2002). On the origin of northern and southern hemisphere grasslands. *Plant Ecology, 163*, 209-229

Breiman, L. (1996). Bagging Predictors. *Machine Learning, 24*, 123 - 140

Breiman, L. (2001a). Random Forests. *Machine Learning, 45*, 5 - 32

Breiman, L. (2001b). Statistical Modeling: The Two Cultures. *Statistical Science, 16*, 199 - 231

Breiman, L., & Cutler, A. (2004). Random forests tools for predicting and understanding data. In, *Interface workshop-April 2004*

Bremond, L., Boom, A., & Favier, C. (2012). Neotropical C3/C4 grass distributions – present, past and future. *Global Change Biology, 18*, 2324-2334

Brodersen, C.R., & Vogelmann, T.C. (2007). Do epidermal lens cells facilitate the absorptance of diffuse light? *American Journal of Botany, 94*, 1061-1066

Bruzzone, L., Roli, F., & Serpico, S.B. (1995). An extension to multiclass cases of the Jeffreys-Matusita distance. *IEEE Transactions on Geoscience and Remote Sensing, 33*, 1318-1321

Bruzzone, L., & Serpico, S.B. (2000). A technique for feature selection in multiclass problems. *International Journal of Remote Sensing, 21*, 549-563

Buschmann, C., & Nagel, E. (1993). In vivo spectroscopy and internal optics of leaves as basis for remote sensing of vegetation. *International Journal of Remote Sensing, 14*, 711-722

Carter, G.A. (1991). Primary and Secondary Effects of Water Content on the Spectral Reflectance of Leaves. *American Journal of Botany, 78*, 916-924

Carter, G.A., Knapp, A.K., Anderson, J.E., Hoch, G.A., & Smith, M.D. (2005). Indicators of plant species richness in AVIRIS spectra of a mesic grassland. *Remote Sensing of Environment, 98*, 304-316

Carter, G.A., & McCain, D.C. (1993). Relationship of leaf spectral reflectance to chloroplast water content determined using NMR microscopy. *Remote Sensing of Environment, 46*, 305-310

Carter, G.A., & Miller, R.L. (1994). Early detection of plant stress by digital imaging within narrow stress-sensitive wavebands. *Remote Sensing of Environment, 50*, 295-302

Ceccato, P., Flasse, S., Tarantola, S., Jacquemoud, S., & Grégoire, J.-M. (2001). Detecting vegetation leaf water content using reflectance in the optical domain. *Remote Sensing of Environment, 77*, 22-33

Chamaillé-Jammes, S., & Bond, W.J. (2010). Will global change improve grazing quality of grasslands? A call for a deeper understanding of the effects of shifts from C_4 to C_3 grasses for large herbivores. *Oikos, 119*, 1857-1861

Chan, J.C.-W., & Paelinckx, D. (2008). Evaluation of random forest and adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment, 112*, 2999-3011

Chazdon, R.L. (1978). Ecological Aspects of the Distribution of C4 Grasses in Selected Habitats of Costa Rica. *Biotropica, 10*, 265-269

Chen, D.-X., Hunt, H.W., & Morgan, J.A. (1996). Responses of a C3 and C4 perennial grass to CO_2 enrichment and climate change: Comparison between model predictions and experimental data. *Ecological Modelling, 87*, 11-27

Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., & Eklundh, L. (2004). A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. *Remote Sensing of Environment, 91*, 332-344

Chen, J.M. (1996). Evaluation of vegetation indices and a modified simple ratio for boreal applications. *Canadian Journal of Remote Sensing, 22*, 229-242

Chi, M., & Bruzzone, L. (2007). Semi-supervised classification of hyperspectral images by svms optimized in the primal. *IEEE Transactions on Geoscience and Remote Sensing, 45*, 1870-1880

Chi, M., Feng, R., & Bruzzone, L. (2008). Classification of hyperspectral remote-sensing data with primal SVM for small-sized training dataset problem. *Advances in Space Research, 41*, 1793-1799

Cho, M.A., Debba, P., Mathieu, R., Naidoo, L., van Aardt, J., & Asner, G.P. (2010). Improving Discrimination of Savanna Tree Species Through a Multiple-Endmember Spectral Angle Mapper Approach: Canopy-Level Analysis. *Geoscience and Remote Sensing, IEEE Transactions on, 48*, 4133-4142

Cho, M.A., Mathieu, R., Asner, G.P., Naidoo, L., van Aardt, J., Ramoelo, A., Debba, P., Wessels, K., Main, R., Smit, I.P.J., & Erasmus, B. (2012). Mapping tree species composition in South African savannas using an integrated airborne spectral and LiDAR system. *Remote Sensing of Environment, 125*, 214-226

Cho, M.A., Skidmore, A., Corsi, F., van Wieren, S.E., & Sobhan, I. (2007). Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares regression. *International Journal of Applied Earth Observation and Geoinformation, 9*, 414-424

Cho, M.A., & Skidmore, A.K. (2009). Hyperspectral predictors for monitoring biomass production in Mediterranean mountain grasslands: Majella National Park, Italy. *International Journal of Remote Sensing, 30*, 499-515

Cho, M.A., Skidmore, A.K., & Atzberger, C. (2008). Towards red-edge positions less sensitive to canopy biophysical parameters for leaf chlorophyll estimation using properties optique spectrales des feuilles (PROSPECT) and scattering by arbitrarily inclined leaves (SAILH) simulated data. *International Journal of Remote Sensing, 29*, 2241-2255

Chopping, M., Su, L., Laliberte, A., Rango, A., Peters, D.P.C., & Kollikkathara, N. (2006). Mapping shrub abundance in desert grasslands using geometric-optical modeling and multi-angle remote sensing with CHRIS/Proba. *Remote Sensing of Environment, 104*, 62-73

Clevers, J.G.P.W., van der Heijden, G.W.A.M., Verzakov, S., & Schaepman, M.E. (2007). Estimating grassland biomass using SVM band shaving of hyperspectral data. *Photogrammetric Engineering & Remote Sensing, 73*, 1141-1148

Cochrane, M.A. (2000). Using vegetation reflectance variability for species level classification of hyperspectral data. *International Journal of Remote Sensing, 21*, 2075 - 2087

Conklin-Brittain, N.L., Dierenfeld, E.S., Wrangham, R.W., Norconk, M., & C., S.S. (1999). Chemical Protein Analysis: A Comparison of Kjeldahl Crude Protein and Total Ninhydrin Protein from Wild, Tropical Vegetation. *Journal of Chemical Ecology, 25*, 2601-2622

Corson, M.S., Rotz, C.A., & Skinner, R.H. (2007). Evaluating warm-season grass production in temperate-region pastures: A simulation approach. *Agricultural Systems, 93*, 252-268

Craine, J.M., Ballantyne, F., Peel, M., Zambatis, N., Morrow, C., & Stock, W.D. (2009). Grazing and landscape controls on nitrogen availability across 330 South African savanna sites. *Austral Ecology, 34*, 731-740

Crawford, M.M., JiSoo, H., Yangchi, C., & Ghosh, J. (2003). Random forests of binary hierarchical classifiers for analysis of hyperspectral data. In, *Advances in Techniques for Analysis of Remotely Sensed Data, 2003 IEEE Workshop on*, (pp. 337-345). Greenbelt, Maryland, USA: IEEE

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

Curran, P.J. (1994). Imaging spectrometry. *Progress in Physical Geography, 18*, 247-266

Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., & Lawler, J.J. (2007). Random forests for classification in ecology. *Ecology, 88*, 2783-2792

Dalponte, M., Bruzzone, L., Vescovo, L., & Gianelle, D. (2009). The role of spectral resolution and classifier complexity in the analysis of hyperspectral images of forest areas. *Remote Sensing of Environment, 113*, 2345-2355

Danson, F.M. (1995). Developments in the remote sensing of forest canopy structure. In F.M. Danson & S.E. Plummer (Eds.), *Advances in Environmental Remote Sensing*. Chichester: John Wiley & Sons

Danson, F.M., & Plummer, S.E. (1995). Red-edge response to forest leaf area index. *International Journal of Remote Sensing, 16*, 183-188

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

Dash, J., & Curran, P.J. (2004). The MERIS terrestrial chlorophyll index. *International Journal of Remote Sensing, 25*, 5403-5413

Daughtry, C.S.T., Gallo, K.P., Goward, S.N., Prince, S.D., & Kustas, W.P. (1992). Spectral estimates of absorbed radiation and phytomass production in corn and soybean canopies. *Remote Sensing of Environment, 39*, 141-152

Davidson, A., & Csillag, F. (2001). The Influence of Vegetation Index and Spatial Resolution on a Two-Date Remote Sensing-Derived Relation to C4 Species Coverage. *Remote Sensing of Environment, 75*, 138-151

Davidson, A., & Csillag, F. (2003). A comparison of three approaches for predicting C4 species cover of northern mixed grass prairie. *Remote Sensing of Environment, 86*, 70-82

Dengler, N.G., Dengler, R.E., Donnelly, P.M., & Hattersley, P.W. (1994). Quantitative Leaf Anatomy of C3 and C4 Grasses (Poaceae): Bundle Sheath and Mesophyll Surface Area Relationships. *Annals of Botany, 73*, 241-255

Diaz-Uriarte, R., & Alvarez de Andres, S. (2006). Gene Selection and Classification of Microarray Data Using Random Forest. *BMC Bioinformatics, 7*, 3

Djouadi, A., Snorrason, O., & Garber, F. (1990). "The quality of training-sample estimates of the Bhattacharyya coefficient". *IEEE Transactions on Pattern analysis and machine intelligence, 12*, 92-97

Doughty, C., Asner, G., & Martin, R. (2010). Predicting tropical plant physiology from leaf and canopy spectroscopy. *Oecologia*, 1-11

Doughty, C., Asner, G., & Martin, R. (2011). Predicting tropical plant physiology from leaf and canopy spectroscopy. *Oecologia, 165*, 289-299

Douglas J. Levey, Heidi A. Bissell, & O'Keefe, S.F. (2000). Conversion of Nitrogen to Protein and Amino Acids in Wild Fruits. *Journal of Chemical Ecology, 26*, 1749-1763

Duro, D.C., Franklin, S.E., & Dubé, M.G. (2012). A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sensing of Environment, 118*, 259-272

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*, 275-289

Edwards, E.J., Osborne, C.P., Strömberg, C.A.E., Smith, S.A., & Consortium, C.G. (2010). The Origins of C4 Grasslands: Integrating Evolutionary and Ecosystem Science. *Science, 328*, 587-591

Ehleringer, J., & Björkman, O. (1977). Quantum Yields for CO_2 Uptake in C3 and C4 Plants. *Plant Physiology, 59*, 86-90

Ehleringer, J.R., Cerling, T.E., & Flanagan, L.B. (2001). Global changes and the linkages between physiological ecology and ecosystem ecology. In M. Press, N. Huntly & S. Levin (Eds.), *Ecology: Achievement and Challenge* (pp. 115-138): Blackwell, Oxford

Ehleringer, J.R., Cerling, T.E., & Helliker, B.R. (1997). C4 photosynthesis, atmospheric CO_2, and climate. *Oecologia, 112*, 285-299

Ehleringer, J.R., & Monson, R.K. (1993). Evolutionary and Ecological Aspects of Photosynthetic Pathway Variation. *Annual Review Of Ecology And Systematics, 24*, 411-439

Ehleringer, J.R., Sage, R.F., Flanagan, L.B., & Pearcy, R.W. (1991). Climate change and the evolution of C4 photosynthesis. *Trends in Ecology & Evolution, 6*, 95-99

Fava, F., Colombo, R., Bocchi, S., Meroni, M., Sitzia, M., Fois, N., & Zucca, C. (2009). Identification of hyperspectral vegetation indices for Mediterranean pasture characterization. *International Journal of Applied Earth Observation and Geoinformation, 11*, 233-243

Fensholt, R., Huber, S., Proud, S.R., & Mbow, C. (2010). Detecting Canopy Water Status Using Shortwave Infrared Reflectance Data From Polar Orbiting and Geostationary Platforms. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of, 3*, 271-285

Ferwerda, J.G., Skidmore, A.K., & Mutanga, O. (2005). Nitrogen detection with hyperspectral normalized ratio indices across multiple plant species. *International Journal of Remote Sensing, 26*, 4083-4095

Ficken, K.J., Wooller, M.J., Swain, D.L., Street-Perrott, F.A., & Eglinton, G. (2002). Reconstruction of a subalpine grass-dominated ecosystem, Lake Rutundu, Mount Kenya: a novel multi-proxy approach. *Palaeogeography, Palaeoclimatology, Palaeoecology, 177*, 137-149

Fischlin, A., Midgley, G.F., Price, J.T., Leemans, R., Gopal, B., Turley, C., Rounsevell, M.D.A., Dube, O.P., Tarazona, J., & Velichko, A.A. (2007). Ecosystems, their properties, goods, and services. In M.L. Parry, O.F. Canziani, J.P. Palutikof, P.J. van der Linden & C.E. Hanson (Eds.), *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 211-272). Cambridge, UK: Cambridge University Press

Fisher, R.A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics, 7*, 179-188

Foody, G.M., & Dash, J. (2007). Discriminating and mapping the C3 and C4 composition of grasslands in the northern Great Plains, USA. *Ecological Informatics, 2*, 89-93

Foody, G.M., & Dash, J. (2010). Estimating the relative abundance of C3 and C4 grasses in the Great Plains from multi-temporal MTCI data: issues of compositing period and spatial generalizability. *International Journal of Remote Sensing, 31*, 351-362

Fourty, T., & Baret, F. (1997). Vegetation water and dry matter contents estimated from top-of-the-atmosphere reflectance data: A simulation study. *Remote Sensing of Environment, 61*, 34-45

Gamon, J.A. (2008). Tropical remote sensing - opportunities and challenges. In M. Kalacska & G.A. Sanchez-Azofeifa (Eds.), *Hyperspectral remote sensing of tropical and subtropical forests* (pp. 297-304): CRC Press

Gamon, J.A., Serrano, L., & Surfus, J.S. (1997). The photochemical reflectance index: an optical indicator of photosynthetic radiation use efficiency across species, functional types, and nutrient levels. *Oecologia, 112*, 492-501

Gao, B.-c. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment, 58*, 257-266

Gao, X., Huete, A.R., Ni, W., & Miura, T. (2000). Optical-Biophysical Relationships of Vegetation Spectra without Background Contamination. *Remote Sensing of Environment, 74*, 609-620

Gates, D.M., Keegan, H.J., Schleter, J.C., & Weidner, V.R. (1965). Spectral properties of plants. *Applied Optics, 4*, 11-20.

Gausman, H.W. (1985). *Plant leaf optical properties in visible and near-infrared light*. Lubbock, Texas, USA

Ghannoum, O. (2008). C_4 photosynthesis and water stress. *Annals of Botany, 103*, 635-644

Ghannoum, O. (2009). C4 photosynthesis and water stress. *Annals of Botany, 103*, 635-644

Ghimire, B., Rogan, J., & Miller, J. (2010). Contextual land-cover classification: incorporating spatial dependence in land-cover classification models using random forests and the Getis statistic. *Remote Sensing Letters, 1*, 45-54

Ghulam, A., Li, Z.-L., Qin, Q., Tong, Q., Wang, J., Kasimu, A., & Zhu, L. (2007). A method for canopy water content estimation for highly vegetated surfaces-shortwave infrared perpendicular water stress index. *Science in China Series D: Earth Sciences, 50*, 1359-1368

Gomez-Chova, L., Calpe, J., Camps-Valls, G., Martin, J.D., Soria, E., Vila, J., Alonso-Chorda, L., & Moreno, J. (2003). Feature selection of hyperspectral data through local correlation and SFFS for crop classification. In, *Geoscience and Remote Sensing Symposium, 2003. IGARSS '03. Proceedings. 2003 IEEE International* (pp. 555-557)

Goodin, D.G., & Henebry, G.M. (1997). A technique for monitoring ecological disturbance in tallgrass prairie using seasonal NDVI trajectories and a discriminant function mixture model. *Remote Sensing of Environment, 61*, 270-278

Granger, J.E., & Schulze, R.E. (1977). Incoming Solar Radiation Patterns and Vegetation Response: Examples from the Natal Drakensberg. *Vegetatio, 35*, 47-54

Grant, C.C., & Scholes, M.C. (2006). The importance of nutrient hot-spots in the conservation and management of large wild mammalian herbivores in semi-arid savannas. *Biological Conservation, 130*, 426-437

Grant, J., Hopcraft, C., Olff, H., & Sinclair, A. (2010). Herbivores, resources and risks: Alternating regulation along primary environmental gradients in savannas. *Trends in Ecology & Evolution, 25*, 119−128

Grömping, U. (2009). Variable Importance Assessment in Regression: Linear Regression versus Random Forest. *The American Statistician, 63*, 308-319

Guan, L., Liu, L., Peng, D., Hu, Y., Jiao, Q., & Liu, L. (2012). Monitoring the distribution of C3 and C4 grasses in a temperate grassland in northern China using moderate resolution

imaging spectroradiometer normalized difference vegetation index trajectories. *Journal of Applied Remote Sensing, 6*, 063535-063531

Guanter, L., Estellés, V., & Moreno, J. (2007). Spectral calibration and atmospheric correction of ultra-fine spectral and spatial resolution remote sensing data. Application to CASI-1500 data. *Remote Sensing of Environment, 109*, 54-65

Guizhong, L., & Fan, Z. (2007). An Efficient Compression Algorithm for Hyperspectral Images Based on Correlation Coefficients Adaptive Three Dimensional Wavelet Zerotree Coding. In, *Image Processing, 2007. ICIP 2007. IEEE International Conference on* (pp. II - 341-II - 344). San Antonio, TX: IEEE

Guo, B., Gunn, S.R., Damper, R.I., & Nelson, J.D.B. (2006). Band Selection for Hyperspectral Image Classification Using Mutual Information. *Geoscience and Remote Sensing Letters, IEEE, 3*, 522-526

Guo, J., & Trotter, C.M. (2004). Estimating photosynthetic light-use efficiency using the photochemical reflectance index: variations among species. *Functional Plant Biology, 31*, 255-265

Guo, X., Price, K.P., & Stiles, J. (2003). Grasslands discriminant analysis using Landsat TM single and multitemporal data. *Photogrammetric Engineering and Remote Sensing, 69*, 1255−1262

Ham, J., Yangchi, C., Crawford, M.M., & Ghosh, J. (2005). Investigation of the random forest framework for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing, 43*, 492-501

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

Hansey, C., N., Lorenz, A., J. , & de Leon, N. (2010). Cell Wall Composition and Ruminant Digestibility of Various Maize Tissues Across Development. *BioEnergy Research, 3*, 28-37

Harris, A., & Dash, J. (2010). The potential of the MERIS Terrestrial Chlorophyll Index for carbon flux estimation. *Remote Sensing of Environment, 114*, 1856-1862

Hewitson, B.C., Tadross, M.A., Sarr, A., Jain, S., Mdoka, M., Jack, C., McKellar, N., Walawege, R., Intsiful, J., Gutowski, W.J., Crane, R., & Stendel, M. (2005). The development of

regional climate change scenarios for sub-sahara Africa. . In, *Assessment of Impacts and Adaptation to Climate Change Project AF07*. Washington, DC: START secretariat.

Hill, M.J., Aspinall, R.J., & Willms, W.D. (1997). Knowledge-based and inductive modelling of rough fescue (Festuca altaica, F. campestris and F. hallii) distribution in Alberta, Canada. *Ecological Modelling, 103*, 135-150

Hill, T.R. (1996). Description, classification and ordination of the dominant vegetation communinities, Cathedral Peak, KwaZulu-Natal Drakensberg. *South African Journal of Botany, 62*, 263-269

Hoffman, T., & Vogel, C. (2008). Climate Change Impacts on African Rangelands. *Rangelands, 30*, 12-17

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

Hoskuldsson, A. (2003). Analysis of latent structures in linear models. *Journal of Chemometrics, 17*, 630-645

Hughes, G. (1968). On the mean accuracy of statistical pattern recognizers. *Information Theory, IEEE Transactions on, 14*, 55-63

Irisarri, J.G.N., Oesterheld, M., Verón, S.R., & Paruelo, J.M. (2009). Grass species differentiation through canopy hyperspectral reflectance. *International Journal of Remote Sensing, 30*, 5959 - 5975

Ismail, R., & Mutanga, O. (2010). A comparison of regression tree ensembles: Predicting Sirex noctilio induced water stress in Pinus patula forests of KwaZulu-Natal, South Africa. *International Journal of Applied Earth Observation and Geoinformation, 12*, S45-S51

Ismail, R., & Mutanga, O. (2011). Discriminating the early stages of *Sirex noctilio* infestation using classification tree ensembles and shortwave infrared bands. *International Journal of Remote Sensing, 32*, 4249-4266

Jacquemoud, S., & Baret, F. (1990). PROSPECT: A model of leaf optical properties spectra. *Remote Sensing of Environment, 34*, 75-91

Jacquemoud, S., Baret, F., Andrieu, B., Danson, F.M., & Jaggard, K.W. (1995). Extraction of vegetation biophysical parameters by inversion of PROSPECT+SAIL models on sugar beet canopy reflectance data. Application to TM and AVIRIS sensors. *Remote Sensing of Environment, 52*, 163-172

Kavzoglu, T., & Mather, P.M. (2002). The role of feature selection in artificial neural network applications. *International Journal of Remote Sensing, 23*, 2919-2937

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*, 131-145

Kemp, P.R., & Williams III, G.J. (1980). A Physiological Basis for Niche Separation Between Agropyron Smithii (C"3) and Bouteloua Gracilis (C"4). *Ecology, 61*, 846-858

Killick, D.J.B. (1963). *An account of the plant ecology of the Cathedral Peak area of the Natal Drakensberg*

King, R.L., Ruffin, C., LaMastus, F.E., & Shaw, D.R. (1999). The analysis of hyperspectral data using Savitzky-Golay filtering-practical issues. 2. In, *Geoscience and Remote Sensing Symposium, 1999. IGARSS '99 Proceedings. IEEE 1999 International* (pp. 398-400 vol.391)

Knapp, A.K., & Carter, G.A. (1998). Variability in Leaf Optical Properties Among 26 Species from a Broad Range of Habitats. *American Journal of Botany, 85*, 940-946

Knipling, E.B. (1970). Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. *Remote Sensing of Environment, 1*, 155-159

Knox, N.M., Skidmore, A.K., Asner, G.P., Prins, H.H.T., van der Werff, H.M.A., de Boer, W.F., van der Waal, C., de Knegt, H.J., Kohi, E.M., Slotow, R., & Grant, C.C. (2011). Mapping savanna forage quality, in the dry season, using CAO Alpha imagery. *Remote Sensing of Environment, 115*, 1478-1488

Knox, N.M., Skidmore, A.K., Prins, H.H.T., Heitkönig, I.M.A., 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

Knox, N.M., Skidmore, A.K., Schlerf, M., de Boer, W.F., van Wieren, S.E., van der Waal, C., Prins, H.H.T., & Slotow, R. (2010). Nitrogen prediction in grasses: effect of bandwidth and plant material state on absorption feature selection. *International Journal of Remote Sensing, 31*, 691-704

Koger, C.H., Bruce, L.M., Shaw, D.R., & Reddy, K.N. (2003). Wavelet analysis of hyperspectral reflectance data for detecting pitted morningglory (Ipomoea lacunosa) in soybean (Glycine max). *Remote Sensing of Environment, 86*, 108-119

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

Kokaly, R.F., & Clark, R.N. (1999). Spectroscopic Determination of Leaf Biochemistry Using Band-Depth Analysis of Absorption Features and Stepwise Multiple Linear Regression. *Remote Sensing of Environment, 67*, 267-287

Kokaly, R.F., Despain, D.G., Clark, R.N., & Livo, K.E. (2003). Mapping vegetation in Yellowstone National Park using spectral feature analysis of AVIRIS data. *Remote Sensing of Environment, 84*, 437-456

Kotze, D.C., & O'Connor, G.T. (2000). Vegetation variation within and among palustrine wetlands along an altitudinal gradient in KwaZulu-Natal, South Africa. *Plant Ecology, 146*, 77-96

Kruse, J.K., Christians, N.E., & Chaplin, M.H. (2006). Remote Sensing of Nitrogen Stress in Creeping Bentgrass. *Agronomy Journal, 98*, 1640-1645

Kumar, L., Schmidt, K.S., Dury, S., & Skidmore, A.K. (2001). Review of hyperspectral remote sensing and vegetation science. In F. Van Der Meer (Ed.), *Hyperspectral remote sensing*. Dordrecht: Kluwer Academic Press

Lattanzi, F.A. (2010). C3/C4 grasslands and climate change. *Grassland Science in Europe, Volume 15* (pp. 3-13). Duderstadt: Mecke Druck und Verlag

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

Lee, J.B., Woodyatt, A.S., & Berman, M. (1990). Enhancement of high spectral resolution remote-sensing data by a noise-adjusted principal components transform. *Geoscience and Remote Sensing, IEEE Transactions on, 28*, 295-304

Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R News, 2/3*, 18-22

Lichtenthaler, H., Buschmann, C., Döll, M., Fietz, H.J., Bach, T., Kozel, U., Meier, D., & Rahmsdorf, U. (1981). Photosynthetic activity, chloroplast ultrastructure, and leaf characteristics of high-light and low-light plants and of sun and shade leaves. *Photosynthesis Research, 2*, 115-141

Lillesand, T.M., Kiefer, R.W., & Chipman, J.W. (2008). *Remote Sensing and Image Interpretation*: John Wiley & Sons

Liu, L., & Cheng, Z. (2011). Mapping C3 and C4 plant functional types using separated solar-induced chlorophyll fluorescence from hyperspectral data. *International Journal of Remote Sensing, 32*, 9171-9183

Liu, Y., & Rayens, W. (2007). PLS and dimension reduction for classification. *Journal of Computational Statistics & Data Analysis, 22*, 189-208

Lu, S., Shimizu, Y., Ishii, J., Funakoshi, S., Washitani, I., & Omasa, K. (2009). Estimation of abundance and distribution of two moist tall grasses in the Watarase wetland, Japan, using hyperspectral imagery. *ISPRS Journal of Photogrammetry and Remote Sensing, 64*, 674-682

Lumsden, T., Schulze, R., & Hewitson, B. (2009). Evaluation of potential changes in hydrologically relevant statistics of rainfall in Southern Africa under conditions of climate change. *Water SA, 35*, 649-656

Martens, H., & Naes, T. (1989). *Multivariate Calibration*: Wiley, Chichester, UK

McMahon, S.M., Harrison, S.P., Armbruster, W.S., Bartlein, P.J., Beale, C.M., Edwards, M.E., Kattge, J., Midgley, G., Morin, X., & Prentice, I.C. (2011). Improving assessment and modelling of climate change impacts on global terrestrial biodiversity. *Trends in Ecology & Evolution, 26*, 249-259

McNaughton, S.J. (1990). Mineral nutrition and seasonal movements of African migratory ungulates. *Nature, 345*, 613-615

Melgani, F., & Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing, 42*, 1778−1790

Miglani, A., Ray, S., Pandey, R., & Parihar, J. (2008). Evaluation of EO-1 hyperion data for agricultural applications. *Journal of the Indian Society of Remote Sensing, 36*, 255-266

Mitchell, J.J., Glenn, N.F., Sankey, T.T., Derryberry, D.R., Anderson, M.O., & Hruska, R.C. (2011). Spectroscopic detection of nitrogen concentrations in sagebrush. *Remote Sensing Letters, 3*, 285-294

Morris, C.D., Taintoi, N.M., & Boleme, S. (1993). Classification of the eastern alpine vegetation of Lesotho. *African Journal of Range & Forage Science, 10*, 47-53

Mountrakis, G., Im, J., & Ogole, C. (2010). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing, In Press, Corrected Proof*

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

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

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

Mutanga, O., & Skidmore, A.K. (2004c). Narrow band vegetation indices overcome the saturation problem in biomass estimation. *International Journal of Remote Sensing, 25*, 3999-4014

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

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

Mutanga, O., van Aardt, J., & Kumar, L. (2009). Imaging spectroscopy (hyperspectral remote sensing) in southern Africa: an overview. *South African Journal of Science, 105*, 193-198

Myers, D.A., Vogelmann, T.C., & Bornman, J.F. (1994). Epidermal focussing and effects on light utilization in Oxalis acetosella. *Physiologia Plantarum, 91*, 651-656

Myoung, B., Choi, Y.-S., & Park, S. (2011). A review on vegetation models and applicability to climate simulations at regional scale. *Asia-Pacific Journal of Atmospheric Sciences, 47*, 463-475

Nash, J.E., & Sutcliffe, J.V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology, 10*, 282-290

Neilson, R.P. (1993). Transient Ecotone Response to Climatic Change: Some Conceptual and Modelling Approaches. *Ecological Applications, 3*, 385-395

Nicodemus, K., Malley, J., Strobl, C., & Ziegler, A. (2010). The behaviour of random forest permutation-based variable importance measures under predictor correlation. *BMC Bioinformatics, 11*, 110

Nicodemus, K., & Shugart, Y. (2007). Impact of linkage disequilibrium and effect size on the ability of machine learning methods to detect epistasis in case-control studies. *Abstract volume of the Sixteenth Annual Meeting of the International Genetic Epidemiology Society, North Yorkshire, UK, 31*, 611

Niu, S., Liu, W., & Wan, S. (2008). Different growth responses of C3 and C4 grasses to seasonal water and nitrogen regimes and competition in a pot experiment. *Journal of Experimental Botany, 59*, 1431-1439

Niu, S., Yuan, Z., Zhang, Y., Liu, W., Zhang, L., Huang, J., & Wan, S. (2005). Photosynthetic responses of C3 and C4 species to seasonal water variability and competition. *Journal of Experimental Botany, 56*, 2867-2876

Noble, S.D., Brown, R.B., & Crowe, T.G. (2002). The use of spectral properties for weed detection and identification- a review. In. Mansonville, QC, Canada: CSAE/SCGR.

Novozamsky, I., Houba, V.J.G., van Eck, R., & van Vark, W. (1983). A novel digestion technique for multi-element plant analysis. *Communications in Soil Science and Plant Analysis, 14*, 239-248

Ogle, K. (2003). Implications of interveinal distance for quantum yield in C_4 grasses: a modeling and meta-analysis. *Oecologia, 136*, 532-542

Ollinger, S.V. (2011). Sources of variability in canopy reflectance and the convergent properties of plants. *New Phytologist, 189*, 375-394

Osborne, C.P. (2011). The Geologic History of C4 Plants. In A.S. Raghavendra & R.F. Sage (Eds.), *C_4 Photosynthesis and Related CO_2 Concentrating Mechanisms* (pp. 339-357): Springer Netherlands

Oyarzabal, M., Paruelo, J.M., Pino, F.d., Oesterheld, M., & Lauenroth, W.K. (2008). Trait differences between grass species along a climatic gradient in South and North America. *Journal of Vegetation Science, 19*, 183-192

Ozdemir, I., & Karnieli, A. (2011). Predicting forest structural parameters using the image texture derived from WorldView-2 multispectral imagery in a dryland forest, Israel. *International Journal of Applied Earth Observation and Geoinformation, 13*, 701-710

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing, 26*, 217-222

Pal, M., & Foody, G.M. (2010). Feature selection for classification of hyperspectral data by SVM. *IEEE Transactions on Geoscience and Remote Sensing, 48*, 2297-2307

Pearcy, R.W., & Ehleringer, J. (1984). Comparative ecophysiology of C3 and C4 plants. *Plant, Cell & Environment, 7*, 1-13

Peñuelas, J., Filella, I., Biel, C., Serrano, L., & Savé, R. (1993). The reflectance at the 950–970 nm region as an indicator of plant water status. *International Journal of Remote Sensing, 14*, 1887-1905

Peters, D.P.C. (2002). Plant species dominance at a grassland-shrubland ecotone: an individual-based gap dynamics model of herbaceous and woody species. *Ecological Modelling, 152*, 5-32

Peterson, D.L., Price, K.P., & Martinko, E.A. (2002). Discriminating between cool season and warm season grassland cover types in northeastern Kansas. *International Journal of Remote Sensing, 23*, 5015 - 5030

Poorter, H. (1993). Interspecific variation in the growth response of plants to an elevated ambient CO_2 concentration. *Plant Ecology, 104-105*, 77-97

Portigal, F., Holasek, R., Mooradian, G., Owensby, P., Dicksion, M., Fene, M., Elliot, M., Hall, E., & Driggett, D. (1997). Vegetation classification using red-edge first derivative and green peak statistical moment indices with the Advanced Airborne Hyperspectral Imaging System (AAHIS). In, *Proceedings of the Third International Airborne Remote*

Sensing Conference and Exhibition (pp. 789-797). Copenhagen, Denmark: Ann Arbor, MI: ERIM

Poulter, B., Ciais, P., Hodson, E., Lischke, H., Maignan, F., Plummer, S., & Zimmermann, N.E. (2011). Plant functional type mapping for earth system models. *Geosci. Model Dev., 4*, 993-1010

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

Price, C.J. (1994). How unique are spectral signatures? *Remote Sensing of Environment, 49*, 181-186

Price, J.C. (1992). Variability of high-resolution crop reflectance spectra. *International Journal of Remote Sensing, 13*, 2593-2610

Price, K.P., Guo, X., & Stiles, J.M. (2002). Optimal Landsat TM band combinations and vegetation indices for discrimination of six grassland types in eastern Kansas. *International Journal of Remote Sensing, 23*, 5031-5042

Pu, R., & Landry, S. (2012). A comparative analysis of high spatial resolution IKONOS and WorldView-2 imagery for mapping urban tree species. *Remote Sensing of Environment, 124*, 516-533

Pudil, P., Novovičová, J., & Kittler, J. (1994). Floating search methods in feature selection. *Pattern Recognition Letters, 15*, 1119-1125

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

Purevdorj, T., Tateishi, R., Ishiyama, T., & Honda, Y. (1998). Relationships between percent vegetation cover and vegetation indices. *International Journal of Remote Sensing, 19*, 3519 - 3535

R Development Core Team (2010). R: A language and environment for statistical computing, . In, *R Foundation for Statistical Computing, .* Vienna, Austria R Foundation for Statistical Computing,

Rama Rao, N., Garg, P.K., Ghosh, S.K., & Dadhwal, V.K. (2008). Estimation of leaf total chlorophyll and nitrogen concentrations using hyperspectral satellite imagery. *The Journal of Agricultural Science, 146*, 65-75

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

Ramoelo, A., Skidmore, A.K., Schlerf, M., Mathieu, R., & Heitkönig, I.M.a. (2011). Water-removed spectra increase the retrieval accuracy when estimating savanna grass nitrogen and phosphorus concentrations. *ISPRS Journal of Photogrammetry and Remote Sensing, 66*, 408-417

Reich, P.B., Tilman, D., Craine, J., Ellsworth, D., Tjoelker, M.G., Knops, J., Wedin, D., Naeem, S., Bahauddin, D., Goth, J., Bengtson, W., & Lee, T.D. (2001). Do species and functional groups differ in acquisition and use of C, N and water under varying atmospheric CO_2 and N availability regimes? A field test with 16 grassland species. *New Phytologist, 150*, 435-448

Richter, K., Atzberger, C., Hank, T.B., & Mauser, W. (2012a). Derivation of biophysical variables from Earth observation data: validation and statistical measures. *Journal of Applied Remote Sensing, 6*, 063557-063551

Richter, K., Hank, T.B., Vuolo, F., Mauser, W., & D'Urso, G. (2012b). Optimal Exploitation of the Sentinel-2 Spectral Capabilities for Crop Leaf Area Index Mapping. *Remote Sensing, 4*, 561-582

Ricotta, C., Reed, B.C., & Tieszen, L.T. (2003). The role of C3 and C4 grasses to interannual variability in remotely sensed ecosystem performance over the US Great Plains. *International Journal of Remote Sensing, 24*, 4421 - 4431

Roberts, D., Brown, K., Green, R., Ustin, S., & Hinckley, T. (1998). Investigating the Relationship Between Liquid Water and Leaf Area in Clonal Populus. In, *Proceedings of the Seventh Earth Science Airborne Workshop, Jet Propulsion Laboratory* (pp. 335-344). Pasadena, CA, : JPL Publ.

Rock, B.N., Hoshizaki, T., & Miller, J.R. (1988). Comparison of in situ and airborne spectral measurements of the blue shift associated with forest decline. *Remote Sensing of Environment, 24*, 109-127

Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J.P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing, 67*, 93-104

Ross, J.K. (1981). *The Radiation Regime and Architecture of Plant Stands*. Kluwer Boston: Hingham, MA

Rosso, P.H., Ustin, S.L., & Hastings, A. (2005). Mapping marshland vegetation of San Francisco Bay, California, using hyperspectral data. *International Journal of Remote Sensing, 26*, 5169-5191

Rouse, J.W., Haas, R.H., Schell, J.A., & Deering, D.W. (1973). Monitoring vegetation systems in the Great Plains with ERTS. In, *Proceedings of the Third ERTS Symposium* (pp. 309-317). Washington DC

RSI (2010). ENVI (Environment for Visualizing Images) Online Help. ITT Visual Information Solutions. Boulder, USA. In

Ruffin, C., & King, R.L. (1999). The analysis of hyperspectral data using Savitzky-Golay filtering-theoretical basis. 1. In, *Geoscience and Remote Sensing Symposium, 1999. IGARSS '99 Proceedings. IEEE 1999 International* (pp. 756-758 vol.752)

Rutherford, M.C., Midgley, G.F., Bond, W.J., Powrie, L.W., Musil, C.F., Roberts, R., & Allsopp, J. (1999). South African Country Study on Climate Change. In, *Terrestrial Plant Diversity Section, Vulnerability and Adaptation*. Pretoria: Department of Environmental Affairs and Tourism, Pretoria

Sage, R.F., Christin, P.-A., & Edwards, E.J. (2011). The C4 plant lineages of planet Earth. *Journal of Experimental Botany, doi: 10.1093/jxb/err048*

Sage, R.F., & Pearcy, R.W. (1987). The Nitrogen Use Efficiency of C3 and C4 Plants: II. Leaf Nitrogen Effects on the Gas Exchange Characteristics of Chenopodium album (L.) and Amaranthus retroflexus (L.). *Plant Physiology, 84*, 959-963

Savitzky, A., & Golay, M.J.E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry, 36*, 1627-1639

Schaepman, M.E., Ustin, S.L., Plaza, A.J., Painter, T.H., Verrelst, J., & Liang, S. (2009). Earth system science related imaging spectroscopy--An assessment. *Remote Sensing of Environment, 113*, S123-S137

Schlerf, M., Atzberger, C., & Hill, J. (2005). Remote sensing of forest biophysical variables using HyMap imaging spectrometer data. *Remote Sensing of Environment, 95*, 177-194

Schmidt, K.S., & Skidmore, A.K. (2001). Exploring spectral discrimination of grass species in African rangelands. *International Journal of Remote Sensing, 22*, 3421-3434

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

Schmidt, K.S., Skidmore, A.K., Kloosterman, E.H., van Oosten, H., Kumar, L., & Janssen, J.A.M. (2004). Mapping coastal vegetation using an expert system and hyperspectral imagery. *Photogrammetric Engineering and Remote Sensing, 70*, 703-715

Scholes, R. (2006). Assessments of Impacts and Adaptations to Climate Change. In, *Impacts and Adaptations to Climate Change by the Biodiversity Sector in Southern Africa*. Washington, DC, USA: International START-Secretariat

Schulze, R.E., Lumsden, T.G., Horan, M.J., C, Warburton, M., & Maharaj, M. (2005). An assessment of impacts of climate change on agrohydrological responses over Southern Africa. In R.E. Schulze (Ed.), *Climate Change and Water Resources in Southern Africa: Studies on Scenarios, Impacts, Vulnerabilities and Adaptation* (pp. 141-189). Pretoria, South Africa: Water Research Commission

Schuster, W.S., & Monson, R.K. (1990). An examination of the advantages of C3-C4 intermediate photosynthesis in warm environments. *Plant, Cell & Environment, 13*, 903-912

Serpico, S.B., & Bruzzone, L. (2001). A new search algorithm for feature selection in hyperspectral remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing, 39*, 1360−1367

Serrano, L., Filella, I., & Peñuelas, J. (2000). Remote sensing of biomass and yield of winter wheat under different nitrogen supplies. *Crop Sci., 40*, 723-731

Serrano, L., Peñuelas, J., & Ustin, S.L. (2002). Remote sensing of nitrogen and lignin in Mediterranean vegetation from AVIRIS data: Decomposing biochemical from structural signals. *Remote Sensing of Environment, 81*, 355-364

Shafri, M.Z.H., & Yusof, M.R.M. (2009). Trends and Issues in Noise Reduction for Hyperspectral Vegetation Reflectance Spectra. *European Journal of Scientific Research, 29*, 404-410

Sieben, E., Morris, C., Kotze, D., & Muasya, A. (2010). Changes in plant form and function across altitudinal and wetness gradients in the wetlands of the Maloti-Drakensberg, South Africa. *Plant Ecology, 207*, 107-119

Sims, D.A., & Gamon, J.A. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment, 81*, 337-354

Sinclair, T.R., Hoffer, R.M., & Schreiber, M.M. (1971). Reflectance and Internal Structure of Leaves from Several Crops During a Growing Season1. *Agronomy Journal, 63*, 864-868

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

Slaton, M.R., Hunt, E.R., Jr., & Smith, W.K. (2001). Estimating Near-Infrared Leaf Reflectance from Leaf Structural Characteristics. *American Journal of Botany, 88*, 278-284

Smith, A.M., & Blackshaw, R.E. (2003). Weed–crop discrimination using remote sensing: a detached leaf experiment1. *Weed Technology, 17*, 811-820

Smith, M.O., Ustin, S.L., Adams, J.B., & Gillespie, A.R. (1990). Vegetation in deserts: I. A regional measure of abundance from multispectral images. *Remote Sensing of Environment, 31*, 1-26

Sripada, R.P., Heiniger, R.W., White, J.G., & Meijer, A.D. (2006). Aerial color infrared photography for determining early in-season nitrogen requirements in corn. *Agronomy Journal, 98*, 968-977

Stehman, S.V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment, 62*, 77-89

Stock, W.D., Chuba, D.K., & Verboom, G.A. (2004). Distribution of South African C3 and C4 species of Cyperaceae in relation to climate and phylogeny. *Austral Ecology, 29*, 313-319

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

Strobl, C., Boulesteix, A., & Augustin, T. (2007a). Unbiased Split Selection for Classification Trees Based on the Gini Index. *Computational Statistics & Data Analysis, 52*, 483 - 501

Strobl, C., Boulesteix, A., Zeileis, A., & Hothorn, T. (2007b). Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution. *BMC Bioinformatics, 8*, 25

Strobl, C., & Zeileis, A. (2008). Danger: high Power! – exploring the statistical properties of a test for random forest variable importance. *Proceedings in Computational Statistics, 2*, 59-66

Su, H., Kanemasu, E.T., Ransom, M.D., & Yang, S.-S. (1990). Separability of soils in a tallgrasss prairie using SPOT and DEM data. *Remote Sensing of Environment, 33*, 157-163

Taylor, S., Kumar, L., Reid, N., & Lewis, C.R.G. (2012). Optimal band selection from hyperspectral data for Lantana camara discrimination. *International Journal of Remote Sensing, 33*, 5418-5437

Teeri, J.A., & Stowe, L.G. (1976). Climatic patterns and the distribution of C4 grasses in North America. *Oecologia, 23*, 1-12

Thenkabail, P.S., Enclona, E.A., Ashton, M.S., Legg, C., & De Dieu, M.J. (2004a). Hyperion, IKONOS, ALI, and ETM+ sensors in the study of African rainforests. *Remote Sensing of Environment, 90*, 23-43

Thenkabail, P.S., Enclona, E.A., Ashton, M.S., & Van Der Meer, B. (2004b). Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote Sensing of Environment, 91*, 354-376

Thenkabail, P.S., Smith, R.B., & De Pauw, E. (2000). Hyperspectral Vegetation Indices and Their Relationships with Agricultural Crop Characteristics. *Remote Sensing of Environment, 71*, 158-182

Thulin, S., Hill, M.J., Held, A., Jones, S., & Woodgate, P. (2012). Hyperspectral determination of feed quality constituents in temperate pastures: Effect of processing methods on predictive relationships from partial least squares regression. *International Journal of Applied Earth Observation and Geoinformation, 19*, 322-334

Tieszen, L.L., Reed, B.C., Bliss, N.B., Wylie, B.K., & DeJong, D.D. (1997). NDVI, C3 and C4 Production, and Distributions in Great Plains Grassland Land Cover Classes. *Ecological Applications, 7*, 59-78

Tieszen, L.L., Senyimba, M.M., Imbamba, S.K., & Troughton, J.H. (1979). The distribution of C3 and C4 grasses and carbon isotope discrimination along an altitudinal and moisture gradient in Kenya. *Oecologia, 37*, 337-350

Toutin, T., Schmitt, C.V., & Wang, H. (2012). Impact of no GCP on elevation extraction from WorldView stereo data. *ISPRS Journal of Photogrammetry and Remote Sensing, 72*, 73-79

Townsend, P.A., Foster, J.R., Chastain, R.A., Jr., & Currie, W.S. (2003). Application of imaging spectroscopy to mapping canopy nitrogen in the forests of the central Appalachian Mountains using Hyperion and AVIRIS. *Geoscience and Remote Sensing, IEEE Transactions on, 41*, 1347-1354

Trenholm, L.E., Schlossberg, M.J., Lee, G., Parks, W., & Geer, S.A. (2000). An evaluation of multi-spectral responses on selected turfgrass species. *International Journal of Remote Sensing, 21*, 709 - 721

Treydte, A.C., Heitkönig, I.M.A., & Ludwig, F. (2009). Modelling ungulate dependence on higher quality forage under large trees in African savannahs. *Basic and Applied Ecology, 10*, 161-169

Trollope, W.S.W. (1990). Development of a technique for assessing veld condition in the Kruger National Park using key grass species. *Journal of the Grassland Society of Southern Africa, 7*, 46-51

Ueno, O., Kawano, Y., Wakayama, M., & Takeda, T. (2006). Leaf Vascular Systems in C3 and C4 Grasses: A Two-dimensional Analysis. *Annals of Botany, 97*, 611-621

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

Ustin, S.L., & Gamon, J.A. (2010). Remote sensing of plant functional types. *New Phytologist, 186*, 795-816

van der Meer, F., Jong, S.D., & Bakker, W. (2001). *Imaging spectrometry: basic principles and prospective applications*: Kluwer Academic Publishers

Vandewalle, M., de Bello, F., Berg, M., Bolger, T., Dolédec, S., Dubs, F., Feld, C., Harrington, R., Harrison, P., Lavorel, S., da Silva, P., Moretti, M., Niemelä, J., Santos, P., Sattler, T., Sousa, J., Sykes, M., Vanbergen, A., & Woodcock, B. (2010). Functional traits as

indicators of biodiversity response to land use changes across ecosystems and organisms. *Biodiversity and Conservation, 19*, 2921-2947

Vapnik, V.N. (1995). *The nature of statistical learning theory*. New York: Springer

Vapnik, V.N. (1998). Statistical learning theory. John Wiley and Sons Inc.

Verhoef, W. (1984). Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model. *Remote Sensing of Environment, 16*, 125-141

Vetter, S. (2005). Rangelands at equilibrium and non-equilibrium: recent developments in the debate. *Journal of Arid Environments, 62*, 321-341

Vogelmann, T.C. (1993). Plant Tissue Optics. *Annual Review of Plant Physiology and Plant Molecular Biology, 44*, 231-251

Vogelmann, T.C., & Martin, G. (1993). The functional significance of palisade tissue: penetration of directional versus diffuse light. *Plant, Cell & Environment, 16*, 65-72

von Fischer, J.C., Tieszen, L.L., & Schimel, D.S. (2008). Climate controls on C_3 vs. C_4 productivity in North American grasslands from carbon isotope composition of soil organic matter. *Global Change Biology, 14*, 1141-1155

Vrindts, E., De Baerdemaeker, J., & Ramon, H. (2002). Weed Detection Using Canopy Reflection. *Precision Agriculture, 3*, 63-80

Vyas, D., Krishnayya, N.S.R., Manjunath, K.R., Ray, S.S., & Panigrahy, S. (2011). Evaluation of classifiers for processing Hyperion (EO-1) data of tropical vegetation. *International Journal of Applied Earth Observation and Geoinformation, 13*, 228-235

Walther, G.-R. (2010). Community and ecosystem responses to recent climate change. *Philosophical Transactions B, 365*, 2019-2024

Walther, G.-R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T.J.C., Fromentin, J.-M., Hoegh-Guldberg, O., & Bairlein, F. (2002). Ecological responses to recent climate change. *Nature, 416*, 389-395

Wand, S.J.E., Midgley, G.F., Jones, M.H., & Curtis, P.S. (1999). Responses of wild C4 and C3 grass (Poaceae) species to elevated atmospheric CO_2 concentration: a meta-analytic test of current theories and perceptions. *Global Change Biology, 5*, 723-741

Wang, C., Jamison, B.E., & Spicci, A.A. (2010). Trajectory-based warm season grassland mapping in Missouri prairies with multi-temporal ASTER imagery. *Remote Sensing of Environment, 114*, 531-539

Wang, C., Menenti, M., Stoll, M.-P., Belluco, E., & Marani, M. (2007). Mapping mixed vegetation communities in salt marshes using airborne spectral data. *Remote Sensing of Environment, 107*, 559-570

Wessman, C.A., Bateson, C.A., & Benning, T.L. (1997). DETECTING FIRE AND GRAZING PATTERNS IN TALLGRASS PRAIRIE USING SPECTRAL MIXTURE ANALYSIS. *Ecological Applications, 7*, 493-511

Williams, A.E.P., & Hunt, E.R., Jr. (2004). Accuracy Assessment for Detection of Leafy Spurge with Hyperspectral Imagery. *Journal of Range Management, 57*, 106-112

Winslow, J.C., Hunt Jr, E.R., & Piper, S.C. (2003). The influence of seasonal water availability on global C3 versus C4 grassland biomass and its implications for climate change research. *Ecological Modelling, 163*, 153-173

Wold, S., Sjostrom, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems, 58*, 109-130

Wylie, B.K., Johnson, D.A., Laca, E., Saliendra, N.Z., Gilmanov, T.G., Reed, B.C., Tieszen, L.L., & Worstell, B.B. (2003). Calibration of remotely sensed, coarse resolution NDVI to CO_2 fluxes in a sagebrush-steppe ecosystem. *Remote Sensing of Environment, 85*, 243-255

Xiaolin, W., & Memon, N. (2000). Context-based lossless interband compression-extending CALIC. *IEEE Transactions on Geoscience and Remote Sensing, 9*, 994-1001

Yang, F., Li, J., Gan, X., Qian, Y., Wu, X., & Yang, Q. (2010). Assessing nutritional status of Festuca arundinacea by monitoring photosynthetic pigments from hyperspectral data. *Computers and Electronics in Agriculture, 70*, 52-59

Ye, X., Sakai, K., Sasao, A., & Asada, S. (2008). Potential of airbornre hyperspectral imagery to estimate fruit yield in citrus. *Chemometrics and Intelligent Laboratory Systems, 90*, 132-144

Zengeya, F.M., Mutanga, O., & Murwira, A. (2012). Linking remotely sensed forage quality estimates from WorldView-2 multispectral data with cattle distribution in a savanna landscape. *International Journal of Applied Earth Observation and Geoinformation, In Press:10.1016/j.jag.2012.07.008*

Zhang, J., Wu, J., & Zhou, L. (2011). Deriving vegetation leaf water content from spectrophotometric data with orthogonal signal correction-partial least square regression. *International Journal of Remote Sensing, 32*, 7557–7574

Zwiggelaar, R. (1998). A review of spectral properties of plants and their potential use for crop/weed discrimination in row-crops. *Crop Protection, 17*, 189-206