Exploring the relationship between spectral reflectance and tree species diversity in the savannah woodlands

Sabelo Madonsela

208506831

PhD Thesis

Submitted in fulfilment of the academic requirements for the Doctor of Philosophy degree in the School of Agricultural, Earth and Environmental sciences, University of KwaZulu-Natal

Declaration 1

This research work was conducted in fulfilment of the academic requirements for the Doctor of Philosophy degree in the School of Agricultural, Earth and Environmental sciences, University of KwaZulu-Natal and represents the original work of the first author. Any research work sourced from other researchers and organization is duly acknowledged in text and in the reference lists.

…………………………………………..

Student: Mr Sabelo Madonsela

…………………………………………….

Supervisor: Prof Onisimo Mutanga

………………………………………………

Co-supervisor: Prof Moses A. Cho

………………………………………………

Co-supervisor: Dr Abel Ramoelo

Declaration 2

The following research papers published or in preparation form part of this thesis and were developed and written by the first author under the supervision of the second, third and fourth authors. The fifth authors had assisted the first author with the application of woody canopy cover and radiative transfer models respectively.

Madonsela, S., Cho, M.A., Ramoelo, A. and Mutanga, O., (2017). Remote sensing of species diversity using Landsat 8 spectral variables. *ISPRS Journal of Photogrammetry and Remote Sensing Vol.* 133, pp. 116-127.

Madonsela, S., Cho, M.A., Ramoelo, A., Mutanga, O. and Naidoo, L., (2018). Estimating tree species diversity in the savannah using NDVI and woody canopy cover. *International Journal of Applied Earth Observation and Geoinformation Vol. 66, pp. 106-115.*

Madonsela, S., Cho, M.A., Ramoelo, A., Mutanga, O. and Masemola, C., (*in preparation*). Estimating tree species diversity using biophysical and biochemical properties derived from the inversion of radiative transfer models.

Madonsela, S., Cho, M.A., Ramoelo, A. and Mutanga, O., (*in preparation*). Exploring the relationship between spectral variability and tree species diversity in the savannah woodlands: a case of Spectral Angle Mapper as a new technique for Spectral Variation Hypothesis.

………………………………………………

Student: Mr Sabelo Madonsela

Table of Content

Table of Figures

List of Tables

Abstract

Remote sensing applications in biodiversity research often rely on the establishment of relationships between spectral information and species diversity. Terrestrial plant species are known to vary in their spectral reflectance due to differences in biophysical and biochemical properties including leaf pigment, water content and plant internal structure. These plant properties determine spectral reflectance in the visible, near infrared and shortwave infrared regions. This study aimed to explore the relationship between spectral reflectance and tree species diversity in the savannah woodlands of South Africa.

The study used Landsat-8 spectral variables, woody canopy cover and Radiative Transfer Models derivatives as predictor variables. Meanwhile, common local diversity indices i.e. Shannon index, Simpson index and species richness were used to quantify tree species diversity. Eventually the study applied linear, multivariate and factorial regression models to explore relationships between predictor variables and tree species diversity indices.

Overall the study made four key findings which are relevant to the application of remote sensing in assessing tree species diversity in the southern African savannah. The first finding, which is supported by other studies in different ecosystems, was that remotely sensed productivity indices (e.g. Normalized Difference Vegetation Index [NDVI]) capture the long established productivity-diversity relationship in ecology and therefore can be used to estimate tree species diversity in the savannah woodlands. At the end of the growing season in particular, the results showed that the factorial model based on NDVI and woody canopy cover, which is a proxy for woodland productivity, enhanced the prediction of tree species diversity in the savannah woodlands (r^2 of 0.38; $p < 0.001$).

The second main finding was that spectral variability is related to the diversity of tree species in the savannah woodlands (r^2 of 0.24; $p < 0.05$). The implication of this finding is that the remote sensing model based on spectral variability would be more suitable for modelling tree species diversity in the savannah woodlands particularly at the end of the growing season. This study developed an innovative method to apply Spectral Angle Mapper (SAM) as a measure of spectral variability to extract essential spectral information from the Landsat-8 data and use it to explain tree species diversity in light of Spectral Variation Hypothesis (SVH).

The third finding was that phenology affects the ability of remote sensing to capture spectral variability related to tree species diversity. This means that the implementation of SVH should target a particular phenological period to increase prediction accuracy. In this study the end of the growing season was the most optimal period to use spectral variability to explain tree species diversity in the savannah woodlands. The fourth finding was that radiative transfer modelling has the potential to retrieve useful information about biophysical and biochemical parameters that can be used to explain tree species diversity in the savannah woodlands.

The study recommends further research on the inversion of radiative transfer modelling (RTM) for the retrieval of biochemical and biophysical parameters on high resolution images capable of detecting fine details on the canopy structure and biochemistry. Moreover, SAM was applied for the first time as a multivariate technique and the mean SAM results were satisfactory, particularly at the end of the growing season. Future research should test this technique on high spectral resolution data where reflectance signals could be associated with tree canopies.

Acknowledgements

I would like to humbly express my gratitude to the following institutions, researchers, colleagues and friends for their contribution in the successful execution of the research work presented in this thesis.

- Many thanks to the Council for Scientific and Industrial Research for offering me a PhD studentship and providing the necessary human resources and research facilities. It has been very advantageous being in this institution in terms of research skill development.
- Many thanks to the National Research Foundation for funding my PhD studies through the Professional Development Programme.
- Many thanks to Ezemvelo KwaZulu-Natal Wildlife and South Africa's National Park for granting me access to work in Hluhluwe-Imfolozi Park and Kruger National Park respectively.
- I would like to thank Professor Moses A. Cho for his guidance and supervision throughout the development of research ideas, execution of those ideas and the write-up of this thesis. I consider myself the main beneficiary of your unstated open-door policy. I met Professor Cho in 2011 while collecting field data in the Dukuduku forest for my Honours research project and that's where he saw a potential in me; thank you Professor Cho for believing in me. I will certainly not forget this advice 'Sabelo! You cannot solve all problems. Just identify few research questions and give it your best.' With this thesis I believe I have made my contribution in the application of remote sensing in biodiversity research. Lastly, thanks for sharing *Desiderata* with me; it always comes handy in testing times.
- I would also like to thank Dr Abel Ramoelo for his guidance in the development of research ideas, execution of those ideas and the write-up of this thesis. Those meetings we had in Professor Moses's office and the one-on-one in Kruger National Park helped me unlock my thinking about the applications of remote sensing. While on our way to the field in Eastern Cape, Dr Ramoelo challenged me to consider doing a PhD; it would be a lie to say I liked that challenge but post the completion of MSc work I took up a PhD studentship and he was available to assist me. Thank you Dr. Ramoelo for your commitment towards human capital development.

- Many thanks to Professor Onisimo Mutanga for his supervision of the overall research progress and comments on each chapter that helped improved the quality of this thesis. Professor encouraged me to consider furthering my research career while I was doing Honours degree and facilitated my transfer to the titanic research organization, the CSIR. Thanks Prof! Lastly, it has been a pleasure working with you.
- I would also like to thank Dr Renaud Mathieu for making woody canopy cover map available and the chats we had over my research work. Thank you!
- Many thanks to colleagues Dr Laven Naidoo, Nokuthula Mahlangu and Cecilia Masemola for the chats we had over my research work.
- I would also like to Thokozani Mntungwa for organizing field work logistics and accommodation.
- Many thanks Cecilia Masemola for assisting me with field data collection; it was not easy going through the thorny canopies of savannah and sometimes getting lost. Thank you for your patience. Thanks to Nobuhle Majozi, Phumlani Zwane, Sibonelo Gumede, Joseph Dlamini, Martin Serela, Veli Ndlovu for your assistance while in the field.
- I would like thank Dr Timithy Dube for the motivation through his success and the occasional chats we had over research career.
- May I also thank the following close friends of mine who provided moral support and cared to listen to my PhD narrative; Tenele Sibandze, Lerato Mpheshea and Elelwani Musekwa. Many thanks to Oupa Malahlela for support and the occasional chats over our careers.
- Many thanks to Reyhana Mahomed for correcting grammatical errors in this thesis.
- Many thanks to CSIR cleaning staff for keeping my office clean.

Chapter 1: General introduction

1.1. Background

Understanding the spatial variation in species richness and experimenting with environmental parameters required to predict the distribution of species and levels of richness has been a topical research question (Kerr and Ostrovsky, 2003; Turner et al. 2003; Reyers and McGeoch, 2007). This research direction was prompted by the growing extent and increasing intensity of human activities that resulted in the modification of the environment, loss of habitats and biodiversity (Kerr and Ostrovsky, 2003; Reyers and McGeoch, 2007). Parallel to other ecosystems on earth, African savannahs are experiencing high pressures from human activities and the burden remains with conservation authorities to maintain biodiversity and ensure sustainable use of biodiversity resources (Asner et al. 2009). Tree species richness, abundance and distribution in the savannah landscapes are impacted upon by land use conversions, e.g. to urban or agricultural lands, (Schlesinger et al. 2015), land management decisions (Wessels et al. 2011; Nacoulma et al. 2011), disturbance regimes, e.g. fire, herbivory (Shackleton et al. 1994; Mudongo et al. 2016), and climate change (Stevens et al. 2014).

In this regard, South Africa's National Parks have established a Threshold of Potential Concerns (TPCs) which serves as a monitoring system to identify changes that may impact on key elements of biodiversity (Gillson and Duffin, 2007; Druce et al. 2008). However, the accomplishment of such monitoring systems rests on the availability of spatially detailed and updated information on the distribution patterns and abundance of species (Turner et al. 2003). This places a high necessity for data and techniques that may enable ecologists to proactively detect changes over a large spatial area (Kerr and Ostrovsky, 2003; Ustin and Gamon, 2010). Remote sensing meets these needs as it collects data over large geographical areas on a regular interval and at varying levels of spatial details (Jetz et al. 2016; Kerr and Ostrovsky, 2003; Turner et al. 2002). More interestingly, ecologists have recognized the need to embrace remote sensing science in order to study biodiversity and prepare conservation responses that are commensurate with the scale of conservation (Jetz et al. 2016; Pereira et al. 2013).

The application of remote sensing in biodiversity research is generally based on the analysis of spectral reflectance to establish the relationship between spectral information on the image and landscape tree species diversity (Gould, 2000; Palmer et al. 2002; Foody and

Cutler, 2003; Rocchini, 2007; Parviainen et al. 2010). The analysis of remotely sensed data and inferences drawn thereof should always display concurrence with field measured data for it to be accepted into mainstream ecology and conservation. Therefore, field data remains essential for validating and testing the reliability of the results from remote sensing data (Kerr and Ostrovsky, 2003). To summarize, this approach combines remote sensing with field data in an effort to advance biodiversity research through the provision of spatially-explicit data.

However, a large number of studies explaining species diversity using remotely sensed data have exploited only the red and near-infrared regions of the electromagnetic spectrum through the application of normalized difference vegetation index (NDVI) (Gould, 2000; Oindo and Skidmore, 2002; Fairbanks and McGwire, 2004; He et al. 2009; Parviainen et al. 2010; Pau et al. 2012). The NDVI has been shown to be sensitive to the amount of available energy in an ecosystem detectable as primary productivity which defines variation in plant diversity (Oindo and Skidmore, 2002; Wang et al. 2004; Parviainen et al. 2010; Witman et al. 2008). It is therefore not unexpected that NDVI has often been successful in estimating tree species diversity in different biomes at various scales (Gould, 2000; Oindo and Skidmore, 2002; Parviainen et al. 2010; Pau et al. 2012). In the savannah, the NDVI signal is influenced by woody canopy foliage and grasses. The question is whether tree species diversity is more related to woody canopy cover (i.e. a proxy for woodland productivity) or to the entire productivity of the system i.e. including trees and grass represented by the NDVI?

Despite its success, NDVI only uses the subset of the available spectral bands in the electromagnetic spectrum. The Landsat program, for instance, collects essential spectral information in the visible, near infrared (NIR) and shortwave infrared (SWIR) regions which relates to plant properties including leaf pigment, water content and plant internal structure (Hernandez-Stefanoni et al. 2012; Nagendra et al. 2010). In addition, studies have emphasized the role of spectral information in the shortwave infrared bands when explaining species diversity (Rocchini, 2007) and floristic structure (Thenkabail et al. 2003) using Landsat data. Therefore, it is necessary to advance multivariate techniques capable of exploiting spectral information across the visible, NIR and SWIR regions of the electromagnetic spectrum when explaining tree species diversity using remotely sensed data.

In recent times, Spectral Variation Hypothesis (SVH) has gained momentum as a multi-dimensional analytical approach to establish the relationship between spectral variability on the image and tree species diversity measured on the ground (Rocchini et al. 2004; Rocchini, 2007; Oldeland et al. 2010; Hernández-Stefanoni et al. 2012). The development of SVH follows Palmer's (2002) assertion that spectral variability on the image stems from the spatial heterogeneity of the environment which by default has high species diversity due to the higher number of available niches. Therefore, spectral variability on the image should be considered as a proxy for species diversity (Palmer et al. 2002; Rocchini, 2007; Rocchini et al. 2010). In terms of techniques for implementing SVH, progress has been slow with most studies (Rocchini et al. 2010; Rocchini, 2007; Oldeland et al. 2010) frequently using mean distance from spectral centroid as a measure of spectral variability.

This study will investigate the possibility of using Spectral Angle Mapper (SAM) as a new technique for implementing SVH. SAM is a mathematical algorithm that has been used to selects bands which increase spectral angle between target species (Keshava, 2004; Cho et al. 2010). In the context of SVH high spectral angle between Landsat-8 pixels indicates high environmental heterogeneity which in turn is linked to high tree species diversity. The primary objective is to explore the relationship between spectral variability from Landsat-8 imagery and tree species diversity in the savannah woodlands.

Moreover, Radiative Transfer Modelling (RTM) also presents an opportunity to exploit the entire spectral content from the visible, NIR and SWIR regions of the electromagnetic spectrum for accurate retrieval of biophysical and biochemical parameters such as chlorophyll a and b (*Cab*), leaf area index (LAI) and equivalent water thickness (EWT). For instance Ceccato et al. (2001) observed that the reflectance in the SWIR is not only controlled by EWT but also dry matter and leaf internal structure. Therefore the accurate retrieval of EWT requires a combination of spectral information from the NIR and SWIR. Studies have used different radiative transfer models to retrieve LAI and chlorophyll and estimate fraction of photosynthetically active radiation absorbed by vegetation (Myneni et al. 1997; Asner et al. 1998; Myneni et al. 2002; Zhang et al. 2005). However, to our knowledge no study has tested the utility of RTM derivatives for estimating tree species diversity particularly in African savannahs. Although biophysical and biochemical parameters have been shown to be useful for characterizing tree species diversity in the

Hawaiian tropical dry forest (Carlson et al. 2007) and in the sub-tropical forest of China (Zhao et al. 2016), this absence persists. Notable in the present and the two preceding paragraphs is that the visible, NIR and SWIR regions all contain essential spectral information for explaining tree species diversity and therefore multivariate techniques will be needed to facilitate utilization of spectral information from these regions. The overall aim of this study was to explore the relationship between spectral reflectance and tree species diversity in the savannah woodland belt. The specific research objectives are discussed below.

1.2. Research objectives

- To explore the utility of spectral information across Landsat-8 spectrum for estimating alpha diversity (α-diversity) in the savannah woodland in southern Africa.
- To investigate and define the species diversity index (i.e. Shannon [H'], Simpson [D_2] and species richness [S]) that best relates to spectral variability on the Landsat-8 Operational Land Imager dataset.
- To investigate whether there is a significant interaction between seasonal NDVI and woody cover when estimating tree species diversity.
- To investigate whether there is a significant relationship between woody canopy cover and tree species diversity across savannah woodland belt.
- To investigate the possibility of using RTM derivatives for estimating tree species diversity in the savannah woodlands. Firstly we simulate canopy reflectance using PROSAIL and INFORM, retrieve EWT, Cab, LAI, dry matter, Cbrown and then use these derivatives as predictor variables in the regression models to estimate species diversity.
- To investigate the possibility of using Spectral Angle Mapper (SAM) as a new measure of spectral variability and explore the relationship between spectral variability and tree species diversity in the savannah woodlands.

1.3. Study area

The study area stretches across the KwaZulu-Natal (KZN), Mpumalanga and Limpopo provinces of South Africa, covering the savannah woodland belt (**Figure 1**). The area is divided into two land management regimes i.e. communal areas and protected areas (Kruger National Park, Hluhluwe-Imfolozi Park and other private nature reserves) with differing land use practices. High tree species diversity has been noted in both areas (du Toit et al. 2003; Shackleton, 2000; Dumalisile, 2009).

1.4. General Methods

The study applied linear, multivariate and factorial regression models to explore relationships between Landsat-8 spectral variables and tree species diversity indices. The study also used Radiative Transfer Models to retrieve biophysical and biochemical variables from Landsat-8 data. Common local diversity indices i.e. Shannon index, Simpson index and species richness were used to quantify tree species diversity. The investigations were conducted using Landsat-8 imagery, woody canopy cover derived from Synthetic Aperture Radar (SAR) and field measured tree species data.

1.5. Thesis Outline

The main body of the thesis contains four technical chapters which are presented in a paper format. This means that each of the four chapters has its own abstract, introduction, methods, results, discussion, conclusion sections and the reference list. The reference list for the general introduction and concluding chapters is placed at the end of the thesis. Chapter 2 investigates the utility of spectral information across Landsat-8 spectrum for estimating alpha diversity (α-diversity) in the savannah woodland in southern Africa and also define the species diversity index (i.e. Shannon [H’], Simpson [D2] and species richness [S]) that best relates to spectral variability on the Landsat-8 Operational Land Imager dataset. Chapter 3 investigates whether there is a significant interaction between seasonal NDVI and woody cover when estimating tree species diversity and also investigates whether there is a significant relationship between woody canopy cover and tree species diversity across savannah woodland belt. Chapter 4 investigates the possibility of using RTM derivatives for estimating tree species diversity in the savannah woodlands. Chapter 5 investigates the possibility of using Spectral Angle Mapper (SAM) as a new measure of spectral variability and explore the relationship between spectral variability and tree species diversity in the savannah woodlands. Chapter 6 is the concluding chapter which presents the synthesis of the observations made in the above chapters.

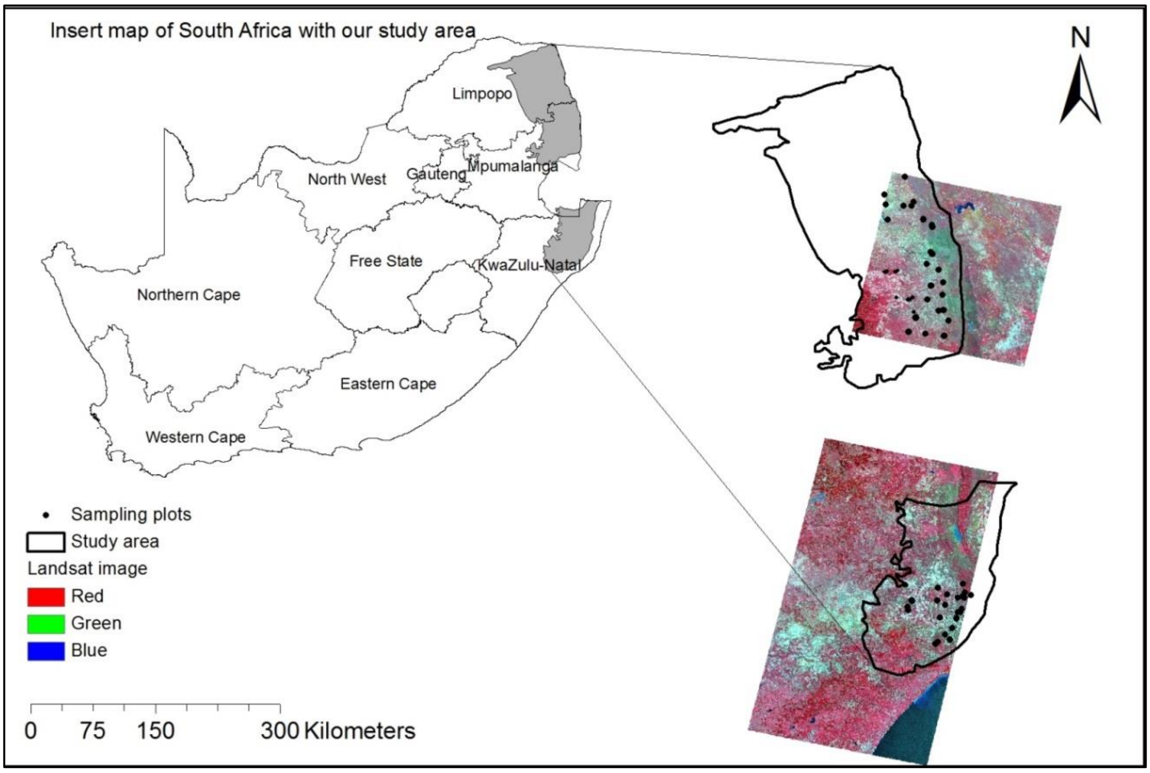

Figure 1 Study area stretching across three provinces of South Africa. Dots on the Landsat imagery are the sampling plots.

Chapter 2: Exploring the utility of spectral information across Landsat-8 spectrum using the Principal Component Analysis (PCA) and estimate alpha diversity (α-diversity) in the savannah woodland in southern Africa

This chapter is based on

Madonsela, S., Cho, M.A., Ramoelo, A. and Mutanga, O., (2017). Remote sensing of species diversity using Landsat 8 spectral variables. *ISPRS Journal of Photogrammetry and Remote Sensing Vol.* 133, pp. 116-127.

.

Abstract

The application of remote sensing in biodiversity estimation has largely relied on the Normalized Difference Vegetation Index (NDVI). The NDVI exploits spectral information from red and near infrared bands of Landsat images and it does not consider canopy background conditions hence it is affected by soil brightness which lowers its sensitivity to vegetation. As such NDVI may be insufficient in explaining tree species diversity. Meanwhile, the Landsat program also collects essential spectral information in the shortwave infrared (SWIR) region which is related to plant properties. The study was intended to: i) explore the utility of spectral information across Landsat-8 spectrum using the Principal Component Analysis (PCA) and estimate alpha diversity (α-diversity) in the savannah woodland in southern Africa, and ii) define the species diversity index (Shannon (H'), Simpson (D_2) and species richness (S) – defined as number of species in a community) that best relates to spectral variability on the Landsat-8 Operational Land Imager dataset. We designed 90m X 90m field plots (n=71) and identified all trees with a diameter at breast height (DbH) above 10cm. H', D_2 and S were used to quantify tree species diversity within each plot and the corresponding spectral information on all Landsat-8 bands were extracted from each field plot. A stepwise linear regression was applied to determine the relationship between species diversity indices (H', D_2 and S) and Principal Components (PCs), vegetation indices and Gray Level Co-occurrence Matrix (GLCM) texture layers with calibration (n=46) and test (n=23) datasets. The results of regression analysis showed that the Simple Ratio Index derivative had a higher relationship with H', D_2 and S (r^2=0.36; r^2=0.41; r^2=0.24 respectively) compared to NDVI, EVI, SAVI or their derivatives. Moreover the Landsat-8 derived PCs also had a higher relationship with H' and D_2 (r^2 of 0.36 and 0.35 respectively) than the frequently used NDVI, and this was attributed to the utilization of the entire spectral content of Landsat-8 data. Our results indicate that: i) the measurement scales of vegetation indices impact their sensitivity to vegetation characteristics and their ability to explain tree species diversity; ii) principal components enhance the utility of Landsat-8 spectral data for estimating tree species diversity and iii) species diversity indices that consider both species richness and abundance (H' and D_2) relates better with Landsat-8 spectral variables.

2.1. Introduction

The savannah biome is characterized by the co-existence of trees and herbaceous vegetation (Scholes and Archer, 1997) and it hosts a large number of floral and faunal diversity (du Toit et al. 2003). Importantly, tree diversity serves many ecological functions in the savannah, e.g. providing habitats and nesting sites to diverse avifaunal species (Dean et al. 1999; Seymour and Dean, 2010); facilitating grass growth and improving grass quality beneath their canopies (Ludwig et al., 2004; Treydte et al. 2007); and serving as food resources to many browsing faunal species (Hempson et al. 2015). Nonetheless, the diversity, abundance and distribution of savannah tree species are impacted by disturbances e.g. the effect of elephants in protected areas (Druce et al. 2008), harvesting for fuelwood (Madubansi and Shackleton, 2006; Matsika et al. 2012) and land use conversion (Schlesinger et al. 2015). Therefore, monitoring the distribution patterns and diversity of tree species remains essential to ensure that disturbances are within the resilience capacity of the ecosystem (Druce et al. 2008). However, the absence of large scale information on tree species distribution upon which management decisions can be based in the African savannah presents a challenge (Asner et al. 2009). The success of any biodiversity monitoring effort depends on the availability of up-to-date and spatially detailed assessments of species richness and distribution at a regional scale (Turner et al. 2003). Space-borne remote sensing meet these needs as it covers large geographical areas on a regular interval and at varying levels of spatial details (Jetz et al. 2016; Kerr and Ostrovsky, 2003). More interestingly, ecologists are recognizing the need to move beyond traditional field-based ecology and embrace remote sensing science in order to prepare conservation responses that are commensurate with the scale of conservation (Jetz et al. 2016; Pereira et al. 2013).

The success of remote sensing applications in biodiversity research hinges more on the spectral resolution of data than spatial resolution (Rocchini et al. 2007; Nagendra et al. 2010). Thenkabail et al. (2003) observed that differences in forest characteristics are better explained by the six bands of Landsat Enhanced Thematic Mapper plus than the four bands of IKONOS data. They attributed 20% of the variability explained by Landsat to two shortwave infrared bands not present in IKONOS. Essentially, the Landsat program collects essential spectral information in the visible, near infrared (NIR) and shortwave infrared

(SWIR) regions which relates to plant properties including leaf pigment, water content and plant internal structure (Hernandez-Stefanoni et al. 2012; Nagendra et al. 2010). As a result, Landsat data performed higher than the high resolution multispectral IKONOS data when estimating forest characteristics.

Nonetheless, most studies have only exploited the red and NIR bands by using normalized difference vegetation index to study species diversity. For instance, Gould (2000) extracted variability from the NDVI image to estimate species richness in the Hood River, central Canadian Arctic. This study excluded non-positive values in the NDVI image to eliminate outliers in the analysis and observed positive correlation between variation on the NDVI image and the species richness. Fairbanks and McGwire (2004) used multi-temporal NDVI to estimate plant species richness in California, USA. They also observed positive relationship with species richness, and attributed it to NDVI sensitivity to abiotic factors impacting species richness. However, Oindo and Skidmore (2002) observed a negative correlation between maximum average NDVI and species richness in Kenya, while NDVI variability had a positive correlation. Meanwhile Parviainen et al. (2010) concluded that using NDVI along with its derivatives produced the best models for estimating species richness in the boreal landscapes. The use of spectral vegetation indices such as NDVI ensures that spectral variability extracted from each plot is mainly due to vegetation characteristics (Viña et al. 2011). It is therefore not surprising that variation in NDVI has been positively related to species diversity.

Whilst the aforementioned studies using NDVI have reported a positive relationship with species diversity, the limitations of NDVI might have suppressed the full extent of landscape variability. NDVI does not consider canopy background conditions hence it is affected by soil brightness which lowers its sensitivity to vegetation (Huete and Jackson, 1988). Moreover, NDVI often shows scaling problem and it saturates in areas of high biomass (Huete et al. 2002; Gitelson, 2004; Main et al. 2011) and may therefore not be sufficient as a means to explain spatial variation in tree species diversity. Meanwhile, enhanced vegetation index (EVI) and simple ratio index (SRI) are not limited to a scale of 0 to 1, and EVI in particular considers canopy background conditions (Huete et al. 2002). This generates the assumption that they might be useful for estimating tree species diversity in semi-arid biome such as savannah. In addition, the mere 30% variation in woody species richness explained by NDVI in Hawaiian dry forests (Pau et al. 2012) bears evidence to the need to move beyond red and NIR bands and explore the utility of the entire spectrum (visible, NIR and SWIR) for estimating tree species diversity.

Moreover, research on the application of remote sensing in biodiversity estimation has frequently relied on univariate regression analysis with limited input variables in terms of predictors (Gould, 2000; Oindo and Skidmore, 2002). Univariate analysis does not fully explore the utility of spectral information content of the remotely sensed image. Despite the limitations of univariate analysis, little has been done to explore the capabilities of multivariate regression models particularly in the African savannah. Multivariate regression analysis presents an opportunity to benefit from the entire spectrum of remote sensing data as more information is analysed simultaneously. Unlike the Landsat derived NDVI which uses only the red and NIR bands, multivariate techniques extract spectral information across the entire spectral regions (the visible, NIR and SWIR) and produce few, uncorrelated principal components which contains all the variability from the original dataset (Jongman et al. 1995; Bro and Smilde, 2014). It is therefore expected that multivariate analysis will demonstrate the utility of satellite remote sensing as a source of information for estimating tree species diversity.

Whilst remote sensing applications in biodiversity estimation has been increasing, minimal attention has been directed to the sensitivity of diversity indices to species distributional patterns. Several studies including Pau et al. (2012); Parviainen et al. (2010) and Gould, (2000) used species richness as a measure of tree species diversity. Species richness only conveys information about the total number of species in a community without due regard to species evenness and abundance (Colwell, 2009). Evenness and abundance relay information regarding the distributional patterns of tree species and thus better reflect the spatial heterogeneity of the landscape (Colwell, 2009). Oldeland et al. (2010) have shown that species abundance has a bearing on the spectral signal captured by the sensor. It is therefore essential that the diversity index used is sensitive to aspects of diversity that impact on the spectral reflectance captured by the remote sensing device. Shannon and Simpson diversity indices both consider richness and evenness (Colwell, 2009; Nagendra, 2002), yet their application with remote sensing data in the African savannah have only been limited to a study by Oldeland et al. (2010).

The two indices have different response to species richness and abundance. The Simpson index is generally influenced by the abundance in the distribution of tree species, while the Shannon index is equally sensitive to both species abundance and rarity of species (Morris et al. 2014). Nonetheless the two indices convey structural information regarding landscape species diversity in terms of dominance and distribution patterns (Morris et al. 2014). The fundamental research question is how does spectral reflectance captured by the Landsat sensor relate to species richness and abundance? The question is of ecological significance as it seeks to advance our ability to estimate spatial patterns of tree species diversity through remote sensing. The aim of the study was to: i) test the assumption that SRI and EVI - which considers canopy background conditions and have a linear relationship with biophysical characteristics of vegetation - might explain tree species diversity better than NDVI; ii) explore the utility of spectral information across Landsat-8 spectrum using PCA and estimate α-diversity in the savannah woodland in southern Africa; and iii) determine the diversity index (H', D_2 and S) that best relates with spectral information on the Landsat-8 dataset.

2.2. Study area

The study area stretches across the KwaZulu-Natal (KZN), Mpumalanga and Limpopo provinces of South Africa, covering the savannah woodland belt (**Figure 2**). The area is divided into two land management regimes i.e. communal areas and protected areas (Kruger National Park, Hluhluwe-Imfolozi Park and other private nature reserves) with differing land use practices. High tree species diversity has been noted in both areas (du Toit et al. 2003; Shackleton, 2000). The savannah woodland is characterized by varying edaphic properties as a result of differential geological substrates and a mountainous terrain, particularly in the KZN region. Topography, rainfall and geology are amongst the key environmental factors that dictate the pattern of tree species diversity (Makhado et al. 2014; Shackleton, 2000).

The northern part of the study area receives low to moderate rainfall and supports the predominance of *Colophospermum mopane* (Makhado et al. 2014). The central part of the study area is dominated by members of the *Combretaceae* (*Terminalia sericea, Combretum collinum, Combretum apiculatum, Combretum zeyheri*) and *Mimosaceae* families (*Acacia nigrescens, Acacia gerradii and Dichrostachys cinerea*), with distribution being controlled by granite and gabbro geological substrates. Other important taxa include *Sclerocarya birrea,* which is widely distributed throughout the region (Eckhardt et al. 2000; du Toit et al. 2003; Shackleton, 2000). The mean annual precipitation ranges from 440mm in the north to 750mm in the south with annual variations around the mean (Makhado et al. 2014; Eckhardt et al. 2000). The month of March marks the end of the growing season while April to November has been described as the dry season in the southern African savannah (Grant and Scholes, 2006; Archibald and Scholes, 2007). Typical of a savannah setting, the vegetation is characterized by a continuous herbaceous layer interspersed by a woody tree cover of varying density depending on the geological substrate. The woody vegetation is characterized by trees of varying heights and crown dimensions (Wessels et al. 2011).

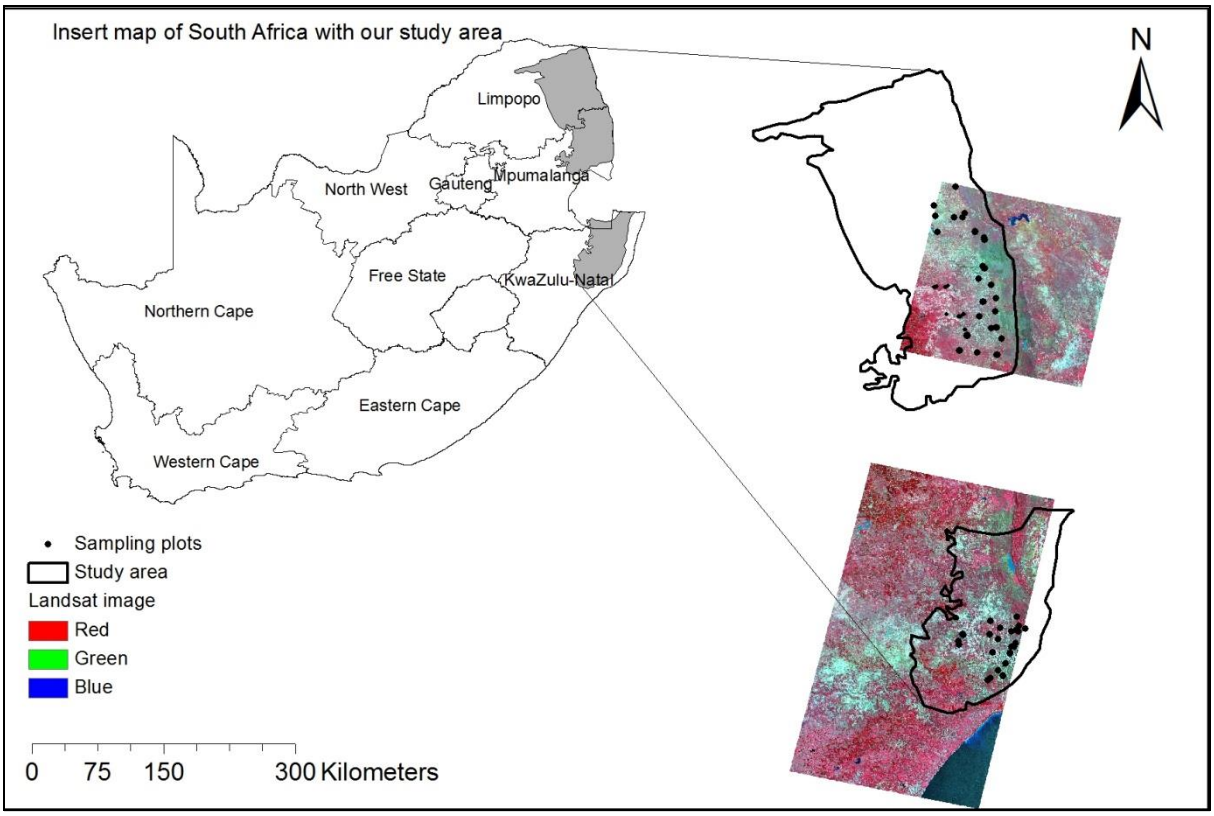

Figure 2 Study area stretching across three provinces of South Africa. Dots on the Landsat imagery are the sampling plots.

2.3. Material and methods

2.3.1. Remote sensing data

The two Landsat-8 Operational Land Imager (Landsat-8 OLI) satellite images were acquired on the 28th and 30th of March 2016. One image covers the KZN portion of the study area while the other images cover the Mpumalanga and Limpopo regions. The month of March marks the end of the growing season (Grant and Scholes, 2006; Madonsela et al. 2017). The study intends to extract vegetation indices for use as predictor variables and it was appropriate to collect the Landsat image when vegetation was still green.

Landsat-8 OLI delivers multi-spectral data with eight bands in the visible, near infrared and shortwave infrared regions of the electromagnetic spectrum. Landsat-8 OLI records data at a moderate spatial resolution of 30m and has a revisit capacity of 16 days. Landsat-8 with its 12-bit quantization of data has improved on the signal-to-noise radiometric performance of the sensor thus increasing its utility for landcover mapping (Pervez et al. 2016). The images were downloaded from the United States Geological Surveys (USGS) download portal (https://earthexplorer.usgs.gov/) with geometric correction already implemented. In addition, a 30m Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM) was acquired from USGS EarthExplorer and used in the atmospheric correction of the KZN Landsat scene. All Landsat images and DEM were projected to the Universal Transverse Mercator (UTM) coordinate system zone 36 south. The Landsat image covering the Mpumalanga and Limpopo regions were atmospherically corrected using ATCOR-2 software since the regions exhibit gentle undulating slopes (Richter and Schläpfer, 2012). The KZN Landsat scene necessitated the use of ATCOR-3 software since the region is mountainous. ATCOR-3 allows for integration of DEM which is useful for the correction of shadow effect on the image depicting mountainous areas (Richter and Schläpfer, 2012).

2.3.2. Field data collection

The study carried out two field campaigns from the 2nd - 27th of November 2015 in KZN and again on the 1st - 19th of March 2016 across Kruger National Park extending over the Mpumalanga and Limpopo provinces. The primary objectives of the field campaigns were to identify tree species within randomly placed sampling plots and quantify the level of diversity in the region using common measures of diversity (H', D_2 and S). Prior to field excursion, we defined the size of field sampling plots using the semi-variogram analysis in

ENVI 4.8 software. Essentially, the semi-variogram quantifies the spatial variability of natural phenomena occurring in space (Fu et al. 2014; Gringarten and Deutsch, 2001). It is computed as follows:

Equation 1

$$y(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2$$

where $y(h)$ is the semi-variance at a given distance h; $z(x_i)$ is the value of the variable Z at location x_i, h is the lag distance and $N(h)$ is the number of pairs of sample points separated by h.

Semi-variance gradually increases as the distance from one location to the next increases until it reaches the range where it starts to level off (Jongman et al. 1995; Gringarten and Deutsch, 2001). A semi-variogram plot is generated by computing variance at different lag distances and a theoretical model such as a spherical or exponential model that is fitted to provide information about spatial structure (Fu et al. 2014). Our study applied semi-variogram analysis to WorldView-2 derived NDVI image to define the scale of spatial variability in tree species richness. The choice to use NDVI was based on an observation that variability in NDVI corresponds to species diversity (Gould, 2000). It was important to use NDVI because it suppresses spectral content from non-vegetated pixels (Viña et al. 2011), and was therefore a viable option to determine pixel variability related to vegetation.

In our analysis, the Worldview-2 image was firstly degraded to a 10m spatial resolution to be compatible with the average tree canopy size in the savannah (Cho et al., 2012) and the generated NDVI image. In ENVI software v4.8, the semi-variogram analysis computed the squared difference between neighbouring pixel values in order to quantify variability. The analysis conducted on Worldview-2 derived NDVI image showed that the scale for tree species variability in the savannah woodland lies at lag distances of 90m to 100m (**Figure 3**). Although semi-variance would seem to be increasing beyond the lag distance of 90m, the increase was not consistent and the lag distance of 90m resulted in plot sizes that are feasible to work on within limited resources. Moreover, the study intended to use Landsat-8 data with 30m pixel resolution hence the plot size of 90m X 90m was opted to ascertain correspondence between field data and pixel spectral content.

The plot size of 90m X 90m was therefore chosen to capture spatial variation in tree species diversity. Stratified random sampling was used to define the placement of sampling plots. The stratification of sampling plots followed four dominant geological formations (granite; siliciclastic; gabbros; granulite) that were observed to have marked influence over vegetation patterns in the study area (du Toit et al., 2003). Plots of 90m x 90m were designed and all trees within the plots with a diameter at breast height (DBH) above 10cm were recorded with the Global Positioning System and species identified. The study collected 5 859 trees belonging to 106 tree species. The field campaigns visited 50 plots distributed across the study area to collect tree species data. A further 26 plots collected under similar conditions in the previous study (Naidoo et al. 2015) were added to our field data. However, five of the total field plots were located on clouded parts of the image and therefore not usable. A total of 71 field plots were used in the analysis.

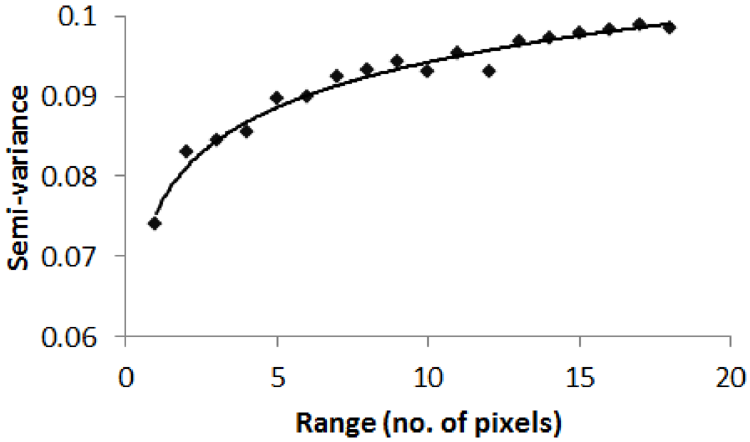

Figure 3 Semi-variogram analysis showing the scale of tree species variability in the savannah woodland.

2.3.3. Data analysis

The quantification of tree species diversity within each sample plot was calculated using the three common measures of local diversity i.e. species richness (*S*), Shannon's diversity (*H'*) and Simpson's dominance (D_2) (**Table 1**). These indices are frequently cited in ecological literature (Lande, 1996; Colwell, 2009; Morris et al. 2014) and were chosen to ensure that the results were comparable with other studies. In addition, *H'* and D_2 considers both species richness (i.e. number of different species) and abundance (i.e. number of individual trees within species) (Colwell, 2009; Morris et al. 2014) and these aspects of diversity have been verified to have a bearing on the spectral signal captured by the remote sensing device (Oldeland, 2010).

Table 1 Alpha diversity indices used in the study and their equations

Species diversity index	Equation	Reference
Species richness	$S=N$	Morris *et al.* (2014)
Shannon index	$H' = -\sum_{i=1}^{s} p_i \text{In}(p_i)$	Shannon, (1948); Morris et al., (2014)
Simpson index	$D_2 = 1/\sum_{i=1}^{s} p_i^2$	Simpson, (1949); Morris *et al.* (2014)

where N is the total number of tree species in a sample; p_i is the proportional abundance of species i relative to the total abundance of all species *S* in a plot; $\text{In}(p_i)$ is the natural logarithm of this proportion.

The nine Landsat pixels falling within sampling plots were identified and the spectral reflectance from all Landsat-8 bands was extracted (**Table 2**). Firstly, the mean, standard deviation and the range statistics within 3x3 pixels were computed from vegetation indices and used as predictor variables. Vegetation indices (VI's) were computed from the blue (452.02 - 512.06 nm), red (635.85 - 673.32 nm) and NIR (850.54 - 878.79 nm) regions of Landsat-8 image (**Table 2**). The range and standard deviation were used as surrogate measures of variability in vegetation characteristics (Viña et al. 2011) and were expected to relate better with tree species diversity. We also computed coefficient of variation (CV) to quantify in percentage the amount of variability captured by each vegetation index. CV was computed as follows:-

Equation 2

$$CV = \frac{\sigma}{\mu} * 100$$

where σ represents the standard deviation of from all samples; μ represents the mean value of vegetation index from all samples.

Further information on spatial variability was extracted in the form of texture using Gray Level Co-occurrence Matrix (GLCM) in ENVI 4.8. Textural properties are indicative of spatial variability (Haralick et al., 1973) and such variability is presumed to reflect greater environmental heterogeneity associated with high assemblage of species diversity (Hernández-Stefanoni et al. 2012). The study used second-order texture measures simply because they account for spatial relations amongst neighbouring pixels and they are therefore more consistent with the aim of the study. We used a 3x3 window size in order to detect fine scale variability (Kelsey and Neff, 2014) consistent with variability defined by semi-variogram analysis. There are three categories within which GLCM texture is computed: i) based on the level of contrast between pixels, we chose dissimilarity; ii) based on pixel organization within a window, we chose entropy; and iii) based pixel statistics, we chose variance (Haralick et al. 1973; Hernández-Stefanoni et al. 2012) (**Table 2**).

Table 2 List of spectral dataset used as predictor variables in the models

Vegetation indices	Equation / Spectral bands	Reference
Normalized Difference Vegetation Index (NDVI)	= (NIR - RED) / (NIR + RED)	Rouse et al., (1973)
Enhanced Vegetation Index (EVI)	=2.5*(NIR-RED)/(NIR+6.0*RED-7.5*BLUE+1.0)	Huete (1999)
Simple Ratio Index (SRI)	= NIR / RED	Tucker, (1979)
Soil Adjusted Vegetation index (SAVI)	= (NIR - RED) / (NIR + RED + L) * (1 + L)	Huete (1988)
Landsat Spectral bands		
Coastal band	434.97 - 450.95 nm	Landsat-8 Data User Handbook (2016)
Blue band	452.02 - 512.06 nm	Landsat-8 Data User Handbook (2016)
Green band	532.74 - 590.07 nm	Landsat-8 Data User Handbook (2016)
Red band	635.85 - 673.32 nm	Landsat-8 Data User Handbook (2016)
Near Infrared band	850.54 - 878.79 nm	Landsat-8 Data User Handbook (2016)
Cirrus band	1363.24 -1383.63 nm	Landsat-8 Data User Handbook (2016)
Shortwave Infrared band-1	1566.5 -1651.22 nm	Landsat-8 Data User Handbook (2016)
Shortwave Infrared band-2	2107.4 - 2294.06 nm	Landsat-8 Data User Handbook (2016)
Gray-Level Co-occurrence Matrix textural layers		
Variance	$= \sum_{i=0}^{G-1}\sum_{i=0}^{G-1}(i-\mu)^2 P(i,j)$	Haralick et al., (1973); Albregtsen, (2008)
Dissimilarity	$= \sum_{i,j=0}^{N-1} Pi,j\lvert i-j\rvert$	Haralick et al., (1973); Beliakov et al., 2008
Entropy	$= -\sum_{i=0}^{G-1}\sum_{i=0}^{G-1} P(i,j)\times\log(P(i,j))$	Haralick et al., (1973); Albregtsen, (2008)

where L represents a constant soil adjustment factor*;* G represents number of gray levels used; N represents the number of distinct gray levels in the quantized image; μ represents the mean value of P; $P(i,j)$represent $(i,j)th$ entry in normalized gray-tone spatial-dependence matrix, = $P(i,j)/R$*; R* represents a normalizing factor.

Moreover, multivariate analysis particularly PCA, was tested as a way of exploring the utility of spectral information from the entire Landsat-8 spectrum (visible, NIR and SWIR) for estimating α-diversity. PCA is a technique that decomposes the original data through linear combination of original variables and produces few principal components (PCs) that best explain the variability in the original data (Bro and Smilde, 2014). To compute PCA, data are prepared in a matrix X with I rows (i = 1,…,l) and J columns and the size will be I x J. The characteristic variables of matrix X are represented by x_j ($j = 1, \ldots, j$) and are all vectors in the I-dimensional space. A linear model of these x variables can be expressed as $t = w_1 \times x_1 + \ldots + w_j \times x_j$, where t represents the new vector in the same space as the x variables. t is the first principal component that explains the most variation in x variables (Bro and Smilde, 2014). The optimal number of PCs is normally defined by PCs that explain over 95% of variability in the original dataset (Thenkabail et al. 2004).

In our application of PCA we firstly normalized the data using autoscaling based on dispersion in ParLes software v3.1 (Rossel, 2008) to cater for differences in scales between variables. Secondly, we plotted the first and the second PCs to detect outliers in the PCs which are defined as samples that behave strangely and have the potential to upset the subsequent analysis if not corrected or removed (Bro and Smilde, 2014). Prior to final removal of outliers, it is recommended to compare the effect on the model before and after removal (Bro and Smilde, 2014). In this study, we only removed outliers in the PCs derived from Landsat-8 spectral bands because they negatively affected the ability of regression model to predict tree species diversity. In order to test different scenarios, principal components were extracted from: i) vegetation indices, ii) Landsat-8 spectral bands, iii) GLCM texture layers and iv) different combinations of all our spectral variables. The PCs were produced using ParLes software v3.1 and then imported into MATLAB software v7.8.0 (R2009a, MathWorks) where bootstrap regression was conducted, and the PCs were used as predictor variables in the stepwise linear regression model. The optimal number of PCs was defined by PCs that explain over 95% of variability in the datasets as reported in the literature (Thenkabail et al. 2004).

In order to assess the precision and the accuracy of the models, the bootstrapping approach was applied in modelling the relationship between spectral variability and species diversity. Firstly, we completed 1000 random permutations of the original data and then split two-thirds of the data for training the models and used the remainder for evaluating the predictive ability of the models. Modelling results are presented in table format in the subsequent section. Two modelling approaches i.e. univariate and multivariate analyses were tested and then followed by comparative analysis of the results. A simple linear regression model was used to investigate the relationship between spectral data as predictor variables and species diversity indices as response variables. The strength of the relationship was assessed using the coefficient of determination (R^2), the *p-value* statistics and the model performance was evaluated using the root mean square error (RMSE).

2.4. Results

2.4.1. Univariate analysis: The relationship between diversity measures and vegetation indices, GLCM layers and Landsat-8 bands

The results of bootstrapped regression analysis demonstrated a significant positive relationship ($p < 0.05$) between vegetation indices and measures of tree species diversity (**Table 3**). In particular, H' and D_2 have demonstrated a higher relationship to vegetation indices (r^2 ranging from 0.26 to 0.29) compared to S (r^2 ranging from 0.21 to 0.23). However, the relationship declined significantly ($p < 0.05$) when derivatives (standard deviation and the range) from NDVI, EVI and SAVI were used as predictors. S had the lowest relationship with derivatives from NDVI, EVI and SAVI (r^2 ranging from 0.0 to 0.03) compared with H' and D_2 (r^2 ranging from 0.10 to 0.20). However derivatives from SRI were an exception and in fact the relationship was significantly improved ($p < 0.05$) when they were used as predictors. SRI derivatives (standard deviation and the range) had the highest relationship with H' (r^2 of 0.36 and 0.34 respectively), D_2 (r^2 of 0.41 and 0.38 respectively) and S (r^2 of 0.24 and 0.22) compared to NDVI, EVI, SAVI and their derivatives. In essence the best model for estimating tree species diversity was derived from the SRI derivative (standard deviation) (**Figure 4**). The SRI standard deviation had the highest relationship with H', D_2 and S confirming its sensitivity to the diversity of tree species in the savannah woodland.

Moreover, H' and D_2 equally showed a higher relationship with vegetation indices (NDVI, EVI, SRI and SAVI) compared to S (**Table 3**). However it was D_2 that had the highest relationship with the high performing SRI derivative (standard deviation) with an r^2 of 0.41. Furthermore, SRI had a higher coefficient of variation (46.6%) compared to NDVI (24%), EVI (33.1%) and SAVI (24.1%). Bootstrapping produced r^2 histograms which verified the precision of our regression models and the robustness of the relationships between vegetation indices and *H′*, D_2 and *S* with mean r^2 ranging from 0.21 to 0.41 (**Figures 5, 6 and 7**). However, regression analysis showed that GLCM texture measures had no relationship with measures of tree species diversity (**Table 4**). In most instances GLCM texture measures maintained the r^2 of less than 0.06 indicating the lack of relationship with either *H′*, D_2 or *S*. It was only entropy derived from NIR and SWIR-2 that had a significant relationship with *S* (r^2 of 0.04; $p < 0.05$) and *H′* (r^2 of 0.05; $p < 0.05$) respectively.

Meanwhile Landsat-8 spectral bands showed a significant negative relationship with measures of tree species diversity (**Table 5**). There was no single Landsat-8 band that consistently outperformed other spectral bands when modelling tree species diversity as measured by *H′*, D_2 and *S*. Noteworthy though, *H′* and D_2 showed a higher relationship with Landsat-8 red band (with r^2 of 0.18 and 0.19 respectively) compared to *S* (with r^2 of 0.14). *S* had a higher relationship with the Landsat-8 coastal band (r^2 of 0.16) compared to other spectral bands. However, the Landsat-8 NIR and cirrus bands were the only spectral bands that did not show a relationship with either *H′*, D_2 or *S*. All the Landsat-8 bands, except the NIR and Cirrus bands, showed a negative relationship with *H′*, D_2 and *S* (**Figure 8**) suggesting a possibility that low diversity areas have low vegetation cover resulting in high signal reflectance across all Landsat-8 spectral bands.

The overall results show that the best models for estimating tree species diversity using Landsat-8 were derived from vegetation indices (SRI derivatives, NDVI, EVI and SAVI). However, it was SRI derivatives models that had significantly lower RMSE ($p < 0.05$) when predicting *H′* and D_2 compared to the regression models from NDVI, EVI and SAVI. While the SRI derivative (standard deviation) had a lower RMSE when predicting *S* than NDVI, EVI and SAVI, the difference was not statistically significant ($p > 0.05$).

Table 3 Relationship observed between three common measures of tree species diversity (*H′*, D_2 and *S*) and spectral variables. The spectral variable statistics were extracted from Landsat derived vegetation index images within 90m X 90m field plot. All computations were drawn from 1000 bootstrapped iterations.

Diversity index	Spectral variable	Average R^2	Confidence interval 95%	*P-value*	RMSE
H′	NDVI (Mean)	0.29	±0.014	0.0005	0.4861
	NDVI (St dev)	0.10	±0.014	0.0167	0.5586
	NDVI (Range)	0.11	±0.014	0.0114	0.5524
	EVI (Mean)	0.29	±0.014	0.0008	0.4869

	EVI (St dev)	0.17	±0.012	0.0063	0.5302
	EVI (Range)	0.17	±0.011	0.0073	0.5286
	SRI (Mean)	0.26	±0.005	0.0006	0.4985
	SRI(St dev)	0.36	±0.010	0.0000	0.4613
	SRI (Range)	0.34	±0.011	0.0000	0.4688
	SAVI (Mean)	0.29	±0.014	0.0007	0.4894
	SAVI (St dev)	0.10	±0.014	0.0118	0.5554
	SAVI (Range)	0.11	±0.014	0.0113	0.5536
D_2	NDVI (Mean)	0.29	±0.012	0.0003	1.8048
	NDVI (St dev)	0.12	±0.015	0.0132	2.0724
	NDVI (Range)	0.11	±0.015	0.0144	2.0761
	EVI (Mean)	0.29	±0.013	0.0003	1.8203
	EVI (St dev)	0.20	±0.013	0.0037	1.9493
	EVI (Range)	0.17	±0.012	0.0061	1.9599
	SRI (Mean)	0.27	±0.012	0.0005	1.8545
	SRI(St dev)	0.41	±0.014	0.0000	1.6668
	SRI (Range)	0.38	±0.014	0.0006	1.7232
	SAVI (Mean)	0.29	±0.012	0.0003	1.8250
	SAVI (St dev)	0.12	±0.015	0.0142	2.0605
	SAVI (Range)	0.11	±0.015	0.0145	2.0569
S	NDVI (Mean)	0.23	±0.013	0.0020	3.4913
	NDVI (St dev)	0.00	±0.005	0.5392	3.9684
	NDVI (Range)	0.00	±0.004	0.4461	3.9580
	EVI (Mean)	0.21	±0.013	0.0027	3.5516
	EVI (St dev)	0.03	±0.011	0.1791	3.9535
	EVI (Range)	0.01	±0.007	0.2128	3.9861
	SRI (Mean)	0.23	±0.015	0.0024	3.5513
	SRI(St dev)	0.24	±0.013	0.0017	3.4732
	SRI (Range)	0.22	±0.012	0.0025	3.5494
	SAVI (Mean)	0.23	±0.013	0.0073	3.5263
	SAVI (St dev)	0.00	±0.004	0.4458	3.9171
	SAVI (Range)	0.00	±0.003	0.4756	3.9552

Table 4 Relationship observed between three common measures of tree species diversity and GLCM texture measures. Texture measures were extracted from Landsat-8 spectral bands within 90m X 90m field plot. All computations were drawn from 1000 bootstrapped iterations.

		Diversity index					
		Shannon index		**Simpson index**		**Species richness**	
Landsat band	**GLCM Texture**	***R^2***	***P-value***	***R^2***	***P-value***	***R^2***	***P-value***
Coastal band	Variance	0.00	0.4840	0.00	0.8435	0.00	0.6140
	Entropy	0.00	0.6761	0.00	0.5548	0.00	0.8204
	Dissimilarity	0.00	0.6859	0.00	0.8539	0.01	0.3971
Blue band	Variance	0.01	0.3345	0.00	0.9080	0.01	0.2834
	Entropy	0.00	0.5339	0.00	0.4646	0.00	0.8969
	Dissimilarity	0.00	0.9567	0.00	0.7771	0.00	0.4579
Green band	Variance	0.01	0.3053	0.00	0.6648	0.00	0.6996
	Entropy	0.00	0.5105	0.00	0.7120	0.00	0.4399
	Dissimilarity	0.00	0.6218	0.00	0.7749	0.03	0.1018

Red band	Variance	0.00	0.5944	0.00	0.5518	0.00	0.5016
	Entropy	0.00	0.8027	0.00	0.8025	0.00	0.6244
	Dissimilarity	0.01	0.3683	0.01	0.6399	0.00	0.9025
NIR band	Variance	0.01	0.4782	0.00	0.4717	0.04	0.0898
	Entropy	0.03	0.1371	0.01	0.3299	0.04	0.0331
	Dissimilarity	0.03	0.1265	0.00	0.7328	0.02	0.2139
Cirrus band	Variance	0.00	0.5135	0.00	0.5334	0.00	0.7506
	Entropy	0.01	0.3044	0.00	0.4709	0.00	0.4826
	Dissimilarity	0.01	0.3534	0.00	0.5385	0.00	0.6906
SWIR-1 band	Variance	0.02	0.1930	0.01	0.3497	0.00	0.9641
	Entropy	0.05	0.0231	0.03	0.1441	0.02	0.1502
	Dissimilarity	0.04	0.0826	0.01	0.2420	0.00	0.4168
SWIR-2 band	Variance	0.00	0.4870	0.00	0.6627	0.00	0.8568
	Entropy	0.02	0.1733	0.00	0.6432	0.00	0.4388
	Dissimilarity	0.00	0.4694	0.00	0.8398	0.00	0.8321

Table 5 Relationship observed between three common measures of tree species diversity and Landsat-8 spectral bands. The mean spectral reflectance was from Landsat bands within 90m X 90m field plot. All computations were drawn from 1000 bootstrapped iterations.

	Diversity index											
	Shannon index				**Simpson index**				**Species richness**			
Landsat band mean reflectance	R^2	**CI 95%**	***P-value***	**RMSE**	R^2	**CI 95%**	***P-value***	**RMSE**	R^2	**CI 95%**	***P-value***	**RMSE**
Coastal band	0.17	±0.014	0.0071	0.5280	0.17	±0.011	0.0060	1.961	0.16	±0.014	0.0079	3.6878
Blue band	0.13	±0.015	0.0127	0.5431	0.14	±0.012	0.0105	1.996	0.13	±0.015	0.0117	3.7537
Green band	0.10	±0.015	0.0173	0.5564	0.11	±0.014	0.0159	2.063	0.09	±0.015	0.0172	3.8579
Red band	0.18	±0.013	0.0060	0.5246	0.19	±0.011	0.0042	1.934	0.14	±0.013	0.0105	3.7192
NIR	0.02	±0.011	0.9999	0.5728	0.02	±0.010	0.9999	2.158	0.00	±0.005	0.9999	3.940
Cirrus	0.01	±0.009	0.9999	0.5767	0.08	±0.002	0.9999	2.134	0.00	±0.008	0.9999	3.991
SWIR-1 band	0.09	±0.016	0.0180	0.5623	0.09	±0.015	0.0193	2.097	0.07	±0.015	0.0212	3.9201
SWIR-2 band	0.12	±0.014	0.0146	0.5489	0.14	±0.013	0.0111	2.019	0.09	±0.015	0.0176	3.8834

CI- Confidence interval

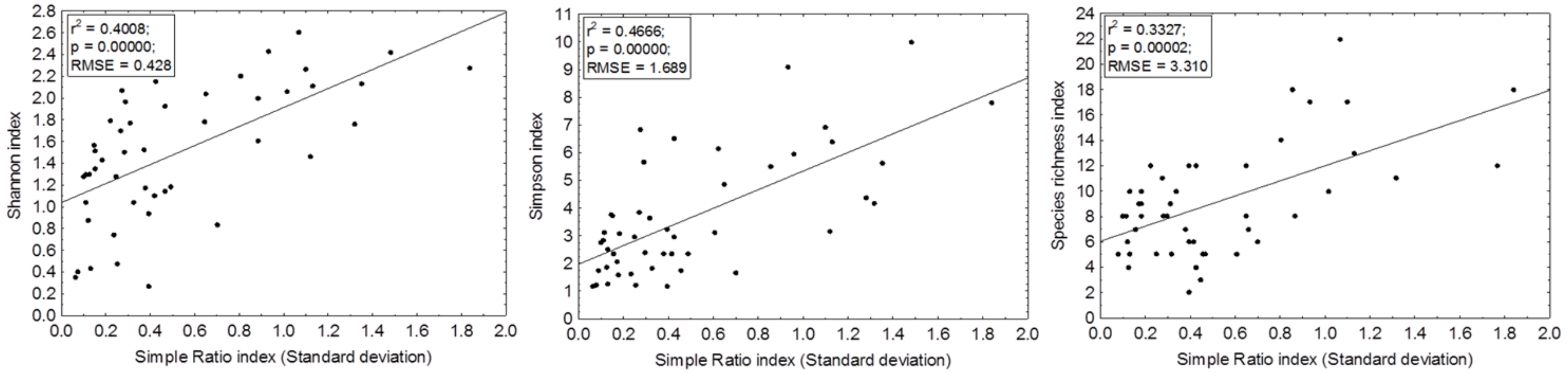

Figure 4 Relationship between Simple Ration Index derivative and on the left) Shannon index; middle) Simpson index; right) Species richness. SRI standard deviation had shown higher positive relationship with tree species diversity and we selected the best model (maximum r^2 with the lowest RMSE from 1000 bootstrapped iterations) to plot the relationship.

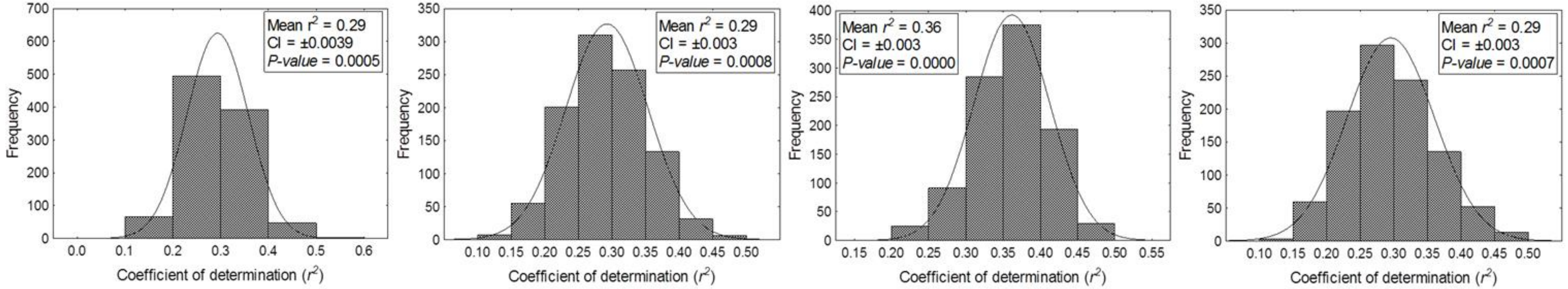

Figure 5 Histograms of bootstrapped r^2 for models involving Shannon index and on the left) mean NDVI; second from left) mean EVI; third from left) SRI standard deviation; fourth from left) mean SAVI.

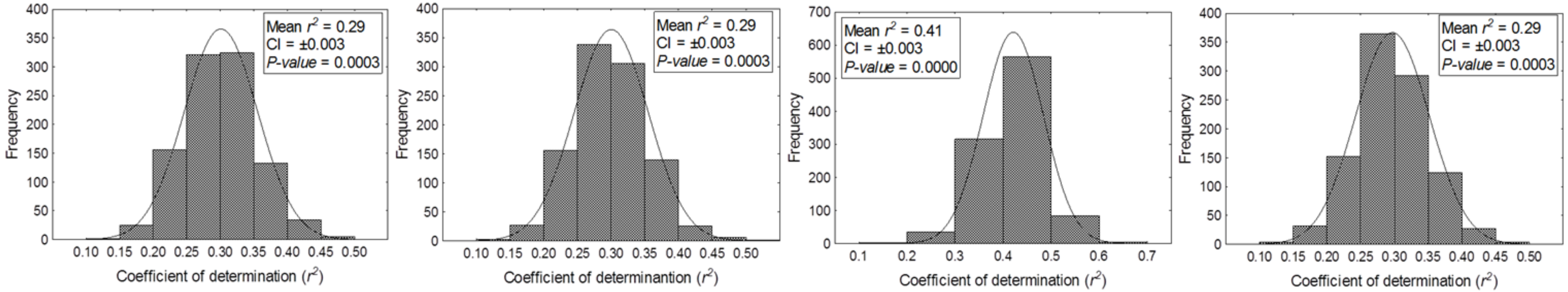

Figure 6 Histograms of bootstrapped r^2 for models involving Simpson index and on the left) mean NDVI; second from left) mean EVI; third from left) SRI standard deviation; fourth from left) mean SAVI.

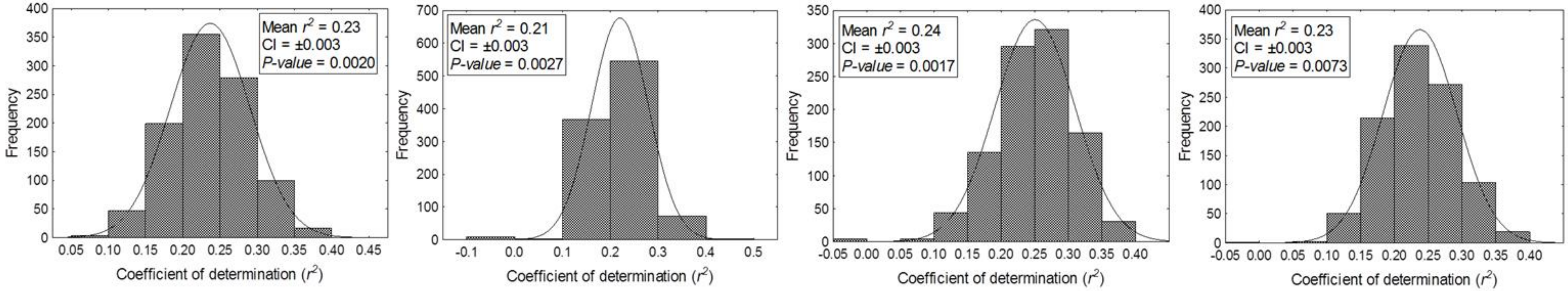

Figure 7 Histograms of bootstrapped r^2 for models involving Species richness and on the left) mean NDVI; second from left) mean EVI; third from left) SRI standard deviation; fourth from left) mean SAVI.

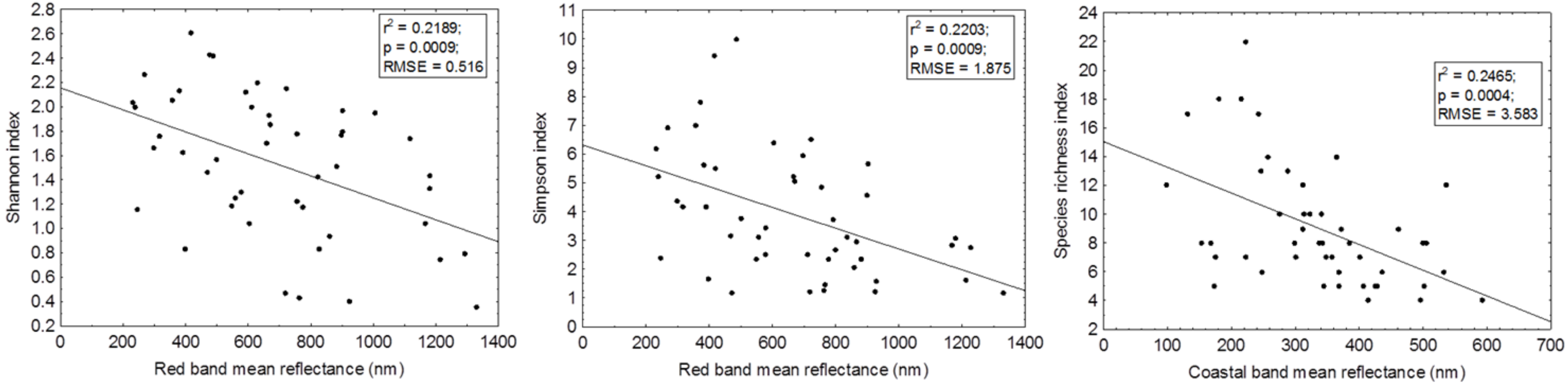

Figure 8 Relationship between red band reflectance and on the left) Shannon index; middle) Simpson index; right) species richness. Red band had shown higher negative relationship tree species diversity than other spectral bands and we selected one best model (maximum r^2 with the lowest RMSE from 1000 bootstrapped iterations) to plot the relationship.

2.4.2. Multivariate analysis

The results of the stepwise linear regression showed that PCs had a significant relationship with all measures of tree species diversity ($p < 0.05$) (**Table 6**). In particular PCs derived from a combination of vegetation indices and Landsat-8 spectral bands had a higher relationship with H' (r^2 of 0.41; $p < 0.05$) and D_2 (r^2 of 0.42; $p < 0.05$) compared to PCs extracted from vegetation indices, Landsat-8 bands, GLCM texture measures separately or any combination of these variables. S had a higher relationship with PCs derived from a combination of Landsat-8 bands and GLCM texture measures (r^2 of 0.27; $p < 0.05$). Moreover the Principal Component Analysis had improved the utility of GLCM texture measures for estimating tree species diversity. PCs derived from GLCM texture measures showed a significant relationship with all measures of tree species diversity ($p < 0.05$) (**Table 6**) and this was a major improvement compared to univariate analysis of GLCM texture measures (**Table 4**). In addition, transforming Landsat-8 spectral bands into PCs improved the explanatory power of Landsat-8 spectral bands. In fact PCs derived from Landsat-8 spectral bands had a higher relationship with H' and D_2 (r^2 of 0.36 and 0.35, respectively) compared to mean NDVI, mean EVI, mean SRI or mean SAVI (r^2 ranging from 0.26 to 0.29).

Comparisons between univariate and multivariate analysis showed that PCs derived from a combination of vegetation indices and Landsat-8 bands predicted H' with significantly lower RMSE ($p = 0.0363$) than the high performing univariate model (SRI standard deviation). However, the same PCs failed to significantly improve the prediction of D_2 and S compared to univariate model derived from SRI standard deviation. SRI model predicted D_2 and S with significantly lower RMSE ($p < 0.05$) compared to PCs. These results suggest that H' is better related to PCs while D_2 relates more with SRI.

Table 6 Relationship observed between PCs and three common measures of tree species diversity (H', D_2 and S). The RMSE indicates predictive performance of stepwise regression models. All computation were drawn from 1000 bootstrap iterations.

Predictor variables	Response variables	PCs explaining over 95%	Average R^2	Confidence interval 95%	*P-value*	RMSE
VIs	H'	2	0.37	±0.016	0.0000	0.459
	D_2	2	0.38	±0.015	0.0000	1.699
	S	2	0.22	±0.017	0.0027	3.627
Landsat bands	H'	3	0.36	±0.016	0.0002	0.480

	D_2	3	0.35	±0.014	0.0002	1.822
	S	3	0.19	±0.015	0.0075	3.826
GLCM texture measures	H'	7	0.21	±0.020	0.0059	0.559
	D_2	7	0.20	±0.019	0.0069	2.093
	S	7	0.22	±0.020	0.0063	3.890
VIs + Landsat	H'	2	0.41	±0.017	0.0001	0.456
	D_2	2	0.42	±0.015	0.0000	1.663
	S	2	0.22	±0.017	0.0035	3.657
VIs + GLCM	H'	6	0.35	±0.020	0.0009	0.523
	D_2	6	0.32	±0.020	0.0017	1.984
	S	6	0.21	±0.016	0.0040	3.647
Landsat + GLCM	H'	6	0.32	±0.022	0.0005	0.516
	D_2	6	0.26	±0.022	0.0012	1.950
	S	6	0.27	±0.021	0.0014	3.523
VIs + Landsat + GLCM layers	H'	5	0.40	±0.016	0.0004	0.487
	D_2	5	0.40	±0.015	0.0004	1.823
	S	5	0.26	±0.017	0.0030	3.655

2.5. Discussion

The significant relationship observed between vegetation indices (NDVI, EVI, SRI and SAVI) and measures of local diversity (H', D_2 and S) suggest that satellite images would be useful for estimating tree species diversity in the savannah woodland. Vegetation indices suppress spectral reflectance from non-vegetative features while enhancing the spectral content from vegetation. Therefore, variability in vegetation indices emanates from a variety of vegetation characteristics, e.g. canopy structure, leaf area index, tree canopy cover and green biomass (Viña et al. 2011; Huete et al. 2002). Furthermore, vegetation indices have been shown to be sensitive to abiotic factors, e.g. rainfall, that impact on tree species diversity (Pau et al. 2012; Oindo and Skidmore, 2002). It is therefore not surprising that mean NDVI, mean EVI, mean SRI and mean EVI had a significant relationship with tree species diversity as measured by H', D_2 and S. The positive linear relationship between vegetation indices and tree species diversity further confirms their sensitivity to abiotic factors impacting tree species diversity in the savannah woodland. Shackleton (2000) observed that plant species richness increase with increasing average annual precipitation in the savannah woodland. For instance, the northern part of the study area with its low to moderate annual rainfall (which is around 440mm per annum), has a low diversity of tree species. The northern part of the study area supports mainly the distribution of *Colophospermum mopane* which has adapted to that environment and this was also observed by Makhado et al. (2014). The diversity of tree species increases with rising annual rainfall towards the southern part of the study area. In essence the linear relationship between vegetation indices and tree species diversity supports the positive productivity-diversity postulation, which states that the relationship between productivity and species diversity follows an environmental gradient (Kirkman et al. 2001; Bai et al. 2007).

Moreover, derivatives from vegetation indices, which were used as a surrogate measure of spatial variability in vegetation characteristics (Viña et al. 2011) also, had a significant positive relationship with tree species diversity. The positive relationship implies that variability in vegetation characteristics is the outcome of high tree species diversity. However, the sensitivity of vegetation indices to variability in vegetation characteristics differs between indices (NDVI, EVI, SRI and SAVI). It was only SRI derivatives that had a higher relationship with tree species diversity compared to mean SRI. Derivatives from NDVI, EVI and SAVI had lower relationship with tree species diversity compared to mean NDVI, mean EVI and mean SAVI respectively. Other studies (Parviainen et al. 2010; Wood et al. 2013) have also made similar observations with derivatives from NDVI. One possible explanation for these differences in sensitivity to vegetation characteristics could be different measurement scales of vegetation indices. SRI has a measurement scale which ranges from 0 to far beyond 1 and this is assumed to enable derivatives from SRI to capture variability much better than NDVI, SAVI and EVI. In this study, SRI had a higher coefficient of variation of 46.6 % compared to NDVI, EVI and SAVI with coefficient of variation of 24.0%, 33.1% and 24.1% respectively. As result, SRI derivatives explained tree species diversity better than NDVI, EVI and SAVI. However, EVI, which also has a measurement scale which ranges from 0 to far beyond 1, had the second highest coefficient of variation (33.1%). Furthermore EVI derivatives also had higher relationship with H' and D_2 (r^2 ranging from 0.17 to 0.20) compared to NDVI or SAVI derivatives (r^2 ranging between 0.10 and 0.11).

These results from SRI and EVI confirm our assertion that measurement scale of vegetation indices impacts their ability to explain tree species diversity.

The significant relationship between VIs and diversity indices confirms the utility of Landsat imagery for practical application in conservation, particularly as a screening tool to identify biodiversity hotspots. However, the success of biodiversity estimation through remotely sensed data would depend largely on the use of spectral variables suited for capturing tree species diversity on the particular landscape. Contrary to observations by Hernández-Stefanoni et al. (2012), the use of GLCM textural measures as a proxy for spatial variability did not show any relationship with tree species diversity measures in the savannah woodlands. In cases where there was a significant relationship it was very low (r^2 of less than 0.06) and cannot be suggested for practical application. The GLCM textural measures quantify variability in reflectance signal between neighbouring pixels (Hernández-Stefanoni et al. 2012) and in the savannah woodlands such variability would always be high due to the heterogeneous structure of vegetation coupled with bare surface ground contribution to reflectance spectra. The small window size (3x3), within which texture measures were computed is sensitive to fine scale variations (Kelsey and Neff, 2014). Unlike vegetation indices which suppress contribution from non-vegetated features, textural properties captures total variation on the image and was not useful for estimating tree species diversity in the savannah. Wood et al. (2013) also observed a very weak correlation between species diversity and image texture in the savannah environment in Fort McCoy Military Installation, USA. The weak correlation was attributed to sparse tree cover in the savannah environment resulting in high textural variability which did not correspond to the tree species diversity of the area.

Meanwhile the untransformed Landsat-8 spectral bands, except the cirrus and NIR bands had shown a significant negative relationship with tree species diversity. Although the relationship was lower compared to that observed with vegetation indices, the results raised an ecological research question. The negative relationship generates an assumption that: i) low diversity plots have low vegetation cover resulting in high spectral signal reflectance; and ii) high diversity plots have high vegetation cover hence low signal reflectance. For instance, Patel et al. (2007) observed that dry vegetation cover has positive correlation with spectral bands in the visible region of electromagnetic spectrum and poor correlation with NIR bands. Positive correlation in the visible region indicates high spectral signal reflectance across all bands and this is typical of dry vegetation due to the background effect as it had dropped its foliage cover (Todd and Hoffer, 1998). The question that arises from this observation, and which will be attended to in future research, is whether vegetation cover is proportionally related to tree species diversity.

Moreover, our study demonstrated that univariate analysis does not fully exploit the information content of remotely sensed data. The application of a multivariate technique, PCA, enabled the utilization of the entire spectral information in the visible, NIR and SWIR regions of Landsat-8 for purpose of estimating tree species diversity. Consequently, the resulting PCs were better than the mean NDVI, mean EVI, mean SRI or mean SAVI in explaining tree species diversity. The Landsat derived PCs contain essential spectral information from the SWIR region which is also related to vegetation properties (Thenkabail et al. 2003; Hernández-Stefanoni et al., 2012). Therefore, the higher explanatory power of PCs over mean NDVI, mean EVI, mean SRI and mean SAVI was attributed to the utilization of the entire spectral content of Landsat-8 data. Consistent with this assertion is the observation by Jakubauskas and Price (1997) that biophysical properties of forest canopy are best explained by a combination of spectral information in the visible and SWIR regions of Landsat-7 Enhanced Thematic Mapper plus image. The observation by Jakubauskas and Price (1997) justifies our assertion that SWIR has essential spectral information useful for characterization of vegetation. However it was the PCs derived from the combination of Landsat spectral bands and vegetation indices that explained H' better than any predictor variable (r^2 of 0.41; $p < 0.05$). The same PCs also had an equally high relationship with D_2 (r^2 of 0.42; $p < 0.05$). The obvious implication is that combining Landsat-8 spectral bands with vegetation indices increase the explanatory power of PCs.

Furthermore, multivariate analysis transformed the GLCM texture measures into useful PCs for explaining tree species diversity. The PCs derived from GLCM texture measures had a significant relationship with tree species diversity although this was not comparable to other predictor variables. However, combining Landsat-8 spectral bands with GLCM textures did not improve the explanatory power of PCs. Overall the results suggest that transforming spectral variables into principal components enhances the utility of Landsat data for tree species diversity estimation. The PCs derived from the combination of Landsat spectral bands and vegetation indices explained 41% of the variability in *H'*, which is comparable to the observation made by Oldeland et al. (2010) using hyperspectral data in the Central Namibian savannah. However, our study only considered tree species diversity whilst the savannah is characterized by the co-existence of trees and grass. Therefore the overall spectral signal captured by Landsat-8 image relates to the total vegetation cover and this is assumed to have contributed to prediction errors observed in the study. Areas with high ratio of grass cover would be susceptible to over prediction. Nonetheless, the fact that *H'* and D_2 had a significant relationship with PCs (r^2 of 0.41 and 0.42; $p < 0.05$) comparable to Oldeland et al. (2010) suggest that the effect of herbaceous vegetation on the spectral signal captured by the Landsat sensor was not dominant.

Moreover, the high performing regression models in this study explained only 41 - 42% variability in tree species diversity. This can be improved with the incorporation of environmental variables known to impact tree species diversity in the savannah. Combining remotely sensed variables with environmental variables have been shown to increase the predictive ability of regression models (Zimmermann et al. 2007; Malahlela et al. 2015). In the southern African savannah, rainfall (Shackleton, 2000) and geology (du Toit et al. 2003) are some of the environmental factors known to impact tree species diversity. Furthermore, our general observation was that species diversity measures that consider both species richness and abundance relate better with vegetation indices and PCs. This is consistent with observations in the literature (Oldeland et al. 2010; Rocchini et al. 2010), which state that abundant tree species make a meaningful contribution in the overall spectral reflectance captured by a remote sensing device and therefore shows a better relationship with vegetation indices and PCs. In addition, this study benefited from ensuring that field plots match Landsat pixel size. Maintaining pixel-field plot correspondence facilitates the extraction of useful spectral information from remotely sensed image which is relevant to field data (Foody and Cutler, 2006).

2.6. Conclusion

The study demonstrated the utility of Landsast-8 spectral data for tree species estimation in the savannah woodland. The application of multivariate technique, PCA, facilitated the use of the entire spectral bands in Landsat-8 and produced PCs which explained *H'* (r^2 of 0.36; $p < 0.05$) and D_2 (r^2 of 0.35; $p < 0.05$) better than NDVI or its derivatives (r^2 ranges from 0.10 to

0.29; $p < 0.05$) which had been used frequently for estimating species diversity (Gould, 2000; Parviainen et al. 2010; Pau et al. 2012). Utilizing the entire spectral information in the Landsat-8 data enhanced our ability to estimate tree species diversity better than NDVI, which is limited to red and NIR regions of Landsat data. Furthermore, deriving PCs from a combination of Landsat-8 spectral data and vegetation indices improved the estimation of tree species diversity and this confirmed that multivariate techniques facilitate maximum exploitation of remotely sensed data for the purpose of biodiversity research. Moreover, the study confirmed our assumption that SRI may useful for estimating tree species diversity. SRI regression models produced results that were comparable to those obtained with PCA variables. Whilst NDVI and its derivatives had a significant relationship with tree species diversity, it was lower compared to SRI derivatives and this was attributed to scale differences between these indices. SRI has measurement scale which ranges from 0 to far beyond 1 and such an open scale facilitated its ability to explain tree species diversity. The NDVI scale problem has long been recognized as limiting in its ability to sense forest canopy variation (Huete et al. 2002) and therefore it is not surprising that NDVI had a lower explanatory power than SRI. The study also showed that H' and D_2 are compatible with Landsat spectral variables. H' and D_2 consider both species richness and abundance and these aspects of biodiversity have been shown to relate well with remotely sensed spectral signal.

In light of the results from the present study, further research on the utility of Landsat-8 for estimating tree species diversity should incorporate environmental variables which are known to impact tree species distribution. Integrated modelling involving remote sensing variables and environmental variables have improved the prediction of invasive species in other studies (Malahlela et al. 2015). Overall, the significant relationship observed between remotely sensed variables and tree species diversity measures confirms the utility of Landsat image for practical application in conservation, particularly as a screening tool to identify biodiversity hotspots. The Landsat imagery covers large geographical areas on regular intervals and may provide useful information that is commensurate with the scale of conservation.

Acknowledgements

The authors would like to express gratitude to the National Research Foundation for funding this research project through the NRF-Professional Development Programme. We would also like to thank Cecilia Masemola, Sbu Gumede, Joseph Dlamini, Martin Sarela and Patrick Ndlovu for their assistance during field work.

References

Albregtsen, F., 2008. Statistical texture measures computed from gray level coocurrence matrices. *Image processing laboratory, department of informatics, University of Oslo*, *5*.

Archibald, S. and Scholes, R.J., 2007. Leaf green-up in a semi-arid African savanna–separating tree and grass responses to environmental cues. *Journal of Vegetation Science*, *18*(4), pp.583-594.

Asner, G.P., Levick, S.R., Kennedy-Bowdoin, T., Knapp, D.E., Emerson, R., Jacobson, J., Colgan, M.S. and Martin, R.E., 2009. Large-scale impacts of herbivores on the structural diversity of African savannas. *Proceedings of the National Academy of Sciences*, *106*(12), pp.4947-4952.

Bai, Y., Wu, J., Pan, Q., Huang, J., Wang, Q., Li, F., Buyantuyev, A. and Han, X., 2007. Positive linear relationship between productivity and diversity: evidence from the Eurasian Steppe. *Journal of Applied Ecology*, *44*(5), pp.1023-1034.

Beliakov, G., James, S. and Troiano, L., 2008, June. Texture recognition by using GLCM and various aggregation functions. In *Fuzzy Systems, 2008. FUZZ-IEEE 2008.(IEEE World Congress on Computational Intelligence). IEEE International Conference on* (pp. 1472-1476).

Barnard, E., Cho, M.A., Debba, P., Mathieu, R., Wessels, K., van Heerden, C., van der Walt, C. and Asner, G.P., 2010, November. Optimizing tree species classification in hyperspectral images. In *Proceedings of the Twenty-First Annual Symposium of the Pattern Recognition Association of South Africa, Stellenbosch, South Africa* (pp. 33-38).

Bro, R. and Smilde, A.K., 2014. Principal component analysis. *Analytical Methods*, *6*(9), pp.2812-2831.

Cho, M.A., Debba, P., Mathieu, R., Naidoo, L., Van Aardt, J. and Asner, G.P., 2010. Improving discrimination of savanna tree species through a multiple-endmember spectral angle mapper approach: Canopy-level analysis. *IEEE Transactions on Geoscience and Remote Sensing*, *48*(11), pp.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. and Erasmus, B., 2012. Mapping tree species composition in South African savannas using an integrated airborne spectral and LiDAR system. *Remote Sensing of Environment*, *125*, pp.214-226.

Colwell, R.K., 2009. Biodiversity: concepts, patterns, and measurement. *The Princeton guide to ecology*, pp.257-263.

Dean, W.R.J., Milton, S.J. and Jeltsch, F., 1999. Large trees, fertile islands, and birds in arid savanna. *Journal of Arid Environments*, *41*(1), pp.61-78.

Druce, D.J., Shannon, G., Page, B.R., Grant, R. and Slotow, R., 2008. Ecological thresholds in the savanna landscape: developing a protocol for monitoring the change in composition and utilisation of large trees. *PloS one*, *3*(12), p.e3979.

du Toit, J.T., Biggs, H.C., Rogers, K.H., 2003. The Kruger Experience: ecology and management of savanna heterogeneity. London: Island Press.
Dumalisile, L., 2009. *Effects of Chromolaena odorata on mammalian biodiversity in Hluhluwe-iMfolozi Park, South Africa* (Doctoral dissertation).

Fairbanks, D.H. and McGwire, K.C., 2004. Patterns of floristic richness in vegetation communities of California: regional scale analysis with multi-temporal NDVI. *Global Ecology and Biogeography*, *13*(3), pp.221-235.

Foody, G.M. and Cutler, M.E., 2006. Mapping the species richness and composition of tropical forests from remotely sensed data with neural networks. *Ecological modelling*, *195*(1), pp.37-42.

Fu, W.J., Jiang, P.K., Zhou, G.M. and Zhao, K.L., 2014. Using Moran's I and GIS to study the spatial pattern of forest litter carbon density in a subtropical region of southeastern China. *Biogeosciences*, *11*(8), pp.2401-2409.

Gitelson, A.A., 2004. Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *Journal of plant physiology*, *161*(2), pp.165-173.

Gould, W., 2000. Remote sensing of vegetation, plant species richness, and regional biodiversity hotspots. *Ecological applications*, *10*(6), pp.1861-1870.

Grant, C.C. and Scholes, M.C., 2006. The importance of nutrient hot-spots in the conservation and management of large wild mammalian herbivores in semi-arid savannas. *Biological Conservation*, *130*(3), pp.426-437.

Gringarten, E. and Deutsch, C.V., 2001. Teacher's aide variogram interpretation and modeling. *Mathematical Geology*, *33*(4), pp.507-534.

Haralick, R.M. and Shanmugam, K., 1973. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6), pp.610-621.

He, K.S., Zhang, J. and Zhang, Q., 2009. Linking variability in species composition and MODIS NDVI based on beta diversity measurements. *acta oecologica*, *35*(1), pp.14-21.

Hempson, G.P., Archibald, S. and Bond, W.J., 2015. A continent-wide assessment of the form and intensity of large mammal herbivory in Africa. *Science*, *350*(6264), pp.1056-1061.

Hernández-Stefanoni, J.L., Gallardo-Cruz, J.A., Meave, J.A., Rocchini, D., Bello-Pineda, J. and López-Martínez, J.O., 2012. Modeling α-and β-diversity in a tropical forest from remotely sensed and spatial data. *International journal of applied earth observation and geoinformation*, *19*, pp.359-368.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X. and Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote sensing of environment*, *83*(1), pp.195-213.

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

Huete, A.R. and Jackson, R.D., 1988. Soil and atmosphere influences on the spectra of partial canopies. *Remote Sensing of Environment*, *25*(1), pp.89-105.

Huete, A., Justice, C. and Van Leeuwen, W., 1999. MODIS vegetation index (MOD13). *Algorithm theoretical basis document*, *3*, p.213.

Jakubauskas, M.E. and Price, K.P., 1997. Emperical relationships between structural and spectral factors of yellowstone lodgepole pine forests. *Photogrammetric engineering and remote sensing*, *63*(12), pp.1375-1380.

Jetz, W., Cavender-Bares, J., Pavlick, R., Schimel, D., Davis, F.W., Asner, G.P., Guralnick, R., Kattge, J., Latimer, A.M., Moorcroft, P. and Schaepman, M.E., 2016. Monitoring plant functional diversity from space. *Nature plants*, *2*(3).

Jongman, R.H., Ter Braak, C.J. and Van Tongeren, O.F. eds., 1995. *Data analysis in community and landscape ecology*. Cambridge University press.

Kelsey, K.C. and Neff, J.C., 2014. Estimates of aboveground biomass from texture analysis of Landsat imagery. *Remote Sensing*, *6*(7), pp.6407-6422.

Kerr, J.T. and Ostrovsky, M., 2003. From space to species: ecological applications for remote sensing. *Trends in Ecology & Evolution*, *18*(6), pp.299-305.

Kirkman, L.K., Mitchell, R.J., Helton, R.C. and Drew, M.B., 2001. Productivity and species richness across an environmental gradient in a fire-dependent ecosystem. *American Journal of Botany*, *88*(11), pp.2119-2128.

Lande, R., 1996. Statistics and partitioning of species diversity, and similarity among multiple communities. *Oikos*, pp.5-13.

Landsat-8 (L8) Data User Handbook, Version 2.0, March 29, 2016.

Ludwig, F., De Kroon, H., Berendse, F. and Prins, H.H., 2004. The influence of savanna trees on nutrient, water and light availability and the understorey vegetation. *Plant Ecology*, *170*(1), pp.93-105.

Madonsela, S., Cho, M.A., Mathieu, R., Mutanga, O., Ramoelo, A., Kaszta, Ż., Van De Kerchove, R. and Wolff, E., 2017. Multi-phenology WorldView-2 imagery improves remote sensing of savannah tree species. *International Journal of Applied Earth Observation and Geoinformation*, *58*, pp.65-73.

Madubansi, M. and Shackleton, C.M., 2006. Changing energy profiles and consumption patterns following electrification in five rural villages, South Africa. *Energy Policy*, *34*(18), pp.4081-4092.

Main, R., Cho, M.A., Mathieu, R., O'Kennedy, M.M., Ramoelo, A. and Koch, S., 2011. An investigation into robust spectral indices for leaf chlorophyll estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*(6), pp.751-761.

Makhado, R.A., Mapaure, I., Potgieter, M.J., Luus-Powell, W.J. and Saidi, A.T., 2014. Factors influencing the adaptation and distribution of Colophospermum mopane in southern Africa's mopane savannas-A review. *Bothalia-African Biodiversity & Conservation*, *44*(1), pp.1-9.
Malahlela, O.E., Cho, M.A. and Mutanga, O., 2015. Mapping the occurrence of Chromolaena odorata (L.) in subtropical forest gaps using environmental and remote sensing data. *Biological Invasions*, *17*(7), pp.2027-2042.

"MATLAB R2009a, the Language of Technical Computing," Mathworks Inc.

Matsika, R., Erasmus, B.F. and Twine, W.C., 2013. A tale of two villages: assessing the dynamics of fuelwood supply in communal landscapes in South Africa. *Environmental Conservation*, *40*(1), pp.71-83.

Morris, E.K., Caruso, T., Buscot, F., Fischer, M., Hancock, C., Maier, T.S., Meiners, T., Müller, C., Obermaier, E., Prati, D. and Socher, S.A., 2014. Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories. *Ecology and evolution*, *4*(18), pp.3514-3524.

Nagendra, H., 2002. Opposite trends in response for the Shannon and Simpson indices of landscape diversity. *Applied Geography*, *22*(2), pp.175-186.

Nagendra, H., Rocchini, D., Ghate, R., Sharma, B. and Pareeth, S., 2010. Assessing plant diversity in a dry tropical forest: Comparing the utility of Landsat and IKONOS satellite images. *Remote Sensing*, *2*(2), pp.478-496.

Naidoo, L., Mathieu, R., Main, R., Kleynhans, W., Wessels, K., Asner, G. and Leblon, B., 2015. Savannah woody structure modelling and mapping using multi-frequency (X-, C-and L-band) Synthetic Aperture Radar data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *105*, pp.234-250.

Oindo, B.O. and Skidmore, A.K., 2002. Interannual variability of NDVI and species richness in Kenya. *International journal of remote sensing*, *23*(2), pp.285-298.

Oldeland, J., Wesuls, D., Rocchini, D., Schmidt, M. and Jürgens, N., 2010. Does using species abundance data improve estimates of species diversity from remotely sensed spectral heterogeneity?. *Ecological Indicators*, *10*(2), pp.390-396.

Parviainen, M., Luoto, M. and Heikkinen, R.K., 2010. NDVI-based productivity and heterogeneity as indicators of plant-species richness in boreal landscapes. *Boreal environment research*, *15*(3).

Patel, N.K., Saxena, R.K. and Shiwalkar, A.J.A.Y., 2007. Study of fractional vegetation cover using high spectral resolution data. *Journal of the Indian Society of Remote Sensing*, *35*(1), pp.73-79.

Pau, S., Gillespie, T.W. and Wolkovich, E.M., 2012. Dissecting NDVI–species richness relationships in Hawaiian dry forests. *Journal of Biogeography*, *39*(9), pp.1678-1686.

Pellegrini, A.F., Socolar, J.B., Elsen, P.R. and Giam, X., 2016. Trade‐offs between savanna woody plant diversity and carbon storage in the Brazilian Cerrado. *Global change biology*, *22*(10), pp.3373-3382.

Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C. and Coops, N.C., 2013. Essential biodiversity variables. *Science*, *339*(6117), pp.277-278.

Pervez, W., Uddin, V., Khan, S.A. and Khan, J.A., 2016. Satellite-based land use mapping: comparative analysis of Landsat-8, Advanced Land Imager, and big data Hyperion imagery. *Journal of Applied Remote Sensing*, *10*(2), pp.026004-026004.

Richter, R. and Schläpfer, D., 2012. Atmospheric/Topographic Correction for Satellite Imagery (ATCOR-2/3 User Guide, Version 8.2 BETA). *German Aerospace Center, Remote Sensing Data Center: Wessling, Germany*.

Rocchini, D., 2007. Effects of spatial and spectral resolution in estimating ecosystem α-diversity by satellite imagery. *Remote Sensing of Environment*, *111*(4), pp.423-434.

Rossel, R.A.V., 2008. ParLeS: Software for chemometric analysis of spectroscopic data. *Chemometrics and intelligent laboratory systems*, *90*(1), pp.72-83.

Rouse Jr, J., Haas, R.H., Schell, J.A. and Deering, D.W., 1974. Monitoring vegetation systems in the Great Plains with ERTS.

Scholes, R.J. and Archer, S.R., 1997. Tree-grass interactions in savannas. *Annual review of Ecology and Systematics*, *28*(1), pp.517-544.

Seymour, C.L. and Dean, W.R.J., 2010. The influence of changes in habitat structure on the species composition of bird assemblages in the southern Kalahari. *Austral Ecology*, *35*(5), pp.581-592.

Shackleton, C.M., 2000. Comparison of plant diversity in protected and communal lands in the Bushbuckridge lowveld savanna, South Africa. *Biological Conservation*, *94*(3), pp.273-285.

Shannon, C.E., 1948. A mathematical theory of communication, Part I, Part II. *Bell Syst. Tech. J.*, *27*, pp.623-656.
Simpson, E.H., 1949. Measurement of diversity. *Nature*.

Thenkabail, P.S., Hall, J., Lin, T., Ashton, M.S., Harris, D. and Enclona, E.A., 2003. Detecting floristic structure and pattern across topographic and moisture gradients in a mixed species Central African forest using IKONOS and Landsat-7 ETM+ images. *International Journal of Applied Earth Observation and Geoinformation*, *4*(3), pp.255-270.

Thenkabail, P.S., Enclona, E.A., Ashton, M.S. and Van Der Meer, B., 2004. Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote sensing of environment*, *91*(3), pp.354-376.

Todd, S.W. and Hoffer, R.M., 1998. Responses of spectral indices to variations in vegetation cover and soil background. *Photogrammetric engineering and remote sensing*, *64*, pp.915-922.

Treydte, A.C., Heitkönig, I.M., Prins, H.H. and Ludwig, F., 2007. Trees improve grass quality for herbivores in African savannas. *Perspectives in Plant Ecology, Evolution and Systematics*, *8*(4), pp.197-205.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment*, *8*(2), pp.127-150.

Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E. and Steininger, M., 2003. Remote sensing for biodiversity science and conservation. *Trends in ecology & evolution*, *18*(6), pp.306-314.

Viña, A., Gitelson, A.A., Nguy-Robertson, A.L. and Peng, Y., 2011. Comparison of different vegetation indices for the remote assessment of green leaf area index of crops. *Remote Sensing of Environment*, *115*(12), pp.3468-3478.

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

Wood, E.M., Pidgeon, A.M., Radeloff, V.C. and Keuler, N.S., 2013. Image texture predicts avian density and species richness. *PloS one*, *8*(5), p.e63211.

Zimmermann, N.E., Edwards, T.C., Moisen, G.G., Frescino, T.S. and Blackard, J.A., 2007. Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. *Journal of applied ecology*, *44*(5), pp.1057-1067.

Chapter 3: Investigation into whether there is a significant interaction between seasonal NDVI and woody cover when estimating tree species diversity

This chapter is based on

Madonsela, S., Cho, M.A., Ramoelo, A., Mutanga, O. and Naidoo, L., (2018). Estimating tree species diversity in the savannah using NDVI and woody canopy cover. *International Journal of Applied Earth Observation and Geoinformation Vol. 66, pp. 106-115.*

Abstract

Remote sensing applications in biodiversity research often rely on the establishment of relationships between spectral information from the image and tree species diversity measured in the field. Most studies have used normalized difference vegetation index (NDVI) to estimate tree species diversity on the basis that it is sensitive to primary productivity which defines spatial variation in plant diversity. The NDVI signal is influenced by photosynthetically active vegetation which, in the savannah, includes woody canopy foliage and grasses. The question is whether the relationship between NDVI and tree species diversity in the savanna depends on the woody cover percentage. This study explored the relationship between woody canopy cover (WCC) and tree species diversity in the savannah woodland of southern Africa and also investigated whether there is a significant interaction between seasonal NDVI and WCC in the factorial model when estimating tree species diversity. To fulfil our aim, we followed stratified random sampling approach and surveyed tree species in 68 plots of 90m X 90m across the study area. Within each plot, all trees with diameter at breast height of >10cm were sampled and Shannon index - a common measure of species diversity which considers both species richness and abundance - was used to quantify tree species diversity. We then extracted WCC in each plot from existing fractional woody cover product produced from Synthetic Aperture Radar (SAR) data. Factorial regression model was used to determine the interaction effect between NDVI and WCC when estimating tree species diversity. Results from regression analysis showed that (i) WCC has a highly significant relationship with tree species diversity ($r^2 = 0.21$; $p < 0.01$), (ii) the interaction between the NDVI and WCC is not significant, however, the factorial model significantly reduced the error of prediction (RMSE = 0.47, $p < 0.05$) compared to NDVI (RMSE = 0.49) or WCC (RMSE = 0.49) model during the senescence period. The result justifies our assertion that combining NDVI with WCC will be optimal for biodiversity estimation during the senescence period.

3.1. Introduction

Savannah ecosystems are characterized by co-occurrence of woody and herbaceous lifeforms (du Toit et al. 2003; Sankaran et al. 2005) and a high assemblage of floral and faunal diversity with important roles in the system (Shackleton, 2000; du Toit et al. 2003). In particular, tree species provides multiple benefits to the savannah ecosystem by maintaining nutrients in the system (Treydte et al. 2007), providing breeding sites for birds (Seymour and Dean, 2010), supporting large faunal species (Hempson et al. 2015) and also acts as a safety net against poverty in the neighbouring communities delivering goods such fuelwood, timber and medicinal products (Shackleton et al. 2007; Matsika et al. 2012). Therefore, the loss of tree species diversity impacts negatively on the functioning of the ecosystem and the benefits it provides. In southern African savannah, trees are heavily impacted upon by elephants (Druce et al. 2008) and human activities (Shackleton, 2000). South Africa's National Park, for instance, have developed Threshold of Potential Concerns (TPCs) which serves as a monitoring system to detect changes that may impact on key elements of biodiversity (Gillson and Duffin, 2007; Druce et al. 2008). The success of such monitoring systems depends on the availability of spatially detailed and updated information on the distribution patterns and abundance of species (Turner et al. 2003). Remote sensing data meet these needs as it covers large geographic areas on a regular interval and at varying levels of spatial details (Jetz et al. 2016; Kerr and Ostrovsky, 2003). Recently, ecologists have embraced remote sensing science in order to study biodiversity and prepare conservation responses to potential threats (Jetz et al. 2016; Pereira et al. 2013).

The application of remote sensing in biodiversity research often relies on establishing relationships between spectral information from the image and tree species diversity measured in the field (Gould, 2000; Parviainen et al. 2010; Hernandez-Stefanoni et al. 2012). Studies have shown that the success of remote sensing application in biodiversity estimation depends highly on the spectral resolution of the data (Thenkabail et al. 2003; Rocchini et al. 2007; Nagendra et al. 2010; Cho et al. 2012). Remote sensing systems e.g. Landsat program collects essential spectral information in the visible, near infrared and middle infrared regions which relates to plant properties including leaf pigment, water content and plant internal structure (Hernandez-Stefanoni et al. 2012; Nagendra et al.

2010). Consequently, the Landsat program has performed comparably or better than high spatial but limited spectral resolution multispectral sensors such as Quickbird and IKONOS when estimating forest characteristics (Thenkabail et al. 2003; Rocchini et al. 2007). For instance Thenkabail et al. (2003) observed that Landsat Thematic Mapper plus explain floristic structure better than IKONOS in Dzanga–Sangha Dense Forest Reserve, Central African Republic and attributed the higher explanatory power from Landsat to two shortwave infrared bands not present in IKONOS.

However, most studies e.g. Gould, (2000); Parviainen et al. (2010); Wood et al. (2013) testing Landsat data for estimating tree species diversity have focused only on the red and near infrared bands present in most remote sensing devices. Vegetation indices particularly the NDVI is derived from these two bands and often showed a positive relationship with species diversity in different biomes (Gould, 2000; He et al. 2009; Parviainen et al. 2010). Our previous study (Madonsela et al. 2017) also observed a significant relationship between mean NDVI and tree species diversity in the savannah biome. In essence, vegetation indices are formulated to suppress spectral reflectance from non-vegetative features while enhancing the spectral content from vegetation (Viña et al. 2006). Moreover, the NDVI is sensitive to essential environmental factors such as rainfall which impact on biodiversity (Pau et al. 2012; Seto et al. 2004; Box et al. 1989). The amount of energy available in an ecosystem detectable with NDVI as primary productivity defines spatial variation in plant diversity (Parviainen et al. 2010; Witman et al. 2008). It is therefore not surprising that NDVI has frequently been successful in estimating tree species diversity in different biomes at various scales (Oindo and Skidmore, 2000; Gould, 2000; Pau et al. 2012; Madonsela et al. 2017b).

The success of Landsat-derived NDVI in estimating tree species diversity raises the question whether tree species diversity is more related to woody canopy cover (i.e. a proxy for woodland productivity) or to the entire productivity trees and grass represented by the NDVI? The research question is informed by the fact that NDVI signal is influenced by photosynthetically active vegetation which, in savannahs, includes woody canopy foliage and grasses. Individually, tree productivity has been shown to be positively related to NDVI (Wang et al. 2004). The question therefore sought to establish the predictive performance of woody vegetation without grass influence. WCC represents the percentage of horizontal vegetated area of the trees (Gonsamo et al. 2013; Naidoo et al. 2015) and is the simplest measure of vegetation structure (Mathieu et al. 2013). While the question of woody cover – tree species diversity has been investigated in North American savannah (Peterson and Reich, 2008), this study extend the question to investigate the interaction between structural variables (woody cover) and NDVI when estimating tree species diversity.

In this study, the woody canopy cover was derived from winter L-band Synthetic Aperture Radar image which interact with vegetation structure i.e. tree trunk and canopy branches. Detailed LiDAR woody cover maps were used as calibration and test data to develop Random Forest model for extrapolating woody cover to the southern African savannah with SAR data (Naidoo et al. 2015). Essentially the tree canopy cover used in this study represents structural information of woody vegetation. Meanwhile the NDVI signal is influenced by tree canopy foliage, underlying grass and canopy background and tends to vary with changes in vegetation phenology. The question is whether combining NDVI from different phenological periods with woody canopy cover in a factorial model improves the estimation of tree species diversity in the savannah woodland. The aim of the study is to investigate whether there is a significant interaction between seasonal NDVI and woody cover when estimating tree species diversity. The study will also investigate whether there is a significant relationship between woody canopy cover and tree species diversity across savannah woodland belt.

3.2. Study area

The study area stretches across the KwaZulu-Natal (KZN), Mpumalanga and Limpopo provinces of South Africa, within the broader savannah woodland belt (**Figure 9**). The area falls within two land management regimes; i) the Kruger National Park (KNP) and Hluhluwe-Imfolozi (HIP) Park which are public nature reserve entities with a mandate to conserve savannah biodiversity and ii) the communal areas adjacent these conservation areas. Typical of the savannah biome, the study area is characterized by the co-occurrence of two lifeforms; the continuous herbaceous layer interspersed by woody vegetation cover (Scholes and Archer, 1997; Sankaran et al. 2005). Fire and rainfall in particular and herbivory are key mechanisms that maintain balanced distributional patterns between these two lifeforms in the savannah in general (Sankaran et al. 2005; Bond et al. 2003). Geologically, the western part of the area is dominated by granite substrate while gabbro substrate dominates in the eastern part. Tree-grass density ratio tends to be defined by these geological structures. Gabbro substrates is characterized by shallow to moderately deep, dark clay soils with high-bulk, nutritious grasses and support few scattered trees mainly *Acacia spp.* (du Toit et al. 2003).

On the contrary, granite substrate is defined by nutrient-poor, shallow to moderately deep sandy soils with gently undulating terrain and it hosts broad-leaved deciduous tree species upslope while fine-leaved species occupy downslope. The granitic substrates are characterized by high species diversity and notable tree species includes *Combretum spp*, *Acacia nigrescens*, *Spirostachys africana* and *Sclerocarya birrea* (du Toit et al. 2003; Eckhardt et al. 2000). The northern portion of the study area is also characterized by the dominance of *Colophospermum mopane* (Makhado et al. 2013; Eckhardt et al. 2000). Meanwhile, the KZN part of the study area is characterized by mountainous terrain with different habitat types supporting a large number of plant species. Typical savannah species includes *Dichrostachys cinerea* and various species of *Euclea* and *Acacia* (Dumalisile, 2009). The area is characterized by north-south rainfall gradient. The mean annual precipitation ranges from 750mm in the southern portion of KNP to 440mm in the north with notable variations around the mean from year to year (Makhado et al. 2013; Eckhardt et al. 2000).

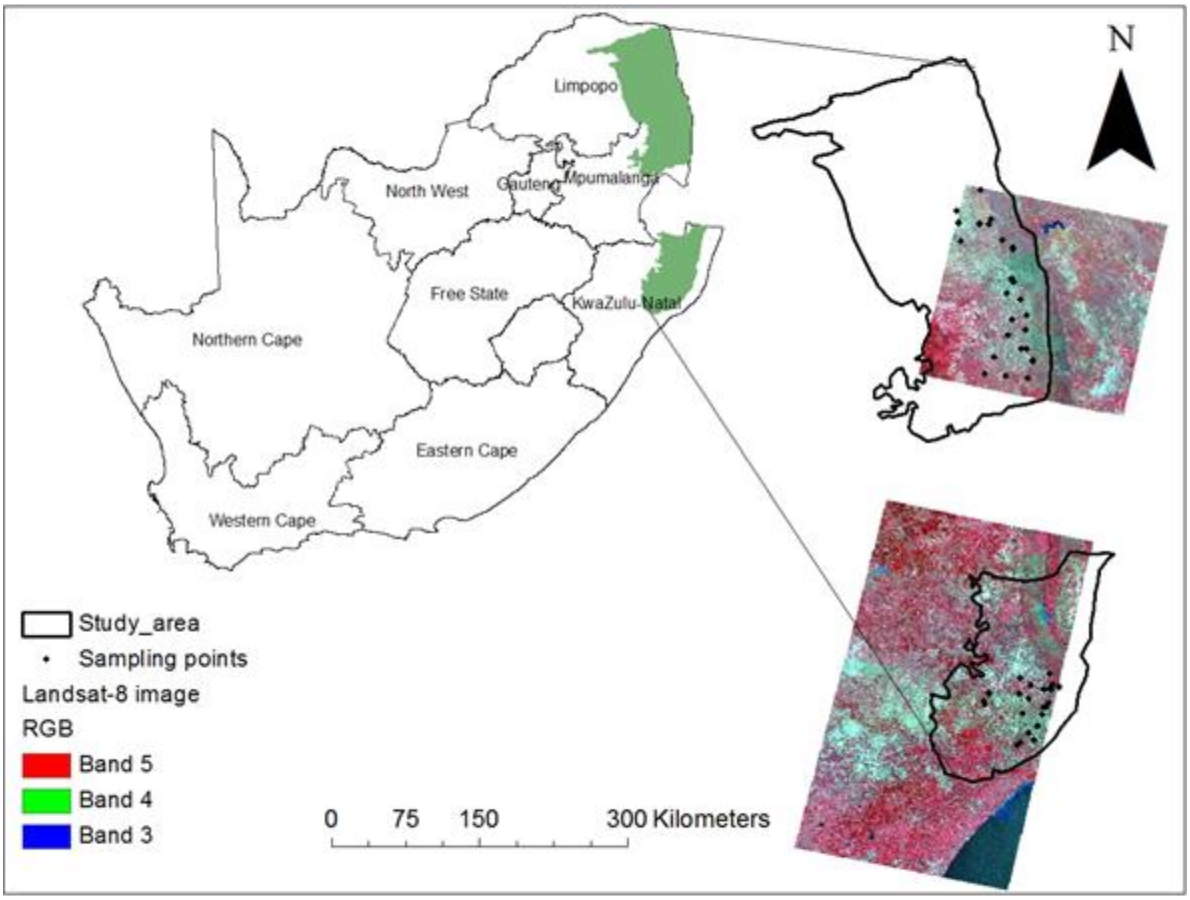

Figure 9 Study area stretching across three provinces of South Africa. The black dots represent the sampling plots

3.3. Material and Methods

3.3.1. Remote sensing data

Four Landsat-8 Operational Land Imager (OLI) satellite images captured in 2016 (28th of March, 29th of April, 31st of May and 24th of July) were downloaded from the United States Geological Surveys (USGS) portal (https://earthexplorer.usgs.gov/). These images were collected in different dates in order to examine the interaction of NDVI and woody canopy cover across different phenological periods. The end of March represents the end of growing season (Grant and Scholes, 2006); April represents transition to senescence (Madonsela et al. 2017a); May represents advanced senescence when most trees starts to drop off leaves and grass will be at their senescent stage (Scholes et al. 2003; Cho et al. 2010); July corresponds to dry season in southern African savannah (du Toit et al. 2003; Kaszta et al. 2016). NDVI was computed from each Landsat image and the Landsat-derived NDVI of March, April, May and July are referred to as $NDVI_{March,}$ $NDVI_{April}$, $NDVI_{May}$ and $NDVI_{July}$ respectively.

Landsat-8 OLI is a multi-spectral sensor with eight spectral bands in the visible, near infrared and shortwave infrared regions of electromagnetic spectrum. Landsat-8 OLI record data at moderate spatial resolution of 30m and has a revisit capacity of 16 days. The 12-bit quantization of data has improved the signal-to-noise radiometric performance of the sensor over its predecessors, thus increasing its usefulness for landcover mapping (Pervez et al. 2016). The Landsat-8 images were downloaded with geometric correction already completed. The Mpumalanga and Limpopo images were atmospherically corrected using the ATCOR-2 software since the area exhibit gently undulating slopes (Richter and Schläpfer, 2012). The KZN Landsat scenes necessitated the use of ATCOR-3 software since the region is mountainous. ATCOR-3 allows for integration of DEM which is useful for the correction of shadow and topographic effects on the image depicting mountainous areas (Richter and Schläpfer, 2012).

In addition, a woody fractional cover map derived from LIDAR and Synthetic Aperture Radar (SAR) data was used to extract the woody canopy cover from each field plot. The canopy cover represents the percentage of horizontal area covered by the vertical projection of woody canopy elements (Gonsamo et al. 2013). The dataset was produced with the 2010 L-band ALOS PALSAR mosaics released by the Japanese Space Agency JAXA, according to the methods detailed in Naidoo et al. (2015) and the National Terrestrial Carbon Sink Assessment (2015). Extensive LiDAR tracks were processed to develop a canopy height model of all woody vegetation above 1m. Detailed LiDAR woody cover maps were derived from the canopy height model at 25m pixel size, and were used as calibration and test data to develop a Random Forest model for extrapolating the woody cover to the South African biome with the dual-polarized (HV, HH) SAR data. The woody fractional cover map was produced at 25m resolution and with a root mean square error of 13.53%.

3.3.2. Field data collection

Field data were collected from the 2nd till the 27th of November 2015 in KwaZulu-Natal and again on the 1st till the 19th of March 2016 across Kruger National Park stretching between Mpumalanga and Limpopo provinces. The principal aim of the field campaign was to identify tree species within randomly placed sampling plots and quantify local species diversity (α-diversity) in the region using the common measure of diversity i.e. Shannon index. Prior to field excursion we defined the size of field sampling plots using semi-variogram analysis in

ENVI 4.8 software. Essentially semi-variogram quantifies the spatial variability of natural phenomenon occurring in space (Fu et al., 2014; Gringarten and Deutsch, 2001). Semi-variogram is computed as follow:

Equation 3

$$y(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2$$

where $y(h)$ is the semi-variance at a given distance h; $z(x_i)$ is the value of the variable Z at location x_i; h is the lag distance and $N(h)$ is the number of pairs of sample points separated by h.

Semi-variance steadily increases as the distance from one location to the next increases till it reaches the range where it starts to level off (Jongman et al. 1995; Gringarten and Deutsch, 2001). Semi-variogram plot is generated by computing variance at different lag distances and a theoretical model such as spherical or exponential model is fitted to provide information about spatial structure (Fu et al. 2014). Our study applied semi-variogram analysis to resampled WorldView-2 derived NDVI image to define the scale of spatial variability in tree species richness. The choice to use NDVI was based on the observation that variability in NDVI is related to species diversity (Gould, 2000).

In our analysis, the Worldview-2 image – covering only a small part of the study area - was firstly resampled to 10m spatial resolution to be compatible with average tree canopy size in the savannah (Cho et al. 2012) and then we generated NDVI image. In ENVI software v4.8 the semi-variogram analysis computed the squared difference between neighbouring pixel values in order to quantify variability. The analysis conducted on Worldview-2 derived NDVI image showed that the scale for tree species variability in the savannah woodland lies at a range of 90m (**Figure 10**). Although semi-variance kept increasing beyond the range, the increase was not consistent and the range of 90m resulted in plot sizes that are feasible to work on within limited resources. Moreover, the study intended to use Landsat data with 30m pixel resolution, hence the plot size of 90m X 90m was considered adequate to ascertain correspondence between field data and spectral data.

The plot size of 90m X 90m was therefore chosen to capture spatial variation in tree species diversity. Stratified random sampling was used to define the placement of sampling plots.

The stratification of sampling plots followed four dominant geological formations (granite; siliciclastic; gabbros; granulite) that were observed to have marked influence over vegetation patterns in the study area (du Toit et al. 2003). Plots of 90m x 90m were designed ensuring that corners of each plot correspond to Landsat pixels by following pre-defined GPS points of each corner. Within the plots all trees with diameter at breast height (DBH) above 10cm were recorded with Global Positioning System and species identified. Eventually we collected 5859 trees belonging to 106 tree species. The field campaign visited 50 plots distributed across the study area and collected tree species data. Further 26 plots collected under similar conditions in the previous study (Naidoo et al. 2015) were added to our field data. However some of these field plots (8 plots) were located on clouded parts of the March and April images and therefore not usable. In total 68 field plots were used in the analysis. We also extracted mean annual rainfall for each plot from the interpolated rainfall data produced by South African National Parks Scientific Services.

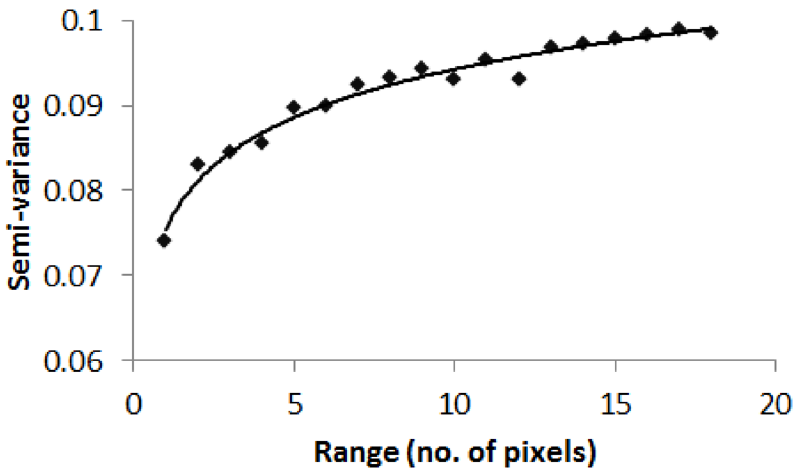

Figure 10 Semi-variogram analysis showing the scale of tree species variability in the savannah woodland

3.3.3. Data analysis

We quantified α-diversity within each plot using the Shannon index (*H'*) which is common measure of diversity in ecological literature (Colwell, 2009; Morris et al. 2014) and was preferred to ensure consistency of our findings with previous studies. *H'* considers both species richness (i.e. number of different tree species) and abundance (i.e. number of individual trees within species) when quantifying species diversity (Shannon and Weaver, 1949; Morris et al. 2014) and these aspects of diversity are considered to have a bearing on the reflectance spectra captured by remote sensing device (Oldeland et al. 2010; Madonsela et al. 2017b). Moreover, Shannon index is considered to be sensitive to vegetation structure (Oldeland et al. 2010). Therefore Shannon index should relate well with spectral data and vegetation structural variables. *H'* is computed as follows:-

Equation 4

$$H' = -\sum_{i=1}^{S} p_i \mathrm{In}(p_i)$$

where p_i is the proportional abundance of species i relative to the total abundance of all species *S* in a plot; $\mathrm{In}(p_i)$ is the natural logarithm of this proportion.

The sampling plots used to collect the tree species data in the field were overlaid on each Landsat-8 NDVI image. We then extracted the mean statistics from each NDVI image corresponding to each plot. The sampling plots were further used to extract mean woody canopy cover from the SAR derived woody cover map. Subsequently, factorial design model was used to analyse interactive effects between NDVI and woody cover when estimating tree species diversity. In essence factorial design model defines the effect of each predictor variable on the response variable. The model also defines the effect of interaction of predictors on the response variable (Gottipati and Mishra, 2010; Dahbi et al. 2015).

In this study, we firstly established 30 random permutations of the original data and then split two-thirds of the data for calibrating the models and used the remainder for evaluating the predictive ability of the models. The study investigated the interaction effect between seasonal NDVI and woody canopy cover when modelling tree species diversity using factorial model. Variance partitioning analysis (VPA) was applied to show the explanatory power of each predictor variable in the factorial model. VPA involved calculating analysis of variance and partitioning the proportion of the sum of squares attributable to each predictor variable and their interaction (NDVI*WCC) relative to the total sum of squares (Watling et al. 2015). We also implemented linear regression model to explore the relationship between WCC and tree species diversity. The strength of the relationship was assessed using coefficient of determination (r^2) and *p-value* statistics and the model performance was evaluated using the root mean square error (RMSE). The best regression models i.e. models with maximum r^2 and the lowest RMSE from 30 bootstrapped iterations, were used to produce tree species diversity maps. The maps covers only the Kruger National Park and this was done deliberately to avoid areas that are affected by human activities.

3.4. Results

3.4.1. Relationship between woody canopy cover and tree species diversity

The results of linear regression models shows that woody canopy cover had a significant but lower relationship with Shannon index (r^2 = 0.13; $p < 0.05$) when compared to end of growing season $\text{NDVI}_{\text{March}}$ (r^2 = 0.24; $p < 0.01$) or transition to senescence $\text{NDVI}_{\text{April}}$ (r^2 = 0.19; $p < 0.01$) (**Table 7**). However, these results were negatively influenced by mono-species stand in the northern part of the study area dominated by *Colophospermum mopane*. For instance, two sample plots located in the northern part of KNP (plot 18 and 19) were exclusively occupied by *Colophospermum mopane* resulting in the lack of diversity. In addition, plot 7 and 13 had high tree species diversity yet woody canopy cover was very low (less than 13% woody canopy cover) presumably because of elephant damage. The removal of these plots as outliers, improved the relationship between woody canopy cover and tree species diversity by 8% (r^2 = 0.21; p <0.01) (**Table 8**). Consequently, the woody canopy cover model explained 21% of tree species diversity, although the improvement was still lower than the variance explained by $\text{NDVI}_{\text{March}}$. The $\text{NDVI}_{\text{March}}$ model explained 33% of tree species diversity after the removal of outliers.

In the April date which represents the transition to senescence, the $\text{NDVI}_{\text{April}}$ model still had a higher explanatory power (r^2 of 0.27; $p < 0.01$) than woody canopy cover model (r^2 = 0.21; p <0.01) (**Table 8**). However, the woody canopy cover model performed approximately the same as senescence season NDVI model (NDVI_{May}) when estimating tree species diversity (r^2 of 0.21 and 0.20 respectively). Moreover, woody canopy cover model performed better than dry season NDVI model ($\text{NDVI}_{\text{July}}$) ($r^2$ of 0.21 and 0.13 respectively) signalling a decline in NDVI performance with changes in phenology.

Meanwhile, the scatterplots show that there is positive linear relationship between the predictor variables (woody canopy cover and seasonal NDVI) and Shannon index (**Fig. 11a, b, c, d and e**). This relationship between the predictor variables and Shannon index is partly controlled by the general rainfall gradient observed in the study area. For instance, woody canopy cover, NDVI and Shannon diversity index all shows an increase with increasing mean annual rainfall (**Fig. 12a, b and c**). It is therefore not surprising that both woody canopy cover and the NDVI had a significant relationship with Shannon index given that they are sensitive to abiotic factors e.g. rainfall impacting on tree species diversity.

Table 7 Relationship observed between Shannon index of diversity and woody canopy cover (WCC) and NDVI from different phenological period prior the removal of outliers. All computations were drawn from 30 bootstrapped iterations.

Response variable	Predictor variables	Average r^2	95%CI	*P-value*	Average RMSE
Shannon index	WCC	0.13	±0.025	0.012	0.537
	$\mathrm{NDVI}_{\mathrm{March}}$	0.24	±0.018	0.003	0.497
	$\mathrm{NDVI}_{\mathrm{April}}$	0.19	±0.026	0.006	0.518
	$\mathrm{NDVI}_{\mathrm{May}}$	0.13	±0.030	0.013	0.533
	$\mathrm{NDVI}_{\mathrm{July}}$	0.06	±0.024	0.021	0.554

CI = confidence interval

Table 8 Relationship observed between Shannon index of diversity and woody canopy cover (WCC) and NDVI from different phenological period after the removal of outliers. All computations were drawn from 30 bootstrapped iterations after removing outliers.

Response variable	Predictor variables	Average r^2	95%CI	*P-value*	Average RMSE
Shannon index	WCC	0.21	±0.015	0.0040	0.487
	$\mathrm{NDVI}_{\mathrm{March}}$	0.33	±0.019	0.0001	0.454
	$\mathrm{NDVI}_{\mathrm{April}}$	0.27	±0.022	0.0008	0.464
	$\mathrm{NDVI}_{\mathrm{May}}$	0.20	±0.019	0.0048	0.492
	$\mathrm{NDVI}_{\mathrm{July}}$	0.13	±0.019	0.0119	0.513

CI = confidence interval

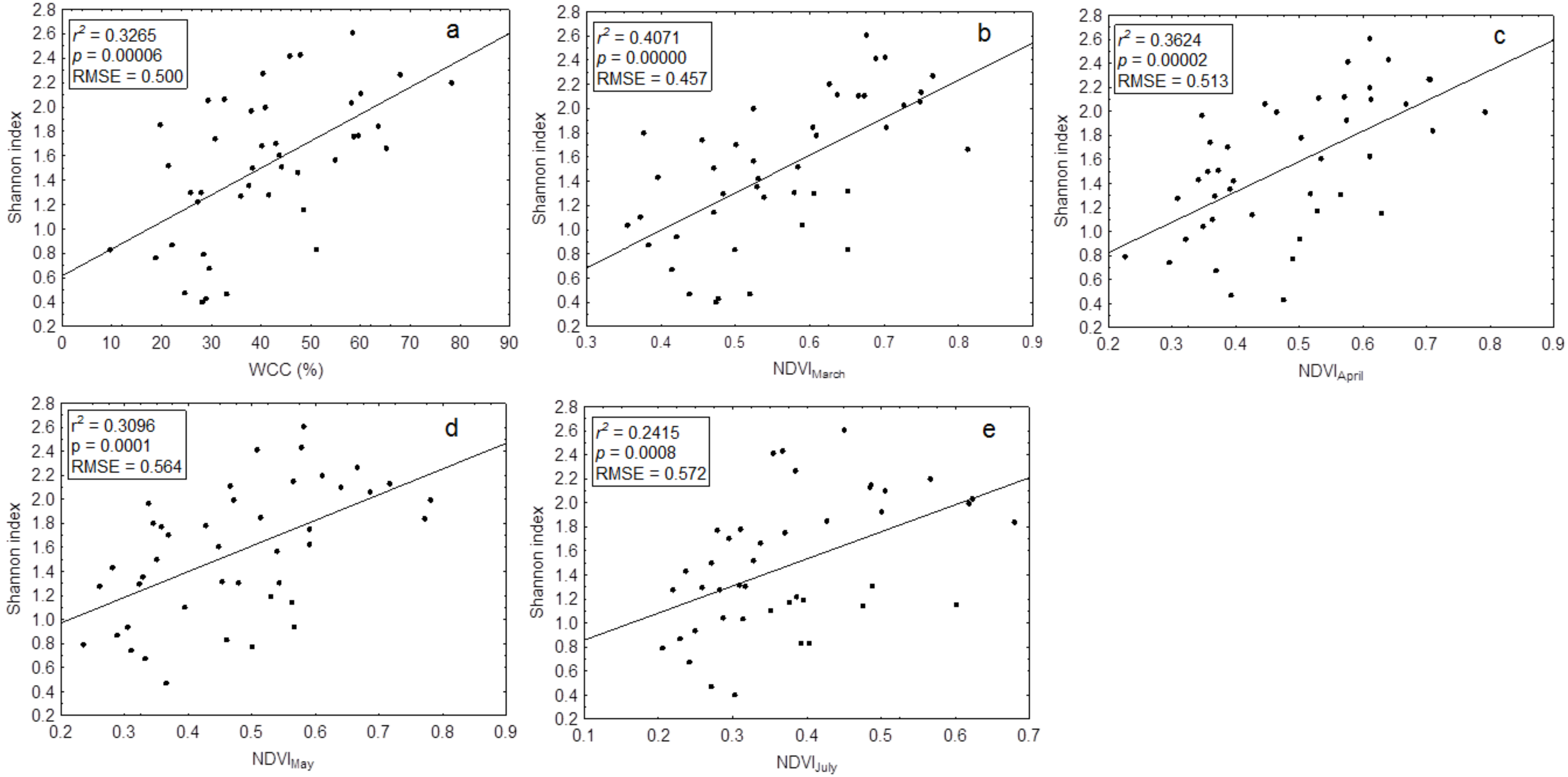

Figure 11 Scatterplot showing linear relationship between predictor variables (woody canopy cover and NDVI) and Shannon index after removing outliers. The scatterplots were selected from the best regression models (maximum r^2 with the lowest RMSE from 30 bootstrapped iterations).

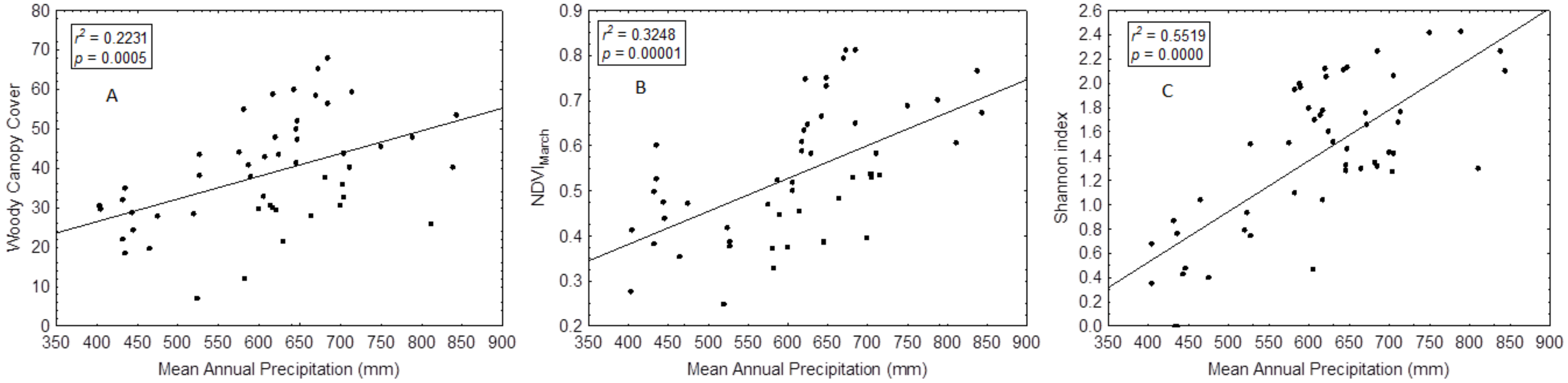

Figure 12 Linear relationship between mean annual precipitation and a) woody canopy cover b) $NDVI_{March}$ and c) Shannon index. The rainfall data covers only the Kruger National Park.

3.4.2. Factorial regression model

The factorial regression models performed better than NDVI or woody canopy cover model when estimating tree species diversity (**Table 7 and 9**). However the results in table 7 and 9 were affected by outliers and the improvements seen with factorial models did not reduce the root mean square error (p >0.05). It was only the combination of $\text{NDVI}_{\text{March}}$ and woody canopy cover in a factorial model that had improved the estimation of Shannon index (r^2 = 0.28; p <0.01) and significantly reduced the RMSE (p < 0.05) when compared to woody cover model (r^2 = 0.13; p <0.05). However the same combination had a higher error of prediction than $\text{NDVI}_{\text{March}}$ model in **Table 7** (average RMSE of 0.508 and 0.497 respectively and the difference was not statistically significant (p > 0.05)).

Post the removal of outliers, factorial model improved estimation of Shannon index while significantly reducing the error of prediction (p <0.01) (**Table 11**). In particular the combination of NDVI_{May} with woody canopy cover in a factorial model had a higher relationship with Shannon index (r^2 = 0.30; p <0.01) than NDVI_{May} model (r^2 = 0.20; p <0.01) or woody cover canopy model (r^2 = 0.21; p <0.01). In addition the same factorial model significantly reduced the error of prediction (p <0.05) compared to NDVI_{May} or woody cover model (**Table 13 and 14**).

However, combining $\text{NDVI}_{\text{July}}$ and woody canopy cover in a factorial model did not significantly reduce the error of prediction when compared to woody canopy cover model (p >0.05) although it had improved the estimation of Shannon index (r^2 = 0.29; p< 0.01). Nonetheless, the same factorial model had significantly lower prediction errors when compared to $\text{NDVI}_{\text{July}}$ model (p <0.01) (**Table 13 and 14**). Combining either $\text{NDVI}_{\text{March}}$ or $\text{NDVI}_{\text{April}}$ with woody canopy cover in a factorial model did not significantly reduce the error of prediction when compared to either $\text{NDVI}_{\text{March}}$ model or $\text{NDVI}_{\text{April}}$ model (p > 0.05). Nonetheless, the same factorial models had significantly lower prediction errors when compared to woody canopy cover model (p <0.001) (**Table 13 and 14**).

The factorial model results present three observations; i) at the end of growing season or during the transition to senescence the NDVI model is optimal for estimating tree species diversity in southern African savannah; ii) during the senescence period combining NDVI_{May} and woody canopy cover significantly improve the estimation of tree species diversity in southern African savannah than either NDVI or woody canopy cover model; iii) woody canopy canopy cover is optimal for estimating tree species diversity during dry season in the savannah.

Concurrent with these observations variance partitioning revealed that $NDVI_{March}$ and $NDVI_{April}$ had a higher explanatory power than woody canopy cover in the factorial model both at the end of growing season (NDVI had r^2 of 0.27 while WCC had r^2 of 0.08) and during transition to senescence (NDVI had r^2 of 0.22 while WCC had r^2 of 0.10) (**Table 12**). Meanwhile, $NDVI_{May}$ and woody canopy cover had approximately equal explanatory power in the model during the senescence period (r^2 of 0.15 and r^2 of 0.12 respectively). During the dry season tree canopy cover had a slightly higher explanatory power than $NDVI_{July}$ (r^2 of 0.14 and r^2 of 0.10 respectively) (**Table 12**). The interaction between NDVI and woody canopy cover had the lowest explanatory power in the factorial model across all phenological periods. **Table 10** presents variance partitioning prior the removal of outliers.

Table 9 Results of factorial regression model (involving NDVI and woody canopy cover (WCC)). All the statistics were drawn from 30 bootstrapped iterations prior the removal of outliers.

Response variable	Factorial models	Average r^2	CI	*P-value*	Average RMSE	CI
Shannon	$X_0 + X_1WCC + X_2NDVI_{March} + X_3WCC*NDVI_{March}$	0.28	±0.018	0.004	0.508	±0.024
	$X_0 + X_1WCC + X_2NDVI_{April} + X_3WCC*NDVI_{April}$	0.24	±0.019	0.021	0.519	±0.025
	$X_0 + X_1WCC + X_2NDVI_{May} + X_3WCC*NDVI_{May}$	0.20	±0.017	0.032	0.533	±0.023
	$X_0 + X_1WCC + X_2NDVI_{July} + X_3WCC*NDVI_{July}$	0.19	±0.019	0.044	0.538	±0.026

CI- confidence interval

Table 10 Regression coefficients and contribution (r^2) of each component of factorial regression (involving NDVI and woody canopy cover (WCC)) after variance partitioning. All the statistics were drawn from 30 bootstrapped iterations prior the removal of outliers.

Phenological period	Effects	Coefficients	CI	Average r^2	CI	*P-value*
March	Intercept	0.409245	±0.191			
	WCC	0.001164	±0.004	0.03	±0.009	0.7156
	NDVI	1.502159	±0.340	0.18	±0.035	0.4463
	WCC*NDVI	0.008538	±0.006	0.07	±0.023	0.6405
April	Intercept	0.493235	±0.175			
	WCC	0.006618	±0.003	0.03	±0.012	0.7345
	NDVI	1.321504	±0.348	0.15	±0.026	0.4590
	WCC*NDVI	0.003845	±0.007	0.06	±0.022	0.6938
May	Intercept	0.515435	±0.182			
	WCC	0.011234	±0.004	0.05	±0.015	0.6472
	NDVI	1.206113	±0.389	0.11	±0.017	0.5171
	WCC*NDVI	-0.002435	±0.007	0.04	±0.021	0.7186
July	Intercept	0.023622	±0.162			
	WCC	0.026949	±0.003	0.09	±0.012	0.2011
	NDVI	2.762607	±0.438	0.07	±0.011	0.2565
	WCC*NDVI	-0.040507	±0.009	0.03	±0.006	0.4385

CI- confidence interval

Table 11 Results of factorial regression model (involving NDVI and woody canopy cover (WCC)). All the statistics were drawn from 30 bootstrapped iterations post the removal of outliers.

Response variable	Factorial models	Average r^2	CI	*P-value*	Average RMSE	CI
Shannon	$X_0 + X_1WCC + X_2NDVI_{March} + X_3WCC*NDVI_{March}$	0.38	±0.018	0.000	0.441	±0.014
	$X_0 + X_1WCC + X_2NDVI_{April} + X_3WCC*NDVI_{April}$	0.34	±0.019	0.001	0.449	±0.014
	$X_0 + X_1WCC + X_2NDVI_{May} + X_3WCC*NDVI_{May}$	0.30	±0.019	0.004	0.466	±0.017
	$X_0 + X_1WCC + X_2NDVI_{July} + X_3WCC*NDVI_{July}$	0.29	±0.020	0.007	0.476	±0.018

CI- confidence interval

Table 12 Regression coefficients and contribution (r^2) of each component of factorial regression (involving NDVI and woody canopy cover (WCC)) after variance partitioning. All the statistics were drawn from 30 bootstrapped iterations post the removal of outliers.

Phenological periods	**Effects**	**Coefficients**	**CI**	**Average r^2**	**CI**	***P-value***
March	Intercept	-0.431219	±0.116			
	WCC	0.020927	±0.002	0.08	±0.009	0.3787
	NDVI	2.836309	±0.222	0.27	±0.017	0.1269
	WCC*NDVI	-0.021416	±0.004	0.03	±0.006	0.5767
April	Intercept	-0.133380	±0.076			
	WCC	0.020699	±0.001	0.10	±0.010	0.3401
	NDVI	2.392799	±0.146	0.22	±0.016	0.1660
	WCC*NDVI	-0.018818	±0.003	0.02	±0.005	0.6314
May	Intercept	-0.087983	±0.093			
	WCC	0.024542	±0.002	0.12	±0.013	0.2632
	NDVI	2.259555	±0.215	0.15	±0.013	0.2272
	WCC*NDVI	-0.024217	±0.004	0.03	±0.004	0.5422
July	Intercept	-0.405418	±0.070			
	WCC	0.035941	±0.001	0.14	±0.011	0.0671
	NDVI	3.674045	±0.191	0.10	±0.011	0.1155
	WCC*NDVI	-0.057939	±0.004	0.05	±0.003	0.2390

CI- confidence interval

Table 13 Results from ANOVA where we compared prediction errors (RMSE) between NDVI model and factorial model. The average RMSE (aRMSE) was obtained from 30 bootstrapped iterations post the removal of outliers. Each time the factorial model runs it combined tree canopy cover and NDVI from March, April, May or July respectively.

aRMSE	**aRMSE**		**aRMSE**	**aRMSE**		**aRMSE**	**aRMSE**		**aRMSE**	**aRMSE**	
$NDVI_{March}$ model	Factorial model	*P-value*	$NDVI_{April}$ model	Factorial model	*P-value*	$NDVI_{May}$ model	Factorial model	*P-value*	$NDVI_{July}$ model	Factorial model	*P-value*
0.454	0.441	0.1021	0.464	0.449	0.1136	0.492	0.466	0.0121	0.512	0.476	0.0011

Table 14 Results from ANOVA where we compared prediction errors (RMSE) between woody canopy cover (WCC) model and factorial model. The average RMSE (aRMSE) was obtained from 30 bootstrapped iterations post the removal of outliers. Each time the factorial model combined tree canopy cover and NDVI from March, April, May or July respectively.

aRMSE	**aRMSE**		**aRMSE**	**aRMSE**		**aRMSE**	**aRMSE**		**aRMSE**	**aRMSE**	
WCC model	Factorial model	*P-value*	WCC model	Factorial model	*P-value*	WCC model	Factorial model	*P-value*	WCC model	Factorial model	*P-value*
0.487	0.441	0.0000	0.487	0.449	0.0000	0.487	0.466	0.0345	0.487	0.476	0.2986

3.4.3. Tree species diversity maps

The tree species diversity maps show diversity pattern that is consistent with our knowledge of the area. Granite substrate hosts high tree species diversity while gabbro substrate has low tree species diversity (du Toit et al. 2003; Cho et al. 2012) and our models predicted a similar pattern of tree diversity (**Figure 13**). This diversity pattern is clearly discernible in figure 5a, b and c with contrasting diversity patterns between granite and gabbro substrate. However, both WCC and factorial models over-predicted tree species diversity in the northern part of KNP which is known to possess low species diversity and support the dominance of *Colophospermum mopane* (Makhado et al. 2013). Over-predictions associated with factorial models were exacerbated by changes in phenology towards senescence and dry season (**Figure 13d-e)**.

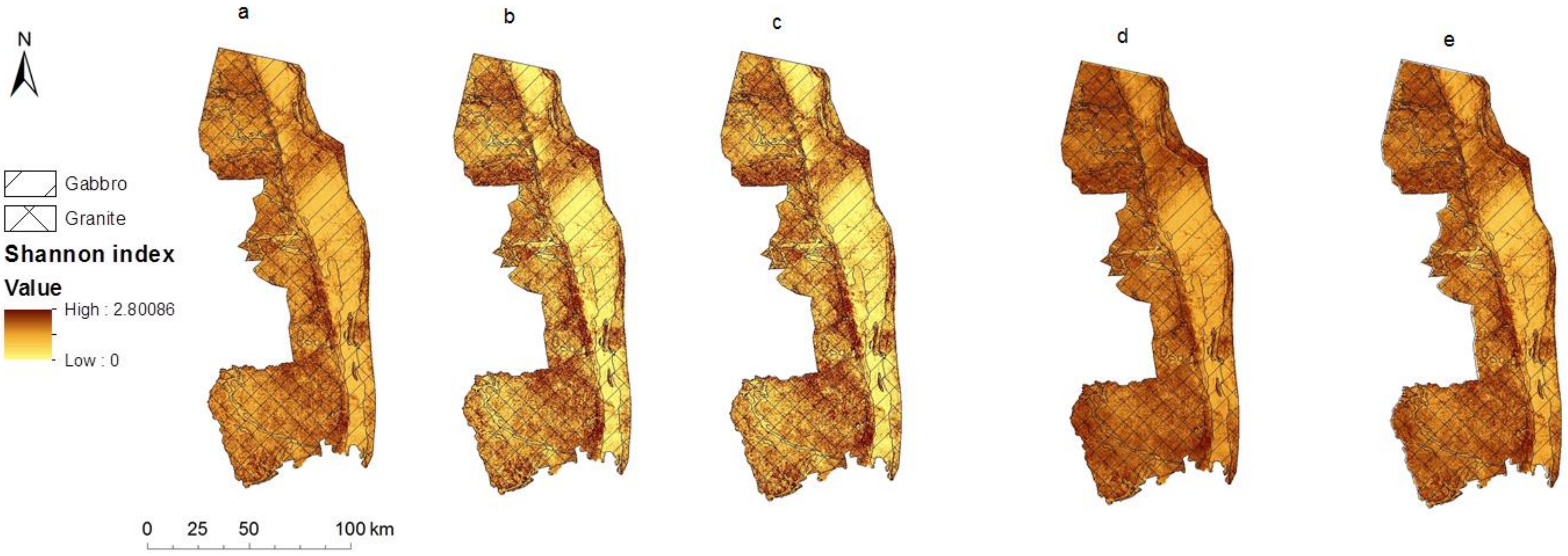

Figure 13 Tree species diversity from the best (a) woody canopy cover (WCC) model, (b) factorial model involving $NDVI_{March}$ and WCC (c) factorial model involving $NDVI_{April}$ and WCC (d) factorial model involving $NDVI_{May}$ and WCC and (e) factorial model involving $NDVI_{July}$ and WCC

3.5. Discussion

The results of our study indicate that Landsat derived NDVI, particularly at the end of the growing season (March), has a higher relationship with tree species diversity when compared to woody canopy cover. During this period the NDVI signal is influenced by woody canopy foliage and also by green herbaceous vegetation which would maintain high green biomass (Grant and Scholes, 2006; Ramoelo et al. 2015). In essence the NDVI signal captures total vegetation productivity within savannah woodland and therefore has a higher explanatory power than woody canopy cover. These results support the argument presented by Parviainen et al. (2010) and Witman et al. (2008) that the amount of energy available in an ecosystem detectable with NDVI as total primary productivity defines spatial variation in plant diversity. Furthermore, the transition to senescence (April) also seen the NDVI maintaining a higher relationship with tree species diversity when compared to woody canopy cover and again this was attributed to grass biomass impacting on the overall NDVI signal. Grass maintains its green biomass post the end of growing season in March (Grant and Scholes, 2006). However, during the transition to senescence the NDVI model had lower predictive power compared to March date (r^2 = 0.27 and r^2 = 0.33 respectively) and this indicates the declining influence of grass and some early senescing tree species e.g. *Acacia nigrescens* (Madonsela et al. 2017a) on the NDVI signal.

It was during the senescence period (May) that the NDVI model predicted tree species diversity in approximately the same way as woody canopy cover. This was not surprising given that the senescence period is characterized by senescent grass (Scholes et al. 2003; Cho et al. 2010) and therefore the NDVI signal was largely influenced by woody canopy foliage and the background. It is possible that the performance of NDVI might have been affected by the background conditions. NDVI does not consider canopy background conditions hence it is affected by soil brightness which lowers its sensitivity to vegetation (Huete and Jackson, 1988). However, the fact that NDVI performed similarly to woody canopy cover and that grass was already senescent (Scholes et al. 2003; Cho et al. 2010) indicates that the influence of woody canopy foliage was more dominant on the NDVI signal. Meanwhile, the dry season (July date) NDVI model had lower predictive power (r^2 of 0.13) compared tree canopy cover (r^2 of 0.21) and this was expected given that during dry season deciduous trees drops their canopy foliage (Tomlinson et al. 2013). These results indicate NDVI sensitivity to seasonal variations in the savannah woodland. As the season and vegetation phenology changes, the ability of the NDVI model to estimate tree species diversity declines effectively presenting NDVI as a phenology constrained predictor of tree species diversity.

Moreover, the fact that NDVI is sensitive to photosynthetically active vegetation which is directly related to rainfall as also observed by Scanlon et al. (2002), Parviainen et al. (2010) and Pau et al. (2012) makes it susceptible to inter-annual rainfall instability. Research (Archibald and Scholes, 2007) has observed that in African savannah large variation in NDVI signal emanates from grass vegetation due to inter-annual variation in grass phenology as a result of rainfall variability. Therefore, the use of NDVI for regularly estimating tree species diversity would always be confronted by this variation in NDVI signal which may lead to under- or over-estimation of tree species diversity particularly at the end of growing season.

Meanwhile, the results from factorial regression showed that the interaction between NDVI and WCC is not significant in explaining tree species diversity. The interaction between NDVI and WCC had the lowest explanatory power in explaining tree species diversity (r^2 of 0.02 – 0.04). Nonetheless, results from factorial model have shown that combining NDVI_{May} with WCC significantly improves the estimation of tree species diversity. Although the combination of either $\text{NDVI}_{\text{March}}$ or $\text{NDVI}_{\text{April}}$ with woody canopy cover in a factorial model had also improved the estimation of tree diversity, it was not significantly better than NDVI model. NDVI_{May} is largely influenced by tree canopy foliage and is not susceptible to grass induced variability since the senescence period (May) is characterized by senescent grass (Scholes et al. 2003; Cho et al. 2010). WCC, on the other hand, carries information related to vegetation structure and variance partitioning showed that WCC and NDVI_{May} were equally essential in explaining tree species diversity. However, in March or April date variance partitioning showed that the NDVI alone was sufficient to explain tree species diversity. These observations justify our assertion that combining NDVI_{May} with WCC should be considered as an alternative for biodiversity estimation during the senescence phenological period. Furthermore, these observations present an opportunity to counter the aforementioned limitations likely to be confronted when using Landsat derived NDVI for biodiversity estimation at the end of growing season. Moreover, the end of growing season is also accompanied by persistent cloud cover making it difficult to obtained cloud-free image in southern African savannah (Kaszta et al. 2016). The low temporal resolution of Landsat sensor exacerbate the issue thus making it necessary to explore alternative phenological period to obtain satellite imagery for studying tree species diversity.

However, woody canopy cover showed a significant positive relationship with tree species diversity (Figure 3a) and this was consistent with observation made Peterson and Reich (2008) in North American savannah. Contrary to NDVI where variation may actually be induced by grass phenology, changes in WCC are indicative of disturbance regimes or absence thereof. For instance, Asner et al. (2009) observed that the exclusion of herbivory in African savannah is associated with an increase in woody canopy cover and diversity of woody vegetation structure. It is not clear if the increase in woody cover impacts negatively or positively on tree species diversity. However, in the absence of herbivory effects or fires in the exclosures where Asner et al. (2009) made these observations, the diversity of woody vegetation structure should be assumed to indicate the presence of high tree species richness with diverse structural arrangements. Concurrent with this assertion Peterson and Reich (2008) observed that the absence of fire disturbances lead to high WCC in North American savannah accompanied by high tree species richness.

The nature of the relationship between woody vegetation cover and tree species diversity has never been established in southern African savannah. In South American savannah Pellegrini et al. (2016) observed that the removal of disturbances in Brazilian Cerrado led to an increase in total woody cover at the expense of endemic tree species adapted to open savannah. Contrary to observation by Pellegrini et al. (2016) our study observed a positive relationship between WCC and tree species diversity suggesting the possibility that the two variables may be linearly related in the southern African savannah. The linear relationship between WCC and tree species diversity can be explained partly by the rainfall gradient which has been observed to have a positive effect on both WCC and tree species diversity in the savannah (Sankaran et al. 2005; Shackleton, 2000). In this study WCC ranged between 6 – 78% and the increase in WCC was often accompanied by high tree species diversity. The question is how much change in woody canopy cover is within the resilience limits of savannah tree species diversity. Knowledge of the lower and upper thresholds in WCC within which tree species diversity thrive should facilitate the use of woody canopy cover for estimating tree species diversity. This assertion is made based on i) the observation made in the present study that woody canopy cover has a positive relationship with tree species diversity and ii) the observation made by Pellegrini et al. (2016) and Peterson and Reich (2008) that changes in savannah woody canopy cover impact on the tree species diversity.

Moreover, our study noted that woody canopy cover embodies the interplay of multiple environmental gradients influencing tree species diversity in the savannah woodland. For instance, we observed that the northern part of KNP, characterized by low mean annual precipitation (<450mm) and high temperatures (Makhado et al. 2013), has woody canopy cover that is below the 40% average woody vegetation cover. Parallel to this observation, the northern part of KNP has low tree species diversity supporting mainly the dominance of *Colophospermum mopane*. Meanwhile the southern portion of KNP has a mean annual precipitation of 750mm (Makhado et al. 2013; Eckhardt et al. 2000) and the granite substrate allows for woody vegetation with its deep-rooted system to have competitive-edge over grass in terms of access to soil moisture (Colgan et al., 2012). Associated with these environmental conditions were moderate to high woody canopy cover ranges (40% to 70%) and relatively high tree species diversity. However, low woody cover was also observed in the southern part of KNP which could be associated with elephant damage known to impact on woody vegetation in the savannah (Cumming et al. 1997; Druce et al. 2008).

The overall impression from these observations is that woody canopy cover may be useful for screening potential diversity hotspots in the southern African savannah. Noteworthy though, WCC model tends to over-predict tree species diversity in mono-species stand such as the northern part of KNP which is dominantly occupied with *Colophospermum mopane*. Over-prediction of tree species diversity in mono-species stand was also observed with NDVI models. However, WCC and NDVI in factorial models tended to perform better when these mono-species stands were removed as outliers. This suggests that remote sensing models based on vegetation productivity and canopy cover will be more suitable for species diverse savannahs. The tree species diversity maps showed diversity patterns that are consistent with our knowledge in southern part of KNP. Granite substrate hosts high tree species diversity while gabbro substrate has low tree species diversity (du Toit et al. 2003; Cho et al. 2012) and our models predicted a similar pattern of tree species diversity (**Figure 13b and c**).

In addition, further research on the utility of WCC and Landsat-8 derived NDVI for estimating tree species diversity in the savannah woodlands should integrate environmental variables which are known to impact tree species distribution. In this study, the highest average variance in tree species diversity explained by our models (WCC, $NDVI_{March}$ and the combination of the two in a factorial model) ranged between r^2 of 0.21 – 0.38. This can be improved with the integration of environmental variables known to impact species diversity e.g. rainfall and geology (Shackleton, 2000 and du Toit et al. 2003). It has been shown in previous studies that integrating environmental variables together with remote sensing variables improves the estimation of plant species (Malahlela et al. 2015).

However, the results of our study are consistent with the observation of Peterson and Reich (2008) in North American savannah where an increase in woody canopy cover was accompanied by high species richness. However, our results contrast with those of Pellegrini et al 2016 in South American savannah. We attributed these contrasting observations between our study and that of Pellegrini et al. (2016) to different climatic conditions prevailing in these savannahs. South American savannah is characterized by mesic conditions with mean annual precipitation of 2500mm which is 750mm above Africa's wettest savannahs (Lehmann et al. 2011; Pellegrini et al. 2016). Such mesic conditions combined with the absence of disturbance regime in South American savannah might be responsible for the transition to closed forest with increasing woody cover displacing savannah endemics (Pellegrini et al. 2016). Meanwhile, our observation and that of Peterson and Reich (2008) relates to semi-arid savannahs. In light of the above, our results suggest that woody canopy cover has a positive relationship to savannah tree species diversity in semi-arid savannahs. However, the above question of how much change in woody canopy cover is within the resilience limits of savannah tree species diversity has to be investigated in order to ascertain our observation.

3.6. Conclusion

In conclusion the study showed a significant positive relationship between WCC and tree species diversity in southern African savannah. The tree species diversity map produced from WCC model showed diversity patterns that are consistent with our knowledge of the area. The ability of WCC to explain tree species diversity highlights the explanatory power of vegetation structural variables. In this study vegetation structural variable was derived from

SAR and LiDAR data in a form of WCC. This also opens up opportunities to further investigate the possibility of using data from SAR and LiDAR sources in biodiversity research. Moreover, the study also showed that the interaction between NDVI and WCC is not significant, however, the factorial model improved the estimation of tree species diversity and significantly reduced the error of prediction (p <0.05) when compared to NDVI or WCC model during the senescence period. Furthermore, the study showed that i) in spite of challenges the NDVI is useful for explaining tree species diversity at the end of growing season and ii) combining NDVI_{May} and WCC in a factorial model improves the estimation of tree species diversity and may counter the challenges associated with the end of growing season.

Acknowledgements

The authors would like to express gratitude to the National Research Foundation for funding this research project through the NRF-Professional Development Programme. We would also like to thank Cecilia Masemola, Sbu Gumede, Joseph Dlamini, Martin Sarela and Patrick Ndlovu for their assistance during field work. We further thank the anonymous reviewers for their comments which improved the content of this paper.

References

Archibald, S. and Scholes, R.J., 2007. Leaf green-up in a semi-arid African savanna–separating tree and grass responses to environmental cues. *Journal of Vegetation Science*, *18*(4), pp.583-594.

Asner, G.P., Levick, S.R., Kennedy-Bowdoin, T., Knapp, D.E., Emerson, R., Jacobson, J., Colgan, M.S. and Martin, R.E., 2009. Large-scale impacts of herbivores on the structural diversity of African savannas. *Proceedings of the National Academy of Sciences*, *106*(12), pp.4947-4952.

Bond, W.J., Midgley, G.F. and Woodward, F.I., 2003. The importance of low atmospheric CO2 and fire in promoting the spread of grasslands and savannas. *Global Change Biology*, *9*(7), pp.973-982.

Box, E.O., Holben, B.N. and Kalb, V., 1989. Accuracy of the AVHRR vegetation index as a predictor of biomass, primary productivity and net CO2 flux. *Plant Ecology*, *80*(2), pp.71-89.

Cho, M.A., Debba, P., Mathieu, R., Naidoo, L., Van Aardt, J. and Asner, G.P., 2010. Improving discrimination of savanna tree species through a multiple-endmember spectral angle mapper approach: Canopy-level analysis. *IEEE Transactions on Geoscience and Remote Sensing*, *48*(11), pp.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.K.J. and Erasmus, B., 2012. Mapping tree species composition in South Africa savannas using an integrated airborne spectral and Lidar system. *Remote Sensing of Environment Vol.* 125, pp. 214-226.

Colgan, M.S., Asner, G.P., Levick, S.R., Martin, R.E. and Chadwick, O.A., 2012. Topo-edaphic controls over woody plant biomass in South African savannas. *Biogeosciences*, *9*, pp.1809-1821.

Colwell, R.K., 2009. Biodiversity: concepts, patterns, and measurement. *The Princeton guide to ecology*, pp.257-263.

Cumming, D.H., Fenton, M.B., Rautenbach, I.L., Taylor, R.D., Cumming, G.S., Cumming, M.S., Dunlop, J.M., Ford, A.G., Hovorka, M.D., Johnston, D.S. and Kalcounis, M., 1997. Elephants, woodlands and biodiversity in southern Africa. *South African Journal of Science*, *93*(5), pp.231-236.

Dahbi, S., El Moussami, H. and Ezzine, L., 2015, October. Multiple regression model for surface roughness using full factorial design. In *Industrial Engineering and Systems Management (IESM), 2015 International Conference on* (pp. 439-444). IEEE.

Druce, D.J., Shannon, G., Page, B.R., Grant, R. and Slotow, R., 2008. Ecological thresholds in the savanna landscape: developing a protocol for monitoring the change in composition and utilisation of large trees. *PloS one*, *3*(12), p.e3979.

du Toit, J.T., Biggs, H.C., Rogers, K.H., 2003. The Kruger Experience: ecology and management of savanna heterogeneity. London: Island Press.

Dumalisile, L., 2009. *Effects of Chromolaena odorata on mammalian biodiversity in Hluhluwe-iMfolozi Park, South Africa* (Doctoral dissertation).

Eckhardt, H.C., Wilgen, B.W. and Biggs, H.C., 2000. Trends in woody vegetation cover in the Kruger National Park, South Africa, between 1940 and 1998. *African Journal of Ecology*, *38*(2), pp.108-115.

Fu, W.J., Jiang, P.K., Zhou, G.M. and Zhao, K.L., 2014. Using Moran's I and GIS to study the spatial pattern of forest litter carbon density in a subtropical region of southeastern China. *Biogeosciences*, *11*(8), pp.2401-2409.

Gillson, L. and Duffin, K.I., 2007. Thresholds of potential concern as benchmarks in the management of African savannahs. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *362*(1478), pp.309-319.

Gonsamo, A., D’odorico, P. and Pellikka, P., 2013. Measuring fractional forest canopy element cover and openness–definitions and methodologies revisited. *Oikos*, *122*(9), pp.1283-1291.

Gould, W., 2000. Remote sensing of vegetation, plant species richness, and regional biodiversity hotspots. *Ecological applications*, *10*(6), pp.1861-1870.

Gottipati, R. and Mishra, S., 2010. Process optimization of adsorption of Cr (VI) on activated carbons prepared from plant precursors by a two-level full factorial design. *Chemical Engineering Journal*, *160*(1), pp.99-107.

Grant, C.C. and Scholes, M.C., 2006. The importance of nutrient hot-spots in the conservation and management of large wild mammalian herbivores in semi-arid savannas. *Biological Conservation*, *130*(3), pp.426-437.

Gringarten, E. and Deutsch, C.V., 2001. Teacher's aide variogram interpretation and modeling. *Mathematical Geology*, *33*(4), pp.507-534.

Hempson, G.P., Archibald, S. and Bond, W.J., 2015. A continent-wide assessment of the form and intensity of large mammal herbivory in Africa. *Science*, *350*(6264), pp.1056-1061.

Hernández-Stefanoni, J.L., Gallardo-Cruz, J.A., Meave, J.A., Rocchini, D., Bello-Pineda, J. and López-Martínez, J.O., 2012. Modeling α-and β-diversity in a tropical forest from remotely sensed and spatial data. *International journal of applied earth observation and geoinformation*, *19*, pp.359-368.

He, K.S., Zhang, J. and Zhang, Q., 2009. Linking variability in species composition and MODIS NDVI based on beta diversity measurements. *acta oecologica*, *35*(1), pp.14-21.

Huete, A.R. and Jackson, R.D., 1988. Soil and atmosphere influences on the spectra of partial canopies. *Remote Sensing of Environment*, *25*(1), pp.89-105.

Jongman, R.H.G., Ter Braak, C.J. and Van Tongeren, O.F. eds., 1995. *Data analysis in community and landscape ecology*. Cambridge university press.

Kaszta, Ż., Van De Kerchove, R., Ramoelo, A., Cho, M.A., Madonsela, S., Mathieu, R. and Wolff, E., 2016. Seasonal Separation of African Savanna Components Using Worldview-2 Imagery: A Comparison of Pixel-and Object-Based Approaches and Selected Classification Algorithms. *Remote Sensing*, *8*(9), p.763.

Kerr, J.T. and Ostrovsky, M., 2003. From space to species: ecological applications for remote sensing. *Trends in Ecology & Evolution*, *18*(6), pp.299-305.

Jetz, W., Cavender-Bares, J., Pavlick, R., Schimel, D., Davis, F.W., Asner, G.P., Guralnick, R., Kattge, J., Latimer, A.M., Moorcroft, P. and Schaepman, M.E., 2016. Monitoring plant functional diversity from space. *Nature plants*, *2*(3).

Lehmann, C.E., Archibald, S.A., Hoffmann, W.A. and Bond, W.J., 2011. Deciphering the distribution of the savanna biome. *New Phytologist*, *191*(1), pp.197-209.

Madonsela, S., Cho, M.A., Mathieu, R., Mutanga, O., Ramoelo, A., Kaszta, Ż., Van De Kerchove, R. and Wolff, E., 2017a. Multi-phenology WorldView-2 imagery improves remote sensing of savannah tree species. *International Journal of Applied Earth Observation and Geoinformation*, *58*, pp.65-73.

Madonsela, S., Cho, M.A., Ramoelo, A. and Mutanga, O., 2017b. Remote sensing of species diversity using Landsat 8 spectral variables. *ISPRS Journal of Photogrammetry and Remote Sensing*, *133*, pp.116-127.

Makhado, R.A., Mapaure, I., Potgieter, M.J., Luus-Powell, W.J. and Saidi, A.T., 2014. Factors influencing the adaptation and distribution of Colophospermum mopane in southern Africa's mopane savannas-A review. *Bothalia-African Biodiversity & Conservation*, *44*(1), pp.1-9.

Malahlela, O.E., Cho, M.A. and Mutanga, O., 2015. Mapping the occurrence of Chromolaena odorata (L.) in subtropical forest gaps using environmental and remote sensing data. *Biological Invasions*, *17*(7), pp.2027-2042.

Mathieu, R., Naidoo, L., Cho, M.A., Leblon, B., Main, R., Wessels, K., Asner, G.P., Buckley, J., Van Aardt, J., Erasmus, B.F. and Smit, I.P., 2013. Toward structural assessment of semi-arid African savannahs and woodlands: The potential of multitemporal polarimetric RADARSAT-2 fine beam images. *Remote Sensing of Environment*, *138*, pp.215-231.

Matsika, R., Erasmus, B.F. and Twine, W.C., 2013. A tale of two villages: assessing the dynamics of fuelwood supply in communal landscapes in South Africa. *Environmental Conservation*, *40*(1), pp.71-83.

Morris, E.K., Caruso, T., Buscot, F., Fischer, M., Hancock, C., Maier, T.S., Meiners, T., Müller, C., Obermaier, E., Prati, D. and Socher, S.A., 2014. Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories. *Ecology and evolution*, *4*(18), pp.3514-3524.

Nagendra, H., Rocchini, D., Ghate, R., Sharma, B. and Pareeth, S., 2010. Assessing plant diversity in a dry tropical forest: Comparing the utility of Landsat and IKONOS satellite images. *Remote Sensing*, *2*(2), pp.478-496.

Naidoo, L., Mathieu, R., Main, R., Kleynhans, W., Wessels, K., Asner, G. and Leblon, B., 2015. Savannah woody structure modelling and mapping using multi-frequency (X-, C-and L-band) Synthetic Aperture Radar data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *105*, pp.234-250.

National Terrestrial Carbon Sink Assessment (2015) Department of Environmental Affairs, Pretoria, South Africa.

Oindo, B.O. and Skidmore, A.K., 2002. Interannual variability of NDVI and species richness in Kenya. *International journal of remote sensing*, *23*(2), pp.285-298.

Oldeland, J., Wesuls, D., Rocchini, D., Schmidt, M. and Jürgens, N., 2010. Does using species abundance data improve estimates of species diversity from remotely sensed spectral heterogeneity?. *Ecological Indicators*, *10*(2), pp.390-396.

Parviainen, M., Luoto, M. and Heikkinen, R.K., 2010. NDVI-based productivity and heterogeneity as indicators of plant-species richness in boreal landscapes. *Boreal environment research*, *15*(3).

Pau, S., Gillespie, T.W. and Wolkovich, E.M., 2012. Dissecting NDVI–species richness relationships in Hawaiian dry forests. *Journal of Biogeography*, *39*(9), pp.1678-1686.

Pellegrini, A.F., Socolar, J.B., Elsen, P.R. and Giam, X., 2016. Trade-offs between savanna woody plant diversity and carbon storage in the Brazilian Cerrado. *Global change biology*, *22*(10), pp.3373-3382.

Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C. and Coops, N.C., 2013. Essential biodiversity variables. *Science*, *339*(6117), pp.277-278.

Peterson, D.W. and Reich, P.B., 2008. Fire frequency and tree canopy structure influence plant species diversity in a forest-grassland ecotone. *Plant Ecology*, *194*(1), pp.5-16.

Pervez, W., Uddin, V., Khan, S.A. and Khan, J.A., 2016. Satellite-based land use mapping: comparative analysis of Landsat-8, Advanced Land Imager, and big data Hyperion imagery. *Journal of Applied Remote Sensing*, *10*(2), pp.026004-026004.

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

Richter, R. and Schläpfer, D., 2012. Atmospheric/Topographic Correction for Satellite Imagery (ATCOR-2/3 User Guide, Version 8.2 BETA). *German Aerospace Center, Remote Sensing Data Center: Wessling, Germany*.

Rocchini, D., 2007. Effects of spatial and spectral resolution in estimating ecosystem α-diversity by satellite imagery. *Remote Sensing of Environment*, *111*(4), pp.423-434.

Sankaran, M., Hanan, N.P., Scholes, R.J. and Ratnam, J., 2005. Determinants of woody cover in African savannas. *Nature*, *438*(7069), p.846.

Scanlon, T.M., Albertson, J.D., Caylor, K.K. and Williams, C.A., 2002. Determining land surface fractional cover from NDVI and rainfall time series for a savanna ecosystem. *Remote Sensing of Environment*, *82*(2), pp.376-388.

Scholes, R.J. and Archer, S.R., 1997. Tree-grass interactions in savannas. *Annual review of Ecology and Systematics*, *28*(1), pp.517-544.

Scholes, R.J., Bond, W.J. and Eckhardt, H.C., 2003. *Vegetation dynamics in the Kruger ecosystem* (pp. 242-262). The Kruger Experience. Island Press.

Seto, K.C., Fleishman, E., Fay, J.P. and Betrus, C.J., 2004. Linking spatial patterns of bird and butterfly species richness with Landsat TM derived NDVI. *International Journal of Remote Sensing*, *25*(20), pp.4309-4324.

Seymour, C.L. and Dean, W.R.J., 2010. The influence of changes in habitat structure on the species composition of bird assemblages in the southern Kalahari. *Austral Ecology*, *35*(5), pp.581-592.

Shackleton, C.M., 2000. Comparison of plant diversity in protected and communal lands in the Bushbuckridge lowveld savanna, South Africa. *Biological Conservation*, *94*(3), pp.273-285.

Shackleton, C.M., Shackleton, S.E., Buiten, E. and Bird, N., 2007. The importance of dry woodlands and forests in rural livelihoods and poverty alleviation in South Africa. *Forest policy and economics*, *9*(5), pp.558-577.

Shannon, C.E. and Weaver, W., 1998. *The mathematical theory of communication*. University of Illinois press.

Thenkabail, P.S., Hall, J., Lin, T., Ashton, M.S., Harris, D. and Enclona, E.A., 2003. Detecting floristic structure and pattern across topographic and moisture gradients in a mixed species Central African forest using IKONOS and Landsat-7 ETM+ images. *International Journal of Applied Earth Observation and Geoinformation*, *4*(3), pp.255-270.

Tomlinson, K.W., Poorter, L., Sterck, F.J., Borghetti, F., Ward, D., Bie, S. and Langevelde, F., 2013. Leaf adaptations of evergreen and deciduous trees of semi-arid and humid savannas on three continents. *Journal of Ecology*, *101*(2), pp.430-440.

Treydte, A.C., Heitkönig, I.M.A., Prins, H.H.T. and Ludwig, F., 2007. Trees enhance grass layer quality in African savannas of distinct rainfall and soil fertility. *Perspectives in Plant Ecology, Evolution and Systematics*, *8*, pp.197-205.

Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E. and Steininger, M., 2003. Remote sensing for biodiversity science and conservation. *Trends in ecology & evolution*, *18*(6), pp.306-314.

Viña, A., Gitelson, A.A., Nguy-Robertson, A.L. and Peng, Y., 2011. Comparison of different vegetation indices for the remote assessment of green leaf area index of crops. *Remote Sensing of Environment*, *115*(12), pp.3468-3478.

Wang, J., Rich, P.M., Price, K.P. and Kettle, W.D., 2004. Relations between NDVI and tree productivity in the central Great Plains. *International Journal of Remote Sensing*, *25*(16), pp.3127-3138.

Watling, J.I., Brandt, L.A., Bucklin, D.N., Fujisaki, I., Mazzotti, F.J., Romañach, S.S. and Speroterra, C., 2015. Performance metrics and variance partitioning reveal sources of uncertainty in species distribution models. *Ecological Modelling*, *309*, pp.48-59.

Witman, J.D., Cusson, M., Archambault, P., Pershing, A.J. and Mieszkowska, N., 2008. The relation between productivity and species diversity in temperate–arctic marine ecosystems. *Ecology*, *89*(sp11).

Wood, E.M., Pidgeon, A.M., Radeloff, V.C. and Keuler, N.S., 2013. Image texture predicts avian density and species richness. *PloS one*, *8*(5), p.e63211.

Chapter 4: Investigating the possibility of using RTM derivatives for estimating tree species diversity in the savannah woodlands

This chapter is based on

Madonsela, S., Cho, M.A., Ramoelo, A., Mutanga, O. and Masemola, C., (*in preparation*). Estimating tree species diversity using biophysical and biochemical properties derived from the inversion of radiative transfer models

Abstract

Remote sensing application in biodiversity has often been based on correlation analysis between spectral vegetation indices particularly the NDVI and species richness. However, the NDVI does not cater for canopy background settings hence it is affected by soil brightness which lowers its sensitivity to vegetation. As such NDVI may be unsatisfactory in explaining tree species diversity. Meanwhile, radiative transfer modelling uses physical laws to reconstruct canopy reflectance and estimate biophysical and biochemical properties affecting the reflectance. This study will investigate the possibility of using radiative transfer model (RTM) derivatives (leaf area index (LAI), Chlorophyll a and b (Cab), equivalent water thickness (EWT) and dry matter) for estimating tree species diversity. To accomplish our objective we followed stratified random sampling and set up field plots (n = 68) of 90m X 90m across the study area. Trees within each plot with diameter at breast height of >10cm were sampled and Shannon index was used to quantify species diversity. Radiative transfer models, PROSAIL and INFORM, were inverted on seasonal Landsat-8 data using a look-up table and bagging algorithm in order to derive leaf area index (LAI), chlorophyll (*Cab*), carotenoid (*Car*), equivalent water thickness (EWT), brown pigment (*Cbrown*) and dry matter (*Cm*) corresponding to each plot. A stepwise linear regression was applied in a bootstrapping approach to determine the relationship between Shannon index and RTM derivatives (LAI, *Cab*, *Car*, EWT, *Cbrown* and *Cm*) with calibration (n = 45) and test (n = 23) datasets. The results of bootstrapped regression analysis showed that RTM derivatives particularly LAI, EWT and *Cbrown* had a significant positive relationship with Shannon index ($p < 0.05$) in the end of the growing season while *Cab* and *Car* had shown no relationship with Shannon index. Contrary to our expectation RTM derivatives did not explain species diversity better than the NDVI model. However, the transformation of these RTM derivatives via Principal Component Analysis into a few principal components improved their ability to explain species diversity better than NDVI particularly at the end of growing season (r^2 of 0.45 and 0.33 respectively; $p < 0.001$). Overall, radiative transfer modelling offers an opportunity to derive several biophysical and biochemical parameters of vegetation which can be useful for explaining species diversity.

4.1. Introduction

Tree species provide many goods and services in the savannah ecosystem such as providing habitats and nesting sites for diverse avifaunal species (Dean et al. 1999; Seymour and Dean, 2010); facilitating grass growth and improving grass quality beneath their canopies (Ludwig et al., 2004; Treydte et al. 2007); and serving as food resources to many browsing faunal species (Hempson et al. 2015). Tree species also acts as a safety net against poverty in the neighbouring communities (Shackleton, 2004; Shackleton et al. 2007). Therefore monitoring the abundance and diversity of savannah tree species remain essential to ensure that the communal use of these biodiversity resources e.g. harvesting for fuelwood (Madubansi and Shackleton, 2006) or their disturbances by elephants in protected areas (Druce et al. 2008) is kept within the resilience capacity of the ecosystem. Successful monitoring of savannah biodiversity is dependent on the availability of up-to-date and spatially detailed information on species richness and distribution at a regional scale (Turner et al. 2003). Space-borne remote sensing meets these demands as it collects data over large geographic areas on a regular interval and at varying levels of spatial details (Jetz et al. 2016; Kerr and Ostrovsky, 2003).

Consequently, a number of studies e.g. Gould, (2000); Oindo and Skidmore (2002) and Parviainen et al. (2010) have employed remotely sensed data to estimate species diversity in different biomes. In particular the normalized difference vegetation index (NDVI) derived from the red and near infrared bands has shown a positive relationship with tree species diversity. The success of NDVI emanates from its sensitivity to abiotic factors such as rainfall which impacts landscape biodiversity (Pau et al. 2012; Shackleton, 2000). Furthermore, the sum of energy available in an ecosystem measurable with NDVI as primary productivity defines the spatial variation of plant and animal diversity (Parviainen et al. 2010; Witman et al. 2008). It is therefore not surprising that the NDVI has successfully explained floral, faunal and avifaunal diversity in different biomes and at various scales (Oindo and Skidmore, 2002; Seto et al. 2004; Pau et al. 2012).

However, NDVI formulation does not cater for canopy background settings hence it is affected by soil brightness which lowers its sensitivity to vegetation (Huete and Jackson, 1988). Therefore, NDVI may be unsatisfactory in explaining tree species diversity. Recently, the spectral variation hypothesis (SVH) has emerged as the basis for explaining species diversity using remotely sensed data. SVH argues that spectral variability on the image emanates particularly from the spatial heterogeneity of the environment which by default has high species diversity due to the higher number of available niches (Rocchini et al. 2010; 2015). SVH therefore employs a multivariate analytical approach to establish a relationship between spectral variability on the image and tree species diversity on the landscape (Rocchini, 2007; Oldeland et al. 2010; Hernández-Stefanoni et al. 2012). However no study has investigated the possibility of using radiative transfer model (RTM) derivatives for estimating tree species diversity particularly in the African savannah.

Firstly, RTM represents physics based modelling which is used to describe canopy reflectance spectra as a function of leaf optical properties, canopy structural attributes, soil background characteristics and viewing geometry variables (Baret et al. 1992; Kötz et al. 2004; Zhang et al. 2005; Darvishzadeh et al. 2008). Hence the transfer of incident radiation within the forest canopy is affected by the distribution of canopy biophysical and biochemical properties relative to each other and the subsequent radiative processes such as multiple scattering (Kötz et al. 2004). Taking into account all these variables, RTM uses physical principles to describe the transfer and interaction of radiation inside the canopy and accurately simulate canopy reflectance (Jacquemoud et al. 1995; Kötz et al. 2004; Darvishzadeh et al. 2008). In essence RTM reveal the relationship that exist between vegetation biophysical and biochemical parameters, observation geometry and canopy reflectance (Darvishzadeh et al. 2008).

The subsequent step entails the retrieval of those parameters that minimize the differences between measured and simulated spectra through inversion of the RTM (Jacquemoud et al. 1995; Atzberger, 2000) and using these parameters to study other ecosystem processes. For instance Kötz et al. (2004) inverted two hybrid canopy reflectance models (GeoSAIL and FLIGHT) to retrieve EWT, LAI and *Cm* and these parameters were considered essential for understanding forest fire risk in the Swiss National Park. Several other studies used different radiative transfer models to retrieve LAI and chlorophyll and estimate the fraction of photosynthetically active radiation absorbed by vegetation (Myneni et al. 1997; Asner et al. 1998; Myneni et al. 2002; Zhang et al. 2005). In this study we investigate the possibility of using RTMs to retrieve biophysical and biochemical parameters (e.g. LAI, *Cab* and EWT) which have been shown to be useful for characterizing tree species diversity (Carlson et al.

2007; Zhao et al. 2016). Unlike NDVI which is limited to red and near infrared (NIR), RTM presents an opportunity to exploit the entire spectral content from the visible, NIR and shortwave infrared (SWIR) regions of the electromagnetic spectrum and accurately retrieve *Cab*, LAI and EWT.

Simultaneous application of spectral information in the visible, NIR and SWIR ensures accurate estimation of plant biophysical and biochemical properties. For instance Ceccato et al. (2001) observed that the reflectance in the SWIR is not only controlled by EWT but also dry matter and leaf internal structure. Therefore the accurate retrieval of EWT requires a combination of spectral information from the NIR and SWIR. Furthermore, Thenkabail et al. (2003) observed that the two SWIR bands present in Landsat program enhances the spectral capacity of Landsat data to explain forest characteristics beyond that of high resolution IKONOS data which is limited to the visible and NIR regions of the electromagnetic spectrum. The aim of this study was to investigate the possibility of using RTM derivatives for estimating tree species diversity in the savannah woodlands. Firstly we simulate canopy reflectance using PROSAIL and INFORM, retrieve EWT, *Cab*, LAI, *Cm*, *Cbrown*, and *Car* and then use these derivatives as predictor variables in the regression models to estimate species diversity.

4.2. Description of study area and field data collection

4.2.1. Study area

The study area stretches across the KwaZulu-Natal (KZN), Mpumalanga and Limpopo provinces of South Africa, covering the savannah woodland belt (**Figure 14**). The area is divided into two land management regimes i.e. communal areas and protected areas (Kruger National Park, Hluhluwe-Imfolozi Park and other private nature reserves) with differing land use practices. High tree species diversity has been noted in both areas (du Toit et al. 2003; Shackleton, 2000). The savannah woodland is characterized by varying edaphic properties as a result of differential geological substrates and a mountainous terrain, particularly in the KZN region. Topography, rainfall and geology are amongst the key environmental factors that dictate the pattern of tree species diversity (Makhado et al. 2014; Shackleton, 2000).

The northern part of the study area receives low to moderate rainfall and supports the predominance of *Colophospermum mopane* (Makhado et al. 2014). The central part of the study area is dominated by members of the *Combretaceae* (*Terminalia sericea, Combretum collinum, Combretum apiculatum, Combretum zeyheri*) and *Mimosaceae* families (*Acacia nigrescens, Acacia gerradii and Dichrostachys cinerea*), with distribution being controlled by granite and gabbro geological substrates. Other important taxa include *Sclerocarya birrea,* which is widely distributed throughout the region (Eckhardt et al. 2000; du Toit et al. 2003; Shackleton, 2000). The mean annual precipitation ranges from 440mm in the north to 750mm in the south with annual variations around the mean (Makhado et al. 2014; Eckhardt et al. 2000). The month of March marks the end of the growing season while April to November has been described as the dry season in the southern African savannah (Grant and Scholes, 2006; Archibald and Scholes, 2007). Typical of a savannah setting, the vegetation is characterized by a continuous herbaceous layer interspersed by a woody tree cover of varying density depending on the geological substrate. The woody vegetation is characterized by trees of varying heights and crown dimensions (Wessels et al. 2011).

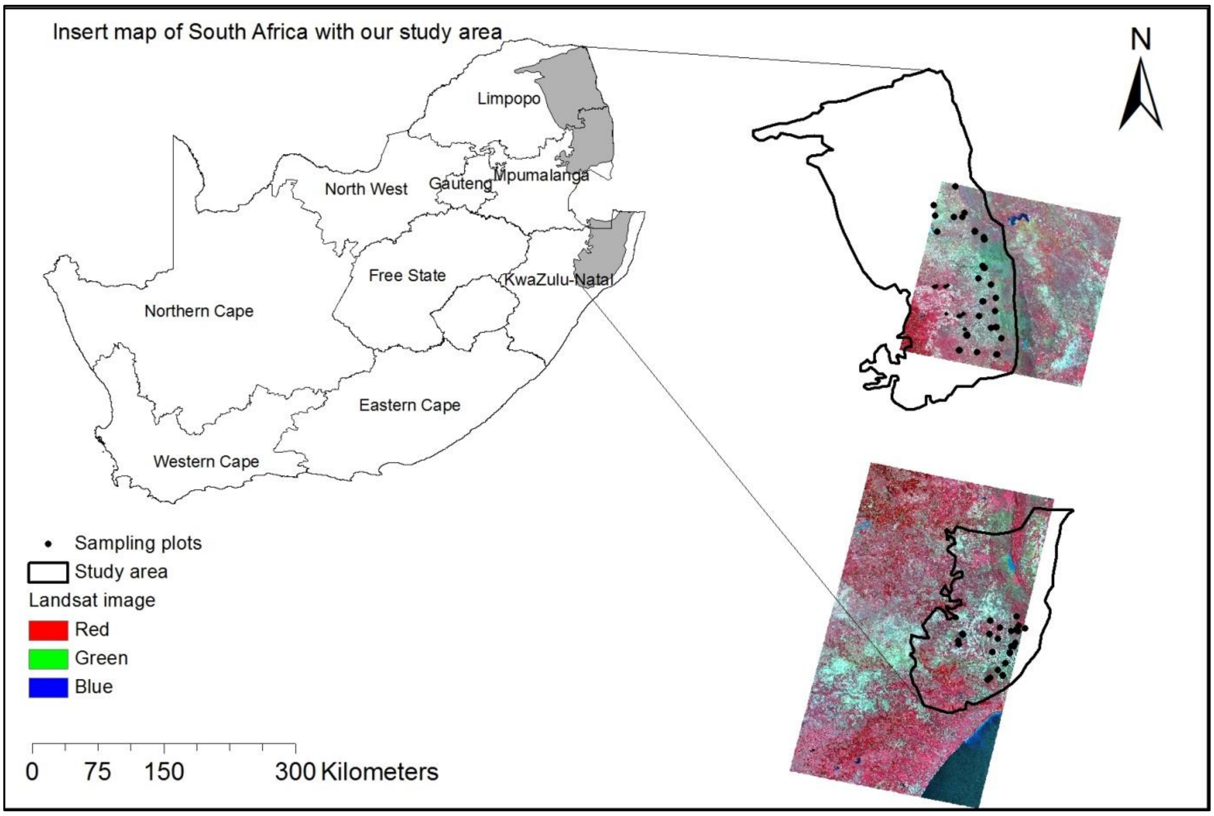

Figure 14 Study area stretching across three provinces of South Africa. Dots on the Landsat imagery are the sampling plots.

4.2.2. Field data collection

Field data was collected on the 2nd till the 27th of November 2015 in KwaZulu-Natal and again on the 1st till the 19th of March 2016 across Kruger National Park stretching between Mpumalanga and Limpopo provinces. The principal aim of the field campaign was to identify tree species within randomly placed sampling plots and quantify local species diversity (α-diversity) in the region using the common measure of diversity i.e. Shannon index. Prior to the field excursion we defined the size of field sampling plots using semi-variogram analysis in ENVI 4.8 software. Semi-variograms measure the spatial variability of natural phenomenon occurring in space (Fu et al. 2014; Gringarten and Deutsch, 2001). A semi-variogram is calculated as follow:-

Equation 5

$$y(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2$$

where $y(h)$ is the semi-variance at a given distance h; $z(x_i)$ is the value of the variable Z at location x_i, h is the lag distance and $N(h)$ is the number of pairs of sample points separated by h.

Semi-variance slowly increases as the distance from one position to the next increases up until it reaches the range where it starts to level off (Jongman et al. 1995; Gringarten and Deutsch, 2001). A semi-variogram plot is produced by calculating variance at different lag distances and a theoretical model such as spherical or exponential model is fitted to provide information about the spatial structure (Fu et al. 2014). Our study applied semi-variogram analysis to degraded WorldView-2 derived NDVI image to define the scale of spatial variability in tree species richness. The choice to use NDVI was based on observation that variability in NDVI corresponds to species diversity (Gould, 2000). It was important to use NDVI because it suppresses spectral content from non-vegetated pixels (Viña et al. 2011) and was therefore a viable option to determine pixel variability related to vegetation.

In our analysis, Worldview-2 image was firstly degraded to 10m spatial resolution to be compatible with average tree canopy size in the savannah (Cho et al. 2012) and then used to generate NDVI image. In ENVI software v4.8 the semi-variogram analysis computed the squared difference between neighbouring pixel values in order to quantify variability. The analysis conducted on the Worldview-2 derived NDVI image showed that the scale for tree species variability in the savannah woodland lies at a lag distances of 90m to 100m (**Figure 15**). Although semi-variance would seem to be increasing beyond the lag distance of 90m the increase was not consistent and the lag distance of 90m resulted in a plot size that were feasible to work on within limited resources. Moreover, the study intended to use Landsat data with a 30m pixel resolution hence the plot size of 90m X 90m was opted to ascertain correspondence between field data and spectral data.

The plot size of 90m X 90m was therefore chosen to capture spatial variation in tree species diversity. Stratified random sampling was used to define the placement of sampling plots. The stratification of sampling plots followed four dominant geological formations (granite; siliciclastic; gabbros; granulite) that were observed to have marked influence over vegetation patterns in the study area (du Toit et al. 2003). Plots of 90m x 90m were designed and all trees within the plots with a diameter at breast height (DBH) above 10cm were recorded with Global Positioning System and the species identified. The field campaign visited 50 plots distributed across the study area to collect tree species data. A further 26 plots, collected under similar conditions in the previous study (Naidoo et al. 2015), were added to our field data. However, some of these field plots (8 plots) were located on clouded parts of the March and April images and therefore were not usable. A total of 68 field plots were used in the analysis.

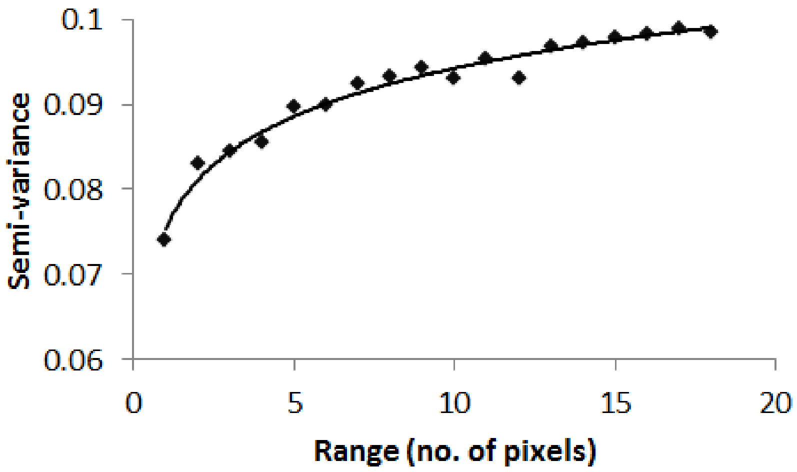

Figure 15 Semi-variogram analysis showing the scale of tree species variability in the savannah woodland

4.3. Materials and methods

4.3.1. Remote sensing data

Four Landsat-8 Operational Land Imager satellite images collected in 2016 (28th of March, 29th of April, 31st of May and 24th of July) were downloaded from United States Geological Surveys (USGS) download portal (https://earthexplorer.usgs.gov/). The dates of image collection coincided with different phenological periods. The end of March represents the end of growing season (Grant and Scholes, 2006; Cho et al. 2010); April captures the transition to senescence (Madonsela et al. 2017) and May captures the advanced senescence stage when most trees starts to drop off leaves. In the month of May grass is already dry (Cho et al. 2010) while July corresponds to the dry season (Kaszta et al. 2016).

Landsat-8 OLI is a multi-spectral sensor with eight spectral bands pioneered in the visible, NIR and SWIR regions of electromagnetic spectrum. However, the study only used the six spectral bands of Landsat-8 i.e. Blue, Green, Red, NIR, SWIR-1 and SWIR-2 which are known to be useful for vegetation studies. Coastal and Cirrus bands are useful for aerosol and water vapour estimation and were not used in the present study. The Landsat-8 OLI records data at a moderate spatial resolution of 30m and has a revisit capacity of 16 days. The 12-bit quantization of data has improved the signal-to-noise radiometric performance of the sensor, in comparison to previous generations, thus increasing its usefulness for landcover mapping (Pervez et al. 2016). Landsat-8 image was downloaded freely from United States Geological Surveys (USGS) download portal with geometric correction already being completed. The image was atmospherically corrected using ATCOR-2 software since the area covered by the image exhibit gently undulating slopes (Richter and Schläpfer, 2012). The KZN Landsat scene necessitated the use of ATCOR-3 software since the region is mountainous. ATCOR-3 allows for integration of a DEM which is useful for the correction of shadow effects on the image which are depicted in mountainous areas (Richter and Schläpfer, 2012). After the pre-processing of Landsat-8 images, nine Landsat pixels corresponding to field sampling plots were identified and the spectral reflectance of the six aforementioned Landsat-8 bands was extracted for use as the measured spectra in the RTM inversion.

4.3.2. Radiative Transfer Models

Two canopy radiative transfer models PROSAIL and INFORM were inverted on the Landsat-8 data. The PROSAIL model is made up of the leaf optical properties model (PROSPECT) (Jacquemoud and Beret, 1990) and the canopy level bidirectional reflectance model (SAIL) (Verhoef 1984, 1985) which incorporates the hotspot effect (SAILH) (Kuusk, 1985). Combining the two models into PROSAIL facilitates the estimation of both leaf and canopy parameters. The latest version of PROSPECT model simulates the leaf transmittance and reflectance as a function of six input variables i.e. the leaf structure, N (unitless); the leaf chlorophyll a+b concentration, *Cab* (μg cm^{-2}); the carotenoids, *Car* (μg cm^{-2}); the dry matter content, *Cm* (g cm^{-2}); the equivalent water thickness, *Cw* (g cm^{-2}) and the leaf brown pigment, *Cbrown* (g cm^{-2}) (Féret et al. 2017). The leaf optical properties calculated by PROSPECT were used as inputs into the SAIL canopy reflectance model. SAIL is a 1-dimensional model used to simulate the bidirectional reflectance factor of a turbid medium plant canopies. The transfer of radiation inside the canopy is calculated using the SAIL model which defines the canopy as a horizontal, homogenous and infinite layer of vegetation composed of Lambertian scatters (Verhoef 1984). In addition to leaf optical properties, the SAIL model needs eight input parameters to simulate the top of the canopy bidirectional reflectance i.e. sun zenith angle, *ts* (deg); sensor viewing angle, to (deg); azimuth angle, phi (deg); fraction of diffuse incoming solar radiation, skyl; background reflectance for each wavelength, rsl; LAI (m^{2} m^{-2}); mean leaf inclination angle, ALA (deg) and hot spot effect, hot (mn^{-1}).

INFORM is a three-dimensional canopy reflectance model (Atzberger, 2000; Schlerf and Atzberger, 2006). In essence INFORM is a hybrid model combining the Forest Light Interaction Model (FLIM) (Rosema et al. 1992), SAIL (Verhoef, 1984, 1985) and PROSPECT (Jacquemoud and Beret, 1990). INFORM simulates reflectance as a function of i) internal leaf parameters; ii) internal canopy parameters and iii) observation viewing geometry parameters (**Table 15**). The leaf optical properties calculated by PROSPECT were further used as inputs into the INFORM canopy reflectance model. Detailed information regarding the technicalities of INFORM may be found in Atzberger, (2000) and Schlerf and Atzberger, (2006).

4.3.2.1. RTM parameterization

In this study the ranges of input parameters describing leaf biochemistry and canopy structure were drawn from the observation made in other studies (Jacquemoud, 1993; Myneni et al. 1997; Ceccato et al. 2001; Privette et al. 2004; Bowyer and Danson, 2004; Main et al. 2011; Féret et al. 2017). Jacquemoud (1993) observed that *N* ranging from 1.5 – 2.5 which represents the internal structure of dicotyledons and in this study a value of 2.1 was assigned to *N*. Meanwhile chlorophyll was observed to range from 15 – 75 μg cm^{-2} amongst savannah trees (Main et al. 2011) and in this study we lowered the minimum value to 0.1 μg cm^{-2} to cater for early senescing tree species e.g. *Acacia nigrescence* (Madonsela et al. 2017). Consistent with Bowyer and Danson (2004) and Zhang et al. (2005) the minimum and maximum values for parameters not well studied in the savannah e.g. leaf dry matter and brown pigment were set to cover as much variability as possible in different biomes on earth. The range for equivalent water content was set between 0.001 g cm^{-2} and 0.047 g cm^{-2} and this was consistent with the observation of Gond et al. (1999) in temperate forests. Although savannahs are different from temperate forests, we found the EWT data applicable to the present study given that the two study areas receive approximately similar mean annual rainfall (Pita et al. 2013; Makhado et al. 2014; Eckhardt et al. 2000) and they are characterized by deciduous trees. We deliberately lowered the minimum EWT value to 0.001 g cm^{-2} to cater for rainfall variability in the savannah. The ranges for carotenoids were set between 1.82 and 10.4. Stem density per hectare and tree height were computed from our field data. LAI ranges were set based on previous studies in the savannah (Privette et al. 2004).

The viewing geometry parameters i.e. sun zenith angle, relative azimuth angle were obtained from the Landsat-8 metadata file. Observer zenith angle was kept at 0^{o} since Landsat images were collected from nadir position. With regards to the diffuse solar radiation, a fixed value of 0.1 was used across all wavelengths as in many similar studies (Schlerf and Atzberger, 2006). To represent soil optical properties, we extracted spectral reflectance of bare areas from Landsat-8 image.

4.3.2.2. Inversion strategy

Biochemical and biophysical attributes of the savannah trees were predicted from the Landsat-8 imagery through the inversion of the canopy reflectance models. Two inversion techniques i.e. Look Up Table (LUT) and Random Forest (RF) were tested to invert PROSAIL and INFORM. The LUT technique requires a large set of input variables which will be used to simulate canopy reflectance in a forward mode. To estimate canopy variables using the LUT technique, the root mean square error (RMSE) is calculated between the measured and simulated reflectance spectra. The estimated solution is regarded as a set of input variables corresponding to the reflectance in the LUT which have the lowest RMSE (Schlerf and Atzberger, 2006; Darvishzadeh et al. 2008). In this study parameter combination was randomly generated and used in the forward modelling. The minimum and maximum values of the model parameters are reported in **Table 15**. RF on the other hand is a machine learning regression algorithm (MLRA) which applies a set of decision trees to enhance prediction accuracy (Breiman, 2001). One major advantage of MLRA is the ability to capture non-linear relationships of image features without prior knowledge of the underlying data distribution. Hence MLRA can be applied without assuming a particular probability density distribution (Verrelst et al. 2015). In this study the inversion of canopy reflectance models using RF followed two steps i) the training of RF model using the simulated reflectance data and ii) the application of the trained RF model to the measured Landsat-8 spectra in order to predict tree canopy variables.

Table 15 Minimum and maximum set of values used to parameterize radiative transfer models.

Parameters	Designation	Unit	Min-Max values	
			PROSAIL	INFORM
Leaf Parameters				
Leaf internal structure	*N*	Unitless	2.1	2.1
Chlorophyll a and b	*Cab*	μg cm^{-2}	0.1 - 75	0.1 - 75
Carotenoids	*Car*	μg cm^{-2}	1.82 – 5.4	1.82 – 5.4
Brown pigment	*Cbrown*	g cm^{-2}	0.1 - 7	0.1 - 7
Equivalent water thickness	*Cw*	g cm^{-2}	0.001 – 0.047	0.001 – 0.047
Dry matter content	*Cm*	g cm^{-2}	0.0014 – 0.05	0.0014 – 0.05
Canopy Parameters				
Leaf area index of a single tree	*LAI*	m2 m-2	0.78 – 4.5	0.78 – 4.5
Leaf area index of the understorey	*LAIu*	m2 m-2		0.45
Average leaf angle	*ALA*	deg	55	55
Stem density per hectare	*SD*	Ha-1		18 - 179
Tree height	*H*	m		3 - 13

Crown diameter	*CD*	m		3 6.5
Viewing Geometry Parameters				
Sun zenith angle		deg	42	42
Azimuth angle		deg	52	52
Observation zenith angle		deg	0	0
Fraction of diffuse radiation	*skyl*	fraction	0.3	0.3

4.3.3. Data analysis

Initially, Shannon index (*H'*) was employed to quantify α-diversity within each field plot. Shannon index is a common diversity index in ecological literature (Colwell, 2009; Morris et al. 2014). *H'* considers both species richness and abundance when quantifying species diversity (Shannon and Weaver, 1949; Morris et al. 2014) and these aspects of diversity are considered to have a bearing on the reflectance spectra captured by the remote sensing device (Oldeland et al. 2010). Therefore Shannon index should be compatible with spectral data. *H'* is computed as follows:

Equation 6

$$H' = -\sum_{i=1}^{S} p_i \mathrm{In}(p_i)$$

where p_i is the proportional abundance of species i relative to the total abundance of all species *S* in a plot; $\mathrm{In}(p_i)$ is the natural logarithm of this proportion.

The RTM derived tree canopy variables (C_w, C_{ab}, C_{ar}, *LAI*, C_m, C_{brown}) were imported into MATLAB software v7.8.0 (R2009a, MathWorks) where bootstrap regression was conducted. These RTM derivatives were used as predictor variables in the linear regression model to predict tree species diversity in the savannah woodland. In order to assess the precision and the accuracy of the models, the bootstrapping approach was applied for modelling the relationship between spectral variability and species diversity. Firstly, we completed 1000 random permutation of the original data and then split two-thirds of the data for training the models and used the remainder for evaluating the predictive ability of the models. Modelling results are presented in table format in the subsequent section. Two modelling approaches i.e. univariate and multivariate analyses were tested and then followed by comparative analysis of the results. A simple linear regression model was used to investigate the relationship between RTM derivatives as predictor variables and tree species diversity indices as response variables. The strength of the relationship was assessed using the coefficient of determination (R^2), the *p-value* statistics and the model performance was evaluated using the root mean square error (RMSE).

4.4. Results

4.4.1. RTM derivatives: PROSAIL

The results of bootstrapped regression analysis showed that RTM derivatives particularly LAI, EWT and *Cbrown* has a significant positive relationship with tree species diversity ($p < 0.05$) (**Table 16**). Dry matter (*Cm*) on the other hand had a significant negative relationship with tree species diversity ($p < 0.05$) while *Cab* and *Car* had shown no relationship with tree species diversity. However, the relationship between RTM derivatives and tree species diversity declined with changing phenology from the end of growing season (March date) to dry season (July date). Moreover, the total variation in species diversity explained RTM derivatives was not significantly higher than that observed with NDVI. In fact EWT was the only RTM derivative which had explained the variation in tree species diversity at approximately the same level as NDVI and this performance was only observed in the end of growing season and during the transition to senescence periods. Meanwhile, during the senescence and dry season, NDVI showed a significantly higher relationship with tree species diversity ($p < 0.05$) than any RTM derivative. Generally, high relationships between predictor variables (NDVI and RTM derivatives) and tree species diversity were observed in the end of growing season (March date). In fact, the best model for estimating tree species diversity came from the equivalent water thickness and NDVI derived from the end of growing season dataset (**Figure 16**).

Table 16 Relationship observed between *H'* and RTM derivatives. RTM derivatives were generated through the inversion of PROSAIL model on Landsat-8 data from different dates. All statistics below were drawn from 1000 bootstrapped iterations.

Date	Diversity index	NDVI and RTM derivatives	Average R^2	Confidence interval 95%	*P-value*	RMSE
March	*H'*	NDVI	0.33	±0.019	0.000	0.454
		LAI	0.19	±0.014	0.000	0.492
		EWT	0.33	±0.017	0.000	0.448
		Cm	0.23	±0.018	0.003	0.479

		Cbrown	0.19	±0.015	0.005	0.491
		Cab	0.00	±0.000	0.045	0.543
		Car	0.00	±0.005	0.033	0.543
April	*H'*	NDVI	0.27	±0.022	0.000	0.464
		LAI	0.12	±0.015	0.015	0.515
		EWT	0.27	±0.016	0.001	0.467
		Cm	0.19	±0.015	0.006	0.494
		Cbrown	0.10	±0.013	0.020	0.519
		Cab	0.00	±0.002	0.031	0.544
		Car	0.00	±0.004	0.032	0.542
May	*H'*	NDVI	0.19	±0.019	0.004	0.492
		LAI	0.04	±0.014	0.024	0.535
		EWT	0.17	±0.013	0.008	0.498
		Cm	0.03	±0.014	0.022	0.535
		Cbrown	0.00	±0.003	0.035	0.541
		Cab	0.00	±0.001	0.021	0.543
		Car	0.13	±0.018	0.012	0.509
July	*H'*	NDVI	0.13	±0.019	0.011	0.513
		LAI	0.00	±0.006	0.031	0.541
		EWT	0.00	±0.000	0.043	0.542
		Cm	0.00	±0.001	0.028	0.547
		Cbrown	0.00	±0.000	NaN	0.544
		Cab	0.03	±0.013	0.023	0.535
		Car	0.00	±0.000	NaN	0.543

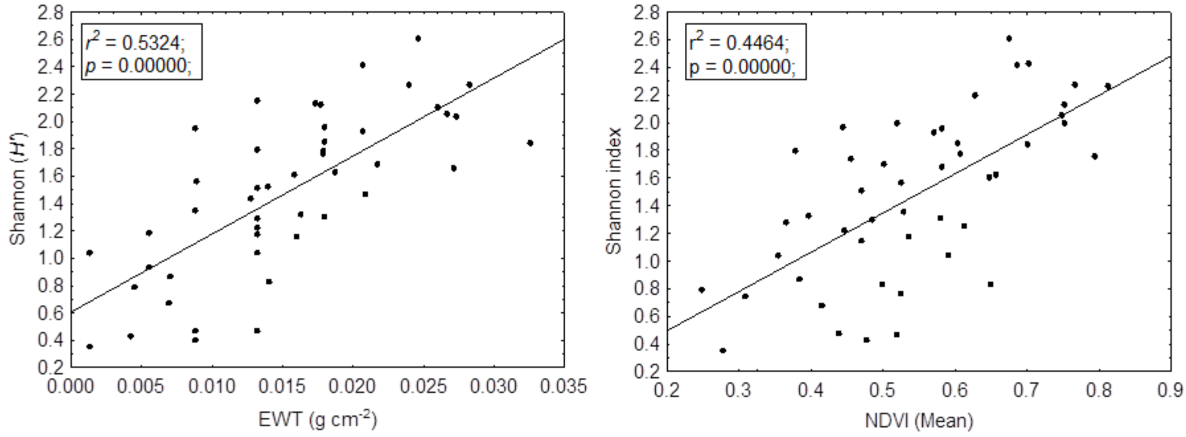

Figure 16 Scatterplot showing linear relationship between predictor variables (EWT and NDVI) and Shannon index. The scatterplots were selected from the best regression models (maximum r^2 with the lowest RMSE from the 1000 bootstrapped iterations).

Further results from the regression analysis showed that transforming RTM derivatives into PCs increase their explanatory power (**Table 17**). PCs associated with March and April dates had a significantly higher relationship with tree species diversity (r^2 of 0.45 and 0.34 respectively; $p < 0.05$) and also lower error of prediction compared to NDVI of the same dates (r^2 of 0.32 and 0.26 respectively; $p < 0.05$). Analysis of variance showed that the improvements in the estimation of tree species diversity using PCs as predictor variables were statistically significant ($p < 0.05$). However, PCs associated with May and July dates had a significant but lower relationship with tree species diversity (r^2 of 0.18 and 0.07 respectively; $p < 0.05$) compared to NDVI of the same dates (r^2 of 0.19 and 0.13 respectively; $p < 0.05$).

Table 17 Relationship observed between *H'* and PCs computed from RTM derivatives. RTM derivatives were generated through the inversion of PROSAIL model on Landsat-8 data from different dates. All statistics below were drawn from 1000 bootstrapped iterations.

Date	Diversity index	PCs explaining over 95%	Average R^2	Confidence interval 95%	*P-value*	RMSE
March	*H'*	5	0.43	±0.016	0.000	0.416
April	*H'*	5	0.33	±0.014	0.000	0.448
May	*H'*	5	0.21	±0.015	0.004	0.487
July	*H'*	6	0.06	±0.017	0.022	0.527

4.4.2. RTM derivatives: INFORM

The results of bootstrapped regression analysis showed that LAI, EWT, *Car, Cbrown* and dry matter derived from the inversion of INFORM had a significant but lower relationship with tree species diversity compared to NDVI (**Table 18**). The relationship deteriorated with changes in phenology from the end of the growing season to the dry season. However, *Cab* had no relationship with tree species diversity throughout all phenological periods. Contrary to the observation above (**Table 17**), the transformation of these biophysical and biochemical parameters through PCA did not improve their explanatory power (**Table 19**).

Table 18 Relationship observed between *H'* and RTM derivatives. RTM derivatives were generated through the inversion of INFORM model on Landsat-8 data from different dates. All statistics below were drawn from 1000 bootstrapped iterations.

Date	Diversity index	NDVI and RTM derivatives	Average R^2	Confidence interval 95%	*P-value*	RMSE
March	*H'*	NDVI	0.33	±0.019	0.000	0.454
		LAI	0.18	±0.017	0.007	0.499
		EWT	0.13	±0.015	0.014	0.511
		Cm	0.15	±0.015	0.011	0.505
		Cbrown	0.16	±0.016	0.009	0.501
		Cab	0.01	±0.007	0.036	0.541
		Car	0.16	±0.015	0.009	0.502
April	*H'*	NDVI	0.27	±0.022	0.000	0.464
		LAI	0.03	±0.013	0.024	0.536
		EWT	0.01	±0.011	0.025	0.537
		Cm	0.00	±0.002	0.029	0.543

		Cbrown	0.18	±0.014	0.007	0.497
		Cab	0.09	±0.017	0.016	0.517
		Car	0.04	±0.014	0.025	0.534
May	*H'*	NDVI	0.19	±0.019	0.004	0.492
		LAI	0.00	±0.006	0.031	0.542
		EWT	0.00	±0.005	0.035	0.542
		Cm	0.01	±0.010	0.026	0.539
		Cbrown	0.02	±0.011	0.026	0.539
		Cab	0.00	±0.003	0.035	0.542
		Car	0.00	±0.006	0.031	0.542
July	*H'*	NDVI	0.13	±0.019	0.011	0.513
		LAI	0.02	±0.009	0.028	0.538
		EWT	0.00	±0.002	0.022	0.543
		Cm	0.00	±0.003	0.033	0.543
		Cbrown	0.02	±0.009	0.028	0.539
		Cab	0.02	±0.010	0.028	0.539
		Car	0.02	±0.011	0.024	0.539

Table 19 Relationship observed between *H'* and PCs computed from RTM derivatives. RTM derivatives were generated through the inversion of INFORM model on Landsat-8 data from different dates. All statistics below were drawn from 1000 bootstrapped iterations.

Date	Diversity index	PCs explaining over 95%	Average R^2	Confidence interval 95%	*P-value*	RMSE
March	*H'*	3	0.18	±0.013	0.0068	0.519
April	*H'*	3	0.06	±0.017	0.0211	0.553
May	*H'*	3	0.01	±0.002	0.0287	0.566
July	*H'*	3	0.02	±0.003	0.0263	0.516

4.5. Discussion

The results of the study support the assertion in the literature (Carlson et al. 2006; Zhao et al. 2016) that LAI, EWT, *Cbrown* and dry matter are useful biochemical and biophysical parameters for explaining species diversity. LAI, EWT and *Cbrown* have all shown a significant positive relationship with Shannon index while dry matter had a negative relationship. The relationship between these biophysical and biochemical parameters and tree species diversity reflects the prevailing environmental conditions. For instance, in southern African savannahs, studies (Scholes et al. 2004; Privette et al. 2004) have shown that LAI is correlated to the rainfall gradient which in turn impact on tree species diversity (Shackleton, 2000). Unger et al. (2013) also observed that in tropical montane forests LAI, stem density and tree species diversity are inversely related to topography. Meanwhile, EWT is a key indicator of vegetation moisture status which has been observed to increase with increasing vegetation density in South American savannah while also showing an inverse relationship to land surface temperatures (Ferreira et al. 2011). In southern African savannah, the northern part of KNP with its low mean annual precipitation (<450mm) and high temperatures (Makhado et al. 2013), has low species diversity and mainly supports the dominance of *Colophospermum mopane*. In light of the evidence that LAI and EWT are related to precipitation (Scholes et al. 2004; Privette et al. 2004; Ferreira et al. 2011), the relationship between these biophysical and biochemical parameters with tree species diversity is governed by environmental variables which also impacts on the diversity of tree species.

Contrary to our expectation though, these biochemical and biophysical parameters were not better than the NDVI model in explaining species diversity in the savannah woodland. In fact *Cab* and *Car* had very low to no relationship with Shannon index throughout all phenological periods and this was attributed to the broad spectral resolution of Landsat-8 which could have led to very poor retrieval of *Cab* and *Car*. The accurate retrieval of leaf pigments i.e. *Cab* and *Car* necessitates high resolution data with spectral bands in the yellow and red-edge regions of the electromagnetic spectrum which have been shown to be sensitive to leaf pigments concentration (Gitelson et al. 2002; Cho and Skidmore, 2006; Mutanga and Skidmore, 2007). It was only the EWT that had explained tree species diversity at approximately the same way as NDVI particularly in the end of growing season (March date) (r^2 of 0.34 and 0.33 respectively) and during the transition to senescence period (r^2 of 0.26 and 0.27 respectively). Meanwhile, the other parameters (LAI, dry matter, C_{brown}) had lower explanatory power than NDVI. In addition, the ability of these parameters to explain species diversity also declined with changes in phenology from the end of the growing season to dry season thus presenting themselves as phenology constrained predictors just like the NDVI.

However, the lower relationship between biochemical parameters (particularly LAI, *Cab*, *Car*, *Cbrown* and dry matter) and Shannon index is not the reflection of poor explanatory power of biochemical parameters in general. In fact the utility of biochemical parameters in explaining species diversity have been shown in Hawaiian dry forest (Carlson et al. 2006) and China's subtropical forest (Zhao et al. 2016). Carlson et al. (2006) and Zhao et al. (2016) relied on hyperspectral sensors to detect biochemical variability between plant species and used it to map species diversity. In this study, LAI, EWT *Cab*, *Car*, *Cbrown* and dry matter were retrieved from Landsat-8 imagery which is a broad multispectral data with 30m spatial resolution. The biophysical data captured by Landsat-8 sensor is largely inadequate since the sensor has medium spatial resolution and coarse spectral bands which renders it insufficient for detecting subtle details on vegetation canopy structure and biochemistry (Carlson et al. 2006; Mutanga et al. 2009). For instance Cho et al. (2014) observed that PROSAIL inversion on high resolution SPOT 6 image produced accurate LAI estimates (RMSE of 0.85) than inversion on MODIS image with a 250m spatial resolution (RMSE of 1.26). Therefore, the lower predictive potential of the retrieved biochemical parameters should be considered as a reflection of Landsat-8 sensor's inadequacy to capture fine details on vegetation canopy structure and biochemistry.

Moreover, the retrieval of LAI, EWT *Cab*, *Car*, *Cbrown* and dry matter relied on the inversion of PROSAIL which is a 1-D model often used to simulate the bidirectional reflectance factor of a turbid medium plant canopies (Jacquemoud et al. 2009; Jacquemoud, 1993). The southern African savannah biome is characterized by heterogeneous vegetation structure (du Toit et al. 2003) and the turbid medium assumption of PROSAIL does not cater for the structural heterogeneity in vegetation canopies (Jacquemoud, 1993; Darvishzadeh et al. 2008; Atzberger et al. 2015). The violation of turbid medium assumption lowers the ability of PROSAIL to accurately simulate canopy reflectance and ultimately introduces bias in the retrieval of biophysical and biochemical variables (Darvishzadeh et al. 2008). In fact this is considered the main limitation of PROSAIL which causes it to perform similarly to empirical models (Jacquemoud et al. 2009). In light of the above, it is possible that the limitations of PROSAIL coupled with Landsat-8 inadequacy to capture fine details on vegetation canopy have lowered the capacity of biochemical parameters to explain species diversity.

However, even the inversion of INFORM which was developed to simulate bidirectional reflectance of forest canopies (Atzberger, 2000; Schlerf and Atzberger, 2006), did not improve the accuracy of retrieval. Hence the retrieved biophysical and biochemical parameters had lower explanatory power than NDVI in all phenological periods. Retrieval of biochemical parameters through inversion of INFORM has only been tested in closed canopy forests (Schlerf and Atzberger, 2006; Ali et al. 2016) and the results of our study suggest that it may not be suited for open canopy environment such as the savannah biome.

Nonetheless, the results of the study showed that radiative transfer modelling offers an opportunity to derive several biophysical and biochemical parameters such as LAI, EWT, *Cbrown* and dry matter which had a significant relationship with species diversity. The application of PCA to transform these biophysical and biochemical parameters into PCs improved their capacity to explain tree species diversity (Table 3). The improvement supports the assertion that while Landsat-8 may not be capable of detecting fine details on vegetation canopies, radiative transfer modelling do offer an opportunity to retrieve useful biochemical and biophysical parameters from the visible, NIR and SWIR regions of the electromagnetic spectrum. This assertion should be tested in future studies using remote sensing data with high resolution spatial and spectral resolution in order to facilitate accurate detect of LAI, EWT *Cab*, *Car*, *Cbrown* and dry matter.

4.6. Conclusion

In conclusion this study demonstrated that useful information on the biophysical and biochemical parameters of vegetation may be extracted from remotely sensed data through radiative transfer modelling. Although the relationship between tree species diversity and EWT, LAI, dry matter and *Cbrown* was not higher than that observed with NDVI, its existence particularly in the end of the growing season suggest that radiative transfer modelling is capable of retrieving useful information on the biophysical and biochemical parameters from the visible, NIR and SWIR regions of the electromagnetic spectrum. In fact the lower relationship that RTM derivatives had with Shannon index when compared with NDVI is attributable to Landsat-8 inadequacy to capture fine details on vegetation canopy structure and biochemistry due to its coarse spectral bands. Hence this study recommends that future studies pursue the research question investigated here on high resolution data which captures subtle biochemical and biophysical differences between tree canopies.

Acknowledgements

The authors would like to express gratitude to the National Research Foundation for funding this research project through the NRF-Professional Development Programme. We would also like to thank Cecilia Masemola, Sbu Gumede, Joseph Dlamini, Martin Sarela and Patrick Ndlovu for their assistance during field work.

References

Ali, A.M., Darvishzadeh, R., Skidmore, A.K. and van Duren, I., 2016. Effects of canopy structural variables on retrieval of leaf dry matter content and specific leaf area from remotely sensed data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*(2), pp.898-909.

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

Atzberger, C., Darvishzadeh, R., Immitzer, M., Schlerf, M., Skidmore, A. and Le Maire, G., 2015. Comparative analysis of different retrieval methods for mapping grassland leaf area index using airborne imaging spectroscopy. *International Journal of Applied Earth Observation and Geoinformation*, *43*, pp.19-31.

Bro, R. and Smilde, A.K., 2014. Principal component analysis. *Analytical Methods*, *6*(9), pp.2812-2831.

Carlson, K.M., Asner, G.P., Hughes, R.F., Ostertag, R. and Martin, R.E., 2007. Hyperspectral remote sensing of canopy biodiversity in Hawaiian lowland rainforests. *Ecosystems*, *10*(4), pp.536-549.

Ceccato, P., Flasse, S., Tarantola, S., Jacquemoud, S. and Grégoire, J.M., 2001. Detecting vegetation leaf water content using reflectance in the optical domain. *Remote sensing of environment*, *77*(1), pp.22-33.

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

Cho, M.A., Debba, P., Mathieu, R., Naidoo, L., Van Aardt, J. and Asner, G.P., 2010. Improving discrimination of savanna tree species through a multiple-endmember spectral angle mapper approach: Canopy-level analysis. *IEEE Transactions on Geoscience and Remote Sensing*, *48*(11), pp.4133-4142.

Cho, M.A., Ramoelo, A. and Holloway, J., 2014, October. Estimating leaf area index (LAI) by inversion of prosail radiative transfer model using SPOT 6 imagery. In *Proceedings of the 10th International Conference of AARSE* (p. 50).

Darvishzadeh, R., Skidmore, A., Schlerf, M. and Atzberger, C., 2008. Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in a heterogeneous grassland. *Remote Sensing of Environment*, *112*(5), pp.2592-2604.

du Toit, J.T., Biggs, H.C., Rogers, K.H., 2003. The Kruger Experience: ecology and management of savanna heterogeneity. London: Island Press.

Ferreira, L.G., Asner, G.P., Knapp, D.E., Davidson, E.A., Coe, M., Bustamante, M.M. and de Oliveira, E.L., 2011. Equivalent water thickness in savanna ecosystems: MODIS estimates based on ground and EO-1 Hyperion data. *International journal of remote sensing*, *32*(22), pp.7423-7440.

Gould, W., 2000. Remote sensing of vegetation, plant species richness, and regional biodiversity hotspots. *Ecological applications*, *10*(6), pp.1861-1870.

Grant, C.C. and Scholes, M.C., 2006. The importance of nutrient hot-spots in the conservation and management of large wild mammalian herbivores in semi-arid savannas. *Biological Conservation*, *130*(3), pp.426-437.

Jacquemoud, S., 1993. Inversion of the PROSPECT+ SAIL canopy reflectance model from AVIRIS equivalent spectra: theoretical study. *Remote sensing of environment*, *44*(2-3), pp.281-292.

Jacquemoud, S., Baret, F., Andrieu, B., Danson, F.M. and Jaggard, K., 1995. Extraction of vegetation biophysical parameters by inversion of the PROSPECT+ SAIL models on sugar beet canopy reflectance data. Application to TM and AVIRIS sensors. *Remote sensing of environment*, *52*(3), pp.163-172.

Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P.J., Asner, G.P., François, C. and Ustin, S.L., 2009. PROSPECT+ SAIL models: A review of use for vegetation characterization. *Remote sensing of environment*, *113*, pp.S56-S66.

Jetz, W., Cavender-Bares, J., Pavlick, R., Schimel, D., Davis, F.W., Asner, G.P., Guralnick, R., Kattge, J., Latimer, A.M., Moorcroft, P. and Schaepman, M.E., 2016. Monitoring plant functional diversity from space. *Nature plants*, *2*(3).

Kerr, J.T. and Ostrovsky, M., 2003. From space to species: ecological applications for remote sensing. *Trends in Ecology & Evolution*, *18*(6), pp.299-305.

Kötz, B., Schaepman, M., Morsdorf, F., Bowyer, P., Itten, K. and Allgöwer, B., 2004. Radiative transfer modeling within a heterogeneous canopy for estimation of forest fire fuel properties. *Remote Sensing of Environment*, *92*(3), pp.332-344.

Ludwig, F., De Kroon, H., Berendse, F. and Prins, H.H., 2004. The influence of savanna trees on nutrient, water and light availability and the understorey vegetation. *Plant Ecology*, *170*(1), pp.93-105.

Madonsela, S., Cho, M.A., Mathieu, R., Mutanga, O., Ramoelo, A., Kaszta, Ż., Van De Kerchove, R. and Wolff, E., 2017. Multi-phenology WorldView-2 imagery improves remote sensing of savannah tree species. *International Journal of Applied Earth Observation and Geoinformation*, *58*, pp.65-73.

Makhado, R.A., Mapaure, I., Potgieter, M.J., Luus-Powell, W.J. and Saidi, A.T., 2014. Factors influencing the adaptation and distribution of Colophospermum mopane in southern Africa's mopane savannas-A review. *Bothalia-African Biodiversity & Conservation*, *44*(1), pp.1-9.

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

Mutanga, O., Van Aardt, J. and Kumar, L., 2009. Imaging spectroscopy (hyperspectral remote sensing) in southern Africa: an overview. *South African Journal of Science*, *105*(5-6), pp.193-198.

Myneni, R.B., Ramakrishna, R., Nemani, R. and Running, S.W., 1997. Estimation of global leaf area index and absorbed PAR using radiative transfer models. *IEEE Transactions on Geoscience and remote sensing*, *35*(6), pp.1380-1393.

Myneni, R.B., Hoffman, S., Knyazikhin, Y., Privette, J.L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G.R. and Lotsch, A., 2002. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote sensing of environment*, *83*(1), pp.214-231.

Privette, J.L., Tian, Y., Roberts, G., Scholes, R.J., Wang, Y., Caylor, K.K., Frost, P. and Mukelabai, M., 2004. Vegetation structure characteristics and relationships of Kalahari woodlands and savannas. *Global Change Biology*, *10*(3), pp.281-291.

Richter, R. and Schläpfer, D., 2012. Atmospheric/Topographic Correction for Satellite Imagery (ATCOR-2/3 User Guide, Version 8.2 BETA). *German Aerospace Center, Remote Sensing Data Center: Wessling, Germany*.

Scholes, R.J., Frost, P.G. and Tian, Y., 2004. Canopy structure in savannas along a moisture gradient on Kalahari sands. *Global Change Biology*, *10*(3), pp.292-302.

Seymour, C.L. and Dean, W.R.J., 2010. The influence of changes in habitat structure on the species composition of bird assemblages in the southern Kalahari. *Austral Ecology*, *35*(5), pp.581-592.

Shackleton, C.M., 2000. Comparison of plant diversity in protected and communal lands in the Bushbuckridge lowveld savanna, South Africa. *Biological Conservation*, *94*(3), pp.273-285.

Staver, A.C., Archibald, S. and Levin, S., 2011. Tree cover in sub-Saharan Africa: Rainfall and fire constrain forest and savanna as alternative stable states. *Ecology*, *92*(5), pp.1063-1072.

Thenkabail, P.S., Hall, J., Lin, T., Ashton, M.S., Harris, D. and Enclona, E.A., 2003. Detecting floristic structure and pattern across topographic and moisture gradients in a mixed species Central African forest using IKONOS and Landsat-7 ETM+ images. *International Journal of Applied Earth Observation and Geoinformation*, *4*(3), pp.255-270.

Treydte, A.C., Heitkönig, I.M.A., Prins, H.H.T. and Ludwig, F., 2007. Trees enhance grass layer quality in African savannas of distinct rainfall and soil fertility. *Perspectives in Plant Ecology, Evolution and Systematics*, *8*, pp.197-205.

Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E. and Steininger, M., 2003. Remote sensing for biodiversity science and conservation. *Trends in ecology & evolution*, *18*(6), pp.306-314.

Unger, M.A.L.T.E., Homeier, J.Ü.R.G.E.N. and Leuschner, C., 2013. Relationships among leaf area index, below-canopy light availability and tree diversity along a transect from tropical lowland to montane forests in NE Ecuador. *Tropical Ecology*, *54*(1), pp.33-45.

Xie, Y., Sha, Z. and Yu, M., 2008. Remote sensing imagery in vegetation mapping: a review. *Journal of plant ecology*, *1*(1), pp.9-23.

Zhang, Q., Xiao, X., Braswell, B., Linder, E., Baret, F. and Moore, B., 2005. Estimating light absorption by chlorophyll, leaf and canopy in a deciduous broadleaf forest using MODIS data and a radiative transfer model. *Remote Sensing of Environment*, *99*(3), pp.357-371.

Zhao, Y., Zeng, Y., Zhao, D., Wu, B. and Zhao, Q., 2016. The optimal leaf biochemical selection for mapping species diversity based on imaging spectroscopy. *Remote Sensing*, *8*(3), p.216.

Chapter 5: Investigating the possibility of using Spectral Angle Mapper (SAM) as a new measure of spectral variability to explore the relationship between spectral variability and tree species diversity in the savannah woodlands

This chapter is based on

Madonsela, S., Cho, M.A., Ramoelo, A. and Mutanga, O., (*in preparation*). Exploring the relationship between spectral variability and tree species diversity in the savannah woodlands: a case of Spectral Angle Mapper as a new technique for Spectral Variation Hypothesis

Abstract

Spectral Variation Hypothesis (SVH) has recently emerged as the multi-dimensional analytical approach to establish relationship between spectral variability from the image and tree species diversity measured on the ground. SVH argue that spectral variability from the image emanates particularly from the spatial heterogeneity of the environment which by default has high species diversity due to the higher number of available niches. The aim of this study is to investigate the possibility of using Spectral Angle Mapper (SAM) as a new measure of spectral variability and explore the relationship between spectral variability and tree species diversity in the savannah woodlands. To accomplish our aim, we designed stratified random sampling approach and surveyed tree species in 68 plots of 90m X 90m across the study area. Within each plot, all trees with diameter at breast height of >10cm were sampled and Shannon index - a common measure of species diversity which considers both species richness and abundance - was used to quantify tree species diversity. The 90 m x 90 m sampling plots used to collect the tree species data in the field were overlaid on each Landsat-8 image from four phenological periods. We then extracted nine pixel spectra corresponding to each field plot and apply SAM as a measure of spectral variability. SAM computes pairwise pixel spectral angle and the large spectral angle denotes high spectral variability. Eventually, we had 36 combinations from which mean SAM and total SAM was computed. The mean SAM produced the average spectral variability from the 36 combinations while the total SAM produced the sum of spectral variability. The study applied linear regression model to explore the relationship between spectral variability on the Landsat-8 imagery and tree species diversity as measured by Shannon index. The results of the regression model showed a significant but low relationship between Landsat-8 spectral variability as measured by total SAM and tree species diversity (r^2 of 0.12; $p< 0.05$) in the end of growing season. Using mean SAM as a measure of spectral variability increased the relationship between Landsat-8 spectral variability and tree species diversity in the end of the growing season (r^2 of 0.24; $p< 0.01$) while no improvements were observed in other phenological periods. The results showed that the application of mean SAM, in particular, facilitates the extraction of useful spectral information from Landsat-8 image particularly in the end of growing season. The strength of mean SAM lies in its ability to offset extreme pixel values. Hence spectral variability measured with mean SAM had higher relationship to tree species diversity.

5.1. Introduction

The diversity of tree species is an important attribute of the savannah woodland which enables the biome to render many goods and services (du Toit et al. 2003; Shackleton et al. 2005; Treydte et al. 2007). For instance, large trees of *Acacia tortilis* improve nutrient quality of grass beneath their canopies to the benefit of herbivores (Ludwig et al. 2008); *Acacia spp. Combretum apiculatum and Sclerocarya birrea* etc. supports browsing requirement of large African ungulates (Bond and Loffell, 2001; du Toit et al. 2003); *Combretum collinum*, *Diospyros mespiliformis*, *Philenoptera violacea*, *Sclerocarya birrea* and *Ziziphus mucronata* serves as livelihood assets to communities of Limpopo, South Africa as they derive various non-timber forest products of economic value (Shackleton et al. 2007; Shackleton et al. 2005). Therefore, the loss of tree species diversity as a result of increased intensity of elephant damage in protected areas (Druce et al. 2008) and over-harvesting of woody resources threatens the many goods and services that are rendered by the savannah ecosystem (Shackleton et al. 2005; Madubansi and Shackleton, 2006; Matsika et al. 2013). South Africa's National Park has developed a Threshold of Potential Concerns (TPCs) as a means to detect potential threats to key elements of biodiversity (Druce, 2008). The success of such monitoring programme depends on the availability of spatially detailed information on the distribution and abundance of tree species (Turner et al. 2003). Spaceborne remote sensing meets this requirement as it collects data over large geographic area on a regular interval and at varying levels of spatial details (Kerr and Ostrovsky, 2003; Rocchini et al. 2015; Jetz et al. 2016). Recently, ecologists have embraced remote sensing science as a viable means for studying biodiversity over regional scale (Pereira et al. 2013; Jetz et al. 2016).

The application of remote sensing in biodiversity estimation is based on establishing the relationship between spectral information on the image and the species diversity measured in the field (Foody and Cutler, 2003; Carlson et al. 2007; Nagendra et al. 2010). Frequently, studies have exploited the red and near infrared (NIR) regions of electromagnetic spectrum through the normalized difference vegetation index (NDVI) to explain species diversity (Gould, 2000; Oindo and Skidmore, 2002; Fairbanks and McGwire, 2004; He et al. 2009; Parviainen et al. 2010; Pau et al. 2012). The use of NDVI is based on the understanding that it is sensitive to primary productivity which defines variation in plant diversity (Oindo and

Skidmore, 2002; Parviainen et al. 2010; Witman et al. 2008). However, the aforementioned studies used only the subset of spectral data acquired by remote sensing devices. The NDVI is limited to the small part of the visible and NIR regions while the shortwave infrared (SWIR) region also contains essential spectral information useful for explaining forest characteristics (Thenkabail et al. 2003; Rocchini, 2007). Carlson et al. (2007) demonstrated that spectral bands centred at 530, 720, 1 201 and 1 523nm contains biochemical information necessary for explaining woody species richness in Hawaiian rainforests. Several studies have attributed the success of Landsat data in explaining species diversity (Rocchini, 2007; Stickler and Southworth, 2008) and floristic structure (Thenkabail et al. 2003) to its shortwave infrared bands.

Therefore, there is a need for multivariate techniques that uses spectral information across the visible, NIR and SWIR regions of the electromagnetic spectrum to explain species diversity. Recently, Spectral Variation Hypothesis (SVH) has emerged as the multi-dimensional analytical approach to establish relationship between spectral variability on the image and tree species diversity measured on the ground (Rocchini, 2007; Oldeland et al. 2010; Hernández-Stefanoni et al. 2012). SVH argue that spectral variability on the image emanates particularly from the spatial heterogeneity of the environment which by default has high species diversity due to the higher number of available niches (Palmer et al. 2002; Rocchini et al. 2010). In particular different plant species display different spectral responses to light in the electromagnetic spectrum linked to their biochemical and biophysical attributes (Nagendra 2001; Nagendra and Rocchini, 2008; Asner and Marin 2009; Cho et al. 2010). As such spectral variability on the image has been considered a proxy for species diversity in light of the SVH (Palmer et al. 2002; Rocchini, 2007; Rocchini et al. 2010).

However, SVH has seldom been explored within multi-phenological space despite phenology being an important variable affecting plant species spectral response (Gilmore et al. 2008; Hill et al. 2010; Madonsela et al. 2017a). Phenological variations between tree species enhance spectral variability. This take place when tree species are at different phenological stages in the same image or when phenological changes occur at different rates between images over a growing season (Hill et al. 2010; Madonsela et al. 2017a). In light of the fact that SVH concept is based on spectral variability, it is necessary to test its potential within the multi-phenological space.

Previous studies that have successfully established a relationship between spectral variability and field measured species diversity used mean distance from spectral centroid as a measure of spectral variability (Rocchini, 2007; Oldeland et al. 2010). Mean distance to spectral centroid is a multivariate technique enabling the exploitation of spectral information across the visible, NIR and SWIR regions of electromagnetic spectrum (Rocchini et al. 2010). This study will investigate the possibility of using Spectral Angle Mapper (SAM) as a new measure of spectral variability and explore the relationship between spectral variability and tree species diversity in the savannah woodlands. SAM is a mathematical algorithm that has been used to selects bands which increase spectral angle between target species (Keshava, 2004; Cho et al. 2010). SAM has been defined as the angle (θ) between two spectra $S_i = S_{i1},\ldots,S_{iL}$ and $s_j = s_{j1},\ldots,s_{jL}$

$$SAM\left(s_i,\, s_j\right) = \theta\left(s_i,\, s_j\right) = cos^{-1}\left(\frac{\sum_{l=1}^{L} s_{il}s_{jl}}{\left[\sum_{l=1}^{L} s_{il}^2\right]^{1/2}\left[\sum_{l=1}^{L} s_{jl}^2\right]^{1/2}}\right) \quad \text{Equation 1}$$

where L is the number of bands. SAM computes pairwise pixel spectral angle using equation 1 and the large spectral angle denotes high spectral variability (Keshava, 2004; Cho et al. 2010). In the context of SVH high spectral angle between Landsat-8 pixels indicates high environmental heterogeneity which in turn is linked to high tree species diversity. The aim of this study is i) to investigate the possibility of using Spectral Angle Mapper (SAM) as a new measure of spectral variability and explore the relationship between spectral variability and tree species diversity in the savannah woodlands and ii) to investigate whether multi-phenological spectral variability improves the estimation of tree species diversity in the savannah woodlands.

5.2. Study area

The study was conducted within the broader savannah woodland belt stretching across the KwaZulu-Natal (KZN), Mpumalanga and Limpopo provinces of South Africa (**Figure 17**). The area is characterized by two land management regimes; i) the public nature reserve entities with a mandate to conserve savannah biodiversity i.e. Kruger National Park (KNP) and

Hluhluwe-Imfolozi (HIP) Park and ii) the communal areas adjacent these conservation areas. Typical of the savannah biome, two vegetation lifeforms characterized the study area; the continuous grass layer interspersed by woody vegetation cover (Scholes and Archer, 1997; Sankaran et al., 2005). Fire and rainfall in particular and herbivory have been identified as key mechanisms maintaining balanced distributional patterns between these two lifeforms (Sankaran et al. 2005; Bond et al. 2003).

In terms of geological formations, granite substrate dominates in the western part while gabbro substrate dominates in the eastern part. Gabbro substrates is characterized by shallow to moderately deep, dark clay soils with high-bulk, nutrient rich grasses and supports few scattered trees mostly *Acacia spp.* (du Toit et al. 2003). On the contrary, granite substrate is characterized by nutrient-poor, shallow to moderately deep sandy soils with gently undulating terrain and it hosts broad-leaved deciduous tree species upslope while fine-leaved species occupy downslope. The granitic substrates are characterized by high species diversity and notable tree species includes *Combretum spp*, *Acacia nigrescens*, *Spirostachys africana* and *Sclerocarya birrea* (du Toit et al. 2003; Eckhardt et al. 2000). The northern portion of the study area is also characterized by the dominance of *Colophospermum mopane* (Makhado et al. 2013; Eckhardt et al. 2000). Meanwhile, the KZN part of the study area is characterized by mountainous terrain with different habitat types supporting a large number of plant species. Typical savannah species includes *Dichrostachys cinerea* and various species of *Euclea* and *Acacia* (Dumalisile, 2009). The area is characterized by north-south rainfall gradient. The mean annual precipitation ranges from 750mm in the southern portion of KNP to 440mm in the north with notable variations around the mean from year to year (Makhado et al. 2013; Eckhardt et al. 2000).

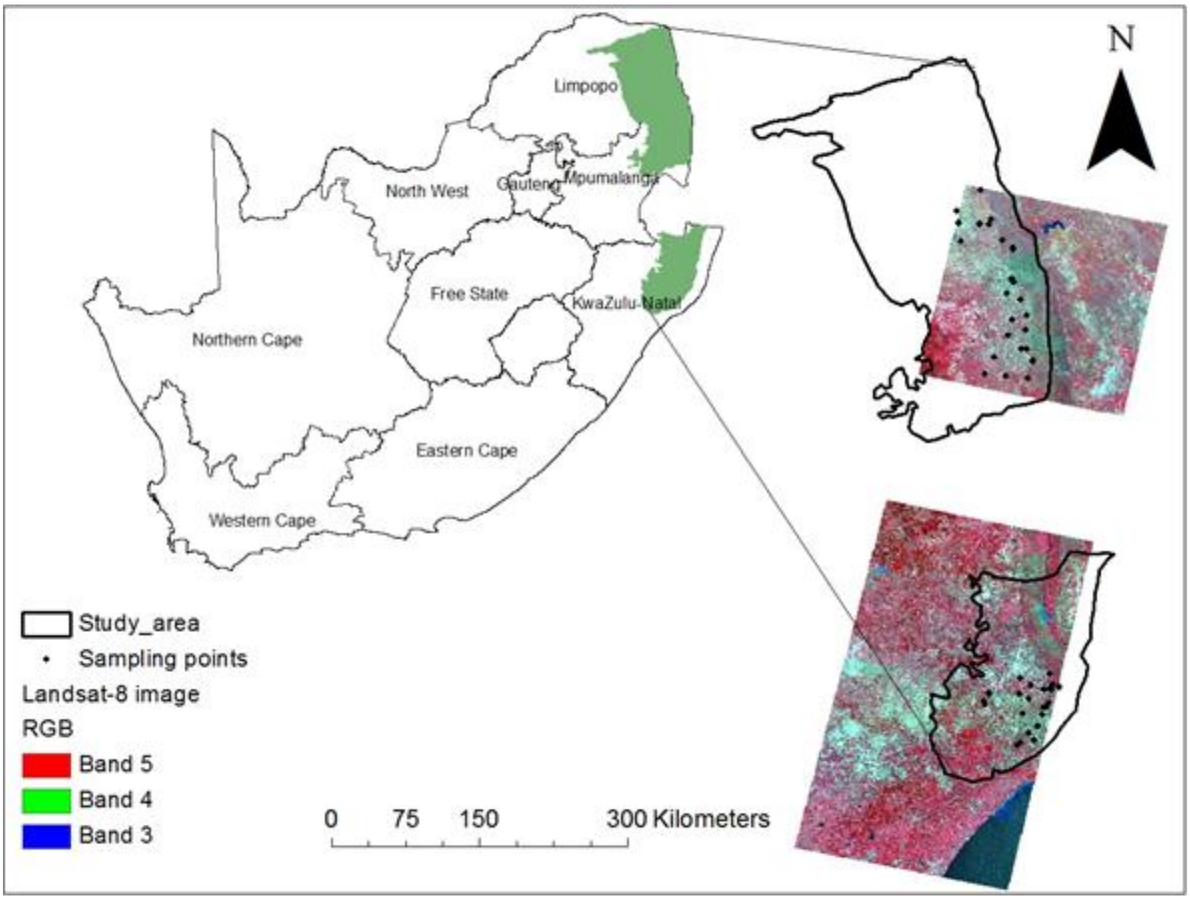

Figure 17 Study area stretching across three provinces of South Africa. The black dots represent the sampling plots

5.3. Materials and methods

5.3.1. Remote sensing data

Four Landsat-8 Operational Land Imager (OLI) satellite images captured in 2016 (28th of March, 29th of April, 31st of May and 24th of July) were downloaded from the United States Geological Surveys (USGS) portal (https://earthexplorer.usgs.gov/). These images were collected on different dates in order to investigate whether multi-phenological spectral variability improves the estimation of tree species diversity in the savannah woodlands. The end of March represents the end of growing season (Grant and Scholes, 2006); April represents transition to senescence (Madonsela et al., 2017); May represents advanced senescence when most trees starts to drop off leaves and grass will be at their senescent stage (Scholes et al. 2003; Cho et al. 2010); July corresponds to dry season in southern African savannah (du Toit et al. 2003; Kaszta et al., 2016). SAM was computed on each Landsat image and the Landsat-derived SAM of March, April, May and July are referred to as SAM_{March}, SAM_{April}, SAM_{May} and SAM_{July} respectively.

Landsat-8 OLI is a multi-spectral sensor with eight spectral bands in the visible, near infrared and shortwave infrared regions of electromagnetic spectrum. Landsat-8 OLI record data at moderate spatial resolution of 30m and has a revisit capacity of 16 days. The 12-bit quantization of data has improved the signal-to-noise radiometric performance of the sensor over its predecessors, thus increasing its usefulness for landcover mapping (Pervez et al., 2016). The Landsat-8 images were downloaded with geometric correction already completed. The Mpumalanga and Limpopo images were atmospherically corrected using the ATCOR-2 software since the area exhibit gently undulating slopes (Richter and Schläpfer, 2012). The KZN Landsat scenes necessitated the use of ATCOR-3 software since the region is mountainous. ATCOR-3 allows for integration of DEM which is useful for the correction of shadow and topographic effects on the image depicting mountainous areas (Richter and Schläpfer, 2012).

5.3.2. Field data collection

Field data were collected from the 2nd till the 27th of November 2015 in KwaZulu-Natal and again on the 1st till the 19th of March 2016 across Kruger National Park stretching between Mpumalanga and Limpopo provinces. The principal aim of the field campaign was to identify tree species within randomly placed sampling plots and quantify local species diversity (α-diversity) in the region using the common measure of diversity i.e. Shannon index. Prior to field excursion we defined the size of field sampling plots using semi-variogram analysis in ENVI 4.8 software. Essentially semi-variogram quantifies the spatial variability of natural phenomenon occurring in space (Fu et al., 2014; Gringarten and Deutsch, 2001). Semi-variogram is computed as follow:

Equation 7

$$y(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2$$

where $y(h)$ is the semi-variance at a given distance h; $z(x_i)$ is the value of the variable Z at location x_i; h is the lag distance and $N(h)$ is the number of pairs of sample points separated by h.

Semi-variance steadily increases as the distance from one location to the next increases till it reaches the range where it starts to level off (Jongman et al., 1995; Gringarten and Deutsch, 2001). Semi-variogram plot is generated by computing variance at different lag distances and a theoretical model such as spherical or exponential model is fitted to provide information about spatial structure (Fu et al., 2014). Our study applied semi-variogram analysis to resampled WorldView-2 derived NDVI image to define the scale of spatial variability in tree species richness. The choice to use NDVI was based on the observation that variability in NDVI is related to species diversity (Gould, 2000).

In our analysis, the Worldview-2 image – covering only a small part of the study area - was firstly resampled to 10m spatial resolution to be compatible with average tree canopy size in the savannah (Cho et al., 2012) and then we generated NDVI image. In ENVI software v4.8 the semi-variogram analysis computed the squared difference between neighbouring pixel values in order to quantify variability. The analysis conducted on Worldview-2 derived NDVI image showed that the scale for tree species variability in the savannah woodland lies at a range of 90m (**Figure 18**). Although semi-variance kept increasing beyond the range, the increase was not consistent and the range of 90m resulted in plot sizes that are feasible to work on within limited resources. Moreover, the study intended to use Landsat data with 30m pixel resolution, hence the plot size of 90m X 90m was considered adequate to ascertain correspondence between field data and spectral data.

The plot size of 90m X 90m was therefore chosen to capture spatial variation in tree species diversity. Stratified random sampling was used to define the placement of sampling plots. The stratification of sampling plots followed four dominant geological formations (granite; siliciclastic; gabbros; granulite) that were observed to have marked influence over vegetation patterns in the study area (du Toit et al., 2003). Plots of 90m x 90m were designed ensuring that corners of each plot correspond to Landsat pixels by following pre-defined GPS points of each corner. Within the plots all trees with diameter at breast height (DBH) above 10cm were recorded with Global Positioning System and species identified. Eventually we collected 5859 trees belonging to 106 tree species. The field campaign visited 50 plots distributed across the study area and collected tree species data. Further 26 plots collected under similar conditions in the previous study (Naidoo et al., 2015) were added to our field data. However some of these field plots (8 plots) were located on clouded parts of the March and April images and therefore not usable. In total 68 field plots were used in the analysis. We also extracted mean annual rainfall for each plot from the interpolated rainfall data produced by South African National Parks Scientific Services.

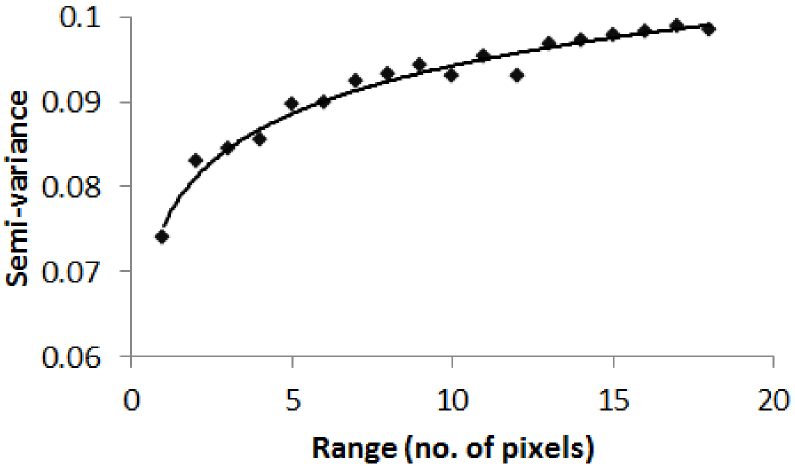

Figure 18 Semi-variogram analysis showing the scale of tree species variability in the savannah woodland

5.3.3. Data analysis

We quantified α-diversity within each plot using the Shannon index (*H'*) which is common measure of diversity in ecological literature (Colwell, 2009; Morris et al., 2014) and was preferred to ensure consistency of our findings with previous studies. *H'* considers both species richness (i.e. number of different tree species) and abundance (i.e. number of individual trees within species) when quantifying species diversity (Shannon and Weaver, 1949; Morris et al. 2014) and these aspects of diversity are considered to have a bearing on the reflectance spectra captured by remote sensing device (Oldeland et al. 2010; Madonsela et al. 2017b). Moreover, Shannon index is considered to be sensitive to vegetation structure (Oldeland et al. 2010). Therefore Shannon index should relate well with spectral data and vegetation structural variables. *H'* is computed as follows:-

Equation 8

$$H' = \quad -\sum_{i=1}^{s} p_i \mathrm{In}(p_i)$$

where p_i is the proportional abundance of species i relative to the total abundance of all species *S* in a plot; $\mathrm{In}(p_i)$ is the natural logarithm of this proportion.

The 90 m x 90 m sampling plots used to collect the tree species data in the field were overlaid on each Landsat-8 image. We then extracted nine pixel spectra corresponding to each field plot and apply SAM as a measure of spectral variability (equation 1). SAM computes pairwise pixel spectral angle using equation 1 and the large spectral angle denotes high spectral variability (Keshava, 2004; Cho et al. 2010). Eventually, we had 36 combinations from which mean SAM and total SAM was computed. The mean SAM produced the average spectral variability from the 36 combinations while the total SAM produced the sum of spectral variability. In this study, we firstly established 1000 random permutations of the original data and then split two-thirds of the data for calibrating the models and used the remainder for evaluating the predictive ability of the models. The study implemented linear regression model to explore the relationship between spectral variability on the Landsat-8 imagery as computed by SAM techniques and tree species diversity as measured by Shannon index. The strength of the relationship was assessed using coefficient of determination (r^2) and *p-value* statistics and the model performance was evaluated using the root mean square error (RMSE).

5.4. Results

5.4.1. Relationship between Landsat-8 spectral variability and tree species diversity

The results of the regression models showed a significant but low relationship between Landsat-8 spectral variability as measured by total SAM and tree species diversity (**Table 20**). In the end of growing season (March) and during the transition to senescence (April), Landsat-8 spectral variability as measured by total SAM explained only 12% of tree species diversity. The relationship between Landsat-8 spectral variability and tree species diversity declined to non-existent with changes in phenology from the end of growing season to dry season (July). However, using mean SAM as a measure of spectral variability increased the relationship between Landsat-8 spectral variability and tree species diversity in the end of the growing season (r^2 of 0.24; $p< 0.05$) while no improvements were observed in other phenological periods (**Table 21**). The results show that remote sensing models based on spectral variability are not affected negatively by mono-species stands when estimating tree species diversity in the savannah woodlands. Contrary to NDVI and Woody Canopy Cover (WCC) models (Madonsela et al. 2018), the removal of mono-species stand as outliers led to the decline in the relationship between Landsat-8 spectral variability and tree species diversity (**Table 22 and 23**).

Table 20 Relationship between Landsat-8 spectral variability as measured by total SAM and tree species diversity as measured by Shannon index (*H'*) prior the removal of mono-species stand as an outlier. Statistics drawn from 1000 bootstrapped iterations.

Predictor variables	Response variable	r^2	CI	P-value	RMSE
SAM_{March}	*H'*	0.12	±0.017	0.014	0.536
SAM_{April}	*H'*	0.12	±0.016	0.015	0.537
SAM_{May}	*H'*	0.06	±0.015	0.023	0.554
SAM_{July}	*H'*	0.01	±0.008	0.026	0.568

Table 21 Relationship between spectral variability as measured by mean SAM and tree species diversity as measured by Shannon index (*H'*) prior the removal of mono-species stand as an outlier. Statistics drawn from 1000 bootstrapped iterations.

Predictor variables	Response variable	r^2	CI	P-value	RMSE
SAM_{March}	*H'*	0.24	±0.014	0.002	0.499
SAM_{April}	*H'*	0.12	±0.016	0.014	0.535
SAM_{May}	*H'*	0.06	±0.015	0.022	0.554
SAM_{July}	*H'*	0.01	±0.008	0.029	0.566

Table 22 Relationship between spectral variability as measured by total SAM and tree species diversity as measured by Shannon index (*H'*) after the removal of mono-species stand as an outlier. Statistics drawn from 1000 bootstrapped iterations.

Predictor variables	Response variable	r^2	CI	P-value	RMSE
SAM_{March}	*H'*	0.09	±0.017	0.018	0.52
SAM_{April}	*H'*	0.10	±0.016	0.016	0.519
SAM_{May}	*H'*	0.05	±0.015	0.022	0.531
SAM_{July}	*H'*	0.00	±0.006	0.029	0.543

Table 23 Relationship between spectral variability as measured by mean SAM and tree species diversity as measured by Shannon index (*H'*) after the removal of mono-species stand as an outlier. Statistics drawn from 1000 bootstrapped iterations.

Predictor variable	Response variable	r^2	CI	P-value	RMSE
SAM_{March}	*H'*	0.22	±0.015	0.003	0.481
SAM_{April}	*H'*	0.10	±0.016	0.017	0.519
SAM_{May}	*H'*	0.05	±0.015	0.023	0.53
SAM_{July}	*H'*	0.00	±0.007	0.03	0.541

5.4.2. Relationship between multi-phenological Landsat-8 spectral variability and tree species diversity

The results of the regression models show that there is very low to no relationship (r^2 of 0.06 and below) between multi-phenological Landsat-8 spectral variability and tree species diversity (**Table 24-25**). The removal of mono-species stand as outliers did not improve the relationship between Landsat-8 spectral variability and tree species diversity (**Table 26-27**).

Table 24 Relationship between spectral variability as measured by total SAM and tree species diversity as measured by Shannon index (*H'*) prior the removal of mono-species stand as an outlier. Statistics drawn from 1000 bootstrapped iterations.

Predictor variables	Response variable	r^2	CI	P-value	RMSE
$SAM_{March, April}$	*H'*	0.02	±0.012	0.026	0.562
$SAM_{March, May}$	*H'*	0.06	±0.016	0.021	0.552
$SAM_{March, July}$	*H'*	0.02	±0.013	0.022	0.561
$SAM_{April, May}$	*H'*	0.01	±0.008	0.030	0.567
$SAM_{April, July}$	*H'*	0.00	±0.007	0.027	0.578
$SAM_{May, July}$	*H'*	0.01	±0.010	0.024	0.564
$SAM_{March, April, May}$	*H'*	0.02	±0.011	0.027	0.565
$SAM_{March, April, July}$	*H'*	0.01	±0.009	0.026	0.566
$SAM_{March, May, July}$	*H'*	0.03	±0.014	0.020	0.561
$SAM_{April, May, July}$	*H'*	0.01	±0.007	0.028	0.568
$SAM_{March, April, May, July}$	*H'*	0.00	±0.007	0.029	0.568

Table 25 Relationship between spectral variability as measured by mean SAM and tree species diversity as measured by Shannon index (*H'*) prior the removal of mono-species stand as an outlier. Statistics drawn from 1000 bootstrapped iterations.

Predictor variables	Response variable	r^2	CI	P-value	RMSE
$SAM_{March, April}$	*H'*	0.02	±0.011	0.026	0.563
$SAM_{March, May}$	*H'*	0.06	±0.016	0.020	0.554
$SAM_{March, July}$	*H'*	0.02	±0.012	0.022	0.564
$SAM_{April, May}$	*H'*	0.01	±0.008	0.030	0.566
$SAM_{April, July}$	*H'*	0.00	±0.006	0.029	0.57
$SAM_{May, July}$	*H'*	0.02	±0.011	0.024	0.564
$SAM_{March, April, May}$	*H'*	0.02	±0.010	0.027	0.565
$SAM_{March, April, July}$	*H'*	0.01	±0.009	0.026	0.566
$SAM_{March, May, July}$	*H'*	0.03	±0.014	0.020	0.563
$SAM_{April, May, July}$	*H'*	0.00	±0.008	0.026	0.567

$\text{SAM}_{\text{March, April, May, July}}$	H'	0.01	±0.010	0.025	0.565

Table 26 Relationship between spectral variability as measured by total SAM and tree species diversity as measured by Shannon index (*H'*) after the removal of mono-species stand as an outlier. Statistics drawn from 1000 bootstrapped iterations.

Predictor variables	Response variable	r^2	CI	P-value	RMSE
$\text{SAM}_{\text{March, April}}$	H'	0.00	±0.003	0.032	0.541
$\text{SAM}_{\text{March, May}}$	H'	0.04	±0.014	0.023	0.533
$\text{SAM}_{\text{March, July}}$	H'	0.01	±0.011	0.021	0.541
$\text{SAM}_{\text{April, May}}$	H'	0.00	±0.004	0.033	0.542
$\text{SAM}_{\text{April, July}}$	H'	0.00	±0.003	0.035	0.543
$\text{SAM}_{\text{May, July}}$	H'	0.01	±0.009	0.026	0.54
$\text{SAM}_{\text{March, April, May}}$	H'	0.00	±0.006	0.029	0.54
$\text{SAM}_{\text{March, April, July}}$	H'	0.00	±0.006	0.028	0.542
$\text{SAM}_{\text{March, May, July}}$	H'	0.01	±0.048	0.022	0.538
$\text{SAM}_{\text{April, May, July}}$	H'	0.00	±0.004	0.029	0.541
$\text{SAM}_{\text{March, April, May, July}}$	H'	0.00	±0.004	0.034	0.543

Table 27 Relationship between spectral variability as measured by mean SAM and tree species diversity as measured by Shannon index (*H'*) after the removal of mono-species stand as an outlier. Statistics drawn from 1000 bootstrapped iterations.

Predictor variables	Response variable	r^2	CI	P-value	RMSE
$\text{SAM}_{\text{March, April}}$	H'	0.01	±0.007	0.031	0.541
$\text{SAM}_{\text{March, May}}$	H'	0.04	±0.014	0.024	0.532
$\text{SAM}_{\text{March, July}}$	H'	0.01	±0.011	0.025	0.539
$\text{SAM}_{\text{April, May}}$	H'	0.00	±0.005	0.032	0.542
$\text{SAM}_{\text{April, July}}$	H'	0.00	±0.003	0.032	0.544
$\text{SAM}_{\text{May, July}}$	H'	0.01	±0.009	0.024	0.541
$\text{SAM}_{\text{March, April, May}}$	H'	0.00	±0.006	0.029	0.542
$\text{SAM}_{\text{March, April, July}}$	H'	0.00	±0.005	0.027	0.543
$\text{SAM}_{\text{March, May, July}}$	H'	0.02	±0.011	0.022	0.54
$\text{SAM}_{\text{April, May, July}}$	H'	0.00	±0.005	0.034	0.542
$\text{SAM}_{\text{March, April, May, July}}$	H'	0.00	±0.006	0.031	0.543

5.4.3. Relationship between multiple predictor variables and tree species diversity

The Landsat-8 spectral variability as measured by mean SAM showed higher relationship with tree species diversity in the end of growing season (March) than other phenological periods (**Table 21**) and therefore it was combined with $NDVI_{March}$ which had also shown a similar feat in the previous research work (Madonsela et al. 2018). The result of the regression model shows that combining Landsat-8 spectral variability with NDVI improves the estimation of tree species diversity (**Table 28**). The combination of Landsat-8 spectral variability with NDVI in a multiple regression model significantly reduced the error of prediction (RMSE of 0.461 from 0.499 and $p< 0.001$ using one-way analysis of variance). The removal of mono-species stand as an outlier further reduced the error of prediction (RMSE of 0.432 from 0.461; $p<0.001$) (**Table 29**).

Table 28 Combining mean SAM and NDVI in a multiple regression model to estimate tree species diversity prior the removal of mono-species stand as an outlier. Statistics drawn from 1000 bootstrapped iterations.

Predictor variables	Response variable	r^2	CI	P-value	RMSE
SAM_{March} and $NDVI_{March}$	*H’*	0.37	±0.015	0.000	0.461

Table 29 Combining mean SAM and NDVI in a multiple regression model to estimate tree species diversity after the removal of mono-species stand as an outlier. Statistics drawn from 1000 bootstrapped iterations.

Predictor variables	Response variable	r^2	CI	P-value	RMSE
SAM_{March} and $NDVI_{March}$	*H’*	0.39	±0.015	0.000	0.432

5.5. Discussion

The results of the study show that Landsat-8 spectral variability, particularly in the end of the growing season, has a significant relationship with tree species diversity in the savannah woodlands. These results are consistent with the observations made by Palmer et al. (2002), Rocchini et al. (2004) Rocchini (2007) and Oldeland et al. (2010) that spectral variability on the image is related to field measured species diversity and therefore may be useful as a proxy for species diversity in light of the SVH. However, the present study further explored the relationship over four different phenological periods and observed a declined in the relationship between Landsat-8 spectral variability as measured by mean SAM and tree species diversity with changes in phenology from the end of growing season to dry season.

In the end of the growing season, the southern African savannah is usually characterized by fully foliated canopies (Grant and Scholes, 2006; Cho et al. 2010) which would change phenologically, albeit at different rates, towards senescence and dry season in the period between May and July (Scholes et al. 2003; Cho et al. 2010; Hill et al. 2010). These phenological changes are accompanied by an increased contribution of background and tree bark to the overall reflectance signal captured by the remote sensing device (Cho et al. 2010). This implies that the observed declined in the relationship between Landsat-8 spectral variability as measured by mean SAM and tree species diversity with changes in phenology can be explained partly by the background and tree bark reflectance attenuating the tree species spectral signal captured by Landsat-8 sensor.

Moreover, the medium spatial resolution of Landsat-8 sensor (30 m) meant that each pixel often captures mixed spectral signal from a number of tree canopies leading to difficulties establishing the spectral variability related to tree species diversity (Nagendra and Rocchini, 2008). In the savannah biome, this difficulty is exacerbated by the fact that tree species co-exists with herbaceous vegetation (Scholes and Archer, 1997; Sankaran et al. 2005) giving rise to mixed spectral signal averaged across different vegetation types. Therefore, it should not be surprising that Landsat-8 spectral variability as measured by mean SAM explained only 24% of tree species diversity in the end of growing season and thereafter its explanatory power declined.

In addition, the aforementioned limitation of Landsat-8 sensor and the challenge posed by the savannah biome may be largely responsible for the failure of multi-phenological Landsat-8 spectral variability to explain tree species diversity. It has been established in previous studies where high spatial resolution data was used that phenological variations between tree species enhance spectral variability thus improving tree species classification (Gilmore et al. 2008; Hill et al. 2010; Madonsela et al. 2017a). The co-occurrence of trees and grasses in the savannah poses a challenge for SVH particularly when using medium resolution sensor such as Landsat-8. The two vegetation lifeforms often display different phenological behaviour related to rainfall events (Chidumayo, 2001; Archibald and Scholes, 2007). These differences in phenology coupled with medium spatial resolution of Landsat-8 and background contribution to the overall reflectance have a potential to introduce spectral variability which is not related to tree species diversity. Hence, multi-phenological Landsat-8 spectral variability could not explain tree species diversity.

Despite the challenges, Landsat-8 spectral variability as measured by mean SAM showed a significant relationship with tree species diversity particularly in the end of the growing season. The observed relationship suggests a possibility of a fine scale tree species diversity resulting in spectrally distinct pixels within 90m X 90m field plots. Ecologists have established that biotic and abiotic processes acting within relatively small areas may cause niche differentiation and fine scale species diversity (Peterson and Reich, 2008; Yessoufou et al. 2013). Consistently, figure 2 shows that semi-variance increase exponentially till it reaches the range and this suggests high species diversity at fine scale. Accordingly, and despite the limitations of Landsat-8 sensor, fine scale species diversity appear to have engendered subtle spectral variability within 3 x 3 Landsat-8 pixels related to field measured tree species diversity and this variability was detectable particularly in the end of growing season.

The results of the study showed SAM, specifically mean SAM, as a reasonable technique for implementing SVH. The application of mean SAM as a measure of spectral variability facilitated the extraction of useful Landsat-8 spectral information in the end of growing season and this spectral information explained 24% of tree species diversity. Meanwhile, spectral information extracted using the total SAM explained only 12% of the tree species diversity in the end of growing season which has been the most optimal phenological period for estimating tree diversity using Landsat-8 data (Madonsela et al. 2017b; Madonsela et al. 2018).

In essence SAM computes pairwise pixel spectral angle using equation 1 and the large spectral angle denotes high spectral variability (Keshava, 2004; Cho et al. 2010). In light of SVH high spectral variability would be indicative of high tree species diversity (Palmer et al. 2002; Rocchini et al. 2010). However, in the savannah, as mentioned above, the co-occurrence of two vegetation types with different phenological behaviour may introduce spectral variability not related to tree species diversity. Therefore, measures of spectral variability should be tailored for this challenge particularly when using medium resolution data such as Landsat-8 where non-canopy features could not be easily masked out. In this regard, mean SAM has shown potential because of its ability to offsets the effect of extreme pixel values as it computes the average spectral variability.

Moreover, the results showed that remote sensing models based on spectral variability are not affected negatively by mono-species stands when estimating tree species diversity in the savannah woodlands. A previous study (Madonsela et al. 2018) established that remote sensing models based on either NDVI or WCC tend to over-estimate species diversity in mono-species stands because of their sensitivity to photosynthetically active vegetation and woody cover. Meanwhile, spectral variability emanates from interspecies phenological and biophysical differences (Nagendra 2001; Hill et al. 2010; Cho et al. 2010; Madonsela et al. 2017a). In all probability mono-species stands would have low spectral variability because they share similar phenological and biophysical attributes. Hence remote sensing models based on spectral variability were able to accurately estimate species diversity in mono-species stands.

In short, mean SAM enabled this study to extract spectral information across Landsat-8 spectrum which is useful for explaining tree species diversity. Meanwhile, NDVI exploit spectral information in the red and NIR regions only. However, the strength of NDVI lies in its sensitivity to vegetation productivity which is known to define variation in species diversity (Parviainen et al. 2010; Witman et al. 2008). The results of the study showed that combining the strength of mean SAM with that of NDVI improves the estimation of tree species diversity. The results highlight the need to develop techniques to exploit Landsat-8 data for purpose of explaining tree species diversity.

5.6. Conclusion

This was the first study to test SAM as a measure of spectral variability in light of SVH and the results of the study showed that the application of mean SAM, in particular, facilitates the extraction of useful spectral information from Landsat-8 image. The strength of mean SAM lies in its ability to offset extreme pixel values. Hence spectral variability measured with mean SAM had higher relationship to tree species diversity (r^2 of 0.24; $p < 0.05$) than that measured with total SAM (r^2 of 0.12; $p < 0.05$) in the end of growing season. Furthermore, the results showed that remote sensing models based on spectral variability are not affected negatively by mono-species stands as it was observed with either NDVI or

WCC model (Madonsela et al. 2018). This renders such models more suitable for estimating tree species diversity in the savannah woodlands. Although spectral variability as measured with mean SAM had a significant relationship with tree species diversity, it was still low and future studies should consider testing mean SAM on high resolution data. In this study mean SAM was tested on Landsat-8 data which has coarse spatial resolution averaging across multiple tree canopies, grass and background. This limitation of Landsat-8 coupled with the challenge posed by co-occurrence of two vegetation types in the savannah were attributed for the failure of multi-phenological spectral variability to explain tree species diversity. Nonetheless, the mean SAM technique itself was useful in extracting spectral information from Landsat-8 image particularly in the end of growing season when savannah woodlands are characterized by fully foliated canopies.

Acknowledgements

The authors would like to express gratitude to the National Research Foundation for funding this research project through the NRF-Professional Development Programme. We would also like to thank Cecilia Masemola, Sbu Gumede, Joseph Dlamini, Martin Sarela and Patrick Ndlovu for their assistance during field work.

References

Archibald, S. and Scholes, R.J., 2007. Leaf green-up in a semi-arid African savanna–separating tree and grass responses to environmental cues. *Journal of Vegetation Science*, *18*(4), pp.583-594.

Bond, W.J. and Loffell, D., 2001. Introduction of giraffe changes acacia distribution in a South African savanna. *African Journal of Ecology*, *39*(3), pp.286-294.

Bond, W.J., Midgley, G.F. and Woodward, F.I., 2003. The importance of low atmospheric CO2 and fire in promoting the spread of grasslands and savannas. *Global Change Biology*, *9*(7), pp.973-982.

Carlson, K.M., Asner, G.P., Hughes, R.F., Ostertag, R. and Martin, R.E., 2007. Hyperspectral remote sensing of canopy biodiversity in Hawaiian lowland rainforests. *Ecosystems*, *10*(4), pp.536-549.

Cho, M.A., Debba, P., Mathieu, Naidoo, L., van Aardt J. and Asner, G.P., 2010. Improving discrimination of savannah tree species through a multi-endmember spectral angle mapper approach: canopy-level analysis. *IEEE Transactions on Geoscience and Remote Sensing Vol.* 48(11), pp. 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.K.J. and Erasmus, B., 2012. Mapping tree species composition in South Africa savannas using an integrated airborne spectral and Lidar system. *Remote Sensing of Environment Vol.* 125, pp. 214-226.

Colwell, R.K., 2009. Biodiversity: concepts, patterns, and measurement. *The Princeton guide to ecology*, pp.257-263.

Druce, D.J., Shannon, G., Page, B.R., Grant, R. and Slotow, R., 2008. Ecological thresholds in the savanna landscape: developing a protocol for monitoring the change in composition and utilisation of large trees. *PloS one*, *3*(12), p.e3979.

du Toit, J.T., Biggs, H.C., Rogers, K.H., 2003. The Kruger Experience: ecology and management of savanna heterogeneity. London: Island Press.

Dumalisile, L., 2009. *Effects of Chromolaena odorata on mammalian biodiversity in Hluhluwe-iMfolozi Park, South Africa* (Doctoral dissertation).

Eckhardt, H.C., Wilgen, B.W. and Biggs, H.C., 2000. Trends in woody vegetation cover in the Kruger National Park, South Africa, between 1940 and 1998. *African Journal of Ecology*, *38*(2), pp.108-115.

Fu, W.J., Jiang, P.K., Zhou, G.M. and Zhao, K.L., 2014. Using Moran's I and GIS to study the spatial pattern of forest litter carbon density in a subtropical region of southeastern China. *Biogeosciences*, *11*(8), pp.2401-2409.

Gilmore, M.S., Wilson, E.H., Barrett, N., Civco, D.L., Prisloe, S., Hurd, J.D. and Chadwick, C., 2008. Integrating multi-temporal spectral and structural information to map wetland vegetation in a lower Connecticut River tidal marsh. *Remote Sensing of Environment Vol*. 112, pp. 4048-4060.

Gould, W., 2000. Remote sensing of vegetation, plant species richness, and regional biodiversity hotspots. *Ecological applications*, *10*(6), pp.1861-1870.

Grant, C.C. and Scholes, M.C., 2006. The importance of nutrient hot-spots in the conservation and management of large wild mammalian herbivores in semi-arid savannas. *Biological Conservation*, *130*(3), pp.426-437.

Gringarten, E. and Deutsch, C.V., 2001. Teacher's aide variogram interpretation and modeling. *Mathematical Geology*, *33*(4), pp.507-534.

He, K.S., Zhang, J. and Zhang, Q., 2009. Linking variability in species composition and MODIS NDVI based on beta diversity measurements. *acta oecologica*, *35*(1), pp.14-21.

Hernández-Stefanoni, J.L., Gallardo-Cruz, J.A., Meave, J.A., Rocchini, D., Bello-Pineda, J. and López-Martínez, J.O., 2012. Modeling α-and β-diversity in a tropical forest from remotely sensed and spatial data. *International journal of applied earth observation and geoinformation*, *19*, pp.359-368.

Hill, R.A., Wilson, A.K., George, M. and Hinsley, S.A., 2010. Mapping tree species in temperate deciduous woodland using time-series multi-spectral data. *Applied Vegetation Science*, *13*(1), pp.86-99.

Jetz, W., Cavender-Bares, J., Pavlick, R., Schimel, D., Davis, F.W., Asner, G.P., Guralnick, R., Kattge, J., Latimer, A.M., Moorcroft, P. and Schaepman, M.E., 2016. Monitoring plant functional diversity from space. *Nature plants*, *2*(3).

Jongman, R.H.G., Ter Braak, C.J. and Van Tongeren, O.F. eds., 1995. *Data analysis in community and landscape ecology*. Cambridge university press.

Kaszta, Ż., Van De Kerchove, R., Ramoelo, A., Cho, M.A., Madonsela, S., Mathieu, R. and Wolff, E., 2016. Seasonal Separation of African Savanna Components Using Worldview-2 Imagery: A Comparison of Pixel-and Object-Based Approaches and Selected Classification Algorithms. *Remote Sensing*, *8*(9), p.763.

Kerr, J.T. and Ostrovsky, M., 2003. From space to species: ecological applications for remote sensing. *Trends in Ecology & Evolution*, *18*(6), pp.299-305.

Keshava, N., 2004. Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries. *IEEE Transactions on Geoscience and remote sensing*, *42*(7), pp.1552-1565.

Ludwig, F., De Kroon, H. and Prins, H.H., 2008. Impacts of savanna trees on forage quality for a large African herbivore. *Oecologia*, *155*(3), pp.487-496.

Madonsela, S., Cho, M.A., Mathieu, R., Mutanga, O., Ramoelo, A., Kaszta, Ż., Van De Kerchove, R. and Wolff, E., 2017. Multi-phenology WorldView-2 imagery improves remote sensing of savannah tree species. *International Journal of Applied Earth Observation and Geoinformation*, *58*, pp.65-73.

Madonsela, S., Cho, M.A., Ramoelo, A. and Mutanga, O., 2017. Remote sensing of species diversity using Landsat 8 spectral variables. *ISPRS Journal of Photogrammetry and Remote Sensing*, *133*, pp.116-127.

Madonsela, S., Cho, M.A., Ramoelo, A., Mutanga, O. and Naidoo, L., 2018. Estimating tree species diversity in the savannah using NDVI and woody canopy cover. *International Journal of Applied Earth Observation and Geoinformation*, *66*, pp.106-115.

Madubansi, M. and Shackleton, C.M., 2006. Changing energy profiles and consumption patterns following electrification in five rural villages, South Africa. *Energy Policy*, *34*(18), pp.4081-4092.

Makhado, R.A., Mapaure, I., Potgieter, M.J., Luus-Powell, W.J. and Saidi, A.T., 2014. Factors influencing the adaptation and distribution of Colophospermum mopane in southern Africa's mopane savannas-A review. *Bothalia-African Biodiversity & Conservation*, *44*(1), pp.1-9.

Matsika, R., Erasmus, B.F. and Twine, W.C., 2013. A tale of two villages: assessing the dynamics of fuelwood supply in communal landscapes in South Africa. *Environmental Conservation*, *40*(1), pp.71-83.

McKane, R.B., Johnson, L.C., Shaver, G.R., Nadelhoffer, K.J., Rastetter, E.B., Fry, B., Giblin, A.E., Kielland, K., Kwiatkowski, B.L., Laundre, J.A. and Murray, G., 2002. Resource-based niches provide a basis for plant species diversity and dominance in arctic tundra. *Nature*, *415*(6867), pp.68-71.

Menge, B.A. and Sutherland, J.P., 1976. Species diversity gradients: synthesis of the roles of predation, competition, and temporal heterogeneity. *The American Naturalist*, *110*(973), pp.351-369.

Morris, E.K., Caruso, T., Buscot, F., Fischer, M., Hancock, C., Maier, T.S., Meiners, T., Müller, C., Obermaier, E., Prati, D. and Socher, S.A., 2014. Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories. *Ecology and evolution*, *4*(18), pp.3514-3524.

Mutanga, O., Van Aardt, J. and Kumar, L., 2009. Imaging spectroscopy (hyperspectral remote sensing) in southern Africa: an overview. *South African Journal of Science*, *105*(5-6), pp.193-198.

Nagendra, H., 2001. Using remote sensing to assess biodiversity. *International Journal of Remote Rensing*, *22*(12), pp.2377-2400.

Nagendra, H. and Rocchini, D., 2008. High resolution satellite imagery for tropical biodiversity studies: the devil is in the detail. *Biodiversity and Conservation*, *17*(14), p.3431.

Naidoo, L., Mathieu, R., Main, R., Kleynhans, W., Wessels, K., Asner, G. and Leblon, B., 2015. Savannah woody structure modelling and mapping using multi-frequency (X-, C-and L-band) Synthetic Aperture Radar data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *105*, pp.234-250.

Oindo, B.O. and Skidmore, A.K., 2002. Interannual variability of NDVI and species richness in Kenya. *International journal of remote sensing*, *23*(2), pp.285-298.

Oldeland, J., Wesuls, D., Rocchini, D., Schmidt, M. and Jürgens, N., 2010. Does using species abundance data improve estimates of species diversity from remotely sensed spectral heterogeneity?. *Ecological Indicators*, *10*(2), pp.390-396.

Palmer, M.W., Earls, P.G., Hoagland, B.W., White, P.S. and Wohlgemuth, T., 2002. Quantitative tools for perfecting species lists. *Environmetrics*, *13*(2), pp.121-137.

Parviainen, M., Luoto, M. and Heikkinen, R.K., 2010. NDVI-based productivity and heterogeneity as indicators of plant-species richness in boreal landscapes. *Boreal environment research*, *15*(3).

Pau, S., Gillespie, T.W. and Wolkovich, E.M., 2012. Dissecting NDVI–species richness relationships in Hawaiian dry forests. *Journal of Biogeography*, *39*(9), pp.1678-1686.

Pellegrini, A.F., Socolar, J.B., Elsen, P.R. and Giam, X., 2016. Trade-offs between savanna woody plant diversity and carbon storage in the Brazilian Cerrado. *Global change biology*, *22*(10), pp.3373-3382.

Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C. and Coops, N.C., 2013. Essential biodiversity variables. *Science*, *339*(6117), pp.277-278.

Peterson, D.W. and Reich, P.B., 2008. Fire frequency and tree canopy structure influence plant species diversity in a forest-grassland ecotone. *Plant Ecology*, *194*(1), pp.5-16.

Richter, R. and Schläpfer, D., 2012. Atmospheric/Topographic Correction for Satellite Imagery (ATCOR-2/3 User Guide, Version 8.2 BETA). *German Aerospace Center, Remote Sensing Data Center: Wessling, Germany.*

Rocchini, D., Chiarucci, A. and Loiselle, S.A., 2004. Testing the spectral variation hypothesis by using satellite multispectral images. *Acta Oecologica*, *26*(2), pp.117-120.

Rocchini, D., 2007. Effects of spatial and spectral resolution in estimating ecosystem α-diversity by satellite imagery. *Remote Sensing of Environment*, *111*(4), pp.423-434

Rocchini, D., Balkenhol, N., Carter, G.A., Foody, G.M., Gillespie, T.W., He, K.S., Kark, S., Levin, N., Lucas, K., Luoto, M. and Nagendra, H., 2010. Remotely sensed spectral heterogeneity as a proxy of species diversity: recent advances and open challenges. *Ecological Informatics*, *5*(5), pp.318-329.

Rocchini, D., Boyd, D.S., Féret, J.B., Foody, G.M., He, K.S., Lausch, A., Nagendra, H., Wegmann, M. and Pettorelli, N., 2016. Satellite remote sensing to monitor species diversity: potential and pitfalls. *Remote Sensing in Ecology and Conservation*, *2*(1), pp.25-36.

Scholes, R.J. and Archer, S.R., 1997. Tree-grass interactions in savannas. *Annual review of Ecology and Systematics*, *28*(1), pp.517-544.

Scholes, R.J., Bond, W.J. and Eckhardt, H.C., 2003. *Vegetation dynamics in the Kruger ecosystem* (pp. 242-262). The Kruger Experience. Island Press.

Shackleton, C.M., 2000. Comparison of plant diversity in protected and communal lands in the Bushbuckridge lowveld savanna, South Africa. *Biological Conservation*, *94*(3), pp.273-285.

Shackleton, C.M., Guthrie, G. and Main, R., 2005. Estimating the potential role of commercial over-harvesting in resource viability: a case study of five useful tree species in South Africa. *Land Degradation & Development*, *16*(3), pp.273-286.

Shackleton, C.M., Shackleton, S.E., Buiten, E. and Bird, N., 2007. The importance of dry woodlands and forests in rural livelihoods and poverty alleviation in South Africa. *Forest policy and economics*, *9*(5), pp.558-577.

Shannon, C.E. and Weaver, W., 1998. *The mathematical theory of communication*. University of Illinois press.

Stickler, C.M. and Southworth, J., 2008. Application of multi-scale spatial and spectral analysis for predicting primate occurrence and habitat associations in Kibale National Park, Uganda. *Remote sensing of environment*, *112*(5), pp.2170-2186.

Thenkabail, P.S., Hall, J., Lin, T., Ashton, M.S., Harris, D. and Enclona, E.A., 2003. Detecting floristic structure and pattern across topographic and moisture gradients in a mixed species Central African forest using IKONOS and Landsat-7 ETM+ images. *International Journal of Applied Earth Observation and Geoinformation*, *4*(3), pp.255-270.

Treydte, A.C., Heitkönig, I.M., Prins, H.H. and Ludwig, F., 2007. Trees improve grass quality for herbivores in African savannas. *Perspectives in Plant Ecology, Evolution and Systematics*, *8*(4), pp.197-205.

Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E. and Steininger, M., 2003. Remote sensing for biodiversity science and conservation. *Trends in ecology & evolution*, *18*(6), pp.306-314.

Witman, J.D., Cusson, M., Archambault, P., Pershing, A.J. and Mieszkowska, N., 2008. The relation between productivity and species diversity in temperate–arctic marine ecosystems. *Ecology*, *89*(sp11).

Yessoufou, K., Davies, T.J., Maurin, O., Kuzmina, M., Schaefer, H., Bank, M. and Savolainen, V., 2013. Large herbivores favour species diversity but have mixed impacts on phylogenetic community structure in an African savanna ecosystem. *Journal of Ecology*, *101*(3), pp.614-625.

Chapter 6: Synthesis

Exploring the relationship between spectral reflectance and tree species diversity in the savannah woodland belt

6.1. Introduction

Why do we need to explore the relationship between spectral reflectance and tree species diversity in the savannah woodland belt? Tree diversity serves many ecological functions in the savannah, e.g. providing habitats and nesting sites to diverse avifaunal species (Dean et al. 1999; Seymour and Dean, 2010); facilitating grass growth and improving grass quality beneath their canopies (Ludwig et al. 2004; Treydte et al. 2007); serving as food resources to many browsing faunal species (Hempson et al. 2015) and also acting as a safety net against poverty in the neighbouring communities delivering goods such as fuelwood, medicinal and non-timber forest products (Shackleton et al. 2007; Matsika et al. 2012).

However, the diversity, abundance and distribution of savannah tree species are impacted by disturbances e.g. the effect of elephants in protected areas (Druce et al. 2008), harvesting for fuelwood (Madubansi and Shackleton, 2006; Matsika et al., 2012) and land use conversion (Schlesinger et al. 2015). As such, South Africa's National Parks have developed a Threshold of Potential Concerns (TPCs) which serves as a monitoring system to detect changes that may negatively impact on key elements of biodiversity (Gillson and Duffin, 2007; Druce et al. 2008). Like other monitoring systems elsewhere, the success of TPCs will depend on the availability of up-to-date and spatially-explicit assessments of species richness and distribution at a regional scale (Turner et al. 2003). Presently, the African savannahs are confronted by a challenge of the absence of large scale spatial information on tree species distribution upon which management decision can be based (Asner et al. 2009). Spaceborne remote sensing meets this requirement as it collects data over a large geographic area on a regular interval and at varying levels of spatial details (Kerr and Ostrovsky, 2003; Jetz et al. 2016). Hence this study was aimed at exploring the relationship between spectral reflectance and tree species diversity in the savannah woodland belt. The major contributions of this study to remote sensing application in biodiversity research are discussed in the subsequent sections.

6.2. Remotely sensed productivity indices capture the long established productivity-diversity relationship

Ecologists have long established the relationship between ecosystem productivity and species diversity to be positive, negative or humped shaped (Waide et al. 1999; Kirkman et al. 2001; Bai et al. 2007; Witman et al. 2008). In fact the application of NDVI for estimating species diversity is based on the understanding that NDVI is sensitive to the amount of available energy detectable as primary productivity which defines the variation in plant species diversity. However, few studies have tested the application of remote sensing for biodiversity estimation in southern African savannah at a regional level. For instance Oldeland et al. (2010) and Cho et al. (2012) applied remote sensing for tree species diversity estimation in the savannah, however, these studies were too local to facilitate the development of remote sensing model that can be used to estimate tree species diversity across the savannah woodland belt.

The present study comprehensively explored the relationship between remotely sensed productivity indices and tree species diversity and in keeping with ecological literature, the study showed that remotely sensed productivity indices have a positive relationship with tree species diversity in the savannah woodlands (Table 30) (Madonsela et al. 2017b; Madonsela et al. 2018). The implication of these results is that remotely sensed productivity indices can be used to estimate tree species diversity in the savannah woodlands. Results from the end of growing season in particular, showed that factorial model based on NDVI and woody canopy cover, which is a proxy for woodland productivity, predicts tree species diversity in a manner that is consistent with our knowledge of tree diversity in the savannah woodlands (Figure 19b).

However, productivity models tend to over-predict tree species diversity in mono-species stands such as mopane woodland which is dominantly occupied by *Colophospermum mopane*. The immediate conclusion from these observations is that remote sensing models based on vegetation productivity are more suitable for estimating tree species diversity in species diverse woodlands. Therefore there is a necessity for the development of remote sensing models that are sensitive to mono-species stands given that mopane woodland occupies a large portion of the southern African savannah. The subsequent finding addresses this point.

Table 30 Relationship observed between three common measures of tree species diversity (*H'*, D_2 and *S*) and remotely sensed productivity indices. All computations were drawn from 1000 bootstrapped iterations.

Diversity index	Spectral variable	Average R^2	Confidence interval 95%	*P-value*	RMSE
H'	NDVI (Mean)	0.29	±0.014	0.0005	0.4861
	EVI (Mean)	0.29	±0.014	0.0008	0.4869
	SRI (Mean)	0.26	±0.005	0.0006	0.4985
	SAVI (Mean)	0.29	±0.014	0.0007	0.4894
D_2	NDVI (Mean)	0.29	±0.012	0.0003	1.8048
	EVI (Mean)	0.29	±0.013	0.0003	1.8203
	SRI (Mean)	0.27	±0.012	0.0005	1.8545
	SAVI (Mean)	0.29	±0.012	0.0003	1.8250
S	NDVI (Mean)	0.23	±0.013	0.0020	3.4913
	EVI (Mean)	0.21	±0.013	0.0027	3.5516
	SRI (Mean)	0.23	±0.015	0.0024	3.5513
	SAVI (Mean)	0.23	±0.013	0.0073	3.5263

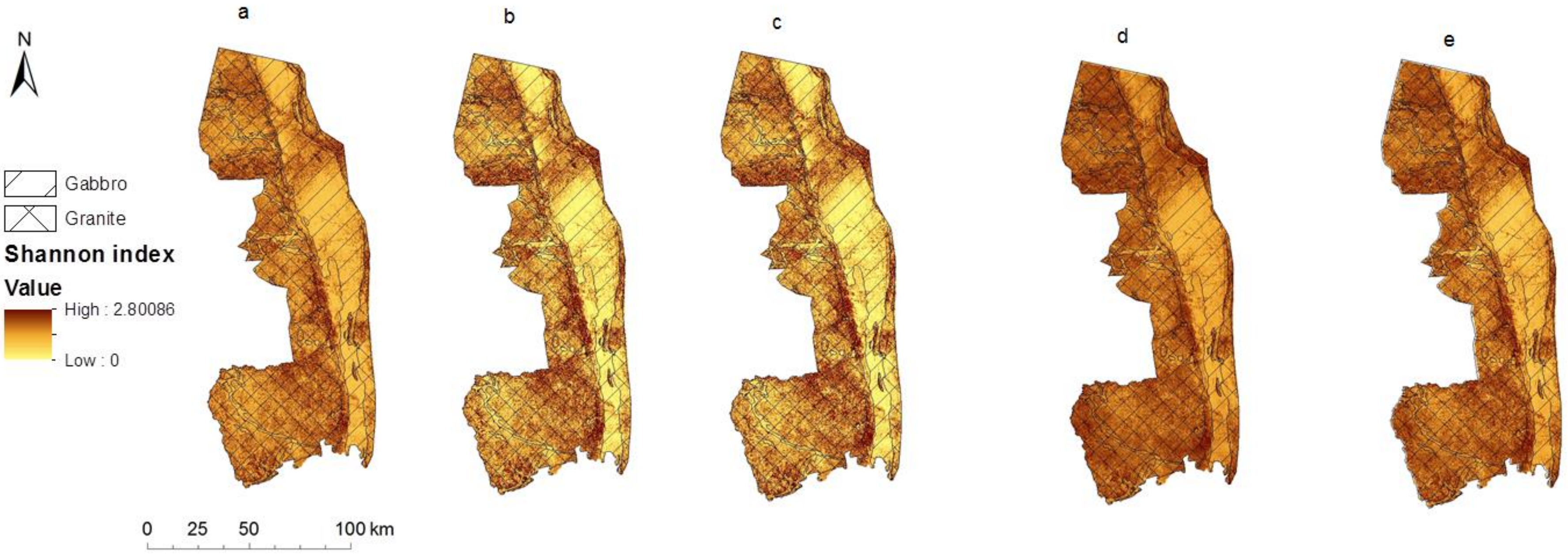

Figure 19 Tree species diversity from the best (a) woody canopy cover (WCC) model, (b) factorial model involving $NDVI_{March}$ and WCC (c) factorial model involving $NDVI_{April}$ and WCC (d) factorial model involving $NDVI_{May}$ and WCC and (e) factorial model involving $NDVI_{July}$ and WCC.

6.3. Remote sensing captures spectral variability related to tree species diversity

Studies have established that different plant species display different spectral responses due to the unique biochemical and biophysical properties between them (Nagendra 2001; Nagendra and Rocchini, 2008; Asner and Marin 2009; Cho et al. 2010) and these are the scientific basis of Spectral Variation Hypothesis (SVH) (Zhao et al. 2016). SVH states that spectral variability on the image emanates from spatial heterogeneity of the environment which by default has high species diversity due to the high availability of niches (Rocchini, 2007; Oldeland et al. 2010; Hernández-Stefanoni et al. 2012). Therefore, spectral variability on the image should be considered as a proxy of species diversity. This study developed an innovative method of applying Spectral Angle Mapper as a measure of spectral variability in order to extract essential spectral information from the Landsat-8 data and use it to explain tree species diversity. Consistent with SVH, SAM computes intra-pixel variability within 3x3 Landsat-8 pixels and relates it to tree species diversity observed within 90m x 90m field plot. The results showed that, at the end of the growing season, mean SAM has a higher relationship with tree species diversity compared to total SAM (r^2 of 0.24; $p< 0.05$ and r^2 of 0.12; $p< 0.05$ respectively) and this was attributed to its ability to offset extreme pixel values as it computes the average spectral variability.

More importantly, mean SAM was not negatively affected by the mono-species stands and that was because it is based on spectral variability. In all likelihood mono-species stands will have low spectral variability because of absence of diversity. In essence remote sensing captures spectral variability related to the diversity of tree species in the savannah woodlands. This implies that remote sensing model based on spectral variability would be more suitable for modelling tree species diversity in the savannah woodlands particularly in the end of growing season. The development of a remote sensing model based on spectral variability which is capable of modelling a mono-species stand is the major contribution of the study given that the previous models based on productivity over-estimated tree diversity in mono-species stands. It would be interesting to test the SAM techniques on high resolution data which captures subtle biochemical and biophysical differences between tree canopies and also verify the results of mean SAM across different phenological stages not tested in this study.

6.4. Phenology affects the ability of remote sensing to capture spectral variability related to tree species diversity

Studies have established that the phenological variation between tree species enhances spectral variability. This occurs when tree species are at different phenological stages on the same image or when phenological changes occur at different rates between images captured across the growing season (Gilmore et al. 2008; Hill et al. 2010; Madonsela et al. 2017a). In light of the above and considering the assertion made by SVH with regards to spectral variability, targeting particular phenological stage at which the remotely sensed data is acquired has consequential effect on the ability of remote sensing model to estimate tree species diversity especially in the savannah woodlands. This study showed that mean SAM has a high relationship with tree species diversity in the end of the growing season and this relationship declined with changes in phenology towards the senescence period and dry season (Table 31). The same pattern of relationship was observed between NDVI and tree species diversity (Madonsela et al. 2018). Nonetheless, the latter was expected given that the amount of photosynthetically active vegetation tends to decline with changes in phenology towards the dry season. In the case of mean SAM, the decline in the relationship with changes in phenology from the end of growing season to transition to senescence (April date) contradicts the findings of Madonsela et al. (2017a) who identified the transition to senescence as the optimal period to benefit from high spectral variability between tree species in the savannah.

Table 31 Relationship between spectral variability as measured by mean SAM and tree species diversity as measured by Shannon index (*H'*). Statistics drawn from 1000 bootstrapped iterations.

Predictor variables	Response variable	r^2	CI	P-value	RMSE
SAM_{March}	*H'*	0.24	±0.014	0.002	0.499
SAM_{April}	*H'*	0.12	±0.016	0.014	0.535
SAM_{May}	*H'*	0.06	±0.015	0.022	0.554
SAM_{July}	*H'*	0.01	±0.008	0.029	0.566

However, the present study and that of Madonsela et al. (2017a) used two different remotely sensed data with the latter benefiting from high spatial resolution WorldView-2 and application specific bands. The present study used medium resolution Landsat-8 with coarse spectral bands. In the savannah, the end of the growing season is associated with fully foliated canopies (Grant and Scholes, 2006; Cho et al. 2010) which would change phenologically, albeit at a different rate, towards senescence and leaf shedding in the period between May and July (Scholes et al. 2003; Cho et al. 2010; Hill et al. 2010). These changes are accompanied by an increased contribution of background and tree bark to the overall reflectance signal captured by the remote sensing sensor (Cho et al. 2010). This implies that the decline in the relationship between mean SAM and tree species diversity can be attributed to background and tree bark effects attenuating the tree species reflectance signal captured by Landsat-8. Therefore, given that Landsat-8 already captures mixed spectral signal, it is advisable to collect it in the end of growing season when canopies are still fully foliated to minimize the effect of background contribution and increase chances of extracting spectral variability related to tree species diversity.

6.5. Radiative transfer modelling envisages the future application of remote sensing in biodiversity research

The aforementioned findings emanate from empirical models which have been criticised as being site, season and sensor specific (Colombo et al. 2003; Cho, 2007; Atzberger et al. 2011). Meanwhile the radiative transfer modelling offers an opportunity to develop stable models for estimating tree species diversity. Radiative transfer models are based on physical processes describing the transfer and interaction of radiation with canopy components at leaf and canopy levels and therefore they establish relationships between vegetation biochemical and biophysical parameters and the canopy reflectance (Schlerf and Atzberger, 2006; Kötz et al. 2004; Darvishzadeh et al. 2008). The present study inverted PROSAIL and INFORM models on Landsat-8 image to retrieve biochemical and biophysical parameters such as EWT, *Cab*, *Car*, LAI, dry matter and *Cbrown* and used these parameters to explain tree species diversity in the savannah woodlands. In particular the EWT, LAI, dry matter and *Cbrown* showed a significant relationship with tree species diversity (Table 32) and this was improved when these parameters were transformed into PCs.

Table 32 Relationship observed between *H'* and RTM derivatives. RTM derivatives were generated through the inversion of PROSAIL model on Landsat-8 data from different dates. All statistics below were drawn from 1000 bootstrapped iterations.

Date	Diversity index	NDVI and RTM derivatives	Average R^2	Confidence interval 95%	*P-value*	RMSE
March	*H'*	NDVI	0.33	±0.019	0.000	0.454
		LAI	0.19	±0.014	0.000	0.492
		EWT	0.33	±0.017	0.000	0.448

		Cm	0.23	±0.018	0.003	0.479
		Cbrown	0.19	±0.015	0.005	0.491
		Cab	0.00	±0.000	0.045	0.543
		Car	0.00	±0.005	0.033	0.543
April	*H’*	NDVI	0.27	±0.022	0.000	0.464
		LAI	0.12	±0.015	0.015	0.515
		EWT	0.27	±0.016	0.001	0.467
		Cm	0.19	±0.015	0.006	0.494
		Cbrown	0.10	±0.013	0.020	0.519
		Cab	0.00	±0.002	0.031	0.544
		Car	0.00	±0.004	0.032	0.542
May	*H’*	NDVI	0.19	±0.019	0.004	0.492
		LAI	0.04	±0.014	0.024	0.535
		EWT	0.17	±0.013	0.008	0.498
		Cm	0.03	±0.014	0.022	0.535
		Cbrown	0.00	±0.003	0.035	0.541
		Cab	0.00	±0.001	0.021	0.543
		Car	0.13	±0.018	0.012	0.509
July	*H’*	NDVI	0.13	±0.019	0.011	0.513
		LAI	0.00	±0.006	0.031	0.541
		EWT	0.00	±0.000	0.043	0.542
		Cm	0.00	±0.001	0.028	0.547
		Cbrown	0.00	±0.000	NaN	0.544
		Cab	0.03	±0.013	0.023	0.535
		Car	0.00	±0.000	NaN	0.543

Although the relationship between tree species diversity and EWT, LAI, dry matter and *Cbrown* was not higher than that observed with NDVI, its existence suggest that radiative transfer modelling has a potential for retrieving useful biophysical and biochemical parameters from remotely sensed data. The significant relationship, particularly at the end of the growing season, between tree species diversity and these parameters retrieved through inversion of PROSAIL and INFROM on medium resolution Landsa-8 image show the potential of derivatives from radiative transfer modelling. The biochemical data captured by Landsat-8 sensor is largely inadequate since the sensor has a coarse spectral band which renders it insufficient for detecting subtle details on vegetation canopy structure and biochemistry (Carlson et al. 2006; Mutanga et al. 2009). For instance Cho et al. (2014) observed that PROSAIL inversion on high resolution SPOT 6 image produced accurate LAI estimates (RMSE of 0.85) than inversion on MODIS image with 250m spatial resolution (RMSE of 1.26).

6.6. Conclusion

This study made major contributions to the application of remote sensing in biodiversity research in the southern African savannah. These include the finding, which is supported by other studies in different ecosystems that i) remotely sensed productivity indices capture the long established productivity-diversity relationship and therefore can be used to estimate tree species diversity in the savannah woodlands; ii) remote sensing captures spectral variability related to the diversity of tree species in the savannah woodlands with implication that remote sensing model based on spectral variability would be more suited for modelling tree species diversity in the savannah woodlands particularly in the end of growing season; iii) tree phenology affects the ability of remote sensing to capture spectral variability related to tree species diversity and therefore the implementation of SVH should target particular phenological period to increase the chances of extracting spectral variability related to tree species diversity and iv) radiative transfer modelling has a potential for retrieving useful information on the biophysical and biochemical parameters that can be used for explaining tree species diversity in the savannah woodlands.

However, the estimation of tree species diversity in the savannah woodlands using remotely sensed is not without challenges. As mentioned above, remotely sensed productivity models tend to over-predict tree species diversity in mono-species stands. Moreover, the present study only considered tree species diversity whilst the savannah is characterized by the co-existence of trees and grass. Therefore, the spectral signal captured by Landsat-8 sensor relates to the total vegetation cover and this is assumed to have contributed to prediction errors observed in the study. Areas with high ratio of grass cover would be susceptible to over prediction.

In addition, Landsat-8 sensor has coarse spatial and spectral resolutions which make it less sensitive to subtle differences between tree species (Carlson et al. 2006; Mutanga et al. 2009). The coarse resolution of Landsat-8 data coupled by the fact that in the savannah tree species co-exists with herbaceous vegetation give rise to mixed spectral signal averaged across different vegetation types. Hence Landsat-8 spectral variability as measured by mean SAM explained only 24% of tree species diversity in the end of growing season and thereafter its explanatory power declined.

Despite these challenges, the study demonstrated the applicability of remote sensing in biodiversity estimation in the savannah woodlands. The availability of new satellite sensors such as Sentinel-2 with high spectral and spatial resolution presents an opportunity to improve the application of remote sensing in biodiversity research. The study recommends further research on testing radiative transfer modelling for retrieval of biochemical and biophysical parameters should consider inverting RTMs on high resolution images capable of detecting fine details on the canopy structure and biochemistry. Moreover, SAM was used for first time as a multivariate technique and mean SAM results were satisfactory particularly in the end of growing season. Future research should test this technique on high spectral resolution data where reflectance signals could be associated with tree canopies. It will also be interesting to test the application of integrated modelling involving remote sensing variables and environmental variables for estimating tree species diversity at regional scale. Integrated modelling involving remote sensing variables and environmental variables have improved the prediction of invasive species in other studies (Malahlela et al., 2015).

References

Ali, A.M., Darvishzadeh, R., Skidmore, A.K. and van Duren, I., 2016. Effects of canopy structural variables on retrieval of leaf dry matter content and specific leaf area from remotely sensed data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*(2), pp.898-909.

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

Atzberger, C., Darvishzadeh, R., Immitzer, M., Schlerf, M., Skidmore, A. and Le Maire, G., 2015. Comparative analysis of different retrieval methods for mapping grassland leaf area index using airborne imaging spectroscopy. *International Journal of Applied Earth Observation and Geoinformation*, *43*, pp.19-31.

Atzberger, C., Richter, K., Vuolo, F., Darvishzadeh, R. and Schlerf, M., 2011, October. Why confining to vegetation indices? Exploiting the potential of improved spectral observations using radiative transfer models. In *Proc. SPIE*(Vol. 8174, p. 81740Q).

Bai, Y., Wu, J., Pan, Q., Huang, J., Wang, Q., Li, F., Buyantuyev, A. and Han, X., 2007. Positive linear relationship between productivity and diversity: evidence from the Eurasian Steppe. *Journal of Applied Ecology*, *44*(5), pp.1023-1034.

Carlson, K.M., Asner, G.P., Hughes, R.F., Ostertag, R. and Martin, R.E., 2007. Hyperspectral remote sensing of canopy biodiversity in Hawaiian lowland rainforests. *Ecosystems*, *10*(4), pp.536-549.

Ceccato, P., Flasse, S., Tarantola, S., Jacquemoud, S. and Grégoire, J.M., 2001. Detecting vegetation leaf water content using reflectance in the optical domain. *Remote sensing of environment*, *77*(1), pp.22-33.

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

Cho, M.A., 2007. *Hyperspectral remote sensing of biochemical and biophysical parameters: the derivate red-edge" double-peak feature", a nuisance or an opportunity?*.

Colombo, R., Bellingeri, D., Fasolini, D. and Marino, C.M., 2003. Retrieval of leaf area index in different vegetation types using high resolution satellite data. *Remote Sensing of Environment*, *86*(1), pp.120-131.

Dean, W.R.J., Milton, S.J. and Jeltsch, F., 1999. Large trees, fertile islands, and birds in arid savanna. *Journal of Arid Environments*, *41*(1), pp.61-78.

Druce, D.J., Shannon, G., Page, B.R., Grant, R. and Slotow, R., 2008. Ecological thresholds in the savanna landscape: developing a protocol for monitoring the change in composition and utilisation of large trees. *PloS one*, *3*(12), p.e3979.

du Toit, J.T., Biggs, H.C., Rogers, K.H., 2003. The Kruger Experience: ecology and management of savanna heterogeneity. London: Island Press.

Dumalisile, L., 2009. *Effects of Chromolaena odorata on mammalian biodiversity in Hluhluwe-iMfolozi Park, South Africa* (Doctoral dissertation).

Fairbanks, D.H. and McGwire, K.C., 2004. Patterns of floristic richness in vegetation communities of California: regional scale analysis with multi-temporal NDVI. *Global Ecology and Biogeography*, *13*(3), pp.221-235.

Foody, G.M. and Cutler, M.E., 2006. Mapping the species richness and composition of tropical forests from remotely sensed data with neural networks. *Ecological modelling*, *195*(1), pp.37-42.

Gillson, L. and Duffin, K.I., 2007. Thresholds of potential concern as benchmarks in the management of African savannahs. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *362*(1478), pp.309-319.

Gilmore, M.S., Wilson, E.H., Barrett, N., Civco, D.L., Prisloe, S., Hurd, J.D. and Chadwick, C., 2008. Integrating multi-temporal spectral and structural information to map wetland vegetation in a lower Connecticut River tidal marsh. *Remote Sensing of Environment Vol*. 112, pp. 4048-4060.

Gould, W., 2000. Remote sensing of vegetation, plant species richness, and regional biodiversity hotspots. *Ecological applications*, *10*(6), pp.1861-1870.

He, K.S., Zhang, J. and Zhang, Q., 2009. Linking variability in species composition and MODIS NDVI based on beta diversity measurements. *acta oecologica*, *35*(1), pp.14-21.

Hernández-Stefanoni, J.L., Gallardo-Cruz, J.A., Meave, J.A., Rocchini, D., Bello-Pineda, J. and López-Martínez, J.O., 2012. Modeling α-and β-diversity in a tropical forest from remotely sensed and spatial data. *International journal of applied earth observation and geoinformation*, *19*, pp.359-368.

Hill, R.A., Wilson, A.K., George, M. and Hinsley, S.A., 2010. Mapping tree species in temperate deciduous woodland using time-series multi-spectral data. *Applied Vegetation Science*, *13*(1), pp.86-99.

Jetz, W., Cavender-Bares, J., Pavlick, R., Schimel, D., Davis, F.W., Asner, G.P., Guralnick, R., Kattge, J., Latimer, A.M., Moorcroft, P. and Schaepman, M.E., 2016. Monitoring plant functional diversity from space. *Nature plants*, *2*(3).
Jongman, R.H., Ter Braak, C.J. and Van Tongeren, O.F. eds., 1995. *Data analysis in community and landscape ecology*. Cambridge University press.

Kerr, J.T. and Ostrovsky, M., 2003. From space to species: ecological applications for remote sensing. *Trends in Ecology & Evolution*, *18*(6), pp.299-305.

Kirkman, L.K., Mitchell, R.J., Helton, R.C. and Drew, M.B., 2001. Productivity and species richness across an environmental gradient in a fire-dependent ecosystem. *American Journal of Botany*, *88*(11), pp.2119-2128.

Ludwig, F., De Kroon, H., Berendse, F. and Prins, H.H., 2004. The influence of savanna trees on nutrient, water and light availability and the understorey vegetation. *Plant Ecology*, *170*(1), pp.93-105.

Madonsela, S., Cho, M.A., Mathieu, R., Mutanga, O., Ramoelo, A., Kaszta, Ż., Van De Kerchove, R. and Wolff, E., 2017a. Multi-phenology WorldView-2 imagery improves remote sensing of savannah tree species. *International Journal of Applied Earth Observation and Geoinformation*, *58*, pp.65-73.

Madonsela, S., Cho, M.A., Ramoelo, A. and Mutanga, O., 2017b. Remote sensing of species diversity using Landsat 8 spectral variables. *ISPRS Journal of Photogrammetry and Remote Sensing*, *133*, pp.116-127.

Madonsela, S., Cho, M.A., Ramoelo, A., Mutanga, O. and Naidoo, L., 2018. Estimating tree species diversity in the savannah using NDVI and woody canopy cover. *International Journal of Applied Earth Observation and Geoinformation*, *66*, pp.106-115.

Madubansi, M. and Shackleton, C.M., 2006. Changing energy profiles and consumption patterns following electrification in five rural villages, South Africa. *Energy Policy*, *34*(18), pp.4081-4092.

Malahlela, O.E., Cho, M.A. and Mutanga, O., 2015. Mapping the occurrence of Chromolaena odorata (L.) in subtropical forest gaps using environmental and remote sensing data. *Biological Invasions*, *17*(7), pp.2027-2042.

Matsika, R., Erasmus, B.F. and Twine, W.C., 2013. A tale of two villages: assessing the dynamics of fuelwood supply in communal landscapes in South Africa. *Environmental Conservation*, *40*(1), pp.71-83.

Mudongo, E., Fynn, R., Bonyongo, M.C., 2016. Influence of fire on woody vegetation density, cover and structure at Tiisa Kalahari Ranch in western Botswana. *Grassland Science Vol*. 62 (1), pp. 3-11.

Mutanga, O., Van Aardt, J. and Kumar, L., 2009. Imaging spectroscopy (hyperspectral remote sensing) in southern Africa: an overview. *South African Journal of Science*, *105*(5-6), pp.193-198.

Myneni, R.B., Hoffman, S., Knyazikhin, Y., Privette, J.L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G.R. and Lotsch, A., 2002. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote sensing of environment*, *83*(1), pp.214-231.

Nacoulma, B., Marie, I., Schumann, K.,Traore, S., Bernhardt-Romermann, M., Hahn, K., Wittig, R. and Thiombiano, A., 2011. Impacts of land-use on West African savanna vegetation: a comparison between protected and communal area in Burkina Faso. *Biodiversity and Conservation Vol*. 20 (14), pp. 3341-3362.

Nagendra, H., 2002. Opposite trends in response for the Shannon and Simpson indices of landscape diversity. *Applied Geography*, *22*(2), pp.175-186.

Nagendra, H., Rocchini, D., Ghate, R., Sharma, B. and Pareeth, S., 2010. Assessing plant diversity in a dry tropical forest: Comparing the utility of Landsat and IKONOS satellite images. *Remote Sensing*, *2*(2), pp.478-496.

Oindo, B.O. and Skidmore, A.K., 2002. Interannual variability of NDVI and species richness in Kenya. *International journal of remote sensing*, *23*(2), pp.285-298.

Oldeland, J., Wesuls, D., Rocchini, D., Schmidt, M. and Jürgens, N., 2010. Does using species abundance data improve estimates of species diversity from remotely sensed spectral heterogeneity?. *Ecological Indicators*, *10*(2), pp.390-396.

Palmer, M.W., Earls, P.G., Hoagland, B.W., White, P.S. and Wohlgemuth, T., 2002. Quantitative tools for perfecting species lists. *Environmetrics*, *13*(2), pp.121-137.

Parviainen, M., Luoto, M. and Heikkinen, R.K., 2010. NDVI-based productivity and heterogeneity as indicators of plant-species richness in boreal landscapes. *Boreal environment research*, *15*(3).

Pau, S., Gillespie, T.W. and Wolkovich, E.M., 2012. Dissecting NDVI–species richness relationships in Hawaiian dry forests. *Journal of Biogeography*, *39*(9), pp.1678-1686.

Pellegrini, A.F., Socolar, J.B., Elsen, P.R. and Giam, X., 2016. Trade-offs between savanna woody plant diversity and carbon storage in the Brazilian Cerrado. *Global change biology*, *22*(10), pp.3373-3382.

Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C. and Coops, N.C., 2013. Essential biodiversity variables. *Science*, *339*(6117), pp.277-278.

Reyers, B. and McGeoch, M.A., 2007. A biodiversity monitoring framework for South Africa: progress and directions. *South African Journal of Science*, *103*(7-8), pp.295-300.

Rocchini, D., Chiarucci, A. and Loiselle, S.A., 2004. Testing the spectral variation hypothesis by using satellite multispectral images. *Acta Oecologica*, *26*(2), pp.117-120.

Rocchini, D., 2007. Effects of spatial and spectral resolution in estimating ecosystem α-diversity by satellite imagery. *Remote Sensing of Environment*, *111*(4), pp.423-434.

Rocchini, D., Balkenhol, N., Carter, G.A., Foody, G.M., Gillespie, T.W., He, K.S., Kark, S., Levin, N., Lucas, K., Luoto, M. and Nagendra, H., 2010. Remotely sensed spectral heterogeneity as a proxy of species diversity: recent advances and open challenges. *Ecological Informatics*, *5*(5), pp.318-329.

Scholes, R.J. and Archer, S.R., 1997. Tree-grass interactions in savannas. *Annual review of Ecology and Systematics*, *28*(1), pp.517-544.

Seymour, C.L. and Dean, W.R.J., 2010. The influence of changes in habitat structure on the species composition of bird assemblages in the southern Kalahari. *Austral Ecology*, *35*(5), pp.581-592.

Shackleton, C.M., 2000. Comparison of plant diversity in protected and communal lands in the Bushbuckridge lowveld savanna, South Africa. *Biological Conservation*, *94*(3), pp.273-285.

Shannon, C.E., 1948. A mathematical theory of communication, Part I, Part II. *Bell Syst. Tech. J.*, *27*, pp.623-656.
Simpson, E.H., 1949. Measurement of diversity. *Nature*.

Schlesinger, J.; Drescher, A.; Shackleton, C. M.,2015. Socio-spatial dynamics in the use of wild natural resources: Evidence from six rapidly growing medium-sized cities in Africa. *Applied Geography Vol*. 56 Pg. 107-115.

Stevens, N., Swemmer, A.M. and Ezzy, L., 2014. Investigating potential determinants of the distribution limits of a savanna woody plant: Colophospermummopane. *Journal pf Vegetation Science Vol*. 25 (2), pp. 363-373.

Thenkabail, P.S., Hall, J., Lin, T., Ashton, M.S., Harris, D. and Enclona, E.A., 2003. Detecting floristic structure and pattern across topographic and moisture gradients in a mixed species Central African forest using IKONOS and Landsat-7 ETM+ images. *International Journal of Applied Earth Observation and Geoinformation*, *4*(3), pp.255-270.

Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E. and Steininger, M., 2003. Remote sensing for biodiversity science and conservation. *Trends in ecology & evolution*, *18*(6), pp.306-314.

Ustin, S.L. and Gamon, J.A., 2010. Remote sensing of plant functional types. *New Phytologist*, *186*(4), pp.795-816.

Zhang, Q., Xiao, X., Braswell, B., Linder, E., Baret, F. and Moore, B., 2005. Estimating light absorption by chlorophyll, leaf and canopy in a deciduous broadleaf forest using MODIS data and a radiative transfer model. *Remote Sensing of Environment*, *99*(3), pp.357-371.

Zhao, Y., Zeng, Y., Zhao, D., Wu, B. and Zhao, Q., 2016. The optimal leaf biochemical selection for mapping species diversity based on imaging spectroscopy. *Remote Sensing*, *8*(3), p.216.

Waide, R.B., Willig, M.R., Steiner, C.F., Mittelbach, G., Gough, L., Dodson, S.I., Juday, G.P. and Parmenter, R., 1999. The relationship between productivity and species richness. *Annual review of Ecology and Systematics*, *30*(1), pp.257-300.

Wang, J., Rich, P.M., Price, K.P. and Kettle, W.D., 2004. Relations between NDVI and tree productivity in the central Great Plains. *International Journal of Remote Sensing*, *25*(16), pp.3127-3138.

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

Witman, J.D., Cusson, M., Archambault, P., Pershing, A.J. and Mieszkowska, N., 2008. The relation between productivity and species diversity in temperate–arctic marine ecosystems. *Ecology*, *89*(sp11).