This article was published in ISPRS
http://www.sciencedirect.com/science/article/pii/S0924271617301612

Remote sensing of species diversity using Landsat 8 spectral variables

Sabelo Madonsela[1,2], Moses Cho[1,2,3], Abel Ramoelo[1,2,4], Onisimo Mutanga[2].

[1]Earth Observation Research Group, Natural Resources and Environment, Council for Scientific and Industrial Research (CSIR), Pretoria, South Africa.
[2]School of Agric. Earth and Environmental Sciences, University of KwaZulu-Natal (UKZN), Pietermaritzburg, South Africa.
[3]Department of Plant Science, University of Pretoria
[4]Risk and Vulnerability Assessment Centre, University of Limpopo, Sovenga, South Africa.

Keywords: PCA, NDVI, Landsat-8, savannah, tree species diversity.

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.

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. 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). 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 wet 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).

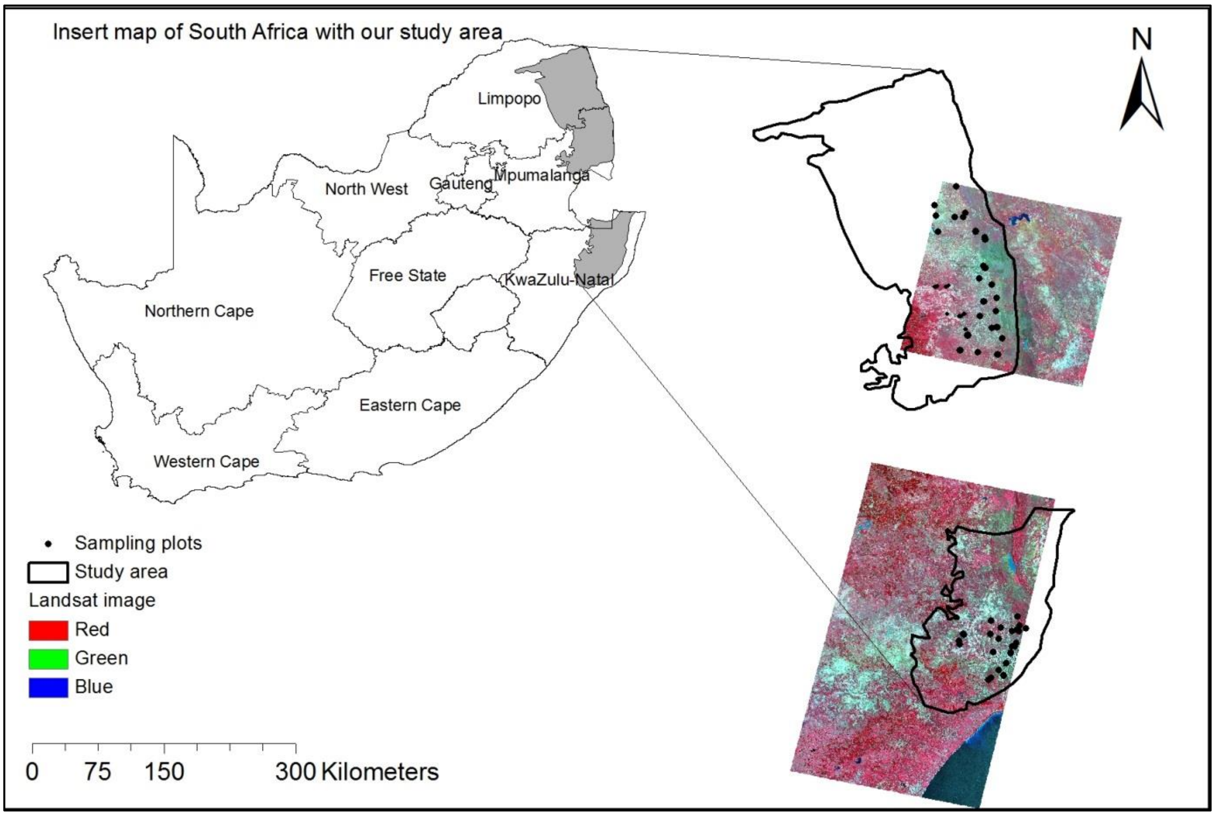

Figure 1 Study area stretching across three provinces of South Africa. Dots on the Landsat imagery are the sampling plots.

3. Material and methods

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 wet season and is characterized as a peak productivity period (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).

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 2**). 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.

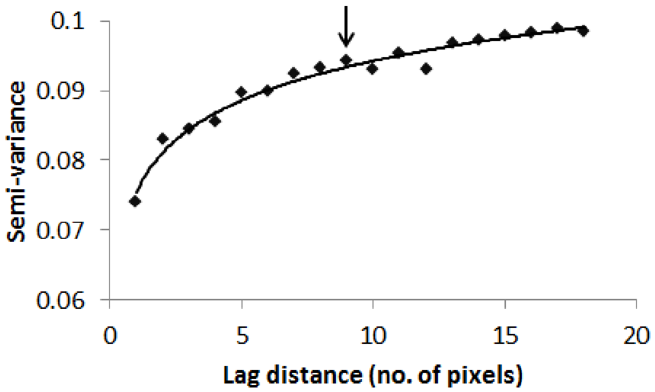

Figure 2 Semi-variogram analysis showing the scale of tree species variability in the savannah woodland.

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*$_2$) (see 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	$H' = -\sum_{i=1}^{S} p_i \mathrm{In}(p_i)$	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; $\mathrm{In}(p_i)$ 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	$= \sum_{i,j=0}^{N-1} Pi,j\|i-j\|$	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 $(j = 1, \dots, j)$ and are all vectors in the I-dimensional space. A linear model of these x variables can be expressed as $t = w_1 \times x_1 + \dots + w_j \times x_j$, 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).

4. Results

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 3**). 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 4, 5 and 6**). 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 7**) 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	±0.003	0.0005	0.4861
	NDVI (St dev)	0.10	±0.003	0.0167	0.5586
	NDVI (Range)	0.11	±0.003	0.0114	0.5524
	EVI (Mean)	0.29	±0.003	0.0008	0.4869
	EVI (St dev)	0.17	±0.003	0.0063	0.5302
	EVI (Range)	0.17	±0.003	0.0073	0.5286
	SRI (Mean)	0.26	±0.003	0.0006	0.4985
	SRI(St dev)	0.36	±0.003	0.0000	0.4613
	SRI (Range)	0.34	±0.003	0.0000	0.4688
	SAVI (Mean)	0.29	±0.003	0.0007	0.4894
	SAVI (St dev)	0.10	±0.003	0.0118	0.5554
	SAVI (Range)	0.11	±0.003	0.0113	0.5536
D_2	NDVI (Mean)	0.29	±0.003	0.0003	1.8048
	NDVI (St dev)	0.12	±0.004	0.0132	2.0724

	NDVI (Range)	0.11	±0.004	0.0144	2.0761
	EVI (Mean)	0.29	±0.003	0.0003	1.8203
	EVI (St dev)	0.20	±0.003	0.0037	1.9493
	EVI (Range)	0.17	±0.003	0.0061	1.9599
	SRI (Mean)	0.27	±0.003	0.0005	1.8545
	SRI(St dev)	0.41	±0.003	0.0000	1.6668
	SRI (Range)	0.38	±0.003	0.0006	1.7232
	SAVI (Mean)	0.29	±0.003	0.0003	1.8250
	SAVI (St dev)	0.12	±0.004	0.0142	2.0605
	SAVI (Range)	0.11	±0.004	0.0145	2.0569
S	NDVI (Mean)	0.23	±0.003	0.0020	3.4913
	NDVI (St dev)	0.00	±0.001	0.5392	3.9684
	NDVI (Range)	0.00	±0.001	0.4461	3.9580
	EVI (Mean)	0.21	±0.003	0.0027	3.5516
	EVI (St dev)	0.03	±0.003	0.1791	3.9535
	EVI (Range)	0.01	±0.002	0.2128	3.9861
	SRI (Mean)	0.23	±0.003	0.0024	3.5513
	SRI(St dev)	0.24	±0.003	0.0017	3.4732
	SRI (Range)	0.22	±0.003	0.0025	3.5494
	SAVI (Mean)	0.23	±0.003	0.0073	3.5263
	SAVI (St dev)	0.00	±0.001	0.4458	3.9171
	SAVI (Range)	0.00	±0.001	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**	***R²***	***P-value***	***R²***	***P-value***	***R²***	***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	±0.003	0.0071	0.5280	0.17	±0.002	0.0060	1.961	0.16	±0.003	0.0079	3.6878
Blue band	0.13	±0.004	0.0127	0.5431	0.14	±0.003	0.0105	1.996	0.13	±0.004	0.0117	3.7537
Green band	0.10	±0.004	0.0173	0.5564	0.11	±0.003	0.0159	2.063	0.09	±0.004	0.0172	3.8579
Red band	0.18	±0.003	0.0060	0.5246	0.19	±0.002	0.0042	1.934	0.14	±0.003	0.0105	3.7192
NIR	0.02	±0.002	0.9999	0.5728	0.02	±0.002	0.9999	2.158	0.00	±0.001	0.9999	3.940
Cirrus	0.01	±0.002	0.9999	0.5767	0.01	±0.002	0.9999	2.134	0.00	±0.001	0.9999	3.991
SWIR-1 band	0.09	±0.004	0.0180	0.5623	0.09	±0.005	0.0193	2.097	0.07	±0.004	0.0212	3.9201
SWIR-2 band	0.12	±0.004	0.0146	0.5489	0.14	±0.003	0.0111	2.019	0.09	±0.004	0.0176	3.8834

CI- Confidence interval

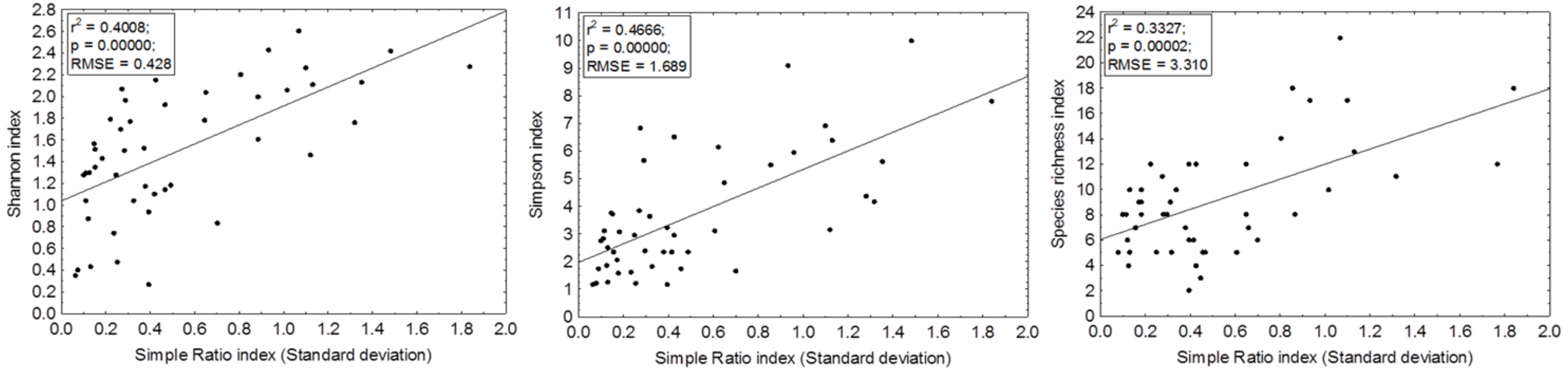

Figure 3 Relationship between Simple Ration Index derivative and on the left) Shannon index; middle) Simpson index; right) Species richness. SRI standard deviation had shown higher positive relationship with tree species diversity and we selected the best model (maximum *r^2* with the lowest RMSE from 1000 bootstrapped iterations) to plot the relationship.

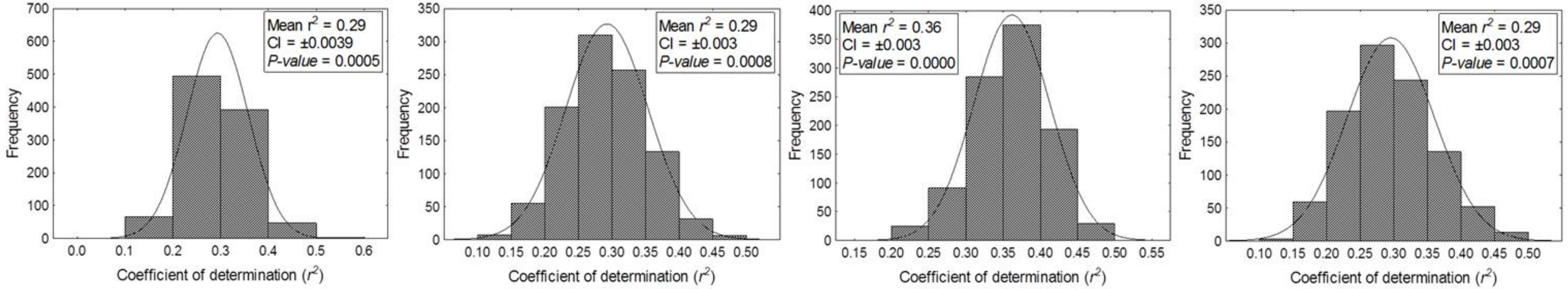

Figure 4 Histograms of bootstrapped r2 for models involving Shannon index and on the left) mean NDVI; second from left) mean EVI; third from left) SRI standard deviation; fourth from left) mean SAVI.

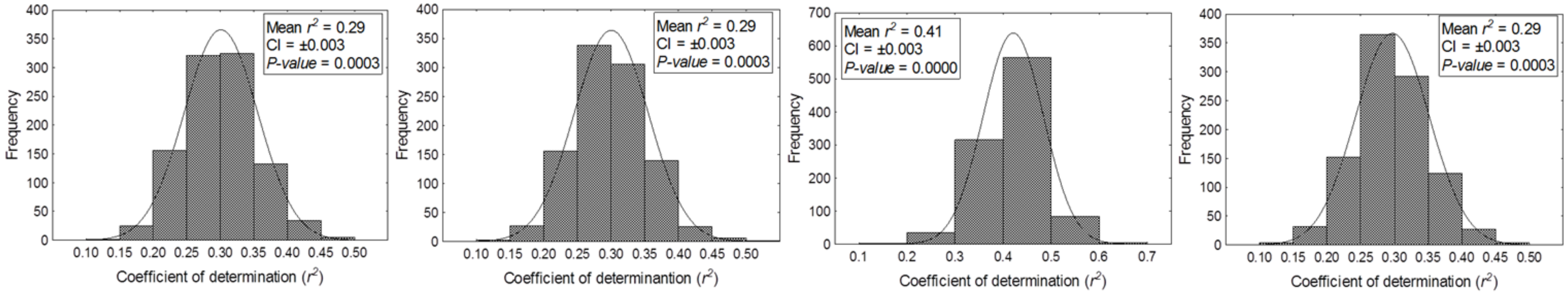

Figure 5 Histograms of bootstrapped r2 for models involving Simpson index and on the left) mean NDVI; second from left) mean EVI; third from left) SRI standard deviation; fourth from left) mean SAVI.

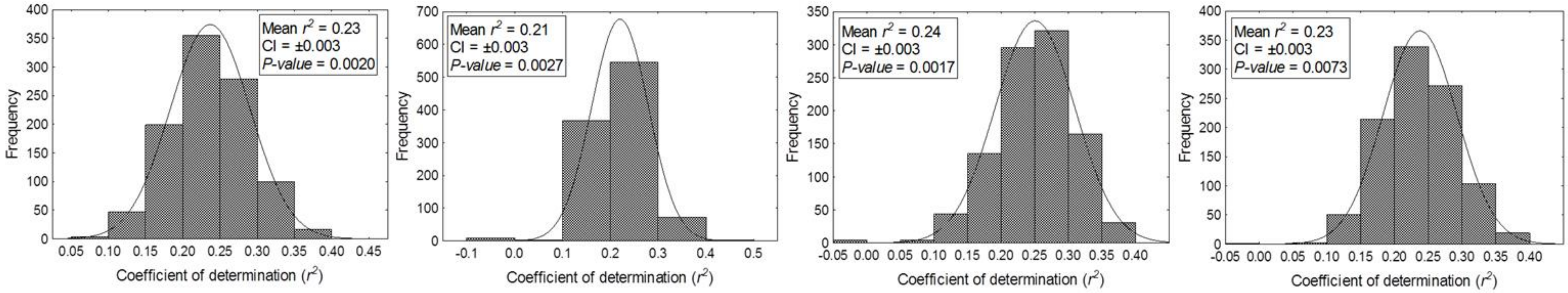

Figure 6 Histograms of bootstrapped r2 for models involving Species richness and on the left) mean NDVI; second from left) mean EVI; third from left) SRI standard deviation; fourth from left) mean SAVI.

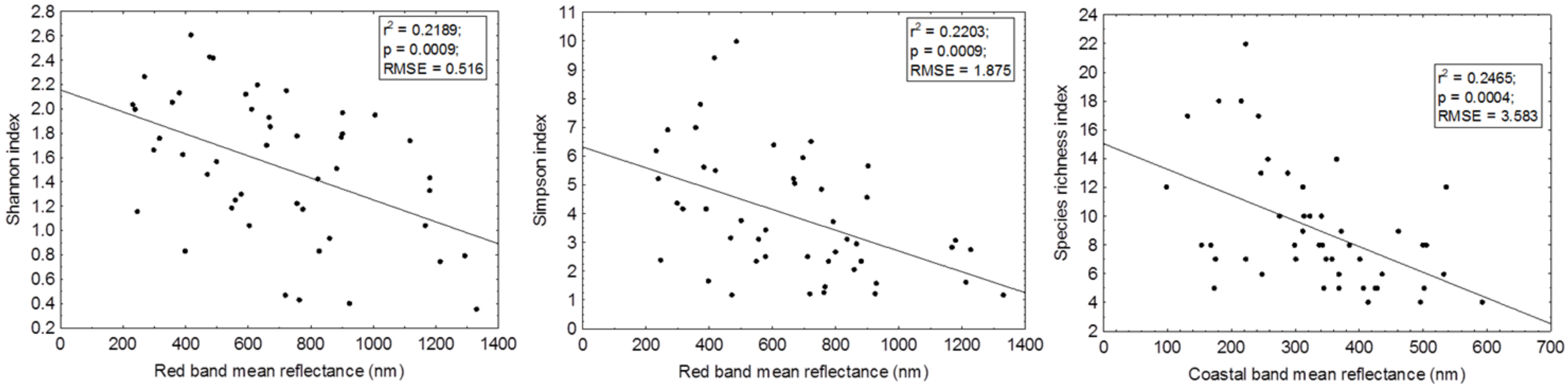

Figure 7 Relationship between red band reflectance and on the left) Shannon index; middle) Simpson index; right) species richness. Red band had shown higher negative relationship tree species diversity than other spectral bands and we selected one best model (maximum r^2 with the lowest RMSE from 1000 bootstrapped iterations) to plot the relationship.

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	±0.003	0.0000	0.459
	D_2	2	0.38	±0.003	0.0000	1.699
	S	2	0.22	±0.003	0.0027	3.627
Landsat bands	*H'*	3	0.36	±0.003	0.0002	0.480

	D_2	3	0.35	±0.003	0.0002	1.822
	S	3	0.19	±0.004	0.0075	3.826
GLCM texture measures	H'	7	0.21	±0.003	0.0059	0.55`9
	D_2	7	0.20	±0.004	0.0069	2.093
	S	7	0.22	±0.006	0.0063	3.890
VIs + Landsat	H'	2	0.41	±0.003	0.0001	0.456
	D_2	2	0.42	±0.003	0.0000	1.663
	S	2	0.22	±0.005	0.0035	3.657
VIs + GLCM	H'	6	0.35	±0.006	0.0009	0.523
	D_2	6	0.32	±0.006	0.0017	1.984
	S	6	0.21	±0.004	0.0040	3.647
Landsat + GLCM	H'	6	0.32	±0.004	0.0005	0.516
	D_2	6	0.26	±0.004	0.0012	1.950
	S	6	0.27	±0.005	0.0014	3.523
VIs + Landsat + GLCM layers	H'	5	0.40	±0.005	0.0004	0.487
	D_2	5	0.40	±0.004	0.0004	1.823
	S	5	0.26	±0.005	0.0030	3.655

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).

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 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. We further thank the anonymous reviewers for their comments which improved the quality of the paper.

References

1. Albregtsen, F., 2008. Statistical texture measures computed from gray level coocurrence matrices. *Image processing laboratory, department of informatics, University of Oslo*, *5*.
2. 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.
3. 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.
4. 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.
5. 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).
6. 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).
7. Bro, R. and Smilde, A.K., 2014. Principal component analysis. *Analytical Methods*, *6*(9), pp.2812-2831.
8. 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.
9. 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.

10. Colwell, R.K., 2009. Biodiversity: concepts, patterns, and measurement. *The Princeton guide to ecology*, pp.257-263.
11. 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.
12. 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.
13. du Toit, J.T., Biggs, H.C., Rogers, K.H., 2003. The Kruger Experience: ecology and management of savanna heterogeneity. London: Island Press.
14. 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.
15. 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.
16. 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.
17. 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.
18. Gould, W., 2000. Remote sensing of vegetation, plant species richness, and regional biodiversity hotspots. *Ecological applications*, *10*(6), pp.1861-1870.
19. 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.
20. Gringarten, E. and Deutsch, C.V., 2001. Teacher's aide variogram interpretation and modeling. *Mathematical Geology*, *33*(4), pp.507-534.
21. Haralick, R.M. and Shanmugam, K., 1973. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6), pp.610-621.
22. 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.
23. 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.
24. 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.

25. 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.
26. Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). *Remote sensing of environment*, *25*(3), pp.295-309.
27. 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.
28. Huete, A., Justice, C. and Van Leeuwen, W., 1999. MODIS vegetation index (MOD13). *Algorithm theoretical basis document*, *3*, p.213.
29. 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.
30. 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).
31. Jongman, R.H., Ter Braak, C.J. and Van Tongeren, O.F. eds., 1995. *Data analysis in community and landscape ecology*. Cambridge University press.
32. 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.
33. 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.
34. 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.
35. Lande, R., 1996. Statistics and partitioning of species diversity, and similarity among multiple communities. *Oikos*, pp.5-13.
36. Landsat-8 (L8) Data User Handbook, Version 2.0, March 29, 2016.
37. 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.
38. 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.
39. 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.

40. 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.
41. 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.
42. 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.
43. "MATLAB R2009a, the Language of Technical Computing," Mathworks Inc.
44. 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.
45. 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.
46. Nagendra, H., 2002. Opposite trends in response for the Shannon and Simpson indices of landscape diversity. *Applied Geography*, *22*(2), pp.175-186.
47. 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.
48. 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.
49. 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.
50. 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.
51. 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).
52. 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.

53. 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.
54. 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.
55. 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.
56. 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.
57. 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*.
58. 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.
59. Rossel, R.A.V., 2008. ParLeS: Software for chemometric analysis of spectroscopic data. *Chemometrics and intelligent laboratory systems*, *90*(1), pp.72-83.
60. Rouse Jr, J., Haas, R.H., Schell, J.A. and Deering, D.W., 1974. Monitoring vegetation systems in the Great Plains with ERTS.
61. Scholes, R.J. and Archer, S.R., 1997. Tree-grass interactions in savannas. *Annual review of Ecology and Systematics*, *28*(1), pp.517-544.
62. 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.
63. 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.
64. Shannon, C.E., 1948. A mathematical theory of communication, Part I, Part II. *Bell Syst. Tech. J.*, *27*, pp.623-656.
65. Simpson, E.H., 1949. Measurement of diversity. *Nature*.
66. 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.
67. 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.

68. 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.
69. 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.
70. Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment*, *8*(2), pp.127-150.
71. 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.
72. 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.
73. 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.
74. 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.
75. 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.