Estimating tree species diversity in the savannah using NDVI and woody canopy cover

Sabelo Madonsela[1,2], Moses Azong Cho[1,2,3], Abel Ramoelo[1,2,4], Onisimo Mutanga[2], Laven Naidoo[1],

[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, South Africa.

[4]Risk and Vulnerability Assessment Centre, University of Limpopo, Sovenga, South Africa.

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.

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. 2017).

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.

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 1). 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).
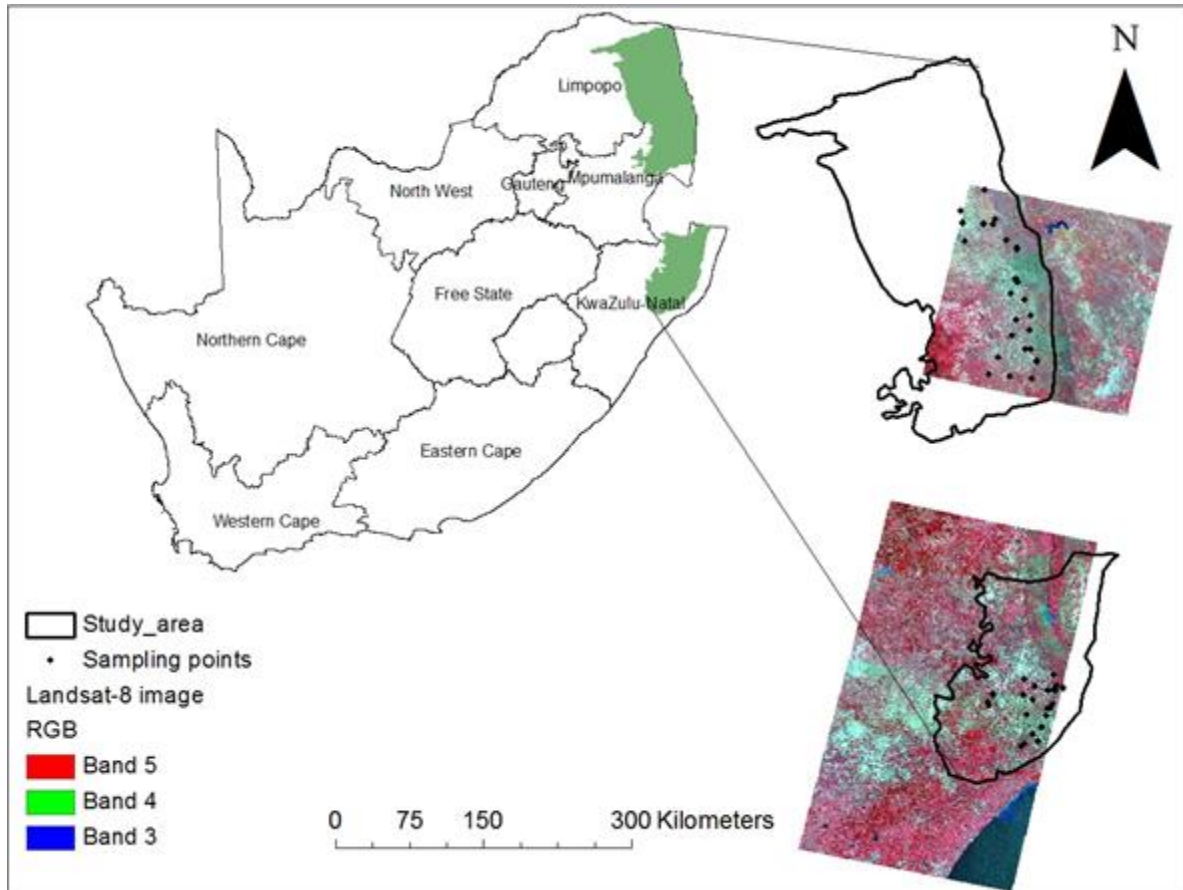
Figure 1 Study area stretching across three provinces of South Africa. The black dots represent the sampling plots

3. Material and Methods

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 peak productivity (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). 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%.

3.2 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:-

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

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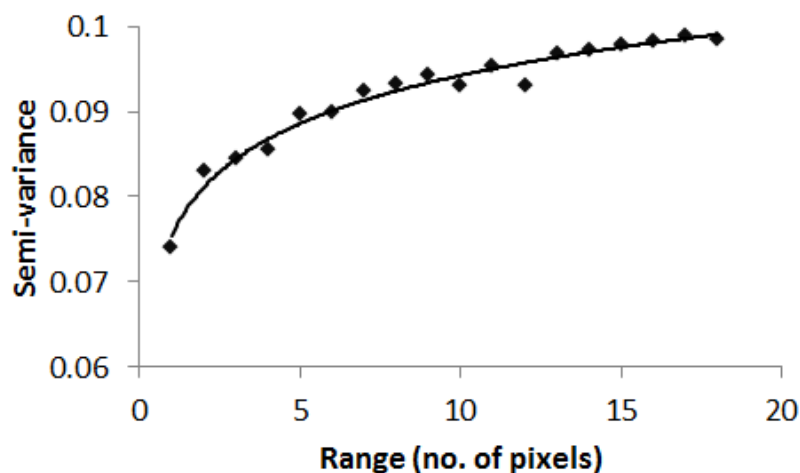
Figure 2 Semi-variogram analysis showing the scale of tree species variability in the savannah woodland

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. 2017). 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:-

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

where p_i is the proportional abundance of species i relative to the total abundance of all species S in a plot; $\ln(p_i)$ 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.

4. Results

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 wet season NDVI$_{March}$ ($r^2 = 0.24$; $p < 0.01$) or transition to senescence NDVI$_{April}$ ($r^2 = 0.19$; $p < 0.01$) (**Table 1**). 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 2**). Consequently, the woody canopy cover model explained 21% of tree species diversity, although the improvement was still lower than the variance explained by NDVI$_{March}$. The NDVI$_{March}$ model explained 33% of tree species diversity after the removal of outliers.

In the April date which represents the transition to senescence, the NDVI$_{April}$ 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 2**). 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 (NDVI$_{July}$) (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. 3a, 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. 4a, 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 1 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
	$NDVI_{March}$	0.24	±0.018	0.003	0.497
	$NDVI_{April}$	0.19	±0.026	0.006	0.518
	$NDVI_{May}$	0.13	±0.030	0.013	0.533
	$NDVI_{July}$	0.06	±0.024	0.021	0.554

CI = confidence interval

Table 2 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
	$NDVI_{March}$	0.33	±0.019	0.0001	0.454
	$NDVI_{April}$	0.27	±0.022	0.0008	0.464
	$NDVI_{May}$	0.20	±0.019	0.0048	0.492
	$NDVI_{July}$	0.13	±0.019	0.0119	0.513

CI = confidence interval
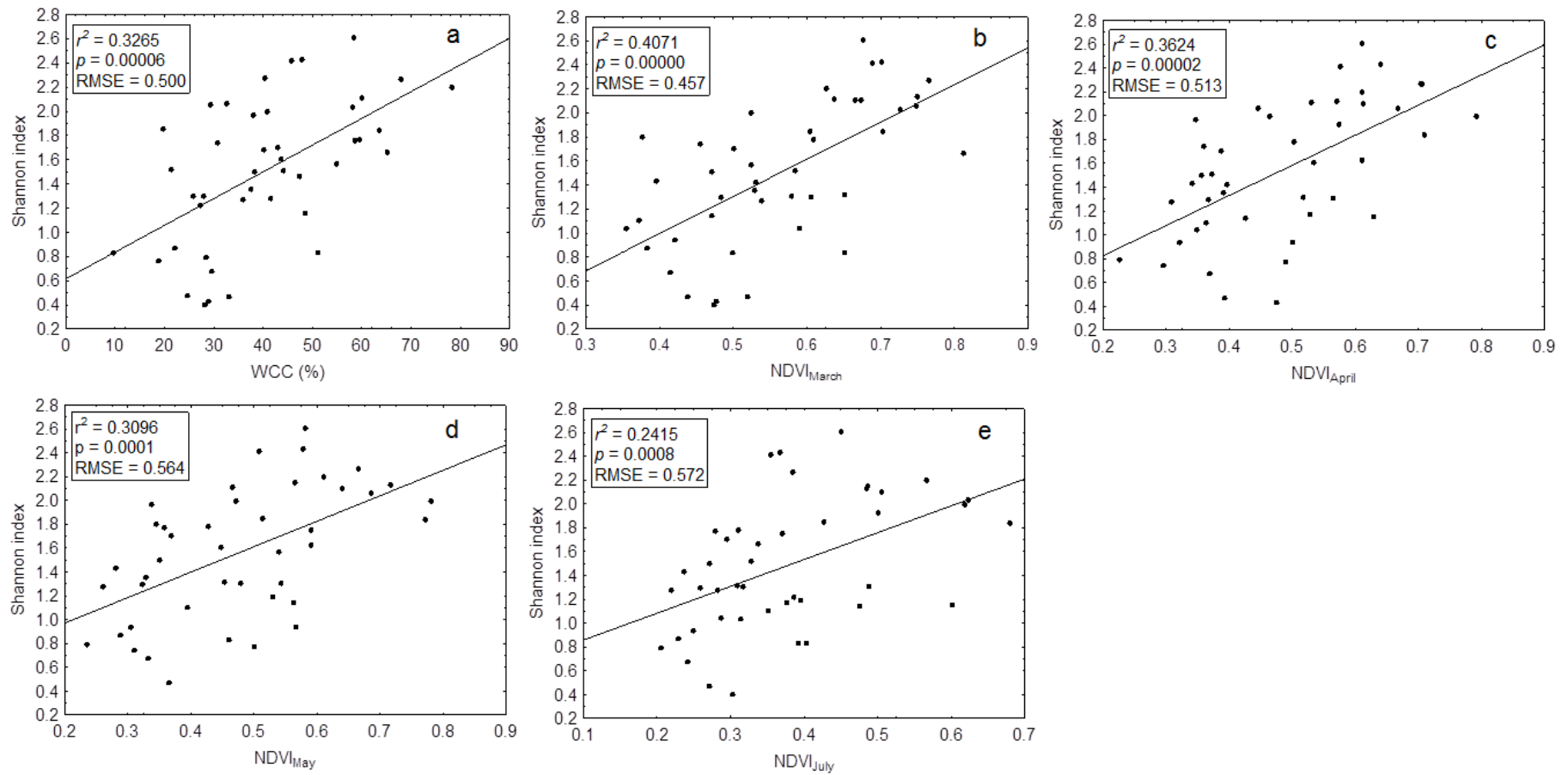
Figure 3 Scatterplot showing linear relationship between predictor variables (woody canopy cover and NDVI) and Shannon index after removing outliers. The scatterplots were selected from the best regression models (maximum r^2 with the lowest RMSE from 30 bootstrapped iterations).
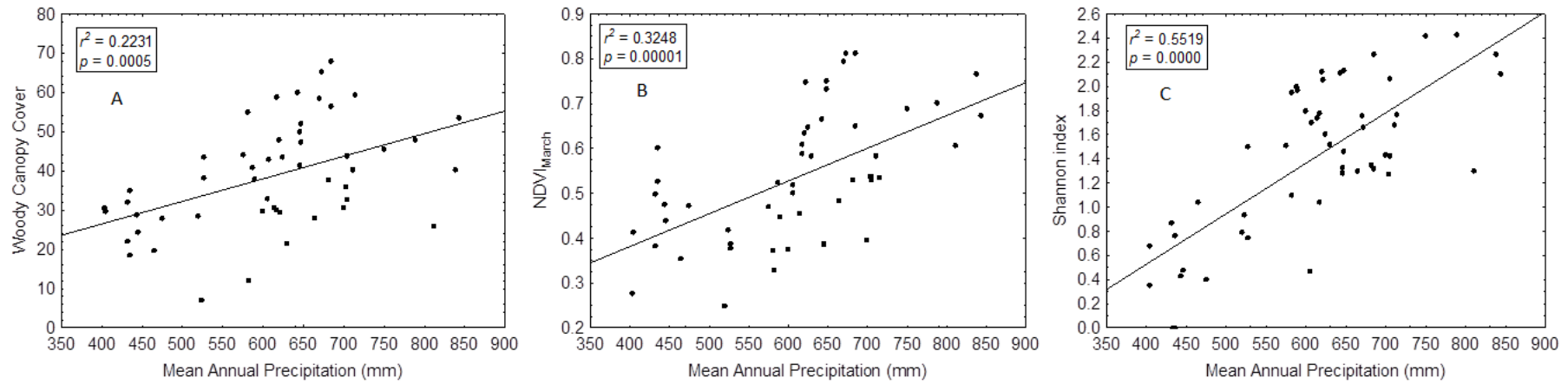
Figure 4 Linear relationship between mean annual precipitation and a) woody canopy cover b) NDVI$_{March}$ and c) Shannon index. The rainfall data covers only the Kruger National Park.

4.2 Factorial regression model

The factorial regression models performed better than NDVI or woody canopy cover model when estimating tree species diversity (**Table 1 and 3**). However the results in table 1 and 3 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 $NDVI_{March}$ 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 $NDVI_{March}$ model in **Table 1** (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 5**). 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 7 and 8**).

However, combining $NDVI_{July}$ 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 $NDVI_{July}$ model (p <0.01) (**Table 7 and 8**). Combining either $NDVI_{March}$ or $NDVI_{April}$ with woody canopy cover in a factorial model did not significantly reduce the error of prediction when compared to either $NDVI_{March}$ model or $NDVI_{April}$ 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 7 and 8**).

The factorial model results present three observations; i) at the end of wet 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 wet 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 6**). 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 6**). The interaction between NDVI and woody canopy cover had the lowest explanatory power in the factorial model across all phenological periods. **Table 4** presents variance partitioning prior the removal of outliers.

Table 3 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 4 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 5 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 6 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 7 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 8 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

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 5**). 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 5d-e)**.
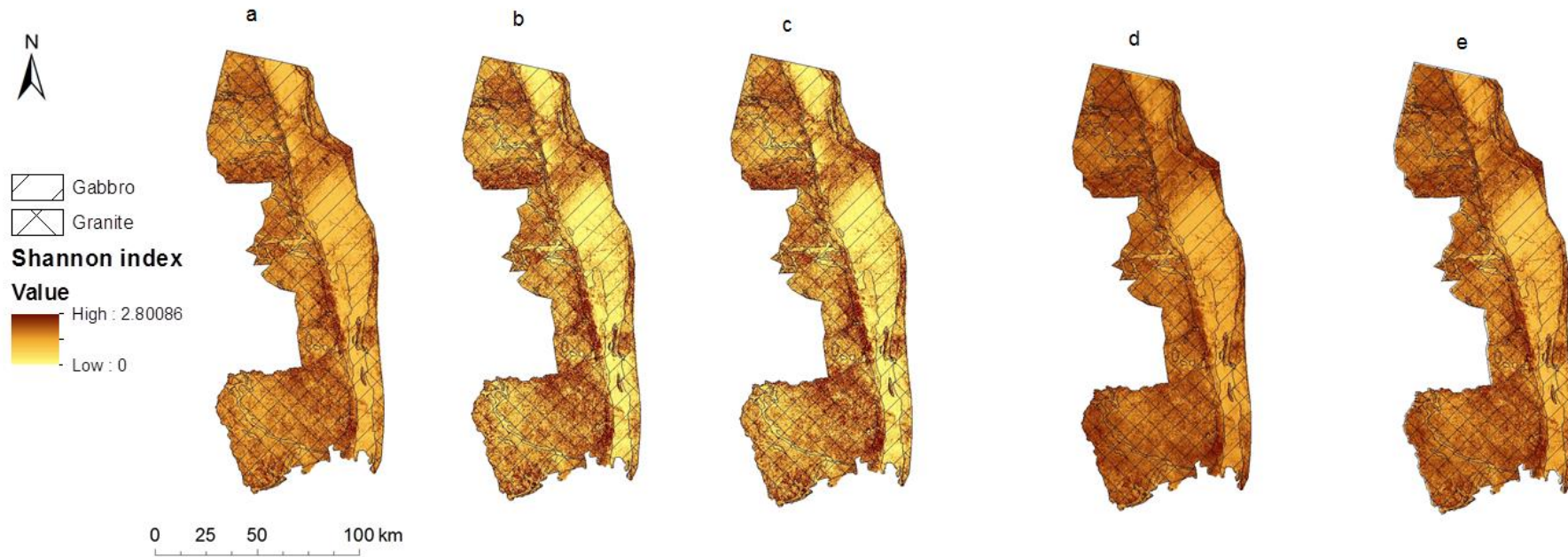
Figure 5 Tree species diversity from the best (a) woody canopy cover (WCC) model, (b) factorial model involving NDVI$_{March}$ and WCC (c) factorial model involving NDVI$_{April}$ and WCC (d) factorial model involving NDVI$_{May}$ and WCC and (e) factorial model involving NDVI$_{July}$ and WCC

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 wet 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., 2017) 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 during the wet 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 $NDVI_{March}$ or $NDVI_{April}$ 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 during wet season. Moreover, wet 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 5b 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.

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 during wet 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 wet season.

Acknowledgements

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., 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.

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.