# MODELLING AND PREDICTING MEASURES OF TREE SPECIES DIVERSITY USING AIRBORNE LASER SCANNING DATA IN MIOMBO WOODLANDS OF TANZANIA

**Mauya, Ernest William**

*College of Forestry, Wildlife and Tourism, Sokoine University of Agriculture, Department of Forest Engineering and Wood Sciences, P.O. Box 3012, Morogoro, Tanzania.*

*Corresponding author, e-mail: mauya@sua.ac.tz

## ABSTRACT

In the recent decade, remote sensing techniques had emerged as one among the best options for quantification of measures of tree species diversity. In this study, potential of using remotely sensed data derived from airborne laser scanning (ALS) for predicting tree species richness and Shannon diversity index was evaluated. Two modelling approaches were tested: linear mixed effects modelling (LMM), by which each of the measures was modelled separately, and the k-nearest neighbour technique (k-NN), by which both measures were jointly modelled (multivariate approach). For both methods, the effect of vegetation type on the prediction accuracies of tree species richness and Shannon diversity index was tested. Separate predictions for richness and Shannon diversity index using LMM resulted in relative root mean square errors (RMSEcv) of 40.7%, and 39.1%, while for the k-NN they were 41.4% and 39.1%, respectively. Inclusion of dummy variables representing vegetation types to the LMM improved the prediction accuracies of tree species richness (RMSEcv = 40.2%) and Shannon diversity index (RMSEcv = 38.0%). The study concluded that ALS data has a potential for modelling and predicting measures of tree species diversity in the miombo woodlands of Tanzania.

**Key words:** Airborne laser scanning, biodiversity, miombo woodland, Liwale-Tanzania, k-NN, tree species diversity.

## INTRODUCTION

Tropical forests ecosystems host at least two thirds of the earth's terrestrial biodiversity and provide significant human benefits at local and regional scales through the provision of economic goods and services (Gardner *et al.* 2009). Nevertheless, the biodiversity of tropical forests is declining rapidly due to conversion of forests to permanent agricultural land, climate change, induced fires, and unsustainable logging practices (Chidumayo 2013, Bellard *et al.* 2012). Other consequences of these changes include increasing amount of carbon emission from tropical forests as compared to other forest types globally. To mitigate this development, and to conserve the beneficial ecosystem services that tropical forests can provide, Reducing Emissions from Deforestation and Forest Degradation (REDD+) has been negotiated as an international effort under the United Nations Framework Convention on Climate Change (UNFCCC) since 2005. REDD+ has now become the most active global land use policy program (Matthews and van Noordwijk 2014) in the tropical countries. Although the primary objective of the REDD+ framework is climate change mitigation through enhancement of forest carbon stock and sustainable forest management, co-benefits from biodiversity conservation are also expected (Imai *et al.* 2014). However, despite this considerable potential, methods for monitoring forest biodiversity are lacking in the context of REDD+ (Ehara *et al.* 2014, Roe *et al.* 2013, Gardner *et al.* 2012). Efforts have so far focused on the establishment of methods for accounting large-scale forest carbon stock and changes, with little emphasis on forest biodiversity assessment (Imai *et al.* 2014). Such methods are also important for other subsidiary bodies like the Convention on Biological Diversity (Chandra and Idrisova 2011) and Millennium Ecosystem Assessment (Mooney *et al*. 2004).

Assessments of biodiversity for management, conservation planning or policy development (e.g., REDD+), however, poses a number of challenges, especially at larger scales due to the broad, multi-dimensional and multi-scale characteristics associated with such assessments (Boutin *et al*. 2009). Unlike in forest carbon assessments, where agreed and consistent metrics exist (e.g., carbon density or emissions in metric tons per hectare), there is lack of consensus on what to monitor in biodiversity assessments (Pereira *et al*. 2013). To account for this, the emphasis has turned to acquiring data for estimation of indicators that aggregate and synthesize information for describing multiple aspects of biodiversity simultaneously (Winter *et al*. 2011). Species diversity is one of the most intuitive and widely adopted indicators of biodiversity because it is strongly correlated with other biodiversity attributes such as genetic diversity and ecosystem functioning (Pereira and Cooper 2006; Colwell and Coddington 1994; Chiarucci and Palmer 2006). Thus, in many studies that aimed at describing biodiversity of forest ecosystems, tree species diversity has been used as an indicator of forest biodiversity (e.g., Barna and Bosela 2015, Schmidt *et al.* 2015, Vanderhaegen *et al*. 2015, Shirima *et al*. 2015). Additionally, from cost-benefit and time-efficiency perspectives, tree species diversity is considered as a more viable option for biodiversity assessment since it can easily be derived from existing datasets such as national forest inventories (Chirici *et al*. 2012).

Species richness and species evenness are two basic measures describing species diversity of different taxa (Magurran 1988). Species richness, computed as the total number of species in a community, has frequently been used as a measure of tree species diversity in different forest types (e.g., Chiarucci and Palmer 2006, McRoberts and Meneguzzo 2005), especially for those cases that involves comparisons of conservation values of different sites. Species evenness (sometimes known as equitability) is describing the way abundance is distributed among individual species in a community. A biological community, in which all the species are represented by the same number of individuals, has high species evenness while a community in which a few species are represented by many individuals and the other species are represented by few individuals, has a low species evenness. A number of indices taking into account both species richness and evenness have also been proposed for measuring species diversity. In this category, the Shannon diversity index (e.g., Magurran 1988) has frequently been used in many studies on tree species diversity (Shirima *et al.* 2015, Borah *et al.* 2015, Nadeau and Sullivan 2015).

The computation and reliability of measures for tree species richness and evenness, or for the two combined, in estimating tree species diversity rely heavily on accurate field-based information. Traditionally, assessment of tree species diversity has been done by using ground-based surveys. However, for larger spatial scales ground field-based surveys are impractical due to huge amount of data to be collected. Current remote sensing techniques may close this gap by providing spatially continuous and time-series information that can be used to describe tree species diversity (Müller and Vierling 2014, Fricker *et al*. 2015). Moreover, remote sensing allows frequently repeated recording of environmental information, and may thereby provide time-series information that are essential for monitoring and understanding how tree species diversity respond to different environmental factors over time (Leutner *et al*. 2012).

Airborne laser scanning (ALS) is a remote sensing technique that recently has gained wide acceptance in ecologically based studies due to its ability to quantify the three-dimensional (3D) structure of forests, which is of particular interest in characterizing measures of tree species diversity and other taxa in the forest ecosystem (Müller and Vierling 2014, Fricker *et al*. 2015). However, ALS cannot measure directly the measures of tree species diversity, thus applications of ALS involve the development of statistical models or classifiers that relate the ALS structural metrics to measures of tree species diversity derived from field plots. By using such models or classifiers along with statistical sampling estimators (e.g., Chao and Shen 2003), estimates of tree species diversity may be produced for relevant geographical areas of interest. Models may also be used for developing tree species distribution maps that can support decision-making and conservation planning. Several studies in temperate and boreal forests, have used ALS data for modelling and predicting measures of tree species diversity (Leutner *et al*. 2012, Simonson *et al*. 2012, Ceballos *et al*. 2015, Hernández-Stefanoni *et al*. 2014) together with other taxa such as beetles (Müller and Brandl 2009), spiders (Vierling *et al*. 2011) and birds (Vogeler *et al*. 2014, Lindberg *et al*. 2015). Parametric, and to lesser extent non-parametric methods (e.g., random forests), have been used in these studies. For example, a study by Leutner *et al*. (2012) conducted in the temperate montane forests of Germany used non parametric- random forests to compare the ability of LiDAR and hyperspectral remote sensing data in predicting tree species richness. Their results concluded that for modelling tree species richness, LiDAR predictors were the best choice with R2 of 0.3. On the other hand, a study by Hernández-Stefanoni *et al*. (2014) conducted in Mexico used parametric method in particular the ordinary least square regression to model the relationship between tree species richness and LiDAR metrics for different field plot sizes. Their results indicated high potential of using LiDAR data for predicting tree species richness especially for the larger plot sizes where R2 increased from 0.32 to 0.67 for the large plot size.

The non-parametric approach k-nearest neighbors (k-NN) have widely been used in modelling and estimation of different forest attributes such as volume and aboveground biomass (AGB) when using remotely sensed data (McRoberts *et al*. 2015). Among the desirable features of k-NN are the ability to handle multivariate data and non-linear and diverse relationships between dependent and independent variables (Eskelson *et al*. 2009). The multivariate feature of the k-NN technique makes it very useful in ecological applications because management decisions frequently require consistent information on multiple parameters and estimates. However, to our knowledge no study has to date attempted to use k-NN for modelling and prediction of measures of tree species diversity using ALS data. Of particular interests is the tropical forests, where irrespective of the methods used, only a few studies have quantified the relationship between measures of tree species diversity and ALS data (Müller and Vierling 2014). Lack of ALS data and a complex structure due to the large number of tree species in tropical forests, as compared to temperate and boreal forests, are among the possible reasons for that fewer studies of this kind have been carried out in tropical forests. Furthermore, the few studies that do exist (e.g., Fricker *et al*. 2015, Hernández-Stefanoni *et al*. 2014), have not attempted to analyze the influence of vegetation types/forest types on the relationship between measures of tree species diversity and ALS data. Vegetation types are considered as important sources of forest structural variation (Swatantran *et al.* 2011), which would also affect the prediction accuracy of tree species diversity when using ALS data. Stratification and post-stratification of forest inventory information has been suggested as a viable means to reduce prediction errors due to structural variations (Latifi *et al*. 2015), particularly in the studies related to prediction of different forest attributes using ALS data. However, such studies are still limited in the aspects of tree species diversity as compared to AGB and volume.

The overall objective of this study was to assess if ALS data can be used to predict measures of tree species diversity in miombo woodlands, forests and other vegetation types surrounding the woodlands in Liwale District, Tanzania. The specific objectives were to: 1) examine the performance of parametric and non-parametric methods for predicting measures of tree species diversity using ALS data; and 2) assess the prediction accuracy of measures of tree species diversity across vegetation types.

## MATERIALS AND METHODS

### Study area description

The study site is located in Liwale District, Lindi Region (Fig. 1) and it occupies an estimated area of about 15,867 km$^2$ in the south-eastern part of Tanzania. The altitude ranges from 360 to 900 meters above sea level. The area experiences an annual rainfall ranging from 600 mm to 1000 mm, mostly between November and early April. The average annual temperature ranges between 20$^o$C and 30$^o$C (LDC 2014). The vegetation type that dominates the study area is miombo woodlands, characterized by the presence of three genera; *Brachystegia*, *Jubernardia*, and *Isoberlinia* from the family *Fabaceae* and sub-family *Caesalpinioideae*. Even though the area is mainly dominated by miombo woodlands, forests and other vegetation types that were neither woodlands nor forests, were also found in the study area.

### Data collection and processing

#### *Field data*

In this study, field plots that were established and measured by the National Forest Resources Monitoring and Assessment (NAFORMA) program in 2011 were used.

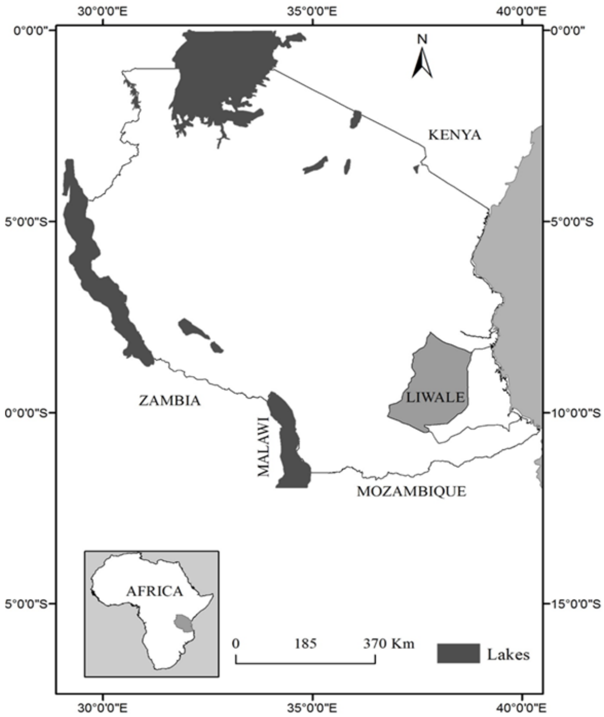



**Figure. 1. Location of the study area in Liwale District, southern Tanzania.**

NAFORMA is the first ground based national forest inventory of Tanzania that was conducted from 2009 to 2014 covering different vegetation types across the country (MNRT 2015). In the study area (i.e., Liwale) the initial measurement by NAFORMA was completed in June 2011. Eight months later (February 2012) all the plots were revisited to ensure temporal consistency with the ALS data acquisition (February/March 2012) but also to accurately record the positions of the plots using survey grade Global Positioning System and the Global Navigation Satellite System receivers.

#### *Sampling design*

The sampling design adopted by NAFORMA is systematic double sampling for stratification with individual plots allocated in clusters (Fig. 2). The details of the planning of this design are given in Tomppo *et al*. (2014); Mauya *et al.* (2015).

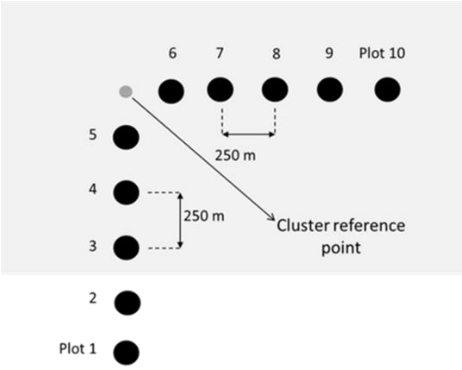


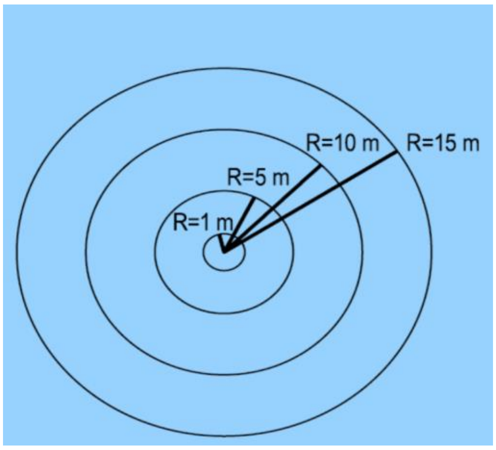

**Figure. 2. The structure of a NAFORMA cluster. The field sample plots are represented by black dots, and the ALS coverage is shown with gray shading. Note that the ALS measurements do not cover the entire cluster.**

### *Plot data*

The NAFORMA cluster structure has a total of ten plots per cluster (MNRT 2011), however during the re-measurement two plots in each cluster were omitted (Fig. 2), because they were outside the corridors designated for ALS data acquisition. Thus, eight plots per cluster were re-measured in the field. Differential Global Navigation Satellite Systems (dGNSS) were used to positions the center of each sample plot. Two Topcon Legacy 40-channels dual frequency receivers observing both pseudo range and carrier phase of the Global Positioning System (GPS) and the Global Navigation Satellite System (GLONASS) were used as rover and base station, respectively.

The re-measurements of the NAFORMA plots used in the current study followed the NAFORMA field protocol. The plots had a concentric circular design with 15 m radius. Within each plot, diameter at breast (dbh) for the trees was measured using calipers or diameter tape following the concentric plot design described in Fig. 3. Height measurements were done for every fifth tally tree in the cluster by using a Suunto hypsometer. Identification of species names of every recorded tree in the plot was done by the professional botanist from Tanzania Forest Research Institute (TAFORI). The identification was also guided by the previous tree species information recorded by NAFORMA in 2011 to ensure consistency between the two measurements and that the same names were used for every measured tree. For all the trees that were identified, the species names were further confirmed by using NAFORMA checklist to ensure that botanic names are correctly recorded. This was done in the Herbarium at Lushoto Silviculture Research Centre before further analysis.

On each plot, information on vegetation types were also recorded. For this study, all the dataset was grouped into three distinct vegetation types (NAFORMA broad categories); woodlands, forest, and other vegetation types as described in Table 1, and further elaborated in MNRT (2011). This information was used to stratify the plots into more homogeneous ecological groups with the objective of helping to explain the variation in tree species diversity better.




1) Within 1 m radius; all trees with dbh > 1cm were recorded
2) Within 5 m radius; all trees with dbh > 5 cm were recorded
3) Within 10 m radius; all trees with dbh >10 cm were recorded
4) Within 15 m radius; all trees with dbh > 20 cm were recorded

**Figure. 3. NAFORMA concentric sample plot design**

**Table 1. Brief description of the structure and biological values of the three vegetation types.**

| Vegetation types | Description | Biodiversity |
|---|---|---|
| Forest | The forest category referred to in this study comprised of low land forests.<br>• They occur at an elevation less than 800m.<br>• They are characterized by semi ever green closed and dense canopy. | High biodiversity values |
| Woodlands | • Generally, the canopy cover for the woodlands is less dense as compared to forest.<br>• Trees are well spaced with short trunked and spreading canopies.<br>• Most of the woodlands are deciduous. | Moderate biodiversity value |
| Other vegetation types | • This category, refers to bushland, grassland and cultivated land.<br>• The canopy cover in these categories is sparser as compared to forest and woodlands.<br>• Detail description of individual categories are given in MNRT (2011). | Less biodiversity value |

## Calculation of the measures of tree species diversity

In this study, tree species richness and Shannon diversity index were considered as the commonly used measures of tree species diversity. Calculation of the two measures of tree species diversity was done for each sample plot, and only trees with dbh ≥ 5 cm were considered. Tree species richness (S) was calculated by summing the number of tree species in each of the plot i, while Shannon diversity index (H') was calculated as:

$$H' = -\sum_{i=1}^{S} p_i \ln p_i$$

Where: $p_i$ = the proportion abundance of the $i^{th}$ species relative to the total abundance of all the species ($S$) sampled in the plot $i$

$l_n$ = the natural logarithm.

Shannon diversity index assumes that the $p_i$*'s* are population parameters where all the species in the population are known. However, in practice, the reported Shannon diversity index is only an estimator of the population and is biased because the number of species observed in the sample (e.g., field plot) is less than species in the population. But, if sampling is adequate, this bias is considered to be small (Oldeland *et al.* 2010).

All the calculations and plotting were done by using vegan library (Oksanen *et al*. 2011) in R statistical software following the procedure by Kindt and Coe (2005). Summary statistics for the tree species richness, and Shannon diversity index over the different vegetation types are presented in Table 2.

**Table 2. Number of field sample plots (n), total tree species richness, mean, standard error (Se) of tree species richness and Shannon diversity index for different vegetation types.**

| Vegetation type [a] | n | Tree species richness | | | Shannon diversity index | |
|---|---|---|---|---|---|---|
| | | Total | Mean | Se | Mean | Se |
| **Woodlands** | 387 | 239 | 5.36 | 0.12 | 1.20 | 0.02 |
| **Forest** | 40 | 90 | 6.68 | 0.36 | 1.29 | 0.07 |
| **Agriculture and other cover types** | 57 | 79 | 3.23 | 0.31 | 0.73 | 0.08 |
| **All** | 484 | 270 | 5.22 | 0.12 | 1.15 | 0.02 |

### ALS data

The ALS data were acquired in the period between 10 February and 7 March 2012. Thirty-two parallel strips with an average width of 1374 m, were systematically distributed over the study area in east-west direction. The ALS strips were spaced 5 km apart, following the NAFORMA 5×5 km grid. A Leica ALS 70 airborne laser sensor (Leica Geosystems AG, Switzerland), carried by a Cessna 404 aircraft was used for data acquisition. The measurements were acquired from an average flying altitude of approximately 1200 m above the ground, at an average ground speed of 77.2 $ms^{-1}$. The sensor was operated at a pulse repetition frequency of 193 kHz, with a scan rate of 36.5 Hz. The beam divergence was 0.28 m rad which produced an average footprint size on the ground of about 34 cm. The average pulse density was 1.8 $m^{-2}$.

The data were initially processed by the contractor (TerraTec AS, Norway), where the first step was to classify echoes into ground- and vegetation echoes. Then, a terrain model was built using the progressive Triangular Irregular Network (TIN) densification algorithm (Axelsson 2000) implemented in TerraScan software (Anon 2012). The heights above ground (the TIN surface) were calculated for all vegetation echoes by subtracting the respective xy-corresponding TIN heights from the echo height values. Up to four echoes were registered per pulse and three echo categories classified as "single", "first of many", and "last of many" were used. The "single" and "first of many" echoes were pooled into one dataset denoted as "first" echoes, and correspondingly, the "single" and "last of many" echoes were pooled into a dataset denoted as "last" echoes.

Separate height distributions were derived from the first and last ALS echoes and used to extract ALS metrics for each sample plot. A threshold of 1.3 m above ground was used to separate canopy echoes. Below this height, echoes were considered to have been reflected from shrubs, grass, or ground, i.e., non-tree objects. Height variables for first and last echoes including maximum value (MaxF and MaxL), mean value (MeanF and MeanL), coefficient of variation (CVF and CVL), and percentiles at 10% intervals labeled PF0, PF10, …, PF90 (first echoes) and PL0, PL10, …, PL90 (last echoes), were computed. Furthermore, several measures of canopy density were derived. Canopy density was calculated for 10 different vertical layers according to Næsset (2004). The height of each layer was defined as one tenth of the distance between the 95 percentile and the lowest canopy height, i.e., 1.3 m (Gobakken *et al.* 2012). Canopy densities were then computed as the proportions of ALS echoes above fraction #0, 1, …, 9 to total number of echoes, and denoted TF0 (>1.3 m), TF1, …, TF9 for the first echoes and TL0, TL1, …, TL9 for the last echoes.

### Statistical analyses

#### *Outline of the analyses*

Parametric and non-parametric methods were used to develop statistical relationship between the measures of tree species diversity and the ALS metrics. Specifically, linear mixed effects models (LMMs) and the *k*-NN approach were used. The analyses were performed based on the following steps below:

1. Explanatory variables for predicting each of the measures of tree species diversity (i.e., tree species richness and Shannon diversity index) were selected separately for fitting LMMs.
2. Separate LMMs for tree species richness and Shannon diversity index were fitted.
3. Multivariate approach was then used for selecting explanatory variables and

predicting measures of tree species diversity using *k*-NN techniques.

4. Both of the two methods were evaluated and compared using measures of reliability derived from the cross validation.
5. Effects of vegetation types on the prediction accuracy of the measures of tree species diversity for each of the method was assessed.

***Parametric method***

*Model development (LMMs)*

When modelling measures of tree species diversity using ALS data previous studies (e.g., Fricker *et al*. 2015, Hernández-Stefanoni *et al*. 2014, Wolf *et al*. 2012) have mostly used ordinary least square regression (OLS). However, in this study the sampling design characteristics employed by NAFORMA imposes a hierarchical data structure where the field plots are nested within the clusters. Theoretically this may induce a spatial dependency among the plots measured from the same cluster, and thus it is likely that the basic assumptions of uncorrelated error terms might not hold. LMM in this case was potentially an ideal tool for development of predictive models that account for dependence of the plots within the clusters, but also for ensuring that the modelling procedure adheres to the sampling design. LMM essentially consists of two main parts i.e., a fixed effect- and random effect part. The fixed effects are common to all subjects, while random effect parameter is specific to each subject (Pinheiro and Bates 2000).

Model development for each of the measures of tree species diversity started with variable selection, i.e., separate explanatory variables were selected for richness and Shannon diversity index. In both cases a subset regression technique was used for variable selection. Specifically, reg-subsets function implemented in the leaps Package of the R software (Team 2014), was used. The model statistics used to determine the best subsets was Bayesian Information Criterion (BIC). The selection of the variables was limited to the best combinations of four or fewer variables in order to avoid multicollinearity among candidate predictors. Variance inflation factor (VIF) values for each of the parameters (βs) were computed, VIF values greater than 10 was considered as an indication of multicollinearity. Variable combinations that yielded VIF-values higher than this threshold were excluded from the model (O'Brien 2007).

To account for heteroscedasticity and non-linearity, the model for tree species richness was fitted with natural-log transformation of the response variable and non-transformed predictor variables. Such model form has been used by Hernández-Stefanoni *et al*. (2014) with good results when modelling and predicting tree species richness using ALS data. For Shannon diversity index non-transformed model was used. All the models were fitted using restricted maximum likelihood estimation (REML) procedure in lme4 package (Pinheiro *et al*. 2007) of the R software (Team 2014). Quality of the model fits (coefficient of determination) were assessed by using pseudo R-square ($R^2$) computed as the square of the Pearson correlation coefficient between observed and predicted values.

*Accuracy assessment*

In order to assess the prediction accuracy of tree species richness and Shannon diversity index when using ALS data, leave one out cross validation (LOOCV) at the cluster level was applied, to ensure that hierarchical data structure was preserved during re-sampling (i.e., leave one cluster out). The predicted values from LOOCV were corrected for biasness due to logarithmic transformation using the method suggested by Snowdon (1991). The accuracies of the prediction from the LOOCV for tree species richness and Shannon diversity index were evaluated by using relative root mean square error (RMSEcv%):

$$\mathrm{RMSE_{CV}}\% = \frac{\sqrt{\sum_{i=1}^{n}(y_i - \hat{y})^2/n}}{\bar{y}} \times 100$$

where $y_i$ and $\hat{y}$ denote observed and predicted tree species richness and Shannon diversity index respectively, for plot *i*, and $\bar{y}$ denotes their mean field observed value for all plots.

***Non-parametric method***

*k*-NN *imputation*

Non-parametric methods have gained wide acceptance in ecologically based studies given their unique ability to account for complex relationships and spatial patterns as compared to the traditional probability based approaches (Drew *et al*. 2010). Furthermore, these methods allow univariate and multivariate predictions of both continuous and categorical variables. In this study, the *k*-NN imputations method carried out with the package *yaImpute* (Crookston and Finley 2008), was used. In *k*-NN language and set up the dataset should be distinguished between reference and target sets. The population units for which observations of both response and explanatory variables are available is labeled reference set; the set of the population units for which only the explanatory variables are available is termed as the target set. In this study, the reference set contained both measures of tree species diversity (i.e., tree species richness and Shannon diversity index) and the ALS metrics, while the target set contained only the ALS metrics.

In typical *k*-NN imputation, the dependent variable for the target observation is predicted by means of finding its k nearest neighbour observations in the reference dataset and assigning the value of the variable to be the weighted averages of the values of the neighbours. Nearness of the observations is measured with the independent variables and is defined in terms of weighted Euclidean distance. The principle behind *k*-NN imputation as it is applied in this study is further explained in Eskelson *et al*. (2009), McRoberts (2012) and its use for forest parameter estimation in *yaImpute* statistical environment can be found in Hudak *et al*. (2008).

In order to identify predictor variables for predicting tree species richness and Shannon diversity index simultaneously, multivariate variable selection was firstly done by using the *VarSelection* function in the R *yaImpute* package. Model fitting and imputation were then performed by using *yai* and *impute* functions within the *yaImpute* package. In this package, any *k* number of reference observations can be selected to impute the target. The *k* values were tested from 1 to 10 and selected the value with lowest RMSEcv from cross validation. LOOCV at the cluster level was used, where one cluster at time was used as the target set while the remaining clusters were used at the reference set. The imputed values from the LOOCV were used to compute RMSEcv% as described in the equation and used to compare with LMM.

***Effect of vegetation types on the prediction accuracy***

To account for the variability in prediction accuracy that might be caused by the differences in vegetation types, LMMs that relates each of the measures of tree species diversity and ALS data, were firstly fitted with dummy variable that specify vegetation types and evaluated. Secondly, vegetation specific LMMs for each of the measures of tree species diversity were fitted and evaluated. Same procedure described above for variable selection and accuracy assessment was applied. Lastly, multivariate k-NN imputation with $k = 1$ was also applied for each of the vegetation type. RMSEcv% from the LOOCV for each of the method was calculated to assess the variability in prediction accuracy across different vegetation types.

## RESULTS

**Parametric method**

Based on the results from best subset regression predictor variables derived from both the first and last echoes were selected.

The number of predictors for the individual models were four consisting of both canopy height and canopy density metrics. For all the selected variables, the VIF values were < 10, indicating acceptable levels of multicollinearity. Similarly, the parameter estimates for LMMs fitted for each of the measures of tree species diversity (Table 3), were significantly different from zero ($p < 0.05$). The LMMs explained relatively more of the variation in tree species richness as compared to Shannon diversity index. However, results from the LOOCV indicated lower RMSEcv% values for Shannon diversity index as compared to tree species richness (Table 4).

**Table 3: Parameter estimates of LMMs for tree species richness and Shannon diversity index.**

| Tree species richness | | | Shannon diversity index | | |
|---|---|---|---|---|---|
| Predictor variables | Parameter estimates | Standard error | Predictor variables | Parameter estimates | Standard error |
| Intercept | 0.6372 | 0.0872 | Intercept | 0.4850 | 0.0788 |
| MaxF | 0.0346 | 0.0072 | MaxF | 0.0262 | 0.0064 |
| PF20 | -0.0782 | 0.0115 | PF10 | -0.0518 | 0.0123 |
| TL0 | -0.8061 | 0.1769 | TF2 | 1.4667 | 0.1894 |
| TF2 | 2.0465 | 0.2002 | TL0 | -0.7453 | 0.1662 |

PF20 = Percentiles of the first echo canopy heights for 20% (m);
TF2 = Canopy densities corresponding to the proportion of first echoes above fraction #2;
TL0 = Canopy densities corresponding to the proportion of last echoes above fraction # 0 (1.3m);
MaxF = Maximum of the canopy height distributions of the first echoes.

**Table 4. Pseudo-$R^2$ ($R^2$), absolute root mean square error (RMSE), and relative root mean square error ($RMSE_{CV}$%) from the LOOCV for predicted tree species richness and Shannon diversity index using LMMs and *k*-NN.**

| Measures of tree species diversity | LMMs | | | | *k*-NN | | |
|---|---|---|---|---|---|---|---|
| | Predictor variables[a] | $R^2$ | $RMSE_{CV}$ | $RMSE_{CV}$ (%) | Predictor variables[a] | $RMSE_{CV}$ | $RMSE_{CV}$ (%) |
| Tree species richness | MaxF, PF20, TL0, TF2 | 0.46 | 2.12 | 40.7 | TF2, TF8, PF10, PL60, TL8, PL0, PF60, TF0 PF90, CVL, PF20 | 2.15 | 41.2 |
| Shannon diversity index | MaxF, PF10, TL0, TF2 | 0.39 | 0.45 | 39.1 | TF2, TF8, PF10, PL60 TL8, PL0, PF60, TF0 PF90, CVL, PF20 | 0.46 | 40.0 |

[a] PF10, PF20, PF60, PF90 = Percentiles of the first echo canopy heights for 10%, 20%,60% and 90% (m);
PL60 = Percentiles of the last echo canopy heights for 60% (m);
TF2, TF8 = Canopy densities corresponding to the proportion of first echoes above fraction #2 and #8;
TL0, TL8, = Canopy densities corresponding to the proportion of last echoes above fraction #0 (1.3m), and #8;
MaxF = Maximum of the canopy height distributions of the first echoes;
CVL = Coefficient of variations of the last echo laser canopy heights.

### Non-parametric

Multivariate predictions for both tree species richness and Shannon diversity index were performed using *k*-NN imputation models (Table 4). A total of eleven predictor variables were selected from the multivariate variable selection procedure. The imputation for both tree species richness and Shannon diversity index resulted in into lowest values of RMSEcv% when using imputation model with 10 neighbours. The RMSEcv% for Shannon diversity index was relatively low as compared to tree species richness. Comparing the two methods (i.e., LMMs and *k*-NN), the RMSEcv% for both tree species richness and Shannon diversity index were slightly lower for the LMMs compared to the *k*-NN.

**Effect of vegetation types on prediction accuracy**

The results suggest that vegetation types affect the relationship between the measures of tree species diversity and the ALS data. In the first approach, where the model fitted with dummy variables, indicated that for both tree species richness and Shannon diversity index, the parameter estimates for the dummy variables were significantly different from zero ($p < 0.05$) (Table 5). The standard error of parameter estimates for the LMMs with dummy variables were relatively low as compared to the LMMs without dummy variables. This is further shown by the results from the LOOCV where the RMSEcv% of the LMMs with dummy variables (Table 6) is relatively small as compared to the model without dummy variable. In the second approach where vegetation specific models were fitted, the influence of vegetation was proved to affect the prediction accuracy. Different predictor variables in each of the vegetation types were selected from the best subset procedure (Table 6). The variability explained by the vegetation specific LMMs ranged from 0.34 to 0.53 for tree species richness and from 0.32 to 0.47 for Shannon diversity index (Table 6). The results from the LOOCV indicated variation in prediction accuracy across the different vegetation types. Lowest RMSEcv% for predicting tree species richness across vegetation types was obtained in woodlands, while for Shannon diversity index the lowest value was obtained in forest (Table 6). Variations in prediction accuracy as measured by the RMSEcv% among the vegetation types were also observed when using *k*-NN (Table 7).

**Table 5. Parameter estimates of LMMs for tree species richness and Shannon diversity index with dummy variables**

| Tree species richness | | | Shannon diversity index | | |
|---|---|---|---|---|---|
| Predictor[a] variables | Parameter estimates | Standard error | Predictor variables | Parameter estimates | Standard error |
| Intercept | 0.3377 | 0.0971 | Intercept | 0.2798 | 0.0909 |
| MaxF | 0.0315 | 0.0070 | MaxF | 0.0234 | 0.0063 |
| PF20 | -0.0760 | 0.0111 | PF10 | -0.0492 | 0.0120 |
| TL0 | -0.7082 | 0.1755 | TF2 | 1.3551 | 0.1859 |
| TF2 | 1.8926 | 0.1918 | TL0 | -0.6592 | 0.1701 |
| Forest | 0.4229 | 0.1183 | Forest | 0.2682 | 0.1117 |
| Woodlands | 0.4283 | 0.0796 | Woodlands | 0.2982 | 0.0748 |

[a] PF20 = Percentiles of the first echo canopy heights for 20% (m);
TF2 = Canopy densities corresponding to the proportion of first echoes above fraction #2;
TL0 = Canopy densities corresponding to the proportion of last echoes above fraction #0 (1.3m);
MaxF = Maximum of the canopy height distributions of the first echoes;
Forest and woodlands = Vegetation types according to MNRT (2011)

**Table 6. Predictor variables, number of field sample plots (n), Pseudo-$R^2$ ($R^2$), absolute root mean square error ($RMSE_{CV}$), and relative root mean square error (RMSEcv%) of LMM for tree species richness and Shannon diversity index across different vegetation types**

| Measures of tree species diversity | Vegetation type[a] | Predictor variables[b] | n | $R^2$ | RMSEcv | RMSEcv (%) |
|---|---|---|---|---|---|---|
| Tree species richness | Woodlands | MaxF, PF20, TF3, TL0 | 387 | 0.41 | 2.08 | 38.9 |
| | Forest | PF20, TF9, PL50 | 40 | 0.34 | 2.79 | 41.8 |
| | Other cover types | MaxF, MeanF, TF1, TL0 | 57 | 0.36 | 1.84 | 57.0 |
| | All | MaxF, PF20, TL0, TF2, DUMMY | 484 | 0.45 | 2.10 | 40.2 |
| Shannon diversity index | Woodlands | CVF, TF4, PF40 | 387 | 0.32 | 0.42 | 35.0 |
| | Forest | PF40, TF8, MaxL, TL0 | 40 | 0.47 | 0.35 | 27.0 |
| | other cover types | MaxL, TF2 | 57 | 0.39 | 0.53 | 73.0 |
| | All | MaxF, PF10, TL0, TF2, DUMMY | 484 | 0.38 | 0.44 | 38.0 |

[a] Vegetation types according to MNRT (2011).
[b] PF10, PF20, PF40 = Percentiles of the first echo canopy heights for 10%, 20%, 40% (m);
PL50 = Percentiles of the last echo canopy heights for 50% (m);
TF1, TF2, TF3, TF4, TF8, TF9 = Canopy densities corresponding to the proportion of first echoes above fraction #1, # 2, #3, #4, #8, and #9;
TL0 = Canopy densities corresponding to the proportion of last echoes above fraction #0 (1.3m);
MeanF and MeanL = Arithmetic mean of first and last echo laser canopy heights (m);
MaxF and MaxL = Maximum of first and last echo laser canopy heights (m);
CVF and CVL = Coefficient of variations for the first and last echo respectively;
DUMMY= vegetation types (0 = Other cover types, 1 = Forest, 2 = Woodlands = 3).

**Table 7. Predictor variables, number of field sample plots (n), absolute root mean square error (RMSEcv), and relative root mean square error (RMSEcv%) of the multivariate *k*-NN imputations for tree species richness and Shannon diversity index across different vegetation types.**

| Measures of tree species diversity | Vegetation type [a] | Predictor variables | n | $RMSE_{CV}$ | $RMSE_{CV}$ (%) |
|---|---|---|---|---|---|
| Tree species richness | Woodlands | TF2, PL40, PF0, PF90, TF4 | 387 | 2.84 | 53.0 |
| | Forest | PL70, PF40, TL9, TL0, PF60, TF8, TF6, MaxL, TL2, PF90, TL3, PL10 | 40 | 2.99 | 44.8 |
| | Other cover types | TL0, TF1, PF80, PL60, YTF8 | 57 | 2.93 | 90.7 |
| Shannon diversity index | Woodlands | TF2, PL40, PF0, PF90, TF4 | 387 | 0.59 | 49.2 |
| | Forest | PL70, PF40, TL9, TL0, PF60, TF8, TF6, MaxL, TL2, PF90, TL3, PL10 | 40 | 0.50 | 39.0 |
| | Other cover types | TL0, TF1, PF80, PL60, TF8 | 57 | 0.72 | 99.0 |

[a] Vegetation types according to MNRT (2011).
[b] PF10, PF40, PF90, PF80, PF90 = Percentiles of the first echo canopy heights for 10%, 40%, 90%, 80%, and 90% (m);
PL10, PL40, PL60, PL70 = Percentiles of the last echo canopy heights for 10%, 40%, 60, and 70% (m);
TF1, TF2, TF4, TF8, TF9 = Canopy densities corresponding to the proportion of first echoes above fraction #1, # 2, # 3, #4, #8, and #9;
TL0, TL2, TL3, and TL9 = Canopy densities corresponding to the proportion of last echoes above fraction #0 (1.3m), #2, #3, and #9;
MaxL = Maximum of last echo laser canopy heights (m).

## DISCUSSION

The main objective of the study was to examine the usefulness of ALS data for modelling and predicting measures of tree species diversity in miombo woodlands of Tanzania. More specifically the performance of parametric and non-parametric methods for modelling and predicting tree species richness and Shannon diversity index using ALS data, were tested and evaluated. Overall, the study has showed that, ALS data can be used for modelling and predicting tree species richness and Shannon diversity index. Both tree species richness and Shannon diversity index are often used as measures of tree species diversity in remote sensing literature (e.g., Leutner *et al*. 2012, Laurin *et al*. 2014).

Based on the results from LMMs, more variations were explained by the ALS data when predicting tree species richness as compared to Shannon diversity index. Similar findings have been reported by Leutner *et al*. (2012) who assessed the potential of ALS and hyperspectral data for predicting plant species richness and Shannon diversity index in the temperate forests of Germany. The reason for this difference is unclear and opens questions for further investigations. However, irrespective of this difference, in an ecological context, Shannon diversity index is considered as a more appropriate measure of tree species diversity as compared to tree species richness, because it has the attributes of both richness and evenness. Besides that, Shannon diversity index is considered to be a better indicator for describing forest structure when using remotely sensed data as compared to tree species richness (Foody and Cutler 2003). Use of Shannon diversity index in remote sensing based studies has also been suggested by others using both ALS and other remotely sensed data (e.g., Oldeland *et al*. 2010, Laurin *et al*. 2014)

Assessment of the performance of the two prediction methods, i.e., LMMs and the *k*-NN, has shown that both of the methods can reliably be used for predicting the measures of tree species diversity. The choice between them may therefore depend on the objective. For example, for making spatially consistent predictions of the multiple measures of tree species diversity, it is more reasonable to consider the use of *k*-NN. This is in-line with earlier studies that reported the strength of the *k*-NN over the parametric based methods (e.g., Moeur and Stage 1995, McRoberts *et al*. 2002). LMM being a parametric method is more applicable if the interest is to examine the relationship between individual measures of tree diversity and the ALS data, which is important for deriving ecological interpretation of relationship between the measures of tree species diversity and the respective ALS metrics.

The assessments of the effects of vegetation types on the prediction accuracy of both LMMs and *k*-NN have shown that vegetation types had significant impact on the prediction accuracy of the measures of tress species diversity when using ALS data. The parameter estimates of the dummy variables representing vegetation types were significantly different from zero when incorporated in the LMMs for predicting tree species richness and Shannon diversity index. This indicates that different vegetation types have different tree species diversity and spatial structure which entirely affect the relationship between the measures of tree species diversity and the ALS measurement. The importance of accounting for the effects of the vegetation types in the LMMs was also demonstrated by the results from LOOCV where the LMMs with dummy variables turned out to have lower RMSEcv% as compared to the LMMs without dummy variable. The variability imposed by different vegetation types was further indicated by different prediction accuracies when testing and evaluating both LMMs and *k*-NN across different vegetation types. As compared to LMMs, the RMSEcv% values for the *k*-NN were relatively higher across different vegetation types. This might be attributed to the relatively low number of observations, since imputations methods are generally more sensitive to the number of observations as compared to parametric methods. The variability in prediction accuracy obtained across vegetation types, implies that it is important to control for the effect of vegetation types when making large scale estimation by post stratifying the area into different vegetation types.

Generally, the findings of this study in terms of model quality criterions such as $R^2$ and RMSEcv% are in accordance with other published studies that have attempted to use ALS data for predicting measures of tree species diversity in the tropical forests (Fricker *et al.* 2015, Wolf *et al.* 2012). For example, Wolf *et al.* (2012) reported $R^2$ of 0.48 when predicting tree species richness in the Neotropical forests of Panama. However, some of the studies from the tropical forests have reported relatively better results (e.g., Hernández-Stefanoni *et al*. 2014, Laurin *et al*. 2014) than what is obtained from this study. This is not surprising since in most cases the predictive ability of the remotely sensed data varies with the ecosystem and biogeographical context (Camathias *et al.* 2013). Furthermore, the differences in factors such as number of species, field design, data characteristics, and scales of reporting may attribute to the differences in prediction accuracy among the studies. It is therefore, important to mention some possible sources of errors that might have attributed to relatively low prediction accuracy in this study. The plot size used was relatively small compared to remote sensing based studies in ecology, where one hectare plot size has commonly been used (e.g., Asner and Mascaro 2014, Asner *et al*. 2012, Mascaro *et al*. 2011). In ALS-based studies, smaller plots have been reported to be a source of model errors due to the poor overlap between the field- and ALS-data. Thus, larger plots would most likely have helped in reducing the residual errors of the models. The effect of field plot size on ALS-based inventories is extensively analyzed in AGB related studies in the tropical forests (e.g., Mascaro *et al*. 2012, Mauya *et al*. 2015) and same pattern in tree species diversity related studies would be expected. For example, Hernández-Stefanoni *et al.* (2014) compared the prediction accuracy of tree species richness at different plot size ranging from 400 $m^2$ to 2200 $m^2$ and indicated that the $R^2$ increased from 0.32 to 0.67. Another source of error that might have affected the accuracy of the model, is the NAFORMA subsampling procedure illustrated in Fig. 3. Since not all the trees within the plots were considered during measurement, we may expect a slight mismatch with what has been captured by the ALS data. Furthermore, with the NAFORMA subsampling procedure, it is possible that there is under-estimation of tree species diversity, particularly when using Shannon diversity index, which theoretically assumes that all the species are represented in the sample.

While this study focused on the use of ALS alone, future studies should attempt to combine ALS with other remotely sensed data that have large spatial coverage such as hyperspectral images. This will likely also improve the prediction accuracy of the models, as the two types of sensors will complement each other, especially when modelling multi-layered forests (e.g., Laurin *et al.* 2014; Dalponte *et al.* 2008). Hyperspectral data are useful in providing detailed surface reflectance characteristics, while ALS data adds detailed three-dimensional positioning information. Moreover, this study focused on the commonly used ALS metrics that have widely been proved efficient in the prediction of various biophysical parameters including tree species diversity. However, further studies should be devoted in exploring other types of metrics such as texture based metrics that might improve the prediction accuracy of the models.

## CONCLUSION

In this study the potential of using ALS data for predicting measures of tree species diversity in the miombo woodlands of southern Tanzania has been analyzed. LMMs and k-NN imputations were tested and evaluated, and the results suggested that both approaches are promising tools for modelling and predicting measures of tree species diversity. The prediction accuracies of tree species richness and Shannon diversity index were affected by the differences in structural properties attributed by the vegetation types. Fusion of ALS data with other remote sensing techniques such as hyperspectral dataset is considered as an avenue for future research that might improve the prediction accuracy of trees species richness and Shannon diversity index.

## ACKNOWLEDGEMENTS

The financial support for this research was provided by the project entitled “Enhancing the Measuring, Reporting and Verification (MRV) of forests in Tanzania through the application of advanced remote sensing techniques”. I am also grateful to the field team in Tanzania for the field data collection and Terratec, Norway, for collecting and processing of the ALS data.

## REFERENCES


Anon. 2012. TerraScan user's guide. Jyvaskyla, Finland: Terrasolid Oy. Available from http://www.terrasolid.com/download/tscan.pdf. 62. Accessed 23rd Aug 2014.

Asner, G.P. & Mascaro, J. 2014. Mapping tropical forest carbon: Calibrating plot estimates to a simple LiDAR metric. Remote Sensing of Environment, 140, 614-624.

Asner, G.P., Mascaro, J., Muller-Landau, H.C., Vieilledent, G., Vaudry, R. & Rasamoelina, M. 2012. A universal airborne LiDAR approach for tropical forest carbon mapping. Oecologia, 168 (4), 1147-1160.

Axelsson, P. 2000. DEM generation from laser scanner data using adaptive TIN models. International Archives of Photogrammetry and Remote Sensing, 33 (B4/1; PART 4), 111-118.

Barna, M. & Bosela, M. 2015. Tree species diversity change in natural regeneration of a beech forest under different management. Forest Ecology and Management, 342 (0), 93-102.

Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W. & Courchamp, F. 2012. Impacts of climate change on the future of biodiversity. Ecology letters, 15 (4), 365-377.

Borah, M., Das, D., Kalita, J., Deka Boruah, H.P., Phukan, B. & Neog, B. 2015. Tree species composition, biomass and carbon stocks in two tropical forest of Assam. Biomass and Bioenergy, 78 (0), 25-35.

Boutin, S., Haughland, D.L., Schieck, J., Herbers, J. & Bayne, E. 2009. A new approach to forest biodiversity monitoring in Canada. Forest Ecology and Management, 258, S168-S175.

Camathias, L., Bergamini, A., Küchler, M., Stofer, S. & Baltensweiler, A. 2013. High-resolution remote sensing data improves models of species richness. Applied Vegetation Science, 16 (4), 539-551.

Ceballos, A., Hernández, J., Corvalán, P. & Galleguillos, M. 2015. Comparison of Airborne LiDAR and Satellite Hyperspectral Remote Sensing to Estimate Vascular Plant Richness in Deciduous Mediterranean Forests of Central Chile. Remote Sensing, 7 (3), 2692-2714.

Chandra, A. & Idrisova, A. 2011. Convention on Biological Diversity: a review of national challenges and opportunities for implementation. Biodiversity and Conservation, 20 (14), 3295-3316.

Chao, A. & Shen, T.-J. 2003. Nonparametric estimation of Shannon’s index of diversity when there are unseen species in sample. Environmental and ecological statistics, 10 (4), 429-443.

Chiarucci, A. & Palmer, M. 2006. The inventory and estimation of plant species richness. Encyclopedia of life support systems, developed under the auspices of the UNESCO. Encyclopedia of Life Support Systems Publishers, Oxford.

Chidumayo, E.N. 2013. Forest degradation and recovery in a miombo woodland landscape in Zambia: 22 years of observations on permanent sample plots. Forest Ecology and Management, 291 (0), 154-161.

Chirici, G., McRoberts, R.E., Winter, S., Bertini, R., Brändli, U.-B. & Asensio, I.A. 2012. National forest inventory contributions to forest biodiversity monitoring. Forest Science, 58 (3), 257-268.

Colwell, R.K. & Coddington, J.A. 1994. Estimating terrestrial biodiversity through extrapolation. Philosophical Transactions of the Royal Society B: Biological Sciences, 345 (1311), 101-118.

Crookston, N.L. & Finley, A.O. 2008. Yaimpute: An r package for knn imputation. Journal of Statistical Software, 23 (10), 1-16.

Dalponte, M., Bruzzone, L. & Gianelle, D. 2008. Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas. Geoscience and Remote Sensing, IEEE Transactions on, 46 (5), 1416-1427.

Drew, C.A., Wiersma, Y.F. & Huettmann, F. 2010. Predictive species and habitat modeling in landscape ecology: concepts and applications. Springer Science & Business Media.

Ehara, M., Hyakumura, K. & Yokota, Y. 2014. REDD+ initiatives for safeguarding biodiversity and ecosystem services: harmonizing sets of standards for national application. Journal of Forest Research, 19 (5), 427-436.

Eskelson, B.N., Temesgen, H., Lemay, V., Barrett, T.M., Crookston, N.L. & Hudak, A.T. 2009. The roles of nearest neighbor methods in imputing missing data in forest inventory and monitoring databases. Scandinavian Journal of Forest Research, 24 (3), 235-246.

Foody, G.M. & Cutler, M.E. 2003. Tree biodiversity in protected and logged Bornean tropical rain forests and its measurement by satellite remote sensing. Journal of Biogeography, 30 (7), 1053-1066.

Fricker, G.A., Wolf, J.A., Saatchi, S.S. & Gillespie, T.W. 2015. Predicting spatial variations of tree species richness in tropical forests from high resolution remote sensing. Ecological Applications.

Gardner, T.A., Barlow, J., Chazdon, R., Ewers, R.M., Harvey, C.A. & Peres, C.A. 2009. Prospects for tropical forest biodiversity in a human modified world. Ecology Letters, 12 (6), 561-582.

Gardner, T.A., Burgess, N.D., Aguilar-Amuchastegui, N., Barlow, J., Berenguer, E. & Clements, T. 2012. A framework for integrating biodiversity concerns into national REDD+ programmes. Biological Conservation, 154, 61-71.

Gobakken, T., Næsset, E., Nelson, R., Bollandsås, O.M., Gregoire, T.G. & Ståhl, G. 2012. Estimating biomass in Hedmark County, Norway using national forest inventory field plots and airborne laser scanning. Remote Sensing of Environment, 123, 443-456.

Hernández-Stefanoni, J.L., Dupuy, J.M., Johnson, K.D., Birdsey, R., Tun-Dzul, F. & Peduzzi, A. 2014. Improving species diversity and biomass estimates of tropical dry forests using airborne LiDAR. Remote Sensing, 6 (6), 4741-4763.

Hudak, A.T., Crookston, N.L., Evans, J.S., Hall, D.E. & Falkowski, M.J. 2008 Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. Remote Sensing of Environment, 112 (5), 2232-2245.

Imai, N., Tanaka, A., Samejima, H., Sugau, J.B., Pereira, J.T. & Titin, J. 2014. Tree community composition as an indicator in biodiversity monitoring of REDD+. Forest Ecology and Management, 313 (0), 169-179.

Kindt, R. & Coe, R. 2005. Tree diversity analysis: a manual and software for common statistical methods for ecological and biodiversity studies. World Agroforestry Centre.

Latifi, H., Fassnacht, F.E., Hartig, F., Berger, C., Hernández, J. & Corvalán, P. 2015. Stratified aboveground forest biomass estimation by remote sensing data. International Journal of Applied Earth Observation and Geoinformation, 38 (0), 229-241.

Laurin, G.V., Chan, J.C.-W., Chen, Q., Lindsell, J.A., Coomes, D.A. & Guerriero, L. 2014. Biodiversity mapping in a tropical west african forest with airborne hyperspectral data. PloS one, 9 (6), e97910.

LDC. 2014. Liwale Distric Socio-Economic Profile. Unpublished Report, p. 56.

Leutner, B.F., Reineking, B., Müller, J., Bachmann, M., Beierkuhnlein, C. & Dech, S. 2012. Modelling forest α-diversity and floristic composition—On the added value of LiDAR plus hyperspectral remote sensing. Remote Sensing, 4 (9), 2818-2845.

Lindberg, E., Roberge, J.-M., Johansson, T. & Hjältén, J. 2015. Can Airborne Laser Scanning (ALS) and Forest Estimates Derived from Satellite Images Be Used to Predict Abundance and Species Richness of Birds and Beetles in Boreal Forest? Remote Sensing, 7 (4), 4233-4252.

Magurran, A.E. 1988. Ecological diversity and its measurement. Princeton university press Princeton.

Mascaro, J., Asner, G.P., Dent, D.H., DeWalt, S.J. & Denslow, J.S. 2012. Scale-dependence of aboveground carbon accumulation in secondary forests of Panama: A test of the intermediate peak hypothesis. Forest Ecology and Management, 276, 62-70.

Mascaro, J., Detto, M., Asner, G.P. & Muller-Landau, H.C. 2011. Evaluating uncertainty in mapping forest carbon with airborne LiDAR. Remote Sensing of Environment, 115 (12), 3770-3774.

Matthews, R.B. & van Noordwijk, M. 2014. From euphoria to reality on efforts to reduce emissions from deforestation and forest degradation (REDD+). Mitigation and Adaptation Strategies for Global Change, 19 (6), 615-620.

Mauya, E.W., Ene, L.T., Bollandsås, O.M., Gobakken, T., Næsset, E. & Malimbwi, R.E. 2015. Modelling aboveground forest biomass using airborne laser scanner data in the miombo woodlands of Tanzania. Carbon balance and management, 10 (1), 28.

Mauya, E.W., Hansen, Endre Hofstad, Gobakken, T., Bollandsås, O.M., Malimbwi, R. & Næsset, E. 2015. Effects of field plot size on prediction accuracy of aboveground biomass in airborne laser scanning-assisted inventories in tropical rain forests of Tanzania. Carbon balance and management, 10 (10), 14.

McRoberts, R.E. 2012. Estimating forest attribute parameters for small areas using nearest neighbors techniques. Forest Ecology and Management, 272, 3-12.

McRoberts, R.E. & Meneguzzo, D.M., 2007. Estimating tree species richness from forest inventory plot data. In: McRoberts, Ronald E.; Reams, Gregory A.; Van Deusen, Paul C.; McWilliams, William H., eds. Proceedings of the seventh annual forest inventory and analysis symposium; October 3-6, 2005; Portland, ME. Gen. Tech. Rep. WO-77. Washington, DC: US Department of Agriculture, Forest Service: 275-279. (Vol. 77)

McRoberts, R.E., Næsset, E. and Gobakken, T. 2015. Optimizing the k-Nearest Neighbors technique for estimating forest aboveground biomass using airborne laser scanning data. Remote Sensing of Environment, 163, 13-22.

McRoberts, R.E., Nelson, M.D. & Wendt, D.G. 2002. Stratified estimation of forest area using satellite imagery, inventory data, and the k-Nearest Neighbors technique. Remote Sensing of Environment, 82 (2), 457-468.

MNRT. 2011. National Forestry Resources Monitoring and Assessment of Tanzania (NAFORMA). Field Manual Biophysical survey. Forestry and Beekeeping Division, Dar es Salaam, Tanzania.

MNRT. 2015. National Forest Resources Monitoring and Assessment of Tanzania Mainland (NAFORMA). Main Results. Unpublished report, p. 106.

Moeur, M. & Stage, A.R. 1995. Most similar neighbor: an improved sampling inference procedure for natural resource planning. Forest Science, 41 (2), 337-359.

Mooney, H.A., Cropper, A. & Reid, W. 2004.The millennium ecosystem assessment: what is it all about? Trends in Ecology & Evolution, 19 (5), 221-224.

Müller, J. & Brandl, R. 2009. Assessing biodiversity by remote sensing in mountainous terrain: the potential of LiDAR to predict forest beetle assemblages. Journal of Applied Ecology, 46 (4), 897-905.

Müller, J. & Vierling, K. 2014. Assessing Biodiversity by Airborne Laser Scanning. In Forestry Applications of Airborne Laser Scanning, Springer, pp. 357-374.

Nadeau, M.B. & Sullivan, T.P. 2015. Relationships between Plant Biodiversity and Soil Fertility in a Mature Tropical Forest, Costa Rica. International Journal of Forestry Research, 2015.

Næsset, E. 2004. Practical large-scale forest stand inventory using a small-footprint airborne scanning laser. Scandinavian Journal of Forest Research, 19 (2), 164-179.

O'brien, R.M. 2007. A caution regarding rules of thumb for variance inflation factors. Quality & quantity, 41 (5), 673-690.

Oksanen, J., Blanchet, F., Kindt, R., Legendre, P., Minchin, P. & O'Hara, R. 2011. Vegan: community ecology package version 2.0-2. R CRAN package.

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

Pereira, H.M. & Cooper, H.D. 2006. Towards the global monitoring of biodiversity change. Trends in Ecology & Evolution, 21 (3), 123-129.

Pereira, H.M., Ferrier, S., Walters, M., Geller, G., Jongman, R. & Scholes, R. 2013. Essential biodiversity variables. Science, 339 (6117), 277-278.

Pinheiro, J.C. and Bates, D.M. 2000. Mixed-effects models in S and S-PLUS. Springer.

Roe, S., Streck, C., Pritchard, L. & Costenbader, J. 2013. Safeguards in REDD+ and forest carbon standards: A review of social, environmental and procedural concepts and application. Climate Focus.

Schmidt, M., Veldkamp, E. and Corre, M.D. 2015. Tree species diversity effects on productivity, soil nutrient availability and nutrient response efficiency in a temperate deciduous forest. Forest Ecology and Management, 338 (0), 114-123.

Shirima, D.D., Totland, Ø., Munishi, P.K.T. & Moe, S.R. 2015. Relationships between tree species richness, evenness and aboveground carbon storage in montane forests and miombo woodlands of Tanzania. Basic and Applied Ecology, 16 (3), 239-249.

Simonson, W.D., Allen, H.D. & Coomes, D.A. 2012. Use of an airborne lidar system to model plant species composition and diversity of Mediterranean oak forests. Conservation Biology, 26 (5), 840-850.

Snowdon, P. 1991. A ratio estimator for bias correction in logarithmic regressions. Canadian Journal of Forest Research, 21 (5), 720-724.

Swatantran, A., Dubayah, R., Roberts, D., Hofton, M. & Blair, J.B. 2011. Mapping biomass and stress in the Sierra Nevada using lidar and hyperspectral data fusion. Remote Sensing of Environment, 115 (11), 2917-2930.

Team, R.C. 2014. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2012. Open access available at: http://cran.r-project. org.

Tomppo, E., Malimbwi, R., Katila, M., Mäkisara, K., Henttonen, H.M. & Chamuya, N. 2014. A sampling design for a large area forest inventory: case Tanzania. Canadian Journal of Forest Research, 44 (8), 931-948.

Vanderhaegen, K., Verbist, B., Hundera, K. & Muys, B. 2015. REALU vs. REDD+: Carbon and biodiversity in the Afromontane landscapes of SW Ethiopia. Forest Ecology and Management, 343 (0), 22-33.

Vierling, K., Bässler, C., Brandl, R., Vierling, L., Weiß, I. & Müller, J. 2011. Spinning a laser web: predicting spider distributions using LiDAR. Ecological Applications, 21 (2), 577-588.

Vogeler, J.C., Hudak, A.T., Vierling, L.A., Evans, J., Green, P. & Vierling, K.T. 2014. Terrain and vegetation structural influences on local avian species richness in two mixed-conifer forests. Remote Sensing of Environment, 147, 13-22.

Winter, S., McRoberts, R.E., Chirici, G., Bastrup-Birk, A., Rondeux, J. & Brändli, U.B. 2011. The need for harmonized estimates of forest biodiversity indicators. In National Forest Inventories: Contributions to Forest Biodiversity Assessments, Springer, pp. 1-23.

Wolf, J.A., Fricker, G.A., Meyer, V., Hubbell, S.P., Gillespie, T.W. & Saatchi, S.S. 2012. Plant species richness is associated with canopy height and topography in a neotropical forest. Remote Sensing, 4 (12), 4010-4021.