Tanzania Journal of Science 50(1): 19-35, 2024
ISSN 0856-1761, e-ISSN 2507-7961

Modelling and Mapping of Forest Biodiversity Indicators Using Sentinel-2 and PlanetScope Remotely Sensed Data

Diana L. Tesha[1]*, Ernest W. Mauya[2] and Sami D. Madundo[2]
[1]*Department of Ecosystems and Conservation, Sokoine University of Agriculture, P.O. Box 3010, Morogoro, Tanzania*
[2]*Department of Forest Engineering and Wood Sciences, Sokoine University of Agriculture, P.O. Box 3014, Morogoro, Tanzania*
*Corresponding author, email: diana.tesha@sua.ac.tz
Received 17 Nov 2023, Revised 23 Mar 2024, Accepted 29 Mar 2024, Publ. 30 Apr 2024
https://dx.doi.org/10.4314/tjs.v50i1.2

Abstract

Tropical montane forests harbour exceptionally high biodiversity yet face severe threats from human activities. Assessing forest biodiversity over large areas is crucial yet extremely challenging. Remote sensing provides an efficient monitoring solution, but few studies have focused on Tanzania's diverse, montane forests. We collected field data on tree species composition within 159 plots across montane forests in Tanzania's West Usambara region. We calculated species richness, evenness, and Shannon diversity index as indicators of tree diversity. Using Sentinel-2 and PlanetScope satellite imagery, we derived spectral, textural, and vegetation index predictors to model these indices via generalized additive models and extreme gradient boosting. PlanetScope-based XGBoost models performed best, explaining 19.7% of variation in Shannon diversity. Incorporating textural predictors further improved model accuracy. Despite inherent challenges in modelling complex tropical forests, our findings demonstrate promising potential of Sentinel-2 and PlanetScope for regional biodiversity monitoring where field surveys are limited. Further research could enhance these initial results by leveraging higher resolution data and increasing field sampling for effective monitoring of tropical biodiversity.

Keywords: Tree species richness; montane forests; Eastern Arc Mountains; GAM, XGBoost

Introduction

Tropical montane forests including the West Usambara Mountains of Tanzania offer various ecosystem services while harbouring exceptionally high biodiversity. They face severe threats from human activities like agricultural expansion and illegal logging (Arroyo-Rodríguez et al., 2020). In many jurisdictions, public forest authorities are requested to monitor biodiversity and report their management efforts to maintain or improve biodiversity to various bodies (Storch et al. 2023). Such bodies include the Division for Sustainable Development Goals (DSDG), the IUCN's Post-2020 Global Biodiversity Framework (Strategic Plan for Biodiversity 2011–2020), and the United Nations Convention on Biological Diversity (Storch et al. 2023).

Reaching the conservation and protection targets require accurate information on the state of the indicators of forest biodiversity such as species richness, evenness, and diversity at different geographical scales. This is challenging in most tropical montane forest environments, given the large coverage associated with many tree species and dense forests located in less accessible and difficult terrain areas (Beyene et al. 2020). Due to the difficulties, field based assessments are rather difficult, cover limited areas (Mauya et al. 2015) and expensive (Corte et al. 2020, Goodbody et al. 2019) which calls the need for integrating them with remote sensing approaches.

Freely available Sentinel-2 and recently released high-resolution PlanetScope satellite data provide valuable opportunities for forest biodiversity monitoring. Sentinel-2 offers open-access optical imagery with enhanced spectral, spatial, and temporal resolution compared to past sensors (Drusch et al., 2012). PlanetScope likewise captures key spectral bands at very high 4.77 m resolution and monthly revisit rate (Poortinga et al., 2021). Combined capabilities enable modelling ecosystem processes and mapping indicators like tree diversity for conservation support (Potapov et al. 2008, Wang and Gamon 2019).

Extreme gradient boosting (XGBoost) and generalized additive models (GAMs) are two common modelling techniques for leveraging satellite remote sensing data to assess biodiversity. XGBoost is an ensemble method combining multiple decision trees with gradient boosting for high performance prediction. It can handle sparse, complex ecological data and scale efficiently to massive datasets (Schratz et al. 2021). GAMs provide flexible nonlinear regression using data-driven smoothing splines, easily incorporating nonlinear relationships common in ecological systems (Wang and Gamon 2019). When derived vegetation indices, spatial textures, and spectral data serve as predictor variables, these methods show utility for modelling patterns in tree diversity, species composition, habitat heterogeneity and other indicators relevant to conservation planning.

While remote sensing has demonstrated potential for assessing biodiversity in tropical forests globally (Abbas et al. 2020), few studies have focused specifically on the diverse, montane forests of Tanzania. In this study, we demonstrate the applicability of Sentinel-2's spectral resolution and PlanetScope's high revisit frequency for mapping key biodiversity indicators like tree species diversity and habitat heterogeneity across extensive, challenging terrain in the West Usambara Mountains. Therefore, this study specifically aimed to (i) quantify the indicators of forest biodiversity (i.e., species richness, evenness, and Shannon diversity) of the study site, (ii) model and predict the biodiversity indicators using semi-parametric and non-parametric models for Sentinel-2 and PlanetScope data, (iii) create a spatial map for each forest biodiversity indicator, and (iv) assess the gain in precision of each remote sensing data (relative efficiency) compared to that of the field-based inventory alone.

Materials and Methods

Study area

Two forests, Shagayu Forest Reserve (SFR) and Magamba Nature Forest Reserve (MNFR) were selected from the West Usambara Montane forest block. This block is part of a collection of isolated mountains known as the Eastern Arc Mountains (EAMs), which span from south-eastern Kenya to south-central Tanzania (Figure 1). There are many protected forests within these blocks, including nature and forest reserves, which are recognized as having extreme global biological importance (Burgess et al. 2007). The SFR is positioned at 4° 31' 0" S and 38° 16' 59" E and has an estimated elevation between 1340 and 2150m above sea level. In comparison, the MNFR is positioned at 4°40' S and 38°15' E, and its altitude varies between 1650 and 2300 m above sea level. West Usambara mountain forests are acknowledged for their copious amounts of precipitation that foster a diverse range of flora and fauna. This area exhibits a bimodal rainfall distribution pattern consisting of two annual rainy seasons. The long rainy season typically starts in March and lasts until May, whereas the short rainy season occurs between October and December (Lovett 1996).

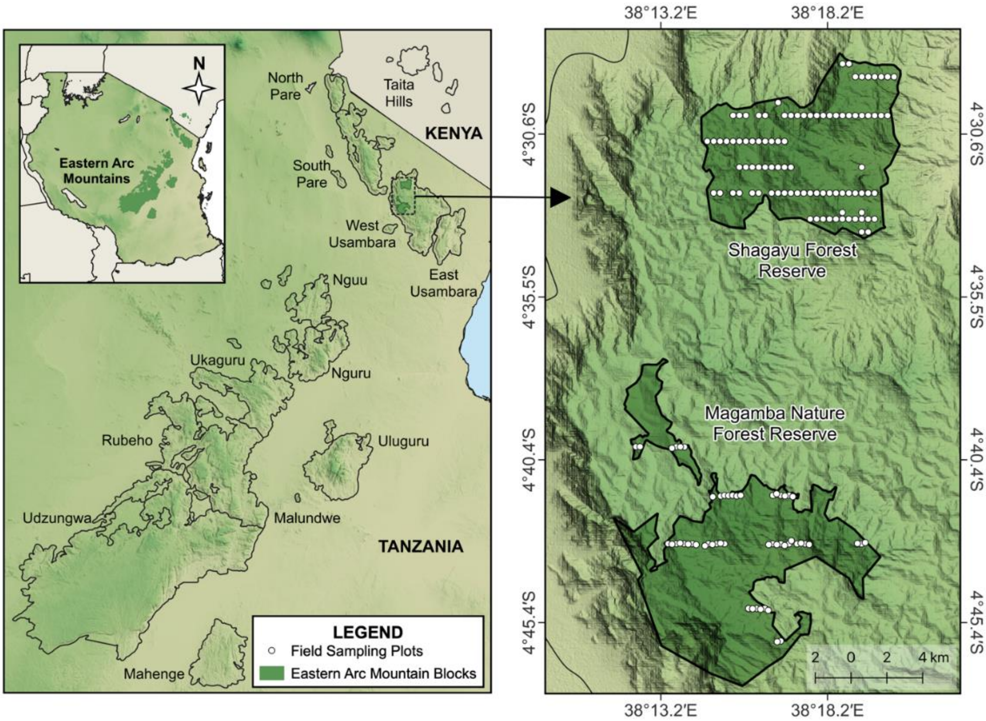

Figure 1: Location of the study forests, in the Eastern Arc Mountains of Tanzania.

Sampling design and data collection

A two-phase, systematic sampling design was used in this study. Grids (225 × 450 m in the MNFR and 350 × 700 m in the SFR) were established during the first phase, with each intersection being a sampling plot. During field expedition, second-phase plots were selected for accessibility.

Field data measurement

A total of 159 circular field sampling plots (radius = 15 m) were established across both forests (MNFR = 55; SFR = 104). In each plot, individual trees with a diameter at breast height (DBH) ≥ 5 cm were recorded and identified to the species level. The geographical location and elevation were also recorded using a handheld GPS (Garmin 78). Finally, three diversity indices that considered the total number of species and their abundances were computed from the tree species information using the following equations:

(i) Species richness (S), was determined as the total number of unique tree species recorded per sampling plot;

(ii) Pileou's evenness (J), also known as the equitability index, measures the evenness of individual tree species distribution among taxa and was calculated as the ratio between the Shannon diversity index and the logarithm of the species richness (eqn. 1).

$$J = {}^{H}\!/_{\log(S)} \quad (1)$$

(iii) The Shannon index (H), which considers the number of individuals and the number of taxa (eqn. 2). Where p_i is the proportion of abundance of each species relative to the total abundance per sampling plot.

$$H = -\sum_{i=1}^{n} p_i \cdot \ln(p_i) \quad (2)$$

Remote sensing data

Two Level 1C Sentinel-2 image tiles (acquired on March 12, 2019, and April 16, 2019) were downloaded from the Copernicus

Open Access Hub (https://scihub.copernicus.eu/dhus/#/home). Level 1C top of the atmosphere (TOA) reflectance data were subsequently processed to Level-2A via the European Space Agency's (ESA) Sen2Cor algorithm (Louis et al. 2016) to obtain bottom of the atmosphere (BOA) reflectance images using the "sen2r" package (Ranghetti et al. 2020). Only the 10 m and 20 m spatial resolution bands were used in this study, and the 20 m bands were resampled to a 10 m resolution using bilinear interpolation (Li et al. 2020) to ensure spatial coherence. Image mosaicking was also performed, because two tiles were required to cover the study sites. PlanetScope imagery were downloaded from https://www.planet.com/basemaps. The imagery did not require pre-processing as it was provided in an analysis-ready form. Therefore, the data were only required to be re-projected to Arc 1960 UTM 37/S along with the pre-processed Sentinel-2 data.

The atmospherically corrected images from each sensor were used to compute the selected vegetation indices (Appendix 1) using the "RStoolbox" package (Hamzehpour et al. 2019) implemented in the R statistical software. We included five broadband optical vegetation indices and three narrowband indices specific to Sentinel-2 data. Sentinel-2 narrowband indices were utilized to assess the tree diversity modelling capability. Furthermore, the grey level co-occurrence matrix (GLCM) textural metrics 'mean', 'variance' and 'dissimilarity' (Haralick 1979), were also computed for all the spectral bands and indices, using the "glcm" package (Zvoleff 2020) in R (Appendix 2). Texture metrics were computed using a 3 × 3 window for each sensor band and vegetation index.

Statistical analysis

Variable selection

To identify key predictors for modelling forest biodiversity indicators, we utilized the variable selection package VSURF (Genuer et al. 2015) in R. This approach leverages the machine learning algorithm random forests to rank predictor importance through an iterative process of creating multiple random models and assessing mean decrease in accuracy when a given variable is excluded. We grouped the top 30 predictors selected by VSURF into four categories: 1) original Sentinel-2 and PlanetScope spectral bands, providing surface reflectance information; 2) derived vegetation indices like NDVI, sensitive to canopy properties; 3) image textures capturing spatial patterns; and 4) all variables (bands, textures and vegetation indices) combined. This allowed comparison of different predictor sets for modelling tree species diversity and other indicators relevant to tropical montane forest conservation. The VSURF selection process identified key spectral bands, indices, and textures related to vegetation characteristics in these complex forest environments.

Model development

Semi-parametric and non-parametric statistical modelling approaches were used to estimate tree diversity. Details of each approach are provided below.

Extreme gradient boosting (XGBoost) is a boosting algorithm based on gradient-boosting decision trees and random forest methods. In very large-scale data training, it is a versatile and highly scalable tree-structure enhancement model that can handle sparse data, significantly increase algorithm performance, and reduce computational memory. The R package "xgboost" (Chen and Guestrin 2016) was used to implement the XGBoost.

The GAMs were fitted for each satellite sensor using a Gaussian error distribution and logarithmic link function to relate plot-level diversity with remote sensing data. This model form is preferred because it offers acceptable estimates when true zeros are present in the tree diversity estimate, which has continuous positive values. The R package "mgcv" (Wood and Wood 2015) was used to perform the GAM regression. Each predictor variable that entered the model received a smoothing spline with a smoothing parameter, k, of 3.

$$ln(y_i) = b_0 - \sum_{i=1}^{n} f_i(x_i) \qquad (3)$$

where y_i is the ground reference diversity index value, b_0 is a constant term (intercept), and $f_i(x_i)$ (i= 1, 2,…, n) is the smoothing function for each independent variable.

Modelling was carried out on the pre-identified predictor variables from both sensors to determine the best model to explain the prediction accuracy of H, S, and J in the study area.

Accuracy assessment

To assess the accuracy and generalizability of the models, k-fold cross-validation (k=10) was implemented to facilitate a comprehensive evaluation of model performance and enhance the reliability of the estimated tree diversity indices. Three criteria for model validation and selection were chosen and computed to identify the best models: adjusted coefficient of determination (R^2), mean absolute error (MAE), and relative root-mean-square error (rRMSE). A model was considered to be the best if it had a relatively high R^2 and lower MAE and rRMSE values computed from the predictions of the 10-fold cross-validation, as presented below:

$$R^2 = \frac{(n-1)\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{(n-2)\sum_{i=1}^{n}(y_i-\bar{y})^2} \quad (4)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (5)$$

$$rRMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i-y_i)^2}{n-2}} \quad (6)$$

where $\hat{y}_i$ and y_i are the predicted and observed values for each biodiversity indicator for the i^{th} plot respectively, n is the number of observations, and $\bar{y}$ is observed mean of the respective biodiversity indicator.

Relative efficiency

To quantify potential improvements in precision from incorporating remote sensing data, we calculated relative efficiency (RE) as the ratio of variances between estimates with and without satellite data (Eqn 7). Specifically, RE compares the variance of biodiversity indicators estimated using remote sensing data (VAR_{RS}) to the variance using field plot data alone (VAR_{FD}). An RE value greater than 1 indicates the variance is lower (precision is higher) when integrating remote sensing versus using field plots alone. For example, an RE of 2 suggests the satellite-enhanced estimate could achieve the same level of precision as doubling the number of field plots sampled. This metric demonstrates the potential of remote sensing predictors from Sentinel-2 and PlanetScope to reduce uncertainty in biodiversity indicator estimates across landscapes compared to field surveys alone. Higher RE values highlight situations where integrating satellite data can strengthen precision for informing conservation and management decisions in tropical montane forests.

$$RE = VAR_{RS}/VAR_{FD} \quad (7)$$

Diversity mapping

To map biodiversity indicators across the study area, we used the raster package (Hijmans et al. 2013) in R to apply the optimal models to predict each indicator's values for all pixels from the remote sensing data. Specifically, we performed spatial prediction using the highest performing model for each indicator - either linear regression or random forest regression based on accuracy assessment. This generated continuous raster maps representing predicted tree species richness, diversity, and evenness across the landscape. We optimized spatial prediction by leveraging the full coverage and resolution of the Sentinel-2 and PlanetScope datasets through the fitted models. The resulting maps provide visualizations of spatial variability, patterns, and estimated values for key forest biodiversity indicators across inaccessible montane terrain. These high-resolution biodiversity distributions can support conservation planning and monitoring in the region.

Results

Tree species diversity

Descriptive statistics of the diversity indices are presented in Table 1. Species richness exhibited an average value of 14 per plot, indicating a relatively diverse species composition within the study area. Evenness, with an average value of 0.84, suggested a balanced distribution of tree species abundance, highlighting a more equitable representation of different taxa. The calculated Shannon diversity index averaged

2.14 per plot, signifying a moderate to high level of overall tree diversity. Overall, Shagayu FR is faring better across all three assessed diversity indices as it has higher species richness, slightly higher evenness, and notably higher Shannon diversity compared to the Magamba NFR. This indicates the tree communities in the Shagayu FR are more diverse and evenly distributed. The lower indices for the Magamba NFR may suggest higher disturbance or threats to biodiversity in that reserve.

Table 1: Descriptive statistics of forest biodiversity indicators across the study forests.

Indicator	Forest	n	Minimum	Maximum	Mean ± CI
Shannon	Magamba NFR	55	0.86	2.94	1.94 ± 0.10^{a}
	Shagayu FR	104	0.00	3.11	2.24 ± 0.09^{b}
	Overall	159	0.00	3.11	2.14 ± 0.04^{bc}
Evenness	Magamba NFR	55	0.41	0.97	0.83 ± 0.02^{a}
	Shagayu FR	104	0.10	0.97	0.84 ± 0.02^{a}
	Overall	159	0.10	0.97	0.84 ± 0.01^{a}
Richness	Magamba NFR	55	3	22	11 ± 0.91^{a}
	Shagayu FR	104	1	33	15 ± 0.98^{b}
	Overall	159	1	33	14 ± 0.67^{bc}

Different lowercase letters in rows indicate significantly different means (independent sample t-test, $p < 0.05$). n = number of plots, CI = Confidence interval (95%).

Model performance

Twenty-four (24) models were developed to predict tree species diversity indices using Sentinel-2 and PlanetScope predictor variables. The best model fit for predicting Shannon's diversity index was obtained using the XGBoost statistical approach with PlanetScope texture variables. The model had an R^2 of 0.2, MAE of 18%, and rRMSE of 23.52%. The best model for predicting tree species richness was obtained using the GAM statistical approach with PlanetScope texture variables (R^2 = 0.193, MAE = 30.72%, rRMSE = 38.91%).

Tree species evenness was best predicted using the XGBoost model with combined PlanetScope variables (R^2 = 0.115, MAE = 9.2%, rRMSE = 13.8%). Generally, combining all predictor variables (bands, textures, and vegetation indices) for each sensor (i.e., Sentinel-2 and PlanetScope) improved the predictive power of all the models (Figure 2). The selected predictor variables and performance criteria for each model are listed in Table 2.

Table 2: Performance of the GAM and XGBoost diversity models fitted with predictors from two satellite sensors. Bold values indicate the best diversity index model.

Satellite	Index	Category	Predictors	GAM rRMSE	R^2	MAE	XGB rRMSE	R^2	MAE
PlanetScope	Evenness	Bands	B, G, NIR, R	0.14	0.05	0.10	0.15	0.06	0.10
		VegIndices	DVI, GNDVI, NDVI, RVI, EVI	0.14	0.04	0.09	0.14	0.10	0.09
		Textures	EVI.mea, NIR.mea, NIR.mea	0.14	0.07	0.10	0.14	0.11	0.08
		Combined	EVI, DVI.var, NDVI, EVI.mea, RVI	0.14	0.05	0.09	**0.14**	**0.11**	**0.09**
	Richness	Bands	B, G, NIR, R	0.40	0.18	0.32	0.41	0.11	0.32
		VegIndices	DVI, GNDVI, NDVI, RVI, EVI	0.40	0.17	0.31	0.42	0.09	0.34
		Textures	DVI.vaR, EVI.con	**0.39**	**0.19**	**0.01**	0.39	0.17	0.31
		Combined	DVI, NIR, DVI.var, EVI.con, G.con	0.40	0.18	0.31	0.40	0.18	0.31
	Shannon	Bands	B, G, NIR, R	0.25	0.09	0.19	0.25	0.14	0.19
		VegIndices	DVI, GNDVI, NDVI, RVI, EVI	0.25	0.11	0.19	0.25	0.12	0.19
		Textures	EVI.con, NIR.mea, NIR.var, RVI.var	0.24	0.15	0.18	**0.24**	**0.20**	**0.18**
		Combined	EVI.con, NIR.var, NDVI.con, NIR.mea, NIR, RVI.var, RVI.mea, G.con	0.25	0.14	0.19	0.24	0.17	0.18
Sentinel-2	Evenness	Bands	B04, B03, B12, B11	**0.15**	**0.10**	**0.10**	0.16	0.07	0.10
		VegIndices	NDRE.1, GNDVI	0.13	0.06	0.09	0.14	0.07	0.10
		Textures	NDVI.var, B04.con, GNDVI.mea, B06.con, RVI.var	0.14	0.04	0.09	0.15	0.05	0.10
		Combined	GNDVI, B04.con, GNDVI.mea, NDVI.var	0.14	0.04	0.09	0.14	0.08	0.10
	Richness	Bands	B07, B06, B05, B11, B08, B04	0.41	0.12	0.32	0.42	0.08	0.33
		VegIndices	NDRE.1, GNDVI, NDRE.2, CLRE, EVI	0.42	0.07	0.33	0.43	0.05	0.34
		Textures	B06.var, B07.var, B11.mea, RVI.con, B8A.var, B08.mea, B03.con, B08.var	0.40	0.15	0.32	0.40	0.13	0.32
		Combined	B06, B07, B11, B07.var, RVI.con, B08.mea, GNDVI, B11.con, B03.con	0.42	0.14	0.33	**0.40**	**0.16**	**0.31**

Shannon	Bands	B06, B07, B8A, B11	0.25	0.13	0.19	0.26	0.08	0.20
	VegIndices	NDRE.1, GNDVI	0.25	0.08	0.19	0.26	0.08	0.20
	Textures	B04.var, B04.con, B06.con, B12.var, GNDVI.mea, CLRE.con, DVI.con	0.25	0.09	0.19	0.25	0.15	0.19
	Combined	B11.var, B06, B04.var, GNDVI, CLRE.con	0.25	0.14	0.19	**0.25**	**0.15**	**0.19**

Note: Sentinel-2 bands; B03 – green band, B04 – red band, B05 – red-edge band 1, B06 – red-edge band 2, B07 – red-edge band 3, B08 – near infrared band, B8A – narrow near infrared band, B11 – shortwave infrared band 1, B12 – shortwave infrared band 2. PlanetScope bands: B – blue band, G – green band, R – red band, NIR – near infrared band. Vegetation indices: DVI – difference vegetation index, GNDVI – green normalized difference vegetation index, NDVI – normalized difference vegetation index, RVI – ratio vegetation index, EVI – enhanced vegetation index, NDRE.1 – normalized difference red-edge 2, NDRE.2 – normalized difference red-edge 2, CLRE – chlorophyll red edge vegetation index. Textures: mea – mean texture, con – contrast texture, var – variance texture.

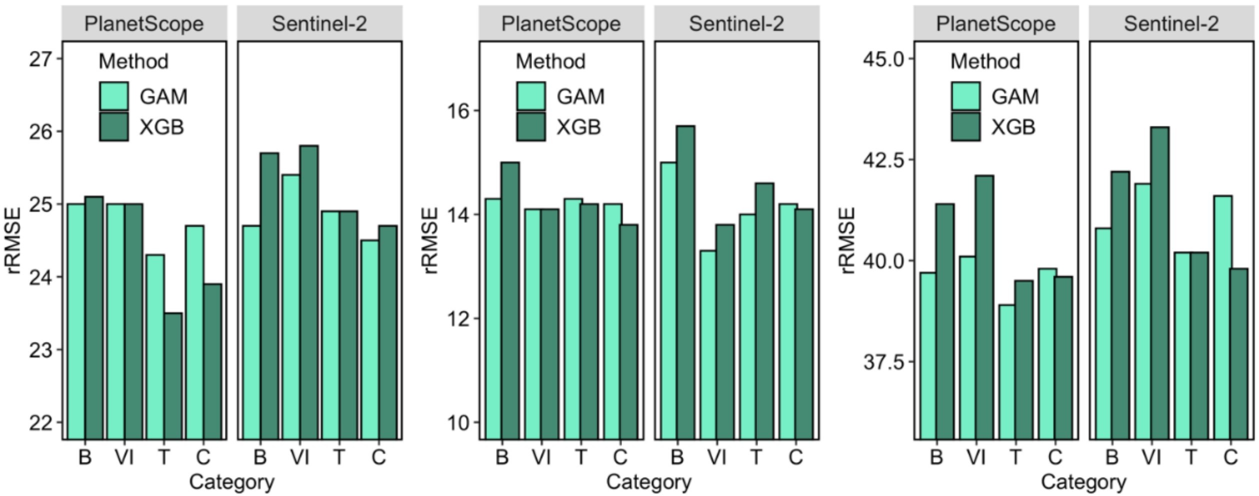

Figure 2: Performance (rRMSE) of GAM and XGBoost models fitted with predictors from the two satellite sensors.

Scatterplots were constructed to further demonstrate the relationship between the observed and predicted indices (Figure 3).

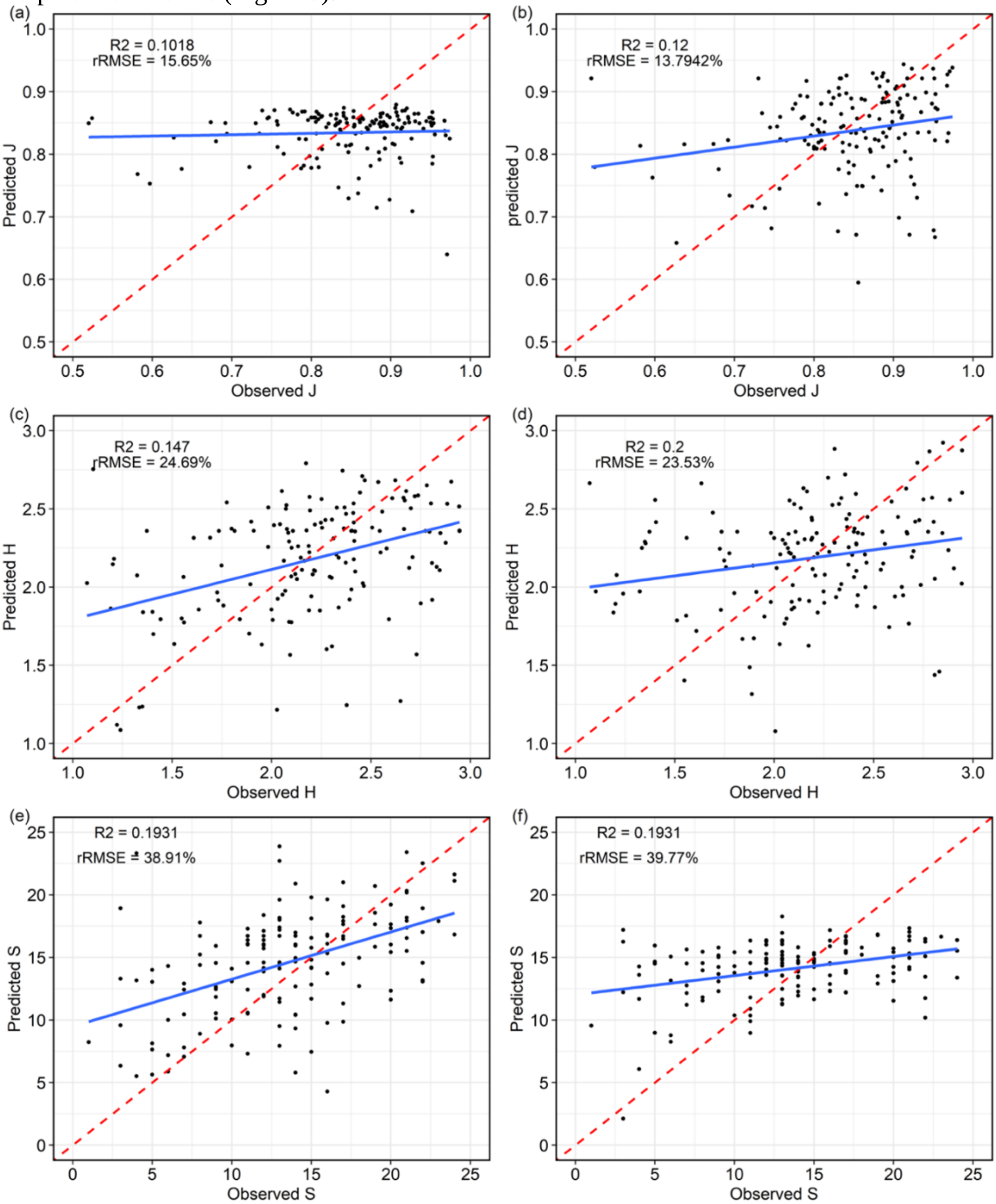

Figure 3: Scatter plots showing the relationship between the predicted and observed species evenness (a-b), Shannon diversity (c-d), and richness (e-f) for Sentinel-2 and PlanetScope, respectively.

Diversity mapping

The best models were used to generate spatial prediction maps for the Shannon's diversity index, tree species evenness, and species richness across the entire forest area (Figure 4-6).

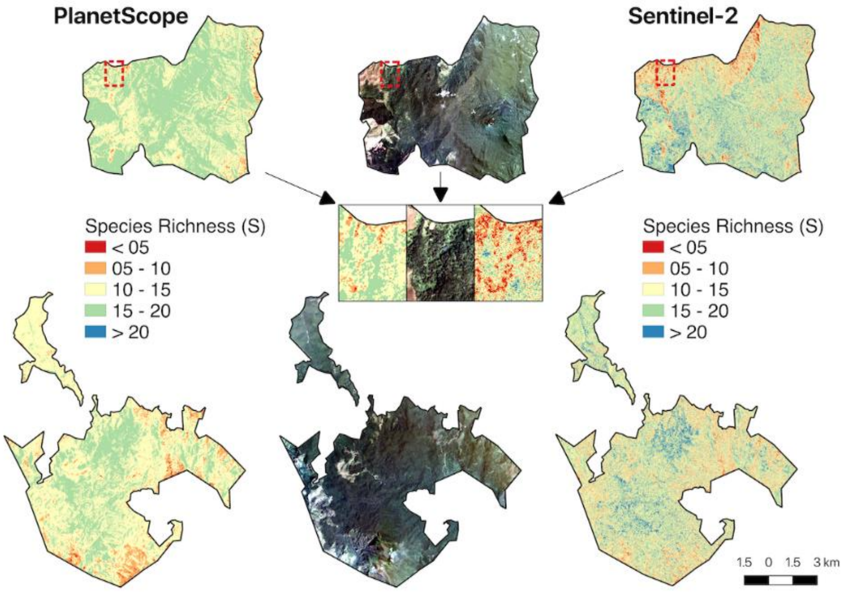

Figure 4: Prediction maps for tree species richness PlanetScope and Sentinel-2.

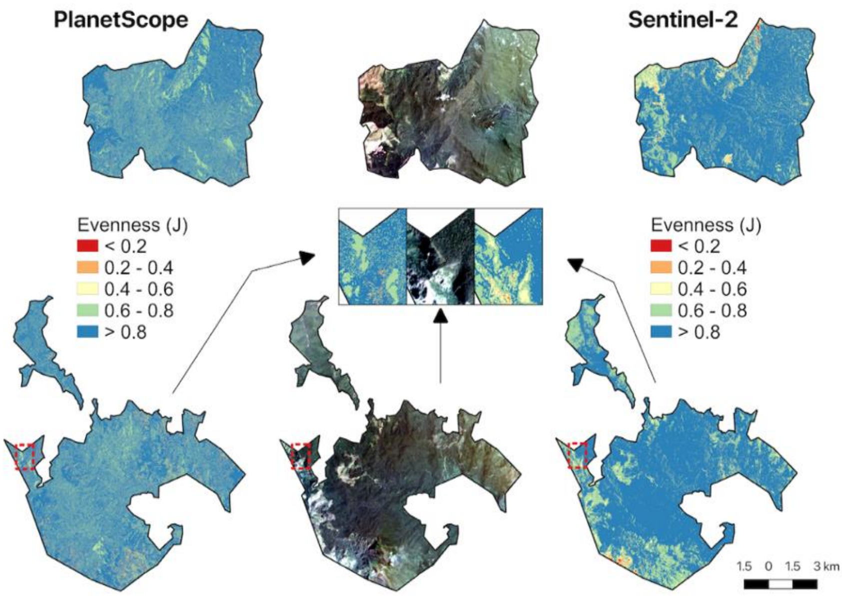

Figure 5: Prediction maps for tree species evenness from PlanetScope and Sentinel-2.

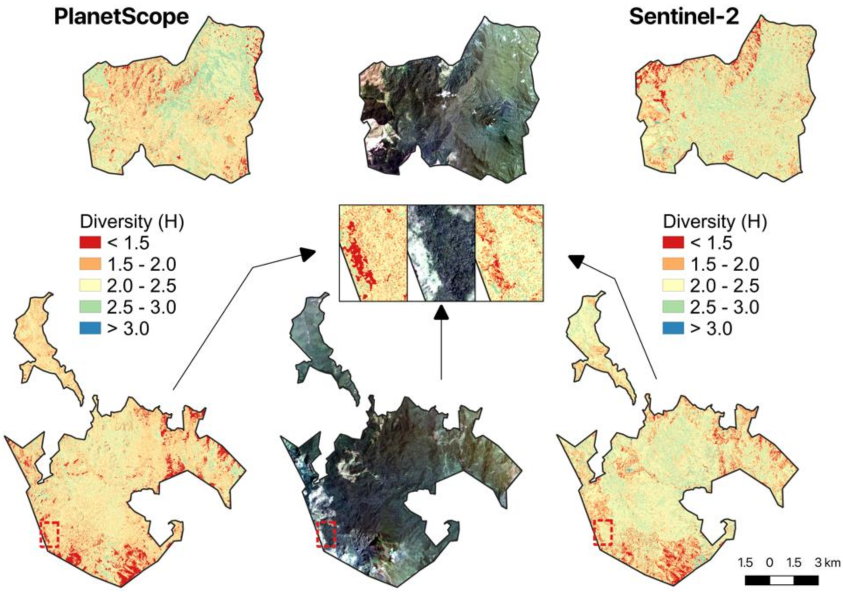

Figure 6: Prediction maps for Shannon's diversity index from PlanetScope and Sentinel-2.

Relative efficiency of the remote sensing data

An analysis of the relative efficiency of predicting tree species diversity revealed notable differences between the PlanetScope and Sentinel-2 datasets in the West Usambaras. The results indicated that the PlanetScope dataset exhibited a higher efficiency in predicting tree species diversity than Sentinel-2, as evidenced by the relative efficiency (RE) values for tree species richness and Shannon's diversity index, which were 1.47 and 2.01, respectively. These findings suggest that the PlanetScope dataset provides accurate and reliable predictions of these diversity measures. Conversely, the Sentinel-2 dataset demonstrated efficiency in predicting tree species evenness, as indicated by the RE value of 1.21, as shown in Figure 7. Generally, irrespective of the marginal differences between the sensors, the RE values indicate that remotely sensed data enhanced the precision of tree species diversity estimates as compared to conventional field-based methods.

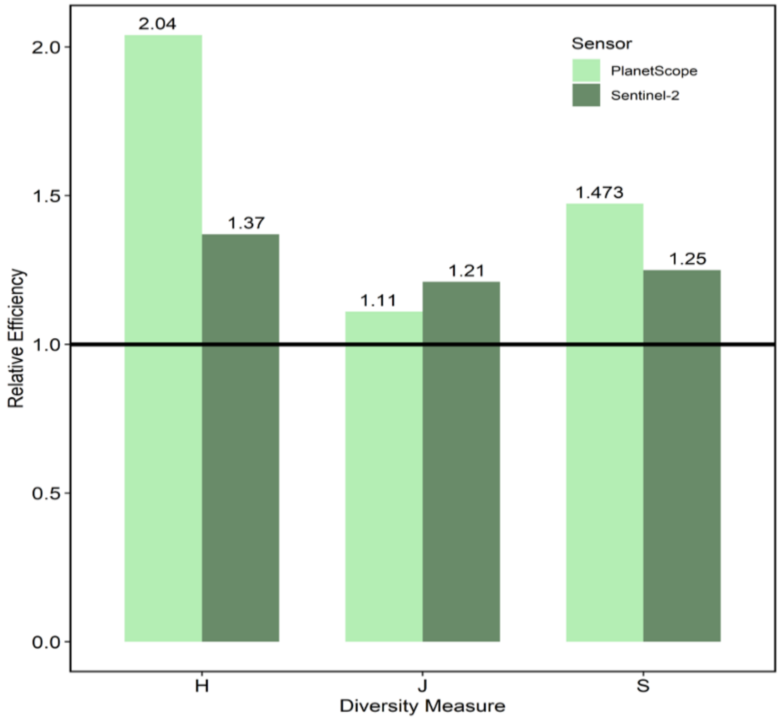

Figure 7: Graph showing relative efficiency of PlanetScope and Sentinel-2 in prediction of tree species diversity (H), evenness (J) and richness (S).

Discussion

The aim of this study was to assess the utility of Sentinel-2 and PlanetScope remotely sensed data for large-scale estimation of tropical forest biodiversity indicators. The predictors used in the estimations were spatial bands, vegetation indices, textures, and a combination of the three. The combination of all the three predictors improved the predictive power of the models. Various studies (Metcalfe et al. 2015, Rovero et al. 2014) have shown different contributions of predictors to improving the efficiency of model prediction power. Spatial resolution has also shown to affect the efficiency of model predictions as elaborated by Wulder et al. (2004), Potapov et al., (2008) and Getzin et al. (2012). The results of this study showed that the models from PlanetScope, a high-resolution sensor, had higher prediction power and overall prediction accuracy.

Higher spatial resolution of PlanetScope may have enabled a better separation of tree species with varying canopy greenness, which plays a significant role in predicting tree species diversity (Wu et al. 2021). Additionally, high-resolution textures derived from PlanetScope imagery provide a more detailed representation of vegetation structural components, leading to greater model precision (Mauya and Madundo 2022). These findings align with those of previous studies (Gyamfi-Ampadu et al. 2021, Baloloy et al. 2018), which reported a similar performance of PlanetScope imagery for estimating tree species diversity and other vegetation properties.

Compared with similar studies (Wang and Gamon 2019, Guisan and Zimmermann 2000), the results of this study showed relatively lower R^2 values. However, this does not indicate poor performance of the proposed models. Lopatin et al. (2016), using LiDAR technology reported $R^2 = 0.33$ for predicting tree species richness using Random Forest model. Irrespective of the model choice and sensor type, the results align with the results from this study, considering LiDAR's higher resolution.

Additionally, the models rRMSE and MAE shows that the models have relatively lower prediction errors, and hence, good performance.

Among the three diversity measures, the predictive power of the models using PlanetScope was the best for estimating tree species diversity. This can mainly be attributed to the high spatial resolution of the PlanetScope sensors compared to the Sentinel-2 sensors, which enhances the identification of plant features. Sentinel-2 only explained a small percentage of the variance in diversity across the study sites, whereas similar results were reported by Ma et al. (2019). PlanetScope data have shown ability to accurately estimate tree species richness and Shannon's diversity index as compared to Sentinel-2 data, which have been shown to underestimate species richness and Shannon's diversity index. However, Sentinel-2 variables were able to better estimate species evenness than PlanetScope variables.

The XGBoost model has shown a higher utility in estimating tree species diversity than the GAM. Schratz et al. (2021) has shown that, in terms of performance, XGBoost outperforms GAM in many benchmarks, particularly when dealing with high-dimensional datasets (e.g. ecological data). In this study, the datasets can be regarded as highly dimensional given the number and types of variables used for the development of predictive models.

A comparison between the produced diversity prediction maps and near real-time imagery showed that the prediction accuracy using data from the two sensors matched the tree species diversity in the study area (Figure 4 to 6). Therefore, the developed models can be used to predict tree species diversity in tropical mountain forests. PlanetScope data have been shown to enhance the performance of the models and produce more accurate prediction maps than Sentinel-2 data main attribution being its high spatial resolution (4.77 m) in contrast to Sentinel-2 sensors (10 m) (Vizzari 2022, Mauya and Madundo 2022).

Prediction maps for tree species richness, species evenness, and Shannon diversity index showed a general trend of lower values in areas located near the forest border. This matches the actual situation at the study sites, which is mainly caused by the human over-exploitation of forest resources especially in borderline areas. This shows that the prediction models efficiency as they could effectively reflect the conditions present in the study sites. Generally, the use of remote sensing techniques has been shown to improve the estimation of forest biodiversity indicators, as the calculated relative efficiency values were all greater than one. This suggests that using remote sensing in forest inventories is more efficient than relying solely on field-based estimates (Puliti et al. 2017, Ene et al. 2017). To achieve a similar level of precision as a pure field-based estimate that employs simple random sampling, the sample size for the field-based inventory must be increased by a factor equivalent to the value of RE (Næsset et al. 2016). This increase in sample size would significantly impact the costs associated with field inventory.

Conclusion

The results of this study showed that the integration of field-based approaches and remote sensing techniques can facilitate accurate large-scale estimation of forest biodiversity indicators for dense tropical mountain forests. In addition, the results demonstrated the contribution of high spatial resolution to the accurate estimation and mapping of various attributes of forest biodiversity indicators. It has also been observed from the tree species diversity prediction maps that areas located in the forest peripherals have low tree species richness and diversity, implying a high level of destruction of forest ecosystems in the study area through various human activities. Therefore, to ensure that ecological systems within the forest and nature reserves remain intact, it is necessary to put in place and implement conservation measures that are more rigorous and efficient for biodiversity conservation.

Acknowledgements

Forest inventory data collection was supported by the Eastern Arc Mountain Conservation Endowment Fund through the project "Meta-Analysis of Forest Carbon Information Using Existing Data and Remote Sensing Techniques". We also thank the Tanzania Forest Fund and the Regional Centre for Mapping of Resources for Development/Global Monitoring for Environment and Security Africa (RCMRD/GMES Africa) for their financial support.

Conflict of interest

The authors declare no conflicts of interest.

References

Abbas S, Wong MS, Wu J, Shahzad N and Muhammad IS 2020 Approaches of satellite remote sensing for the assessment of above-ground biomass across tropical forests: Pan-tropical to national scales. *Remote Sens* 1220: 20.

Arroyo-Rodríguez V, Fahrig L, Tabarelli M, Watling JI, Tischendorf L, Benchimol M, Cazetta E, Faria D, Leal IR, and Melo FP 2020 Designing optimal human-modified landscapes for forest biodiversity conservation. *Ecology Letters*, *23*(9), 1404–1420.

Baloloy AB, Blanco AC, Candido CG, Argamosa RJL, Dumalag JBLC, Dimapilis LLLC and Paringit EC 2018 Estimation of mangrove forest aboveground biomass using multispectral bands, vegetation indices and biophysical variables derived from optical satellite imageries: Rapideye, planetscope and sentinel-2. ISPRS Ann Photogramm *Remote Sens Spatial Inf Sci* 4: 29–36.

Barnes EM, Clarke TR, Richards SE, Colaizzi PD, Haberland J, Kostrzewski M, Waller P, Choi CRE, Thompson T, Lascano RJ, Li Hand Moran MS 2000 Coincident detection of crop water stress, nitrogen status and canopy density using ground based multispectral data. *Proc 5th Int Conf Precis Agric*

Beyene SM, Hussin YA, Kloosterman HE and Ismail MH 2020 Forest inventory and aboveground biomass estimation with terrestrial LiDAR in the tropical forest of Malaysia. *Can J Remote Sens* 46(2): 130–145.

Burgess Nd, Butynski TM, Cordeiro NJ, Doggart NH, Fjeldså J, Howell KM, Kilahama FB, Loader SP, Lovett JC and Mbilinyi B 2007 The biological importance of the Eastern Arc Mountains of Tanzania and Kenya. *Biol.Conserv.* 134(2): 209–231.

Chen T and Guestrin C 2016 Xgboost: A scalable tree boosting system. Proc 22nd *Acm Sigkdd Int Conf Knowledge Discovery Data Mining*: 785–794.

Corte APD, Souza DV, Rex FE, Sanquetta CR, Mohan M, Silva CA, Zambrano AMA, Prata G, Alves de Almeida DR, Trautenmüller JW, Klauberg C, de Moraes A, Sanquetta MN, Wilkinson Band Broadbent EN 2020 Forest inventory with high-density UAV-Lidar: Machine learning approaches for predicting individual tree attributes. *Comput. Electron Agric.* 179: 105815.

Dainelli R, Toscano P, Di Gennaro SF and Matese A 2021 Recent advances in unmanned aerial vehicle forest remote sensing-A systematic review. Part I: A general framework. *Forests* 12(3): 3.

Drusch M, Del Bello U, Carlier S, Colin O, Fernandez V, Gascon F, Hoersch B, Isola C, Laberinti P, and Martimort P 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, *120*, 25–36.

Ene LT, Næsset E, Gobakken T, Bollandsås OM, Mauya, EW and Zahabu E 2017 Large-scale estimation of change in aboveground biomass in miombo woodlands using airborne laser scanning and national forest inventory data. *Remote Sens. Environ.* 188: 106–117.

Genuer R, Poggi JMandTuleau-Malot C 2015 VSURF: An R package for variable selection using random forests. *R J.* 7(2): 19–33.

Getzin S, Wiegand K and Schöning I 2012 Assessing biodiversity in forests using very high-resolution images and unmanned aerial vehicles. *Methods Ecol. Evol.* 3(2): 397–404.

Gitelson AA, Gritz YandMerzlyak MN 2003 Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *J. Plant Physiol.*(1603): 271–282.

Gitelson AA, Kaufman YJ and Merzlyak MN 1996 Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.* 58: 289–298.

Gitelson A and Merzlyak MN 1994 Spectral reflectance changes associated with autumn senescence of Aesculus hippocastanum L and Acer platanoides L leaves. Spectral features and relation to chlorophyll estimation. *J. Plant Physiol.* 143(3): 286–292.

Goodbody TRH, Coops NC and White JC 2019 Digital aerial photogrammetry for updating area-based forest inventories: A review of opportunities, challenges, and future directions. *Curr. Forestry Rep.* 5(2): 55–75.

Guisan A and Zimmermann NE 2000 Predictive habitat distribution models in ecology. *Ecol. Modell.* 135(2–3): 147–186.

Gyamfi-Ampadu E, Gebreslasie M and Mendoza-Ponce A 2021 Evaluating multi-sensors spectral and spatial resolutions for tree species diversity prediction. *Remote Sens.* 13(5): 1033.

Hamzehpour N, Shafizadeh-Moghadam H and Valavi R 2019 Exploring the driving forces and digital mapping of soil organic carbon using remote sensing and soil texture. *Catena* 182: 104141.

Haralick RM 1979 Statistical and structural approaches to texture. *Proc IEEE* 67(5): 786–804.

Hijmans RJ, Van Etten J, Mattiuzzi M, Sumner M, Greenberg JA, Lamigueiro OP, Bevan A, Racine EB and Shortridge A 2013 Raster package in R. Version. https://mirrors.sjtug.sjtu.edu.cn/cran/web/packages/raster/raster.pdf

Hui QL and Huete A 1995 A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. *IEEE Trans Geosci. Remote Sens.* 33(2): 457–465.

Li Z, Zhang HK, Roy DP, Yan Land Huang H 2020 Sharpening the Sentinel-2 10 and 20 m Bands to Planetscope-0 3 m Resolution. *Remote Sens.* 12(15): 2406.

Lopatin J, Dolos K, Hernández HJ, Galleguillos Mand Fassnacht FE 2016 Comparing generalized linear models and random forest to model vascular plant species richness using LiDAR data in a natural forest in central Chile. *Remote Sens. Environ.* 173: 200–210.

Louis J, Debaecker V, Pflug B, Main-Knorn M, Bieniarz J, Mueller-Wilm U, Cadau E and Gascon F 2016 Sentinel-2 Sen2Cor: L2A processor for users. *Proc Living Planet Symp* 2016: 1–8.

Lovett JC 1996 Elevational and latitudinal changes in tree associations and diversity in the Eastern Arc mountains of Tanzania. *J. Trop. Ecol.* 12(5): 629–650.

Ma X, Mahecha MD, Migliavacca M, van der Plas F, Benavides R, Ratcliffe S, Kattge J, Richter R, Musavi T and Baeten L 2019 Inferring plant functional diversity from space: The potential of Sentinel-2. *Remote Sens. Environ.* 233: 111368.

Mauya EW and Hansen EH, Gobakken T, Bollandsås OM, Malimbwi RE, 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 Manage.* 10: 1–14.

Mauya EW and Madundo S 2022 Modelling Above Ground Biomass Using Sentinel 2 and Planet Scope Data in Dense Tropical Montane Forests of Tanzania. *Tanz. J. Forestry Nature Conserv.* 91(1): 132–153.

Metcalfe P, Beven Kand Freer J 2015 Dynamic TOPMODEL: A new implementation in R and its sensitivity to time and space steps. *Environ Model Softw.* 72: 155–172.

Næsset E, Ørka HO, Solberg S, Bollandsås OM, Hansen EH, Mauya E, Zahabu E, Malimbwi R, Chamuya N, Olsson HandGobakken T 2016 Mapping and estimating forest area and aboveground biomass in miombo woodlands in Tanzania using data from airborne laser scanning, TanDEM-X, RapidEye, and global forest maps: A comparison of estimated precision. *Remote Sens. Environ.* 175: 282–300.

Pearson RL and Milton LD 1972 Remote mapping of standing crop biomass for estimation of the productivity of the shortgrass prairie, Pawnee National Grasslands, Colorado. *Proc Eighth Int. Symp. Remote Sens. Environ.* October 2–6, 1972.

Poortinga A, Thwal NS, Khanal N, Mayer T, Bhandari B, Markert K, Nicolau AP, Dilger J, Tenneson K and Clinton N 2021 Mapping sugarcane in Thailand using transfer learning, a lightweight convolutional neural network, NICFI high resolution satellite imagery and Google Earth Engine. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, *1*, 100003.

Potapov P, Yaroshenko A, Turubanova S, Dubinin M, Laestadius L, Thies C, Aksenov D, Egorov A, Yesipova Y, Glushkov I, Karpachevskiy M, Kostikova A, Manisha A, Tsybikova E and Zhuravleva I 2008 Mapping the World's Intact Forest Landscapes by Remote Sensing. *Ecology and Society*, *13*(2).

Puliti S, Ene LT, Gobakken T and Næsset E 2017 Use of partial-coverage UAV data in sampling for large scale forest inventories. *Remote Sens. Environ.* 194: 115–126.

Ranghetti L, Boschetti M, Nutini F and Busetto L 2020 "sen2r": An R toolbox for automatically downloading and pre-processing Sentinel-2 satellite data. *Comput. Geosci.* 139: 104473.

Richardson AJ and Wiegand CL 1977 Distinguishing vegetation from soil background information. *Photogramm Eng. Remote Sens.* 43(12): 1541–1552.

Rouse JW Jr, Haas RH, Schell JA and Deering DW 1974 Monitoring Vegetation Systems in the Great Plains with Erts, NASA Special Publication. Proc Third Earth Resources Technology Satellite- 1 Symp, 309–317.

Rovero F, Menegon M, FjeldsAa J, Collett L, Doggart N, Leonard C, Norton G, Owen N, Perkin A and Spitale D 2014 Targeted vertebrate surveys enhance the faunal importance and improve explanatory models within the Eastern Arc Mountains of Kenya and Tanzania. *Divers Distrib.* 20(12): 1438–1449.

Schratz P, Muenchow J, Iturritxa E, Cortés J, Bischl B and Brenning A 2021 Monitoring forest health using hyperspectral imagery: Does feature selection improve the performance of machine-learning techniques? *Remote Sens.* 13(23): 4832.

Storch F, Boch S, Gossner MM, Feldhaar H, Ammer C, Schall P, Polle A, Kroiher F, Müller J and Bauhus J 2023 Linking structure and species richness to support forest biodiversity monitoring at large scales. *Ann. Forest Sci.* 80(1): 1–17.

Vizzari M 2022 PlanetScope, Sentinel-2, and Sentinel-1 Data Integration for Object-Based Land Cover Classification in Google Earth Engine. *Remote Sens.* 14(11): 2628.

Wang R and Gamon JA 2019 Remote sensing of terrestrial plant biodiversity. *Remote Sens. Environ.* 231: 111218.

Wood S and Wood MS 2015 Package 'mgcv'. R Package Version 1(29): 729.

Wu S, Wang J, Yan Z, Song G, Chen Y, Ma Q, Deng M, Wu Y, Zhao Y and Guo Z 2021 Monitoring tree-crown scale autumn leaf phenology in a temperate forest with an integration of PlanetScope and drone remote sensing observations. *ISPRS J. Photogramm Remote Sens.* 171: 36–48.

Wulder MA, Hall RJ, Coops NCand Franklin SE 2004 High Spatial Resolution Remotely Sensed Data for Ecosystem Characterization. *BioScience* 54(6): 511–521.

Zvoleff A 2020 Glcm: Calculate textures from grey-level Co-occurrence matrices (GLCMs) version 1.6.5 from CRAN. CRAN Package 'glcm'.

Appendices

Appendix 1: Description of vegetation indices used as predictor variables for tree species diversity index modelling.

Index	Name	Expression	Sensor	Reference
DVI	Difference Vegetation Index	NIR-Red	S-2, PS	Richardson and Wiegand (1977)
EVI	Enhanced Vegetation Index	2.5[(NIR-Red)/(NIR+2.4Red + 1)]	S-2, PS	Hui and Huete (1995)
GNDVI	Green Normalized Difference Vegetation Index	(NIR - Green)/(NIR + Green)	S-2, PS	Gitelson et al. (1996)
NDVI	Normalized Difference Vegetation Index	(NIR - Red)/(NIR + Red)	S-2, PS	Rouse et al. (1974)
RVI	Ratio Vegetation Index	NIR/Red	S-2, PS	Pearson and Milton (1972)
CLRE	Chlorophyll Red-Edge	(RE3/RE1) - 1	S-2	Gitelson et al. (2003)
ND-RE1	Normalized Difference Red Edge	(RE2 – RE1)/(RE2 + RE1)	S-2	Gitelson and Merzlyak (1994)
ND-RE2	Normalized Difference Red Edge	(RE3 – RE1)/(RE3 + RE1)	S-2	Barnes et al. (2000)

Appendix 2: General description of grey-level co-occurrence matrix (GLCM) texture metrics used in this study.

Texture	Expression	Expression
Mean (mea)	$\mu_i = \sum_{i,j=0}^{N-1} iP_{i,j}$	Mean of grey level (GL) distribution of the image.
Variance (var)	$\sum_{i,j=0}^{N-1} iP_{i,j}\,(i - \mu_i)^2$	GLCM variance is a measure of the dispersion of GL distribution
Contrast (con)	$\sum_{i,j=0}^{N-1} iP_{ij}\,(i - j)^2$	Contrast indicates the amount of local GL variation in an image. Large values indicate the presence of edges, noise or wrinkled features.