Modelling Above Ground Biomass Using Sentinel 2 and Planet Scope Data in Dense Tropical Montane Forests of Tanzania

E.W. Mauya and S. Madundo

Department of Forest Engineering and Wood Sciences
College of Forestry, Wildlife and Tourism, Sokoine University of Agriculture
Morogoro, Tanzania

Correspondence: mauya@sua.ac.tz

ABSTRACT

Forest biomass estimation using field -based inventories at a large scale is challenging and generally entails large uncertainty in tropical regions. In this study, we investigated the performance of Sentinel 2 and Planet Scope data for above ground biomass (AGB) modelling, in the tropical rainforest of Tanzania. A total of 296 field inventory plots were measured across the west Usambara mountain forests. The results showed that, Sentinel 2-based model fitted using GLMs had better performance (cvRMSEr = 67.00 % and pseudo-R^2 = 20%) as compared to Planet Scope-based models (cvRMSEr = 72.1 % and pseudo-R^2 = 5.2%). Overall GLMs resulted into models with less prediction errors in contrast to random forest when using Sentinel 2 data. However, for the Planet Scope, there was marginal improvement when using random forest (cvRMSEr = 72.0%). Models that incorporated texture variables produced better prediction accuracy as compared to those with band values and indices only. The study has shown that, Sentinel 2 and Planet Scope remotely sensed data can be used to develop cost-effective method for AGB estimation in tropical rainforests of Tanzania.

Key words: Biomass – GLMs - Planet Scope - Sentinel-2 - Random Forest – Texture.

INTRODUCTION

Estimation of forest biomass and carbon stock is important for quantification of the roles of forests as carbon sources or sinks, and for supporting sustainable forest management (Mauya *et al*. 2015a, Temesgen *et al*. 2015). In the recent decades, the concern about global climate change, has further highlighted the need for developing efficient methods for estimating and reporting forest biomass and carbon stocks at local, national, continental and global scale (Fawzy *et al*. 2020). One of the notable forest-based climate change mitigation programme under the United Nations Framework Conversion on Climate change, is a programme on Reducing Emission From Deforestation and Forest Degradation, through conservation, sustainable management of forests and enhancement of forest carbon stocks in developing countries (REDD+) (Mauya and Madundo 2021). This programme requires accurate information on forests biomass and carbon stock as a basis for its implementation and monitoring. Moreover, forest biomass is also recognized by the Global Climate Observing Systems as an Essential Climate Variable (Duncanson *et al*. 2019) and its systematic characterization is important for reporting on afforestation, reforestation, and deforestation categories globally (Herold *et al*. 2019).

Field-based sample survey such as national forest inventory, had traditionally been used to provide estimates of aboveground biomass (AGB) at regional and national scales (Naesset *et al*. 2016). However, the wall-to-wall estimation of AGB over large areas by field-based measurements require a dense network of inventory plots to reach good accuracies and precision (Mitchell *et al*. 2017). Using remote sensing-assisted forest inventories, in such areas is therefore, the

most practical option given the ability of remotely sensed data to account for limitations related to sample size, time lines, expenses, and accessibility (Koch 2015). Remotely sensed data provide also a synoptic view over large areas and greatly enhance the precision and usefulness of the conventional field-based methods (Sinha *et al*. 2015). However, there are still challenges in selecting the appropriate remote sensing data, variables, and modeling algorithms for different ecological environments to produce intended results.

To date, a variety of remotely sensed data including; Light Detection and Ranging (LiDAR), Radio Detection and Ranging (RADAr) and Optical Remote Sensing have been used for estimating and mapping AGB in different forest types. Of these methods, LiDAR system has been proven to have an excellent ability in predicting and estimating AGB with better precision in comparison to using radar and optical data (Tian *et al*. 2013). However, the limitations of data availability, high cost and huge data volume (Zhao et al. 2016a), impede its wider application in estimating forest AGB in lower- and middle-income countries with larger coverage of forest. Landsats has been traditionally the first-hand choice alternative to commercial remotely sensed data, and it has widely been applied globally for AGB estimation and mapping across different forests types. This is mainly because of their medium spatial resolution, relatively large coverages, and freely and long history time series data availability since the 1980s (Boyd and Danson 2005, Wulder *et al*. 2011).

The launching of the Copernicus program of the European Space Agency had further increased the global repository of open access data with more important developments in spatial, temporal, and radiometric resolution (Astola *et al*. 2019, Li *et al*. 2021). For example; the improved spatial resolution from 30 m of Landsat 8 to 10 m of Sentinel-2 makes a big difference for the operational actors which enables estimation of variables (e.g., AGB per ha) at the lower scale levels of forest plots and stands. Sentinel-2 (particularly A and B) has more spectral bands (13 Sentinel-2 vs. 7 Landsat-8 bands), including three Vegetation Red Edge (VRE) and one Narrow Near Infrared (NNIR) bands (Forkuor *et al*. 2018, Biswas *et al*. 2020). The VRE bands are expected to contribute to improved AGB estimation and mapping (Qiu *et al*. 2017). Furthermore, recently, high-resolution imagery (Planet Scope) has been made available to the public through the Norway's International Climate and Forest Initiative (NICFI), with Kongsberg Satellite Services (KSAT) and its partners Planet and Airbus (Poortinga *et al*. 2021). Planet Scope data are specifically intended to enhance forest monitoring in the tropical countries for REDD+ implementation and sustainable forest management at large. Despite the freely availability of both Sentinel-2 and Planet Scope data, fewer studies have reported their capability in estimating and mapping AGB in the tropical rainforests. Thus, understanding their performance and contributions in enhancing the precision of AGB estimates relatively to the conventional field-based methods, will set a baseline information for developing robust method for estimating AGB at different spatial scales in the dense tropical forests.

Like any other remotely sensed data, Sentinel 2 and Planet scope, do not directly measure AGB from the air or space, but rather rely on the empirical models developed by linking the information derived from ground field measurements (i.e., AGB/ha) at a plot or stand level and the corresponding remotely sensed predictor variables derived from the same spatial scale as the field plot (Vafaei *et al*. 2018). Such models are used to provide pixel-wise predictions of the respective AGB over the entire area of interest covered by the remotely sensed data (e.g. Gizachew *et al*. 2016, Jha *et al*. 2021, Li *et al*. 2021). Thus, the quality of the model is of fundamentally important in deriving precise estimates of

AGB when using remotely sensed assisted forest inventory. To achieve this, a number of statistical methods including parametric, semi-parametric and non-parametric have been applied with varying levels of accuracy (Chen *et al*. 2018, Nuthammachot *et al*. 2018, Ahmad *et al*. 2021, Cosenza *et al*. 2021, Jiang *et al*. 2021). Performance of these methods varies with data types, forest types, forest structure and sample size (Fassnacht *et al*. 2014). Therefore, it is important to compare these methods in order to select appropriate algorithm for establishing AGB estimation models when using new remotely sensed data (Li *et al*. 2020). Such studies are however limited in the tropical rainforest especially those using Sentinel-2 and planet-scope imageries.

Selection of the predictor variable is another key parameter that affects the quality of the model when relating AGB and remotely sensed predictor variables (Adame-Campos *et al*. 2019). In the context of optical remotely sensed, the commonest predictor variables which have been applied in different studies include, reflectance values, vegetation indices and texture variables (Dang *et al*. 2019, Jha *et al*. 2021). Among all predictors, textural variables, had been reported to improve the accuracy of AGB prediction models across different forest types, partly because of their strong correlation with different forest structure attributes including AGB (Pandit *et al*. 2020). Therefore, while the main objective of this study is to predict AGB using Sentinel-2 and Planet Scope data, specifically the study aimed to evaluate the importance of different statistical methods and predictor variables on modelling and predicting AGB using Sentinel-2 and Planet Scope. Potential gain in precision of remotely sensed AGB estimation compared to pure field measurements were also quantified.

MATERIALS AND METHODS

Study area

This study was conducted in west Usamabara mountains block (4° 25'-5° 07' S and 38° 10'-38° 35' E) located in Northern Tanzania (Figure 1). These mountains are part of the widely known Eastern Arc Mountains (EAMs) which are group of isolated mountains stretching from Southeast Kenya to the Makambako gap in south central Tanzania (Figure 1). The west Usambara montains block (WUMB) are considered as the large upland block in the northern part of the Eastern Arc range which covers about 2,200 km^2 and rising from 408 to 2294 m.a.s.l.

The WUMB are found mainly in Lushoto District, but a smaller area extends to Korogwe District. The climate in WUMB is oceanic with bimodal rainfall, partly determined by their proximity to the Indian Ocean and the equator. Rainfall peaks in April and November. The maximum mean annual rainfall is 2000 mm in the wettest areas, falling to less than 600 mm in the rain shadow areas (Lovett 1996). Temperatures are higher on the lower parts (25-27° C mean monthly) and lower on the plateau (13-18°C mean monthly). The minimum and maximum temperatures are 13° C and 27° C, respectively. Extreme temperatures (7° C during cold seasons and 30° C during hot seasons) have been recorded (Msuya and Kideghesho 2009).

This study was conducted in five forests namely: Magamba nature forest reserve, Shagayu forest reserve, Ndelemai forest reserve, Balangai forest reserve, Mahezangulu and Kisimagonja forest reserve located within the WUMB (Figure 2). The areas and elevational ranges of the forests are presented in Table 1.

These forests have vegetation types ranging from lowland, intermediate (sub-montane) and highland (montane) evergreen forests. Common tree species are *Newtonia buchananii, Parinari excelsa, Albizia*

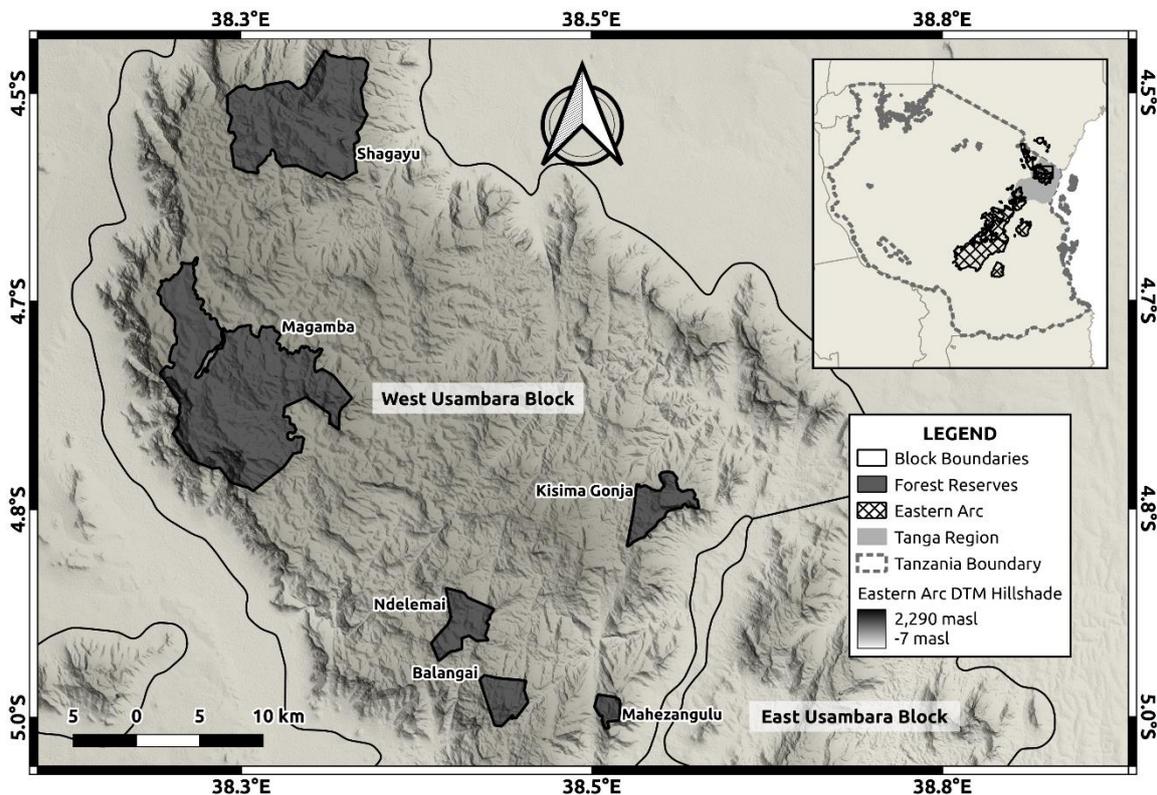*gummifera, Ocotea usambarensis and Allanblackia stuhlmannii.*

Figure 1. Location of forests within WUMB

Table 1: **Name, area and elevation ranges of the individual forests within WUMB**

Forest name	Area	Elevation ranges
Magamba	9283	1650-2300
Shagayu	7830	1400-2100
Ndelemai	1421	1422-1790
Balangai	992	1440-1760
Mahezangulu	322	1400-1750
Kisimagonja	1440	1400-1765

Data Collection

Sampling design

Systematic sampling design was used in all the five forests with slight variations from one forest to another. In Magamba Nature Forest Reserve a systematic grid of **225 x 900** was established, field plots were populated to cover the entire area with inter plot distance of 225m. A subsample of 55 circular plots were selected and measured in the field. We selected 55field plots, due to high travel costs and long walking distances in the steep and rough terrain which would nearly not permit to cover all the field plots on the 225x 900 grid. However, to ensure that all the variations of AGB on the entire forest were covered, we developed a sampling strategy that ensured the entire altitudinal variations existing in the forest were included. In Shagayu forest reserve, a systematic grid of **700x 350**, was established and intensified with plots at a distance of **350** m apart. The initial plan was to measure all the plots, but given the difficult terrain, some of the plots were inaccessible. Thus, 99 field circular plots of 15 m radius were measured in Shagayu forest reserve. For Balangai, Ndelemai, Mahezangulu and Kisimagonja forest reserves, the sample plots were established at a grids of 700 x 350 m. All the plots were measured in the field with inter-plot distance of 350m in each of the forest as presented in Table 2.
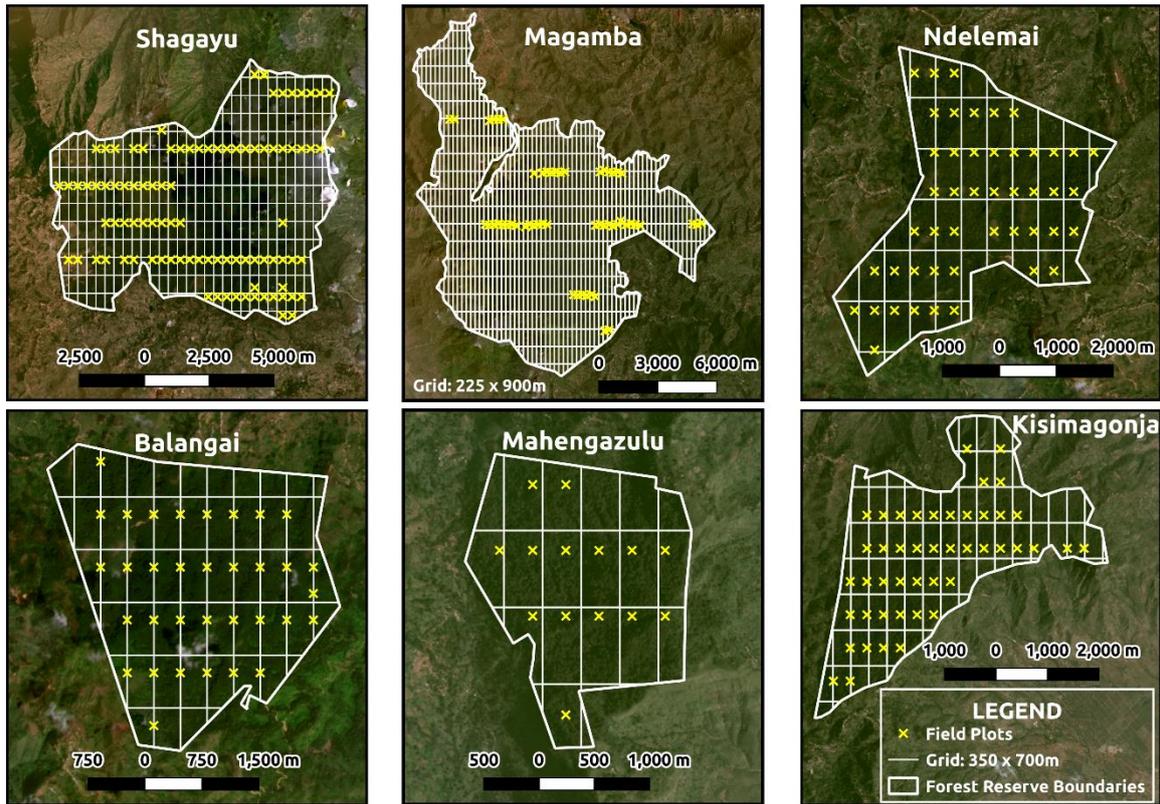
Figure 2. Location of field sample plots for each forest within WUMB

Field data

Field data were collected between March and June 2020. Hand held Global Positioning System (GPS) was used to navigate to the center of the field plot using the pre-defined coordinates obtained from the sampling grid, and recorded. In Magamba, Shagayu, Mahezangulu and Kisimagonja, circular plots of 15 m radius were established. On each plot, diameter at breast height (*dbh*) for all trees with *dbh* larger than 5cm were measured using caliper and their scientific and local names recorded. In Balangai and Ndelemai, diameter at breast height (*dbh*) was measured using caliper following the lower *dbh* thresholds in accordance with the concentric circle plot design (Vesa *et al*. 2010). The radii of the concentric circles were 2, 5, 10, and 15m. Trees with *dbh* $\geq$ 1, $\geq$ 5, $\geq$ 10, and $\geq$ 20 cm, respective to these concentric plots were measured. However, in our analysis we considered only trees with *dbh* greater than 5, because such trees are

considered to store substantial amount of carbon as compared to smaller trees. Across all the six sites, three trees (i.e., larger, medium and small) in each sample plot were measured for height using Vertex Hypsometer. The heights of the remaining trees were predicted using diameter-height model that was developed based on the sample trees. A number of model forms for *diameter–height* relationship were tested using non-linear mixed effect approach implemented in lmfor package (Mehtatalo and Mehtatalo 2015), of the R statistical software Best model fit, judged by the Akaike information criterion (AIC), was obtained using the model form by Näslund (1936).

Field estimates of AGB

AGB was calculated for each individual tree using the local allometric model developed by Masota *et al*. (2016) with both *dbh* and *height* as predictor variables. Using models with both *dbh* and height has been reported

to moderate the effect of large *dbh*-values on AGB estimates as compared to models with *dbh* only. The individual tree AGB were then summed to obtain total AGB for the respective plot and finally up scaled to per ha

values by dividing with plot area. The descriptive statistics of AGB are presented in Table 1 and the distribution across individual forests is shown in Figure 3.

Table 2. Descriptive statistics for AGB (Mg/ha)

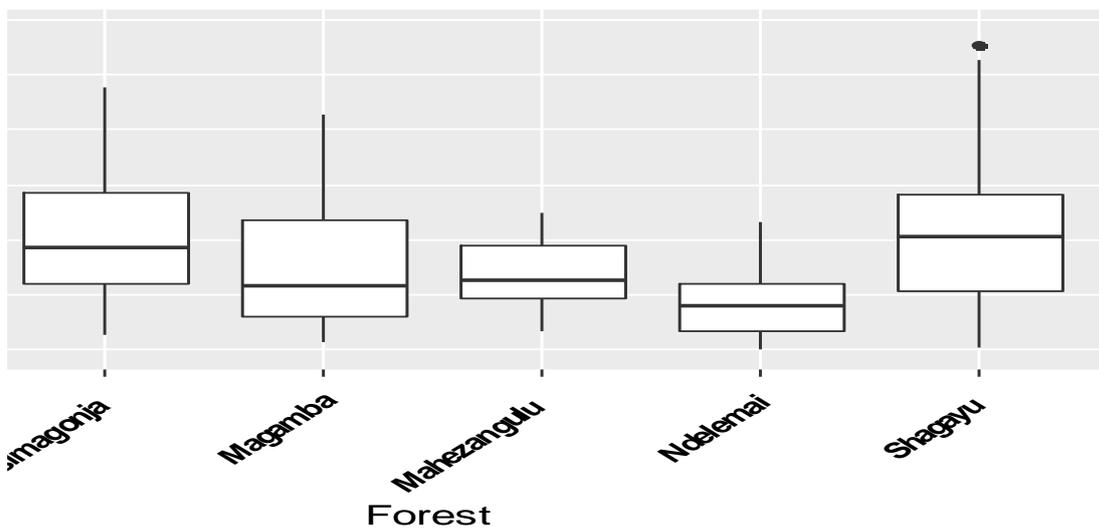| Forest	Number of plots	Minimum	Maximum	Mean	Standard deviation
Balangai	34	7.41	627.33	178.92	112.29
Kisimagonja	46	42.30	716.46	310.94	172.59
Magamba	55	22.66	641.74	230.01	168.81
Mahezangulu	14	53.54	372.74	200.52	93.55
Ndelemai	48	1.31	347.05	129.64	96.79
Shagayu	99	4.60	1125.91	326.05	215.23
All	296	1.31	1125.91	251.17	184.63

Figure 3. AGB distribution for the forests of WUMB. The high dots represent maximum value, the solid middle bar is the median value and lower dot is lower value

Remotely sensed data

Sentinel 2

Cloud- and shadow-free Sentinel-2 bottom of atmosphere (L2A) mosaic for the entire of the WUMB was obtained from the *Sentinel Hub* (Kirches 2018). The mosaics contained the Blue (B02), Green (B03), Red (B04), Red Edge 1 (B05), Red Edge 2 (B06), Red Edge 3 (B07), Near Infrared (B08), Narrow Near Infrared (B08), Shortwave Infrared 2 (B11), Short Wave Infrared 3 (B12). All the bands were ordered at the spatial resolution of 10 m, indicating that, bands with original resolution coarser than 10 m (all bands except B2-B4) were re-sampled to 10 m using the nearest neighbour method, as described in the S2 Global Mosaic User Manual (https://usermanual.readthedocs.io/en/latest/). In additional to the spectral bands, we derived the vegetation indices from the sentinel 2 optical bands using RStoolbox package (Benjamin *et al*. 2019) implemented in the R statistical software. The calculated vegetation indices were specifically chosen to evaluate the potential of the bands in operating in NIR and the red edge spectrum as bands operating at these wavelengths have been found affective in predicting forest characteristics in various studies (e.g. Chen *et al*. 2018, Mauya *et al*. 2019, Malhi *et al*. 2021, Theofanous *et al*. 2021). The vegetation indices were computed as indicated in Table 3. The

Grey Level Co-Occurrence Matrix (GLCM) textural metrics "mean", "variance" and "dissimilarity" (Haralick *et al.* 1973), were also computed for all the spectral bands and indices, using GLCM package (Zvoleff 2020) in R software. In computation of the texture metrics, we selected a window size of 3×3 pixels to ensure that it is closely matching and comparable with the pixels size of the input data (i.e.,10 m): larger window sizes are unlikely to reveal textural vegetation patterns that are relevant for successfully estimating AGB.

Table 3. Names, equations and sources of the vegetation indices derived from Sentinel 2 data

Vegetation Indices	Name	Equations	References
CLG	Green-band Chlorophyll Index	(B07)/(B03-1)	(Gitelson et al. 2003)
NDVI	Normalized Difference Vegetation Index	(B08-B04)/(B08+B04)	(Rouse et al. 1974)
CLRE	Red-edge-band Chlorophyll Index	(B07)/(B05-1)	(Gitelson et al. 2003)
GNDVI	Green Normalized Difference Vegetation Index	(B08-BO3)/(B08+B03)	(Gitelson et al. 1996)
NBRI	Normalized Burn Ratio Index	(B08-B12)/(B08+B12)	(García and Caselles 1991)
NDREI1	Normalized Difference Red Edge Index1	(B06-B05)/(B06+B05)	(Gitelson and Merzlyak 1994)
NDREI2	Normalized Difference Red Edge Index 2	(B07-B05)/(B07+B05)	(Barnes et al. 2000)
RE-NDVI_705	Red-Edge Normalized Difference Vegetation index 3	(B08-B05)/(B08+B05)	(Puletti et al. 2018)
RE-NDVI_740	Red-Edge Normalized Difference Vegetation index 1	(B08-B06)/(B08+B06)	(Fernández-Manso et al. 2016)
RE-NDVI_783	Red-Edge Normalized Difference Vegetation index 2	(B08-B07)/(B08+B07)	(Huete et al. 1997)
SAVI	Soil Adjusted Vegetation Index	$\frac{(B08-B04)*(1+L)}{(BO8+B04+L)}$	(Huete 1988)
EVI	Enhanced Vegetation Index	2.5*(B08-B04)/(BO8+2.4*B04+1)	(Jiang et al. 2008)

Planet Scope

Planet Scope Surface Reflectance Mosaics covering our area of interest was downloaded from https://www.planet.com/basemaps/#/mosaic/planet_medres_visual_2021-06_mosaic/zoom/2.57). The mosaics were optimized for scientific and quantitative analysis with minimum effects of atmospheric sensor characteristics and other artifacts caused by haze, light and topography (Poortinga *et al*. 2021). The mosaics contained four bands which are; Red, Green, Blue and Near-Infrared both with spatial resolution of 4.77m per pixel. We used Bi-Annual mosaic acquired in June 2019 in order to have a closer match with the Sentinel 2 data for comparisons of the performance of the two optical remotely sensed data. For each band, we computed average, mode and standard deviation of the reflectance values. NDVI was also calculated based on the formula by Rouse *et al*. (1974). For each band as well for the NDVI layer, we calculated various texture metrics, including; Mean, Variance, Homogeneity, Contrast, Dissimilarity, Entropy, Second Moment and Correlation were computed using GLCM package in R software.

Extraction of the remotely sensed explanatory variables

In order to ensure spatial overlap between the measured AGB at the 15 m radius field plot and the information acquired from the remotely sensed data, we firstly overlaid the field plots on the remotely sensed data mosaics (i.e. bands, vegetation indices, and texture layers). Secondly, we made a sample plot polygon (i.e., buffer) with the radius of 15 m and extracted the area weighted means

of the pixel values intersecting with the sample plot polygons using the *extract function* in raster package of R software (Hijmans *et al*. 2015). The extracted values for each dataset were then grouped into different categories. For Sentinel-2, we grouped into: 1) band values; 2) vegetation indices; 3) texture of band values; 4) texture of indices; and 5) combination of all variables. For Planet Scope we had: 1) band values; 2) texture of NDVI; and 3) combination of band values and texture of NDVIs.

Model development

Generalized Linear Models (GLMs) and Random forest were used to develop AGB predictive models using remotely sensed data. The models were developed for each dataset (i.e., Sentinel-2 and Planet Scope), as well for each category of predictor variables described above. Three key steps were followed in the modelling processes which include: 1) Variable selection; 2) model development/ fitting; and 3) model validation. The details for each step are described below.

Variable selection and Model fitting

GLMs

A generalized linear Model with gamma distribution and logarithmic link functions (Zuur *et al*. 2013) was used to develop models relating AGB at the plot levels and the remote sensing predictor variables. To ensure that we develop robust models for different datasets and predictor variables, we first performed variable selection as one of the critical steps in modelling the combination of field and remote sensing data. Candidate predictor variables, were selected using "*regsubset*" function implemented in "*leaps*" package (Lumley and Lumley 2013) of the R software. The "*regsubset*" regression performs "all subsets" where all possible variable combinations are considered and ranked based on different scoring criteria ("AIC", "BIC", " Mallow's *Cp* statistics " etc.). In this

study we used BIC, a combination of predictors that minimizes the BIC over all possible subsets, was considered as the best subset for model development. The variable selection was repeated for each category of predictor variables. The best subsets were then used to fit the models and the variables were further assessed based on their significance (i.e., $p<0.05$) and variance inflation factor (VIF). Predictor variables with VIF values greater than 10 were regarded as an indication of multi-collinearity problems (Nelson *et al*. 2017), and were trimmed out from the model.

Random Forest

Random forest (RF) is a non-parametric regression method, which is developed, based on the regression trees algorithm, where predictor variables are split to grow a number of nodes to select the best predictor variable. About two-thirds of the samples (in-bag samples) are used to train the trees and remaining one third (out-of-bag (OOB) samples) are used in an internal cross-validation technique for estimating the OOB error (Belgiu and Drăguţ 2016). The principal behind random forest regression as it is applied in this study is explained in Breiman (2001) and its use for modelling and prediction of forest tree attributes has widely been reported (Hayashi *et al*. 2014, Vafaei *et al*. 2018). A key advantage however, in random forest, is that, greater number of predictor variables of various types (categorical, continuous, binary) can be handled and the relative importance of each predictor variable can be estimated during the model calibration process. Furthermore, RF had ability to identify complex nonlinear relationships between response and predictor variables (Fassnacht *et al*. 2014).

In this analysis, development of the random forest model was done using the selected predictor variables from the VSURF package implement in R software (Genuer *et al*. 2015). The VSURF algorithm first filters out unimportant predictor variables based on random forest mean variable importance

values. Then, an iterative optimization is conducted to select the variables most suitable for predicting the response variable. VSURF suggests two sets of variables, one optimized for interpretation (i.e., some predictors may be redundant but equally important for predicting the response) and another one optimized for prediction (i.e. focusing solely on obtaining a possibly high model fit) (Genuer *et al*. 2015). Here, we selected the variable subset optimized for prediction accuracy and fitted the model with the number of trees (ntrees) fixed to 500 and the mtry-parameter to number of predictors / 3. With this, we followed the findings of earlier investigations which stated that these standard settings for mtry and ntrees obtain good accuracies in most cases (Oshiro *et al*. 2012, Probst *et al*. 2019, Fassnacht *et al*. 2021).

Accuracy assessment

GLMs

Two step approach was done to validate and evaluate the GLMs. In the first step, we evaluated the fits of the models by calculating Akaike's Information Criterion (AIC), root mean square error (RMSE) and relative value (RMSEr) based on the predictions from the model (i.e. internal self-validation). The RMSE and RMSEr, were calculated using the equations below.

$$RMSE = \sqrt{\sum_{I=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n}} \qquad (1)$$

and

$$RMSE_r = \frac{RMSE}{\bar{y}} \times 100\% \qquad (2)$$

Where y_i and $\hat{y}_i$ denote field measured AGB and predicted AGB for plot i, respectively, and $\bar{y}$ denotes mean of the field measured AGB for all plots.

In the second step, to enable comparison among the models developed from different groups of predictor variables, and to understand the models' performance on other datasets as well to compare the performance with the non-parametric method, the models were cross validated using a *k*-fold cross validation. This approach involves randomly dividing the data into k approximately equal folds or groups. Each of these folds is then treated as a validation set in k different iterations. We used a *k*-value of 10 since it's been widely used and shown empirically to yield test error rate estimates that suffer neither from excessively high bias nor from very high variance (James *et al*. 2013, Kuhn and Johnson 2013). The 10-fold cross-validation involves splitting the dataset into 10-subsets. In each fold, one subset is held out for checking the model performance (i.e., the validation set), while the model is trained on all other subsets (i.e. 9). The process is done repeatedly until all the subsets have been used as the validation dataset. The predicted values from all the folds were finally compiled into a table and used to estimate cross validated RMSE (cvRMSE) and RMSEr (cvRMSEr), using the equations presented above, now with predicted values from the 10-fold cross validation.

Random forest

Random forest models were evaluated using both the internal boot strap procedure as well as the *k*-fold cross validation. In the first place, the predictions from the OOB samples were used to compute RMSE and RMSEr. To confirm this, and to have fair comparison with the parametric method, we evaluated the models using the k-fold procedure described above. The predictions from k-fold were then used to compute cvRMSE and RMSEr.

Efficiency of remotely sensed assisted AGB estimation

In order to estimate the relative efficiency (RE) of using remotely sensed assisted AGB estimation as compared to pure field estimates, we computed the variance of both field and remotely sensed AGB estimates, i.e.,

$$RE = \frac{\hat{V}_{field}}{\hat{V}_{remotely\ sensed}} \qquad (3)$$

Where $\widehat{V}_{field}$ is the variance of the pure field based AGB, which is computed using equation below:

$$\widehat{V}_{field} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n(n-1)}. \tag{4}$$

$\widehat{V}_{remotely\ sensed}$ is the variance of the remotely sensed AGB estimation for specific predictor category and data source. This was estimated using the variance estimator of the so-called generalized regression estimator presented in equation 5 below.

$$\widehat{V}_{remotely\ sensed} = \frac{\sum_{i=1}^n (\hat{e}_i - \underline{e})^2}{n(n-1)} \tag{5}$$

Where $\hat{e}_i = y_i - \hat{y}_i$ is the model prediction residual for plot i and $\underline{e} = \frac{\sum_{i=1}^n \hat{e}_i}{n}$ is the mean residual for all plots. Standard error (SE) was computed as the square root of the variance estimates. Values of RE greater than 1.0 indicates higher efficiency of remotely sensed assisted estimates than field-based estimates for a given data source.

AGB prediction map

Finally, the best parametric and non–parametric model was used to predict AGB over the entire area of interest. Since our interest was to see only the spatial prediction of AGB, we used the native resolution of each image. The mean and standard deviation of the AGB predictions from the map were then computed to get an indication of its variability relative to the mean and standard deviation of the field based AGB measurement.

RESULTS

Performance of Parametric Method (GLMs)

Sentinel 2

Models comprising of band values, indices, texture of band values, texture of indices and their combination were developed. The number of variables for the models were ranging from one to four. For all the models, the parameter estimates were significantly different from zero (p<0.05) and the VIF values were <10, indicating acceptable levels of multicollinearity. The AIC values for the models were ranging from 3777 to 3819, with the lowest values obtained using a model with the combination of texture metrics of CLG, NBRI, RE.NDVI as well as BO8 values located in the NIR (Table 4). This implies that, there is improvement in model fits when combining texture and band values. This is further shown by the results from the cross validation where the cvRMSEr of the best model dropped by 3% as compared to the model with band values only (Table 4).

Planet Scope

Two sets of models were developed from Planet Scope data with a maximum of two variables. The best model with lowest AIC values comprised of texture metrics of the NDVI. Combination of all variables i.e bands and texture of indices, resulted into the selection of only textures-based variables (Table 4). This implies that, texture variables have strong statistical relationship with AGB as compared to band values. Results from the cross validation indicated that, the cvRMSEr for best Planet Scope model was 72.1% which was relatively higher as compared to the best model derived from the Sentinel 2 data when using parametric methods (Table 4).

Table 4. Performance of GLMs fitted with predictors from Sentinel 2 and Planet Scope

Data source	Predictor category	Predictors[a]	Calibration				Validation-k-fold	
			AIC	pseudo R^2 (%)	RMSE (Mg/ha)	RMSEr (%)	cvRMSE (Mg/ha)	cvRMSEr (%)
Sentinel2	Band	BO3	3798	7.6	174.5	69.5	175.6	70.0
	Indices	CLG	3819	2.0	183.0	72.93	184.0	73.3
	Texture of Bands	B8A_mean, B03_dis	3809	5.0	179.1	71.3	180.9	72.0
	Texture of indices	CLG_3x3_dis, NBRI_3x3_dis,	3801	7.3	175.5	69.9	177.1	70.6
	All	**CLG_3x3_dis, NBRI_3x3_dis, RE.NDVI_783_3x3_dis, B08**	**3776.6**	**20.0**	**166**	**66**	**168.1**	**67.00**
Planet Planet	Band Values	NIR_avg	3814	3.0	181	72.3	182.3	72.6
	Texture of indices	**NDVI_second moment, NDVI_corr**	**3808**	**5.2**	**179**	**71.6**	**180.8**	**72.1**
	All	Only the above texture variables were selected						

Notes: [a]BO3= Band 03, CLG = Green-band Chlorophyll Index, B8A_mean = Mean texture metric of Band 8A, B03_dis = Dissimilarity texture metric of Band 03, = CLG_3x3_dis= Dissimilarity texture metric of Green-band Chlorophyll Index, NBRI_3x3_dis = Dissimilarity texture metric of Normalized Burn Ratio Index, CLG_3x3_dis = Dissimilarity texture metric of Green-band Chlorophyll Index, RE.NDVI_783_3x3_dis = Dissimilarity texture metric of Red-Edge Normalized Difference Vegetation index 2, B08 = Band 08, NIR_avg = Average value of the Near Infrared, NDVI_second moment = Second moment texture metric of the Normalized Difference Vegetation Index, NDVI_corr = Correlation texture metric of the Normalized Difference Vegetation Index.

Performance of Non –parametric

Sentinel 2

Random forest best regression models, with different sets of predictor variables were developed. Unlike, the parametric method, the best model comprised of the texture of metrics of BO3, B12, BO4, BO5 and BO6.The cvRMSEr for the best random forest (RF) model was slightly higher when compared to the best GLM.

Planet Scope

Based on the cross validation results, model developed using the texture variables of the NDVI turned out to be the best with the RMSEr of 69.8 and cvRMSEr of 72.2 (Table 5). The values were slightly lower compared to the values obtained using the parametric method. Furthermore, compared to Sentinel 2 random forest models, the value is slightly lower (Table 5).

Table 5. Performance of random forest models fitted with predictors from sentinel2 and planet sat data sources.

Data source	Predictor category	Predictors[a]	Out of the bag		Validation-k-fold	
			RMSE (Mg/ha)	RMSEr (%)	cvRMSE (Mg/ha)	cvRMSEr (%)
Sentinel 2	Band Values	BO4, BO12, BO3	182	72.6	181.2	72.2
	Indices	CLRE, NDREI2, RE.NDVI_705, NDREI1, SAVI, GNDVI	192	76.6	191.9	76.5
	Texture of Band values	**B03_3x3_var, B12_3x3_dis, B04_3x3_dis, B04_3x3_var, B05_3x3_dis, B06_3x3_var**	**175.4**	**70.0**	**178.3**	**71.1**

Data source	Predictor category	Predictors[a]	Out of the bag		Validation-k-fold	
			RMSE (Mg/ha)	RMSEr (%)	cvRMSE (Mg/ha)	cvRMSEr (%)
	Texture of indices	CLG_3x3_mea, GNDVI_3x3_mea, NDVI_3x3_dis, CLG_3x3_dis, NDREI2_3x3_var, EVI_3x3_mea, CLRE_3x3_var, NDVI_3x3_mea, GNDVI_3x3_var	186.3	74.0	182.8	72.8
	All	BO4, B12, B04_3x3_var, B03_3x3_var CLG_3x3_mea	178.3	71.0	185.2	73.8
Planet	Band Values	Green_avg, NIR_mod, Blue_mod, NIR_avg, Red_sd	186.2	74.1	182.3	72.7
	All	**NIR_mod, Green_avg, Red_sd, Blue_mod NIR_avg, NDVI_corr NDVI_secmom, NDVI_entropy**	**175.2**	**69.8**	**180.7**	**72.0**

Notes: [a]BO4 = Band 04. , BO12= Band 12 , BO3 = Band 03, B03_3x3_var = Variance texture metric of Band 03, B12_3x3_dis = Dissimilarity texture metric of Band 12, B04_3x3_dis = Dissimilarity texture metric of band 04, B04_3x3_var = Variance texture metric of Band 04, B05_3x3_dis = Dissimilarity texture metric of Band 05, B06_3x3_var = Variance texture metric of Band 06, CLG_3x3_mea = Mean texture metric of Green-band Chlorophyll Index, GNDVI_3x3_mea = Mean texture metric of Green Normalized Difference Vegetation Index, CLG_3x3_dis = Dissimilarity texture metric of Green-band Chlorophyll Index, NDREI2_3x3_var = Variance texture metric of the Normalized Difference Red Edge Index2, EVI_3x3_mea = Mean texture metric of Enhanced Vegetation Index, CLRE_3x3_var = Variance texture metric of Red-edge-band Chlorophyll Index , NDVI_3x3_mea = Mean texture metric of Normalized Difference Vegetation Index, GNDVI_3x3_var = Variance texture metric of Green Normalized Difference Vegetation Index, CLG_3x3_mea = Mean texture metric of Green-band Chlorophyll Index, Green_avg = Average value of Green band, NIR_mod = Mode value of Near Infrared band, Blue_mod = Mode value of Blue band, NIR_avg, = Average value of the Near Infrared band, Red_sd = Standard deviation value of the Red band.

Relative efficiency

Relative efficiency was computed based on best models for each of the dataset. the results showed that the RE values for both of the datasets were >1, indicating that there is gain in precision of the estimates when using either sentinel-2 or planet sat data for AGB estimation. For sentinel 2, the R.E value was 1.2 while for Planet Scope the R.E value was 1.1. This implies that, the efficiency of sentinel 2 is 20% and for planet sat is 10% greater than field-based inventory.

Prediction Maps

AGB prediction maps in Mg/ha based on Sentinel 2 images were developed using best GLMS and random forest model (Figure 4). The mean AGB prediction obtained from the map was 243.729 Mg/ha when using GLM and 257.47 Mg/ha when using the random forest model. Both of the values were close to mean AGB obtained using the field based data (Table 1). The standard deviation of the AGB predictions from the map when using GLM was 183.70 and 84.81 when using random forest
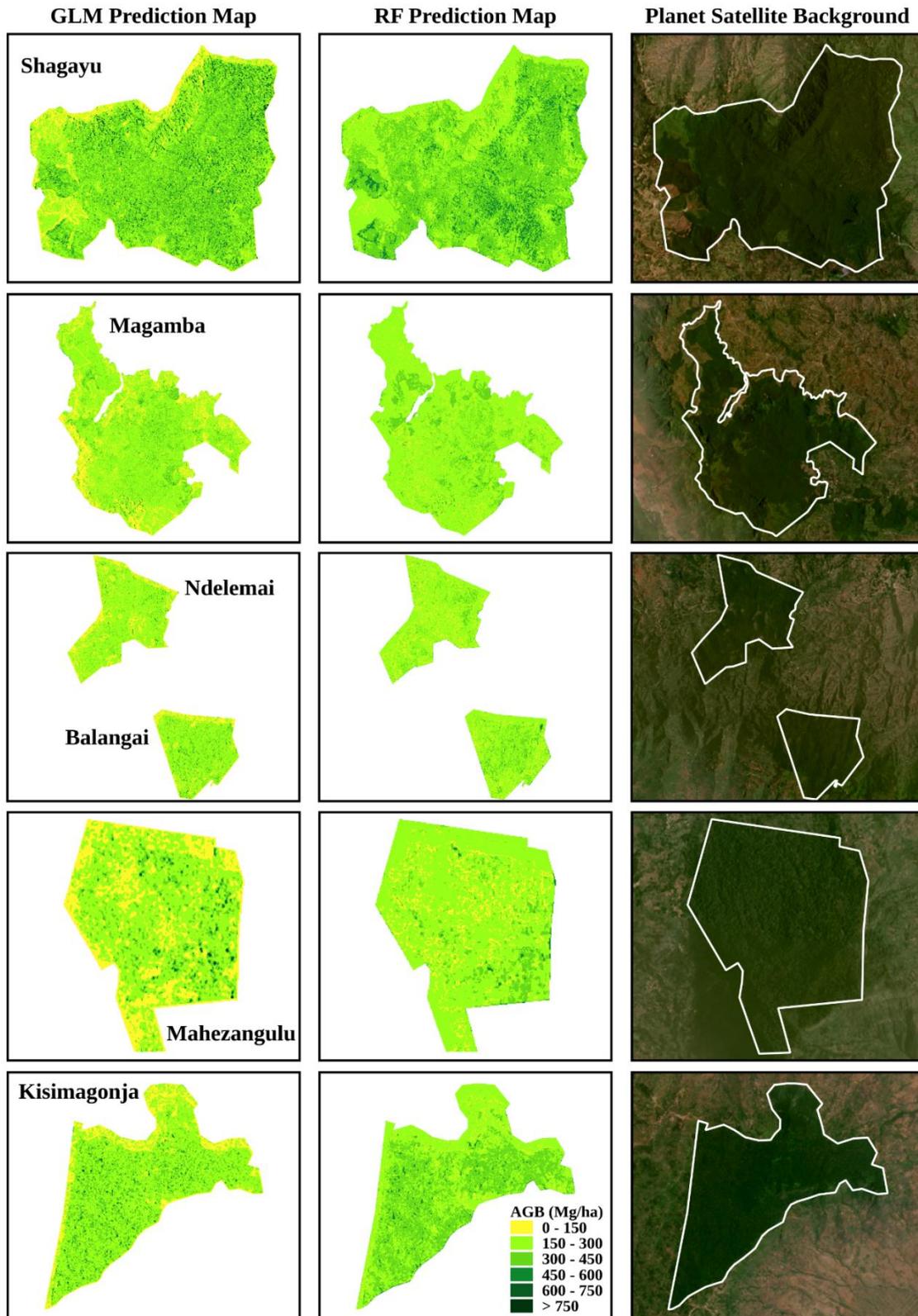
Figure 4. AGB prediction maps for each of the forest based on GLM and RF predictions as well as the planet satellite background

DISCUSSION

In this study, we demonstrated the potential of Sentinel 2 and Planet Scope remotely sensed data for modelling, predicting and mapping AGB in the tropical mountain forests of west Usambara in Tanzania. To our

understanding, this was the first study to be conducted in this biome, using freely available remotely sensed data. Many of the previous studies (e.g. Munishi and Shear 2004, Mbwambo *et al*. 2012, Mauya and Madundo 2021) used only in-situ field based measurements. This study used a comprehensive datasets of 296 field-based plots distributed across the west Usambara montain forests, to calibrate the remotely sensed data. Generally, our results have shown that, Sentinel 2 and Planet Scope data can be used to develop cost-effective methods for AGB estimation within the context of tropical rainforests of Tanzania, at a varying level of accuracy and precision.

The pseudo $-R^2$ and cvRMSEr values for both Sentinel 2 and Planet Scope, are comparable to other studies reported from the tropical and subtropical forests (e.g. Cutler *et al*. 2012, Ghosh and Behera 2018, Pandit *et al*. 2018, 2020, Taddese *et al*. 2020), with slight variations caused by complex forest structure in our study area. Comparing the two data sources, Sentinel-2 resulted into models with less prediction error as compared to Planet Scope data, especially when using the GLMS for developing the statistical models. Similar result has been reported by Taddese *et al*. (2020) who compared Sentinel-2 and Planet Scope data for AGB estimation in the dry Afromontane forests of Ethiopia. Good performance of Sentinel-2 in estimating AGB is widely reported even in other studies which attempted to compare Sentinel-2 with other optical remotely sensed data such as Landsat 8 (e.g. Astola *et al*. 2019, Jha *et al*. 2021). This might be attributed by the rich spectral information of Sentinel-2 covering 13 bands including visible, near-infrared, red-edge, and shortwave infrared wavelengths which are highly correlated with AGB and other forest structure attributes (Li *et al*. 2021).

The prediction accuracy of the AGB models across the two datasets was however varying depending on the sets of the predictors selected from the variable selection process, where for all the data sources inclusion of the texture variables in the model resulted into better AGB prediction accuracy. These results agrees to a large extent with the results from others studies (e.g. Eckert 2012, Vashum and Jayakumar 2012, Ghosh and Behera 2018), which concluded that inclusion of texture based variables in the AGB models improves the predictions accuracy of AGB as compared to reflectance variables only. This is because texture measurement can maximize the discrimination of spatial information independently from the tone and increases the biomass range that can be measured as well as reduce those forest structural differences, which are independent of biomass (Nichol and Sarker 2010). This may partially compensate some of the band saturation problem and hence increasing the correlation between fields measured AGB and the texture variables (Kelsey and Neff 2014). Furthermore, the heterogeneous nature of the forest structural attributes in our study sites caused by difference in age, tree species, disturbances history, height structures, and phenological phases of different tree species, may have contributed to the strong relationship between AGB and texture based variables (e.g. Lu *et al*. 2005, Gao *et al*. 2018).

For Sentinel 2, the best model comprised of Band 08 operating on the NIR and textures of CLR, NBR and CLR which are derived from NIR, red edge, and SWIR wave lengths. NIR domains had been found in other studies to have strong relationship with AGB mainly because of their low red reflectance in dense green vegetation as well as higher penetrating capacity through forested canopies (Bannari *et al*. 1995, Barbosa *et al*. 2014, Kelsey and Neff 2014, Forkuor *et al*. 2020). Likewise, SWIR domain had been widely reported to be important in in AGB modeling (López-Serrano *et al*. 2016) due to the fact that it is more sensitive to moisture and shade components inherent in the forest stand structure and that atmospheric

conditions have less impact on spectral signatures and hence reducing noises in the models (Gao *et al*. 2018). Furthermore, SWIR has higher data saturations value which makes it more suitable for AGB modelling and estimation in dense tropical vegetation with higher average AGB values (Zhao *et al*. 2016b). Same pattern was noted for Planet Scope data, where the best model comprised of the texture of NDVI derived from the NIR and Red, which are commonly reported to have strong relationship with forest attributes.

To demonstrate further the potential of remotely sensed data for larger scale mapping, our study generated a prediction map (Figure 4) based on the GLMs and the random forest models developed using Sentinel-2 data. The average prediction values of the AGB from the map was closer to the field based measurements as well with values reported from other studies based on field inventories (e.g. Mbwambo *et al*. 2012, Mauya and Madundo 2021). GLMs was able to capture most of the variations of AGB in our study area, where sites closer to the borders were indicated to have lower AGB values as compared to the sites far from the borders. This aligned very well with the sidewise Planet base map which shows less crown covers in the borders as compared to the middle parts of the forests. This implies that Sentinel-2 and Planet Scope data can support larger scale forest AGB estimation in our study area with reasonable accuracy and precisions. This is further shown in the RE values, where for the both datasets RE values were greater than one, i.e., 1.2 for Sentinel-2 and 1.1 for the Planet Scope. This implies that, remotely sensed assisted forest inventory is more efficient as compared to pure-field based estimates. To achieve similar precision of a pure field-based estimate relying on simple random sampling, would mean to increase the sample size for the field-based inventory by a factor equivalent to the value of RE, which would have a substantial effect on field inventory costs. Comparing with other studies reported

previously in the rainforests and miombo woodlands using different sets of remotely sensed data, the R.E obtained from this study are relatively smaller. For example, Mauya *et al*. (2015b) reported R.E values of 7.7 when estimating AGB using ALS data with field plot size of 3000 $m^{2.}$ in tropical rainforests of Tanzania. Likewise, Naesset *et al*. (2016) in miombo woodlands of Tanzania reported RE values of 3.6 for ALS, 3.3 for RapidEye 2.8 for InSAR, and 1.3-1.4 for Landsat and PALSAR. Recently, Taddese *et al*. (2020) reported R.E values 1. 4, 1.37 and 1.68 for Landsat 8 and Sentinel 2 for estimating AGB in the dry Afromontane forest of south central Ethiopia. The results from the latter study by Taddese *et al*. (2020), is comparable to our results, the marginal differences might be attributed by more complex forest structure in our study area. Though ALS data resulted into a higher R.E values, the limited spatial and temporal coverage as well as the costs of acquiring and processing ALS data, however, limit their extensive applicability to larger areas. Thus, irrespectively of the relative less accurate results as compared to ALS, our study portrays Planetscope and Sentinel 2 as promising remotely sensed datasets for AGB modelling, estimation and in the tropical rainforests of Tanzania. The two datasets are current freely available and all the computation was done using free statistical software R which makes our approach easily transferable across regions. This is especially useful for the ongoing REDD+ initiatives in Tanzania as well as for general forest management practices. Further studies are however encouraged to look more on the best way of optimizing the efficiency of the two data sources in AGB estimations. Studies on synergistic use of optical and SAR data as well as studies on effect of statistical modelling methods on the prediction accuracy (i.e. parametric vs. non-parametric) may be considered. Likewise, future studies may look further on other key parameters which affect the prediction accuracy e.g. plot sizes, number of field plots as well as stratification by individual forests.

CONCLUSION

To conclude, our study has demonstrated that Sentinel 2 and Planet Scope remotely sensed data can be used to develop cost-effective method for AGB estimation within the context of tropical rainforests of Tanzania. Sentinel 2 AGB predictive models developed using parametric method resulted into a better accuracy as compared those developed using random forest regression. For Planet Scope, AGB models developed using random forest regression resulted into slightly better accuracy as compared to those based on parametric method. Overall, Sentinel 2 resulted into a model with better prediction accuracy as compared to Planet Scope. However, the difference was marginal, and thus our results on this should not be conclusive, rather more efforts should be devoted for comparative tests with more optimized parameters which affect the prediction accuracy. Models developed based on texture variables resulted in better prediction accuracy for both of the data sources and statistical methods. The predicted AGB map based on Sentinel 2 data was able to capture the AGB variability over our study area and further indicated that most part of the WUMFs are characterized by the AGB values of 150-300 Mg/ha. Both Sentinel 2 and Planet Scope improved the precisions of AGB estimates for about 20% and 10% respectively as compared to pure field based measurement. Further studies on the analysis of the cost efficiency of remotely sensed assisted forest assisted inventory versus pure field based methods are highly recommended. Contribution of the environmental and climatic variables in the prediction accuracy of the models may also be investigated.

ACKNOWLEDGMENT

REFERENCES

Adame-Campos, R.L., Ghilardi, A., Gao, Y., Paneque-Gálvez, J. & Mas, J.F. 2019. Variables Selection for Aboveground Biomass Estimations Using Satellite Data: A Comparison between Relative Importance Approach and Stepwise Akaike's Information Criterion. ISPRS International Journal of Geo-Information **8**:245.

Ahmad, A., Gilani, H. & Ahmad, S.R. 2021. Forest Aboveground Biomass Estimation and Mapping through High-Resolution Optical Satellite Imagery - A Literature Review. Forests **12**:914.

Astola, H., Häme, T,. Sirro, L., Molinier, M. & Kilpi, J. 2019. Comparison of Sentinel-2 and Landsat 8 imagery for forest variable prediction in boreal region. Remote Sensing of Environment **223**:257-273.

Bannari, A., Morin, D., Bonn, F. & Huete, A. 1995. A review of vegetation indices. Remote sensing reviews **13**:95-120.

Barbosa, J., Broadbent, E. & Bitencourt, M. 2014. Remote sensing of aboveground biomass in tropical secondary forests: A review. International Journal of Forestry Research **2014**.

Barnes, E., Clarke, T., Richards, S., Colaizzi, P., Haberland, J., Kostrzewski, M., Waller, P., Choi, C., Riley, E. and Thompson, T. 2000. Coincident detection of crop water stress, nitrogen status and canopy density using ground based multispectral data.*in* Proceedings of the Fifth International Conference on

Precision Agriculture, Bloomington, MN, USA.

Belgiu, M., & Drăguţ, L. 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing **114**:24-31.

Benjamin, L., Ned, H. & Jakob, S. 2019. RStoolbox: tools for remote sensing data analysis. R package version 0.2 **6**.

Biswas, S., Huang, Q., Anand, A., Mon, M.S., Arnold, F.E. & Leimgruber, P. 2020. A multi sensor approach to forest type mapping for advancing monitoring of sustainable development goals (SDG) in Myanmar. Remote Sensing **12**:3220.

Boyd, D., & Danson, F. 2005. Satellite remote sensing of forest resources: three decades of research development. Progress in Physical Geography **29**:1-26.

Breiman, L. 2001. Random forests. Machine learning **45**:5-32.

Chen, L., Ren, C., Zhang, B., Wang, Z. & Xi, Y. 2018. Estimation of forest above-ground biomass by geographically weighted regression and machine learning with sentinel imagery. Forests **9**:582.

Cosenza, D.N., Korhonen, L., Maltamo, M., Packalen, P., Strunk, J.L., Næsset, E., Gobakken, T., Soares, P. & Tomé, M. 2021. Comparison of linear regression, k-nearest neighbour and random forest methods in airborne laser-scanning-based prediction of growing stock. Forestry: An International Journal of Forest Research **94**:311-323.

Cutler, M., Boyd, D., Foody, G. & Vetrivel, A. 2012. Estimating tropical forest biomass with a combination of SAR image texture and Landsat TM data: An assessment of predictions between regions. ISPRS Journal of Photogrammetry and Remote Sensing **70**:66-77.

Dang, A.T.N., Nandy, S., Srinet, R., Luong, N.V., Ghosh, S. & Kumar, A.S. 2019. Forest aboveground biomass estimation using machine learning regression algorithm in Yok Don National Park, Vietnam. Ecological Informatics **50**:24-32.

Duncanson, L., Armston, J., Disney, M., Avitabile, V., Barbier, N., Calders, K., Carter, S., Chave, J., Herold, M. and Crowther, T.W. 2019. The importance of consistent global forest aboveground biomass product validation. Surveys in geophysics **40**:979-999.

Eckert, S. 2012. Improved forest biomass and carbon estimations using texture measures from WorldView-2 satellite data. Remote Sensing **4**:810-829.

Fassnacht, F., F. Hartig, H. Latifi, C. Berger, J. Hernández, P. Corvalán, and B. Koch. 2014. Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass. Remote Sensing of Environment **154**:102-114.

Fassnacht, F.E., Poblete-Olivares, J., Rivero, L., Lopatin, J., Ceballos-Comisso, A. & Galleguillos, M. 2021. Using Sentinel-2 and canopy height models to derive a landscape-level biomass map covering multiple vegetation types. International Journal of Applied Earth Observation and Geoinformation **94**:102236.

Fawzy, S., Osman, A.I., Doran, J. & Rooney, D.W. 2020. Strategies for mitigation of climate change: a review. Environmental Chemistry Letters:1-26.

Fernández-Manso, A., Fernández-Manso, O. & Quintano, C. 2016. SENTINEL-2A red-edge spectral indices suitability for discriminating burn

severity. International Journal of Applied Earth Observation and Geoinformation **50**:170-175.

Forkuor, G., Dimobe, K., Serme, I. & Tondoh, J.E. 2018. Landsat-8 vs. Sentinel-2: examining the added value of sentinel-2's red-edge bands to land-use and land-cover mapping in Burkina Faso. GIScience & remote sensing **55**:331-354.

Forkuor, G., Zoungrana, J.B.B., Dimobe, K., Ouattara, B., Vadrevu, K.P. & Tondoh, J.E. 2020. Above-ground biomass mapping in West African dryland forest using Sentinel-1 and 2 datasets-A case study. Remote Sensing of Environment **236**:111496.

Gao, Y., Lu, D., Li, G., Wang, G., Chen, Q., Liu, L. & Li, D. 2018. Comparative analysis of modeling algorithms for forest aboveground biomass estimation in a subtropical region. Remote Sensing **10**:627.

García, M.L. & Caselles, V., 1991. Mapping burns and natural reforestation using Thematic Mapper data. Geocarto International **6**:31-37.

Genuer, R., Poggi, J.M. & Tuleau-Malot, C. 2015. VSURF: an R package for variable selection using random forests. The R Journal **7**:19-33.

Ghosh, S.M., & Behera, M.D. 2018. Aboveground biomass estimation using multi-sensor data synergy and machine learning algorithms in a dense tropical forest. Applied Geography **96**:29-40.

Gitelson, A., & Merzlyak, M.N. 1994. Spectral reflectance changes associated with autumn senescence of Aesculus hippocastanum L. and Acer platanoides L. leaves. Spectral features and relation to chlorophyll estimation. Journal of plant physiology **143**:286-292.

Gitelson, A.A., Gritz, Y. & Merzlyak, M.N. 2003. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. Journal of plant physiology **160**:271-282.

Gitelson, A.A., Kaufman, Y.J. & Merzlyak, M.N. 1996. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. Remote Sensing of Environment **58**:289-298.

Gizachew, B., Solberg, S., Næsset, E., Gobakken, T., Bollandsås, O.M., Breidenbach, J., Zahabu, E. & Mauya, E.W. 2016. Mapping and estimating the total living biomass and carbon in low-biomass woodlands using Landsat 8 CDR data. Carbon balance and management **11**:1-14.

Haralick, R.M., Shanmugam, K., & Dinstein, I.H. 1973. Textural features for image classification. IEEE Transactions on systems, man, and cybernetics:610-621.

Hayashi, R., Weiskittel, A., & Sader, S. 2014. Assessing the feasibility of low-density LiDAR for stand inventory attribute predictions in complex and managed forests of northern Maine, USA. Forests **5**:363-383.

Herold, M., Carter, S., Avitabile, V., Espejo, A.B., Jonckheere, I., Lucas, R., McRoberts, R.E., Næsset, E., Nightingale, J. & Petersen, R. 2019. The role and need for space-based forest biomass-related measurements in environmental management and policy. Surveys in geophysics **40**:757-778.

Hijmans, R.J., Van Etten, J., Cheng, J., Mattiuzzi, M., Sumner, M., Greenberg, J.A., Lamigueiro, O.P., Bevan, A., Racine, E.B. & Shortridge, A. 2015. Package 'raster'. R package **734**.

Huete, A. R. 1988. A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment **25**:295-309.

Huete, A., Liu, H., Batchily, K. & Van Leeuwen, W. 1997. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. Remote Sensing of Environment **59**:440-451.

James, G., Witten, D., Hastie, T. & Tibshirani, R. 2013. An introduction to statistical learning. Springer.

Jha, N., Tripathi N.K., Barbier, N., Virdis, S.G., Chanthorn, W., Viennois, G., Brockelman, W.Y., Nathalang, A., Tongsima, S. & Sasaki, N. 2021. The real potential of current passive satellite data to map aboveground biomass in tropical forests. Remote Sensing in Ecology and Conservation.

Jiang, F., Zhao, F., Ma, K., Li, D. & Sun, H. 2021. Mapping the Forest Canopy Height in Northern China by Synergizing ICESat-2 with Sentinel-2 Using a Stacking Algorithm. Remote Sensing **13**:1535.

Jiang, Z., Huete A.R., Didan, K. & Miura. T. 2008. Development of a two-band enhanced vegetation index without a blue band. Remote Sensing of Environment **112**:3833-3845.

Kelsey, K.C., & Neff, J.C. 2014. Estimates of aboveground biomass from texture analysis of Landsat imagery. Remote Sensing **6**:6407-6422.

Kirches, G. 2018. Sentinel 2 Global Mosaics: Copernicus Sentinel-2 Global Mosaic (S2GM) within the Global Land Component of the Copernicus Land Service. Algorithm Theoretical Basis Document JRC: European Commission.

Koch, B. 2015. Remote Sensing supporting national forest inventories NFA. FAO Knowledge Reference for National Forest Assessments; FAO: Rome, Italy:77-92.

Kuhn, M. & Johnson, K. 2013. Applied predictive modeling. Springer.

Li, C., Zhou, L. & Xu, W. 2021. Estimating Aboveground Biomass Using Sentinel-2 MSI Data and Ensemble Algorithms for Grassland in the Shengjin Lake Wetland, China. Remote Sensing **13**:1595.

Li, Y., Li, M., Li, C. & Liu, Z. 2020. Forest aboveground biomass estimation using Landsat 8 and Sentinel-1A data with machine learning algorithms. Scientific reports **10**:1-12.

López-Serrano, P.M., López-Sánchez, C.A., Alvarez-Gonzalez, J.G. 7 Garcia-Gutierrez, J. 2016. A comparison of machine learning techniques applied to Landsat-5 TM spectral data for biomass estimation. Canadian Journal of Remote Sensing **42**:690-705.

Lovett, J.C. 1996. Elevational and latitudinal changes in tree associations and diversity in the Eastern Arc mountains of Tanzania. Journal of Tropical Ecology:629-650.

Lu, D., Batistella, M. & Moran, E. 2005. Satellite estimation of aboveground biomass and impacts of forest stand structure. Photogrammetric Engineering & Remote Sensing **71**:967-974.

Lumley, T. & Lumley, M.T. 2013. Package 'leaps'. Regression subset selection. Thomas Lumley Based on Fortran Code by Alan Miller. Available online: http://CRAN. R-project. org/package= leaps (Accessed on 18 March 2018).

Malhi, R.K.M., Anand, A., Srivastava, P.K., Chaudhary, S.K., Pandey, M.K., Behera, M.D., Kumar, A., Singh, P. & Kiran, G.S. 2021. Synergistic evaluation of Sentinel 1 and 2 for biomass estimation in a tropical

forest of India. Advances in Space Research.

Masota, A., Bollandsås, O., Zahabu, E. & Eid, T. 2016. 4 Allometric Biomass And Volume Models For Lowland And Humid Montane Forests. Allometric Tree Biomass and Volume Models in Tanzania:35.

Mauya, E. & Madundo, S. 2021. Aboveground biomass and carbon stock of usambara tropical rainforests in Tanzania. Tanzania Journal of Forestry and Nature Conservation **90**:63-82.

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

Mauya, E.W., Hansen, E.H., Gobakken, T., Bollandsås, O.M., Malimbwi, R.E. & Næsset, E. 2015b. 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.

Mauya, E.W., Koskinen, J., Tegel, K., Hämäläinen, J., Kauranne, T. & Käyhkö, N. 2019. Modelling and predicting the growing stock volume in small-scale plantation forests of Tanzania using multi-sensor image synergy. Forests **10**:279.

Mbwambo, L., Eid, T. Malimbwi, R.E., Zahabu, E., Kajembe, G. & Luoga, E. 2012. Impact of decentralised forest management on forest resource conditions in Tanzania. Forests, Trees and Livelihoods **21**:97-113.

Mehtatalo, L., & Mehtatalo, M.L. 2015. Package 'lmfor'.

Mitchell, A.L., Rosenqvist, A. & Mora, B. 2017. Current remote sensing approaches to monitoring forest degradation in support of countries measurement, reporting and verification (MRV) systems for REDD+. Carbon balance and management **12**:1-22.

Msuya, T.S. & Kideghesho, J.R. 2009. The role of traditional management practices in enhancing sustainable use and conservation of medicinal plants in West Usambara Mountains, Tanzania. Tropical Conservation Science **2**:88-105.

Munishi, P. & Shear, T. 2004. Carbon storage in Afromontane rain forests of the Eastern Arc mountains of Tanzania: their net contribution to atmospheric carbon. Journal of Tropical Forest Science:78-93.

Naesset, E., Ørka, H.O., Solberg, S., Bollandsås, O.M., Hansen, E.H., Mauya, E.W., Zahabu, E. Malimbwi, R.E. Chamuya, N. & Olsson, H. 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 Sensing of Environment **175**:282-300.

Näslund, M. 1936. Skogsförsöksanstaltens gallringsförsök i tallskog.

Nelson, R., Margolis, H., Montesano, P., Sun, G., Cook, B., Corp, L., Andersen, H.E., deJong, B., Pellat, F.P. & Fickel, T. 2017. Lidar-based estimates of aboveground biomass in the continental US and Mexico using ground, airborne, and satellite observations. Remote Sensing of Environment **188**:127-140.

Nichol, J.E. & Sarker, M.L.R. 2010. Improved biomass estimation using the texture parameters of two high-

resolution optical sensors. IEEE Transactions on Geoscience and Remote Sensing **49**:930-948.

Nuthammachot, N., Phairuang, W., Wicaksono, P. & Sayektiningsih, T. 2018. Estimating aboveground biomass on private forest using Sentinel-2 imagery. Journal of Sensors **2018**.

Oshiro, T.M., Perez, P.S. & Baranauskas, J.A. 2012. How many trees in a random forest? Pages 154-168 *in* International workshop on machine learning and data mining in pattern recognition. Springer.

Pandit, S., Tsuyuki, S. & Dube, T. 2018. Estimating above-ground biomass in sub-tropical buffer zone community forests, Nepal, using Sentinel 2 data. Remote Sensing **10**:601.

Pandit, S., Tsuyuki, S. & Dube. T. 2020. Exploring the inclusion of Sentinel-2 MSI texture metrics in above-ground biomass estimation in the community forest of Nepal. Geocarto International **35**:1832-1849.

Poortinga, A., Thwal, N. ., Khanal, N., Mayer, T., Bhandari, B., Markert, K., Nicolau, A.P., Dilger, J., Tenneson, K. & 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:100003.

Probst, P., Wright, M.N. & Boulesteix, A.L. 2019. Hyperparameters and tuning strategies for random forest. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **9**:e1301.

Puletti, N., Chianucci, F. & Castaldi, C. 2018. Use of Sentinel-2 for forest classification in Mediterranean environments. Ann. Silvic. Res **42**:32-38.

Qiu, S., He, B., Yin, C. & Liao, Z. 2017. Assessments of Sentinel 2 vegetation red-edge spectral bands for improving land cover classification. Proc. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci **42**:1055-1059.

Rouse, J.W., Haas, R.H., Schell, J.A. & Deering, D.W. 1974. Monitoring vegetation systems in the Great Plains with ERTS. NASA special publication **351**:309.

Sinha, S., Jeganathan, C., Sharma, L. & Nathawat, M. 2015. A review of radar remote sensing for biomass estimation. International Journal of Environmental Science and Technology **12**:1779-1792.

Taddese, H., Asrat, Z., Burud, I., Gobakken, T., Ørka, H.O., Dick, Ø.B. & Næsset, E. 2020. Use of Remotely Sensed Data to Enhance Estimation of Aboveground Biomass for the Dry Afromontane Forest in South-Central Ethiopia. Remote Sensing **12**:3335.

Temesgen, H., Affleck, D., Poudel, K., Gray, A. & Sessions, J. 2015. A review of the challenges and opportunities in estimating above ground forest biomass using tree-level models. Scandinavian Journal of Forest Research **30**:326-335.

Theofanous, N., Chrysafis, I., Mallinis, G., Domakinis, C., Verde, N. & Siahalou, S. 2021. Aboveground Biomass Estimation in Short Rotation Forest Plantations in Northern Greece Using ESA's Sentinel Medium-High Resolution Multispectral and Radar Imaging Missions. Forests **12**:902.

Tian, S., Tanase, M.A., Panciera, R., Hacker, J. & Lowell, K. 2013. Forest biomass estimation using radar and LiDAR synergies. Pages 2145-2148 *in* 2013 IEEE International Geoscience and Remote Sensing Symposium-IGARSS. IEEE.

Vafaei, S., Soosani, J., Adeli, K., Fadaei, H., Naghavi, H., Pham, T.D. & Tien Bui, D. 2018. Improving accuracy estimation of forest aboveground biomass based on incorporation of ALOS-2 PALSAR-2 and sentinel-2A imagery and machine learning: a case study of the Hyrcanian forest area (Iran). Remote Sensing **10**:172.

Vashum, K.T. & Jayakumar, S. 2012. Methods to estimate above-ground biomass and carbon stock in natural forests-a review. J. Ecosyst. Ecogr **2**:1-7.

Vesa, L., Malimbwi, R.E., Tomppo, E., Zahabu, E., Maliondo, S., Chamuya, N., Nssoko, E., Otieno, J., Miceli, G. & Kaaya, A. 2010. NAFORMA Field Manual: biophysical survey. Dar es Salaam, Tanzania: Forestry and Beekeeping Division, Ministry of Natural Resources and Tourism **96**.

Wulder, M.A., White, J.C.' Masek, J.G. Dwyer, J. & Roy, D.P. 2011. Continuity of Landsat observations: Short term considerations. Remote Sensing of Environment **115**:747-751.

Zhao, P., Lu, D., Wang, G., Liu, L., Li, D., Zhu, J. & Yu, S. 2016a. Forest aboveground biomass estimation in Zhejiang Province using the integration of Landsat TM and ALOS PALSAR data. International Journal of Applied Earth Observation and Geoinformation **53**:1-15.

Zhao, P., Lu, D., Wang, G., Wu, C., Huang, Y. & Yu, S. 2016b. Examining spectral reflectance saturation in Landsat imagery and corresponding solutions to improve forest aboveground biomass estimation. Remote Sensing **8**:469.

Zuur, A.F., Hilbe, J.M. & Ieno, E.N. 2013. A Beginner's Guide to GLM and GLMM with R: A Frequentist and Bayesian Perspective for Ecologists. Highland Statistics Limited.

Zvoleff, A. 2020. Package 'glcm'.