



Research Article

Estimating and Mapping Forest Carbon Sequestration in the Mau Forest Complex Using Sentinel-2 Remote Sensing and Machine Learning

Fred Junior Onyango* , Charity Wamwea , Amos Kipkorir Langat

Department of Actuarial Science and Statistics, Jomo Kenyatta University of Agriculture and Technology (JKUAT), Juja, Kenya

Abstract

Accurate forest-carbon estimates are essential for climate mitigation, restoration planning and monitoring, reporting and verification, yet conventional field inventories are expensive and spatially discontinuous. This study evaluated whether freely available Sentinel-2 imagery combined with statistical and machine-learning models can estimate and map forest carbon across the Mau Forest Complex, Kenya. Existing geo-referenced forest inventory data comprising 3,907 measured trees were linked to seven Sentinel-2 vegetation indices: Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Soil-Adjusted Vegetation Index (SAVI), Normalized Difference Moisture Index (NDMI), Normalized Burn Ratio (NBR), red-edge Chlorophyll Index (CI_{re}) and Red-Edge Normalized Difference Vegetation Index (RENDVI). Above-ground biomass was estimated using a pan-tropical allometric equation; below-ground biomass, carbon stock and CO₂ equivalent were subsequently derived using IPCC factors. Ordinary linear regression, a linear mixed-effects model, Random Forest and Extreme Gradient Boosting (XGBoost) were compared using a 70:30 training-test split, with five-fold cross-validation for machine-learning hyperparameter tuning. All vegetation indices were positively associated with above-ground biomass ($p < 0.001$), with CI_{re} (Pearson $r = 0.530$), RENDVI ($r = 0.516$) and NDMI ($r = 0.513$) among the strongest correlates. On the independent test set, XGBoost performed best ($R^2 = 0.473$, RMSE = 0.424, MAE = 0.327), followed by Random Forest ($R^2 = 0.468$, RMSE = 0.426, MAE = 0.329), the mixed-effects model ($R^2 = 0.454$, RMSE = 0.431, MAE = 0.333) and linear regression ($R^2 = 0.436$, RMSE = 0.438, MAE = 0.340). XGBoost reduced RMSE by 3.3% relative to the linear baseline. Residual spatial autocorrelation was negligible (Moran's $I = 0.059$). Spatial prediction identified higher carbon densities in western and south-western portions of the Complex and lower values in more disturbed eastern blocks. The study contributes a reproducible, low-cost benchmark of linear, mixed-effects and ensemble models for spatial forest-carbon monitoring in a heterogeneous African montane forest; the modest gain from nonlinear models shows that optical indices alone limit accuracy, so future improvements will likely require structural information from radar or LiDAR.

Keywords

Mau Forest Complex, Above-ground Biomass, Carbon Sequestration, Sentinel-2, Random Forest, XGBoost, Remote Sensing, REDD+

*Correspondence: Fred Junior Onyango (onyangofredjunior@gmail.com)

Received: 8 September 2026; Accepted: 27 September 2026; Published: 23 October 2026



Copyright: © The Author(s), 2026. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Forests are central to the terrestrial carbon cycle because they remove atmospheric carbon dioxide through photosynthesis and retain carbon in living biomass, dead organic matter and soils. Tropical forests are especially important because of their high productivity and large carbon pools, but their mitigation function is vulnerable to deforestation, degradation and disturbance [8, 16, 17]. Reliable estimation of forest biomass and carbon is consequently important not only for ecological research but also for national greenhouse-gas inventories, REDD+ programmes and Nationally Determined Contributions [11].

The Mau Forest Complex is Kenya's largest indigenous montane forest ecosystem and one of the country's major water towers. It supports biodiversity, regulates regional hydrology and stores substantial quantities of carbon. At the same time, agricultural encroachment, settlement, illegal logging and other disturbances have altered forest condition across the landscape. Earlier work in the Mau ecosystem has documented above-ground biomass through field inventory [18] and examined forest-change detection and carbon trading [27]. Forest-cover area alone cannot fully describe these changes: a forest may remain mapped as forest while losing biomass and carbon through degradation. Management therefore requires estimates of carbon condition and its spatial heterogeneity, not only estimates of forest extent.

Traditional forest inventories remain the reference approach for estimating biomass at plot level. Diameter at breast height, tree height and wood density can be combined through allometric equations [4] to estimate above-ground biomass with strong biological grounding. However, repeated field inventories are labour-intensive, expensive and spatially sparse. They are therefore difficult to use for frequent wall-to-wall monitoring across large, heterogeneous and topographically complex landscapes.

Satellite remote sensing offers a scalable complement to field measurement. Sentinel-2 is particularly attractive for tropical forest applications because it provides freely available multispectral observations [9], including red-edge and shortwave-infrared bands that are sensitive to chlorophyll, canopy structure, moisture and disturbance. Conventional indices such as NDVI remain useful indicators of greenness, but they can saturate in dense canopies [13, 25]. Moisture-sensitive and red-edge indices may therefore contain additional information relevant to biomass estimation [10, 13].

Machine-learning methods can further exploit satellite information by modelling nonlinear relationships and interactions that are difficult to represent with conventional regression. Random Forest [3] is robust to noise and nonlinearities, while XGBoost [5] combines sequential tree boosting with regularisation and often provides strong predictive performance. A recent review of machine-learning and remote-sensing approaches to forest carbon-stock estimation found Random Forest to be the most frequently applied algorithm and

XGBoost to perform best in most studies in which it was compared with other methods [26]. Nevertheless, the added value of flexible algorithms depends on the information content of the predictors. Comparing them directly with linear and mixed-effects baselines is therefore important for determining whether model complexity produces meaningful gains.

Evidence from Sentinel-2 studies in tropical and montane forests is nevertheless mixed. In the Yayu afro-montane forest of Ethiopia, Sentinel-2 red-edge and biophysical variables predicted above-ground biomass well (R^2 of about 0.74) from a small plot network [24], whereas in the Usambara mountains of Tanzania Sentinel-2 models explained only about 20% of the variance in above-ground biomass and generalized linear models produced smaller prediction errors than Random Forest [21, 22]. Elsewhere, Random Forest was selected as the most suitable algorithm for aboveground carbon estimation in Yunnan Province, China [6], and XGBoost has been used with Landsat and national forest inventory data, with kriging of residuals improving accuracy [19], and with Sentinel-2 to map forest structure [14]. Global and pan-tropical biomass and carbon maps [1, 2, 17, 29] provide valuable context but are coarse relative to plot and stand scales, and they can diverge substantially from plot-based estimates [23]; relationships between vegetation indices and above-ground biomass have also been examined in Nepal [28] and Thailand [30]. Together, this literature shows that the value of Sentinel-2 indices and of flexible algorithms depends strongly on forest structure, plot size and sampling design, which motivates a like-for-like benchmark in a heterogeneous Kenyan montane forest complex.

Despite growing use of remote sensing for forest-cover monitoring in Kenya, high-resolution carbon mapping that integrates Sentinel-2 vegetation indices, field-derived biomass and modern machine learning remains limited for the Mau Forest Complex. This study addressed that gap by linking existing forest inventory observations to Sentinel-2 predictors, comparing four modelling approaches, and using the best-performing model to generate spatially explicit carbon estimates. The contributions of the study are fourfold. First, it links a large tree-level inventory (3,907 trees) to seven Sentinel-2 indices representing greenness, canopy background, moisture, disturbance and red-edge responses in a heterogeneous montane forest complex. Second, it benchmarks linear, mixed-effects and ensemble models on a common independent test set, which shows how much of the achievable accuracy is attributable to algorithm choice rather than to predictor information. Third, it applies a block-stratified offset correction and residual spatial-autocorrelation diagnostics to the plot-pixel linkage and the predictions. Fourth, it separates spatial carbon mapping from temporal inference by treating the 2020-2024 comparison explicitly as exploratory, and it documents a reproducible open-data workflow relevant to REDD+ and MRV. The objectives were to: (i) quantify relationships between

above-ground biomass and satellite-derived vegetation indices; (ii) fit linear, mixed-effects, Random Forest and XGBoost models; (iii) compare their predictive performance and reliability; and (iv) map spatial patterns and exploratory temporal variation in carbon across the Mau Forest Complex.

2. Materials and Methods

2.1. Study Area

The study was conducted in the Mau Forest Complex in the western highlands of Kenya, approximately between 0°10'S and 0°45'S and 35°15'E and 36°00'E. The Complex covers about 273,300 ha and spans rolling highlands, ridges and river valleys at elevations of approximately 1,800-3,000 m above sea level. The climate is humid tropical montane, with annual rainfall of roughly 1,200-2,000 mm and a bimodal rainfall regime. The forest comprises Afro-montane evergreen forest, bamboo, secondary forest and restoration areas, creating substantial variation in canopy condition, disturbance history and biomass.

2.2. Field Inventory Data and Carbon Estimation

Field reference information came from an existing forest-restoration monitoring dataset in the South-West Mau block and related sampling strata. The dataset contained measurements from 3,907 individual trees distributed across geo-referenced restoration plots. Recorded structural attributes included diameter at breast height (DBH), total tree height and species information. Species-specific wood-density values were assigned using established reference databases.

Above-ground biomass (AGB) was estimated using the pan-tropical allometric framework of Chave et al. [4], incorporating DBH, tree height and wood density. Below-ground biomass (BGB) was calculated using a root-to-shoot ratio of 0.24. Total biomass was the sum of AGB and BGB. Carbon stock was estimated as 47% of total biomass, and carbon dioxide equivalent (CO₂e) was obtained using a molecular-weight conversion factor of 3.67, consistent with IPCC guidance [15]. Plot-level carbon estimates formed the response used in predictive modelling.

2.3. Sentinel-2 Imagery and Vegetation Indices

Sentinel-2 surface-reflectance imagery [9] was processed in Google Earth Engine [12]. Preprocessing included atmospheric correction as represented in the surface-reflectance product, cloud masking, harmonisation and resampling. Seven vegetation indices were derived (Table 1): NDVI, EVI, SAVI, NDMI, NBR, CI_{re} and RENDVI. These indices were selected to represent complementary dimensions of vegetation greenness, canopy background, moisture, disturbance, chlorophyll

concentration and red-edge response [13, 25].

Table 1. Sentinel-2 vegetation indices used as predictors and their primary sensitivities.

Index	Primary sensitivity
NDVI	Vegetation greenness and canopy density
EVI	Greenness with atmospheric/background correction
SAVI	Vegetation signal with soil-background adjustment
NDMI	Canopy moisture status
NBR	Moisture and disturbance/burn response
CI _{re}	Red-edge chlorophyll response
RENDVI	Red-edge vegetation structure/chlorophyll

2.4. Spatial Harmonisation and Plot-pixel Linkage

Satellite predictors were spatially linked to geo-referenced field observations. To reduce systematic spatial bias associated with illumination, atmospheric and adjacency effects, the study applied a block-stratified additive offset correction aligned with forest stations, agro-ecological zones and restoration-management blocks. For each block, an additive offset was estimated from the mean difference between reference and raw satellite-derived index values at plot locations. Blocks with fewer than three reference plots borrowed information from adjacent blocks using inverse-distance weighting. This transparent correction was selected because it could be implemented consistently within the Google Earth Engine workflow and because it preserved the ecological stratification of the sampling design.

2.5. Statistical Analysis and Predictive Modelling

Descriptive statistics were used to summarise structural variables, carbon metrics, and Sentinel-2 vegetation indices. Continuous variables were summarised using the mean, standard deviation, median, minimum, and maximum, as appropriate. The arithmetic mean was defined as:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

where x_i represents the value of the variable for observation i and n is the total number of observations.

The sample standard deviation was calculated as:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

where s represents the standard deviation and $\bar{x}$ is the sample mean.

Pearson and Spearman correlation coefficients were used to quantify linear and monotonic relationships, respectively, between above-ground biomass (AGB) and the Sentinel-2 vegetation indices. Pearson's correlation coefficient was calculated as:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

where x_i and y_i represent paired observations and $\bar{x}$ and $\bar{y}$ are their respective sample means.

Spearman's rank correlation coefficient was calculated as:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

where d_i represents the difference between the ranks of the paired observations and n is the number of paired observations.

Variance inflation factors (VIFs) were calculated to characterise multicollinearity among the Sentinel-2 predictors:

$$VIF_j = \frac{1}{1 - R_j^2}$$

where R_j^2 is the coefficient of determination obtained by regressing predictor j on all the remaining predictors. Higher VIF values indicated greater multicollinearity among predictors.

The modelling dataset was divided using stratified random sampling into 70% training and 30% independent test data, preserving representation across restoration stages and ecological conditions. The training dataset was used for model development and hyperparameter tuning, whereas the test dataset was reserved for independent evaluation of predictive performance.

Four models were evaluated: ordinary least squares (OLS) linear regression, a linear mixed-effects model with forest block as a random intercept, Random Forest regression, and XGBoost regression.

The OLS model was specified as:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \varepsilon_i$$

where Y_i is the observed response for observation i , β_0 is the intercept, $\beta_1, \dots, \beta_p$ are regression coefficients, $X_{i1}, \dots, X_{ip}$ represent the Sentinel-2 predictors, and ε_i is the random error term.

To account for potential differences among forest blocks, the linear mixed-effects model included a forest-block-specific random intercept:

$$Y_{ik} = \beta_0 + \beta_1 X_{i1k} + \dots + \beta_p X_{ipk} + u_k + \varepsilon_{ik}$$

where u_k represents the random effect associated with forest block k . The random effects and residual errors were assumed to follow normal distributions:

$$u_k \sim N(0, \sigma_u^2)$$

$$\varepsilon_{ik} \sim N(0, \sigma^2)$$

Mixed-effects model parameters were estimated using restricted maximum likelihood (REML).

Random Forest regression [3] was used to capture nonlinear relationships and interactions among the vegetation indices. Predictions were obtained by averaging predictions across T regression trees:

$$\hat{Y}_i^{\text{RF}} = \frac{1}{T} \sum_{t=1}^T f_t(X_i)$$

where $f_t(X_i)$ is the prediction from tree t for observation i and T is the total number of trees.

XGBoost [5] was implemented as a sequential ensemble of regression trees. Its prediction can be represented as:

$$\hat{Y}_i = \sum_{k=1}^K f_k(X_i)$$

where K is the number of boosting iterations and f_k represents the regression tree fitted at iteration k .

Random Forest and XGBoost hyperparameters were selected using five-fold cross-validation within the training dataset. Variable importance was evaluated using permutation importance for Random Forest and gain-based importance for XGBoost.

Predictive performance was evaluated exclusively on the held-out 30% test dataset using the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and prediction bias.

The coefficient of determination was calculated as:

$$R^2 = 1 - \frac{\sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - \bar{Y})^2}$$

where Y_i is the observed value, $\hat{Y}_i$ is the predicted value, and $\bar{Y}$ is the mean of the observed values.

RMSE was calculated as:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum (Y_i - \hat{Y}_i)^2}$$

MAE was calculated as:

$$\text{MAE} = \frac{1}{n} \sum |Y_i - \hat{Y}_i|$$

Prediction bias was calculated as:

$$\text{Bias} = \frac{1}{n} \sum (\hat{Y}_i - Y_i)$$

Higher R^2 and lower RMSE and MAE values indicated better predictive performance, while bias values closer to zero indicated less systematic over- or under-prediction.

Spatial dependence remaining in the residuals of the best-performing model was assessed using Moran's I with inverse-distance spatial weights:

$$I = \frac{n}{W} \times \frac{\sum \sum w_{ij}(e_i - \bar{e})(e_j - \bar{e})}{\sum (e_i - \bar{e})^2}$$

where e_i and e_j are residuals at locations i and j , w_{ij} represents the spatial weight between the two locations, $\bar{e}$ is the mean residual, and:

$$W = \sum \sum w_{ij}$$

A Moran's I value close to zero was interpreted as evidence of little residual spatial autocorrelation.

2.6. Spatial Prediction and Exploratory Temporal Assessment

The best-performing model was applied to the complete Sentinel-2 predictor surfaces to generate continuous estimates of forest carbon density. For a spatial location s , predicted carbon can be expressed generally as:

$$\hat{C}(s) = \hat{f}(X_1(s), X_2(s), \dots, X_p(s))$$

where $\hat{C}(s)$ represents predicted carbon density at location s , $\hat{f}$ represents the fitted predictive model, and $X_1(s), \dots, X_p(s)$ represent the Sentinel-2 vegetation indices at that location.

The model was applied across the Sentinel-2 surfaces to generate wall-to-wall predictions at 10 m spatial resolution. Predicted carbon was subsequently summarised by forest block to identify areas characterised by relatively high carbon stocks and areas potentially requiring greater conservation or restoration attention.

Temporal patterns from 2020 to 2024 were explored using plot-year cohorts and interpolated carbon surfaces. The difference in predicted carbon between two periods was expressed as:

$$\Delta C(s) = \hat{C}_{t_2}(s) - \hat{C}_{t_1}(s)$$

where $\hat{C}_{t_1}(s)$ and $\hat{C}_{t_2}(s)$ represent predicted carbon density at location s during the earlier and later periods, respectively.

For descriptive purposes, the annualised rate of change was expressed as:

$$r_C(s) = \frac{\hat{C}_{t_2}(s) - \hat{C}_{t_1}(s)}{t_2 - t_1}$$

Positive values indicated apparent carbon gains, whereas negative values indicated apparent carbon losses.

Importantly, different plots contributed observations in different years rather than the same plots being repeatedly measured from 2020 to 2024. The temporal component was therefore interpreted as an exploratory space-for-time comparison rather than a direct longitudinal estimate of carbon sequestration. Consequently, the calculated changes describe landscape-level temporal patterns and should not be interpreted as definitive biological sequestration rates. Estimation of true temporal sequestration would require repeated field measurements at the same geo-referenced plots over time.

2.7. Ethics and Reproducibility

The study used open-access Sentinel-2 imagery and secondary forest-inventory data collected under applicable research permissions and data-sharing arrangements. No primary human-subject data were collected. Satellite processing and modelling workflows were implemented using reproducible scripts in Google Earth Engine and statistical software, including R and Python.

3. Results

The results are presented in the order of the study objectives: the structure and carbon distribution of the field data (Section 3.1), the distribution of the Sentinel-2 indices (Section 3.2), their relationships with above-ground biomass (Section 3.3; objective i), model performance (Section 3.4; objectives ii and iii), spatial patterns (Section 3.5; objective iv) and the exploratory temporal comparison (Section 3.6).

3.1. Forest Structure and Carbon Distribution

The field measurements showed marked heterogeneity in forest structure and carbon storage (Table 2). Mean DBH was 7.80 cm (SD 4.74; median 6.20; range 3.00–29.30), while mean tree height was 7.20 m (SD 4.64) and mean wood density was 0.612 g cm⁻³ (SD 0.175). Biomass and carbon variables were strongly right-skewed. Mean AGB per plot was 1,142.1 kg, whereas the median was only 291.6 kg; carbon stock averaged 665.6 kg with a median of 169.9 kg. The skewness of AGB, BGB, total biomass, carbon stock and CO₂e was 5.56, indicating that a relatively small number of high-biomass observations contributed disproportionately to total carbon storage.

Table 2. Descriptive statistics of field-measured structural variables and derived biomass and carbon metrics.

Variable	Mean	SD	Min	Median	Max	Skew
DBH (cm)	7.80	4.74	3.00	6.20	29.30	1.69
Tree height (m)	7.20	4.64	1.50	6.36	20.34	0.55
Wood density (g cm ⁻³)	0.612	0.175	0.335	0.591	1.014	0.33
AGB per plot (kg)	1,142.1	2,666.1	6.3	291.6	31,148.0	5.56
BGB per plot (kg)	274.1	639.9	1.5	70.0	7,475.5	5.56
Total biomass (kg)	1,416.2	3,305.9	7.9	361.5	38,623.5	5.56
Carbon stock (kg)	665.6	1,553.8	3.7	169.9	18,153.1	5.56
CO ₂ e (kg)	2,442.9	5,702.4	13.6	623.6	66,621.7	5.56

3.2. Distribution of Sentinel-2 Vegetation Indices

NDVI was high across the sampled forest (Table 3; mean 0.814, SD 0.073), consistent with generally dense vegetation

but also suggesting saturation in high-biomass areas. Moisture-sensitive indices showed wider ecological variation: NDMI ranged from -0.156 to 0.536 and NBR from -0.074 to 0.786. CIre had the largest numerical dispersion (mean 2.756, SD 0.671), reflecting strong variation in red-edge chlorophyll response.

Table 3. Descriptive statistics of the Sentinel-2 vegetation indices at the sampled locations.

Index	Mean	SD	Min	Median	Max
NDVI	0.814	0.073	0.276	0.833	0.907
EVI	0.475	0.079	0.148	0.483	0.718
SAVI	0.448	0.065	0.160	0.456	0.618
NDMI	0.306	0.081	-0.156	0.324	0.536
NBR	0.586	0.101	-0.074	0.616	0.786
CIre	2.756	0.671	0.517	2.784	4.970
RENDVI	0.569	0.067	0.197	0.580	0.713

3.3. Relationships Between AGB and Vegetation Indices

All seven indices were positively correlated with AGB at $p < 0.001$ (Figure 2). CIre showed the strongest Pearson correlation ($r=0.530$), followed by RENDVI ($r=0.516$) and NDMI

($r=0.513$). The correlation matrix (Figure 1) also showed very strong intercorrelation among several spectral predictors, including NDMI-NBR ($r \approx 0.98$), CIre-RENDVI ($r \approx 0.98$), EVI-SAVI ($r \approx 0.99$) and NDVI-NBR ($r \approx 0.97$). Correspondingly, VIF values were high (Table 4), ranging from 29.3 for NDVI to 107.8 for SAVI. This severe multicollinearity limits coefficient-level interpretation in conventional regression but is less problematic for tree-based prediction.

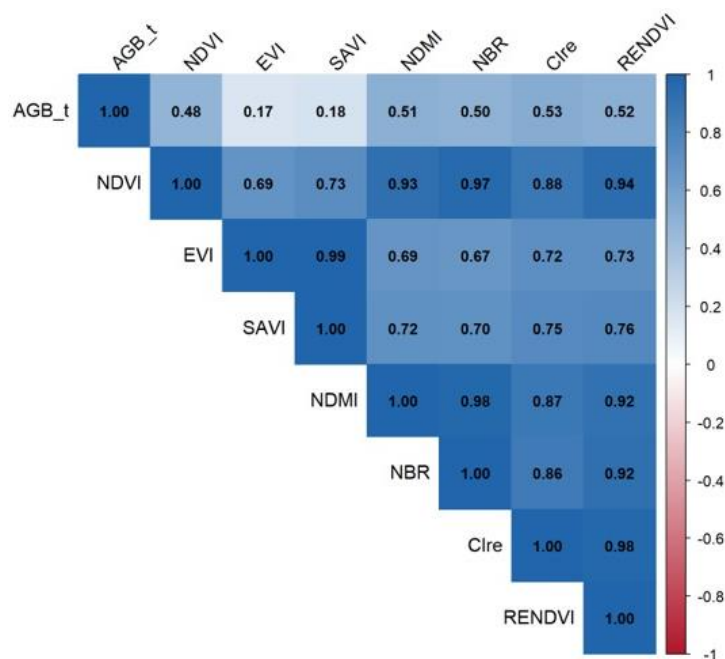


Figure 1. Correlation matrix among above-ground biomass and Sentinel-2 vegetation indices.

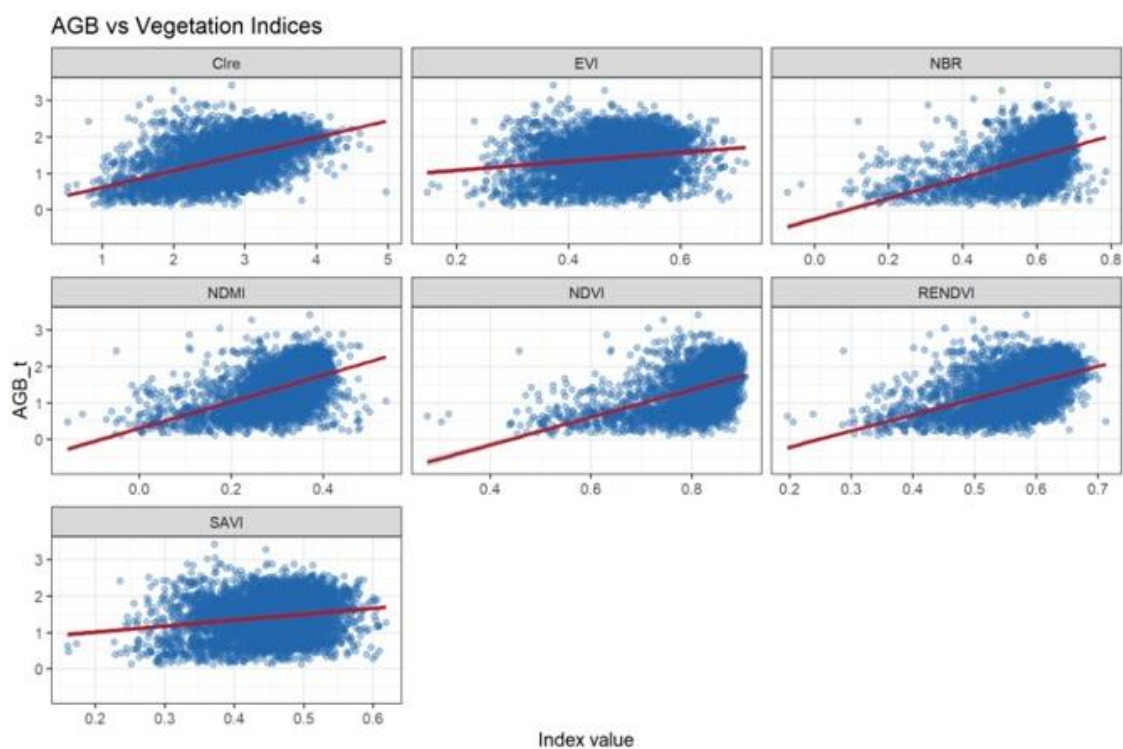


Figure 2. Bivariate relationships between above-ground biomass and the seven vegetation indices.

Table 4. Variance inflation factors (VIFs) of the Sentinel-2 vegetation indices.

Index	VIF
SAVI	107.8
RENDVI	101.9
EVI	95.6

Index	VIF
NBR	59.2
CIre	52.8
NDMI	30.0
NDVI	29.3

3.4. Model Tuning, Importance and Predictive Performance

The tuned Random Forest used 500 trees, mtry=2, the extratrees split rule and a minimum node size of 10. The tuned XGBoost model used eta=0.03, maximum depth=3, minimum child weight=1, subsample=0.70, column subsample=0.85, gamma=0, lambda=1 and 245 boosting rounds. In the mixed-

effects model, the between-block standard deviation was 0.146 and the residual standard deviation was 0.424, confirming non-negligible block-level variability.

Variable-importance rankings differed slightly by algorithm but converged on the importance of moisture-sensitive and red-edge information (Table 5). Random Forest ranked NBR first, followed by NDMI, CIre and RENDVI. XGBoost ranked NDMI first, followed by NBR, SAVI and CIre. NDVI was the least important predictor in XGBoost, consistent with saturation in dense forest.

Table 5. Variable importance of the Sentinel-2 indices in the Random Forest (permutation importance) and XGBoost (gain) models.

Rank	Random Forest permutation importance	XGBoost gain
1	NBR (0.118)	NDMI (0.278)
2	NDMI (0.111)	NBR (0.228)
3	CIre (0.098)	SAVI (0.164)
4	RENDVI (0.095)	CIre (0.107)
5	NDVI (0.070)	EVI (0.085)
6	SAVI (0.056)	RENDVI (0.081)
7	EVI (0.048)	NDVI (0.057)

On the independent test set, all models produced moderate predictive accuracy (Table 6). XGBoost achieved the best performance ($R^2=0.473$, RMSE=0.4237, MAE=0.3272), followed closely by Random Forest ($R^2=0.468$, RMSE=0.4258, MAE=0.3286). The mixed-effects model ($R^2=0.454$,

RMSE=0.4311) improved on the linear baseline ($R^2=0.436$, RMSE=0.4382). Relative to OLS, XGBoost reduced RMSE by 3.3%, Random Forest by 2.8% and the mixed-effects model by 1.6%. Bias was close to zero for all four models.

Table 6. Predictive performance of the four models on the independent test set (30% hold-out).

Model	R^2	RMSE	MAE	Bias	RMSE change vs OLS
XGBoost	0.473	0.4237	0.3272	+0.006	-3.3%
Random Forest	0.468	0.4258	0.3286	+0.005	-2.8%
Linear mixed-effects	0.454	0.4311	0.3334	+0.002	-1.6%
OLS baseline	0.436	0.4382	0.3395	-0.001	0.0%

Moran’s I for XGBoost test-set residuals was 0.059, indicating negligible residual spatial autocorrelation. This suggests that the vegetation-index predictors together with the block-stratified correction captured most of the systematic spatial structure available in the modelling data.

3.5. Spatial Distribution of Predicted Carbon

Application of XGBoost produced a continuous predicted

carbon surface with pronounced spatial heterogeneity (Figure 3). High-carbon areas, commonly around 90-115 t C ha⁻¹, formed a broad belt along western and south-western portions of the study landscape and occurred in additional localised patches. Lower predicted carbon was more common in eastern areas associated with greater historical disturbance.

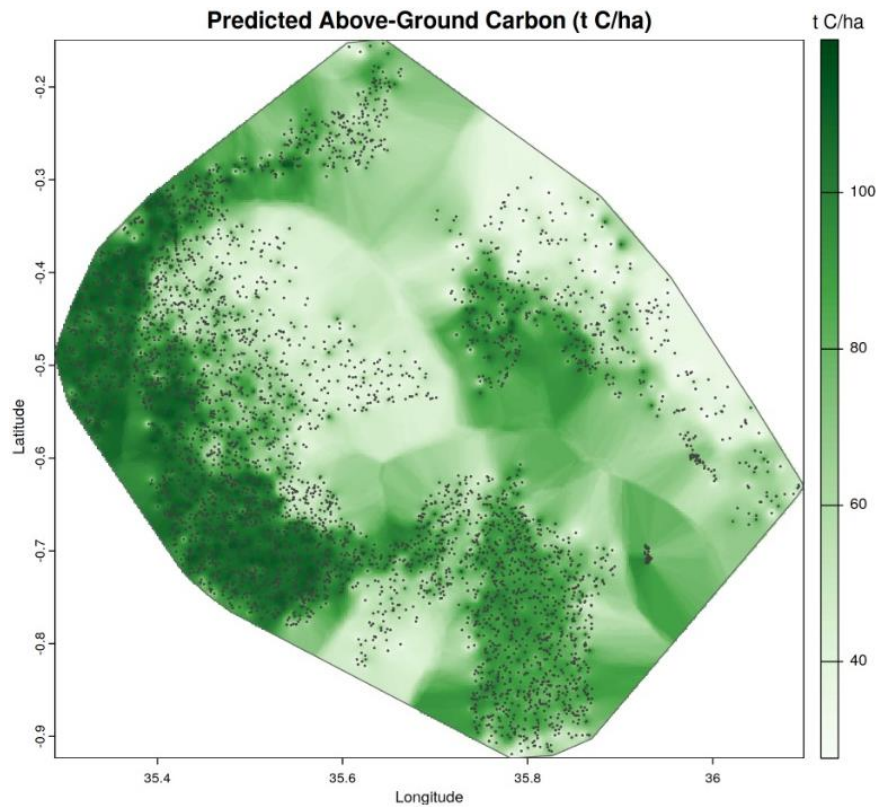


Figure 3. Predicted above-ground carbon density across the Mau Forest Complex study landscape.

Block-level aggregation reinforced this west-east pattern (Table 7). Transmara had the highest mean predicted carbon per plot (0.96 t), followed by Southern Mau (0.85 t) and Ol

Pusimoru (0.84 t). Eastern Mau (0.68 t) and Mau Narok (0.66 t) ranked lowest. The landscape median carbon density was approximately 85-90 t C ha⁻¹.

Table 7. Mean predicted carbon by forest block (t per plot) and rank.

Forest block	Mean predicted carbon (t per plot)	Rank
Transmara	0.96	1
Southern Mau	0.85	2
Ol Pusimoru	0.84	3
Maasai Mau	0.81	4
South West Mau	0.81	5
Western Mau	0.80	6

Forest block	Mean predicted carbon (t per plot)	Rank
Eastern Mau	0.68	7
Mau Narok	0.66	8

3.6. Exploratory Temporal Pattern, 2020–2024

Median predicted carbon density across plot-year cohorts remained broadly stable at approximately 85–90 t C ha⁻¹, with a shallow dip around 2022 and recovery by 2024. A simple cohort-level trend was slightly negative (about −0.48 t C ha⁻¹ yr⁻¹; $p \approx 0.05$), while differencing the interpolated 2020 and 2024 surfaces yielded a mean change of approximately −0.30 t C ha⁻¹ yr⁻¹, with about 44% of the mapped area showing gains. However, these estimates must not be interpreted as repeated-measurement sequestration rates because the yearly cohorts comprised different plot locations. The most defensible conclusion is that landscape carbon was broadly stable over the study window, with localised gains and losses.

4. Discussion

4.1. Principal Findings and Their Relation to Previous Studies

This study evaluated whether open Sentinel-2 vegetation indices and machine learning can estimate and map forest carbon in a heterogeneous montane forest complex. The results show that the indices carry useful but incomplete information: all seven were positively related to above-ground biomass, yet the best model explained less than half of the variance in the independent test data ($R^2 = 0.473$). Three findings are particularly important: carbon storage was highly uneven and right-skewed; moisture-sensitive and red-edge indices carried more predictive information than broad-band greenness; and machine-learning models improved on linear baselines only marginally. Each finding is interpreted below in the light of previous work.

The strong right skew in AGB and carbon (skewness 5.56; Table 2) is ecologically plausible. Tropical forest carbon is often concentrated in a small number of large stems, whereas regenerating and disturbed stands contain many small stems; the median DBH in this restoration-monitoring dataset was only 6.2 cm. A skewed response is also difficult for any regression model to reproduce at its extremes, because predictions tend to regress toward the mean, over-predicting low values and under-predicting the highest values. A similar smoothing effect was reported for Sentinel-2 mapping in Tanzanian montane forest, where the predicted biomass map showed a lower standard deviation than the field data [21]. Such compression matters for carbon accounting because the

few high-biomass observations contribute disproportionately to total stocks.

4.2. Predictive Information in Sentinel-2 Indices

CI_{re}, RENDVI, NDMI and NBR repeatedly emerged as informative predictors in both the correlation and the variable-importance analyses (Section 3.3 and Table 5). Red-edge indices respond to chlorophyll concentration and canopy properties and are less prone to saturation in dense vegetation than red/near-infrared indices such as NDVI [13, 25], while NDMI and NBR incorporate shortwave-infrared reflectance that is sensitive to canopy water content, disturbance and structural condition. This agrees with Muhe and Argaw [24], who found that a red-edge index (IRECI) was the only conventional vegetation index strongly correlated with above-ground biomass in the Yayu afro-montane forest of Ethiopia, and with evidence that narrow-band and red-edge indices reduce saturation at high canopy density [25]. It differs in one respect: Muhe and Argaw [24] reported their strongest correlations ($r = 0.65$ – 0.74) for biophysical variables such as leaf area index and the fraction of absorbed photosynthetically active radiation, which were not tested here and should be evaluated in the Mau Forest Complex. The correlations in the present study were moderate (r of about 0.5), and differences in plot design, sample size and forest structure between the two studies may contribute to this contrast.

The predictor set was, however, severely collinear (VIF 29.3–107.8; Table 4), as expected because several indices are transformations of overlapping Sentinel-2 bands. For explanatory regression this complicates the interpretation of individual coefficients; Muhe and Argaw [24], for example, excluded collinear variables before fitting their regression model. Here all seven indices were retained so that the four algorithms could be compared on identical inputs. Tree-based ensembles accommodate redundant predictors more readily than OLS, but variable importance should be interpreted by spectral family (red-edge, moisture/shortwave-infrared, greenness) rather than as a causal ranking of individual indices. Variable selection or dimensionality reduction is a natural extension for future work.

4.3. Model Performance and the Limits of Optical Predictors

XGBoost achieved the lowest RMSE and highest R^2 , with Random Forest close behind (Table 6). This ordering is consistent with a recent review in which XGBoost outperformed

other algorithms in most studies where it was compared [26], with applications of XGBoost to biomass and carbon mapping from Landsat and inventory data [19] and from Sentinel-2 [14], and with the selection of Random Forest for aboveground carbon estimation in Yunnan Province, China [6]. However, the 3.3% RMSE reduction relative to OLS is much smaller than the gains sometimes attributed to machine learning, and the RMSE values of all four models lie within a narrow range (0.424–0.438). Differences among models were not tested formally for statistical significance, so the ranking should be interpreted cautiously and the models regarded as broadly similar in accuracy.

The moderate accuracy is in line with other Sentinel-2-only studies in East African montane forests, although reported values vary widely and are not strictly comparable because of differences in metrics, plot size and validation design: pseudo- R^2 values of about 0.20 with generalized linear models and Random Forest in the Usambara mountains [21, 22] and an R^2 of about 0.74 from a small plot network in Ethiopia [24]. In Tanzania, generalized linear models produced smaller prediction errors than Random Forest when only Sentinel-2 predictors were used [22], which supports the view that algorithm flexibility adds little when predictor information is limited. The mixed-effects model here improved slightly on OLS, and its between-block standard deviation (0.146) confirms non-negligible block-level variability. The block-stratified offset correction probably removed part of the systematic spatial variation before model fitting, leaving a relationship that was only moderately nonlinear, and residual spatial autocorrelation was negligible (Moran's $I = 0.059$). By contrast, Li et al. [19] found that kriging of residuals improved XGBoost biomass estimates, which suggests that residual spatial structure can be important in other settings.

More fundamentally, all four models used the same seven optical indices. Passive optical data are sensitive to canopy-surface properties but provide limited direct information on vertical structure, tree height and woody volume, so the moderate R^2 values are consistent with an information ceiling imposed by the predictors rather than a failure of machine learning. Combining sensors is the most promising route: the review by Nguyen and Saha [26] identified multi-sensor approaches combining Sentinel-1, Sentinel-2 and LiDAR as especially effective, and machine-learning retrieval of tree height from SAR is being developed in preparation for the ESA BIOMASS mission [7]. Integrating GEDI or other LiDAR measurements, Sentinel-1 SAR, topography, texture and multi-temporal features is therefore likely to give larger gains than further increasing model complexity on the same optical indices.

4.4. Spatial Patterns, Monitoring Relevance and Temporal Interpretation

Western and south-western blocks emerged as the principal carbon reservoirs, while eastern blocks had lower predicted

carbon (Table 7 and Figure 3). The east-west contrast is consistent with the interpretation that disturbance and accessibility reduce carbon density, as reported for the Yayu forest, where accessible buffer and transition zones held lower biomass than the protected core zone [24]. These patterns are model predictions, however, and the inventory used for calibration comes from a restoration-monitoring programme centred on the South-West Mau block and related strata; its representativeness for the eastern blocks should be verified with independent field data before predictions there are used for decisions. Existing global and pan-tropical carbon maps [1, 2, 17, 29] are much coarser than the 10 m surfaces produced here and can diverge substantially from plot-based estimates [23], which underlines the value of locally calibrated products with explicit uncertainty. The predicted surfaces also complement earlier forest-change and carbon-trading work in the Mau Forest Complex [27] by describing carbon condition within forest that remains mapped as forest. They are best used to prioritise protection of carbon-rich forest, to target assisted natural regeneration and enrichment planting, and to focus field verification where predicted carbon is low or uncertain.

For REDD+ and MRV, the main contribution is not the choice of XGBoost but the reproducible end-to-end workflow: field-based biomass estimation, open Sentinel-2 imagery, transparent preprocessing and bias correction, independent model validation and wall-to-wall spatial prediction. Cheaper and more transparent inventories can increase confidence in forest-carbon projects [20], and because the imagery and major analytical tools are openly accessible, the framework can support periodic updates at lower cost than repeated full field inventories. Field plots remain essential for calibration and validation, and the current accuracy ($R^2 = 0.47$) is suitable for screening and prioritisation rather than for stand-alone carbon accounting.

The exploratory 2020–2024 analysis should be interpreted cautiously. The yearly data represent different plot cohorts rather than repeated observations at fixed locations, so the small negative central trend and the 2020–2024 difference surface cannot establish net forest decline or a true annual sequestration rate. Change estimation of that kind requires repeated measurements, as illustrated by the use of successive national forest inventories to model biomass change in China [19]. A defensible monitoring design for the Mau Forest Complex would use fixed plots, repeated field measurements and satellite extractions from identical locations and seasons. The present analysis is best viewed as evidence of broad landscape stability with localised gains and losses, and as a demonstration of the workflow required for future repeated monitoring.

4.5. Strengths and Limitations

A major strength of the study is the integration of a substantial field dataset (3,907 trees) with freely available satellite imagery and multiple modelling paradigms evaluated on an in-

dependent test set. The use of five-fold cross-validation, explicit linear and mixed-effects baselines, variable-importance analysis and spatial residual diagnostics improves transparency and interpretability. The workflow is also reproducible and operationally relevant.

Several limitations remain. Sentinel-2 is an optical sensor and is vulnerable to cloud contamination and spectral saturation in dense canopies. Structural predictors from LiDAR or radar were not included. Model performance therefore reflects the information available in optical vegetation indices. The temporal analysis used different plot cohorts across years rather than repeated measurements at fixed locations, limiting inference about true annual sequestration. Finally, model quality depends on the accuracy, spatial coverage and representativeness of the underlying field inventory and on the assumptions used to convert field measurements into biomass and carbon.

4.6. Implications for Policy and Forest Management

- 1) Integrate routinely updated Sentinel-2 products into Mau Forest monitoring to complement field inventories and identify emerging degradation.
- 2) Prioritise protection of high-carbon western and south-western forest blocks while targeting lower-carbon eastern blocks for restoration, assisted natural regeneration and enrichment planting.
- 3) Use spatial carbon products as decision-support layers within REDD+ and national MRV systems, with transparent documentation of calibration, validation and uncertainty, recognising that the current accuracy (R^2 of about 0.47) supports screening and prioritisation rather than stand-alone carbon accounting.
- 4) Develop a fixed-plot longitudinal monitoring network and combine Sentinel-2 with Sentinel-1 SAR, GEDI/LiDAR and topographic predictors to improve structural sensitivity and quantify true sequestration rates.

5. Conclusions

Combining field inventory data, Sentinel-2 vegetation indices and machine learning provides a practical, reproducible approach for estimating and mapping relative forest carbon condition across the Mau Forest Complex. Red-edge and moisture-sensitive indices were the most informative predictors, and XGBoost gave the best held-out performance ($R^2 = 0.473$, RMSE = 0.424), but its advantage over linear regression ($R^2 = 0.436$, RMSE = 0.438) was modest, indicating that predictor information rather than algorithm choice limits accuracy. The predicted surfaces identified carbon-rich western and south-western blocks and lower-carbon eastern blocks that may warrant restoration attention, and residual spatial autocorrelation was negligible. The 2020–2024 comparison was exploratory and does not provide sequestration

rates. The findings support the use of open satellite data and reproducible workflows as a low-cost complement to field inventories for REDD+ and MRV, but the present accuracy is not sufficient for stand-alone carbon accounting; substantial gains will likely depend on integrating structural information from radar and LiDAR and on repeated measurements at fixed plots.

Abbreviations

AGB	Above-Ground Biomass
BGB	Below-Ground Biomass
CIre	Red-Edge Chlorophyll Index
CO _{2e}	Carbon Dioxide Equivalent
DBH	Diameter at Breast Height
EVI	Enhanced Vegetation Index
GEDI	Global Ecosystem Dynamics Investigation
IPCC	Intergovernmental Panel on Climate Change
LiDAR	Light Detection and Ranging
MAE	Mean Absolute Error
MRV	Monitoring, Reporting and Verification
NBR	Normalized Burn Ratio
NDMI	Normalized Difference Moisture Index
NDVI	Normalized Difference Vegetation Index
OLS	Ordinary Least Squares
REDD+	Reducing Emissions from Deforestation and Forest Degradation
REML	Restricted Maximum Likelihood
RENDVI	Red-Edge Normalized Difference Vegetation Index
RMSE	Root Mean Square Error
SAR	Synthetic Aperture Radar
SAVI	Soil-Adjusted Vegetation Index
VIF	Variance Inflation Factor
XGBoost	Extreme Gradient Boosting

Acknowledgments

The study builds on forest restoration monitoring data and open Sentinel-2 imagery. The authors acknowledge the institutions and data custodians that supported access to the field information and satellite-processing infrastructure.

Author Contributions

Fred Junior Onyango: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Visualization, Writing – original draft

Charity Wamwea: Methodology, Supervision, Validation, Writing – review & editing

Amos Kipkorir Langat: Resources, Supervision, Validation, Writing – review & editing

Data Availability Statement

The analytical workflow was implemented using reproducible Google Earth Engine, R and Python scripts. Availability of the underlying field inventory remains subject to the permissions and data-sharing conditions of the original data custodians.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Avitabile, V. et al. (2016). An integrated pan-tropical biomass map using multiple reference datasets. *Global Change Biology*, 22(4), 1406-1420. <https://doi.org/10.1111/gcb.13139>
- [2] Baccini, A. et al. (2012). Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. *Nature Climate Change*, 2(3), 182-185. <https://doi.org/10.1038/nclimate1354>
- [3] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>
- [4] Chave, J. et al. (2014). Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biology*, 20(10), 3177-3190. <https://doi.org/10.1111/gcb.12629>
- [5] Chen, T., Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. <https://doi.org/10.1145/2939672.2939785>
- [6] Cheng, F., Ou, G., Wang, M., Liu, C. (2024). Remote sensing estimation of forest carbon stock based on machine learning algorithms. *Forests*, 15(4), 681. <https://doi.org/10.3390/f15040681>
- [7] Colverd, G., Takami, J., Schade, L., Bot, K., Gallego-Mejia, J. A. (2024). 3D-SAR tomography and machine learning for high-resolution tree height estimation. *arXiv preprint arXiv:2409.05636*. <https://doi.org/10.48550/arXiv.2409.05636>
- [8] Dixon, R. K., Solomon, A. M., Brown, S., Houghton, R. A., Trexler, M. C., Wisniewski, J. (1994). Carbon pools and flux of global forest ecosystems. *Science*, 263(5144), 185-190. <https://doi.org/10.1126/science.263.5144.185>
- [9] Drusch, M. et al. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, 120, 25-36. <https://doi.org/10.1016/j.rse.2011.11.026>
- [10] Dube, T., Mutanga, O. (2015). Evaluating the utility of the medium-spatial resolution Landsat 8 multispectral sensor in quantifying aboveground biomass in uMgeni catchment, South Africa. *ISPRS Journal of Photogrammetry and Remote Sensing*, 101, 36-46. <https://doi.org/10.1016/j.isprsjprs.2014.11.001>
- [11] Food and Agriculture Organization of the United Nations (2018). The state of the world's forests 2018 - Forest pathways to sustainable development. FAO, Rome.
- [12] Gorelick, N. et al. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18-27. <https://doi.org/10.1016/j.rse.2017.06.031>
- [13] Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83, 195-213. [https://doi.org/10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2)
- [14] Illarionova, S., Tregubova, P., Shukhratov, I., Shadrin, D., Efimov, A., Burnaev, E. (2024). Advancing forest carbon stocks' mapping using a hierarchical approach with machine learning and satellite imagery. *Scientific Reports*, 14, 21032. <https://doi.org/10.1038/s41598-024-71133-8>
- [15] Intergovernmental Panel on Climate Change (2006). 2006 IPCC Guidelines for National Greenhouse Gas Inventories, Volume 4: Agriculture, Forestry and Other Land Use. Institute for Global Environmental Strategies, Hayama, Japan.
- [16] Intergovernmental Panel on Climate Change (2021). Climate change 2021: The physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press. <https://doi.org/10.1017/9781009157896>
- [17] Kindermann, G., McCallum, I., Fritz, S., Obersteiner, M. (2008). A global forest growing stock, biomass and carbon map based on FAO statistics. *Silva Fennica*, 42(3), 387-396. <https://doi.org/10.14214/sf.244>
- [18] Kinyanjui, M. J., Latva-Kärä P., Kenyanya, O., Pellikka, P. (2014). An inventory of the above ground biomass in the Mau Forest Ecosystem. *Open Journal of Ecology*, 4(10), 662-676. <https://doi.org/10.4236/oje.2014.410052>
- [19] Li, Y., Li, M., Wang, Y. (2022). Forest aboveground biomass estimation and response to climate change based on remote sensing data. *Sustainability*, 14(21), 14222. <https://doi.org/10.3390/su142114222>
- [20] Lütjens, B., Liebenwein, L., Kramer, K. (2019). Machine learning-based estimation of forest carbon stocks to increase transparency of forest preservation efforts. *NeurIPS 2019 Workshop on Tackling Climate Change with Machine Learning*. arXiv: 1912.07850. <https://doi.org/10.48550/arXiv.1912.07850>
- [21] Madundo, S. D., Mauya, E. W., Lolila, N. J., Mchelu, H. A. (2022). Modelling and mapping forest above-ground biomass using earth observation data. *International Journal of Natural Resource Ecology and Management*, 7(1), 15-21. <https://doi.org/10.11648/j.ijnrem.20220701.13>
- [22] Mauya, E. W., Madundo, S. (2021). Modelling above ground biomass using Sentinel 2 and PlanetScope data in dense tropical montane forests of Tanzania. *Tanzania Journal of Forestry and Nature Conservation*. <https://www.ajol.info/index.php/tjfn/article/view/221552>

- [23] Mitchard, E. T. A. et al. (2014). Markedly divergent estimates of Amazon forest carbon density from ground plots and satellites. *Global Ecology and Biogeography*, 23(8), 935-946. <https://doi.org/10.1111/geb.12168>
- [24] Muhe, S., Argaw, M. (2022). Estimation of above-ground biomass in tropical afro-montane forest using Sentinel-2 derived indices. *Environmental Systems Research*, 11, 5. <https://doi.org/10.1186/s40068-022-00250-y>
- [25] Mutanga, O., Skidmore, A. K. (2004). Narrow band vegetation indices overcome the saturation problem in biomass estimation. *International Journal of Remote Sensing*, 25(19), 3999-4014. <https://doi.org/10.1080/01431160310001654923>
- [26] Nguyen, A., Saha, S. (2024). Machine learning and multi-source remote sensing in forest carbon stock estimation: A review. *arXiv preprint arXiv: 2411.17624*. <https://doi.org/10.48550/arXiv.2411.17624>
- [27] Otieno, K. O. (2016). Forestry carbon sequestration and trading: A case of Mau Forest Complex in Kenya. Master's thesis, University of Gävle, Sweden. <https://www.diva-portal.org/smash/get/diva2:882491/FULLTEXT02.pdf>
- [28] Poudel, A. et al. (2022). Analyzing the relationship between above-ground biomass and vegetation indices in Nepal. *Journal of Forest and Natural Resource Management*, 3(1), 53-67.
- [29] Saatchi, S. S. et al. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences*, 108(24), 9899-9904. <https://doi.org/10.1073/pnas.1019576108>
- [30] Sangpradid, S., Khamyong, S., Wachrinrat, C. (2022). Assessment of aboveground biomass using satellite-based vegetation indices and geographically weighted regression in Thailand. *International Journal of Conservation Science*, 13(1), 49-60.