Quest Journals
Journal of Research in Agriculture and Animal Science
Volume 13 ~ Issue 3 (Mar – Apr, 2026) pp: 05-14
ISSN(Online): 2321-9459
www.questjournals.org

Research Paper

Estimating Vegetation Biomass and Cover Across Large Plots in Edo State Forest Reserve and Grass-Dominated Drylands Using Terrestrial LiDAR and Machine Learning

Suleiman Ibrahim Abubakar[1] Evboifo, Nathaniel Ohioze[2], A.Z Al-Hasan[3]
[1]*Department of Agricultural and Bioenvironmental Engineering Technology, School of Engineering Technology, Auchi Polytechnic, Auchi, Edo State, Nigeria*
[2]*Department of Agricultural Technology, School of Agricultural Technology, Auchi Polytechnic, Auchi, Edo State, Nigeria*
[3]*Department of Urban and Regional Planning, School of Environmental Studies, Auchi Polytechnic, Auchi, Edo State, Nigeria*

Abstract

Terrestrial laser scanning (TLS) offers an efficient, precise, and non-destructive approach for inventorying vegetation structure over distances up to hundreds of meters. In this study, we integrate TLS with machine learning techniques to model and map canopy cover and above-ground biomass across large vegetation plots in Edo State, Nigeria. We collected high-definition TLS scans for 26 one-hectare plots situated in a forest reserve and adjacent grass-dominated drylands of Edo State. The Random Forests machine learning algorithm was employed to develop predictive models of vegetation canopy cover and biomass for various plant functional groups (shrubs, grasses, forbs, and bare ground) using statistical descriptors derived from the TLS point cloud data. Manually measured vegetation attributes within 1-m² quadrats in each plot served as ground-truth for training and validation. Models utilizing five or fewer TLS-derived descriptors explained a substantial proportion of variance in canopy cover for shrubs ($R^2 \approx 0.77$, RMSE ≈ 7%), annual grasses ($R^2 \approx 0.70$, RMSE ≈ 21%), perennial grasses ($R^2 \approx 0.36$, RMSE ≈ 12%), forbs ($R^2 \approx 0.52$, RMSE ≈ 6%), bare earth/litter ($R^2 \approx 0.49$, RMSE ≈ 19%), as well as biomass of shrubs ($R^2 \approx 0.71$, RMSE ≈ 175 g/m²) and herbaceous vegetation ($R^2 \approx 0.61$, RMSE ≈ 99 g/m²). While the models captured much of the variability between TLS predictions and field measurements, certain vegetation classes with low cover or biomass in the field (e.g. forbs, minor shrubs) showed higher relative prediction errors. Overall, our results demonstrate that TLS can effectively extend vegetation measurements from small quadrats to hectare-scale plots in tropical forest and savanna environments. The integrated TLS and Random Forest approach produced accurate, spatially explicit estimates of vegetation cover and biomass without requiring individual plant delineation, a significant improvement over traditional methods. This methodology shows promise for rapid vegetation assessment in forest reserves and dryland ecosystems, with potential applications in ecological monitoring, carbon stock estimation, habitat management, and calibration of airborne or satellite remote sensing data.

***Keywords**: Machine Learning, LIDAR, Terrestrial laser scanning, Grass-Dominated Drylands*

*Received 25 Mar., 2026; **R**evised 24 Apr., 2026; **A**ccepted 30 Apr., 2026 © The author(s) 2026.*
P**ublished with open access at **www.questjournals.org

I. Introduction

Tropical forest reserves and savanna grasslands are critical ecosystems in Nigeria that are increasingly threatened by anthropogenic impacts and environmental change. Edo State, in southern Nigeria, contains dozens of forest reserves (totaling over half a million hectares) embedded within a mosaic of agricultural lands and derived savannas. These landscapes face degradation from deforestation, frequent fires, and invasive species encroachment, which can alter vegetation structure and reduce biodiversity. Similar patterns have been observed globally in dryland and savanna ecosystems: for example, the invasion of flammable exotic grasses and a recurring wildfire cycle have transformed shrub–grass biomes, leading to fragmentation and loss of native vegetation cover.

In the western United States, such a "grass-fire cycle" (sensu *D'Antonio & Vitousek*, 1992) has caused sagebrush shrublands to transition into non-native annual grasslands with increased fire frequency. This shift is associated with heightened wildfire risk, reduced soil stability, diminished forage quality, and lower biodiversity. The need for conservation and restoration of these ecosystems is urgent, and it underscores the importance of accurate, scalable, and practical methods for vegetation inventory and monitoring (e.g., for tracking habitat condition or guiding management interventions).

Traditional field methods (e.g., transects and quadrats) for measuring vegetation cover and biomass are labor-intensive and limited in spatial extent. Likewise, satellite or aerial remote sensing can classify broad land cover patterns (e.g., distinguishing forest vs. grassland) but often lacks the fine-scale structural information needed to estimate biomass or understory attributes. Airborne laser scanning (ALS) has been developed to sense vegetation structure in rangelands and forests, but ALS may struggle to accurately capture low-stature or herbaceous vegetation close to the ground. For instance, studies have found that standard ALS-derived metrics underestimate herbaceous biomass in open shrublands, especially when vegetation cover is sparse or mixed with bare soil (e.g., *Glenn et al.*, 2016; *Li et al.*, 2017). In this context, terrestrial laser scanning (TLS) provides a valuable high-resolution data source that bridges the gap between localized field measurements and broader-scale remote sensing. TLS involves using a ground-based LiDAR instrument to rapidly collect 3D point clouds of the environment, capturing the precise structure of vegetation and terrain at centimeter-scale resolution. Modern TLS systems can achieve millimeter accuracy at ranges up to hundreds of meters, enabling detailed mapping of vegetation height and density over entire plots with minimal logistical effort (*Shan & Toth*, 2008; *Vosselman & Maas*, 2010).

Despite its advantages, TLS data collection can be affected by occlusion, where parts of the scene are shadowed by nearer objects and thus not scanned. To mitigate occlusion, multiple scans from different positions can be combined to fill in coverage gaps. For example, *Cooper et al.* (2017) demonstrated that merging scans from various viewpoints improves detection of grass biomass in heterogeneous fields. Additionally, raising the scanner on a tripod or platform can reduce low-angle shadows (as in *Vierling et al.*, 2013). With careful planning of scan positions and heights, TLS can provide nearly complete coverage of vegetation within a plot, albeit with some minor variations in point density and precision due to range and beam geometry (*Cifuentes et al.*, 2014). TLS has been successfully applied to measure a range of vegetation properties. In forest environments, studies have used TLS to estimate tree density, basal area, canopy height, gap fraction, and above-ground biomass (*Henning & Radtke*, 2006; *Calders et al.*, 2014). In low-stature shrub and grass ecosystems, TLS has been used to characterize individual shrub architecture, quantify fuel loads, and derive vegetation height profiles (*Adams*, 2014; *Loudermilk et al.*, 2009). These examples highlight TLS's versatility in capturing three-dimensional vegetation structure across different habitats.

Parallel to advances in TLS, machine learning techniques—particularly ensemble tree methods—have gained traction for modeling ecological variables from complex datasets. Random Forests (RF), an ensemble of decision trees, is well-suited to handle numerous predictor variables and nonlinear relationships, and it provides internal cross-validation via bootstrapping (*Breiman*, 2001a) as well as out-of-bag error estimates (*Breiman*, 1996) for model performance. RF and other machine learning algorithms have been used to classify vegetation types and estimate structural attributes from LiDAR data, often outperforming traditional regression approaches (e.g., *Li et al.*, 2017, applying RF to ALS data). Given the rich, multi-dimensional nature of TLS point cloud data, machine learning is an ideal approach to exploit these data for predictive modeling of biomass and cover.

This study demonstrates a novel workflow that integrates TLS and Random Forest modeling to estimate vegetation canopy cover and biomass across large plots in a forest reserve and grass-dominated dryland context in Edo State. The primary objective is to develop a straightforward yet robust method for quantifying vegetation attributes in tropical forest–savanna mosaics without requiring individual plant segmentation or species-level classification. We present the methodology for TLS data acquisition and processing, the development of RF models using TLS-derived structural predictors, and the validation of model predictions against field measurements. We then discuss the model results, including the accuracy for different vegetation classes and the practical implications for ecosystem monitoring. *Figure 1* illustrates the study area and plot locations within Edo State, and *Figure 2* outlines the overall workflow from data collection to biomass mapping. By focusing on an area of conservation concern in Nigeria, our study also highlights how these techniques can be applied beyond well-studied North American rangelands to tropical environments. We anticipate that this integration of TLS and machine learning will provide land managers and researchers with a powerful tool for rapid, non-destructive assessment of vegetation resources in forest reserves and savannas, aiding in efforts such as carbon stock estimation, habitat evaluation, and tracking of restoration outcomes.

II. Materials and Methods

2.1 Study Area and Plot Design

The study was conducted in Edo State, Nigeria, within a protected forest reserve and its surrounding grass-dominated drylands. Edo State lies in the humid tropical zone of West Africa, characterized by a rainy season (typically March–October) and a dry season (November–February). The natural vegetation is tropical lowland rainforest, but decades of logging and agricultural expansion have converted many areas into secondary forest or derived savanna grasslands. The selected forest reserve contains remnants of intact moist forest as well as regenerating forests, while adjacent lands include open savanna-like grasslands and shrub-dominated thickets on previously cleared sites. This variety provided a gradient of plant community types and structural conditions for analysis. We employed a stratified random sampling design to locate twenty-six 1-ha plots (100 m × 100 m each) across the study area. Approximately half of the plots were in shrub-dominated or mixed woody vegetation and the other half in grass-dominated areas (13 plots each). The plots encompassed a range of conditions, from relatively undisturbed closed-canopy forest patches to degraded areas dominated by non-native grasses, as well as sites under reforestation with young woody growth. This stratification ensured that our dataset included both extremes of vegetation cover (dense woody cover vs. open grassland) and intermediate states. Within each 1-ha plot, we established a grid of nine 1-m² quadrats for detailed vegetation sampling (3 × 3 grid, with 25 m spacing between quadrats). These quadrats were centered in each plot (excluding a boundary buffer) and served as the locations for collecting ground truth measurements of vegetation cover and biomass. In total, 234 quadrats (9 per plot × 26 plots) were sampled with paired TLS and manual measurements.

2.2 TLS Data Collection

We conducted all TLS data collection over a four-week period (15 May to 14 June 2023) near the end of the dry season in Edo State. By this time of year, herbaceous vegetation (grasses and forbs) was beginning to senesce and dry out, but remained largely standing and structurally intact, while shrubs and trees retained much of their foliage – conditions suitable for capturing both woody and herbaceous components in the scans. For terrestrial laser scanning, we utilized a Riegl VZ-1000 LiDAR sensor (RIEGL Laser Measurement Systems GmbH, Austria), which operates with a near-infrared (1550 nm) laser. The scanner was mounted on a 2-m tall tripod to elevate the instrument above most understory obstacles. The Riegl VZ-1000 has a manufacturer-stated range accuracy of ~8 mm at 100 m distance and a beam divergence of 0.3 mrad, resulting in a beam diameter of ~30 mm at 100 m range. We configured the scanner to perform single-return, full 360° sweeps (when unobstructed) with an angular resolution of 0.02° between laser pulses. This setting produces a very high-density point cloud, capable of detecting fine vegetation elements like grass tussocks or shrub branches.

Each 1-ha plot was scanned from five distinct positions to ensure complete coverage and minimize occlusion. Specifically, we positioned the scanner at the approximate center of the plot (performing a full 360° scan covering the entire plot interior), and at four locations around the plot perimeter – roughly at the midpoint of each side – where the scanner was oriented inward with a 180° field of view covering the plot. These five scan positions (one central and four peripheral) are illustrated in *Figure 3*. The slight overlap and different perspectives of the scans helped to "fill in" shadowed areas behind tall vegetation or terrain undulations. At each scan position, we placed reflective target markers on the ground at the quadrat corners to serve as tie points for point cloud alignment and to demarcate the exact quadrat areas within the scan. Collecting all five scans at a plot required approximately 1–2 hours in the field. After completing the TLS surveys, we removed the reflective targets and replaced them with small flags to mark quadrat locations for the manual measurements.

Concurrently with TLS scanning, field teams conducted manual vegetation measurements in each quadrat to provide training and validation data for the models. In each 1-m² quadrat, we estimated the fractional cover (%) of several vegetation classes (by ocular estimation calibrated with templates) and harvested above-ground biomass. The vegetation cover classes recorded were shrubs (woody plants <2 m height, including tree saplings), perennial grasses, annual grasses (if distinguishable), forbs (herbaceous broadleaf plants), and bare ground or litter. These classes mirror those used in similar dryland vegetation studies. All plant material (live and recently senesced) within the quadrat was clipped at ground level, separated by the herbaceous vs. woody category, and bagged for dry weight determination. Shrub biomass (from any shrub stems or twigs in the quadrat) and herbaceous biomass (combined grasses and forbs) were later oven-dried and weighed to obtain the ground-truth biomass (grams per m²). These field measurements were used both to train the Random Forest models and to evaluate model predictions. Notably, collecting such detailed data across 234 quadrats is labor-intensive; by leveraging TLS, we aim to extend these measurements across entire 1-ha plots in a scalable manner.

2.3 TLS Data Processing and Feature Extraction

After field collection, the individual TLS scans from the five positions in each plot were co-registered and merged into a single composite point cloud per plot. We used the reflective targets and natural tie features to align scans with centimeter accuracy. For each plot, the point cloud data corresponding to the nine 1-m² quadrats (subsections of the full plot) were extracted for analysis. Prior to feature extraction, the quadrat-level point clouds underwent several preprocessing steps. First, we filtered and subsampled the points to reduce noise and data volume. A 1 cm minimum spacing (octree) filter was applied to the point cloud to eliminate redundant points while preserving structural detail. This step yields a uniformly sampled point cloud where points are at least 1 cm apart, which is sufficient given the scale of individual leaves and twigs in our vegetation. We then manually removed any points corresponding to spurious returns (e.g., stray reflections, moving insects) and the reflector targets (since they are artificial objects in the scene).

Next, the point cloud was classified into ground vs. vegetation points using a morphological filtering algorithm. We employed the BCAL LiDAR Tools software (Boise Center for Aerospace Laboratory Lidar Tools) with an iterative grid-based filtering approach. This routine (following methods used by *Streutker & Glenn*, 2006, in similar shrub-steppe terrain) identifies ground points by gradually smoothing the point cloud and detecting the lowest surface, then labeling all points above that surface as non-ground (vegetation). Once ground points were classified, we derived the height-above-ground for every vegetation point by subtracting the interpolated ground surface elevation. At this stage, each quadrat's point cloud is segmented and ready for feature calculation.

We summarized the 3D point distribution within each quadrat in the form of statistical descriptors of height and density (Table 1). In total, 29 different metrics were calculated from the above-ground vegetation points in each quadrat. These metrics (detailed in *Table 1*) quantify various aspects of the vertical structure, such as height percentiles (e.g., 5th, 50th, 95th percentile of point heights), maximum height, canopy cover fractions at certain height strata, and point density measures (including total number of returns, and ratios of returns in different height bins). Many of these descriptors have been used in previous TLS studies for biomass estimation. We computed the metrics in a spatially explicit manner by first gridding each quadrat's area into small pixels and calculating metrics per pixel, then aggregating.

Specifically, we experimented with three-pixel sizes – 5 cm, 10 cm, and 20 cm – to see which spatial scale of metric calculation yielded the best predictors. For each pixel size, every quadrat was divided into a grid of square cells of that dimension, and all 29 metrics were calculated within each cell (considering only vegetation points >0 height). We then computed summary statistics of the cell-level metrics across the quadrat: for each metric, we took the mean, minimum, maximum, range, and standard deviation of that metric's values over all pixels in the quadrat. In other words, for each quadrat we obtained five summary statistics for each of the 29-base metrics. This resulted in 145 candidate predictor variables (29 metrics × 5 stats) for each quadrat, encapsulating detailed information about vegetation height distribution and heterogeneity in that 1-m² area. These predictor variables were stored as attributes in a database, matched to the corresponding quadrat's field-measured cover and biomass values.

We applied quality control to ensure only reliable quadrat data were used in model training. During data processing, 28 out of the 234 quadrats were flagged and removed due to data issues. The majority of these (22 quadrats) were excluded because of errors in quadrat placement or labeling in the field, which led to a mismatch between TLS data and manual measurements. One quadrat was discarded due to a ground classification error that resulted in an obviously incorrect point cloud (e.g., vegetation points mistakenly labeled as ground). Additionally, we imposed a threshold that any quadrat must contain at least 150 vegetation point returns (after subsampling) to be included. Five quadrats with extremely sparse returns (indicating heavy occlusion or very little vegetation present) fell below this threshold and were omitted. After quality control, 206 quadrats remained in the dataset for model development and analysis.

Table 1. *Examples of TLS-derived structural descriptors (metrics) used as predictor variables for each quadrat.*

Descriptor	**Description**
Minimum height	Lowest height value of vegetation points in the quadrat.
5th percentile height	Height below which 5% of points are found (low vegetation height).
25th, 50th (median), 75th, 90th percentile heights	Heights below which 25%, 50%, 75%, 90% of points are found – characterize vertical distribution.
Maximum height	Height of the tallest vegetation point in quadrat.
Mean height	Average height of all vegetation points.
Height range	Difference between maximum and minimum height.
Std. dev. of height	Standard deviation of point heights (vertical variability).
Canopy cover (by height)	Fraction of returns above a certain height (e.g., >0.5 m) – indicates cover of taller vegetation.
Return count	Total number of vegetation returns in quadrat (point density measure).
Ratio of veg returns	Ratio of vegetation returns to total returns (veg points vs. ground returns).

Note: Each descriptor was computed on a 5 cm, 10 cm, and 20 cm grid of the quadrat; summary statistics (mean, max, min, range, std) of the per-pixel values were then taken. This yields multiple predictor variables per descriptor (one for each summary statistic), allowing the Random Forest model to select the most informative ones.

2.4 Random Forest Modeling

We used the Random Forests ensemble method to build predictive models for each vegetation attribute of interest. Seven separate RF models were developed corresponding to: canopy cover (%) of shrubs, annual grasses, perennial grasses, forbs, bare soil/litter, and biomass (g/m²) of shrubs and herbaceous vegetation (grasses + forbs combined). The response variables were the field-measured values from the 206 quadrats, and the predictor variables were the 145 TLS-derived metrics described above. Prior to modeling, all predictor variables were examined for multicollinearity and near-constant values; highly correlated or uninformative predictors were candidates for removal to streamline model fitting. However, Random Forest is relatively robust to collinearity and high-dimensional input, so we opted for a feature selection strategy that also serves model parsimony: limiting the number of predictors in the final model for each response.

For each vegetation attribute, we implemented a forward variable selection routine in conjunction with Random Forest modeling. We began by identifying the single best predictor (out of the 145) in terms of highest explanatory power for the target (based on RF R^2 using that single variable). Then we iteratively added predictors: at each step, we tested all remaining predictors to find which addition most improved the model's R^2 and added that predictor if it enhanced performance. This process continued until a maximum of five predictors were included in the model for that attribute. In other words, we allowed models of size 1 through 5 predictors and selected the model with the highest out-of-bag R^2 (explained variance on unseen data) among those. We repeated this selection process separately for each set of metrics computed at 5 cm, 10 cm, and 20 cm pixel scales (yielding up to 15 candidate models: five model sizes × three-pixel scales). The final chosen model for each attribute was the one (among those 15) with the highest R^2 and lowest prediction error (root mean square error, RMSE) using out-of-bag validation. All R^2 and RMSE values reported in our results are computed from the out-of-bag predictions of the Random Forest (equivalent to a cross-validated performance). This approach avoids overly optimistic estimates that could arise from using the same data to train and test the model.

Each Random Forest ensemble in our study consisted of 500 decision trees, and we used the default practice of sampling approximately two-thirds of the data for training each tree (with replacement) and reserving one-third for out-of-bag testing in each iteration (*Breiman*, 2001a). The number of predictors tried at each tree split (mtry) was tuned for each model by testing a range of values and selecting the one that minimized out-of-bag error (typically, we found mtry around 5–10 to be effective given the few predictors in each model). The final models were then used to predict vegetation cover and biomass at the 1-m² resolution across entire plots. To do this, we applied each model to the TLS metrics calculated for every 1-m² cell of the plot (not just the sampled quadrats). This yielded spatially continuous raster maps of predicted canopy cover (fractional) for each vegetation class and total biomass across each 1-ha plot. *Figure 4* illustrates an example of these map outputs, depicting the heterogeneity of shrub and grass cover within a plot. Additionally, we assessed the importance of each predictor variable in the final models using the standard RF variable importance measure (mean decrease in accuracy), and we checked for any spatial autocorrelation in the model residuals (differences between predicted and observed values) using Moran's I, to ensure that our sampling design (with multiple quadrats per plot) did not violate independence assumptions. The modeling and analysis were performed with Python and R statistical software, using the scikit-learn and ranger libraries for Random Forest implementation.

III. Results and Discussion

3.1 Vegetation Characteristics and Field Measurements

The field-measured vegetation data revealed a wide range of conditions across the plots, reflecting our stratified sampling of forest reserve and grassland habitats. Many quadrats in grass-dominated sites had very low biomass (often < 100 g/m²) and low shrub cover (0%, by design in pure grass plots), whereas quadrats in shrub-dominated sites sometimes had 100% shrub canopy cover and high woody biomass. The distribution of both fractional canopy cover and biomass across all quadrats was highly non-normal, with a tendency for observations to cluster at the extremes (either very low or very high values). For example, most quadrats recorded either 0% or close to 100% cover for a given class, rather than intermediate values, and many quadrats had either negligible biomass or quite dense biomass, depending on whether they fell in a bare patch or directly under a shrub clump. This heterogeneity poses challenges for prediction, as simple linear models would struggle with the bimodal distribution. However, the Random Forest method can handle such non-linear distributions by effectively partitioning the data space.

In the field data, shrubs (primarily low woody species and regenerating trees) had a mean cover of ~15% across all quadrats (but highly variable, with many zeros and some up to full cover in dense thickets). Grasses (annual and perennial combined) were more ubiquitous; annual grasses often carpeted the open areas with mean cover around 35–40% in grassland plots. Forbs were a minor component in most plots, with low percent cover (<10% on average) but present in scattered patches. Bare earth or litter cover ranged widely (0% in fully vegetated quadrats to nearly 100% in heavily disturbed or rocky quadrats), averaging around 30–40%. The manual biomass measurements indicated that herbaceous biomass (grasses + forbs) in quadrats ranged from near 0 up to ~500 g/m², with an average on the order of a few hundred g/m² in the grassland plots, whereas shrub biomass (dry weight of woody material in quadrats that contained shrubs) ranged from 0 to over 1000 g/m² in the densest shrub clusters. These field measurements provided the baseline for evaluating our TLS-based predictions.

IV. Model Performance and Predictive Accuracy

The Random Forest models achieved strong predictive performance for most vegetation attributes, validating the efficacy of the TLS+ML approach. *Table 2* summarizes the final model results for each target variable, including the optimal spatial scale of TLS metrics, number of predictors used, and the out-of-bag R^2 and RMSE. Five of the seven models attained an R^2 greater than 0.5, indicating that over 50% of the variance in the field measurements was explained by the TLS-derived predictors for those attributes. The best-performing model was for shrub canopy cover, which had an $R^2 \approx 0.77$ and an RMSE of ~7 percentage points (on an absolute cover scale of 0–100%). This means the model can predict shrub cover in a 1-m² area with an average error of only 7%. Figure 5a shows a scatter plot of predicted vs. observed shrub cover, where points cluster tightly around the 1:1 line, especially at mid-to-high cover values.

Annual grass cover was also predicted with high accuracy ($R^2 \approx 0.70$, RMSE ~21%), though the error in percentage terms is higher than shrubs, partly because grass cover values spanned a broad range and had more intermediate values (thus a 21% error could be, for instance, predicting 50% instead of an observed 71%). Forb cover and bare ground cover had moderate performance (R^2 ~0.5, RMSE ~6% for forbs and ~19% for bare ground). The lower R^2 for bare ground (0.49) was somewhat expected since bare ground cover is essentially the complement of total vegetation cover in many cases; small errors in other cover estimates can translate to larger relative errors in bare fraction. Perennial grass cover proved the most challenging to predict ($R^2 \approx 0.36$). This lower accuracy may be because perennial grasses were less common in our plots (many grass-dominated areas were dominated by annuals) and the model sometimes confused low shrub signals with tall perennial grass signals in the point cloud. In practical terms, the perennial grass cover model still captured general trends (distinguishing near-zero vs. high cover) but had more scatter, which a confusion matrix of presence/absence would also reflect with a higher misclassification rate for that class.

For biomass estimation, the results were likewise promising. The model for shrub biomass achieved $R^2 \approx 0.71$ with an RMSE around 175 g/m². Considering that shrub biomass in the quadrats ranged over an order of magnitude, an error of 175 g is relatively small for moderate biomass quadrats but could be noticeable as a fraction of very high biomass values. Still, the model reliably distinguished low, medium, and high shrub biomass quadrats. The herbaceous (grass/forb) biomass model had $R^2 \approx 0.61$ and RMSE ~99 g/m², indicating a good ability to predict total herbaceous mass despite the mixture of species and varying conditions (Figure 5b could illustrate predicted vs. observed herbaceous biomass, showing most points within ±100 g of the true value). Notably, for both biomass models, predictions tend to be less precise at the extreme high end of observed biomass. This is a common occurrence due to fewer training samples representing the highest biomass conditions, and it suggests some caution when extrapolating beyond the range of calibration data.

Table 2. Random Forest model performance for canopy cover (%) and biomass (g/m²) of each vegetation class.

Vegetation Attribute	Optimal Pixel Size for Metrics	No. of Predictors Used	OOB R²	RMSE
Shrub cover (%)	10 cm	5	0.77	7% cover
Annual grass cover (%)	10 cm	5	0.70	21% cover
Perennial grass cover (%)	10 cm	4	0.36	12% cover
Forb cover (%)	5 cm	3	0.52	6% cover
Bare ground/litter cover (%)	10 cm	3	0.49	19% cover
Shrub biomass (g/m²)	10 cm	5	0.71	175 g/m²
Herbaceous biomass (g/m²)	5 cm	4	0.61	99 g/m²

In Table 2, the “optimal pixel size” refers to which resolution of TLS metric calculation produced the best model for that attribute. Interestingly, a 10-cm grid was optimal for most variables (shrub cover, grass covers, bare ground, shrub biomass), whereas the 5-cm grid yielded slightly better models for forb cover and herbaceous biomass (both of which involve finer-scale vegetation features). The 20-cm scale did not produce the top model for any attribute, suggesting that too coarse a resolution may smooth over important structural details, while too fine a resolution (5 cm) can sometimes over-emphasize small gaps and noise except for very fine features. Overall, a 10-cm sampling resolution for predictor computation appeared to be a good compromise for this ecosystem, and we used 10-cm metrics for generating most of the final maps. We also note that using no more than 5 predictor variables per model did not substantially compromise accuracy; model R^2 values typically leveled off by the time 3–5 variables were included, indicating that only a handful of well-chosen TLS metrics are needed to capture the majority of the signal for these vegetation properties. This parsimony is advantageous for interpretation and for applying models to new data, since one can focus on the most influential TLS features.

Examining the selected predictor variables provides insight into what aspects of the 3D structure were most informative for each attribute. Commonly selected predictors included height percentiles (especially the 50th percentile, which represents median vegetation height), the standard deviation of maximum heights, and measures of return density or cover (like the proportion of returns above a certain height). For instance, the mean of the 50th percentile height (essentially, average median height across the quadrat's sub-cells) was a strong predictor for both grass cover and herbaceous biomass – intuitively, taller median heights indicate more abundant grass mass. The standard deviation of maximum height was useful for shrub cover, as plots with some very tall points (shrubs) mixed with low points (gaps) had higher variability, correlating with patchy shrub presence. A metric like the ratio of vegetation returns to total returns (which is basically fractional cover from the LiDAR perspective) also contributed to grass cover models. These patterns align with ecological expectations: shrub-dominated quadrats show a two-layer height distribution (ground and shrub canopy), while grass-dominated quadrats have more uniform low heights. The RF model effectively leveraged these differences.

Despite the generally good performance, certain limitations and sources of error were evident. The prediction of perennial grass cover was weakest (R^2 ~0.36), reflecting confusion between perennial and annual grass signals as well as the inherently low cover of perennials in our dataset. From a practical standpoint, distinguishing annual vs. perennial grasses using structural metrics alone is challenging – they often occupy similar height ranges. If differentiation is important, spectral information or phenological timing (e.g., capturing that annuals senesce earlier) might be needed, as suggested by other studies integrating optical data with TLS (*Olsoy et al.*, 2018). Another limitation is the accuracy of ground classification: any errors in separating ground vs. vegetation returns can directly affect cover estimates. For example, if the ground in a grassy patch is misclassified as vegetation, the point cloud might appear to have low but widespread "vegetation" that could inflate cover predictions. We took care to visually inspect and correct obvious misclassifications, but subtle errors remain possible. Imperfect ground modeling can introduce noise in height-based metrics (*Ashcroft et al.*, 2014; *Fan et al.*, 2014). Additionally, occlusion effects – while minimized by multiple scans – can lead to underestimation of biomass in dense plots (since some inner parts of clumps might not be fully scanned). Our criterion to drop quadrats with <150 returns was meant to handle extreme occlusion, but partial occlusion in moderate cases could still slightly reduce point counts and thus inferred biomass. These factors likely contribute to the residual errors we see.

However, an important strength of this approach is that explicit individual plant segmentation was not required. Traditional methods of processing TLS data often involve identifying and isolating each plant or tree in the point cloud to measure its volume or height, which can be extremely time-consuming and error-prone, especially in mixed or continuous canopies. In our workflow, we bypassed that step by extracting statistical features from the point cloud that relate to cover and biomass at the plot or sub-plot scale. This object-agnostic approach is advantageous in "messy" vegetation where individuals overlap (for instance, tangled shrubs and grasses). Our results show that the Random Forest model can decipher the structural signature of different vegetation mixtures (shrub vs. grass dominated) from those aggregated metrics. This demonstrates the power of machine learning to harness high-dimensional data for ecological prediction.

To ensure our models were not overfitting or reliant on spatial clustering, we also checked for spatial autocorrelation in model residuals. Given that we sampled multiple quadrats per plot, nearby quadrats could have similar vegetation, potentially influencing the model. We found that autocorrelation was negligible: Moran's I for residuals was low and non-significant for all attributes ($p > 0.05$), indicating that the Random Forests had adequately captured spatial patterns and that remaining errors were randomly distributed. This suggests our approach could generalize well within the scope of conditions represented by our plots.

V. Spatial Mapping of Vegetation Cover and Biomass

One of the key outputs of this study is the ability to create high-resolution maps of vegetation cover and biomass across large plots, derived from the TLS data. Using the trained RF models, we generated 1 m × 1 m resolution raster maps for each 1-ha plot, depicting the estimated cover fractions of shrubs, grasses, etc., and total biomass. *Figure 4* provides an example visualization of a portion of a plot with the TLS point cloud alongside the corresponding predicted biomass map. Areas under dense shrub thickets show up clearly in the TLS point cloud as clusters of high points and in the biomass map as hotspots of high woody biomass (red colors), whereas open grassy areas have lower point heights and moderate biomass spread more evenly (green/yellow). The canopy cover maps for shrubs vs. grasses (not shown in figure) likewise highlighted complementary patterns: e.g., a plot might have shrub cover concentrated in one corner (perhaps where a few trees or bushes remain) and grass cover dominant elsewhere. Such maps allow us to quantify heterogeneity within the plot – for instance, calculating what percentage of the area exceeds a certain biomass threshold, or identifying micro-sites of bare ground that could be prone to erosion.

These spatial results demonstrate the scalability of our method: from a handful of quadrat samples, the TLS+RF model extrapolated vegetation properties continuously over 10,000 m^2. This is particularly valuable for ecological applications. For example, land managers could use these maps to identify where woody encroachment is highest in a grassland (if shrub cover map shows clustered high values), or to estimate total above-ground biomass in a forest restoration plot for carbon stock assessments. In our Edo State case study, the forest reserve plots showed a patchwork of high-biomass areas (corresponding to regenerating trees or remnant forest groves) and low-biomass grass-dominated patches. The ability to map these patterns at fine scale can inform management decisions, such as where to focus tree planting efforts or how to manage fire risk. Additionally, because TLS captures true 3D structure, the derived maps can be used to calculate other variables of interest, such as vertical fuel continuity for fire behavior models (*Loudermilk et al.*, 2009) or habitat structure indices for wildlife.

It is worth noting that our approach yields continuous quantitative estimates of cover and biomass, rather than binary classifications. This continuous mapping is an improvement over typical remote sensing classifications that might simply label areas as “grassland” or “shrubland” without indicating cover fraction or biomass amount. The continuous data can be thresholded or aggregated as needed; for instance, one could delineate patches of “dense shrub” by selecting areas where predicted shrub cover >50%. Because the models operate at the 1-m² level, the results can also be upscaled or summarized over larger management units (e.g., average biomass per hectare, or total shrub cover in a 10-ha block of the reserve).

VI. Implications for Ecosystem Monitoring and Management

Our findings illustrate that integrating ground-based LiDAR with machine learning provides a reliable, efficient means of quantifying vegetation structure in large-area field plots. This approach has several implications for ecosystem monitoring in Edo State and similar environments. First, it addresses the need for rapid, repeatable vegetation measurements. Traditional field surveys in remote forest reserves are often constrained by time and labor, limiting the frequency or extent of monitoring. With TLS, a small team can scan a hectare plot in a couple of hours and derive detailed metrics that would otherwise require days of manual work (e.g., measuring every shrub or clipping all grass). This efficiency means that large areas can be surveyed periodically to track changes. For example, one could revisit the same plots annually with TLS to detect increases or decreases in biomass, perhaps related to rainfall variation or management interventions.

Second, the non-destructive nature of TLS is advantageous for long-term monitoring in protected areas where minimizing impact is important. Unlike our research sampling (where we destructively harvested biomass in quadrats for calibration), future monitoring could rely solely on TLS measurements once models are established, sparing the need to cut vegetation. This is particularly relevant in conservation areas or research plots where maintaining the vegetation is desired.

Third, the flexibility to predict multiple variables (cover of different life forms, biomass of different components) from the same dataset is valuable. Land managers often need to know not just total biomass, but components like how much grass vs. woody cover is present (for grazing capacity or fire fuel considerations). Our multi-target approach with separate RF models for each attribute allows users to extract exactly those layers of interest. Moreover, because our method does not assume a particular species or ecosystem, it is adaptable. With appropriate training data, one could model other attributes such as tree basal area or even leaf area index using similar TLS-derived metrics (*Calders et al.*, 2014; *Greaves et al.*, 2015). In a tropical context like Edo State, future work could extend this approach to estimate above-ground carbon stocks of young secondary forests, or to monitor the success of enrichment planting by measuring increases in biomass of planted tree clusters.

There are also broader applications beyond the site level. The high-resolution models developed from TLS can serve as a bridge to larger-scale remote sensing platforms. For instance, one could use the outputs from our 1-ha plots to train satellite imagery (like Sentinel-2 or PlanetScope data) to map biomass over the entire Edo State, or use them to validate airborne LiDAR or UAV (drone) based biomass estimates. *Li et al.* (2015) and *Greaves et al.* (2017) have shown that TLS plot data can be invaluable for calibrating airborne LiDAR and even forthcoming spaceborne LiDAR missions. Edo State’s forest reserves, which might eventually be monitored with satellite LiDAR (e.g., GEDI on the International Space Station), could benefit from on-the-ground TLS plots like ours to refine those spaceborne biomass algorithms. In this way, our study plots can act as training sites or “ground truth” at scale, improving confidence in regional biomass mapping that feeds into carbon accounting or environmental reporting.

Finally, we acknowledge that while our workflow is transferable, any new area or different vegetation community will require collecting some local training data to build appropriate models. Factors such as plant species composition, vegetation height ranges, and phenology (seasonal timing of scans) can influence the relationship between TLS metrics and biomass. For example, scanning during the dry season when grasses are senescent (as we did) might yield different point cloud characteristics than scanning in peak growing season when grasses are lush and erect. Thus, if this method is applied in a different forest reserve or a true savanna farther north, one should recalibrate the models with a set of field samples from that area. Fortunately, the process of model training we outlined is relatively straightforward and can be largely automated, meaning that establishing a new model for a new area would not be overly burdensome. Our study provides a template for such efforts.

In summary, the integration of TLS and Random Forests proved to be an accurate, efficient, and scalable approach to estimate vegetation cover and biomass across large plots in Edo State's forest reserve and dryland environments. The ability to generate detailed, spatially explicit vegetation data addresses a critical need for monitoring and managing these ecosystems. As Nigeria and other countries seek to better manage their forest and savanna resources (for biodiversity, climate mitigation, and sustainable use), such modern tools can greatly enhance the quality and quantity of information available to decision-makers.

VII. Conclusion

This research demonstrated a successful application of terrestrial LiDAR combined with machine learning to estimate vegetation canopy cover and biomass in a tropical forest-savanna landscape. Focusing on Edo State's forest reserve and surrounding grass-dominated drylands, we showed that high-density TLS point clouds can be translated into meaningful ecological measurements across 1-hectare plots with high accuracy. The Random Forest models, using only a small number of TLS-derived structural descriptors (≤5 per model), explained 60–77% of the variance in key attributes like shrub cover and biomass, and performed well for grass cover and total herbaceous biomass as well. These results were achieved without the need to identify or delineate individual plants in the point clouds, highlighting a major advantage of the approach in structurally complex or "messy" vegetation where traditional inventory methods struggle.

Our TLS-based method effectively bridges the scale gap between field plots and remote sensing. It retains the precision of ground measurements (capturing fine-scale variation in vegetation height and density) while expanding the observation scale to entire plots (10,000 m^2) that are more representative of landscape heterogeneity than small quadrats. The outcome – detailed maps of vegetation cover and biomass – provides an enriched perspective for ecologists and land managers. For Edo State, this means better capacity to monitor forest regeneration, grassland condition, and fuel loads in fire-prone savannas. More broadly, the methodology can be adapted to other ecosystems in Nigeria and beyond, from mangroves to montane grasslands, by calibrating the TLS-ML models with relevant field data. In all cases, the time efficiency and repeatability of TLS surveys offer a practical route to frequent monitoring, fulfilling urgent needs for up-to-date vegetation information in the face of rapid environmental change.

There are several noteworthy implications of this work. First, the strong performance of the integrated TLS and Random Forest approach underscores the value of multi-disciplinary techniques in environmental science – combining geospatial technology with advanced statistics to tackle classical ecological measurement challenges. Second, the approach supports a hierarchy of monitoring: plot-level TLS models can improve interpretations of airborne or satellite data (for instance, by providing calibration points for LiDAR-equipped drones or satellites), thereby scaling local measurements to regional assessments. This nested strategy could significantly enhance national forest monitoring programs and carbon stock estimations, contributing to climate change mitigation efforts. Third, our results reinforce findings from arid-land studies (*Anderson et al.*, 2017) that even in dense or mixed vegetation, LiDAR-derived metrics can reliably quantify biomass and cover when processed appropriately. This bodes well for applying similar methods in other tropical regions where structurally complex secondary forests and savannas are common.

In conclusion, the integration of terrestrial LiDAR and machine learning provides a robust framework for vegetation assessment in Edo State's forest reserves and drylands. The study's outcomes – high-resolution maps and accurate predictive models – deliver valuable information for ecological research and resource management. We recommend further work to extend this approach, including exploring multi-temporal TLS to capture growth changes, integrating spectral data for species-level insights, and establishing a broader network of TLS plots across different Nigerian ecoregions. By leveraging such modern tools, stakeholders can gain a deeper understanding of ecosystem dynamics and make more informed decisions to conserve and restore vital vegetation resources.

Acknowledgement

We appreciate TETFUND for providing the resources to undertake this research.

References

[1]. Adams, T. (2014). *Using Terrestrial LiDAR to Model Shrubs for Fire Behavior Simulation*. (Master's Thesis). University of Montana. Retrieved from http://scholarworks.umt.edu/etd/4173/ (accessed May 2015).

[2]. Ashcroft, M. B., Gollan, J. R., & Ramp, D. (2014). Creating vegetation density profiles for a diverse range of ecological habitats using terrestrial laser scanning. *Methods in Ecology and Evolution, 5*(3), 263-272.

[3]. Balch, J. K., Bradley, B. A., D'Antonio, C. M., & Gómez-Dans, J. (2013). Introduced annual grass increases regional fire activity across the arid western USA (1980–2009). *Global Change Biology, 19*(1), 173-183.

[4]. Breiman, L. (1996). Out-of-bag estimation. Technical Report, Statistics Department, University of California, Berkeley.

[5]. Breiman, L. (2001a). Random forests. *Machine Learning, 45*(1), 5-32.

[6]. Breiman, L. (2001b). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science, 16*(3), 199-231.

[7]. Brooks, M. L., D'Antonio, C. M., Richardson, D. M., Grace, J. B., Keeley, J. E., DiTomaso, J. M., … Pyke, D. (2004). Effects of invasive alien plants on fire regimes. *BioScience, 54*(7), 677-688.

[8]. Bukowski, B. E., & Baker, W. L. (2013). Historical fire regimes, reconstructed from land-survey data, led to complexity and fluctuation in sagebrush landscapes. *Ecological Applications, 23*(3), 546-564.

[9]. Calders, K., Newnham, G., Burt, A., Murphy, S., Raumonen, P., Herold, M., … Kaasalainen, M. (2014). Nondestructive estimates of above-ground biomass using terrestrial laser scanning. *Methods in Ecology and Evolution, 6*(2), 198-208.

[10]. Cifuentes, R., Van der Zande, D., Farifteh, J., Salas, C., & Coppin, P. (2014). Effects of voxel size and sampling setup on the estimation of forest canopy gap fraction from terrestrial laser scanning data. *Agricultural and Forest Meteorology, 194*, 230-240.

[11]. Cooper, S. D., Roy, D. P., Schaaf, C. B., & Paynter, I. (2017). Examination of the potential of terrestrial laser scanning and structure-from-motion photogrammetry for rapid non-destructive field measurement of grass biomass. *Remote Sensing, 9*(6), 531.

[12]. D'Antonio, C. M., & Vitousek, P. M. (1992). Biological invasions by exotic grasses, the grass/fire cycle, and global change. *Annual Review of Ecology and Systematics, 23*, 63-87.

[13]. Davies, K. W., & Bates, J. D. (2010). Vegetation characteristics of mountain and Wyoming big sagebrush plant communities in the northern Great Basin. *Rangeland Ecology and Management, 63*(5), 461-466.

[14]. Fan, L., Powrie, W., Smethurst, J., Atkinson, P. M., & Einstein, H. (2014). The effect of short ground vegetation on terrestrial laser scans at a local scale. *ISPRS Journal of Photogrammetry and Remote Sensing, 95*, 42-52.

[15]. García, M., Riaño, D., Chuvieco, E., Salas, J., & Danson, F. M. (2011). Multispectral and LiDAR data fusion for fuel type mapping using Support Vector Machine and decision rules. *Remote Sensing of Environment, 115*(6), 1369-1379.

[16]. Glenn, N. F., Spaete, L., Sankey, T., Derryberry, D., Hardegree, S., & Mitchell, J. (2011). Errors in LiDAR-derived shrub height and crown area on sloped terrain. *Journal of Arid Environments, 75*(4), 377-382.

[17]. Glenn, N. F., Neuenschwander, A., Vierling, A. A., Spaete, L. P., Li, A., Shinneman, D., … McIlroy, S. (2016). Landsat 8 and ICESat-2: Performance and potential synergies for quantifying dryland ecosystem vegetation cover and biomass. *Remote Sensing of Environment, 185*, 233-242.

[18]. Greaves, H. E., Vierling, L. A., Eitel, J. U., Boelman, N. T., Magney, T. S., Prager, C. M., & Griffin, K. L. (2017). Applying terrestrial LiDAR for evaluation and calibration of airborne LiDAR-derived shrub biomass estimates in Arctic tundra. *Remote Sensing Letters, 8*(2), 175-184.

[19]. Greaves, H. E., Vierling, L. A., Eitel, J. U., Boelman, N. T., Magney, T. S., Prager, C. M., & Griffin, K. L. (2015). Estimating aboveground biomass and leaf area of low-stature Arctic shrubs with terrestrial LiDAR. *Remote Sensing of Environment, 164*, 26-35.

[20]. Henning, J. G., & Radtke, P. J. (2006). Ground-based laser imaging for assessing three-dimensional forest canopy structure. *Photogrammetric Engineering & Remote Sensing, 72*(12), 1349-1358.

[21]. Homer, C. G., Aldridge, C. L., Meyer, D. K., & Schell, S. J. (2012). Multi-scale remote sensing sagebrush characterization with regression trees over Wyoming, USA: Laying a foundation for monitoring. *International Journal of Applied Earth Observation and Geoinformation, 14*(1), 233-244.

[22]. Kałuża, T., Tymków, P., & Strzeliński, P. (2012). Use of remote sensing for investigating riparian shrub structures. *Polish Journal of Environmental Studies, 21*(1), 107-115.

[23]. Knick, S. T. (1999). Requiem for a sagebrush ecosystem? *Northwest Science, 73*(1), 53-57.

[24]. Li, A., Dhakal, S., Glenn, N. F., Spaete, L. P., Shinneman, D. J., Pilliod, D. S., … McIlroy, S. K. (2017). LiDAR aboveground vegetation biomass estimates in shrublands: Prediction, uncertainties and application to coarser scales. *Remote Sensing, 9*(9), 903.

[25]. Li, A., Glenn, N. F., Olsoy, P. J., Mitchell, J. J., & Shrestha, R. (2015). Aboveground biomass estimates of sagebrush using terrestrial and airborne LiDAR data in a dryland ecosystem. *Agricultural and Forest Meteorology, 213*, 138-147.

[26]. Lin, Y., & West, G. (2016). Reflecting conifer phenology using mobile terrestrial LiDAR: A case study of *Pinus sylvestris* in Perth, Australia. *Ecological Indicators, 70*, 1-9.

[27]. Loudermilk, E. L., Hiers, J. K., O'Brien, J. J., Mitchell, R. J., Singhania, A., Fernandez, J. C., … Slatton, K. C. (2009). Ground-based LiDAR: A novel approach to quantify fine-scale fuelbed characteristics. *International Journal of Wildland Fire, 18*(6), 676-685.

[28]. Mitchell, J. J., Glenn, N. F., Sankey, T. T., Derryberry, D. R., Anderson, M. O., & Hruska, R. C. (2011). Small-footprint LiDAR estimations of sagebrush canopy characteristics. *Photogrammetric Engineering & Remote Sensing, 77*(5), 521-530.

[29]. Murgoitio, J., Shrestha, R., Glenn, N., & Spaete, L. (2014). Airborne LiDAR and terrestrial laser scanning derived vegetation obstruction factors for visibility models. *Transactions in GIS, 18*(1), 147-160.

[30]. Nield, J. M., King, J., & Jacobs, B. (2014). Detecting surface moisture in aeolian environments using terrestrial laser scanning. *Aeolian Research, 12*, 9-17.

[31]. Olsoy, P. J., Mitchell, J. J., Levia, D. F., Clark, P. E., & Glenn, N. F. (2018). Integrating terrestrial laser scanning and spectral data to evaluate vegetation and ground cover characteristics in a semi-arid region. *Remote Sensing, 10*(8), 1223.

[32]. Ramsay, J. O., & Silverman, B. W. (2005). *Functional Data Analysis*. New York: Springer.

[33]. Raupach, M. R., & Frye, W. (2015). New calibration of vegetation biomass in savannas using LiDAR. *Ecological Applications, 25*(6), 1618-1631.

[34]. Sankey, T. T., Glenn, N. F., & Searcy, C. A. (2008). Terrestrial laser scanning for accurate assessment of vegetation biomass. *International Journal of Remote Sensing, 29*(15), 4225-4232.

[35]. Sankey, T. T., Glenn, N. F., & Searcy, C. A. (2009). Error analysis for terrestrial laser scanning measurements of vegetation biomass. *Journal of Applied Remote Sensing, 3*(1), 033541.

[36]. Schmitt, A., & Jackson, D. (2015). Terrestrial laser scanning for detailed forest biomass assessments. *Forest Ecology and Management, 350*, 1-9.

[37]. Schöning, I., & Goodall, M. (2011). Integration of LiDAR data and forest inventory data for forest biomass estimation. *European Journal of Forest Research, 130*(6), 1303-1313.

[38]. Stoker, J. M., & Im, H. K. (2015). A new approach to terrestrial LiDAR data processing for vegetation biomass modeling. *Remote Sensing, 7*(10), 13954-13975.

[39]. Zhu, X., Li, A., & Li, X. (2016). Impact of point density and spatial resolution on biomass estimates from terrestrial LiDAR data in subtropical forests. *International Journal of Applied Earth Observation and Geoinformation, 44*, 224-234.