remote sensing

MDPI

Article

Integration of Structural Characteristics from GEDI Waveforms for Improved Forest Type Classification

Mary M. McClure [1], Satoshi Tsuyuki [2,*] and Takuya Hiroshima [2]

[1] Independent Researcher; mmmcclu2@alumni.u-tokyo.ac.jp
[2] Department of Global Agricultural Sciences, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo 113-8657, Japan; hiroshim@g.ecc.u-tokyo.ac.jp
* Correspondence: tsuyuki@fr.a.u-tokyo.ac.jp; Tel.: +81-3-5841-7509

Abstract: Forest types correspond to differences in structural characteristics and species composition that influence biomass and biodiversity values, which are essential measurements for ecological monitoring and management. However, differentiating forest types in tropical regions remains a challenge. This study aimed to improve forest type extent mapping by combining structural information from discrete full-waveform LiDAR returns with multitemporal images. This study was conducted in a tropical forest region over complex terrain in north-eastern Tanzania. First, structural classes were generated by applying time-series clustering algorithms. The results showed four different structural clusters corresponding to forest types, montane–humid forest, montane–dry forest, submontane forest, and non-forest, when using the Kshape algorithm. Kshape considers the shape of the full-sequence LiDAR waveform, requiring little preprocessing. Despite the overlap amongst the original clusters, the averages of structural characteristics were significantly different across all but five metrics. The labeled clusters were then further refined and used as training data to generate a wall-to-wall forest cover type map by classifying biannual images. The highest-performing model was a KNN model with 13 spectral and 3 terrain features achieving 81.7% accuracy. The patterns in the distributions of forest types provide better information from which to adapt forest management, particularly in forest–non-forest transitional zones.

Keywords: waveform LiDAR; forest structure; clustering; time-series

check for updates

Citation: McClure, M.M.; Tsuyuki, S.; Hiroshima, T. Integration of Structural Characteristics from GEDI Waveforms for Improved Forest Type Classification. *Remote Sens.* **2024**, *16*, 4776. https://doi.org/10.3390/rs16244776

Academic Editor: Brenden E. McNeil

Received: 12 November 2024
Revised: 14 December 2024
Accepted: 19 December 2024
Published: 21 December 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Mapping the extent of different forest types is valuable for developing forest management plans, inferring biodiversity, and estimating above-ground biomass. However, differentiating forest types is still a challenge due to similarities in spectral signatures, topographic effects in mountainous terrain and seasonal variability. LiDAR (Light Detection and Ranging), hyperspectral sensors, and high-resolution imagery complimented with extensive field inventory have improved our ability to distinguish and monitor forest types. However, the application of these systems was previously limited by high operational costs and sparse spatial coverage [1]. To overcome these limitations, NASA launched the Global Ecosystem Dynamics Investigation (GEDI) mission in 2018. GEDI offers high-resolution LiDAR data in 25 m noncontiguous footprints, enabling more detailed insights into the vertical structure of forests [2]. In contrast to previous satellite-based datasets like ICESat, GEDI has been calibrated for vegetation monitoring. This capability is essential for accurately modeling canopy height and estimating above-ground biomass, two integral parameters for understanding ecosystem functions, such as CO_2 sequestration rates, which are influenced by forest type. GEDI's datasets, particularly the L1B geolocated waveforms and L2A canopy height metrics, have demonstrated efficacy in terrestrial monitoring, providing insights into canopy height, biomass, and species diversity.

For example, Yun et al. (2023) used GEDI L1B data, along with airborne LiDAR, to classify forest overstory and understory layers. Their study demonstrated the potential

of L1B data in providing insights into the multi-layer structure of forests by extracting waveform metrics such as multiple waveform peaks, canopy cover, and terrain slope [3]. Hakkenberg et al. (2023) investigated the relationships between forest structural metrics derived from GEDI spaceborne LiDAR and Airborne Laser Scanning (ALS) data with species diversity by comparing these metrics against nearby field inventory data. Their study showed significant associations between GEDI-derived structural attributes and alpha, beta, and gamma diversity, thus enhancing our understanding of how LiDAR technology can be used for biodiversity assessment across different biomes [4]. Furthermore, Dwiputra et al. (2023) used GEDI L1B-derived metrics, along with L2A metrics, to improve land cover classification. The authors first extracted relevant structural parameters from the data sets and then labeled the GEDI footprint using very-high-resolution imagery. They found that, when these structural parameters were included with optical data, the classification accuracy significantly improved [5].

Despite these advancements, there remains a need to explore unsupervised learning methods in characterizing forest structure, which can be particularly valuable in regions lacking labeled datasets. Traditional approaches have relied on supervised techniques, necessitating concurrent or nearby ground truth data. Furthermore, by employing unsupervised techniques, we can potentially find patterns in LiDAR waveforms that correlate with different vegetation structures otherwise missed when relying on optical data alone. A novel study aimed at integrating small-footprint airborne LiDAR-derived metrics in land cover classification with unsupervised learning was conducted by Moran et al. (2018) in Montana, USA. They used metrics from LiDAR to segment structural classes by applying random forest models to find similarities between LiDAR metrics and then performed hierarchical clustering to group the data by the estimated similarities. Finally, they used the cluster labels to conduct a supervised classification of land cover [6]. Another study by Alexander et al. (2018) conducted in Sumatra, Indonesia, used ALS data to characterize forest patches by extracting gridded structural metrics from a canopy height model. Then, they clustered these metrics using kMedoids and found correlations between field-identified forest types and these clusters, proving the effectiveness of ALS in mapping tropical forest types [7]. To our knowledge, unsupervised applications with large-footprint LiDAR are scarcer in the literature and have not yet been conducted with GEDI L1B.

This study is building from the earlier works on unsupervised vegetation mapping to investigate the potential of time-series clustering methods to classify GEDI L1B waveform returns. Our hypothesis is that waveforms of similar vegetation will show distinctive structural patterns, allowing for unsupervised grouping of forest types. These groups can be combined with optical data to create wall-to-wall maps of forest structural types, enhancing our ability to map and monitor forest ecosystems. This method is advantageous in that it leverages both structural and spectral characteristics with limited labeled data.

The facet of unsupervised learning that this study explores is time-series (TS) clustering. TS clustering partitions chronological data so that sequences within each group exhibit similar patterns, unlike traditional clustering methods, which typically handle static data points [8]. TS clustering can be broadly grouped into three types: whole TS clustering, subsequence TS clustering, and time-point clustering. The whole TS approach treats each TS as a single entity and uses specific distance measures along with traditional clustering algorithms to group similar series. Common distance measures include Dynamic Time Warping (DTW) for series of unequal lengths, Euclidean distance for equal-length series, or shape-based similarity measures to capture the unique characteristics of temporal data [8,9]. The subsequence TS clustering method involves dividing a time-series into smaller subsequences and clustering these segments, which is useful for identifying recurring patterns within a single time-series. Finally, time-point clustering identifies patterns and similarities at specific moments in time across multiple series, which can be particularly useful for detecting anomalies or changes in data behavior over time [10]. This study opted to use the whole-TS approach, making the assumption that the global shape of the sequence (whole waveform), rather than recurrent patterns within the waveform is most applicable for

distinguishing vegetation structural classes. Currently, metrics derived from the waveform after rigorous preprocessing, such as wavelength, energy, and amplitudes, are correlated with the vegetation structure (relative heights, vegetation cover, volume, etc.) and then used to characterize landscape vegetation [2,11]. Applications of TS analysis of LiDAR data in vegetation studies have been limited. However, Vaughn et al. (2011) used the Fourier transformation method on contiguous ALS data to differentiate tree species in the Pacific Northwest. Their study successfully classified three hardwood species with an overall accuracy of 75%, identifying key frequencies that aided in species differentiation [12].

Given that earlier studies have proven that features from waveforms can help to characterize landscapes, TS clustering may be a useful tool for automatic forest and land cover segmentation. This method retains information of the entire wavelength otherwise lost when extracting specific metrics, while clustering allows us to draw quick insights into locations for which there is little data. Another advantage of TS clustering specifically is that they are designed to handle noisy data that could reduce the need for waveform decomposition to perform rapid classification tasks [10,13].

The purpose of this research was twofold. First, investigating, as a proof of concept, the efficacy of unsupervised TS clustering methods for grouping GEDI LiDAR waveforms by forest type. The tasks to achieve this purpose were set as follows:

1. Comparing clustering algorithms via internal validation indices;
2. Validating by means of multiple comparison analysis and cross-checking with field inventory.

The second purpose was to generate a forest-type map incorporating structural and optical data by:

1. Integrating structural characteristics by correlating the clusters with wall-to-wall raster data by machine-learning models;
2. Validating the classification map against forest inventory data.

2. Materials and Methods

2.1. Study Area and Inventory Data

This study was conducted in Same, Tanzania, centering on Kwizu Forest Reserve. Although a lesser-known forest, this reserve is in the Eastern Arc Mountains, which is renowned as a biodiversity hot spot [14–16]. Kwizu forest was officially gazetted as a production forest in 1953, but in the 1990s, harvesting in Kwizu forest was banned in an effort to maintain the functional integrity of the water catchment [17]. In 1998, the management transitioned to joint forest management (JFM) between the surrounding villages and the district council [17].

According to the World Database for Protected Areas (WDPA), the reserve encompasses approximately 2981 hectares [18]. It is important to note that there is also agriculture, agroforestry, and settlement land use within the reserve bounds. The southern border of the reserve is adjacent to villages where the forest cover transitions to open forest—agroforestry land use—in the highlands and non-forest—grain and legume cropping systems—in the lowlands. The northern border of the reserve transitions into Mkomazi National Park (Figure 1).

The forest inventory was collected in July of 2023. The inventory included measurements of Diameter at Breast Height (DBH), tree Height (H), Crown Diameter (CD), and species identification. There were 32 plots with a concentric circle design of 2 m, 6 m, and 12.5 m radii to match the area covered by a GEDI footprint. The plots in the reserve were stratified by elevation and a preliminary model of biomass to capture the compositional and structural heterogeneity of the reserve. The limitations of this study included the sample size, and the uneven allocation of validation plots concentrated in the southern portion of the study area. Therefore, it was necessary to incorporate LiDAR and optical datasets for a more comprehensive assessment of the entire reserve.

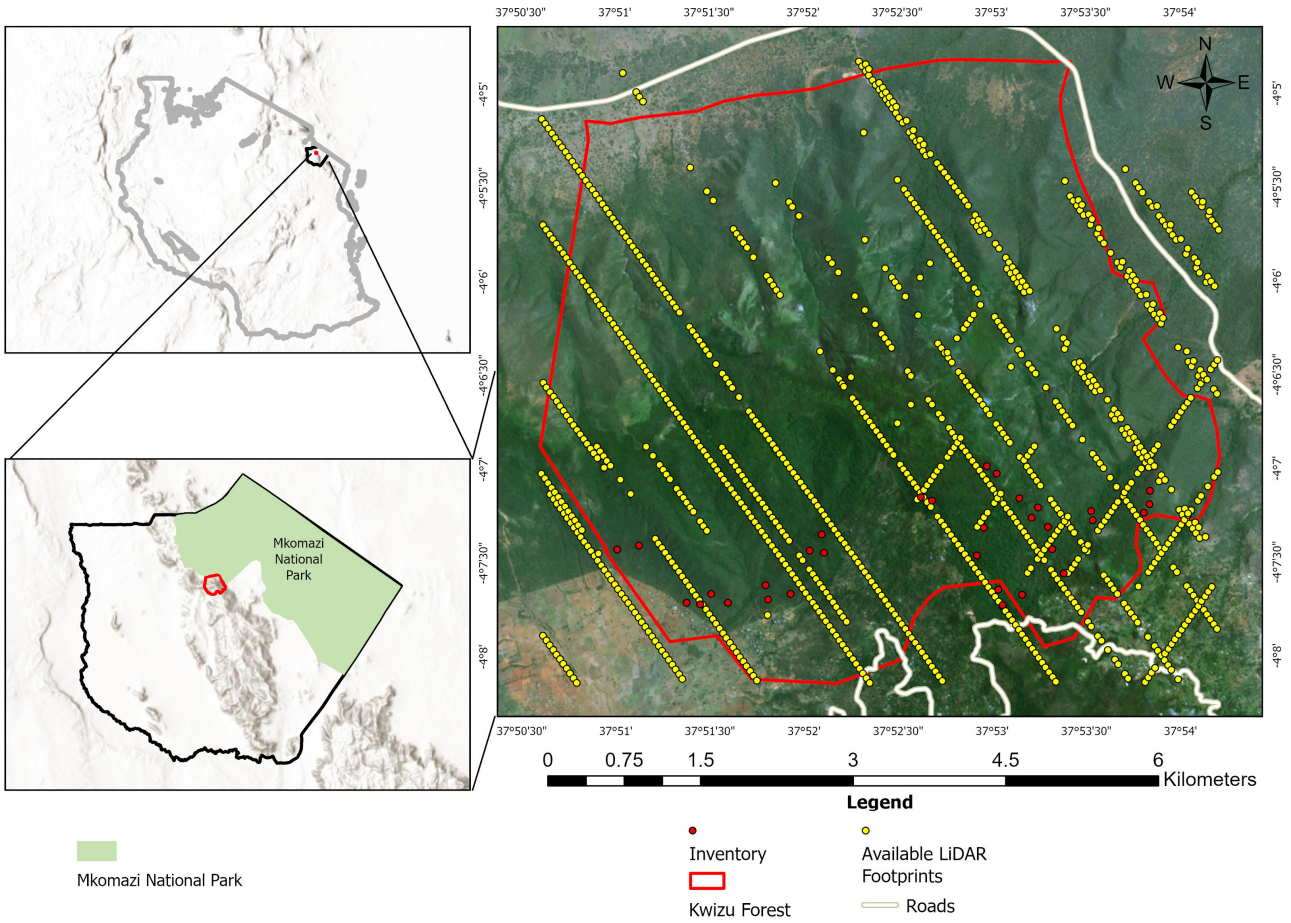


Figure 1. Location and extent of the study area within Same District, Tanzania. The red border represents the forest reserve boundary provided by WDPA [18]. The yellow points represent the available geolocated footprints within the bounding box of the reserve after filtering by quality flags, while the red points represent the locations of forest inventory plots.

According to Lovett et al. (2014) and Kajembe et al. (2008), regions within Kwizu Forest reserve meet the criteria for tropical montane forests—elevation between 1200–1800 m and annual rainfall up to 1200 mm—and dry submontane forests—800–1200 m in elevation and annual rainfall less than 1000 mm [17,19]. Kwizu forest reserve ranges in elevation between 730 m and almost 1900 m a.s.l., while precipitation varies from around 350 mm to 1200 mm annually [17]. In addition to the elevation and rainfall criteria of Lovett et al. (2014), several studies on forests in the Eastern Arc Mountains have shown how species composition changes in relation to microclimatic factors: temperature, soil, and distance to anthropogenic activity [20,21]. Kwizu forest also exhibited high variation in vegetation composition and structure (Tables 1 and 2) that led to further dividing the montane region into montane–dry and montane–humid sub-groups.

Table 1. Visuals of different forest types within Kwizu Forest reserve.

Description	Image 1	Image 2
Image 1: Profile of montane humid forest with multilayer closed canopy. Image 2: Understory conditions of montane–humid forest.		
Image 1: Montane–dry forest with multilayer canopy consisting of taller thin trees reaching the upper layer of the canopy and many short trees that make up the understory. Image 2: Montane–dry forest in which there is a mixture of *Euphorbia* sp. and *Albizia* sp. at 1300 m in elevation.		
Image 1: Submontane forest with open canopy in which few trees have reached heights greater than 5 m. Image 2: The top layer of the canopy is open and comprises *Adansonia digitata* and *Euphorbia* sp. while the understory is dominated by shrubs.		

Table 2. Inventory summary of structural characteristics by forest.

Forest Type	Maximum Tree Height (m)	Mean Tree Height (m)	Mean DBH (cm)	Average Basal Area	Biomass (tons/ha)	S-H Diversity Index	Elevation (m a.s.l.)
Montane–humid	29 ± 3.6	13 ± 2	28.3 ± 6.8	2.20 ± 0.78	325 ± 102	1.42 ± 0.37	1558 ± 122
Montane–dry	17 ± 3.1	9 ± 1.5	19.2 ± 2.8	1.00 ± 0.20	152 ± 33	1.22 ± 0.49	1319 ± 177
Submontane	15 ± 5.2	5 ± 1	12.7 ± 2.2	0.39 ± 0.08	95 ± 31	1.24 ± 0.59	1139 ± 111

Grouping the forest inventory plots into different forest types involved partitioning the area by an elevation threshold of 1200 m. Then, the montane forest was further divided into humid and dry forest by considering the structural and species composition in terms of Dominant Height (DH) and the species Importance Value Index (IVI) per plot. The IVI

considers the density of a particular tree species in a plot compared with the total number of trees and the basal area of tree species relative to the total basal area of the plot according to Equations (1)–(3) [22]:

$$\text{relative density} = \frac{\text{number of observations of species i}}{\text{number of all observations in the plot}} \quad (1)$$

$$\text{relative basal area} = \frac{\text{basal area of species i}}{\text{total basal area of all obsevations in the plot}} \quad (2)$$

$$\text{IVI} = \frac{\text{relative density}_{\text{i}} + \text{relative basal area}_{\text{i}}}{2} \times 100 \quad (3)$$

The characteristics of the group of plots labeled as montane–humid forest included a higher frequency of evergreen species: *Leptoncycia usambarensis* and *Xymalos monospora,* and deciduous species: *Newtonia buchananii.* This forest type was generally located at higher elevations where precipitation was greater for both rainy seasons. In more than half of these plots, the dominant species by DH or IVI were *Newtonia buchananii* and *Ficus sp.;* however, this area had higher biodiversity values on average, as shown by the Shannon–Wiener Diversity Index (SH) in Table 2. The forest structure was influenced by more observations of large trees than other areas, as indicated by the higher average DBH and average heights (Table 2). The plots labeled as montane–dry forest were located in mid-altitude regions, where *Albizia gummifera*, *Albizia schimperiana*, and *Albizia anthelmintica* were the most dominant species by height or IVI. The upper layer of the canopy in this region reached 17 m on average (Table 2), much lower than that of montane–humid forest plots. Finally, the submontane plots were dominated almost exclusively by *Euphorbia bussei* (a succulent tree), apart from one plot, which had a large baobab tree (*Adansonia digitata*). Structurally, this area had the lowest average tree height and few observations of large trees (Table 2).

2.2. Remote Sensing Data Collection

This study used GEDI datasets L1B, L2A, and L2B for analysis and a combination of raster datasets and forest inventory for classification and validation. The GEDI L1B dataset corresponded to the geolocated return waveform data provided as binned amplitudes [2]. GEDI L2A data are the waveforms processed to relative height, RH 1–100, which are the heights at different percentages of returned energy [11]. GEDI L2A was further processed to extract metrics such as canopy cover, plant area index, plant volume, etc., and were available in the GEDI L2B dataset [23]. The GEDI datasets were downloaded from NASA's Earthdata repository (https://search.earthdata.nasa.gov/search accessed on 17 January and 10 August 2024) [24].

There were seventeen available granules intersecting the study boundaries. These granules were further subset following the procedures described in the processing tutorials provided by the Land Processes Distributed Active Archive Center, which can be found at (https://github.com/nasa/GEDI-Data-Resources accessed on 17 January 2024 [25]. The procedure included spatial sub-setting (clipping the transects to the bounding box) and quality filtering the footprints. Quality filtering involved removing shots where the 'quality_flag' was 0, the 'degrade_flag' was 1, the slope of the footprint was 30 degrees or greater, and the sensitivity was >0.95. After initial clustering trials, the sensitivity threshold was increased to >0.98 for elevations greater than 1200 m to account for the denser vegetation in this area; this decreased the number of observations from 712 to 495. The sensitivity and slope thresholds—which influence ground detection in GEDI data products—were based on recommendations from previous studies [26–28]. Footprints that fell outside of the boundaries of the reserve were not removed to retain a higher sample size and to quickly assess if the clustering methods were distinguishing non-vegetated areas from vegetated areas via visual inspection, as several other studies have proven that non-vegetated land has a distinct waveform shape [2,5].

The L2A and L2B datasets provide a comprehensive list of waveform-derived features that do not require further processing. However, several additional metrics were calculated from the "significant part" of the L1B dataset based on a literature review. The "significant" part of the waveform refers to the portion of the waveform between the ground location determined by the location of the lowest mode ('elev_lowestmode') until the canopy top indicated by the location of the mode of highest return ('elev_highestreturn') [2,11]. The full set of features used for characterizing waveforms can be found in Table A1 of Appendix A. Processing the L1B waveforms and clustering analysis were conducted using Jupyter Notebook; however, several features (fslope and roughness) were calculated using the waveformlidar package in RStudio [29].

In addition to the GEDI datasets, a Digital Elevation Model (DEM) from the Shuttle Radar Topography Mission (SRTM) and spectral data from three different image providers: PlanetScope-Norway's International Climate and Forests Initiative (NICFI) Satellite Data Program, European Space Agency's Sentinel 2 imagery, and the USGS's Landsat 8 OLI imagery, were used for classification [30–33]. A summary of the datasets can be found in Table 3. The DEM and spectral datasets are publicly available in the Google Earth Engine (GEE) data catalog [34]. The spectral images were grouped by dates 1 January–30 June 2023 as a proxy for the rainy season and 1 July–31 December 2023 as a proxy for the dry season. For the NICFI data set, images with evident cloud cover over large areas of the study site were removed manually, leaving one image in the dry season (October) and three images in the wet season, which were composited using the median value of the collection. Sentinel 2 images were filtered by cloud cover less than 30%, cloud-masked using the cloud probability score, and finally composited by the median values to reduce atmospheric effects from haze. There were 18 images available for the wet season and 24 images for the dry season; the image acquisition times varied between 7:26 and 7:58 AM. The cloud probability score from the complimentary "S2 _CLOUD_PROBABILITY" data was used to mask clouds in the sentinel images before creating the composite images. Landsat 8 images were also filtered by cloud cover less than 30%, cloud-masked using the quality assessment bands, and then composited using the '.qualityMosaic()' function with the Normalized-Difference Vegetation Index (NDVI) to reduce haze. Landsat 8 only had two images available (one in February and one in March) that met the cloud filtering criteria. The acquisition times for these images were 7:38 AM and 7:37 AM according to the image metadata. All optical data were resampled to 30 m resolution to match the DEM using nearest neighbor resampling in GEE. The raster data sets were used for correlating the noncontiguous structural clusters with wall-to-wall coverage to generate a complete forest cover map over the study area.

Table 3. Comparison of the raster datasets used for classification.

Data Type	Source: GEE Data Catalogue	Sensor	Spatial Resolution	Spectral Range
Multispectral Image	"projects/planet-nicfi/assets/basemaps/africa"	SuperDove	4.77 m	Visible, Near-Infrared
Multispectral Image	"COPERNICUS/S2_SR_HARMONIZED" & "COPERNICUS/S2_CLOUD_PROBABILITY"	Sentinel 2 MultiSpectral Instrument	10–20 m	Visible, Near-Infrared, and Shortwave Infrared
Multispectral Image	'LANDSAT/LC08/C02/T1_L2'	Landsat 8: Operational Land Imager	30 m	Visible, Near-Infrared, and Shortwave Infrared
Digital Elevation Model	"USGS/SRTMGL1_003"	SRTM: Radar interferometer	30 m	Elevation

2.3. Clustering Analysis

Clustering analysis was conducted in the Python environment with tslearn and sklearn libraries, which provide an extensive list of unsupervised learning algorithms [35,36]. As highlighted by Lima et al. (2023), data normalization was a necessary step to ensure

meaningful comparisons of time-series data. Thus, the waveforms were normalized using the 'TimeSeriesScalerMeanVariance()' function from the tslearn library, which standardizes each waveform individually so that it has a mean of zero and unit standard deviation; this minimizes the effects of amplitude outlier values [37]. The assumption in this scaling method is that the shape of the data is more important than the range [37]. Although there are many clustering algorithms, partitioning-based algorithms, Kmeans and Kshape, with attention to the differences in distance metrics, DTW and shape-based distance (SBD), were tested. Kmeans was selected as the most comparable algorithm to Kshape as both use an iterative refinement procedure to assign clusters [13]. The TS distance metrics were chosen as they were based on two different underlying principles, the dynamic programming principle for DTW and cross-correlation for SBD [38].

Challenges in unsupervised learning include determining the number of appropriate clusters and hyperparameter tuning. To address these challenges, an iterative approach was taken, whereby different parameters and cluster sizes were evaluated using the silhouette score and elbow method to optimize partitioning. The silhouette score assesses the level of agreement of an observation and its cluster by calculating the average distance between a data point and the other points in its cluster, and the distance of the point to the nearest other cluster [39]. A score closer to 1 indicates perfect clustering, scores close to 0 mean that the clusters may be very close or overlapping, and a negative score indicates that the observation was misassigned [39]. The optimal combination and number of clusters were determined by ensuring that each cluster was positive and near the average silhouette score (no one cluster performed particularly poorly, while another performed well). The elbow method involves plotting the explained variance or within-cluster sum of squares as a function of the number of clusters; the optimal number of clusters is the number in which the explained variance becomes marginal [39]. If both methods were not in agreement for the optimal number of classes, then iterations for both optimal classes were run and the number of groups was determined by comparing two additional metrics: the Davies–Boulding (D-B) Index and Calinski–Harabasz Index (C-H). The Davies–Bouldin index compares the average similarity ratio of each cluster with its most similar cluster and describes the overall cluster separation and compactness, with lower values indicating better clustering [40]. The C-H Index (Variance Ratio Criterion) is the ratio of the sum of between-cluster dispersion to within-cluster dispersion [41]. Knowledge of the study area was also applied to lower computation time by limiting the number of possible clusters (2–10) based on general land cover types.

2.4. Clustering Evaluation

Several approaches were used to evaluate the quality of clusters in this analysis. Although the cluster indices described in Section 2.3 may indicate the quality of clusters in terms of separation and cohesion, it does not confirm that the clusters reflect patterns in biophysical characteristics. Thus, the clusters were characterized by the extracted waveform features shown in Table A1 in Appendix A and multiple comparison analysis was performed using a Kruskal–Wallis test with Bonferroni correction followed by a post hoc Dunn's Test to determine which clusters were significantly different per feature. The clusters were then mapped to visually inspect any spatial patterns that may correspond to vegetation or land cover types and to associate clusters with forest inventory.

2.5. Forest Type Classification

Different machine learning algorithms, Random Forest (RF), gradient boost, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Voronoi polygon clustering, were tested for mapping. The models were trained on the clustered GEDI footprints. The training set included 70% of the clustered GEDI footprints, with the remaining 30% used for validation along with the field inventory data.

Bands from the optical datasets: NICFI, Sentinel 2 and Landsat 8, and DEM were used to correlate the structural clusters with raster data so that a wall-to-wall forest type map

could be created. Spectral indices from the optical datasets were calculated to enhance differences in forest cover types and environmental features; a list of the indices tested is provided in Table A2 of Appendix A. In addition to optical data, topographic features: elevations, slope, hill shade, and aspect, were derived from the DEM and incorporated in the classification models.

Features for each model were selected following three different methodologies; the first: running an initial gradient-boost model with all available bands and 100 decision trees to rank the features by importance, the second: performing recursive feature elimination iterating from 5–20 bands, and the third: forward feature selection retaining 5–20 bands. The best model was determined by overall accuracy.

3. Results

3.1. Clustering Results

The optimal number of structural clusters was four or five depending on the algorithm applied. The study area included some regions that were under non-forest land cover and three potential forest types; agroforestry was considered forest, despite having a different land use than reserved forest. Therefore, the number of optimal clusters determined from internal validation scores was reasonable. A summary of the clustering evaluation metrics can be found in Table 4. In general, the internal validation indices: silhouette score, Davies–Bouldin, and Callinski–Harabasz, were low for the TS machine learning methods for both algorithms. Silhouette scores less than 0.5 indicated that clusters were similar to each other with potential overlap. Kshape performed slightly better in terms of internal validation, as shown by the higher silhouette and C-H scores and lower D-B score in comparison with the Kmeans algorithm with DTW distance. K-Shape is designed to cluster time series based on their overall shape, using normalized cross-correlation as a similarity measure. This approach is time-invariant, shifting the time series for better alignment [38]. This algorithm focuses on grouping by the underlying patterns, making it more applicable, where the shape of the waveform is correlated with the structure of the vegetation. In contrast, DTW is time-warping in that it stresses or compresses the series to find the lowest-cost distance between the two data sets, but may not necessarily consider the global shape [9]. Kshape was used for subsequent analysis, given the slightly better clustering index scores, and for its computation efficiency over DTW.

Table 4. Internal evaluation metrics of the time-series clustering algorithms.

Clustering Algorithm	Number of Clusters	Silhouette Score	Davies–Bouldin Index	Callinski–Harabasz Index
Time-Series Kmeans (DTW)	5	0.15	2.16	150.79
K-Shape (shape-based)	4	0.23	1.67	295.20

3.1.1. Multiple Comparison Analysis of Metrics

Due to the indication of overlap in the clusters, it was necessary to determine if the clusters represented statistically different groups in terms of structural characteristics. This check was performed via multiple comparison analysis using the Kruskal–Wallis test with Bonferroni correction. This test was chosen as it is considered more conservative in that it is more likely to give a false negative than a false positive [42]. The waveform metrics found in Table A1 (Appendix A) were calculated from the waveform itself or extracted from the associated L2A and L2B GEDI datasets by the shot identification number. The averages of the structural metrics for each cluster were used for the statistical comparison as a more practical and intuitive way of understanding the differences across groups.

Despite the low internal validation scores and some spatial overlap, most features showed highly significant results, with p-values well below the standard significance threshold of 0.05, even after applying Bonferroni correction, indicating clear structural difference across cluster averages. Only five features: '0–2 m', 'Plant Volume 12', 'Plant

Volume 13', 'Cover_z 11', and 'Cover_z 12', were not found to be significantly different for any group. The feature '0–2 m' in this context refers to the number of detected peaks between the ground and two meters above the ground. The number of modes between the ground and two meters above the ground likely represented the ground and/or the first layer of vegetation. The features 'Cover_z 11' and 'Cover_z 12' from the GEDI L2B dataset refer to the estimated vegetation cover in the eleventh and twelfth vertical bin, respectively, which would be approximately 55 m and 60 m from the ground. 'Plant Volume 12' and 'Plant Volume 13' are also from the L2B dataset and refer to the estimated volume of plant material in the vertical bins 12 and 13 (approximately 60 m and 65 m from the ground). Given that the maximum height of trees in this area was around 30 m (Table 2), vertical bins 11–13 were likely insignificant because few observations were present across all groups. A post hoc Dunn's test was conducted to further examine which specific groups were significantly different per feature, as shown in Figure 2. Waveform characteristics: TotalAmp (sum of amplitude values), energy (sum of amplitudes within the 'significant' part of the waveform), and wavelength (length of returns within the significant portion of the waveform) had exceptionally low p-values (in the range of $10e^{-130}$) and higher η^2 values (above 0.6). This suggests that these features explain a large portion of the variance in the outcome of the clustering.

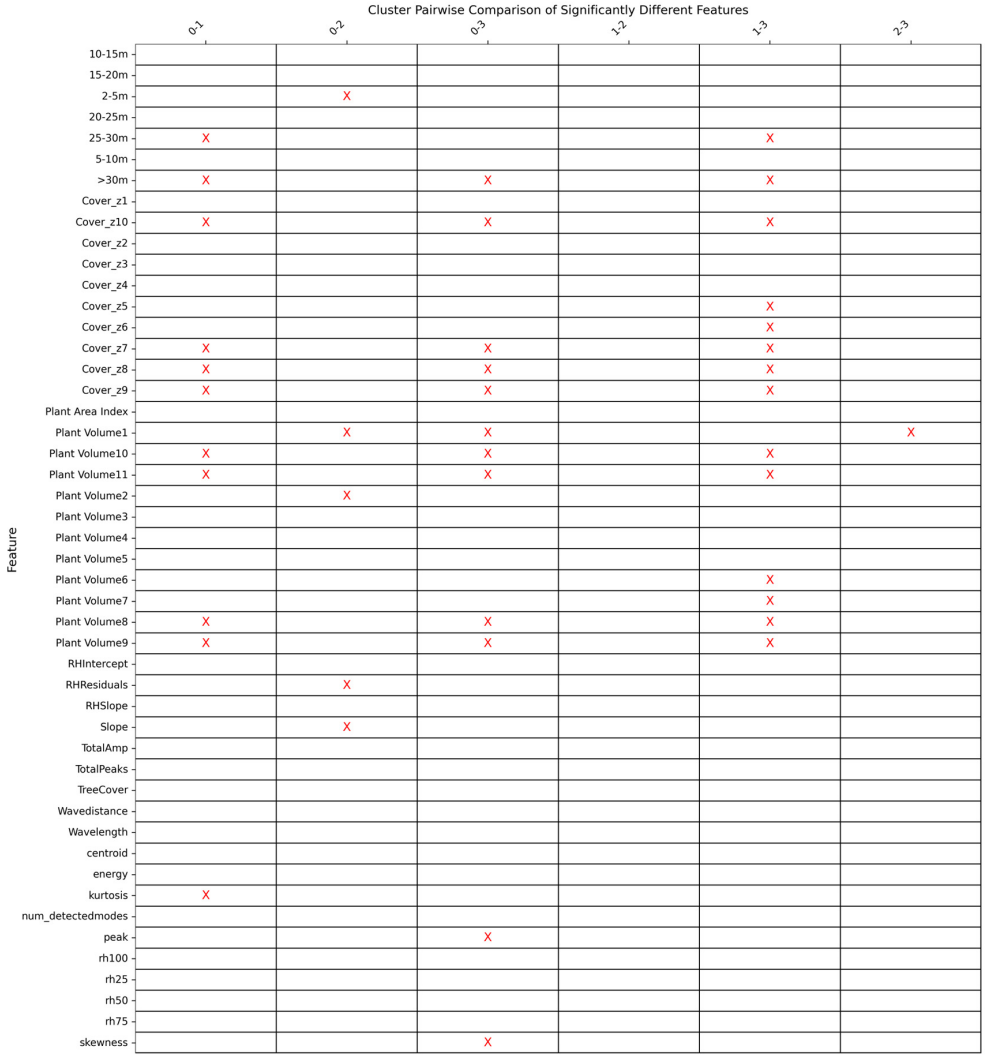


Figure 2. Dunn's test results showing the significant and nonsignificant (denoted by an X) features between cluster pairs. The full table of the Kruskal–Wallis test results can be found in Appendix C. Feature explanations can be found in Appendix A (Table A1).

Other waveform characteristics, such as the centroid and peak (value of the maximum amplitude), were found to be highly significant (adjusted p-value of $8.16e^{-127}$ and $4.26e^{-56}$, respectively), while skewness and kurtosis were less so. In terms of the derived biophysical metrics, relative height rh100 (described in Appendix A) showed a strong significance (e.g., rh100: H = 609.60, $p = 8.35e^{-102}$) and was significantly different across all groups. Relative heights rh25, rh50, and rh75 were also significant, suggesting the importance of the vertical structure in distinguishing groups. However, it is important to note that the wavelength, wave distance, and rh values were directly correlated (Figure 3).

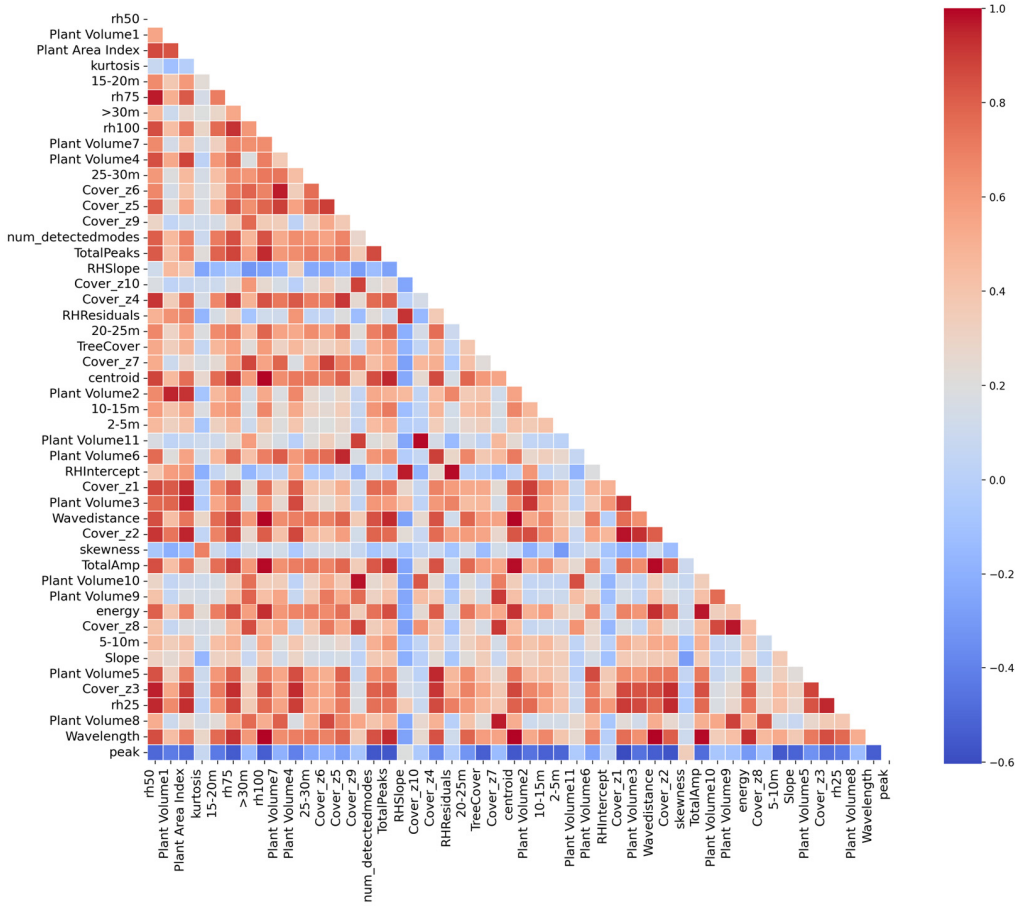


Figure 3. Correlation heat map among structural features. The plot shows high correlations between waveform metrics such as wavelength, wave distance, and centroid with derived vegetation metrics, such as the relative heights. Plant cover, volume, metrics, and rh100 showed a strong negative correlation with 'peak' amplitude, but a positive correlation with 'TotalAmp', likely caused by the scattering of signal in denser, taller vegetation.

Features related to waveform shape complexity, and thus structural complexity, i.e., 20–25 and 25–30 (number of peaks 15–20 and 20–25 m from the ground respectively), also showed high significance (adjusted p-values: 9.72^{-95} and 1.99^{-91}, respectively), suggesting that the upper canopy structure varied among groups. Vegetation structure metrics from L2B that had exceptionally high H and η^2 statistics and low p-values were Plant Volume 5–7. These metrics are binned values of estimated plant volume corresponding to the estimated plant volume at heights between 25–35 m, associated with the top layer of the canopy for humid and dry montane forest types. Plant Volume 10 and 11 at heights greater than 50 m from the ground were more likely due to waveform processing errors (poor ground detection), rather than actual vegetation differences, while Plant Volume 12 and 13 were not significant. Furthermore, values independent of the waveform return: terrain slope and tree cover estimated from Landsat products, were also found to be highly significant

across groups, suggesting consistency with external analysis and topographical variation in forest composition.

Overall, the analysis demonstrated that most features had significant differences between at least two groups and thus, on average, significant differences in the vegetation structure with which the clusters coincided. The correlation among features shown in Figure 3 also suggested that fewer metrics that target height, density and complexity variations, such as wavelength (or relative heights), number of modes, and maximum amplitude may be most valuable features for distinguishing different forest types.

3.1.2. Comparison with Inventory Data

Based on the characteristics of the cluster means and proximity to forest inventory, each structural cluster was assigned to a forest type. Field inventory data were matched to the nearest GEDI footprint within a 300 m search radius using the 'near' tool in ArcGIS Pro 3.1.2 software. This returned 27 of 32 possible matches. Clusters 0, 1, 2, and 3 were assigned montane–dry forest, non-forest, montane–humid forest, and submontane forest respectively. The corresponding LiDAR observations and the means of their clusters were compared with the means of the inventory data that were matched to the same cluster to check for consistency in structure and proximity. Table 5 shows a comparison of GEDI-based structural parameters in comparison with the forest inventory, which helped to determine which forest type each cluster was associated with.

Table 5. Comparison of cluster labels with nearest inventory plot forest type.

Method	Cluster Labels	Matched Plots	Forest Type Majority Matched	Cluster RH100 (m)	Cluster Elevation (m a.s.l)	Cluster Count	Inventory Tree Height (m)
		Count	Mode	Mean	Mean	Count	Max
Kshape Clustering	0	4/8	Dry	19 ± 5.7	1274 ± 227	390	17 ± 3.1
	2	10/12	Humid	29 ± 8.5	1413 ± 213	204	29 ± 3.6
	3	4/7	Submontane	12 ± 4.4	1136 ± 295	296	15 ± 5.2

Several sources of error during clustering were caused by highly variable topography; in areas with gullies and steep slopes, some LiDAR footprints had uncharacteristically long waveforms, which tended to lead to a misassignment of that shot to the cluster associated with humid montane forest. Meanwhile, uncharacteristically shorter waveforms occurred over dense vegetation due to the inability of LiDAR to penetrate dense vegetation to reach the true ground. The third source of confusion during clustering occurred from LiDAR observations that covered canopy gaps or riparian areas. Footprints in canopy gaps, while technically being in forest regions, had structures similar to non-forest areas. Although this information may be useful for detecting canopy gaps, for the purposes of general forest classification mapping, these footprints acted as outliers. GEDI footprints that covered forest areas near water sources were assigned to either humid or dry forest, which may have been accurately capturing a fluctuation in vegetation structure caused by changes in growing conditions (water availability), but caused confusion when creating a generalized map forest cover map. Illustrations of the aforementioned complications can be found in Table A3 of Appendix B.

As previously mentioned, to prepare the clusters for classification mapping, it was necessary to reduce erroneous data points caused by poor ground detection and canopy gaps. The clusters were further refined by sampling optical and ancillary features—NDVI bands for dry and wet seasons, elevation, and slope—and selecting the 90 closest observations to each cluster's centroid. This approach was chosen over increasing the slope threshold or beam type to avoid disproportionately removing observations of any one cluster. It also ensured that the clusters were structurally, spectrally, and topographically similar, thereby mitigating anomalies. Table 6 shows the characteristics of the clusters post-filtering. After increasing the filtering criteria, 346 footprints remained within the region of interest to be

used for forest type mapping and validation, together with the field inventory plots. The distribution of shots after filtering can be found in Appendix B Figure A1.

Table 6. Comparison of clusters with inventory post-filtering.

Method	Cluster Labels	Matched Plots	Forest Type Majority Matched	Cluster RH100 (m)	Cluster Elevation (m a.s.l)	Cluster Count	Inventory Tree Height (m)
		Count	Mode	Mean	Mean	Count	Max
Kshape Clustering	0	4/8	Dry	15 ± 4.5	1123 ± 80	90	17 ± 3.1
	2	8/12	Humid	30 ± 10.2	1322 ± 170	90	29 ± 3.6
	3	4/7	Submontane	9 ± 2.2	940 ± 102	90	15 ± 5.2

3.2. Forest Type Classification Results

The results of each classification algorithm's highest-scoring iteration can be found in Table 7. Among the tested classification algorithms, the K-Nearest Neighbors (KNN) model with NICFI imagery performed the best, achieving the highest overall accuracy of 81.7% (Kappa = 0.78; Table 7). This result indicates the suitability of high-resolution, multi-temporal data combined with topographic features for forest type classification.

Table 7. Summary of model accuracy assessment and features.

Algorithm	NICFI Imagery and DEM		Sentinel 2 Imagery and DEM		Landsat 8 Imagery and DEM	
	Overall Accuracy	Kappa	Overall Accuracy	Kappa	Overall Accuracy	Kappa
Random Forest	80.2%	0.73	76.8%	0.68	72.9%	0.61
SVM	78.6%	0.71	74.4%	0.65	68.8%	0.55
Gradient Boost	76.3%	0.68	76.0%	0.67	73.9%	0.63
KNN	81.7%	0.75	68.0%	0.56	68.8%	0.54
Voronoi Polygons	46.8%	23.7%				

RF and SVM models achieved slightly lower accuracies of 80.2% and 78.6%, respectively, for NICFI data. The final model included a total of 16 features (selected by recursive feature elimination), including: three topographic features (elevation, slope, and aspect), four indices from the wet-season composite (MSAVI, VARI, SIPI, and R), eight indices from the dry-season composite (YI, R, NDVI, MSAVI, ARVI, B, CIG, and G), and one texture feature (NDVI_variance). The retained bands emphasized the importance of incorporating multi-season images and topography in order of separate forest classes. The final model included features that enhanced the detection of healthy vegetation (MSAVI, NDVI, and ARVI), stressed or dry vegetation (SIPI and YI), and environmental gradients (slope, aspect, and elevation) as shown in Table 8.

Table 8. Features in highest-performing models.

	NICFI Imagery and DEM	Sentinel 2 Imagery and DEM	Landsat 8 Imagery & DEM
Wet Season Spectral Bands and Indices	R MSAVI, VARI, SIPI	B11, B4, B5, B6, BSI, EVI, NDBI, NDVI, NDII	B1, B2, B3, B4, B5, B6, B7 BSI, EVI, NBR, NDBI, NDMI, NDVI, NDII
Dry season spectral bands and indices	B, R, G NDVI, MSAVI, AR-VI, CIG, YI NDVI_variance	B2, B4, B7, B6, B8A NDII, NDBI	
Topographic	slope, elevation, aspect	aspect, elevation, hill shade, slope	aspect, elevation, hill shade, slope

The performance of RF, gradient boost, and SVM, while generally robust, may have been hindered by the relatively small training size and the high dimensionality of features. Tree-based models could also have struggled to model the differences between forest types due to overfitting or less effective handling of multitemporal data. Given the random sampling of features as part of the algorithms' mechanisms. KNN relies directly on proximity in feature space, which may have allowed it to better leverage local patterns in feature space and directly compare multi-temporal feature vectors. However, the difference between RF and KNN was marginal.

The results from the Sentinel 2 and Landsat imagery models did not perform as well as those from the NICFI imagery, despite having better spectral resolution. The highest-achieving Sentinel 2 model was a gradient boost decision tree model with 20 bands (4 topographic, 9 wet-season, and 7 dry-season indices). Once again, the value of incorporating multi-date features was also reflected in the Sentinel 2 model, as it incorporates phenological differences between evergreen and deciduous forest types. The low performance of the Sentinel 2 model, in comparison with NICFI, can be attributed loss of temporal detail by creating composites. A coarser spatial resolution could have also affected the performance of the Landsat and Sentinel 2 models compared with NICFI models, despite resampling to 30 m resolution.

From the confusion matrix for the best-performing KNN model (Table 9), the class with the highest user accuracy (85.7%), producer accuracy (90.0%), and F1 score (87.8%) was non-forest, which was expected as it was the most structurally and spectrally different from the other classes. The montane–dry forest type performed lowest in terms of user accuracy (74.7%) and F1 score (78.0%). The error stemmed from a misclassification of montane–dry forest as montane–humid forest, indicating an overestimation of montane–humid forest on the map. The confusion between montane–humid and montane–dry classes was in part due to the similar spectral signatures caused by mixed species composition. For instance, while montane–humid forests are dominated by evergreen species, they also contain deciduous species, such as *Newtonia buchananii*, which reaches the upper layer of the canopy, leading to similar spectral responses during both seasons. Furthermore, these two groups had observations in close spatial proximity, which may indicate transitional areas that are more difficult to separate. Finally, the submontane forest class had the lowest producer accuracy, indicating an underestimation of submontane area in the classification. The confusion in the submontane class was mostly with the montane–dry forest class; this confusion was also likely due to the similar species compositions, as both were highly deciduous. The confusion between montane–dry and both montane–humid and submontane classes led to the lowest F1 score (78.0%) for montane–dry forest.

Table 9. Confusion matrix from the random forest classification.

Forest Type	Montane–Humid	Montane–Dry	Submontane	Non-Forest	Total	User Accuracy	F1 Score
Montane–Humid	33	5	0	1	39	84.6%	83.5%
Montane–Dry	6	32	5	0	43	74.4%	78.0%
Submontane	1	2	24	1	28	85.7%	80.0%
Non-forest	0	0	3	18	21	85.7%	87.8%
Total	40	39	32	20	Total	Overall Accuracy	Kappa
Producer's Accuracy	82.5%	82.1%	75.0%	90.0%	131	81.7%	78.5%

Reducing the resolution (Figure 4) of the map was intended to account for the well-documented ten-meter or higher geolocation accuracy. However, resampling came at a loss of detail of strips of riverine forest and canopy gaps particularly at higher elevations. According to the classified map, 39.3% of the reserve area (1172 ha) was montane–humid

forest (including agroforesty land use), 31.9% (952 ha) was montane–dry forest, 20.5% (613 ha) was submontane, and 8.1% (243 ha) was non-forest land cover.

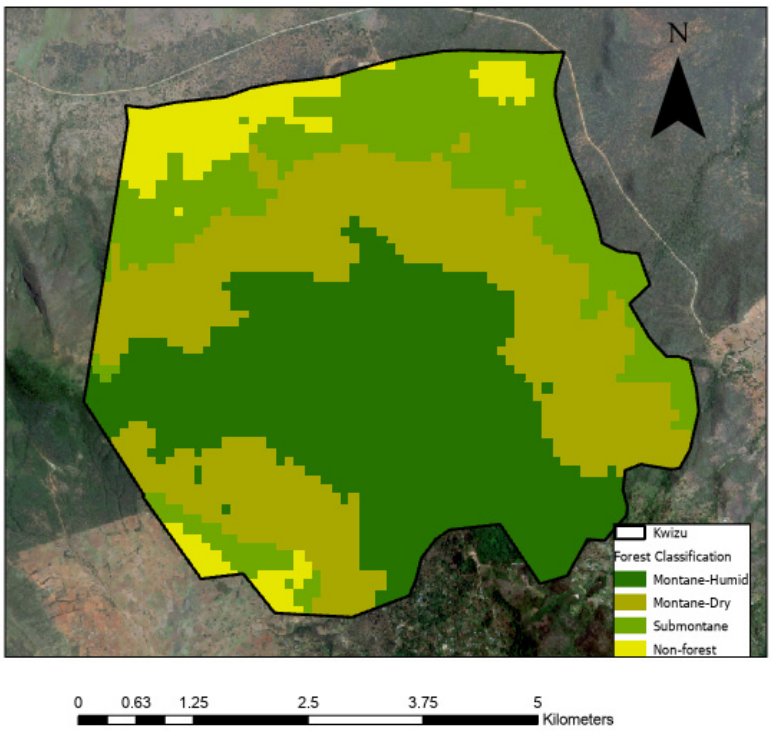


Figure 4. Classified forest map of Kwizu Forest reserve. The forest type classification map was resampled to 100 m resolution.

4. Discussion

4.1. Clustering

The results from the time-series clustering analysis, specifically using the Kshape algorithm, demonstrated its capability to identify meaningful structural groupings within the GEDI LiDAR data with minimal preprocessing. Kshape's ability to find similarities in the overall waveform shape, rather than relying on exact alignment, allowed for better differentiation of clusters. For example, in Figure 5, the three forest classes, humid (2), dry (0), and submontane (3), were most noticeably different in the width at base, which became increasingly narrow. The width at base represented the length of the significant part of the waveform (first to last mode), which was directly correlated with the relative height measurements. This corresponded well with the forest inventory data that had a decreasing trend in average height values from humid to dry to submontane forests.

Cluster 1 was associated with non-forest areas and had the highest amplitude and narrowest width, which aligns with the literature [5]. This sharp peak suggested stronger returns from the ground, characteristic of areas with little vegetation. In contrast, Cluster 3, representing submontane forests, had a lower average amplitude, attributed to signal scattering caused by vegetation cover, as well as longer wavelengths attributable to taller vegetation. The submontane and montane–dry forest types had very similar structures overall, but varied by total amplitude and slightly by the amplitude of the max peak. Lastly, another distinction between the montane–humid (2) and montane–dry (0) forest clusters was the number of peaks along the wavelengths from different canopy layers.

These findings suggest that Kshape's underlying mechanism of preserving waveform shape through cross-correlation was sensitive to changes in both density and structural variation, offering an advantage over traditional feature extraction methods when inventory data were limited. However, for this approach to be effective, it requires selecting high-quality LiDAR returns achieved through more strict terrain and sensitivity thresholds, if

the number of available footprints allows. In addition to slope and sensitivity thresholds, selecting power beams, rather than incorporating both power and coverage beams, may reduce errors related to different sensor types.

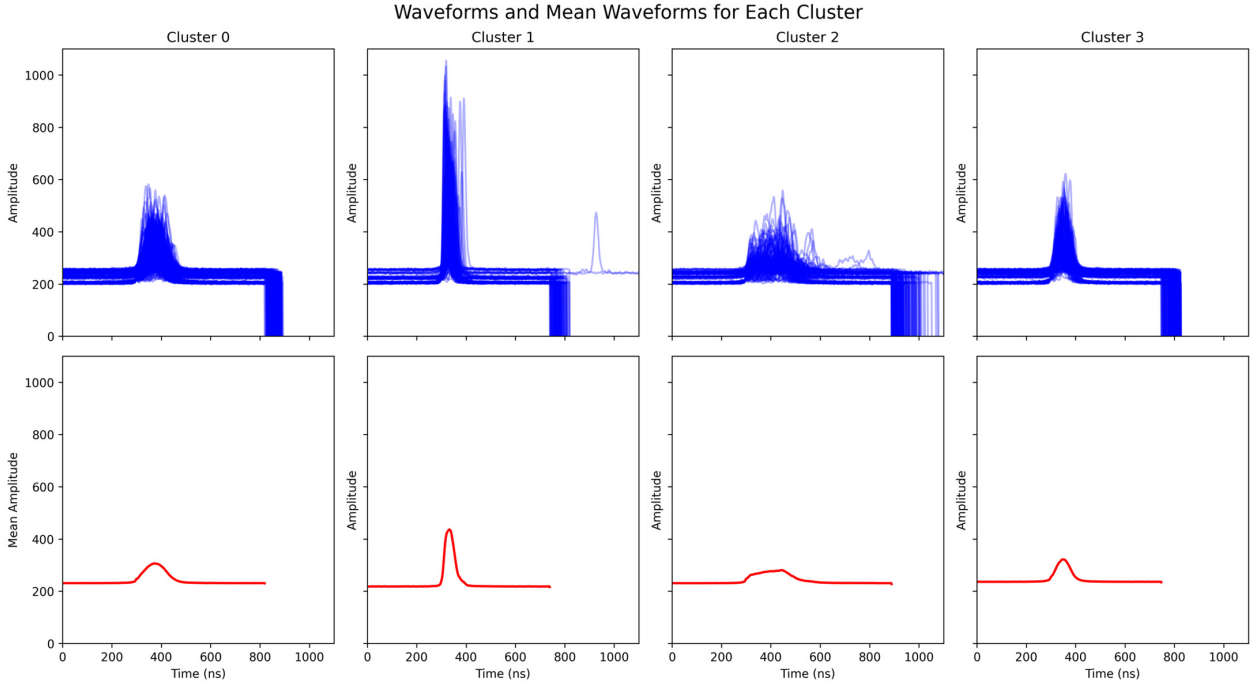


Figure 5. Individual waveforms (blue) and mean waveform (red) of each cluster. The mean waveform was the average amplitude value at each time point for all the observations in the cluster. Cluster 0 was associated with montane–dry forest, cluster 1 with non-forest, 2 with montane–humid forest, and 3 with submontane forest.

4.2. Classification

The overall accuracy of the highest-performing forest classification was 81.7%, achieved with 16 features and a KNN model. The median spectral composites used for analysis included four bands and thus only spectral indices that incorporated bands: R, G, B, and N were used with the NICFI imagery. Although this research achieved better results in terms of overall accuracy when using multitemporal images from the NICFI dataset, future researchers may benefit from using sensors, such as Landsat or Sentinel 2, that include short-wave infrared bands, which may help to spectrally dissociate each class. The lower accuracy with Sentinel 2 and Landsat datasets in this study area was attributed to the persistent haze during the image acquisition periods that made it difficult to create multidate composites free from atmospheric effects and reduced resolution of spectral bands. Compositing in itself also reduces the temporal detail. Furthermore, more advanced classification and time series algorithms may also benefit future research.

The classification results showed varying degrees of confusion between the forest types, particularly between dry and humid montane forests and dry and submontane. Part of the confusion between dry forest and the two other forest types was caused by the similarity in spectral reflectance, especially in transitional zones, where species overlapped. Spectral similarity may have been exacerbated by the ten-meter or higher geolocation error reported for GEDI's observations that could have caused cluster types to be mismatched with the raster data. For dry and submontane forest types, clusters were also structurally similar, as shown in Figure 5 (clusters 0 and 3 respectively) and, although on average significantly different, there was likely overlap or misclassification in the initial clustering step, resulting in similarities in training data that lead to confusion in classification. Furthermore, dry forest observations were occasionally near to observations of both montane and submontane forest types. The

transition from one forest type to the next is a gradual process and does not necessarily occur exactly at elevation and precipitation thresholds; therefore some confusion was to be expected. In addition, the region was topographically complex; differences in terrain and distances to water sources may have created niche growing conditions.

4.3. Class Distribution

The distribution of forest types shown in Figure 4 was consistent with literature for forests in the Eastern Arc Mountains, where higher elevations were occupied by montane–humid forest and transitioned to dry and submontane forest at lower elevations [20,21,43]. One noticeable limitation, however, was that open-canopy agroforestry areas were not distinguished from montane–humid forest. A higher number of LiDAR observations over the agroforestry area and increasing the number of clusters could potentially allow for further separation into a new class. Field visits revealed that agroforestry systems included a diverse mix of older, naturally occurring large trees, timber species (e.g., *Grevillea robusta* and *Eucalyptus* sp.), and fruit-bearing species, such as loquat, mango, and avocado. This heterogeneous composition created complex vertical structures within LiDAR footprints over tree-dominated areas. Again, a higher number of observations in this region would be needed to find significant differences between land uses.

The submontane clusters used to map the region showed a concentration near the forest edge and steep northern-facing slopes. Inventory data from the submontane forest plots in the southern part of the reserve had shorter average and maximum tree heights due to fewer observations of large trees when compared with the other forest types. LiDAR footprints associated with submontane forest (Tables 1 and 5) had lower, open canopies. However, several species found in the upper layer of the submontane plots were also found in the understory of montane–dry forest plots, indicating potential degradation of montane–dry forest, rather than a transition to a new forest in the southern portion of the reserve. The change in structure between these two forest types could be influenced by forest edge effects. A road lies directly north of the forest reserve (Figure 1), while farmland borders the south side of the forest, creating abrupt changes in land cover. As documented in several studies, these created edges cause different lighting, temperature, and moisture conditions that, in turn, lead to differences in species composition and structure [43–45]. Ewers and Banks-Leite (2013) highlight how fragmentation weakens the microclimatic buffering capacity of forest interiors, making edges more vulnerable to external climatic fluctuations [46]. In the northern portion of the reserve, edge effects may be exacerbated by the steeper slopes and higher solar radiation than those in the southern portion of the reserve, which can affect soil moisture [47]. In addition to environmental changes, anthropogenic influence may also be higher at the edge of the forest, where tree harvesting is easier to access. This can influence composition through the selection of more desirable species and influence structures by creating openings after harvest [48]. Furthermore, Alignier and Deconchat (2013) found that edge effects on vegetation vary according to factors such as edge orientation, adjacent land use, and topography, which may support the differences in the extent effects between the northern and southern edges of the reserve [49].

Detecting differences in the structure and composition of a forest area is important for monitoring ecosystem function. For example, Smith et al. (2018) demonstrated that edge effects can significantly alter carbon dynamics, depending on biome [50]. The authors found that carbon stock declined near edges, consistent with the observations found in this study. They argue that not accounting for these differences impacts the estimates of CO_2 sequestration necessary for setting emission-reduction targets. Other forest functions, such as soil erosion prevention and water and nutrient cycling, are also affected by changes in forest composition and structure that have profound impacts locally. Loss of deep-rooted species and vegetation in general has been strongly associated with higher runoff, especially in montane regions [51].

5. Conclusions

The main purposes of this study were to ascertain if unsupervised TS methods could distinguish forest types from GEDI waveforms and then to integrate these structural differences into a forest classification map. Our findings demonstrated that unsupervised TS clustering, particularly using shape-based distance metrics such as Kshape, was effective in grouping LiDAR footprints into classes that captured changes in forest structure, meeting the first objective. Kshape uses normalized cross-correlation as a similarity measure that preserves the global structure of a time series, which may have contributed to its higher performance over other TS methods [38]. An advantage of Kshape clustering was its retention of complete waveform information, which would otherwise be lost when only specific features were extracted from the waveform. This approach may better capture the slight details in structure and require less preprocessing than current waveform feature extraction methods, making it suitable for rapid assessments. Alternatively, TS clustering may be valuable for data exploration by providing insights into the natural groupings and significantly different characteristics among vegetation types. This information is valuable for refining feature selection in supervised LiDAR classification processes. However, it is important to acknowledge that there was a lower performance observed in internal validation indices, suggesting that the clusters had overlapped. In part, this accurately explains the reality on the ground. Stark boundaries from one forest type to another are unlikely and usually occur gradually due to micro-climatic changes, anthropogenic influence, and transitions in forest successional stages [52].

On the other hand, this method was still sensitive to measurement errors in clustering. Sources of error stemmed from poor ground detection in densely vegetated areas, terrain slope, geolocation accuracy, and spectral similarity among clusters, consequently, influencing the clustering and classification results. The impacts of these errors can reduce the accuracy of forest extent mapping, leading to skewed assessment of resources, functional integrity, or severity of degradation.

To mitigate these effects, the GEDI shots were filtered by incorporating terrain and spectral information. Thus, the usefulness of this approach for forest type classification may be limited without further research into improving initial clustering, prior to filtering, or a more universal and effective means of removing unreliable measurements in mountainous environments. Applying stricter slope and sensitivity thresholds and using observations from power beams, rather than coverage beams, may help to ensure quality measurements in areas in which there is an ample number of observations. This method may be more appropriate in regions with less topographic variation or larger areas with a greater sample size. Additionally, the structural clusters may be more effective in providing insight into general growth and degradation patterns (structural changes) in other study areas, rather than inferring compositional changes. For a more general forest structure map, GEDI shots in canopy gaps and gullies would not need to be as heavily filtered. Despite the errors, the structural clusters were similar to observations from forest inventory and distribution patterns reported in the literature.

Due to the noncontiguous nature of GEDI LiDAR samples, incorporating the structural details of the LiDAR shots was performed indirectly by correlating the structural clusters with spectral and topographic features. Multidate satellite imagery with a KNN classifier was the most accurate (81.7% overall accuracy) in capturing differences between forest types.

Forest type mapping provides structural and compositional differences within a forest system, which can be valuable in practical forest management policies. Targeted interventions can be designed based on the specific ecological and structural characteristics of each forest type. For example, humid montane forests, with their higher biodiversity and larger carbon storage potential, may require stricter protection measures and efforts to mitigate the encroachment of agroforestry land use. Dry montane forests might benefit from restoration initiatives to address degradation at forest edges and enhance ecosystem services. Initiatives should focus on community benefit sharing from carbon credits programs to offset the opportunity cost of conservation and incentivize the joint-forest

management programs previously established. Such tailored approaches ensure more effective conservation and sustainable management of these ecosystems.

Author Contributions: Conceptualization, M.M.M. and S.T.; methodology, M.M.M.; software, M.M.M.; validation, M.M.M.; formal analysis, M.M.M.; investigation, M.M.M.; resources, M.M.M.; data curation, M.M.M.; writing—original draft preparation, M.M.M.; writing—review and editing, S.T. and T.H.; visualization, M.M.M.; supervision, S.T. and T.H.; project administration, S.T. and T.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The majority of the data used in this study are freely and publicly available. Forest inventory data can be shared upon request.

Acknowledgments: We would like to thank the Same District Council Office of Natural Resources for their insights into the region and assistance in data collection that made this research possible.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Waveform and Spectral Features

Table A1. List of metrics derived from the GEDI L1B, L2A, and L2B data sets related to the biophysical characteristics of underlying vegetation.

Features	Description	Data
RH25, RH50, RH75, RH100	Relative heights at 25%, 50%, 75%, and 100% of the returned energy. Measures of vertical structure.	GEDI L2A
num_detectedmodes	Number of peaks within the return waveform. Measure of canopy and understory complexity.	GEDI L2A
Cover *	Total canopy cover	GEDI L2B
cover_z	Cumulative canopy cover from height z to ground (z = 0), in which z has been processed into 5 m vertical bins, where z = 1 refers to 5 m from the ground, z = 2 refers to 10 m from the ground, and so forth [31].	GEDI L2B
pavd_z	Plant area volume distribution at regular height intervals	GEDI L2B
fhd_normal *	Foliage height diversity, measure of canopy complexity related to biodiversity (calculated as the vertical foliage profile normalized by the pai)	GEDI L2B
pai	Total plant area index	GEDI L2B
Tree Cover	Landsat canopy cover * for comparison only	GEDI L2B
Max_amp	Maximum amplitude value	GEDI L2A
RHslope, RH intercept, RHResiduals	Slope and intercept of the line: $rh\_x - \sqrt{\{x\}} \mid x = [0, 100]$	Derived from GEDI L2A [5]
skewness	Waveform shape	Derived from L1B
kurtosis	Waveform shape	Derived from L1B
centroid	Waveform shape	Derived from L1B
'0–2 m', '2–5 m', '5–10 m', '10–15 m', '15–20 m', '20–25 m', '25–30 m', '>30 m'	Number of peaks at different height intervals (after thresholding amplitudes by mean + sd)	Derived from L1B [5]
Wavelength	Number of returns within the significant portion of the waveform	Derived from L1B
Wavedistance	Calculated distance of the significant portion of the waveform (Number of amplitudes * 0.15m for 1 ns time bins)	Derived from L1B [29]
fslope *	Front slope, the slope from the beginning of the waveform to the first significant peak	Derived from L1B [29]
rough *	Distance from the beginning of the waveform to the first peak	Derived from L1B [29]
TotalAmp	Sum of amplitudes	Derived from L1B
Slope	Slope of the terrain	SRTM

The features denoted with an asterisk * were removed during preprocessing due to missing values.

Table A2. Optical and terrain features tested in classification models.

Data Set	Band or Index	Formula
SRTM	Elevation, slope, hill shade, aspect	ee.Terrain.slope() ee.Terrain.hillshade() ee.Terrain.aspect()
NICFI	R, G, B, N	RED, GREEN, BLUE, NIR
Sentinel 2	B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12	BLUE, GREEN, RED, RE1, RE2, RE3, NIR, RE4, SWIR1, SWIR 2
Landsat 8	B1, B2, B3, B4, B5, B6, B7	Ultra blue, BLUE, GREEN, RED, NIR, SWIR1, SWIR2
NICFI, Landsat 8, Sentinel 2	Normalized-Difference Vegetation Index (NDVI)	$NDVI = \frac{NIR-RED}{(NIR+RED)}$
NICFI	Normalized-Difference Water Index (NDWI)	$NDWI = \frac{GREEN-NIR}{(GREEN+NIR)}$
NICFI	Structure Intensive Pigment Vegetation Index (SIPI)	$SIPI = \frac{NIR-BLUE}{(\mathrm{NIR}-\mathrm{RED})}$
NICFI, Landsat 8, Sentinel 2	Atmospherically Resistant Vegetation index (ARVI)	$ARVI = \frac{NIR-(2*RED)+BLUE}{(NIR+(2*RED)+BLUE)}$
NICFI	Modified Soil-Adjusted Vegetation Index (MSAVI)	$MSAVI2 = \frac{2*NIR+1-\sqrt{\left((2*NIR+1)^2-8*(NIR-RED)\right)}}{2}$
NICFI, Landsat 8, Sentinel 2	Visible Atmospherically Resistant Index (VARI)	$VARI = \frac{GREEN-RED}{GREEN+RED-BLUE}$
NICFI	Triangular Greenness Index (TGI)	$TGI = \frac{120*(RED-BLUE)-190*(RED-GREEN)}{2}$
NICFI	Normalized Difference Yellowness Index (NDYI)	$YI = \frac{GREEN-BLUE}{GREEN+BLUE}$
NICFI	Chlorophyll Index Green (CIG)	$CIG = \left(\frac{NIR}{GREEN}\right) - 1$
NICFI	Nonlinear Vegetation Index	$NLI = \frac{NIR^2-RED}{NIR^2+RED}$
NICFI dry (NDVI band)	Grey-scale co-occurrence matrix	Variance *, correlation, contrast, inverse difference moment
Landsat 8, Sentinel 2	Bare Soil Index (BSI)	$BSI = \frac{((\mathrm{RED}+\mathrm{SWIR})-(\mathrm{NIR}+\mathrm{BLUE}))}{((\mathrm{RED}+\mathrm{SWIR})+(\mathrm{NIR}+\mathrm{BLUE}))}$
Landsat 8, Sentinel 2	Enhanced Vegetation Index (EVI)	$EVI = \frac{2.5*(NIR-RED)}{(NIR+6*RED-7.5*BLUE+1)}$
Landsat 8, Sentinel 2	Normalized-Difference Built Index (NDBI)	$NDBI = \frac{SWIR1-NIR}{SWIR1+NIR}$
Landsat 8, Sentinel 2	Normalized-Difference Shortwave Infrared Index (NDII)	$NDII = \frac{SWIR2-NIR}{SWIR2+NIR}$
Landsat 8, Sentinel 2	Normalized Burned Ratio (NBR)	$NBR = \frac{NIR-SWIR2}{NIR+SWIR2}$
Landsat 8, Sentinel 2	Normalized Difference Moisture Index (NDMI)	$NDMI = \frac{NIR-SWIR1}{NIR+SWIR1}$
Sentinel 2	Red-Edge Normalized-Difference Vegetation Index (REND)	$REND = \frac{RE2-RE1}{RE2+RE1}$

* RE1 and RE2: Sentinel 2 B5 and B6, respectively.

Appendix B. Filtered GEDI Footprints and Terrain

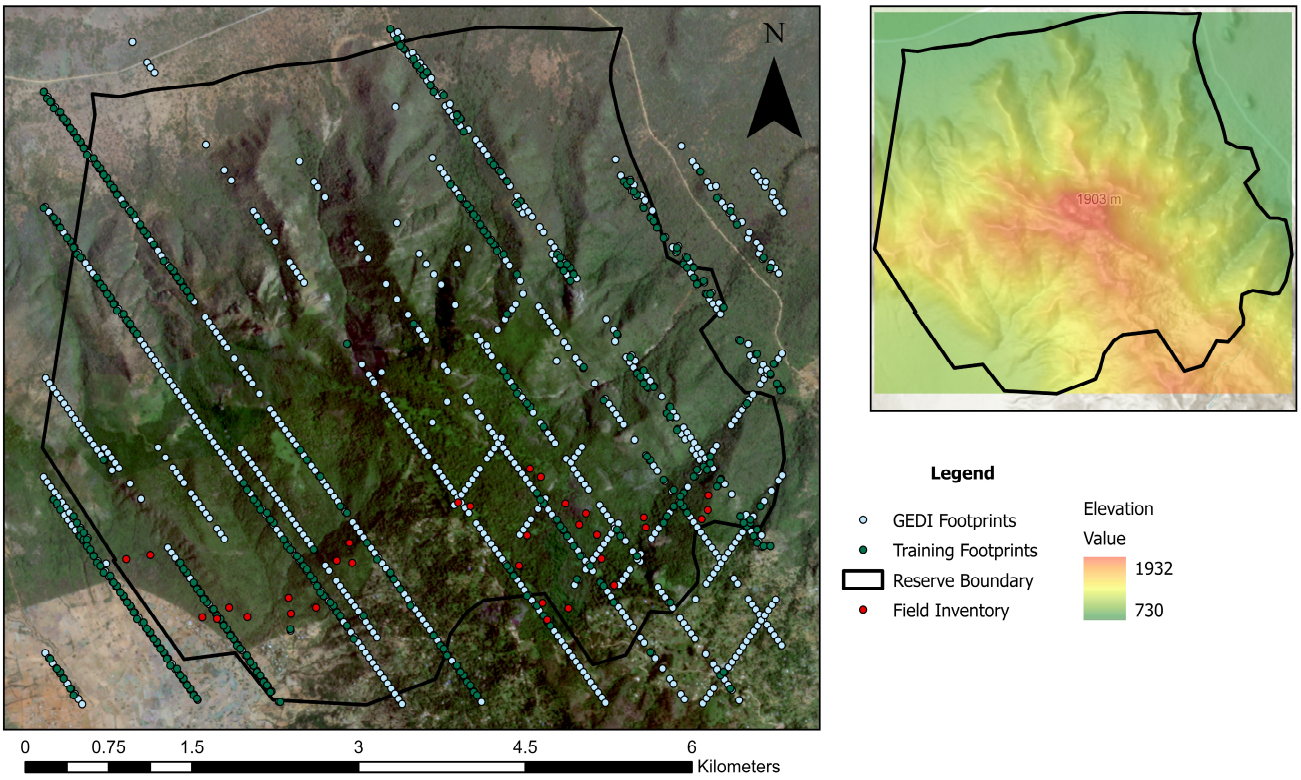


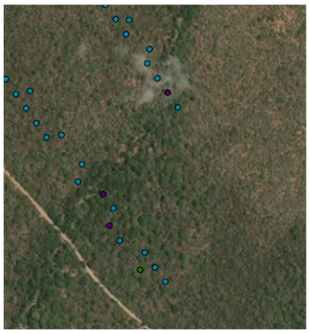
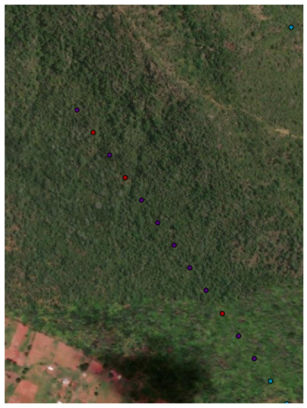
Figure A1. The locations of filtered GEDI footprints overlayed on the study area in light blue, the green points represent the footprints that were extracted for the training data. The locations of the field inventory are shown in red. The underlying image is an RGB composite image from January 2023 at 4.77 m resolution from Norway's International Climate and Forest Initiative (NICFI), available in Google Earth Engine (GEE). To the right is a map of elevation gradient overlaid on hill shade to better illustrate the topographic complexity.

The terrain in Kwizu Forest reserve is complex. The elevation changes between approximately 730 m and 1932 m.

Table A3. Sources of error and anomalies in clustering.

	Description of Error or Anomaly
	LiDAR footprints overlaying gaps in the forest canopy were classified as non-forest areas. While this demonstrates the effectiveness of time-series clustering to distinguish between forest and non-forest areas, this level of detail was confounding for creating general forest-type extent maps.

Table A3. *Cont.*

	Description of Error or Anomaly
	Changes in vegetation in areas close to water lines were often taller and denser leading to a cluster assignment of dry forest (in this example), despite the larger region being characteristic of submontane forest. Once again, this demonstrates the effectiveness of the GEDI LiDAR and time-series clustering to capture these nuances, but may be too detailed for a 100 m forest type distribution map.
	The slope in this area ranged between 27 and 30 degrees, distorting the waveform for the two red points, which had abnormally long waveforms.
	The blue circles represent submontane forest clusters which, in this case, were inaccurate as the clusters lay over dense tall forest at the at the top of the mountain in the humid forest region. These shots were removed after increasing the sensitivity threshold. The sensitivity metric is intended to indicate the sensors' ability to reach the ground; in this case, it is likely that the LiDAR signal did not reach the forest floor, shortening the waveform.

Appendix C. Multiple Comparison Analysis Results

Table A4. Multiple-comparison analysis: Kruskal–Wallis with Bonferroni correction.

Feature	H-Statistic	*p*-Value	Eta-Squared	Adjusted *p*-Value	Significant
Slope	260.82	$2.99e^{-56}$	0.261	$1.59 - e^{-54}$	TRUE
num_detectedmodes	357.28	$3.0e^{-77}$	0.358	$2.09e^{-75}$	TRUE
rh25	174.45	$1.39e^{-37}$	0.173	$7.3e^{-36}$	TRUE
rh50	342.74	$5.56e^{-74}$	0.344	$2.95e^{-72}$	TRUE
rh75	473.17	$3.1e^{-102}$	0.475	$1.65e^{-100}$	TRUE
rh100	609.60	$8.4e^{-132}$	0.613	$4.43e^{-130}$	TRUE
Peak	268.08	$8.04e^{-58}$	0.268	$4.26e^{-56}$	TRUE
RHSlope	233.71	$2.18e^{-50}$	0.233	$1.15e^{-48}$	TRUE
RHIntercept	81.76	$1.29e^{-17}$	0.080	$6.81e^{-16}$	TRUE
RHResiduals	37.45	$3.69e^{-08}$	0.035	$1.96e^{-06}$	TRUE
Skewness	111.75	$4.6e^{-24}$	0.110	$2.44e^{-22}$	TRUE

Table A4. *Cont.*

Feature	H-Statistic	*p*-Value	Eta-Squared	Adjusted *p*-Value	Significant
Kurtosis	83.97	$4.31e^{-18}$	0.082	$2.29e^{-16}$	TRUE
Energy	565.75	$2.7e^{-122}$	0.569	$1.42e^{-120}$	TRUE
Centroid	594.54	$1.5e^{-128}$	0.598	$8.16e^{-127}$	TRUE
0–2 m	11.78	$8.16e^{-3}$	0.009	$4.33e^{-01}$	FALSE
2–5 m	148.79	$4.8e^{-32}$	0.147	$2.54e^{-30}$	TRUE
5–10 m	215.68	$1.73e^{-46}$	0.215	$9.15e^{-45}$	TRUE
10–15 m	337.10	$9.26e^{-73}$	0.338	$4.91e^{-71}$	TRUE
15–20 m	363.22	$2.04e^{-78}$	0.364	$1.08e^{-76}$	TRUE
20–25 m	446.54	$1.83e^{-96}$	0.448	$9.72e^{-95}$	TRUE
25–30 m	431.26	$3.75e^{-93}$	0.433	$1.99e^{-91}$	TRUE
>30 m	303.97	$1.37e^{-65}$	0.304	$7.29e^{-64}$	TRUE
Wavelength	610.05	$6.7e^{-132}$	0.614	$3.54e^{-130}$	TRUE
Wave distance	610.05	$6.7e^{-132}$	0.614	$3.54e^{-130}$	TRUE
TotalPeaks	517.90	$6.3e^{-112}$	0.521	$3.34e^{-110}$	TRUE
TotalAmp	617.06	$2e^{-133}$	0.621	$1.07e^{-131}$	TRUE
TreeCover	338.70	$4.18e^{-73}$	0.339	$2.22e^{-71}$	TRUE
Plant Area Index	260.45	$3.59e^{-56}$	0.260	$1.90e^{-54}$	TRUE
Plant Volume 1	117.90	$2.19e^{-25}$	0.116	$1.16e^{-23}$	TRUE
Plant Volume 2	183.24	$1.76e^{-39}$	0.182	$9.34e^{-38}$	TRUE
Plant Volume 3	277.62	$6.93e^{-60}$	0.278	$3.67e^{-58}$	TRUE
Plant Volume 4	377.35	$1.78e^{-81}$	0.379	$9.44e^{-80}$	TRUE
Plant Volume 5	436.82	$2.33e^{-94}$	0.439	$1.24e^{-92}$	TRUE
Plant Volume 6	471.20	$8.3e^{-102}$	0.473	$4.41e^{-100}$	TRUE
Plant Volume 7	449.83	$3.55e^{-97}$	0.452	$1.88e^{-95}$	TRUE
Plant Volume 8	380.03	$4.69e^{-82}$	0.381	$2.49e^{-80}$	TRUE
Plant Volume 9	187.63	$1.98e^{-40}$	0.187	$1.05e^{-38}$	TRUE
Plant Volume 10	77.25	$1.19e^{-16}$	0.075	$6.33e^{-15}$	TRUE
Plant Volume 11	27.24	$5.25e^{-06}$	0.025	$2.78e^{-04}$	TRUE
Plant Volume 12	11.63	$8.77e^{-3}$	0.009	$4.65e^{-01}$	FALSE
Plant Volume 13	3.87	0.27	0.001	1.00	FALSE
Cover_z1	260.41	$3.66e^{-56}$	0.260	$1.94e^{-54}$	TRUE
Cover_z2	343.27	$4.28e^{-74}$	0.344	$2.27e^{-72}$	TRUE
Cover_z3	426.60	$3.83e^{-92}$	0.428	$2.03e^{-90}$	TRUE
Cover_z4	463.16	$4.6e^{-100}$	0.465	$2.43e^{-98}$	TRUE
Cover_z5	478.87	$1.8e^{-103}$	0.481	$9.59e^{-102}$	TRUE
Cover_z6	451.05	$1.93e^{-97}$	0.453	$1.02e^{-95}$	TRUE
Cover_z7	380.11	$4.51e^{-82}$	0.381	$2.39e^{-80}$	TRUE
Cover_z8	187.63	$1.98e^{-40}$	0.187	$1.05e^{-38}$	TRUE
Cover_z9	77.25	$1.19e^{-16}$	0.075	$6.33e^{-33}$	TRUE
Cover_z10	27.24	$5.25e^{-06}$	0.025	$2.78e^{-04}$	TRUE
Cover_z11	11.63	$8.77e^{-3}$	0.009	$4.65e^{-1}$	FALSE
Cover_z12	3.87	0.276115	0.001	1.00	FALSE

References

1. Musthafa, M.; Singh, G.; Kumar, P. Comparison of Forest Stand Height Interpolation of GEDI and ICESat-2 LiDAR Measurements over Tropical and Sub-Tropical Forests in India. *Environ. Monit. Assess.* **2022**, *195*, 71. [CrossRef] [PubMed]
2. GEDI Science Team. *GEDI L1B User Guide: For SDPS PGEVersion 3 (P003) of GEDI L1B Data*; Version 1.0; NASA Goddard Space Flight Center and University of Maryland: Greenbelt, MD, USA, 2020. Available online: https://lpdaac.usgs.gov/documents/590/GEDIL01_User_Guide_V1.pdf (accessed on 10 January 2024).
3. Yun, Z.; Zheng, G.; Monika Moskal, L.; Li, J.; Gong, P. Stratifying Forest Overstory and Understory Using the Global Ecosystem Dynamic Investigation Laser Scanning Data. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *124*, 103538. [CrossRef]
4. Hakkenberg, C.R.; Atkins, J.W.; Brodie, J.F.; Burns, P.; Cushman, S.; Jantz, P.; Kaszta, Z.; Quinn, C.A.; Rose, M.D.; Goetz, S.J. Inferring Alpha, Beta, and Gamma Plant Diversity across Biomes with GEDI Spaceborne Lidar. *Environ. Res. Ecol.* **2023**, *2*, 035005. [CrossRef]
5. Dwiputra, A.; Coops, N.C.; Schwartz, N.B. GEDI Waveform Metrics in Vegetation Mapping—A Case Study from a Heterogeneous Tropical Forest Landscape. *Environ. Res. Lett.* **2023**, *18*, 015007. [CrossRef]

6. Moran, C.J.; Rowell, E.M.; Seielstad, C.A. A Data-Driven Framework to Identify and Compare Forest Structure Classes Using LiDAR. *Remote. Sens. Environ.* **2018**, *211*, 154–166. [CrossRef]
7. Alexander, C.; Korstjens, A.H.; Usher, G.; Nowak, M.G.; Fredriksson, G.; Hill, R.A. LiDAR Patch Metrics for Object-Based Clustering of Forest Types in a Tropical Rainforest. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *73*, 253–261. [CrossRef]
8. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006; ISBN 978-0-387-31073-2.
9. Zhao, J.; Itti, L. ShapeDTW: Shape Dynamic Time Warping. *Pattern Recognit.* **2018**, *74*, 171–184. [CrossRef]
10. Aghabozorgi, S.; Seyed Shirkhorshidi, A.; Ying Wah, T. Time-Series Clustering—A Decade Review. *Inf. Syst.* **2015**, *53*, 16–38. [CrossRef]
11. Hofton, M.; Blair, J.B.; Story, S.; Yi, D. *Algorithm Theoretical Basis Document (ATBD) for GEDI Transmit and Receive Waveform Processing for L1 and L2 Products*; Version 1.0; Goddard Space Flight Center: Greenbelt, MD, USA, 2019.
12. Vaughn, N.R.; Moskal, L.M.; Turnblom, E.C. Fourier transformation of waveform lidar for species recognition. *Remote Sens. Lett.* **2011**, *2*, 347–356. [CrossRef]
13. Paparrizos, J.; Gravano, L. K-Shape: Efficient and Accurate Clustering of Time Series. In Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Melbourne, Australia, 31 May—4 June 2015. [CrossRef]
14. Doggart, N.; Leonard, C.; Perkin, A.; Menegon, M.; Rovero, F. *The Vertebrate Biodiversity and Forest Condition of the North Pare Mountains*; TFCG Technical Paper No. 17; Tanzania Forest Conservation Group: Dar es Salaam, Tanzania, 2008.
15. Finch, J.; Marchant, R.; Courtney Mustaphi, C.J. Ecosystem Change in the South Pare Mountain Bloc, Eastern Arc Mountains of Tanzania. *Holocene* **2017**, *27*, 796–810. [CrossRef]
16. Baker, M. A Survey of the Avifauna of Chome Forest Reserve, South Pare Mountains, North-East Tanzania. East African Cross-Border Biodiversity Project. 2001. Available online: https://www.easternarc.or.tz/wp-content/uploads/2020/03/Chome-avifauna-survey_Baker-2001.pdf (accessed on 19 March 2023).
17. Kajembe, G.C.; Nduwamungu, J.; Luoga, E.J. The Impact of Commnity-Based Forest Management and Joint Forest Managent on the Forest Resource Base and Local People's Livelihoods: Case Studies from Tanzania. *Commons South. Afr. CASS/PLAAS* **2008**, *8*, 1–17.
18. Explore the World's Protected Areas. Available online: https://www.protectedplanet.net/en/thematic-areas/wdpa?tab=WDPA (accessed on 10 August 2024).
19. Lovett, J.C.; Ruffo, C.K.; Gereau, R.E.; Sørensen, L.; Lovett, J.; Taplin, J. Field Guide to the Moist Forest Trees of Tanzania; Published online: ResearchGate. 2014. Available online: https://www.researchgate.net/publication/265247417 (accessed on 10 January 2024).
20. Lolila, N.J.; Shirima, D.D.; Mauya, E.W. Tree Species Composition along Environmental and Disturbance Gradients in Tropical Sub-Montane Forests, Tanzania. *PLoS ONE* **2023**, *18*, e0282528. [CrossRef] [PubMed]
21. Tesha, D.L.; Mauya, E.W.; Madundo, S.D.; Emily, C.J. Role of Topography, Soil and Climate on Forest Species Composition and Diversity in the West Usambara Montane Forests of Tanzania. *Folia Oecologica* **2023**, *50*, 105–118. [CrossRef]
22. Tolangara, A.; Ahmad, H.; Liline, S. The Composition and Important Value Index of Trees for Wildlife Feed in Bacan Island, South Halmahera. *IOP Conf. Ser. Earth Environ. Sci.* **2019**, *276*, 012037. [CrossRef]
23. GEDI L2B Product Data Dictionary. Available online: https://lpdaac.usgs.gov/documents/980/gedi_l2b_dictionary_P003_v2.html (accessed on 28 October 2023).
24. Earthdata Search | Earthdata Search. Available online: https://search.earthdata.nasa.gov/search (accessed on 10 August 2024).
25. GitHub—Nasa/GEDI-Data-Resources: This Repository Provides Guides, Short How-Tos, and Tutorials to Help Users Access and Work with Data from the Global Ecosystem Dynamics Investigation (GEDI) Mission. Available online: https://github.com/nasa/GEDI-Data-Resources (accessed on 17 January 2024).
26. Schneider, F.D.; Ferraz, A.; Hancock, S.; Duncanson, L.I.; Dubayah, R.O.; Pavlick, R.P.; Schimel, D.S. Towards Mapping the Diversity of Canopy Structure from Space with GEDI. *Environ. Res. Lett.* **2020**, *15*, 115006. [CrossRef]
27. Kutchartt, E.; Pedron, M.; Pirotti, F. Assessment of Canopy and Ground Height Accuracy From GEDI LiDAR Over Steep Mountain Areas. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2022**, *V-3-2022*, 431–438. [CrossRef]
28. Adam, M.; Urbazaev, M.; Dubois, C.; Schmullius, C. Remote Sensing Accuracy Assessment of GEDI Terrain Elevation and Canopy Height Estimates in European Temperate Forests: Influence of Environmental and Acquisition Parameters. *Remote Sens.* **2020**, *12*, 3948. [CrossRef]
29. Zhou, T.; Popescu, S. Waveformlidar: An R Package for Waveform LiDAR Processing and Analysis. *Remote Sens.* **2019**, *11*, 2552. [CrossRef]
30. Farr, T.G.; Rosen, P.A.; Caro, E.; Crippen, R.; Duren, R.; Hensley, S.; Kobrick, M.; Paller, M.; Rodriguez, E.; Roth, L.; et al. The Shuttle Radar Topography Mission. *Rev. Geophys.* **2007**, *45*, RG2004. [CrossRef]
31. KSAT; Planet; Airbus. NICFI User Guide v2. NICFI Data Program. June 2021. Available online: https://www.planet.com/nicfi/ (accessed on 7 August 2024).
32. European Space Agency (ESA); Copernicus. *Harmonized Sentinel-2 MSI: MultiSpectral Instrument, Surface Reflectance (HARMONIZED)*; European Union: Brussels, Belgium; ESA: Paris, France; Copernicus: Göttingen, Germany, 2024.
33. U.S. Geological Survey. *Landsat 8 Collection 2 Level-1 Data*; United States Geological Survey: Sioux Falls, SD, USA, 2021. Available online: https://www.usgs.gov/landsat-missions/landsat-data-access (accessed on 12 December 2024).

34. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [CrossRef]
35. Tavenard, R.; Faouzi, J.; Vandewiele, G.; Divo, F.; Androz, G.; Holtz, C.; Payne, M.; Yurchak, R.; Rußwurm, M. Tslearn, A Machine Learning Toolkit for Time Series Data. *J. Mach. Learn. Res.* **2020**, *21*, 1–6.
36. Pedregosa Fabianpedregosa, F.; Michel, V.; Grisel Oliviergrisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Vanderplas, J.; Cournapeau, D.; Pedregosa, F.; Varoquaux, G.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830. [CrossRef]
37. Lima, F.T.; Souza, V.M.A. A Large Comparison of Normalization Methods on Time Series. *Big Data Res.* **2023**, *34*, 100407. [CrossRef]
38. Li, Y.; Shen, D.; Nie, T.; Kou, Y. A New Shape-Based Clustering Algorithm for Time Series. *Inf. Sci.* **2022**, *609*, 411–428. [CrossRef]
39. Rousseeuw, P.J. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. [CrossRef]
40. Davies, D.L.; Bouldin, D.W. A Cluster Separation Measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **1979**, *PAMI-1*, 224–227. [CrossRef]
41. Caliński, T.; Harabasz, J. A Dendrite Method Foe Cluster Analysis. *Commun. Stat.* **1974**, *3*, 1–27. [CrossRef]
42. Lee, S.; Lee, D.K. What Is the Proper Way to Apply the Multiple Comparison Test? *Korean J. Anesthesiol.* **2018**, *71*, 353. [CrossRef]
43. Addo-Fordjour, P.; Abrokwah, J.; Arko, A.P.; Pappoe, N.A.; Yawson, D.; Yeboah, N.; RobertTsiquaye, A.; Puobe, R.N.; Anane-Frimpong, K.; Fosu, S.A.; et al. Effects of linear edges on tree communities and soil properties in a moist semi-deciduous forest in Ghana. *Plant Ecol.* **2024**, *225*, 345–359. [CrossRef]
44. Bauer, L.; Huth, A.; Bogdanowski, A.; Müller, M.; Fischer, R. Edge Effects in Amazon Forests: Integrating Remote Sensing and Modelling to Assess Changes in Biomass and Productivity. *Remote Sens.* **2024**, *16*, 501. [CrossRef]
45. Nunes, M.H.; Vaz, M.C.; Camargo, J.L.C.; Laurance, W.F.; Andrade, A.; Vicentini, A.; Laurance, S.; Raumonen, P.; Jackson, T.; Zuquim, G.; et al. Edge Effects on Tree Architecture Exacerbate Biomass Loss of Fragmented Amazonian Forests. *Nat. Commun.* **2023**, *14*, 44004. [CrossRef] [PubMed]
46. Ewers, R.M.; Banks-Leite, C. Fragmentation impairs the microclimate buffering effect of tropical forests. *PLoS ONE* **2013**, *8*, e58093. [CrossRef]
47. Gallardo-Cruz, J.A.; Pérez-García, E.A.; Meave, J.A. β-Diversity and vegetation structure as influenced by slope aspect and altitude in a seasonally dry tropical landscape. *Landsc. Ecol.* **2009**, *24*, 473–482. [CrossRef]
48. Hall, J.; Burgess, N.D.; Lovett, J.; Mbilinyi, B.; Gereau, R.E. Conservation Implications of Deforestation across an Elevational Gradient in the Eastern Arc Mountains, Tanzania. *Biol. Conserv.* **2009**, *142*, 2510–2521. [CrossRef]
49. Alignier, A.; Deconchat, M. Patterns of forest vegetation responses to edge effect as revealed by a continuous approach. *Ann. For. Sci.* **2013**, *70*, 601–609. [CrossRef]
50. Smith, I.A.; Hutyra, L.R.; Reinmann, A.B.; Marrs, J.K.; Thompson, J.R. Piecing together the fragments: Elucidating edge effects on forest carbon dynamics. *Front. Ecol. Environ.* **2018**, *16*, 213–221. [CrossRef]
51. Mmbando, G. Hydrological Sensitivity of the Mkomazi River Basin (Tanzania) to Climate Change. Ph.D. Thesis, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany, 2015.
52. Gu, Z.; Cao, S.; Sanchez-Azofeifa, G.A. Using LiDAR Waveform Metrics to Describe and Identify Successional Stages of Tropical Dry Forests. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *73*, 482–492. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.