Abstract: In recent years, with the growing environmental concern regarding climate change, there has been a search for efficient alternatives in indirect methods for the quantification of biomass and forest carbon stock. In this article, we seek to obtain pioneering results of biomass and carbon estimates from forest inventory data and LiDAR technology in a dry tropical forest in Brazil. We use forest inventory data in two areas together with data from the LiDAR flyby, generating estimates of local biomass and carbon levels obtained from local species. We approach three types of models for data analysis: Multiple linear regression with principal components (PCA), conventional multiple linear regression and stepwise multiple linear regression. The best fit total above ground biomass (TAGB) and total above ground carbon (TAGC) model was the stepwise multiple linear regression, concluding, then, that LiDAR data can be used to estimate biomass and total carbon in dry tropical forest, proven by an adjustment considered in the models employed, with a significant correlation between the LiDAR metrics. Our finding provides important information about the spatial distribution of TAGB and TAGC in the study area, which can be used to manage the reserve for optimal carbon sequestration.

Keywords: Caatinga vegetation; aboveground biomass; carbon stocks; allometry; statistical models

1. Introduction

The increase in the carbon dioxide concentration (CO₂) in the atmosphere in recent decades and its consequences for the environment have been attracting the attention of society and are being addressed as a matter of global concern [1]. The high CO₂ concentration in the atmosphere is worrying, as it generates an increase in the greenhouse effect, and consequently causes global warming [2–4]. In this scenario, dry tropical forest areas play an important role, and the Caatinga vegetation in northeastern Brazil significantly contribute to the global carbon cycle through aboveground biomass and carbon stock [5,6].

There is currently a rich ongoing discussion among scientists around the world about the main tools and methods for generating measures to mitigate climate change. The first guiding question to be answered is “how do we measure the impacts of global climate change and how can we slow its progress?” The second key question is “what tools and methods should be used to ensure a reliable estimate?” In many cases, the variables best
known for generating rapid responses for dry tropical forest ecosystems are aboveground biomass and carbon.

However, the destructive sampling of trees is a limiting factor for calibrating statistical models, mainly due to the high cost involved in field work [7]. These methods are currently based on forest inventory data using carbon factors and equations which transform the biometric parameters of the forest, such as diameter at breast height (DBH) and total height (Ht) of individuals, in estimating the carbon stock contained in aboveground forest biomass [8].

Remote sensing techniques combined with optical sensors have recently been presented as a viable alternative for estimating the biomass and carbon stock in planted and natural forests [9,10]. Among the current remote sensing techniques, laser tillering, also known as light detection and ranging (LiDAR), has prominently emerged in the forest scenario [11], by which biomass and carbon stock estimates can be systematically and efficiently obtained in the field [7].

LiDAR is an active system, and its principle consists of emitting a laser pulse which interacts with an object on the Earth’s surface and subsequently returns to the sensor in a given time interval. The technology makes it possible to accurately reproduce digital terrain models (DTMs, models which enable describing the elevation of land free of objects), digital surface models (DSMs, models which enable describing the elevation of the terrain including the objects present), and digital height models (DHMs, models which describe the height of all objects, with the cloud points referring to the ground normalized to zero).

Although LiDAR technology is an efficient alternative and widely applied in forest inventory in countries like the United States, Finland, and Sweden, there are still obstacles to its use in other countries such as Brazil. The limitations are not due to the functioning of the technology itself, but mainly because it is still an emerging technology in Brazil [9]. The elaboration of processing methodologies aimed at Brazilian needs is still recent, and in this respect the execution of this work is mainly justified by the search for a scientific technical advance which provides developing routines which can assist in LiDAR data acquisition and processing, and in turn seeks to efficiently attain aboveground biomass and carbon stock estimates in a Brazilian dry tropical forest.

In particular, LiDAR technology has the ability to directly measure the vegetation attributes (metrics) on a vertical scale with high precision, and therefore a system can be developed to sample the biomass and carbon stock of the trees in situ in environmental gradients, providing a potential solution to outstanding problems related to forest biomass and aboveground carbon stock [12,13]. Biomass and carbon estimates at local and regional levels, as well as the spatialization of these variables using maps, can provide an overview of biodiversity and forest structure [14–16]. This information is extremely important for the Caatinga vegetation domain in Pernambuco for possible payments for environmental services and other projects aimed at reducing emissions from deforestation and forest degradation (REDD+).

Due to the great importance of Caatinga forest resources, quantifying and mapping biomass and carbon stock using LiDAR metrics are key factors to meet the legal aspects concerning sustainable management, mainly reconciling sustainable wood production and stock maintenance of carbon in the area. This task is one of the main long-term planning tools, because in addition to dimensioning the forest’s stock and productivity, it generates information which will direct ecosystem maintenance through conservation and/or preservation [17].

In this sense, this work was developed with the intention of generating information on biomass and carbon stock using LiDAR metrics in different dry tropical forest areas in the municipality of Floresta, Pernambuco. Specifically, it is intended to: (a) Estimate the total biomass and carbon stock for the parcels inventoried in two areas using a local allometric model; (b) use the aerial LiDAR system to generate attributes in plots in the different dry forest areas inventoried in Pernambuco; (c) develop an allometric model for estimating biomass and carbon stock using LiDAR metrics for the different dry forest areas
inventoried in Pernambuco; (d) generate biomass and carbon stock maps for the different dry forest areas inventoried in Pernambuco.

2. Materials and Methods
2.1. Study Area

The work was carried out in two semiarid areas of Itapemirim Farm. Its extension is approximately 60 km$^2$ (Figure 1D), located in the municipality of Floresta in the São Francisco mesoregion in Pernambuco, northeast Brazil (8°30′37″ S and 37°59′07″ W).

![Figure 1](image-url)

Figure 1. Coverage of the study area: (A, B, D), and profile photo in Floresta (C), in the hinterland of Pernambuco, Brazil. Source: [18].

The two areas make up different structures due to the history of use. The first area is called “Transposição” (considered preserved with 55 years of small anthropogenic disturbances) and had 40 permanent plots systematically arranged in a square (20 × 20 m). The second area is called “Correntão” and followed the same protocol as the previous area. However, this area has been in regeneration for 29 years and is considered to have a high degree of disturbance due to the history of land use with eucalyptus plantations (Figure 2).

The vegetation in these areas is predominantly Caatinga (dry tropical forest), meaning savanna–steppe characterized by shrub–tree vegetation along with the presence of cacti and herbaceous strata [19]. The climate is BSh according to the Köppen classification, a hot semiarid region with an average annual precipitation of approximately 400 to 500 mm, with a rainy period from January to April, and an average annual temperature of 26.1 °C. Its distributions of temperature and precipitation throughout the year studied (2014) in the municipalities belonging to the study area are shown in Figure 3. The municipality has an area of 3643.97 Km$^2$ and an average altitude of 323 m [20].
In this study area, at least five main species occur, such as catingueira (*Poincianella bracteosa* (Tul.) LP Queiroz), jurema-de-embira (*Mimosa ophthalmocentra* Mart. Ex Benth.), pereiro (*Aspidosperma pyrifolium* Mart.), aroeira (*Myracrodium urundeuva* (Engl.) Fr. All.) and mororó (*Bauhinia cheilanta* (Bong). Steud.).
Figure 3. Distribution of air temperature and precipitation over the year 2014 in the study area by the nearest weather station. Source: Agritempo adjusted. Source: [21].

2.2. Estimation of the Biomass/Carbon Stock in the Field

Plots, which had already been inventoried, were used to facilitate logistics and data collection in each area (Figure 2). A total of 40 plots in each area have been monitored since 2008. They are systematically distributed and each is 20 × 20 m (400 m²). They are 80 m apart, 50 m from the edge. All arboreal individuals with their diameter measured at 1.30 m above the ground (DBH) ≥ 6.0 cm were identified, labeled and measured, and the total heights (Ht) were measured by a clinometer.

Local biomass estimates in each plot in the different areas were generated from a previously developed local equation [18], with subsequent conversion to carbon stock (Mg·ha⁻¹). Thus, we were able to perform local estimates of the biomass logarithm using DBH and Ht logarithms as independent variables, as shown in the equation below:

\[
\text{TAGB} = \exp(-3.5336 + 1.9126 \times \ln(\text{DBH}) + 1.2438 \times \ln(\text{Ht}))
\]

where: DBH is the diameter of the tree at breast height (1.30 m) in cm; Ht is the total height of the tree (m). This equation was developed for the site and reported an Akaike information criterion (AIC) value of 573.77; an adjusted determination coefficient (R²Adj) of 0.90; root mean square error (RMSE) of 18.2%; and bias of 0.20%.

Next, the estimated biomass was converted using the average carbon fraction (CF ≈ 48%) of the Caatinga woody species [22] for carbon stock estimates (Mg·ha⁻¹). In addition, it has traditionally been assumed that the carbon content of a tree’s dry biomass is 50% for estimating carbon stocks for sites [23–25], but it should be emphasized that the carbon fraction of the wood may exhibit some small variations between species [26]. Thus, the carbon stock is assumed as follows:

\[
\text{TAGC} = \text{TAGB} \times \text{CF}
\]

where: TAGC is the estimated total aboveground carbon stock (Mg·ha⁻¹); TAGB is the total estimated aboveground biomass (Mg·ha⁻¹); CF is the carbon fraction (48%).

In summary, the forest inventory data analyzed for this study as well as the biomass and carbon predictions were for 2014 and are summarized in Table 1. The choice of this measurement period was defined according to the same LiDAR flyover year in the areas.
Table 1. Descriptive values (mean and standard deviation) and number of individuals and total scheme 2014. Please define abbreviations in the table. DBH (cm) is diameter breast height (1.30 m above the ground); Ht (m) is the total height of the tree or shrub; AGB (Mg ha\(^{-1}\)) and AGC (Mg ha\(^{-1}\)) are above-ground biomass and above-ground carbon respectively.

| Areas      | DBH (cm)   | Ht (m)     | AGB (Mg ha\(^{-1}\)) | AGC (Mg ha\(^{-1}\)) | N° PITS | N° Ind. | N° Stem |
|------------|------------|------------|-----------------------|-----------------------|---------|---------|---------|
| Transposição | 1.91       | 29.92      | 3.85                  | 2.37                  | 1.3     | 9.0     | 1.3     |
| Correntão   | 2.06       | 52.20      | 11.44                 | 5.4                   | 1.7     | 7.5     | 1.7     |

2.3. Estimation of the Biomass/Carbon Stock by LiDAR Data

The LiDAR data used in this study were made available by the Pernambuco Three-Dimensional Program (PE3D) as part of the Pernambuco Water Sustainability Program (PSHPE) with the objective of mapping the entire territory of the state of Pernambuco using its services covering aerophotogrammetric and laser profiling. They were collected in August 2014 using a Leica ALS50 system coupled to a BEM-810 C-Seneca II-Prefix PT-RQA aircraft, and their characteristics are shown below (Table 2).

Table 2. Details of acquisition of LiDAR data.

| Attribute                  | Values       |
|----------------------------|--------------|
| LiDAR system               | ALS-50 Leica|
| Flight altitude (m)        | 3.068        |
| Data acquisition           | 10 August 2014|
| Opening angle (°)          | 34.5         |
| Scanner frequency (Kz; kHz)| 36.8 kHz     |
| Pulse density (pulses·m\(^{-2}\)) | 0.5          |
| Datum                      | Sirgas 2000  |

After acquiring the point clouds of the areas to be analyzed on the Pernambuco Tridimensional Program website (http://www.pe3d.pe.gov.br/mapa.php, accessed on 5 August 2016), it was necessary to change the format from “.xyzi” to “.las” using the LAS Utility software. Thus, a descriptive report with several important characteristics of the LiDAR data set was produced with the Fusion3.8 software program using the “Catalog” tool.

Therefore, the returns which were on the soil surface (points on the soil surface) were filtered from the LiDAR point cloud using the “Ground Filter” tool. The next step for both areas was to obtain the digital terrain model (DTM) using the “Grid Surface Create” tool and the digital surface model (DSM) using the “Canopy Model” tool.

It was first necessary to normalize the data in order to carry out the subsequent analyses with the metrics of only the trees. This task was performed by subtracting the DSM data from the DTM data using the “Clipdata” tool. The next step was to obtain the canopy height digital model (CHM) with the aid of the “Canopy Model” tool to obtain the metric value for each plot of the two areas, for which it was necessary to perform clipping by plot using the shape file of the plots with the point cloud using the “Polyclip DATA” tool.

The LiDAR metrics calculate a series of estimates of descriptive statistical parameters of the LiDAR point cloud and in this study were generated from the “Cloud Metrics” tool. Thus, a total of 26 metrics were generated at the end of the data processing by the point clouds in each sample unit in the different areas. These metrics are the most used in biomass and carbon estimation studies, categorized according to their origin and calculated symbology (Table 3).
Table 3. List of LiDAR metrics evaluated in the study, obtained from the “Cloud Metrics” tool of Fusion v. 3.8.

| Category                                      | LiDAR Metrics            | Symbology       |
|-----------------------------------------------|--------------------------|-----------------|
| Height                                         |                          |                 |
| Maximum height                                | Elev.maximum             |                 |
| Minimum height                                | Elev.minimum             |                 |
| Mean height                                   | Elev.mean                |                 |
| Modal height                                  | Elev.mode                |                 |
| Standard deviation of heights                 | Elev.stdev               |                 |
| Height variation coefficient                  | Elev.CV                  |                 |
| Height asymmetry                              | Elev.skewness            |                 |
| Kurtosis of height                            | Elev.kurtosis            |                 |
| Median of absolute deviations from the general mean | Elev.MAD.median          |                 |
| 1st percentile of height                      | Elev.P01                 |                 |
| 5th percentile of height                      | Elev.P05                 |                 |
| 10th percentile of height                     | Elev.P10                 |                 |
| 20th percentile of height                     | Elev.P20                 |                 |
| 25th percentile of height                     | Elev.P25                 |                 |
| 30th percentile of height                     | Elev.P30                 |                 |
| 40th percentile of height                     | Elev.P40                 |                 |
| 50th percentile of height                     | Elev.P50                 |                 |
| 60th percentile of height                     | Elev.P60                 |                 |
| 70th percentile of height                     | Elev.P70                 |                 |
| 75th percentile of height                     | Elev.P75                 |                 |
| 80th percentile of height                     | Elev.P80                 |                 |
| 90th percentile of height                     | Elev.P90                 |                 |
| 95th percentile of height                     | Elev.P95                 |                 |
| 99th percentile of height                     | Elev.P99                 |                 |
| Canopy density                                | Canopy relief ratio \(^1\) | Canopy.relief.ratio |
| Percentage of all returns above 1.30          | Percentage.all.returns.above.1.30 |

\(^1\) Canopy relief ratio = \((H_{mean} - H_{min})/(H_{max} - H_{min})\).

Some of the main metrics used in predicting biomass and carbon are described below:

Elev.maximum = maximum height: This is the highest value found in the measurement range in meters within each sample unit, considering variations at each meter along the walking axis.

Elev.mean = mean height: This is the mean value of the highest points, considering variations every meter in the measurement range in meters within each sample unit (Equation (3)).

\[
Elev.mean = \frac{1}{n} \times \sum_{i=1}^{n} h_i \quad (3)
\]

Elev.stddev = Standard deviation of height in the LiDAR point cloud:

\[
Elev.stddev = \sqrt{\frac{1}{n-1} \times \sum_{i=1}^{n} (h_i - h_{med})^2} \quad (4)
\]

where: hmean = mean height of the point cloud.

Elev.CV = height variation coefficient in the LiDAR point cloud:

\[
h_{cv} = \frac{h_{desv}}{h_{med}} \quad (5)
\]
Height percentiles in the LiDAR point cloud \((h_{pi})\): The \(i\)-th percentile of \(n\) points traditionally represented in the LiDAR point cloud, ordered in height values corresponding to the value which occupies the \(K\) position of the data set, as in the following equation:

\[
K = \frac{h_{pi}(n + 1)}{100} \tag{6}
\]

where: \(K = \text{value that occupies the } i\text{-th percentile in height in the point cloud; } h_{pi} = \text{i-th percentile in height in the point cloud.}\)

2.4. Modeling the Biomass/Carbon Stock Using LiDAR Data

The “R Project for Statistical Computing” Lidar Data_Analysis \cite{27} and ArcGIS 10® software programs were used to construct, validate and apply predictive models and generate representative biomass and forest carbon maps in the different areas.

The LidarData_Analysis Tools software program was developed by the USDA Forest Service—Remote Sensing Applications Center, written in the Python language, and works as an interface of R. Lidar Data_Analysis Tools was designed to streamline the statistical regression analysis process involving LiDAR metrics generated by the North American Forest Service, and in fact works as a graphical interface to access the statistical modeling packages available in R which simplify processing large volumes of data \cite{9}.

Next, three data analysis approaches were used to construct the biomass allometric models per hectare according to the LiDAR metrics for the two areas: Multiple linear regression, stepwise multiple linear regression and multiple linear regression with principal components—PCA.

First, traditional modeling was used employing multiple linear regression. Thus, it is assumed that there is a linear relationship between a \(Y\) variable (biomass; carbon) and \(k\) independent variables, \(x_j (j = , \ldots , k = \text{LiDAR point cloud metrics})\). The mathematical model which expresses the equation of multiple linear regression has the following form:

\[
Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \ldots + \beta_kX_k + \epsilon \tag{7}
\]

where: \(Y = \text{TAGB (Mg·ha}^{-1})\); \(\beta_0 = \text{intercept on the Y axis; } \beta_i = \text{slope of the } i\text{-th explanatory variable; } k = \text{number of explanatory variables; } \epsilon = \text{random error.}\)

Regression analysis was used with an emphasis on solving most of the forest problems, especially when it is intended to obtain estimates of forest parameters through biometric relationships. A careful analysis was performed in selecting the best metrics from the LiDAR point cloud candidates for modeling among the metrics generated by the LiDAR data processing for the construction the models. According to \cite{9}, Pearson’s linear correlation test \((r^2)\) should first be applied for this selection to obtain the correlation between the predictive variables and to evaluate the possible existence of collinearity between them. Variables with \(R^2 > 0.9\) were excluded from the analysis to avoid the presence of collinearity.

Second, the stepwise technique was applied using the regsubsets function of the “leaps” package in R to obtain subsets of independent variables which are candidates for composing the definitive biomass and carbon models in the different areas. This method performs an exhaustive search to select the best combinations of independent variables by minimizing Akaike’s information criterion (AIC), and rearranging them into subsets which may later give rise to the selected models. Regsubsets require the use of a maximum number (represented by the nvmax argument) of independent variables for the growing construction of these subsets.
Third, the principal component technique (PCA) was applied to the selected LiDAR metrics, and the metrics most likely to contribute to developing the model were identified by inspecting the eigenvectors on each principal component. Then, the metrics with the highest load on the PCs were used as input variables in multivariate linear regression models that predicted biomass per hectare.

An example of using PCA including the equations used to obtain the eigenvalues, eigenvectors and principal components (PC) can be found in [27]. PCA was applied in the present study to the selected LiDAR metrics using the prcomp function of the statistics package in R. A correlation matrix derived from the LiDAR metrics provided the basis for calculating eigenvalues and eigenvectors and for the subsequent determination of PC scores. Each score represented a transformed metric from the linear combination of LiDAR metrics. Differences in the contribution of each LiDAR metric to the variability in the data set, as well as the similarity in the calculated metrics [9], can be established by analyzing the eigenvectors and the PC score.

2.5. Evaluation of Models

The model parameters were estimated using the ordinary least squares (OLS) method in all the modeling methods described. The parameters were generally calculated using all plots sampled in each area and will be assumed to be the true parameters that represent the biomass and carbon stock at each location.

For each of the criteria established for the biomass estimation, the obtained equations were analyzed using comparisons of statistical criteria obtained according to the following equations:

$$R^2_{aj} = R^2 - \left[ \frac{k - 1}{n - k} \right] \times \left( 1 - R^2 \right)$$  \hspace{1cm} (8)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n}}$$  \hspace{1cm} (9)

In which $Y_i$ is the replied variable (biomass and/or carbon) observed in the field (i); $\hat{Y}_i$ is the estimate (biomass and/or carbon); k is the number of parameters; and n is the total number of observations.

2.6. Generation of Stock Maps Using LiDAR Data

After selecting the best model, the AsciiGrid Input function, available in yaInpute in R package [28], was used to view the estimates generated by the model on a map. The map expresses a grid where each cell represents a 5 × 5 m grid colored according to the biomass and the estimated carbon content for that cell. The adopted methodology can be seen in the organization chart below (Figure 4).
The average minimum and maximum elevation values which correspond to the tree heights obtained by the LiDAR metrics ranged from 1.49 to 4.77 m with an average of 3.06 m for the Correntão area. There is little difference in the minimum elevations per plot (1.39 m) for the Transposição area, but this area had the lowest average elevation (2.88 m) and the highest maximum elevation (5.05 m). The highest biomass and total carbon concentration per hectare was observed in the Correntão area, with values ranging from 0.61 to 129 Mg·ha⁻¹, with an average of 24.93. Total above ground biomass (TAGB) estimates for all plots in the Transposição area ranged from 1.22 to 29.37 Mg·ha⁻¹, with an average value of 9.32 Mg·ha⁻¹.

The component loads (correlations between each variable and each principal component), the eigenvalues and the variation percentage of the principal components for the LiDAR metrics for the Transposição and Correntão areas are shown in Tables 4 and 5, respectively. The principal component analysis of the Transposição metrics produced three principal components which synthesized 85.6% of the variability in the data. The first principal component (PC1) of the metrics in this area had a high load of all variables with maximum contribution from the average increase (0.996). The principal component analysis for the metrics of the Correntão area also produced three principal components which only represented a small amount of variance were not used in the regression analysis.
Table 4. Component loads, eigenvalues and percentage of variation of the main principal components (PCs) for the LiDAR metrics in the Transposição area.

| Principal Component | Elev. Minimum | Elev. Maximum | Elev. Mean | Elev. Mode | Elev. Stdddev | Elev. CV | Elev. Skewness | Elev. Kurtosis | Elev. MAD. Median | Elev. P01 | Auto Values | Var (%) |
|---------------------|--------------|--------------|-----------|-----------|--------------|---------|---------------|---------------|----------------|-----------|-----------|---------|
| PC1                 | 0.285        | 0.678        | 0.996     | 0.759     | 0.719        | 0.333   | −0.231        | −0.192        | 0.689          | 0.420    | 12.463    | 51.929  |
| PC2                 | −0.395       | 0.621        | 0.070     | −0.020    | 0.675        | 0.902   | 0.738         | 0.151         | 0.590          | −0.508   | 5.841     | 76.266  |
| PC3                 | 0.622        | 0.281        | −0.009    | −0.070    | 0.016        | 0.597   | 0.672         | −0.203        | 0.662          | 2.255    | 85.663    |         |

Table 5. Component loads, eigenvalues and percentage of variation of the main components (PCs) for the LiDAR metrics in the Correntão area.

| Principal Component | Elev. Minimum | Elev. Maximum | Elev. Mean | Elev. Mode | Elev. Stdddev | Elev. CV | Elev. Skewness | Elev. Kurtosis | Elev. MAD. Median | Elev. P01 | Auto Values | Var (%) |
|---------------------|--------------|--------------|-----------|-----------|--------------|---------|---------------|---------------|----------------|-----------|-----------|---------|
| PC1                 | 0.019        | 0.825        | 0.999     | 0.709     | 0.708        | 0.158   | −0.192        | 0.170         | 0.557          | 0.305    | 13.233     | 55.136  |
| PC2                 | −0.130       | 0.423        | 0.018     | −0.302    | 0.675        | 0.943   | 0.708         | −0.179        | 0.633          | −0.234   | 4.119     | 72.299  |
| PC3                 | 0.765        | −0.283       | 0.016     | −0.106    | 0.074        | 0.110   | −0.103        | −0.805        | 0.359          | 0.636    | 2.524     | 82.815  |

The results of the step-by-step regression analysis for each area are shown in Table 6. The variables which did not significantly contribute to the TAGB prediction for each area were simultaneously eliminated during the analysis. The determination coefficient ($R^2_{adj}$) of the models for the Transposição area varied from 0.17 to 0.42 and the RMSE varied from 3.18 to 5.99 Mg·ha$^{-1}$. Models with higher $R^2$ values and lower RMSE indicate better TAGB prediction. Models developed using the principal component technique of the LiDAR metrics for each area generally had an unsatisfactory performance.

Table 6. Multiple linear regression models adjusted for the biomass estimate, obtained by LiDAR data. The definition of the variables is described in Table 3.

| Area            | Biomass Predictive Models                                                                 | $R^2_{adj}$ | RMSE |
|-----------------|------------------------------------------------------------------------------------------|-------------|------|
| Transposição    | $\text{TAGB} = -86.809 + 33.295 \times (\text{Elev. minimum}) + 5.446$ (Elev. maximum) + 195.226 (Elev. mean) + 3.774 (Elev. mode) − 92.658 (Elev. stdddev) + 206.851 (Elev. CV) + 13.627 (Elev. skewness) − 1.734 (Elev. kurtosis) + 24.360 (Elev. MAD. median) + 29.676 (Elev. P01) − 25.707 (Elev. P10) − 64.704 (Elev. P20) + 49.118 (Elev. P25) − 26.958 (Elev. P30) − 44.133 (Elev. P50) − 21.226 (Elev. P60) − 11.419 (Elev. P75) 2.295 (Elev. P80) − 27.855 (Elev. P90) − 15.740 (Elev. P95) + 98.142 (Canopy. relief. ratio) + 0.024 (Percentage all. returns above 1.30) | 0.1924      | 3.18 |
| Correntão       | $\text{TAGB} = -21.08 + 35.756 \times (\text{Elev. minimum}) + 119.784$ (Elev. mean) + 4.582 (Elev. mode) − 63.752 (Elev. stdddev) + 101.103 (Elev. CV) + 27.823 (Elev. P01) − 17.626 (Elev. P10) − 29.152 (Elev. P20) − 44.745 (Elev. P50) − 18.032 (Elev. P90) + 143.152 (Canopy. relief. ratio) + 0.024 (Percentage all. returns above 1.30) + 519.378 (Canopy. relief. ratio) − 0.103 (Percentage all. returns above 1.30) | 0.4239      | 3.51 |
| Correntão       | $\text{TAGB} = 9.145 + 0.607 \times (\text{Dim. 1}) + 1 \times (\text{Dim. 3})$ | 0.1723      | 5.99  |
Table 6. Cont.

| Area          | Biomass Predictive Models                              | $R^2_{adj}$ | RMSE  |
|---------------|-------------------------------------------------------|-------------|-------|
| Stepwise regression | $\text{TAGB} = -269.86 + 145.44 \text{ (Elev. maximum)} - 402.19\text{ (Elev. mean)} + 440.13 \text{ (Elev.CV)} - 53.26 \text{ (Elev.kurtosis)} + 88.49$ | 0.533       | 14.76 |
| PCA regression | $\text{TAGB} = 30.270 - 6.465 \text{ (Dim.3)} + 93.09 \text{ (Elev.P01)} + 165.73 \text{ (Elev.P30)} - 67.6$ | 0.09621     | 28.45 |

The models which combined the step-by-step selection of the metrics showed better results in both areas compared to models which use the traditional multiple regression modeling, although the prediction error of the multiple model was slightly lower. Based on the adjustment statistics, the stepwise regression model was considered the most suitable for predicting TAGB for both areas and this is illustrated in Figure 5.

![Figure 5](image)

**Figure 5.** Predictions of the best allometric equations for the Correntão area (a) and Transposição (b) using the LiDAR metrics selected step by step in relation to the observed values. The solid black line indicates the fit of the equation and the red dotted line indicates the ideal fit in a 1:1 ratio.

The most significant predictor variables for the model for the Transposição area, and therefore the most suitable for predicting local TAGB, were Elev.minimum, Elev.maximum, Elev.mean and Elev.P01; the last two were found in all tested models. The final model for the Correntão area predicts that the total biomass stock may be more significant with the metrics of Elev.maximum, Elev.mean, Elev.CV, Elev.kurtosis, Elev.P01, Elev.P10, Elev.P30, Elev.P75 and Canopy.relief.ratio.

The stepwise regression model was applied to map the density of TAGB in the study areas (Figure 6). The TAGB map was converted to carbon stock/Mg·ha$^{-1}$ (TAGC) maps to predict the total stock in the areas using the average carbon fraction (0.48), as estimated in this study. The estimated average density of TAGB and TAGC for the Transposição area was $9.2 \pm 6.1$ Mg·ha$^{-1}$ and $4.4 \pm 14$ Mg·ha$^{-1}$, respectively. Most plots in both areas had a low density of TAGB above 30 Mg·ha$^{-1}$ (85.8%), storing approximately 44.1% of the TAGC stock. About 1.2% of the forest stand in the area in both locations had a TAGB density greater than 50 Mg·ha$^{-1}$. Areas with TAGB below 100 Mg·ha$^{-1}$ cover about 99% of the forest stand. Low-density areas of TAGB and TAGC, as shown in the maps, are potential areas for assisted regeneration and enrichment planting to increase the carbon sequestration capacity of the forest stand.
When comparing the areas, it is noted that the most preserved area (Transposição) presented a lower carbon stock than the most degraded area (Correntão) (Figure 6). This can be explained in the inventories carried out in the degraded area, which showed individuals with greater dominance than in the preserved area, strongly influencing the predicted biomass and carbon values.

These results are not only absolute numbers of the inventoried variables, but also other variables obtained from the LiDAR technology and should be considered for better interpretation of the results obtained in this work, namely: the density of pulses emitted by LiDAR, the characteristics of the vegetation trunks/stems studied and the effect of water stress on leaves.

4. Discussion

The preliminary results of this study indicate that the LiDAR metrics provided reliable estimates of TAGB and TAGC for the two study areas. The results show a good statistical relationship between field biomass data and height (elevation) metrics, suggesting that they are an important predictor of local biomass, especially when selected using the stepwise method. However, the relationship was weaker in both areas when all metrics were incorporated into the traditional modeling process or when using regression with principal components. Only the individual categories of the height metrics (minimum, average and maximum) explained more than 40% of the variance. This suggests that the incorporation of height data may be necessary to improve the prediction of TAGB and TAGC.

The $R^2$ and RMSE values of the best-fit models vary in similar studies which used LiDAR data and metrics to estimate and map the aboveground biomass of trees in dry forests. The authors of [29] reported $R^2$ values ranging from 0.36 to 0.71 and RMSE values from 99 to 175 Mg·ha$^{-1}$ in Idaho (USA) based on machine learning data. In a dry tropical forest in Mexico, [30] reported $R^2$ of 0.77 and RMSE from 21.6 to 25.7 Mg·ha$^{-1}$ based on linear regression and LiDAR data. Additionally, in Idaho (USA), [11] reported $R^2$ of 0.87 and RMSE of 3.59 kg based on linear regression and data on percentage plant cover derived from other variables.
from ALS (airborne laser scanning). The authors of [31] reported $R^2$ ranging from 0.38 to 0.64 in the Amazon based on mixed-effect models and LiDAR data from agroforestry systems. The biomass estimates by the tested methodologies showed satisfactory and promising results in all of these cases.

The best-fit TAGB model consisted of 10 LiDAR predictor variables for the Transposição area and nine LiDAR predictor variables for the Correntão area. These variables indicate that the vertical profile of vegetation at different elevations (minimum, medium and maximum heights) are potential predictors for TAGB. TAGB estimation and mapping for the studied forest was based on the best-fit model.

The spatial distribution pattern of TAGB is related to the structure and composition of the forest landscape, where low-TAGB areas correspond to low-forest cover areas, while high-biomass areas correspond to densely forested areas. The overall average TAGB of 32 Mg·ha$^{-1}$ for two areas is comparable and is within the range of values reported for mature dry tropical forests in previous studies, indicating that the modeling approach used in this study provides reasonable predictions of TAGB [15,32,33].

Uncertainties in the TAGB estimates in this study could result from errors associated with field measurements, however, plot TAGB estimates are based on a location-specific allometric equation with the breadth of data covering a wide sample in the model development, showing an accurate TAGB estimate. Any errors associated with the TAGB estimate at the plot level using a pantropical equation would propagate to the TAGB regression model, and therefore to the TAGB estimates [34,35]. However, another possible source of error in biomass predictions may be the density of points adopted, as 0.5 pulses/m$^2$ were adopted in the data used, while carbon quantification works generally work with 4 to 25 points [4,16,36,37]. Through tests with point density, [9] found that its reduction can result in a decrease in $R^2$ and an increase in RMSE, in addition to increasing the variance of AGB estimates. The authors of [38] obtained bias in height estimates which translated into errors of 80–125 Mg·ha$^{-1}$ when the operator worked with the pulse density below 4 m$^2$.

Pulse densities from moderate to high changes are invariant and affect results relatively little; however, once the pulse drops to 1/m$^2$, it makes metrics related to coverage (canopy coverage, tree density and shrub coverage) more sensitive to changes in this density [39]. Thus, the increase in RMSE can be explained by the less accurate classification of soil returns [9]. However, it is worth noting that the low density of points can also be related to good results, depending on the variable to be studied, such as canopy metrics [40] or volume [41]. In addition, the TAGC for the study area was estimated based on the TAGB and the average carbon fraction of the trees; any errors in the TAGB estimate would extend to the TAGC estimates. However, the results of this study indicate that the TAGB model provided reliable TAGB estimates for the study area.

5. Conclusions

According to the results found in this work, we concluded that LiDAR data can be used for estimating biomass and total carbon in dry tropical forest, as confirmed by an adjustment considered in the models used, with good correlation between the LiDAR metrics and the biomass data observed in the field. More specifically, we have the following conclusions and recommendations for future work:

- Using a stepwise method to reduce the metrics proved to be more effective for better adjustment of the models;
- The LiDAR metrics which were most present in the models were: Elev.minimum, Elev.maximum, Elev.mean and Elev.P01, with the last two being found in all models;
- The most preserved area had less carbon stock than the most degraded area, this occurrence can be explained in the inventories carried out in the area that presented the largest number of bole measured at ground level (DGL: diameter at ground level) in the area degraded than in the preserved area, strongly influencing the estimated carbon values in the areas;
• The pulse density, even though it is not a variable within the models, indirectly influenced the accuracy of the models, therefore, it is recommended that data be tested with a higher pulse density in future works.

• The model is limited to the TAGB estimate in the study area and may not be suitable for application in other forests. This is due to differences in forest structure, species composition, vegetation vigor and impacts of atmospheric conditions and soil moisture and precipitation.

• New studies are recommended to assess the transferability of the model to other protected forests with same forest structure and species composition. Other studies that will test the ability of non-parametric algorithms (such as random forest) to develop TAGB estimation models for the study area, in comparison with linear regression analysis, are also recommended.

• Our preliminary results provided important information on the spatial distribution of TAGB and TAGC in the study area, which can be used to manage the reserve for optimal carbon sequestration.

Author Contributions: C.P.d.O. and R.L.C.F. planned the study and wrote the manuscript. C.P.d.O., R.B.d.L., J.A.A.d.S. and E.A.S. participated in the processing LiDAR data, calculation and modeling of the data and wrote the manuscript. E.A.S., J.D.S.d.L., A.F.d.S., N.A.T.d.S., C.L.S.-M.d.M., M.M.d.L.P. and I.J.C.L. collected and processed the data inventory. They also commented on the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Coordination for the Improvement of Higher Education Personnel-Brazil (CAPES), Financial Code 001, National Council for Scientific and Technological Development (CNPq-303991/2016-0; 400540/2016-9), Foundation for the Support of Science and Technology of the State of Pernambuco and Federal Rural University of Pernambuco (FACEPE and UFRPE-IBPG-0827-5.02/14).

Data Availability Statement: The data can be consulted by email request to the lead authors of the project.

Acknowledgments: We thank the forestry workers and German Cespedez, who is responsible for the data access to the area for developing this work. The authors are grateful for the financial support granted by the National Council for Scientific and Technological Development (CNPq) by the productivity bag of the second and third co-authors and project leaders. Special thanks to the University of the State of Amapá for financially supporting the researchers with the fees of publication of the manuscript.

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

References
1. Crowther, T.W.; Glick, H.B.; Covey, K.R.; Bettigole, C.; Maynard, D.S.; Thomas, S.M.; Smith, J.R.; Hintler, G.; Duguid, M.C.; Amatulli, G.; et al. Mapping Tree Density at a Global Scale. Nature 2015, 525, 201–205. [CrossRef]
2. Achard, F.; House, J.I. Reporting Carbon Losses from Tropical Deforestation with Pan-Tropical Biomass Maps. Environ. Res. Lett. 2015, 10, 101002. [CrossRef]
3. Brahma, B.; Nath, A.J.; Sileshi, G.W.; Das, A.K. Estimating Biomass Stocks and Potential Loss of Biomass Carbon through Clear-Felling of Rubber Plantations. Biomass Bioenergy 2018, 115, 88–96. [CrossRef]
4. Coomes, D.A.; Dalponte, M.; Jucker, T.; Asner, G.P.; Banin, L.F.; Burslem, D.F.R.P.; Lewis, S.L.; Nilus, R.; Phillips, O.L.; Phua, M.-H.; et al. Area-Based vs Tree-Centric Approaches to Mapping Forest Carbon in Southeast Asian Forests from Airborne Laser Scanning Data. Remote Sens. Environ. 2017, 194, 77–88. [CrossRef]
5. Althoff, T.D.; Menezes, R.S.C.; Pinto, A.D.S.; Pareyn, F.G.C.; de Carvalho, A.L.; Martins, J.C.R.; de Carvalho, E.X.; da Silva, A.S.A.; Dutra, E.D.; Sampaio, E.V.D.S.B. Adaptation of the Century Model to Simulate C and N Dynamics of Caatinga Dry Forest before and after Deforestation. Agric. Ecosyst. Environ. 2018, 254, 26–34. [CrossRef]
6. Sampaio, E.V.; Silva, G.C. Biomass Equations for Brazilian Semiarid Caatinga Plants. Acta Bot. Bras. 2005, 19, 935–943. [CrossRef]
7. Duncanson, L.; Huang, W.; Johnson, K.; Suanantran, A.; McRoberts, R.E.; Dubayah, R. Implications of Allometric Model Selection for County-Level Biomass Mapping, Carbon Balance Manag. 2017, 12, 1–11. [CrossRef]
8. Somogyi, Z.; Cienciala, E.; Mäkipää, R.; Muukkonen, P.; Lehtonen, A.; Weiss, P. Indirect Methods of Large-Scale Forest Biomass Estimation. Eur. J. For. Res. 2007, 126, 197–207. [CrossRef]
9. Silva, C.; Hudak, A.; Vierling, L.; Klauberg, C.; Garcia, M.; Ferraz, A.; Keller, M.; Eitel, J.; Saatchi, S. Impacts of Airborne Lidar Pulse Density on Estimating Biomass Stocks and Changes in a Selectively Logged Tropical Forest. Remote Sens. 2017, 9, 1068. [CrossRef]

10. Silva, C.A.; Klauberg, C.; Hudak, A.T.; Vierling, L.A.; Liesenberg, V.; Carvalho, S.P.C.E.; Rodriguez, L.C.E. A Principal Component Approach for Predicting the Stem Volume in Eucalyptus Plantations in Brazil Using Airborne LiDAR Data. Forestry 2016, 89, 422–433. [CrossRef]

11. Li, A.; Dhakal, S.; Glenn, N.; Spaete, L.; Shinneman, D.; Pilliod, D.; Arkle, R.; McIlroy, S. Lidar Aboveground Vegetation Biomass Estimates in Shrublands: Prediction, Uncertainty and Application to Coarser Scales. Remote Sens. 2017, 9, 903. [CrossRef]

12. Avitabile, V.; Herold, M.; Henry, M.; Schmullius, C. Mapping Biomass with Remote Sensing: A Comparison of Methods for the Case Study of Uganda. Carbon Balance Manag. 2011, 6, 7. [CrossRef]

13. Saatchi, S.; Malhi, Y.; Zutta, B.; Buermann, W.; Anderson, L.O.; Araujo, A.M.; Phillips, O.L.; Peacock, J.; ter Steege, H.; Lopez Gonzalez, G.; et al. Mapping Landscape Scale Variations of Forest Structure, Biomass, and Productivity in Amazonia. Biogeosci. Discuss. 2009, 6, 5461–5505. [CrossRef]

14. Becknell, J.M.; Keller, M.; Piotto, D.; Longo, M.; dos-Santos, M.N.; Scarranello, M.A.; de Oliveira Cavalcante, R.B.; Porder, S. Landscape-Scale Lidar Analysis of Aboveground Biomass Distribution in Secondary Brazilian Atlantic Forest. Biotropica 2018, 50, 520–530. [CrossRef]

15. Martinuzzi, S.; Gould, W.A.; Vierling, L.A.; Hudak, A.T.; Nelson, R.F.; Evans, J.S. Quantifying Tropical Dry Forest Type and Succession: Substantial Improvement with LiDAR. Biotropica 2013, 45, 135–146. [CrossRef]

16. Nelson, R.; Margolis, H.; Montesano, P.; Sun, G.; Cook, B.; Corp, L.; Andersen, H.-E.; dejong, B.; Pellat, F.P.; Fickel, T.; et al. Lidar-Based Estimates of Aboveground Biomass in the Continental US and Mexico Using Ground, Airborne, and Satellite Observations. Remote Sens. Environ. 2017, 188, 127–140. [CrossRef]

17. Mohebian, P.M.; Aguilar, F.X. Beneath the Canopy: Tropical Forests Enrolled in Conservation Payments Reveal Evidence of Less Degradation. Ecol. Econ. 2018, 143, 64–73. [CrossRef]

18. Oliveira, C.P.; Ferreira, R.L.C.; Silva, J.A.A.; Lima, R.B.; Silva, E.A.; Júnior, F.T.A.; Silva, A.F.; Lucena, J.; Santos, N.A.T.; Lopes, I.J.C.; et al. Prediction of Biomass in Dry Tropical Forests: An Approach on the Importance of Total Height in the Development of Local and Pan-tropical Models. J. Sustain. For. 2021, 31, 1–16. [CrossRef]

19. Instituto Brasileiro de Geografia e Estatística-IBGE. Coordenação de Recursos Naturais e Estudos Ambientais. 2012; Manual Técnico da Vegetação Brasileira. Available online: https://biblioteca.ibge.gov.br/visualizacao/livros/liv63011.pdf (accessed on 22 March 2017).

20. Empresa Brasileira de Pesquisa Agropecuária-Embrapa. Zoneamento Agroecológico do Estado de Pernambuco–ZAPE. 2007. Available online: http://www.uep.cnps.embrapa.br/zape (accessed on 15 April 2017).

21. Sistema de Monitoramento Agrometeorológico-Agritempo. Estações Meteorológicas para o Estado de PE. Available online: https://www.agritempo.gov.br/agritempo/jsp/estacao/index.jsp?siglaUF=PE&lang=pt_br (accessed on 10 April 2017).

22. Lana, M.D.; Ferreira, R.L.C.; Silva, J.A.A.; Duda, G.P.; Cespedes, G.H.G. Carbon Content in Shrub-tree Species of the Caatinga. Floresta e Ambiente 2019, 26, e20170617. [CrossRef]

23. Brown, S.; Lugo, A.E. The storage and production of organic matter in tropical forests and their role in the global carbon cycle. Biotropica 1982, 14, 161–187. [CrossRef]

24. Roy, J.; Saugier, B.; Mooney, H.A. Global Terrestrial Productivity, 1st ed.; Academic Press: San Diego, CA, USA, 2001; pp. 559–573.

25. Malhi, Y.; Baker, T.R.; Phillips, O.L.; Almeida, S.; Alvarez, E.; Arroyo, L.; Chave, J.; Czimczik, C.I.; Di Fiore, A.; Higuchi, N.; et al. The above-ground coarse wood productivity of 104 Neotropical forest plots. Biotropica 1982, 14, 161–187. [CrossRef]

26. Elias, M.; Potvin, C. Assessing inter- and intra-specific variation in trunk carbon concentration for 32 neotropical tree species. Can. J. For. Res. 2003, 33, 1039–1045. [CrossRef]

27. Jensen, J.R. Introductory Digital Image Processing: A Remote Sensing Perspective, 4th ed.; Brigham Young University: Provo, Utah; Prentice Hall: Upper Saddle River, NJ, USA, 2016; pp. 1–400.

28. Crookston, N.L.; Finley, A. Yaimpute: An R package for k-NN imputation. J. Stat. Softw. 2008, 23, 1–16. [CrossRef]

29. Anderson, K.E.; Glenn, N.F.; Spaete, L.P.; Shinneman, D.J.; Pilliod, D.S.; Arkle, R.S.; McIlroy, S.K.; Derryberry, D.R. Estimating Vegetation Biomass and Cover across Large Plots in Shrub and Grass Dominated Drylands Using Terrestrial Lidar and Machine Learning. Ecol. Indic. 2018, 84, 793–802. [CrossRef]

30. Hernández-Stefanoni, J.L.; Johnson, K.D.; Cook, B.D.; Dupuy, J.M.; Birdsey, R.; Peduzzi, A.; Tun-Dzul, F. Estimating Species Richness and Biomass of Tropical Dry Forests Using LISSAR during Leaf-on and Leaf-off Canopy Conditions. Appl. Veg. Sci. 2015, 18, 724–732. [CrossRef]

31. Chen, Q.; Lu, D.; Keller, M.; dos-Santos, M.; Bolle, E.; Feng, Y.; Wang, C. Modeling and Mapping Agroforestry Aboveground Biomass in the Brazilian Amazon Using Airborne Lidar Data. Remote Sens. 2015, 8, 21. [CrossRef]

32. Cheng, H.; A. Oktari, H.; Nasset, E.; Eit, T.; Gobakken, T. Influence of Plot Size on Efficiency of Biomass Estimates in Inventories of Dry Tropical Forests Assisted by Photogrammetric Data from an Unmanned Aircraft System. Remote Sens. 2017, 9, 610. [CrossRef]

33. Naveen Kumar, J.; Arunkumar, K.S.; Sundarapandian, S. Biomass and Carbon Stocks of a Tropical Dry Forest of the Javadi Hills, Eastern Ghats, India. Carbon Manag. 2017, 8, 351–361. [CrossRef]
34. Chave, J.; Réjou-Méchain, M.; Búrquez, A.; Chidumayo, E.; Colgan, M.S.; Delitti, W.B.C.; Duque, A.; Eid, T.; Fearnside, P.M.; Goodman, R.C.; et al. Improved Allometric Models to Estimate the Aboveground Biomass of Tropical Trees. *Glob. Chang. Biol.* **2014**, *20*, 3177–3190. [CrossRef] [PubMed]
35. Chave, J.; Andalo, C.; Brown, S.; Cairns, M.A.; Chambers, J.Q.; Eamus, D.; Fölster, H.; Fromard, F.; Higuchi, N.; Kira, T.; et al. Tree Allometry and Improved Estimation of Carbon Stocks and Balance in Tropical Forests. *Oecologia* **2005**, *145*, 87–99. [CrossRef] [PubMed]
36. Silva, C.A.; Klauberg, C.; Hudak, A.T. Mapeamento de Estoques de Carbono Acima Do Solo Utilizando Dados LiDAR Em Plantações de Eucalyptus Spp No Estado de São Paulo, Brasil. *Sci. For.* **2014**, *42*, 14.
37. Figueiredo, E.O.; d’Oliveira, M.V.N.; Braz, E.M.; de Almeida Papa, D.; Fearnside, P.M. LIDAR-Based Estimation of Bole Biomass for Precision Management of an Amazonian Forest: Comparisons of Ground-Based and Remotely Sensed Estimates. *Remote Sens. Environ.* **2016**, *187*, 281–293. [CrossRef]
38. Leitold, V.; Keller, M.; Morton, D.C.; Cook, B.D.; Shimabukuro, Y.E. Airborne Lidar-Based Estimates of Tropical Forest Structure in Complex Terrain: Opportunities and Trade-Offs for REDD+. *Carbon Balance Manag.* **2015**, *10*, 1–12. [CrossRef] [PubMed]
39. Magnusson, M.; Fransson, J.E.S.; Holmgren, J. Effects on Estimation Accuracy of Forest Variables Using Different Pulse Density of Laser Data. *For. Sci.* **2007**, *53*, 619–626. [CrossRef]
40. Thomas, V.; Treitz, P.; McCaughey, J.H.; Morrison, I. Mapping Stand-Level Forest Biophysical Variables for a Mixedwood Boreal Forest Using Lidar: An Examination of Scanning Density. *Can. J. For. Res.* **2006**, *36*, 34–47. [CrossRef]
41. Takahashi, T.; Awaya, Y.; Hirata, Y.; Furuya, N.; Sakai, T.; Sakai, A. Stand Volume Estimation by Combining Low Laser-Sampling Density LiDAR Data with QuickBird Panchromatic Imagery in Closed-Canopy Japanese Cedar (*Cryptomeria japonica*) Plantations. *Int. J. Remote. Sens.* **2010**, *31*, 1281–1301. [CrossRef]