Article

Improving Species Diversity and Biomass Estimates of Tropical Dry Forests Using Airborne LiDAR

José Luis Hernández-Stefanoni [1],*, Juan Manuel Dupuy [1], Kristofer D. Johnson [2], Richard Birdsey [2], Fernando Tun-Dzul [1], Alicia Peduzzi [2], Juan Pablo Caamal-Sosa [3], Gonzalo Sánchez-Santos [3] and David López-Merlín [3]

[1] Centro de Investigación Científica de Yucatán, A.C. Unidad de Recursos Naturales, Calle 43 # 130. Colonia Chuburná de Hidalgo, Mérida, Yucatán C.P. 97200, Mexico; E-Mails: jmdupuy@cicy.mx (J.M.D.); fjtun@cicy.mx (F.T.-D.)

[2] U.S. Forest Service, Northern Research Station, Newtown Square, PA 19073, USA; E-Mails: kristoferdjohnson@fs.fed.us (K.D.J.); rbirdsey@fs.fed.us (R.B.); apeduzzi@fs.fed.us (A.P.)

[3] Proyecto México-Noruega, Comisión Nacional Forestal, Col. Del Carmen Coyoacán, Coyoacán D.F. C.P. 04100, Mexico; E-Mails: jcaamalsosa@gmail.com (J.P.C.-S.); bedxemex@hotmail.com (G.S.-S.); davidlopezmerlin@gmail.com (D.L.-M.)

* Author to whom correspondence should be addressed; E-Mail: jl_stefanoni@cicy.mx; Tel.: +52-999-942-8330 (ext. 372).

Abstract: The spatial distribution of plant diversity and biomass informs management decisions to maintain biodiversity and carbon stocks in tropical forests. Optical remotely sensed data is often used for supporting such activities; however, it is difficult to estimate these variables in areas of high biomass. New technologies, such as airborne LiDAR, have been used to overcome such limitations. LiDAR has been increasingly used to map carbon stocks in tropical forests, but has rarely been used to estimate plant species diversity. In this study, we first evaluated the effect of using different plot sizes and plot designs on improving the prediction accuracy of species richness and biomass from LiDAR metrics using multiple linear regression. Second, we developed a general model to predict species richness and biomass from LiDAR metrics for two different types of tropical dry forest using regression analysis. Third, we evaluated the relative roles of vegetation structure and habitat heterogeneity in explaining the observed patterns of biodiversity and biomass, using variation partition analysis and LiDAR metrics. The results showed that with increasing plot size, there is an increase of the accuracy of biomass estimations. In contrast, for species

richness, the inclusion of different habitat conditions (cluster of four plots over an area of 1.0 ha) provides better estimations. We also show that models of plant diversity and biomass can be derived from small footprint LiDAR at both local and regional scales. Finally, we found that a large portion of the variation in species richness can be exclusively attributed to habitat heterogeneity, while biomass was mainly explained by vegetation structure.

Keywords: above-ground biomass; biodiversity; habitat heterogeneity; LiDAR; vegetation structure

---

1. Introduction

Tropical forests are one of the most diverse terrestrial communities in the world. They provide goods and ecological services to human populations, store more carbon than any other terrestrial biome and play a crucial role for mitigating future global warming [1]. Nevertheless, the rate of forest loss by natural disturbances and human interventions has increased dramatically. Tropical deforestation currently represents about 8% of annual global greenhouse gas emissions, although historically, this rate has ranged from 15%–25% [2,3]. Tropical forest clearing brings declines in biological diversity, environmental functions and forest products [4]. Thus, accurate estimates of the spatial distribution of plant diversity and biomass are needed to support policies that are designed to maintain terrestrial biodiversity and tropical forest carbon stocks. Remote sensing is becoming widely used for supporting such activities [5,6]. Understanding and identifying the main factors that affect biodiversity and biomass is critical to mapping these variables [7]. This would allow us to develop appropriate methods for predicting these variables if we found some parameters or indicators measured from remotely sensed data that can be used as proxies for these factors [8,9].

There are several limitations of using remote sensing data to quantify important vegetation structure parameters, such as diversity and biomass [10]: first, the use of medium or coarse spatial resolutions to estimate biodiversity; a single pixel often encompasses a number of different individual plants belonging to different species. Thus, each pixel corresponds to a mixed averaged signature, leading to difficulties in individual species identification. However, increasing spatial resolution is often accompanied by reducing other sensor properties. For example, a recent study [11] compared the ability of medium (Landsat) and high- (IKONOS) resolution satellite imagery for assessing plant diversity in a tropical dry Indian forest and found that Landsat performed better than IKONOS. This was due to a scale factor, since it is not possible to directly measure habitat heterogeneity at high spatial resolution. Thus, medium resolutions allow the development of meaningful measures of landscape heterogeneity that are more related to species diversity. A second major limitation is the inability of different vegetation indices (such as NDVI) to detect changes in areas of high biomass above a threshold level, which makes it difficult to estimate biomass and species richness in certain areas, due to the saturation of the signal [12]. Finally, there is often poor availability of cloud-free optical imagery in the tropics.

Recent studies have found that LiDAR (light detection and ranging) can be a powerful predictor of different vegetation attributes, such as height, basal area, stem density and other vegetation structure

parameters [13–15]; yet, LiDAR has been used only rarely to estimate plant species diversity [16,17]. This sensor uses laser pulses to directly measure ground and vegetation height, as well as the vertical distribution of intercepted surfaces, making it an ideal tool for mapping vegetation structure with no saturation at high biomass values [18]. Thus, LiDAR measurements have been shown to produce more accurate estimates of vegetation structure parameters than other remotely sensed data, because LiDAR has the ability to penetrate tropical forest canopies and to detect three-dimensional forest structures [19]. Above-ground biomass is related to several vegetation structure parameters, including diameter, height and basal area [20]. Furthermore, many studies have demonstrated a strong relationship between above-ground biomass and LiDAR measurements in different ecosystems, ranging from conifers to tropical forest [18,21].

Other studies have explored the potential of LiDAR to model the assemblage composition and diversity of insects, spiders and birds [22–24]. However, few studies have explicitly analyzed the relationships between plant species diversity and LiDAR measurements [16,17]. Here, we evaluated the potential of LiDAR to map the spatial distribution of species richness in a tropical dry forest based on two main factors related to diversity. First, we tested whether LiDAR can be used to predict plant diversity based on vegetation structure. LiDAR measurements are well related to estimates of vegetation structure parameters [19], and these parameters are associated with different groups of species. For example, pioneer species commonly grow in open areas, while non-pioneer species are established almost entirely beneath the forest canopy [25]. Moreover, the species richness of tropical dry forests increases from young to old stands [26,27]. Second, we evaluated whether species richness is related to habitat complexity or heterogeneity. Different measures of variability of remotely sensed data have been proposed and used to measure habitat heterogeneity, such as the variance of vegetation indices like NDVI [28], the spectral variability derived from the mean of spectral values in a multi-dimensional system of bands [29] and variability in the reflectance values among pixels using the texture of remotely sensed imagery [30]. It is generally assumed that greater habitat heterogeneity allows for a higher number of species to coexist [31]. Here, we propose to use the variance of LiDAR metrics as a proxy of habitat heterogeneity.

The goal of this study was to evaluate the accuracy of predicting the spatial distribution of species richness and above-ground biomass using airborne LiDAR in two tropical dry forests of the Yucatan Peninsula. The design and aims of the study address some of the research shortcomings that we have outlined previously. To this end, we established three specific objectives.

The first objective was to evaluate the effect of using different plot sizes and plot designs for improving the prediction accuracy of the species richness and biomass of tropical dry forests. There is a tendency for errors to decrease in biomass estimates with increasing plot size, because large plots reduce the likelihood of edge effects, which occur when the canopy of trees are found along the plot boundary. There is also a decrease of errors in general, because large plots capture an adequate amount of structural variability in the field [32]. Thus, our first prediction was that the accuracy of estimation for above-ground biomass (AGB) and species richness would increase as the plot size increased [33]. In terms of plot design, having a cluster of separated plots allows for capturing habitat heterogeneity, possibly resulting in higher correlations with species richness, but could also increase errors associated with edge effects, thereby potentially decreasing the accuracy of AGB estimates. Consequently, our

second prediction was that having a cluster of separated plots would improve estimates of species richness, but not of AGB.

The second objective was to develop a general model to estimate species richness and above-ground biomass from LiDAR data for two different types of tropical dry forests. We expected that vegetation structure and species composition would vary with climate and anthropogenic disturbance regimes in the studied area [27] and that relationships between LiDAR metrics and forest characteristics are site dependent [34]. It is important to evaluate the accuracy of a general model encompassing different vegetation types, for possible applications at a regional level.

The third objective was to use LiDAR metrics as surrogates for both vegetation structure and habitat heterogeneity to evaluate the relative roles of these two factors in explaining the observed patterns of biodiversity and biomass. AGB increases rapidly with stand age in the Yucatan Peninsula [35,36], and we hypothesized that vegetation structure is the most important variable affecting above-ground biomass due to its relationship with stand age. We also hypothesized that species richness is influenced mostly by habitat heterogeneity, since plant diversity has been reported to be strongly correlated with environmental variables, such as soil fertility and proximity to seed sources [37].

2. Data and Methods

2.1. Site Descriptions

We acquired LiDAR imagery and collected field data from two different regions of the Yucatan Peninsula—the Kiuic site located in the southern part of the State of Yucatan (89°32′W–89°34′W, 20°04′N–20°06′N) and the Felipe Carrillo Puerto (FCP) site situated in the middle portion of the State of Quintana Roo (88°03′W–88°05′W, 19°28′N–19°30′N) (Figure 1). The Kiuic site lies within a private protected area, while the FCP site consists mostly of communal land. Both sites are covered with a tropical dry forest and have a tropical warm climate, with summer rain and a dry season from November to April, and a mean annual temperature of about 26 °C. However, there is considerable variation in the precipitation, topography and land-use history between both sites that confers differences in species composition and vegetation structure. On the Kiuic site, mean annual precipitation ranges between 1000 and 1100 mm. The landscape consists of Cenozoic limestone hills with a moderate slope (10°–25°) alternating with flat areas, and the elevation ranges from 60 to 180 m [38]. The area is dominated by seasonally dry semi-deciduous tropical forests (50%–75% of species drop their leaves during the dry season) of different ages of abandonment after traditional slash-and-burn agriculture. The forest has a relatively low canopy stature (8–13 m) with a few prominent trees attaining 15–18 m in the oldest (60–70 year old) stands. The most abundant species in this forest are *Neomillspaughia emarginata*, *Gymnopodium floribundum*, *Bursera simaruba*, *Piscidia piscipula* and *Lysiloma latisiliquum*. The FCP site has fairly flat topography and mean annual rainfall between 1000 and 1300 mm [39]. The landscape is dominated by seasonally dry semi-evergreen tropical forest (25%–30% of species drop their leaves during the dry season), which grows up to 25 m tall and is a structurally complex community with two or three canopy layers, consisting mostly of trees, where the most abundant species are *Manilkara zapota*, *Vitex gaumeri*, *Bursera simaruba*,

Metopium brownei and *Cecropia obtusifolia*. The dominant land use is pastures for cattle raising, although traditional swidden agriculture is also practiced, both leading to a mosaic of open fields and vegetation in different successional stages.

Figure 1. Location of the study sites and field samples: (**A**) Kiuic and (**B**) Felipe Carrillo Puerto (FCP).
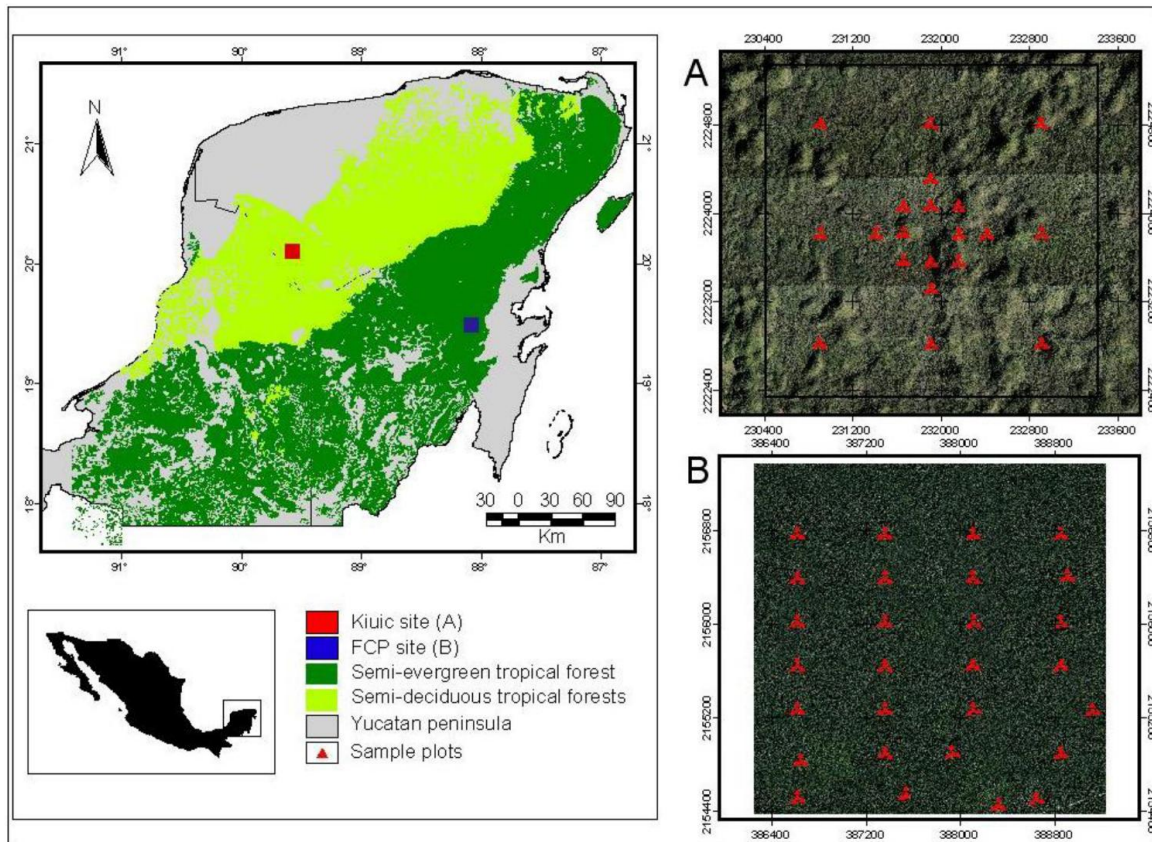
2.2. *Species Richness and Biomass Data*

Field data for both sites were recorded from a systematic plant survey conducted during the rainy season of 2013 in an area of 9 km^2. At the Kiuic site, clusters of sample plots were located systematically around an eddy covariance flux tower, while at the FCP site, the clusters of plots were located on a fixed grid of evenly-spaced sample locations, having in total 20 and 28 sample clusters of plots for the Kiuic and FCP sites, respectively (Figure 1). The cluster plot design was based on the field data layout used for Mexico's National Forest Inventory (INFyS) [40]. Each cluster consists of 4 circular plots of 400 m^2 each, with a radius of 11.28 m. The plots are distributed over an area of 1.0 ha, representing a sample of the conditions within this area. Plot 1 is located in the center of the cluster, whereas Plots 2, 3 and 4 are located 38.6 m at azimuths of 0°, 120° and 240° from the center of Plot 1. All plots were located on the ground with a Garmin GPS unit. The precision of the X and Y coordinates of center plots was estimated from different measurements of the position, and the mean location errors were less than 3 m. In each plot, all woody plants >7.5 cm in DBH (diameter at breast height: 1.3 m) were sampled. In addition to the 400 m^2 area and as a modification of the original INFyS design, we added another larger concentric area of 1000 m^2 (17.84 m radius), centered at

Plot 1, where all woody plants >20.0 cm in DBH were registered. We measured the diameter of all stems and the height of all individuals. We calculated the number of all woody plant species per sample site (*i.e.*, species density sensu [41]), as a measure of local or α species diversity, and above-ground biomass for different plot areas and spatial arrangements: individual 400 m^2 plots, 1000 m^2 plots and the whole cluster (2200 m^2). To calculate the biomass from tree diameter (and height), two allometric equations developed for tropical forests of Mexico were employed: one for trees ≥10 cm in DBH [42] modified by [43] and the other for trees <10 cm in DBH [44]. To calculate liana biomass, we used the allometric equation reported by [45], while for palms ≥10 cm in DBH, we used the equation developed by [46].

2.3. LiDAR Data Processing

The LiDAR coverage data for the Kiuic and FCP sites were acquired in August of 2012 and January of 2013, respectively, by a private contractor, CartoData [47], operating a Cessna T202 aircraft. The LiDAR data were collected using an airborne laser scanner, RIEGL-QV-480 LiDAR, equipped with a NovAtel GPS/IMU and a 16-mpx RGB nadir looking camera. The system was operated at an average height of 396.2 m above ground level, a 30 ° field of view and a pulse repetition frequency of 200 kHz, for which the aircraft maintained a ground speed between 80 and 90 kph. Flights had an approximate overlap of 50% between adjacent flight lines, averaged more than 5 pulses per square meter and included up to 5 returns for each pulse.

LiDAR data were processed using FUSION software [48]. Using the X, Y coordinates and the radius for each field plot, the clouds of points were clipped to correspond with the area of plots (400 and 1000 m^2). Before applying the clipping process, the data were normalized to the ground surface, in order to express the returns in terms of heights above the ground instead of elevation above sea level. Then, a set of 62 LiDAR metrics were calculated from the cloud of points within each of the 400 and 1000-m^2 plots. For the entire cluster of sample plots (2200 m^2), the mean and standard deviation values were calculated for the metrics derived from all three 400-m^2 plots and the 1000-m^2 plot.

The LiDAR metrics were used as the predictor variables in the models for estimating the spatial distribution of species richness and biomass. Such metrics belonged to two categories. The first group was based on height statistics and includes mean, maximum and minimum elevation, the variability of return heights (variance, coefficient of variation), statistics to quantify location (percentiles 1, 5, 10, …, 100 and L-moments), among others. The second group included canopy density metrics and was used to evaluate the amount of vegetation cover. A threshold of 1.5 m as a minimum height above ground was used to reduce the noise within the near-ground cloud of returns caused by low vegetation and imperfections of the ground. A canopy threshold height of 4.0 m was used to compute LiDAR canopy cover metrics. A list of the metrics is shown in Table A1; for a detailed description and the equations used to calculate the LiDAR metrics, see [48].

2.4. Data Analysis

Ordinary least squares (OLS) multiple regression analysis was used to model the statistical relationship between the response variable (species richness or above-ground biomass) and explanatory variables (LiDAR metrics) at each of three sample areas (400, 1000 and 2200 m^2). The

dependent variables were formally tested for normality and homoscedasticity, while the response variables were transformed as needed with $1/x$, $\log 10(x)$, $\log 10(x + 1)$ and $\mathrm{sqrt}(x)$ to meet linearity assumptions [48]. None of the LiDAR metrics (predictors) required transformation once the distributions of response variables were normalized. All multiple regression analyses were carried out using stepwise forward selection. Multicollinearity between predictor variables can cause problems in multivariable modeling; therefore, the explanatory variables considered for the analysis were either uncorrelated or expressed only small collinearity, with a variance inflation factor less than 2.0 [49].

Second, a multiple linear regression technique was used to develop a regional model to predict the species richness and biomass of the tropical dry forest in the Yucatan from LiDAR metrics. These models used the best model selected once the goodness of fit was applied for each variable, and then, they were applied to the combined data sets from both sites (Kiuic and FCP).

The performance of the different models was assessed by leave-one-out cross-validation. In this procedure, one observation is temporally removed from the data set, and the remaining sampling plots are used to fit the model. Then, coefficients obtained are applied to this datum in order to produce a predicted value. The cross-validation yields a list of estimated values of species richness and biomass paired to those obtained from the observed sampling plots. Predicted values were also back-transformed to original values as needed and corrected for bias introduced during the back-transformation process using a method suggested by [50]. The predicted and observed values of species richness and biomass were compared using the coefficient of determination (R^2), the root mean square error (RMSE) and the agreement coefficient (AC) proposed by [51]. The last quantitative measure of agreement has the ability to provide metrics that are bounded by the fixed minimum and maximum values and are standardized to non-dimensional units, so the units of measurement of observed and predicted values do not affect the value. *AC* is calculated as:

$$AC = 1 - \frac{\mathrm{SSD}}{SPOD} \tag{1}$$

where *SSD* is the sum of squared differences:

$$SSD = \sum_{i=1}^{n} \left(X_i - Y_i \right)^2 \tag{2}$$

and *SPOD* is the sum of potential difference:

$$SPOD = \sum_{i=1}^{n} \left(\left| \bar{X} - \bar{Y} \right| + \left| X_i - \bar{X} \right| \left(\left| \bar{X} - \bar{Y} \right| + \left| Y_i - \bar{Y} \right| \right) \right) \tag{3}$$

where $\bar{X}$ and $\bar{Y}$ are the mean values of the observed (*X*) and predicted (*Y*) values of the variable, respectively. The AC values range from <0 to 1, where *AC* = 1 means perfect agreement between observed and predicted values, and values less than or equal to zero indicate no agreement.

Considering that the third objective of this study was to evaluate the relative contribution of vegetation structure (mean values of LiDAR metrics calculated in 4 plots) and habitat heterogeneity (standard deviation values of LiDAR metrics calculated in 4 plots) to overall variation in species richness and biomass, a combination of multiple regression and variation partitioning methods was used [52]. The general procedure involves the following steps. First, a model of multiple regressions between response variables and mean values of LiDAR metrics per cluster was fitted. This model

represents the variability explained by vegetation structure and the variation explained jointly by vegetation structure data and habitat heterogeneity (a + b). Dependent variables (species richness and above-ground biomass) were formally tested for the normality and homogeneity of variances in the residuals [52]. These variables were transformed with $1/x$, $\log10(x)$, $\log10(x + 1)$ and $\mathrm{sqrt}(x)$, as necessary to meet linearity assumptions [49]. Second, a multiple regression model using habitat heterogeneity (standard deviation values of LiDAR metrics calculated in 4 plots per cluster) was fitted to the response variables. This second model represents habitat heterogeneity plus the variation explained jointly by vegetation structure and habitat heterogeneity (b + c). Then, linear trends were checked by conducting a regression analysis of response variables with the *X* and *Y* spatial locations of each site. In the case of significant linear trends, detrended residuals were used as response variables for both previous models. Thirdly, the total amount of variation explained (a + b + c) was calculated by combining the two previous multiple regression models into an overall regression model using exclusively significant selected variables. All multiple regression analyses were carried out using forward selection. Finally, variation partitioning was performed to determine the relative importance of vegetation structure variables (a = (a + b + c) − (b + c)), habitat heterogeneity (c = (a + b + c) − (a + b)) and shared variation (b = (a + b + c) − (a) − (c)) on species richness and biomass [52].

3. Results

3.1. Patterns of Species Richness and Biomass

A total of 5843 individuals belonging to 152 plant species were recorded in the 20 clusters of four plots in the Kiuic site, whereas 10,301 individuals belonging to 144 plant species were recorded in the 28 clusters of four plots in the FCP site. The number of sample plots employed provided adequate representations of species richness at the landscape level, both for the Kiuic and for the FCP sites, as shown in [27,53], through the use of species accumulation curves. Both species richness and AGB were consistently higher in FCP than in Kiuic across sampling areas (Table 1). Species richness also consistently increased as the sampled area increased, whereas AGB failed to show a clear trend with the total area sampled (Table 1).

Table 1. Key statistics of the field data for the tropical dry forests in the Yucatan peninsula.

Site and Number of Plots	Sampled Area (m^2)	Mean (SE)	
		Species Richness	**Biomass (Ton/ha)**
Kiuic (20)	400	22.75 (1.00)	147.20 (14.99)
	1000	23.55 (1.06)	109.71 (9.59)
	2200	46.20 (1.71)	179.16 (8.16)
FCP (28)	400	26.93 (1.09)	270.27 (16.12)
	1000	29.46 (1.22)	376.77 (20.91)
	2200	58.25 (1.35)	351.81 (27.15)

3.2. Effects of Plot Size and Plot Design on Species Richness and Biomass

LiDAR showed a low association with species richness for the 400 and 1000-m^2 sample areas. The R^2 values ranged from 0.18 to 0.19 and the AC values were all negative, indicating no agreement between observed and predicted values of species richness (Figures 2 and 3). However, predictions of species richness that considered the entire cluster of four plots (2200 m^2) improved the accuracy of predictions compared to the analysis using the individual plots. The R^2 values increased to 0.39 and 0.49, respectively, for the FCP and Kiuic sites. Similarly, the AC values were 0.11 and 0.29, respectively, showing an agreement between observed and predicted values of species richness. Multiple regression models consistently retained some LIDAR metrics based on percentiles of height for the 400 and 1000-m^2 sample areas, whereas for the cluster of plots, species richness was mainly explained by the standard deviation of the LiDAR metrics in the four plots in both landscapes (Tables 2 and 3, respectively). These results suggest that a cluster-plot design can improve the accuracy of the predictions of species richness, likely because it is able to capture habitat heterogeneity within the sampled area.
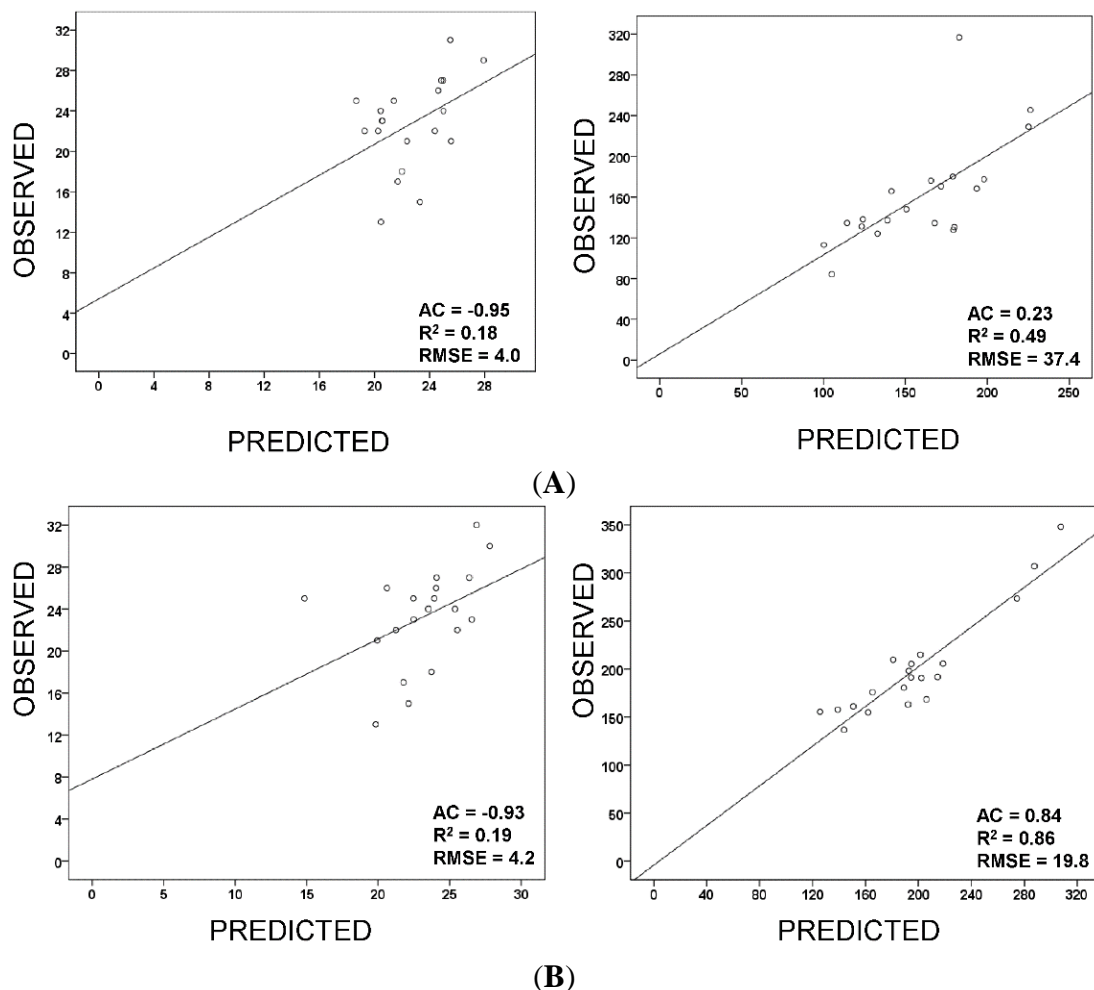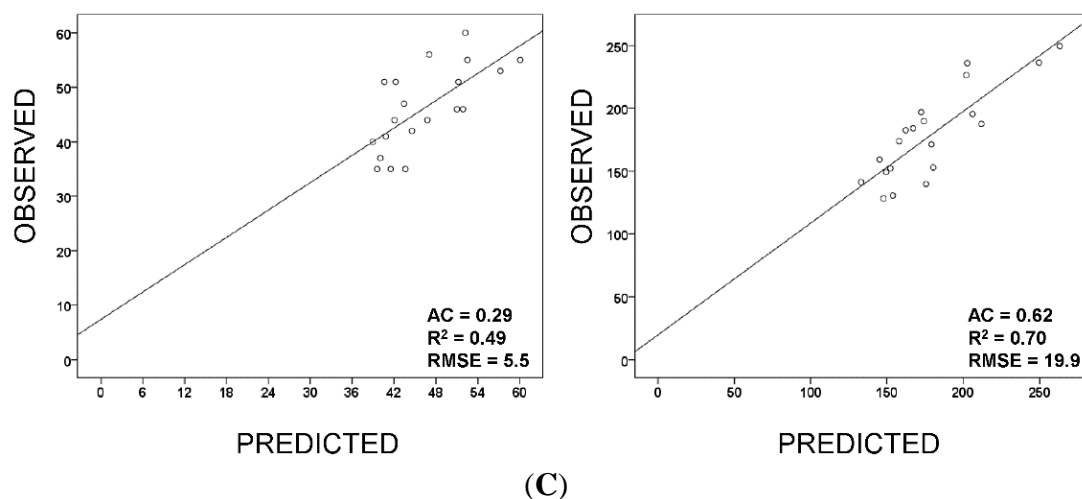
Figure 2. The results of cross-validation analyses used to compare the performance of observed and predicted values of (**left panel**) species richness and (**right panel**) above-ground biomass (MG/ha) in the Kiuic site. (**A**) Plots of 400 m^2; (**B**) Plots of 1000 m^2; (**C**) Plots of 2200 m^2. *R*^{2} is the determination coefficient; RMSE is the root mean square error, and AC is the agreement coefficient.

(**A**)

(**B**)

Figure 2. *Cont.*

(**C**)

Figure 3. The results of cross-validation analyses used to compare the performance of observed and predicted values of (**left panel**) species richness and (**right panel**) above-ground biomass (MG/ha) in the FCP site. (**A**) Plots of 400 m^2; (**B**) Plots of 1000 m^2; (**C**) Plots of 2200 m^2. R^2 is the determination coefficient; RMSE is the root mean square error, and AC is the agreement coefficient.
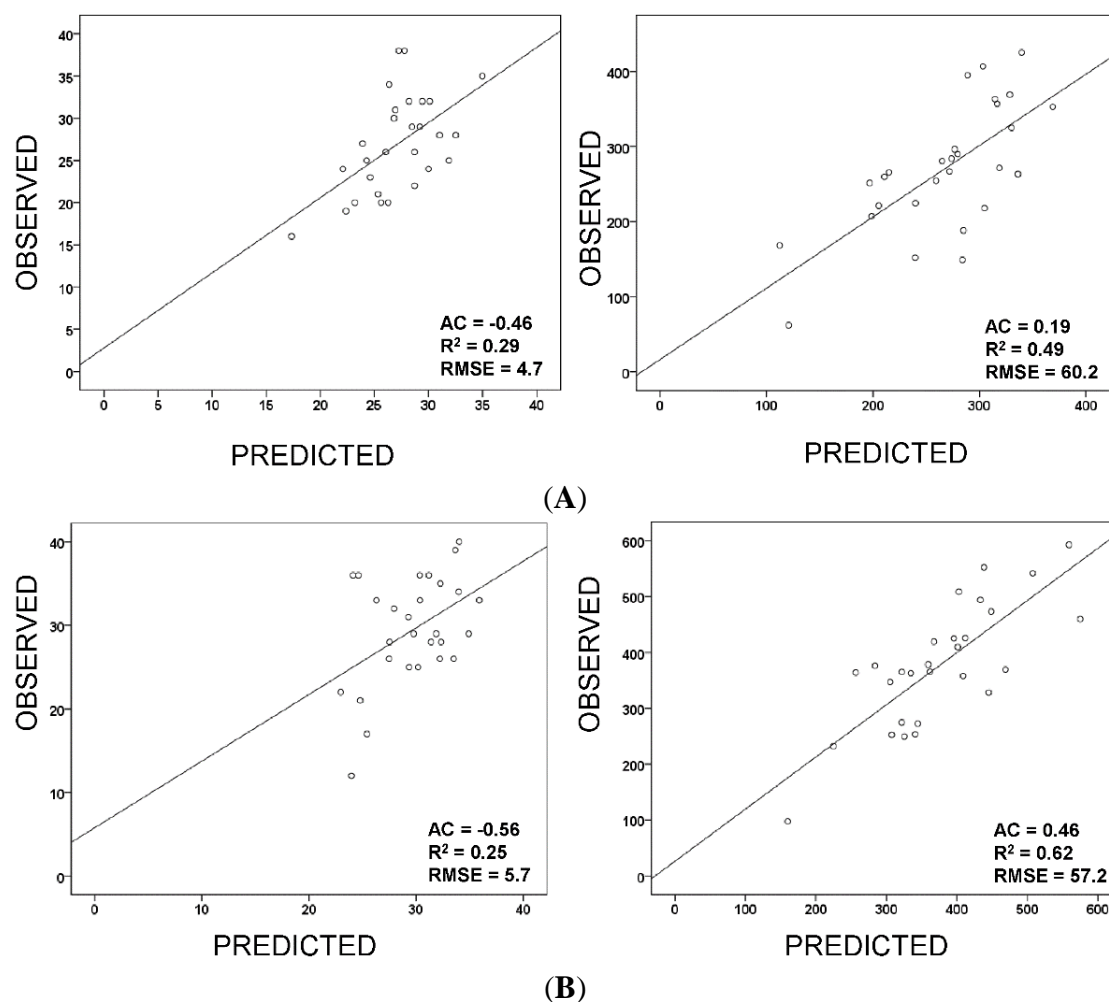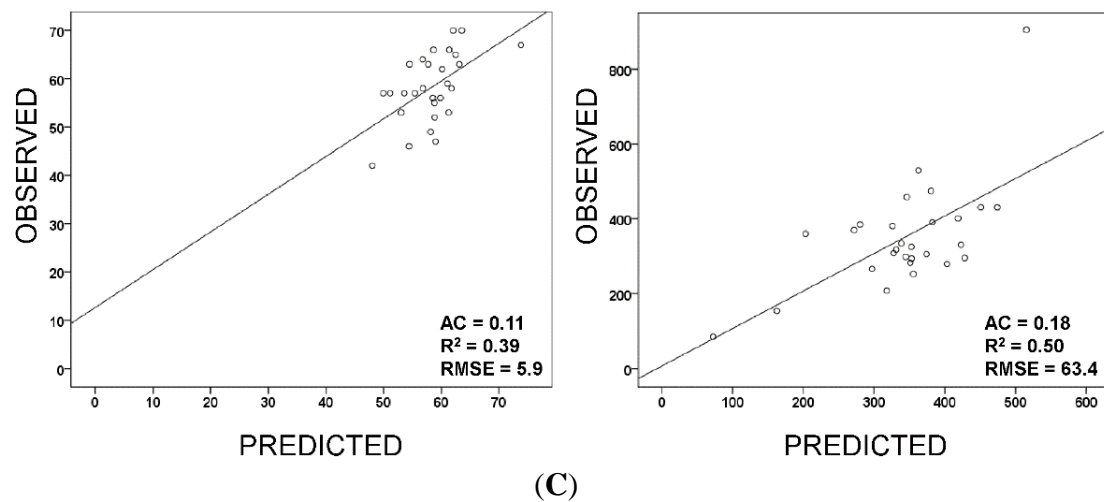

(**A**)

(**B**)

Figure 3. *Cont.*

(**C**)

Table 2. Summary statistics of multiple linear regressions of species richness and biomass with LiDAR metrics in the Kiuic site, using different sample areas.

Dependent Variable	Sampled Area (m²)	Predictor Variables ***	Parameters Estimate (SE)	R² MODEL
Species richness	400	Intercept	12.72 (3.54) *	0.32
		Elev P10	1.98 (0.68) *	
	1000	Intercept	3.30 (6.19)	0.37
		Elev P90	1.55 (0.47) *	
	2200	Intercept	31.81 (9.21) *	0.67
		MEAN Elev Maximum	1.11 (0.65) **	
		STD Elev L4	137.97 (60.92) *	
		STD canopy relief ratio	−216.50 (57.05) *	
Sqrt Biomass	400	Intercept	10.38 (0.51) *	0.59
		Elev L3	−14.33 (2.79) *	
	1000	Intercept	52.52 (7.33) *	0.89
		Percentage of first returns above 4.0	−0.65 (0.09) *	
		(All returns above mean)/(Total firs returns) ×100	0.12 (0.03) *	
		Elev P50	1.64 (0.15) *	
	2200	Intercept	7.20 (2.15) *	0.78
		MEAN Elev P80	0.63 (0.19) *	
		MEAN Elev kurtosis	−19.55 (8.87) *	
		STD Elev P10	1.09 (0.55) **	

Notes: * Variables included in the model with $p < 0.01$; ** variables included in the model with $p < 0.05$; *** Elev P10, P50, P80 and P90 = 10, 50, 80 and 90 of the percentile value of height; Elev Maximum = maximum value of height; Elev L3 and L4 = L3 and L4 moment value of height,; Elev Kurtosis = kurtosis value of height; MEAN and STD = mean and standard deviation values of the metrics for the cluster of four plots.

Table 3. Summary statistics of multiple linear regressions of species richness and biomass with LiDAR metrics in the FCP site, using different sample areas.

Dependent Variable	Sampled Area (m²)	Predictor Variables ***	Parameters Estimate (SE)	R^2 MODEL
Species richness	400	Intercept	45.16 (8.92) *	
		Elev P90	−3.47 (0.87) *	0.41
		(All returns above mean)/(Total firs returns) ×100	0.44 (0.11) *	
	1000	Intercept	69.79 (15.83) *	
		Elev P99	−3.70 (1.01) *	0.38
		(All returns above mean)/(Total firs returns) ×100	0.18 (0.06) *	
	2200	Intercept	75.91 (8.97)	
		MEAN Elev Kurtosis	−8.70 (3.17) *	
		STD Return 4 count above 1.50	0.57 (0.24) *	0.49
		STD Elev MAD mode	8.70 (3.25) *	
		STD Elev P60	−4.60 (2.36) **	
Sqrt Biomass	400	Intercept	5.52 (1.81)*	0.58
		Elev P60	0.98 (0.16) *	
	1000	Intercept	−937.56 (434.85) *	
		Percentage of all returns	0.33 (0.05) *	0.73
		Elev minimum	627.66 (290.50) *	
	2200	Intercept	−3.26 (6.58)	
		MEAN Elev P99	1.31 (0.38) *	0.63
		STD Return 3 count above 1.50	0.03 (0.01) *	
		STD Return 1 count above 1.50	−0.004 (0.002) *	

Notes: * Variables included in the model with $p < 0.01$; ** variables included in the model with $p < 0.05$; *** Elev P60, P90 and P99 = 60, 90 and 99 of the percentile value of height; Elev Minimum = minimum value of height; Elev MAD mode = median of the absolute deviations from the overall mode of height; Elev Kurtosis = kurtosis value of height; MEAN and STD = mean and standard deviation values of the metrics for the cluster of four plots.

We found substantial improvement of the prediction accuracy of ABG as the plot size increased from 400 to 1000 m². The R^2 and AC values increased, respectively, from 0.49 and 0.23 to 0.86 and 0.84 for the Kiuic site (Figure 2) and from 0.49 and 0.19 to 0.62 and 0.46 for the FCP site. Meanwhile the RMSE decreased from 37.4 to 19.8 and from 60.2 to 57.2 for Kiuic and FCP, respectively. However, the accuracy of predictions did not improve for the 2200-m² sample area, compared with the 1000-m² area. On the contrary, the R^2 decreased from 0.86 to 0.70 and from 0.62 to 0.50, respectively, for Kiuic and FCP sites. These results suggest that although the sampled area increased when we considered the cluster of four plots, this also increased the edge effect, leading to greater errors for biomass estimation.

AGB was explained by point density measures and height metrics for the 400 and 1000-m² sample areas, whereas for the cluster of plots, biomass was explained by both the mean and standard deviation of the LiDAR metrics in the cluster of four plots in the Kiuic and FCP sites (Tables 2 and 3).

3.3. Regional Model to Predict Species Richness and Biomass from LiDAR

Our general model, incorporating the sample plots of both sites, explained 46% and 62% of the variation of species richness and biomass, respectively (Figure 4). This general model was built using the best model obtained for each explanatory variable (species richness and biomass) when evaluating the sites individually. Multiple regression models indicated that species richness was mainly explained by the standard deviation of the LiDAR metrics in the cluster of four plots, while above-ground biomass was best explained by point density and height metrics for the plot of 1000 m^2 (Table 4).

Figure 4. The results of cross-validation analyses used to compare the performance of observed and predicted values of (**left**) species richness and (**right**) above-ground biomass (MG/ha) in the entire area (samples from both Kiuic and FCP sites). R^2 is the determination coefficient; RMSE is the root mean square error, and AC is the agreement coefficient.
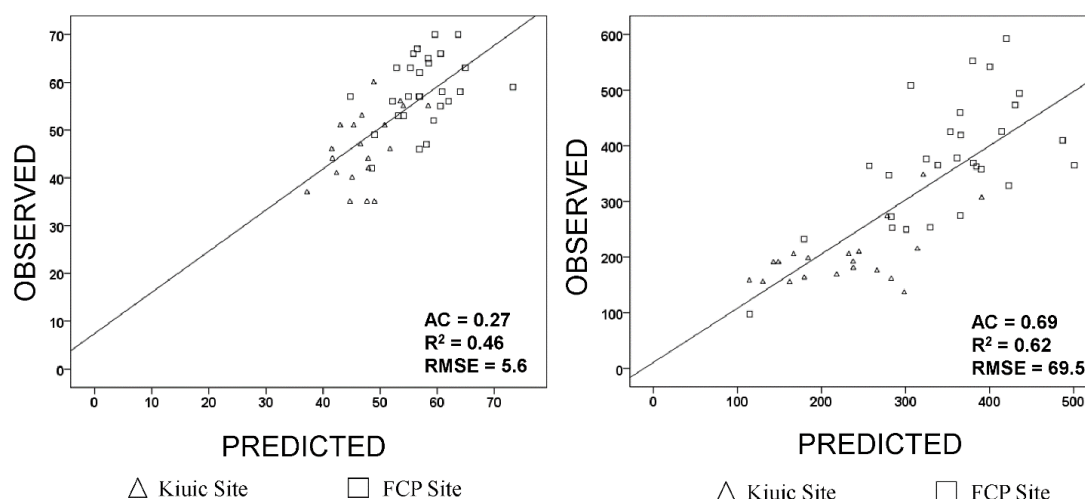

Table 4. Summary statistics of multiple linear regressions of the number of species and biomass with LiDAR metrics for a regional model.

| Dependent Variable	Sampled Area (m^2)	Predictor Variables ***	Parameters Estimate (SE)	R^2 MODEL
Species Richness	2200	Intercept	48.07 (5.54) *	0.59
		MEAN Elev variance	0.60 (0.36) **	
		STD return four count above 1.50	1.02 (0.26) *	
		STD First returns above mean	−0.015 (0.005) *	
		STD Elev variance	1.32 (0.65) *	
		STD First returns above mode	−0.006 (0.003) **	
Sqrt Biomass	1000	Intercept	−1.00 (2.63)	0.71
		Elev P90	1.21 (0.18) *	
		Return three count above 1.50	0.004 (0.001) *	
		Percentage all returns above mean	−0.04 (0.23) **	

Notes: * Variables included in the model with $p < 0.01$; ** variables included in the model with $p < 0.05$; *** Elev P90 = 90 of the percentile value of height; Elev variance = variance value of height; MEAN and STD = mean and standard deviation values of the metrics for the cluster of four plots.

Comparing the variation explained by the general model with that of each site independently, the general model differed from the models obtained at each site by 3%–7% for species richness and by 0–24% for above-ground biomass.

3.4. Variation Partitioning of Species Richness and Biomass

The amount of variation explained by vegetation structure (mean values of LiDAR metrics in the cluster of four plots) and habitat heterogeneity (standard deviation values of LiDAR metrics in the cluster of four plots) differed between species richness and biomass (Figure 4). The total variation explained by the models was consistently higher for biomass (63%–79%) compared to species richness (49%–67%) (Figure 5). Variation partitioning revealed that habitat heterogeneity was the single most important factor, accounting for 42% and 27% of the total variation in species richness, respectively, for the Kiuic and FCP sites. In contrast, for biomass, the combined effect of vegetation structure and habitat heterogeneity was the most important factor. However vegetation structure was more important than habitat heterogeneity, accounting for 16 and 20% of the total variation in stand biomass for the Kiuic and FCP sites, respectively (Figure 5).

Figure 5. Partitioning of the variation in (**left panel**) species richness and (**right panel**) above-ground biomass using mean (Struct) and standard deviation values (Hab_H) of LiDAR metrics in the cluster of four plots. Struct is vegetation structure, and Hab_H is habitat heterogeneity. (**A**) Kiuic Site; (**B**) FCP Site.
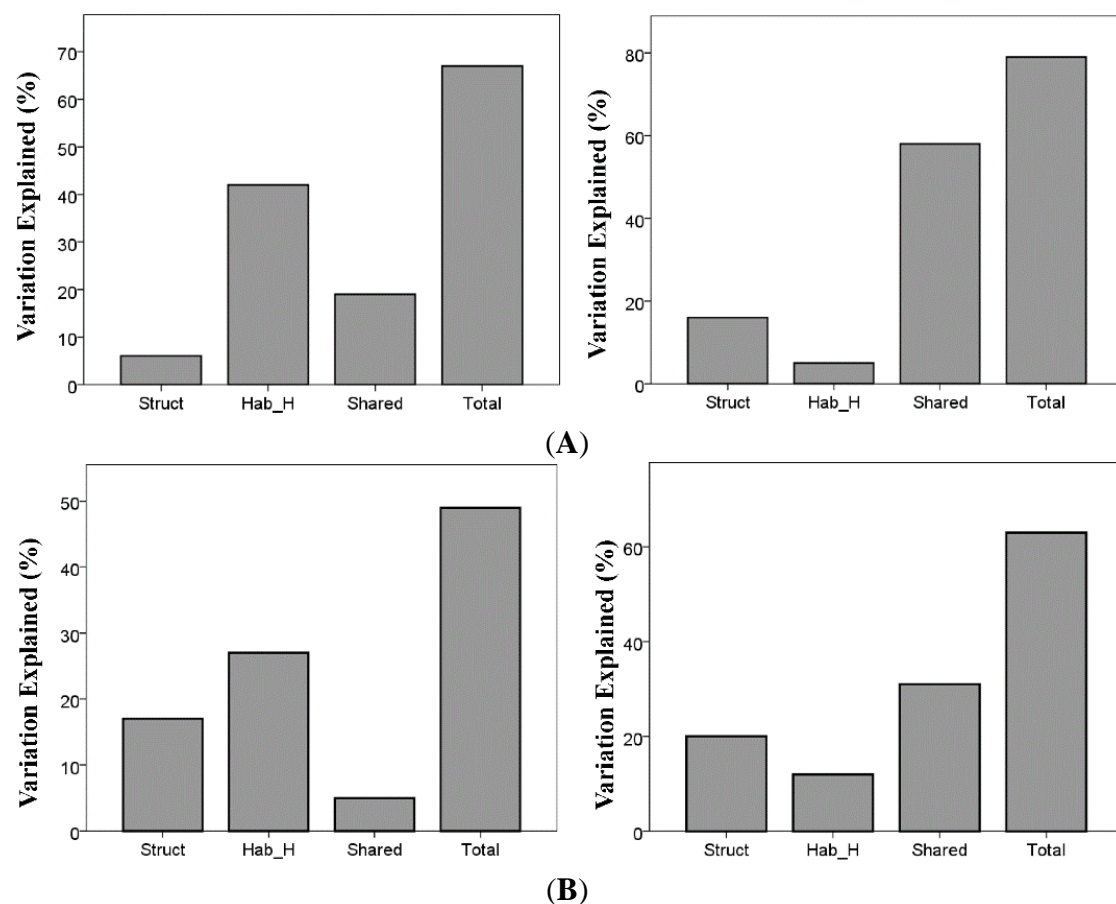
4. Discussion

4.1. Effects of Plot Size and Plot Design on Species Richness and Biomass Estimations

Our results suggest that plot size and plot spatial arrangement, respectively, strongly influence the accuracy of estimates of AGB and species richness obtained from LiDAR. For AGB, the R^2 values in the cross-validation procedure increased from 0.49 to 0.86 and from 0.49 to 0.62 in the Kiuic and FCP sites, respectively, when comparing the 400-m^2 and the 1000-m^2 plot sizes. However, a further increase in the total sample area from 1000 m^2 to 2200 m^2 did not result in an increase of the R^2 values. Thus, our results indicate that plot size, rather than total sample area, is critical for improving LiDAR-derived biomass estimations. There are several reasons why increasing the field plot size should be considered when estimating ABG from airborne LiDAR data. First, there are errors in the field estimates of AGB when small plots are used in inventories, due to the potential over- or under-representation of rare large trees in small areas [32]. Second, increasing the plot size allows more overlap between ground plots and LiDAR data, thereby reducing the potential errors associated with inaccurate GPS locations collected at the center of the plots [54]. Third, a larger plot has a lower perimeter-to-area ratio, resulting in fewer potential edge-related issues, and more accurate LiDAR metrics describing vertical structure [33]. The latter point can also help explain why increasing the total sample area to 2200 m^2 resulted in a decrease, rather than an increase, in the R^2 values compared to the 1000-m^2 plots, since this increase in area was accompanied by a substantial increase in the perimeter and the potential errors associated with edge effects [33], when comparing a cluster of four plots to a single large plot.

Although implementing larger plot sizes increases the accuracy of biomass predictions, this also increases the cost. Thus, for determining an optimal plot size for mapping biomass using airborne LiDAR, it is necessary to quantify acceptable levels of error and cost. The study by [54] evaluated different plot sizes (314, 707, 1257 and 1964 m^2) for estimating biomass from LiDAR and showed that R^2 increased from 0.82 to 0.88, with an asymptotic non-linear trend, suggesting that little improvement is expected for plots larger than 1257 m^2. Our predictive R^2 values are comparable to those of [54], providing reliable estimates for the AGB, and suggest that it is not necessary to increase the plot size much beyond the size of the larger plot we used.

In contrast to estimating biomass, there was no benefit to increasing the plot size from 400 to 1000 m^2 to model species richness; instead, the ability to capture the local variability of the landscape using a cluster of four plots proved to be far more efficient. Habitat heterogeneity has been frequently associated with species richness. Several authors have reported strong associations between spectral heterogeneity (as a proxy of habitat heterogeneity) and species richness [28–30,55]. Using the variance of height on LiDAR metrics within plots of 400 and 1000 m^2, the R^2 values were increased from 0.19 to 0.49 and from 0.25 to 0.39 in the Kiuic and FCP sites, respectively, when comparing the 1000-m^2 plot and the cluster of four plots (2200 m^2), suggesting that habitat heterogeneity is a scale-dependent proxy of species richness [56]. In other words, the inclusion of different habitat conditions at a larger scale (cluster of four plots over an area of 1.0 ha) may reveal a relationship between species richness and habitat heterogeneity that is not captured at the local, single-plot scale [56]. For example, the variety of land cover classes within clusters of four plots may be

correlated with species richness, reflecting different topographic conditions and stages of forest succession, as also found by [27,57], respectively. Both studies were performed in tropical dry forests of the Yucatan Peninsula.

Besides plot design, the type of species diversity measurement used may also affect the estimation accuracy. In particular, the Shannon diversity index, which takes into account relative species abundance, may improve the ability to detect local species diversity by remotely sensed data, as suggested by [58]. This is mainly due to the fact that this index is less affected by the presence of rare species than species richness. However, the occurrence of rare species is one of the most frequently used criterion for selecting and prioritizing habitat sites for preservation [59,60]. This was the reason why we selected the number of species as our estimate of alpha diversity.

Species richness and biomass were shown to be strongly related to LiDAR data in the studied area. Specifically, regression results suggest that species richness is mainly related to the standard deviation of LiDAR metrics in the four plots: standard deviation of the elevation L4, canopy relief ratio and elevation MAD (median of absolute deviations) mode. The positive coefficient of the standard deviation of LiDAR metrics indicates that greater habitat heterogeneity (more variability of topography and tree height) promotes a greater number of species to be present within the area of the clusters of plots. On the other hand, above-ground biomass was mainly related to point density (all returns above mean/total first returns $\times 100$, the percentage of all returns) and height metrics (elevation P50, elevation minimum). We found a positive relationship between point density and height metrics in most of the regression models for estimating biomass, meaning that biomass increases with taller trees and more canopy cover.

4.2. Regional Model to Predict Species Richness and Biomass

The accuracy of the predictions of species richness and biomass from LiDAR metrics varied between the studied sites. The R^2 values were consistently higher for the Kiuic site (0.49 and 0.86 for species richness and AGB, respectively), compared to FCP (0.39 and 0.62). Different studies have shown differences in associations between LiDAR metrics and forest structure parameters across sites, such as that of Drake *et al.* [14] comparing a seasonal moist tropical forest in Panama and a wet tropical forest in Costa Rica. These differences could be due to a combination of between-site differences in forest structure, resulting from environmental conditions, land use changes and the limited precision of LiDAR footprints. The denser, taller and multilayered canopy in the FCP site may represent more difficult conditions than those of the Kiuic site for acquiring accurate GPS positions [61]. Our general model explained 46% and 62% of the variation of species richness and biomass, respectively, and showed some differences compared with the individual models for each site. Between-site variations in the relationships among LiDAR metrics and species richness and biomass may be related to the level of variation explained by the general *vs.* site-specific models. However, our biomass regional models have a similar accuracy ($R^2 = 0.62$) when compared with other regional tropical forest models ($R^2 = 0.56$ to 0.80) [34,62].

4.3. Factors Related to Species Richness and Biomass Estimations

The results of this study suggest that the standard deviations of LiDAR metrics can be used as indicators of habitat heterogeneity within stands and, therefore, of species richness [31]. A greater variety of canopy and subcanopy LiDAR returns reflects the greater variability of topography and tree height (possibly linked to forest succession) and conditions that provide opportunities for different species to be present within the sample area. The question remains if there is a limit to how many species can be detected with this approach. For the range of conditions that exist for our two studied sites, where species richness is relatively high, the current approach seems adequate.

The ability to predict biomass and species richness, however, is based on the combination of two main aspects that contribute to the predictive value of the models. The first aspect is related to the structural attributes of vegetation, *i.e.*, canopy cover, tree diameter, basal area and tree height, which are indicators of the structural complexity of the forest [63] and are related to the species richness and biomass present in the stand, since this fosters different groups of species [26] and biomass levels [35,36]. The second aspect is based on habitat heterogeneity, *i.e.*, the heterogeneity within and among stands caused by disturbance and topography. Although this second aspect is considered a relevant surrogate of biodiversity [31], habitat fragmentation substantially reduces forest biomass [64], possibly due to enhanced tree mortality and the proliferation of disturbance-adapted species, such as lianas [65]. Moreover, basal area and tree height differ between hills and flat areas [27].

An important finding of our study is that the structural complexity of vegetation and habitat heterogeneity, measured through airborne LiDAR data, were significant predictors of species richness and biomass. A large portion (27%–42%, see Figure 5) of the variation in species richness can be attributed exclusively to habitat heterogeneity, whereas a much smaller fraction of the variation (5%−19%) was explained solely by the structure of vegetation. An opposite pattern to that of species richness was found for biomass, since the variation in biomass was mainly explained by vegetation structure (16%−20%) compared with habitat heterogeneity (5%−12%). However, shared variation (31%−58%) was the most important determinant of the biomass. These results are consistent with recent findings in tropical forests showing that biomass is mainly explained by the structure of vegetation [35,36], whereas species richness is strongly affected by factors not directly measured by LiDAR and related to the variability of habitat conditions, such as soil fertility and other environmental components [37,66].

5. Conclusions

We presented in this study a potentially useful approach for mapping the number of species and biomass based on LiDAR data (as surrogates of environmental factors). The results showed that increasing plot size, rather than total sample area, is better for LiDAR-derived biomass estimations. In contrast, the inclusion of different habitat conditions (as in the case of the clusters of four plots over an area of 1.0 ha) allows better species richness estimations. We also showed that vegetation structure and habitat heterogeneity, represented with LiDAR data, may contribute significantly to our understanding of how diversity and AGB are maintained in any given area. An important finding of our study is that a large portion of the variation in species richness can be exclusively attributed to the

habitat heterogeneity. An opposite pattern was found for biomass, since the variation of biomass was mainly explained by vegetation structure.

Finally, a strong limitation faced by conservation biologists and managers of natural resources is the lack of continuous information concerning species distribution patterns [67]. In addition to providing guidance regarding the selection and effectiveness of protected natural areas, precise biodiversity maps produced by accurate modeling can also help to assess species responses to global climate change. In the same way, mapping the spatial distribution of above-ground biomass through remote sensing will translate into better estimates of carbon stocks at broad scales, a requirement of a deforestation-reduction program, such as a REDD+ (Reducing Emissions −of green-house gasses− from Deforestation and Degradation, plus enhancing forest carbon stocks). We have shown that models of plant biodiversity and biomass can be derived from small footprint LiDAR at both local and regional scales in the tropical dry forest in the Yucatan Peninsula. The main limitation to expanding these models to produce regional or national maps in Mexico is a lack of wall-to-wall LiDAR data. Nonetheless, recent LiDAR acquisitions by the NASA Goddard's LiDAR, Hyperspectral and Thermal (G-LiHT) team and more acquisitions planned by the United States Agency for International Development Mexican REDD+ (USAID M-REDD+) program will fill more gaps and should provide more opportunities for future research in this area. The challenge remains to scale these estimations with other wall-to-wall data, such as Landsat, MODIS or RapidEye.

Author Contributions

José Luis Hernández-Stefanoni conceived of the research and designed the experiments. José Luis Hernández-Stefanoni, Juan Manuel Dupuy and Fernando Tun-Dzul analyzed the data. José Luis Hernández-Stefanoni, Juan Manuel Dupuy, Kristofer D. Johnson, Richard Birdsey and Alicia Peduzzi wrote the paper. Juan Pablo Caamal-Sosa, Gonzalo Sánchez-Santos and David López-Merlín processed the field data samples. All authors shared equally in the editing of the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Thomson, A.; Calvin, K.; Chini, L.; Hurtt, G.; Edmonds, J.; Ben Bond-Lamberty, B.; Frolking, S.; Wise, M.A.; Janetos, A.C. Climate mitigation and the future of tropical landscapes. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 19633–19638.
2. Fearnside, P.M.; Laurance, W.F. Tropical deforestation and greenhouse gas emissions. *Ecol. Appl.* **2004**, 14,982–986.
3. Le Quéré, C.; Peters, G.P.; Andres R.J.; Andrew, R.M.; Boden, T.; Ciais, P.; Friedlingstein, P.; Houghton, R.A.; Marland, G.; Moriarty, R.; *et al.* Global carbon budget 2013. *Earth Syst. Sci. Data Discuss.* **2013**, *6*, 689–760.
4. Fahrig, L. Effects of habitat fragmentation on biodiversity. *Annu. Rev. Ecol. Evol. Syst.* **2003**, *34*, 487–515.
5. Gibbs, H.K.; Brown, S.; Niles, J.O.; Foley, J.A. Monitoring and estimating tropical forest carbon stocks: Making REDD a reality. *Environ. Res. Lett.* **2007**, *2*, doi:10.1088/1748-9326/2/4/045023.
6. Gillespie, T.W.; Foody, G.M.; Rocchini, D.; Giorgi, A.P.; Saatchi, S. Measuring and modelling biodiversity from space. *Prog. Phys. Geogr.* **2008**, *32*, 203–221.
7. St-Louis, V.; Pidgeon, A.M.; Clayton, M.K.; Locke, B.A.; Bash, D.; Radeloff, V.C. Satellite image texture and a vegetation index predict avian biodiversity in the Chihuahuan Desert of New Mexico. *Ecography* **2009**, *32*, 468–480.
8. Turner, W.; Spector, S.; Gardiner, N.; Fladeland, M.; Sterling, E.; Steiniger, M. Remote sensing for biodiversity science and conservation. *Trends Ecol. Evol.* **2003**, *18*, 306–314.
9. Hernández-Stefanoni, J.L.; Gallardo-Cruz, J.A.; Meave, J.A.; Dupuy, J.M. Combining geostatistical models and remotely sensed data to improve tropical plant richness mapping. *Ecol. Indic.* **2011**, *11*, 1046–1056.
10. Sanchez-Azofeifa, G.A.; Castro-Esau, K.L.; Kurz, W.A.; Joyce, A. Monitoring carbon stocks in the tropics and the remote sensing operational limitations: From local to regional projects. *Ecol. Appl.* **2009**, *19*, 480–494.
11. Nagendra, H.; Rocchini, D.; Ghate, R.; Sharma, B.; Pareeth, S. Assessing plant diversity in a dry tropical forest: Comparing the utility of Landsat and IKONOS satellite images. *Remote Sens.* **2010**, *2*, 478–496.
12. Helmer, E.; Brandeis, T.; Lugo, A.; Kennaway, T. Factors influencing spatial pattern in tropical forest clearance and stand age: Implications for carbon storage and species diversity. *J. Geophys. Res.* **2008**, *113*, doi:10.1029/2007JG000568.
13. Næsset, E.; Okland, T. Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve. *Remote Sens. Environ.* **2002**, *79*, 105–115.
14. Drake, J.B.; Knox, R.G.; Dubayah, R.O.; Clark, D.B.; Condit, R.; Blair, J.B.; Hofton, M. Above-ground biomass estimation in closed canopy Neotropical forests using LiDAR remote sensing: Factors affecting the generality of relationships. *Glob. Ecol. Biogeogr.* **2003**, *12*, 147–159.

15. Nelson, R.; Ranson, K.J.; Sun, G.; Kimes, D.S.; Kharuk, V.; Montesano, P. Estimating Siberian timber volume using MODIS and ICEsat/GLAS. *Remote Sens. Environ.* **2009**, *113*, 691–701.

16. Simonson, W.D.; Allen, H.D.; Coomes, D.A. Use of an airborne lidar system to model plant species composition and diversity of mediterranean oak forests. *Conserv. Biol.* **2012**, *26*, 840–850.

17. Leutner, B.F.; Reineking, B.; Müller, J.; Bachmann, M.; Beierkuhnlein, C.; Dech, S.; Wegmann, M. Modelling forest α-diversity and floristic composition—On the added value of LiDAR plus hyperspectral remote sensing. *Remote Sens.* **2012**, *4*, 2818–2845.

18. Lefsky, M.A.; Harding, D.J.; Keller, M.; Cohen, W.B.; Caraba-jal, C.C.; Espirito-Santo, F.D.; Hunter, M.O.; de Oliveira, R. Estimates of forest canopy height and above-ground biomass using ICESat. *Geophys. Res. Lett.* **2005**, *32*, doi:10.1029/2005GL023971.

19. Lefsky, M.A.; Cohen, W.B.; Spies, R.A. An evaluation of alternate remote sensing products for forest inventory, monitoring, and mapping of Douglas-fir forests in western Oregon. *Can. J. For. Res.* **2001**, *31*, 78–87.

20. Chave, J.; Andalo, C.; Brown, S.; Cairns, M.A.; Chambers, J.Q.; Eamus, D.; Folster, H.; Fromard, F.; Higuchi, N.; Kira, T.; *et al.* Tree allometry and improved estimation of carbonstocks and balance in tropical forests. *Oecologia* **2005**, *145*, 87–99.

21. Asner, G.P.; Hughes, R.F.; Varga, T.A.; Knapp, D.E.; Kennedy-Bowdoin, T. Environmental and biotic controls over aboveground biomass throughout a tropical rain forest. *Ecosystems* **2009**, *12*, 261–278.

22. Goetz, S.; Steinberg, D.; Dubayah, R.; Blair, B. Laser remote sensing of canopy habitat heterogeneity as a predictor of bird species richness in an eastern temperate forest, USA. *Remote Sens. Environ.* **2007**, *108*, 254–263.

23. Muller, J.; Brandl, R. Assessing biodiversity by remote sensing in mountainous terrain: The potential of LiDAR to predict forest beetle assemblages. *J. Appl. Ecol.* **2009**, *46*, 897–905.

24. Vierling, K.T.; Bassler, C.; Brandl, R.; Vierling, L.A.; Weib, I.; Muller, J. Spinning a laser web: Predicting spider distributions using LiDAR. *Ecol. Appl.* **2011**, *21*, 577–588.

25. Whitmore, T.C. Canopy gaps and the mayor groups of forest trees. *Ecology* **1989**, *70*, 536–538.

26. Lebrija-Trejos, E.; Pérez-García, E.A.; Meave, J.A.; Bongers, F.; Poorter, L. Functional traits and environmental filtering drive community assembly in a species-rich tropical landscape. *Ecology* **2010**, *91*, 386–398.

27. Dupuy, J.M.; Hernández-Stefanoni, J.L.; Hernández-Juárez, R.A.; Tetetla-Rangel, E.; López-Martínez, J.O.; Leyequién-Abarca, E.; Tun-Dzul, F.J.; May-Pat, F. Patterns and correlates of tropical dry forest structure and composition in a highly replicated chronosequence in Yucatan, Mexico. *Biotropica* **2012**, *44*, 151–162.

28. Gillespie, T.W. Predicting woody-plant species richness in tropical dry forests: A case study from South Florida, USA. *Ecol. Appl.* **2005**, *15*, 27–37.

29. Rocchini, D. Effects of spatial and spectral resolution in estimating ecosystem α-diversity by satellite imagery. *Remote Sens. Environ.* **2007**, *111*, 423–34.

30. Hernández-Stefanoni, J.L.; Gallardo-Cruz, J.A.; Meave, J.A.; Rocchini, D.; Bello-Pineda, J.; López-Martínez, J.O. Modelling α and β-diversity in a tropical forest from remotely sensed and spatial data. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *19*, 359–368.

31. Palmer, M.W.; Earls, P.G.; Hoagland, B.W.; White, P.S.; Wohlgemuth, T. Quantitative tools for perfecting species lists. *Environmetrics* **2002**, *13*, 121–137.

32. Chave, J.; Condit, R.; Aguilar, S.; Hernandez, A.; Lao, S.; Perez, R. Error propagation and scaling for tropical forest biomass estimates. *Philos. Trans. R. Soc. Lond. B* **2004**, *359*, 409–420.

33. Mascaro, J.; Detteo, M.; Asner, G.P.; Muller-Landau, H.C. Evaluating uncertainty in mapping forest carbon with airborne LiDAR. *Remote Sens. Environ.* **2011**, *115*, 3770–3774.

34. Vincent, G.; Sabatier, D.; Blanc, L.; Chave, J.; Weissenbacher, E.; Pelissier, R.; Fonty, E.; Molino, J.-F.; Couteron, P. Accuracy of small footprint airborne LiDAR in its predictions of tropical moist forest stand structure. *Remote Sens. Environ.* **2012**, *125*, 23–33.

35. Read, L.; Lawrence, D. Recovery of biomass following shifting cultivation in dry tropical forests of the Yucatán. *Ecol. Appl.* **2003**, *13*, 85–97.

36. Eaton, J.M.; Lawrence, D. Loss of carbon sequestration potential after decades of shifting cultivation in the Southern Yucatán. *For. Ecol. Manag.* **2009**, *258*, 949–958.

37. Chazdon, R.L.; Letcher, S.G.; Breugel, M.V.; Martínez-Ramos, M.; Bongers, F.; Finegan, B. Rates of change in tree communities of secondary Neotropical forests following major disturbances. *Philos. Trans. R. Soc. B* **2007**, *362*, 273–289.

38. Flores, J.; Espejel, I. *Tipos de Vegetación de la Península de Yucatán*; Universidad Autónoma de Yucatán, Dirección General de Extensión: Yucatán, México, 1994.

39. Cabrera-Cano, E.; Sousa-Sánchez, M.; Téllez-Valdés, O. *Imágenes de la Flora Quintanarroense*; Centro de Investigaciones de Quintana Roo: Mexico City, Mexico, 1982.

40. Manuales Para el Levantamiento de Datos en Campo del Inventario Nacional Forestal y de Suelos. Available online: http://www.cnf.gob.mx:8080/snif/portal/infys/temas/documentos-metodologicos (accessed on 13 May 2014).

41. Gotelli, N.J.; Colwell, R.K. Quantifying biodiversity: Procedures and pitfalls in the measurement and comparison of species richness. *Ecol. Lett.* **2001**, *4*, 379–391.

42. Cairns, M.A.; Olmsted, I.; Granados, J.; Argaez, J. Composition and aboveground tree biomass of a dry semi-evergreen forest on Mexico's Yucatan Peninsula. *For. Ecol. Manag.* **2003**, *186*, 125–132.

43. Urquiza-Haas, T.; Dolman, P.M.; Peres, C.A. Regional scale variation in forest structure and biomass in the Yucatan Peninsula, Mexico: Effects of forest disturbance. *For. Ecol. Manag.* **2007**, *247*, 80–90.

44. Hughes, R.F.; Kauffman, J.B.; Jaramillo-Luque, V.J. Biomass, carbon, and nutrient dynamics of secondary forests in a humid tropical region of México. *Ecology* **1999**, *80*, 1892–1907.

45. Chave, J.; Condit, R.S.; Lao, S.; Caspersen, J.; Foster, R.B.; Hubbell, S.P. Spatial and temporal variation of biomass in a tropical forest: Results from a large census plot in Panama. *J. Ecol.* **2003**, *91*, 240–252.

46. Frangi, J.L.; Lugo, A.E. Ecosystem dynamics of a sub-tropical floodplain forest. *Ecol. Monogr.* **1985**, *55*, 351–369.

47. CartoData. Available online: http://www.cartodata.com (accessed on 13 May 2014).

48. Software for LIDAR Data Analysis and Visualization. Available online: http://www.fs.fed.us/eng/rsac/fusion (accessed on 13 May 2014).

49. Zar, J.H. *Biostatistical Analysis*; Prenctice Hall: Upper Saddle River, NJ, USA, 1999.

50. Miller, D.M. Reducing transformation bias in curve fitting. *Am. Stat.* **1984**, *38*, 124–126.

51. Ji, L.; Gallo, K. An agreement coefficient for image comparison. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 823–833.

52. Borcard, D.; Legendre, P.; Avois-Jacquet, C.; Tuomisto, H. Dissecting the spatial structure of ecological data at multiple scales. *Ecology* **2004**, *85*, 1826–1832.

53. Hernández-Stefanoni, J.L.; Ponce-Hernández, R. Mapping the spatial distribution of plant diversity indices using multi-spectral satellite image classification and field measurements. *Biodivers. Conserv.* **2004**, *13*, 2599–2621.

54. Frazer, G.W.; Magnussen, S.; Wulder, M.A.; Niemann, K.O. Simulated impact of sample plot size and co-resgistration error on the accuracy and uncertainty of LiDAR-derived estimates of forest stand biomass. *Remote Sens. Environ.* **2001**, *111*, 636–649.

55. Levin, N.; Shmida, A.; Levanoni, O.; Tamari, H.; Kark, S. Predicting mountain plant richness and rarity from space using satellite-derived vegetation indices. *Divers. Distrib.* **2007**, *13*, 1–12.

56. Nagendra, H. Using remote sensing to assess biodiversity. *Int. J. Remote Sens.* **2001**, *22*, 2377–2400.

57. López-Martínez, J.O.; Hernández-Stefanoni, J.L.; Dupuy, J.M.; Meave, J.A. Partitioning the variation of woody plant β-diversity in a landscape of secondary tropical dry forests across spatial scales. *J. Veg. Sci.* **2013**, *24*, 33–45.

58. Oldeland, J.; Wesuls, D.; Rocchini, D.; Schmidt, M.; Jrgens, N. Does using species abundance data improve estimates of species diversity from remotely sensed spectral heterogeneity? *Ecol. Indic.* **2010**, *10*, 390–396.

59. Prendergast, J.R.; Quinn, R.M.; Lawton, J.H.; Eversham, B.C.; Gibbons, D.W. Rare species, the coincidennce of diversity hotspots and conservation strategies. *Nature* **1993**, *365*, 335–337.

60. Rossi, E.; Kuitunen, M. Ranking of habitats for the assessment of ecological impact in land use planning. *Biol. Conserv.* **1996**, *77*, 227–234.

61. D'Oliveira, M.V.N.; Reutebuch, S.E.; McGaughey, R.J.; Andersen, H.E. Estimating forest biomass and identifying logging areas using airborne scanning LiDAR in Antimary State Forest, Acre State, Western Brazilian Amazon. *Remote Sens. Environ.* **2012**, *124*, 479–491.

62. Asner, G.P.; Mascaro, J.; Muller-Landau, H.C.; Vieilledent, G.; Vaudry, R.; Rasamoelina, M.; Hall, J.S.; van Breugel, M. A universal airborne LiDAR approach for tropical forest carbon mapping. *Oecologia* **2012**, *168*, 1147–1160.

63. McElhinny, C.; Gibbons, P.; Bracks, C.; Bauhus, J. Forest and wood stand structural complexity: Its definition and measurement. *For. Ecol. Manag.* **2005**, *218*, 1–24.

64. Nascimento, H.E.M.; Laurence, W.F. Biomass dynamics in Amazonian forest fragments. *Ecol. Appl.* **2004**, *14*, 127–138.

65. D'Angelo, S.A.; Andrade, A.C.S.; Laurance, S.G.; Laurance, W.F.; Mesquita, R.C.G. Inferred causes of tree mortality in fragmented and intact Amazonian forest. *J. Trop. Ecol.* **2004**, *20*, 243–246.

66. Jones, M.M.; Tuomisto, H.; Borcard, D.; Legendre, P.; Clark, D.B.; Olivas, P.C. Explaining variation in tropical plant community composition: Influence of environmental and spatial data quality. *Oecologia* **2008**, *155*, 593–604.

67. Araújo, M.B.; Guisán, A. Five (or so) challenges for species distribution modelling. *J. Biogeogr.* **2006**, *33*, 1677–1688.

Appendix

Table A1. List of LiDAR metric used in this study.

Category	Metric
Height statistics	Minimum
	Maximum
	Mean
	Median (output as 50th percentile)
	Mode
	Standard deviation
	Variance
	Coefficient of variation
	Interquartile distance
	Above mean
	Above mode
	Skewness
	Kurtosis
	AAD (average absolute deviation)
	L-moments (L1, L2, L3, L4)
	L-moment skewness
	L-moment kurtosis
	(1st, 5th, 10th, 20th, 25th, 30th, 40th, 50th, 60th, 70th, 75th, 80th, 90th, 95th, 99th percentiles)
Canopy density	Total number of returns
	Count of returns by return number (1, 2, 3, 4, 5, 6, 7, 8, 9, other)
	Percentage of first returns above a specified height (canopy cover estimate)
	Percentage of first returns above the mean height/elevation
	Percentage of first returns above the mode height/elevation
	Percentage of all returns above a specified height
	Percentage of all returns above the mean height/elevation
	Percentage of all returns above the mode height/elevation
	Number of returns above a specified height/total first returns $\times 100$
	Number of returns above the mean height/total first returns $\times 100$
	Number of returns above the mode height/total first returns $\times 100$
	Number of first returns above mean
	Number of first returns above mode
	Number of returns above mean
	Number of returns above MODE
	Total number of 1st returns
	Total number of returns