**Title.** Land-use history shapes tree aboveground biomass patterns in tropical montane cloud forest landscapes

**Authors**

**Affiliations**

**Keywords**

**Abstract**

**Introduction**

How does small-scale farming shapes regional patterns of AGB in tropical montane cloud forest across environmental gradients?

1. How much AGB TMCFs hold and where is it allocated?
2. How does land-use shape AGB spatial patterns along an environmental gradient?
3. What is the relationship between AGB and tree diversity in TMCF?

Acronyms: TMCF, FI

**Methods**

*Study Area*

We delimited a study area following a TMCF regionalization conducted by Toledo-Aceves *et al.* in 2011 based on geomorphology, forest cover, watershed margins, rivers, and cultural differences (such as presence of indigenous groups). We focused on the Northern Mountains of Oaxaca (NMO), a region that harbors some of the most biodiverse forests in Mexico including the largest and most continuous TMCF, and where forest conservation is considered a critical priority (Toledo-Aceves et al., 2011). In the NMO, TMCF are found on hillslopes and humid ravines with frequent fog and drizzle. The mountains impose an environmental gradient that goes from warmer to cooler temperatures as elevation increases. Soils in these forests usually develop from the weathering of metamorphic rocks and volcanic outcrops, they tend to be deep, and rich in clay and organic matter (Torres, 2004). We defined the distribution of TMCF within NMO using the official map of vegetation and land-use series V published by the National Institute of Statistic and Geography (INEGI) in 2013 (CITA) (Figure 1).

*Data Collection and Processing*

We gathered and integrated information on forest structure and composition, environment, topography, and land-use, from different sources into a single dataset. The main data source for this work is the publicly available Mexican National Forest Inventory database, which contains information on forest structure and composition, as well as forest disturbance. From this database we estimated forest structural attributes, AGB, and land-use variables. To complement this data, we obtained information on mean annual precipitation and temperature from WorldClim (CITA). Lastly, we retrieved topographical information from NASA’s Shuttle Radar Topography Mission digital elevation data. A detailed description of the foregoing variables can be found below (see also Table S1 for a summary of the variables used in this study).

*Forest Inventory Sites.* FI data collection was carried out between 2009 and 2014 following a hierarchical nested sampling design with 1-ha circular sites as the main sampling unit. All sites were established 25 km apart from each other in a grid-like fashion. Within each site, four circular plots of 400 m2 were established. One in the center of the site, and the other three in a north, southeast, and southwest direction, respectively, at 45.14 m from the central plot. All trees, lianas, shrubs, palm trees and ferns within the plots with a diameter at breast height (DBH) larger than 7.5 cm were sampled for height, DBH, basal area (BA), and taxonomic identification. Information about the geographic location, vegetation type, and land ownership of each site was also documented (CONAFOR, 2018). For selecting FI sites relevant to our study, we performed a spatial intersection in QGIS 3.16 between FI sites and the NMO using FI sites’ geographic coordinates and the NMO shapefile we acquired from the National Commission of Biodiversity (CONABIO) GeoPortal (CITAS).

*Forest Structural Attributes and Aboveground Biomass Estimation.* Based on FI raw data, we derived three structural attributes at plot level that were then averaged by site: (1) stem density, *i.e.*, the number of trees per hectare; (2) basal area, defined as the sum of the cross-sectional surface area of trees per hectare; and (3) Lorey’s height, which is a measure of forest stand height weighted by its basal area.

To estimate AGB, we first calculated each individual’s AGB using 47 different allometric equations (Table S2). Whenever allometric equations were available in the scientific literature at species or genus level we would use them. Otherwise, we used the following generic allometric equation (eq. 1) developed by Chave *et al.* (2014(2015?)) for tropical trees based on tree wood density (*ρ*), height (*H*) and DBH (*D*):

|  |  |
| --- | --- |
|  | (1) |

We corrected taxonomic names collating a list of species with the Taxonomic Name Resolution Service and searched for the wood density value of each species or its closest relative in global wood density databases using the R package BIOMASS (Réjou‐Méchain, 2017).

We calculated AGB per plot adding up the biomass of each individual tree and AGB per site averaging plot’s AGB. There is always some uncertainty inherent to upscaling biomass estimates from trees to forest stands that arises from the propagation of errors in field data collection, allometric equations, wood density estimates, and forest variation. To account for this uncertainty, we estimated plot AGB standard deviation following error propagation through a Monte Carlo scheme, assuming 95% of the samples have a low diameter error and the remaining 5% a high diameter error (close to 5 cm), and a height error of 10%, as suggested in Chave et al. (2004). To estimate site’s AGB error, we assumed standard error independence and used the following equation (eq. 2):

|  |  |
| --- | --- |
|  | (2) |

*Tree Diversity.* We calculated species richness, *i.e.*, the total number of species, and Shannon (H) diversity index with the following equation (eq. 3):

|  |  |
| --- | --- |
|  | (3) |

Where *pi* is the proportion of species *I* and *S* is the number of species.

*Environmental and Topographic Variables*. We calculated annual precipitation extracting WorldClim bio12 values per plot using the package *raster* in R. When plot’s annual precipitation values within a single site varied, we calculated a mean annual precipitation value per site. WorldClim’s annual mean temperature (bio1) was also extracted following the same procedure. Then, we extracted slope (in degrees) and aspect values for each plot from NASA’s Shuttle Radar Topography Mission digital elevation data (~30 m resolution), using Google Earth Engine (Farr et al., 2007).

*Land-use Variables*. To quantify land-use in TMCF we used three variables: forest disturbance related to agricultural activities, forest disturbance related to cattle grazing activities, and a land-use gradient we built on information about the landscape composition of each site (*i.e.*, the proportion of different types of land cover within a site). We assessed FI sites’ forest disturbance related to agricultural and grazing activities at the time of data collection using FI’s disturbance data set, which contains information about the cause and severity of vegetation disturbance. The causes of disturbance are classified in 11 classes: fires, hurricanes, floods, roads, logging, land-use change, grazing, pests and diseases, electrical cables, mining, and urbanization. The severity of disturbance is classified in a four-category nominal scale: very low, low, medium, and high severity. Both the cause and severity of disturbance where qualitatively assessed during field data collection (CONAFOR, 2018). Agriculture and grazing activities are reported within the categories of land-use change, grazing, logging, and fires with labels such as ‘clearing for growing coffee’, ‘swidden agriculture’, or ‘conversion from forest to cattle ranch’. We reviewed all recorded disturbance causes in the database, identified, and extracted data related to agricultural and cattle grazing activities. Then, we assigned each site a disturbance severity value from 0, when no disturbance was reported, to 4, indicating high severity disturbance. Whenever a site presented more than one reported disturbance related to agriculture or grazing, we averaged the disturbance severity value.

To assess the landscape composition of each site in terms of the proportion of mature forest, secondary forest, and agricultural or grazing lands present in the site, we first identified an approximate successional state of FI plots. To do so, we classified FI plots with a k-means analysis using structural attributes, including tree height, DBH, and tree density. K-means is a non-hierarchical cluster analysis where the user defines the initial number of centers. We run the analysis using two, three, four and five initial centers with 25 random sampling sets each. Then, we compared 30 indices to define the best number of clusters and chose the one that was better supported by most indices. Most of the indices suggested three clusters as the best classification. Thus, from this analysis, we obtained three structural clusters: the first one groups together plots with very low tree density, low basal area, and low tree height; the second cluster groups together forest plots with high tree density, and medium basal area and tree height; and a third one groups together plots with very high basal area and tree height, but medium tree density (Figure S1). Other studies conducted in TMCF in Oaxaca have shown that tree height, DBH, and tree density change through time after disturbance and are useful for estimating an approximate stage of forest succession in forest-agriculture mosaics (Velasco-Murguía et al., 2021). In general, TMCFs after disturbance show an increase in tree density, height, and basal area. Over time, tree height and basal area continues to increase but stem density decreases. This transition usually happens around 50 years after disturbance and differentiates young forest from mature forest (del Castillo, 2015). Because the three clusters we obtained from the non-hierarchical cluster analysis follow this general trend, we assigned approximate successional stages to each cluster as follows: we defined cluster one as very young forest, cluster two as young forest, and cluster three as mature forest. It is important to note that forest succession is a continuum and a complex process. Here, we classified forest succession in discrete categories as a methodological approach conducted for the sake of the analysis. This approach has proven to be useful for understanding forest ecosystem dynamics elsewhere (CITAS).

Once plots were classified in three successional stages, we assessed the composition of very young, young, and mature forests in each site. Interestingly, some of the sites have plots that fall across different structural classes, showing the patchiness in these forest-agriculture mosaic landscapes. To describe this patchiness, we defined a landscape composition value (eq. 4) adding up all successional stages within a site and normalizing the value as follows:

|  |  |
| --- | --- |
|  | (4) |

Where SS is the sum of the successional stage categories of the plots in a site. Successional stage categories can take the following values: very young forest = 1; young forest = 2, and mature forest = 3. Thus, considering there are four plots in each site, the minimum possible SS value is always 4, and the maximum is always 12. This way, a landscape composition value from 0 to 1 is assigned to all sites, where 0 represents sites dominated by mature forests, 1 represents sites dominated by very young forest stands, and everything in between are sites with a combination of forests at different successional stages. Thus, this variable represents a land-use gradient where sites dominated by very young forests (*i.e.*, sites closer to 1) show a greater degree of forest disturbance. It is important to note that other causes of forest disturbance unrelated to land-use can result in values closer to 1, such as pest outbreaks. However, agriculture and cattle ranching expansion have been identified as main causes of forest loss in TMCF (CITAS), and there is a correlation between our landscape composition variable and the presence of forest disturbance related to agricultural and grazing activities reported in the FI disturbance database (Figure S2). Therefore, other sources of disturbance were excluded from the analysis.

*Data Selection and Quality Control*

Because TMCF has a scattered distribution along the NMO, not all sites within this region correspond to our study system. To filter FI sites further, we used the following criteria: (1) sites should be within an elevation range between 1,000 and 2,800 m asl; (2) sites should receive at least 1,000 mm of annual precipitation; (3) sites should be described as cloud forest in the vegetation type column of the FI database; and (4) all sites must have epiphytes.

We performed data quality control homogenizing missing data values, correcting taxonomic and places names, removing diacritics, and filtering out rows with missing information. Additionally, we removed plots with many unidentified species or where most trees were dead. Because some plots are in places impossible to collect field data (such as very steep ravines), not all sites have four sampled plots. To avoid a biased sampling design, we selected only sites where four plots were sampled. After applying these filters, we ended up with a dataset of 40 sites located between 16.8938 and 18.6155 degrees N and -95.6699 and -97.0214 degrees W, within the distribution of TMCF (Figure 1).

*Statistical Analysis*

We performed statistical analysis at two sampling levels. On the one hand, we used sites (n= 40) as our main unit of analysis to assess the amount of AGB in TMCF, and its variation across space in relation to tree diversity, environmental and land-use gradients. On the other, we used plots (n= 160; a smaller sampling unit) for understanding the contribution of tree size to AGB and stem density, as well as changes in structural attributes in forest at different successional stages. First, we gained a general sense of the amount of AGB in TMCF and its variation performing basic summary statistics at site level and estimating the correlation between all variables. Then, we conducted analyses to answer our three main questions. The following sections describe these analyses further. All statistical analyses were performed in R version 4.1.1 (2021).

*AGB distribution in TMCF*. We explored the contribution of tree size to AGB and stem density following a similar approach as Cuni-Sanchez et al. (2021) and classified trees in six size classes based on their DBH: ≤10 cm, 10-20 cm, 20-30 cm, 30-40 cm, 40-50 cm, >50 cm. Then, we calculated the proportion of AGB and stem density per size class in each plot and performed an analysis of variance (ANOVA) to assess the difference in the contribution of tree size to AGB and stem density. Additionally, we computed a two-way ANOVA to evaluate the differences in the contribution of tree size classes within and among forests at different successional stages. Finally, we carried out four more ANOVA to test whether tree density, tree height, basal area and AGB are significantly different in forests at different successional stages.

*AGB patterns along environmental and land-use gradients*. We explored the relationships among tree biomass, diversity, environmental variables, and land-use in TMCF sites using multiple linear regression models. First, to reduce the number of variables, we computed a principal component analysis (PCA) of altitude, precipitation, temperature, and slope. Aspect is not a continuous variable and lacks variation in our dataset as most sites are facing either south or west, and thus, was excluded from the analysis. The first principal component (PC1) explained 68% of the variation and is correlated to temperature, precipitation, and altitude. Thus, PC1 represents an environmental gradient from warmer and moister sites at lower elevations to cooler and drier sites at higher elevations (Figure S3). Therefore, we selected PC1 and slope as our environmental predictors. We fitted a model with AGB as the response variable and tree diversity, environmental gradient, slope, and the three land-use variables as predictors. Because data is not normally distributed, we log-transformed AGB. Then, following an approach similar to Tredennick et al. (2021) for model exploration, we performed variable selection by comparing the full model against a series of reduced models in which each predictor is dropped in a stepwise fashion. We selected the best model comparing their adjusted R2, Mallows' Cp (CP), and Bayesian Information Criterion (BIC). To test for spatial autocorrelation, we calculated the Moran’s I of the residuals of the final model using a neighboring distance of 10, 25, and 50 km, and visualized the spatial distribution of residuals with a map.

*AGB relationship to tree diversity.* To understand the relationship between AGB and tree diversity in TMCF, we performed a series of nonparametric regression analysis. Nonparametric regressions allow to test relationships between two or more variables without assuming linearity. Given tree diversity and AGB in our dataset show a positive but week correlation, we could not assume a linear relationship. Here, we assessed the relationship between AGB and stem density with Shannon diversity index and species richness at site level fitting locally weighted regression (loess) curves. Similarly, we performed these nonparametric regression analyses at the plot level. Although we do not have specific ages after disturbance for assessing how diversity and AGB change over time, taking a space-by-time approach we evaluated these trends using the three-category successional stages we assigned to each plot.

**Results**

*AGB allocation in TMCF*

Average tree AGB in TMCF in the NMO is 137.49 ± 121.29 Mg ha-1 and it ranges from as low as 8.266 Mg ha-1 to as high as 414.524 Mg ha-1, showing the wide variation of AGB existing in these landscapes (Table 1). In fact, out of the 160 plots analyzed, 30 are outside this range. Specifically, 17 plots show values below 8.26 Mg ha-1, several of which were completely devoid of trees larger than 7.5 cm of DBH at the time of data collection. In contrast, 13 plots exhibit higher values than the highest averaged AGB found at site level, some of them even surpassing 500 Mg ha-1. Variation in structural attributes within and among sites is also large (Table 1). For instance, average stem density in all 40 sites is 641.56 ha-1 but some sites have as few as 81.25 trees ha-1.

The large variation in AGB and structural attributes found in TMCF sites results from the diversity in landscape composition found in these forest-agriculture mosaics and it is driven by the successional stage of each plot. In very young forests, tree density is low, and trees are short and thin. As forests develop, all these structural attributes increase. Thus, young forests show larger trees and higher stem density. In mature forests, the density of trees decreases as trees become even taller and bigger (Figure 2). As a result, AGB in forests at different successional stages is significantly different (ANOVA, p< 0.001, Table S3). In very young forests, AGB averages only 33.38 Mg ha-1, in young forests this value notably increases to 151.69 Mg ha-1, and in mature forests it reaches 354.86 Mg ha-1 (Table 2). The ANOVA we performed to analyze structural differences among forests at different successional stages showed that all attributes are statistically different in all three categories (very young, young, and mature forests).

In these landscapes there is a large proportion of small trees (DBH < 20 cm), and trees with a DBH greater than 30 cm are uncommon (Figures 3 and S5). In most plots, trees between 10 and 20 cm of DBH represent almost 50% of the total number of trees. In contrast, trees with a DBH larger than 50 cm were found only in 51 out of 160 plots. When present, these large trees rarely account for more than 10% of the total number of trees. Despite they are less abundant, large trees contribute the most to total AGB, particularly the largest ones (DBH > 50 cm, Figures 3 and S5). Interestingly, this pattern changes in forest plots at different successional stages (Figure 3). Although the proportion of stems of different size remains somewhat constant across very young, young, and mature forests, their contribution to AGB varies widely (Two-way ANOVA, p < XX, Table S5). In very young and mature forests, largest trees are the ones contributing the most to total AGB, but in young forest the contribution to AGB is equally distributed across all tree size classes. In fact, in young forests, trees with a DBH smaller than 10 cm is the only tree size class showing a statistically significant different contribution to total AGB because their contribution is lower than any other tree size class (Table S5). In this case, largest trees do not stand out as big contributors to total AGB. In contrast, in very young forests, trees with a DBH larger than 40 and 50 cm show a statistically significant difference contribution to AGB than the rest tree size classes, because they are the largest contributors to total AGB. Likewise, in mature forests, trees with a DBH larger than 50 cm show a statistically significant large contribution to total AGB when compared to smaller tree size classes (Table S5). On the other hand, smaller trees are more relevant for AGB in very young and young forests than in mature forests. In fact, while the average contribution to AGB of trees with a DBH smaller than 10 cm in very young and young forests ranges from 6 to 14%, in mature forests this proportion barely reaches 1% (Figure 3).

The patterns of tree size contribution to the total number of stems and AGB shown by forests at different successional stages can be explained by the progression of structural attributes over time (Figure 2). In particular, the number of large trees present in each successional stage seems to play a fundamental role in defining AGB allocation in a forest plot. Very young forests have few large trees and small AGB. Thus, the few but large trees that exist in these plots represent a large proportion of the total amount of AGB. In young forests, the number of large trees is still small but there is more total AGB because tree density is higher. Hence, all small and big trees are important contributors to total AGB. Finally, in mature forests, the number of large trees is high and that of small trees low. Tree density decreases but AGB is large, which makes large trees the main contributors to total AGB (Figure 3).

*Patterns of Tree AGB Across Land-use and Environmental Gradients*

Tree AGB in TMCF sites is controlled mainly by land-use, and secondly by environmental factors. Here, we assessed both land-use and environmental factors with multiple linear regression models. Through a stepwise model selection process, we found the best model as the one having the lowest BIC, large explanatory power (high R2), and where all predictors are statistically significant (Table 3). The best model includes three variables: landscape composition, forest disturbance related to agriculture and slope gradient (p < 0.05, adjusted R2= 0.811, Table 4, Figure S6). The level of land-use, represented here in the landscape composition variable, controls AGB the most. This variable has a strong relationship with AGB (p < 0.05, adjusted R2= 0.729, Figure 4). As land-use increases, the proportion of very young forests in the landscape grows and, thus, AGB steadily decreases. Although forest disturbance driven by agriculture is related to our landscape composition variable, adding this variable to the model improves its explanatory power and it seems to be the second most relevant predictor of AGB (Table 3). Similarly, adding slope gradient improves the linear regression model. However, from the three predictors, slope gradient is the least influential in determining AGB patterns in TMCF. The residuals of this model do not show spatial autocorrelation (Figure S6). We computed Moran’s I statistic for neighborhoods of 10, 25, and 50 km of distance and all of them resulted non-significant (p= 0.517, p= 0.604, p= 0.187, respectively).

Interestingly, environmental gradient is not included as a relevant predictor in the best model. The second best model includes environmental gradient as a fourth predictor, but this variable does not seem to fundamentally control AGB in TMCF landscapes within the NMO (Table 3). The environmental gradient variable (which includes temperature, precipitation, and elevation) has a statistically significant relationship with AGB (p < 0.05, adjusted R2= 0.185, Figure 5) where warmer and more humid sites at lower elevations (closer to 1,000 m asl) exhibit smaller AGB than sites at higher elevations that have a cooler and drier climate. Additionally, this environmental gradient is related to the land-use gradient. Sites at lower elevations show larger land use and forest disturbance than sites at higher elevations (p < 0.05, adjusted R2= 0.372, Figure 5). The fact that land-use exerts a strong effect on AGB, in addition to be related to the environmental gradient, results in the latter being only marginally relevant when both predictors are considered.

*AGB and Tree Diversity*

We analyzed the relationship between AGB and tree diversity at site and plot levels. Although we were expecting a linear positive relationship between these variables, they showed a positive but weak correlation with correlation coefficients near 0.3 (Figure S4). We addressed this issue using nonparametric regressions, where linearity is not assumed. Using Shannon diversity index (H) and species richness, we found that tree diversity increases with AGB until sites reach approximately 200 Mg ha-1 (Figure 6). Sites with more than 200 Mg ha-1 tend to show less diversity. This probably occurs in forests where large dominant trees ─which are the ones contributing greatly to the total amount of AGB in TMCF─ are more abundant, decreasing evenness and diversity in a site. Indeed, when we analyzed the relationship between AGB and tree diversity across plots at different successional stages, we found that tree diversity increases over time but reaches a limit, where tree diversity slightly decreases in mature forests (Figure 6c and d). Thus, sites with a larger composition of mature forests (i.e., with landscape composition values closer to 0) show greater AGB but not necessarily greater tree diversity (Figure 4).

**Discussion**

TMCF sites are found along two compounding gradients: an environmental one, going from warmer and wetter areas at lower elevations to cooler and drier ones at higher elevations; and a land-use gradient, from higher to lower forest disturbance related to agricultural and cattle grazing activities.

Spatial heterogeneity caused by small-scale forest disturbance determines AGB patterns at landscape and regional scales.

The effect of environmental variables on tree AGB and diversity is evident once land-use is taken into consideration.

It is important to note that only trees larger than 7.5 cm of DBH are considered in this analysis, and shrubs, lianas, palm trees and ferns were excluded from both AGB and diversity estimates.

**Conclusions**

**Tables**

Table 1. Summary statistics of structural attributes, tree aboveground biomass, and tree diversity in Tropical Montane Cloud Forest sites (n= 40).

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Min ± S.E.** | **Max ± S.E.** | **Mean ± S.D.** |
| Stem density (tree ha-1) | 81.25 ± 15.72 | 1806.25 ± 444.10 | 641.56 ± 355.23 |
| Basal area (m ha-1) | 2.26 ± 0.64 | 51.08 ± 6.57 | 20.42 ± 13.48 |
| Lorey’s height (m) | 5.92 ± 0.68 | 24.91 ± 1.00 | 12.46 ± 4.63 |
| Aboveground biomass (Mg ha-1) | 8.26 ± 1.02 | 414.52 ± 19.29 | 137.49 ± 121.29 |
| Species richness | 1.75 ± 0.25 | 9 ± 1 | 4.4 ± 1.63 |
| Shannon diversity index | 0.33 ± 0.21 | 1.75 ± 0.25 | 1.06 ± 0.36 |

Table 2. Structural attributes and tree aboveground biomass (AGB) in forest plots (n= 160) in three different successional stages: very young, young, and mature forests. Differences in structural attributes and AGB between different successional stages are all statistically significant (ANOVA, p < 0.05, Table S3, Figure 2).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Very Young (n= 69)** | | **Young (n= 62)** | | **Mature (n= 29)** | |
| **Mean** | **S.D.** | **Mean** | **S.D.** | **Mean** | **S.D.** |
| Stem density (tree ha-1) | 275.000 | 166.550 | 986.694 | 444.380 | 775.862 | 313.141 |
| Basal area (m ha-1) | 6.875 | 4.432 | 24.756 | 9.336 | 43.422 | 13.288 |
| Lorey’s height (m) | 9.376 | 3.550 | 11.764 | 2.366 | 21.314 | 3.124 |
| AGB (Mg ha-1) | 33.381 | 32.631 | 151.691 | 92.042 | 354.866 | 151.605 |

Table 3. Results of stepwise model selection process comparing seven multiple linear regression models of aboveground biomass as a function of land-use, environmental, and species diversity variables. Shown are the number of predictors considered in each model ordered from most to least relevant (top to bottom), as well as their statistical significance (\*\*\*p < 0.001, \*\*p < 0.01, \*p<0.05, ◦p<0.10), adjusted R2, Mallows' Cp (CP), and Bayesian Information Criterion (BIC). Shown in bold are the highest adjusted R2, and lowest CP and BIC.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Predictors | Number of predictors considered in the model | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Landscape composition | \*\*\* | \*\*\* | \*\*\* | \*\*\* | \*\*\* | \*\*\* | \*\*\* |
| Disturbance by agriculture |  | \*\* | \*\* | \*\*\* | \*\*\* | \*\*\* | \*\*\* |
| Slope |  |  | \*\* | \* | \*\* | \*\* | \*\* |
| Environmental gradient |  |  |  | 0.16 | 0.13 | ◦ | 0.19 |
| Disturbance by grazing |  |  |  |  | 0.16 | ◦ | ◦ |
| Shannon diversity index |  |  |  |  |  | 0.24 | 0.24 |
| Species richness |  |  |  |  |  |  | 0.48 |
| Adjusted R2 | 0.7294 | 0.7691 | 0.8117 | 0.8167 | 0.8217 | **0.8237** | 0.8210 |
| CP | 21.4582 | 13.7513 | 5.8754 | **5.8568** | 5.8764 | 6.5081 | 8.0000 |
| BIC | -45.9589 | -49.6720 | **-55.2514** | -53.7532 | -52.3369 | -50.2971 | -47.2385 |

Table 4. Results of the multiple linear regression selected as the best model to explain aboveground biomass patterns in TMCF. This model includes three predictors: slope gradient, landscape composition, and disturbance by agricultural activities.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Coefficient** | **S.E.** | **t** | **p** |
| Slope gradient | 0.027 | 0.009 | 3.064 | 0.004 \*\* |
| Landscape composition | -2.425 | 0.266 | -9.088 | < 0.001 \*\*\* |
| Disturbance by agriculture | -0.279 | 0.080 | -3.478 | 0.001 \*\* |
| b= 5.406; F= 57.07; df (3, 36); p < 0.001; adjusted R2= 0.811 | | | | |
| \*\*\*p < 0.001, \*\*p < 0.01, \*p<0.05, ◦p<0.10 | | | | |

**Figures**

Figure 1. Map

|  |
| --- |
| Gráfico, Gráfico de cajas y bigotes  Descripción generada automáticamente |
| Figure 2. Comparison of a) stem density, b) Lorey’s height, c) basal area, and d) aboveground biomass between very young (VY, shown in orange), young (Y, shown in purple), and mature (M, shown in green) forest plots in TMCF (n= 160). Asterisks indicate statistically significant differences resulted from Tukey HSD tests as follows: \*\*\*\*p < 0.0001, \*\*\*p < 0.001, \*\*p < 0.01, \*p<0.05. Results of ANOVA and Tukey HSD test analyses can be found in Table S3. |

|  |
| --- |
|  |
| Figure 3. Contribution of tree size categories to a) stem density and b) aboveground biomass (AGB) in very young (VY, shown in orange), young (Y, shown in purple), and mature (M, shown in green) forest plots in TMCF (n= 160). Trees were categorized in six size classes based on their DBH. Asterisks indicate statistically significant differences between forests at different successional stages within each tree size class tested with one-way ANOVAs and Tukey tests (results shown in Table S6). Statistical significance: \*\*\*\*p < 0.0001, \*\*\*p < 0.001, \*\*p < 0.01, \*p<0.05. Non-significant differences are not shown. Letters indicate statistically significant differences between tree size classes within very young (VY, shown in orange), young (Y, shown in purple), and mature (M, shown in green) forests resulted from one-way ANOVAs and Tukey tests (results shown in Table S7). Boxes sharing a letter are not statistically different. |

|  |
| --- |
| Gráfico, Gráfico de dispersión  Descripción generada automáticamente |
| Figure 4. Landscape composition relationship with a) aboveground biomass (linear regression, p < 0.05, adjusted R2= 0.729, shown in black dashed line), and b) tree diversity (linear regression, p= 0.041, adjusted R2=0.081, shown in black dashed line) estimated with Shannon diversity index (H) in TMCF sites (n= 40). Locally weighted regression (loess) curves are shown in dashed gray lines. Landscapes dominated by mature forests have low landscape composition values and those dominated by very young forests have high landscape composition values. See text for details on how the landscape composition variable was calculated. |

|  |
| --- |
| Gráfico, Gráfico de dispersión  Descripción generada automáticamente |
| Figure 5. Linear regressions curves between a) aboveground biomass (p < 0.05, adjusted R2= 0.185), and b) landscape composition (p < 0.05, adjusted R2= 0.372) as a factor of environmental gradient in TMCF sites (n= 40). Negative values in the environmental gradient show warmer and more humid sites at lower elevations and positive values represent cooler and drier sites at higher elevations. See text for details on how the environmental gradient variable was calculated. |

|  |
| --- |
| Gráfico, Gráfico de dispersión  Descripción generada automáticamente |
| Figure 6. Locally weighted regression (loess) curves (dashed gray lines) showing non-linear relationships between aboveground biomass (AGB) and tree diversity in TMCF sites (n= 40, panels a and b) and forest plots (n= 160, panels c and d), measured with Shannon diversity index (H) and species richness (S). Forest successional stage is displayed as follows: very young (VY) in orange, young (Y) in purple, and mature (M) forests in green. |

**Supporting Information**

Table S1. Variables used in this study at three sampling units: site, plot, and tree.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Level** | **Variable** | **Units** | **Explanation** | **Source** |
| Site | Site ID | - | Unique identifier of site | FI |
| Longitude | Decimal degrees | Geographic coordinate at site centroid | FI |
| Latitude | Decimal degrees | Geographic coordinate at site centroid | FI |
| Altitude | m | Elevation above sea level at site centroid | FI |
| Temperature | Degrees C | Annual mean temperature | WorldClim |
| Precipitation | mm | Annual mean precipitation | WordClim |
| Slope | Degrees | Average slope | Calculated from NASA’s Shuttle Radar Topography Mission digital elevation data (~30 m resolution), averaged by plot |
| Aspect | Degrees | Average aspect | Calculated from NASA’s Shuttle Radar Topography Mission digital elevation data (~30 m resolution), averaged by plot |
| Plot number | Plot | Number of plots in site (goes from 1 to 4) | FI (edited after data quality control) |
| Tree number | Tree | Site’s average number of trees measured | Derived from FI raw data, averaged by plot |
| Tree density | stems/ha | Average number of trees per area in site | Derived from FI raw data, averaged by plot |
| Basal area | m/ha | Site’s average basal area | Derived from FI raw data, averaged by plot |
| Tree height | m | Site’s average tree height | Derived from FI raw data, averaged by plot |
| Lorey’s height | m | Mean tree height weighted by their basal area | Derived from FI raw data, averaged by plot |
| Aboveground biomass (AGB) | Mg/ha | Site’s average AGB | Averaged by plot, calculated with allometric equations using FI raw data |
| Shannon (H) | bits | Diversity index | Calculated using FI raw data, averaged by plot |
| Species richness | Species | Site’s average number of species | Calculated using FI raw data, averaged by plot |
| Landscape composition | - | Site’s patchiness, goes from 0, when all plots in site are mature forests, to 1, when all plots in site are very young forests or agricultural lands | Calculated using plot’s successional stage |
|  | Disturbance by agriculture | - | Forest disturbance related to agricultural activities, aggregated, and averaged by site; it goes from 0 when no disturbance was detected to 4 when disturbance was severe | Derived from FI disturbance database |
|  | Disturbance by grazing | - | Forest disturbance related to cattle grazing, aggregated, and averaged by site; it goes from 0 when no disturbance was detected to 4 when disturbance was severe | Derived from FI disturbance database |
| Plot | Plot ID | - | Unique identifier of plot | FI |
| Latitude | Decimal degrees | Geographic coordinate | FI |
| Longitude | Decimal degrees | Geographic coordinate | FI |
| Altitude | m | Elevation above sea level | FI |
| Slope | Degrees | Hillslope steepness | Calculated from NASA’s Shuttle Radar Topography Mission digital elevation data (~30 m resolution) |
| Aspect | Degrees | Direction that the slope faces (a.k.a., exposure) | Calculated from NASA’s Shuttle Radar Topography Mission digital elevation data (~30 m resolution) |
| Epiphytes | - | Whether epiphytes are present in plot. This variable was only used to select cloud forest sites. | FI |
| Tree number | tree | Total number of trees measured in plot | Calculated using FI raw data |
| Tree density | stems/ha | Number of trees per area | Calculated using FI raw data |
| Tree height | m | Plot’s average tree height | Calculated using FI raw data |
| Lorey’s height | m | Tree’s height weighted by their basal area | Calculated using FI raw data |
| Basal area | m/ha | Sum of tree’s basal area in relation to plot’s area | Calculated using FI raw data |
| Aboveground biomass (AGB) | Mg/ha | Sum of tree’s AGB per area | Calculated with allometric equations using FI raw data |
| Shannon (H) | bits | Diversity index | Calculated using FI raw data with package vegan in R |
| Species richness | Species | Total number of species in plot | Calculated using FI raw data |
| Successional stage | - | Whether very young, young, or mature forest | Calculated using FI raw data with a non-hierarchical cluster analysis (k-means) |
| Tree | Site | - | Unique identification of site | FI |
| Plot\_id | - | Unique identification of plot | FI |
| Species | - | Taxonomic name | FI corrected with Taxonomic Name Resolution Service |
| Family | - | Taxonomic family | FI corrected with Taxonomic Name Resolution Service with BIOMASS function correctTaxo |
| Common name | - | Common name | FI |
| Status | - | Alive or dead | FI |
| Life form | - | Tree, shrub, palm tree, fern, or liana | FI |
| Height | m | Individual total height | FI |
| Diameter at breast height (DBH) | cm | Diameter of trunk at 1.3 m from the ground | FI |
| Basal area | m | Cross sectional area of trunk at 1.3 m from the ground | FI |
| Mean wood density | g/cm3 | Wood density as recorded in scientific literature | Calculated with BIOMASS function getWoodDensity |
| Aboveground biomass (AGB) | Mg | Dry mass of the aboveground component (i.e., excluding roots) of plants | Calculated with allometric equations |

Table S2. Generic and specific allometric equations used in this study to estimate aboveground biomass.

|  |  |  |
| --- | --- | --- |
| **Species** | **Allometric equation** | **Reference** |
| *Abies sp.* | [0.0754]\*[DBH^2.513] | Avedaño et al., 2009 |
| *Alchornea latifolia* | [Exp[-3.363]\*[DBH^2.2714]\*[TH^0.4984] | Aquino-Ramírez et al., 2015 |
| *Alnus acuminata* | [Exp[-2.14]\*[DBH^2.23]] | Acosta-Mireles et al, 2002 |
| *Alnus jorullensis* | [0.0195]\*[DBH^2.7519] | Carrillo et al., 2014 |
| *Brosimum alicastrum* | [0.479403]\*[DBH^2.0884] | Rodríguez-Laguna et al., 2008 |
| *Cecropia obtusifolia* | [[0.000022]\*[D^1.9]\*[H]] + [[-0.56 + 0.02[D^2] + 0.04[H]]/10^3] | Hughes et al., 1999 |
| *Citrus sp.* | [-6.64]+[0.279\*BA]+[0.000514\*BA^2] | Schroth et al., 2002 |
| *Clethra sp.* | [Exp[-1.90]\*[DBH^2.15]] | Acosta et al., 2002 |
| *Clethra hartwegii* | [Exp[-1.90]\*[DBH^2.15]] | Acosta et al., 2002 |
| *Clethra mexicana* | [0.4632]\*[DBH^1.8168] | Acosta et al., 2011 |
| *Clethra pringlei* | [0.067833]\*[DBH^2.50972] | Rodríguez et al., 2006 |
| *Cordia alliodora* | [10^-0.755]\*[DBH^2.072] | Segura et al., 2006 |
| *Cupressus lusitanica* | [0.5266]\*[DBH^1.7712] | Vigil, 2010 |
| *Dendropanax arboreus* | [0.037241]\*[DBH^2.99585] | Rodríguez-Laguna et al., 2008 |
| *Eugenia sp.* | [0.4600]+[[0.0370]\*[DBH^2]\*TH] | Cairns et al., 2003 |
| *Fraxinus uhdei* | [362.129]\*[[3.1416]\*[[[[DBH^2]/4]]^1.100]] | Cano, 1994 |
| *Heliocarpus appendiculatus* | [[Exp[4.9375]] \* [[DBH^2]^1.0583]] \* [1.14]/ 1000000 | Hughes et al., 1999 |
| *Inga sp.* | [Exp[-1.76]\*[DBH^2.26]] | Acosta et al., 2002 |
| *Inga vera* | [Exp[-1.76]\*[DBH^2.26]] | Acosta et al., 2002 |
| *Inga punctata* | [Exp[-3.363]\*[DBH^2.4809]\*[TH^0.4984] | Aquino-Ramírez et al., 2015 |
| *Juglans olanchana* | [10^-1.417]\*[DBH^2.755] | Segura et al., 2006 |
| *Juniperus flaccida* | [0.209142]\*[DBH^1.698] | Rodríguez et al., 2009 |
| *Liquidambar sp.* | [Exp[-2.22]\*[DBH^2.45]] | Acosta et al., 2002 |
| *Liquidambar styraciflua* | [0.180272]\*[DBH^2.27177] | Rodríguez et al., 2006 |
| *Nectandra ambigens* | [[Exp[4.9375]]\*[[DBH^2]^1.0583]]\*[1.14]/1000000 | Hughes et al., 1999 |
| *Pinus sp.* | [0.058]\*[[[DBH^2]\*TH]^0.919] | Ayala, 1998 |
| *Pinus ayacahuite* | [0.058]\*[[[DBH^2]\*TH]^0.919] | Ayala, 1998 |
| *Pinus devoniana* | [0.182]\*[DBH^1.936] | Méndez et al., 2011 |
| *Pinus herrerae* | [0.1354]\*[DBH^2.3033] | Návar, 2009 |
| *Pinus leiophylla* | [[Exp^-3.549]\*[DBH^2.787]]] | Návar, 2009 |
| *Pinus oocarpa* | [0.058]\*[[[DBH^2]\*TH]^0.919] | Ayala, 1998 |
| *Pinus patula* | [0.0514]\*[DBH^2.5222] | Pacheco, 2011 |
| *Pinus pseudostrobus* | [0.058]\*[[[DBH^2]\*TH]^0.919] | Ayala, 1998 |
| *Prunus persica* | [Exp[-2.76]\*[DBH^2.37]] | Acosta, 2003 |
| *Psidium guajava* | [0.246689]\*[DBH^2.24992] | Rodríguez-Laguna et al., 2008 |
| *Quercus sp.* | [0.1269]\*[DBH^2.5169] | González, 2008 |
| *Quercus candicans* | [[Exp[-4.775313]\*[DBH^1.798292]\*[TH^1.570775]]+[[Exp[-3.547008]\*[DBH^2.593972]]+[[Exp[-4.752007]\*DBH^2]] | Cortés-Sánchez et al., 2019 |
| *Quercus crassifolia* | [0.283]\*[[[DBH^2]\*TH]^0.807] | Ayala, 1998 |
| *Quercus laurina* | [0.283]\*[[[DBH^2]\*TH]^0.807] | Ayala, 1998 |
| *Quercus obtusata* | [[exp[-3.53684]\*[DBH^2.043763]\*[TH^0.759522]]+[[Exp[-5.803952]\*[DBH^2\*TH]^1.224292]]+[[Exp[-6.181035]\*[DBH^2.488617]] | Cortés-Sánchez et al., 2019 |
| *Quercus peduncularis* | [Exp[-2.27]\*[DBH^2.39]] | Acosta, et al., 2002 |
| *Quercus rugosa* | [0.283]\*[[[DBH^2]\*TH]^0.807] | Ayala, 1998 |
| *Trema micrantha* | [-2.305 + 2.351 \* ln[DBH]] \* 1.033 | Van Breugel et al., 2011 |
| *Trichilia havanensis* | [0.130169]\*[DBH^2.34924] | Rodríguez-Laguna et al., 2008 |
| *Trichospermum mexicanum* | [0.449]\*[DBH^2]-33.565 | Montes de Oca-Cano et al., 2020 |
| *Zanthoxylum sp.* | [0.00166]\*[DBH^3.6586] | Manzano, 2010 |
| Tropical trees | 0.0673 \* (WD \* H \* DBH^2)^0.976 | Chave et al., 2014 |

Table S3. Results of ANOVA and Tukey HSD test on stem density, Lorey’s height, basal area, and aboveground biomass (AGB) between forest plots at different successional stages (very young (VY), young (Y), and mature (M); n= 160). Significant p-values are shown in bold.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Response | ANOVA | | Tukey HSD | | | |
| F | *p* | Comparison | Estimate | 95% CI | *p* |
| Stem density | 80.86 | **<2e-16** | VY-Y | 711.69 | (576.74, 846.64) | **9.50e-14** |
| VY-M | 500.86 | (330.19, 671.52) | **2.85e-10** |
| Y-M | -210.83 | (-384.32, -37.33) | **1.27e-02** |
| Lorey’s height | 157.90 | **<2e-16** | VY-Y | 2.38 | (1.12, 3.65) | **4.63e-05** |
| VY-M | 11.93 | (10.33, 13.54) | **9.41e-14** |
| Y-M | 9.54 | (7.92, 11.17) | **9.41e-14** |
| Basal area | 197.50 | **<2e-16** | VY-Y | 17.88 | (14.32, 21.43) | **9.81e-14** |
| VY-M | 36.54 | (32.04, 41.04) | **9.41e-14** |
| Y-M | 18.66 | (14.09, 23.24) | **1.22e-13** |
| AGB | 135.70 | **<2e-16** | VY-Y | 118.30 | (81.62, 154.99) | **6.50e-12** |
| VY-M | 321.48 | (275.08, 367.88) | **9.41e-14** |
| Y-M | 203.17 | (156.00, 250.34) | **1.21e-13** |

Table S4. Results of ANOVA and Tukey HSD test on tree size contribution to stem density and aboveground biomass (AGB) between tree size classes in TMCF plots (n= 160). Trees were classified in six size classes according to their DBH as follows: class 1: DBH < 10 cm; class 2: DBH 20-30 cm; class 3: DBH 30-40 cm; class 4: DBH 40-50 cm; class 5: DBH 50-60 cm; class 6: DBH > 60 cm. Significant p-values are shown in bold.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Response | ANOVA | | Tukey HSD | | | |
| F | *p* | Comparison | Estimate | 95% CI | *p* |
| Contribution to stem density proportion | 129 | **<2e-16** | 1-2 | 0.144 | (0.09, 0.19) | **1.75e-10** |
| 1-3 | -0.123 | (-0.17, -0.07) | **1.92e-10** |
| 1-4 | -0.193 | (-0.24, -0.14) | **1.75e-10** |
| 1-5 | -0.223 | (-0.28, -0.16) | **1.75e-10** |
| 1-6 | -0.236 | (-0.30, -0.17) | **1.75e-10** |
| 2-3 | -0.267 | (-0.31, -0.22) | **1.75e-10** |
| 2-4 | -0.337 | (-0.38, -0.28) | **1.75e-10** |
| 2-5 | -0.368 | (-0.42, -0.30) | **1.75e-10** |
| 2-6 | -0.380 | (-0.44, -0.31) | **1.75e-10** |
| 3-4 | -0.069 | (-0.12, -0.01) | **2.69e-03** |
| 3-5 | -0.100 | (-0.16, -0.03) | **7.57e-05** |
| 3-6 | -0.112 | (-0.17, -0.03) | **1.77e-05** |
| 4-5 | -0.030 | (-0.09, 0.03) | 0.75 |
| 4-6 | -0.043 | (-0.11, 0.02) | 0.46 |
| 5-6 | -0.012 | (-0.08, 0.06) | 0.99 |
| Contribution to AGB proportion | 34.13 | **<2e-16** | 1-2 | 0.19 | (0.12, 0.25) | **1.75e-10** |
| 1-3 | 0.18 | (0.11, 0.24) | **1.76e-10** |
| 1-4 | 0.18 | (0.10, 0.25) | **2.11e-10** |
| 1-5 | 0.23 | (0.14, 0.31) | **1.76e-10** |
| 1-6 | 0.36 | (0.27, 0.45) | **1.75e-10** |
| 2-3 | -0.01 | (-0.07, 0.05) | 0.99 |
| 2-4 | -0.01 | (-0.08, 0.05) | 0.99 |
| 2-5 | 0.03 | (-0.04, 0.12) | 0.82 |
| 2-6 | 0.17 | (0.08, 0.26) | **4.92e-07** |
| 3-4 | -0.00 | (-0.07, 0.07) | 1.00 |
| 3-5 | 0.04 | (-0.03, 0.13) | 0.60 |
| 3-6 | 0.18 | (0.09, 0.27) | **1.33e-07** |
| 4-5 | 0.04 | (-0.04, 0.13) | 0.63 |
| 4-6 | 0.18 | (0.09, 0.28) | **4.11e-07** |
| 5-6 | 0.13 | (0.03, 0.24) | **2.70e-03** |

Table S5. Results of two-way ANOVA on tree size contribution to stem density and aboveground biomass (AGB) between tree size classes, forest successional stage (very young (VY), young (Y), and mature (M)), and their interactions in TMCF plots (n= 160). Trees were classified in six size classes according to their DBH as follows: class 1: DBH < 10 cm; class 2: DBH 20-30 cm; class 3: DBH 30-40 cm; class 4: DBH 40-50 cm; class 5: DBH 50-60 cm; class 6: DBH > 60 cm. Significant p-values are shown in bold. To see Tukey HSD results see full table in Github repository.

|  |  |  |  |
| --- | --- | --- | --- |
| Response | Factor? | F | *p* |
| Contribution to stem density proportion | Size class | 144.038 | **< 2e-16** |
| Successional stage | 19.048 | **9.43e-09** |
| Size class\*Successional stage | 4.641 | **2.15e-06** |
| Contribution to AGB proportion | Size class | 41.454 | **< 2e-16** |
| Successional stage | 53.949 | **< 2e-16** |
| Size class\*Successional stage | 3.804 | **5.35e-05** |

Table S6. Results of one-way ANOVA and Tukey HSD test on tree size class contribution to stem density and aboveground biomass (AGB) between forest plots at different successional stages (very young (VY), young (Y), and mature (M)) within each tree size class. Trees were classified in six size classes according to their DBH as follows: class 1: DBH < 10 cm; class 2: DBH 20-30 cm; class 3: DBH 30-40 cm; class 4: DBH 40-50 cm; class 5: DBH 50-60 cm; class 6: DBH > 60 cm. Significant p-values are shown in bold.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Response |  | ANOVA | | Tukey HSD | | | |
| F | *p* | Comparison | Estimate | 95% CI | *p* |
| Successional Stage |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

Table S7. One way ANOVA and Tukey test of tree size contribution to AGB and stem density in forest plots at different successional stages between tree size classes.

|  |
| --- |
| Mapa  Descripción generada automáticamente |
| Figure S1. Non-hierarchical cluster analysis – plots categories in three ‘successional stages’ |

|  |  |
| --- | --- |
|  |  |
| Figure S2. Correlation matrices. | |
|  | |
| Figure S3. PCA of environmental variables | |

|  |
| --- |
| a) |
| b) |
| Figure S4. Contribution of tree size categories to a) stem density and b) aboveground biomass in TMCF sites (n= 40). Trees were classified in six classes according to their DBH as follows: class 1: DBH < 10 cm; class 2: DBH 20-30 cm; class 3: DBH 30-40 cm; class 4: DBH 40-50 cm; class 5: DBH 50-60 cm; class 6: DBH > 60 cm. Letters indicate statistically significant differences between tree size classes assessed with a one-way ANOVA and Tukey test (results shown in Table S4). Boxes sharing a letter are not statistically different. |

|  |
| --- |
|  |
| Figure S5. Contribution of tree size categories to a) stem density and b) aboveground biomass in very young (VY, shown in orange), young (Y, shown in purple), and mature (M, shown in green) forest plots in TMCF (n=160). Trees were categorized in six size classes based on their DBH. Letters indicate statistically significant differences between tree size classes and forest succession assessed with a two-way ANOVA and Tukey test (results shown in Table S5). Boxes sharing a letter are not statistically different. Forest successional stage is shown in colors as follows: very young (VY) in orange, young (Y) in purple, and mature (M) forests in green. |

|  |  |
| --- | --- |
| a) | b) |
| c) | d) |
| Figure S6. Partial residual plots of best model and map of residuals with Moran’s I | |