**Title.** Land-use intensity determines aboveground biomass distribution in tropical montane cloud forest across spatial scales

**Authors**

**Affiliations**

**Keywords:** forest-agriculture mosaics; tropical mountains; forest disturbance; environmental gradient; tree diversity

**Abstract**

**Introduction**

Tropical forests play a fundamental role in the carbon cycle because they contain ~25% of the carbon in the terrestrial biosphere (Bonan, 2008). This carbon is stored in the living biomass of trees and other understory vegetation mainly allocated above the soil in stems, branches, and leaves, in what is known as aboveground biomass (AGB) (Gibbs et al., 2007). Because AGB represents the main carbon pool in tropical forests, it also determines the amount of carbon loss to the atmosphere when these forests are disturbed (Houghton, Hall, & Goetz, 2009)(Pan?). Thus, AGB is considered an essential climate variable and an important input to Earth system models (REFs). Moreover, AGB is related to forests structure and composition, and research has found it has a positive relationship with tree diversity suggesting an interesting synergy between carbon storage and biodiversity (Poorter et al., 2015), two of the most threatened ecosystem features by current global environmental change.

The precise quantification of AGB in tropical forests remains a challenge, particularly in tropical montane forests (located above 1,000 m asl) where carbon has been historically understudied and underestimated (Cuni-Sanchez et al., 2021; Spracklen & Righelato, 2014). Although several studies have been conducted for estimating AGB at local (refs), regional (refs), and global scales (Saatchi, 2011; avitabile et al 2016), most of them focus on lowland tropical forests, leaving the magnitude, patterns, and drivers of AGB in tropical montane forests unknown (refs). Estimating AGB in tropical mountains is difficult because field data is sparse and remote sensing approaches are challenged by rugged terrain (refs). Additionally, estimates have large uncertainties due to AGB significant spatial variation driven mainly by two reasons: (1) environmental factors changing with elevation and (2) forest disturbance (Houghton et al., 2009) (more refs).

The environmental gradient imposed by elevation influences AGB distribution in tropical mountains (refs). As elevation increases, temperatures decrease and forests are exposed to frequent cloud cover and fog (Gotsch, Asbjornsen, & Goldsmith, 2020). This elevation gradient shapes forest structure and composition and gives place to a distinctive ecosystem known as tropical montane cloud forest (TMCF), whose main feature is to be persistently immerse in ground-level clouds. Fog, decreasing temperatures, waterlogged soils, and nutrient limitation are all features found in TMCF that generally restrain primary productivity (Fahey et al., 2016; Letts & Mulligan, 2005). Thus, research has found that AGB in TMCF is usually lower than in their lowland counterparts and, more general, that AGB in tropical mountains declines with elevation (Asner et al., 2014; Spracklen & Righelato, 2014). However, AGB patterns along elevation transects can be complex depending on how environmental factors other than temperature change with elevation, as well as their relative roles in shaping forest structural attributes, such as stem density, tree height, wood density and basal area (Clark, Hurtado, & Saatchi, 2015). In fact, examples showing AGB increasing with elevation and TMCF exhibiting surprisingly high AGB have been described (e.g. Cuni Sanchez 2017, others), suggesting more research is needed to understand how AGB changes along different environmental gradients.

The other large source of variation in AGB patterns in tropical mountains is forest disturbance. AGB decreases with forest disturbance but increases immediately after if the conditions allow new vegetation to establish and surviving trees to keep growing. In TMCF, conversion to croplands and grazing lands for cattle represents the main source of disturbance (refs). When croplands and cattle ranches are permanently established, AGB is also permanently removed, causing large carbon losses. However, in places where swidden agriculture is practiced and cash crops are grown in agroforestry systems, as in the case of many tropical mountains, vegetation is allowed to regrow after disturbance resulting in AGB increases and, consequently, carbon gains. The coexistence of different agricultural systems and landholding sizes in tropical mountains (from small-scale farming to big cash crop plantations) results in patchy landscapes with forest at different successional stages surrounded by agricultural and grazing lands. These highly heterogeneous landscapes have been described as forest-agriculture mosaics and exhibit substantial AGB variation across space (refs).

Despite the fundamental role that land use plays in shaping carbon distribution in tropical mountains, land use effect on AGB patterns remains poorly understood (Erb et al., 2018). Considering more than 60% of forested ecosystems around the globe experience some sort of land use (Grantham et al., 2020), this represents a fundamental knowledge gap. Furthermore, this lack of understanding probably represents a large source of uncertainty in AGB estimates because AGB is usually measured at local scales in forests considered undisturbed and then extrapolated to regional scales without explicitly accounting for the effect of land use across scales.

With the aim of contributing to a better understanding of the relative roles of environmental factors and land use on shaping AGB patterns across scales, here, we focus on a TMCF region located in southern Mexico that is subjected to both an environmental gradient and a land-use intensity gradient. Three research questions guide our study: (1) How much AGB does TMCF in this region hold and how is it spatially distributed? (2) How is AGB related to tree diversity in the study region? and (3) How does land-use intensity shapes AGB patterns along an elevation gradient? We hypothesize first, that AGB in this TMCF will be larger than reported in regional and global estimates as has been the case for other cloud forests studied recently (refs), and that its spatial distribution will be driven by forest structural attributes, particularly by tree height, as in other TMCF described in the Neotropics (asner). Secondly, we expect to find a positive correlation between tree diversity and AGB as has been found in lowland tropical forests (Porter et al.,) and in Mexican TMCF along land-use intensity gradients (Vizcaino-Bravo). Finally, we predict that environmental factors and land use will have a compounding effect on AGB along the elevation gradient where both land-use intensity and cooler temperatures led by elevation will decrease AGB.

**Methods**

*Study Area*

We delimited a study area following a tropical montane cloud forest (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. 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). This region has a long history of land use with swidden agriculture, some permanent agricultural and grazing lands, and several types of coffee farms (including sun and shaded coffee) interspersed with forests, creating a forest-agriculture mosaic. To define the distribution of TMCF within NMO we used 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, climate, 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 (FI) 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 temperature and precipitation 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 systematic 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 (Figure 1). All trees, lianas, shrubs, palm trees and ferns within the plots with a diameter at breast height (DBH) larger than 7.5 cm were taxonomically identified and sampled for height, DBH, and basal area (BA). Information about the geographic location, vegetation type, and land ownership of each site was also documented. Additionally, signs of forest disturbance were assessed in each site and recorded (CONAFOR, 2018). For selecting FI sites relevant to our study, we performed a spatial intersection in QGIS 3.16 between FI sites and a NMO shapefile, the latter acquired from the National Commission of Biodiversity (CONABIO) GeoPortal (CITAS).

It is important to note that sampling was not directed towards undisturbed or ‘old-growth’ forests. On the contrary, this systematic sampling enabled the collection of data on a diversity of landscapes, many of them mosaics of different land cover classes including agricultural and grazing lands, coffee farms, and forests at different successional stages. Therefore, the Mexican FI provides a unique opportunity to test the effect of landscape composition and land use on forest structure and composition. This sampling design also made possible an analysis across spatial scales. Here, we processed data at two sampling levels. On the one hand, we used 400 m2 plots as our smallest sampling unit to describe forest structural attributes and tree diversity. With this information we associated each plot to an approximate successional stage and estimated AGB. On the other hand, we used 1-ha sites to understand AGB spatial variation in relation to environmental factors and land-use intensity.

*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.

We used allometric equations to calculate the AGB of every alive tree measured in 160 plots in 40 FI sites within our study region (Figure 1). A total of 4,106 trees belonging to 148 species were recorded. To correct for possible typos and identify synonyms in taxonomic names we collated our list of species with the Taxonomic Name Resolution Service using the correctTaxo function in R package BIOMASS (Réjou‐Méchain, 2017). Then, we searched for all possible allometric equations published in the scientific literature that would match our species list. We found 47 allometric equations described at species or genus levels (Table S2 and references therein) with which we estimated the AGB of 2,700 trees. For estimating AGB of the remaining trees whose allometric equation has not been described, we used a generic allometric equation developed by Chave et al. (2015) for tropical trees based on tree wood density (*ρ*), height (*H*) and DBH (*D*):

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

All allometric equations we used estimate AGB with a combination of trees’ DBH and height, except for Chave et al.’s generic equation (eq.1) that also requires a wood density value. DBH and height were measured in the field and are available in the FI database. We searched for the wood density value of each species or its closest relative in global wood density databases using the function BIOMASS::getWoodDensity, which provides a wood density value per tree and its associated standard deviation (calculated with repeated measurements of wood density at species, genus or family levels).

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 AGB standard deviation at plot level following error propagation through a Monte Carlo statistical simulation informed by field data using the BIOMASS::AGBmonteCarlo function. To do so, we used wood densities standard deviations, and assumed 95% of field data samples have a low DBH error and the remaining 5% a high DBH error (close to 5 cm), and that all field data samples have a height error of 10%, as suggested in Chave et al. (2004). To estimate AGB error at site level, we assumed standard error independence between plots and used the following equation (eq. 2):

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

*Tree diversity.* We used field measurements of species richness (*S*), *i.e.*, the total number of species, and species abundance to calculate Shannon (*H*) diversity index (REF) in each plot with the following equation (eq. 3):

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

Where *pi* is the proportion of species *i* and *S* is the number of species. We obtained mean *S* and *H* values per site averaging plots’ *S* and *H*. To obtain total *S* per site, we added up all species sampled in the four plots within a site, and calculated *H* again using this combined species pool. In this way, we got *S* and *H* at plot level, as well as average *H*, average *S*, total *H* and total *S* at site level.

*Environmental variables*. We focus on climate and topography and their interactions as key environmental variables moderating the effect of land-use intensity on species diversity and ecosystem AGB. We extracted annual precipitation and mean annual temperature values at site level from WorldClim (bio12 and bio1, respectively) at ~1-km spatial resolution, using the package *raster* in R (ref). 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). We averaged plot values to obtain slope, and aspect at site level.

*Land use variables*. To quantify the effect of land use on AGB and tree diversity we used three variables: (1) forest disturbance related to agricultural activities, (2) forest disturbance related to cattle grazing activities, and (3) a land-use intensity gradient we built assessing the landscape composition of each site (*i.e.*, the proportion of different types of land cover within a site) after a forest succession categorization at plot level (described below). We assessed FI sites’ forest disturbance related to agricultural and grazing activities at the time of data collection using FI disturbance data set, which contains information about the cause and severity of vegetation disturbance at site level (REF). The causes of disturbance are classified in 11 classes: fires, hurricanes, floods, roads, logging, land-use change, grazing, pests and diseases, power lines, 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.

For building a land-use intensity gradient we first identified the approximate successional stage of each forest stand at plot level for later assessing the proportion of forests at different successional stages present in each site. To assign each forest stand a successional stage, we classified all FI plots with a k-means analysis using their structural attributes, including tree height, DBH, and stem density. K-means is a non-hierarchical cluster analysis where the user defines the initial number of centers. We run the analysis with two, three, four and five initial centers and 25 random sampling sets each using R package stats (Figure S1). Then, we compared 30 indices to define the best number of clusters and chose the one that was better supported by most indices (ref nbclust). Most of the indices suggested three clusters as the best classification. Thus, from this analysis, we classified plots in three 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 2a). Other studies conducted in forest-agriculture mosaics in Mexican TMCF 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 (Velasco-Murguía et al., 2021). In general, immediately after croplands are abandoned, TMCFs naturally regenerate showing an increase in tree density, height, and basal area. Over time, tree height and basal area continue to increase but stem density decreases, representing a transition from young to mature forest that occurs around 50 years after disturbance (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 after del Castillo (2015) as follows: cluster one was defined as young fallows, DESCRIPTION; cluster two as young forest, DESCRIPTION; and cluster three as mature forest, DESCRIPTION. We would like to acknowledge that forest succession is a continuum and a complex process (REFS). Here, however, 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 young fallows, young and mature forests in each site. Interestingly, most of the sites have plots that fall across different successional stages, showing the patchiness in these forest-agriculture mosaic landscapes. To describe this patchiness, we assigned a value from 1 to 3 to each successional stage where young fallows = 1, young forest = 2, and mature forest = 3. Then, we defined a site forest proportion value (eq. 4) adding up the values of all successional stages within a site and normalizing the value to get a number from 0 to 1.

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

*SS* is the sum of the successional stage categories of the plots in a site, *minSS* is the minimum possible *SS* value present in a site and *maxSS* is the maximum possible *SS* value present in a site. Considering there are four plots in each site, *minSS* is always 4 (when all plots within a site are young fallows), and *maxSS* is always 12 (when all plots within a site are mature forests). This way, a forest proportion value from 0 to 1 is assigned to all sites, where 0 represents sites dominated by young fallows, 1 represents sites dominated by mature forests, and everything in between are sites with a combination of forests at different successional stages. Assuming sites dominated by young fallows experience greater intensity of land use and sites where most plots are classified as mature forest have experienced less land use, we estimated a land-use intensity gradient using the inverse of our forest proportion variable (eq.5). Similar approaches have been used by XPerfectoXVandermeerXEtc to describe land-use intensity in tropical landscapes, where forest cover is inversely related to agricultural intensification and, consequently, can be used as a proxy for quantifying land use at landscape scales.

|  |  |
| --- | --- |
|  | (5) |

It is important to note that other causes of forest disturbance unrelated to direct land-use can result in land-use intensity gradient values closer to 1, such as pest outbreaks. However, agriculture and cattle ranching expansion have been identified as main causes of forest disturbance in TMCF (Calderon-Aguilera et al., 2012; MORE), and there is a positive correlation between our land-use intensity 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 FI sites within this region correspond to our study system. Thus, we used the following key features of TMCF stated in the scientific literature as criteria to filter FI sites further: (1) sites should be within an elevation range of 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 (Fahey et al., 2016; Jardel-Pelaez, et al., 2014; Scatena et al., 2011; Torres, 2004).

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 160 plots within 40 sites located between 16.89 and 18.61 degrees N and -95.66 and -97.02 degrees W, within the distribution of TMCF (Figure 1).

*Statistical Analyses*

We deployed well-established statistical methods to answer each of our three main questions using two main sampling levels: plot level (n= 160), representing fairly homogeneous forest stands in an area of 400 m2; and site level (n = 40), conformed by 4 plots, representing the heterogeneous nature of these landscapes in area of 1 ha. Thus, most of the values given at site level are averages of values found at the plot level. Both questions one and two make use of plot-level and site-level data. However, because question three is concerned with patterns at larger scales (landscape to regional), it is the only one we answered with analyses carried out solely at the site level. All statistical analyses were performed in R version 4.1.1 (2021) and are described in the following sections.

*Q1. How much AGB does TMCF in this region hold and how is it spatially distributed?*

To gain a general sense of the amount of AGB in TMCF and its variation we performed basic summary statistics of all variables at site level and estimated the correlation between them. Because the plots within FI sites are at different successional stages, we were able to assess how AGB and forest structural attributes change in young fallows, young and mature forests. To test whether stem density, tree height, basal area, wood density and AGB are significantly different in forests at different successional stages, we carried out five one-way analyses of variance (ANOVA).

We analyzed which structural attribute better explained AGB distribution in these forests as well as AGB distribution across trees of different size. Research shows that in some TMCF tree height is more strongly related to AGB than other structural attributes (asner), in others, basal area and the density of very large trees play a larger role in explaining AGB distribution (cuni). We analyzed the relationship between AGB and stem density, tree height, basal area, and wood density at plot level (n= 160) using linear regressions and visualized them fitting locally weighted regression (loess) curves after log-transforming AGB.

To explore the contribution of tree size to AGB and stem density we employed 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, and >50 cm. Then, we calculated the proportion of stem density and AGB represented by each tree size class in every forest plot. We used one-way ANOVAs to assess whether tree size classes contribute to stem density and AGB in statistically different proportions. To test if the contribution of tree size classes to stem density and AGB varies between forests at different successional stages we conducted a couple of two-way ANOVAs using size class, forest successional stage, and the interaction between size class and successional stage as explanatory variables. Additionally, we conducted one-way ANOVAs to test whether the contribution of each tree size class to stem density and AGB is statistically different between forest plots at different successional stages, as well as to test whether the contribution to stem density and AGB varies between each tree size class within young fallows, young and mature forests. All statistically significant ANOVAs were followed by post-hoc Tukey tests.

*Q2. How is AGB related to tree diversity in the study region?*

To understand the relationship between AGB and tree diversity in TMCF, we performed a series of nonparametric regression analyses. 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 weak correlation, we could not assume a linear relationship. Here, we assessed the relationship between AGB and diversity using Shannon diversity index (H) and species richness (S) in TMCF plots (n= 160) and sites (n = 40) fitting loess curves. Analyzing these relationships at both sampling units offered the possibility to test, first, whether the relationship between AGB and diversity is scale-dependent (as some studies have shown e.g., Porter), and whether this relationship changes in forest stands at different successional stages. Thus, we visualized the results at plot (smaller scale) and site (larger scale) levels, considering the successional stage of each plot.

*Q3. How does land-use intensity shapes AGB patterns along an elevation gradient?*

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 environmental 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). Slope is not related to PC1 and thus is not represented in this environmental gradient. 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 statistic 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.

**Results**

*AGB magnitude and distribution in TMCF in the NMO*

Average tree AGB in TMCF sites in the NMO is 137.49 ± 121.29 Mg ha-1 and it ranges from as low as 8.26 to as high as 414.52 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. 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 while others almost reach 2,000 trees ha-1 (Table 1). Similarly, the variation in tree height is large, going from about 5 to 25 m (Table 1).

The large variation in AGB and structural attributes found in TMCF sites stems from the diversity in landscape composition found in these forest-agriculture mosaics and it is driven by the successional stage of each plot. In young fallows, tree density is low, and trees are short and thin (Table 2 and Figure 2). As forest succession develops, all these structural attributes increase. Thus, young forests show larger trees and higher stem density than young fallows (Table 2 and Figure 2). In mature forests, the density of trees decreases as trees become even taller and bigger (Table 2 and Figure 2). As a result, AGB in forests at different successional stages is significantly different (ANOVA, p< 0.001, Figure 2f Table S3). In young fallows, 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). On average mature forests hold more than twice the biomass found in young forests, and about ten times more than young fallows. However, both young and mature forests contain similar proportions of the total measured AGB, due to the dissimilarity in the number of plots classified in each successional stage, accounting for about 47 and 43%, respectively. In contrast, young fallows account for the remaining 10% of the total AGB measured in these 160 plots. The ANOVAs we performed to analyze structural differences among forests at different successional stages show that all attributes are statistically different in all three categories (young fallows, young forest, and mature forest) (Table S3). Wood density is the only structural attribute that is not statistically different between young and mature forests, although it is significantly lower in young fallows (Figure 2 and Table S3).

The structural attribute that better explains AGB distribution is basal area (linear regression, p < 0.001, adjusted R2= 0.86, Figure 3b). Despite tree height and stem density have a positive relationship with AGB, larger values show ample variation and data spreads out at the higher end of the curves, resulting in poor linear fits (Figure 3). Log-transforming AGB improves data visualization in these cases, highlighting that the relationships of stem density and tree height with AGB in TMCF is not linear (Figure?). In fact, plots with the highest stem densities are not the ones showing greater AGB. On the contrary, AGB is greater in plots with stem densities closer to 1,000 trees ha-1 (Figure 3a). In contrast, the highest trees are found in plots with the greatest AGB, but variation in these plots is large and some show high AGB magnitudes (greater than 400 Mg ha-1) with trees shorter than 20 m (Figure 3c). Wood density does not show a clear relationship with AGB, suggesting there are trees with high and low density across all plots (Figure 3d). There are not plots with high AGB whose trees show only low wood density, but this trend is probably controlled, at least partially, by the number of trees in each plot.

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 (Figure S4a and Table S4). 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, FigureS4b and Table S4). Interestingly, this pattern changes in forest plots at different successional stages. The two-way ANOVAs on tree size contribution to stem density and AGB show that both size class and successional stage are statistically significant, as well as the interaction between them (Table S5 and Figure S5). Although the proportion of stems across tree size classes remains somewhat constant between young fallows, young forests, and mature forests (Figure 4a), their contribution to AGB is variable (Figure 4b, Table S6 and Table S7). In mature forests, larger trees (DBH > 50 cm) stand out as the main contributors to total AGB despite their low abundance. However, in young and very young forests the contribution to total AGB is very similar across all size classes. In all cases, small trees (DBH < 10 cm) contribute significantly less to total AGB despite representing a considerable proportion of stem density in all three successional stages, especially in mature forests, where the proportion of AGB represented by the smallest trees barely reaches 1% (Figure 4b).

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). Young fallows have small AGB and small number of trees. Thus, the few but large trees that do exist in these plots represent a large proportion of the total amount of AGB (although this is not statistically different from the contribution to AGB by other size classes). In young forests the number of large trees is also small, but stem density in general is higher than in young fallows. Because there are many small to medium trees, they represent the bulk proportion of AGB, and very large trees do not stand out as significant contributors to total AGB. In contrast, in mature forests the number of trees is lower but larger trees are more common, and thus larger trees stand out as the main contributors to AGB (Figure 4).

*Relationship between AGB and tree diversity*

We found a total of 148 tree species in the region, being *Quercus*, *Saurauia*, and *Pinus* the most abundant genus. 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 show a positive but weak correlation with correlation coefficients near 0.3 (Figure S2). We addressed this issue using nonparametric regressions, where linearity is not assumed. Using Shannon diversity index (H) and species richness (S), we found that at site level H is very variable and does not have a clear relationship with AGB. S, on the other hand, increases with AGB until sites reach approximately 200 Mg ha-1 and then decreases at higher AGB values (Figure 5a and 5b). Thus, sites with more than 200 Mg ha-1 tend to show lower tree diversity. This probably occurs in forests where large dominant trees ─which are the main contributors to total AGB─ are more abundant, decreasing evenness and diversity in a site.

When we analyzed the relationship between AGB and tree diversity at plot level, we found a clearer relationship between H and AGB, and a similar trend than S at site level (Figure 5c and 5d). In both cases, tree diversity increases until reaching a point (at about 200 Mg ha-1) where diversity tends to decrease at higher AGB values. The analysis at plot level also shows a trend between AGB and tree diversity over different successional stages. Tree diversity increases in plots going from young fallows to young forests, but it slightly decreases in mature forests (Figure 5c and 5d). Thus, sites with a larger composition of mature forests show greater AGB but not necessarily greater tree diversity, suggesting these two variables follow slightly different trends in forest-agriculture mosaics.

*Patterns of AGB along 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: land-use intensity gradient, forest disturbance related to agriculture, and slope (p < 0.05, adjusted R2= 0.811, Table 4, Figure S6). From the three variables, land-use intensity controls AGB the most, showing a strong relationship with AGB (linear regression, p < 0.05, adjusted R2= 0.729, Figure 6a). Adding forest disturbance driven by agriculture to the model improves its explanatory power and it seems to be the second most relevant predictor of AGB (Table 3). Similarly, adding slope improves the linear regression model and it is the third most relevant variable in all models. However, from the three predictors included in the first model, slope is the least influential in determining AGB patterns in TMCF (Table 4). The residuals of this model do not show spatial autocorrelation (Figure S6d). 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, Table S8).

The second best model includes six predictors (Table 3). It adds environmental gradient, disturbance by grazing, and Shannon diversity index (H) to the three most relevant predictors included in the first model (i.e., land-use intensity, disturbance by agriculture, and slope). Although this model shows high R2 and the lowest CP, one of the predictors, H, is only marginally significant. Moreover, it only shows a slightly better R2 than the best model, a higher BIC value, and variables that are not always significant, as in the case of land-use intensity, disturbance by agriculture, and slope. Despite adding H to the model improves its explanatory power, both H and S are the least relevant predictors of AGB. This resonates with the non-linear relationship between tree diversity and AGB explained in the previous section, and it is supported by the dissimilar behavior of the relationships between land-use intensity and AGB and the former with H (Figure 6), where we show the strong relationship between land-use intensity and AGB and that the linear regression between land-use intensity gradient and H is not significant (p= 0.70).

Interestingly, environmental gradient (which comprises temperature, precipitation, and elevation) is not included as a relevant predictor in the best model. Although the second best model does environmental gradient, this variable does not seem to fundamentally control AGB in TMCF landscapes within the region and it is significant only when disturbance by grazing and tree diversity are also included (Table 3). The environmental gradient variable has a statistically significant relationship with AGB (p < 0.05, adjusted R2= 0.185, Figure 7a) 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 relatively cooler and drier climate. Additionally, this environmental gradient is related to the land-use intensity 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 7b). 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.

**Discussion**

* AGB is stored in very large trees in mature forests but in medium trees in young forests.
* Variation in AGB allocation arises from structural differences: mature forests in large trees, secondary forests in high stem density
* Secondary forests can hold relatively large amounts of AGB (150 Mg ha-1 on average) but the ones storing more AGB are mature forests.
* 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 (within sites of 1 ha) determines AGB patterns at landscape and regional scales.
* The effect of climate on AGB is overshadowed by land use even at regional scales.
* AGB and tree diversity do not have a linear relationship, which means they behave in a different way as landscape composition changes. -- forest disturbance by land use probably has different effects on these two ecosystem services.
* 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. -> diversity index values can change due to inclusion of trees smaller than DBH=7.5cm (like in Mejia et al., 2004).
* Structural attributes fall within ranges reported for other TMCF in Mexico, and the world?
* AGB values found in this region are underestimated in global maps of AGB
* . As elevation increase, forests experience a series of environmental features that generally result in lower primary productivity, such as frequent cloud cover and fog, waterlogged soils, and lower temperatures (Gotsch et al., 2020). Thus, tropical forests located above 1,000 m asl are expected to show less AGB than their lowland counterparts.

**Conclusions**

**Tables**

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

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Min ± S.E.** | **Max ± S.E.** | **Mean ± S.D.** |
| ***Forest structural attributes and AGB*** | | | |
| 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 |
| Average wood density (g cm-3) | 0.32 ± 0.02 | 0.76 ± 0.01 | 0.56 ± 0.14 |
| Aboveground biomass (Mg ha-1) | 8.26 ± 1.02 | 414.52 ± 19.29 | 137.49 ± 121.29 |
| ***Tree diversity*** | | | |
| Species richness (average) | 1.75 ± 0.25 | 9 ± 1 | 4.4 ± 1.63 |
| Species richness (total) |  |  |  |
| Shannon (average) | 0.33 ± 0.21 | 1.75 ± 0.25 | 1.06 ± 0.36 |
| Shannon (total) |  |  |  |
| ***Environmental variables*** | | | |
| Mean annual temperature (◦C) |  |  |  |
| Annual precipitation (mm) |  |  |  |
| Slope (degrees) |  |  |  |
| Elevation (m asl) |  |  |  |
| ***Land-use variables*** | | | |
| Disturbance by agriculture |  |  |  |
| Disturbance by cattle grazing |  |  |  |
| Land-use intensity |  |  |  |

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

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **F (n= 69)** | | **Y (n= 62)** | | **M (n= 29)** | |
| **Mean** | **S.D.** | **Mean** | **S.D.** | **Mean** | **S.D.** |
| Stem density (tree ha-1) | 275.00 | 166.55 | 986.69 | 444.38 | 775.86 | 313.14 |
| Basal area (m ha-1) | 6.87 | 4.43 | 24.75 | 9.33 | 43.42 | 13.28 |
| Lorey’s height (m) | 9.37 | 3.55 | 11.76 | 2.36 | 21.31 | 3.12 |
| Average wood density (g cm-3) | 0.49 | 0.13 | 0.57 | 0.08 | 0.57 | 0.08 |
| AGB (Mg ha-1) | 33.38 | 32.63 | 151.69 | 92.04 | 354.86 | 151.60 |

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 |
| Land-use intensity | \*\*\* | \*\*\* | \*\*\* | \*\*\* | \*\*\* | \*\*\* | \*\*\* |
| Disturbance by agriculture |  | \*\* | \*\* | \*\*\* | \*\*\* | \*\*\* | \*\*\* |
| Slope |  |  | \*\* | \* | \*\* | \*\* | \*\* |
| Environmental gradient |  |  |  | 0.16 | 0.13 | \* | 0.10 |
| Disturbance by grazing |  |  |  |  | 0.16 | \* | \* |
| Shannon diversity index |  |  |  |  |  | ◦ | ◦ |
| Species richness |  |  |  |  |  |  | 0.30 |
| Adjusted R2 | 0.729 | 0.769 | 0.811 | 0.816 | 0.821 | 0.833 | **0.834** |
| CP | 26.089 | 17.600 | 8.928 | 8.746 | 8.606 | **7.089** | 8.000 |
| BIC | -45.958 | -49.672 | **-55.251** | -53.753 | -52.336 | -52.689 | -50.339 |

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, land-use intensity gradient, and disturbance by agricultural activities.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Coefficient** | **S.E.** | **t** | **p** |
| Slope | 0.027 | 0.009 | 3.064 | 0.004 \*\* |
| Land-use intensity | -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. Forest Inventory sites (n= 40, black points) within Tropical Montane Cloud Forest (TMCF) distribution (in green) in the Northern Mountains of Oaxaca (NMO), located in the south of Mexico. A zoomed-in site shows the hierarchical nested sampling design carried out by the Forest Inventory (FI) where four plots of 400 m2 where established. |

|  |
| --- |
| Gráfico, Gráfico de cajas y bigotes  Descripción generada automáticamente |
| Figure 2. a) Results of k-means analysis visualized with a PCA showing three clusters representing young fallows (F, in orange), young forest (Y, in purple), and mature forest (M, in green) in TMCF plots (n= 160) that were classified based on their structural attributes (stem density, basal area, and Lorey’s height) measured in the field. Comparison of b) stem density, c) Lorey’s height, d) basal area, e) wood density, and f) aboveground biomass (AGB) between young fallows (F), young forest (Y), and mature forest (M). Boxes cover the interquartile range (IQR), the horizontal line within boxes shows the median, and values 1.5 times larger or smaller than the IQR are shown in dark gray points. Asterisks indicate statistically significant differences resulted from Tukey HSD tests as follows: \*\*\*\*p < 0.0001, \*\*\*p < 0.001, \*\*p < 0.01, \*p<0.05, and ns represents a non-significant difference. Results of ANOVAs and Tukey HSD tests can be found in Table S3.   |  | | --- | | Gráfico, Gráfico de dispersión  Descripción generada automáticamente | | Figure 3. | |

|  |
| --- |
| Gráfico, Gráfico de cajas y bigotes  Descripción generada automáticamente |
| Figure 4. Contribution of tree size categories to a) stem density and b) aboveground biomass (AGB) in young fallows (F, shown in orange), young forest (Y, shown in purple), and mature forest (M, shown in green) plots in TMCF (n= 160). Trees were categorized in six size classes based on their DBH. Boxes cover the interquartile range (IQR), the horizontal line within boxes shows the median, and values 1.5 times larger or smaller than the IQR are shown in dark gray points. 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 young fallows (F, shown in orange), young forest (Y, shown in purple), and mature forest (M, shown in green) 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 5. 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: young fallows (F) in orange, young (Y) in purple, and mature (M) forests in green. |

|  |
| --- |
| Gráfico, Gráfico de dispersión  Descripción generada automáticamente |
| Figure 6. Relationship of land-use intensity with a) aboveground biomass (mean ± SE) and b) tree diversity estimated with Shannon diversity index (H) in TMCF sites (n= 40) overlaid by their linear regression curves (black lines). Landscapes dominated by mature forests are at the lower end of the land-use intensity gradient and those dominated by very young forests have high land-use intensity values. See text for details on how the land-use intensity gradient was calculated. |

|  |
| --- |
| Gráfico, Gráfico de dispersión  Descripción generada automáticamente |
| Figure 7. Linear regression curves between a) aboveground biomass (AGB) (p < 0.05, adjusted R2= 0.185), and b) land-use intensity (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. |

**Supporting Information**

**SI Tables**

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 (n= 160) at different successional stages (young fallows (F), young forest (Y), and mature forest (M)). Significant p-values are shown in bold.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Response | ANOVA | | Tukey HSD | | | |
| F | *p* | Comparison | Estimate | 95% CI | *p* |
| Stem density | 80.86 | **<2e-16** | F-Y | 711.69 | (576.74, 846.64) | **9.50e-14** |
| F-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** | F-Y | 2.38 | (1.12, 3.65) | **4.63e-05** |
| F-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** | F-Y | 17.88 | (14.32, 21.43) | **9.81e-14** |
| F-M | 36.54 | (32.04, 41.04) | **9.41e-14** |
| Y-M | 18.66 | (14.09, 23.24) | **1.22e-13** |
| Wood density | 9.77 | **<0.001** | F-Y | 0.07 | (0.03, 0.12) | **3.74e-4** |
| F-M | 0.08 | (0.02, 0.13) | **2.87e-3** |
| Y-M | 0.005 | (-0.05, 0.06) | 0.96 |
| AGB | 135.70 | **<2e-16** | F-Y | 118.30 | (81.62, 154.99) | **6.50e-12** |
| F-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 10-20 cm; class 3: DBH 20-30 cm; class 4: DBH 30-40 cm; class 5: DBH 40-50 cm; class 6: DBH > 50 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 (young fallows (F), young forest (Y), and mature forest (M)), and their interaction 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 10-20 cm; class 3: DBH 20-30 cm; class 4: DBH 30-40 cm; class 5: DBH 40-50 cm; class 6: DBH > 50 cm. Significant p-values are shown in bold. Tukey HSD results in Github repository.

|  |  |  |  |
| --- | --- | --- | --- |
| Response | Explanatory | 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 on the contribution to stem density and aboveground biomass (AGB) in each tree size class between forest plots at different successional stages (young fallows (F), young forest (Y), and mature forest (M)). Trees were classified in six size classes according to their DBH as follows: class 1: DBH < 10 cm; class 2: DBH 10-20 cm; class 3: DBH 20-30 cm; class 4: DBH 30-40 cm; class 5: DBH 40-50 cm; class 6: DBH > 50 cm. Significant p-values (i.e., p < 0.05) are shown in bold. Tukey HSD results in Github repository

|  |  |  |  |
| --- | --- | --- | --- |
| ANOVA | | | |
| Response | | F | p |
| Contribution to stem density within size classes between forest successional stage | Size class 1 | 21.28 | **9.05e-09** |
| Size class 2 | 2.20 | 0.11 |
| Size class 3 | 2.15 | 0.12 |
| Size class 4 | 8.17 | **5.29e-04** |
| Size class 5 | 7.93 | **9.22e-04** |
| Size class 6 | 6.84 | **2.00e-03** |
| Contribution to AGB within size classes between forest successional stage | Size class 1 | 11.54 | **2.33e-05** |
| Size class 2 | 12.39 | **1.03e-05** |
| Size class 3 | 12.66 | **9.89e-06** |
| Size class 4 | 16.41 | **7.37e-07** |
| Size class 5 | 26.12 | **9.67e-09** |
| Size class 6 | 3.37 | **0.04** |

Table S7. Results of one-way ANOVA on the contribution to stem density and aboveground biomass (AGB) in forest plots at different successional stages (young fallows (F), young forest (Y), and mature forest (M)) between tree size classes. Trees were classified in six size classes according to their DBH as follows: class 1: DBH < 10 cm; class 2: DBH 10-20 cm; class 3: DBH 20-30 cm; class 4: DBH 30-40 cm; class 5: DBH 40-50 cm; class 6: DBH > 50 cm. Significant p-values (i.e., p < 0.05) are shown in bold. Tukey HSD results in Github repository

|  |  |  |  |
| --- | --- | --- | --- |
| ANOVA | | | |
| Response | | F | *p* |
| Contribution to stem density within forest successional stage between size classes | F | 18.96 | **2.09e-15** |
| Y | 115.06 | **5.18e-64** |
| M | 54.38 | **4.47e-32** |
| Contribution to AGB within forest successional stage between size classes | F | 11.89 | **4.44e-10** |
| Y | 21.31 | **6.95e-18** |
| M | 46.52 | **6.47e-29** |

Table S8. Moran’s I statistics for multiple linear regression selected as the best model to explain aboveground biomass patterns in TMCF for neighborhoods of 10, 25, and 50 km of distance between sites (n = 40). This model includes three predictors: slope gradient, landscape composition, and disturbance by agricultural activities.

|  |  |  |
| --- | --- | --- |
| Distance between sites (km) | Moran’s I | *p* |
| 10 | -0.049 | 0.51 |
| 25 | -0.05 | 0.60 |
| 50 | 0.02 | 0.18 |

**SI Figures**

|  |
| --- |
| Mapa  Descripción generada automáticamente |
| Figure S1. K-means cluster analysis on structural attributes (stem density, Lorey’s height, and basal area) of TMCF plots showing two (to left), three (top right), four (bottom left) and five (bottom right) clusters. By comparing the four possible classifications with 30 indices, the three cluster classification was selected as the best one based on the majority rule. These three clusters match the expected structure found in young fallows, young forest, and mature forest in TMCFs. |

|  |  |
| --- | --- |
|  |  |
| Figure S2. Correlation matrices showing the correlation between all pair of variables used in this study at site level (n= 40) in a graphic (left) and a numeric (right) way. Red colors indicate negative relations and blue indicate positive relations. | |

|  |
| --- |
|  |
| Figure S3. Principal component analysis (PCA) of climatic (temperature and precipitation) and topographic variables (slope and elevation) showing the two first principal components (PC1 and PC2). PC1 and PC2 explain 92% of the variation. PC1 explains 68% of variation and shows that there is a positive correlation between temperature and precipitation and a negative correlation between these two and elevation. Thus, PC1 represents an environmental gradient from warmer and wetter sites at lower elevations (negative values) to cooler and drier sites at higher elevations (positive values). Slope is not correlated to the other three variables, and it is the variable driving PC2. |

|  |
| --- |
| a) |
| b) |
| Figure S4. Contribution of tree size categories to a) stem density and b) aboveground biomass (AGB) 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 10-20 cm; class 3: DBH 20-30 cm; class 4: DBH 30-40 cm; class 5: DBH 40-50 cm; class 6: DBH > 50 cm. Letters indicate statistically significant differences between tree size classes assessed with a one-way ANOVA and Tukey HSD test (results shown in Table S4). Boxes sharing a letter are not statistically different. |

|  |
| --- |
| Gráfico, Gráfico de cajas y bigotes  Descripción generada automáticamente |
| Figure S5. Contribution of tree size categories to a) stem density and b) aboveground biomass (AGB) in young fallows (F, shown in orange), young forest (Y, shown in purple), and mature forest (M, shown in green) plots in TMCF (n=160). Trees were categorized in six size classes based on their DBH. Boxes cover the interquartile range (IQR), the horizontal line within boxes shows the median, and values 1.5 times larger or smaller than the IQR are shown in dark gray points. Letters indicate statistically significant differences between tree size classes and forest succession assessed with a two-way ANOVA and Tukey test (ANOVA results shown in Table S5). Boxes sharing a letter are not statistically different. |

|  |  |
| --- | --- |
| a) | b) |
| c) | d) |
| Figure S6. Partial residual plots of multiple linear regression selected as the best model to explain aboveground biomass patterns in TMCF based on R2, Mallows' Cp (CP), andBayesian Information Criterion (BIC) with a stepwise model selection process. Plots show the three predictors included in the model: a) slope, b) land-use intensity gradient, and c) disturbance by agricultural activities. Panel d displays the model residuals according to their geographic location to show there is no spatial autocorrelation between them (Moran’s I statistics shown in Table S8). | |