**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?

How much AGB can TMCF hold and how is it distributed (big trees / tree density / etc)

Interaction between environmental gradient and AGB

Interaction between AGB and diversity

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 that country 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 1 for a summary of the variables).

*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 main sampling unit. 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. All sites were established 25 km apart from each other in a grid-like fashion. For selecting FI sites relevant to our study, we performed an intersection of sites within the NMO using the open software QGIS 3.16.

*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 2). 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 developed by Chave *et al.* (2014) for tropical trees based on tree wood density, height and DBH:

AGB = 0.0673 \* (WD \* H \* D^2)^0.976

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. Functions to correct taxonomic names and search for wood density values are available in 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). We assumed standard error independence to estimate site’s AGB error, and used the following equation:

AGBsite = (Eplot1^2+Eplot2^2+Eplot3^2+Eplot4^2)^1/2

*Tree Diversity.* We calculated species richness, *i.e.*, the total number of species, Shannon (H) and Simpson (D) diversity indices with the following equations, respectively:

H = -sum p\_i log(p\_i )

1-D, where D = sum p\_i^2

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 went over 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 (Supporting Information-Figure 1). 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 averaging and normalizing all successional stages within each site as follows:

Landscape composition = 1 – (sum(SS) - min SS / max SS - min SS)

Where SS is the successional stage category and can take the following values: very young forest = 1; young forest = 2, and mature forest = 3. 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 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 the region (CITAS), so other sources of disturbance were excluded from the analysis. Moreover, 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, which supports our approach.

*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 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*

**Results**

Tree AGB and Diversity in TMCF

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

**Discussion**

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.

**Conclusions**

**Tables**

Table 1. Variables used in this study

Table 2. Allometric equations

Table 3. Structural attributes, tree AGB and diversity

Table 4. Multiple regression model-AGB

Table 5. Multiple regression model- diversity

Table 6. GAM

**Figures**

Figure 1. Map

Figure 2. Structural attributes and tree size contribution to AGB and stem density.

Figure 3. Relationship between forest structure and diversity

Figure 4. Relationship between AGB and landscape composition (mosaic), and diversity and landscape composition (mosaic)

Figure 5. GAM result or surface plot

**Supporting Information**

Non-hierarchical cluster analysis – plots categories in three ‘successional stages’

PCA of environmental variables