**Title.**

**Abstract.**

**Introduction**

We are modeling for exploration.

Goals: to understand the effect of forest conversion to agricultural lands over space (landscape and regional scales) and time (forest succession patterns). How does forest disturbance interact with environmental variables known to shape biomass and diversity such as elevation, precipitation, and aspect?

**Methods**

1. **Study Area**
   1. **Tropical Montane Cloud Forest**

In Mexico, TMCF is distributed discontinuously along the country’s main mountain ranges at elevations that go from 600 to 2,800 m asl (Jardel-Peláez et al. 2014). Features that characterize this ecosystems−besides frequent cloud cover−, include high precipitation, usually above 1,000 mm per year, and abundant presence of epiphytes. To select FI sites within TMCF, we first delimited a study area following a cloud forest regionalization conducted by Toledo-Aceves *et al.* in 2011 based on geomorphology, forest cover, watershed margins, rivers, and cultural differences (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). Shapefiles of these TMCF regions are available through CONABIO’s GeoPortal and FI database provides geographic coordinates of all sites and plots sampled. Using QGIS 3.16 we selected all FI sites within the NMO (n= 730).

* 1. **Tropical Montane Cloud Forest Sites Selection and Description**

FI data collection was carried out between 2009 and 2014 following a systematic sampling method where all sites are equidistant. Each site is a 1 ha-circular area where four circular plots of 400 m2 are 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). Height, diameter at breast height (DBH), basal area (BA), canopy area, and taxonomic identification of all trees with a DBH larger than 7.5 cm were recorded in all plots. Shrubs, palms, and tree ferns were also measured. In total, 145,544 individuals were sampled in the NMO, including trees−dead and alive−, shrubs, tree ferns, palm trees and lianas. Information about the location (state, municipality, geographic coordinates, and altitude), land tenure (whether the site is on federal, private, communal or *ejido* land), epiphyte presence and forest disturbance of each site is also included in the FI data base.

Because TMCF has a scattered distribution along the NMO, not all selected sites 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. Although TMCF in Mexico can be found at lower elevations, internationally, 1,000 m asl is often used as a threshold between lowland and montane forests and thus sites below this threshold were excluded from this study; (2) sites should receive at least 1,000 mm annual precipitation; (3) sites should be described as *‘bosque mesófilo de montaña’* either by the Mexican National Forestry Commission (CONAFOR) or the National Institute of Statistics and Geography (INEGI); (4) all sites must have epiphytes. Information about altitude, vegetation type, and epiphytes is included in FI database. 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 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 performed data quality control homogenizing missing data values, correcting names, removing diacritics, and filtering out rows with missing information. We removed plots with many unidentified species or where most trees were dead. After applying these filters, we ended up with a dataset of 7,196 sampled trees, in 302 plots corresponding to 107 sites located between 16.82808 and 18.75114 degrees N and -95.25731 and -97.06556 degrees W, within the distribution of TMCF (Figure 1). Because some plots are in places where is impossible to collect field data (such as very steep ravines), not all sites have four plots. Out of 107 sites, 83 have at least two plots, 67 have at least three, and 45 have all four plots. Table 1 provides a list of all variables used in this study, indicating the source of each variable, aggregated in three levels: individual, plot, and site.

1. **Forest Structure in FI Plots**

Forest structural attributes, species composition and diversity can provide information about the state of a forest stand, such as its successional stage, and the effects of disturbance on relevant ecological features like carbon storage and biodiversity. Based on FI raw data, we derived three structural attributes: (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 (Table 1). This information was later used to estimate forest aboveground biomass (AGB) and an approximate successional stage of each site.

**2.1. Aboveground Biomass**

AGB is defined as the dry mass of the aboveground component (i.e., excluding roots) of live or dead plants and it is expressed in mass per unit area. AGB of tropical trees represents a large carbon pool, as such, it is recognized as an essential climate variable and an important input to Earth system models (CITAS). We calculated AGB using allometric equations. Allometric equations are derived through the comparison of destructive measurements and tree’s structural attributes such as height and DBH (Gibbs et al., 2007). They are usually species- or ecosystem-specific, although some generic equations have been developed (e.g., Chave et al., 2005 and Brown, 1997). When available, we used allometric equations at species or genus level published in the scientific literature (Table 2). Otherwise, we used a generic allometric equation developed by Chave *et al.* (2014) for tropical trees. To calculate AGB we used the R package BIOMASS (Réjou‐Méchain, 2017). First, we corrected taxonomic names using the correctTaxo function, which collates a species list with the Taxonomic Name Resolution Service (<https://tnrs.biendata.org/about/>). Then, we estimated wood density using the getWoodDensity function which searches the wood density value of each species or its closest relative in global wood density databases. Finally, using mean wood density values, DBH, and height, we estimated AGB with the function computeAGB, that uses the following equation (Eq. 4 in Chave et al., 2014): AGB = 0.0673 \* (WD \* H \* D^2)^0.976. 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 using the function AGBmonteCarlo 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). Assuming plot AGB standard error independence, we estimated site’s AGB error with the following equation: AGBsite = (Eplot1^2+Eplot2^2+Eplot3^2+Eplot4^2)^1/2

1. **Forest Composition and Diversity in FI plots**

To analyze forest composition and diversity, we constructed a species abundance matrix with the package *vegan* in R, where rows correspond to sites and columns to species. We calculated species richness, Shannon (H) and Simpson (D) diversity indices by plot and site using the function diversity in the package *vegan*. Species richness is the total number of species per site. Shannon and Simpson diversity indices are defined by the following equations, respectively:

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

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

We compared species composition among sites using a hierarchical cluster analysis and a principal component analysis using only sites where four plots were sampled and information on disturbance was complete (n= 40). Additionally, we computed rank-abundance curves with the R package geoveg. Rank-abundance curves provide a visual display of relative species abundance and thus, are useful for comparing species richness and evenness across sites.

1. **Forest Disturbance and Forest Succession**

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, when even though the cause of impact is present, the quality and quantity of forest resources is not affected; low, when the negative effects caused to forest resources are not permanent, and these can be recovered without human intervention; medium, when the damage to forest resources is not permanent but human intervention is required to reverse the degradation process in addition to eliminating the causes; and high severity, when vegetation requires the implementation of extensive restoration measures over a considerable period of time. Both the cause and severity of disturbance where qualitatively assessed during field data collection (CONAFOR, 2018). We identified and extracted data related to agricultural and cattle grazing activities from the disturbance data set. Agriculture and grazing activities were reported within the categories of land-use change, grazing, logging, and fires. 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 identify an approximate successional state of FI sites, we analyzed patterns in forest structural attributes. First, we classified FI plots with a non-hierarchical cluster analysis using structure attributes, including tree height, DBH, and tree density, and visualized this classification with a PCA using R packages stats, factoextra, and vegan. To define the best number of clusters we used R package NbClust. NbClust compares 30 indices and suggests the best number of clusters based on the majority rule. From this analysis, we obtained three structural clusters. A first cluster groups together plots with very low tree density, low basal area, and low tree height; the second cluster groups together 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 2). 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, TMCF 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 stands (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 mosaic value averaging and normalizing all successional stages within each site. We assigned a value from 0 to 1 to all sites, where 0 represents sites where all plots are very young forest stands, 1 represents sites where all plots are mature forests, and everything in between are sites with a combination of forests at different successional stages.

1. **Statistical Analyses: Patterns of Tree Biomass and Diversity in Relation to Environmental Variables, Forest Succession and Forest Disturbance**

We explored the relationships among tree biomass, diversity, environmental variables, and forest disturbance and patchiness in forest-agriculture mosaics using multiple linear regression models. First, we fitted a model with AGB as the response variable and tree diversity, environmental factors, forest disturbance and patchiness as predictors. Because data is not normally distributed, we log-transformed AGB. Then, following a similar approach 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 carried out model selection using the R function regsubsets from the package leaps. To test for spatial autocorrelation, we calculated the Moran’s I of the residuals of the final model and visualize the spatial distribution of residuals with a map. We repeated this approach using Simpson diversity index as predictor variable instead of tree AGB.

Next, we filtered out sites where disturbance was recorded and very young forest was prominent (i.e., sites where mosaic < 0.5) to test the relationship between tree AGB and diversity with environmental factors without the effect of land-use. We performed two multiple regression models with all environmental variables as predictor variables, one using AGB as response variable, and another one with tree diversity as response variable. As with previous models, we calculated Moran’s I of the residuals and mapped them to test for spatial autocorrelation.

To explore the effect of disturbance and patchiness on community composition we conducted a redundancy analysis (RDA) where mosaic, grazing, and agriculture constrained the ordination of species. Additionally, we visualize the relationship between disturbance and patchiness adding these variables to the hierarchical cluster analysis and PCA we performed with all sites with four plots (see methods section 3 above). Finally, we conducted another RDA using sites with only forest (i.e., sites where mosaic > 0.05 and no disturbance was recorded) and environmental factors, to test the effect of environment on community composition without the direct effect of land-use.

**Results**

1. **Tree AGB and Diversity in TMCF**

Tree AGB in TMCF in the NMO ranges from as low as 5.4234 Mg/ha to as high as 414.5241 Mg/ha in sites where at least three plots were sampled (Table 3). However, from the 302 plots analyzed, 50 plots exhibited values below or above this range. Specifically, 33 plots showed values lower than 5.4234 Mg/ha, several of which were completely devoid of trees larger than 7.5 cm of DBH. Additionally, 17 plots exhibited higher values than the highest averaged AGB reported in sites with more than three sampled plots, some of them surpassing 500 Mg/ha and a couple of them reaching 700 Mg/ha. Variation in structural attributes and AGB within and among sites is large. For instance, average stem density in all 67 sites was 605.47 but some sites had as few as 16.67 trees/ha.

|  |  |  |
| --- | --- | --- |
| 16.67 ± 8.33 | 1806.25 ± 444.10 | 605.47 ± 358.82 |

in 107 sites. A total of 275 tree species were recorded in all 107 sites.

1. **Effects of Land-use on Tree AGB and Diversity**

Some plots in the young forest cluster may actually be mature forests but ended up in the young forest class due to having trees of smaller stature (which is not rare in TMCF).

1. **Environmental Drivers of Tree AGB and Diversity**

**Tables**

Table 1. Variables used in this study aggregated by site, plot, and tree.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Level** | **Variable** | **Units** | **Explanation** | **Source** |
| Site | Site ID | - | Unique identifier of site | FI |
| Date | - | Date of field data collection | FI |
| State | - | State where the site is located | FI |
| Municipality | - | Municipality where the site is located | FI |
| Land tenure | - | Whether private property, communal, or *ejido* land | 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 |
| Simpson (D) | Probability | 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 |
| Mosaic | - | Site’s patchiness, goes from 0, when all plots in site are very young forests or agricultural lands, to 1, when all plots in site are mature forests | Calculated using plot’s successional stage |
| 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 | 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.  H = -sum p\_i log(b) p\_i | Calculated using FI raw data |
| Simpson (D) | Probability | Diversity index.  1-D, where D = sum p\_i^2 | Calculated using FI raw data |
| 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 |
| 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 data |
|  | 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 data |
| Tree | State | - |  | FI |
| Municipality | - |  | FI |
| 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 | 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 2. Generic and specific allometric equations used in this study to calculate AGB.

|  |  |  |
| --- | --- | --- |
| **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 \* D^2)^0.976 | Chave et al., 2014 |

Table 3. Structural attributes, tree AGB and tree diversity in FI sites with at least three sampled plots (n = 67) in the TMCF of the NMO.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Min ± s.e.** | **Max ± s.e.** | **Mean ± s.d.** |
| Stem density | 16.67 ± 8.33 | 1806.25 ± 444.10 | 605.47 ± 358.82 |
| Basal area | 1.56 ± 1.08 | 51.08 ± 6.57 | 18.22 ± 12.85 |
| Lorey’s height | 5.80 ± 0.24 | 24.91 ± 1.00 | 11.64 ± 4.26 |
| Aboveground biomass | 5.423 ± 0.57 | 414.52 ± 19.29 | 111.25 ± 108.73 |
| Species richness | 2 | 18 | 7 |
| Simpson diversity index | 0 | 0.79 | 0.53 |
| Shannon diversity index | 0 | 1.83 | 1.03 |

**Figures**

|  |
| --- |
|  |
| Figure 1. Map of TMCF distribution (teal) in the NMO (gray) and location of FI sites. FI sites sampling design (top right). |

|  |
| --- |
|  |
| Figure 2. PCA showing the three successional stages resulted from the non-hierarchical cluster analysis. The three clusters are defined by tree density, basal area, and tree height, where group 1 (black hollow points) has very little trees and basal area, group 2 (red points) has higher tree density, but trees are not very big, and group 3 (green points) where plots have big trees. |

----------------------------------------------------------------------------------------------------------------------------------- We assessed forest disturbance through time with Normalized Difference Vegetation Index (NDVI) time series. NDVI normalizes the difference between red light and near infrared light reflection using the following equation: (BNIR - BRed)/(BNIR + BRed), where low NDVI values indicate stressed or dead vegetation and high NDVI values indicate healthy vegetation. We extracted NDVI time series per plot (n= 302) from atmospherically corrected surface reflectance images acquired by Landsat-5 TM and Landsat-7 ETM+ sensors from 1993 to 2020 using Google Earth Engine (GEE). Landsat data over our study region before 1993 is very sparse and impossible to use in time series analysis (Solórzano et al., 2020). We applied a cloud mask using the CFMask algorithm and computed NDVI in GEE (Zhu et al., 2015). We imported NDVI time series to R and conducted satellite intercalibration comparing monthly NDVI averages between Landsat-5 TM and Landsat-7 ETM+ in months where both satellites acquired data (Landsat 5 and 7 overlap #YEARS; SUP MAT). We detected forest clearing events over time using the algorithm Breaks For Additive Seasonal Trend (BFAST) implemented in the R package *BFAST.* This algorithm decomposes time series into trend, seasonal, and reminder components, and identifies changes fitting the data iteratively to a piecewise linear trend and seasonal model. BFAST also characterizes the magnitude and direction of change, where breaks in the trend component represent abrupt changes such as forest clearing events (Verbesselt et al., 2010). To detect breaks, we used an ordinary least squares-moving sums of residuals (OLS-MOSUM) model, with a level of 0.05 and a harmonic seasonal trend. To avoid overestimating breaks, we XSophia’s work-reviewed time series by handX and defined a magnitude threshold of XX. We defined this magnitude threshold by means of comparing high-resolution images (google earth pro/planet) from dates before and after BFAST break dates in XX30?XX time series. In this way, we verified X#X breakpoints as forest clearing events with an overall accuracy of XX%. Breaks with magnitudes < XX were not identified as forest clearing events in high resolution images and thus were excluded from further analyses. XNote: decide -and describe- what to do in plots that are mostly devoted to agriculture (the ones where there’s too much disturbance to detect breaks)X. Finally, we estimated the number of breakpoints and the number of years elapsed between FI field data collection and the most recent break detected. Because FI field data collection begun in 2009 and spans 5 years, all NDVI time series are between 16-21 years long and disturbances that occurred before 1993 were not detected.

Forest species composition also changes after disturbance. In Mexican TMCF, young forest stands tend to be dominated by few pioneer species such as *Pinus pseudostrobus*, *P. leiophylla*, *P. chiapensis*, *Alnus acuminata*, and *Liquidambar styraciflua*. Over time, shade-tolerant species replace pioneer species and both evenness and diversity increase (Velasco-Murguia et al., 2021, MOOORE). We compared species abundance and richness of sites with different mosaic values using rank abundance curves. We conducted these analyses using R package geoveg.